

**TRƯỜNG ĐẠI HỌC BÁCH KHOA HÀ NỘI**  
**VIỆN TOÁN ỨNG DỤNG VÀ TIN HỌC**  
————— o0o —————



**ĐỒ ÁN I**

**BÀI TOÁN NHẬN DIỆN KHUÔN MẶT**

**Giáo viên hướng dẫn: Ts. TRẦN NGỌC THĂNG**

**Sinh viên thực hiện : TRỊNH HOÀNG ĐỨC  
PHẠM NGỌC QUANG ANH  
ĐẶNG HỮU TÚ**

**Lớp : KSTN TOÁN TIN - K60**

**HÀ NỘI - 2019**

# Mục lục

<b>1</b>	<b>Mô hình nhận diện khuôn mặt</b>	<b>4</b>
1.1	Detect face . . . . .	4
1.1.1	CNN . . . . .	4
1.1.2	MT-CNN . . . . .	6
1.2	Recognize face . . . . .	10
1.2.1	Inception Resnet v1 . . . . .	10
<b>2</b>	<b>Kết quả</b>	<b>11</b>
<b>3</b>	<b>Kết luận</b>	<b>12</b>
	Tài liệu tham khảo	13

# Mở đầu

aa

# Chương 1

## Mô hình nhận diện khuôn mặt

Mô hình nhận diện khuôn mặt được chia làm hai bước:

- **Detect face**: nhận diện khu vực khuôn mặt có trong ảnh. Mô hình nhận diện được dùng ở đây là **MT-CNN**

- **Recognize face**: lấy đầu vào là khuôn mặt được nhận diện trong ảnh từ kết quả của phần detect face. Sau đó, cho kết quả này qua một mô hình CNN để lấy ra đặc trưng khuôn mặt. Đặc trưng này sẽ được so khớp với các đặc trưng của đối tượng muốn so khớp được lưu sẵn trong cơ sở dữ liệu. Từ đó, cho ra kết quả ảnh đầu vào chứa đối tượng nào. Mô hình để trích rút đặc trưng khuôn mặt ở đây là **Inception ResNet V1**

### 1.1 Detect face

#### 1.1.1 CNN

Mạng CNN là mô hình được sử dụng phổ biến trong nhận diện đối tượng. Điểm nổi bật của mô hình này so với mô hình mạng neural thông thường là có thể trích rút ra đặc trưng chính của đối tượng trong quá trình training.

Công việc trích rút đặc trưng chính được liên quan trực tiếp đến hai phần là **convolution** và **pooling**.

#### Convolution

Ở pha **convolution**, các tầng ẩn được gọi là **convolution layer**. Các **convolution layer** này là tập các **feature map** và mỗi feature map chính là đầu ra của tầng trước và là đầu vào của tầng sau trong mô hình.

Để có feature map, ta cần có **convolution layer**. Ở đây, **convolution layer** (gọi tắt là **filter**) là một ma trận sẽ quét lần lượt qua ma trận dữ liệu đầu vào (từ trái qua phải, từ trên xuống dưới) và nhân tương ứng từng giá trị của ma trận đầu vào với giá trị trong **filter**, đưa qua hàm kích hoạt (tanh, relu, elu, ...). Kết quả của quá trình này chính là

feature map.

Input và Filter

1	1	1	0	0
0	1	1	1	0
0	0	1	1	1
0	0	1	1	0
0	1	1	0	0

Input

1	0	1
0	1	0
1	0	1

Filter / Kernel

Kết quả

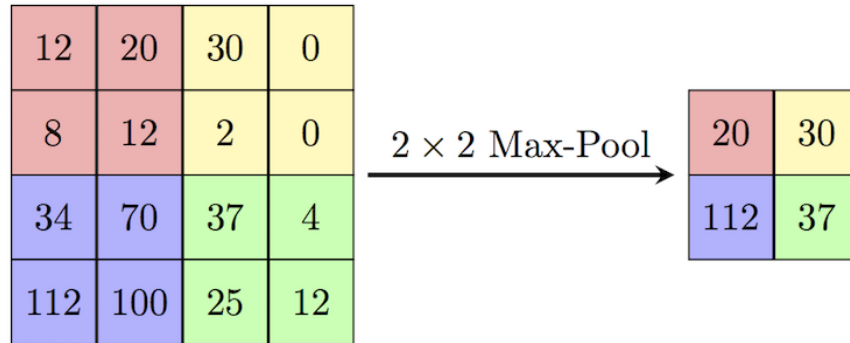
1	1	1	0	0
0	1	1	1	0
0	0	1x1	1x0	1x1
0	0	1x0	1x1	0x0
0	1	1x1	0x0	0x1

4	3	4
2	4	3
2	3	4

## Pooling

Mục đích của **pooling** là làm giảm số chiều của ma trận, từ đó giảm thời gian tính toán và còn tránh được overfit.

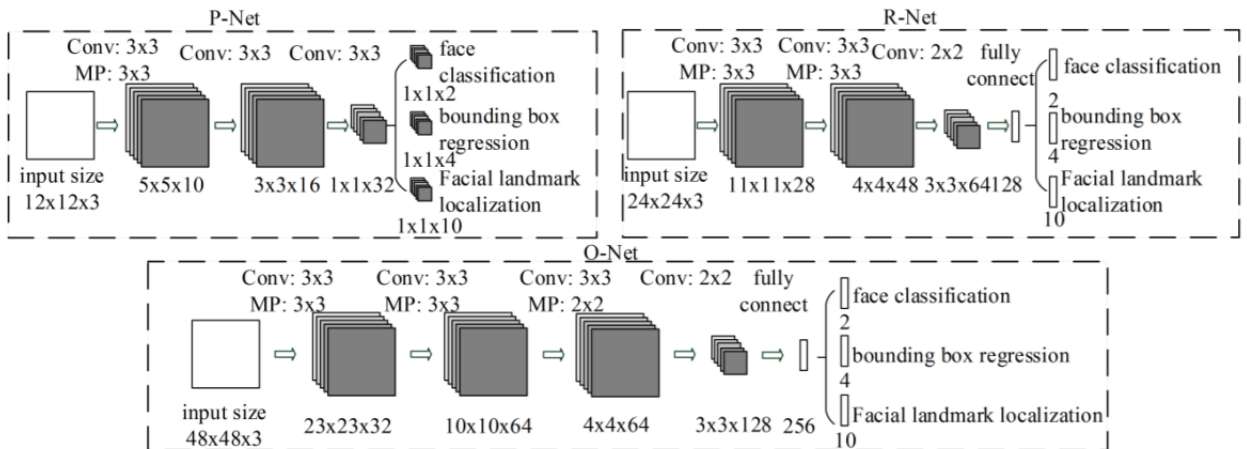
Kiểu **pooling** được sử dụng phổ biến nhất là **max pooling** với từng ô  $2 \times 2$



### 1.1.2 MT-CNN

**MT-CNN** là mô hình mạng **neural** phát triển từ **CNN**. Mô hình này hoạt động rất mạnh trong việc nhận diện khuôn mặt. Đặc trưng của nó là dùng tăng độ chính xác cho việc nhận diện mặt bằng cách nhận diện thêm năm điểm đặc trưng trên khuôn mặt bao gồm: hai mắt, mũi và hai khóe miệng.

Kiến trúc của mạng **MT-CNN** là dùng lần lượt ba mô hình P-net, O-net, R-net.



**P-net**

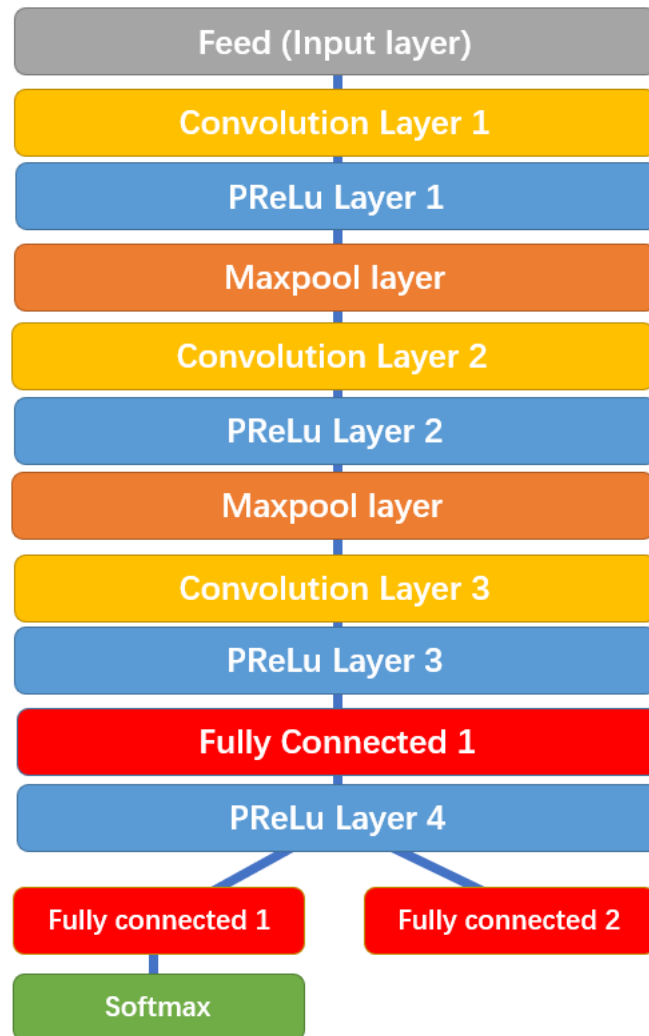
Tại P-Net, thuật toán sử dụng 1 kernel 12x12 chạy qua mỗi bức hình để tìm kiếm khuôn mặt.



Sau lớp convolution thứ 3, mạng chia thành 2 lớp. Convolution 4-1 đưa ra xác suất của một khuôn mặt nằm trong mỗi bounding boxes, và Convolution 4-2 cung cấp tọa độ của các bounding boxes. Đến đây, ta đã có cơ bản các **bounding box** có xác suất cao là chứa khuôn mặt.

**R-net**

R-Net có cấu trúc tương tự với P-Net. Tuy nhiên sử dụng nhiều layer hơn. Tại đây, network sẽ sử dụng các bounding boxes đc cung cấp từ P-Net và tính chỉnh là tọa độ.



Tương tự R-Net chia ra làm 2 layers ở bước cuối, cung cấp 2 đầu ra đó là tọa độ mới của các bounding boxes, cùng độ tin tưởng của nó.

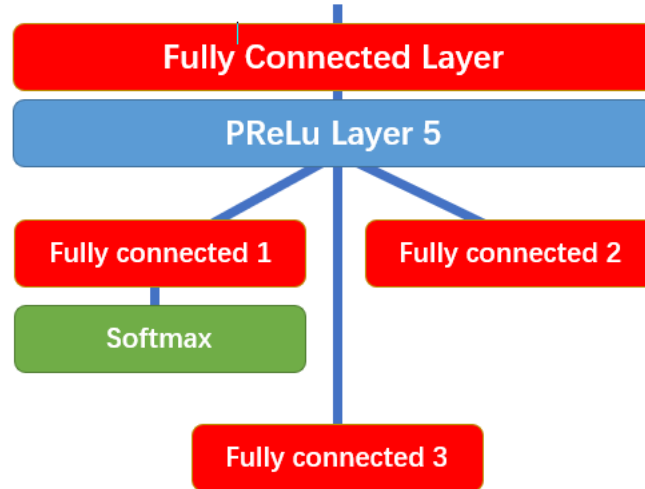


### O-net

O-Net lấy các bounding boxes từ R-Net làm đầu vào và đánh dấu các tọa độ của các mốc trên khuôn mặt.



Ở bước này, thuật toán đưa ra 3 kết quả đầu ra khác nhau bao gồm: xác suất của khuôn mặt nằm trong bounding box, tọa độ của bounding box và tọa độ của các mốc trên khuôn mặt (vị trí mắt, mũi, miệng).



## 1.2 Recognize face

Đầu vào của recognize là ma trận ảnh khuôn mặt được lấy từ phần nhận diện khuôn mặt. Ma trận ảnh này sẽ qua mô hình mạng **Inception ResNet v1** để được một vector đặc trưng với số chiều cố định. Sau đó, ta so khớp vector này với các vector đặc trưng đã được ghi sẵn trong cơ sở dữ liệu để cho ra kết quả.

### 1.2.1 Inception Resnet v1

## Chương 2

### Kết quả

## Chương 3

## Kết luận

aaa

# Tài liệu tham khảo

- [1] On Certain Formal Properties of Grammars  
*Noam Chomsky. In Information and Control 2, page 137 - 167, 1959.*
- [2] Otomat và ngôn ngữ hình thức  
*Đoàn Văn Ban, Bộ môn Khoa học máy tính, Khoa Công nghệ thông tin, Đại học Thái Nguyên.*
- [3] Introduction to the Theory of Computation  
*Michael Sipser. Cengage Learning, Boston, third edition, 2013.*
- [4] A Second Course in Formal Language  
*Jeffrey Shallit. Cambridge University Press, 2009.*