**Danh mục bảng viết tắt:**

|  |  |  |
| --- | --- | --- |
| **STT** | **Từ Viết Tắt** | **Ý Nghĩa Đầy Đủ** |
| 1 | MTCNN | Multi-task Cascaded Convolutional Networks: Mạng Nơ-Ron Tích Chập nối tầng đa nhiệm |
| 2 | SVM | Support Vector Machine: Máy vector hỗ trợ |
| 3 | CNN | Convolutional Neural Network: Mạng Nơ-Ron Tích Chập |
| 4 | DNN | Deep Neural Network: Mạng Nơ-Ron sâu |

**Chương 1: Giới thiệu**

Nhận dạng khuôn mặt (Face recognition) là một chủ đề nghiên cứu thuộc lĩnh vực thị giác máy tính (Computer vision) đã được phát triển từ đầu những năm 90 của thế kỷ trước. Cho tới hiện nay, đây vẫn là một chủ đề nghiên cứu mở nhận được sự quan tâm của nhiều nhà nghiên cứu từ nhiều lĩnh vực nghiên cứu khác nhau như nhận dạng mẫu (Pattern recognition), học máy (Machine learning), thống kê (Statistic), sinh trắc học (Biometrics). Điều này là do có rất nhiều ứng dụng thực tế cần tới một hệ thống nhận dạng khuôn mặt, từ các hệ thống quản lý đăng nhập đơn giản cho tới các ứng dụng giám sát tại các địa điểm công cộng (Public Area Surveillance) hoặc quản lý dân số (Population Management) và pháp lý (Forensics). Bên cạnh đó, so với các hệ thống nhận dạng dựa trên các đặc điểm sinh trắc học khác của con người như nhận dạng mống mắt và vân tay (Fingerprint and iris), dáng đi (Gait recognition), nhận dạng khuôn mặt có nhiều ưu điểm:

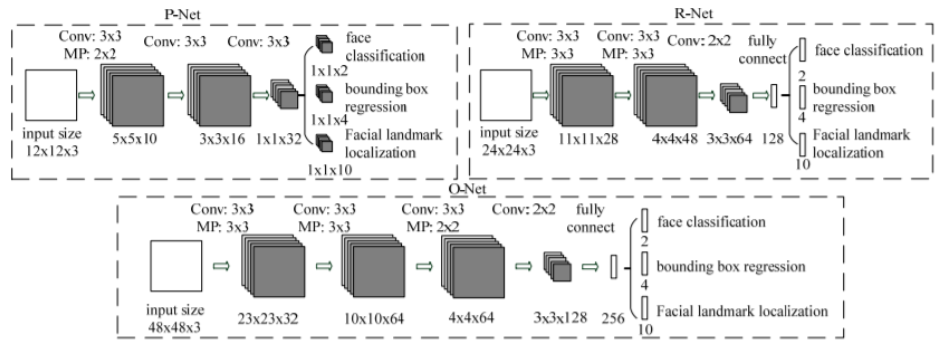
* Một hệ thống nhận dạng khuôn mặt không đòi hỏi sự tương tác trực tiếp giữa đối tượng và nhận dạng hệ thống.
* Việc thu nhận dữ liệu (các bức ảnh về khuôn mặt con người) cho quá trình nhận dạng một con người dễ thực hiện hơn so với việc thu nhận các đặc điểm sinh trắc học khác (như thu nhận dấu vân tay, …)
* Dữ liệu về khuôn mặt phổ biến hơn so với các đặc trưng khác do sự bùng nổ các mạng xã hội (Facebook, Twitter, Yahoo, …), các dịch vụ chia sẻ dữ liệu đa phương tiện (Youtube, Vimeo, …) và sự phát triển nhanh chóng của các thiết bị thu nhận hình ảnh.
* Từ khuôn mặt của một người, ta có thể khai thác nhiều thông tin liên quan chứ không chỉ là danh tính của người đó, ví dụ như giới tính, màu da, hướng nhìn, chủng tộc, hành vi, sức khỏe, độ tuổi, cảm xúc, …

**Chương 2: Các kỹ thuật trong hệ thống nhận dạng khuôn mặt**

1. Nhận dạng khuôn mặt với mạng MTCNN và FaceNet
   1. MTCNN (Multi-task Cascaded Convolutional Networks)

MTCNN hoạt động theo 3 bước, mỗi bước có một mạng neural riêng lần lượt là: P-Net, R-Net và O-net.

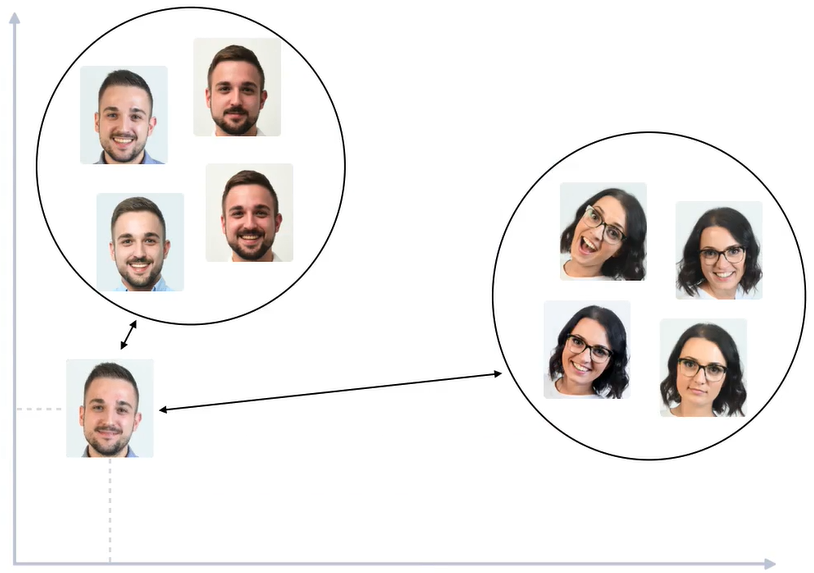
* P-Net: Proposal network, dự đoán các vùng trong bức ảnh có thể là khuôn mặt (trong đó có nhiều vùng không phải khuôn mặt).
* R-Net: Refine network, sử dụng đầu ra của P-Net để loại bỏ các vùng không phải khuôn mặt.
* O-Net: Output network, sử dụng đầu ra của R-Net để đưa ra kết quả cuối cùng với 5 facial landmark (2 mắt + mũi + 2 khóe miệng).

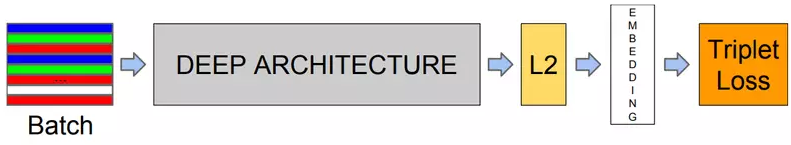


* 1. FaceNet (A Unified Embedding for Face Recognition and Clustering)

[FaceNet](https://arxiv.org/abs/1503.03832) là một mạng thần kinh sâu được sử dụng để trích xuất các đặc điểm từ hình ảnh khuôn mặt của một người. Nó được xuất bản vào năm 2015 bởi các nhà nghiên cứu của Google, Schroff et al.

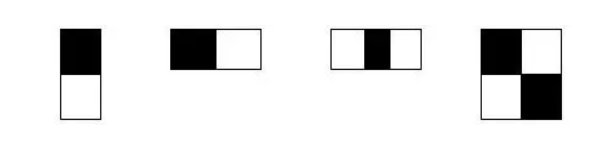
FaceNet lấy hình ảnh khuôn một người và trả về một vector có 128 features cho một khuôn mặt. Sau đó dung SVM để phân nhóm các vector đó vào các nhóm để biết vector đó là mặt của ai.

Quá trình training:

* Sử dụng một tập Dataset với rất nhiều các cá thể người khác nhau, mỗi cá thể có một số lượng ảnh nhất định.
* Xây dựng một mạng DNN dùng để làm Feature Extractor cho Dataset trên, kết quả là 1 embedding 128-Dimensions. Trong paper có 2 đại diện mạng là Zeiler&Fergus và InceptionV1.
* Huấn luyện mạng DNN để kết quả embedding có khả năng nhận diện tốt, bao gồm 2 việc là sử dụng l2l2 normalization (Khoảng cách Euclide) cho các embeddings đầu ra và tối ưu lại các parameters trong mạng màng bằng Triplet Loss.
* Hàm Triplet Loss sẽ sử dụng phương pháp Triplet Selection, lựa chọn các embeddings sao cho việc học diễn ra tốt nhất.
  1. MTCNN kết hợp FaceNet

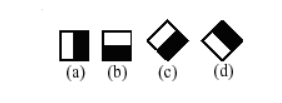
MTCNN và FaceNet là 2 mạng rất nổi tiếng trong việc xử lý bài toán Face Recognition nói chung. Và việc kết hợp giữa chúng, khi đầu vào là ảnh/video với rất nhiều người và trong hoàn cảnh thực tế, sẽ đưa ra được kết quả khá tốt. Khi đó, MTCNN sẽ đóng vai trò là Face Detection/Alignment, giúp cắt các khuôn mặt ra khỏi khung hình dưới dạng các tọa độ bounding boxes và chỉnh sửa / resize về đúng shape đầu vào của mạng FaceNet. Còn FaceNet sẽ đóng vai trò là mạng Feature Extractor + Classifier cho từng bounding boxes, đưa ra embedding và tiền hành phân biệt và nhận dạng các khuôn mặt.

1. Sử dụng AdaBoost
   1. Đặc trưng của Haar-Like

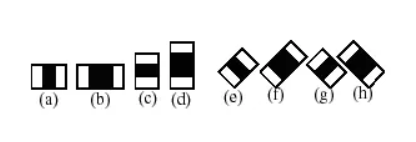
Mỗi đặc trưng Haar-Like là sự kết hợp của hai hay ba hình chữ nhật trắng hay đen như trong hình sau:

Để sử dụng các đặc trưng này vào việc xác định khuôn mặt người, 4 đặc trưng Haar-Like cơ bản được mở rộng ra và được chia làm 3 tập đặc trưng như sau:

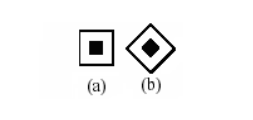
* + Đặc trưng cạnh (edge feature)

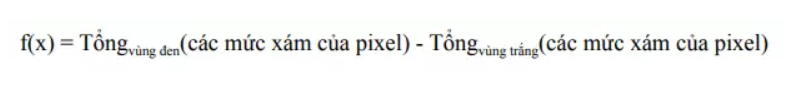


* + Đặc trưng đường (line feature)

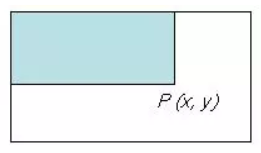


* + Đặc trưng xung quanh tâm (center-surround features)

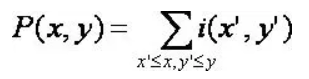


Dùng các đặc trưng trên, ta có thể tính được các giá trị của đặc trưng Haar-Like là sự chênh lệch giữa tổng của các pixel của vùng đen và vùng trắng như trong công thức sau:

Viola và Joines đưa ra một khái niệm gọi là Integral Image, là một mảng 2 chiều với kích thước bằng với kích thước của ảnh cần tính đặc trưng Haar-Like, với mỗi phần tử của mảng này được tính bằng cách tính tổng của điểm ảnh phía trên (dòng-1) và bên trái (cột-1) của nó



.

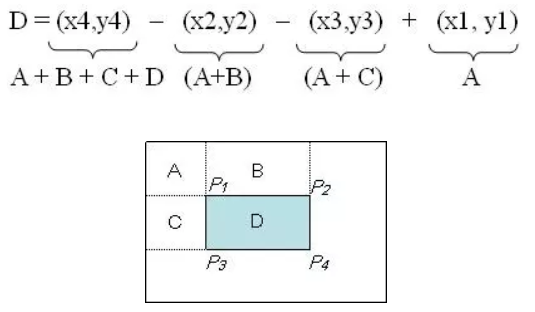
Công thức tính Intergral Image:

Sau khi tính được Integral Image, việc tính tổng các giá trị mức xám của một vùng bất kỳ nào đó trên ảnh thực hiện rất đơn giản theo cách sau:

Giả sử ta cần tính tổng giá trị mức xám của vùng D như hình dưới, ta có thể tính được như sau:

D = A + B + C + D – (A+B) – (A+C) + A

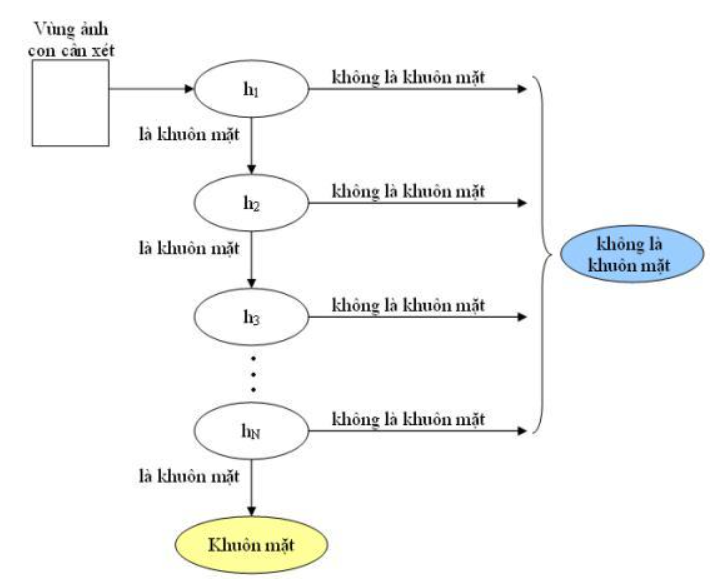
Với A + B + C + D chính là giá trị tại điểm P4 trên Integral Image, tương tự như vậy A+B là giá trị tại điểm P2, A+C là giá trị tại điểm P3, và A là giá trị tại điểm P1. Vậy ta có thể viết lại biểu thức tính D ở trên như sau:



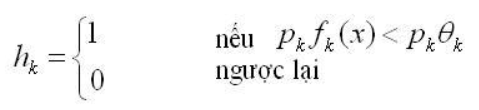
* 1. AdaBoost

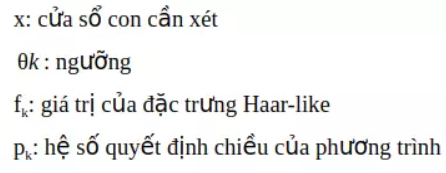
AdaBoost là một bộ phân loại mạnh phi tuyến phức dựa trên hướng tiếp cận boosting được Freund và Schapire đưa ra vào năm 1995. Adaboost cũng hoạt động trên nguyên tắc kết hợp tuyến tính các weak classifiers để hình thành một trong các classifiers.

Viola và Jones dùng AdaBoost kết hợp các bộ phân loại yếu sử dụng các đặc trưng Haar-like theo mô hình phân tầng (cascade) như sau:

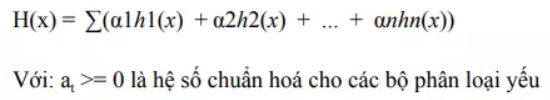


Trong đó, **h(k)** là các bộ phân loại yếu, được biểu diễn như sau:

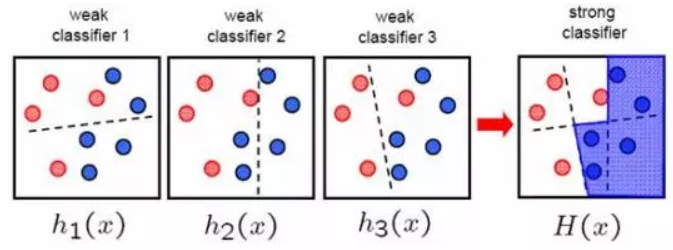


Với:

AdaBoost sẽ kết hợp các bộ phân loại yếu thành bộ phân loại mạnh như sau:



Đây là hình ảnh minh họa việc kết hợp các bộ phân loại yếu thành bộ phân loại mạnh:



* 1. Sơ đồ nhận diện khuôn mặt

