

# GIẢI PHÁP ỨNG DỤNG KỸ THUẬT HỌC SÂU TRONG NHẬN DẠNG MẶT NGƯỜI

Hoàng Văn Dũng, Đặng Văn Đạt  
Trường Đại học Quảng Bình

**Tóm tắt.** Với sự phát triển nhanh chóng của trí tuệ nhân tạo trong những năm gần đây, việc ứng dụng trí tuệ nhân tạo vào các lĩnh vực an ninh, bảo mật, y học, giáo dục... ngày càng phổ biến. Một trong những hướng phát triển mạnh của trí tuệ nhân tạo hiện nay là kỹ thuật học sâu. Đây là một lĩnh vực chuyên sâu trong Machine learning, được ứng dụng để giải quyết các vấn đề thực tế bằng cách khai thác các mạng thần kinh nhân tạo và mô phỏng việc đưa ra các quyết định của con người. Trong nghiên cứu này, chúng tôi đề xuất giải pháp sử dụng mạng học sâu để trích xuất đặc trưng khuôn mặt nhằm phục vụ cho việc đánh giá mức độ tương đồng nhằm nhận dạng ra định danh người tương ứng. Giải pháp sử dụng mạng neural sâu trong quá trình nhận dạng khuôn mặt mang lại độ chính xác đạt 95,3%, độ chính xác tăng từ 8% đến 10% so với các phương pháp truyền thống.

**Từ khóa:** Kỹ thuật học sâu, nhận dạng mặt người, mạng neural nhân tạo, mạng neuraltích chập

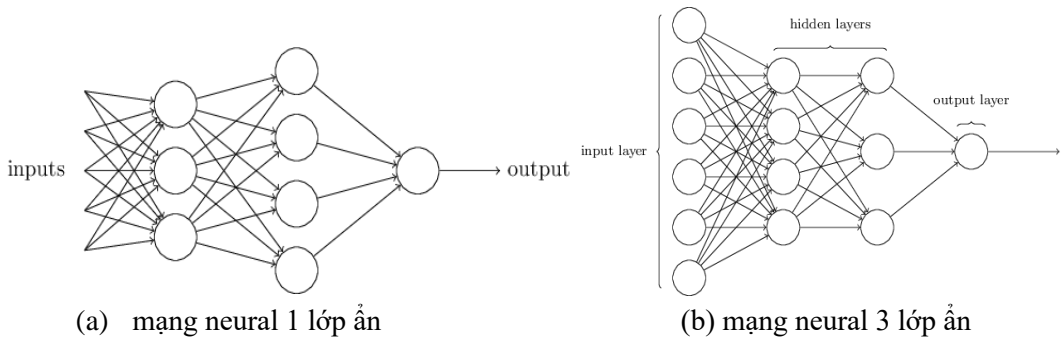
## 1. GIỚI THIỆU

Trong những năm gần đây, kỹ thuật học sâu (Deep learning) được nhiều nhà khoa học nghiên cứu phát triển và ứng dụng sâu rộng trong nhiều lĩnh vực như tìm kiếm sự sai khác giữa các văn bản, phát hiện gian lận, phát hiện spam, nhận dạng chữ viết tay, nhận dạng hình ảnh, giọng nói... góp phần quan trọng trong việc hỗ trợ con người trong nhiều lĩnh vực đời sống. Tại một số quốc gia phát triển, áp dụng giải pháp kỹ thuật học sâu cho hệ thống nhận dạng khuôn mặt để quản lý các hoạt động thanh toán điện tử, giám sát an ninh mang lại hiệu quả và độ chính xác cao. Hiện nay, lĩnh vực nghiên cứu học sâu được nhiều nhà khoa học quan tâm nghiên cứu và công bố các bài báo liên quan, như trong [1-4]. Ở Việt Nam, giải pháp ứng dụng kỹ thuật học sâu trong nhận dạng mặt

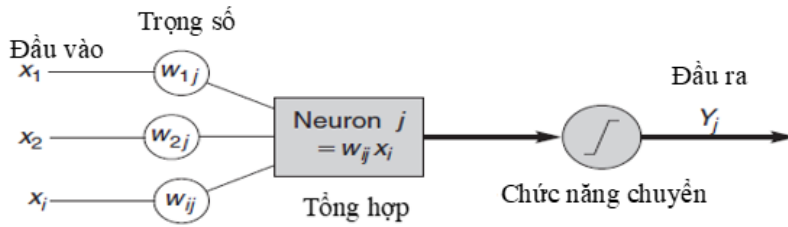
người chưa được áp dụng vào thực tiễn, các tài liệu nghiên cứu của các nhà khoa học trong nước về giải pháp này còn hạn chế. Trong bài báo này, chúng tôi đưa ra một giải pháp ứng dụng kỹ thuật học sâu trong nhận dạng mặt người nhằm nghiên cứu và ứng dụng kỹ thuật học sâu để xây dựng hệ thống nhận dạng mặt người, qua đó báo cáo kết quả thực nghiệm và đánh giá độ chính xác của giải pháp.

## 2. MẠNG NEURAL TÍCH CHẬP HỌC SÂU

Trong cách tiếp cận truyền thống, mạng neural nhân tạo (còn gọi là mạng neural) là một mô hình toán học hay mô hình tính toán được xây dựng dựa trên các mạng neural sinh học. Mạng neural nhân tạo truyền thống có thể được thể hiện như hình sau:



**Hình 1.** Một số kiểu mạng neural [1]

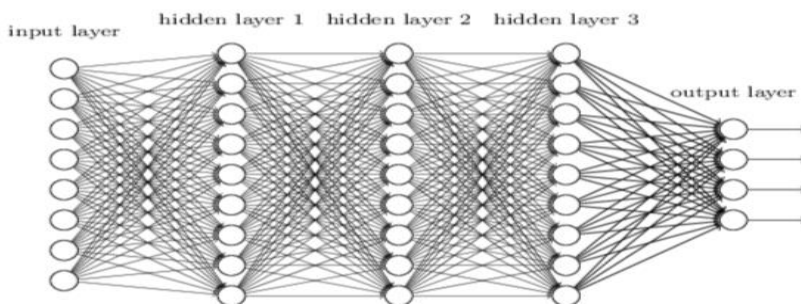


**Hình 2.** Quá trình xử lý thông tin của một ANN [3]

Quá trình xử lý thông tin của một mạng neural nhân tạo được khái quát như sau:

Mạng neural nhân tạo được huấn luyện (Training) hay được học (Learning) theo 2 kỹ thuật cơ bản đó là học có giám sát (Supervised learning) và học không giám sát (Unsupervised learning).

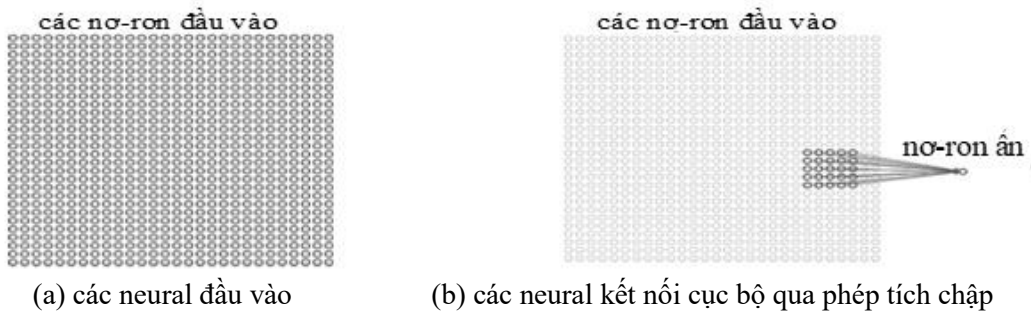
Với cách tiếp cận mới, người ra xây dựng mạng với rất nhiều lớp ẩn được gọi là mạng neural sâu. Mạng neural có các lớp mạng liền kề được kết nối đầy đủ với nhau, cụ thể mỗi neural trong mạng được kết nối với tất cả các neural trong lớp liền kề [1].



**Hình 3.** Mạng kết nối đầy đủ với ba lớp ẩn

Mạng neural tích chập sử dụng ba ý tưởng cơ bản: trường tiếp nhận cục bộ, trọng số chia sẻ và tổng hợp với phạm vi

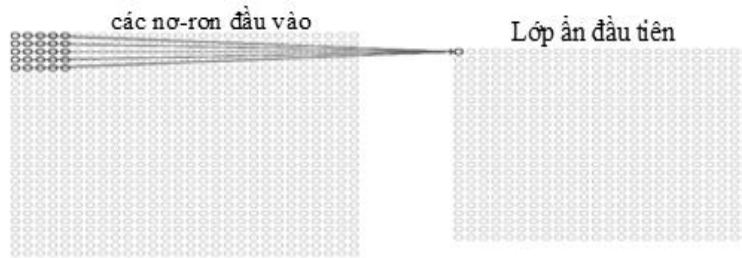
cục bộ nhằm giảm chi phí tính toán và không gian tham số của mạng.



**Hình 4.** Mạng neural tích chập [1]

Quá trình tích chập được thực hiện như sau: Mặt nạ tích chập được “trượt” trên toàn bộ hình ảnh đầu vào để tính giá

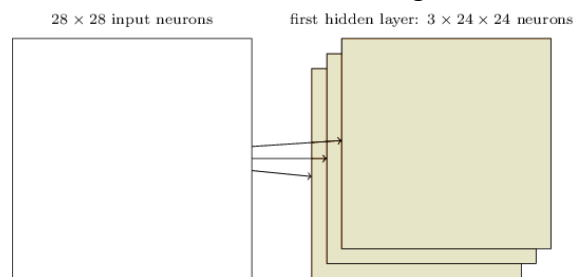
trị tích chập. Bắt đầu với một trường tiếp nhận cục bộ ở góc trên bên trái:



**Hình 5.** Các neural đầu vào kết nối cục bộ với các neural lớp ẩn đầu tiên [1].

Sau đó, “trượt” mặt nạ tích chập sang phải một neural để kết nối với một neural ẩn thứ hai.

Để nhận dạng hình ảnh, cần nhiều hơn một ánh xạ đặc trưng. Do đó, một lớp tích chập hoàn chỉnh bao gồm các ánh xạ đặc trưng khác nhau:

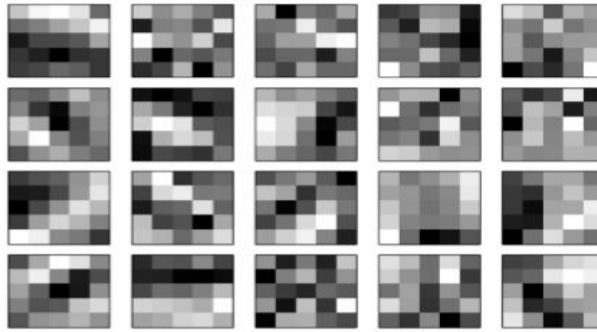


**Hình 6.** Ánh xạ đặc trưng vào lớp ẩn [1]

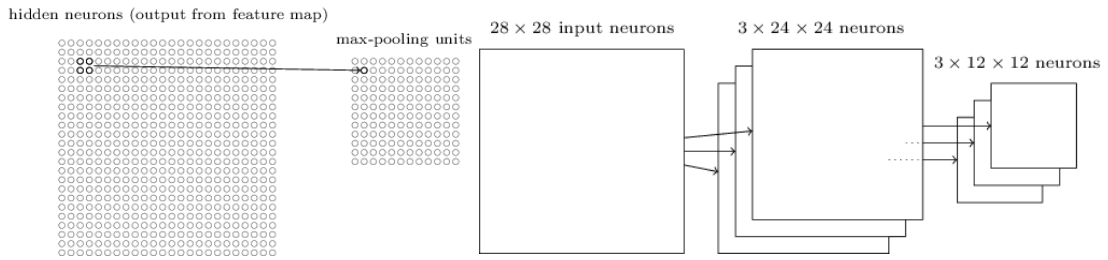
Dưới đây là một số các đặc trưng thu được sau khi huấn luyện:

Lớp làm giảm kích thước: Trong max-pooling, một đơn vị tổng hợp là kết quả đầu ra kích hoạt tối đa trong 2x2 khu

vực đầu vào (Hình 8a). Áp dụng max-pooling cho mỗi ánh xạ đặc trưng riêng biệt. Nếu có ba ánh xạ đặc trưng, các lớp tích chập và max-pooling kết hợp (Hình 8b).



**Hình 7.** Ảnh xạ đặc trưng trong các hình ảnh khác nhau [1]



(a) Kiểu lớp Max-pooling

(b) Thực hiện tích chập sau đó sử dụng pooling nhằm giảm kích thước mẫu

**Hình 8.** Một số kiểu lớp trong mạng [1]

### 3. GIẢI PHÁP NHẬN DẠNG MẶT NGƯỜI

Giải pháp nhận dạng khuôn mặt được thực thi qua 2 giai đoạn:

#### Giai đoạn 1: Giai đoạn thu thập cơ sở dữ liệu và huấn luyện

Bước 1: Phát hiện khuôn mặt. Các khuôn mặt được đánh dấu từ các hình ảnh video qua các thuật toán phát hiện khuôn mặt. Ở đây chúng tôi sử dụng thuật toán phát hiện nhanh khuôn mặt trong bài báo [7], giải pháp này đã được cài đặt sẵn trong các thư viện xử lý ảnh OpenCV, Matlab.

Bước 2: Trích xuất đặc trưng. Ở đây học sâu sẽ tiến hành xác định phần nào là quan trọng nhất của khuôn mặt để tiến hành tính toán đo lường. Bằng việc sử dụng mạng neural sâu (DCNN) được huấn luyện để học các đặc trưng riêng biệt bằng

cách sử dụng 128 phần tử cho mỗi khuôn mặt. Tức là mỗi khuôn mặt sẽ được biểu diễn bằng một vector đặc trưng với 128 giá trị số. Việc này được thực hiện thông qua giá trị đầu ra của lớp kết nối đầy đủ cuối của mạng học sâu thông qua hàm activate. Nghĩa là trong trường hợp này chúng tôi không sử dụng DCNN để phân lớp dữ liệu mà chỉ dùng để trích xuất đặc trưng phục vụ cho quá trình nhận dạng so khớp ở giai đoạn 2.

Bước 3: Lưu trữ mô hình và đặc trưng vào cơ sở dữ liệu. Bước này thực hiện lưu trữ các tham số trong kiến trúc mạng sau khi huấn luyện (mô hình) và bộ vector đặc trưng của các khuôn mặt của mỗi đối tượng.

#### Giai đoạn 2: Nhận dạng

Bước 1: Phát hiện khuôn mặt: Tìm kiếm các khuôn mặt có ảnh cần truy xuất.

Đánh dấu vị trí của khuôn mặt trong hình tương tự bước 1 ở giai đoạn 1.

Bước 2: Tiền xử lý: Thực hiện các quá trình lọc nhiễu, chuẩn hóa hình ảnh.

Bước 3: Trích xuất đặc trưng: Sử dụng mô hình mạng neural sâu để trích xuất vector đặc trưng với 128 giá trị đặc trưng cho từng khuôn mặt.

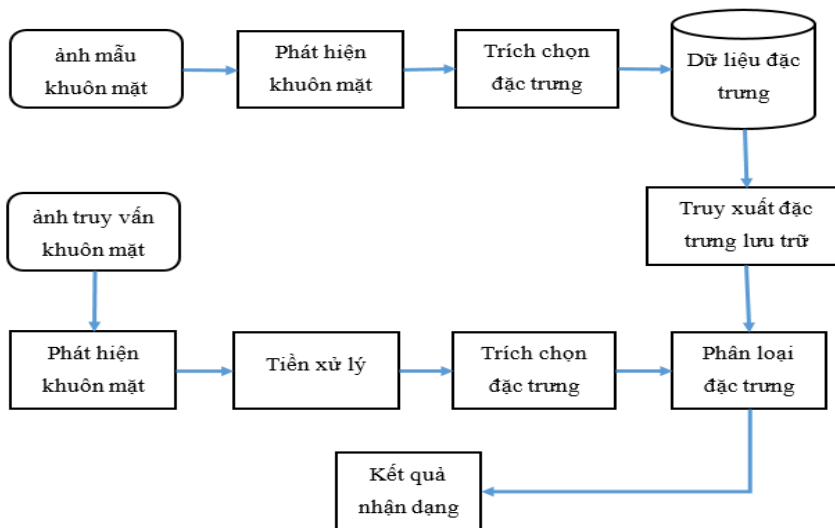
Bước 4: Nhận dạng: Vector đặc trưng khuôn mặt được so sánh với các vector tính năng được lưu trữ trong cơ sở dữ liệu bằng phương pháp Facenet [8] với độ đo tương tự bằng hàm khoảng cách Euclide. Kết quả trả về là tên của người cần truy vấn bằng nhận dạng khuôn mặt.

### Quá trình huấn luyện mạng neural tích chập sâu

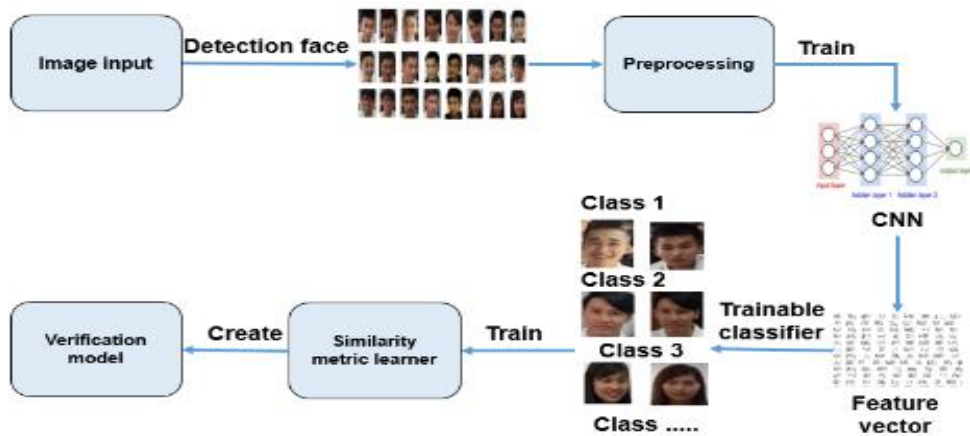
Quá trình huấn luyện mạng neural tích chập sâu được thực hiện qua các tiến trình sau: Đầu tiên cung cấp các hình ảnh đầu vào với hình ảnh đáp ứng có chứa đối tượng khuôn mặt. Sau khi thu thập các tập dữ liệu đầu vào, quá trình phát hiện khuôn mặt trong ảnh được thực thi. Kết thúc quá trình phát hiện khuôn mặt ta sẽ thu được

các hình ảnh được cắt từ tập dữ liệu ảnh đầu vào. Quá trình phát hiện được diễn ra liên tục đảm bảo tìm kiếm toàn bộ các khuôn mặt có trong ảnh.

Sau khi thu được các hình ảnh khuôn mặt được cắt từ dữ liệu ảnh, ảnh sẽ chuyển đến quá trình tiền xử lý nhằm thực hiện các quá trình lọc nhiễu, chuẩn hóa hình ảnh. Kết thúc quá trình tiền xử lý. Các hình ảnh khuôn mặt được đưa vào đào tạo mô hình mạng neural tích chập (CNN). Tại đây khuôn mặt được đào tạo qua các lớp trong mô hình CNN. Các khuôn mặt sau khi được đào tạo qua các lớp sẽ thu được các vec tơ đặc trưng, trên cơ sở các vec tơ đặc trưng, chuyển đến quá trình đào tạo phân lớp. Quá trình phân lớp được thực thi, ứng với mỗi lớp sẽ là tập hợp đặc trưng khuôn mặt cho chính đối tượng trong lớp đó. Kết thúc quá trình phân lớp, các chỉ số và giá trị thu được qua quá trình đào tạo được lưu lại trong file mô hình đào tạo làm cơ sở cho các quá trình nhận dạng và trích xuất các nhãn.



**Hình 9.** Giải pháp nhận dạng khuôn mặt



Hình 10. Quá trình huấn luyện mạng neural tích chập sâu

#### 4. THỰC NGHIỆM VÀ ĐÁNH GIÁ

**Ngôn ngữ và công cụ lập trình:** trong quá trình thực hiện thực nghiệm chúng tôi đã sử dụng ngôn ngữ lập trình python phiên bản python 3.6 để viết mã và chạy chương trình. Công cụ hỗ trợ viết mã python là IDE Pycharm kết hợp với thư viện mã nguồn mở Tensorflow[5, 9] và OpenCV [6, 10].

##### Quá trình thực nghiệm:



Hình11. Hình ảnh ví dụ từ tập dữ liệu ảnh khuôn mặt cho 8 danh tính

Dữ liệu ảnh sau khi thu thập được đưa vào quá trình phát hiện khuôn mặt để trích xuất ra vùng mặt nhằm tạo mẫu. Quá trình này được thực thi bằng cách chạy file “detector.py” của thư viện OpenCV. Kết thúc quá trình chạy file thực thi thu được các hình ảnh khuôn mặt được cắt trong tập

**Thu thập và xử lý dữ liệu:** Trong phần này có thể tiến hành thu thập hình ảnh khuôn mặt bằng cách tải xuống các file hình ảnh từ mạng xã hội facebook, instagram,.. hoặc thu thập bằng cách chụp ảnh trực tiếp. Trong nghiên cứu này, chúng tôi tiến hành thu thập ảnh bằng cách chụp ảnh các thành viên trong lớp học phục vụ tạo mẫu dữ liệu phục vụ huấn luyện và đánh giá.

dữ liệu ảnh ban đầu. Trong thực nghiệm này, các mẫu ảnh khuôn mặt được trích ra và chuẩn hóa về kích thước size (182x182) và được lưu trữ thành các file ảnh khuôn mặt tạo thành bộ cơ sở dữ liệu huấn luyện và đánh giá. Một số ví dụ minh họa trong Hình 11.

**Huấn luyện mô hình:** Sau khi hoàn tất các công đoạn trong xử lý trước dữ liệu, tiến hành huấn luyện mô hình bằng cách sử dụng hàm huấn luyện mạng học sâu “training.py” trong môi trường Tensorflow. Hoàn tất quá trình huấn luyện, mô hình được lưu trữ lại và phục vụ cho ứng dụng sau này.

**Nhận dạng:** Quá trình nhận dạng theo cách tiếp cận trước đây bằng cách sử dụng hàm có sẵn “recognition.py” trong môi trường Tensorflow. Với phương pháp

đề xuất chúng tôi trích xuất tại lớp kết nối đầy đủ để thu được vector 128 đặc trưng và đưa vào bộ so sánh Facenet để nhận dạng ra danh tính người. Kết quả nhận dạng khuôn mặt ra người tương ứng được khoanh vùng và hiển thị tên phía dưới vùng đánh dấu khuôn mặt. Quá trình nhận dạng được thực hiện trên máy tính cấu hình thấp không có hỗ trợ GPU nhằm phù hợp với ứng dụng thực tế trong trường hợp không có hệ thống máy tính cấu hình cao.

### Đánh giá kết quả

**Bảng 1.** Thống kê kết quả quá trình thử nghiệm hệ thống nhận dạng khuôn mặt sử dụng kỹ thuật Deeplearning

Lần thử	Phương pháp	Số mẫu	Số nhận dạng đúng	Số nhận dạng sai	Tỷ lệ chính xác
1	Phương pháp PCA	600	494	106	82,33%
	Phương pháp LBP	600	491	109	81,83%
	Phương pháp đề xuất	600	560	40	93,33%
2	Phương pháp PCA	1000	841	159	84,1%
	Phương pháp LBP	1000	858	142	85,8%
	Phương pháp đề xuất	1000	942	58	94,2%
3	Phương pháp PCA	1200	1021	179	85,03%
	Phương pháp LBP	1200	1024	176	85,33%
	Phương pháp đề xuất	1200	1144	56	95,3%

#### Đánh giá:

- Tỷ lệ nhận dạng chính xác ở mức tốt trong các môi trường và ánh sáng tốt.

- Tốc độ xử lý: tốc độ xử lý nhận dạng 5,2 hình/giây với cấu hình máy tính Core I5, 4GB RAM.

### 5. KẾT LUẬN

Trong bài viết này, chúng tôi đã trình bày giải pháp ứng dụng kỹ thuật học sâu trong nhận dạng mặt người. Khác với cách tiếp cận trước, trong nghiên cứu này chúng tôi sử dụng kỹ thuật CNN để trích xuất đặc trưng phục vụ bộ so sánh đánh giá độ tương tự để xác định danh người qua khuôn mặt. Dữ liệu thực nghiệm được xây dựng trên

ảnh thật của một nhóm người trong lớp. Kết quả thực nghiệm đánh giá chúng tôi đưa ra một số kết luận như sau: về tốc độ xử lý trong quá trình nhận dạng đạt mức độ tiệm cận chấp nhận được với ứng dụng thực, độ chính xác khi áp dụng giải pháp đề xuất có độ chính xác tăng từ 8% đến 10% so với các kỹ thuật nhận dạng thông thường được cài đặt mặc định trong thư viện xử lý ảnh. Từ thực nghiệm và đánh giá chúng tôi nhận thấy giải pháp ứng dụng kỹ thuật học sâu trong trích xuất đặc trưng và so sánh để nhận dạng khuôn mặt mang lại độ tin cậy và chính xác cao, có thể phát triển hoàn chỉnh để áp dụng vào các hệ thống giám sát

an ninh, bảo mật của các cơ quan, tổ chức, cá nhân có nhu cầu sử dụng.

## **TÀI LIỆU THAM KHẢO**

- [1] Michael A. Nielsen (2013), “Neural Networks And Deep Learning”, Determination Press.
- [2] A. Krizhevsky, I. Sutskever, and G. E. Hinton “ImageNet classification with deep convolutional neural networks”, In NIPS, pp. 1106–1114, 2012.
- [3] Szegedy, Christian, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. “Going deeper with convolutions.” Cvpr, 2015.
- [4] Simonyan, Karen, and Andrew Zisserman, “Very deep convolutional networks for large-scale image recognition”, preprint arXiv:1409.1556, 2014.
- [5] Abadi, Martín, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, “TensorFlow: A System for Large-Scale Machine Learning”, InOSDI, vol. 16, pp. 265-283, 2016.
- [6] Robert Laganière, “OpenCV Computer Vision Application Programming” Cookbook Second Edition, Packt Publishing Ltd, 2014.
- [7] Viola, Paul, and Michael J. Jones, “Robust real-time face detection.” International journal of computer vision, vol.57(2), pp.137-154 (2004).
- [8] Schroff, Florian, Dmitry Kalenichenko, and James Philbin. “Facenet: A unified embedding for face recognition and clustering.” InProceedings of the IEEE conference on computer vision and pattern recognition, pp. 815-823. 2015.
- [9] [https://www.tensorflow.org/api\\_guides/python](https://www.tensorflow.org/api_guides/python)
- [10] [http://opencv-python-tutroals.readthedocs.io/en/latest/py\\_tutorials](http://opencv-python-tutroals.readthedocs.io/en/latest/py_tutorials)



## **A SOLUTION FOR FACIAL RECOGNITION BASED ON DEEP LEARNING**

**Abstract.** *Recent years, artificial intelligence has been growing rapidly. Artificial intelligence based smart applications have been increasing popular such as monitoring security, medical image processing, e-education, etc. One of the most important trend of artificial intelligence is deep learning (DL) technique. It is a particularly important field in machine learning. DL has been successfully applied to solve the practical applications. In this paper, a solution for facial recognition based on deep learning technique is investigated and presented for improving accuracy and efficiency of the recognition systems. The proposed approach based on the deep convolutional neural network achieved 95.3% accuracy, which is higher 8% to 10% precision rate than state of the arts.*

**Keywords:** *Deep learning, facial recognition, artificial neural network, convolutional neural network.*

### **Liên hệ:**

TS. Hoàng Văn Dũng

*Phòng Đào tạo, Trường Đại học Quảng Bình*

*312 Lý Thường Kiệt, Đồng Hới, Quảng Bình*

*Email: zunghv@gmail.com*