



Ngày nhận bài: 05/7/2026; Ngày thẩm định: 12/8/2026; Ngày nhận đăng: 13/8/2026.

NGHIÊN CỨU XÂY DỰNG MÔ HÌNH THỬ NGHIỆM HỆ THỐNG NHẬN DẠNG KHUÔN MẶT THỜI GIAN THỰC PHỤC VỤ KIỂM SOÁT RA VÀO THÔNG MINH TẠI HỌC VIỆN PHÒNG CHÁY CHỮA CHÁY VÀ CỨU NẠN CỨU HỘ

Thượng tá, TS NGUYỄN TÀI NĂNG LƯỢNG

Khoa Khoa học cơ bản và Ngoại Ngữ, Học viện Phòng cháy chữa cháy và Cứu nạn cứu hộ

*Tác giả liên hệ: Nguyễn Tài Năng Lượng (Email: tailuong2003@gmail.com)

Tóm tắt: Kiểm soát ra vào thủ công bằng thẻ từ, sổ ghi chép tại các cơ sở đào tạo Công an nhân dân đang bộc lộ hạn chế về tốc độ, độ chính xác và khả năng lưu trữ dữ liệu. Bài báo đề xuất kiến trúc và xây dựng mô hình thử nghiệm (pilot) hệ thống nhận dạng khuôn mặt thời gian thực phục vụ kiểm soát ra vào thông minh tại Học viện Phòng cháy chữa cháy và Cứu nạn cứu hộ (PCCC&CNCH), trên cơ sở các mô-đun phát hiện khuôn mặt, nhận dạng khuôn mặt và theo dõi đối tượng. Mô hình được triển khai tại khu vực kiểm soát giả lập với 02 camera IP Full HD, máy chủ AI trang bị GPU NVIDIA RTX 4060, trên bộ dữ liệu 120 người (20 cán bộ, 100 học viên), khoảng 1.200 ảnh đăng ký và 35 video thu thập trong 05 ngày. Kết quả đạt độ chính xác trung bình 96,8%, Điểm F1 96,2%, tốc độ xử lý 21 khung hình/giây, bước đầu khẳng định tính khả thi của giải pháp và là cơ sở để tiếp tục hoàn thiện trước khi xem xét triển khai toàn trường.

Từ khóa: nhận dạng khuôn mặt, mô hình thử nghiệm, kiểm soát ra vào thông minh, tích hợp hệ thống, trường đại học thông minh.

Abstract: Manual access control using magnetic cards and physical logbooks at People's Public Security educational institutions is increasingly revealing limitations regarding speed, accuracy, and data storage capacity. This paper proposes the architecture and development of a pilot real-time face recognition system for smart access control at the Fire and Rescue Academy, based on face detection, face recognition, and object tracking modules. The model was deployed in a simulated control area equipped with two Full HD IP cameras and an AI server powered by an NVIDIA RTX 4060 GPU. The dataset comprised 120 individuals (20 staff members and 100 cadets), consisting of approximately 1,200 registered images and 35 videos collected over a five-day period. The experimental results achieved an average accuracy of 96.8%, an F1-score of 96.2%, and a processing speed of 21 frames per second (FPS). These initial findings demonstrate the feasibility of the proposed solution and serve as a foundation for further refinement prior to full-scale campus deployment.

Keywords: face recognition, pilot model, smart access control, system integration, smart higher educational institution.

1. Đặt vấn đề

Các cơ sở đào tạo thuộc lực lượng Công an nhân dân, trong đó có Học viện PCCC&CNCH, có số lượng học viên, cán bộ đông đảo, thường xuyên ra vào khuôn viên phục vụ học tập, huấn luyện và công tác. Phương thức kiểm soát truyền thống bằng thẻ từ, sổ ghi chép hoặc kiểm tra trực tiếp của bảo vệ đang bộc lộ hạn chế về tốc độ xử lý, khả năng lưu vết lịch sử và nguy cơ giả mạo danh tính.

Sự phát triển của trí tuệ nhân tạo, đặc biệt là học sâu và thị giác máy tính, đã giúp công nghệ nhận dạng khuôn mặt đạt độ chính xác cao, xử lý theo thời gian thực, mở ra khả năng ứng dụng rộng rãi trong kiểm soát ra vào tại trường học, doanh nghiệp và khu công nghiệp [2], [5], [6]. Xuất phát từ yêu cầu thực tiễn về bảo đảm an ninh, an toàn tại Học viện PCCC&CNCH - nơi có đặc thù trang phục, phương tiện bảo hộ trong huấn luyện thực hành, việc nghiên cứu, thử nghiệm ứng dụng công nghệ này là cần thiết, làm cơ sở để học viện cân nhắc lộ trình triển khai trên quy mô rộng hơn.

Nhận dạng khuôn mặt là quá trình xác định hoặc xác thực danh tính dựa trên đặc điểm khuôn mặt, xây dựng trên nền tảng học sâu, cho độ chính xác vượt trội so với các phương pháp xử lý ảnh truyền thống [6]. Một hệ thống phục vụ kiểm soát ra vào thường gồm ba thành phần kỹ thuật cốt lõi sau.

Phát hiện khuôn mặt (Face Detection): xác định vị trí khuôn mặt trong ảnh hoặc khung hình video bằng khung giới hạn (bounding box), làm đầu vào cho các bước xử lý tiếp theo.

Nhận dạng khuôn mặt (Face Recognition): sử dụng mô hình học sâu (ví dụ InsightFace/ArcFace [2]) để chuyển khuôn mặt đã phát hiện thành vector đặc trưng (embedding), so sánh với dữ liệu đã đăng ký bằng độ tương đồng Cosine để xác định danh tính; Vector đặc trưng khuôn mặt không chứa trực tiếp ảnh gốc và được sử dụng để so sánh nhận dạng. Tuy nhiên, đây vẫn là dữ liệu sinh trắc học, cần được mã hóa, phân quyền truy cập và quản lý an toàn nhằm bảo đảm quyền riêng tư và an toàn thông tin.

Theo dõi đối tượng (Object Tracking): gán và duy trì mã định danh theo dõi (Track ID) cho từng người xuất hiện liên tục trên video, giúp hệ thống

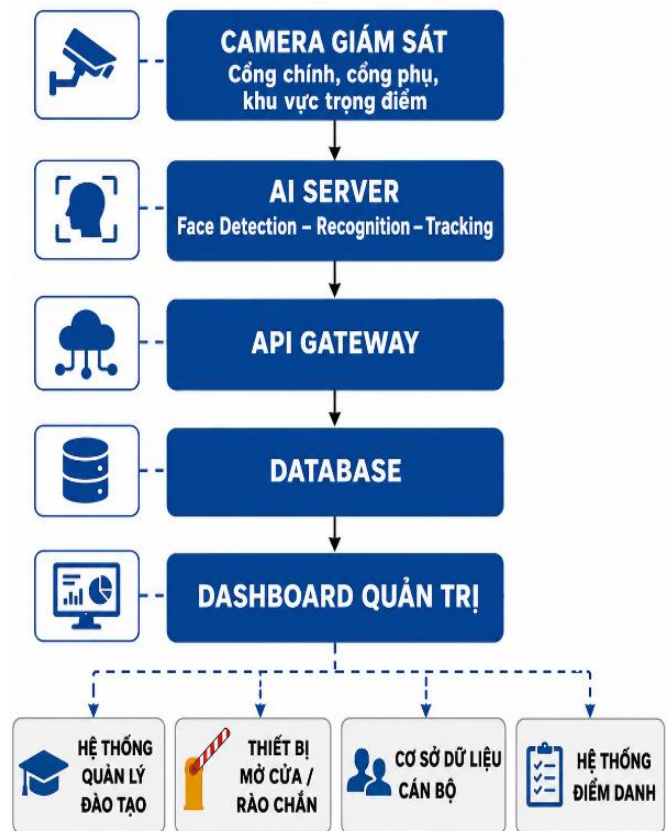
tránh nhận dạng lặp lại, giảm tải tính toán và bảo đảm xử lý thời gian thực khi lưu lượng ra vào lớn [7].

Mặc dù nhiều nghiên cứu đã đạt độ chính xác cao trong nhận dạng khuôn mặt, phần lớn tập trung vào cải tiến thuật toán hoặc đánh giá trên các bộ dữ liệu chuẩn. Các nghiên cứu về xây dựng hệ thống hoàn chỉnh phục vụ kiểm soát ra vào trong môi trường trường đại học, đặc biệt đối với cơ sở đào tạo thuộc lực lượng Công an nhân dân, vẫn còn hạn chế. Đây chính là khoảng trống mà nghiên cứu này hướng tới.

2. Đề xuất kiến trúc hệ thống và khả năng tích hợp

2.1. Kiến trúc hệ thống

Trên cơ sở ba thành phần kỹ thuật nêu trên, tác giả đề xuất kiến trúc hệ thống xử lý xuyên suốt từ thu nhận hình ảnh đến vận hành, quản lý nghiệp vụ, minh họa tại (hình 1).



Hình 1: Kiến trúc mô hình thử nghiệm hệ thống nhận dạng khuôn mặt.

Camera giám sát được bố trí tại cổng chính, cổng phụ và các khu vực cần kiểm soát, thu nhận hình ảnh liên tục và truyền về máy chủ AI (AI Server) - thành phần đảm nhiệm toàn bộ khối xử lý trí tuệ nhân tạo (phát hiện, nhận dạng, theo dõi), quyết định độ chính xác và tốc độ xử lý của hệ thống. Trong mô hình

thử nghiệm, mô-đun phát hiện khuôn mặt sử dụng thuật toán SCRFD để xác định vị trí khuôn mặt trên từng khung hình. Sau đó, mô-đun nhận dạng sử dụng ArcFace (InsightFace) để trích xuất vector đặc trưng khuôn mặt và so sánh với cơ sở dữ liệu bằng độ tương đồng cosine (Cosine Similarity) nhằm xác định danh tính. Đối tượng sau khi được nhận dạng sẽ được theo dõi bằng thuật toán ByteTrack, giúp duy trì Track ID giữa các khung hình liên tiếp, hạn chế việc nhận dạng lặp và giảm tải tính toán. API Gateway đóng vai trò trung gian tiếp nhận và điều phối kết quả nhận dạng. Thành phần này đồng thời chuẩn hóa giao tiếp giữa AI Server với các hệ thống nghiệp vụ khác, giúp tăng khả năng mở rộng và bảo đảm an toàn dữ liệu. Cơ sở dữ liệu (Database) lưu trữ vector đặc trưng khuôn mặt, hồ sơ cán bộ - học viên và lịch sử ra vào. Giao diện quản trị (Dashboard) cung cấp nền Web cho cán bộ trực ban giám sát, tra cứu, thống kê và cấu hình camera. Từ lớp Dashboard, hệ thống được thiết kế để kết nối tiếp với Hệ thống quản lý đào tạo, Thiết bị mở cửa/rào chắn, Cơ sở dữ liệu cán bộ và Hệ thống điểm danh.

2.2. Khả năng tích hợp hệ thống

Trên cơ sở kiến trúc đề xuất, hệ thống có thể tích hợp với một số hệ thống nghiệp vụ hiện có của học viện như:

- Phần mềm quản lý đào tạo - đồng bộ lịch học, lịch huấn luyện để đối chiếu dữ liệu ra vào;
- Cơ sở dữ liệu học viên, cán bộ - liên thông hồ sơ, ảnh đăng ký phục vụ xác thực, phân quyền theo vị trí công tác;
- Thư viện và ký túc xá - xác thực bằng khuôn mặt thay thẻ từ, quản lý giờ giấc sinh hoạt nội trú;
- Bãi xe thông minh - kết hợp nhận dạng khuôn mặt với nhận dạng phương tiện;
- Cổng thông tin điện tử và hệ thống camera an ninh sẵn có - tra cứu thông tin ra vào và tận dụng hạ tầng hiện hữu, giảm chi phí đầu tư mới.

Việc thiết kế Cổng giao tiếp API (API Gateway) như một lớp trung gian độc lập (Mục 2.1) là điều kiện kỹ thuật để hệ thống từng bước kết nối với các phần mềm, cơ sở dữ liệu nêu trên mà không phải thay đổi lõi xử lý AI, phù hợp với lộ trình chuyển đổi số theo từng giai đoạn của Học viện.

3. Mô hình thử nghiệm và kết quả đánh giá

3.1. Mô hình thử nghiệm

Để đánh giá tính khả thi, tác giả triển khai thử nghiệm tại khu vực kiểm soát giả lập, sử dụng 02 camera IP Full HD và máy chủ AI trang bị GPU NVIDIA RTX 4060. Cấu hình chi tiết được trình bày tại Bảng 1.

Bảng 1: Cấu hình mô hình thử nghiệm.

Thành phần	Cấu hình
Camera	02 x Camera IP Full HD 1080P
Máy chủ AI	Intel Core i7-13700, RAM 32 GB
Bộ xử lý đồ họa GPU	RTX 4060 8 GB
Hệ điều hành	Ubuntu Server 24.04
Thư viện học sâu	PyTorch [3]
Cơ sở dữ liệu (Database)	PostgreSQL
Khung phát triển Web (Web Framework)	FastAPI [4]
Bộ nhớ đệm (Cache)	Redis
Nền tảng đóng gói (Container)	Docker

Bộ dữ liệu thử nghiệm được thu thập trong 05 ngày, gồm 120 người (20 cán bộ, 100 học viên). Mỗi người được đăng ký khoảng 10 ảnh khuôn mặt, thu được khoảng 1.200 ảnh đăng ký phục vụ xây dựng cơ sở dữ liệu nhận dạng. Dữ liệu đánh giá gồm 35 video độc lập, không sử dụng các ảnh đăng ký, được ghi hình trong các điều kiện ánh sáng bình thường, ánh sáng yếu, người đeo kính và người đeo khẩu trang. Tổng số lượt nhận dạng được sử dụng để đánh giá là 6.800 lượt.

3.2. Kết quả đánh giá

Các chỉ tiêu đánh giá được tính trên toàn bộ dữ liệu kiểm tra. Độ chính xác trung bình (Average Accuracy) được xác định theo tỷ lệ số lượt nhận dạng đúng trên tổng số lượt nhận dạng. Chỉ tiêu Precision (Độ chính xác dự đoán), Recall (Độ bao phủ) và F1-

score (Điểm F1) được tính theo các công thức đánh giá phổ biến trong bài toán phân loại. Tốc độ xử lý 21 FPS được đo trên một luồng camera Full HD, sử dụng GPU NVIDIA RTX 4060. Sau thử nghiệm, mô hình đạt được các kết quả như Bảng 2.

Bảng 2: Kết quả đánh giá mô hình.

Chỉ tiêu	Kết quả
Độ chính xác trung bình (Average Accuracy)	96,8%
Độ chính xác dự đoán (Precision)	96,5%
Độ bao phủ (Recall)	95,9%
Điểm F1 (F1-score)	96,2%
Tốc độ xử lý (FPS)	21
Thời gian phản hồi	0,6 giây

Mô hình hoạt động ổn định trong điều kiện ánh sáng bình thường, đáp ứng yêu cầu xử lý thời gian thực ở quy mô thử nghiệm. Kết quả thử nghiệm cho thấy hệ thống đạt độ chính xác trung bình 96,8% trong điều kiện ánh sáng bình thường. Khi người sử dụng đeo kính, độ chính xác đạt khoảng 95,6%; khi đeo khẩu trang đạt khoảng 91,8%; trong điều kiện ánh sáng yếu đạt khoảng 90,9%. Như vậy thử nghiệm cho thấy mô hình đáp ứng yêu cầu nhận dạng thời gian thực trong phạm vi thử nghiệm. Sai số chủ yếu phát sinh trong điều kiện ánh sáng yếu và khi khuôn mặt bị che khuất một phần, phù hợp với đặc điểm chung của các hệ thống nhận dạng khuôn mặt dựa trên học sâu.

3.3. Hạn chế và định hướng phát triển

Do đây là nghiên cứu ở giai đoạn thử nghiệm, kết quả còn một số hạn chế: chưa đánh giá trên quy mô toàn trường; chưa thử nghiệm qua nhiều mùa, điều kiện thời tiết; chưa tích hợp cơ chế chống giả mạo khuôn mặt (Face Anti-Spoofing); chưa đánh giá khả năng nhận dạng bằng camera hồng ngoại trong điều kiện thiếu sáng hoàn toàn; và chưa đánh giá hoạt động khi lưu lượng ra vào đồng thời quá lớn, ví dụ giờ cao điểm. Những hạn chế này là cơ sở để xác định rõ phạm vi, mức độ tin cậy của mô hình và định hướng cho các nghiên cứu tiếp theo.

Trong thời gian tới, tác giả dự kiến hoàn thiện hệ thống theo các hướng: tích hợp cơ chế chống giả mạo khuôn mặt; mở rộng nhận dạng đa camera, xử lý đồng thời nhiều luồng video; nghiên cứu triển khai theo hướng trí tuệ nhân tạo tại thiết bị biên (Edge AI) để giảm tải máy chủ trung tâm; tích hợp chính thức với hệ thống quản lý đào tạo và cơ sở dữ liệu dùng chung của học viện; tích hợp mô-đun quản lý phương tiện cho bãi xe thông minh; và xây dựng chức năng điểm danh tự động dựa trên dữ liệu nhận dạng khuôn mặt. Việc triển khai theo lộ trình từng bước sẽ giúp học viện kiểm soát rủi ro, tối ưu chi phí đầu tư và nâng cao tính khả thi khi triển khai thực tế.

4. Kết luận

Trong bối cảnh chuyển đổi số và yêu cầu ngày càng cao đối với công tác bảo đảm an ninh, an toàn tại các cơ sở đào tạo của lực lượng Công an nhân dân, việc nghiên cứu, thử nghiệm ứng dụng công nghệ nhận dạng khuôn mặt thời gian thực vào kiểm soát ra vào là hướng đi phù hợp với xu thế xây dựng trường đại học thông minh [1].

Bài báo đã tổng quan các thành phần kỹ thuật cốt lõi của công nghệ nhận dạng khuôn mặt, trên cơ sở đó đề xuất kiến trúc hệ thống cùng khả năng tích hợp với nhiều hệ thống nghiệp vụ của Học viện PCCC&CNCH. Với bộ dữ liệu 120 người, khoảng 1.200 ảnh đăng ký và 35 video thu thập trong 5 ngày, mô hình thử nghiệm đạt các chỉ tiêu đáng ghi nhận (độ chính xác trung bình 96,8%, F1-score 96,2%, tốc độ xử lý 21 khung hình/giây), bước đầu khẳng định tính khả thi của giải pháp.

Bên cạnh kết quả đạt được, nghiên cứu còn một số hạn chế về quy mô và điều kiện thử nghiệm như đã trình bày ở Mục 3.3. Trên cơ sở đó, tác giả đề xuất lộ trình triển khai theo từng giai đoạn: hoàn thiện mô hình thử nghiệm; mở rộng dần quy mô dữ liệu và phạm vi lắp đặt; tích hợp với các hệ thống quản lý của học viện; và xem xét triển khai đồng bộ trên quy mô toàn trường. Việc tiếp tục nghiên cứu, hoàn thiện hệ thống sẽ góp phần nâng cao hiệu quả quản lý học viên, cán bộ, tăng cường an ninh, an toàn và thúc đẩy chuyển đổi số tại Học viện PCCC&CNCH. ■

TÀI LIỆU THAM KHẢO

1. Bộ Công an (2021), *Quyết định số 2612/QĐ-BCA ngày 20/4/2021 về Đề án chuyển đổi số trong Công an nhân dân đến năm 2025, định hướng đến năm 2030*, Hà Nội.
2. J. Deng, J. Guo, N. Xue, & S. Zafeiriou, 2019, *ArcFace: Additive Angular Margin Loss for Deep Face Recognition*. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), tr. 4690–4699.
3. A. Paszke, S. Gross, F. Massa, *et al.*, 2019, *PyTorch: An Imperative Style, High-Performance Deep Learning Library*. Advances in Neural Information Processing Systems (NeurIPS), tr. 32.
4. R. E. Bogachev, N. V. Zarikovskaya, 2021, *Overview of the capabilities of the FastAPI web framework for implementing the backend of web systems*, Proceedings of the All-Russian Scientific Conference on Achievements of Science and Technology. DOI: 10.47813/dnit.2021.2.289-293.
5. Y. Kortli, M. Jridi, A. Al Falou, & M. Atri, 2020, *Face Recognition Systems: A Survey*. Sensors, vol.20(2), tr. 342.
6. M. Wang, W. Deng, 2021, *Deep Face Recognition: A Survey*, Neurocomputing, vol.429, tr. 215-244.
7. Y. Zhang, P. Sun, Y. Jiang, *et al.*, 2022, *ByteTrack: Multi-Object Tracking by Associating Every Detection Box*, Proceedings of the European Conference on Computer Vision (ECCV).
8. J. Li, Y. Wang, C. Wang, *et al.*, 2021, *SCRFD: Sample and Computation Redistribution for Efficient Face Detection*. arXiv:2105.04714.