

MẠNG NƠ-RON TÍCH CHẬP (CNN) VÀ MẠNG VGG16 TRÊN HỆ THỐNG NHẬN DẠNG KHUÔN MẶT THỜI GIAN THỰC

CONVOLUTION NEURON NETWORK (CNN) AND VGG16 NETWORK ON REAL-TIME FACE RECOGNITION SYSTEM

Viên Thanh Nhã^{1,*}, Tiếp Sỹ Minh Phụng¹,
Phạm Thành Công², Nguyễn Thanh Tùng², Lê Đình Phú Cường³

TÓM TẮT

Nhận dạng khuôn mặt người được coi là một chủ đề được nghiên cứu nhiều trong lĩnh vực thị giác máy tính. Khả năng tự động nhận dạng và xác thực khuôn mặt người bằng hình ảnh thời gian thực là một khía cạnh quan trọng trong giám sát, bảo mật và các lĩnh vực liên quan khác. Có các ứng dụng riêng biệt giúp xác định các cá nhân tại các địa điểm cụ thể giúp phát hiện những kẻ xâm nhập. Nhận dạng thời gian thực là bắt buộc cho các mục đích giám sát. Một số phương pháp học máy cùng với bộ phân loại được sử dụng để nhận dạng khuôn mặt. Công việc này giới thiệu một hệ thống Nhận dạng khuôn mặt thời gian thực mới. Quá trình này được chia thành bốn bước chính: (1) thu thập cơ sở dữ liệu, (2) nhận dạng khuôn mặt để xác định những người cụ thể và (3) đánh giá hiệu suất. Tại bước đầu tiên, hệ thống thu thập 7890 khuôn mặt trong thời gian thực của 14 khuôn mặt người bằng camera có độ phân giải 1920*1080. Bước thứ hai, thuật toán nhận dạng khuôn mặt người tại thời gian thực được sử dụng để nhận dạng khuôn mặt với một cơ sở dữ liệu đã biết. Đối với Nhận dạng khuôn mặt trong thời gian thực, VGG-16 với học chuyển giao và mạng nơ-ron tích chập (CNN) được sử dụng. Hệ thống đề xuất này được thực hiện bằng thư viện Keras. Cuối cùng, hiệu suất của hai bộ phân loại này được đánh giá qua các thông số như: độ lặp lại, độ thu hồi, điểm F1 và độ chính xác.

Từ khóa: Học sâu; mạng nơ-ron tích chập (CNN); VGG16; xác thực khuôn mặt; hình ảnh khuôn mặt thời gian thực.

ABSTRACT

Face Recognition is considered as a heavily studied topic in computer vision field. The capability to automatically identify and authenticate human's faces using real-time images is an important aspect in surveillance, security, and other related domains. There are separate applications that help in identifying individuals at specific locations which help in detecting intruders. The real-time recognition is mandatory for surveillance purposes. A number of machine learning methods along with classifiers are used for the recognition of faces. This work introduces a new real time face Recognition system. The process is broken into four major steps: (1) database collection, (2) face recognition to identify particular persons and (3) Performance evaluation. For the first step, the system collects 7890 faces in real time for 14 persons using a camera with resolution of 1920*1080. Second step, efficient real time face recognition algorithm is then used to recognize faces with a known database. For real time face Recognition, VGG-16 with Transfer Learning and Convolutional Neural Network (CNN) are used. This proposed system is implemented using Keras framework. Lastly the performance of these two classifiers is measured using of precision, recall, F1- score and accuracy.

Keywords: Deep learning; Convolutional Neural Network (CNNs); VGG16; face authentication; real time face images.

¹Phân hiệu Miền Nam, Trường Đại học Thủy Lợi

²Trung tâm Công nghệ thông tin, Trường Đại học Công nghiệp Hà Nội

³Trường Đại học Yersin Đà Lạt

*Email: vienthanhnha@tlu.edu.vn

Ngày nhận bài: 20/02/2022

Ngày nhận bài sửa sau phản biện: 20/4/2022

Ngày chấp nhận đăng: 27/6/2022

1. GIỚI THIỆU

Nhận dạng khuôn mặt là một chủ đề có liên quan trong lĩnh vực sinh trắc học. Nó đang được sử dụng trong một số ứng dụng ngày nay do sự xuất hiện của nó trong những năm gần đây. Các vấn đề và sự phát triển trong nhận dạng khuôn mặt đã và đang thu hút rất nhiều nhà khoa học làm việc trong lĩnh vực thị giác máy tính, nhận dạng mẫu và sinh trắc học. Các thuật toán nhận dạng khuôn mặt khác nhau đã được sử dụng trong các ứng dụng đa dạng như lập chỉ mục và nén video thuộc lĩnh vực sinh trắc học. Các khái niệm nhận dạng khuôn mặt có thể được sử dụng để phân loại nội dung đa phương tiện và giúp tìm kiếm nhanh chóng các tài liệu mà người dùng cuối quan tâm. Một cơ chế nhận dạng khuôn mặt toàn diện có thể được hỗ trợ trong các lĩnh vực như giám sát và khoa học pháp y. Nó cũng có thể được sử dụng trong các lĩnh vực thực thi pháp luật và xác thực hệ thống bảo mật và ngân hàng. Ngoài ra, nó cũng cung cấp quyền kiểm soát và quyền truy cập ưu tiên vào các khu vực được bảo mật và người dùng được ủy quyền. Các vấn đề về nhận dạng khuôn mặt thậm chí còn có ý nghĩa hơn sau khi khủng bố gia tăng đột biến trong những năm gần đây. Nó làm giảm phần lớn nhu cầu về mặt khẩu và có thể cung cấp bảo mật nâng cao. Đối

với điều này, nhận dạng khuôn mặt nên được sử dụng với các cơ chế bảo mật bổ sung.

Mặc dù nhận dạng khuôn mặt phát triển nhanh chóng và được đánh giá là một cơ chế xác thực quan trọng, các thuật toán được sử dụng để nhận dạng khuôn mặt vẫn chưa được phát triển đáng kể. Đã gần hai thập kỷ kể từ khi nhận dạng khuôn mặt xuất hiện nhưng một hệ thống toàn diện có khả năng tạo ra kết quả tích cực trong điều kiện thời gian thực vẫn chưa được phát triển.

Các hệ thống nhận dạng khuôn mặt hiện đại dành cho các môi trường phức tạp đã thu hút sự chú ý rất lớn trong những thập kỷ gần đây. Hệ thống nhận dạng khuôn mặt được tự động hóa là một công nghệ đang phát triển thu hút được nhiều sự quan tâm. Tồn tại các thuật toán thông thường khác nhau, được sử dụng để phát triển ảnh màu và ảnh tính. Độ phức tạp của dữ liệu được tăng lên trong hình ảnh màu khi các pixel được ánh xạ vào không gian chiều cao. Điều này làm giảm đáng kể độ chính xác và hiệu quả xử lý của nhận dạng khuôn mặt [1]. Trong những năm gần đây, người ta suy ra rằng học sâu hoạt động tốt hơn rất nhiều đối với các mẫu lớn. Ngược lại, người ta cũng quan sát thấy rằng các cơ chế học máy thông thường có thể hoạt động tối ưu trong các bộ dữ liệu tương đối nhỏ hơn.

Nghiên cứu này đề xuất hướng xử lý hình ảnh màu để nhận dạng và phát hiện khuôn mặt với độ chính xác cao trong thời gian thực. CNN (Mạng nơ-ron tích chập) cùng với VGG-16 đã được chúng tôi sử dụng để nâng cao độ chính xác của nhận dạng. Hệ thống sử dụng học sâu để nhận dạng khuôn mặt một cách chính xác. Hệ thống này sẽ có khả năng nhận dạng nhiều khuôn mặt hơn có thể được sử dụng để tìm kiếm nghi phạm vì các sai sót được giảm đáng kể.

2. NHỮNG NGHIÊN CỨU LIÊN QUAN

Các phương pháp nhận dạng khuôn mặt thông thường phụ thuộc rất nhiều vào các tính năng như bộ mô tả kết cấu và các cạnh. Những tính năng này được kết hợp với một số phương pháp học máy nhất định như PCA - Phân tích thành phần chính, SVM - Máy vectơ hỗ trợ và LDA - Phân tích phân biệt tuyến tính. Sự phức tạp liên quan đến các phương pháp này khiến các nhà khoa học tập trung vào các phương pháp khác như phương pháp bất biến chiếu sáng [5, 6], phương pháp bất biến đặt ra [4] và phương pháp bất biến tuổi [2, 3].

Apoorva và cộng sự [7] đã triển khai bộ phân loại Haar sử dụng camera giám sát để nhận dạng khuôn mặt. Hệ thống có bốn bước tuần tự bao gồm: (1) đào tạo hình ảnh thời gian thực (2) nhận dạng khuôn mặt với sự trợ giúp của bộ phân loại Haar (3) so sánh hình ảnh thời gian thực với hình ảnh được chụp từ máy ảnh (4) thể hệ kết quả theo sự so sánh. Haar được sử dụng để phát hiện khuôn mặt một cách mạnh mẽ trong các tình huống thời gian thực. Nhận diện khuôn mặt sử dụng một thuật toán được gọi là xếp tầng Haar. Các khuôn mặt có thể được theo dõi và nhận dạng trong nền tảng Thư viện CV Mở. Trong trường hợp này, mức độ chính xác là khá cao. Một hệ thống được gọi là Aadhar đã được Ấn Độ áp dụng để công nhận các công

dân. Nếu cơ sở dữ liệu này được sử dụng làm cơ sở dữ liệu cho người dân, thì người dân địa phương và người nước ngoài có thể dễ dàng nhận ra. Thông tin này cuối cùng có thể được sử dụng để xác định người đó có phải là tội phạm hay không.

Raj [8] đã đề xuất một PCA để nhận dạng khuôn mặt trong thời gian thực. Hệ thống được đề xuất hiệu quả vì nó sử dụng hệ thống nhận dạng khuôn mặt dựa trên C++ và OpenCV. Hệ thống sử dụng các bộ phân loại khoảng cách khác nhau để trích xuất đối tượng địa lý. Các bộ phân loại khoảng cách chính được áp dụng ở đây là khoảng cách Mahalanobis, Khoảng cách Manhattan và khoảng cách Euclidian. Tỷ lệ nhận dạng lý tưởng đạt được khi sử dụng khoảng cách Mahalanobis được ước tính là khoảng 92,3%. Điều này tốt hơn khi so sánh với 73,1% thu được khi sử dụng PCA bình thường. Hiệu suất cao hơn hệ thống của chúng phản ứng với các truy vấn nhận dạng khuôn mặt trong khoảng thời gian ước tính là 0,2 giây.

Bah và Ming [9] đề xuất một LBP (Mẫu nhị phân cục bộ) để nhận dạng khuôn mặt. Điều này được sử dụng cùng với các phương pháp xử lý hình ảnh khác như Cân bằng biểu đồ, Bộ lọc song phương, Trộn hình ảnh và Điều chỉnh độ tương phản để cải thiện độ chính xác tổng thể của hệ thống. Các mã LBP được cải thiện ở đây do đó hiệu suất của hệ thống được nâng cao. Theo kết quả thử nghiệm, phương pháp này đáng tin cậy, mạnh mẽ và chính xác. Nó có thể được sử dụng trong môi trường thực tế như một hệ thống quản lý chấm công.

Kumar và cộng sự [10] đã đề xuất hệ thống thác AdaBoost với Haar để nhận dạng khuôn mặt trong thời gian thực. Hệ thống này thu gọn phần lớn phương sai. Để xác định khuôn mặt người, thác Haar được sử dụng cùng với AdaBoost. LDA và PCA nhanh giúp nhận dạng khuôn mặt dễ dàng. Các khuôn mặt phù hợp được sử dụng để đánh dấu sự tham dự trong phòng thí nghiệm. Là một hệ thống sinh trắc học, đây đang ngày càng được chấp nhận như một hệ thống chấm công thời gian thực sử dụng các thuật toán nhanh chóng và hiệu quả. Tỷ lệ chính xác của phương pháp này cũng tốt đáng kể.

Shieh và cộng sự [11] đã đề xuất một PCA với Tối ưu hóa đám đông hạt SVM để tạo ra các hệ thống thời gian thực nhằm nhận dạng khuôn mặt. Đa số các ứng dụng tương tác giữa người và robot sử dụng phương pháp dựa trên PCA do khả năng giảm kích thước của nó. Ở đây, PSO thực hiện lựa chọn tính năng trong khi SVM hoạt động như một chức năng phù hợp của PSO đối với các vấn đề phân loại. Kết quả chỉ ra rằng hệ thống được đề xuất đơn giản hóa các tính năng một cách toàn diện trong khi đạt được độ chính xác phân loại cao.

Shubha và Meenakshi [12] đề xuất một LBP nhận dạng khuôn mặt thời gian thực. Hình ảnh khuôn mặt được thể hiện bằng cách sử dụng thông tin về kết cấu và hình dạng. Để đại diện cho khuôn mặt một cách toàn diện, vùng mặt được chia thành các phần khác nhau. Biểu đồ LBP sau đó được tách ra và được kết hợp thành một biểu đồ duy nhất.

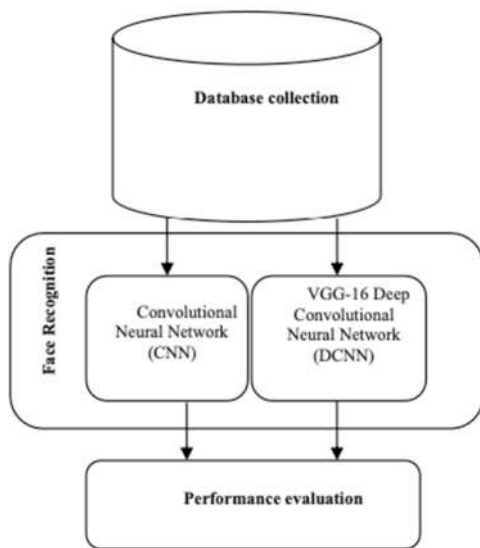
Sau đó, nhận dạng khuôn mặt được sử dụng bộ phân loại Hàng xóm Gần nhất. Việc xác thực thuật toán được thực hiện bằng cách tạo ra một mô hình nguyên mẫu sử dụng máy tính bảng đơn Raspberry Pi và MATLAB. Kết quả chỉ ra rằng tỷ lệ nhận dạng khuôn mặt của thuật toán LBP tương đối cao hơn khi so sánh với các phương pháp tiếp cận khác.

Zhang và cộng sự [13] đã đưa ra một thuật toán mạnh mẽ và hiệu quả có thể thực hiện nhận dạng khuôn mặt ở những nền phức tạp. Nó được thực hiện với sự trợ giúp của một chuỗi các chế độ xử lý tín hiệu bao gồm PCA, Haar like feature, LBP, cascade classifier và Ada Boost. Thuật toán này sử dụng bộ phân loại tăng để đào tạo các bộ dò tìm mắt và khuôn mặt có độ chính xác. Các đặc điểm trên khuôn mặt được trích xuất bằng cách sử dụng bộ mô tả LBP có thể phát hiện khuôn mặt một cách nhanh chóng. Phát hiện mắt được thực hiện bằng cách sử dụng thuật toán cũng giúp giảm tỷ lệ nhận diện khuôn mặt giả. Thuật toán PCA được sử dụng để nhận dạng khuôn mặt một cách chính xác. Thuật toán nhận dạng khuôn mặt được đào tạo bằng cách sử dụng cơ sở dữ liệu lớn có hình ảnh của khuôn mặt và không phải khuôn mặt. Tỷ lệ chính xác của các thuật toán đối với nhận dạng khuôn mặt cao.

3. PHƯƠNG PHÁP ĐỀ XUẤT

Hệ thống này được chia thành ba bước chính như sau:

1. Thu thập cơ sở dữ liệu
2. Nhận dạng khuôn mặt để xác định những người cụ thể
3. Đánh giá hiệu suất.



Hình 1. Hệ thống nhận diện khuôn mặt

Đối với bước đầu tiên, hệ thống thu thập các khuôn mặt trong thời gian thực. Cơ sở dữ liệu bao gồm 14 người khác nhau ở dạng 1056 hình ảnh có độ phân giải 112*92. Trong bước thứ hai, Mạng nơ-ron hội tụ (CNN) và Mạng nơ-ron hội tụ sâu VGG-16 (DCNN) được giới thiệu để nâng cao độ chính xác của nhận dạng. Cuối cùng, đánh giá kết quả của hai bộ phân loại này được thực hiện bằng cách sử dụng độ

chính xác, thu hồi, điểm F1, độ chính xác. Trong thời gian thực, các bộ phân loại này nhận dạng khuôn mặt với độ chính xác cao. Các khối xây dựng nhận dạng được hiển thị trong hình 1.

3.1. Tiền xử lý ảnh đầu vào

Bước này chúng tôi sử dụng Real-time camera có độ phân giải 1920*1080 để quay lại hình ảnh khuôn mặt ở nhiều góc độ khác nhau trong thời gian 5 giây. Sau đó chúng tôi tiến hành tách các khung hình ở trong video thành các ảnh đơn. Sau đó chúng tôi thực hiện trích xuất đặc trưng của khuôn mặt bằng cách sử dụng thư viện OpenCV4 và bộ phân loại haarcascade_frontalface và ta sẽ được bộ dữ liệu khuôn mặt ở nhiều góc độ khác nhau với kích cỡ 224*224. Sau đó ta thực hiện chia tập dữ liệu này thành 2 tập huấn luyện (train) và kiểm tra (test) bằng cách sử dụng thư viện split-folders, ta sẽ có được tập dữ liệu huấn luyện như hình 2.



Hình 2. Tập dữ liệu huấn luyện

3.2. Nhận diện khuôn mặt với CNN

Convolutional Neural Network (CNN) hay còn được gọi là mạng Nơ-ron tích chập là một trong những mô hình của Deep Learning. Tác dụng của thuật toán này chính là tạo ra những hệ thống thông minh, có sự phản ứng với độ chính xác cao. Ví dụ như Facebook, Google, đã đưa vào sản phẩm của mình chức năng nhận diện khuôn mặt,... Ứng dụng cơ bản nhất của thuật toán này là phân lớp, tức là phân biệt hoặc là cái này hoặc là cái kia, tức là khi đưa hình ảnh vào máy tính, nó sẽ là các điểm ảnh hai chiều và điều CNN thực hiện đó là khi các điểm ảnh thanh đối thì máy tính vẫn biết được đó là hình ảnh gì.

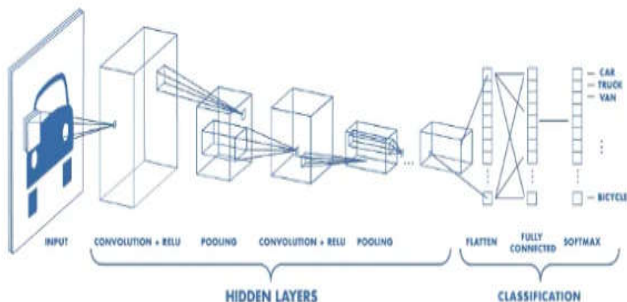
CNNs được chia thành 3 chiều: rộng, cao, sâu. Các Nơ-ron trong mạng không liên kết hoàn toàn với toàn bộ Nơ-ron kế đến mà chỉ liên kết tới một vùng nhỏ. Cuối cùng, một tầng đầu ra được tối giản thành vec-tơ của giá trị xác suất.

CNNs gồm 2 thành phần:

Phần tầng ẩn hay phần rút trích đặc trưng: Trong phần này, mạng sẽ tiến hành tính toán hàng loạt phép tích chập

(Convolutional layer) và phép hợp nhất (pooling) để phát hiện các đặc trưng.

Phần phân lớp: Tại phần này, một số lớp các liên kết đầy đủ (Fully Connected) sẽ đóng vai trò như một bộ phân lớp các đặc trưng đã rút trích trước đó. Tầng này sẽ đưa ra xác suất của một đối tượng trong hình 3.

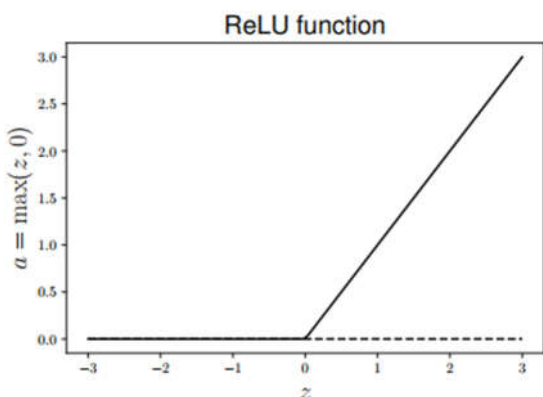


Hình 3. Cấu trúc của CNN (Nguồn <https://www.mathworks.com>)

Lớp tích chập (convolution): Trong CNN, qua ma trận mặt thời gian thực đầu vào, một ma trận còn được gọi là hạt nhân được chuyển qua để tạo ra một bản đồ đặc trưng, được sử dụng trong lớp tiếp theo. Một phép toán trong toán học được gọi là tích chập được thực hiện bằng cách sử dụng Kernel trượt qua ma trận mặt thời gian thực đầu vào. Trên mỗi vị trí, nhân ma trận khuôn mặt được thực hiện và thêm tập hợp kết quả vào bản đồ tính năng cuối cùng. Ví dụ: hãy sử dụng xem xét bộ lọc hạt nhân 2 Chiều là K và đầu vào hình ảnh 2 Chiều là I. Trong trường hợp này, hình ảnh phức hợp được tính như thể hiện trong công thức (1):

$$(i, j) = \sum_m \sum_n (m, n)(i - m, j - n) \tag{1}$$

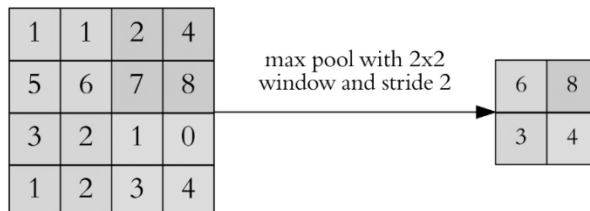
Các hàm kích hoạt phi tuyến tính (ReLU): Nút sau lớp tích chập được gọi là hàm kích hoạt. Đơn vị tuyến tính đã chỉnh lưu (ReLU) có thể được coi là hàm tuyến tính từng phần, sẽ cung cấp đầu ra dưới dạng đầu vào nếu nó là số dương, hoặc nếu không đầu ra sẽ được cho bằng không. Biểu thức của hàm ReLU là $R(z) = \max(0, z)$ hàm và hình ảnh đạo hàm của nó được thể hiện trong hình 4.



Hình 4. Hàm kích hoạt ReLU

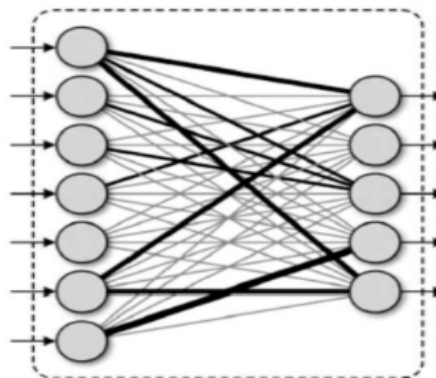
Lớp tổng hợp (Pooling): Nhược điểm lớn của đầu ra bản đồ tính năng của lớp CNN là nó ghi lại tất cả các tính năng ở vị trí chính xác của đầu vào. Điều này cho thấy rằng khi xoay, cắt xén hoặc bất kỳ thay đổi nào khác đối với đầu vào,

hình ảnh khuôn mặt sẽ tạo ra một bản đồ đặc điểm hoàn toàn khác. Để khắc phục vấn đề này, lấy mẫu theo các lớp phức hợp khác nhau được thông qua. Lấy mẫu giảm trong trường hợp này đạt được bằng cách triển khai một lớp gộp sau lớp phi tuyến tính. Chủ yếu có hai loại chức năng tổng hợp trong CNN được thể hiện trong hình 5. Trong công việc này, tính năng tổng hợp tối đa được tuân theo để nhận dạng hình ảnh khuôn mặt.



Hình 5. Lớp Pooling (Nguồn: <https://www.Viblo>)

Lớp được kết nối đầy đủ (FC): Đầu ra cuối cùng của Lớp tổng hợp cuối cùng đóng vai trò là đầu vào của Lớp được kết nối đầy đủ trong CNN. Một hoặc nhiều lớp có thể có trong này. Kết nối đầy đủ cho thấy, mọi nút trong lớp ban đầu được kết nối với mọi nút của lớp tiếp theo như minh họa trong hình 6.



Hình 6. Lớp kết nối đầy đủ

Tiến hành phân loại khuôn bằng bằng đoạn mã như hình 7.

```
import tensorflow as tf

from tensorflow.keras import datasets, layers, models

model = models.Sequential()
model.add(layers.Conv2D(32, (5, 5), activation='relu', input_shape=(224, 224, 3)))
model.add(layers.MaxPooling2D((2, 2)))
model.add(layers.Flatten())
model.add(layers.Dense(10))
model.summary()
model.compile(optimizer='adam',
              loss=tf.keras.losses.SparseCategoricalCrossentropy(from_logits=True),
              metrics=['accuracy'])

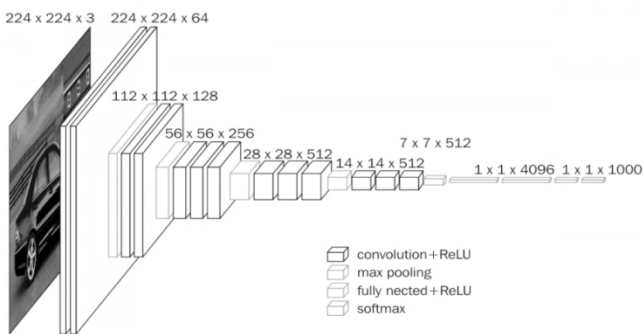
history = model.fit(train_images, train_labels, epochs=10,
                  validation_data=(test_images, test_labels))
```

Hình 7. Phân loại khuôn mặt với CNN

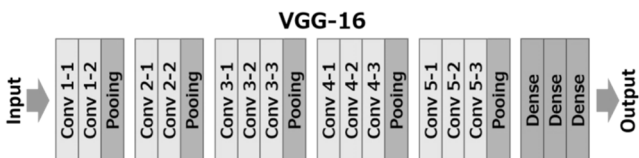
3.3. Nhận diện khuôn mặt với VGG16 và học chuyển giao

VGG16 được coi là mô hình CNN nâng cao AlexNet bằng cách thay thế các bộ lọc có kích thước hạt nhân khổng lồ (11 và 5 trong hai lớp phức hợp đầu tiên) có nhiều bộ lọc kích thước hạt nhân 3x3 tuần tự. Hình ảnh của các khuôn mặt được chuyển qua các lớp phức hợp với các bộ lọc có trường tiếp nhận tối thiểu: 3x3 (đây là kích thước nhỏ nhất

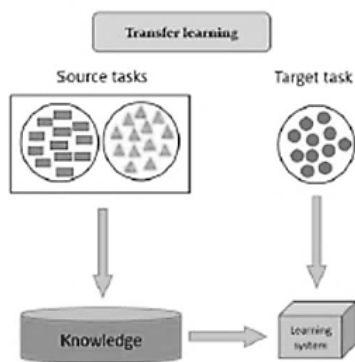
cần thiết để thu thập khái niệm lên / xuống, trái / phải và trung tâm). Nó sử dụng các bộ lọc tích chập 1x1 trong một cấu hình có thể được xem như là chuyển đổi tuyến tính các kênh đầu vào khác nhau. Tổng hợp không gian được thực hiện bằng cách sử dụng 5 lớp tổng hợp tối đa theo sau một vài lớp tích chập (mọi lớp tích chập không theo sau bằng tổng hợp tối đa). Sự sắp xếp tương tự của các lớp hồ bơi tối đa và các lớp tích tụ được thực hiện nhất quán trong toàn bộ kiến trúc. Ở phần cuối, nó có 2 lớp Được kết nối đầy đủ ngay sau đó là softmax cho đầu ra. 16 trong VGG16 chỉ ra rằng nó có 16 lớp với các trọng lượng liên quan. Học chuyển tiếp có thể cải thiện hiệu suất học tập đáng kể. Ý tưởng cốt lõi đằng sau việc học chuyển tiếp là mượn các hình ảnh khuôn mặt được dán nhãn để đạt được hiệu suất cao hơn trong lĩnh vực quan tâm cụ thể. Việc học chuyển giao chiết xuất kiến thức từ các nhiệm vụ nguồn trước đây sang các nhiệm vụ đích mà ở đó, nhiệm vụ sau có rất ít hình ảnh khuôn mặt được sử dụng để nhận dạng khuôn mặt tại hình 10 [14].



Hình 8. Mô hình VGG16 (Nguồn <https://neurohive.io/en/>)



Hình 9. Mô hình VGG16 tóm tắt (Nguồn <https://neurohive.io/en/popular-networks/vgg16/>)



Hình 10. Hình minh họa học chuyển giao

Ở bài nghiên cứu này ở đầu ra tác giả chỉ sử dụng 1 lớp kết nối đầy đủ với 1024 đơn vị và 1 lớp Dropout để tránh tình trạng overfitting có thể xảy ra.

```

from tensorflow.keras.preprocessing.image import ImageDataGenerator
#vgg16 model
classifier_vgg16 = VGG16(input_shape=(224,224,3),include_top=False,weights='imagenet')
for layer in classifier_vgg16.layers:
    layer.trainable = False

#adding extra layers for our class/images
main_model = classifier_vgg16.output
main_model = Flatten()(main_model)
main_model = Dense(1024,activation='relu',name='fc1')(main_model)
main_model = Dropout(0.5)(main_model)
main_model = Dense(14,activation='softmax',name='prediction')(main_model)

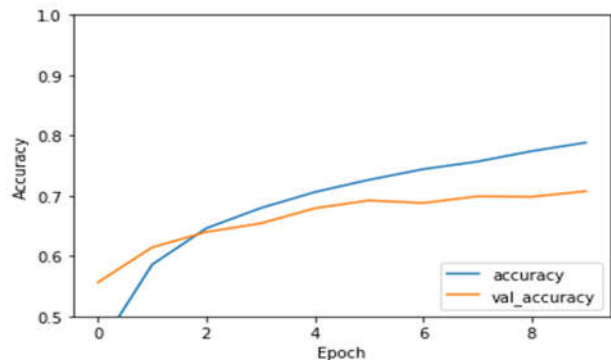
#compiling
model = Model(inputs = classifier_vgg16.input , outputs = main_model)
model.compile(optimizer='adam', loss='categorical_crossentropy', metrics=['accuracy'])

print(model.summary())
    
```

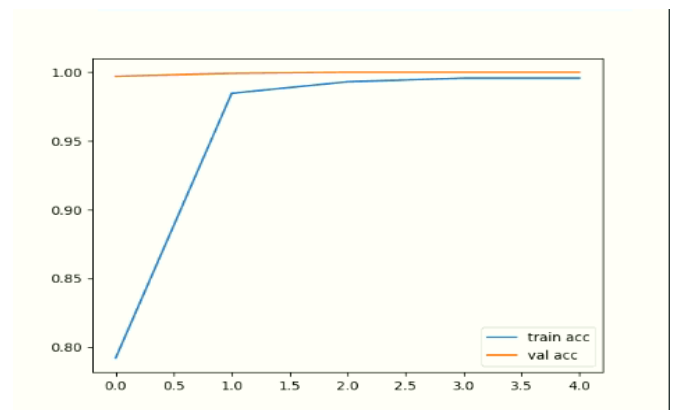
Hình 11. Định nghĩa mô hình VGG16 của nhóm tác giả

4. THỰC NGHIỆM VÀ KẾT QUẢ

Quá trình thử nghiệm và huấn luyện được thực hiện trên hệ thống máy chủ với cấu hình bộ xử lý như: CPU Intel CoreI9-10900K, bộ nhớ 64Gb RAM và 8Gb GPU RTX 2070 Super Ti. Hệ thống được cài đặt hệ điều hành Ubuntu, môi trường Python 3.6, các frameworks và thư viện cơ bản cho học máy (Machine Learning) như numpy, matplotlib, tensorflow, keras,... thuận lợi cho việc tổ chức dữ liệu phục vụ chạy thử nghiệm và lưu trữ kết quả. Theo đó, chương trình thử nghiệm của chúng tôi được xây dựng trên môi trường Python và sử dụng frameworks của tensorflow với giao diện thư viện keras, đây là thư viện cung cấp các tính năng khá mạnh mẽ cho xử lý ảnh và cho mô hình Neural network.



Hình 12. Biểu đồ CNN



Hình 13. Biểu đồ VGG16

Kết quả quá trình huấn luyện VGG16 trên tập dữ liệu được thể hiện trong hình 13. Đồ thị thể hiện khi huấn luyện dữ liệu sử dụng VGG16. Kết quả thể hiện được hiệu quả trên số epoch và độ chính xác trong quá trình huấn luyện. Đây là kết quả 5 lần chạy thử nghiệm. Kết quả trên tập dữ liệu để cho kết quả độ chính xác phân lớp (accuracy) tốt hơn so với mô hình VGG16.

5. KẾT LUẬN

Thuật ngữ Sinh trắc học xác định DNA của một cá nhân cùng với các khía cạnh khác như đặc điểm khuôn mặt, hình dạng bàn tay của họ,... Ngoài ra, các khía cạnh hành vi như chữ ký tay, giọng nói và tổ hợp phím cũng được xem xét. Trong nhiều trường hợp, tính năng nhận dạng khuôn mặt ngày càng được chấp nhận và hoan nghênh hơn trong các công nghệ dựa trên số liệu sinh học. Điều này giúp đo lường dữ liệu tự nhiên của một cá nhân. Công việc này đưa ra nhận dạng khuôn mặt thời gian thực bằng cách sử dụng các phương pháp phân loại. Hệ thống được đề xuất bao gồm ba bước chính bao gồm (1) thu thập hình ảnh khuôn mặt (2) so sánh hình ảnh khuôn mặt thời gian thực được đào tạo thông qua hai bộ phân loại như CNN và VGG16 với học chuyển giao (3). So sánh kết quả liên quan đến chỉ số độ chính xác, trong thời gian thực, bộ phân loại CNN và VGG16 nhận dạng khuôn mặt với độ chính xác cao hơn. Cả hai bộ phân loại đều được thực hiện theo cách tuần tự. Đối với mô hình VGG16 được thực hiện dựa trên việc học chuyển giao. Chuyển giao học tập có ý định trích xuất thông tin từ một số nhiệm vụ nguồn và áp dụng nó cho một nhiệm vụ đích. Vì vậy, VGG16 cho độ chính xác được cải thiện hơn so với bộ phân loại CNN. Bộ phân loại được triển khai với 7890 hình ảnh khuôn mặt của 14 người khác nhau. Hệ thống được đề xuất có thể nhận dạng thành công 14 khuôn mặt người khác nhau, có thể hữu ích trong việc tìm kiếm nghi phạm vì độ chính xác của nó cao hơn nhiều so với các phương pháp khác. Dựa trên kết quả nghiên cứu trên ta có thể hoàn toàn xây dựng các ứng dụng nhận diện khuôn mặt cho đối tượng sinh viên, giảng viên. Đây là cơ sở để chúng tôi phát triển tiếp và xây dựng ứng dụng cho bài toán thực tiễn như hệ thống điểm danh khuôn mặt các sinh viên trong lớp học, hệ thống nhận diện cán bộ tại cơ quan, hệ thống chấm công, hệ thống nhận diện dân cư...

TÀI LIỆU THAM KHẢO

- [1] Zhu X., Ramanan D., 2012. *Pose Estimation and Landmark Localization in the Wild*. IEEE Conference on Computer Vision and Pattern Recognition, Providence, pp. 2879-2886.
- [2]. Park U., Y. Tong, A. K. Jain, 2010. *Age-invariant face recognition*. IEEE transactions on pattern analysis and machine intelligence, vol. 32, no. 5, pp. 947-954.
- [3]. Li Z., U. Park, A. K. Jain, 2011. *A discriminative model for age invariant face recognition*. IEEE transactions on information forensics and security, vol. 6, no. 3, pp. 1028-1037.

- [4]. Ding C., D. Tao, 2016. *A comprehensive survey on pose-invariant face recognition*. ACM Transactions on intelligent systems and technology (TIST), vol. 7, no. 3, 37, 2016.
- [5]. Liu D.H., K.M. Lam, L.S. Shen, 2005. *Illumination invariant face recognition*. Pattern Recognition, vol. 38, no. 10, pp. 1705-1716.
- [6]. Tan X., B. Triggs, 2010. *Enhanced local texture feature sets for face recognition under difficult lighting conditions*. IEEE transactions on image processing, vol. 19, no. 6, pp. 1635-1650.
- [7]. Apoorva P., H.C. Impana, S.L. Siri, M.R. Varshitha, B. Ramesh, 2019. *Automated Criminal Identification by Face Recognition using Open Computer Vision Classifiers*. IEEE 3rd International Conference on Computing Methodologies and Communication (ICCMC), pp. 775-778.
- [8]. Raj D., 2011. *A realtime face recognition system using pca and various distance classifiers*. CS676: Computer Vision and Image Processing, pp. 1-11.
- [9]. Bah S.M., F. Ming, 2020. *An improved face recognition algorithm and its application in attendance management system*. Array.
- [10]. Kumar K.S., V.B. Semwal, R.C. Tripathi, 2011. *Real time face recognition using adaboost improved fast PCA algorithm*. arXiv preprint arXiv:1108.1353.
- [11]. Shieh M.Y., J.S. Chiou, Y.C. Hu, K.Y. Wang, 2014. *Applications of PCA and SVM-PSO based real-time face recognition system*. Mathematical Problems in Engineering.
- [12]. Shubha P., M. Meenakshi, 2019. *Human Face Recognition Using Local Binary Pattern Algorithm- Real Time Validation*. International Conference On Computational Vision and Bio Inspired Computing, pp. 240-246.
- [13]. Zhang X., T. Gonnot, J. Saniie, 2017. *Real-time face detection and recognition in complex background*. Journal of Signal and Information Processing, Vol.8, No.2, pp.99-112.
- [14]. Tammina S., 2019. *Transfer learning using VGG-16 with Deep Convolutional Neural Network for Classifying Images*. International Journal of Scientific and Research Publications, Volume 9, Issue 10, pp.143-150.

AUTHORS INFORMATION

Vien Thanh Nha¹, Tiep Sy Minh Phung¹, Pham Thanh Cong², Nguyen Thanh Tung², Le Dinh Phu Cuong³

¹Thuyloi University -Southern Campus

²Center of Information Technology, Hanoi University of Industry

³Yersin University