



(12) **BẢN MÔ TẢ GIẢI PHÁP HỮU ÍCH THUỘC BẰNG ĐỘC QUYỀN
GIẢI PHÁP HỮU ÍCH**

(19) **Cộng hòa xã hội chủ nghĩa Việt Nam (VN)
CỤC SỞ HỮU TRÍ TUỆ**

(11)



2-0003910

(51)

(13) **Y**

(21) 2-2024-00012

(22) 30/09/2020

(67) 1-2020-05614

(45) 27/01/2025 442

(43) 25/01/2021 394

(73) **TẬP ĐOÀN CÔNG NGHIỆP - VIỄN THÔNG QUÂN ĐỘI (VN)**

Lô D26 Khu đô thị mới Cầu Giấy, phường Yên Hòa, quận Cầu Giấy, thành phố Hà Nội.

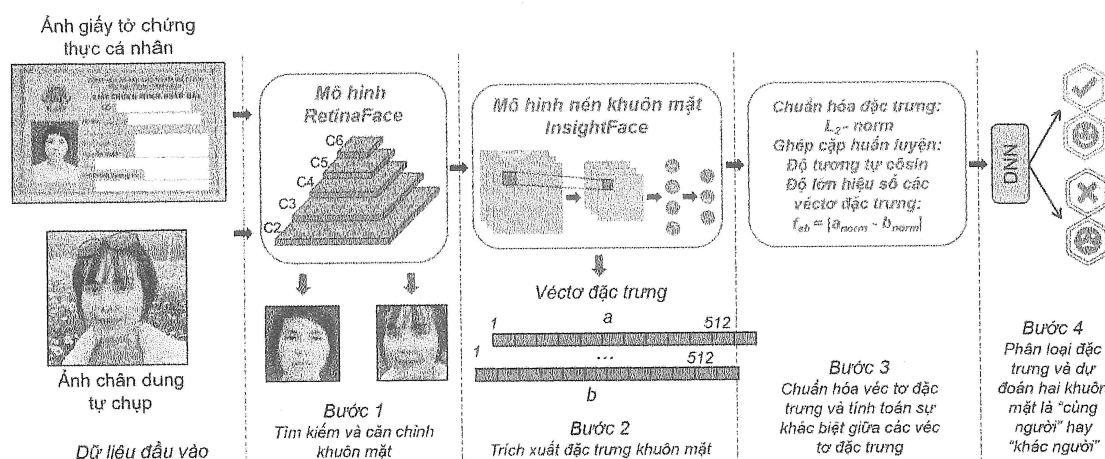
(72) Vũ Thị Hạnh (VN); Nguyễn Mạnh Quý (VN); Hoàng Ngọc Dương (VN); Hoàng Trung Hiếu (VN); Nguyễn Văn Nam (VN).

(74) Công ty TNHH NACILAW (NACILAW)

(54) **PHƯƠNG PHÁP ĐỐI SÁNH ẢNH KHUÔN MẶT GIỮA ẢNH TRÊN GIẤY TỜ
CHỨNG THỰC CÁ NHÂN VÀ ẢNH CHÂN DUNG TỰ CHỤP SỬ DỤNG CÔNG
NGHỆ TRÍ TUỆ NHÂN TẠO**

(21) 2-2024-00012

(57) Giải pháp hữu ích đề cập đến phương pháp đối sánh ảnh khuôn mặt giữa ảnh trên giấy tờ chứng thực cá nhân và ảnh chân dung tự chụp sử dụng trí tuệ nhân tạo với độ chính xác cao và thời gian tối ưu. Phương pháp dựa trên các mô hình học sâu tiên tiến được sử dụng trong công nghệ nhận diện khuôn mặt để tìm kiếm và trích xuất đặc trưng khuôn mặt, kết hợp cùng với các thuật toán tiền xử lý ảnh đầu vào, tính toán độ tương tự cosin, và bộ phân loại đặc trưng dựa trên mạng nơron sâu DNN. Để thực hiện được phương pháp theo như giải pháp hữu ích, cần trải qua các bước: bước 1: tìm kiếm và căn chỉnh khuôn mặt trong ảnh; bước 2: trích xuất đặc trưng khuôn mặt; bước 3: chuẩn hóa véc tơ đặc trưng và tính toán sự khác biệt giữa các véc tơ đặc trưng; bước 4: phân loại đặc trưng và dự đoán hai khuôn mặt có phải cùng một người hay không.



Hình 1

Lĩnh vực kỹ thuật được đề cập

Giải pháp hữu ích cập đến phương pháp đối sánh ảnh khuôn mặt giữa ảnh trên giấy tờ chứng thực cá nhân và ảnh chân dung tự chụp sử dụng công nghệ trí tuệ nhân tạo. Cụ thể, phương pháp được đề cập sử dụng trong lĩnh vực nhận dạng và định danh khuôn mặt.

Tình trạng kỹ thuật của giải pháp hữu ích

Phương pháp đối sánh ảnh khuôn mặt giữa ảnh trên giấy tờ chứng thực cá nhân (chứng minh nhân dân, căn cước công dân, hộ chiếu - sau đây gọi chung là “ID” (*Identity document*)) và ảnh chân dung tự chụp được áp dụng trong sản phẩm liên quan tới định danh người dùng điện tử (*electronic Know Your Customer - eKYC*) của các dịch vụ viễn thông, ngân hàng, tổ chức tín dụng, vv, hoặc kiểm tra hộ chiếu điện tử của hành khách ở các sân bay (*ePassport*), vv. Thông thường, công việc đối sánh ảnh khuôn mặt từ ảnh trên ID với người mang giấy tờ ID đó được thực hiện trực tiếp bằng mắt thường của các nhân viên giao dịch trong ngân hàng, cán bộ hải quan, công an, vv. Phương pháp đối sánh ảnh khuôn mặt đóng góp trong việc tự động hóa để tiết kiệm thời gian, nhân lực và chi phí cho các hoạt động liên quan đến thủ tục xác minh danh tính.

Việc so khớp ảnh trên ID và ảnh chân dung tự chụp có một số khó khăn nhất định bao gồm chất lượng ảnh chụp trên ID thấp, thay đổi gương mặt do tuổi tác, trang điểm, thiếu tập dữ liệu huấn luyện đa dạng do đây là dữ liệu bảo mật cá nhân. Gần đây, cùng với sự phát triển của công nghệ trí tuệ nhân tạo, đã có một số phương pháp đối sánh ảnh khuôn mặt ra đời sử dụng các kỹ thuật học máy (*machine learning*) và học sâu (*deep learning*). Trong giải pháp hữu ích này, chúng tôi đề xuất một phương pháp mới sử dụng công nghệ trí tuệ nhân tạo cho vấn đề đối sánh ảnh khuôn mặt cải thiện được các khó khăn hiện có để đạt được độ chính xác cao hơn, thời gian xử lý tối ưu hơn, đặc biệt là áp dụng hiệu quả trên tập dữ liệu gương mặt của người Việt Nam.

Bản chất kỹ thuật của giải pháp hữu ích

Mục đích của giải pháp hữu ích là đề xuất phương pháp đối sánh ảnh khuôn mặt giữa ảnh trên giấy tờ chứng thực cá nhân và ảnh chân dung tự chụp phục vụ cho các hệ thống xác minh danh tính với độ chính xác cao và thời gian tối ưu.

Để đạt được mục đích trên, phương pháp này gồm các bước:

Bước 1: tìm kiếm và căn chỉnh khuôn mặt trong ảnh; bước này được thực hiện dựa trên mô hình học sâu RetinaFace để xác định vị trí gương mặt trong ảnh và căn chỉnh gương mặt về phương thẳng đứng; kết hợp cùng một số kỹ thuật tiền xử lý ảnh đầu vào để khắc phục chất lượng ảnh khi ảnh được chụp từ nhiều thiết bị đa dạng, góc chụp và cự li chụp khác nhau.

Bước 2: trích xuất đặc trưng khuôn mặt; bước này được thực hiện dựa trên mô hình học sâu nén khuôn mặt Insightface để nén ảnh khuôn mặt đầu vào thành các vectơ đặc trưng dùng để đối sánh.

Bước 3: chuẩn hóa vectơ đặc trưng và tính toán sự khác biệt giữa các vectơ đặc trưng; bước này được thực hiện dựa trên phép chuẩn hóa L_2 -norm, sau đó ghép cặp để huấn luyện mô hình phân loại ở bước 4 dựa trên thuật toán tính độ tương tự cosin (*cosine similarity*) và tính độ lớn của hiệu các vectơ đặc trưng sau khi chuẩn hóa và ghép cặp.

Bước 4: phân loại đặc trưng và dự đoán hai khuôn mặt có phải cùng một người hay không; bước này được thực hiện dựa trên một mạng nơron sâu (*Deep neuron network - DNN*) để học các đặc trưng dùng cho việc đối sánh; khi đưa hai ảnh khuôn mặt đầu vào ở bước 1, đầu ra ở bước 4 trả về xác suất trùng khớp (%) của hai ảnh đó và đưa ra kết quả dự đoán.

Mô tả vắn tắt các hình vẽ

Hình 1: hình ảnh mô tả các bước trong phương pháp đối sánh ảnh khuôn mặt giữa ảnh trên giấy tờ chứng thực cá nhân và ảnh chân dung tự chụp sử dụng công nghệ trí tuệ nhân tạo

Hình 2: hình ảnh mô tả bước tìm kiếm và căn chỉnh khuôn mặt trong ảnh

Hình 3: hình ảnh mô tả bước phân loại đặc trưng và dự đoán hai khuôn mặt có phải cùng một người hay không

Mô tả chi tiết giải pháp hữu ích

Các bước và trình tự các bước của phương pháp đối sánh ảnh khuôn mặt giữa ảnh trên giấy tờ chứng thực cá nhân và ảnh chân dung tự chụp phục vụ cho các hệ thống định danh người dùng điện tử được mô tả trong tham chiếu Hình 1. Dữ liệu đầu vào là các ảnh trên ID và ảnh chân dung tự chụp được chụp từ nhiều thiết bị đa dạng và từ nhiều góc độ, cự li khác nhau.

Tuy nhiên, cũng cần phải nói thêm rằng trong sáng chế này, các thuật ngữ “RetinaFace”, “Insightface”, “ L_2 -norm”, “ResNet”, “ResNet100”, “ArcFace”, “Sigmoid”, “Cross-entropy”, “Adam” là các danh từ riêng (tên của mô hình, thuật toán, hàm số, vv) chưa có từ tiếng Việt thay thế nên xin được để ở dạng chuẩn quốc tế.

Các bước của phương pháp được thực hiện như sau:

Bước 1: tìm kiếm và căn chỉnh khuôn mặt trong ảnh.

Nhiệm vụ của bước này là tìm kiếm được vùng chứa gương mặt trong ảnh, căn chỉnh khuôn mặt về phương thẳng đứng, và cắt ra vùng chứa gương mặt đó. Trình tự của bước này bao gồm hai pha được mô tả trong Hình 2.

- Pha 1: sử dụng mô hình học sâu RetinaFace để tìm kiếm ra vùng chứa khuôn mặt (*bounding box*) và xác định năm vùng bộ phận (*landmarks*) của khuôn mặt bao gồm hai mắt, mũi và hai khóe miệng.

- Pha 2: dựa trên các vùng bộ phận khuôn mặt (*landmarks*) đã xác định được để căn chỉnh lại khuôn mặt theo phương thẳng đứng và cắt ra vùng khuôn mặt đã được căn chỉnh.

Do dữ liệu đầu vào là các ảnh trên ID và ảnh chân dung tự chụp được chụp từ nhiều thiết bị đa dạng và từ nhiều góc độ, cự li khác nhau nên chất lượng ảnh đầu vào tốt hay xấu tùy thuộc vào thiết bị sử dụng và điều kiện chụp ảnh (phông nền, độ sáng, vv). Vì vậy, cần kết hợp với các kỹ thuật tiền xử lý ảnh đầu vào phù hợp, đặc biệt là với ảnh trên ID do ảnh

này chứa gương mặt có thể rất nhỏ và mờ. Với các ảnh như vậy, kỹ thuật tiền xử lý được mô tả như sau:

- Kỹ thuật 1 sử dụng cho ảnh trên ID bị nhỏ và mờ: sử dụng mô hình học sâu RetinaFace để tìm kiếm ra vùng chứa khuôn mặt (*bounding box*) ban đầu từ giấy tờ chứng thực cá nhân, coi vùng chứa khuôn mặt này như một ảnh đầu vào mới với mục đích xác định năm vùng bộ phận (*landmarks*) chính xác hơn.

- Kỹ thuật 2 sử dụng cho ảnh chân dung tự chụp: với mục đích giảm thời gian xử lý và đạt được độ chính xác cao hơn trong việc xác định vị trí khuôn mặt với nhiều cự li chụp khác nhau, kỹ thuật lựa chọn (*grid search*) kích cỡ ảnh đầu vào phù hợp với mô hình RetinaFace được sử dụng. Cụ thể, nhiều kích cỡ (*multi-scale*) chiều rộng ảnh được đề xuất như [800; 700; 500; 300] điểm ảnh (*pixels*) để thay đổi (*resize*) lại ảnh đầu vào và lựa chọn kích cỡ tốt nhất.

Bước 2: trích xuất đặc trưng khuôn mặt.

Nhiệm vụ của bước này là nén ảnh khuôn mặt đầu vào (chính là ảnh khuôn mặt đã được căn chỉnh và cắt ra từ bước 1) thành các vectơ đặc trưng đầu ra để phục vụ vấn đề đối sánh khuôn mặt. Một mạng nơron tích chập (*Convolutional neural network - CNN*) được đề xuất sử dụng ở bước này là mô hình học sâu nén khuôn mặt Insightface. Mô hình Insightface được xây dựng dựa trên mô hình cơ sở là ResNet100 như một mạng nhúng, sau tầng CNN cuối cùng của mạng ResNet, kỹ thuật chuẩn hóa theo lô (*batch normalization*) và học bỏ qua (*drop-out*) được áp dụng, kết quả đầu ra của mô hình này là vectơ đặc trưng khuôn mặt có chiều là 1×512 . Đặc biệt, mô hình InsightFace sử dụng hàm giá trị ArcFace (*ArcFace loss*) để trích chọn được những đặc trưng tốt nhất dùng để phân loại trong các bài toán nhận diện khuôn mặt.

Bước 3: chuẩn hóa vectơ đặc trưng và tính toán sự khác biệt giữa các vectơ đặc trưng.

Các vectơ đặc trưng khuôn mặt được chuẩn hóa theo L_2 -norm như công thức sau (đây là một công thức phổ biến có sẵn được sử dụng trong nhiều lĩnh vực đa dạng; tuy nhiên, công thức được nêu ở đây để chỉ ra cách áp dụng trong vấn đề của giải pháp hữu ích):

$$\|x\|_2 = \sqrt{(\sum_{i=1}^n x_i^2)}; \quad \hat{x} = \frac{x}{\|x\|_2} \quad (1)$$

trong đó x là một vectơ đặc trưng, $x \in \mathbb{R}^n$, và $n = 512$ chính là đầu ra của mô hình nén khuôn mặt InsightFace. Sau khi các vectơ đặc trưng được chuẩn hóa, sự khác biệt giữa các khuôn mặt được tính toán dựa trên độ lớn của hiệu hai vectơ đặc trưng giữa ảnh trên ID và ảnh chân dung tự chụp của “cùng người” và “khác người”. Giả sử, hai vectơ đặc trưng là a and b , độ lớn hiệu số được tính theo công thức sau:

$$f_{ab} = |a_{norm} - b_{norm}| \quad (2)$$

trong đó a_{norm} và b_{norm} là các vectơ sau khi chuẩn hóa. f_{ab} sẽ được dùng làm đầu vào của bước 4.

Việc ghép cặp để tính hiệu số f_{ab} và sử dụng cho việc huấn luyện mô hình phân loại ở bước 4 dựa trên thuật toán tính độ tương tự cosin. Ý tưởng sử dụng thuật toán này dựa trên việc khi giảm thiểu được sự khác biệt (f_{ab}) giữa ảnh trên ID và ảnh chân dung tự chụp từ cùng một người, đồng thời tăng cường sự khác biệt (f_{ab}) giữa ảnh trên ID và ảnh chân dung tự chụp của những người khác, thì khả năng phân loại hai nhãn “cùng người” và “khác người” sẽ tăng lên. Thuật toán tính độ tương tự cosin đo độ giống nhau giữa hai vectơ của một không gian tích bên trong. Độ tương tự giữa chúng được đo dựa trên tính toán cosin góc giữa hai vectơ đó. Cho hai vectơ A và B , độ tương tự cosin được biểu diễn theo công thức sau:

$$similarity = \cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}} \quad (3)$$

trong đó $\cos(\theta)$ là độ tương tự cosin, và A_i và B_i lần lượt là các thành phần thứ i của hai vectơ A và B .

Bước 4: phân loại đặc trưng và dự đoán hai khuôn mặt có phải cùng một người hay không.

Sau khi đã ghép cặp các vectơ dựa trên thuật toán đo độ tương tự cosin và tính toán các độ lớn hiệu số f_{ab} , các hiệu số này sẽ được dùng để huấn luyện mô hình phân loại. Một mạng nơron sâu (DNN) với ba tầng được đề xuất sử dụng ở bước này, mô tả của mạng này tham chiếu ở Hình 3. Đầu vào của mạng các vectơ có chiều 1×512 . Tầng thứ hai của mạng được xây dựng với 16 nơron để học các đặc trưng từ đầu vào với hàm kích hoạt biến đổi tuyến tính hiệu chỉnh (ReLU - Rectified Linear Unit), đầu ra của mạng là một nơron với hàm kích hoạt là Sigmoid, mục đích là trả về xác suất đối sánh khuôn mặt thuộc về “cùng người” hay “khác người”. Các tham số và siêu tham số (*hyper-parameters*) bao gồm số lượt huấn luyện toàn bộ dữ liệu là 200 (*epoch*), tốc độ học (*learning rate*) là 10^{-3} , kích thước lô khi huấn luyện (*batch size*) là 64, hệ số chính quy hóa L_2 (L_2 regularization) là $\lambda = 10^{-2}$, hàm mất mát (*loss*) là Cross-entropy, hàm tối ưu hóa (*optimizer*) là Adam với các tham số: tốc độ học (*learning rate*) là 10^{-3} , các tốc độ phân rã (*exponential decay*) lần lượt là $\beta_1 = 0.9$, $\beta_2 = 0.999$, và hệ số $\epsilon = 10^{-7}$ để ngăn chặn các lỗi chia cho 0.

Sau khi mạng nơron sâu (DNN) trả về xác suất hai khuôn mặt khớp nhau là bao nhiêu phần trăm (%), một ngưỡng phân loại sẽ được đặt ra để dự đoán ảnh hai khuôn mặt là “cùng người” hay “khác người”. Trong giải pháp hữu ích này, ngưỡng được đề xuất dựa trên thực nghiệm là 60%.

Giải pháp hữu ích được mô tả chi tiết bằng cách sử dụng các phương án được mô tả ở trên. Tuy nhiên, rõ ràng là đối với người hiểu biết trung bình trong lĩnh vực giải pháp hữu ích không bị giới hạn ở phương án được mô tả trong phần mô tả giải pháp hữu ích. giải pháp hữu ích có thể được thực hiện ở chế độ cải biến hoặc thay đổi mà không nằm ngoài phạm vi giải pháp hữu ích được xác định bởi các điểm yêu cầu bảo hộ. Vì vậy, những gì được mô tả trong phần mô tả giải pháp hữu ích chỉ nhằm mục đích minh họa, và sẽ không áp đặt bất kỳ giới hạn nào đối với giải pháp hữu ích.

Yêu cầu bảo hộ

1. Phương pháp đối sánh ảnh khuôn mặt giữa ảnh trên giấy tờ chứng thực cá nhân và ảnh chân dung tự chụp sử dụng công nghệ trí tuệ nhân tạo, bao gồm các bước sau:

bước 1: tìm kiếm và căn chỉnh khuôn mặt trong ảnh;

tại bước này, sử dụng mô hình học sâu retinaface để tìm kiếm ra vùng chứa khuôn mặt (*bounding box*) và xác định năm vùng bộ phận (*landmarks*) để căn chỉnh lại khuôn mặt theo phương thẳng đứng và cắt ra vùng khuôn mặt đã được căn chỉnh; do dữ liệu đầu vào là các ảnh trên giấy tờ chứng thực cá nhân (chứng minh nhân dân, căn cước công dân, hộ chiếu) và ảnh chân dung tự chụp được chụp từ nhiều thiết bị đa dạng và từ nhiều góc độ, cự li khác nhau nên chất lượng ảnh đầu vào đa dạng; một số kỹ thuật tiền xử lý ảnh đầu vào được áp dụng, gồm:

- kỹ thuật 1 sử dụng mô hình học sâu retinaface để tìm kiếm ra vùng chứa khuôn mặt (*bounding box*) ban đầu từ giấy tờ chứng thực cá nhân, coi vùng chứa khuôn mặt này như một ảnh đầu vào mới với mục đích xác định năm vùng bộ phận (*landmarks*) chính xác hơn;

- kỹ thuật 2 sử dụng cho ảnh chân dung tự chụp: với mục đích giảm thời gian xử lý và đạt được độ chính xác cao hơn trong việc xác định vị trí khuôn mặt với nhiều cự li chụp khác nhau, kỹ thuật lựa chọn (*grid search*) kích cỡ ảnh đầu vào phù hợp với mô hình retinaface được sử dụng; cụ thể, nhiều kích cỡ (*multi-scale*) chiều rộng ảnh được đề xuất như [800; 700; 500; 300] điểm ảnh (*pixels*) để thay đổi (*resize*) lại ảnh đầu vào và lựa chọn kích cỡ tốt nhất;

bước 2: trích xuất đặc trưng khuôn mặt;

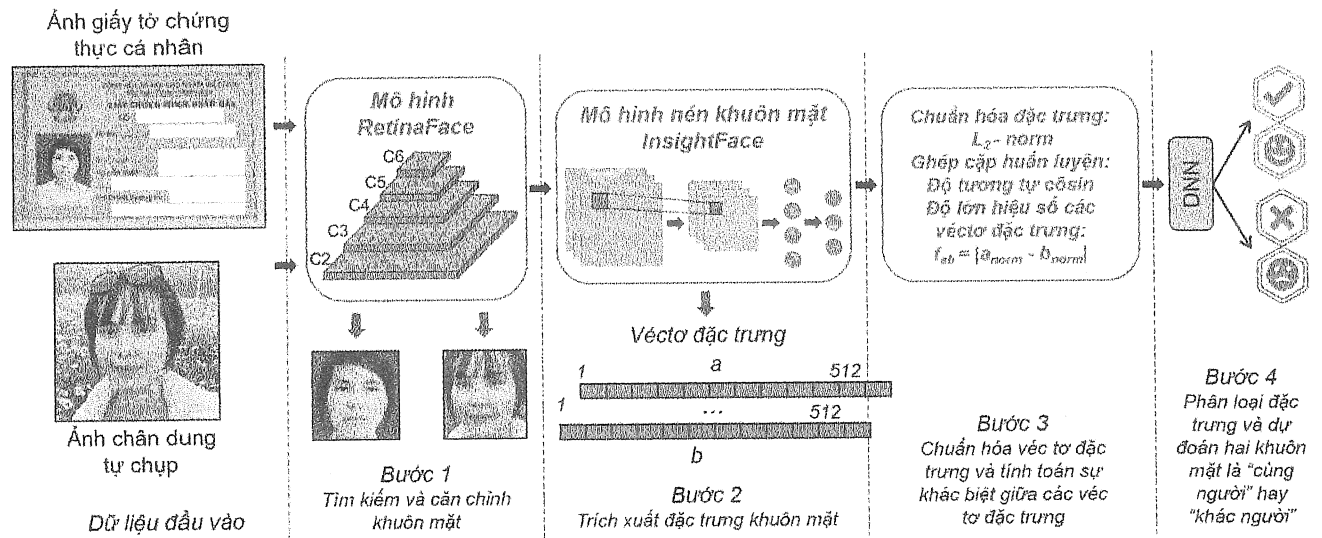
tại bước này, một mạng nơ-ron tích chập (*CNN*) được đề xuất sử dụng là mô hình học sâu nén khuôn mặt insightface với chức năng là nén ảnh khuôn mặt đầu vào thành các vectơ đặc trưng đầu ra có chiều là 1×512 để phục vụ cho bài toán đối sánh khuôn mặt;

bước 3: chuẩn hóa vectơ đặc trưng và tính toán sự khác biệt giữa các vectơ đặc trưng;

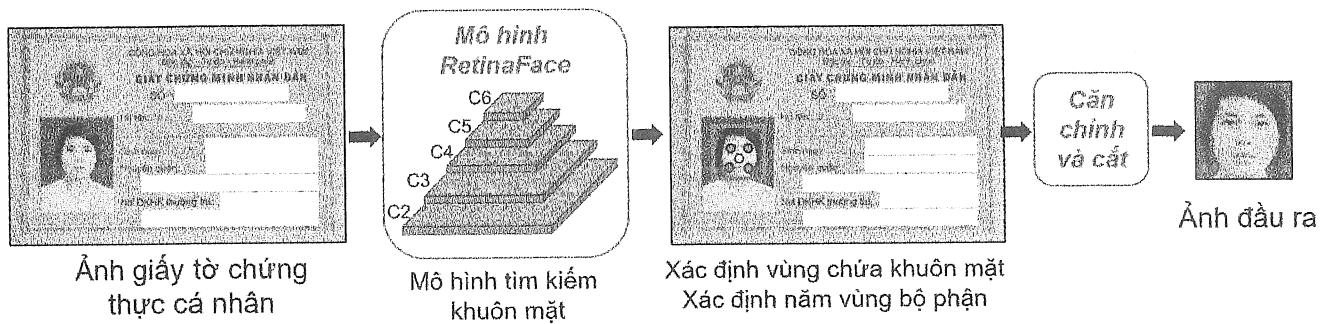
tại bước này, kỹ thuật chuẩn hóa L_2 -norm được áp dụng cho các vectơ đặc trưng, sau đó ghép cặp để huấn luyện mô hình phân loại dựa trên thuật toán tính độ tương tự cosin (*cosine similarity*) và tính độ lớn của hiệu các vectơ đặc trưng sau khi chuẩn hóa và ghép cặp.

bước 4: phân loại đặc trưng và dự đoán hai khuôn mặt có phải cùng một người hay không;

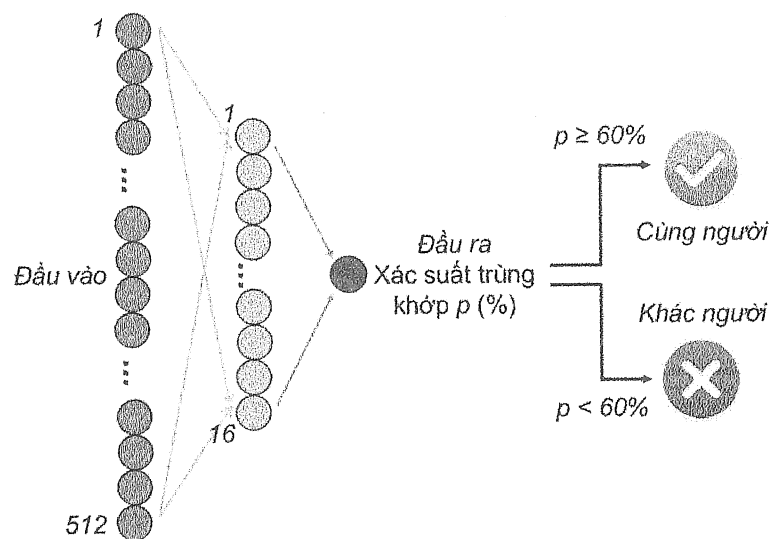
tại bước này, một mạng nơron sâu (*DNN*) với ba tầng được đề xuất sử dụng: đầu vào của mạng là các vectơ đặc trưng có chiều là 1×512 , tầng thứ hai của mạng được xây dựng với 16 nơron để học các đặc trưng từ đầu vào với hàm kích hoạt biến đổi tuyến tính hiệu chỉnh (*ReLU*), đầu ra của mạng là một nơron với hàm kích hoạt là Sigmoid, mục đích là trả về xác suất hai khuôn mặt khớp nhau là bao nhiêu %; tiếp theo, một ngưỡng phân loại sẽ được đặt ra để dự đoán ảnh hai khuôn mặt là “*cùng người*” hay “*khác người*”; để phân loại đặc trưng và dự đoán hai khuôn mặt có phải cùng một người hay không; một mạng nơron sâu (*DNN*) gồm ba tầng được huấn luyện với các tham số và siêu tham số (*hyper-parameters*) như sau: số lượt huấn luyện trên toàn bộ dữ liệu là 200 (*epoch*), tốc độ học (*learning rate*) là 10^{-3} , kích thước huấn luyện theo lô (*batch size*) là 64, hệ số chính quy hóa L_2 (*L_2 regularization*) là $\lambda = 10^{-2}$, hàm mất mát (*loss*) là cross-entropy, hàm tối ưu hóa (*optimizer*) là adam với các tham số: tốc độ học (*learning rate*) là 10^{-3} , các tốc độ phân rã (*exponential decay*) lần lượt là $\beta_1 = 0,9$, $\beta_1 = 0,999$, và hệ số $\varepsilon = 10^{-7}$ để ngăn chặn các lỗi chia cho 0.



Hình 1



Hình 2



Hình 3