Nhận dạng khuôn mặt trong video bằng mạng nơ ron tích chập

Đoàn Hồng Quang*, Lê Hồng Minh, Thái Doãn Nguyên

Trung tâm Công nghệ Vi điện tử và Tin học, Viện Ứng dụng Công nghệ

Ngày nhận bài 8/7/2019; ngày chuyển phản biện 11/7/2019; ngày nhận phản biện 12/8/2019; ngày chấp nhận đăng 22/8/2019

Tóm tắt:

Deep Learning là thuật toán dựa trên một số ý tưởng từ não bộ tới việc tiếp thu nhiều tầng biểu đạt, cả cụ thể lẫn trừu tượng, qua đó làm rõ nghĩa của các loại dữ liệu. Deep Learning được ứng dụng trong nhận diện hình ảnh, nhận diện giọng nói, xử lý ngôn ngữ tự nhiên. Hiện nay rất nhiều các bài toán nhận dạng sử dụng Deep Learning, vì nó có thể giải quyết các bài toán với số lượng lớn các biến, tham số kích thước đầu vào lớn với hiệu năng cũng như độ chính xác vượt trội so với các phương pháp phân lớp truyền thống, xây dựng những hệ thống thông minh với độ chính xác cao. Trong bài báo này, các tác giả nghiên cứu mạng nơ ron tích chập (CNN - Convolutional Neural Network) là một trong những mô hình Deep Learning tiên tiến cho bài toán nhận dạng khuôn mặt từ video.

Từ khóa: **mạng nơ ron học sâu, mạng nơ ron tích chập, nhận dạng khuôn mặt.**

Chỉ số phân loại: **1.2**

Giới thiệu

CNN là một trong những mô hình mạng Học sâu phổ biến nhất hiện nay [1-3], có khả năng nhận dạng và phân loại hình ảnh với độ chính xác rất cao, thậm chí còn tốt hơn con người trong nhiều trường hợp. Mô hình này đã và đang được phát triển, ứng dụng vào các hệ thống xử lý ảnh lớn của Facebook, Google hay Amazon… cho các mục đích khác nhau, như các thuật toán gắn thẻ tự động, tìm kiếm ảnh hoặc gợi ý sản phẩm cho người tiêu dùng.

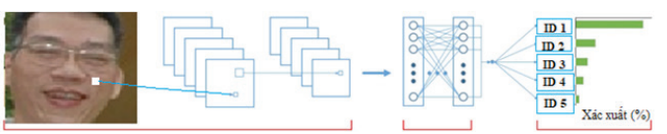
Sự ra đời của mạng CNN là dựa trên ý tưởng cải tiến cách thức các mạng nơ ron nhân tạo truyền thống [4] học thông tin trong ảnh. Do sử dụng các liên kết đầy đủ giữa các điểm ảnh vào node, các mạng nơ ron nhân tạo truyền thẳng (Feedforward Neural Network) [5-7] bị hạn chế rất nhiều bởi kích thước của ảnh, ảnh càng lớn thì số lượng liên kết càng tăng nhanh, kéo theo sự bùng nổ khối lượng tính toán. Ngoài ra, sự liên kết đầy đủ này cũng là sự dư thừa với mỗi bức ảnh, các thông tin chủ yếu thể hiện qua sự phụ thuộc giữa các điểm ảnh với những điểm xung quanh nó mà không quan tâm nhiều đến các điểm ảnh ở cách xa nhau. Mạng CNN với kiến trúc thay đổi, có khả năng xây dựng liên kết chỉ sử dụng một phần cục bộ trong ảnh kết nối đến node trong lớp tiếp theo thay vì toàn bộ ảnh như trong mạng nơ ron truyền thẳng.

Trong bài viết này, chúng tôi nghiên cứu về mạng CNN [2] sử dụng mô hình VGG16 ứng dụng trong việc xây dựng hệ thống nhận dạng khuôn mặt tự động từ video.

Mạng nơ ron CNN - VGG16

Kiến trúc mạng CNN

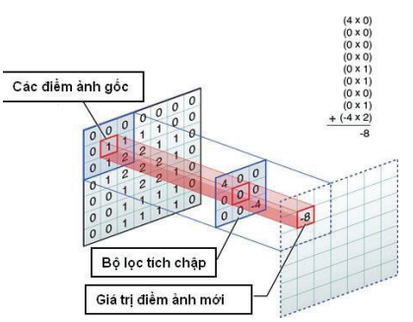
Hình 1 trình bày một kiến trúc mạng CNN, các lớp cơ bản trong một mạng CNN bao gồm: lớp tích chập (Convolutional); lớp kích hoạt phi tuyến ReLU (Rectified Linear Unit); lớp lấy mẫu (Pooling); lớp kết nối đầy đủ (Fully connected) được thay đổi về số lượng và cách sắp xếp để tạo ra các mô hình huấn luyện phù hợp cho từng bài toán khác nhau.



Các lớp tích chập (Convolutional), kích hoạt phi tuyến ReLU và lấy mẫu (Pooling) | Các lớp kết nối đầy đủ (Fully connected) | Phân loại

Hình 1. Kiến trúc cơ bản của một mạng CNN.

- Lớp tích chập: đây là thành phần quan trọng nhất trong mạng CNN, thể hiện sự liên kết cục bộ thay vì kết nối toàn bộ các điểm ảnh. Các liên kết cục bộ được tính toán bằng phép tích chập giữa các giá trị điểm ảnh trong một vùng ảnh cục bộ với các bộ lọc filters có kích thước nhỏ.



Hình 2. Bộ lọc tích chập được sử dụng trên ma trận điểm ảnh.

*Tác giả liên hệ: Email: daohaoquang@gmail.com.

Face recognition in video using convolutional neural network

Hong Quang Doan* Hong Minh Le, Doan Nguyen Thai

Center for Micro Electronics and Information Technology, National Center for Technological Progress

Received 8 July 2019; accepted 22 August 2019

Abstract:

Deep Learning is an algorithm based on some ideas from the brain to absorb many layers of expression, both concrete and abstract, thereby clarifying the meaning of data types. Deep Learning is applied in image recognition, speech recognition, natural language processing. Currently, many identification problems are solved by deep learning based methods thanks to its ability to solve problems in a large number of variables, large input size with superior performance and accuracy as compared to traditional classification methods, and its ability to build intelligent systems with high accuracy. In this article, the authors conducted a study into the convolutional neural network (CNN) which is one of the advanced deep learning models for the problem of facial recognition from video.

Keywords: **convolutional neural network, deep learning, face recognition.**

Classification number: **1.2**

Trong hình 2, bộ lọc được sử dụng là một ma trận có kích thước 3x3, bộ lọc này dịch chuyển lần lượt qua từng vùng ảnh đến khi hoàn thành quét toàn bộ bức ảnh, tạo ra một bức ảnh mới có kích thước nhỏ hơn hoặc bằng với kích thước ảnh đầu vào. Kích thước này được quyết định tùy theo kích thước các khoảng trắng được thêm ở viền bức ảnh gốc và được tính theo công thức sau:

$$O = \frac{i + 2 * p - k}{s} + 1 \quad (1)$$

Trong đó: O: kích thước ảnh đầu ra; *i*: kích thước ảnh đầu vào; p: kích thước khoảng trắng phía ngoài viền của ảnh gốc; k: kích thước bộ lọc; s: bước trượt của bộ lọc.

Như vậy, sau khi đưa một bức ảnh đầu vào cho lớp tích chập nhận được kết quả đầu ra là một loạt ảnh tương ứng với các bộ lọc đã được sử dụng để thực hiện phép tích chập. Các trọng số của các bộ lọc này được khởi tạo ngẫu nhiên trong lần đầu tiên và sẽ được cập nhật trong quá trình huấn luyện.

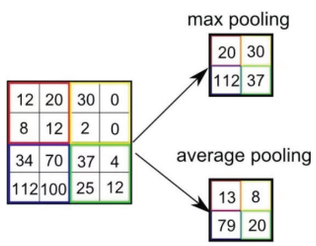
- Lớp kích hoạt phi tuyến ReLU: được xây dựng để đảm bảo tính phi tuyến của mô hình huấn luyện sau khi đã thực hiện một loạt các phép tính toán tuyến tính qua các lớp tích chập. Lớp kích hoạt phi tuyến sử dụng các hàm kích hoạt phi tuyến như ReLU hoặc sigmoid, tanh… để giới hạn phạm vi biên độ cho phép của giá trị đầu ra. Trong số các hàm kích hoạt này, hàm ReLU được chọn do cài đặt đơn giản, tốc độ xử lý nhanh mà vẫn đảm bảo được tính toán hiệu quả. Phép tính toán của hàm ReLU chỉ đơn giản là chuyển tất cả các giá trị âm thành giá trị 0.

Lớp ReLU được áp dụng ngay phía sau lớp tích chập, với đầu ra là một ảnh mới có kích thước giống với ảnh đầu vào, các giá trị điểm ảnh cũng hoàn toàn tương tự, trừ các giá trị âm đã bị loại bỏ.

$$f(x) = max(0, x) \quad (2)$$

- Lớp lấy mẫu: được đặt sau lớp tích chập và lớp ReLU để làm giảm kích thước ảnh đầu ra trong khi vẫn giữ được các thông tin quan trọng của ảnh đầu vào. Việc giảm kích thước dữ liệu có tác dụng làm giảm được số lượng tham số cũng như tăng hiệu quả tính toán. Lớp lấy mẫu cũng sử dụng một cửa sổ trượt để quét toàn bộ các vùng trong ảnh như lớp tích chập, và thực hiện phép lấy mẫu thay vì phép tích chập, sẽ chọn lưu lại một giá trị duy nhất đại diện cho toàn bộ thông tin của vùng ảnh đó.

Hình 3 thể hiện các phương thức lấy mẫu thường được sử dụng nhất hiện nay, đó là Max Pooling (lấy giá trị điểm ảnh lớn nhất) và Avarage Pooling (lấy giá trị trung bình của các điểm ảnh trong vùng ảnh cục bộ).



Hình 3. Phương thức Avarage Pooling và Max Pooling.

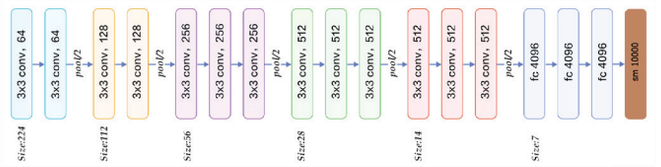
Như vậy, với mỗi ảnh đầu vào được đưa qua lấy mẫu sẽ thu được một ảnh đầu ra tương ứng, có kích thước giảm xuống đáng kể nhưng vẫn giữ được các đặc trưng cần thiết cho quá trình tính toán và nhận dạng.

- Lớp kết nối đầy đủ: được thiết kế tương tự như trong mạng nơ ron truyền thống, tất cả các điểm ảnh được kết nối đầy đủ với node trong lớp tiếp theo.

So với mạng nơ ron truyền thống [4], các ảnh đầu vào của lớp này đã có kích thước được giảm bớt rất nhiều, đồng thời vẫn đảm bảo các thông tin quan trọng của ảnh cho việc nhận dạng. Do vậy, việc tính toán nhận dạng sử dụng mô hình truyền thẳng đã không còn phức tạp và tốn nhiều thời gian như trong mạng nơ ron truyền thống.

Xây dựng mô hình mạng

Hình 4 trình bày một cấu trúc VGG16 ứng dụng vào bài toán nhận dạng khuôn mặt trong video.



Hình 4. Kiến trúc mạng VGG16.

Tổng tham số trong mô hình là 138.357.544, các tham số trong mỗi lớp của mô hình mạng như sau:

*** Ảnh đầu vào**:

Đầu vào: ảnh với kích thước 224 x 224 x 3 = 150K (3 tương ứng với 3 màu: đỏ, xanh lục, xanh lam trong hệ màu RGB thông thường).

*** Lớp 1** (Tích chập):
- Số bộ lọc: 64
- Kích thước bộ lọc: 3 x 3 x 64
- Bộ nhớ: 224 x 224 x 64 = 3,2M
- Số lượng tham số: (3 x 3 x 3) x 64 = 1.728

*** Lớp 2** (Tích chập):
- Đầu vào: 224 x 224 x 64
- Số bộ lọc: 64
- Kích thước bộ lọc: 3 x 3 x 64
- Bộ nhớ: 224 x 224 x 64 = 3,2M
- Số lượng tham số: (3 x 3 x 64) x 64 = 36.864

*** Lớp chuyển tiếp sang lớp 3** (Lấy mẫu):
- Size = (2,2)
- Stride = 2
- Padding = 0
- Bộ nhớ: 112 x 112 x 64 = 800K

Kích thước đầu ra của dữ liệu giảm 1/2, từ (224 x 224 x 3) xuống (112 x 112 x 3), và chiều sâu được giữ nguyên

*** Lớp 3** (Tích chập):
- Đầu vào: 112 x 112 x 3
- Số bộ lọc: 128
- Kích thước bộ lọc: 3 x 3 x 128
- Bộ nhớ: 112 x 112 x 128 = 1,6M
- Số lượng tham số: (3 x 3 x 64) x 128 = 73.728

*** Lớp 4** (Tích chập):
- Đầu vào: 112 x 112 x 3
- Số bộ lọc: 128
- Kích thước bộ lọc: 3 x 3 x 128
- Bộ nhớ: 112 x 112 x 128 = 1,6M
- Số lượng tham số: (3 x 3 x 128) x 128 = 147.456

*** Lớp chuyển tiếp sang lớp 5** (Lấy mẫu):
- Size = (2,2)
- Stride = 2
- Padding = 0
- Bộ nhớ: 56 x 56 x 128 = 400K

Kích thước đầu ra của dữ liệu giảm 1/2, từ (112 x 112 x 3) xuống (56 x 56 x 3), và chiều sâu được giữ nguyên

***Lớp 5** (Tích chập):
- Đầu vào: 56 x 56 x 3
- Số bộ lọc: 256
- Kích thước bộ lọc: 3 x 3 x 256
- Bộ nhớ: 56 x 56 x 256 = 800K
- Số lượng tham số: (3 x 3 x 128) x 256 = 294.912

***Lớp 6** (Tích chập):
- Đầu vào: 56 x 56 x 3
- Số bộ lọc: 256
- Kích thước bộ lọc: 3 x 3 x 256
- Bộ nhớ: 56 x 56 x 256 = 800K
- Số lượng tham số: (3 x 3 x 256) x 256 = 589.824

*** Lớp 7** (Tích chập):
- Đầu vào: 56 x 56 x 3
- Số bộ lọc: 256
- Kích thước bộ lọc: 3 x 3 x 256
- Bộ nhớ: 56 x 56 x 256 = 800K
- Số lượng tham số: (3 x 3 x 256) x 256 = 589.824

*** Lớp chuyển tiếp sang lớp 8** (Lấy mẫu):
- Size = (2,2)
- Stride = 2
- Padding = 0
- Bộ nhớ: 28 x 28 x 256 = 200K

Kích thước đầu ra của dữ liệu giảm 1/2, từ (56 x 56 x 3) xuống (28 x 28 x 3), và chiều sâu được giữ nguyên

*** Lớp 8** (Tích chập):
- Đầu vào: 28 x 28 x 3
- Số bộ lọc: 512
- Kích thước bộ lọc: 3 x 3 x 512
- Bộ nhớ: 28 x 28 x 512 = 400K
- Số lượng tham số: (3 x 3 x 256) x 512 = 1.179.648

*** Lớp 9** (Tích chập):
- Đầu vào: 28 x 28 x 3
- Số bộ lọc: 512
- Kích thước bộ lọc: 3 x 3 x 512
- Bộ nhớ: 28 x 28 x 512 = 400K
- Số lượng tham số: (3 x 3 x 512) x 512 = 2.359.296

*** Lớp 10** (Tích chập):
- Đầu vào: 28 x 28 x 3
- Số bộ lọc: 512
- Kích thước bộ lọc: 3 x 3 x 512
- Bộ nhớ: 28 x 28 x 512 = 400K
- Số lượng tham số: (3 x 3 x 512) x 512 = 2.359.296

*** Lớp chuyển tiếp sang lớp 11** (Lấy mẫu):
- Size = (2,2)
- Stride = 2
- Padding = 0
- Bộ nhớ: 14 x 4 x 512 = 100K

Kích thước đầu ra của dữ liệu giảm 1/2, từ (28 x 28 x 3) xuống (14 x 14 x 3), và chiều sâu được giữ nguyên

*** Lớp 11** (Tích chập):
- Đầu vào: 14 x 14 x 3
- Số bộ lọc: 512
- Kích thước bộ lọc: 3 x 3 x 512
- Bộ nhớ: 14 x 14 x 512 = 100K
- Số lượng tham số: (3 x 3 x 512) x 512 = 2.359.296

*** Lớp 12** (Tích chập):
- Đầu vào: 14 x 14 x 3
- Số bộ lọc: 512
- Kích thước bộ lọc: 3 x 3 x 512
- Bộ nhớ: 14 x 14 x 512 = 100K
- Số lượng tham số: (3 x 3 x 512) x 512 = 2.359.296

*** Lớp 13** (Tích chập):
- Đầu vào: 14 x 14 x 3
- Số bộ lọc: 512
- Kích thước bộ lọc: 3 x 3 x 512
- Bộ nhớ: 14 x 14 x 512 = 100K
- Số lượng tham số: (3 x 3 x 512) x 512 = 2.359.296

*** Lớp chuyển tiếp sang lớp 14** (Lấy mẫu):
- Size = (2,2)
- Stride = 2
- Padding = 0
- Bộ nhớ: 7 x 7 x 512 = 25K

Kích thước đầu ra của dữ liệu giảm 1/2, từ (14 x 14 x 3) xuống (7 x 7 x 3), và chiều sâu được giữ nguyên

*** Lớp 14** (Kết nối đầy đủ):
- Đầu vào: 1 x 1 x 4.096
- Bộ nhớ: 4.096K
- Số lượng tham số: 7 x 7 x 512 x 4.096 = 102.760.448

*** Lớp 15** (Kết nối đầy đủ):
- Đầu vào: 1 x 1 x 4.096
- Bộ nhớ: 4.096K
- Số lượng tham số: 4.096 x 4.096 = 16.777.216

*** Lớp 16** (Kết nối đầy đủ):
- Đầu vào: 1 x 1 x 4.096
- Bộ nhớ: 1.000K
- Số lượng tham số: 4.096 x 1.000 = 4.096.000

Học chuyển giao và tinh chỉnh mô hình huấn luyện

Là quá trình khai thác, tái sử dụng các tri thức đã được học tập bởi một mô hình huấn luyện trước đó vào giải quyết một bài toán mới mà không phải xây dựng mô hình huấn luyện khác từ đầu. Hiện nay, phương pháp phổ biến thường được áp dụng khi huấn luyện mô hình với một bộ CSDL tương đối nhỏ là sử dụng Học chuyển giao để tận dụng một mạng đã được huấn luyện trước.

CNN đã được huấn luyện trước đó với bộ dữ liệu rất lớn như ImageNet (1,2 triệu ảnh với 1.000 nhãn đánh dấu). Phương pháp này sử dụng mạng CNN theo hai cách chính như sau:

Mạng CNN này chỉ được sử dụng như một bộ trích chọn đặc trưng cho bộ CSDL huấn luyện mới, bằng cách thay thế các lớp kết nối đầy đủ ở cuối mô hình mạng và giữ cố định các tham số cho toàn bộ các lớp còn lại của mô hình. Thực hiện tối ưu, tinh chỉnh (Fine-tune) một vài hoặc tất cả các lớp trong mô hình mạng.

Việc tái sử dụng mạng CNN là dựa trên các đặc trưng được học trong các lớp đầu của mạng là các đặc trưng chung nhất với phần lớn bài toán (ví dụ đặc trưng về cạnh, hình khối hay các khối màu…). Các lớp sau đó của mạng CNN sẽ nâng dần độ cụ thể, riêng biệt của các chi tiết phục vụ cho bài toán nhận dạng cần giải quyết. Do đó, hoàn toàn có thể tái sử dụng lại các lớp đầu của mạng CNN mà không phải mất nhiều thời gian và công sức huấn luyện từ đầu.

Có 2 loại học chuyển giao:

Feature Extractor: sau khi lấy ra các đặc điểm (tai, mũi, tóc…) của ảnh bằng việc sử dụng ConvNet của mô hình được huấn luyện trước, sẽ dùng phân loại tuyến tính (Linear SVM, Softmax Classifier...) để phân loại ảnh.

Fine-tuning: sau khi lấy ra các đặc điểm của ảnh bằng việc sử dụng CNN của mô hình được huấn luyện trước, thì sẽ coi đây là đầu vào của CNN mới bằng cách thêm các lớp tích chập và lớp kết nối đầy đủ.

Kết quả nhận dạng khuôn mặt bằng mô hình VGG16

* Nguồn dữ liệu dùng huấn luyện mô hình thử nghiệm được thu thập trên Internet:

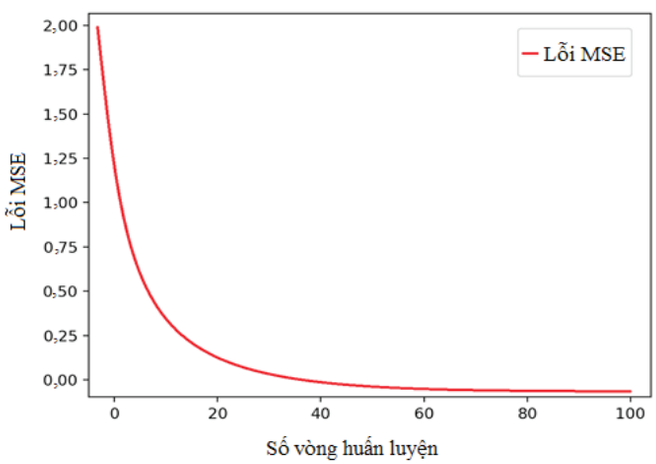
- Tổng số người: 2.622
- Tổng số khuôn mặt: 1.200.000

* Khởi tạo các thông số để huấn luyện mạng:
- Tốc độ học: 0,25
- Hệ số quán tính: 0,3
- Sai số cực tiểu: 0,00001
- Số lần học tối đa: 100 vòng

* Huấn luyện mạng:

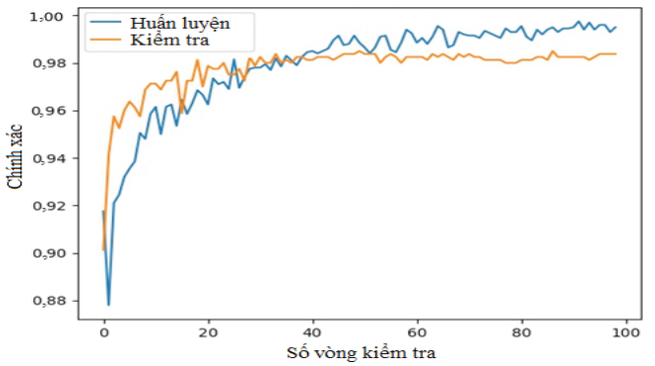
Môi trường được sử dụng để huấn luyện mô hình nhận dạng là Windows Server 2012, ngôn ngữ Python phiên bản 3.7.1 với framework dùng cho huấn luyện mô hình là Caffe, card đồ họa Nvidia 1080ti, trong khoảng 20 ngày huấn luyện.

Huấn luyện mạng: lấy 80% mẫu dữ liệu để huấn luyện mạng, kết quả huấn luyện thể hiện ở hình 5 với sai số MSE là 0,002 qua 100 vòng huấn luyện.



Hình 5. Đồ thị lỗi huấn luyện với mẫu dữ liệu.

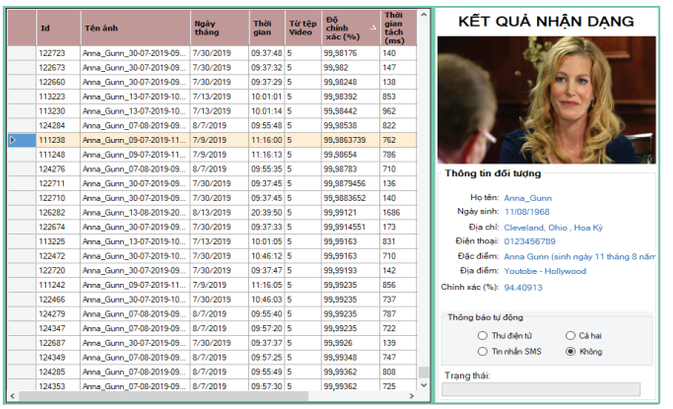
* Kiểm tra mạng: lấy 20% mẫu dữ liệu còn lại để kiểm tra mạng với các bộ trọng số đã được huấn luyện, kết quả kiểm tra với lỗi MSE là 0,0001 (hình 6).



Hình 6. Đồ thị lỗi kiểm tra.

* Nhận dạng:

Dữ liệu video nhận dạng được lấy trên youtube với kết quả cao như trong hình 7, và một số kết quả được thống kê trong bảng 1.



Hình 7. Nhận dạng khuôn mặt trong video.

Bảng 1. Kết quả thử nghiệm nhận dạng khuôn mặt từ video.

Id	Tên ảnh	Thời gian	Từ tệp video	Độ chính xác (%)
111240	Anna_Gunn_09-07-2019-11-15-30.jpg	11:15:30	5	98,18
111241	Anna_Gunn_09-07-2019-11-15-31.jpg	11:15:31	5	97,45
111242	Anna_Gunn_09-07-2019-11-15-32.jpg	11:15:32	5	96,95
111243	Anna_Gunn_09-07-2019-11-15-33.jpg	11:15:33	5	98,23
111244	Anna_Gunn_09-07-2019-11-15-33.jpg	11:15:33	5	98.22
111245	Anna_Gunn_09-07-2019-11-15-34.jpg	11:15:34	5	99.63
111246	Anna_Gunn_09-07-2019-11-15-35.jpg	11:15:35	5	97.72
111247	Anna_Gunn_09-07-2019-11-15-36.jpg	11:15:36	5	95.55
111248	Anna_Gunn_09-07-2019-11-15-37.jpg	11:15:37	5	94.66
111249	Anna_Gunn_09-07-2019-11-15-38.jpg	11:15:38	5	97.33
111250	Anna_Gunn_09-07-2019-11-15-38.jpg	11:15:38	5	99,24
111251	Anna_Gunn_09-07-2019-11-15-39.jpg	11:15:39	5	99,02
111252	Anna_Gunn_09-07-2019-11-15-40.jpg	11:15:40	5	95,61
111253	Anna_Gunn_09-07-2019-11-15-41.jpg	11:15:41	5	99,84
111254	Anna_Gunn_09-07-2019-11-15-42.jpg	11:15:42	5	94,32
111255	Anna_Gunn_09-07-2019-11-15-45.jpg	11:15:45	5	91,48
111256	Anna_Gunn_09-07-2019-11-15-48.jpg	11:15:48	5	92,60
111257	Anna_Gunn_09-07-2019-11-15-49.jpg	11:15:49	5	93,23
111258	Anna_Gunn_09-07-2019-11-15-50.jpg	11:15:50	5	99,78
111259	Anna_Gunn_09-07-2019-11-15-51.jpg	11:15:51	5	98,64
111260	Anna_Gunn_09-07-2019-11-15-52.jpg	11:15:52	5	98,12
111261	Anna_Gunn_09-07-2019-11-15-53.jpg	11:15:53	5	99,70
111262	Anna_Gunn_09-07-2019-11-15-54.jpg	11:15:54	5	99,73
111263	Anna_Gunn_09-07-2019-11-15-55.jpg	11:15:55	5	99,92
111264	Anna_Gunn_09-07-2019-11-15-56.jpg	11:15:56	5	99,93
111265	Anna_Gunn_09-07-2019-11-15-56.jpg	11:15:56	5	99,77
111266	Anna_Gunn_09-07-2019-11-15-59.jpg	11:15:59	5	99,56
111267	Anna_Gunn_09-07-2019-11-16-00.jpg	11:16:00	5	99,97
111268	Anna_Gunn_09-07-2019-11-16-01.jpg	11:16:01	5	99,85

Kết luận

Mô hình mạng VGG16 với kiến trúc thay đổi, khả năng xây dựng liên kết chỉ sử dụng một phần cục bộ trong ảnh kết nối đến node trong lớp tiếp theo thay vì toàn bộ ảnh như trong mạng nơ ron truyền thăng, làm tăng khả năng xử lý và đạt tỷ lệ cao trong phân loại ảnh.

Độ chính xác nhận dạng khuôn mặt của mô hình trong điều kiện lý tưởng đã đạt hoặc vượt qua cả con người. Tuy nhiên, do các yếu tố khác nhau như ánh sáng, góc, biểu hiện và tuổi tác, làm giảm độ chính xác của quá trình nhận dạng. Trong thời gian tới, các tác giả sẽ tập trung vào việc xây dựng và bổ sung các tập thuộc tính để nâng cao độ chính xác của quá trình nhận dạng.

Từ những kết quả đã thử nghiệm của mô hình cho thấy, có thể xây dựng các ứng dụng dựa trên phân loại và nhận dạng khuôn mặt, như: hệ thống chấm công tự động, điểm danh tự động trong các cơ sở đào tạo, và các hệ thống kiểm soát an ninh, phòng chống tội phạm.

TÀI LIỆU THAM KHẢO

[1] A. Canziani, A. Paszke and E. Culurciello (2016), "An analysis of deep neural network models for practical applications", *arXiv preprint arXiv:1605.07678.ss*

[2] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, T. Darrell Caffe (2014), "Convolutional Architecture for Fast Feature", *Embedding arXiv:1408.5093.*

[3] Y. Sun, X. Wang and X. Tang (2014), "Deep learning face representation by joint identification-verification", *CoRR, abs/1406.4773.*

[4] Đoàn Hồng Quang, Lê Hồng Minh, Chu Anh Tuấn (2015), "Nhận dạng bàn tay bằng mạng nơ ron nhân tạo", *Tuyển tập báo cáo Diễn đàn "Đổi mới - Chìa khóa cho sự phát triển bền vững"*, Viện Ứng dụng Công nghệ, Bộ Khoa học và Công nghệ.

[5] Đoàn Hồng Quang, Lê Hồng Minh (2014), "Dùng RFNN kết hợp khử mùa và khử xu hướng để dự báo chỉ số giá vàng trên thị trường", *Tuyển tập báo cáo Diễn đàn "Đổi mới - Chìa khóa cho sự phát triển bền vững"*, Viện Ứng dụng Công nghệ, Bộ Khoa học và Công nghệ.

[6] Nguyễn Quang Hoan, Đoàn Hồng Quang (2014), "Dự báo chỉ số giá chứng khoán bằng RFNN", *Tạp chí Khoa học và Công nghệ, Trường Đại học Sư phạm Kỹ thuật Hưng Yên*, **1**, tr.52-56.

[7] Nguyễn Quang Hoan, Dương Thu Trang, Đoàn Hồng Quang (2018), "Dự báo số học sinh nhập trường bằng mạng nơ ron nhân tạo", *Tạp chí Khoa học và Công nghệ, Trường Đại học Sư phạm Kỹ thuật Hưng Yên*, **18**, tr.1-8.