Điểm Danh Tự Động Dựa Trên Mô Hình Mạng Nơ-Ron Tích Chập Xếp Tầng Đa Nhiệm Và Kỹ Thuật Triplet Loss

Automatic Attendance based on Multi-Task Cascaded Convolutional Neural Network Model and Triplet Loss Technique

Lê Thị Thu Nga[1], Nguyễn Văn Châu[2], Nguyễn Xuân Pha[3]

[1,2,3] Trường Đại học Công nghệ thông tin và Truyền thông Việt - Hàn,
Đại học Đà Nẵng, Việt Nam
{lttnga, nvchau.17it3, nxpha}@vku.udn.vn

Tóm tắt. Mạng nơ-ron tích chập xếp tầng đa nhiệm MTCNN (Multi-Task Cascaded Convolutional Neural Networks) là mô hình học sâu hiện đại, cho phép xác định khuôn mặt ở nhiều góc nghiêng khác nhau, ngay cả trong trong điều kiện thiếu sáng và một phần khuôn mặt bị che khuất. Bài báo này, chúng tôi đề xuất giải pháp điểm danh tự động bằng cách sử dụng mô hình MTCNN nhằm xác định vị trí khuôn mặt, đồng thời kết hợp kỹ thuật Triplet Loss để nhận diện danh tính đối tượng điểm danh. Kỹ thuật căn chỉnh khuôn mặt cũng được áp dụng nhằm gia tăng độ chính xác của nhận diện. Thực nghiệm cho thấy với sự kết hợp mô hình và các kỹ thuật này, tỉ lệ nhận diện đạt 80-95%, kể cả trong điều kiện thiếu sáng, góc nghiêng hay một phần khuôn mặt bị che khuất.

Từ khóa: điểm danh tự động, nhận diện khuôn mặt, mạng nơ-ron tích chập.

Abstract. The Multi-Task Cascaded Convolutional Neural Networks (MTCNN) is a modern deep learning model that allows faces identified at many different views, even in low light and part of the face is obscured. This article proposes the solution of automatic attendance by using the MTCNN model to determine faces and the Triplet Loss technique to identify objects. A face alignment technique has also been applied to increase the accuracy of recognition. The experiment shows that with the combination of the MTCNN model and the Triplet Loss technique, the recognition rate reaches 80-95% even in low light conditions, view or part of the face is obscured.

Keywords: Automatic attendance, face recognition, neural network.

1 Giới thiệu

Điểm danh là công việc thường xuyên, hằng ngày tại các trường học, cơ quan, nhà máy. Tuy nhiên, hầu hết các đơn vị này vẫn đang thực hiện bằng tay hoặc bán tự động thông qua việc quan sát và ghi nhận sự có mặt của người học, nhân viên, công nhân,… Với phương pháp thủ công truyền thống, việc điểm danh bằng trực giác thường tốn nhiều thời gian, không tránh khỏi mạo danh, sai sót và đôi khi mang lại sự khó chịu cho những người tham gia điểm danh. Ngày nay, với sự phát triển vượt bậc của trí tuệ nhân tạo (AI - Artificial intelligence), các hệ thống điểm danh tự động bằng vân tay, mống mắt hay khuôn mặt ngày càng hoàn thiện và đang dần được đưa vào sử dụng, giúp quá trình quản lý nhân sự dễ dàng, nhanh chóng và chính xác hơn; đồng thời mang lại sự thoải mái và nâng cao hiệu quả trong công tác quản lý con người.

Mỗi khuôn mặt đều có nhiều đặc trưng, những phần lồi lõm tạo nên các điểm nút của khuôn mặt. Công nghệ nhận diện khuôn mặt có khả năng xác định, xác nhận một người từ hình ảnh kỹ thuật số hoặc từ một khung hình trong video. Đây là phương pháp xác minh danh tính một người dựa vào những đặc trưng trên khuôn mặt của người đó, phân biệt được ngay cả với các trường hợp song sinh [1,2]. Do đó, ngoài việc

ứng dụng trong điểm danh để quản lý nhân sự, nhận diện khuôn mặt còn là sự lựa chọn trong các lĩnh vực an ninh, bảo mật, giao dịch.

Có nhiều phương pháp phát hiện khuôn mặt. Phương pháp Haar-like Adaboost (HA) xác định khuôn mặt dựa trên sự kết hợp của 4 thành phần: Haar-like xác định đặc trưng, Integral Image tính toán các đặc trưng, bộ lọc Adaptive Boost và Cascade nhằm tăng tốc độ phân loại [3]. Phương pháp này cho kết quả nhận diện nhanh nhưng dễ bị ảnh hưởng bởi ánh sáng môi trường xung quanh và chỉ phù hợp với khuôn mặt ở góc chính diện [3,4]. Phương pháp Histogram of Oriented Gradians (HOG) tuy ít bị ảnh hưởng bởi ánh sáng môi trường nhưng cho kết quả không tốt đối với một phần khuôn mặt bị phủ lấp [5]. Tiếp cận Deformable Part Models (DPM), một dạng của mô hình Markov ẩn, cũng đã thu được hiệu suất vượt trội, tuy nhiên mô hình này yêu cầu chi phí tính toán cao, đặc biệt là trong giai đoạn huấn luyện [6].

Mạng nơ-ron tích chập CNN (Convolutional Neural Networks) là mô hình học sâu (Deep Learning) hiệu quả, được dùng trong nhiều bài toán phát hiện và nhận diện khuôn mặt, phân tích video, ảnh MRI,…Hầu hết các CNN đều thích hợp và giải quyết tốt các bài toán dạng này [7-9]. Mạng MTCNN được phát triển từ CNN [10]. Mô hình này cho phép xác định khuôn mặt ở nhiều góc độ khác nhau, ít bị ảnh hưởng bởi ánh sáng của môi trường xung quanh và nhận diện ngay cả trong trường hợp một phần khuôn khuôn mặt bị che khuất [10-12].

Trong bài báo này, chúng tôi đề xuất giải pháp điểm danh tự động bằng cách sử dụng mô hình MTCNN nhằm xác định vị trí khuôn mặt, đồng thời kết hợp kỹ thuật Triplet Loss để nhận diện danh tính đối tượng điểm danh, kể cả trong điều kiện thiếu sáng, góc nghiêng hay một phần khuôn mặt bị che khuất. Kỹ thuật căn chỉnh khuôn mặt cũng được áp dụng nhằm gia tăng độ chính xác của kết quả nhận diện. Việc điểm danh dựa trên nhận diện khuôn mặt được thực hiện hoàn toàn tự động thông qua hình ảnh nhận được từ camera. Kết quả nhận diện được xuất trực tiếp lên website của hệ thống quản lý sinh viên.

Phần còn lại của bài báo bao gồm: Phần 2 trình bày mạng CNN, mô hình MTCNN và kỹ thuật Triplet Loss được sử dụng trong hệ thống điểm danh nhằm phát hiện và nhận diện khuôn mặt; Phần 3 đề xuất giải pháp và mô hình hệ thống; kết quả thực nghiệm cũng được đưa ra trong Phần 4; và cuối cùng, Phần 5 là kết luận của bài báo.

2 Phái hiện và nhận diện khuôn mặt

2.1 Mạng nơ-ron tích chập

Mô hình học sâu của mạng nơ-ron tích chập CNN rất thích hợp cho các bài toán với dữ liệu là ảnh hoặc video số. Bốn loại lớp chính trong CNN gồm (Fig.1)[13]:

- **Lớp tích chập** (Convolutional layer): có chức năng phát hiện các đặc trưng như góc, cạnh, màu sắc, texture,… của đối tượng thông qua các bộ lọc.
- **Lớp kích hoạt phi tuyến** (Nonlinear layer): đặt sau lớp tích chập nhằm đảm bảo tính phi tuyến của mô hình huấn luyện. Hàm PReLU thường được chọn dùng vì tính toán nhanh, đơn giản, hạn chế tình trạng triệt tiêu gradient và cho kết quả tốt.
- **Lớp co** (Pooling layer): nằm sau lớp kích hoạt phi tuyến, nhằm giảm kích thước của ảnh đầu ra nhưng vẫn giữ được thông tin quan trọng của ảnh vào, giảm thời gian huấn luyện. Các lớp Pooling thường dùng là Max-pooling và Average-pooling.
- **Lớp kết nối đầy đủ** (Fully connected layer): tương tự mạng nơ-ron truyền thống, lớp này chuyển ma trận đặc trưng ở lớp co thành vector chứa xác suất của các đối tượng cần được dự đoán.

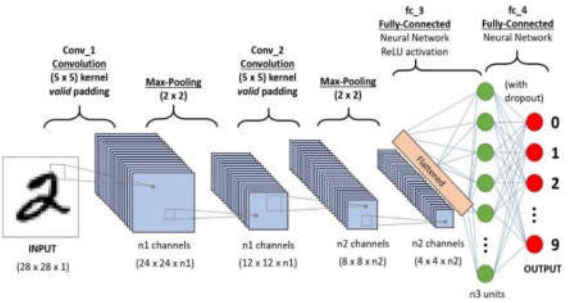


Fig. 1. Mô hình mạng nơ-ron tích chập (CNN).

2.2 Phát hiện khuôn mặt dùng MTCNN

Mô hình MTCNN được sử dụng để phát hiện khuôn mặt. Mô hình này hoạt động theo ba bước, mỗi bước dùng một CNN riêng, các CNN này lần lượt là: P-Net, R-Net và O-Net (Fig.2). Mỗi CNN có cấu trúc khác nhau, đảm nhiệm vai trò khác nhau trong mô hình. Đầu ra của MTCNN là vị trí khuôn mặt và các điểm nút trên khuôn mặt như: mắt, mũi, miệng. Cấu trúc cụ thể và chức năng của các CNN trong MTCNN như sau [10,11] (Fig.2): P-Net (Proposal Network) là CNN với 3 lớp tích chập và 1 lớp co, nhằm xác định vùng chứa khuôn mặt; R-Net (Refine Network) là CNN với 3 lớp tích chập, 2 lớp co và 1 lớp kết nối đầy đủ. R-Net sử dụng đầu ra của P-Net để loại bỏ các vùng không phải khuôn mặt; O-Net (Output Network) là CNN với 4 lớp tích chập, 3 lớp co và 1 lớp kết nối đầy đủ. O-Net sử dụng đầu ra của R-Net để đưa ra kết quả cuối cùng với 5 điểm nút trên khuôn mặt, bao gồm: 2 điểm mắt, 1 điểm mũi và 2 điểm khóe miệng.

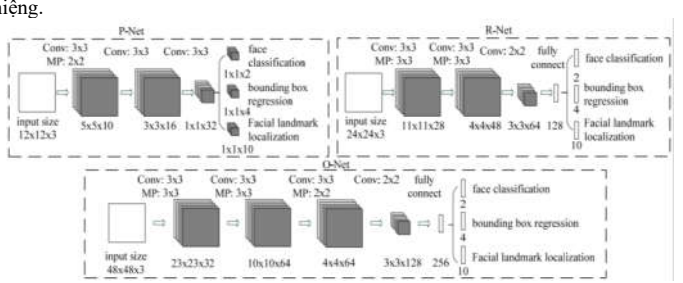


Fig. 2. Kiến trúc P-Net, R-Net và O-Net trong mô hình MTCNN [10].

2.3 Nhận diện khuôn mặt với kỹ thuật Triplet Loss

Sau khi phát hiện khuôn mặt bằng MTCNN, tiếp theo là tiền xử lý và trích chọn đặc trưng của ảnh. Trong bước này, mỗi ảnh sẽ được đưa vào một CNN đã được huấn luyện để trích xuất các đặc trưng của bức ảnh đó. Kết quả đầu ra của CNN này được biểu diễn dưới dạng một vector đặc trưng (feature vector). Sau khi thu được vector đặc trưng, dùng kNN, SVM, hoặc so sánh khoảng cách để tìm "cụm" mà vector đặc trưng đó thuộc về, từ đó suy ra danh tính, nhận diện người cần điểm danh. Trong bài báo này, chúng tôi sử dụng kỹ thuật so sánh khoảng cách để đo lường sự khác biệt giữa hai vector đặc trưng tương ứng với hai ảnh của khuôn mặt. Khoảng cách d giữa hai vector đặc trưng $x1$ và $x2$ được xác định như sau:

$$d(x1, x2) = ||f(x1) - f(x2)||_2^2 \quad (1)$$

Trong đó, hàm *f(x)* tương tự như phép biến đổi trong lớp kết nối đầy đủ của CNN để tạo tính phi tuyến và giảm chiều dữ liệu, thông thường là 128. Khi $x1$và $x2$ là đặc trưng của cùng một người, khoảng cách

*d(x*1*, x*2*)* sẽ có giá trị nhỏ. Ngược lại, khi *x*1và *x*2 là đặc trưng của hai người khác nhau, khoảng cách này sẽ có giá trị lớn.

Fig. 3. Khoảng cách dùng triplet loss trước và sau huấn luyện.

Với mỗi bức ảnh đầu vào, thông qua việc đo lường sự khác biệt từ công thức (1), mô hình huấn luyện chỉ học được một trong hai khả năng là sự giống nhau nếu chúng cùng một nhãn hoặc khác nhau nếu chúng khác nhãn, mà không học được cùng lúc sự giống và khác nhau trên cùng một lượt huấn luyện. Kỹ thuật Triplet Loss cho phép học được đồng thời sự giống nhau giữa hai bức ảnh cùng nhãn và phân biệt các bức ảnh không cùng nhãn. Kỹ thuật này sử dụng hàm hàm mất mát *triplet loss* để đánh giá khoảng cách giữa các vector đặc trưng đại diện cho bộ ba ảnh A (Anchor face), P (Positive faces) và N (Negative face). Mục tiêu của hàm *triplet loss* là tối thiểu hóa khoảng cách giữa hai ảnh khi chúng là Negative và tối đa hóa khoảng cách khi chúng là Positive. Gọi $\alpha \in R^+$ là tham số giới hạn, hàm *triplet loss* được định nghĩa như sau:

$$l(A, P, N) = max(\mathrm{d(A, P)} - \mathrm{d(A, N)} + \alpha,\ 0) \qquad (2)$$

Khi huấn luyện mô hình với *triplet loss*, luôn phải xác định trước cặp ảnh (A,P) thuộc về cùng một người. Ảnh N sẽ được lựa chọn ngẫu nhiên từ các bức ảnh thuộc các nhãn còn lại. Như vậy khi áp dụng *triplet loss* vào CNN, sẽ phân biệt tốt các ảnh Negative rất giống ảnh Positive, đồng thời các ảnh thuộc cùng một nhãn sẽ trở nên gần nhau hơn trong không gian chiều euclidean.

3 Điểm danh dựa trên MTCNN và Triplet Loss

Trong phần này, chúng tôi đề xuất xây dựng hệ thống điểm danh tự động dựa trên mô hình MTCNN và kỹ thuật Triplet Loss đã được trình bày chi tiết ở mục 2. Sơ đồ hệ thống điểm danh đề xuất được mô tả ở Fig.4.

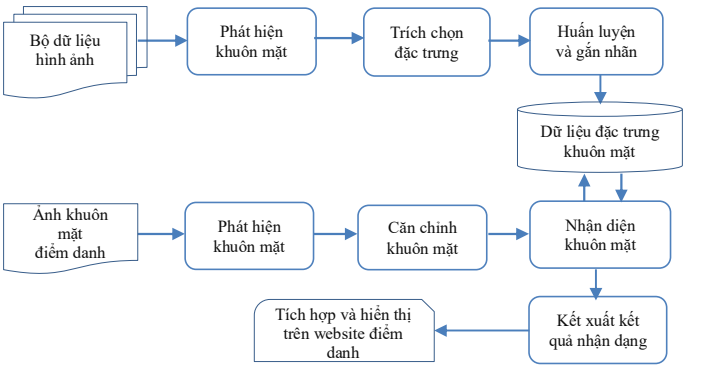


Fig. 4. Sơ đồ hệ thống điểm danh tự động.

Các công việc chính bao gồm:

- **Thu thập bộ dữ liệu khuôn mặt** (dataset): sử dụng webcam của máy tính hoặc có thể là hình ảnh từ nhiều nguồn khác nhau. Ảnh được thu thập cần đảm bảo điều kiện ánh sáng, các góc nhìn khác nhau của khuôn mặt.
- **Phát hiện khuôn mặt từ bộ dữ liệu ảnh** (face detection): sử dụng mô hình MTCNN như đã trình bày ở mục 2.
- **Căn chỉnh khuôn mặt** (face alignment): sử dụng phương pháp căn chỉnh 2D dựa vào các điểm nút thu được sau bước O-Net của quá trình phát hiện khuôn mặt dùng mô hình MTCNN. Các mốc trên khuôn mặt (đặc biệt là vùng mắt) được xoay, dịch chuyển và điều chỉnh tỷ lệ của khuôn mặt về cùng một kích thước.
- **Trích chọn đặc trưng và gắn nhãn** (pre-train): bộ dữ liệu khuôn mặt sẽ được chia theo từng thư mục tương ứng với hình ảnh của từng đối tượng sinh viên. Hệ thống sẽ tiến hành quét qua toàn bộ ảnh trong các thư mục và tìm kiếm khuôn mặt có trong ảnh, cắt lấy khuôn mặt và đưa kích thước về 160x160 pixel. Sau đó tiến hành trích rút đặc trưng của từng khuôn mặt, áp dụng kỹ thuật Triplet Loss và gắn nhãn cho từng khuôn mặt, nhãn sẽ được lấy theo tên thư mục chứa ảnh.
- **Nhận diện khuôn mặt** (face recognition): sau khi phát hiện, căn chỉnh và trích rút đặc trưng khuôn mặt. Các đặc trưng này sẽ được lưu vào vector đặc trưng 128 chiều để so sánh với bộ dữ liệu các đặc trưng đã có nhằm tìm ra khuôn mặt giống nhất và gắn nhãn cho ảnh. Với một hệ thống điểm danh tự động như trong nhà trường, việc loại trừ các trường hợp điểm danh hộ hay đi học thuê là vô cùng cần thiết. Khuôn mặt lạ xuất hiện trong ảnh được gọi là các "unknown", không có trong bộ dữ liệu được huấn luyện từ trước. Chúng tôi sử dụng phép đo độ tương tự cosine (cosine similarity) để xác định các khuôn mặt "unknown" dựa trên ngưỡng cho phép (threshold).
- **Kết xuất kết quả nhận dạng và tích hợp vào website điểm danh**: khi đã nhận diện được các khuôn mặt của sinh viên và loại bỏ các khuôn mặt lạ, hệ thống sẽ lấy định danh của khuôn mặt để tiến hành điểm danh với danh sách sinh viên trong cơ sở dữ liệu. Danh sách điểm danh của sinh viên sẽ được cập nhật lên website của hệ thống quản lý sinh viên một cách tự động.

4 Kết quả thực nghiệm

Trong bài báo này, chúng tôi sử dụng bộ dữ liệu huấn luyện gồm 4815 hình ảnh của 10 sinh viên với số lượng ảnh của mỗi sinh viên là khác nhau, dao động từ 200 đến 600 ảnh cho mỗi sinh viên. Máy tính với bộ xử lý Intel core i7-8750H, VGA Nvdia GTX 1050Ti. Thời gian xử lý được ghi nhận như sau:

- **Thời gian huấn luyện**: thời gian cho cả quá trình phát hiện khuôn mặt, trích xuất đặc trưng và huấn luyện bộ dữ liệu là 2231,3 giây (≈ 37,2 phút).
- **Thời gian nhận diện**: thời gian trung bình cho một bức ảnh bao gồm phát hiện khuôn mặt, căn chỉnh và nhận diện phụ thuộc vào số lượng khuôn mặt suất hiện trong hình. Với ảnh có chứa duy nhất một khuôn mặt, thời gian mất khoảng 5 giây. Với ảnh chứa 10 khuôn mặt, mất khoảng 8 giây.

Sau đây là một số kết quả thu được từ hệ thống điểm danh tự động dựa trên mô hình MTCNN và kỹ thuật Triplet Loss mà chúng tôi đã ghi nhận được trong quá trình thực nghiệm:

4.1 Hiệu quả của mô hình MTCNN trong phát hiện khuôn mặt

(a) *(b)*

Fig. 5. Hiệu quả của Haar-like Adaboost (a) và MTCNN (b)

Bằng cách thực hiện lần lượt ba bước P-Net, R-Net và O-Net trong mô hình MTCNN, chúng tôi đã phát hiện được được khung hình và 5 điểm nút quan trọng trên khuôn mặt. Các thông tin thu được sẽ là dữ liệu đầu vào quan trong cho các bước căn chỉnh, trích chọn đặc trưng,.. ở các bước tiếp theo. Để thấy được ưu điểm của mô hinh MTCNN trong việc phát hiện khuôn mặt, chúng tôi cũng thử với phương pháp sử dụng Haar-like Adaboost [3]. Kết quả cho thấy, Haar-like Adaboost chỉ hiệu quả với các khuôn mặt chính diện. Trong trường hợp một phần khuôn mặt bị che khuất (như mắt kính của sinh viên nữ bên phải), hay với góc nghiêng mạnh (như trường sinh viên nam bên trái), Haar-like Adaboost đã nhận diện nhầm vùng tai, điều này thật sự không tốt khi sử dụng cho việc pre-train(Fig.5a). Với MTCNN, tất cả các khuôn mặt có trong hình đều được phát hiện, ngay cả khuôn mặt nghiêng hay có vật cản (Fig.5b)

4.2 Cải thiện độ chính xác bằng căn chỉnh khuôn mặt và kỹ thuật Triplet Loss

Chúng tôi sử dụng phương pháp căn chỉnh 2D, dựa vào hai điểm nút là trí mắt trái và mắt phải thu được từ MTCNN, tính điểm giữa và tiến hành xoay khuôn mặt sao cho hai mắt cùng nằm trên hàng ngang (Fig.6).

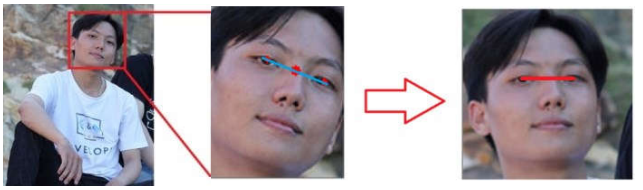


Fig. 6. Trước và sau khi căn chỉnh khuôn mặt

Để thấy được hiệu quả của căn chỉnh khuôn mặt trong việc nhận diện, chúng tôi đã tiến hành thực nghiệm như sau: Đối với bộ dữ liệu huấn luyện, với bộ dữ liệu 4815 hình ảnh của 10 sinh viên, chúng tôi huấn liệu thành 2 bộ dữ liệu riêng biệt, một bộ căn chỉnh khuôn mặt trước khi huấn luyện, bộ còn lại sử dụng hình ảnh gốc; Đối với ảnh sử dụng để nhận diện, cũng sẽ thử nghiệm lần lượt qua căn chỉnh và không qua căn chỉnh.

Như vậy sẽ có 4 trường hợp thử nghiệm: bộ dữ liệu không, bộ dữ liệu căn chỉnh, ảnh nhận dạng không căn chỉnh, ảnh nhận dạng căn chỉnh. Ảnh thử nghiệm là khuôn mặt của cùng một người trên hai bối cảnh khác nhau về: góc chụp, độ nghiêng của khuôn mặt và điều kiện ánh sáng (Fig.7). Ảnh thứ nhất có điều kiện ánh sáng đầy đủ, góc chụp từ dưới lên, khuôn mặt nghiêng mạnh về bên phải và không bị che khuất (Fig.7a), các kết quả thu được thể hiện ở Table 1. Ảnh thứ hai có điều kiện ánh sáng tại vùng mặt tương đối kém, góc chụp thẳng, khuôn mặt tương đối thẳng và không bị che khuất (Fig.7b), các kết quả thu được thể hiện ở Table 2.

(a) *(b)*

Fig. 7. Nhận diện khuôn mặt của cùng một người trên hai bối cảnh khác nhau về: góc chụp, độ nghiêng của khuôn mặt và điều kiện ánh sáng

Table 1. Tỉ lệ chính xác đối với nhận diện ảnh thứ nhất (Fig.9a)

Trường hợp	Bộ dữ liệu không căn chỉnh	Bộ dữ liệu căn chỉnh	Ảnh nhận dạng không căn hình	Ảnh nhận dạng căn chỉnh	Tỉ lệ chính xác(%)
1	✓		✓		79.22

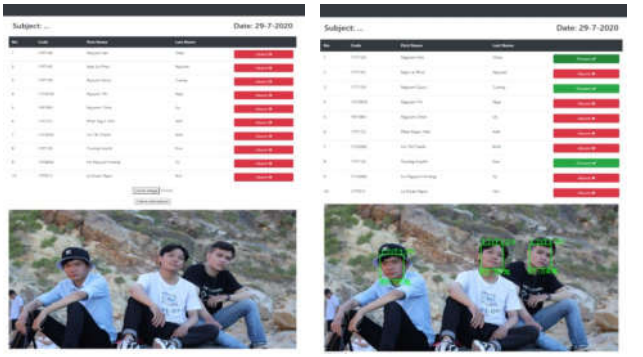
2	✓			✓	82.44
3		✓	✓		74.11
4		✓		✓	85.13

Table 2. Tỉ lệ chính xác đối với nhận diện ảnh thứ hai (Fig.9b)

Trường hợp	Bộ dữ liệu không căn chỉnh	Bộ dữ liệu căn chỉnh	Ảnh nhận dạng không căn hình	Ảnh nhận dạng căn chỉnh	Tỉ lệ chính xác(%)
1	✓		✓		88.61
2	✓			✓	91.51
3		✓	✓		74.65
4		✓		✓	87.00

Sau khi thử nghiệm với 2 hình ảnh qua 4 trường hợp, chúng tôi nhận thấy: Với ảnh thứ nhất, độ chính xác cao nhất đạt được trong trường hợp bộ dữ liệu đã căn chỉnh và ảnh nhận dạng cũng qua căn chỉnh (85.13%). Như vậy, đối với những khuôn mặt có độ nghiêng mạnh, việc sử dụng bộ dữ liệu đã căn chỉnh và hình ảnh đầu vào qua căn chỉnh cho ta kết quả tốt nhất. Với ảnh thứ hai, độ chính xác cao nhất đạt được trong trường hợp bộ dữ liệu không ăn chỉnh và ảnh nhận dạng qua căn chỉnh (91.51%). Như vậy, có thể thấy bộ dữ liệu đã căn chỉnh khuôn mặt chưa thực sự hiệu quả với các trường hợp khuôn mặt khác nhau.

4.3 Tích hợp kết quả nhận dạng vào website điểm danh



(a) *(b)*

Fig. 8. Danh sách trước (*a*) và sau khi điểm danh (*b*)

Sau khi nhận diện được các khuôn mặt của sinh viên và loại bỏ các khuôn mặt lạ, hệ thống sẽ lấy định danh của khuôn mặt để tiến hành điểm danh với danh sách sinh viên trong cơ sở dữ liệu. Danh sách điểm danh của sinh viên sẽ được cập nhật lên website quản lý sinh viên một cách tự động (Fig.8).

5 Kết luận

Mạng nơ-ron tích chập xếp tầng đa nhiệm MTCNN cho phép xác định khuôn mặt ở nhiều góc nghiêng khác nhau, ngay cả trong trong điều kiện thiếu sáng và một phần khuôn mặt bị che khuất. Trong bài báo này, chúng tôi đã đề xuất giải pháp điểm danh tự động dùng mô hình MTCNN nhằm phát hiện khuôn mặt, đồng thời kết hợp kỹ thuật Triplet Loss để nhận diện đối tượng. Kỹ thuật căn chỉnh khuôn mặt cũng

được áp dụng nhằm gia tăng độ chính xác của nhận diện. Kết quả thực nghiệm trên hệ thống điểm danh đề xuất

- Về khả năng phát hiện khuôn mặt, hệ thống phát hiện khá tốt hầu hết các trường hợp, kể cả trong điều kiện thiếu sáng, góc nghiêng, hay có vật che khuất như kính mắt,…
- Về khả năng nhận dạng, hệ thống đạt kết quả từ 80-95% đối với các khuôn mặt thẳng và điều kiện ánh sáng thích hợp, đạt 70-85% đối với các khuôn mặt nghiêng hoặc thiếu sáng.
- Về khả năng loại trừ các khuôn mặt "unknown", kết quả đạt khoảng 50-60% khuôn mặt lạ được phát hiện trong quá trình thử nghiệm.

Hệ thống điểm danh hoạt động ổn định. Giao diện được xây dựng trên nền web là một lợi thế vì tính đơn giản và tiện lợi. Không chỉ dừng lại ở việc điểm danh, hệ thống nhận dạng khuôn mặt còn có thể được sử dụng trong các hệ thống mở khóa, thanh toán, hay truy tìm tội phạm,…

References

1. Xiao Zhang, Zhiyuan Fang, Yandong Wen, Zhifeng Li, and Yu Qiao: *Range Loss for Deep Face Recognition with Long-Tail*. In: IEEE Conf. on Computer Vision and Pattern Recognition (2017).
2. M. T. Pham, Y. Gao, V. D. D. Hoang, and T. J. Cham: *Fast polygonal integration and its application in extending haar-like features to improve object detection*. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 942-949 (2010).
3. Paul Viola and Michael Jones: *Rapid Object Detection using a Boosted Cascade of Simple Features*. Computer Vision and pattern Recognition (2001).
4. J. H. Shah, M. Sharif, M. Raza, M. Murtaza, Saeed-Ur-Rehman: *Robust Face Recognition Technique under Varying Illumination*. In: Journal of Applied Research and Technology, vol. 13, issue 1, pp. 97-105, (2015).
5. Q. Zhu, M. C. Yeh, K. T. Cheng, and S. Avidan: *Fast human detection using a cascade of histograms of oriented gradients*. In: IEEE Computer Conference on Computer Vision and Pattern Recognition, pp. 1491-1498 (2006).
6. Ross Girshick, Forrest Iandola, Trevor Darrell, Jitendra Malik: *Deformable Part Models are Convolutional Neural Networks*, Computer Vision and Pattern Recognition (2014).
7. Jiankang Deng, Jia Guo, Niannan Xue, Stefanos Zafeiriou: ArcFace: *Additive Angular Margin Loss for Deep Face Recognition*, InsightFace, FaceSoft (2016).
8. H. Li, Z. Lin, X. Shen, J. Brandt, and G. Hua: *A convolutional neural network cascade for face detection*. In: Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA (2015).
9. Denton, E., Zaremba, W., Bruna, J., LeCun, Y., Fergus, R.: *Exploiting linear structure within convolutional networks for efficient evaluation*. In: Advances in Neural Information Processing Systems (2014)
10. K. Zhang, Z. Zhang, Z. Li, and Y. Qiao: *Joint face detection and alignment using multitask cascaded convolutional networks*. In: IEEE Signal Processing Letters, vol. 23, no. 10, pp. 1499–1503 (2016).
11. X. LiuZhijun, X. Xiang, Z. Jianrong and GuQi Xu: *Driver Fatigue Detection Using Multitask Cascaded Convolutional Networks*, Advances in Information and Communication Technology,(IFIPAICT, volume 510 (2017).
12. S. Zhang, X. Zhu, Z. Lei, H. Shi, X. Wang, and S. Z. Li: FaceBoxes: *A CPU real-time face detector with high accuracy*. In: Proceedings of IEEE International Joint Conference on Biometrics, pp. 1–9, IEEE, Denver, CO, USA, (2017).
13. H. Li, Z. Lin, X. Shen, J. Brandt, and G. Hua*: A convolutional neural network cascade for face detection*. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 5325-5334 (2015).