

ỨNG DỤNG QUẢN LÝ VÀ CHẤM CÔNG BẰNG NHẬN DẠNG KHUÔN MẶT DỰA TRÊN TRÍ TUỆ NHÂN TẠO

Nguyễn Đức Nhật Quang*, Phan Văn Cường,
Ngô Văn Tiến Đạt, Phan Hải Phong

Khoa Điện, Điện tử và Công nghệ vật liệu, Trường Đại học Khoa học, Đại học Huế

*Email: ndnquang@hueuni.edu.vn

Ngày nhận bài: 19/9/2023; ngày hoàn thành phản biện: 28/9/2023; ngày duyệt đăng: 4/12/2023

TÓM TẮT

Nhận dạng khuôn mặt là một bài toán quan trọng trong trí tuệ nhân tạo. Chấm công bằng nhận dạng khuôn mặt với độ chính xác cao, giúp tiết kiệm thời gian và chi phí so với các phương pháp khác. Trong nghiên cứu này, nhóm tác giả ứng dụng thuật toán FaceNet kết hợp với mạng nơ-ron tích chập đa nhiệm (Multi-task Cascaded Convolutional Networks – MTCNN) để phát hiện và xác định khuôn mặt trong hệ thống chấm công. FaceNet được sử dụng để so sánh và nhận dạng khuôn mặt, trong khi MTCNN sử dụng học sâu để phát hiện khuôn mặt, xác định điểm đặc trưng và giới hạn khuôn mặt. Ngoài ra, nhóm tác giả còn thiết kế giao diện ứng dụng thân thiện với người dùng, giúp quản lý và chấm công trực quan hơn. Kết quả thử nghiệm cho thấy hiệu quả nhận dạng khuôn mặt với độ chính xác lên đến 96%, đồng thời giảm ảnh hưởng của yếu tố khách quan như ánh sáng, góc độ.

Từ khóa: attendance, face recognition, facenet, MTCNN

1. MỞ ĐẦU

Trong bối cảnh của thời đại số hóa và sự phát triển mạnh mẽ của công nghệ thông tin, việc quản lý nhân sự một cách hiệu quả và chính xác đang trở thành một thách thức quan trọng đối với các tổ chức và doanh nghiệp. Hiện nay, các phương pháp chấm công phổ biến được sử dụng trong các tổ chức và doanh nghiệp bao gồm chấm công bằng tay (chấm công thủ công), chấm công bằng thẻ từ và chấm công bằng vân tay [1]. Mặc dù đã phục vụ mục đích quản lý nhân sự trong thời gian dài nhưng các phương pháp này đều gặp phải một số hạn chế như gian lận, sai sót do nhập liệu, không đảm bảo tính chính xác và hiệu quả chưa cao. Để giải quyết những hạn chế này, hệ thống quản lý và chấm công bằng công nghệ nhận dạng khuôn mặt [2] đã xuất hiện như một giải pháp đột phá. Công nghệ nhận dạng khuôn mặt hoạt động một cách tự động và chính xác, có khả năng

ghi nhận thời gian làm việc của nhân viên, đồng thời mang lại nhiều lợi ích vượt trội cho việc quản lý nhân sự.

Hệ thống này hoạt động bằng cách sử dụng thiết bị chụp ảnh hoặc camera để quét và ghi lại hình ảnh khuôn mặt của nhân viên. Sau đó, hệ thống sử dụng các thuật toán thông minh để phân tích và so sánh các đặc điểm độc nhất của khuôn mặt với cơ sở dữ liệu đã được lưu trữ trước đó. Khi nhận dạng thành công, hệ thống sẽ tự động ghi nhận thời gian và thông tin liên quan, như thời gian đến làm, thời gian ra về và các hoạt động khác.

Hệ thống quản lý và chấm công bằng công nghệ nhận dạng khuôn mặt mang lại nhiều lợi ích đáng kể. Đầu tiên, nó giúp tiết kiệm thời gian và công sức của cả nhân viên và người quản lý, vì quá trình chấm công trở nên tự động và nhanh chóng. Thứ hai, nó giảm thiểu sai sót trong quá trình chấm công, đảm bảo tính chính xác và minh bạch. Thứ ba, nó tăng cường an ninh và kiểm soát truy cập bằng cách loại bỏ khả năng sử dụng thẻ hoặc mật khẩu giả mạo. Cuối cùng, nó cung cấp cho người quản lý thông tin liên quan đến chấm công một cách thuận tiện và dễ dàng.

Với nhận thức về xu hướng phát triển của công nghệ, nghiên cứu này tập trung vào "Hệ thống quản lý và chấm công bằng công nghệ nhận dạng khuôn mặt" và trình bày hai chức năng chính: chấm công nhân viên bằng nhận dạng khuôn mặt và quản lý trên ứng dụng.

2. PHƯƠNG PHÁP NGHIÊN CỨU

Nhận dạng khuôn mặt là một lĩnh vực quan trọng của Trí tuệ nhân tạo (Artificial Intelligence – AI) và đã được ứng dụng rộng rãi trong nhiều lĩnh vực như an ninh, giám sát, chấm công, giải trí, y tế và nhiều hơn nữa. Cùng với sự phát triển không ngừng của công nghệ, các phương pháp và thuật toán để nhận dạng khuôn mặt ngày càng được cải tiến và hoàn thiện. Nhận dạng khuôn mặt được mô tả qua các bước phát hiện khuôn mặt, chuẩn hoá và trích chọn đặc trưng để đưa ra kết quả.

Hiện nay, có ba phương pháp tiếp cận phổ biến trong bài toán phát hiện và nhận dạng khuôn mặt, bao gồm: thuật toán Haar Cascade [3], mạng nơ-ron tích chập (Convolutional Neural Network – CNN) [4] và phân tích đặc trưng (Feature Analysis). Trong đó:

- Thuật toán Haar Cascade hoạt động bằng cách sử dụng một tập hợp các đặc trưng kiểu Haar để phát hiện sự hiện diện của khuôn mặt trong một hình ảnh. Các đặc trưng kiểu Haar là các mẫu chữ nhật đơn giản có thể được sử dụng để phân biệt giữa các vùng khác nhau của một hình ảnh dựa trên giá trị độ sáng của các điểm ảnh (pixel). Thuật toán này có khả năng phát hiện khuôn mặt với độ chính xác cao.

- CNN hoạt động bằng cách sử dụng một mạng nơ-ron sâu với các lớp tích chập để trích xuất các đặc trưng của khuôn mặt từ một hình ảnh. Các lớp tích chập trong CNN sử dụng bộ lọc để quét qua các vùng của hình ảnh và tạo ra các bản đồ đặc trưng tương ứng. Bằng cách kết hợp các lớp tích chập với các lớp kích hoạt phi tuyến, CNN có khả năng học và phân loại các đặc trưng phức tạp của khuôn mặt với độ chính xác cao.
- Thuật toán phân tích đặc trưng là một phương pháp nhận dạng khuôn mặt dựa trên việc phân tích các đặc trưng của khuôn mặt. Thuật toán này sử dụng một tập hợp các bộ lọc đơn giản để phát hiện các đặc trưng của khuôn mặt trong một hình ảnh. Các bộ lọc này bao gồm các bộ lọc đường thẳng, hình chữ nhật và hình tam giác. Qua quá trình huấn luyện và tối ưu hóa, thuật toán phân tích đặc trưng có thể phát hiện và phân loại khuôn mặt với độ chính xác cao.

Tuy nhiên, các phương pháp trên vẫn tồn tại một số nhược điểm như tốc độ xử lý chậm, độ chính xác thấp khi khuôn mặt bị che khuất, không đủ ánh sáng hoặc góc nghiêng của khuôn mặt, khó tích hợp vào các ứng dụng thực tế và có độ bảo mật không cao.

Do đó, trong nghiên cứu này, nhóm tác giả đã đề xuất sử dụng Mạng nơ-ron tích chập đa nhiệm (Multi-task Cascaded Convolutional Networks – MTCNN) [5] kết hợp với thuật toán FaceNet [5] làm phương pháp tiếp cận mới và cải tiến. Trong đó, MTCNN sử dụng CNN để phát hiện khuôn mặt và tạo ra các véc-tơ đặc trưng biểu diễn vị trí khuôn mặt. FaceNet sử dụng mạng CNN để nhúng hình ảnh vào không gian Euclid và tạo ra các biểu diễn đặc trưng để nhận dạng khuôn mặt. Sự kết hợp giữa hai thuật toán này giúp khắc phục các nhược điểm của các phương pháp truyền thống và mang lại hiệu suất nhận dạng khuôn mặt cao hơn.

2.1. Phát hiện khuôn mặt bằng MTCNN

Mạng nơ-ron tích chập đa nhiệm là một phương pháp phổ biến được sử dụng để phát hiện khuôn mặt trong bức ảnh. Nó bao gồm ba mạng nơ-ron tích chập được xếp chồng lên nhau để phát hiện và xác định vị trí khuôn mặt. Mỗi mạng nơ-ron tích chập trong MTCNN có cấu trúc và vai trò khác nhau trong quá trình phát hiện khuôn mặt. Khi sử dụng, MTCNN cho phép tạo ra nhiều bản sao của ảnh đầu vào với các kích thước khác nhau để sử dụng làm các dữ liệu đầu vào. Kết quả đầu ra của MTCNN là một véc-tơ đặc trưng biểu diễn vị trí khuôn mặt được xác định trên bức ảnh, bao gồm các đặc điểm như mắt, mũi, miệng, và nhiều hơn nữa.

Quá trình phát hiện khuôn mặt bằng MTCNN được thực hiện thông qua ba bước với ba mạng nơ-ron tích chập riêng biệt tương ứng cho mỗi bước là mạng đề xuất

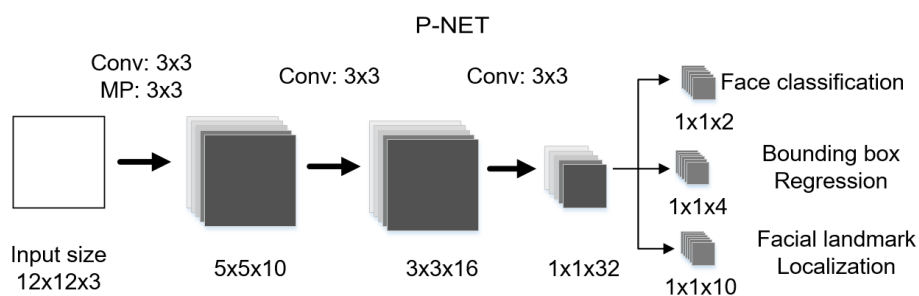
(Proposal Network – P-Net), mạng lọc (Refine Network – R-Net) và mạng đầu ra (Output Network – O-Net), cụ thể:

Bước 1: Mạng đề xuất

Trong bước này, mạng đề xuất (Proposal Network – P-Net) được sử dụng để thu được các cửa sổ chứa khuôn mặt và các véc-tơ hồi quy để điều chỉnh vị trí của các cửa sổ đó. Tiếp theo, các cửa sổ chứa khuôn mặt được hiệu chuẩn dựa trên các véc-tơ hồi quy. Cuối cùng, những cửa sổ xếp chồng nhau tại một vùng được hợp nhất thành một cửa sổ. Kết quả của P-Net là các cửa sổ có khả năng chứa khuôn mặt.

P-Net trong MTCNN sử dụng một kiến trúc CNN đơn giản gồm 3 lớp tích chập (Convolutional – CONV) và 1 lớp gộp cực đại (Max Pooling – MP) để thực hiện nhiệm vụ phát hiện khuôn mặt (Hình 1). Đầu vào của P-Net là một cửa sổ trượt có kích thước $12 \times 12 \times 3$. Trong đó:

- Số 3 tương ứng với 3 kênh màu: đỏ, xanh lục, xanh lam trong hệ màu RGB.
- 12×12 : kích thước của cửa sổ hoặc vùng mà mô hình sẽ quét qua bức ảnh. Kích thước 12×12 nghĩa là cửa sổ này là một hình vuông có chiều rộng và chiều cao đều là 12 điểm ảnh.



Hình 1. Mạng đề xuất (P-Net) [8].

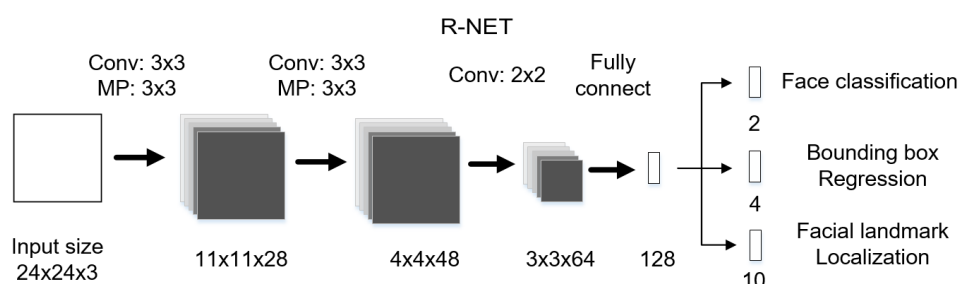
P-Net gồm 3 cụm như sau:

- Cụm thứ nhất có 2 bộ lọc kích thước 1×1 để nhận dạng khuôn mặt. Cụm này được sử dụng để xác định xem cửa sổ hiện tại có chứa khuôn mặt hay không.
- Cụm thứ hai có 4 bộ lọc kích thước 1×1 để đóng khung 4 vị trí hộp giới hạn (bounding box). Cụm này được sử dụng để xác định vị trí và kích thước của khuôn mặt trong cửa sổ.
- Cụm thứ ba có 10 bộ lọc kích thước 1×1 để đóng khung 10 vị trí trên khuôn mặt. Cụm này được sử dụng để xác định các điểm chính trên khuôn mặt như mắt, mũi, miệng, v.v.

P-Net cho phép xác định vị trí của khuôn mặt và các đặc điểm trên khuôn mặt trong cửa sổ trượt. Các cụm này được sử dụng để tiếp tục quá trình phát hiện khuôn mặt và điều chỉnh vị trí trong các bước tiếp theo của MTCNN.

Bước 2: Mạng lọc

Tất cả các cửa sổ chứa khuôn mặt từ bước trước sẽ được sàng lọc bằng cách đưa vào mạng lọc (Refine Network – R-Net) như trên **Hình 2**. R-Net tiếp tục lọc bỏ các cửa sổ không chứa khuôn mặt và áp dụng các véc-tơ hồi quy để điều chỉnh vị trí của các cửa sổ còn lại.



Hình 2. Mạng lọc (R-Net) [8].

R-Net sử dụng kiến trúc CNN gồm: 3 lớp tích chập, 2 lớp gộp cực đại và 1 lớp kết nối đầy đủ (Fully Connected – FC). Đầu vào cửa sổ trượt với kích thước $24 \times 24 \times 3$. Trong đó:

- Số 3 tương ứng với 3 kênh màu: đỏ, xanh lục, xanh lam trong hệ màu RGB.
- 24×24 : kích thước của cửa sổ hoặc vùng mà mô hình sẽ quét qua bức ảnh. Kích thước 24×24 nghĩa là cửa sổ này là một hình vuông có chiều rộng và chiều cao đều là 24 điểm ảnh.

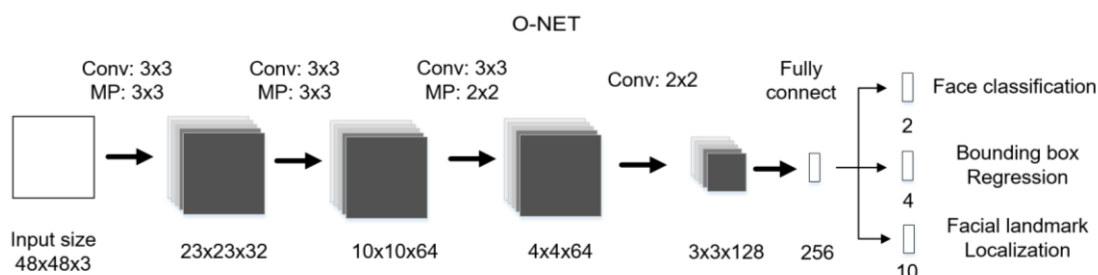
R-Net phân thành 3 cụm:

- Cụm thứ nhất có 2 lớp nhận dạng khuôn mặt. Cụm này được sử dụng để xác định xem cửa sổ trượt hiện tại có chứa khuôn mặt hay không.
- Cụm thứ hai có 4 lớp đánh dấu vị trí hộp giới hạn. Cụm này được sử dụng để xác định vị trí và kích thước của khuôn mặt trong cửa sổ trượt.
- Cụm thứ ba có 10 lớp vị trí khuôn mặt. Cụm này được sử dụng để xác định các điểm chính trên khuôn mặt như mắt, mũi, miệng, v.v.

Các cụm này cung cấp thông tin quan trọng để tiếp tục quá trình phát hiện khuôn mặt và điều chỉnh vị trí trong bước cuối cùng của MTCNN.

Bước 3: Mạng đầu ra

Bước cuối cùng sử dụng mạng đầu ra (Output Network – O-Net) như trên **Hình 3** để lọc kết quả một lần nữa và đánh dấu vị trí của năm điểm chính trên khuôn mặt: hai mắt, mũi và hai góc miệng.



Hình 3. Mạng đầu ra (O-Net) [8].

O-Net sử dụng CNN gồm: 4 lớp tích chập, 3 lớp tổng hợp tối đa và 1 lớp kết nối đầy đủ. Đầu vào của cửa sổ trượt có kích thước $48 \times 48 \times 3$. Trong đó:

- Số 3 tương ứng với 3 kênh màu: đỏ, xanh lục, xanh lam trong hệ màu RGB.
- 48×48 : kích thước của cửa sổ hoặc vùng mà mô hình sẽ quét qua bức ảnh. Kích thước 48×48 nghĩa là cửa sổ này là một hình vuông có chiều rộng và chiều cao đều là 48 điểm ảnh.

O-Net phân thành 3 cụm:

- Cụm thứ nhất có 2 lớp nhận dạng khuôn mặt. Cụm này được sử dụng để xác định xem cửa sổ trượt hiện tại có chứa khuôn mặt hay không.
- Cụm thứ hai có 4 lớp đánh dấu vị trí hộp giới hạn. Cụm này được sử dụng để xác định vị trí và kích thước của khuôn mặt trong cửa sổ trượt.
- Cụm thứ ba có 10 lớp vị trí đặc điểm trên khuôn mặt. Cụm này được sử dụng để xác định các điểm chính trên khuôn mặt như mắt, mũi, miệng, v.v.

Ứng dụng MTCNN giúp xác định khuôn mặt trong bức ảnh tốt hơn (độ chính xác cao, phát hiện đa nhiệm, hiệu suất trong thời gian thực cao) so với các phương pháp khác [6] nhờ vào việc sử dụng các mạng nơ-ron tích chập phức tạp để phát hiện và xác định vị trí khuôn mặt cùng với các đặc điểm trên khuôn mặt.

Như vậy, mỗi bước trong MTCNN sử dụng một mạng CNN riêng biệt để thực hiện các nhiệm vụ phát hiện khuôn mặt và điều chỉnh vị trí. Các kết quả cuối cùng là các cửa sổ chứa khuôn mặt và các thông tin về vị trí của các đặc điểm trên khuôn mặt như mắt, mũi, miệng. MTCNN cho phép xác định khuôn mặt trong bức ảnh và chuẩn hoá vị trí khuôn mặt để sử dụng cho các bước tiếp theo trong quá trình nhận dạng khuôn mặt.

2.2. Nhận dạng khuôn mặt bằng FaceNet

FaceNet là một thuật toán hỗ trợ việc nhận dạng và phân cụm khuôn mặt, sử dụng một mạng CNN và cho phép giảm số chiều dữ liệu của véc-tơ đặc trưng (thường là 128 chiều). FaceNet có thể được coi là một dạng Siam Network [7], biểu diễn véc-tơ đặc trưng của các bức ảnh trong một không gian Euclid n chiều (thường là 128 chiều). Việc biểu diễn này tuân theo quy tắc rằng nếu khoảng cách giữa các véc-tơ nhúng càng nhỏ, thì mức độ tương đồng giữa chúng càng lớn và ngược lại. Tập hợp các véc-tơ này được sử dụng làm dữ liệu đầu vào cho hàm mất mát để đánh giá khoảng cách giữa chúng.

FaceNet sử dụng một mạng CNN, được biểu diễn bằng hàm $f(x)$, để nhúng hình ảnh X vào không gian Euclid d chiều. Mục tiêu là tạo ra một biểu diễn sao cho khoảng cách giữa các hình ảnh của cùng một người không bị ảnh hưởng bởi các điều kiện bên ngoài, và khoảng cách giữa các khuôn mặt giống nhau của cùng một người là nhỏ, trong khi khoảng cách giữa các ảnh khác nhau sẽ có khoảng cách lớn.

Hàm $f(x) \in R^d$ có chức năng biểu diễn ảnh X vào không gian Euclid d chiều. Tại đây, bao gồm 3 bức ảnh: Anchor (ảnh gốc), Positive (ảnh gần giống với ảnh gốc) và Negative (ảnh khác với ảnh gốc). Sau khi biểu diễn vào không gian Euclid thì tương ứng với 3 bức ảnh trên là X_a, X_p, X_n . Để nhận dạng tốt thì khoảng cách từ X_a tới X_p phải nhỏ hơn khoảng cách từ X_a tới $X_n : d(X_a, X_p) < d(X_a, X_n)$. Do đó, dẫn đến biểu thức (1) (với g là giá trị biên):

$$\|f(X_a) - f(X_p)\|_2^2 + g < \|f(X_a) - f(X_n)\|_2^2 \quad (1)$$

Lúc đó hàm Triplet Loss sẽ có dạng như sau:

$$L(X_a, X_p, X_n) = \sum \|f(X_a) - f(X_p)\|_2^2 - \|f(X_a) - f(X_n)\|_2^2 + g \quad (2)$$

Khi huấn luyện mô hình Siam network với triplet loss cần phải xác định trước cặp ảnh (X_a, X_p) thuộc về cùng một người. Ảnh X_n là ảnh khác với ảnh gốc của người đó thường sẽ được lựa chọn ngẫu nhiên từ các bức ảnh thuộc các lớp còn lại. Do đó, tập hợp ảnh X_n thường được thu thập nhiều hơn 1 bức ảnh/1 người để có thể chuẩn bị được tập dữ liệu huấn luyện. Nếu 1 người chỉ có 1 ảnh thì có thể đưa những tập dữ liệu như vậy làm bộ ảnh X_n khi huấn luyện.

Như đã nêu trên có thể thấy, khi sử dụng triplet loss vào các mô hình CNN có thể tạo ra các véc-tơ đặc trưng tốt nhất cho mỗi một bức ảnh. Các véc-tơ đặc trưng này sẽ cho phép phân biệt rõ các ảnh Negative (ảnh khác với ảnh gốc) rất giống ảnh Positive (ảnh gần giống với ảnh gốc). Hơn nữa, khoảng cách giữa các bức ảnh thuộc cùng một lớp sẽ trở nên gần nhau hơn trong không gian chiếu Euclid.

Tuy vậy, việc sử dụng bộ ba như trên sẽ khiến cho quá trình hội tụ chậm. Do đó, cần chọn bộ ba thích hợp trong quá trình huấn luyện để cải thiện được hiệu suất và độ chính xác của mô hình.

Để khắc phục được việc hội tụ chậm, thường sẽ chọn bộ ba sai số sao cho khoảng cách giữa ảnh gốc và ảnh gần với ảnh gốc (ảnh của cùng 1 người) là lớn nhất và khoảng cách giữa ảnh gốc và ảnh của người khác là bé nhất:

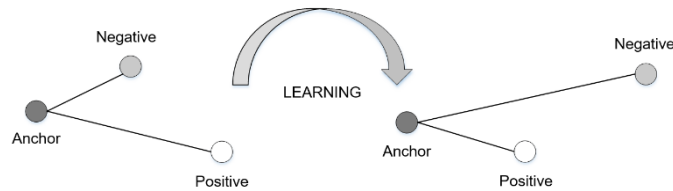
$$\begin{aligned} & \arg \max(\|f(X_a) - f(X_p)\|_2^2) \\ & \arg \max(\|f(X_a) - f(X_n)\|_2^2) \end{aligned} \quad (3)$$

Việc chọn hình ảnh như trên có thể xảy ra trường hợp:

$$\|f(X_a) - f(X_p)\|_2^2 > \|f(X_a) - f(X_n)\|_2^2 \quad (4)$$

Lúc này ta sẽ huấn luyện làm sao cho biểu thức (4) trở về biểu thức (1). Việc huấn luyện sẽ giúp khoảng cách giữa hai ảnh của cùng 1 người là nhỏ nhất và ngược lại ảnh của 2 người sẽ có khoảng cách là lớn nhất.

Hình 4 minh họa về quá trình sau huấn luyện. Trong đó, việc lựa chọn bộ ba sai số sẽ ảnh hưởng đến hiệu quả của mô hình, nếu giá trị bộ ba sai số được xác định tốt thì quá trình hội tụ khi huấn luyện sẽ nhanh hơn và kết quả sẽ cho độ chính xác cao hơn. Việc lựa chọn ngẫu nhiên bộ ba sai số có thể dẫn tới mô hình huấn luyện không thể hội tụ.

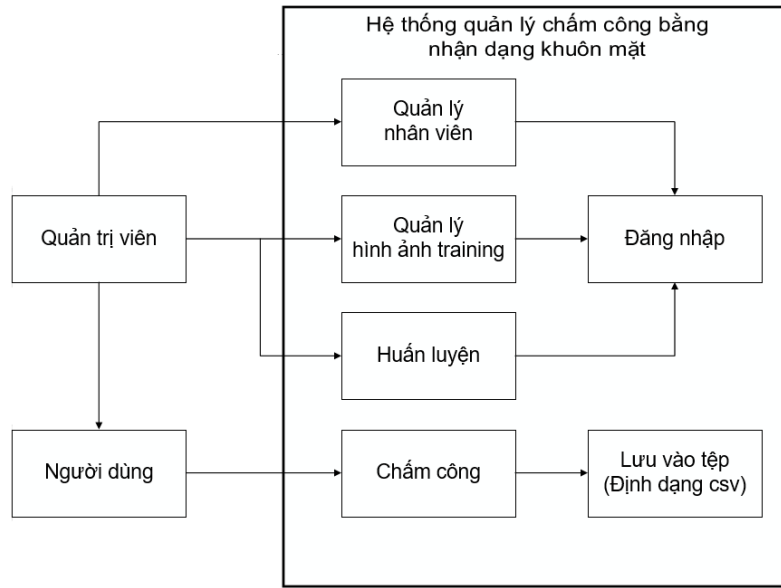


Hình 4. Minh họa về quá trình sau huấn luyện [9].

2.3. Hệ thống chấm công bằng nhận dạng khuôn mặt.

Nhóm tác giả tiến hành xây dựng hệ thống nhận dạng trên cơ sở ứng dụng MTCCN và FaceNet sử dụng mạng nơ-ron tích chập [8].

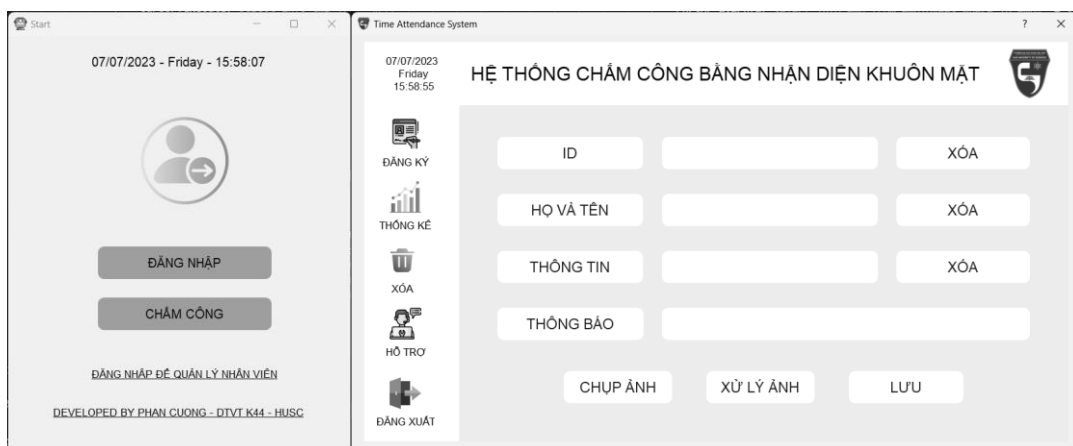
Hệ thống được thiết kế gồm 02 phân quyền chính là Quản trị viên và Người dùng (**Hình 5**). Phân quyền Người dùng có thể thực hiện chức năng: Chấm công (nhận dạng trực tiếp sử dụng camera), lưu dữ liệu vào tệp (định dạng *.csv). Phân quyền quản trị thừa kế từ phân quyền người dùng và có thêm các chức năng: Quản lý nhân viên, quản lý hình ảnh và huấn luyện. Chức năng của hệ thống được mô tả thông qua sơ đồ sau:



Hình 5. Sơ đồ khối hệ thống quản lý và chấm công bằng công nghệ nhận diện khuôn mặt

Bên cạnh đó, hệ thống sử dụng Google Drive lưu trữ ảnh huấn luyện và TensorFlow (một thư viện và nền tảng học máy mã nguồn mở do Google phát triển để xây dựng, đào tạo và triển khai các mô hình học sâu) để thực hiện các phép tính toán và sử dụng Google Colaboratory (một môi trường trực tuyến miễn phí cho máy học và khoa học dữ liệu) của Google để huấn luyện dữ liệu. Hệ thống có chức năng quản lý nhân viên gồm: thêm, xóa, thống kê ngày đi làm, giờ đi làm của nhân viên.

Giao diện GUI (Graphical User Interface) được sử dụng để giúp người dùng dễ dàng sử dụng phần mềm và hướng tới các đối tượng không chuyên về công nghệ, mang lại trải nghiệm thân thiện và gần gũi (**Hình 6**).



Hình 6. Giao diện GUI

Để thêm mới thông tin người dùng, quản trị viên cần thực hiện các bước sau:

- Bước 1: Nhập các thông tin của nhân viên bao gồm: ID, Họ và tên, Thông tin. Sau đó nhấn nút chụp ảnh, hệ thống sẽ kiểm tra xem nhân viên đã tồn tại trong hệ thống chưa nếu đã tồn tại thì sẽ hiển thị thông báo nhân viên đã tồn tại, còn không thì cửa sổ chụp ảnh sẽ được mở lên và tiến hành chụp ảnh nhân viên để thêm vào cơ sở dữ liệu.
- Bước 2: Xử lý ảnh, ảnh sau khi được chụp thì cần xử lý cho phù hợp với yêu cầu của hệ thống.
- Bước 3: Lưu, tiến hành lưu dữ liệu của nhân viên vào hệ thống gồm ID, Họ và tên, Thông tin và ảnh của nhân viên sau khi được xử lý.

Ngoài ra, hệ thống còn có chức năng thống kê thời gian đi làm của nhân viên, xóa nhân viên và gọi hỗ trợ kỹ thuật nếu phần mềm gặp trục trặc trong quá trình hoạt động.

3. KẾT QUẢ VÀ THẢO LUẬN

Để huấn luyện mô hình, nhóm tác giả sử dụng tập dữ liệu khuôn mặt người Việt từ Google Images - một dịch vụ tìm kiếm của Google cho phép người dùng tìm hình ảnh trên các trang web, gồm 24120 khuôn mặt của 1020 người. Đặc điểm của bộ dữ liệu này các ảnh của một người được thu thập tại các thời kỳ, hoàn cảnh khác nhau. Nhóm tác giả lấy ảnh của 1000 người để làm tập dữ liệu huấn luyện mô hình. Mô hình được huấn luyện với tham số và thời gian huấn luyện mô hình mất 120 phút.

Tập dữ liệu để kiểm tra bao gồm 50 người với 1020 bức ảnh. Nhóm tác giả sử dụng 720 ảnh để huấn luyện và 300 ảnh để kiểm thử (tất cả các ảnh được điều chỉnh về kích thước 160x160) cho cả 2 thuật toán: Haar Cascade (thuật toán được phát triển cho các đối tượng cụ thể được đào tạo trước như khuôn mặt) và FaceNet (sử dụng MTCNN để phát hiện khuôn mặt). Kết quả thực hiện nhận dạng của 2 phương pháp trên được thể hiện ở *Bảng 1*. Dữ liệu thực tế từ

Bảng 1 cho thấy, thời gian nhận dạng trung bình của phương pháp Haar Cascade nhanh hơn khoảng 64%. Tuy nhiên, phương pháp nhận dạng khuôn mặt sử dụng thuật toán FaceNet kết hợp với MTCNN để phát hiện khuôn mặt lại chú trọng đến việc nâng cao kết quả nhận dạng chính xác hơn, lên đến 96%.

Bảng 1. Kết quả nhận dạng khuôn mặt.

Phương pháp	Số ảnh huấn luyện	Số ảnh kiểm tra	Số ảnh nhận dạng đúng	Số ảnh nhận dạng sai	Thời gian nhận dạng trung bình (giây)	Hiệu suất (%)
Haar Cascade [3]	720	300	276	24	0,28	92
FaceNet kết hợp với MTCNN	720	300	288	12	0,44	96

Khi khuôn mặt được nhận dạng đúng thì sẽ lưu ngày giờ đi làm của nhân viên vào tệp tin với định dạng *.csv đảm bảo sự công bằng và minh bạch bằng cách xuất file Excel.

Bài báo nghiên cứu xây dựng hệ thống nhận dạng khuôn mặt trên cơ sở áp dụng MTCNN và thuật toán FaceNet để phát hiện và nhận dạng khuôn mặt, cho phép cải thiện độ chính xác khi nhận dạng. Kết quả thực nghiệm cho thấy, hệ thống có thể áp dụng đối với việc chấm công trong thực tế doanh nghiệp. Nâng cao hiệu quả chấm công mang lại sự chính xác.

4. KẾT LUẬN

Bài báo nghiên cứu, xây dựng ứng dụng nhận dạng khuôn mặt trên cơ sở áp dụng MTCNN kết hợp với thuật toán FaceNet để phát hiện và nhận dạng khuôn mặt, cho phép cải thiện độ chính xác khi nhận dạng. Kết quả thực nghiệm cho thấy, hệ thống có thể áp dụng đối với việc chấm công trong thực tế doanh nghiệp. Nâng cao hiệu quả chấm công, mang lại độ chính xác cao.

LỜI CẢM ƠN

Nghiên cứu này được thực hiện trong khuôn khổ đề tài nghiên cứu khoa học sinh viên cấp cơ sở ĐHKH2023B-04.

TÀI LIỆU THAM KHẢO

- [1] R. Memane, P. Jadhav, J. Patil, S. Mathapati and A. Pawar (2022). Attendance Monitoring System Using Fingerprint Authentication, *2022 6th International Conference On Computing, Communication, Control And Automation (ICCUBEA)*, Pune, India, pp. 1-6.
- [2] A. Yadav, A. Sharma and S. S. Yadav (2022). Attendance Management System Based on Face Recognition Using Haar-Cascade, *2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*, Greater Noida, India, pp. 1972-1976.

-
- [3] S. S. Ali, J. H. Al' Ameri and T. Abbas (2022). Face Detection Using Haar Cascade Algorithm, *2022 Fifth College of Science International Conference of Recent Trends in Information Technology (CSCTIT)*, Baghdad, Iraq, pp. 198-201
- [4] R. Xin, J. Zhang and Y. Shao (2020). Complex network classification with convolutional neural network, in *Tsinghua Science and Technology*, vol. 25, no. 4, pp. 447-457.
- [5] K. K. Suhas, M. S. Prakash, Y. Venkatesh and P. M. Latha (2023). Multi-Task Cascaded Convolutional Neural Networks (MTCNN) based Driver Drowsiness Detection, *2023 7th International Conference on Computing Methodologies and Communication (ICCMC)*, Erode, India, pp. 790-793.
- [6] M. H. Robin, M. M. Ur Rahman, A. M. Taief and Q. Nahar Eity (2020). Improvement of Face and Eye Detection Performance by Using Multi-task Cascaded Convolutional Networks, *2020 IEEE Region 10 Symposium (TENSYMP)*, Dhaka, Bangladesh, pp. 977-980.
- [7] Y. Liu, G. Chang, G. Fu, Y. Wei, J. Lan and J. Liu (2022). Self-Attention based Siamese Neural Network recognition Model, *2022 34th Chinese Control and Decision Conference (CCDC)*, Hefei, China, pp. 721-724.
- [8] S. Qi, X. Zuo, W. Feng and I. G. Naveen (2022). Face Recognition Model Based On MTCNN And Facenet, *2022 IEEE 2nd International Conference on Mobile Networks and Wireless Communications (ICMNWC)*, Tumkur, Karnataka, India, pp. 1-5.
- [9] Z. Yang, W. Ge and Z. Zhang (2020). Face Recognition Based on MTCNN and Integrated Application of FaceNet and LBP Method, *2020 2nd International Conference on Artificial Intelligence and Advanced Manufacture (AIAM)*, Manchester, United Kingdom, pp. 95-98.

APPLICATION FOR MANAGEMENT AND ATTENDANCE TRACKING WITH FACE RECOGNITION BASED ON ARTIFICIAL INTELLIGENCE

**Nguyen Duc Nhat Quang*, Phan Van Cuong,
Ngo Van Tien Dat, Phan Hai Phong**

Faculty of Electronics, Electrical Engineering and Material Technology,
University of Sciences, Hue University

*Email: ndnquang@hueuni.edu.vn

ABSTRACT

Face recognition is an important problem in artificial intelligence. Attendance using high-precision face recognition saves time and costs compared to other methods. In this research, the authors apply the FaceNet algorithm combined with the Multi-task Cascaded Convolutional Networks (MTCNN) to detect and identify faces in the attendance system. FaceNet is used for face comparison and recognition, while MTCNN uses deep learning to detect faces, identify key points, and bound the faces. Additionally, the authors design a user-friendly application interface to make management and attendance tracking more visual. The test results show the effectiveness of face recognition with an accuracy of up to 96%, while reducing the impact of subjective factors such as lighting and angles.

Keywords: attendance, face recognition, facenet, MTCNN.



Nguyễn Đức Nhật Quang sinh ngày 08/10/1992 tại Thừa Thiên Huế. Năm 2015, ông tốt nghiệp kỹ sư chuyên ngành Điện tử - Viễn thông, Trường Đại học Khoa học, Đại học Huế. Năm 2020, ông nhận bằng thạc sĩ chuyên ngành Khoa học máy tính và Kỹ thuật thông tin (CSIE) tại Trường Đại học Quốc gia Thành Công (NCKU), Đà Loan. Hiện nay, ông đang công tác tại Khoa Điện, Điện tử và Công nghệ vật liệu, Trường Đại học Khoa học, Đại học Huế.

Lĩnh vực nghiên cứu: Thiết kế vi mạch số, Trí thông minh nhân tạo (AI), Internet vạn vật kết nối (IoT), Hệ thống nhúng.



Phan Văn Cường sinh ngày 19/10/2002 tại Hà Tĩnh. Hiện nay, ông đang là sinh viên năm 4 chuyên ngành Điện tử - Viễn thông, Trường Đại học Khoa học, Đại học Huế.

Lĩnh vực nghiên cứu: Quản trị mạng, Kỹ thuật đảo ngược (Reverse Engineering)



Ngô Văn Tiến Đạt sinh ngày 16/08/2002 tại Thừa Thiên Huế. Hiện nay, ông đang là sinh viên năm 4 chuyên ngành Điện tử - Viễn thông, Trường Đại học Khoa học, Đại học Huế.

Lĩnh vực nghiên cứu: Trí thông minh nhân tạo (AI), Mạch tích hợp quang tử (PIC), Quản trị mạng.



Phan Hải Phong sinh năm 1982 tại Thừa Thiên Huế. Ông tốt nghiệp Đại học ngành Vật lý năm 2004 tại trường Đại học Khoa học, ĐH Huế; nhận bằng Thạc sĩ (2010) và Tiến sĩ (2019) về chuyên ngành Kỹ thuật điện tử tại trường Đại học Công nghệ, ĐH Quốc Gia Hà Nội. Hiện nay, ông đang công tác tại Khoa Điện, Điện tử và Công nghệ vật liệu, trường Đại học Khoa Học, Đại học Huế.

Lĩnh vực nghiên cứu: thiết kế vi mạch, hệ thống nhúng, hệ thống tích hợp trên vi mạch.

