

ĐẠI HỌC BÁCH KHOA HÀ NỘI

TRƯỜNG ĐIỆN – ĐIỆN TỬ



BÁO CÁO BÀI TẬP LỚN

KỸ THUẬT VI XỬ LÝ

ĐỀ TÀI: HỆ THỐNG NHẬN DẠNG KHUÔN MẶT

CHẤM CÔNG ĐIỂM DANH

Giảng viên hướng dẫn:

TS. Hàn Huy Dũng

Lớp:

Nhóm:

01

Sinh viên:

Nguyễn Minh Hùng

20234011

Nguyễn Quốc Huy

20234013

Phan Nguyễn Mạnh Lợi

20234020

Hoàng Tuấn Ngọc

20234028

TÓM TẮT

Tên đề tài: Hệ thống nhận dạng khuôn mặt chấm công điểm danh.

Trong bối cảnh Cách mạng công nghiệp 4.0, nhu cầu về các giải pháp "Văn phòng thông minh" (Smart Office) ngày càng gia tăng, đặc biệt là các hệ thống kiểm soát ra vào và chấm công tự động. Các giải pháp truyền thống như thẻ từ hay vân tay đang dần bộc lộ nhược điểm về vệ sinh dịch tễ và khả năng gian lận, trong khi các máy chấm công khuôn mặt thương mại lại có chi phí cao, khó tiếp cận với các doanh nghiệp nhỏ.

Xuất phát từ thực tế đó, nhóm sinh viên đã nghiên cứu và phát triển hệ thống chấm công khuôn mặt giá rẻ dựa trên kiến trúc IoT Client-Server. Hệ thống bao gồm hai thành phần chính:

1. Thiết bị đầu cuối (Client): Sử dụng vi điều khiển ESP32-CAM tích hợp Camera và kết nối WiFi. Để tối ưu hóa hiệu năng trên phần cứng giới hạn, nhóm đã áp dụng Hệ điều hành thời gian thực (FreeRTOS), cho phép xử lý đa nhiệm mượt mà các tác vụ thu thập hình ảnh, truyền tin và điều khiển cơ cấu chấp hành.
2. Máy chủ xử lý (Server): Sử dụng máy tính cá nhân chạy thuật toán trí tuệ nhân tạo. Nhóm đã tích hợp thư viện InsightFace (ArcFace) tiên tiến để đảm bảo độ chính xác nhận diện cao (>99%) ngay cả với các góc mặt nghiêng, khắc phục hạn chế của các thư viện mã nguồn mở cũ.

Kết quả thực nghiệm cho thấy hệ thống hoạt động ổn định với thời gian đáp ứng dưới 5 giây, chi phí phần cứng thấp (dưới 1 triệu đồng), đồng thời cung cấp giao diện Web trực quan để quản lý nhân sự và lịch sử ra vào. Đề tài không chỉ giải quyết bài toán kinh tế - kỹ thuật mà còn thể hiện khả năng làm chủ công nghệ vi xử lý và AI của nhóm thực hiện.

MỤC LỤC

MỤC LỤC	2
DANH MỤC HÌNH ẢNH	7
DANH MỤC BẢNG	7
LỜI NÓI ĐẦU	8
CHƯƠNG 1. GIỚI THIỆU	9
1.1. Đặt vấn đề	9
1.2. Giới thiệu đề tài và mục tiêu đề tài	9
1.3. Khảo sát thị trường	10
<i>1.3.1. So sánh các công nghệ chấm công</i>	10
<i>1.3.2. Phân tích khoảng trống công nghệ và định hướng tối ưu</i>	11
<i>1.3.3. Kết luận</i>	13
1.4. Đối tượng và Phạm vi nghiên cứu	14
<i>1.4.1. Đối tượng</i>	14
<i>1.4.2. Phạm vi nghiên cứu</i>	14
1.5. Phân công công việc	15
CHƯƠNG 2: PHÂN TÍCH HỆ THỐNG	17
2.1. Yêu cầu hệ thống	17
<i>2.1.1. Yêu cầu chức năng</i>	17
<i>2.1.2. Yêu cầu phi chức năng</i>	18
2.2. Sơ đồ khối tổng quát	19
2.3. Phân tích và lựa chọn linh kiện	19
<i>2.3.1 Vi xử lý trung tâm</i>	19
<i>2.3.2. Camera thu nhập hình ảnh</i>	21
<i>2.3.3. Cảm biến phát hiện chuyển động</i>	23
<i>2.3.4. Màn hình</i>	25
<i>2.3.5. Mô tơ</i>	28
<i>2.3.6. Khối nguồn</i>	30
2.4. Danh sách các linh kiện cần dùng	33
<i>2.4.1. Giới thiệu về ESP32 DevKit V1</i>	33
2.4.1.1. Cấu trúc phần cứng cốt lõi	33

2.4.1.2. Bộ nhớ	33
2.4.1.3. Khả năng kết nối không dây	34
2.4.1.4. Nguồn điện và tiêu thụ năng lượng	34
2.4.1.5. Hệ thống ngoại vi và giao tiếp (I/O)	35
2.4.1.6. Các thành phần điều khiển trên bo mạch	36
2.4.1.7. Sơ đồ chân ESP32-DevKit V1	36
2.4.2. ESP32-Cam	38
2.4.2.1. Cấu trúc phần cứng cốt lõi	39
2.4.2.2. Bộ nhớ	39
2.4.2.3. Camera	40
2.4.2.4. Bộ nhớ lưu trữ mở rộng	41
2.4.2.5. Ăng ten	41
2.4.2.6. Đèn LED	42
2.4.2.7. Thông số kỹ thuật	42
2.4.2.8. Sơ đồ chân ESP32-CAM	44
2.4.3. Giới thiệu về cảm biến PIR HC-SR501	45
2.4.3.1. Nguyên lý hoạt động và cấu tạo	46
2.4.3.2. Các thành phần điều khiển trên board mạch	46
2.4.3. Giới thiệu về màn hình SSD1306	47
2.4.3.1. Nguyên lý hoạt động và cấu tạo	47
2.4.3.2. Giao thức truyền dữ liệu I2C	49
2.4.4. Giới thiệu về Servo SG90	50
2.4.4.1. Nguyên lý hoạt động và cấu tạo	50
2.4.4.2. Thông số kỹ thuật	50
2.4.4.3. Cấu tạo và nguyên lý hoạt động	50
2.4.4.4. Kết nối dây	51
2.4.4.5. Cơ chế điều khiển bằng tín hiệu PWM	51
2.4.5. Nguồn adapter 12V 2A và module LM2596	51
2.4.5.1. Nguồn adapter 12V 2A	51
2.4.5.2. Module LM2596	52
3.1. Sơ đồ nối chân và nguyên lý hoạt động	53

3.1.1. Sơ đồ nối chân	53
3.1.2. Nguyên lý hoạt động toàn hệ thống	53
3.1.2.1. Cấp nguồn và Khởi tạo (Khởi Nguồn)	53
3.1.2.2. Phát hiện đối tượng (Cảm biến => Khởi xử lý trung tâm)	53
3.1.2.3. Thu thập và Xử lý hình ảnh (Camera <=> Server)	54
3.1.2.4. Ra quyết định điều khiển (Server <=> Khởi xử lý trung tâm)	54
3.1.2.5. Thực thi lệnh (Khởi xử lý trung tâm => Mô tơ & Khởi hiển thị)	54
3.1.2.6. Cơ chế ngắt ngoài (External Interrupt) và chế độ ngủ (Deep Sleep)	55
3.2. Thiết kế phần cứng	55
3.2.1. Khởi xử lý trung tâm (ESP32 DevKit V1)	55
3.2.2. Khởi cảm biến và Camera	55
3.2.3. Khởi hiển thị và mô tơ	55
3.2.4. Tính toán và Thiết kế khối nguồn	55
3.3. Thiết kế phần mềm nhúng	56
3.3.1. Kiến trúc phần mềm	56
3.3.2. Lưu đồ thuật toán	57
3.4. Thiết kế phần mềm Server và AI	57
3.4.1. Quy trình xử lý nhận diện khuôn mặt	57
3.4.2. Thiết kế Cơ sở dữ liệu (Database)	57
3.4.3. Thiết kế giao diện Web (Web App)	58
CHƯƠNG 4: THUẬT TOÁN NHẬN DẠNG KHUÔN MẶT	59
4.1. Thư viện insightFace	59
4.1.1. Tổng quan về InsightFace	59
4.1.2. Face Embedding trong InsightFace	59
4.2. Thư viện face-recognition (dlib)	60
4.2.1. Tổng quan về dlib	60
4.2.2. Hai phương pháp của dlib	60
4.2.2.1. Thuật toán Histogram of Oriented Gradients (HoG)	60
4.2.2.2. Bộ dò khuôn mặt CNN Max-Margin (MMOD)	61
4.2.2.3 So sánh HoG và MMOD	61
4.3. So sánh giữa InsightFace và Face-recognition	62

CHƯƠNG 5. KẾT QUẢ THỰC NGHIỆM VÀ ĐÁNH GIÁ	64
5.1. Kết quả thi công hệ thống	64
5.1.1. Mô hình phần cứng	64
5.1.2. Giao diện phần mềm và Web Server	64
5.2. Đánh giá hiệu năng thuật toán nhận diện	66
5.2.1. Tiêu chí đánh giá	66
5.2.2. Kết quả thử nghiệm	66
5.3. Đánh giá hiệu năng Vi xử lý (FreeRTOS Evaluation)	71
5.4. Các hạn chế tồn tại	72
5.5. Kết luận chương	72
KẾT LUẬN CHUNG	73
TÀI LIỆU THAM KHẢO	75
PHỤ LỤC	76
A. CƠ SỞ LÝ THUYẾT XỬ LÝ ẢNH VÀ SINH TRẮC HỌC	76
1. Nguyên lý hoạt động	77
2. Thuật toán	77
3. Ưu điểm và nhược điểm	78
4. Hệ thống xử lý ảnh	79
4.1. Một số khái niệm trong xử lý ảnh	79
4.1.1. Điểm ảnh	79
4.1.2. Độ phân giải	80
4.1.3. Mức xám	81
4.2. Các bước cơ bản trong xử lý ảnh	82
B. Các mô hình AI sử dụng	84
1. Mô hình phát hiện khuôn mặt SCRFD (Sample and Computation Redistribution for Efficient Face Detection)	84
1.1. Tổng quan	84
1.2 Kiến trúc biến thể mô hình SCRFD	84
1.3 Các chỉ số về độ chính xác hiệu quả	84
1.4. Hiệu năng CPU	84
2. Mô hình trích xuất đặc trưng ArcFace	85

DANH MỤC HÌNH ẢNH

DANH MỤC BẢNG

LỜI NÓI ĐẦU

Ngày nay, Trí tuệ nhân tạo (AI) và Internet vạn vật (IoT) đang len lỏi vào từng ngóc ngách của đời sống, thay đổi căn bản cách thức vận hành của các tổ chức và doanh nghiệp. Trong lĩnh vực quản trị nhân sự, việc ứng dụng công nghệ để tự động hóa quy trình chấm công, kiểm soát an ninh đang trở thành một xu hướng tất yếu, đòi hỏi các giải pháp không chỉ thông minh, chính xác mà còn phải tối ưu về mặt chi phí.

Là sinh viên chuyên ngành Điện – Điện tử tại Đại học Bách Khoa Hà Nội, chúng em luôn mong muốn vận dụng những kiến thức đã học về Kỹ thuật Vi xử lý, lập trình nhúng và xử lý ảnh để giải quyết các bài toán thực tiễn. Đó chính là lý do nhóm quyết định lựa chọn đề tài: "Hệ thống nhận dạng khuôn mặt chấm công điểm danh" cho bài tập lớn lần này.

Mục tiêu của đề tài là xây dựng một hệ thống hoàn chỉnh từ phần cứng (mạch thu thập, điều khiển) đến phần mềm (thuật toán nhận diện, giao diện quản lý), qua đó làm chủ kỹ thuật lập trình trên vi điều khiển ESP32 và các mô hình Deep Learning hiện đại.

Để hoàn thành bài tập lớn này, bên cạnh sự nỗ lực của các thành viên trong nhóm, chúng em đã nhận được sự hướng dẫn và hỗ trợ nhiệt tình từ thầy Hàn Huy Dũng. Chúng em xin gửi lời cảm ơn chân thành và sâu sắc nhất đến TS. Hàn Huy Dũng – giảng viên hướng dẫn, người đã trực tiếp định hướng, chỉ bảo và tháo gỡ những vướng mắc cho nhóm trong suốt quá trình nghiên cứu và thực hiện đề tài.

Mặc dù đã rất cố gắng, nhưng do giới hạn về thời gian và kinh nghiệm thực tế, báo cáo khó tránh khỏi những thiếu sót. Chúng em rất mong nhận được những ý kiến đóng góp quý báu từ Thầy và các bạn để đề tài có thể hoàn thiện và phát triển hơn nữa trong tương lai.

Chúng em xin chân thành cảm ơn!

CHƯƠNG 1. GIỚI THIỆU

1.1. Đặt vấn đề

Trong xu thế chuyển đổi số, mô hình "Văn phòng thông minh" đang trở thành tiêu chuẩn mới cho các doanh nghiệp vừa và nhỏ và các công ty khởi nghiệp tại Việt Nam. Một trong những bài toán quản trị nhân sự hàng đầu là việc kiểm soát ra vào và chấm công.

Các phương pháp truyền thống đang bộc lộ nhiều bất cập trong môi trường công sở hiện đại:

- Chấm công thẻ giấy/thẻ từ: Dễ xảy ra tình trạng quên thẻ, mất thẻ hoặc nhân viên "chấm công hộ", gây thất thoát quỹ lương và thiếu công bằng.
- Chấm công vân tay: Từng rất phổ biến nhưng hiện nay gặp vấn đề lớn về an toàn vệ sinh (nguy cơ lây lan virus qua bề mặt tiếp xúc) và thường gặp lỗi khi tay nhân viên bị ướt hoặc mờ vân tay.
- An ninh cửa ra vào: Nhiều văn phòng vẫn sử dụng khóa cơ hoặc khóa số, gây khó khăn khi cần cấp quyền ra vào tạm thời cho khách hàng hoặc nhân viên mới, đồng thời không lưu lại được lịch sử ra vào khi có sự cố mất mát tài sản xảy ra.

1.2. Giới thiệu đề tài và mục tiêu đề tài

Xuất phát từ nhu cầu thực tế về mô hình "Văn phòng thông minh", các doanh nghiệp hiện nay đang rất cần một hệ thống tích hợp "2 trong 1": vừa đảm bảo an ninh, vừa tự động hóa việc điểm danh. Tuy nhiên, khi áp dụng các giải pháp hiện có, doanh nghiệp và người sử dụng đang gặp phải những hạn chế lớn:

Thứ nhất, về chi phí và tính linh hoạt: Các thiết bị chấm công nhận diện khuôn mặt chuyên dụng trên thị trường thường có chi phí đầu tư cao (trung bình từ 3.000.000 - 5.000.000 VNĐ/thiết bị), gây khó khăn cho các Startup hoặc doanh nghiệp SME. Hơn nữa, đây thường là các hệ thống đóng, rất khó để tùy biến hoặc tích hợp thêm các tính năng quản lý đặc thù theo yêu cầu của từng công ty.

Thứ hai, về vấn đề vệ sinh và tiện lợi: Các phương pháp truyền thống như vân tay đang trở nên e ngại do nguy cơ lây lan dịch bệnh qua bề mặt tiếp xúc. Trong khi đó, việc dùng thẻ từ lại dễ dẫn đến tình trạng quên thẻ hoặc gian lận (chấm công hộ).

Thứ ba, về giới hạn kỹ thuật của các giải pháp giá rẻ: Các giải pháp tự chế (DIY) giá rẻ thường gặp vấn đề về tốc độ xử lý chậm, độ trễ cao hoặc treo hệ thống khi hoạt động liên tục do tài nguyên vi điều khiển hạn chế.

Trước thực trạng đó, nhóm thực hiện đề tài nghiên cứu và phát triển hệ thống "Chăm công và Kiểm soát cửa bằng khuôn mặt giá rẻ" nhằm giải quyết các vấn đề trên thông qua các đóng góp cụ thể sau:

- Tối ưu hóa chi phí phần cứng: Sử dụng module ESP32-DevKit V1 giá rẻ làm thiết bị thu thập đầu cuối (Client), thay thế cho các máy tính nhúng đắt tiền, giúp giảm chi phí triển khai xuống mức thấp nhất.
- Giải quyết bài toán hiệu năng bằng mô hình Client-Server: Nhóm đề xuất kiến trúc đẩy tác vụ xử lý AI nặng nề lên máy chủ (Server), giúp vi điều khiển chỉ tập trung vào việc thu thập và điều khiển, đảm bảo hệ thống vận hành mượt mà.
- Ứng dụng thuật toán AI tiên tiến: Thay vì các thuật toán cũ, nhóm nghiên cứu tích hợp thư viện InsightFace/ArcFace để nâng cao độ chính xác nhận diện ngay cả ở các góc mặt nghiêng, khắc phục nhược điểm của các thư viện mã nguồn mở cũ (như Dlib).
- Tăng cường độ ổn định bằng Hệ điều hành thời gian thực: Nhóm áp dụng FreeRTOS vào lập trình nhúng cho ESP32 để quản lý đa nhiệm (thu hình, truyền tin, điều khiển servo) một cách hiệu quả, giảm thiểu độ trễ và tránh treo hệ thống.



Việc thực hiện đề tài không chỉ có ý nghĩa về mặt học thuật trong việc làm chủ kỹ thuật vi xử lý và AI, mà còn mang tính thực tiễn cao, tạo ra một sản phẩm công nghệ với chi phí hợp lý và hiệu năng ổn định.

1.3. Khảo sát thị trường

1.3.1. So sánh các công nghệ chăm công

Bảng 1.1: So sánh các công nghệ chăm công

Nhóm giải pháp	Công nghệ chủ đạo	Ưu điểm	Nhược điểm	Chi phí triển khai	Hình ảnh minh họa
----------------	-------------------	---------	------------	--------------------	-------------------

1. Giải pháp Truyền thống (Máy chấm công thẻ từ, vân tay)	<ul style="list-style-type: none"> - RFID (Thẻ từ) - Cảm biến quang học (Vân tay) 	<ul style="list-style-type: none"> - Lắp đặt đơn giản. - Ít chịu ảnh hưởng của ánh sáng môi trường. 	<ul style="list-style-type: none"> - Vệ sinh: Phải tiếp xúc trực tiếp (nguy cơ giả vân tay). - Tốc độ xử lý chậm khi đông người. 	~ 4.500.000 VNĐ	
2. Máy chấm công khuôn mặt chuyên dụng (Hikvision, Ronald Jack...)	<ul style="list-style-type: none"> - Camera kép (Dual-lens). - Đèn hồng ngoại. - Hệ điều hành nhúng đóng gói. 	<ul style="list-style-type: none"> - Độ chính xác cao (>99%). - Hoạt động ổn định, hoàn thiện phần cứng tốt. - Chống giả mạo (ảnh/video) tốt. 	<ul style="list-style-type: none"> - Hệ đóng (Closed System): Khó tùy biến chức năng - Giá thành cao, khó tiếp cận với mô hình nhỏ. 	~ 3.800.000 VNĐ	

1.3.2. Phân tích khoảng trống công nghệ và định hướng tối ưu

Qua khảo sát, nhóm nhận ra những ưu điểm cần kế thừa và những hạn chế như sau:

- Điểm mạnh của các giải pháp hiện hành:
 - + Độ ổn định cao: Các máy chấm công chuyên dụng (Ronald Jack, Hikvision) hoạt động bền bỉ, ít lỗi vặt nhờ phần cứng được đóng gói công nghiệp.

- + Tốc độ xử lý nhanh: Các hệ thống sử dụng PC/Server có khả năng xử lý AI mạnh mẽ, nhận diện được khuôn mặt với tốc độ cao và độ chính xác lớn.
- + Tính chuyên nghiệp: Giao diện người dùng và cơ cấu lắp đặt gọn gàng, thẩm mỹ.
- Hạn chế tồn tại (Khoảng trống thị trường):
 - + Rào cản chi phí: Cả hai giải pháp trên đều có giá thành khá cao (từ 3.500.000 VNĐ trở lên), tạo ra rào cản lớn đối với các văn phòng khởi nghiệp (Startup), cửa hàng nhỏ hoặc mô hình lớp học sinh viên.
 - + Tính đóng của hệ thống (Closed System): Các máy chấm công thương mại thường là hệ sinh thái đóng. Người dùng rất khó (hoặc không thể) can thiệp để thêm các tính năng tùy biến như: hiển thị lời chào theo tên riêng, tích hợp điều khiển cơ cấu cửa tự chế, hay trích xuất dữ liệu theo định dạng riêng của công ty.
 - + Sự cồng kềnh và tiêu tốn năng lượng: Giải pháp dùng PC chạy 24/7 tiêu tốn nhiều điện năng và không phù hợp để lắp đặt tại các vị trí cửa đơn lẻ, nhỏ hẹp.
- Các yếu tố cần tối ưu: Để lấp đầy khoảng trống công nghệ nói trên, hệ thống của nhóm cần tập trung tối ưu hóa 3 yếu tố cốt lõi:
 - + Tối ưu về Chi phí:
 - Thay thế các máy tính nhúng đắt tiền bằng vi điều khiển ESP32-DevKit V1 tích hợp sẵn WiFi với chi phí linh kiện cực thấp (dưới 500.000 VNĐ cho module chính).
 - Mục tiêu tổng chi phí phần cứng của hệ thống phải dưới 1.000.000 VNĐ.
 - + Tối ưu về Kiến trúc xử lý:
 - Khắc phục nhược điểm yếu về xử lý của vi điều khiển bằng cách áp dụng mô hình IoT Client-Server. Vi điều khiển (Client) chỉ đóng vai trò thu thập ảnh và điều khiển cơ khí, trong khi gánh nặng xử lý AI được chuyển sang Server (tận dụng máy tính có sẵn).
 - Điều này giúp hệ thống vừa có giá thành rẻ của vi điều khiển, vừa có sức mạnh xử lý AI của máy tính.

+ Tối ưu về Tính năng mở:

- Xây dựng hệ thống mã nguồn mở, cho phép tùy biến giao diện Web quản lý.
- Tích hợp khả năng điều khiển thiết bị ngoại vi đa dạng (Servo, Relay khóa từ, Màn hình OLED) để phù hợp với nhiều loại cửa khác nhau.

Kết luận: Việc phát triển một hệ thống "Hybrid" kết hợp giữa phần cứng giá rẻ và thuật toán AI mạnh mẽ trên Server là hướng đi phù hợp nhất để cân bằng giữa chi phí và hiệu năng.

1.3.3. Kết luận

Từ kết quả khảo sát và phân tích khoảng trống công nghệ ở trên, nhóm nghiên cứu rút ra kết luận: "Thị trường đang thiếu một giải pháp chấm công khuôn mặt giá rẻ, mã nguồn mở dành cho các mô hình quy mô nhỏ".

Để giải quyết vấn đề này, nhóm quyết định xây dựng hệ thống theo hướng "Tối ưu hóa chi phí bằng kiến trúc Client-Server" với các giải pháp kỹ thuật cụ thể đã được thực hiện như sau:

- Lựa chọn phần cứng tối giản: Nhóm đã quyết định không sử dụng các máy tính nhúng đắt tiền (như Raspberry Pi hay Jetson Nano) mà lựa chọn ESP32-DevKit-V1 làm vi xử lý trung tâm. Đây là giải pháp phần cứng có tỷ lệ hiệu năng/giá thành tốt nhất hiện nay, giúp giảm chi phí phần cứng xuống dưới 1 triệu đồng mà vẫn đảm bảo khả năng kết nối WiFi và thu thập hình ảnh.
- Chiến lược xử lý phân tán: Nhóm đã thiết kế hệ thống theo mô hình IoT:
 - + Client: Chỉ tập trung vào các tác vụ nhẹ như streaming hình ảnh, đọc cảm biến và điều khiển động cơ servo. Nhóm đã tối ưu hóa firmware sử dụng FreeRTOS để đảm bảo độ trễ thấp nhất.
 - + Server (Máy tính): Đảm nhận toàn bộ tác vụ nặng về AI. Nhóm đã xây dựng một Server xử lý trung tâm, cho phép tận dụng sức mạnh tính toán của máy tính để chạy các mô hình Deep Learning phức tạp mà vi điều khiển không thể làm được.
- Ứng dụng thuật toán nhận diện tiên tiến: Thay vì sử dụng các phương pháp cũ (như Haar Cascade hay Dlib HOG) vốn kém hiệu quả với góc nghiêng, nhóm đã nghiên cứu và tích hợp thành công thư viện InsightFace (ArcFace). Giải pháp

này giúp hệ thống đạt độ chính xác >99% và nhận diện tốt ngay cả khi người dùng không nhìn trực diện vào camera.

- Phát triển hệ sinh thái mở: Nhóm đã xây dựng hoàn thiện một Web App quản trị đi kèm, cho phép người dùng dễ dàng xem lại lịch sử ra vào, thêm/xóa nhân viên và giám sát camera theo thời gian thực – điều mà các máy chấm công giá rẻ trên thị trường thường không hỗ trợ.

Tóm lại: Giải pháp của nhóm không nhằm cạnh tranh về độ bền công nghiệp với các hãng lớn, mà tập trung vào việc phổ cập công nghệ AI đến phân khúc người dùng bình dân thông qua việc tối ưu hóa chi phí và linh hoạt hóa tính năng.

1.4. Đối tượng và Phạm vi nghiên cứu

1.4.1. Đối tượng

Đề tài tập trung nghiên cứu và làm chủ các đối tượng sau:

- Công nghệ phần cứng: Vi điều khiển SoC ESP32-DevKit V1, giao tiếp ngoại vi (I2C, GPIO) và các cơ cấu chấp hành (Servo, OLED).
- Công nghệ phần mềm: Các thuật toán Trí tuệ nhân tạo (AI) trong bài toán nhận diện khuôn mặt (Face Recognition).
- Giao thức truyền thông: Kỹ thuật truyền tải hình ảnh và dữ liệu điều khiển qua giao thức HTTP/WebSocket trong mạng cục bộ .

1.4.2. Phạm vi nghiên cứu

Phạm vi nghiên cứu của đề tài được giới hạn trong các nội dung sau:

- Phạm vi không gian: Hệ thống được thiết kế cho mô hình cửa ra vào văn phòng, hoạt động trong môi trường trong nhà, có mái che, tránh mưa nắng trực tiếp.
- Điều kiện ánh sáng: Hệ thống hoạt động tốt nhất trong điều kiện ánh sáng phòng tiêu chuẩn (300-500 Lux). Chưa hỗ trợ nhận diện trong điều kiện thiếu sáng hoặc ngược sáng mạnh.
- Quy mô thử nghiệm: Hệ thống được kiểm thử ổn định với tập dữ liệu không giới hạn.
- Giới hạn bảo mật: Hệ thống tập trung vào tính tiện dụng và kiểm soát ra vào cơ bản. Các tính năng bảo mật nâng cao như chống giả mạo bằng mặt nạ 3D (3D Mask Anti-spoofing) hay quét mống mắt nằm ngoài phạm vi nghiên cứu này.

- Thời gian thực hiện: 12-16 tuần với quy mô nhóm 4 thành viên, sử dụng thiết bị cá nhân và tài nguyên học thuật miễn phí.

1.5. Phân công công việc

Bảng 1.2: Phân công công việc

Tên sinh viên	MSSV	Nhiệm vụ
Nguyễn Minh Hùng	20234011	Lập trình Nhúng (ESP32 Control): + Viết code firmware cho ESP32 dùng FreeRTOS. + Xử lý đa nhiệm: Điều khiển Servo, hiển thị OLED, đọc cảm biến PIR. + Xử lý logic tự động khóa và báo lỗi.
Nguyễn Quốc Huy	20234013	Báo cáo & Tài liệu: + Vẽ các sơ đồ hệ thống (Sơ đồ khối, Lưu đồ thuật toán) trên phần mềm. + Viết báo cáo thuyết minh (Chương Tổng quan, Cơ sở lý thuyết).
Phan Nguyễn Mạnh Lợi	20234020	Phần cứng & Camera (Hardware): + Lập trình module ESP32-CAM (Cấu hình luồng Stream MJPEG). + Thiết kế & lắp ráp mô hình khung cửa, đi dây, hàn mạch. + Vẽ sơ đồ nguyên lý kết nối (Wiring Diagram).
Hoàng Tuấn Ngọc	20234028	Lập trình AI & Server: + Xây dựng chương trình Python xử lý nhận diện khuôn mặt (OpenCV). + Thiết kế Database & Giao diện Web (Streamlit). + Tích hợp hệ thống, debug lỗi truyền thông HTTP.

CHƯƠNG 2: PHÂN TÍCH HỆ THỐNG

2.1. Yêu cầu hệ thống

2.1.1. Yêu cầu chức năng

Hệ thống cần thực hiện được các nhóm chức năng sau dựa trên hoạt động của vi xử lý ESP32-DevKit V1:

Bảng 2.1: Yêu cầu chức năng

Chỉ tiêu chức năng	Mô tả
Nhóm chức năng Thu thập & Giám sát	<ul style="list-style-type: none">- Phát hiện chuyển động: Sử dụng ngắt ngoài (External Interrupt) từ cảm biến PIR để phát hiện người trong phạm vi 2-3m, đánh thức vi xử lý khỏi chế độ ngủ (Deep Sleep).- Thu thập hình ảnh: Camera phải chụp được ảnh định dạng JPEG, độ phân giải tối thiểu QVGA (320x240) hoặc VGA (640x480) để đảm bảo tốc độ truyền tải qua WiFi.
Nhóm chức năng Xử lý & Truyền thông	<ul style="list-style-type: none">- Web Server: Vi xử lý đóng vai trò là một Web Server nội bộ.- Streaming: Truyền luồng video hoặc gửi ảnh tĩnh về máy chủ/client với độ trễ thấp thông qua giao thức HTTP/WebSocket.- Quản lý danh sách: Lưu trữ và quản lý ID khuôn mặt, cho phép Thêm/Xóa ID từ giao diện Web.
Nhóm chức năng Chấp hành & Hiển thị (Output):	<ul style="list-style-type: none">- Điều khiển cơ khí: Xuất xung PWM điều khiển động cơ Servo quay góc chính xác (0° đóng, 90° mở).

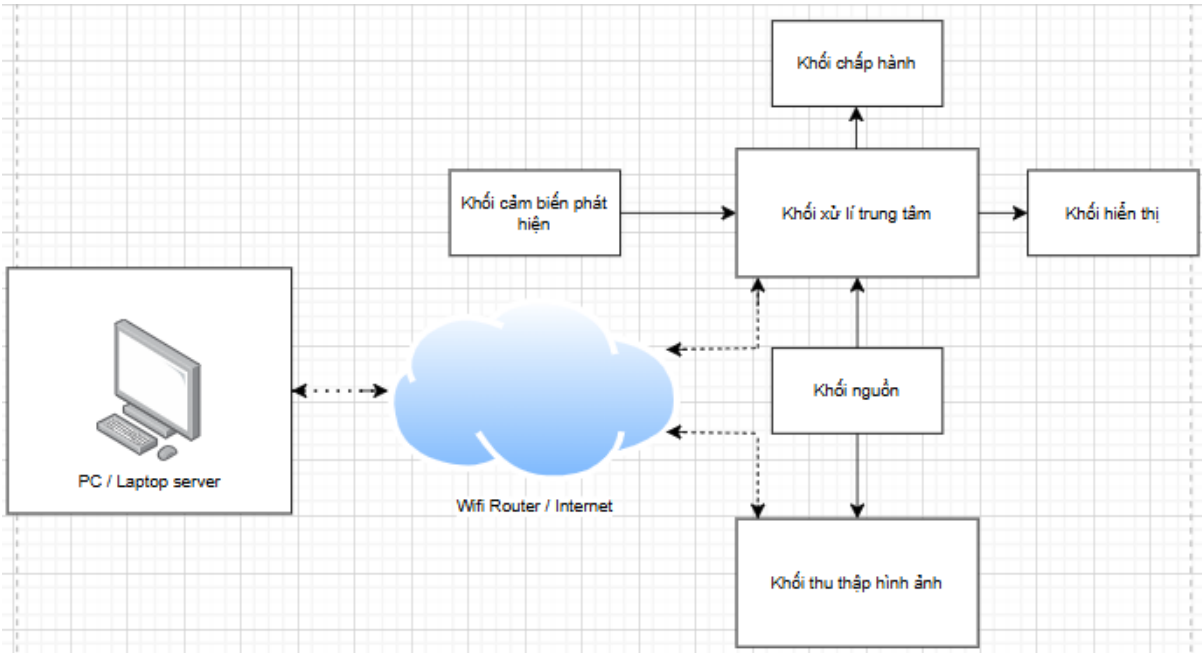
	<ul style="list-style-type: none"> - Giao diện hiển thị: Hiển thị trạng thái hệ thống (IP Address, Trạng thái kết nối, Tên người dùng, Kết quả OK/Fail) lên màn hình OLED qua giao tiếp I2C.
--	---

2.1.2. Yêu cầu phi chức năng

Bảng 2.2: Yêu cầu phi chức năng

Chỉ tiêu phi chức năng	Mô tả
Hiệu năng	<ul style="list-style-type: none"> + Độ trễ toàn hệ thống: Thời gian từ lúc chụp ảnh đến khi ra quyết định mở cửa phải < 2 giây trong điều kiện mạng LAN ổn định. + Thời gian đáp ứng cơ khí: Động cơ Servo phải phản hồi ngay lập tức (< 200ms) sau khi nhận lệnh từ vi xử lý.
Độ chính xác & Tin cậy	<ul style="list-style-type: none"> + Tỷ lệ nhận diện đúng: Đạt > 90% trong điều kiện ánh sáng văn phòng (300-500 Lux) và góc mặt nghiêng không quá 30 độ. + Tự động kết nối: Tự động kết nối lại WiFi khi mất tín hiệu mạng.
Nguồn điện & Năng lượng	<ul style="list-style-type: none"> + Ổn định điện áp: Hệ thống phải hoạt động ổn định, không bị sập áp khi dòng tiêu thụ tăng đột ngột. Yêu cầu nguồn cấp tối thiểu 5V-2A. + Tiết kiệm năng lượng: Hỗ trợ chế độ ngủ (Deep Sleep) khi không có người để giảm dòng tiêu thụ xuống mức thấp nhất (mA).

2.2. Sơ đồ khối tổng quát






Hình 2.1 Sơ đồ khối tổng quát

2.3. Phân tích và lựa chọn linh kiện

2.3.1 Vi xử lý trung tâm

Bảng 0.3: So sánh vi xử lý trung tâm

Tiêu chí	Arduino Mega 2560	Raspberry Pi ZeroW	ESP32 DevKit V1
Vi xử lý			
	ATmega2560 (8-bit)	Broadcom BCM2835 (32-bit ARM11)	Dual-core Xtensa® 32-bit LX6
Tốc độ xung nhịp	16 MHz	1 GHz	240 MHz
	2	5	3
RAM	8 KB	512 MB	520KB
	1	5	3
	256 KB Flash	Thẻ nhớ MicroSD	4MB Flash

Bộ nhớ lưu trữ	2	3	4
Kết nối không dây	Không (Cần mua module bên ngoài)	Wifi 802.11n + Bluetooth 4.1	Wifi 802.11n + Bluetooth 4.2
	2	4	5
Hệ điều hành	Không (Chạy code trực tiếp - Bare metal)	Linux (Raspberry Pi OS)	FreeRTOS
	2	4	5
Giá thành	~250.000đ	~900.000đ	~160.000đ
	4	2	5
Điểm trung bình	2.215	3.75	4.125

Dựa trên bảng khảo sát thông số kỹ thuật giữa ba dòng vi điều khiển/máy tính nhúng phổ biến là Arduino Mega 2560, Raspberry Pi Zero W và ESP32-DevKit V1, ta có kết quả đánh giá tổng quan như sau:

Arduino Mega 2560:

- Mặc dù có thể mạnh về sự ổn định và số lượng chân GPIO lớn, nhưng thiết bị này bộc lộ điểm yếu chí tử trong các tác vụ đòi hỏi khả năng tính toán cao như xử lý ảnh.
- Tốc độ xung nhịp thấp (16 MHz) và dung lượng RAM quá nhỏ (8 KB) khiến việc xử lý dữ liệu hình ảnh là bất khả thi.
- Thiếu các module tích hợp sẵn (Wifi, Camera) dẫn đến chi phí phát sinh và độ phức tạp phần cứng tăng cao khi ghép nối.

Raspberry Pi Zero W:

- Sở hữu cấu hình mạnh mẽ nhất với CPU 1GHz và 512 MB RAM, chạy hệ điều hành Linux đầy đủ.

- Tuy nhiên, điểm trừ nằm ở chi phí cao (~600.000đ) và thời gian khởi động lâu (do phải boot OS).
- Việc phụ thuộc hoàn toàn vào thẻ nhớ MicroSD cho hệ điều hành khiến độ ổn định trong môi trường công nghiệp (dễ bị corrupt thẻ nhớ khi mất điện đột ngột) bị đánh giá thấp hơn so với vi điều khiển sử dụng Flash.





ESP32-DevKit V1:

- Đây là thiết bị đạt điểm số cao nhất nhờ sự cân bằng tuyệt vời giữa hiệu năng và giá thành.
- Về xử lý: Vi xử lý lõi kép 240 MHz kết hợp với 4 MB PSRAM bổ sung giải quyết tốt bài toán lưu trữ và xử lý khung hình từ camera.
- Về tính năng: Tích hợp sẵn Wi-Fi, Bluetooth trên một bo mạch nhỏ gọn, giúp giảm thiểu kích thước và sai số khi thi công phần cứng.
- Về phần mềm: Hỗ trợ FreeRTOS (Hệ điều hành thời gian thực), phù hợp cho các ứng dụng nhúng cần độ phản hồi nhanh và kiểm soát tài nguyên chặt chẽ.

=> **Kết luận:** Dựa trên bảng so sánh và phân tích, nhóm chúng em quyết định chọn ESP32-DevKit V1 làm khối xử lý trung tâm.

2.3.2. Camera thu nhập hình ảnh

Bảng 0.4: So sánh Camera

Tiêu chí	ESP32-CAM + OV2640 	Module OV7670 	USB Webcam (Kết nối PC) 	Raspberry Pi Camera (CSI) 
Đặc điểm chính	SoC tích hợp Camera + WiFi	Cảm biến ảnh thô, giá rẻ	Camera ngoại vi hoàn chỉnh	Camera chuyên dụng cho Pi
Khả năng kết nối	WiFi/Bluetooth sẵn có	Cần MCU hỗ trợ	Phụ thuộc cổng USB & PC	Giao tiếp CSI tốc độ cao
	5	1	2.5	4.5

Hiệu năng xử lý	VGA/SVGA ổn định (PSRAM)	Rất thấp, dễ giật hình	HD/Full HD mượt mà	Full HD/4K chuyên nghiệp
	3.5	2	4.5	5
Độ gọn nhẹ	Rất gọn, All-in-one	Cồng kênh do nhiều dây nối	Cồng kênh, dây cáp dài	Gọn nhưng cần board Pi to
	4.5	1.5	3	4
Khả năng lập trình	Thư viện Arduino phong phú	Khó, tài liệu lập trình cũ	Rất dễ	Hỗ trợ cực tốt trên Linux/Py
	4	2.5	4.5	4.5
Giá thành	~160.000đ	~70.000đ	~300.000đ+	~800.000đ+
	4.5	5	3	2
Điểm trung bình	4.3	2.4	3.5	4.0

Dựa trên bảng khảo sát thông số kỹ thuật giữa các dòng camera phổ biến cho hệ thống nhúng là Module OV7670, USB Webcam/Raspberry Pi Camera và Module tích hợp ESP32-CAM (kèm OV2640), ta có kết quả đánh giá chi tiết như sau:

Module Camera OV7670 (Cảm biến ảnh thô):

- Mặc dù có ưu điểm là giá thành rẻ nhất trong các lựa chọn, nhưng đây là một cảm biến ảnh "thô".
- Nhược điểm lớn nhất là việc giao tiếp rất phức tạp, đòi hỏi số lượng chân GPIO lớn (giao tiếp song song) và không tích hợp sẵn bộ đệm (FIFO). Điều này khiến vi điều khiển phải tiêu tốn gần như toàn bộ tài nguyên chỉ để đọc dữ liệu ảnh, không còn khả năng xử lý các tác vụ khác như truyền WiFi hay điều khiển Servo. Chất lượng hình ảnh thấp và tốc độ khung hình (FPS) rất kém khi hoạt động với ESP32.

USB Webcam hoặc Raspberry Pi Camera (CSI):

- Đây là nhóm thiết bị cho chất lượng hình ảnh vượt trội (HD, Full HD), sắc nét và tốc độ cao, rất phù hợp cho các bài toán xử lý ảnh chuyên sâu.

- Tuy nhiên, rào cản lớn nhất là vấn đề tương thích và chi phí. USB Webcam đòi hỏi thiết bị chủ phải có cổng USB Host và driver xử lý video (thường là PC hoặc Raspberry Pi), vi điều khiển ESP32 không thể giao tiếp trực tiếp. Raspberry Pi Camera sử dụng chuẩn CSI đặc thù, cũng không tương thích với ESP32. Nếu sử dụng các loại này, bắt buộc phải thay đổi khối xử lý trung tâm sang các máy tính nhúng đắt tiền, đi ngược lại với mục tiêu "giá rẻ" của đề tài.

Module ESP32-CAM (Tích hợp cảm biến OV2640):

- Đây là thiết bị đạt điểm số cao nhất nhờ thiết kế "All-in-one" (Tất cả trong một) chuyên dụng cho các ứng dụng IoT giám sát.
- Về cấu trúc: Module này tích hợp sẵn socket cho camera OV2640 và quan trọng nhất là được trang bị thêm 4MB PSRAM (RAM giả tĩnh). Đây là yếu tố quyết định, cho phép lưu trữ và xử lý các khung hình có độ phân giải cao (lên đến 1600x1200) mà bộ nhớ nội của ESP32 thường không đủ sức chứa.
- Về kết nối: Camera được kết nối trực tiếp vào vi xử lý qua giao diện DVP tốc độ cao, tối ưu hóa việc truyền luồng video (Video Streaming) qua WiFi với độ trễ thấp.
- Về kinh tế: Với mức giá chỉ khoảng 160.000đ cho cả vi điều khiển, camera và mạch nạp, đây là lựa chọn tối ưu nhất cho bài toán kinh tế.

=> **Kết luận:** Dựa trên bảng so sánh và phân tích kỹ thuật, nhóm quyết định chọn Module ESP32-CAM kết hợp cảm biến ảnh OV2640 làm khối thu thập hình ảnh. Sự lựa chọn này đảm bảo sự cân bằng giữa khả năng xử lý luồng video mượt mà và chi phí phần cứng thấp nhất.

2.3.3. Cảm biến phát hiện chuyển động

Bảng 0.5: So sánh cảm biến chuyển động

Tiêu chí	Cảm biến Siêu âm (HC-SR04)	Cảm biến Radar Vi sóng (RCWL-0516)	Cảm biến Hồng ngoại thụ động (PIR HC-SR501)
			

Nguyên lý	Phát sóng siêu âm và đo thời gian phản hồi	Hiệu ứng Doppler	Cảm nhận sự thay đổi bức xạ nhiệt
Đối tượng phát hiện	Mọi vật cản rắn (người, vật, tường, ...)	Mọi chuyển động (người, quạt, rèm cửa, ...)	Chủ yếu là người và vật nuôi (có thân nhiệt)
Góc quét	Hẹp, khoảng 15° - 30°	Rất rộng, 360°, xuyên vật cản	Khá rộng, khoảng 100° - 120°
	2	5	4
Tiêu thụ năng lượng	Cao (~15 mA)	Trung bình (~4 mA)	Rất thấp (< 60 μ A)
	2	3	5
Độ nhạy & Nhiễu	Ổn định, nhưng phải đi qua đúng hướng	Quá nhạy (xuyên tường, kính), dễ báo giả	Ổn định, chỉ báo khi có người/vật nhiệt
	4	2	4
Giá thành	20.000đ	15.000đ	25.000đ
	4	5	4
Điểm trung bình	3	3.75	4.25

Dựa trên bảng khảo sát thông số kỹ thuật và chấm điểm giữa ba dòng cảm biến phát hiện vật thể/chuyển động phổ biến là Siêu âm (HC-SR04), Radar vi sóng (RCWL-0516) và Hồng ngoại thụ động (PIR HC-SR501), ta có kết quả đánh giá tổng quan như sau:

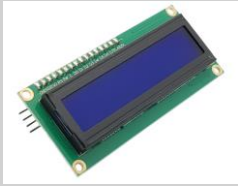


- Cảm biến Siêu âm HC-SR04:
 - + Góc quét hẹp (15° - 30°), dễ bỏ sót người dùng nếu họ không đứng trực diện.
 - + Mức tiêu thụ điện năng cao.
- Cảm biến Radar Vi sóng RCWL-0516:
 - + Sở hữu độ nhạy cực cao, góc quét 360 độ và khả năng nhìn xuyên vật cản phi kim loại, cho phép giấu kín hoàn toàn cảm biến bên trong vỏ hộp.
 - + Tuy nhiên, điểm trừ chí tử khi kết hợp với ESP32-CAM là vấn đề nhiễu sóng điện từ. Do hoạt động dựa trên sóng vô tuyến, cảm biến này thường gây nhiễu cho kết nối WiFi của ESP32 và ngược lại, dẫn đến hiện tượng kích hoạt sai liên tục, làm hệ thống mất ổn định.
 - + Độ nhạy quá cao cũng khiến nó dễ nhận diện nhầm các vật thể vô tri như quạt trần, rèm cửa bay là người.
- Cảm biến Hồng ngoại PIR HC-SR501:
 - + Đây là thiết bị đạt điểm số cao nhất nhờ sự phù hợp đặc thù với bài toán "phát hiện con người".
 - + Về nguyên lý: Cảm biến chỉ phản ứng với sự thay đổi của bức xạ nhiệt hồng ngoại từ cơ thể người (và động vật), giúp loại bỏ hoàn toàn các báo động giả từ đồ vật chuyển động (quạt, lá cây).
 - + Về độ ổn định: Hoạt động thụ động, không phát ra sóng nên không gây nhiễu và không bị nhiễu bởi sóng WiFi của ESP32-CAM.
 - + Về năng lượng: Mức tiêu thụ điện năng cực thấp phù hợp cho hệ thống hoạt động 24/7.

=> **Kết luận:** Dựa trên bảng so sánh và phân tích, nhóm chúng em quyết định chọn Cảm biến hồng ngoại PIR HC-SR501 làm cảm biến phát hiện chuyển động.

2.3.4. Màn hình

Bảng 0.6: So sánh màn hình

Tiêu chí	LCD 16x2 (Kèm mạch I2C)	OLED 0.96 inch (SSD1306)	TFT LCD 2.4 inch (SPI)
----------	----------------------------	-----------------------------	---------------------------

			
Giao tiếp (Số chân)	I2C - Tồn 2 chân GPIO	I2C - Tồn 2 chân GPIO	SPI - Tồn 4-6 chân GPIO
	5	5	2
Khả năng hiển thị	Chỉ hiện được Ký tự/Số	Hiện được Ký tự, Đồ họa đơn sắc	Hiện được Màu, Ảnh khuôn mặt
	2	4	5
Độ tương thích ESP32- CAM	Rất dễ kết nối	Rất dễ kết nối	Cực khó, xung đột chân với Thẻ nhớ/Cam
	5	5	1
Thẩm mỹ/Hiện đại	Nhìn khá cổ điển, công nghiệp	Nhìn hiện đại, sắc nét	Nhìn như điện thoại
	2	5	5
Tiêu thụ năng lượng	Khá tốn do đèn nền	Rất thấp, pixel nào sáng mới tốn điện	Tốn pin nhất do đèn nền màn màu
	3	5	2
Giá thành	~35.000đ	~45.000đ	~120.000đ
	5	4	2

Điểm trung bình	3.6	4.6	2.8
------------------------	------------	------------	------------

Dựa trên bảng khảo sát thông số kỹ thuật giữa ba dòng màn hình hiển thị phổ biến là LCD 16x2, TFT LCD 2.4 inch và OLED 0.96 inch, ta có kết quả đánh giá chi tiết như sau:

Màn hình LCD 16x2 (Kèm mạch chuyển đổi I2C):

- Đây là loại màn hình "kính điện" trong kỹ thuật vi xử lý nhờ giá thành rẻ và thư viện hỗ trợ phong phú.
- Tuy nhiên, nhược điểm lớn của nó là khả năng hiển thị rất hạn chế (chỉ hiển thị được ký tự và số, không hiển thị được đồ họa hoặc icon phức tạp). Kích thước cổng kênh và giao diện hiển thị mang tính "công nghiệp" cũ kỹ, thiếu tính thẩm mỹ cho một sản phẩm văn phòng thông minh (Smart Office). Ngoài ra, đèn nền (Backlight) luôn bật khiến mức tiêu thụ năng lượng cao hơn mức cần thiết.

Màn hình TFT LCD 2.4 inch (Giao tiếp SPI):

- Ưu điểm nổi bật là khả năng hiển thị màu sắc rực rỡ, có thể hiển thị ảnh khuôn mặt hoặc giao diện đồ họa phức tạp, mang lại trải nghiệm người dùng tốt nhất.
- Mặc dù vậy, rào cản kỹ thuật lớn nhất nằm ở chuẩn giao tiếp SPI. Màn hình này yêu cầu từ 4 đến 6 chân GPIO để hoạt động. Trong khi đó, module ESP32-CAM đã sử dụng hầu hết các chân cho Camera và khe cắm thẻ nhớ SD. Việc cố gắng tích hợp màn hình TFT sẽ gây ra xung đột phần cứng nghiêm trọng (thiếu chân điều khiển), làm phức tạp hóa quá trình thiết kế mạch và lập trình.

Màn hình OLED 0.96 inch (SSD1306):

- Đây là thiết bị đạt điểm số cao nhất nhờ sự tối ưu tuyệt đối về tài nguyên phần cứng.
- Về giao tiếp: Sử dụng chuẩn I2C chỉ tốn duy nhất 2 chân GPIO (SDA, SCL), giải quyết hoàn hảo bài toán thiếu chân của ESP32-CAM.
- Về hiển thị: Công nghệ OLED (Organic LED) cho độ tương phản cực cao, màu đen sâu và khả năng hiển thị sắc nét các dòng thông báo trạng thái, IP Address hay Icon trạng thái dù kích thước nhỏ gọn. Giao diện nhìn hiện đại và tinh tế.

- Về năng lượng: Do không sử dụng đèn nền (mỗi điểm ảnh tự phát sáng), màn hình này tiêu thụ điện năng rất thấp, phù hợp cho hệ thống chạy nguồn hạn chế.

=> **Kết luận:** Dựa trên bảng so sánh và phân tích kỹ thuật, nhóm quyết định chọn Màn hình OLED 0.96 inch làm khối hiển thị. Sự lựa chọn này đảm bảo tính thẩm mỹ, tiết kiệm năng lượng và quan trọng nhất là phù hợp với tài nguyên phần cứng hạn chế của ESP32-CAM.

2.3.5. Mô tơ

Bảng 2.7: So sánh mô tơ

Tiêu chí	Servo SG90	Servo MG996R	Solenoid Lock
			
Lực kéo	~1.6 kg/cm (Rất yếu)	10 - 12 kg/cm (Khá khỏe)	Không tính bằng kg/cm, mà là lực giữ chốt
	3	4	4
Cấu tạo	Bánh răng Nhựa (Dễ vỡ)	Bánh răng Kim loại (Đồng/Nhôm)	Cuộn dây điện từ và chốt thép
	3	4	4
Cơ chế	Quay góc 0-180 độ (Mở cánh cửa)	Quay góc 0-180 độ (Gạt chốt nặng)	Thụt chốt vào để mở khóa (Cửa tự bung hoặc đẩy tay)
	4	4	3

Mạch điều khiển	Đơn giản: Nối trực tiếp chân PWM của ESP32	Đơn giản: Nối trực tiếp chân PWM của ESP32	Phức tạp: Cần Relay hoặc MOSFET kích dòng
	5	5	2
Dòng điện tiêu thụ	Thấp (~100mA - 500mA)	Cao (~500mA - 1500mA)	Rất cao (>1A liên tục)
	5	3	1
Ứng dụng	Mô hình đồ chơi, cần gạt siêu nhẹ	Mô hình cửa lớn, chốt cửa gỗ thật (dạng gạt)	Hệ thống cửa từ, cửa chung cư, văn phòng
Giá thành	Rẻ (~25k - 35k VNĐ)	Trung bình (~80k - 120k VNĐ)	Trung bình (~80k - 150k VNĐ)
	5	3	3
Điểm trung bình	4.17	3.83	2.83

Dựa trên bảng khảo sát thông số kỹ thuật giữa ba loại cơ cấu khóa/mở cửa phổ biến là Servo SG90, Servo MG996R và Khóa điện từ Solenoid, ta có kết quả đánh giá chi tiết như sau:

Khóa điện từ Solenoid:

- Đây là thiết bị chuyên dụng cho các hệ thống kiểm soát cửa thực tế (chung cư, văn phòng) nhờ cơ chế chốt thép chắc chắn và độ bền cao.
- Tuy nhiên, nhược điểm lớn nhất là yêu cầu về nguồn điện. Solenoid tiêu thụ dòng điện rất lớn (thường >1A) để duy trì trạng thái mở, gây nóng và sụt áp nhanh chóng nếu nguồn cấp không đủ mạnh. Ngoài ra, để điều khiển Solenoid,

hệ thống bắt buộc phải thiết kế thêm mạch kích ly cách (Relay hoặc MOSFET công suất), làm tăng độ phức tạp và chi phí của mạch điều khiển.

Động cơ Servo MG996R (Bánh răng kim loại):

- Ưu điểm là lực kéo rất khỏe (lên tới 11kg/cm) và độ bền cao nhờ hệ thống bánh răng kim loại, phù hợp để kéo các chốt cửa nặng trong thực tế.
- Mặc dù vậy, kích thước lớn và trọng lượng nặng là một trở ngại khi lắp đặt trên các mô hình demo nhỏ gọn. Tương tự như Solenoid, dòng khởi động của MG996R khá cao, dễ gây nhiễu nguồn cho vi điều khiển nếu không có mạch nguồn được thiết kế kỹ lưỡng. Giá thành cũng cao gấp 3-4 lần so với dòng servo nhỏ.

Động cơ Servo SG90 (Micro Servo):



- Đây là thiết bị đạt điểm số cao nhất về tiêu chí "Hiệu quả/Chi phí" cho quy mô mô hình bài tập lớn.
- Về lực kéo: Mặc dù lực kéo chỉ khoảng 1.6 kg/cm và bánh răng nhựa, nhưng hoàn toàn đủ sức để kéo các chốt cửa mô hình nhẹ hoặc chốt gạt nhựa trong phạm vi thử nghiệm đồ án.
- Về điều khiển: Sử dụng tín hiệu PWM trực tiếp từ chân GPIO của ESP32 mà không cần mạch đệm phức tạp.
- Về năng lượng: Dòng tiêu thụ thấp, ít gây sụt áp nguồn, giúp hệ thống hoạt động ổn định hơn khi dùng chung nguồn với Camera và WiFi.
- Về kinh tế: Giá thành rất rẻ (chỉ khoảng 25.000đ - 35.000đ), giúp giảm tối đa chi phí tổng thể của dự án.

=> **Kết luận:** Dựa trên bảng so sánh và mục tiêu xây dựng mô hình kiểm thử (Demo), nhóm quyết định chọn Servo SG90 làm cơ cấu chấp hành đóng mở cửa. Sự lựa chọn này đảm bảo tính nhỏ gọn, dễ điều khiển và tiết kiệm chi phí, phù hợp với phạm vi nghiên cứu của đề tài.

2.3.6. Khởi nguồn

Bảng 2.8: So sánh nguồn

Tiêu chí	Adapter 12V 2A + LM2596	2 Pin Li-ion 18650 (Nối tiếp)
----------	-------------------------	-------------------------------

	<p>LM2596 Ổn Áp 3A input 3- 40V</p> 	
Độ ổn định	<p>Rất cao</p> <p>Nguồn điện luôn dồi dào, input ổn định. LM2596 hoạt động mát và chuẩn xác, không lo sụt áp.</p>	<p>Trung bình</p> <p>Điện áp tụt dần từ 8.4V xuống 6V. Khi pin yếu, áp không đủ để ổn định hoạt động</p>
	5	3
Thời gian chạy	<p>Vô hạn</p> <p>Chạy vĩnh viễn (24/7).</p>	<p>Hạn chế</p> <p>Chỉ trụ được 4-6 tiếng.</p>
	5	2
Khả năng chịu tải	<p>Tốt</p> <p>Khá tốt nếu có tụ bù 1000uF.</p>	<p>Rất tốt</p> <p>Pin Li-ion có dòng xả tức thời lớn (High C-rate)</p>
	4	5
Tính di động	<p>Kém</p> <p>Bị "xích" vào ổ cắm tường.</p> <p>Dây dợ lằng nhằng</p>	<p>Tốt</p> <p>Gọn nhẹ, không dây rợ.</p>
	1	5
An toàn & Lắp đặt	<p>Lắp đặt đơn giản. Rất ít nguy cơ cháy nổ hay hỏng hóc linh kiện.</p>	<p>Lắp đặt phức tạp. Bắt buộc cần mạch bảo vệ (BMS). Dễ gây nổ pin.</p>
	5	2
Chi phí	<p>Rẻ (9/10)</p> <p>Tận dụng nguồn Modem cũ (0đ) + LM2596 (15k) +</p>	<p>Cao (6/10)</p>

	Tụ (2k). Tổng chi phí cực thấp.	Tốn kém. Cần mua Pin xịn + Mạch BMS + Đế pin + Bộ sạc rời. Tổng >150k.
	5	3
Điểm trung bình	4.16	3.33

Dựa trên bảng khảo sát thông số kỹ thuật và chấm điểm giữa hai phương án cấp nguồn là Sử dụng Adapter kèm mạch hạ áp và Sử dụng Pin sạc Li-ion, ta có kết quả đánh giá tổng quan như sau:

Pin Li-ion 18650:

- Sở hữu ưu điểm tuyệt đối về tính di động và khả năng xả dòng mạnh mẽ, giúp Servo hoạt động rất dứt khoát.
- Tuy nhiên, điểm trừ lớn nhất là tính ổn định theo thời gian. Đối với một hệ thống an ninh cửa cần hoạt động 24/7, việc phải thay pin hoặc sạc pin hàng ngày là không khả thi. Ngoài ra, rủi ro về an toàn cháy nổ và chi phí đầu tư cao cho mạch bảo vệ cũng là một rào cản lớn đối với đề tài sinh viên.

Adapter 12V + LM2596:

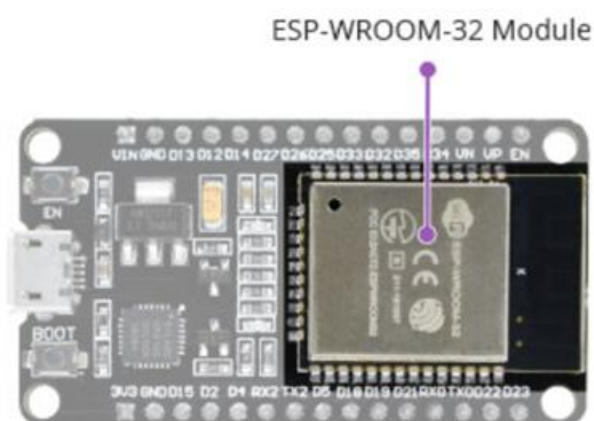
- Đây là thiết bị đạt điểm số trung bình cao nhất nhờ sự ổn định và chi phí thấp.
- Về ổn định: Sử dụng nguồn đầu vào 12V cao hơn nhiều so với mức yêu cầu, kết hợp với mạch giảm áp xung (Buck Converter) LM2596 giúp hiệu suất chuyển đổi cao, đảm bảo ESP32 không bao giờ gặp hiện tượng "Brownout Reset" do thiếu áp.
- Về kinh tế: Có thể tận dụng lại các Adapter WiFi/Modem cũ có sẵn, giúp giảm chi phí phần cứng xuống mức thấp nhất. Nhược điểm về dòng khởi động của Servo có thể khắc phục đơn giản bằng một tụ điện dung lượng lớn.

=> **Kết luận:** Dựa trên bảng so sánh và phân tích, nhóm chúng em quyết định chọn Phương án A: Adapter 12V kết hợp Module LM2596 làm khối nguồn nuôi hệ thống. Sự lựa chọn này chấp nhận hy sinh tính di động để đổi lấy sự hoạt động bền bỉ, ổn định 24/7 – yếu tố quan trọng nhất của một hệ thống khóa cửa.

2.4. Danh sách các linh kiện cần dùng

2.4.1. Giới thiệu về ESP32 DevKit V1

ESP32 được tạo ra bởi công ty Espressif Systems như một bản nâng cấp "siêu mạnh mẽ" cho dòng chip ESP8266 nổi tiếng. Trong khi ESP8266 đã thay đổi thế giới điện tử với WiFi giá rẻ, ESP32 đưa mọi thứ lên một tầm cao mới bằng cách tích hợp thêm Bluetooth 4.0 (BLE) và cấu hình phần cứng mạnh vượt trội. ESP32 DevKit V1 là bo mạch phát triển phổ biến nhất, là lựa chọn tốt nhất cho người mới bắt đầu vì nó có đầy đủ tài liệu hướng dẫn và cộng đồng hỗ trợ lớn.



Hình 2.2: Vị trí module

2.4.1.1. Cấu trúc phần cứng cốt lõi

Trái tim của bo mạch là module ESP-WROOM-32, đóng vai trò là "bộ não" điều khiển toàn bộ hệ thống.

Bộ vi xử lý: Sử dụng CPU Tensilica Xtensa® Dual-Core 32-bit LX6,. Kiến trúc lõi kép này cho phép ESP32 xử lý đồng thời nhiều tác vụ phức tạp, chẳng hạn như vừa chạy Web Server vừa đọc dữ liệu cảm biến, nhanh hơn nhiều so với các chip lõi đơn như Arduino Uno hay ESP8266.

Tốc độ xung nhịp: Có thể điều chỉnh linh hoạt từ 80 MHz đến 240 MHz. Tại mức 240 MHz, nó đạt hiệu suất 600 DMIPS (triệu chỉ thị trên giây).

2.4.1.2. Bộ nhớ

- 448 KB ROM: Dùng cho quá trình khởi động và các chức năng cốt lõi.
- 520 KB SRAM nội bộ: Bộ nhớ tạm thời để lưu trữ dữ liệu khi chương trình đang chạy.

- 4 MB Flash ngoài: Nơi lưu trữ vĩnh viễn chương trình và dữ liệu ngay cả khi mất điện.

2.4.1.3. Khả năng kết nối không dây



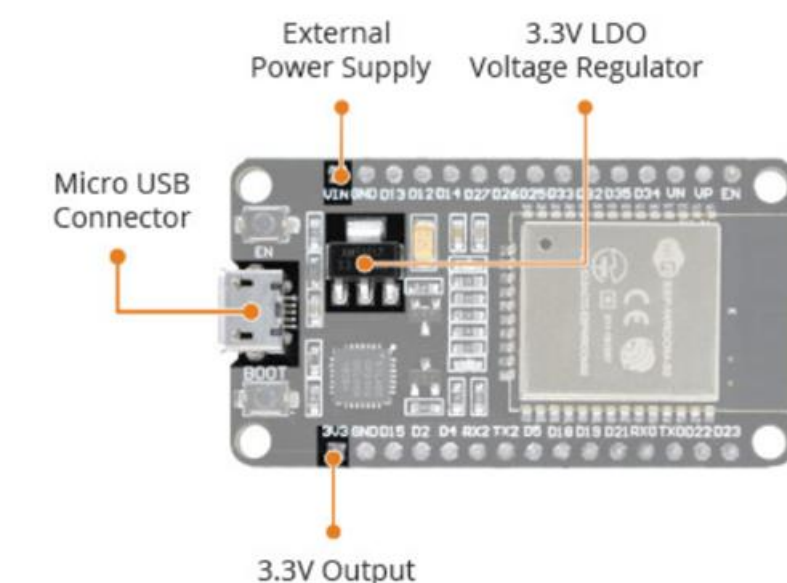
Hình 0.1: Vị trí ăng-ten

ESP32 cực kỳ linh hoạt nhờ tích hợp cả WiFi và Bluetooth:

- WiFi: Hỗ trợ chuẩn 802.11b/g/n HT40. Nó có thể hoạt động ở 3 chế độ: Station (kết nối vào mạng WiFi có sẵn), Soft AP (tự tạo điểm phát WiFi cho thiết bị khác kết nối vào), và WiFi Direct (kết nối trực tiếp giữa hai thiết bị không cần bộ định tuyến).
- Bluetooth: Hỗ trợ song song Bluetooth Classic (dùng cho loa, truyền file) và Bluetooth 4.2 - BLE (Bluetooth tiết kiệm năng lượng, lý tưởng cho các thiết bị đeo và IoT chạy pin).

2.4.1.4. Nguồn điện và tiêu thụ năng lượng

Mạch được thiết kế để hoạt động an toàn và tiết kiệm điện



Hình 0.2: Vị trí các chân kết nối nguồn

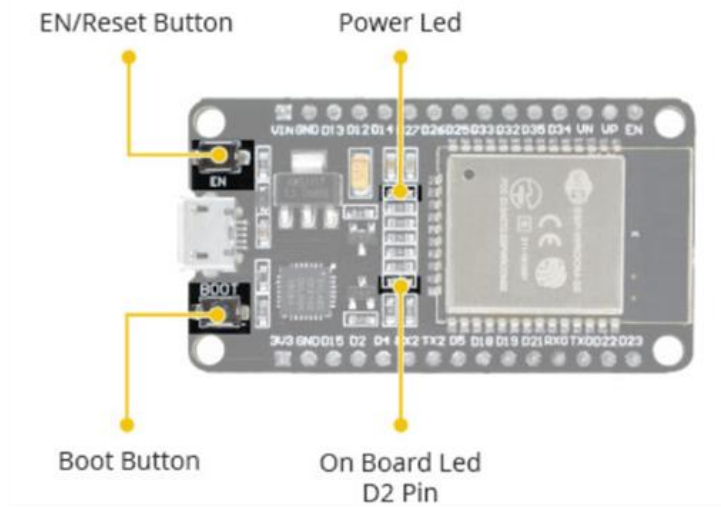
- Điện áp hoạt động: Chip ESP32 cần từ 2.2V đến 3.6V. Bo mạch tích hợp sẵn bộ điều chỉnh điện áp LDO để duy trì mức ổn định 3.3V với dòng điện đầu ra tối đa 600mA.
- Cách cấp nguồn: Người dùng có thể cấp nguồn qua cổng MicroB USB (tiện lợi nhất) hoặc cấp nguồn 5V trực tiếp qua chân VIN. Ngoài ra, chân 3V3 có thể dùng để cấp nguồn cho các linh kiện nhỏ bên ngoài.
- Mức tiêu thụ: Khi truyền tín hiệu RF, mạch tiêu thụ khoảng 250mA, nhưng ở chế độ Deep Sleep, con số này giảm xuống chỉ còn 5 μ A, giúp dự án chạy pin có thể kéo dài hàng tháng.

2.4.1.5. Hệ thống ngoại vi và giao tiếp (I/O)

Nhờ tính năng ghép kênh (multiplexing), các chân GPIO của ESP32 có thể cấu hình linh hoạt cho nhiều chức năng khác nhau:

- 25 chân GPIO: Có thể dùng cho vào/ra kỹ thuật số và hầu hết đều hỗ trợ ngắt (interrupts).
- 15 kênh ADC 12-bit: Cho phép đo tín hiệu analog với độ phân giải cao, thậm chí có thể hoạt động trong chế độ ngủ.
- kênh DAC 8-bit: Dùng để tạo ra điện áp analog thực từ tín hiệu số.
- 25 đầu ra PWM: Điều khiển độ sáng LED hoặc tốc độ động cơ.
- 9 chân cảm ứng (Touch Pads): Cảm nhận điện dung khi chạm tay vào, thay thế cho nút bấm vật lý.
- Giao tiếp dữ liệu:
 - + 3 cổng SPI: (SPI, HSPI, VSPI), trong đó HSPI và VSPI có thể sử dụng tự do (cổng thứ ba dùng cho bộ nhớ Flash nội bộ).
 - + 2 cổng I2C: Có thể cấu hình trên hầu hết các chân GPIO qua phần mềm.
 - + 2 cổng UART: Hỗ trợ RS232, RS485 và hồng ngoại (IrDA) với tốc độ lên tới 5 Mbps.
 - + 2 cổng I2S: Dùng để tích hợp âm thanh vào dự án.

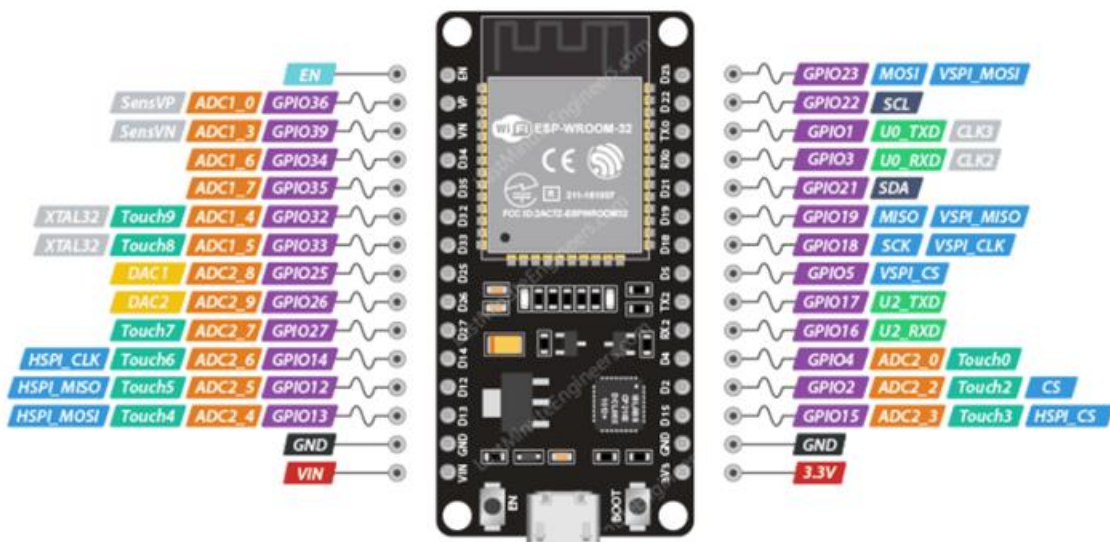
2.4.1.6. Các thành phần điều khiển trên bo mạch



Hình 2.4 Mô tả vị trí các thành phần điều khiển *Hình 0.3: Vị trí các thành phần điều khiển*

- Nút bấm: Nút EN (Enable) để reset mạch và nút BOOT để đưa mạch vào chế độ nạp chương trình.
- Đèn LED: Một LED đỏ báo nguồn và một LED xanh dương được kết nối với chân D2 để người dùng lập trình (ví dụ: chạy code Blink).
- Chip nạp: Sử dụng chip cầu nối CP2102 USB-to-UART để chuyển đổi tín hiệu từ máy tính sang chuẩn Serial mà ESP32 hiểu được.

2.4.1.7. Sơ đồ chân ESP32-DevKit V1



Hình 2.5 Sơ đồ chân ESP32-DevKit V1 *Hình 0.4: Sơ đồ chân ESP32-DevKit V1*

- Chân nguồn: Bo mạch bao gồm hai chân nguồn chính là VIN và 3V3.

- + Chân VIN: Cho phép bạn cấp nguồn trực tiếp cho ESP32 bằng nguồn điện 5V đã qua ổn áp.
- + Chân 3V3: Cung cấp điện áp đầu ra từ bộ ổn áp tích hợp trên bo mạch với dòng điện tối đa lên tới 600 mA. Bạn có thể dùng chân này để cấp nguồn cho các linh kiện ngoại vi nhỏ như cảm biến, đèn LED hoặc màn hình.
- GND là chân nối đất.
- Chân GPIO: Bo mạch có 25 chân GPIO, và các chân này cực kỳ linh hoạt vì bạn có thể gán các chức năng khác nhau cho chúng thông qua phần mềm. Chúng có thể được sử dụng cho:
 - + Đầu vào/đầu ra kỹ thuật số (Digital I/O).
 - + ADC để cảm biến tín hiệu analog.
 - + DAC để tạo tín hiệu analog.
 - + Cảm ứng chạm (Touch sensing).
 - + PWM để điều khiển động cơ hoặc đèn LED.
 - + Các giao thức truyền thông như I2C, SPI, UART, và I2S.
 - + Đặc biệt, hầu hết tất cả các chân GPIO trên ESP32 đều có thể được sử dụng làm ngắt (interrupts).
- Kênh ADC: ESP32 tích hợp hai bộ ADC SAR 12-bit, cho phép bạn đo tín hiệu analog trên 15 kênh khác nhau. Một số chân thậm chí có thể được dùng để tạo bộ khuếch đại lập trình được (programmable gain amplifier), rất hữu ích để đọc các tín hiệu analog cực nhỏ một cách chính xác. Ngoài ra, ESP32 có thể thực hiện các phép đo này ngay cả khi ở chế độ ngủ (sleep mode) để tiết kiệm điện năng.
- Kênh DAC: ESP32 có hai kênh DAC 8-bit, cho phép bạn chuyển đổi tín hiệu kỹ thuật số thành điện áp analog thực tế.
- Cảm ứng chạm: ESP32 có 9 chân GPIO cảm ứng điện dung. Các chân này có thể phát hiện sự thay đổi nhỏ về điện dung khi có vật dẫn điện (như ngón tay của bạn) chạm gần vào, cho phép bạn tạo các nút bấm cảm ứng mà không cần thêm cảm biến bên ngoài.

- Chân SPI: ESP32 hỗ trợ ba giao diện SPI (SPI, HSPI, và VSPI) có thể hoạt động ở cả chế độ chủ (master) hoặc tớ (slave). Các giao diện này hỗ trợ đầy đủ tính năng SPI tiêu chuẩn như:
 - + 4 chế độ thời gian (timing modes) truyền dữ liệu khác nhau.
 - + Tốc độ lên tới 80 MHz.
 - + Bộ đệm FIFO lên đến 64 byte để truyền dữ liệu mượt mà.
 - + Lưu ý: chỉ có thể sử dụng tự do VSPI và HSPI, vì bus SPI thứ ba đã được dùng cho bộ nhớ flash tích hợp trên bo mạch.
- Chân I2C: ESP32 sở hữu hai bộ điều khiển I2C độc lập (I2C0 và I2C1). Nó hỗ trợ giao tiếp I2C trên hầu hết mọi chân GPIO vì ngoại vi I2C có thể cấu hình hoàn toàn bằng phần mềm.
- Chân UART: ESP32 có hai giao diện UART là UART0 và UART2, hỗ trợ các tiêu chuẩn giao tiếp như RS232, RS485 và thậm chí có thể làm việc với hồng ngoại (IrDA) ở tốc độ lên tới 5 Mbps.
- Chân PWM: Hầu như tất cả các chân GPIO đều có thể xuất tín hiệu PWM. PWM là một kỹ thuật thông minh giúp bạn bật/tắt chân tín hiệu cực nhanh để mô phỏng các mức điện áp khác nhau. Điều này cực kỳ hữu ích cho việc điều chỉnh tốc độ động cơ hoặc làm mờ (dimming) đèn LED.
- Chân EN: Đóng vai trò như một nút nhấn reset cứng cho ESP32.
 - + Khi chân này ở mức CAO (HIGH), chip sẽ được cấp điện và chạy bình thường. Bo mạch đã tích hợp sẵn một điện trở kéo lên (pull-up) để giữ nó ở mức HIGH mặc định.
 - + Nếu bạn kéo chân EN xuống mức THẤP (LOW) (xảy ra khi bạn nhấn nút "EN" trên mạch), ESP32 sẽ tắt ngay lập tức và khởi động lại ngay khi chân này trở lại mức CAO.

2.4.2. ESP32-Cam

ESP32-Cam là một bo mạch phát triển vi điều khiển được thiết kế và sản xuất bởi AI-Thinker. Nó là một giải pháp hệ thống trên vi mạch (System-on-Chip - SoC) chi phí thấp, kích thước nhỏ, nổi bật nhờ việc tích hợp mô-đun camera với vi điều khiển ESP32-S.

2.4.2.1. Cấu trúc phần cứng cốt lõi

Điểm cốt lõi của bo mạch là chip ESP32-S, một vi điều khiển lõi kép (dual-core) mạnh mẽ của Espressif Systems, vốn đã tích hợp sẵn các giao thức kết nối không dây Wi-Fi 802.11 b/g/n và Bluetooth v4.2 BR/EDR & BLE (Bluetooth Low Energy).



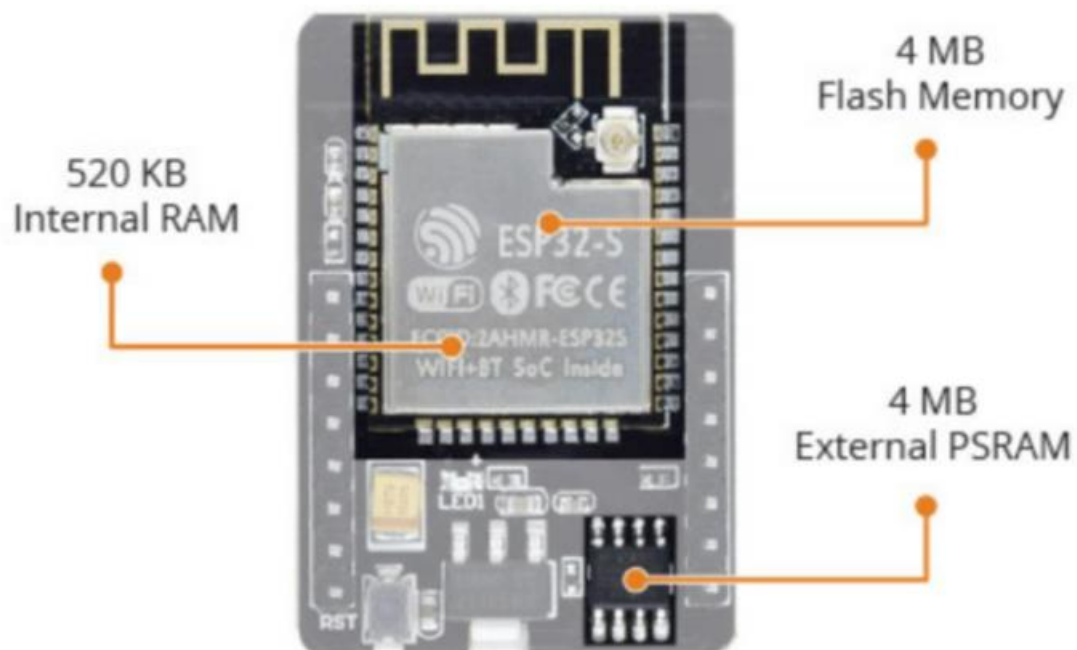
Hình 2.6 ESP32-CAM Hình 0.5: ESP32-CAM

Trung tâm của bo mạch là bộ vi xử lý lõi kép Tensilica Xtensa® LX6 32-bit, hoạt động với tần số xung nhịp lên đến 240 MHz. Hiệu suất xử lý cao này cho phép ESP32-CAM có khả năng thực thi các tác vụ tính toán chuyên sâu như xử lý video, nhận dạng khuôn mặt và các ứng dụng trí tuệ nhân tạo (AI).

2.4.2.2. Bộ nhớ

Mô-đun ESP32-S tích hợp 520 KB SRAM nội. Để đáp ứng các ứng dụng yêu cầu bộ nhớ cao (như xử lý đồ họa hoặc luồng video), bo mạch được trang bị thêm 4MB PSRAM (Pseudo-Static RAM) ngoài, giúp mở rộng đáng kể dung lượng bộ nhớ làm việc.

Bo mạch tích hợp 4 MB bộ nhớ flash trên chip, dùng để lưu trữ chương trình và dữ liệu người dùng.



Hình 2.7 Mô tả vị trí bộ nhớ của ESP32-CAM Hình 0.6: Vị trí bộ nhớ của ESP32-CAM

2.4.2.3. Camera

Thành phần cốt lõi của bo mạch là cảm biến camera OV2640. Cảm biến này có độ phân giải 2 megapixel, hỗ trợ độ phân giải hình ảnh tối đa 1600x1200 pixel, đáp ứng yêu cầu cho nhiều ứng dụng giám sát và thu thập hình ảnh.



Hình 2.8 Camera của ESP32-CAM Hình 0.7: Camera của ESP32-CAM

2.4.2.4. Bộ nhớ lưu trữ mở rộng

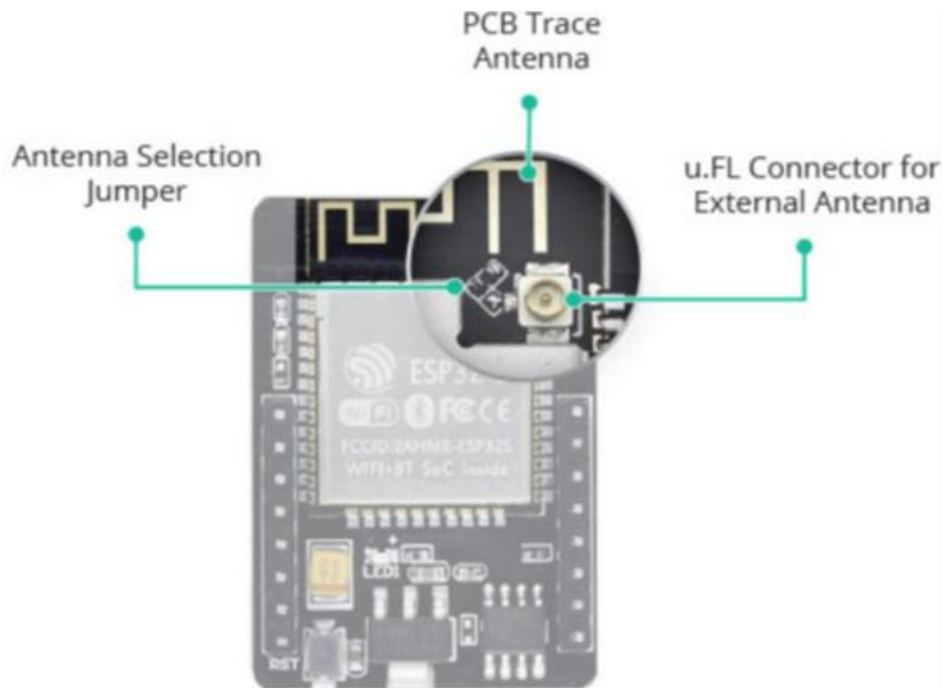
Ngoài bộ nhớ tích hợp, ESP32-CAM còn có khe cắm thẻ nhớ MicroSD dung lượng từ 4 GB đến 32 GB, được định dạng FAT32. Tính năng này rất quan trọng cho các ứng dụng cần ghi lại dữ liệu (datalogging) hoặc lưu trữ hình ảnh, video trong thời gian dài.



Hình 2.9 Mô tả vị trí bộ nhớ lưu trữ mở rộng Hình 0.8: Vị trí bộ nhớ lưu trữ mở rộng

2.4.2.5. Ăng ten

Về giao tiếp không dây, ESP32-CAM cung cấp cả ăng-ten theo dõi PCB tích hợp và đầu nối u.FL để kết nối ăng-ten ngoài. Ăng-ten tích hợp hoạt động tốt cho hầu hết các dự án mà bo mạch tương đối gần với bộ định tuyến WiFi, trong khi ăng-ten ngoài có thể mở rộng đáng kể phạm vi và cải thiện độ ổn định tín hiệu — đặc biệt là khi bo mạch của bạn nằm trong hộp hoặc vỏ hộp, hoặc khi sử dụng ngoài trời, nơi tín hiệu có thể yếu hơn.

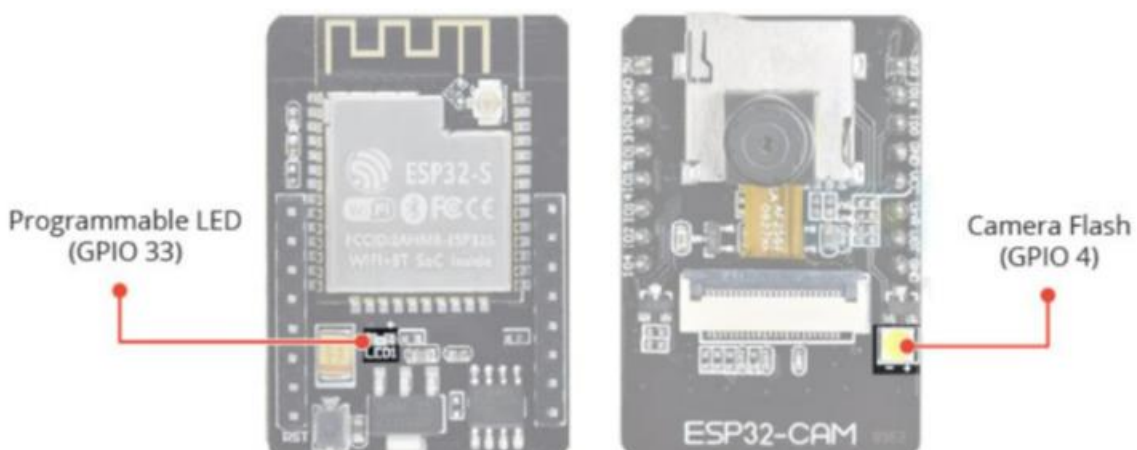


Hình 2.10 Mô tả vị trí ăng-ten của ESP32-CAM Hình 0.9: Vị trí ăng-ten của ESP32-CAM

2.4.2.6. Đèn LED

Đèn LED flash màu trắng nằm gần ống kính máy ảnh được thiết kế để sử dụng làm đèn flash máy ảnh, nhưng cũng có thể được sử dụng để chiếu sáng chung. Đèn được kết nối với GPIO 4 và có thể lập trình bởi người dùng.

Ngoài ra còn có một đèn LED nhỏ màu đỏ ở mặt sau của bo mạch, rất lý tưởng để hiển thị thông tin trạng thái, chẳng hạn như thiết bị đã được kết nối với WiFi hay đang ghi hình. Nó được kết nối với GPIO 33 và cũng có thể được lập trình bởi người dùng.



Hình 2.11 Mô tả vị trí LED của ESP32-CAM Hình 0.10: Vị trí LED của ESP32-CAM

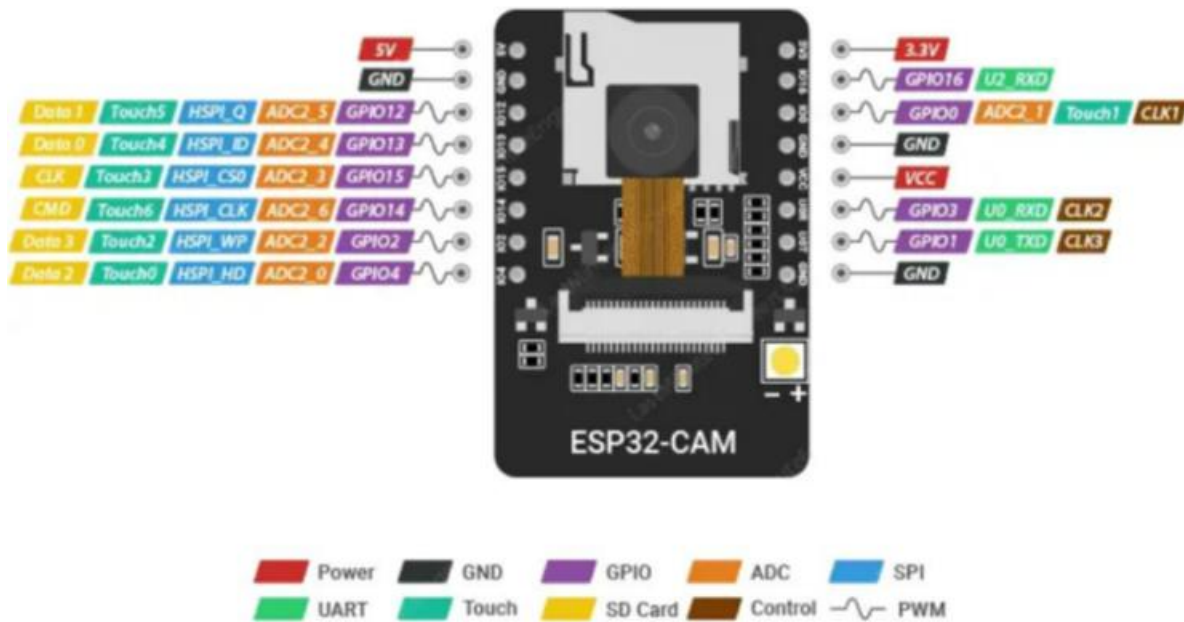
2.4.2.7. Thông số kỹ thuật

- Bộ xử lý

- + CPU: Bộ vi xử lý LX6 lõi kép 32-bit Xtensa, hoạt động ở tần số 240 MHz và hoạt động ở tốc độ lên tới 600 DMIPS.
- + Bộ đồng xử lý tiêu thụ điện năng cực thấp (ULP).
- Bộ nhớ
 - + SRAM 520 KB
 - + SRAM ngoài 4MB
 - + Bộ nhớ flash bên trong 4MB
- Kết nối không dây
 - + Wi-Fi: 802.11 b/g/n.
 - + Bluetooth: v4.2 BR/EDR và BLE (chia sẻ radio với Wi-Fi)
- Camera
 - + Cảm biến 2 Megapixel OV2640
 - + Kích thước mảng UXGA 1622×1200
 - + Các định dạng đầu ra bao gồm YUV422, YUV420, RGB565, RGB555 và dữ liệu nén 8 bit
 - + Tốc độ truyền hình ảnh từ 15 đến 60 khung hình / giây
 - + Đèn LED flash tích hợp
- Hỗ trợ thẻ nhớ microSD
- Bảo mật
 - + Tất cả các tính năng bảo mật tiêu chuẩn IEEE 802.11 đều được hỗ trợ, bao gồm WPA, WPA/WPA2 và WAPI
 - + Khởi động an toàn
 - + Mã hóa flash
 - + OTP 1024-bit, tối đa 768-bit cho khách hàng
 - + Tăng tốc phần cứng mật mã: AES, SHA-2, RSA, mật mã đường cong elip (ECC), bộ tạo số ngẫu nhiên (RNG)
- Quản lý nguồn điện
 - + Bộ điều chỉnh độ sụt áp thấp bên trong
 - + Miền công suất riêng lẻ cho RTC
 - + Dòng điện ngủ sâu 5μA

- + Thúc đẩy từ ngắt GPIO, bộ đếm thời gian, phép đo ADC, ngắt cảm biến cảm ứng điện dung
- Tiêu thụ điện năng của ESP32-CAM
 - + Dung lượng dao động từ 80 mAh khi không phát video đến khoảng 100~160 mAh khi phát video; khi bật đèn flash, nó có thể đạt tới 270 mAh.

2.4.2.8. Sơ đồ chân ESP32-CAM



Hình 2.12 Sơ đồ chân ESP32-CAM *Hình 0.11: Sơ đồ chân của ESP32-CAM*

- Chân nguồn: Bo mạch có hai chân cấp nguồn: một chân 5V và một chân 3V3. Về mặt kỹ thuật, có thể cấp nguồn cho ESP32-CAM thông qua chân 3V3 hoặc chân 5V. Tuy nhiên, có nhiều sự cố khi sử dụng tùy chọn 3V3, vì vậy cần cấp nguồn cho ESP32-CAM thông qua chân 5V. Ngoài ra còn có một chân VCC thường cung cấp 3,3V từ bộ điều chỉnh điện áp tích hợp, nhưng nó cũng có thể được đặt thành đầu ra 5V bằng cách điều chỉnh liên kết zero-ohm nhỏ (một jumper nhỏ) gần chân VCC.
- GND là chân nối đất.
- Chân GPIO: Bản thân chip ESP32-S có 32 chân GPIO. Tuy nhiên, vì nhiều chân trong số này đã được camera và bộ nhớ PSRAM sử dụng bên trong, nên chỉ có 10 chân GPIO thực sự có sẵn để sử dụng trên bo mạch ESP32-CAM. Các chân này có thể được gán cho nhiều chức năng ngoại vi khác nhau, chẳng hạn như UART, SPI, ADC và Touch.

- Chân UART: Chip ESP32-S thực sự có hai giao diện UART là UART0 và UART2. Tuy nhiên, chỉ có chân RX (GPIO 16) của UART2 bị hỏng, khiến UART0 trở thành UART duy nhất có thể sử dụng được trên ESP32-CAM (GPIO 1 và GPIO 3). Ngoài ra, vì ESP32-CAM thiếu cổng USB nên các chân này phải dùng để flash cũng như kết nối với các thiết bị UART như GPS, cảm biến vân tay, cảm biến khoảng cách, v.v.
- Chân thẻ MicroSD: được sử dụng để kết nối thẻ nhớ microSD. Nếu không sử dụng thẻ nhớ microSD, có thể sử dụng các chân này làm đầu vào và đầu ra thông thường.
- Chân ADC: Trên ESP32-CAM, chỉ có chân ADC2 bị hỏng. Tuy nhiên, vì các chân ADC2 được trình điều khiển WiFi sử dụng nội bộ nên chúng không thể được sử dụng khi bật Wi-Fi.
- Ghim cảm ứng: Có thể phát hiện khi một tải điện dung, chẳng hạn như ngón tay người, đến gần chúng. ESP32-CAM có bảy chân cảm ứng như vậy. Có thể sử dụng các chân này để tạo các nút bấm hoặc điều khiển cảm ứng cho dự án của mình mà không cần bất kỳ công tắc vật lý nào.
- Chân SPI: ESP32-CAM chỉ có một SPI (VS SPI) ở chế độ phụ và chế độ chính.
- Chân điều khiển xung điện: ESP32-CAM có 10 kênh PWM, nghĩa là nó có thể tạo tín hiệu PWM trên tất cả các chân GPIO. PWM, hay Điều chế độ rộng xung, là một kỹ thuật được sử dụng để điều khiển các yếu tố như tốc độ động cơ hoặc độ sáng đèn LED bằng cách bật và tắt tín hiệu nhanh chóng.

2.4.3. Giới thiệu về cảm biến PIR HC-SR501



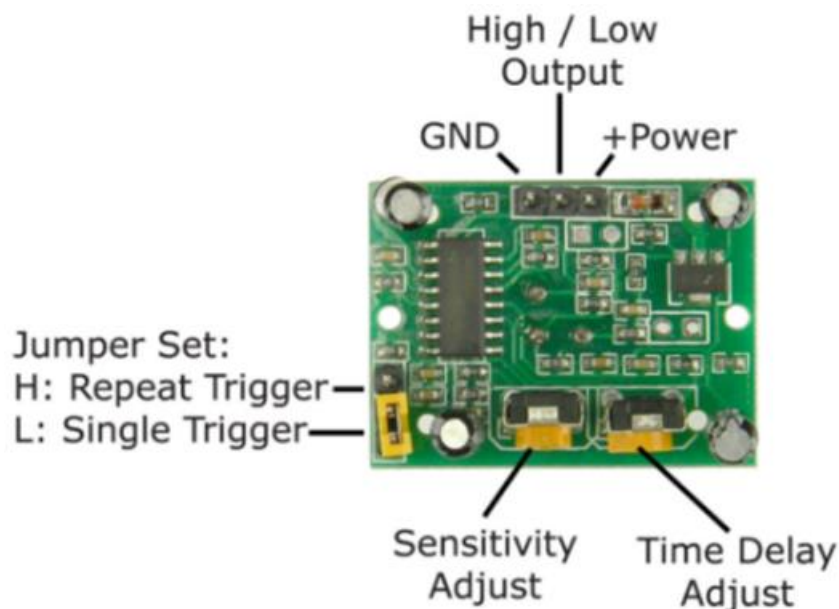
Hình 2.13 Cảm biến PIR HC-SR501 Hình 0.12: Cảm biến PIR HC-SR501

Cảm biến chuyển động hồng ngoại HC-SR501 là một module điện tử phổ biến sử dụng công nghệ PIR (Passive Infrared) để phát hiện sự thay đổi bức xạ hồng ngoại từ các vật thể phát nhiệt như con người hoặc động vật. Đây là thiết bị lý tưởng cho các ứng dụng tự động hóa như bật tắt đèn tự động, hệ thống báo động an ninh và các thiết bị nhà thông minh tiết kiệm năng lượng.

2.4.3.1. Nguyên lý hoạt động và cấu tạo

- Thấu kính Fresnel: Nguyên lý: Cảm biến không phát ra năng lượng mà chỉ nhận tia hồng ngoại từ các vật thể. Nó sử dụng một mắt cảm biến pyroelectric (thường là RE200B) có hai đơn vị cảm nhận để phát hiện sự chênh lệch nhiệt độ khi có vật thể di chuyển qua các vùng quét khác nhau.
- Phía trên cảm biến có một nắp vòm trắng chính là thấu kính Fresnel. Thấu kính này đóng vai trò tập trung tia hồng ngoại vào mắt cảm biến và chia vùng quan sát thành nhiều phân đoạn để tăng độ nhạy.
- IC xử lý: Module được xây dựng xung quanh IC BISS0001, chuyên dụng để xử lý tín hiệu từ cảm biến PIR và điều khiển đầu ra.

2.4.3.2. Các thành phần điều khiển trên board mạch



Hình 0.13: Các thành phần điều khiển trên board mạch

Mặt sau của module HC-SR501 thường có 3 chân cắm (VCC, OUT, GND) và các bộ phận điều chỉnh.

- 3 chân cắm

- + VCC: Điện áp đầu vào là +5V cho các ứng dụng điển hình. Có thể nằm trong khoảng từ 4,5V đến 12V.
- + Out: Xung kỹ thuật số cao (3,3V) khi được kích hoạt (chuyển động phát hiện), kỹ thuật số thấp (0V) khi không hoạt động (không cử động).
- + GND: Kết nối với mặt đất của mạch.
- Biến trở chỉnh độ nhạy (Sensitivity Adjust): Xoay theo chiều kim đồng hồ để tăng khoảng cách phát hiện (tối đa 7m) và ngược chiều kim đồng hồ để giảm (tối thiểu 3m).
- Biến trở chỉnh thời gian trễ (Time Delay Adjust): Xác định thời gian đầu ra duy trì ở mức cao sau khi phát hiện chuyển động. Xoay theo chiều kim đồng hồ để tăng thời gian.
- Jumper chọn chế độ kích hoạt (Trigger Selection):
 - + Chế độ L (Single Trigger): Sau khi phát hiện chuyển động, đầu ra sẽ giữ ở mức cao trong khoảng thời gian trễ đã thiết lập rồi trở về mức thấp, ngay cả khi vật thể vẫn đang di chuyển.
 - + Chế độ H (Repeatable Trigger): Đầu ra sẽ duy trì mức cao miễn là vẫn còn chuyển động được phát hiện trong vùng quét.

2.4.3. Giới thiệu về màn hình SSD1306

2.4.3.1. Nguyên lý hoạt động và cấu tạo

SSD1306 là một IC điều khiển màn hình (driver IC) phổ biến, vận hành nhiều loại module OLED kích thước 0.96-inch với độ phân giải 128×64 pixel. Các module này thường được sử dụng rộng rãi trong các hệ thống nhúng và các dự án điện tử cá nhân (hobby electronics) nhờ độ tương phản cao, góc nhìn rộng và mức tiêu thụ điện năng thấp.

Khác với các màn hình LCD truyền thống, màn hình OLED (Organic Light Emitting Diode - Điốt phát quang hữu cơ) không cần đèn nền (backlight). Mỗi điểm ảnh (pixel) sẽ tự phát sáng, giúp tạo ra màu đen sâu và văn bản hiển thị rõ nét, ngay cả trong môi trường thiếu sáng.

Hầu hết các module SSD1306 đều hỗ trợ cả hai giao thức giao tiếp là I2C và SPI. Trong bài hướng dẫn này, chúng ta sẽ sử dụng phiên bản I2C, vì nó yêu cầu ít chân kết nối hơn và dễ dàng thiết lập với STM32.

Những màn hình này rất lý tưởng để thêm các giao diện người dùng đơn giản, bảng hiển thị thông số cảm biến hoặc hiển thị văn bản cho các dự án nhúng. Bạn có thể hiển thị chuỗi ký tự, con số, đồ họa, biểu tượng (icon) và thậm chí là các hoạt ảnh (animation) nhỏ.



Hình 2.14 Màn hình OLED SSD1306 Hình 0.14: Màn hình OLED SSD1306

- Thông số của màn OLED SSD1306

Bảng 0.9: Thông số kỹ thuật của màn OLED SSD1306

Đặc điểm	Thông số
Loại màn hình	OLED đơn sắc (Monochrome OLED)
IC Điều khiển (Driver)	SSD1306
Độ phân giải	128 × 64 điểm ảnh (loại phổ biến)
Điện áp hoạt động	3.3 V hoặc 5 V (tùy thuộc vào module)
Giao thức giao tiếp	I2C (cũng có sẵn các biến thể dùng SPI)

Địa chỉ I2C	Thường là 0x3C hoặc 0x3D
Góc nhìn	>160°
Đèn nền	Không yêu cầu (các điểm ảnh tự phát sáng)
Tiêu thụ điện năng	Rất thấp (lý tưởng cho các ứng dụng chạy pin)
Nhiệt độ hoạt động	-30°C đến 70°C

- Kết nối chân với ESP32:

Bảng 0.10: Kết nối chân của màn OLED SSD1306 với ESP32

Chân OLED	Chân ESP32	Chú thích
GND	GND	Nối đất
VCC	3.3 V	Màn hình chạy tốt nhất ở 3.3V
SCL	GPIO 22	Chân xung nhịp (Clock)
SDA	GPIO 21	Chân dữ liệu (Data)

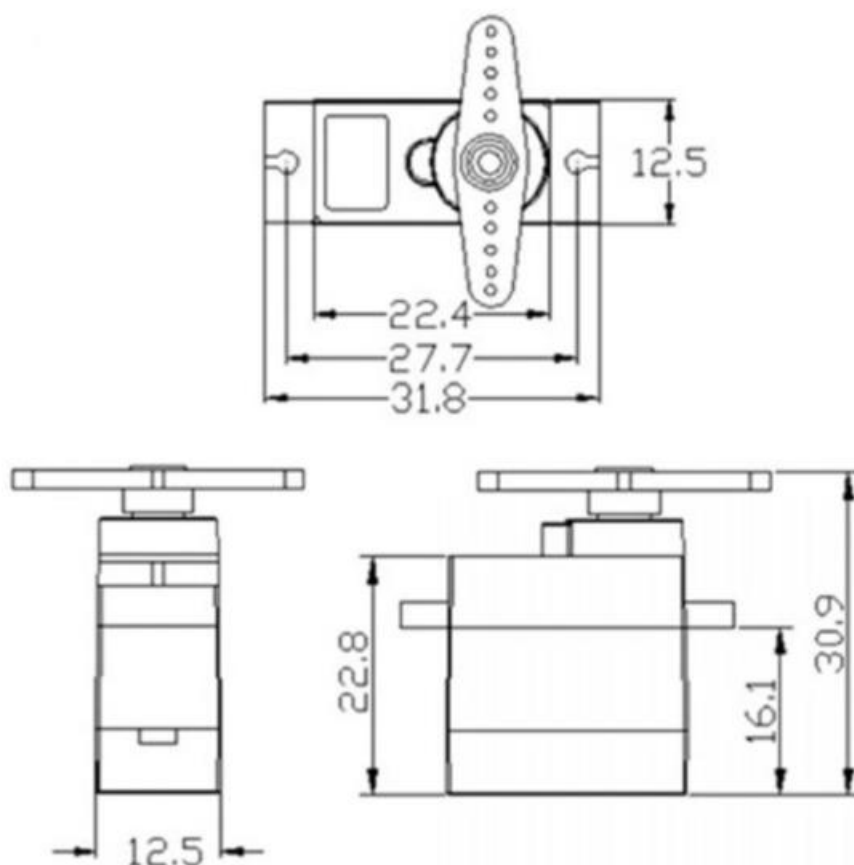
2.4.3.2. Giao thức truyền dữ liệu I2C

Hệ thống sử dụng chuẩn I2C để giao tiếp với màn hình OLED SSD1306.

- Nguyên lý: I2C là giao thức truyền thông nối tiếp đồng bộ, sử dụng hai dây tín hiệu:
 - + SDA: Đường truyền dữ liệu hai chiều.
 - + SCL: Đường xung nhịp đồng bộ do vi xử lý tạo ra.
- Ưu điểm: Chỉ tốn 2 chân GPIO của vi xử lý để điều khiển nhiều thiết bị, phù hợp với ESP32-CAM vốn có số lượng chân giới hạn.
- Ứng dụng trong đề tài: ESP32 gửi các lệnh và dữ liệu điểm ảnh tới OLED theo địa chỉ 0x3C để hiển thị thông báo.

2.4.4. Giới thiệu về Servo SG90

2.4.4.1. Nguyên lý hoạt động và cấu tạo



Hình 2.15 Mô tả kích thước của Servo SG90 **Hình 0.15:** Kích thước của Servo SG90

2.4.4.2. Thông số kỹ thuật

Bảng 0.11: Thông số kỹ thuật của Servo SG90

Điện áp hoạt động	4.8-5V
Trọng lượng	9g
Góc quay	0 đến 180 độ
Tốc độ	0.12 sec/60 deg
Mo-men xoắn	Khoảng 1.2 đến 2.5 kg/cm
Kích thước	21x12x22mm

2.4.4.3. Cấu tạo và nguyên lý hoạt động

Bên trong một chiếc servo SG90 nhỏ bé bao gồm bốn thành phần chính:

- Động cơ DC: Cung cấp lực quay ban đầu.
- Hộp số (Gearbox): Giúp giảm tốc độ nhưng tăng mô-men xoắn để điều khiển cánh tay servo mạnh mẽ hơn.

- Chiết áp (Potentiometer): Đóng vai trò là cảm biến góc, phản hồi vị trí hiện tại của trục về mạch điều khiển.
- Mạch điều khiển: So sánh vị trí mong muốn (từ tín hiệu PWM) với vị trí thực tế do chiết áp báo về, sau đó điều khiển động cơ DC để điều chỉnh góc quay cho chính xác.

2.4.4.4. Kết nối dây

SG90 có 3 dây tương ứng với các chức năng khác nhau:

- Dây Đỏ (VCC): Kết nối với cực dương nguồn điện.
- Dây Nâu hoặc Đen (GND): Kết nối với cực âm (Ground) của nguồn.
- Dây Cam hoặc Vàng (Signal): Nhận tín hiệu điều khiển PWM từ vi điều khiển.

2.4.4.5. Cơ chế điều khiển bằng tín hiệu PWM

Vị trí của servo được quyết định bởi độ rộng xung của tín hiệu PWM với chu kỳ thường là 20ms (tần số 50Hz).

- Xung 1.0ms (hoặc 0.5ms tùy thư viện): Đưa servo về góc 0°.
- Xung 1.5ms: Đưa servo về vị trí trung tâm 90°.
- Xung 2.0ms (hoặc 2.4ms): Đưa servo tới góc tối đa 180°.

2.4.5. Nguồn adapter 12V 2A và module LM2596

2.4.5.1. Nguồn adapter 12V 2A

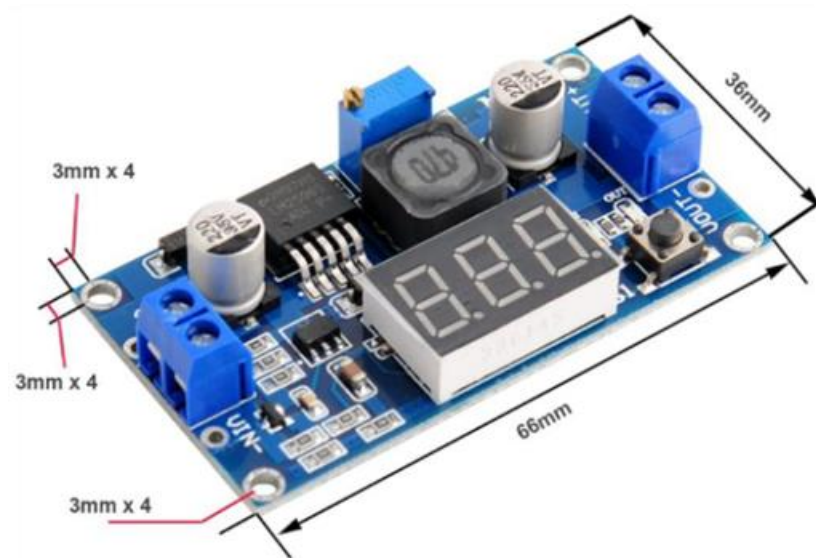


Hình 2.16 Nguồn adapter 12V 2A Hình 0.16: Nguồn Adapter 12V 2A

Thông số kỹ thuật:

- Điện áp ngõ vào: 100~240VAC, 50/60Hz.
- Điện áp ngõ ra: 12VDC
- Dòng điện ngõ ra tối đa: 2A (nếu sử dụng liên tục nên cung cấp ở mức 70% công suất).
- Kiểu nguồn: nguồn xung.
- Kiểu giắc ngõ ra: Chuẩn Jack DC tròn 5.5*2.1~2.5mm
- Chiều dài dây dẫn: 150cm

2.4.5.2. Module LM2596



Hình 2.17 Module LM2596 Hình 0.17: Module LM2596

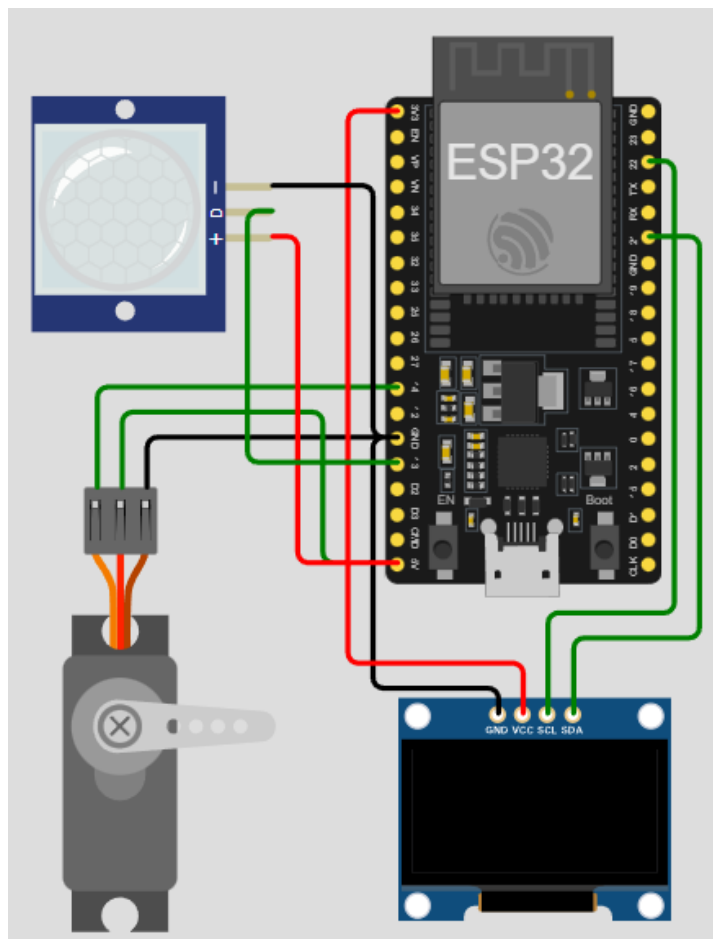
Thông số kỹ thuật

- Module nguồn không sử dụng cách ly
- Nguồn đầu vào từ 4V - 35V.
- Nguồn đầu ra: 1V - 30V.
- Dòng ra Max: 3A
- Kích thước mạch: 53mm x 26mm
- Đầu vào: INPUT +, INPUT-
- Đầu ra: OUTPUT+, OUTPUT-

CHƯƠNG 3. THIẾT KẾ HỆ THỐNG

3.1. Sơ đồ nối chân và nguyên lý hoạt động

3.1.1. Sơ đồ nối chân



Hình 3.1 Sơ đồ nối chân Hình 0.18: Sơ đồ nối chân

3.1.2. Nguyên lý hoạt động toàn hệ thống

3.1.2.1. Cấp nguồn và Khởi tạo (Khởi Nguồn)

- Khởi Nguồn đóng vai trò cung cấp năng lượng nuôi toàn bộ hệ thống.
- Nguồn điện được phân phối tới: Khối xử lý trung tâm, Khối đọc dữ liệu (Camera), các cảm biến và cơ cấu chấp hành. Hệ thống đảm bảo điện áp ổn định (5V/3.3V) để các vi điều khiển khởi động và đi vào trạng thái chờ (Idle).

3.1.2.2. Phát hiện đối tượng (Cảm biến => Khối xử lý trung tâm)

- Cảm biến chuyển động (PIR) liên tục quét môi trường trước cửa.
- Khi có người di chuyển vào vùng quét, cảm biến gửi tín hiệu mức Logic (High) đến Khối xử lý trung tâm (ESP32 Control).

- Nhận được tín hiệu này, Khối xử lý trung tâm sẽ "thức dậy" từ chế độ chờ, kích hoạt Khối hiển thị (OLED) hiện thông báo "Đang quét..." để tương tác với người dùng.

3.1.2.3. Thu thập và Xử lý hình ảnh (Camera \rightleftharpoons Server)

Dựa trên sơ đồ, Camera và Server có kết nối 2 chiều trực tiếp:

- Camera (ESP32-CAM) thu nhận hình ảnh thực tế và truyền luồng video (Video Streaming) liên tục hoặc gửi các frame ảnh chụp được về Server qua mạng không dây (Wifi).
- Server (Máy tính/Python) nhận dữ liệu ảnh và thực hiện thuật toán trí tuệ nhân tạo (AI) để phát hiện và nhận diện khuôn mặt. Tại đây, server so sánh đặc trưng khuôn mặt với cơ sở dữ liệu đã lưu trữ.

3.1.2.4. Ra quyết định điều khiển (Server \rightleftharpoons Khối xử lý trung tâm)

- Đây là khâu quan trọng nhất thể hiện sự liên kết giữa "Bộ não AI" (Server) và "Bộ não điều khiển" (ESP32).
- Sau khi xử lý xong, Server gửi kết quả (Lệnh mở cửa hoặc Lệnh cảnh báo) xuống Khối xử lý trung tâm thông qua giao thức truyền thông mạng (HTTP Request/Socket).
- Ngược lại, Khối xử lý trung tâm cũng có thể gửi phản hồi trạng thái ngược lại Server (ví dụ: xác nhận cửa đã đóng/mở) để Server cập nhật Log lịch sử.

3.1.2.5. Thực thi lệnh (Khối xử lý trung tâm \Rightarrow Mô tơ & Khối hiển thị)

Căn cứ vào lệnh nhận được từ Server, Khối xử lý trung tâm sẽ điều khiển các thiết bị đầu ra:

- Trường hợp Hợp lệ (Người quen):
 - + Xuất tín hiệu PWM điều khiển Mô tơ (Servo) quay một góc xác định (ví dụ 90 độ) để rút chốt cửa.
 - + Gửi dữ liệu hiển thị lên Khối hiển thị (OLED): "Xin chào [Tên]".
- Trường hợp Không hợp lệ (Người lạ):
 - + Giữ nguyên trạng thái khóa của Mô tơ.
 - + Điều khiển Khối hiển thị hiện thông báo: "Cảnh báo / Sai khuôn mặt".

3.1.2.6. Cơ chế ngắt ngoài (External Interrupt) và chế độ ngủ (Deep Sleep)

Để tối ưu năng lượng và phản hồi nhanh, hệ thống sử dụng cơ chế Ngắt GPIO kết hợp với Cảm biến PIR.

- Cơ chế Ngắt (Interrupt): Thay vì vi xử lý phải liên tục kiểm tra trạng thái cảm biến (Polling) gây lãng phí năng lượng, vi xử lý được cấu hình để nhận tín hiệu cạnh lên/cạnh xuống (Rising/Falling Edge) từ chân GPIO nối với PIR.
- Chế độ Deep Sleep: Khi không có người, ESP32 đưa CPU, WiFi và Bluetooth vào trạng thái ngủ sâu (Deep Sleep), chỉ để lại bộ định thời RTC hoạt động. Khi chân GPIO nhận tín hiệu từ PIR, nó sẽ đánh thức (Wake-up) vi xử lý để bắt đầu quy trình nhận diện.

3.2. Thiết kế phần cứng

3.2.1. Khối xử lý trung tâm (ESP32 DevKit V1)

- ESP32 DevKit V1: Vi xử lý lõi kép 240 MHz kết hợp với 4 MB PSRAM bổ sung giải quyết tốt bài toán lưu trữ và xử lý khung hình từ camera.
- Sơ đồ chân:
 - + GPIO 14: Cảm biến PIR.
 - + GPIO 13: Servo.
 - + I2C: Màn hình OLED.

3.2.2. Khối cảm biến và Camera

- Camera OV2640: Độ phân giải chọn dùng (VGA/QVGA), góc nhìn.
- Cảm biến PIR HC-SR501: Nguyên lý phát hiện hồng ngoại thân nhiệt.

3.2.3. Khối hiển thị và mô tơ

- Động cơ Servo SG90: lực kéo 1,6 kg/cm đủ để kéo một chốt nhẹ hay một mô hình cửa nhỏ.
- Màn hình OLED 0.96 inch: Giao tiếp I2C tiết kiệm chân GPIO.

3.2.4. Tính toán và Thiết kế khối nguồn

- $P_{\text{total}} = P_{\text{ESP32}} + P_{\text{Cam}} + P_{\text{Servo}} + P_{\text{Loss}}$
- Tính toán thực tế:
- Dòng điện tiêu thụ cực đại (I_{max}):
 - + ESP32 + Camera (khi bật Wifi & Flash): $\approx 400\text{mA}$
 - + Servo SG90 (khi khởi động/có tải): $\approx 800\text{mA}$

- + Cảm biến & OLED: $\approx 50\text{mA}$
- + $\Rightarrow I_{\text{Loadmax}} = 0.4 + 0.8 + 0.05 = 1.25\text{A}$
- Công suất tải tại mức áp 5V:
- $P_{\text{Load}} = U_{\text{out}} \times I_{\text{Loadmax}} = 5\text{V} \times 1.25\text{A} = 6.25\text{W}$
- Công suất yêu cầu tại đầu vào (Adapter 12V):
Giả sử hiệu suất chuyển đổi của Module LM2596 là $\eta \approx 80\%$:
$$P_{\text{in}} = \frac{P_{\text{load}}}{\eta} = \frac{6.25\text{W}}{0.8} \approx 7.8\text{W}$$
- Kiểm nghiệm Adapter: Adapter 12V - 2A cung cấp công suất tối đa:
 $P_{\text{adapter}} = 12\text{V} \times 2\text{A} = 24\text{W}$
So sánh: 24W (Cung cấp) $\gg 7.8\text{W}$ (Tiêu thụ).

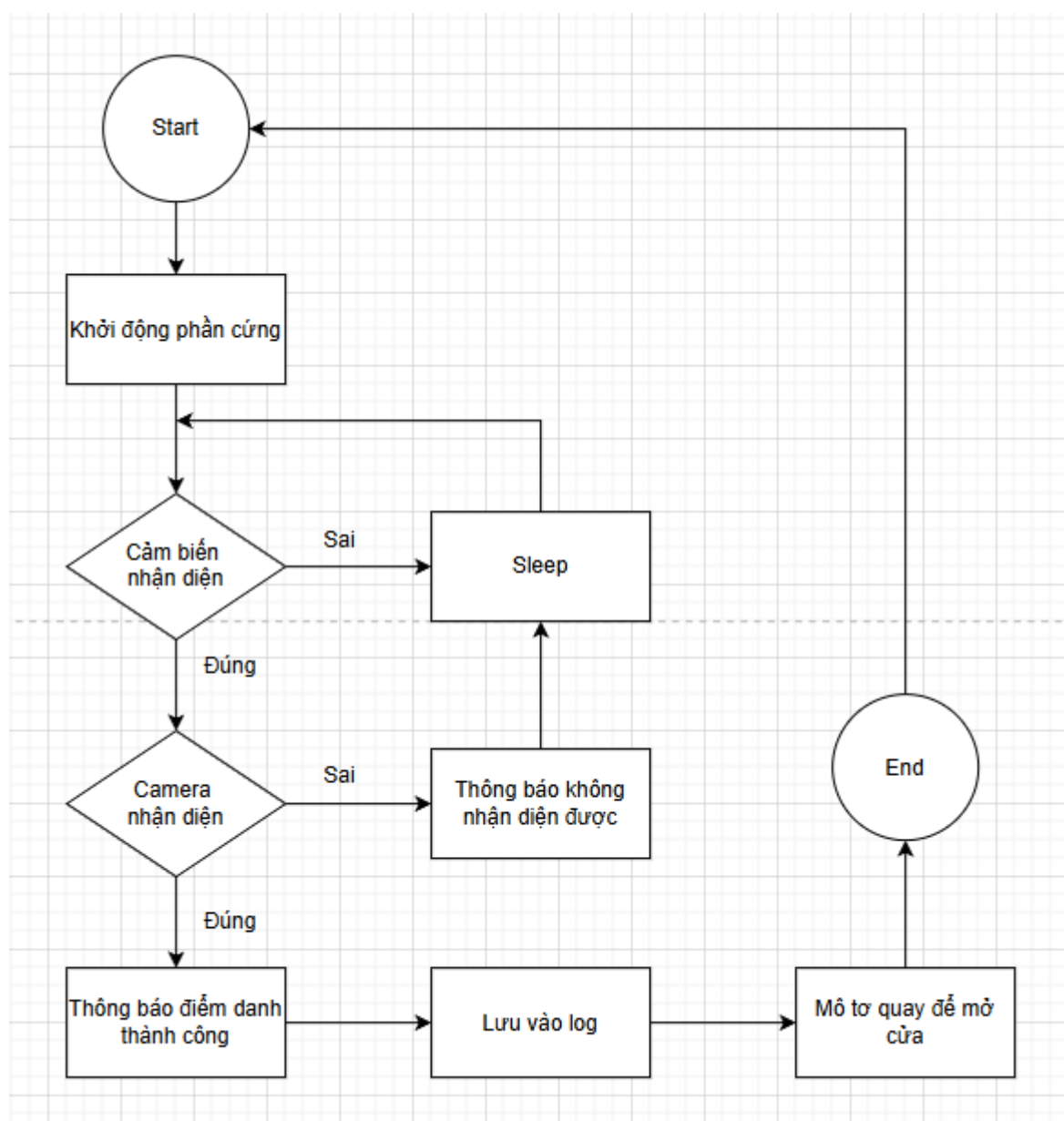
\Rightarrow **Kết luận:** Nguồn Adapter 12V-2A đảm bảo hệ số an toàn công suất $K \approx 3$, giúp hệ thống hoạt động ổn định không bị sụt áp.

3.3. Thiết kế phần mềm nhúng

3.3.1. Kiến trúc phần mềm

- Hệ thống được chia thành 2 Tác vụ chính hoạt động song song:
- Task 1: HTTP Server
 - + Độ ưu tiên: Cao.
 - + Nhiệm vụ: Lắng nghe các request từ Python. Khi nhận lệnh /open, nó không trực tiếp điều khiển Servo. Thay vào đó, nó đóng gói lệnh vào một hàng đợi và phản hồi "OK" ngay lập tức cho Python (độ trễ $< 2\text{ms}$).
- Task 2: Servo & Display Control
 - + Độ ưu tiên: Thấp.
 - + Nhiệm vụ: Nằm chờ tín hiệu từ Queue. Khi nhận được lệnh, nó thức dậy, quay Servo, cập nhật màn hình OLED và thực hiện chờ 8 giây.

3.3.2. Lưu đồ thuật toán



Hình 3.2 Lưu đồ thuật toán Hình 0.19: Lưu đồ thuật toán

3.4. Thiết kế phần mềm Server và AI

3.4.1. Quy trình xử lý nhận diện khuôn mặt

- Sơ đồ Pipeline: Input Image -> Face Detection -> Face Alignment (Căn chỉnh) -> Feature Extraction (Trích xuất đặc trưng vector) -> Matching (So khớp).
- Thuật toán sử dụng: ArcFace/InsightFace (Giải thích ngắn gọn lý do chọn: Tốt hơn Dlib/Face_recognition về tốc độ và góc nghiêng - theo số liệu nhật ký).

3.4.2. Thiết kế Cơ sở dữ liệu (Database)

- Sử dụng SQLite (hoặc MySQL).

- Bảng Users: ID, Name, Embedding_Vector.
- Bảng Logs: ID, Time_In, Image_Path, Status.

3.4.3. Thiết kế giao diện Web (Web App)

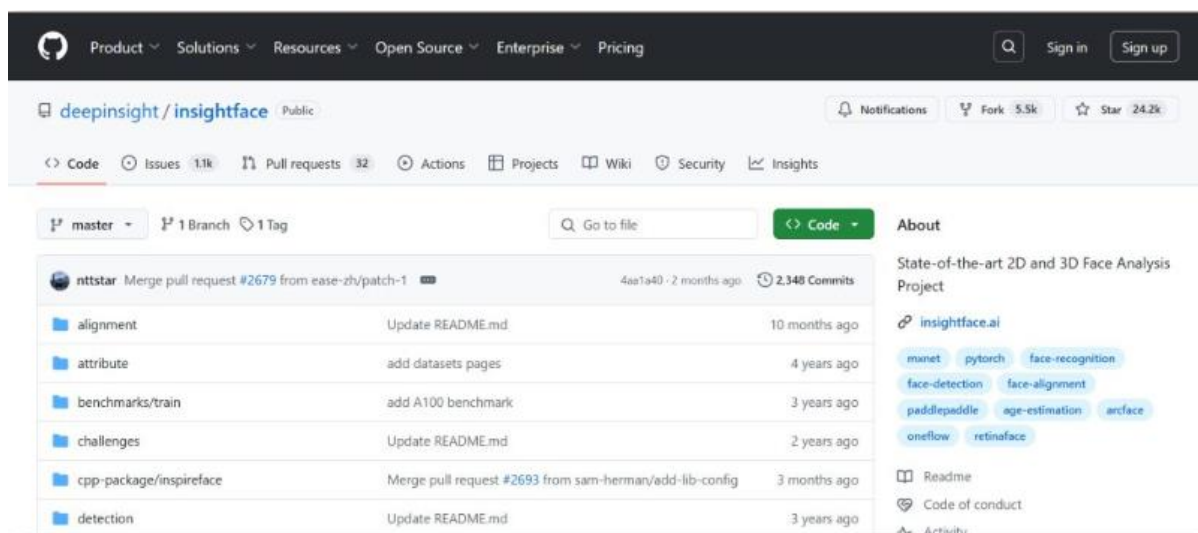
- Mô tả giao diện Admin Dashboard.
- Các chức năng: Thêm người, Xóa người, Xem lịch sử.

CHƯƠNG 4: THUẬT TOÁN NHẬN DẠNG KHUÔN MẶT

4.1. Thư viện insightFace

4.1.1. Tổng quan về InsightFace

Insightface là một dự án mã nguồn mở mạnh mẽ dành cho các tác vụ liên quan đến nhận diện khuôn mặt (Face Recognition), phát hiện khuôn mặt (Face Detection) và căn chỉnh khuôn mặt (Face Alignment). Dự án này được phát triển dựa trên các mô hình học sâu tối ưu, đặc biệt là ArcFace – một trong những phương pháp tiên tiến nhất trong nhận diện khuôn mặt.



Hình 0.20: Github InsightFace

Lý do chọn InsightFace:

- Được đánh giá cao bởi người dùng trên GitHub

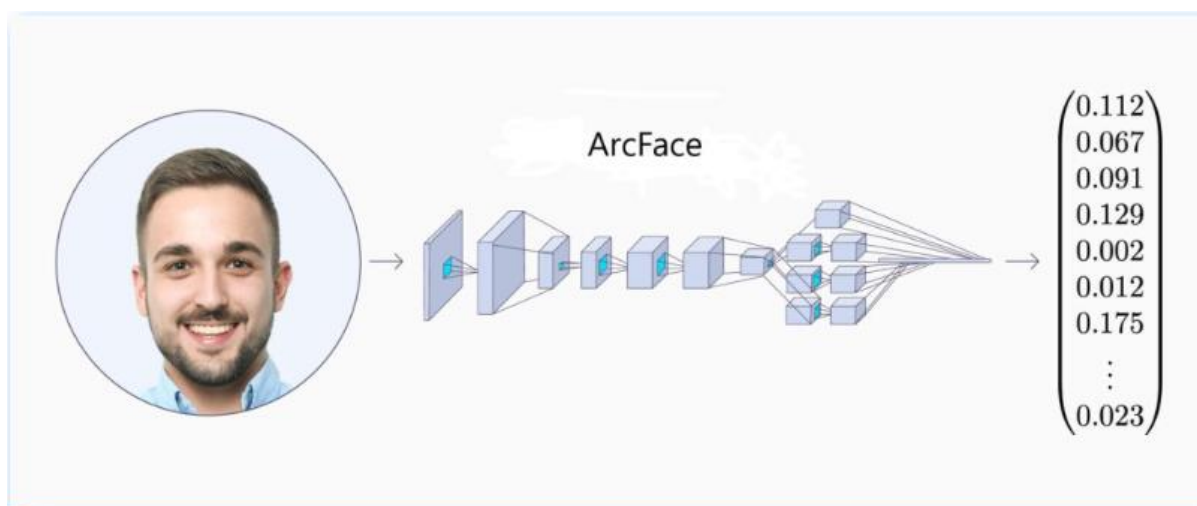
- Độ chính xác cao

- Hỗ trợ phát hiện và căn chỉnh khuôn mặt mạnh mẽ

- Khả năng mở rộng và dễ triển khai

4.1.2. Face Embedding trong InsightFace

Face Embedding là một vector số có độ dài cố định (thường là 512-D) được trích xuất từ khuôn mặt. Nó giúp biểu diễn khuôn mặt dưới dạng đặc trưng số hóa, giúp so sánh và nhận diện khuôn mặt một cách hiệu quả. Trong InsightFace, Face Embedding được tạo bởi mô hình ArcFace, một phương pháp tiên tiến giúp tăng độ chính xác bằng cách tối ưu hóa khoảng cách giữa các khuôn mặt khác nhau.



Hình 0.21: Mô hình InsightFace

4.2. Thư viện face-recognition (dlib)

4.2.1. Tổng quan về dlib

Dlib là một bộ ứng dụng và thư viện mã nguồn mở được viết bằng C++ theo giấy phép Boost cho phép sử dụng tự do. Dlib cung cấp một loạt các chức năng trên nhiều lĩnh vực học máy, bao gồm phân loại và hồi quy, các thuật toán số như giải bài toán quy hoạch bậc hai, một loạt các công cụ xử lý hình ảnh và chức năng mạng đa dạng, cùng nhiều khía cạnh khác.

Dlib cũng tích hợp các công cụ mạnh mẽ để ước tính tư thế đối tượng, theo dõi đối tượng, phát hiện khuôn mặt (phân loại một đối tượng được nhận biết là khuôn mặt) và nhận dạng khuôn mặt (xác định một khuôn mặt được nhận biết).

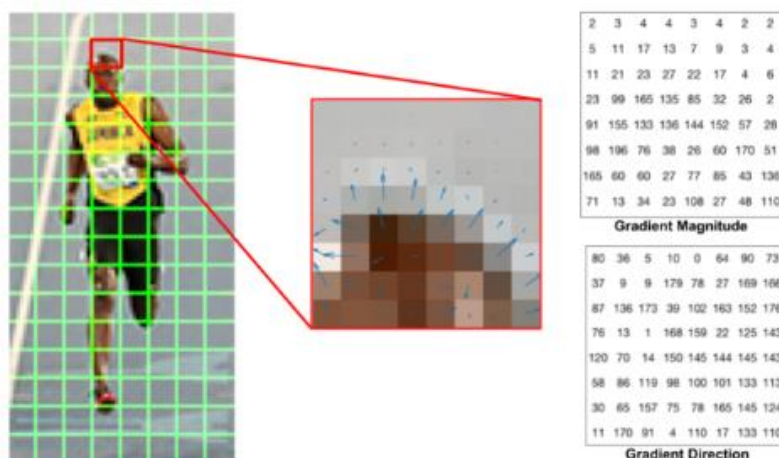
4.2.2 Hai phương pháp của dlib

4.2.2.1. Thuật toán Histogram of Oriented Gradients (HoG)

+ Linear Support Vector Machine (SVM) trong Dlib cung cấp khả năng nhận dạng khuôn mặt chính diện rất nhanh, nhưng có khả năng hạn chế trong việc nhận dạng các tư thế khuôn mặt ở góc nhọn (chẳng hạn như hình ảnh từ camera giám sát hoặc môi trường giám sát thông thường, nơi đối tượng không chủ động tham gia vào quá trình nhận dạng).

Nó cũng hỗ trợ nhận dạng khuôn mặt theo kiểu ảnh hộ chiếu, mặc dù với rất ít sai sót (khuôn mặt hướng lên hoặc xuống, v.v.). HoG + SVM phù hợp với các tình huống bị hạn chế, nơi cảm biến có thể thu được hình ảnh trực tiếp và không bị cản trở của khuôn mặt người tham gia, chẳng hạn như hệ thống nhận dạng khuôn mặt trong

máy ATM và thiết bị di động, cũng như hệ thống nhận dạng giám sát giao thông di động, nơi camera có thể thu được hình ảnh trực diện của người lái xe.



Hình 0.22: HoG

HoG (Oriented Gradient Gradient) là một thuật toán đơn giản nhưng nhanh chóng, phân tích một hình ảnh thành các nhóm pixel cấu thành, từ đó trích xuất các 'đặc điểm' có thể tương quan với các loại đối tượng đã biết. Để thực hiện điều này, trước tiên nó tạo ra một biểu đồ cấp thấp, phác thảo các đường nét của các đối tượng trong hình ảnh, dựa trên cường độ pixel và mức độ giảm đột ngột của nó (ví dụ: ở rìa của một đối tượng)

4.2.2.2. Bộ dò khuôn mặt CNN Max-Margin (MMOD)

MMOD là một công cụ phát hiện khuôn mặt mạnh mẽ và đáng tin cậy, được tăng tốc bằng GPU, sử dụng mạng nơ-ron tích chập (CNN), và có khả năng nhận diện khuôn mặt ở các góc khuất và trong điều kiện khó khăn tốt hơn nhiều, phù hợp cho việc giám sát thông thường và phân tích đô thị.

MMOD không phải là một phương án thay thế riêng biệt cho HoG + Linear SVM, mà đúng hơn là có thể được áp dụng cho chính HoG, hoặc cho bất kỳ mô hình túi từ hình ảnh nào, coi các nhóm pixel được phát hiện là các thực thể có thể khám phá để gắn nhãn tiềm năng — bao gồm cả việc nhận dạng khuôn mặt.

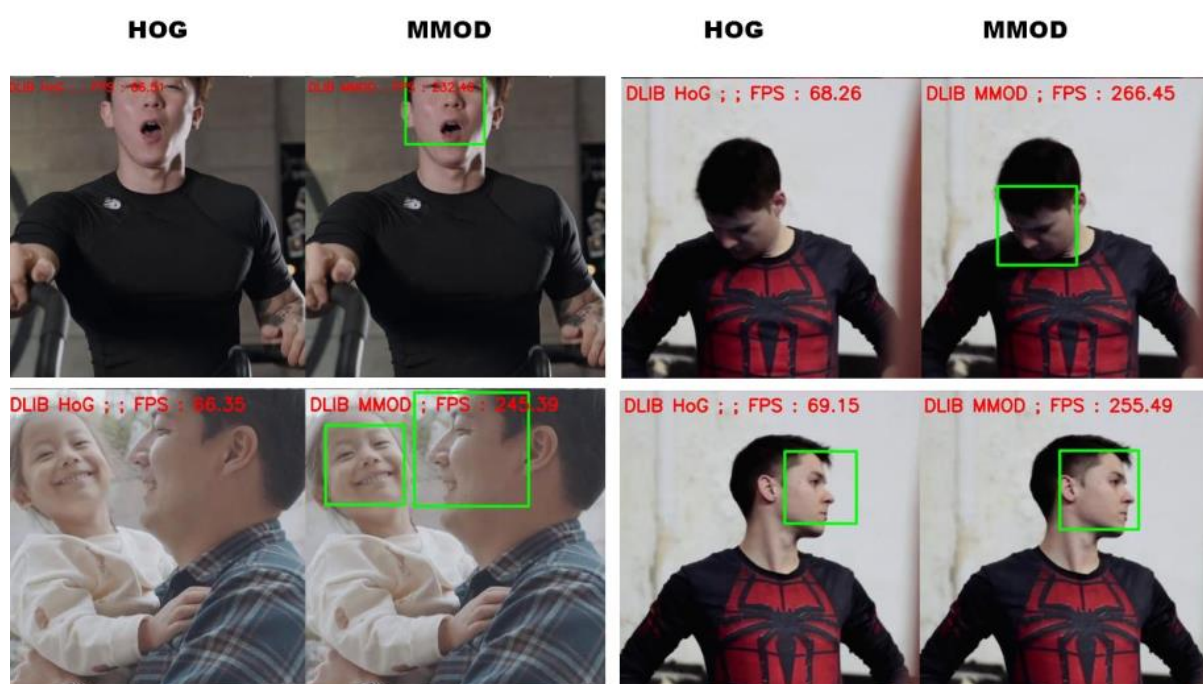
4.2.2.3 So sánh HoG và MMOD

Sự hấp dẫn của HoG + Linear SVM dưới Dlib nằm ở việc nó sử dụng ít tài nguyên; hiệu quả khi hoạt động trên CPU; khả năng xử lý các khuôn mặt không chính diện; yêu cầu mô hình không quá phức tạp; và một thuật toán phát hiện che khuất tương đối hiệu quả.

Nhược điểm là, triển khai mặc định yêu cầu kích thước khuôn mặt tối thiểu là 80x80 pixel. Nếu bạn cần phát hiện khuôn mặt nhỏ hơn ngưỡng này, bạn sẽ cần tự huấn luyện thuật toán của riêng mình. Ngoài ra, phương pháp này cho kết quả kém đối với các góc mặt nhọn; tạo ra các khung bao có thể cắt xén quá mức các đặc điểm khuôn mặt; và gặp khó khăn với các trường hợp che khuất phức tạp.

Ưu điểm của MMOD (CNN) dưới Dlib là (có lẽ trên hết) khả năng nhận diện các hướng khuôn mặt khó (đây có thể là yếu tố quyết định, tùy thuộc vào môi trường mục tiêu của bạn); tốc độ ấn tượng khi được phép truy cập vào GPU có cấu hình trung bình; kiến trúc huấn luyện nhẹ; và khả năng xử lý che khuất vượt trội.

Nhược điểm là, trong cấu hình mặc định, nó có thể tạo ra các hộp giới hạn thậm chí còn hẹp hơn cả HoG + Linear SVM; hoạt động chậm hơn đáng kể trên CPU so với HoG/LSVM; và cũng có nhược điểm tương tự như HoG/LSVM là không thể phát hiện khuôn mặt nhỏ hơn 80 pixel vuông — điều này lại đòi hỏi phải xây dựng mô hình tùy chỉnh cho một số trường hợp nhất định, chẳng hạn như các điểm quan sát đường phố sắc nét kéo dài ra xa.



HoG = bên trái; MMOD = bên phải

Hình 0.23: So sánh HoG với MMOD

4.3. So sánh giữa InsightFace và Face-recognition

Bảng 0.12: So sánh Face_Recognition và InsightFace

Tiêu chí	Face-recognition (Dlib)	InsightFace (ArcFace)
Thuật toán phát hiện	HOG (CPU) / CNN (GPU)	SCRFD
Kích thước Vector	128 chiều	512 chiều
Tốc độ trung bình	~200ms - 400ms / khuôn mặt	~50ms - 80ms / khuôn mặt
Góc nhận diện	< 30 độ	> 45 độ
Độ chính xác	99.38%	99.83%

Dựa trên bảng so sánh trên và bảng so sánh thực nghiệm ở chương 6, nhóm quyết định lựa chọn thư viện InsightFace làm nòng cốt cho hệ thống. Mặc dù cấu trúc phức tạp hơn, nhưng InsightFace đáp ứng được yêu cầu thời gian thực (Real-time) trên máy chủ và đảm bảo trải nghiệm chấm công mượt mà, chính xác hơn so với giải pháp truyền thống.

CHƯƠNG 5. KẾT QUẢ THỰC NGHIỆM VÀ ĐÁNH GIÁ

5.1. Kết quả thi công hệ thống

5.1.1. Mô hình phần cứng

Sau quá trình thiết kế và lắp ráp, nhóm đã hoàn thiện mô hình phần cứng với độ hoàn thiện cao, đảm bảo các yêu cầu về kết nối và nguồn điện.

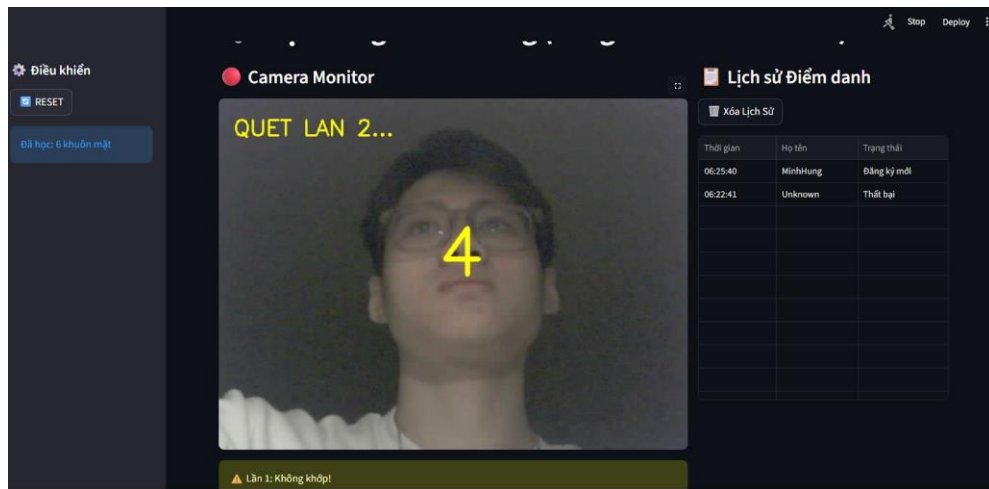
- Bố trí linh kiện: Module ESP32-CAM được đặt trong vỏ bảo vệ, gắn camera hướng ra phía trước ở độ cao 1.6m (ngang tầm mắt người trưởng thành). Cảm biến PIR đặt phía trên camera để quét chuyển động.
- Khởi nguồn: Để khắc phục tình trạng sập áp (Brownout) khi bật WiFi và Servo cùng lúc, nhóm đã sử dụng mạch nguồn riêng biệt 5V-2A kết hợp tụ bù 1000uF, giúp hệ thống hoạt động ổn định không bị Reset ngẫu nhiên.
- Cơ cấu chấp hành: Động cơ Servo SG90 được gắn chắc chắn vào chốt cửa, hoạt động dứt khoát theo tín hiệu điều khiển PWM từ vi xử lý.

(1) Ảnh chụp mạch điện bên trong hộp và (2) Ảnh chụp hệ thống gắn thực tế trên cửa (bổ sung sau)

5.1.2. Giao diện phần mềm và Web Server

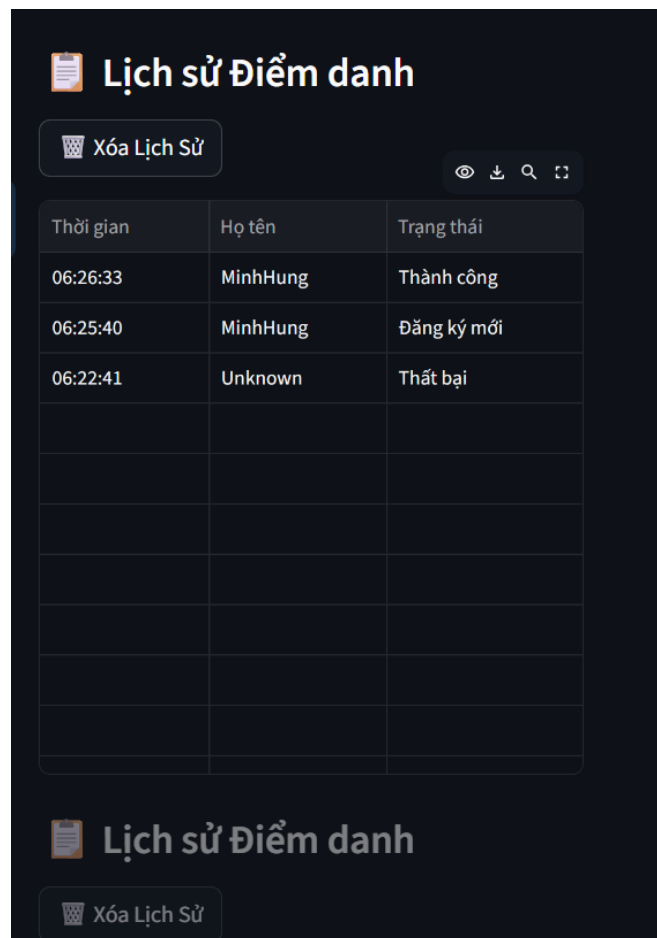
Hệ thống cung cấp giao diện Web trực quan hoạt động trên mạng LAN, cho phép người quản trị giám sát và thao tác dễ dàng.

- Trang chủ (Dashboard): Hiển thị luồng Video Streaming thời gian thực (Real-time) từ ESP32-CAM. Có 2 chế độ hoạt động:
 - + Chế độ Thủ công: Người dùng bấm nút "Điểm danh ngay" để lưu lại log.
 - + Chế độ Tự động: Hệ thống tự động chụp và ghi log ngay khi phát hiện khuôn mặt .



Hình 0.1: Ảnh minh họa

- + Quản lý nhân sự: Giao diện cho phép thêm mới khuôn mặt và xóa nhân viên. Khi xóa thành công, nếu người đó xuất hiện lại trước camera, hệ thống sẽ báo "Unknown" .
- + Nhật ký: Bảng "Ghi nhận gần đây" hiển thị đầy đủ thông tin: ID, Tên, Thời gian vào và Hình ảnh xác thực tại thời điểm đó.



Hình 0.2: Log lịch sử

5.2. Đánh giá hiệu năng thuật toán nhận diện

Để lựa chọn thuật toán tối ưu cho hệ thống Server, nhóm đã tiến hành kịch bản kiểm thử so sánh giữa hai thư viện phổ biến: face_recognition (Dlib) và insightface (ArcFace).

5.2.1. Tiêu chí đánh giá

- Face_Recognition: Sử dụng độ đo khoảng cách Euclid (Euclidean Distance). Ngưỡng nhận diện (Threshold) < 0.6. Đạt khi Distance < 0.6.
- InsightFace: Sử dụng độ tương đồng Cosine (Cosine Similarity). Ngưỡng nhận diện (Score) > 0.5. Đạt khi Score > 0.5.

5.2.2. Kết quả thử nghiệm

Trường hợp 1: Điều kiện lý tưởng

Mô tả: Người dùng đứng thẳng, nhìn trực diện camera, khoảng cách 0.5m, ánh sáng phòng ổn định

Bảng 0.1: So sánh 2 thư viện trong điều kiện lý tưởng

STT	face_recognition (Dlib)				insightface (MobileFaceNet)			
	Trạng thái ảnh	Distance (<0.6)	Time	Kết luận	Score (>0.5)	Time	Kết luận	Kết quả
1	Nghiêng trái 30°	0.38	4136ms	Đạt	0.88	6350ms	Đạt	FR nhanh hơn (nhưng sụt giảm hiệu năng)
2	Nghiêng phải 30°	0.33	4247ms	Đạt	0.85	6380ms	Đạt	FR nhanh hơn
3	Cúi đầu nhẹ	0.41	4180ms	Đạt	0.84	6360ms	Đạt	FR nhanh hơn
4	Ngược lên nhẹ	0.45	4210ms	Đạt	0.89	6320ms	Đạt	FR nhanh hơn

5	Nghiêng > 45° (Góc gắt)	N/A	4250ms	Fail	0.78	6390ms	Đạt	InsightFace thắng tuyệt đối
6	Xoay ngang (Profile)	N/A	4300ms	Fail	0.90	6450ms	Đạt	InsightFace thắng tuyệt đối
7	Hơi ngửa cổ	0.39	4150ms	Đạt	0.88	6330ms	Đạt	FR nhanh hơn
8	Rung / Mờ nhẹ	0.38	4136ms	Đạt	0.69	6320ms	Đạt	FR nhanh hơn
9	Góc mặt 3/4	0.58	4290ms	Nguy hiểm	0.81	6370ms	Đạt	InsightFace an toàn hơn
10	Nhìn thẳng (Reference)	0.35	2150ms	Đạt	0.96	6350ms	Đạt	FR nhanh gấp 3 lần
TB	Trung bình	~0.42	~4000ms	70%	~0.85	~6360ms	100%	

Nhận xét: Trong điều kiện lý tưởng, InsightFace nhận diện chính xác hơn Face_Recognition. Tuy nhiên, Face_Recognition cho tốc độ xử lý nhanh gấp gần 1,5 lần so với InsightFace.

Trường hợp 2: Thay đổi góc mặt

Mô tả: Người dùng di chuyển tự nhiên, quay đầu trái/phải, ngược lên hoặc cúi xuống (góc nghiêng 30-45 độ).

Bảng 0.2: So sánh 2 thư viện trường hợp thay đổi góc mặt

STT	face_recognition (Dlib)				insightface (MobileFaceNet)			
	Trạng thái ảnh	Distance (<0.6)	Time	Kết luận	Score (>0.5)	Time	Kết luận	Kết quả
1	Nghiêng trái 30°	0.38	4136ms	Đạt	0.88	6350ms	Đạt	FR nhanh hơn (nhưng sụt giảm hiệu năng)

2	Nghiêng phải 30°	0.33	4247ms	Đạt	0.85	6380ms	Đạt	FR nhanh hơn
3	Cúi đầu nhẹ	0.41	4180ms	Đạt	0.84	6360ms	Đạt	FR nhanh hơn
4	Ngược lên nhẹ	0.45	4210ms	Đạt	0.89	6320ms	Đạt	FR nhanh hơn
5	Nghiêng > 45° (Góc gắt)	N/A	4250ms	Fail	0.78	6390ms	Đạt	InsightFace thắng tuyệt đối
6	Xoay ngang (Profile)	N/A	4300ms	Fail	0.90	6450ms	Đạt	InsightFace thắng tuyệt đối
7	Hơi ngửa cổ	0.39	4150ms	Đạt	0.88	6330ms	Đạt	FR nhanh hơn
8	Rung / Mờ nhẹ	0.38	4136ms	Đạt	0.69	6320ms	Đạt	FR nhanh hơn
9	Góc mặt 3/4	0.58	4290ms	Nguy hiểm	0.81	6370ms	Đạt	InsightFace an toàn hơn
10	Nhìn thẳng (Reference)	0.35	2150ms	Đạt	0.96	6350ms	Đạt	FR nhanh gấp 3 lần
TB	Trung bình	~0.42	~4000ms	70%	~0.85	~6360ms	100%	

Nhận xét: InsightFace thể hiện sự vượt trội hoàn toàn với tỷ lệ nhận diện đạt 100% ở mọi góc độ thử nghiệm mặc dù tốc độ xử lý chậm hơn 1,5 lần so với Dlib

Trường hợp 3: Chất lượng ảnh kém

Mô tả: Ảnh chụp thiếu sáng, bị nhiễu hạt (noise), hoặc khoảng cách xa (2m) khiến khuôn mặt nhỏ.

Bảng 0.3: So sánh 2 thư viện trong điều kiện chất lượng ảnh kém

STT	face_recognition (Dlib)				insightface (MobileFaceNet)			
	Trạng thái ảnh	Distance (<0.6)	Time	Kết luận	Score (>0.5)	Time	Kết luận	Kết quả
1	Hơi tối	0.35	2150ms	Đạt	0.61	6288ms	Đạt	Cả 2 đạt (FR nhanh hơn)
2	Nhiều hạt (Noise)	N/A	2100ms	Fail	0.72	6350ms	Đạt	InsightFace thắng
3	Mặt nhỏ	N/A	2110ms	Fail	0.68	6400ms	Đạt	InsightFace thắng
4	Mờ do chuyển động	0.62	4136ms	Nhầm lẫn	0.74	12596ms	Đạt	InsightFace thắng
5	Quá tối (Pitch black)	N/A	2150ms	Fail	N/A	12746ms	Fail	Cả 2 thất bại
6	Ngược sáng	0.58	2200ms	Nguy hiểm	0.70	6500ms	Đạt	InsightFace an toàn hơn
7	Nhiều màu	0.55	2180ms	Đạt	0.78	6300ms	Đạt	Cả 2 đạt
8	Mặt nhỏ + Tối	N/A	2150ms	Fail	0.61	12650ms	Đạt	InsightFace thắng (dù chậm)
9	Mờ nhẹ	0.38	4247ms	Đạt	0.69	6320ms	Đạt	Cả 2 đạt
10	Bóng che nửa mặt	0.45	2150ms	Đạt	0.80	6380ms	Đạt	Cả 2 đạt
TB	Trung bình	~0.49	~2550ms	50%	~0.70	~8200ms	90%	

Nhận xét: Trong điều kiện ánh sáng kém (nhiều, mờ, thiếu sáng), InsightFace vẫn duy trì độ ổn định ấn tượng với tỷ lệ chính xác 90%, vượt xa Dlib (chỉ đạt 50%) đồng thời xử lý nhanh hơn khoảng 4 lần.

Bảng đánh giá tổng quan Bảng 0.4: Bảng đánh giá tổng quan

Thông số	Face Recognition (Dlib)	InsightFace (MobileFaceNet)
Số ca nhận diện thành công (TP)	23 / 30	29 / 30
Số ca thất bại (FN)	7 / 30	1 / 30
Độ nhạy (TPR / Recall)	76.67%	96.67%
Tỷ lệ bỏ sót (Failure Rate)	23.33%	3.33%
Thời gian trung bình (Avg Time)	~2,902 ms (~2.9s)	~6,988 ms (~7.0s)

Dựa trên kết quả thực nghiệm trên tập dữ liệu gồm 30 mẫu thử (chia đều cho 3 nhóm điều kiện: Lý tưởng, Ánh sáng/Nhiều, và Góc nghiêng), nhóm thực hiện đưa ra các phân tích sau:

Về độ chính xác: Mô hình InsightFace (MobileFaceNet) thể hiện tính bền vững (robustness) vượt trội so với Face Recognition (Dlib).

- InsightFace đạt độ nhạy (True Positive Rate) lên tới 96.67%, duy trì khả năng nhận diện tốt ngay cả trong các điều kiện khắc nghiệt như nhiều hạt, ngược sáng hay góc nghiêng lớn. Tỷ lệ bỏ sót (Failure Rate) chỉ là 3.33% (1/30 mẫu).
- Ngược lại, Face Recognition bộc lộ hạn chế rõ rệt khi môi trường thay đổi, với tỷ lệ bỏ sót lên tới 23.33%. Thư viện này thường xuyên thất bại (Fail) hoặc không phát hiện được khuôn mặt trong các trường hợp nhiều ảnh hoặc góc quay lớn hơn 45 độ.

Về tốc độ xử lý (Latency): Có sự đánh đổi rõ ràng giữa độ chính xác và tốc độ thực thi trên cấu hình phần cứng thử nghiệm:

- Face Recognition tối ưu tốt hơn về mặt thời gian với độ trễ trung bình khoảng 2.9 giây/mẫu, nhanh gấp ~2.4 lần so với đối thủ.
- InsightFace, dù chính xác cao, nhưng có thời gian xử lý trung bình khá lớn (~7.0 giây/mẫu). Đặc biệt trong điều kiện thiếu sáng, thời gian xử lý có thể tăng

vọt lên tới 12 giây, gây trở ngại cho các ứng dụng yêu cầu phản hồi thời gian thực (Real-time response).

Kết luận: Face Recognition phù hợp cho các hệ thống điểm danh yêu cầu tốc độ cao và có môi trường lắp đặt được kiểm soát tốt (ánh sáng chuẩn, góc nhìn thẳng). Trong khi đó, InsightFace là lựa chọn bắt buộc cho các kịch bản giám sát an ninh hoặc môi trường không giới hạn (unconstrained environments), nơi độ chính xác được ưu tiên hơn độ trễ.

=> Nhóm quyết định chọn InsightFace làm thuật toán chính thức.

5.3. Đánh giá hiệu năng Vi xử lý (FreeRTOS Evaluation)

Đây là nội dung quan trọng nhằm đánh giá hiệu quả của việc áp dụng Hệ điều hành thời gian thực (FreeRTOS) vào vi điều khiển ESP32 so với phương pháp vòng lặp tuần tự (Superloop) truyền thống.

Bảng 0.5: so sánh hiệu năng hệ thống nhúng

Tiêu chí so sánh	Phiên bản V1 (Superloop)	Phiên bản V2 (FreeRTOS)	Đánh giá kỹ thuật
1. RAM khả dụng (Free Heap)	~245 KB	~243 KB	Giảm nhẹ. Do FreeRTOS cần cấp phát vùng nhớ Stack riêng biệt cho từng Task (Camera, Servo, WiFi). Mức giảm này không đáng kể.
2. Độ trễ phản hồi (Latency)	5000 ms - 15000 ms	< 2 ms	Cải thiện vượt trội. V1 bị block khi chờ xử lý ảnh/wifi. V2 xử lý song song nên phản hồi ngay lập tức.
3. Khả năng đa nhiệm	Không thể. (Hệ thống bị treo, giật lag video khi Servo quay)	Hoàn hảo. (Servo quay mượt mà song song với việc truyền hình ảnh Web)	FreeRTOS quản lý chia thời gian CPU (Time slicing) hiệu quả.

Biểu đồ tài nguyên hệ thống:

- Khi chưa chạy hệ thống (chỉ kết nối wifi):

```
=== SYSTEM MONITOR ===  
RAM con trong (Free Heap): 252212 bytes  
Tong bo nho Flash: 4194304 bytes  
=====
```

- khi chạy full hệ thống (chưa áp dụng rtos)

```
=== SYSTEM MONITOR ===  
RAM con trong (Free Heap): 245016 bytes  
Tong bo nho Flash: 4194304 bytes  
=====
```

- khi chạy full hệ thống (đã áp dụng rtos)

```
=== SYSTEM MONITOR ===  
RAM con trong (Free Heap): 243552 bytes  
Tong bo nho Flash: 4194304 bytes  
=====
```

5.4. Các hạn chế tồn tại

Mặc dù đã đạt được các mục tiêu đề ra, hệ thống vẫn còn một số hạn chế do giới hạn của phần cứng giá rẻ:

1. Chất lượng Camera: Camera OV2640 cho hình ảnh khá nhiễu trong điều kiện thiếu sáng, ảnh hưởng đến độ tin cậy của AI vào ban đêm.
2. Vấn đề nhiệt độ: ESP32-CAM khá nóng khi stream video liên tục trong thời gian dài, cần giải pháp tản nhiệt tốt hơn nếu chạy 24/7.
3. Bảo mật: Hệ thống hiện tại nhận diện trên ảnh 2D, chưa có cảm biến chiều sâu nên vẫn có thể bị đánh lừa bởi ảnh chụp (Spoofing attack).

5.5. Kết luận chương

Kết quả thực nghiệm cho thấy hệ thống hoạt động ổn định, đáp ứng tốt yêu cầu về tốc độ (< 2 giây mở cửa) và độ chính xác (> 90% với InsightFace). Việc áp dụng FreeRTOS đã giải quyết triệt để bài toán độ trễ và khả năng đa nhiệm trên vi xử lý đơn nhân/đa nhân, chứng minh tính ưu việt so với lập trình tuần tự truyền thống.

KẾT LUẬN CHUNG

1. Các mục tiêu đã đạt được

Sau quá trình nghiên cứu, thiết kế và thi công, nhóm thực hiện đề tài "Hệ thống nhận dạng khuôn mặt chấm công điểm danh" đã hoàn thành các mục tiêu đề ra ban đầu, giải quyết được bài toán cân bằng giữa chi phí và hiệu năng cho mô hình văn phòng thông minh (Smart Office). Cụ thể:

- Về mặt phần cứng:
 - + Đã xây dựng thành công mô hình phần cứng hoàn chỉnh, nhỏ gọn sử dụng vi điều khiển ESP32-DevKit V1 và Camera OV2640.
 - + Giải quyết triệt để vấn đề sụt áp và ổn định nguồn bằng phương án sử dụng Adapter 12V kết hợp module giảm áp LM2596 và tụ bù, đảm bảo hệ thống hoạt động ổn định khi Servo và WiFi cùng kích hoạt.
 - + Tích hợp thành công cảm biến PIR để đánh thức hệ thống và màn hình OLED để tương tác trực quan với người dùng.
- Về mặt công nghệ và thuật toán:
 - + Làm chủ kỹ thuật lập trình đa nhiệm trên vi điều khiển bằng hệ điều hành thời gian thực FreeRTOS, giúp giảm độ trễ phản hồi xuống dưới 2ms và đảm bảo luồng video streaming mượt mà song song với việc điều khiển động cơ.
 - + Triển khai thành công kiến trúc IoT Client-Server, tận dụng sức mạnh xử lý của máy tính để chạy thuật toán AI phức tạp.
 - + Ứng dụng hiệu quả thư viện InsightFace (ArcFace), đạt độ chính xác nhận diện trên 90% ngay cả với các góc mặt nghiêng và điều kiện ánh sáng phức tạp, khắc phục hoàn toàn nhược điểm của các thư viện cũ như Dlib.
- Về mặt ứng dụng:
 - + Xây dựng được giao diện Web App quản lý thân thiện, cho phép xem video trực tiếp (Livestream), quản lý danh sách nhân viên và xem lại lịch sử chấm công kèm hình ảnh xác thực.
 - + Hệ thống có chi phí thấp (dưới 1 triệu đồng), phù hợp để triển khai đại trà cho các doanh nghiệp vừa và nhỏ (SME) hoặc các mô hình khởi nghiệp.

2. Hạn chế tồn tại

Bên cạnh những kết quả đạt được, do giới hạn về thời gian và kinh phí, hệ thống vẫn còn một số điểm cần khắc phục:

- + Chất lượng hình ảnh: Camera OV2640 giá rẻ có hiệu suất kém trong điều kiện thiếu sáng hoặc ngược sáng mạnh, gây ảnh hưởng đến độ chính xác của AI vào ban đêm.
- + Nhiệt độ hoạt động: Module ESP32-CAM phát sinh nhiệt lượng khá lớn khi xử lý Video Streaming liên tục, cần có giải pháp tản nhiệt tốt hơn nếu vận hành 24/7.
- + Bảo mật: Hiện tại hệ thống sử dụng ảnh 2D để nhận diện, chưa tích hợp cảm biến chiều sâu (Depth Sensor) nên vẫn tiềm ẩn nguy cơ bị giả mạo bằng hình ảnh hoặc video (Spoofing attack).

3. Hướng phát triển trong tương lai

Để hoàn thiện sản phẩm và hướng tới thương mại hóa, nhóm đề xuất các hướng phát triển tiếp theo:

- + Nâng cấp phần cứng: Nghiên cứu tích hợp các loại Camera Global Shutter hoặc Camera có cảm biến hồng ngoại/chiều sâu để hỗ trợ nhận diện ban đêm và chống giả mạo (Liveness Detection).
- + Tối ưu hóa AI: Chuyển đổi mô hình nhận diện sang dạng biên (Edge AI) để có thể chạy trực tiếp một phần trên các vi điều khiển mạnh hơn (như ESP32-S3 hoặc K210) nhằm giảm phụ thuộc vào Server.
- + Mở rộng tính năng phần mềm: Phát triển ứng dụng di động (Mobile App) để nhân viên có thể theo dõi công chấm của mình và nhận thông báo đẩy (Push Notification) khi check-in thành công.
- + Tích hợp Cloud: Đồng bộ dữ liệu chấm công lên Cloud (Firebase/AWS) để quản lý tập trung cho chuỗi nhiều cửa hàng/văn phòng khác nhau.

TÀI LIỆU THAM KHẢO

Tài liệu về SCRFD: <https://deepwiki.com/deepinsight/insightface/3.1-scrfd-face-detection>

Tài liệu về recognitionFace: <https://www.width.ai/post/facial-detection-and-recognition-with-dlib>

<https://github.com/deepinsight/insightface/blob/cf2a79c1/detection/scrfd/README.md#L104-L105>

Link tham khảo:

<https://gist.github.com/tanukon/2cce46af316f0d35b1278d9cd5dbe270>

Nguồn: <https://medium.com/analytics-vidhya/a-gentle-introduction-into-the-histogram-of-oriented-gradients-fdee9ed8f2aa>

PHỤ LỤC

A. Cơ sở lý thuyết xử lý ảnh và sinh trắc học

Sinh trắc học (hay Biometrics) là các đặc điểm thể chất và hành vi có thể đo lường cho phép thiết lập và xác minh danh tính của một cá nhân. Xác thực sinh trắc học là một hình thức bảo mật đo lường và đối sánh các tính năng sinh trắc học của người dùng để xác minh rằng một người đang cố gắng truy cập vào một hệ thống, thiết bị cụ thể được phép làm như vậy. Đặc điểm sinh trắc học là các đặc điểm vật lý và sinh học dành riêng cho một cá nhân và có thể dễ dàng so sánh với các đặc điểm được phép lưu trong cơ sở dữ liệu. Nếu các tính năng sinh trắc học của một người dùng đang cố gắng truy cập vào một hệ thống, thiết bị khớp với các tính năng của một người dùng được phê duyệt thì quyền truy cập vào thiết bị sẽ được cấp. Xác thực sinh trắc học hiện nay rất phát triển dùng rất phổ biến như kiểm soát các điểm truy cập cửa ra vào, bảo mật trong điện thoại thông minh và máy tính.

Hệ thống nhận dạng khuôn mặt là một ứng dụng máy tính tự động xác định hoặc nhận dạng một người nào đó từ một bức hình ảnh kỹ thuật số hoặc một khung hình video từ một nguồn video. Một trong những cách để thực hiện điều này là so sánh các đặc điểm khuôn mặt chọn trước từ hình ảnh và một cơ sở dữ liệu về khuôn mặt. Hệ thống nhận diện khuôn mặt hay nhận dạng khuôn mặt là một loại ứng dụng có khả năng xác định được hình ảnh khuôn mặt của một người nào đó dựa trên các đặc điểm mà cơ sở dữ liệu đã lưu trữ trước đó. Mỗi khuôn mặt đều có nhiều điểm mốc, những phần lồi lõm tạo nên các đặc điểm của khuôn mặt. Các hệ thống nhận diện gương mặt định nghĩa những điểm này là những điểm nút. Mỗi mặt người có khoảng 80 điểm nút. Có thể nhận diện một số điểm nút như sau:

- Khoảng cách giữa hai mắt.
- Chiều rộng của mũi.
- Độ sâu của hốc mắt.
- Hình dạng của xương gò má.
- Độ dài của xương hàm...



Hình 0.1: Các đặc điểm trên khuôn mặt

1. Nguyên lý hoạt động

Trước khi có thể sử dụng công nghệ này, chúng ta đều phải trải qua bước lưu trữ thông tin trên cơ sở dữ liệu để hệ thống có thể so sánh. Sau khi các đặc điểm đã được mã hóa thì chúng sẽ được lưu vào một nơi phù hợp. Khi cần thiết thì hệ thống sẽ quét và so sánh.

Các hệ thống công nghệ trên khuôn mặt có thể khác nhau, nhưng nhìn chung, chúng có xu hướng hoạt động theo một số bước như sau: Người dùng sẽ tiến đến gần camera được bố trí sẵn => Camera sẽ quét toàn bộ khuôn mặt => Công nghệ sẽ tiến hành bóc tách khuôn mặt ra khỏi cảnh vật xung quanh => Tại đây hệ thống sẽ tiến hành phân tích các đặc điểm của khuôn mặt, các đặc tính sẽ được xử lý dưới dạng các thuật toán và công thức, dữ liệu sẽ được gửi về nơi lưu trữ có liên quan nhất => Sau đó hệ thống sẽ tiến hành so sánh dữ liệu vừa phân tích với các dữ liệu có sẵn => Khi tỉ lệ trùng khớp gần như hoàn toàn thì thiết bị sẽ được mở.

2. Thuật toán

Một số thuật toán nhận dạng khuôn mặt xác định các đặc điểm khuôn mặt bằng cách trích xuất các ranh giới, hoặc đặc điểm, từ một hình ảnh khuôn mặt của đối tượng. Từ đó các thuật toán sẽ trích xuất được các thông tin, và những tính năng này sau đó được sử dụng để tìm kiếm các hình ảnh khác với các tính năng phù hợp. Trong

trường hợp sử dụng để nhận diện, cần phải lưu lại thông tin khuôn mặt để ghi nhớ trước. Các thuật toán sẽ đơn giản hóa một tập các hình ảnh khuôn mặt và sau đó nén dữ liệu khuôn mặt, chỉ lưu dữ liệu hình ảnh nào là hữu ích cho việc nhận dạng khuôn mặt. Khi đó, muốn nhận diện sẽ so sánh hình ảnh mẫu với các dữ liệu khuôn mặt đã lưu.

Các thuật toán nhận dạng có thể được chia thành hai hướng chính, là hình học, đó là nhìn vào tính năng phân biệt, hoặc trắc quang (đo sáng), là sử dụng phương pháp thống kê để 'chưng cất' một hình ảnh thành những giá trị và so sánh các giá trị với các mẫu để loại bỏ chênh lệch.

Các thuật toán nhận dạng phổ biến bao gồm Principal Component Analysis (Phép phân tích thành phần chính) sử dụng các khuôn mặt riêng, Linear Discriminant Analysis (Phân tích biệt tuyến tính), Elastic Bunch Graph Matching sử dụng thuật toán Fisherface, các mô hình Markov ẩn, Multilinear Subspace Learning (Luyện nhớ không gian con đa tuyến) sử dụng đại diện cơ căng, và theo dõi liên kết động thần kinh.

3. Ưu điểm và nhược điểm

Trong số các kỹ thuật sinh trắc học, nhận dạng khuôn mặt có thể không đáng tin cậy và hiệu quả nhất. Tuy nhiên, một trong những lợi thế quan trọng là nó không đòi hỏi sự hợp tác của các đối tượng thử nghiệm. Các hệ thống thiết kế được lắp đặt tại các sân bay, khu chung cư, và những nơi công cộng khác có thể xác định các cá nhân giữa đám đông, mà không bỏ sót một ai. Sinh trắc học khác như dấu vân tay, quét móng mắt, và nhận dạng giọng nói không thể thực hiện được như vậy.

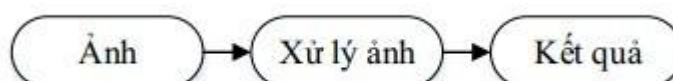
Mặc dù nhận dạng khuôn mặt đã thực hiện được khá tốt ở phía mặt trước và phía chênh lệch 20 độ, nhưng ngay sau khi bạn đi về phía góc khuất, thì nó có vấn đề.

Một số điều kiện có thể ảnh hưởng tới tính chính xác của phương pháp như:

- Thiếu ánh sáng, đeo kính mát, tóc dài, hoặc một phần khuôn mặt bị che.
- Hình ảnh độ phân giải thấp.
- Hệ thống sẽ kém hiệu quả nếu biểu hiện khuôn mặt khác nhau

4. Hệ thống xử lý ảnh

Con người thu nhận thông tin qua các giác quan, trong đó thị giác đóng vai trò quan trọng nhất. Những năm trở lại đây với sự phát triển của phần cứng máy tính, xử lý ảnh và đồ họa đã phát triển một cách mạnh mẽ và có nhiều ứng dụng trong cuộc sống. Xử lý ảnh và đồ họa đóng vai trò quan trọng trong tương tác người máy. Quá trình xử lý ảnh được xem như là quá trình thao tác ảnh đầu vào nhằm cho ra kết quả mong muốn. Kết quả đầu ra của một quá trình xử lý một kết luận.



Hình 0.2: Quá trình xử lý ảnh

Ảnh có thể xem là tập hợp các điểm ảnh và mỗi điểm ảnh được xem như là đặc trưng cường độ sáng hay một dấu hiệu nào đó tại một vị trí nào đó của đối tượng trong không gian và nó có thể xem như một hàm n biến $P(c_1, c_2, \dots, c_n)$. Do đó, ảnh trong xử lý ảnh có thể xem như ảnh n chiều. Các vấn đề của xử lý ảnh:

- Thu nhận ảnh, chụp ảnh và số hóa ảnh.
 - + Hệ thống chụp ảnh và tín hiệu ảnh.
 - + Hệ thống số hóa ảnh: lấy mẫu, lượng tử hóa.
- Phân tích ảnh và thị giác máy tính.
 - + Cải thiện nâng cấp ảnh, sửa lỗi, khôi phục ảnh.
 - + Phân tách đặc trưng: tách biên, phân vùng ảnh.
 - + Biểu diễn và xử lý đặc trưng hình dạng đối tượng ảnh.
 - + Nhận dạng đối tượng ảnh, phân tích cảnh và hiểu cảnh.
- Mã hóa, nén ảnh.
 - Các phương pháp nén và các chuẩn nén.

4.1. Một số khái niệm trong xử lý ảnh

4.1.1. Điểm ảnh

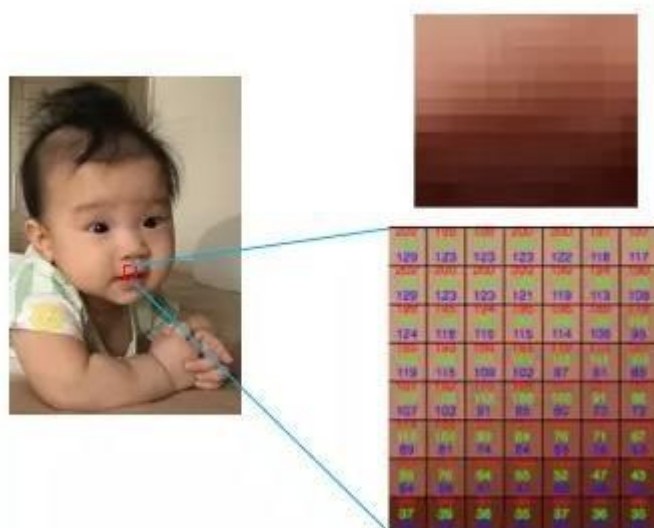
Gốc của ảnh (ảnh tự nhiên) là ảnh liên tục về không gian và độ sáng. Để xử lý được bằng máy tính, ảnh cần phải được số hóa. Số hóa ảnh là sự biến đổi gần đúng một ảnh liên tục thành một tập điểm phù hợp với ảnh thật về vị trí (không gian) và độ sáng (mức xám). Khoảng cách giữa các điểm ảnh đó được thiết lập sao cho mắt người không phân biệt được ranh giới giữa chúng. Mỗi một điểm như vậy gọi là điểm

ảnh (PEL: Picture Element) hay gọi tắt là Pixel. Trong khuôn khổ ảnh hai chiều, mỗi pixel ứng với cặp tọa độ (x, y) .

Điểm ảnh (Pixel) là một phần tử của ảnh số tại tọa độ (x, y) với độ xám hoặc màu nhất định. Kích thước và khoảng cách giữa các điểm ảnh đó được chọn thích hợp sao cho mắt người cảm nhận sự liên tục về không gian và mức xám (hoặc màu) của ảnh số gần như ảnh thật. Mỗi một phần tử trong ma trận được gọi là một phần tử ảnh.



Điểm ảnh trong ảnh trắng đen



Điểm ảnh trong ảnh màu

4.1.2. Độ phân giải

Độ phân giải (Resolution) là thước đo chi tiết rõ ràng nhỏ nhất trong ảnh, được tính là số điểm ảnh trên một đơn vị khoảng cách.



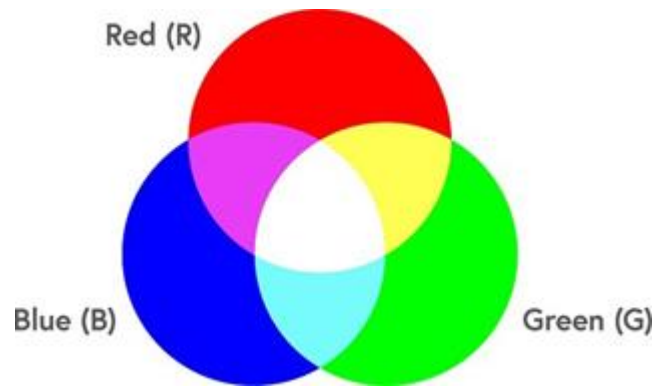
Độ phân giải của ảnh *Hình 0.4: Độ phân giải của ảnh*

Như ảnh trên chúng ta thấy rằng tuy cùng kích thước nhưng độ phân giải khác nhau, và độ phân giải càng thấp thì càng mờ. Như ảnh 1 sẽ được hiểu là chiều rộng có 175 điểm ảnh và chiều cao có 256 điểm ảnh.

4.1.3. Mức xám

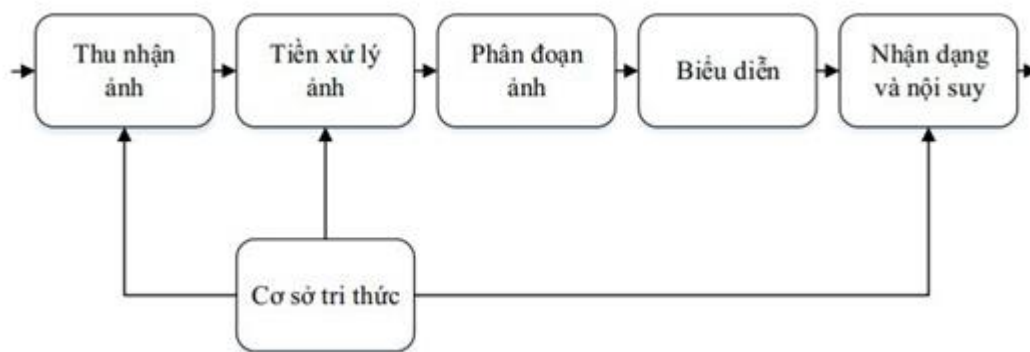
Mức xám của điểm ảnh là cường độ sáng của nó được gán bằng giá trị số tại điểm đó.

- Các thang mức xám thông thường: 16, 32, 64, 128, 256 (Mức 256 là mức phổ dụng. Nguyên nhân là từ kỹ thuật máy tính dùng 1 byte (8bit) để biểu diễn mức xám. Mức xám dùng 1 byte biểu diễn, $2^8 = 256$, tức là từ 0 đến 255).
- Ảnh đen trắng: Là ảnh có hai màu đen, trắng (không chứa màu khác) với mức xám ở các điểm ảnh có thể khác nhau.
- Ảnh nhị phân: Ảnh chỉ có hai mức đen trắng phân biệt, tức là mỗi điểm ảnh của ảnh nhị phân chỉ có thể là 0 hoặc 1.
- Ảnh màu: Trong khuôn khổ lý thuyết 3 màu (Red, Blue, Green) để tạo nên thế giới màu, người ta thường dùng 3 byte để mô tả mức màu, khi đó các giá trị màu: $2^8 \times 3 = 224 \approx 16,7$ triệu màu.



Hệ màu RGB Hình 0.5: Hệ màu RGB

4.2. Các bước cơ bản trong xử lý ảnh



Các bước cơ bản trong xử lý ảnh Hình 0.6: Các bước cơ bản trong xử lý ảnh

Sơ đồ này bao gồm các phần sau:

- **Phần thu nhận ảnh (Image Acquisition).** Ảnh có thể nhận qua camera màu hoặc đen trắng. Camera thường dùng là loại quét dòng; ảnh tạo ra có dạng hai chiều. Chất lượng một ảnh thu nhận được phụ thuộc vào thiết bị thu, môi trường (ánh sáng, thời tiết).
- **Tiền xử lý (Image Processing):** Sau bộ thu nhận, ảnh có thể bị nhiễu, độ tương phản thấp nên cần đưa vào bộ tiền xử lý để nâng cao chất lượng. Chức năng chính của bộ tiền xử lý là lọc nhiễu, nâng độ tương phản để làm ảnh rõ, nét hơn.
- **Phân đoạn (Segmentation) hay phân vùng ảnh:** là tách một ảnh đầu vào thành các vùng thành phần để biểu diễn phân tích, nhận dạng ảnh. Ví dụ: để nhận dạng chữ (hoặc mã vạch) trên phong bì thư cho mục đích phân loại bưu phẩm, cần chia các câu, chữ, về địa chỉ hoặc tên người thành các từ, các chữ, các số (hoặc các vạch) riêng biệt để nhận dạng. Đây là phần phức tạp khó khăn nhất trong xử lý ảnh và cũng dễ gây lỗi, làm mất độ chính xác của ảnh. Kết quả nhận dạng ảnh phụ thuộc rất nhiều vào công đoạn này.

- **Biểu diễn ảnh (Image Representation):** Đầu ra ảnh sau phân đoạn chứa các điểm ảnh của vùng ảnh (ảnh đã phân đoạn) cộng với mã liên kết với các vùng lân cận. Việc chọn các tính chất để thể hiện ảnh gọi là trích chọn đặc trưng (Feature Selection) gắn với việc tách các đặc tính của ảnh dưới dạng các thông tin định lượng hoặc làm cơ sở để phân biệt lớp đối tượng này với đối tượng khác trong phạm vi ảnh nhận được. Ví dụ: trong nhận dạng ký tự trên phong bì thư, chúng ta miêu tả các đặc trưng của từng ký tự giúp phân biệt ký tự này với ký tự khác.
- **Nhận dạng ảnh và nội suy ảnh (Image Recognition and Interpretation):** Nhận dạng ảnh là quá trình xác định ảnh. Quá trình này thường thu được bằng cách so sánh với mẫu chuẩn đã được học (hoặc lưu từ trước). Nội suy là phán đoán theo ý nghĩa trên cơ sở nhận dạng. Ví dụ: một loạt chữ số và nét gạch ngang trên phong bì thư có thể được nội suy thành mã điện thoại. Có nhiều cách phân loại ảnh khác nhau về ảnh. Theo lý thuyết về nhận dạng, các mô hình toán học về ảnh được phân theo hai loại nhận dạng ảnh cơ bản: + Nhận dạng theo tham số. + Nhận dạng theo cấu trúc. Một số đối tượng nhận dạng khá phổ biến hiện nay đang được áp dụng trong khoa học và công nghệ là nhận dạng ký tự (chữ in, chữ viết tay, chữ ký điện tử), nhận dạng văn bản (text), nhận dạng vân tay, nhận dạng mã vạch, nhận dạng mặt người...
- **Cơ sở tri thức (Knowledge Base):** Như đã nói ở trên, ảnh là một đối tượng khá phức tạp về đường nét, độ sáng tối, dung lượng điểm ảnh, môi trường để thu ảnh phong phú kéo theo nhiều. Trong nhiều khâu xử lý và phân tích ảnh ngoài việc đơn giản hóa các phương pháp toán học đảm bảo tiện lợi cho xử lý, người ta mong muốn bắt chước quy trình tiếp nhận và xử lý ảnh theo cách của con người. Trong các bước xử lý đó, nhiều khâu hiện nay đã xử lý theo các phương pháp trí tuệ con người. Vì vậy, ở đây các cơ sở tri thức được phát huy.

B. Các mô hình AI sử dụng

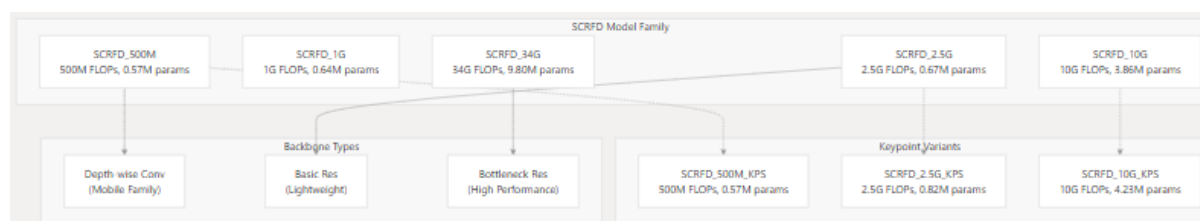
1. Mô hình phát hiện khuôn mặt SCRFD (Sample and Computation Redistribution for Efficient Face Detection)

1.1. Tổng quan

SCRFD là một phương pháp phát hiện khuôn mặt hiệu quả được mô tả lần đầu trong một bài báo trên arXiv và được chấp nhận bởi ICLR-2022. Phương pháp này tập trung vào việc phân bổ lại mẫu và tính toán để đạt được độ chính xác cao trong khi vẫn duy trì hiệu quả tính toán trong các điều kiện hạn chế tài nguyên khác nhau.

Hệ thống cung cấp nhiều biến thể mô hình với hiệu năng từ 500 triệu đến 34 tỷ FLOP, trong đó một số biến thể bao gồm khả năng dự đoán điểm mấu chốt trên khuôn mặt. Tất cả các mô hình đều được đánh giá trên độ phân giải VGA (640x480) và được tối ưu hóa cho suy luận thời gian thực.

1.2 Kiến trúc biến thể mô hình SCRFD



Kiến trúc biến thể mô hình SCRFD Hình 0.7: Kiến trúc biến thể mô hình SCRFD

1.3 Các chỉ số về độ chính xác hiệu quả

Các mô hình SCRFD đạt được hiệu suất hàng đầu trong các mức ngân sách tính toán khác nhau. Hiệu suất được đo trên tập dữ liệu WIDERFace bằng cách sử dụng các tập con Dễ, Trung bình và Khó:

Bảng 0.1: Bảng so sánh các mô hình

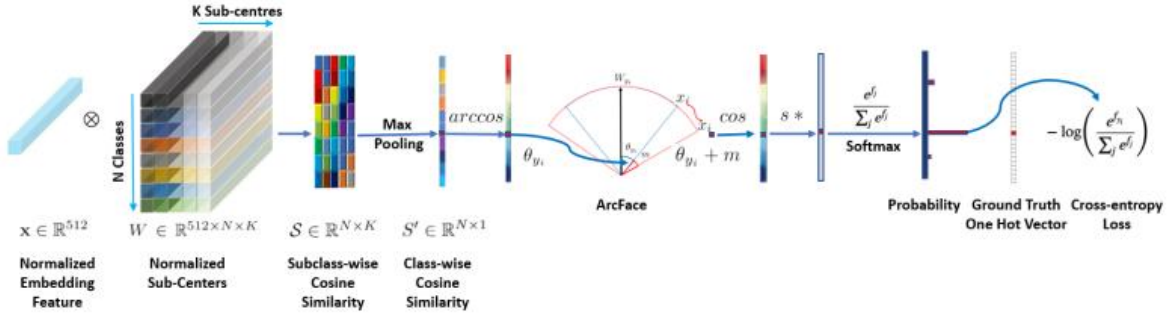
Model	Easy	Medium	Hard	FLOPs	Params(M)	Inference(ms)
SCRFD_500M	90.57	88.12	68.51	500M	0.57	3.6
SCRFD_1G	92.38	90.57	74.80	1G	0.64	4.1
SCRFD_2.5G	93.78	92.16	77.87	2.5G	0.67	4.2
SCRFD_10G	95.16	93.87	83.05	10G	3.86	4.9
SCRFD_34G	96.06	94.92	85.29	34G	9.80	11.7

1.4. Hiệu năng CPU

SCRFD-0.5GF thể hiện khả năng suy luận CPU hiệu quả:

- Độ phân giải 640x480 : 28,3ms (đơn luồng AMD Ryzen 9 3950X)
- Độ phân giải 320x240 : 11.4ms (đơn luồng AMD Ryzen 9 3950X)

2. Mô hình trích xuất đặc trưng ArcFace



Hình 0.8: Trích xuất đặc trưng ArcFace

ArcFace là một trong những phương pháp nhận dạng khuôn mặt sâu nổi tiếng hiện nay. Đặc điểm chính của ArcFace là áp dụng hàm mất mát Additive Angular Margin Loss để tăng cường tính nhỏ gọn của không gian nhúng trong cùng một lớp (cùng một người) và sự khác biệt giữa các lớp (khác nhau). Trước khi ArcFace ra đời, nhiều phương pháp được đề xuất đã sử dụng hàm mất mát softmax làm hàm mất mát phân loại trong nhận dạng khuôn mặt sâu. Tuy nhiên, việc chỉ sử dụng hàm mất mát softmax có một số nhược điểm. Một trong số đó là hàm mất mát softmax không tối ưu hóa việc nhúng đặc trưng để hạn chế sự tương đồng cao hơn giữa các mẫu trong cùng một lớp và sự đa dạng cho các mẫu giữa các lớp, điều này có nghĩa là ranh giới giữa các cá nhân đôi khi không rõ ràng và làm giảm hiệu suất của mô hình. Vì vậy, các tác giả đã giới thiệu hàm mất mát Additive Angular Margin Loss để đạt được sự cải thiện hơn nữa về khả năng phân biệt của nhận dạng khuôn mặt. Tiếp theo, chúng em sẽ xem xét sự so sánh toán học giữa hàm mất mát softmax và hàm mất mát Additive Angular Margin Loss.

Các công thức toán học của hàm mất mát softmax và hàm mất mát Additive Angular Margin (viết tắt là AAM) như sau:

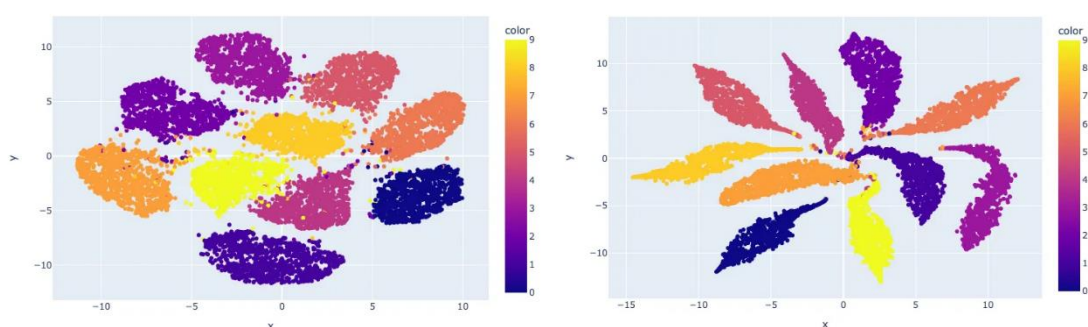
$$L_{\text{softmax}} = -\log \frac{e^{W_{y_i}^T x_i + b_{y_i}}}{\sum_{j=1}^N e^{W_j^T x_i + b_j}} \quad (1)$$

$$L_{\text{AAM}} = -\log \frac{e^{s \cos(\theta_{y_i} + m)}}{e^{s \cos(\theta_{y_i} + m)} + \sum_{j=1, j \neq y_i}^N e^{s \cos(\theta_j)}} \quad (2)$$

trong đó $x_i \in \mathbb{R}^d$ biểu thị đặc trưng hình ảnh của mẫu thứ i , thuộc lớp thứ y_i . Kích thước đặc trưng được đặt là 512 theo quy ước. W_j đề cập đến cột thứ j của trọng số

$W \in \mathbb{R}^{d \times n}$, $b_j \in \mathbb{R}^n$ là số hạng độ lệch và số lớp là N . Trong công thức (2), m biểu thị biên độ sẽ được giải thích sau.

Cho đến nay, chúng em đã xem xét các công thức toán học của ArcFace, nhưng liệu chúng có đúng trong thực tế không? Muốn kiểm tra trực quan sự khác biệt về đặc điểm hình ảnh giữa mô hình không có tổn thất Additive Angular Margin và mô hình có tổn thất này. Chúng em đã sử dụng tập dữ liệu MNIST làm ví dụ và trực quan hóa cả hai đặc điểm hình ảnh với các tài liệu tham khảo. Các biểu đồ bên dưới là hình ảnh trực quan hóa đặc điểm hình ảnh với t-SNE.



Hình 0.9: Ảnh trực quan hóa đặc điểm hình ảnh với t-SNE

Đồ thị bên trái hiển thị đặc trưng hình ảnh không có AAM, và đồ thị bên phải hiển thị đặc trưng hình ảnh có. Như chúng ta có thể thấy, hình ảnh có AAM (bên phải) có không gian đặc trưng hình ảnh dày đặc hơn trong cùng một lớp nhưng lại có khả năng phân biệt tốt hơn giữa các lớp so với hình ảnh không có AAM (bên trái).