

NHẬN DIỆN GÓC XOAY KHUÔN MẶT SỬ DỤNG MẠNG NƠN TÍCH CHẬP

Bùi Vũ Long – 19146205
Trí Tuệ Nhân Tạo – Lớp 05CLC
Khoa Đào Tạo Chất Lượng Cao

Tóm tắt: Mạng nơon tích chập là một thuật toán Deep Learning có thể lấy hình ảnh đầu vào, gán độ quan trọng (các trọng số - weights và độ lệch - bias có thể học được) cho các đặc trưng/đối tượng khác nhau trong hình ảnh và có thể phân biệt được từng đặc trưng/đối tượng này với nhau. Công việc tiền xử lý được yêu cầu cho mạng nơon tích chập thì ít hơn nhiều so với các thuật toán phân loại khác. Trong các phương thức sơ khai, các bộ lọc được thiết kế bằng tay (hand - engineered), với một quá trình huấn luyện để chọn ra các bộ lọc/đặc trưng phù hợp thì mạng nơon tích chập lại có khả năng tự học để chọn ra các bộ lọc/ đặc trưng tối ưu nhất. Trong bài viết lần này chúng ta sẽ đi áp dụng mạng nơon tích chập vào việc nhận diện góc xoay khuôn mặt.

Từ khoá: Mạng nơon tích chập, nhận diện góc xoay khuôn mặt, Deep Learning, Convolutional Neural Network (CNN)

I. Giới thiệu.

1. Nhận dạng khuôn mặt

Hệ thống nhận dạng khuôn mặt là một ứng dụng máy tính tự động xác định hoặc nhận dạng một người nào đó từ một bức hình ảnh kỹ thuật số hoặc một khung hình video từ một nguồn video. Một trong những cách để thực hiện điều này là so sánh các đặc điểm khuôn mặt chọn trước từ hình ảnh và một cơ sở dữ liệu về khuôn mặt.

Hệ thống này thường được sử dụng trong các hệ thống an ninh và có thể được so sánh với các dạng sinh trắc học khác như các hệ thống nhận dạng vân tay hay tròng mắt.

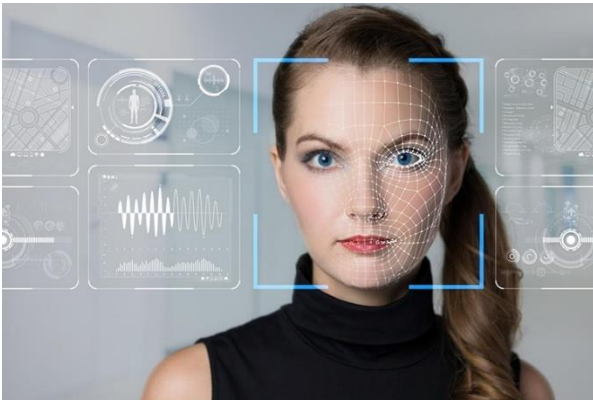
2. Ứng dụng nhận dạng góc xoay trong nhận dạng khuôn mặt

Trong việc điểm danh, chấm công: Nhờ nhận dạng được khuôn mặt người đang không để đúng hướng, khiến cho việc quét khuôn mặt khó khăn, hệ thống

sẽ đưa ra cảnh báo và thống báo điều chỉnh khuôn mặt sao cho hợp lý.

Trong máy camera, quay phim chụp hình: Camera sẽ di chuyển theo hướng khuôn mặt người di chuyển, từ đó bắt trọn được mọi khoảnh khắc.

Ngoài ra còn một số ứng dụng khác như: một số hãng xe trang bị để thông báo tài xế ngủ gật; hệ thống bảo mật công nghệ;...



Hình 1. Nhận dạng khuôn mặt.

3. Các phương pháp tiếp cận.

Để xác định được góc xoay khuôn mặt, ta có một số hướng tiếp cận với vấn đề như sau:

Hướng tiếp cận dựa trên tri thức: Trong hướng tiếp cận này, các luật sẽ phụ thuộc rất lớn vào tri thức của những tác giả nghiên cứu về bài toán xác định khuôn mặt người. Đây là hướng tiếp cận dạng top-down. Dễ dàng xây dựng các luật cơ bản để mô tả các đặc trưng của khuôn mặt và các quan hệ tương ứng. Ví dụ, một khuôn mặt thường có hai mắt đối xứng nhau qua trục thẳng đứng ở giữa khuôn

mặt và có một mũi, một miệng. Các quan hệ của các đặc trưng có thể được mô tả như quan hệ về khoảng cách và vị trí. Thông thường các tác giả sẽ trích đặc trưng của khuôn mặt trước tiên để có được các ứng viên, sau đó các ứng viên này sẽ được xác định thông qua các luật để biết ứng viên nào là khuôn mặt và ứng viên nào không phải khuôn mặt. Thường áp dụng quá trình xác định để giảm số lượng xác định sai.

Hướng tiếp cận đặc trưng không thay đổi: Đây là hướng tiếp cận theo kiểu bottom-up. Các tác giả cố gắng tìm các đặc trưng không thay đổi của khuôn mặt người để xác định khuôn mặt người. Dựa trên nhận xét thực tế, con người dễ dàng nhận biết các khuôn mặt và các đối tượng trong các tư thế khác nhau và điều kiện ánh sáng khác nhau, thì phải tồn tại các thuộc tính hay đặc trưng không thay đổi. Có nhiều nghiên cứu đầu tiên xác định các đặc trưng khuôn mặt rồi chỉ ra có khuôn mặt trong ảnh hay không. Các đặc trưng như: lông mày, mắt, mũi, miệng, và đường viền của tóc được trích bằng phương pháp xác định cạnh. Trên cơ sở các đặc trưng này, xây dựng một mô hình thống kê để mô tả quan hệ của các đặc trưng này và xác định sự tồn tại của khuôn mặt trong ảnh. Một vấn đề của các thuật toán theo hướng tiếp cận đặc trưng cần phải điều chỉnh cho phù hợp điều kiện ánh sáng, nhiễu, và bị che khuất. Đôi khi bóng của khuôn mặt sẽ tạo thêm cạnh mới, mà cạnh này lại rõ hơn cạnh thật sự của khuôn mặt,

vì thế nếu dùng cạnh để xác định sẽ gặp khó khăn.

Hướng tiếp cận so khớp mẫu: Trong so khớp mẫu, các mẫu chuẩn của khuôn mặt (thường là khuôn mặt được chụp thẳng) sẽ được xác định trước hoặc xác định các tham số thông qua một hàm. Từ một ảnh đưa vào, tính các giá trị tương quan so với các mẫu chuẩn về đường viền khuôn mặt, mắt, mũi và miệng. Thông qua các giá trị tương quan này mà các tác giả quyết định có hay không có tồn tại khuôn mặt trong ảnh. Hướng tiếp cận này có lợi thế là rất dễ cài đặt, nhưng không hiệu quả khi tỷ lệ, tư thế, và hình dáng thay đổi (đã được chứng minh). Nhiều độ phân giải, đa tỷ lệ, các mẫu con, và các mẫu biến dạng được xem xét thành bất biến về tỷ lệ và hình dáng.

Hướng tiếp cận dựa trên diện mạo: Trái ngược với các phương pháp so khớp mẫu với các mẫu đã được định nghĩa trước bởi những chuyên gia, các mẫu trong hướng tiếp cận này được học từ các ảnh mẫu. Một cách tổng quát, các phương pháp theo hướng tiếp cận này áp dụng các kỹ thuật theo hướng xác suất thống kê và máy học để tìm những đặc tính liên quan của khuôn mặt và không phải là khuôn mặt. Các đặc tính đã được học ở trong hình thái các mô hình phân bố hay các hàm biệt số nên dùng có thể dùng các đặc tính này để xác định khuôn mặt người. Đồng thời, bài toán giảm số chiều thường được quan

tâm để tăng hiệu quả tính toán cũng như hiệu quả xác định.

II. Convolutional Neural Network

Convolutional Neural Network (CNN hoặc ConvNet) được tạm dịch là: Mạng nơ ron tích tụ. Đây được xem là một trong những mô hình của Deep Learning – tập hợp các thuật toán để có mô hình dữ liệu trừu tượng hóa ở mức cao bằng cách sử dụng nhiều lớp xử lý cấu trúc phức tạp.

Ví dụ: Bạn có thể sử dụng loại mạng thần kinh này trong các ứng dụng như: nhận dạng hình ảnh, nhận dạng khuôn mặt và hiện tại đã được đưa vào các nền tảng Facebook, Google,...

1. Feature là gì?

Feature được dịch theo nghĩa tiếng Việt là đặc điểm. Khi sử dụng thuật toán CNN so sánh hình ảnh theo từng mảnh, mỗi mảnh đó được gọi là Feature.

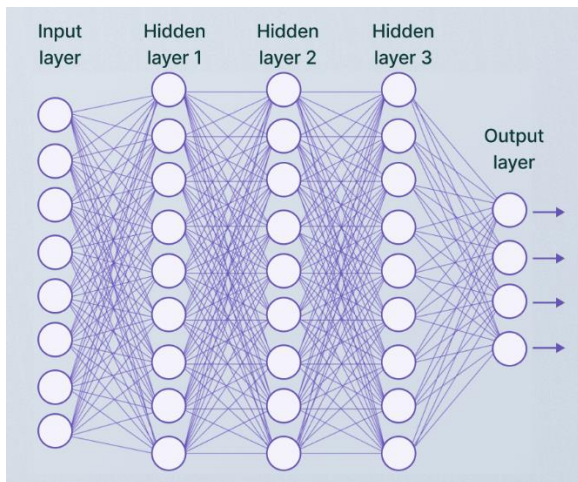
Mỗi Feature được xem như một hình ảnh mini hay gọi là những mảnh hai chiều nhỏ. Các Feature được khớp với những khía cạnh chung trong bức ảnh đó. Nghĩa là Feature sẽ tương ứng với khía cạnh nào đó của hình ảnh và chúng sẽ khớp lại với nhau.

2. Convolutional là gì?

Convolutional được hiểu là tích chập. Xét về cơ bản, khi xem một hình ảnh mới, thuật toán CNN sẽ không nhận biết

được nó ở vị trí nào, các Feature sẽ khớp với nhau ở đâu?

Chính vì vậy, Convolutional sẽ thử chúng với tất cả các vị trí khác nhau và tạo thành một bộ lọc gọi là Filter. Quá trình này được thực hiện thông qua phần toán nơ ron tích chập.



Hình 2. Mạng nơ ron tích chập.

3. Một số lớp cơ bản trong Convolutional Neural Network.

a. *Convolutional Layer*: đại diện CNN thực hiện mọi phép toán.

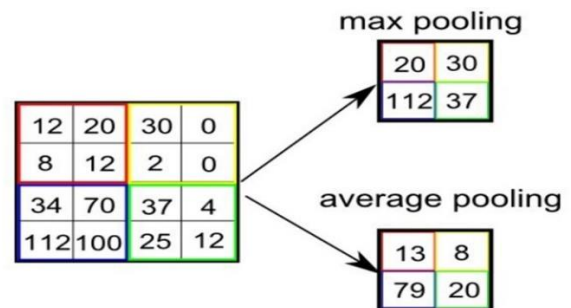
b. *Pooling Layer*:

Khi đầu vào quá lớn, các lớp Pooling Layer sẽ được dịch chuyển vào giữa những lớp Convolutional Layer nhằm giảm các Parameter.

Pooling Layer được biết đến với hai loại phổ biến là: Max Pooling và Average Pooling.

Tại Pooling Layer, khi bạn sử dụng lớp Max Pooling thì số lượng Parameter có

thể sẽ giảm đi. Vì vậy, Convolutional Neural Network sẽ xuất hiện nhiều lớp Filter Map, mỗi Filter Map đó sẽ cho ra một Max Pooling khác nhau.



Hình 3. Lớp Pooling.

c. *Relu Layer*

Đây chính là một hàm kích hoạt trong Neural Network. Chúng ta có thể biết đến hàm kích hoạt này với một tên gọi khác là Activation Function. Nhiệm vụ chính của hàm kích hoạt là mô phỏng lại các Neuron có tỷ lệ truyền xung qua Axon. Trong đó, hàm kích hoạt sẽ bao gồm các hàm cơ bản như: Sigmoid, Tanh, Relu, Leaky Relu, Maxout.

d. *Fully Connected Layer*

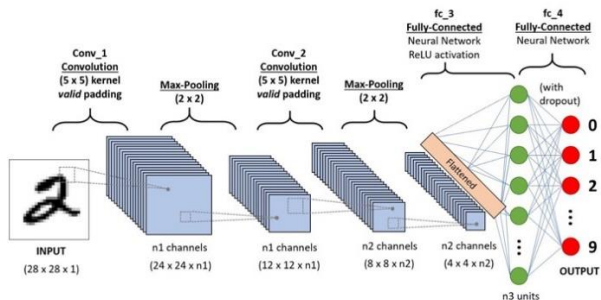
Fully Connected Layer thường sử dụng để đưa ra các kết quả.

Cấu trúc của Convolutional Neural Network

Mạng Convolutional Neural Network là tập hợp nhiều lớp Convolutional chồng lên nhau, sử dụng các hàm Nonlinear Activation và tanh để kích hoạt các trọng số trong các node. Ở mỗi lớp CNN, sau khi được các hàm kích

hoạt sẽ tạo ra các thông tin trừu tượng hơn cho những lớp tiếp theo. Mỗi Layer kết tiếp sẽ là kết quả Convolution từ Layer trước đó nên chúng ta có được các kết nối cục bộ.

Thông qua quá trình huấn luyện mạng, các lớp Layer CNN tự động học các giá trị được thể hiện qua các lớp Filter.



Hình 4. Cấu trúc mạng nơ ron tích chập.

III. Xây dựng kiến trúc mạng CNN nhận dạng góc xoay khuôn mặt.

1. Xây dựng mạng CNN.

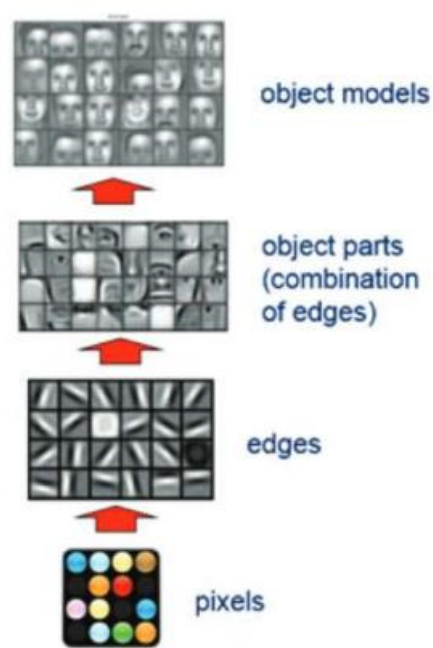
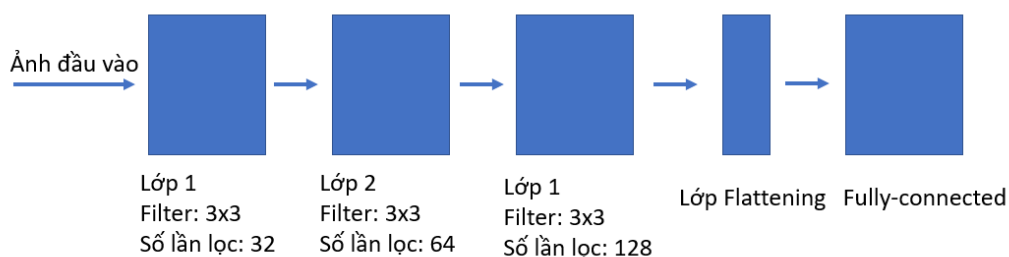
Xây dựng tầng Convolutional

Xây dựng tầng Poolong: Sử dụng 64 bộ lọc filter

Xây dựng hàm biếng đổi tuyến tính

Kết nối Fully-connected.

Cụ thể được mô tả như sau: $[[CONV \rightarrow RELU]*N \rightarrow POOL?]*M \rightarrow [FC \rightarrow RELU]*K \rightarrow FC$



Hình 5. Thứ tự các bộ lọc trong CNN.

2. Huấn luyện mô hình.

Bài toán chỉ áp dụng trong mô hình nhỏ chỉ nhận 9 góc độ khác nhau của khuôn mặt. Sự dụng tập ảnh các góc khuôn mặt do chính em tự tạo. Mô hình dùng 80% ảnh để huấn luyện và 20% ảnh để kiểm chứng.

Bước 1: Xây dựng tập dữ liệu là ảnh góc nghiêng các khuôn mặt, phân loại các thư mục và lưu vào thư mục Train.

Hình 6. Mô hình huấn luyện.



Hình 7. Tập mẫu huấn luyện.

Bước 2: Thay đổi kích thước huấn luyện thành cỡ 200x200

Bước 3: Đọc dữ liệu train

Bước 4: Xây dựng mô hình:

```
model=tf.keras.models.Sequential()
# lớp CNN1
model.add(Conv2D(32,(3,3), activation='relu',input_shape=(200,200,3)))
model.add(MaxPooling2D(2,2))
# lớp CNN2
model.add(Conv2D(64,(3,3), activation='relu'))
model.add(MaxPooling2D(2,2))
# lớp CNN3
model.add(Conv2D(128,(3,3), activation='relu'))
model.add(MaxPooling2D(2,2))
# chuyển dữ liệu đầu ra của mạng CNN từ mảng 2 chiều về mảng 1 chiều
model.add(Flatten())
# lớp ẩn
model.add(Dense(128, activation=tf.nn.relu))
# lớp output
model.add(Dropout(0.2))
model.add(Dense(9, activation=tf.nn.softmax))
model.summary()
```

Hình 8. Xây dựng các lớp huấn luyện.

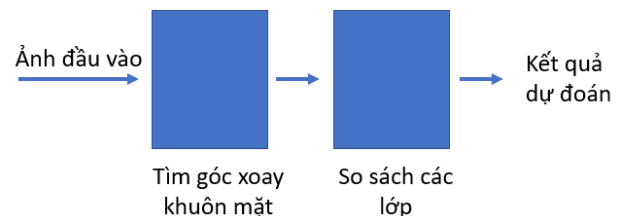
Bước 5: Thực hiện huấn luyện.

```
=====] - 1s 639ms/step - loss: 1.3484 - acc: 0.5000 - val_loss: 1.1122 - val_acc: 0.7500
=====] - 1s 614ms/step - loss: 1.2746 - acc: 0.5938 - val_loss: 0.9950 - val_acc: 0.7969
=====] - 1s 644ms/step - loss: 1.1114 - acc: 0.6562 - val_loss: 1.0691 - val_acc: 0.6406
=====] - 1s 638ms/step - loss: 1.0005 - acc: 0.6562 - val_loss: 0.7976 - val_acc: 0.7656
=====] - 1s 627ms/step - loss: 0.9021 - acc: 0.7188 - val_loss: 0.8973 - val_acc: 0.6719
=====] - 1s 659ms/step - loss: 0.6624 - acc: 0.6875 - val_loss: 0.6750 - val_acc: 0.7812
=====] - 1s 664ms/step - loss: 0.8077 - acc: 0.8438 - val_loss: 0.5969 - val_acc: 0.8906
=====] - 1s 629ms/step - loss: 0.7989 - acc: 0.7188 - val_loss: 0.5474 - val_acc: 0.9219
=====] - 1s 627ms/step - loss: 0.5861 - acc: 0.8750 - val_loss: 0.3302 - val_acc: 0.9688
=====] - 1s 620ms/step - loss: 0.6028 - acc: 0.7500 - val_loss: 0.4189 - val_acc: 0.8906
=====] - 1s 626ms/step - loss: 0.4353 - acc: 0.9375 - val_loss: 0.3916 - val_acc: 0.8906
=====] - 1s 631ms/step - loss: 0.8241 - acc: 0.7812 - val_loss: 0.4621 - val_acc: 0.8906
=====] - 1s 631ms/step - loss: 0.3787 - acc: 0.8438 - val_loss: 0.4431 - val_acc: 0.8594
```

Hình 9. Kết quả huấn luyện.

Bước 6: Lưu kết quả huấn luyện.

3. Xây dựng mô hình dự đoán:



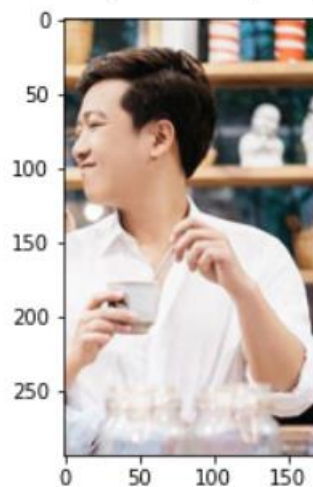
Hình 10. Mô hình dự đoán

Bước 1: Load file huấn luyện.

Bước 2: Đọc ảnh và xử lý.

Bước 3: So sánh và đưa ra kết quả dự đoán

Góc xoay khuôn mặt dự đoán được là: 90



Hình 11. Kết quả dự đoán.

IV. Hướng phát triển.

Sử dụng real-time cho thuật toán.

Tỉ lệ nhận dạng còn phụ thuộc vào điều kiện, môi trường sáng xung quanh.

Chưa nhận diện được toàn bộ hướng khuôn mặt.

Sử dụng kết hợp học tăng cường cùng thuật toán để tăng độ chính xác cho mô hình.

V. Kết luận.

Mô hình tuy đã hoàn thành nhưng vẫn còn nhiều thiếu sót. Chưa áp dụng được realtime vào bài toán và độ nhận diện chính xác chưa cao mặc dù training ra chỉ số xấp xỉ 1.

Theory and Practice, Wiley, [ISBN 978-0-470-51706-2](#), 2009 ([1] TM book)

7. Williams, Mark. "[Better Face-Recognition Software](#)". Truy cập ngày 2 tháng 6 năm 2008.

Liên kết GitHub:

<https://github.com/LongbuiVu/AI-FINAL.git>



Tài liệu tham khảo.

1. "[Facial Recognition Applications](#)". Animetrics.
2. [Face Recognition Technology Overview](#)
3. Bonsor, K. "[How Facial Recognition Systems Work](#)".
4. Smith, Kelly. "[Face Recognition](#)" (PDF).
5. R. Brunelli and T. Poggio, "Face Recognition: Features versus Templates", IEEE Trans. on PAMI, 1993, (15)10:1042-1052
6. R. Brunelli, Template Matching Techniques in Computer Vision: