

ĐỔI MỚI LỚP HỌC TÍCH HỢP THÔNG MINH: KẾT HỢP XÁC THỰC NGƯỜI DÙNG VÀ ĐÁNH GIÁ CẢM XÚC

Đoàn Thiện Minh^{1*}, Phan Mạnh Thường¹, Nguyễn Tấn Lợi¹, Phạm Mạnh Trí¹, Hà Nguyễn Lê Hoàng²,
Nguyễn Thị Vĩnh³

¹Trường Đại học Lạc Hồng, số 10 Huỳnh Văn Nghệ, Bửu Long, Biên Hòa, Đồng Nai, Việt Nam

²Trường Cao đẳng Kỹ thuật Đồng Nai, số 88, Nguyễn Văn Hoa, KP.1, P. Thống Nhất, Biên Hòa, Đồng Nai, Việt Nam

³THPT Nguyễn Thị Minh Khai, khu phố Đông Thành, phường Tân Đông Hiệp, thành phố Dĩ An, Bình Dương, Việt Nam

* Tác giả liên hệ: dtminh@lhu.edu.vn

THÔNG TIN BÀI BÁO

Ngày nhận: 2/7/2024
Ngày hoàn thiện: 8/8/2024
Ngày chấp nhận: 16/9/2024
Ngày đăng: 11/10/2024

TỪ KHÓA

MTCNN;
Mạng lưới Facenet;
Xác thực người dùng;
Nhận dạng cảm xúc;
Mạng nơ-ron tích chập.

TÓM TẮT

Giải pháp đổi mới lớp học tích hợp thông minh trong nghiên cứu sử dụng kỹ thuật MTCNN và FaceNet được ứng dụng để xây dựng hệ thống có khả năng xác thực thông tin người dùng dựa trên khuôn mặt. Hơn thế nữa, ứng dụng có thể thực hiện các tính năng như xác thực điểm danh, phát hiện người dùng trong một nhóm nhiều người, đánh giá cảm xúc người học. Chúng tôi đã so sánh đánh giá mô hình xác thực người dùng cùng một số mô hình như: MTCNN và FaceNet, VGG16 và OpenVino. Kết quả cho thấy, sự kết hợp giữa MTCNN và FaceNet đạt được độ chính xác tối ưu, có thể đến 96,4% cho bài toán xác thực định danh người dùng nhằm phục vụ chức năng điểm danh người học. Bên cạnh đó, để phục vụ chức năng đánh giá cảm xúc người học, nghiên cứu sử dụng mạng nơ ron tích chập (CNN) dựa trên bộ dữ liệu Fer 2013 và Fer Plus để đánh giá các trạng thái cảm xúc người học trong khi thi.

Việc đổi mới hệ thống quản lý lớp học tích hợp thông minh dựa trên việc nhận dạng khuôn mặt sử dụng MTCNN và Face Net cùng với tính năng đánh giá cảm xúc người học trong trường đại học có thể nâng cao hiệu quả giảng dạy, tăng chất lượng đào tạo và theo dõi trạng thái của sinh viên trong quá trình thi.

INNOVATION THE SMART-INTEGRATED CLASSROOM: COMBINING USER AUTHENTICATION AND EMOTION ASSESSMENT

Doan Thien Minh^{1*}, Phan Manh Thuong¹, Nguyen Tan Loi¹, Pham Manh Tri¹, Ha Nguyen Le Hoang²,
Nguyễn Thị Vĩnh³

¹Lac Hong University, 10 Huynh Van Nghe Street, Buu Long, Bien Hoa, Dong Nai, Vietnam

²Dong Nai Technical College, No. 88, Nguyen Van Hoa, KP.1, Thong Nhat Ward, Bien Hoa, Dong Nai, Vietnam

³Nguyen Thi Minh Khai High school, Dong Thanh quarter, Tan Dong Hiep ward, Di An city, Binh Duong, Vietnam

*Corresponding Author: dtminh@lhu.edu.vn

ARTICLE INFO

Received: Jul 2nd, 2024
Revised: Aug 8th, 2024
Accepted: Sep 16th, 2024
Published: Oct 11st, 2024

KEYWORDS

MTCNN;
Facenet;
User Authentication;
Emotion recognition;
Convolutional Neural Networks.

ABSTRACT

The innovation of the smart integrated classroom solution in the research utilizes MTCNN and FaceNet techniques to develop a system capable of verifying user information based on facial recognition. Moreover, the application can perform features such as attendance verification, detecting users in a group of multiple people, and assessing the emotions of learners. We have compared the user verification model with several models such as MTCNN and FaceNet, VGG16, and OpenVino. The results show that the combination of MTCNN and FaceNet achieved optimal accuracy, up to 96.4%, for the task of user identification to serve the attendance function for learners. Additionally, for the emotion assessment function, the study uses convolutional neural networks (CNN) based on the Fer2013 and Fer Plus datasets to evaluate the emotional states of learners during exams.

The innovation of the smart integrated classroom management system, based on facial recognition using MTCNN and FaceNet, along with the emotional assessment feature for learners in universities, can enhance teaching effectiveness, improve training quality, and monitor the status of students during exams.

Doi: <https://doi.org/10.61591/jslhu.18.506>

Available online at: <https://js.lhu.edu.vn/index.php/lachong>

1. ĐẶT VẤN ĐỀ

Việc xác thực và điểm danh người học tại các trường đại học thường phụ thuộc vào các phương pháp truyền thống như định danh thủ công qua đọc tên, quét mã thẻ RFID, hoặc tiên tiến hơn như tại Đại học Lạc Hồng sử dụng mã QR (QR Code-Quick response code). Tuy nhiên, các phương thức xác thực đơn thuần tiềm ẩn nhiều rủi ro dễ dàng có thể xảy ra sai sót, đặc biệt vào các thời gian quan trọng như trong các kỳ thi. Để đạt mục tiêu xác thực chính xác mà nhanh chóng rời khỏi sai lầm trong xác thực thì đòi hỏi cần có phương pháp hiệu quả hơn. Hiện nay, việc sử dụng phương pháp nhận diện khuôn mặt mang lại nhiều lợi ích đáng kể. Lợi ích đầu tiên, tính tiện lợi cao khi không yêu cầu người dùng phải nhớ mật khẩu hay mang theo thẻ. Thay vào đó, chỉ cần một cái nhìn vào camera là có thể xác thực người dùng một cách nhanh chóng và thuận tiện. Điều này không chỉ giảm bớt độ phức tạp mà còn giúp giảm nguy cơ quên mật khẩu, một vấn đề phổ biến trong các hệ thống sử dụng mật khẩu. Thứ hai, tính an toàn và khó sao chép của dữ liệu khuôn mặt là một yếu tố quan trọng. So với thẻ hoặc mật khẩu, dữ liệu khuôn mặt khó bị sao chép và sử dụng một cách trái phép. Công nghệ nhận diện khuôn mặt thường liên quan đến việc sử dụng thông tin khuôn mặt có thể biểu diễn dưới dạng một chuỗi số học khó có thể làm giả, tăng tính an toàn cho hệ thống. Vì thế, giải pháp nhận diện khuôn mặt điểm danh người dùng mong muốn sẽ giúp cho việc điểm danh, xác thực người dùng trong nhóm người được chính xác và dễ dàng hơn.

Bên cạnh việc xác thực người học, một vấn đề quan trọng không kém là đánh giá cảm xúc người học trong quá trình học tập, thi cử. Cụ thể hơn là đánh giá cảm xúc người học trong kỳ thi. Việc đánh giá cảm xúc người học sẽ giúp tăng tính tương tác cá nhân với người học. Việc đánh giá cảm xúc người học sẽ giúp giảng viên, người quản trị hiểu rõ hơn về trạng thái tâm lý của từng người học, từ đó có thể điều chỉnh phương pháp giảng dạy phù hợp, kịp thời đưa ra các giải pháp. Khi biết được cảm xúc của học sinh, giảng viên có thể đưa ra phản hồi ngay lập tức, giúp học sinh cải thiện kịp thời và tránh được sự tích tụ của các vấn đề tâm lý.

Việc đánh giá cảm xúc người học có thể cải thiện chất lượng kỳ thi, giúp giảng viên hiểu rõ hơn về những yếu tố ảnh hưởng đến kết quả thi, chẳng hạn như lo lắng thi cử, áp lực tâm lý. Điều này giúp thiết kế kỳ thi công bằng và chính xác hơn. Hơn thế nữa, thông qua quá trình đánh giá cảm xúc, giảng viên và nhà quản lý giáo dục có thể cung cấp hỗ trợ cá nhân hóa cho các trường hợp người học gặp khó khăn về tâm lý trong quá trình thi cử, giúp họ có điều kiện tốt nhất để thể hiện khả năng của mình. Chính vì những lý do đó, lớp học tích hợp thông minh với nhiều đổi mới như tính năng xác thực định danh người học, xác thực người học trong nhóm người, tính năng đánh giá cảm xúc người học trong kỳ thi đóng vai trò vô cùng quan trọng.

2. CƠ SỞ LÝ THUYẾT VÀ MÔ HÌNH NGHIÊN CỨU

2.1 Xác thực định danh người dùng

Nhận diện và xác thực người dùng dựa trên việc nhận diện khuôn mặt là một bài toán lớn, khi giải quyết ta sẽ chia bài toán này ra nhiều giai đoạn và thành phần nhưng

quan trọng nhất là tập trung vào hai chức năng chính: Phát hiện khuôn mặt (Face Detection) và xác thực khuôn mặt (Face Verification).

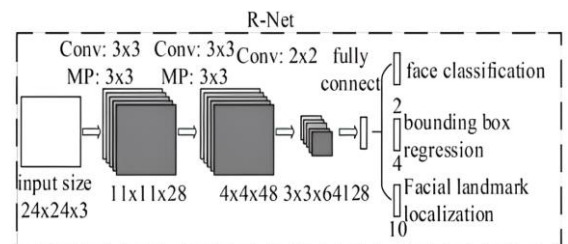
Phát hiện khuôn mặt (Face Detection)

Mục tiêu của bước phát hiện khuôn mặt (Face Detection) là xác định vị trí và kích thước của các khuôn mặt trong ảnh hoặc khung hình video. Mô hình MTCNN (Multi-task Cascaded Convolutional Networks) được sử dụng để thực hiện công việc này. Mô hình được đề xuất năm 2016 bởi Zhang và cộng sự [1]. Mô hình này bao gồm 3 lớp mạng CNN (Cascaded Convolutional Networks) khác biệt xếp chồng và đồng thời hoạt động khi nhận diện khuôn mặt, tương trưng cho 3 lớp mạng CNN chính là P-Net, R-Net và O-Net. Độ chính xác của P-Net, R-Net và O-Net đã được chứng minh tối ưu đối với bài toán nhận diện khuôn mặt người [1,2]. Mỗi lớp mạng có cấu trúc và đảm nhiệm vai trò khác nhau. Đầu ra của MTCNN chính là vị trí khuôn mặt và các điểm trên mặt như: mắt, mũi, miệng... Đầu tiên, ảnh được thay đổi kích thước ảnh gốc thành một loạt các ảnh bản sao với kích thước thu nhỏ dần đều (Image Pyramid). Mạng P-Net được áp dụng cho từng bản sao ảnh này để tìm kiếm các khuôn mặt. Kết quả đầu ra của giai đoạn P-Net là vị trí ước tính của các bounding box (một khung ảnh vừa đủ) chứa khuôn mặt và điểm confident (nhận diện) của mỗi bounding box.



Hình 1. Image Pyramid

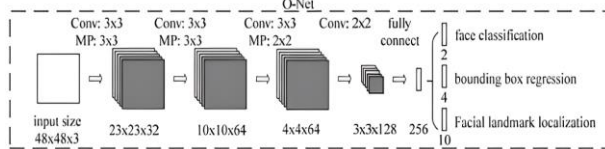
Stage 2: R-Net: Mạng R (Refine Network) là giai đoạn tiếp theo sau P-Net thực hiện các bước như mạng P. Giai đoạn này còn sử dụng một phương pháp gọi là “padding”. Phương pháp padding sẽ giúp chèn thêm các zero-pixels vào các phần còn thiếu của bounding box nếu bounding box bị vượt quá phạm vi của ảnh. Tất cả các bounding box lúc này sẽ được chuyển về cùng một kích thước 24x24, được chuyển vào mạng R, sau đó sẽ nhận được kết quả là những tọa độ mới của các box còn lại và sẽ được sử dụng để đưa vào mạng xử lý tiếp theo, mạng O.



Hình 2. Cấu trúc R-Net

Stage 3: O-Net. Giai đoạn O-Net (Output Network) là giai đoạn cuối cùng trong MTCNN. Mạng O-Net cũng thực hiện tương tự như mạng R-Net nhưng thay đổi kích thước thành 48x48. Kết quả của mạng O-Net không còn chỉ là các tọa độ của các box nữa mà trả về 3 giá trị bao gồm: 4 tọa độ của bounding box; tọa độ 5 điểm landmark

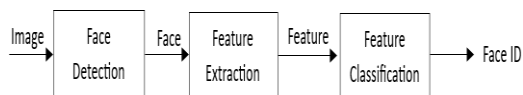
(đặc điểm nhận dạng) trên mặt, bao gồm 2 mắt, 1 mũi, 2 bên cánh môi và điểm confident của mỗi box. Cuối cùng những mạng này sẽ được lưu vào thành 1 dictionary với 3 giá trị kể trên.



Hình 3. Cấu trúc O-Net

Xác thực khuôn mặt (Face Verification):

Vấn đề quan trọng trong bài toán xác thực khuôn mặt là xác định rõ ràng xem hình ảnh hiện tại có khớp với thông tin, khuôn mặt trong hệ thống không. Mô hình trợ giúp giải quyết vấn đề này là FaceNet, được giới thiệu bởi google vào năm 2015 [3]. Model này chỉ cần đưa ảnh vào (quy định kích thước đúng) thì FaceNet sẽ trả về một vector 128 chiều đặc biệt cho một mặt. Sau đó, SVM (Support Vector Machine: là một thuật toán học máy phân loại được sử dụng rộng rãi trong việc giải quyết các bài toán phân loại và hồi quy) để phân loại và nhận diện mặt. Vector này đại diện cho các đặc trưng của khuôn mặt và có thể so sánh với các vector khác để xác định độ tương đồng giữa các khuôn mặt để biết vector đại diện cho khuôn mặt.



Hình 4. Mô hình xác thực khuôn mặt

Mô hình xác thực khuôn mặt cơ bản thường đầu tiên phát hiện khuôn mặt (phát hiện một hoặc nhiều khuôn mặt trong một hình ảnh). Kế tiếp, mô hình trích xuất đặc trưng khuôn mặt (trích xuất các đặc trưng quan trọng nhất từ một hình ảnh khuôn mặt, tiếp theo là phân loại khuôn mặt, việc phân loại khuôn mặt dựa trên các đặc trưng được trích xuất. Trong nghiên cứu này sẽ tập trung vào các phương pháp học sâu phổ biến, cụ thể hơn nghiên cứu thực hiện phát hiện khuôn mặt bằng MTCNN, trích xuất tính năng bằng FaceNet và phân loại bằng Softmax.

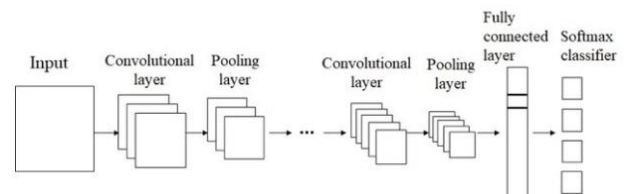
2.2 Phân loại đánh giá cảm xúc

Cảm xúc người học trong khi thi được tiến hành nghiên cứu là các cảm xúc gồm: Angry, Disgust, Fear, Happy, Neutral, Sad, Surprise (Giận dữ, ghê tởm, sợ hãi, vui vẻ, bình thường, buồn bã, ngạc nhiên) được thể hiện trên khuôn mặt người dùng khi họ quét khuôn mặt trên camera. Việc nghiên cứu nhận diện các cảm xúc trên khuôn mặt người dùng qua camera yêu cầu môi trường ánh sáng và điều kiện quét tối ưu. Điều này bao gồm ánh sáng đồng đều, không có bóng đổ mạnh, tránh ánh sáng nền mạnh, và sử dụng ánh sáng tán xạ. Đồng thời, cần đảm bảo điều kiện môi trường ít nhiễu, sử dụng camera chất lượng cao và quét khuôn mặt từ góc độ trực diện.

Mô hình phân loại cảm xúc cũng dựa trên mô hình mạng nơ ron tích chập (Convolutional Neural Network-CNN)[4]. Mạng nơ ron tích chập là một loại mô hình mạng thần kinh đặc biệt giải quyết các vấn đề phức tạp ví dụ như xử lý dữ liệu hình ảnh. Mạng nơ ron tích chập huấn luyện máy tính xử lý dữ liệu dựa trên việc mô phỏng

cách thức hoạt động của mạng nơ ron thần kinh giống não người. Kỹ thuật xử lý của mạng nơ ron tích chập sử dụng các nút hoặc nơ ron liên kết với nhau trong một cấu trúc phân lớp. Mỗi nút trong một lớp được kết nối với tất cả các nút trong lớp tiếp theo và có nhiều lớp ẩn. Mục tiêu mạng nơ ron tích chập nhằm tạo ra một hệ thống huấn luyện máy học hồi và liên tục cải tiến dựa trên dữ liệu huấn luyện.

Convolutional Neural Network bao gồm: các lớp tích chập (convolutional layers), lớp tổng hợp (Pooling layer), lớp kết nối đầy đủ (Fully Connected), các lớp tích chập để trích xuất các đặc trưng quan trọng từ ảnh. Lớp tổng hợp để giảm bớt số lượng tham số khi hình ảnh quá lớn mà vẫn giữ được các thông tin quan trọng. Lớp kết nối đầy đủ ở cuối có nhiệm vụ tối ưu hóa mục tiêu để kết nối tất cả thông tin học được và đưa ra kết luận xử lý như dự đoán kết quả. Mạng nơ ron tích chập (CNN) tạo ra nhiều feature maps mới từ dữ liệu đầu vào. Cụ thể, mỗi lớp tích chập trong CNN tạo ra một tập hợp các feature maps mới dựa trên dữ liệu đầu vào và các bộ lọc.



Hình 5. Mô hình Mạng Nơ ron tích chập cơ bản

3. THỰC NGHIỆM

Bộ dữ liệu huấn luyện xác thực định danh người dùng để huấn luyện gồm hơn 35206 ảnh.

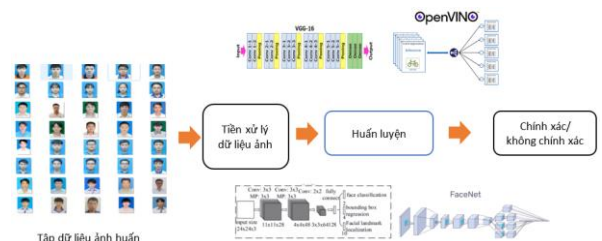
Quy trình huấn luyện mô hình gồm các bước thực hiện:

Bước 1: Chuẩn bị tập dữ liệu ảnh huấn luyện.

Bước 2: Tiền xử lý dữ liệu ví dụ như: xử lý cắt ảnh khuôn mặt và điều chỉnh về kích thước 160x160 pixel.

Bước 3: Huấn luyện ảnh.

Bước 4: Thử nghiệm kết quả huấn luyện của model, sử dụng bộ dữ liệu thử nghiệm để tiến hành kiểm tra độ chính xác của model.



Hình 6. Mô hình phân loại xác thực người dùng

Nhằm chọn ra được mô hình để sử dụng cho việc nhận diện và phân loại xác thực người dùng hiệu quả, nghiên cứu đã thực nghiệm đánh giá bộ dữ liệu trên ba mô hình MTCNN -FaceNet, VGG16, OpenVino.

Mô hình kết hợp MTCNN và FaceNet là gồm có 2 mô hình kết hợp lại với nhau, với nhiệm vụ và chức năng của mỗi mô hình là MTCNN sẽ nhận nhiệm vụ phát hiện khuôn mặt còn FaceNet sẽ nhận nhiệm vụ phân loại nhận diện khuôn mặt được MTCNN phát hiện. MTCNN và

FaceNet cũng được nhóm tác giả nghiên cứu Siyao Qi và đồng tác giả thực hiện cho độ chính xác đến 86.0% [11]. Hiệu quả của mô hình kết hợp Facenet cũng được nhóm nghiên cứu R E Saragih đánh giá khá tốt với độ chính xác cao trong khoảng từ 88%-97% trong bài báo khảo sát về nhận diện khuôn mặt "A Survey of Face Recognition Based on Convolutional Neural Network" [12].

Mô hình VGG16 [8] là mô hình nơ ron tích chập (convolutional neural network) có độ sâu 16 lớp. VGG16 sử dụng kiến trúc có bộ lọc tích chập rất nhỏ (3×3) nhằm đạt được sự cải thiện đáng kể đối với các cấu hình tiên tiến bằng việc đẩy độ sâu lên trọng số 16 lớp. Mô hình VGG16 sử dụng 13 lớp ẩn và 3 lớp kết nối đầy đủ. Mạng được đào tạo trước có thể phân loại hình ảnh thành 1000 loại đối tượng. Mặc dù nhóm tác giả bài báo "Automate facial paralysis detection usingvgg architectures" [8], vận dụng mô hình VGG16 nhằm nắm bắt các mối quan hệ phức tạp trong hình ảnh khuôn mặt để xác định bệnh nhân bị liệt mặt, đem lại độ chính xác khá cao, từ 83,55% tới 92,15% nhưng khi vận dụng mô hình VGG16 trong việc xác định khuôn mặt và phân loại cảm xúc thì mô hình VGG16 không đem lại độ chính xác tốt trong việc nhận dạng khuôn mặt và xác định thông tin như mô hình Open VINO, MTCNN và FACENET. Việc mô hình VGG16 cùng một số mô hình CNN trong quá trình thực nghiệm phát hiện xác thực với độ chính xác trong khoảng 52% - 84% được tổng hợp đáng giá qua bài tổng quan của nhóm tác giả Rahmeh Abou Zafra cùng đồng tác giả [13] cũng như được thực nghiệm bởi nhóm tác giả Tony Gwyn và đồng sự thực nghiệm [14]. Tóm lại, mô hình VGG16 thực sự hiệu quả trong các bài toán về phân loại nhị phân hay phân loại đa nhân nhưng chưa phù hợp so với OpenViNo, MTCNN và FACENET trong việc phát hiện và xác thực khuôn mặt.

OpenVINO (Open Visual Inference and Neural Network Optimization) được phát triển bởi Intel [9,10], được thiết kế để tối ưu hóa và triển khai các mô hình học sâu cho các ứng dụng suy luận trên nhiều loại phần cứng khác nhau. Khi áp dụng trong việc xác thực khuôn mặt, OpenVINO có thể mang lại nhiều lợi ích về hiệu suất và khả năng triển khai. Qua một số thực nghiệm, kết quả thể hiện mô hình OpenVINO là một công cụ mạnh mẽ và linh hoạt, độ chính xác lên đến 98% khi xác thực trong thực nghiệm. Tuy nhiên, một số lý do trong quá trình thực nghiệm khi nghiên cứu ảnh hưởng đến việc triển khai OpenVINO như: phụ thuộc vào phần cứng của Intel, hiệu suất kém trên phần cứng không phải Intel, khó khăn trong việc chuyển đổi mô hình, không phải tất cả các mô hình đều hỗ trợ, hạn chế trong việc tối ưu hóa mô hình: giới hạn tối ưu hóa, mặc dù OpenVINO cung cấp nhiều công cụ tối ưu hóa, không phải lúc nào cũng đạt được hiệu suất tối ưu như mong đợi, đặc biệt đối với các mô hình phức tạp hoặc các lớp tùy chỉnh. Và vì mục tiêu của bài báo sau khi phát hiện khuôn mặt, xác thực thông tin có thể tiếp tục vận dụng mô hình mạng nơ ron tích chập để đánh giá cảm xúc người học. Chính vì vậy, mặc dù OpenVINO sau khi thực nghiệm có độ chính xác cao nhưng nghiên cứu không lựa chọn mô hình Open VINO.



Hình 7. Ứng dụng xác thực người dùng

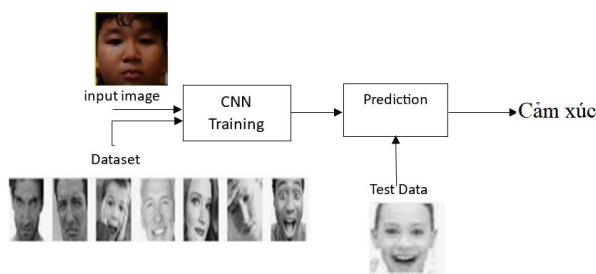
Mặc dù MTCNN và FaceNet trong quá trình thực nghiệm độ chính xác nằm trong khoảng 88%-97% [12], thấp hơn so với Open VINO trong quá trình thực nghiệm nhưng không đáng kể nhưng nghiên cứu lựa chọn MTCNN và FaceNet nhờ một số ưu điểm khi triển khai: MTCNN có độ chính xác cao trong việc phát hiện khuôn mặt, MTCNN với cấu trúc ba giai đoạn (P-Net, R-Net, O-Net) liên tiếp nhau, giúp lọc và tinh chỉnh các bounding box một cách hiệu quả; xác định chính xác các điểm đặc trưng trên khuôn mặt giúp cải thiện độ chính xác cho các bước xử lý tiếp theo; hiệu suất cao và thời gian thực phù hợp cho nghiên cứu; khả năng phát hiện khuôn mặt ở nhiều kích thước và tỷ lệ khác nhau, đảm bảo tính linh hoạt trong các tình huống thực tế. MTCNN hoạt động tốt trong các điều kiện ánh sáng và góc nhìn khác nhau, đảm bảo phát hiện khuôn mặt chính xác ngay cả trong môi trường không lý tưởng. Ngoài ra, FaceNet có khả năng mở rộng cao, dễ dàng tích hợp với các hệ thống và ứng dụng hiện có của nghiên cứu.

Phân loại đánh giá cảm xúc

Cùng mô hình mạng nơ ron tích chập, bộ dữ liệu FER2013 [5][6] và FER Plus [7] được nghiên cứu sử dụng để huấn luyện đánh giá các cảm xúc người học [5]. Bộ dữ liệu FER2013 và bộ dữ liệu FER+ là phần mở rộng của bộ dữ liệu FER gốc. Một số cảm xúc trong cả hai bộ dữ liệu: Angry, Disgust, Fear, Happy, Neutral, Sad, Surprise. Hai bộ dữ liệu có tổng số lượng ảnh hơn 35.887 ảnh. Khi nghiên cứu đánh giá cảm xúc người dùng, nhóm tác giả Brijesh Bakariya cùng đồng sự đã sử dụng mô hình CNN nhận diện các cảm xúc con người có kết hợp sử dụng âm nhạc [5] với độ chính xác là 73,02%. Rahmeh Abou Zafra cùng đồng tác giả cũng đánh giá kết quả trên tập dữ liệu FER2013, CK+ có độ chính xác lần lượt là khoảng từ 60% đến 85% [13].

Bài báo này sử dụng quy trình phân loại ảnh, gồm các bước như: Tập dữ liệu ảnh đầu vào, tiền xử lý, xử lý rút trích đặc trưng và phân loại.

Trạng thái cảm xúc của người học sẽ được đo lường từ dữ liệu hình ảnh thu thập được camera của thiết bị sử dụng trong xuyên suốt quá trình người học tham gia thi. Các trạng thái cảm xúc của người học được đánh giá bao gồm: Tức giận, sợ hãi, vui vẻ, trung lập, buồn, ngạc nhiên,...



Hình 8. Mô hình huấn luyện và phân loại cảm xúc

Để xây dựng ứng dụng web phát hiện cảm xúc khuôn mặt tích hợp trong hệ thống thông qua camera, nhóm nghiên cứu xây dựng hệ thống minh họa hỗ trợ cho việc xử lý dữ liệu đầu vào từ webcam. Quá trình xử lý qua các bước như sau :

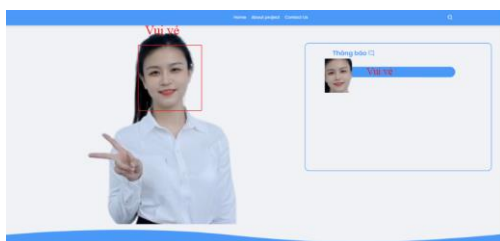
Bước 1: Ảnh đầu vào được chuyển thành đa cấp xám.

Bước 2: Dùng haar cascade để tìm kiếm vùng mặt người trên ảnh đầu vào, sau khi cắt được vùng khuôn mặt thì sẽ được chuyển đến bước 3.

Bước 3: Vùng ảnh mặt người được chuyển đổi về kích thước 48x48.

Bước 4: Ảnh sau khi đã được chuẩn hóa thành kích thước 48x48 đa cấp xám chuyển đổi về miền $[0, 1]$ sau đó đưa vào mô hình CNN đã được huấn luyện để dự đoán nhãn cảm xúc.

Bước 5: Đánh giá xác định cảm xúc người dùng



Hình 9. Ứng dụng phân loại đánh giá cảm xúc

Bài báo xây dựng giải pháp quản lý lớp học tích hợp thông minh có chức năng kết hợp giữa công nghệ nhận diện khuôn mặt và đánh giá cảm xúc người học. Tính mới của bài báo nằm ở việc tích hợp các kỹ thuật học sâu để không những thực hiện nhận diện khuôn mặt, xác thực khuôn mặt, mà còn phân tích được trạng thái cảm xúc của người học trong quá trình học và thi. Bài báo kết hợp việc sử dụng mô hình tiên tiến như MTCNN và FaceNet để xác thực danh tính cùng mô hình mạng nơ ron tích chập CNN để phát hiện, đánh giá cảm xúc người học mang tính chuyên biến mới, mạnh mẽ khác biệt so với các phương pháp truyền thống.

4. KẾT LUẬN

Nghiên cứu đổi mới lớp học tích hợp thông minh có thể xác thực người học và đánh giá cảm xúc của người học trong quá trình thi là tính mới thay thế cho việc ghi nhận thông tin thủ công như sổ đầu bài, hoặc điểm danh thủ công, QR... Việc đổi mới hệ thống quản lý lớp học tích

hợp thông minh dựa trên việc nhận dạng khuôn mặt sử dụng MTCNN và Face Net cùng với tính năng đánh giá cảm xúc người học trong trường đại học giúp công tác đánh giá có thêm số liệu để cải tiến công tác quản lý, từ đó giúp người học có thể học tập hiệu quả hơn, giúp việc đổi mới trong giáo dục thông minh - Smart Education ngày càng hoàn thiện hơn nữa.

5. TÀI LIỆU THAM KHẢO

- [1] Kaipeng zhang và cộng sự, Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks, **2016**.
- [2] XianBen Yang và Wei Zhang, Heterogeneous face detection based on multi-task cascaded convolutional neural network, IET Image Processing, **2021**.
- [3] Pilli Bhavya Sri cùng đồng tác giả, Harnessing Security Improvements Using FaceNet Approach for Face Recognition System, IEEE, **2024**.
- [4] Siyu Jia, Ying Tian, Face detection based on improved multi-task cascaded convolutional neural networks, IAENG international journal of computer science, **2024**.
- [5] Ian j. goodfellow và Đồng tác giả, "Challenges in Representation Learning: A report on three machine learning contests", **2013**.
- [6] Brijesh Bakariya và đồng sự, Facial emotion recognition and music recommendation system using CNN-based deep learning techniques, Evolving Systems 15, Volume 15, pages 641–658, **2024**.
- [7] Emad barsoum và đồng tác giả, "Training Deep Networks for Facial Expression Recognition with Crowd-Sourced Label Distribution", **2016**.
- [8] Abbas Khalifa Nawar, Hadi Raheem Ali, Mothefer Majeed Jahefer, Sabah Abdulazeez Jebur, "Automate facial paralysis detection using vgg architectures", ijciar.v7i1.158, Volume 7, Issue 1, **2024**.
- [9] Alexander demidovskij và đồng tác giả, "openvino Deep Learning Workbench: Towards Analytical Platform for Neural Networks Inference Optimization", Journal of Physics: Conference Series, **2021**.
- [10] Jiawei Huang và Ding Zhou, A scalable real-time computer vision system for student posture detection in smart classrooms, SpringerLink, Volume 29, pages 917–937, **2024**.
- [11] Siyao Qi và đồng tác giả, "Face Recognition Model Based On MTCNN And Facenet", IEEE, **2022**.
- [12] R E Saragih, Q H To, A Survey of Face Recognition Based on Convolutional Neural Network, ndonesian Journal of Information Systems (IJIS) Vol. 4, No. 2, February **2022**.
- [13] Rahmeh Abou Zafra, Lana Ahmad Abdullah, Rouaa Alaraj, Rasha Albezreh, Tarek Barhoum, Khlood Al Jallad, An experimental study in Real-time Facial Emotion Recognition on new 3RL dataset, Computer Vision and Pattern Recognition, **2023**.
- [14] Tony Gwyn, Kaushik Roy, Mustafa Atay, "Face Recognition Using Popular Deep Net Architectures: A Brief Comparative Study", **2021**.