

ỨNG DỤNG DEEP LEARNING: NHẬN DIỆN KHUÔN MẶT ĐỂ XÁC MINH DANH TÍNH SINH VIÊN TRONG PHÒNG THI

APPLICATION OF DEEP LEARNING: FACE RECOGNITION FOR VERIFICATION OF STUDENT IDENTITY IN THE EXAM ROOM

Ngày nhận bài: 30/05/2022

Ngày chấp nhận đăng: 21/06/2022

*Nguyễn Thị Uyên Nhi[✉], Phạm Thị Thanh Hà, Nguyễn Ngọc Quỳnh Anh, Trần Thị Kim
Phú, Đỗ Nguyễn Minh Thư, Nguyễn Thị Phương Uyên*

TÓM TẮT

Nhận diện khuôn mặt là một trong những lĩnh vực quan trọng của thị giác máy tính, nhằm xác minh, định danh người dùng dựa vào hình ảnh hay video. Nhận diện khuôn mặt được ứng dụng trong nhiều lĩnh vực như hệ thống an ninh, hệ thống sinh trắc, điểm danh, chấm công, v.v. Nhiều kỹ thuật nhận diện khuôn mặt đã được nghiên cứu phát triển, trong đó các kỹ thuật học sâu cho độ chính xác vượt trội. Trong bài báo này, một mô hình ứng dụng mạng nơron tích chập CNN được đề xuất nhằm nhận diện khuôn mặt từ hình ảnh để xác minh danh tính sinh viên khi vào phòng thi. Đầu tiên, thuật toán MTCNN được sử dụng để phát hiện khuôn mặt và tiền xử lý dữ liệu, sau đó kết quả sẽ được đưa vào mô hình FaceNet, một mô hình dựa trên mạng CNN của Google, để trích xuất đặc trưng và sử dụng hàm mất mát Triplet để tối ưu hóa việc nhận diện. Bộ ảnh của các sinh viên (STUDUE) được thực hiện cho bài toán đặt ra. Thực nghiệm được thực hiện trên hai tập ảnh Yale và STUDUE cho độ chính xác lần lượt là 92,1% và 88,4%. Kết quả thực nghiệm được so sánh với các công trình nghiên cứu khác trên cùng một tập ảnh, cho thấy tính chính xác và hiệu quả của mô hình đề xuất.

Từ khóa: Nhận diện khuôn mặt; xác minh danh tính sinh viên; CNN, MTCNN; FaceNet; STUDUE.

ABSTRACT

Face recognition is one of the critical areas of computer vision, which aims to verify a person's identity based on images or videos. Face recognition is applied in many fields such as security systems, biometric systems, attendance, etc. Many face recognition techniques have been researched and developed, in which deep learning techniques give outstanding accuracy. This paper proposes a model based on Convolutional Neural Network (CNN) to recognize faces from images to verify student identity when entering the exam room. First, we use the MTCNN algorithm for face detection and data preprocessing. Then, the results will be fed into the FaceNet model, a Google model based on CNN, for feature extraction and use the Triplet loss function to optimize the recognition. The student image dataset (STUDUE) is built for this study. Experiments were performed on the Yale and STUDUE image dataset with the accuracy of 92.1% and 88.4%, respectively. The experimental results are compared with other studies on the same image dataset, showing the accuracy and efficiency of the proposed model.

Keywords: Face recognition; verify student identity; CNN; MTCNN; FaceNet; STUDUE.

1. Giới thiệu

Trong xu thế của cuộc cách mạng công nghiệp 4.0, các thiết bị điện tử như camera, smartphone, tablet, v.v, phát triển mạnh mẽ, khiến cho việc tạo ra ảnh số vô cùng đơn giản và trở nên quen thuộc với người dùng. Theo Oloyede và cộng sự (2020) sự phát triển của trí tuệ nhân tạo và các kỹ thuật học

máy, học sâu, nhận diện khuôn mặt trở thành hệ thống sinh trắc học phổ biến được sử

Nguyễn Thị Uyên Nhi, Phạm Thị Thanh Hà, Nguyễn Ngọc Quỳnh Anh, Trần Thị Kim Phú, Đỗ Nguyễn Minh Thư, Nguyễn Thị Phương Uyên, Khoa Thống kê - Tin học, Trường Đại học Kinh tế - Đại học Đà Nẵng

✉ Email: nhintu@due.edu.vn

dụng để xác định hoặc xác minh một người từ ảnh hay video như giám sát an ninh, xác định danh tính, điểm danh, chấm công, v.v. Hiện nay, có nhiều phương pháp nhận diện khuôn mặt như dựa trên đặc trưng cục bộ (Xiang & cộng sự, 2018; Wijaya & cộng sự, 2018), đặc trưng toàn cục (Zhang & cộng sự, 2019; Al-Dabagh & cộng sự, 2018) hay đặc trưng sâu từ mạng nơ-ron (Hansen & cộng sự, 2018; Ravi & cộng sự, 2020). Tuy nhiên, các phương pháp tiếp cận cục bộ hay toàn cục thường không nhạy cảm với các biến thể (nét mặt, biểu cảm, và tư thế, v.v.) hay thay đổi ánh sáng của môi trường (Ravi & cộng sự, 2020), do đó, hệ thống nhận diện khuôn mặt dựa trên mạng học sâu được phát triển và thu được kết quả đáng kinh ngạc. Mặc dù sử dụng học sâu cho độ chính xác rất cao, nhưng kích thước đặc trưng sâu thường rất lớn (hơn 1000 chiều) làm cho việc huấn luyện, suy luận và nhận diện phức tạp, mất nhiều thời gian (Wu và cộng sự, 2021). Vì vậy, nhiều công nghệ sử dụng học sâu để nhận diện khuôn mặt đã được phát triển, cải tiến. Vào năm 2015, Google đã đề xuất mô hình FaceNet (Schroff & cộng sự, 2015) dựa trên mạng nơ-ron tích chập CNN có độ chính xác nhận diện khuôn mặt rất cao. FaceNet trích xuất đặc trưng khuôn mặt thành một vec-tơ nhúng (embedding vector) nhỏ gọn với chỉ 128 chiều, sử dụng hàm mất mát Triplet loss (Ming & cộng sự, 2017), độ đo Euclid cho việc nhận diện nhanh và chính xác, nhưng vẫn đảm bảo được độ chính xác cao.

Trong xu hướng này, với sự phát triển của giáo dục tại Việt Nam, số lượng sinh viên tại các trường đại học ngày càng gia tăng, dẫn đến công tác quản lý khó khăn, phức tạp, đòi hỏi một hệ thống quản lý thông minh với ứng dụng của trí tuệ nhân tạo. Khi số lượng sinh viên lớn, việc xác minh chính xác được danh tính của sinh viên là một thử thách lớn, mất thời gian, tốn nhân lực. Do đó, bài toán xác

minh danh tính sinh viên khi vào phòng thi của các trường đại học là một bài toán cấp thiết, nhận được nhiều sự quan tâm từ các nhà quản lý. Hiện nay, việc xác minh danh tính sinh viên chủ yếu là dựa vào các loại giấy tờ như thẻ sinh viên, căn cước công dân, v.v. Việc xác minh này đơn giản, ít tốn kém nhưng khó kiểm tra khi sinh viên làm mất/quên giấy tờ, khó kiểm tra độ chính xác nếu sinh viên làm giả giấy tờ để đi thi hộ, mất thời gian, nhân lực để kiểm tra thông tin chính xác, v.v. Vì vậy, cần thiết một phương pháp cho bài toán xác minh danh tính sinh viên khi vào phòng thi tại trường đại học, đảm bảo nhanh chóng về thời gian, tránh gian lận và đạt được độ chính xác cao.

Từ những vấn đề được đặt ra, trong bài báo này chúng tôi đề xuất ứng dụng mô hình FaceNet để nhận diện khuôn mặt nhằm xác minh danh tính sinh viên khi vào phòng thi dựa trên tập ảnh STUDUE được xây dựng dành cho sinh viên trường Đại học Kinh tế. Hệ thống này có thể sử dụng hình ảnh sinh viên được chụp từ camera của smartphone, sau đó nhận diện khuôn mặt và xác minh sinh viên này là ai, có thuộc phòng thi hay không? Việc xác minh này nhanh chóng, chính xác và đảm bảo an ninh, tránh việc gian lận khi thi cử. Đóng góp chính của bài báo bao gồm: (1) Nghiên cứu Deep Learning và mô hình FaceNet cho bài toán nhận diện khuôn mặt nhằm xác minh danh tính sinh viên khi vào phòng thi; (2) Đề xuất mô hình nhằm nhận diện khuôn mặt sinh viên; (3) Xây dựng tập dữ liệu ảnh mới STUDUE của sinh viên trường Đại học kinh tế; (4) Thực nghiệm mô hình nhận diện khuôn mặt đã đề xuất dựa trên tập ảnh Yale và STUDUE.

Phần còn lại của bài báo như sau: trong phần 2, chúng tôi nghiên cứu các công trình liên quan về lĩnh vực nhận diện khuôn mặt; phần 3 trình bày cơ sở lý thuyết và các phương pháp nghiên cứu; phần 4 mô tả các thực nghiệm và một số kết quả cũng như

đánh giá cho mô hình đề xuất. Cuối cùng trong phần 5, chúng tôi trình bày tổng kết cho bài toán đã thực hiện trong bài báo và hướng phát triển tương lai của nghiên cứu.

2. Các công trình nghiên cứu liên quan

Nhận diện khuôn mặt là một vấn đề đầy thách thức trong lĩnh vực phân tích hình ảnh và thị giác máy tính (Oloyede & cộng sự, 2020). Việc bảo mật thông tin đang trở nên rất quan trọng và khó khăn, vì thế hệ thống sinh trắc nhằm đảm bảo an ninh với nhận diện khuôn mặt nhận được nhiều sự quan tâm của các nhóm nghiên cứu trong và ngoài nước.

Trong nước, các phương pháp nhận diện khuôn mặt được nghiên cứu phổ biến trong những năm gần đây. Nhóm nghiên cứu Hồng Quang và Doãn Thái Nguyên (2020) đề xuất phương pháp nhận diện khuôn mặt trong video bằng mạng nơ ron tích chập CNN, cho thấy phương pháp được đề xuất có độ chính xác vượt trội, có khả năng ứng dụng trong thực tiễn. Lê Song Toàn (2020) đã xây dựng hệ thống nhận diện khuôn mặt cho việc check in tại các sự kiện. Tác giả đề xuất phương pháp sử dụng HOG để trích xuất đặc trưng, phát hiện khuôn mặt với MTCNN. Ứng dụng nhận diện khuôn mặt được trích xuất hình ảnh từ webcam và gửi thông tin về sự kiện cho họ. Nhóm nghiên cứu Nguyễn Thanh Hải và cộng sự (2020) đề xuất thuật toán rừng ngẫu nhiên và Haar-Like để trích xuất đặc trưng và lưu trữ dữ liệu cho bài toán điểm danh sinh viên bằng nhận diện gương mặt, cho thấy độ chính xác cao và khả thi khi ứng dụng vào thực tế. Lê Thị Thu Nga và cộng sự (2020) đề xuất phương pháp kết hợp mạng MTCNN và hàm mất mát Triplet Loss nhằm điểm danh tự động. Đồng thời, nhóm tác giả đề xuất hương pháp căn chỉnh khuôn mặt để cho độ chính xác nhận diện cao là 80-95%, kể cả trong điều kiện không thuận lợi về ánh sáng, góc xoay, v.v. Các nghiên cứu

trong nước trong những năm gần đây cho thấy các kỹ thuật nhận diện khuôn mặt đang nhận được nhiều sự quan tâm, nhằm nâng cao hiệu quả nhận diện và ứng dụng trong thực tế.

Ngoài ra, lĩnh vực nhận diện khuôn mặt cũng nhận được rất nhiều sự quan tâm nghiên cứu của các nhà khoa học trên thế giới. Mustafa và cộng sự (2018) đã xây dựng một hệ thống nhận diện khuôn mặt dựa trên bộ phân tích Kernel Discriminant Analysis (KDA) kết hợp với thuật toán phân lớp SVM và k-NN. Nhóm tác giả đã thực nghiệm trên hai bộ dữ liệu Yale và ORL (Our Database of Faces) với độ chính xác lần lượt là 95.25% và 96%. Ali và cộng sự (2019) đã đề xuất các kỹ thuật học sâu (Deep Learning) kết hợp với hàm băm để nhận diện nhanh khuôn mặt thông qua một tập cơ sở dữ liệu lớn khuôn mặt và thực nghiệm trên bộ ảnh khuôn mặt LFW với mã băm có độ dài 64 của 48 mẫu truy vấn. Jose và cộng sự (2019) đề xuất phương pháp nhận diện khuôn mặt đa màn hình camera thông qua hệ thống giám sát dựa trên FaceNet và thuật toán MTNN, nhằm theo dõi đối tượng hoặc nghi phạm. Kết quả nhận diện đạt độ chính xác 97% cho thấy hiệu quả của FaceNet. Nhóm nghiên cứu Anitha (2020) đề xuất hệ thống nhận diện khuôn mặt để chấm công cho các công ty dựa vào thuật toán MTCNN để phát hiện khuôn mặt và mô hình FaceNet để nhận diện cá nhân. Kết quả của hệ thống là thiết thực, đáng tin cậy và loại bỏ sự xáo trộn và mất thời gian của hệ thống chấm công truyền thống.

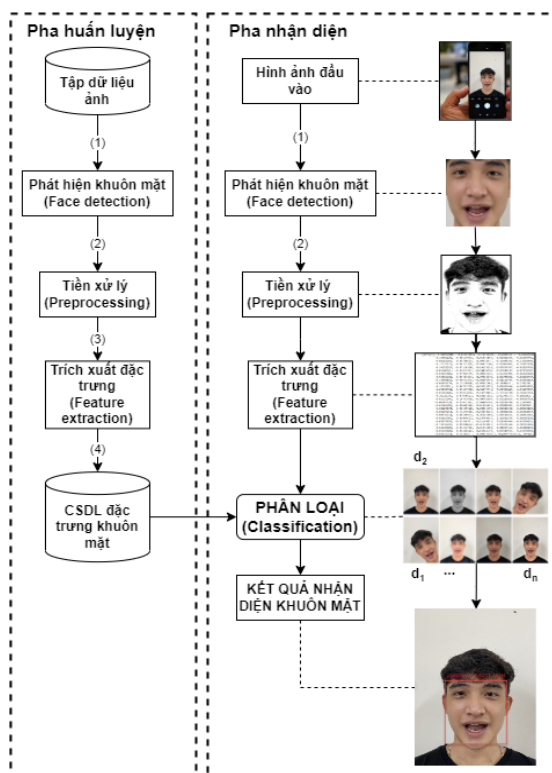
Từ các nghiên cứu trong và ngoài nước cho thấy tính khả thi, cấp thiết của việc nhận diện khuôn mặt trong các bài toán thực tế. Đồng thời, mô hình FaceNet với nhiều ưu điểm trong nhận diện khuôn mặt được ứng dụng trong nhiều lĩnh vực. Đây chính là động lực cho nhóm nghiên cứu phát triển phương pháp nhận diện khuôn mặt nhằm xác minh

danh tính sinh viên khi vào phòng thi dựa trên hình ảnh.

3. Cơ sở lý thuyết và các phương pháp nghiên cứu

3.1. Kiến trúc của hệ thống nhận diện khuôn mặt sinh viên

Trong phần này, kiến trúc của hệ thống nhận diện khuôn mặt (FACE_STUDUE) được thực hiện xây dựng theo hai pha, gồm pha huấn luyện và pha nhận diện, được trình bày trong **Hình 1**.



Hình 1: Kiến trúc của hệ thống nhận diện khuôn mặt

Pha huấn luyện bao gồm bốn giai đoạn như sau:

(1) Với mỗi hình ảnh đã được gán nhãn từ cơ sở dữ liệu, thực hiện quy trình phát hiện khuôn mặt từ thư viện TensorFlow của Multi-Task Cascaded Convolutions Neural (MTCNN) (Ku Hongchang, 2020);

(2) Chuẩn hóa và tiền xử lý hình ảnh khuôn mặt đã được phát hiện trước đó;

(3) Trích xuất đặc trưng khuôn mặt dựa vào mô hình FaceNet;

(4) Lưu trữ đặc trưng đã được trích xuất vào cơ sở dữ liệu đặc trưng khuôn mặt.

Pha nhận diện được thực hiện như sau:

(1) Với một hình ảnh đầu vào được chụp từ camera của smartphone, thực hiện thuật toán phát hiện khuôn mặt;

(2) Chuẩn hóa và tiền xử lý hình ảnh khuôn mặt đã được phát hiện trước đó;

(3) Trích xuất đặc trưng khuôn mặt dựa vào mô hình FaceNet;

(4) So sánh đặc trưng này với các đặc trưng trong cơ sở dữ liệu khuôn mặt đã được huấn luyện trước đó để phân loại;

(5) Cho ra kết quả nhận diện khuôn mặt để xác minh danh tính của sinh viên.

3.2. Tập dữ liệu ảnh

Các bộ dữ liệu được sử dụng thực nghiệm cho hệ thống nhận diện khuôn mặt sinh viên nhằm xác minh danh tính khi vào phòng thi (FACE_STUDUE) bao gồm tập dữ liệu Yale và tập dữ liệu STUDUE.

3.2.1. Tập dữ liệu ảnh Yale

Tập ảnh Yale (Yale Face Dataset original, 1995) được tạo bởi UDSC Computer Vision, Đại học Yale, New Haven, Hoa Kỳ. Bộ dữ liệu này có kích thước 6,4MB chứa 165 hình ảnh khuôn mặt của 15 người, mỗi người có 11 hình ảnh với nhiều góc độ với những trạng thái nét mặt khác nhau ở thang độ xám (Hình 2). Tập dữ liệu Yale là tập ảnh đã được gán nhãn cho từng ảnh, mỗi cá thể gồm nhiều ảnh được lưu trữ trong một thư mục được đặt định danh như subject01, subject02, v.v. Tập ảnh này tập trung vào các biểu cảm khuôn mặt và hướng ánh sáng như: Bình thường, Bất ngờ, Nháy mắt, Buồn ngủ, Buồn, Vui vẻ, Đeo kính, Không đeo kính, Đổ sáng bên trái, Đổ sáng bên phải, Đổ sáng trung tâm.



Hình 2: Minh họa tập dữ liệu ảnh Yale

3.2.2. Tập dữ liệu ảnh STUDUE

a) Thông tin chung

Bộ dữ liệu STUDUE là một bộ dữ liệu ảnh được nhóm nghiên cứu tự thu thập và xử lý từ các sinh viên trường Đại học Kinh Tế - Đại học Đà Nẵng. Tập dữ liệu 770 hình ảnh của 55 sinh viên đến từ các khoa khác nhau, mỗi sinh viên có 14 hình ảnh màu được gán nhãn các biểu cảm. Đồng thời, hình ảnh của mỗi sinh viên được lưu trữ trong cùng một thư mục ảnh, gán nhãn định danh theo cấu trúc: MASV_Họ tên_Lớp_Khoa.

Để tập dữ liệu ảnh có chất lượng tốt, điều kiện của các hình ảnh thu thập là: (1) được chụp từ camera sau của smartphone (độ phân giải từ 750 x 1334 pixels trở lên nhằm hạn chế việc mờ nét), (2) camera được đặt thẳng đứng; (3) đối tượng ở chính giữa bức ảnh, lấy nét tập trung vào giữa khuôn mặt, đủ sáng trung tâm, trong tư thế thẳng đứng; (4) khoảng cách từ đối tượng đến camera từ 1m-2m để đảm bảo ảnh rõ nét khuôn mặt.

b) Thu thập dữ liệu

Quá trình thu thập dữ liệu hình ảnh sinh viên bao gồm các bước:

- Bước 1 - Xác định đối tượng lấy mẫu: là sinh viên của trường Đại học Kinh Tế - Đại học Đà Nẵng.
- Bước 2 - Thiết lập danh sách các trạng thái biểu cảm khuôn mặt với 14 hình thái khác nhau bao gồm: bình thường, cười mỉm, vui, đeo kính, nháy mắt trái, nháy mắt phải, nhắm hai mắt, bất ngờ, tức giận, buồn, quay sang trái, quay sang phải, ngẩng mặt lên, cúi mặt xuống;

- Bước 3 - Tiến hành lấy mẫu ảnh bằng cách chụp hình biểu cảm khuôn mặt của đối tượng, mỗi biểu cảm chụp 3 bức hình. Xác định phong nền chung là màu trắng be (nền tường), ánh sáng ban ngày, tập trung vào chính giữa khuôn mặt. Đồng thời ghi lại các thông tin của đối tượng chụp như Mã sinh viên, họ tên, lớp, khoa;

- Bước 4 - Lưu trữ dữ liệu ảnh và các thông tin sinh viên đã thu.

c) Làm sạch và chuẩn hóa dữ liệu

Dữ liệu sau khi được thu thập sẽ bao gồm 2475 hình ảnh của 55 sinh viên. Đây là các dữ liệu thô chưa được xử lý. Để dữ liệu có thể sử dụng hiệu quả, quá trình làm sạch và chuẩn hóa dữ liệu được mô tả như sau:

- Bước 1 - Phân loại và lọc dữ liệu: Phân loại hình ảnh theo từng sinh viên, lưu trữ theo từng thư mục. Sau đó, thực hiện kiểm soát lỗi, làm sạch dữ liệu: loại bỏ các ảnh nhiễu, mờ, v.v. Với mỗi biểu cảm hay góc máy, giữ lại một hình ảnh tốt nhất.

- Bước 2 - Gán nhãn cho mỗi thư mục ảnh với Mã sinh viên, họ tên, lớp, khoa và mỗi hình ảnh theo từng biểu cảm.

- Bước 3 - Định dạng lại dữ liệu ảnh: Tất cả các hình ảnh được chuyển về theo định dạng *.jpeg, nền màu trắng be, kích thước 2000 (cao) x 1500 (rộng).



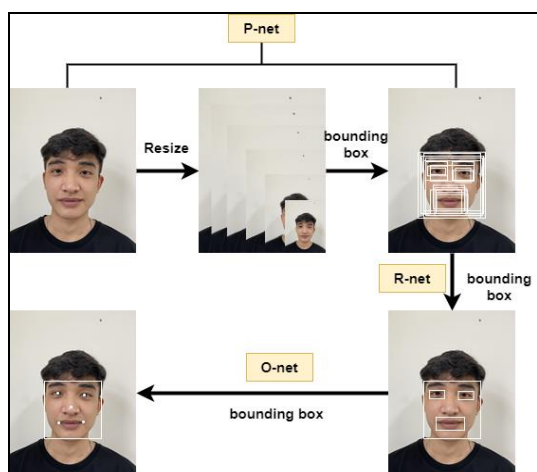
Hình 3: Minh họa tập dữ liệu ảnh STUDUE

Sau quá trình làm sạch và chuẩn hóa dữ liệu, tập ảnh STUDUE bao gồm 770 hình ảnh của 55 sinh viên, mỗi sinh viên có 14 hình ảnh với các biểu cảm khác nhau (Hình 3).

3.3. Phát hiện khuôn mặt và tăng cường dữ liệu ảnh với MTCNN

3.3.1. Phát hiện khuôn mặt

Với một ảnh đầu vào, đầu tiên, hệ thống phải thực hiện phát hiện khuôn mặt (Face detection) với MTCNN (Ku Hongchang, 2020), một mô hình mạng nơ-ron tích hợp CNN hoạt động đa nhiệm. MTCNN là gồm ba lớp Mạng đề xuất P-net (Proposal Network), Mạng tinh chỉnh R-net (Refine Network) và Mạng đầu ra O-net (Output Network). Hình 4 mô tả thuật toán MTCNN.



Hình 4: Cách thức hoạt động của MTCNN

Ban đầu các hình ảnh đầu vào được đưa vào P-Net để điều chỉnh kích thước để phát hiện các khuôn mặt có tất cả các kích thước khác nhau, đồng thời lấy ra các cửa sổ có thể là khuôn mặt và các vector hồi quy trong các cửa sổ đó. Sau đó, các cửa sổ này được sàng lọc thông qua mạng R-Net để loại bỏ phần lớn các cửa sổ không chứa khuôn mặt. Cuối cùng, Mạng đầu ra (O-Net) được sử dụng để chọn lọc kết quả chính xác một lần nữa và đánh dấu tọa độ của năm điểm mốc trên khuôn mặt.

3.3.2. Tăng cường dữ liệu ảnh

Bộ dữ liệu quy mô lớn là điều kiện cần thiết để huấn luyện thành công mạng nơ-ron. Công nghệ tăng cường hình ảnh sử dụng một loạt các thay đổi ngẫu nhiên đối với hình ảnh

đào tạo để tạo ra các mẫu đào tạo tương tự nhưng khác nhau, do đó mở rộng kích thước của tập dữ liệu. Để nâng cao độ chính xác của mô hình khi nhận diện, thư viện MTCNN được sử dụng để tăng cường hình ảnh, được mô tả trong hình 5, cụ thể như sau:

- Chuẩn hóa theo phân phối chuẩn các pixels của ảnh;
- Tạo các ảnh với các góc nghiêng là 20 độ (trái, phải);
- Dịch chuyển ảnh theo rộng;
- Dịch chuyển ảnh theo chiều cao;
- Lật ảnh theo chiều ngang;



Hình 5: Minh họa về tăng cường dữ liệu ảnh

Như vậy, với từ một hình ảnh khuôn mặt được phát hiện, chúng tôi tăng cường thêm 9 hình ảnh, tạo ra tập dữ liệu ảnh cho đầu vào mô hình FaceNet là 7700 ảnh. Mỗi lớp (thư mục ảnh) được chia thành 80-20 cho huấn luyện (training) và thử nghiệm (testing).

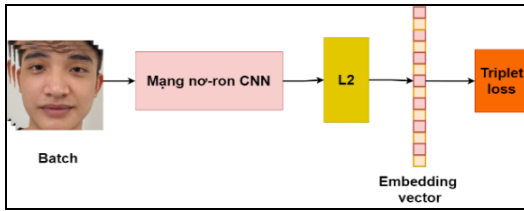
3.4. Ứng dụng mô hình FaceNet cho nhận diện khuôn mặt

Facenet là một hệ thống nhận diện khuôn mặt sử dụng mạng nơ-ron tích hợp CNN, được Google phát triển vào năm 2015. Hệ thống cải thiện hàm mất mát (Loss function) trong mạng nơ-ron, đề xuất hàm mất mát mới dựa trên đo độ tương tự Euclide, và sử dụng Triplet Loss làm hàm mất mát. FaceNet thực hiện trích xuất đặc trưng với vec-tơ embedding 128 chiều và sử dụng Triplet loss để đo lường sự khác biệt, từ đó nhận diện khuôn mặt nhanh chóng và chuẩn xác.

3.4.1. Trích xuất đặc trưng

FaceNet (Anitha G., 2020) sử dụng Mạng nơ-ron tích hợp (CNN) chuyển hình ảnh khuôn mặt của người vào không gian Euclide (tập hợp các điểm hình học) còn được gọi là

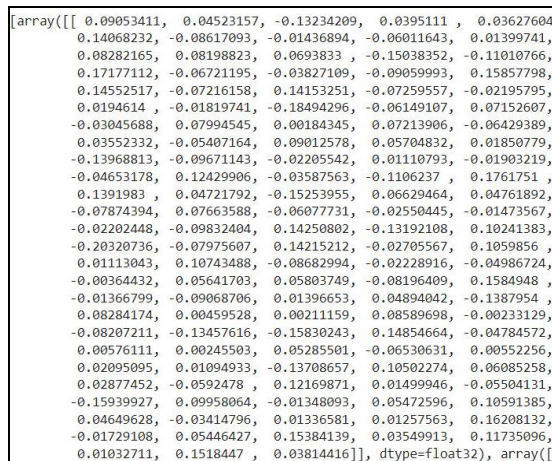
nhúng (embedding). Mô hình trích xuất đặc trưng của FaceNet được mô tả trong Hình 6.



Hình 6: Trích xuất đặc trưng của FaceNet

Với tập các hình ảnh đã được phát hiện khuôn mặt (Batch) sẽ đi vào kiến mạng nơ-ron tích chập, sau đó chuẩn hóa L2 và kết quả là các vec-tơ nhúng (embedding vector) 128 chiều cho các đặc trưng khuôn mặt, cuối cùng được đào tạo bằng cách sử dụng Triplet Loss để tạo embedding vector tốt nhất. Hình 7 là một minh họa cho embedding vector 128 chiều được thực hiện trong bài báo của chúng tôi.

Cấu trúc mạng CNN được sử dụng trong FaceNet là Inception V1 của Google (2014). Inception V1 là một mạng Siam network, loại bỏ đi lớp đầu ra và trích xuất ảnh thành một embedding vector nhỏ gọn, giúp mạng huấn luyện và suy luận nhận diện nhanh hơn.

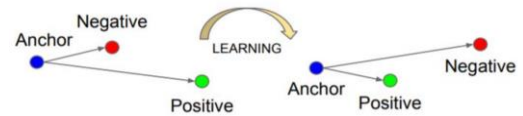


Hình 7: Minh họa embedding vector

3.4.2. Hàm Triplet loss

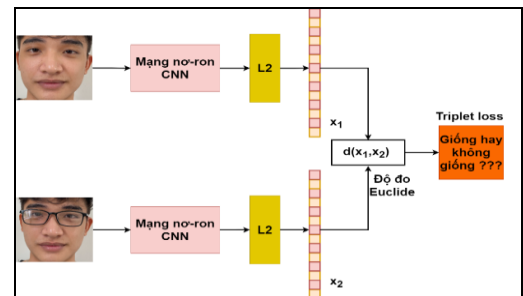
Hàm mất mát là hàm tính toán sự tương đồng hay khác biệt giữa hai hình ảnh dựa vào khoảng cách. Thông thường, trong một lần huấn luyện hàm mất mát chỉ tính được sự giống nhau của hai ảnh nếu nó cùng một lớp

hoặc khác nhau nếu khác lớp. Do đó việc huấn luyện sẽ mất rất nhiều thời gian. Mô hình FaceNet khắc phục vấn đề này khi sử dụng Triplet loss (Schroff, Florian, 2015) trong quá trình huấn luyện (Hình 8), với đầu vào là bộ ba ảnh: ảnh gốc (Anchor), ảnh giống gốc (Positive) và ảnh khác gốc (Negative). Mục tiêu của hàm Triplet loss là tối thiểu hóa khoảng cách giữa 2 ảnh khi chúng là Negative và tối đa hóa khoảng cách khi chúng là Positive.



Hình 8: Quá trình huấn luyện với Triple loss

Triplet loss giúp mô hình giảm thiểu việc nhận diện sai ảnh sai thành đúng, tạo ra các vec-tơ đặc trưng tốt nhất cho mỗi một ảnh. Hình 9 là một ví dụ minh họa về nhận diện khuôn mặt của mô hình FaceNet sử dụng Triplet loss.



Hình 9: Nhận diện khuôn mặt với Triplet loss

Trong hình 9, với hai hình ảnh cần so sánh, qua mô hình FaceNet để trích xuất embedding vector x_1 , x_2 . Nếu hai hình ảnh là cùng một người khoảng cách Euclidean $d(x_1, x_2)$ đạt min, ngược lại nếu khác nhau thì $d(x_1, x_2)$ đạt max.

4. Thực nghiệm và đánh giá kết quả

4.1. Môi trường thực nghiệm

Hệ thống nhận diện khuôn mặt nhằm xác minh danh tính sinh viên khi vào phòng thi được đặt tên là FACE_STUDUE. Hệ thống này áp dụng hai mô hình MTCNN và

FaceNet dựa trên tập ảnh dữ liệu STUDUE để thực hiện nhận dạng khuôn mặt sinh viên trong phòng thi. Thực nghiệm được thực hiện trên ngôn ngữ lập trình Python, OpenCV và thư viện TensorFlow. Cấu hình máy tính của thực nghiệm: Intel(R) Core (TM) i5-7200U, CPU 2,70GHz, RAM 12GB, SSD 232GB, HDD 465GB và hệ điều hành Windows 10 Professional.

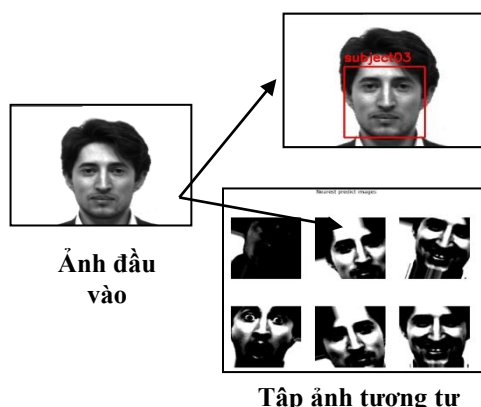
Tập dữ liệu ảnh thực nghiệm là tập ảnh phổ biến Yale và tập ảnh STUDUE, được mô tả cụ thể trong Bảng 1. Tập dữ liệu ảnh được chia 80%-20% cho pha huấn luyện (train) và pha thực nghiệm (test).

Bảng 1. Thông tin các bộ ảnh thực nghiệm

Bộ ảnh	Số ảnh	Số phân lớp	Độ lớn
Yale	165	15	11.2 MB
STUDUE	770	55	496MB

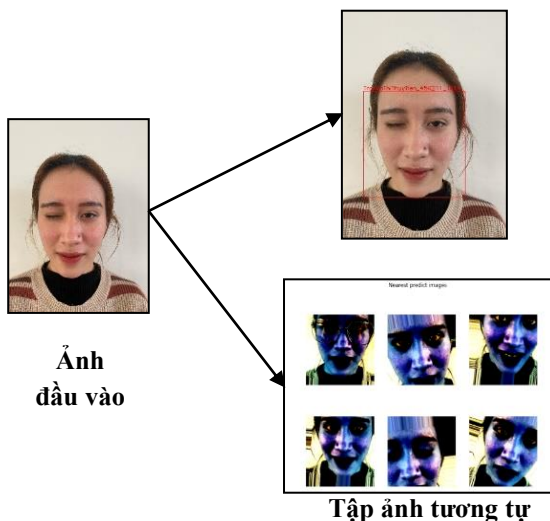
4.2. Thực nghiệm

Với một ảnh đầu vào trong tập ảnh test, hệ thống FACE_STUDUE cho ra một dự báo để nhận diện hình ảnh, đây là những thông tin cơ bản nhân của thư mục có chứa hình ảnh tương tự của ảnh đầu vào. Mỗi dự báo nhận diện có thể đúng hoặc sai, từ đó tính toán được Accuracy của tập ảnh test. Hình 10 là một kết quả thực nghiệm của hệ thống FACE_STUDUE với Yale, cho kết quả nhận diện là khuôn mặt thuộc Subject03 cùng tập ảnh tương tự với ảnh đầu vào.



Hình 10: Một kết quả thực nghiệm của hệ thống FACE_STUDUE trên tập ảnh Yale

Hình 11 là một kết quả thực nghiệm của hệ thống FACE_STUDUE với tập ảnh STUDUE, cho kết quả nhận diện là sinh viên 191121521134_TranVoThuyTien_45K21.1_TKTH.



Hình 11: Một kết quả thực nghiệm của hệ thống FACE_STUDUE trên tập ảnh STUDUE

Với thực nghiệm này, giám thị coi thi có thể xác minh chính xác và nhanh chóng danh tính, thông tin của sinh viên như mã sinh viên, họ tên, lớp, từ đó kiểm tra trong danh sách coi thi để cho sinh viên vào phòng thi.

4.3. Kết quả và đánh giá

4.3.1. Kết quả

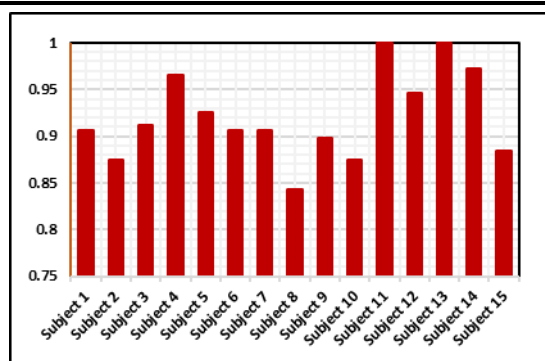
Kết quả thực nghiệm nhận diện khuôn mặt trên bộ ảnh Yale được thể hiện trong Bảng 2 và hình 12, với độ chính xác trung bình theo từng thư mục ảnh (subject), với thời gian nhận diện trung bình là 87ms. Bảng 3 và Hình 13 là kết quả thực nghiệm nhận diện khuôn mặt trên bộ ảnh STUDUE với độ chính xác trung bình theo từng nhóm ảnh, mỗi nhóm gồm 11 sinh viên, được chia ngẫu nhiên. Thời gian nhận diện trung bình của bộ ảnh STUDUE là 156ms.

Bảng 2. Kết quả nhận diện trên bộ ảnh Yale

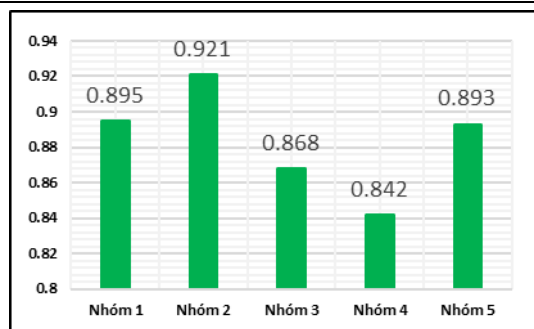
Thư mục	Số ảnh	AVG. Accuracy
subject01	2	0.906
subject02	3	0.875
subject03	3	0.912
subject04	2	0.965
subject05	3	0.926
subject06	1	0.906
subject07	2	0.906
subject08	2	0.843
subject09	2	0.898
subject10	3	0.875
subject11	2	1.00
subject12	3	0.946
subject13	2	1.00
subject14	2	0.973
subject15	1	0.884
Trung bình	33	0.921

Bảng 3. Kết quả nhận diện trên bộ ảnh STUDUE

Thư mục	Số ảnh	AVG. Accuracy
Nhóm 1	34	0.895
Nhóm 2	26	0.921
Nhóm 3	30	0.868
Nhóm 4	32	0.842
Nhóm 5	32	0.893
Trung bình	154	0.884



Hình 12: Độ chính xác nhận diện trên bộ Yale theo từng thư mục



Hình 13: Độ chính xác nhận diện trên bộ STUDUE theo từng nhóm sinh viên

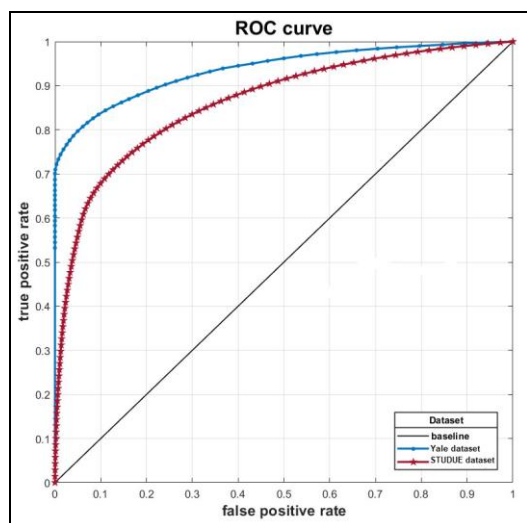
4.3.2. Đánh giá

Từ kết quả trong Bảng 2, Bảng 3, bộ Yale đạt được độ chính xác cao hơn so với bộ STUDUE, do bộ Yale là bộ ảnh phổ biến, được tiền xử lý chuẩn xác hơn và số lượng ảnh cũng ít hơn. Dựa trên các số liệu thực nghiệm, các đồ thị được thực hiện để đánh giá hiệu suất của hệ thống FACE_STUDUE. Hình 12 cho thấy độ chính xác trung bình nhận diện trên bộ Yale là cao, đều lớn hơn 80%, có những subject có độ nhận diện chính xác là 100%. Tuy nhiên, tùy thuộc vào đặc điểm của từng thư mục ảnh, đặc điểm của từng khuôn mặt mà độ chính xác trung bình có thể khác nhau. Hình 13 cho thấy độ chính xác trung bình nhận diện trên bộ STUDUE là cao, đều lớn hơn 84%, cao nhất là nhóm 2 với 92.1%.

Ngoài ra, trong nhiệm vụ nhận diện khuôn mặt, để xác định xem hệ thống có phân loại chính xác hay không thì đường cong ROC (Receiver Operating Characteristic) được thực hiện để đánh giá. Đường cong ROC đại diện cho tỷ lệ dương tính giả (FPR- False Positive Rate), và tỷ lệ dương tính thực (TPR - True Positive Rate).

Hình 14 là đồ thị đường cong ROC của tập ảnh Yale và STUDUE. Dựa vào Hình 14 có thể thấy các điểm trên ROC curve đều nằm trên đường baseline và gần với điểm có tọa độ (0, 1) trên đồ thị (góc trên bên trái) nên hiệu suất phân loại của mô

hình là hiệu quả. ROC của Yale gần tọa độ (0,1) hơn, AUC cũng lớn hơn so với STUDUE, chứng tỏ độ chính xác nhận diện trên tập Yale tốt hơn.



Hình 14: Đồ thị ROC của tập ảnh Yale và STUDUE

Để đánh giá độ chính xác và hiệu quả của hệ thống FACE_STUDUE, chúng tôi so sánh độ chính xác thu được từ thực nghiệm với các phương pháp của các công trình nghiên cứu khác trên cùng tập dữ liệu ảnh trong Bảng 4.

Bảng 4. So sánh kết quả thực nghiệm bộ ảnh Yale với các phương pháp khác

Phương pháp	Accuracy
Yee và cộng sự (2019) [21]	85.13%
Ravi và cộng sự (2020) [22]	74.4%
FACE_STUDUE	92.1%

Các phương pháp khác được dùng để so sánh trên cùng tập ảnh Yale bao gồm: (1) Nhóm nghiên cứu Yee và cộng sự (2019) đề xuất phương pháp nhận diện khuôn mặt sử dụng trích xuất đặc trưng cục bộ với Laplacian và thực nghiệm trên bộ Yale; (2) Ravi và cộng sự (2020) sử dụng LBP để trích xuất đặc trưng cục bộ và SVM cho việc phân loại, nhận diện khuôn mặt. Kết quả trong

Bảng 4 cho thấy, phương pháp đề xuất trong bài báo cho độ chính xác vượt trội hơn các phương pháp khác. Điều này cho thấy việc sử dụng học sâu với mạng CNN và triplet loss của FaceNet cho khả năng nhận diện khuôn mặt hiệu quả. Vì vậy, ứng dụng mô hình FaceNet vào việc nhận diện khuôn mặt sinh viên cho bài toán xác minh danh tính vào phòng thi là khả thi và hiệu quả.

5. Kết luận

Trong bài báo này, một mô hình nhận diện khuôn mặt sinh viên nhằm xác minh danh tính sinh viên vào phòng thi dựa trên mô hình Facenet đã được trình bày và triển khai. Trong đó, với hình ảnh đầu vào, thuật toán MTCNN được sử dụng để phát hiện khuôn mặt và tiền xử lý dữ liệu, sau đó được đưa vào FaceNet để trích xuất đặc trưng với embedding vector 128 chiều và hàm mất mát Triplet loss để nhận chọn vector đặc trưng tốt nhất. FaceNet giúp việc huấn luyện và suy luận các dự báo nhận diện nhanh chóng và độ chính xác cao. Để thực hiện bài toán đề ra, chúng tôi tiến hành xây dựng tập dữ liệu ảnh sinh viên STUDUE và thực nghiệm trên tập ảnh này cùng tập ảnh Yale, với độ chính xác nhận diện lần lượt là 88,4% và 92,1%. So sánh với các kết quả từ các nghiên cứu khác trên cùng tập ảnh Yale cho thấy mô hình đề xuất của chúng tôi cho độ chính xác vượt trội.

Trong các nghiên cứu tương lai, chúng tôi tiếp tục nghiên cứu các phương pháp, thuật toán nhằm cải thiện tốt hơn quá trình nhận diện, bổ sung tập dữ liệu ảnh và xây dựng một hệ thống xác minh danh tính sinh viên, chống giả mạo khi vào phòng thi, kết hợp với các hệ thống khác của trường đại học để tìm kiếm nhanh chóng thông tin sinh viên, giảm giấy tờ, nhân công như kết hợp với hệ thống thư viện, hệ thống đào tạo, khảo thí, v.v.

Lời cảm ơn:

Nghiên cứu này là một phần của đề tài NCKH cấp cơ sở do Trường Đại học Kinh tế - ĐHQĐN tài trợ với mã số đề tài T2022-04-21.

TÀI LIỆU THAM KHẢO

- Al Kobaisi, A., & Wocjan, P. (2019). *MaxHash for Fast Face Recognition and Retrieval*. International Conference on Computational Science and Computational Intelligence (CSCI), 652-656.
- Al-Dabagh, M. Z. N., Alhabib, M. M., & Al-Mukhtar, F. H. (2018). Face recognition system based on kernel discriminant analysis, k-nearest neighbor and support vector machine. *International Journal of Research and Engineering*, 5(3), 335-338.
- Anitha, G., Devi, P. S., Sri, J. V., & Priyanka, D. (2020). Face Recognition Based Attendance System Using Mtcnn and Facenet. *Zeichen Journal.*, 6(1), 189-195.
- Hansen, M. F., Smith, M. L., Smith, L. N., Salter, M. G., Baxter, E. M., Farish, M., & Grieve, B. (2018). Towards on-farm pig face recognition using convolutional neural networks. *Computers in Industry*, 98, 145-152.
- Jose, E., Greeshma, M., Haridas, M. T., & Supriya, M. H. (2019, March). *Face recognition based surveillance system using facenet and mtcnn on jetson tx2*. 5th International Conference on Advanced Computing & Communication Systems (ICACCS).
- Ku, H., & Dong, W. (2020). Face recognition based on mtcnn and convolutional neural network. *Frontiers in Signal Processing*, 4(1), 37-42.
- Lê, T. T. N., Nguyễn, V. C., & Nguyễn, X. P. (2020). *Điểm danh tự động dựa trên mô hình mạng Nơ-Ron tích chập xếp tầng đa nhiệm và kỹ thuật Triplet Loss*. Hội thảo khoa học quốc gia (CITA), 219-226.
- Lê, S. T. (2020). *Xây dựng hệ thống quản lý ảnh và check in sự kiện bằng nhận diện khuôn mặt*. Hội thảo khoa học quốc gia (CITA), 196-203.
- Ming, Z., Chazalon, J., Luqman, M. M., Visani, M., & Burie, J. C. (2017, October). *Simple triplet loss based on intra/inter-class metric learning for face verification*. International Conference on Computer Vision Workshops (ICCVW), 1656-1664.
- Nguyễn, T. H., Trịnh, T. T. L., Trần, B. T., Phan, K. Y. N., Trần, T. Đ., & Nguyễn, T. N. (2020). *Giải pháp điểm danh sinh viên bằng nhận diện gương mặt với đặc trưng Haar-Like kết hợp thuật toán rừng ngẫu nhiên*. Hội thảo khoa học quốc gia (CITA), 179-186.
- Oloyede, M. O., Hancke, G. P., & Myburgh, H. C. (2020). A review on face recognition systems: recent approaches and challenges. *Multimedia Tools and Applications*, 79(37), 27891-27922.
- Quang, H., & Lê Hồng Minh, T. D. N. (2020). Nhận dạng khuôn mặt trong video bằng mạng nơ ron tích chập. *Bản B của Tạp chí Khoa học và Công nghệ Việt Nam*, 62(1), 8-12.
- Ravi, R., & Yadhukrishna, S. V. (2020, March). *A face expression recognition using CNN & LBP*. Fourth International Conference on Computing Methodologies and Communication (ICCMC).

- Schroff, F., Kalenichenko, D., & Philbin, J. (2015). *Facenet: A unified embedding for face recognition and clustering*. In Proceedings of the IEEE conference on computer vision and pattern recognition.
- Wijaya, I. G. P. S., Husodo, A. Y., & Arimbawa, I. W. A. (2018). Real time face recognition based on face descriptor and its application. *Telkomnika*, 16(2), 739-746.
- Wu, C., & Zhang, Y. (2021). Mtcnn and facenet based access control system for face detection and recognition. *Automatic Control and Computer Sciences*, 55(1), 102-112.
- Xiang, Z., Tan, H., & Ye, W. (2018). The excellent properties of a dense grid-based HOG feature on face recognition compared to Gabor and LBP. *IEEE Access*, 6, 29306-29319.
- Yale Face Dataset original, from <http://vision.ucsd.edu/content/yale-face-database>
- Yee, S. Y., Rassem, T. H., Mohammed, M. F., & Awang, S. (2020). *Face recognition using Laplacian completed local ternary pattern (LapCLTP)*. In Advances in electronics engineering.
- Zhang, Y., Xiao, X., Yang, L. X., Xiang, Y., & Zhong, S. (2019). Secure and efficient outsourcing of PCA-based face recognition. *IEEE Transactions on Information Forensics and Security*, 15, 1683-1695.