**Tìm hiểu về Face Detection, Face Recognition và một số model phổ biến dùng để nhận diện khuôn mặt**

1. **Computer Vision.**
2. *Định nghĩa.*

Thị giác máy tính là một mảng trong lĩnh vực AI, cho phép các máy tính hoặc hệ thống tự định nghĩa được các thông tin dựa trên những hỉnh ảnh, video kỹ thuật số, …Tự đó máy tính hoặc hệ thống có thể thực hiện, phản hổi dựa trên các thông tin đã phân tích.

Thị giác máy tính khá giống với thị giác con người, tuy nhiên con người có một thời gian rất dài trong cuộc sống để quan sát và tiếp thu sự vật, sự việc biến đổi, để rồi nhận thức được đó là vật gì, đó là tình huống gì. Còn đối với máy tính những thị giác, võng mạc, hệ thống dây thần kinh được thay bằng các camera, dự liệu và các thuật toán. Quá trình học của máy tính phải tương đối nhanh hơn con người để có thể trả ra kết quả một cách nhanh và chính xác nhất.

1. *Cách hoạt động.*

Cách mà computer vision hoạt động cơ bản là dùng một lượng lớn dữ liệu được đưa vào trong máy tình, từ đó tiến hành phân tích, quá trình này sẽ lặp đi lặp lại cho đến khi nhận ra được sự khác biệt giữa các dữ liệu (bức ảnh).

Và để máy tính thực hiện được việc này cần hai yếu tố chủ yếu hỗ trợ đó là machine learning và mạng nơ tron tích chập CNN.

Với machine learning, may tính sẽ được xây dựng một thuật toán để có thể xử lý được các ảnh đầu vào, nhận biết được ngữ cảnh của các dữ liệu, phân biệt được ngữ cảnh và vật thể trong dữ liệu đầu vào. Sau đó máy tính sẽ đẩy các dữ liệu đó vào một cái mô hình để huấn luyện (hay nói cách khác là để cho máy tính “nhìn”).

Với CNN, CNN là thuật toán trong Deep Learning ra đời năm 1980, của Pháp Yan LeCun. Sau khi đưa dữ liệu vào mô hình huấn luyện, thuật toán CNN sẽ từ các dữ liệu đưa vào trích xuất ra các đặc trưng riêng biệt của từng bức ảnh, cụ thể là máy tính sẽ xem các dữ liệu thành là một tập các pixel được dán nhãn (label). Sử dụng các label để thực hiện tích chập để tự dự đoán ra đối tượng đang “nhìn” là cái gì. Các bức ảnh sẽ được đưa qua nhiều lớp như vậy trong mạng nơ tron, đưa ra được độ chính xác với các dự đoán qua nhiều lần tích chập cho đến khi dữ đoán đưa ra là chính xác.

1. *Ứng dụng.*

Thị giác máy tính là một công nghệ thông minh được ứng dụng rộng rãi có thể kể đến như dùng để Image classification, Object Detection, Object Tracking, Content-base image retrieval, automatics cars, Face Detection, Optical Character Recognition

1. **Face Detection.**
2. *Định nghĩa.*

Phát hiện khuôn mặt là một ứng dụng của Computer Vision, và cũng là một phần không thể thiếu trong quá trình nhận diện khuôn mặt (Face recognition). Phát hiện khuôn mặt sẽ giúp máy tính nhận dạng được khuôn mặt của con người sau khi qua huấn luyện với tập dữ liệu rất lớn về mặt người.

1. *Cách hoạt động.*

Như đã nói các bài toán của Computer Vision là một thuật toán với đầu vào là một lượng lớn dữ liệu được sử dụng, và dữ liệu được sử dụng ở đây là các hình ảnh kỹ thuật số về khuôn mặt người. Các thuật toán phát hiện khuôn mặt luôn bắt đầu với việc tìm kiếm mắt vì nó là dể nhất sau đó sẽ tiếp tục với lông mày, mũi, miệng,... Sau khi đã kiểm tra xong các đặt điểm trên, máy tính sẽ xác nhận đó là khuôn mặt người hay không phải. Mỗi một bức ảnh ta để làm như thế để huấn luyện mô hình từ đó làm tăng thêm đọ chính xác trong việc phát hiện khuôn mặt.

Các phương pháp dùng trong phát hiện khuôn mặt là:

Như nói trên, ta sử dụng các đặt trưng như mắt mũi miệng để thực hiện nhận diện khuôn mặt, đây cũng là một cách phát hiện khuôn mặt tuy nhiên có nhược điểm là sẽ bị ảnh hưởng bởi các nhiễu và ánh sáng. Phương pháp thứ hai sẽ là phát hiện khuôn mặt dựa trên các hình ảnh khuôn mặt mẫu tiêu chuẩn cao mang đi so sánh. Nhưng cách này sẽ bị ảnh hưởng bởi các khuôn mặt có tỷ lệ khác nhau, hình dạng kì lạ, … Phương pháp thức ba, sử dụng các đặc điểm để trích xuất các đặc trưng liên quan đến hình ảnh khuôn mặt.

Một số kĩ thuật cần lưu ý trong phát hiện khuôn mặt như remove background, đây là kỹ thuật phổ biến dùng trong các bài toán xử lý ảnh, xóa đi nền sẽ giúp việc trích xuất trở nên chính xác hơn tránh có nhiễu. Kĩ thuật cần lưu ý nữa là việc phát hiện khuôn mặt trong các video, khi đó khuôn mặt sẽ chuyển động không theo quy luật rất khó phân biệt vậy nên ta cần tính toán ra vùng hoạt động của khuôn mặt để có thể cho ra kết quả tốt nhất.

Một số thuật toán Deep learning dùng để phát hiện khuôn mặt nói riêng và phát hiện vật thế nói chung là CNN, R-CNN, SSD(Single Shot Detection)

1. *Ứng dụng cụ thể.*

Phát hiện khuôn mặt được sử dụng trong nhiều lĩnh vực như bảo mật, trắc sinh học, phát luật, giải trí và thông tin cá nhân. Nó hỗ trợ nhận diện khuôn mặt, nhận diện cảm xúc, phân tích khuôn mặt,…

1. *Ưu điểm và nhược điểm.*

Ưu điểm: tăng cường bảo mật, dể dàng tích hợp với các ứng dụng thực tế, tự động nhận diện được khuôn mặt người.

Nhược điểm: dữ liệu quá lớn, phát hiện có thể không hoàn toàn chính xác bởi các loại khuôn mặt khác biệt lớn hay các góc máy ảnh, có khả năng xâm nhập quyền riêng tư của người khác.

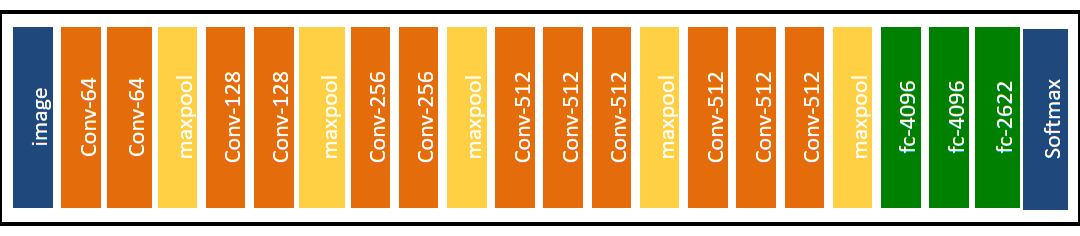
1. **Face Recognition**.

Nhận diện khuôn mặt cùng phát hiện khuôn mặt luôn là đi chung với nhau để tạo ra các ứng dụng thực tế. Tuy nhiên phát hiện khuôn mặt chỉ dừng lại ở mức phát hiện được mặt của con người, còn nhận diện khuôn mặt không chỉ phát hiện được khuôn mặt của con người mà còn phân biệt, nhận diện được đây là người nào, người này có đặc điểm gì khác so với những khuôn mặt khác. Đây là khác biệt duy nhất giữa hai phương pháp, còn lại về thuật toán, ứng dụng thì tương đối giống nhau

1. Một số model phổ biến dùng để nhận diện khuôn mặt.
2. VGG-Face model

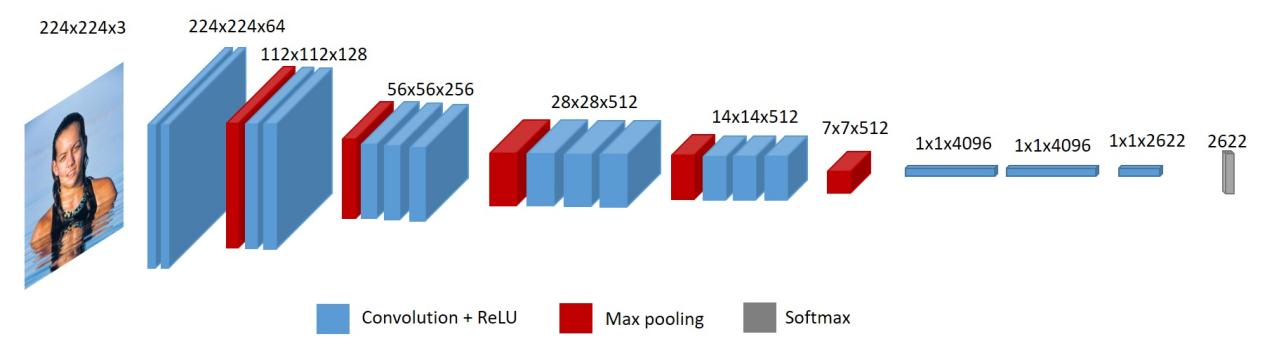
VGG được viết tắt là Visual Geometry Group, VGG-Face được ra đời năm 2015, model dựa trên mạng tích chập CNN để phát triên một mô hình nhận dạng khuôn mặt chuẩn.

Kiến trúc của model VGG-Face là:



Kiến trúc chia là 6 block khác nhau, và cứ sao mỗi lớp Maxpolling sẽ hình thành một block, trong một block có ít nhất 2 lớp tích chập và một lớp maxpooling, riêng block cuối cùng thì bao gồm các lớp fully-connected network và lớp dense với activation function là softmax.

Để có thể hình dung rõ hơn về mô hình ta nhìn hình ảnh sau:



Ban đầu từ một tấm ảnh có dimension là 3, sẽ được qua từng lớp tích chập chiều thứ 3 của tấp ảnh tăng lên từ 3 sang 64, việc tăng số lượng lớp cho chiều thứ 3 giúp việc trích xuất và học dữ liệu thu được trở nên nhiều hơn tăng độ chính xác. Sau đó ta đưa kết quả vào lớp maxpooling, ở lớp này sẽ giúp tăng cường ảnh để việc trích xuất trở nên tốt hơn bằng cách làm kích thước của ảnh nhỏ lại. Nhắc lại, lớp Maxpooling chỉ có tác dụng tăng cường ảnh, không hỗ trợ việc học dữ liệu của model. Cứ trải quả quá trình lặp lại như thế, rồi đưa qua lớp fully-connected để cập nhật các trọng số sau quá trình học, từ đó đúc kết lại chỉ còn một vector một chiều với các giá trị được tính toán sau khi qua hàm activation function là softmax. Hàm loss function được dùng là Cross-entropy loss.

1. Face-Net mode

1. Open-Face model

1. Deep-Face model