

HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG



Nguyễn Xuân Bách

**HỆ THỐNG XÁC THỰC KHUÔN MẶT CHO ỨNG DỤNG DI
ĐỘNG**

Chuyên ngành: Hệ thống thông tin

Mã số: 8480104

TÓM TẮT ĐỀ ÁN TỐT NGHIỆP THẠC SĨ

HÀ NỘI - NĂM 2025

Đề án tốt nghiệp được hoàn thành tại:

HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG

Người hướng dẫn khoa học: TS. Nguyễn Duy Phương
(Ghi rõ học hàm, học vị)

Phản biện 1: PGS.TS. Nguyễn Hà Nam

Phản biện 2: PGS.TS. Nguyễn Trọng Khánh

Đề án tốt nghiệp sẽ được bảo vệ trước Hội đồng chấm đề án tốt nghiệp
thạc sĩ tại Học viện Công nghệ Bưu chính Viễn thông

Vào lúc: 8 giờ 25 ngày 19 tháng 07 năm 2025

Có thể tìm hiểu đề án tốt nghiệp tại:

- Thư viện của Học viện Công nghệ Bưu chính Viễn thông.

MỞ ĐẦU

1. Lý do lựa chọn đề tài

Khái quát vấn đề còn tồn tại:

- Thiếu khả năng xác minh người dùng chính chủ.

- Ứng dụng vào các ứng dụng nhà thông minh

- Trải nghiệm người dùng kém

Lý do lựa chọn đề tài:

- Trong bối cảnh phát triển của công nghệ hiện tại, nhu cầu bảo mật càng trở nên cần thiết.

- Xác thực khuôn mặt giúp nhận diện chính xác, tăng cường bảo mật và mang đến trải nghiệm tiện lợi, nhanh chóng cho người dùng.

- Tích hợp với kịch bản tự động như điều kiện đầu vào giúp tạo ra những kịch bản phù hợp với ngôi nhà

- Vì vậy, đề tài “Hệ thống xác thực khuôn mặt cho ứng dụng di động” được lựa chọn để cung cấp cho các nhà phát triển ứng dụng di động một cách để tích hợp chức năng xác thực khuôn mặt vào ứng dụng của mình, giúp đảm

bảo quyền riêng tư cho người dùng, đồng thời giúp cuộc sống con người dễ dàng hơn.

2. Mục đích, ý nghĩa của đề tài

Mục tiêu của đề án

- Tạo ra được hệ thống xác thực khuôn mặt hiệu quả, tích hợp nhanh chóng và dễ dàng với các ứng dụng di động
- Tích hợp được các công cụ nhận diện, xác thực khuôn mặt vào hệ thống
- Giảm thời gian thực thi và tăng độ chính xác trong việc xác thực khuôn mặt

Các kết quả cần đạt được:

Về mặt lý luận

- Hiểu rõ hơn về các mô hình, phương pháp nhận diện, xác thực khuôn mặt, từ đó có thể cải tiến thuật toán, giúp rút ngắn thời gian thực thi hoặc tăng độ chính xác
- Hiểu rõ hơn về cách vận hành hệ thống

Về mặt thực tiễn

- Tạo ra được một hệ thống xác thực khuôn mặt hiệu quả
- Hệ thống có thể hoạt động tốt trên các loại smart phone khác nhau.
- Đánh giá, cải tiến các mô hình đã sử dụng.

Ý nghĩa của đề tài: Nghiên cứu này có thể đóng góp vào việc phát triển các hệ thống tích hợp chức năng xác thực khuôn mặt, giúp các ứng dụng di động có thêm lớp bảo mật.

Tính cấp thiết, khả thi của đề tài:

- Tính cấp thiết: Để đảm bảo an toàn dữ liệu trong ứng dụng, cần có một hệ thống nhận diện, xác thực khuôn mặt để tích hợp, triển khai nhanh chóng .
- Tính khả thi: Phần lớn các ứng dụng nhắn tin, ứng dụng nhà thông minh hiện tại chưa tích hợp xác thực khuôn mặt, chưa có những ngữ cảnh, tự động theo sự kiện xác thực khuôn mặt.

3. Đối tượng, phương pháp nghiên cứu

Đối tượng nghiên cứu:

- Các mô hình, phương pháp nhận diện, xác thực khuôn mặt, cùng với các ứng dụng có tích hợp chúng trong các chức năng của ứng dụng.

- Các hệ thống tương tự đã có trong thực tế.

Phương pháp lựa chọn: Mô phỏng và thực nghiệm

Nghiên cứu tài liệu và khảo sát

4. Cấu trúc đề tài

Nội dung chính của đề án được chia thành những chương chính như sau:

- Chương 1: Cơ sở lý luận: Chương này tập trung vào việc phân tích các phương pháp, mô hình được sử dụng để xây dựng hệ thống như MTCNN, MobileNetV2, FaceNet

- Chương 2: Xây dựng hệ thống nhận diện khuôn mặt: Chương này tập trung vào việc thiết kế hệ thống, các module trong hệ thống.

- Chương 3: Triển khai và thử nghiệm: Chương này tập trung vào cách triển khai hệ thống lên ứng dụng nhà thông minh và đưa ra các kết quả thử nghiệm.

- Chương 4: Đánh giá và kết luận: Đánh giá được ưu điểm và nhược điểm còn tồn tại của hệ thống, đề xuất các nghiên cứu nâng cấp và cải thiện hệ thống.

CHƯƠNG 1: CƠ SỞ LÝ LUẬN

1.1. Phát hiện và căn chỉnh khuôn mặt bằng MTCNN

1.1.1. Định nghĩa

MTCNN (Mạng tích chập nhiều tác vụ) là một framework học sâu được thiết kế để đồng thời phát hiện khuôn mặt, căn chỉnh và định vị các điểm mốc của khuôn mặt. MTCNN tích hợp chúng bằng cách sử dụng các mạng CNN nối tiếp thống nhất (unified cascaded CNNs), cho phép xử lý hiệu quả và chính xác.

1.1.2. Cấu tạo

MTCNN bao gồm 3 mạng neural tích chập P-Net, R-Net, O-Net được sắp xếp thành từng tầng, mỗi mạng là một giai đoạn riêng biệt trong quá trình phát hiện khuôn mặt:

1.1.3. Nguyên lý hoạt động

MTCNN bắt đầu với việc xây dựng image pyramid, trong đó hình ảnh đầu vào được thay đổi kích thước theo nhiều tỉ lệ.

P-Net tạo ra những vùng ứng viên khuôn mặt.

R-Net tiếp tục loại bỏ các ứng viên không phải là khuôn mặt.

O-Net đảm nhận việc đưa ra binding box và vị trí 5 điểm mốc của khuôn mặt.

1.1.4. Ưu nhược điểm

Ưu điểm:

- MTCNN có độ chính xác phát hiện khuôn mặt đến 98.7% trên tập dữ liệu Winder Face.
- Khả năng phát hiện khuôn mặt cùng với căn chỉnh và xác định các điểm mốc giúp đơn giản hoá cho các tác vụ tiếp theo như nhận diện khuôn mặt, kiểm tra giả mạo khuôn mặt.
- Hoạt động trong thời gian thực.

Nhược điểm:

- Độ phức tạp trong tính toán của MTCNN vẫn còn lớn đối với các ảnh có độ phân giải cao.
- Khó có thể phát hiện được các khuôn mặt khi ở các góc lạ hoặc bị che khuất quá nhiều.

1.1.5. Ứng dụng thực tế

1.2. Nhận diện khuôn mặt với FaceNet và MobileNet

1.2.1. Tổng quan về nhận diện khuôn mặt

1.2.2. MobileNet

1.2.2.1. Tổng quan về MobileNet

1.2.2.2. Depthwise separable convolution

Mô hình MobileNet được xây dựng dựa trên depthwise separable convolution (mạng tích chập phân tách theo chiều sâu) bao gồm 2 bước:

- Depthwise convolution (tích chập theo chiều sâu): là một kiểu tích chập mà áp dụng từng bộ lọc tích chập cho mỗi kênh đầu vào. Tích chập theo chiều sâu chia đầu vào thành các kênh, thực hiện tích chập với từng bộ lọc riêng biệt, sau đó xếp chồng các tích chập với nhau thành đầu ra.

- Pointwise convolution (tích chập điểm): là kiểu tích chập 1×1 : sử dụng kernel 1×1 để kết hợp các đặc trưng từ nhiều kênh của đầu ra của bước trước. Số kênh đầu ra tương đương với số kernel cần có.

Với depthwise separable convolution, MobileNet đã làm giảm được phần lớn chi phí tính toán so với tích chập tiêu chuẩn. Điều này làm nó trở nên phù hợp với các hệ thống có tài nguyên hạn chế.

1.2.2.3. Cấu trúc inverted residual và linear bottlenecks

Khác với residual block truyền thống sử dụng mô hình “rộng-hẹp-rộng”, inverted residual block đảo ngược thành “hẹp-rộng-hẹp”. Tiếp theo, lớp depthwise convolution 3×3 thực hiện lọc đặc trưng không gian, và cuối cùng, lớp projection giảm số kênh thông qua pointwise convolution 1×1 . Kết nối tắt (shortcut) chỉ được áp dụng khi $\text{stride} = 1$ và kích thước đầu vào/ra bằng nhau, giúp tối ưu hóa luồng gradient.

Điểm đột phá thứ hai nằm ở việc loại bỏ hàm kích hoạt phi tuyến (ReLU) trong các lớp bottleneck mỏng. MobileNetV2 chỉ sử dụng hàm tuyến tính ở lớp projection, trong khi giữ ReLU6 cho các lớp mở rộng có chiều cao.

1.2.2.4. Cấu tạo của MobileNetV2

Mạng được thiết kế theo hướng giảm dần độ phân giải và tăng dần số kênh:

- Input: ảnh RGB $224 \times 224 \times 3$
- Conv đầu tiên: 32 kênh, stride 2
- Các bottleneck layer: $16 \rightarrow 24 \rightarrow 32 \rightarrow 64 \rightarrow 96 \rightarrow 160 \rightarrow 320$ kênh
- Conv cuối: 1×1 với 1280 kênh
- Global average pooling
- Lớp phân loại đầu ra (1×1 conv)

1.2.2.5. Tham số thu gọn mô hình

Width Multiplier (α): giảm số kênh (channel) tại tất cả các lớp.

Resolution Multiplier (ρ): giảm độ phân giải của ảnh input.

Hệ số mở rộng (t): làm tăng số kênh đầu vào lên t lần, mở rộng không gian, qua đó giúp mạng học được những đặc trưng phức tạp hơn.

1.2.3. Kiến trúc FaceNet sử dụng MobileNet

1.2.3.1. Tổng quan về FaceNet

1.2.3.2. Kiến trúc của FaceNet

Mạng FaceNet bao gồm một lớp đầu vào theo lô (batch) và một mạng tích chập sâu, sau đó là chuẩn hóa L2 để tạo các vector embedding và cuối cùng tính toán triplet loss để tạo khoảng cách giữa các đối tượng giống nhau càng nhỏ càng tốt, khoảng cách giữa các đối tượng khác nhau càng lớn càng tốt.

Mạng tích chập được sử dụng trong FaceNet là GoogleNet, sử dụng module Inception.

1.2.3.3. FaceNet dựa trên MobileNet

Với việc sử dụng MobileNet thay thế cho GoogleNet trong việc trích xuất đặc trưng khuôn mặt, số lượng tính toán được giảm đáng kể mà vẫn giữ được kết quả tốt.

CHƯƠNG 2: XÂY DỰNG HỆ THỐNG NHẬN DIỆN KHUÔN MẶT

2.1. Kiến trúc hệ thống

Hệ thống nhận diện khuôn mặt này là một giải pháp hữu hiệu được xây dựng bằng FastAPI và ứng dụng các công nghệ học sâu hiện đại. Hệ thống kết hợp các giai đoạn

phát hiện khuôn mặt, phát hiện khuôn mặt giả mạo và nhận dạng khuôn mặt để đảm bảo xác thực người dùng mạnh mẽ và an toàn.

Các thành phần chính bao gồm:

- Lớp API (FastAPI)
- Lớp dịch vụ (FaceRecognitionService)
- Lớp dữ liệu (SQLAlchemy)

Mô hình này giúp phân biệt rõ ràng giữa các vai trò, tăng khả năng kiểm soát và mở rộng sau này.

2.2. Quy trình xử lý

Quy trình xử lý nhận diện khuôn mặt trong hệ thống được chia thành ba giai đoạn chính:

- Phát hiện khuôn mặt (Face Detection)
- Kiểm tra giả mạo khuôn mặt (Liveness Detection)
- Nhận diện khuôn mặt (Face Recognition)

2.3. Cơ sở dữ liệu của hệ thống

Cơ sở dữ liệu của hệ thống gồm một bảng duy nhất là “users”. Bảng gồm có trường id là khóa chính, trường tên, ba trường để lưu trữ vector embedding dạng nhị phân.

Ngoài ra, còn một trường dạng boolean, được sử dụng để kiểm soát việc bật/tắt chức năng nhận diện khuôn mặt cho ứng dụng.

2.4. Các chức năng chính của hệ thống

2.4.1. Đăng ký khuôn mặt

Lưu lại các vector embedding vào cơ sở dữ liệu

2.4.2. Nhận diện khuôn mặt

So sánh các vector embedding từ ảnh với trong cơ sở dữ liệu

2.4.3. Nhận diện khuôn mặt giả mạo

Nêu lý do lựa chọn MobileNetV2 làm backbone cho mô hình nhận diện khuôn mặt giả mạo.

So sánh với các mô hình khác như ResNet và EfficientNet, làm rõ tại sao việc sử dụng MobileNetV2 lại tốt hơn chúng.

2.4.4. Quản lý người dùng và bật/tắt nhận diện khuôn mặt

2.5. Hiệu suất và khả năng mở rộng

Đánh giá về hiệu suất và khả năng mở rộng của hệ thống:

- Hệ thống sử dụng FastAPI – một framework nhẹ và hiệu suất cao – giúp xử lý hàng trăm yêu cầu mỗi giây.
- Việc sử dụng TensorFlow Lite và SQLite giúp hệ thống có thể triển khai trên các thiết bị edge hoặc hệ thống nhúng.
- Kiến trúc phân tầng giúp dễ dàng mở rộng với các tính năng mới như nhận diện biểu cảm, phân tích giới tính, độ tuổi,...

CHƯƠNG 3: TRIỂN KHAI VÀ THỬ NGHIỆM

3.1. Huấn luyện mô hình

3.1.1. Nhận diện khuôn mặt

3.1.1.1. Giai đoạn phát hiện khuôn mặt với MTCNN tối ưu

MTCNN hoạt động thông qua ba mạng con xếp tầng (P-Net, R-Net, O-Net), mỗi mạng giải quyết các nhiệm vụ cụ thể với độ phức tạp tăng dần. P-Net (Proposal Network) xử lý ảnh đầu vào 12×12 pixel để đề xuất các vùng khuôn

mặt tiềm năng thông qua một mạng CNN nhẹ sử dụng depthwise separable convolution.

3.1.1.2. Trích xuất đặc trưng với Fast FaceNet

Thay thế Inception-ResNet bằng MobileNetV2 trong FaceNet mang lại những cải tiến đáng kể về hiệu suất. Kiến trúc MobileNetV2 gồm 15 inverted residual blocks. Lớp projection cuối cùng sử dụng linear activation để tránh mất thông tin trong không gian chiều thấp. Quá trình huấn luyện kết hợp triplet loss.

3.1.1.3. Tích hợp hệ thống end-to-end

Luồng xử lý hoàn chỉnh bao gồm:

- Tiền xử lý ảnh: Chuẩn hóa histogram và cân bằng ánh sáng sử dụng CLAHE.
- Phát hiện đa phương thức: Chạy MTCNN trên ảnh RGB.
- Trích xuất đặc trưng: Fast FaceNet xử lý hình ảnh để lấy được vector embedding khuôn mặt.
- Matching thông minh: sử dụng cosine similarity.

3.1.2. Phát hiện khuôn mặt giả mạo

3.1.2.1. Mô tả bài toán

Bài toán phát hiện giả mạo khuôn mặt được định nghĩa là một bài toán phân loại nhị phân. Cho một ảnh đầu vào I , mô hình cần phải dự đoán nhãn y của ảnh đó, trong đó:

- $y = 1$ nếu ảnh I chứa khuôn mặt của người thật (live).
- $y = 0$ nếu ảnh I là một phiên bản giả mạo (spoof).

3.1.2.2. Mô tả tập dữ liệu

Để huấn luyện và đánh giá mô hình, luận án sử dụng tập dữ liệu CelebA-Spoof, một bộ dữ liệu công khai và quy mô lớn được thiết kế chuyên biệt cho bài toán này.

- Quy mô: Tập dữ liệu bao gồm 100.000 hình ảnh được gán nhãn
- Phân bố lớp: Dữ liệu được chia thành hai lớp: "live" và "spoof".
- Đặc điểm ảnh: Các ảnh trong tập dữ liệu có độ phân giải và điều kiện ánh sáng đa dạng.

- Phân chia dữ liệu: Tập dữ liệu được phân chia thành hai tập con: tập huấn luyện (training set) và tập kiểm định (validation set) theo tỷ lệ 80/20.

3.1.2.3. Tiền xử lý và tăng cường dữ liệu

Chuẩn hóa kích thước và giá trị pixel: Tất cả các ảnh đầu vào, dù có kích thước ban đầu khác nhau, đều được đồng nhất về kích thước 224x224 pixels.

Tăng cường dữ liệu (Data Augmentation): Các phép biến đổi ngẫu nhiên được áp dụng lên ảnh trong quá trình huấn luyện.

3.1.2.4. Kiến trúc mô hình

Sử dụng mô hình nền MobileNetV2 cùng 1 bộ phân loại tùy chỉnh thay cho lớp phân loại cuối cùng.

3.1.2.5. Chiến lược huấn luyện và tùy chỉnh

- Tinh chỉnh (fine - tuning)
- Xử lý mất cân bằng dữ liệu
- Hàm mất mát và trình tối ưu hoá

3.1.2.6. Đánh giá và phân tích

3.2. Triển khai hệ thống nhận diện

3.2.1. API đăng ký khuôn mặt

3.2.2. API nhận diện khuôn mặt

3.2.3. API lấy trạng thái bật/tắt nhận diện khuôn mặt

3.2.4. API cập nhật/thay đổi trạng thái nhận diện khuôn mặt

3.2.5. API xóa khuôn mặt đã lưu

3.3. Triển khai trên ứng dụng ONE Home

3.3.1. Ứng dụng ONE Home

3.3.2. Triển khai trên ứng dụng di động ONE Home

Các chức năng được triển khai trên ứng dụng bao gồm:

- Cài đặt xác thực khuôn mặt
- Bảo mật cho camera:
- Kích hoạt các kịch bản tự động khi camera

nhận diện được khuôn mặt

3.4. Kết quả thử nghiệm

Hiệu Suất Nhận Diện

Trên bộ dữ liệu kiểm thử VGGFace2-MTCNN, mô hình Fast-FaceNet đạt:

- Độ chính xác (Accuracy): ~95%
- Độ nhạy (Recall): ~93%
- Độ chuẩn xác (Precision): ~94%
- F1-score: ~93.5%

Kết quả thực nghiệm trên tập dữ liệu CelebA-Spoof cho thấy mô hình sử dụng MobileNetV2 đạt độ chính xác lên tới 99,4%, đồng thời giảm đáng kể thời gian suy luận (chỉ còn khoảng 35ms mỗi ảnh trên thiết bị phổ thông) và kích thước mô hình (giảm từ 95MB xuống còn 14MB).

Hiệu Năng API

- Thời gian phản hồi trung bình của API là 1.2 giây
- Thời gian để phát hiện khuôn mặt là 0.3 giây
- Thời gian để nhận diện khuôn mặt là 0.8 giây
- Thời gian để kiểm tra liveness là 0.4 giây
- Thời gian để đăng ký khuôn mặt là 4-5 giây

Kết luận

Hệ thống đã đáp ứng được các yêu cầu cơ bản về nhận diện khuôn mặt với độ chính xác cao. Các tính năng chính như liveness detection, face registration và recognition đều hoạt động tốt. Tuy nhiên, vẫn còn một số vấn đề về hiệu suất và bảo mật cần được cải thiện.

3.5. Đánh giá hệ thống

Ưu Điểm của Hệ Thống

- Độ chính xác cao
- Dễ dàng triển khai
- Khả năng mở rộng
- Truy cập từ xa

Hạn chế:

- Chưa có các phương pháp bảo mật dữ liệu.
- Tốc độ xử lý thực tế vẫn chưa cao do sự chậm trễ của server ngrok.
- Độ chính xác của model liveness trong thực tế vẫn còn chưa được tốt.
- Chưa có cơ chế cache cho embedding, có thể gây chậm nếu số lượng user tăng lên.
- Chưa có cơ chế load balancing cho việc xử lý nhiều request đồng thời.

CHƯƠNG 4: KẾT LUẬN VÀ KHUYẾN NGHỊ

Trong đề tài này, học viên đã đề xuất và xây dựng mô hình nhận diện khuôn mặt FaceNet với MobileNet V2 làm xương sống, cùng với đó là mô hình nhận diện khuôn mặt giả mạo cũng với MobileNet V2. Các kết quả thực nghiệm cho thấy hệ thống đã chạy và đạt hiệu suất tốt.

Hệ thống là một giải pháp tốt cho việc xác thực danh tính bằng khuôn mặt, với nhiều tính năng bảo mật và độ chính xác cao. Tuy nhiên, vẫn còn một số điểm cần cải thiện để đáp ứng nhu cầu của hệ thống lớn và đảm bảo hiệu suất tối ưu. Với những cải tiến được đề xuất, hệ thống có tiềm năng trở thành một giải pháp xác thực khuôn mặt mạnh mẽ và đáng tin cậy:

- **Cải thiện hiệu suất:** Thử nghiệm với các kiến trúc mạng tiên tiến hơn và kỹ thuật huấn luyện nâng cao.
- **Tối ưu hóa tốc độ:** Áp dụng các kỹ thuật như quantization và pruning để giảm kích thước mô hình và tăng tốc độ suy luận.

- **Bảo mật dữ liệu:** Triển khai các giải pháp mã hóa và bảo mật cho dữ liệu khuôn mặt nhạy cảm.

- **Mở rộng tính năng:** Phát triển thêm các tính năng như nhận diện cảm xúc, ước tính độ tuổi.

- Cần thêm các cơ chế load balancing, caching để tăng hiệu suất, tốc độ phản hồi của hệ thống.

Hệ thống phù hợp cho:

- Hệ thống xác thực trong doanh nghiệp vừa và nhỏ

- Ứng dụng di động cần xác thực khuôn mặt

- Hệ thống kiểm soát ra vào

- Ứng dụng cần xác thực danh tính với độ bảo mật cao

Tuy nhiên, cần lưu ý về khả năng mở rộng và hiệu suất khi triển khai trong môi trường production với số lượng người dùng lớn.

Với hướng nghiên cứu tiếp theo, chúng ta có thể nghiên cứu, phát triển thêm chức năng nhận diện cảm xúc,

ước tính độ tuổi. Những chức năng mới này đều mang tính ứng dụng cao trong thực tế. Với nhận diện cảm xúc, chúng ta có thể đánh giá được mức độ hài lòng của khách hàng với sản phẩm. Với ước tính độ tuổi, chúng ta có thể áp dụng vào camera để quản lý trẻ nhỏ.