****

TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN

VÀ TRUYỀN THÔNG VIỆT - HÀN

**KHOA KHOA HỌC MÁY TÍNH**

**BÁO CÁO**

**ĐỒ ÁN CHUYÊN NGÀNH 2**

**TÊN ĐỀ TÀI**

**XÂY DỰNG HỆ THỐNG QUẢN LÝ ĐIỂM DANH BẰNG NHẬN DIỆN KHUÔN MẶT**

**Sinh viên thực hiện : Lê Thành An**

**Lê Văn Hiểu**

**Mã sinh viên : 19IT061**

**19IT078**

**Lớp 19IT2**

**Giảng viên hướng dẫn : Ths. Đỗ Công Đức**

***Đà Nẵng, tháng 05 năm 2022***

LỜI CẢM ƠN

Để đồ án này đạt kết quả tốt đẹp, em đã nhận được sự hỗ trợ, giúp đỡ của thầy cô. Với tình cảm sâu sắc, chân thành, cho phép em được bày tỏ lòng biết ơn sâu sắc đến tất cả quý thầy cô đã tạo điều kiện giúp đỡ trong quá trình học tập và nghiên cứu và phát triển đề tài.

Trước hết em xin gửi tới các thầy cô trường Đại Học Công Nghệ Thông Tin Và Truyền Thông Việt Hàn-Đại Học Đà Nẵng lời chào trân trọng, lời chúc sức khỏe và lời cảm ơn sâu sắc. Với sự quan tâm, dạy dỗ, chỉ bảo tận tình chu đáo của thầy cô, đến nay em đã có thể hoàn thành đồ án.

Đặc biệt em xin gửi lời cảm ơn chân thành nhất tới thầy giáo – Ths. Đỗ Công Đức đã quan tâm, giúp đỡ tận tình em hoàn thành tốt đồ án này trong thời gian qua.

Với điều kiện thời gian cũng như kinh nghiệm còn hạn chế, đồ án này không thể tránh được những thiếu sót. Em rất mong nhận được sự chỉ bảo, đóng góp ý kiến của các thầy cô để em có điều kiện bổ sung, nâng cao kiến thức của mình, phục vụ tốt hơn công tác thực tế sau này.

Em xin chân thành cảm ơn!

NHẬN XÉT CỦA GIÁO VIÊN HƯỚNG DẪN

MỤC LỤC

[LỜI CẢM ƠN i](#_Toc134352662)

[NHẬN XÉT CỦA GIÁO VIÊN HƯỚNG DẪN ii](#_Toc134352663)

[MỤC LỤC iii](#_Toc134352664)

[DANH MỤC HÌNH ẢNH v](#_Toc134352665)

[DANH MỤC BẢNG BIỂU vi](#_Toc134352666)

[DANH MỤC CÁC TỪ VIẾT TẮT vii](#_Toc134352667)

[MỞ ĐẦU 1](#_Toc134352668)

[1. Giới thiệu đề tài 1](#_Toc134352669)

[2. Mục tiêu của đề tài 2](#_Toc134352670)

[3. Phương pháp nghiên cứu 2](#_Toc134352671)

[4. Phạm vi nghiên cứu 2](#_Toc134352672)

[5. Nội dung và kế hoạch thực hiện 2](#_Toc134352673)

[6. Bố cục báo cáo 2](#_Toc134352674)

[CHƯƠNG 1 CƠ SỞ LÝ THUYẾT 4](#_Toc134352675)

[1.1 Tổng quan bài toán nhận diện khuôn mặt 4](#_Toc134352676)

[1.1.1 Bài toán nhận diện khuôn mặt 4](#_Toc134352677)

[1.1.2 Một số bài toán nhận dạng khuôn mặt 4](#_Toc134352678)

[1.1.3 Tổng quan kiến trúc 5](#_Toc134352679)

[1.2 Một số lĩnh vực ứng dụng nhận diện khuôn mặt 7](#_Toc134352680)

[1.3 Phương pháp nhận diện khuôn mặt 7](#_Toc134352681)

[1.3.1 Bài toán xác định mặt người: 7](#_Toc134352682)

[1.3.2 Hướng tiếp cận dựa trên tri thức (knowledge-based): 8](#_Toc134352683)

[1.3.3 Hướng tiếp cận dựa trên các đặc trưng không đổi (feature invariant) 9](#_Toc134352684)

[1.3.4 Hướng tiếp cận dựa trên phương pháp đối sánh mẫu (Template matching) 9](#_Toc134352685)

[1.3.5 Hướng tiếp cận dựa trên diện mạo (appearance-based) 10](#_Toc134352686)

[1.4 Hướng tiếp cận 11](#_Toc134352687)

[1.4.1 Mạng Nơ-ron nhân tạo 11](#_Toc134352688)

[1.4.2 Ứng dụng của mạng Nơ-ron trong bài toán xác định khuôn mặt 11](#_Toc134352689)

[1.4.3 Các bước xác định khuôn mặt dựa trên mạng Nơ-ron 12](#_Toc134352690)

[1.5 Giới thiệu mạng Neural 13](#_Toc134352691)

[1.5.1 Phân loại 14](#_Toc134352692)

[1.5.2 Cấu trúc mạng thần kinh tích chập (CNN ) 18](#_Toc134352693)

[1.6 Phát hiện khuôn mặt bằng MTCNN 21](#_Toc134352694)

[1.7 Nhận diện khuôn mặt bằng thuật toán FACENET 23](#_Toc134352695)

[CHƯƠNG 2 PHÂN TÍCH THIẾT KẾ 25](#_Toc134352696)

[2.1 Biểu đồ use case 25](#_Toc134352697)

[2.1.1 Đặc tả usecase 25](#_Toc134352698)

[2.1.2 Biểu đồ hoạt động 27](#_Toc134352699)

[2.2 Cơ sở dữ liệu 27](#_Toc134352700)

[2.2.1 SQL 27](#_Toc134352701)

[2.2.2 Dataset 28](#_Toc134352702)

[2.3 Ngôn ngữ và công cụ hỗ trợ 28](#_Toc134352703)

[2.3.1 Ngôn ngữ 28](#_Toc134352704)

[2.3.2 Công cụ 28](#_Toc134352705)

[CHƯƠNG 3 XÂY DỰNG ỨNG DỤNG 29](#_Toc134352706)

[KẾT LUẬN 30](#_Toc134352707)

[TÀI LIỆU THAM KHẢO 31](#_Toc134352708)

DANH MỤC HÌNH ẢNH

[Hình 1.1 Mô tả hệ thống nhận diện khuôn mặt 4](#_Toc134368174)

[Hình 1.2. Các bước chính trong hệ thống nhận dạng khuôn mặt 5](#_Toc134368175)

[Hình 1.3 Nhận dạng dựa trên tri thức 8](#_Toc134368176)

[Hình 1.4. Nhận diện dựa trên phương thức đối sánh mẫu 10](#_Toc134368177)

[Hình 1.5. Tập ảnh dùng để huấn luyện mặt người 11](#_Toc134368178)

[Hình 1.6. Cấu trúc mạng Neural 13](#_Toc134368179)

[Hình 1.7. Công thức toán học của mạng Neural K 14](#_Toc134368180)

[Hình 1.8. Hoạt động của ANN 15](#_Toc134368181)

[Hình 1.9. Phân loại hình ảnh 16](#_Toc134368182)

[Hình 1.10. RNN và ANN 17](#_Toc134368183)

[Hình 1.11. Cấu trúc mạng tích chập CNN 19](#_Toc134368184)

[Hình 1.12. Nguyên lý bộ lọc tích chập 19](#_Toc134368185)

[Hình 1.13. Kết quả xử lý sau lớp tích chập 20](#_Toc134368186)

[Hình 1.14. Xử lý Max Pooling 20](#_Toc134368187)

[Hình 1.15. Fully Connected Layer 21](#_Toc134368188)

[Hình 1.16. So sánh phương pháp MTCNN 22](#_Toc134368189)

[Hình 1.17. Nguyên lý hoạt động MTCNN 22](#_Toc134368190)

[Hình 1.18. FaceNet process 24](#_Toc134368191)

[Hình 2.1. Biểu đồ use-case 25](#_Toc134368192)

[Hình 2.2. Biểu đồ hoạt động 27](#_Toc134368193)

[Hình 2.3. Cơ sở dữ liệu hệ thống 28](#_Toc134368194)

DANH MỤC BẢNG

[Bảng 1.1 Bảng so sánh các phương pháp 2](#_Toc99389705)

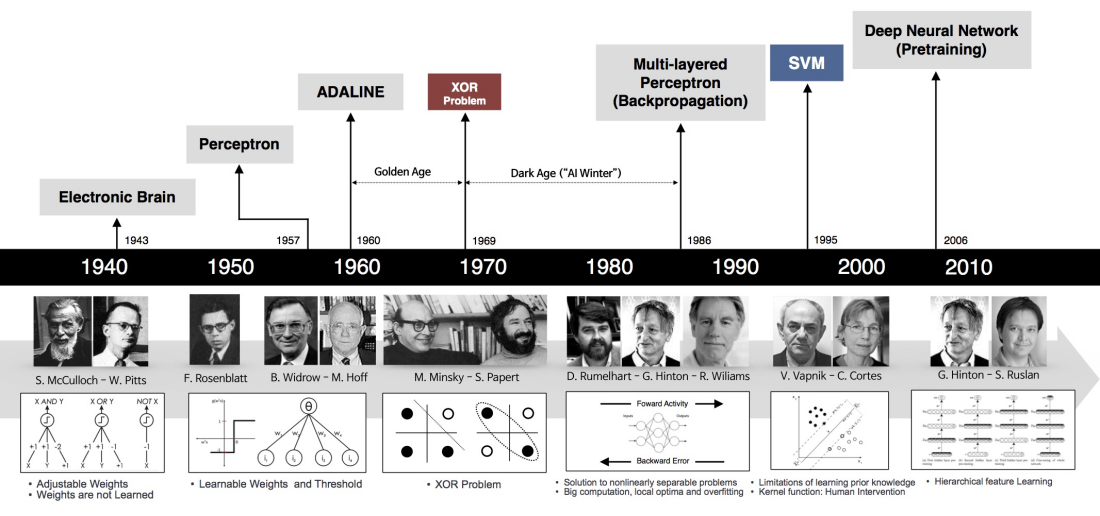
DANH MỤC CÁC TỪ VIẾT TẮT

|  |  |  |
| --- | --- | --- |
| **STT** | **Viết tắt** | **Cụm từ** |
| 1 | CNPM | Công nghệ phần mềm |
| 2 | CNTT | Công nghệ thông tin |
| 3 | 5E | Engagement, Exploration, Explanation, Elaboration, Evaluation |
| 4 | STEM | Science Technology Engineering Mathematics |
|  |  |  |

MỞ ĐẦU

##### Giới thiệu đề tài

Vào đầu những năm 1940, với sự xuất hiện và phát triển mạnh mẽ của thiết bị bán dẫn, linh kiện điện tử và máy tính đã đặt những nền móng đầu tiên cho sự xuất hiện của trí tuệ nhân tạo (AI). Tuy nhiên trí tuệ nhân tạo thời gian này vẫn chưa thật sự có ứng dụng thực tiễn hoặc thành tựu nổi bật nào. Khoảng thời gian từ năm 1960 đến 2000, giới chuyên gia và các nhà nghiên cứu đã phải trải qua hai mùa đông AI (AI Winter), do sự bế tắc và các gián đoạn nghiên cứu xảy ra trong khoảng thời gian này. Sự đột phá bắt đầu từ năm 2006, khi mà Hinton [1] giới thiệu ý tưởng về tiền huấn luyện không giám sát (unsupervised pre-training) thông qua deep belief nets (DBN). Điểm nổi bật trong bài báo này là đã tạo ra được một nơron nhân tạo với nhiều lớp ẩn (hidden layer) thay vì chỉ một lớp như trước đây. Từ thời gian này, neural networks với nhiều lớp ẩn được gọi với cái tên là Deep Learning.



Những cột mốc quan trọng của Deep Learning

Sự phát triển không ngừng của lĩnh vực trí tuệ nhân tạo, máy học và học sâu cũng chỉ để nhằm vào mục đích duy nhất, đó là phục vụ cho nhu cầu lợi ích và cuộc 15 sống của con người. Như giúp thiết bị điện tử nhận diện trắc sinh học (vân tay, quét võng mạc, nhận diện khuôn mặt, …), giúp dự đoán thời tiết, chẩn đoán các loại bệnh hay dịch các ngôn ngữ khác nhau. Để phục vụ những công việc phức tạp, muôn hình vạn trạng của con người, trí tuệ nhân tạo chia ra làm các lĩnh vực để chuyên biệt hóa như Hệ chuyên gia, Cây ngữ nghĩa, Xử lý ngôn ngữ tự nhiên, Robotics, Quy hoạch, Thị giác máy tính,… Tuy nhiên, việc nhận diện gương mặt là một vấn đề không hề đơn giản. Gương mặt mỗi người đều có những đặc trưng riêng biệt, điều này đặt ra các thách thức không nhỏ cho việc huấn luyện mô hình nhận diện. Do đó, công việc này phải thực hiện bởi phương pháp học sâu (deep learning), với dữ liệu đầu vào là ảnh gương mặt, được đưa vào mạng CNN đã được huấn luyện lại bằng các mô hình ResNet, DenseNet và các biến thể, cuối cùng là so sánh các mô hình với nhau và chọn ra mô hình huấn luyện có độ chính xác cao nhất.

Hiện nay việc nhận diện gương mặt đã trở nên cần thiết đối với nhiều ngành nghề, lĩnh vực khác nhau. Việc xây dựng một hệ thống gương mặt có thể đáp ứng được mọi nhu cầu sử dụng vẫn đang là một vấn đề nan giải mặt dù đã tồn tại rất nhiều mô hình nhận diện gương mặt. Để có thể giải bài toán này, ta cần phải thiết kế hệ thống có khả năng linh hoạt cao cũng như phạm vi ứng dụng lớn để người dùng có thể truy cập và sử dụng ở mọi lúc, mọi nơi.

Thị giác máy tính (computer vision) được áp dụng rộng rãi vào trong đời sống của con người. Một trong những ứng dụng phổ biến nhất của thị giác máy tính đó chính là nhận diện gương mặt

##### Mục tiêu của đề tài

* Tìm hiểu và xây dựng được một hệ thống điểm danh bằng nhận diện khuôn mặt.
* Sử dụng ngôn ngữ lập trình Python và các thư viện hỗ trợ để xây dựng ứng dụng.
* Áp dụng, củng cố các kiến thức đã được học.

##### Phương pháp nghiên cứu

##### Phạm vi nghiên cứu

##### Nội dung và kế hoạch thực hiện

* Tìm hiểu cơ sở lý thuyết môn thị giác máy tính.
* Tìm hiểu về Deep Learning.
* Phân tích và thiết kế hệ thống.
* Tìm hiểu ngôn ngữ lập trình Python.
* Xây dựng ứng dụng.

##### Bố cục báo cáo

Sau phần *Mở đầu*, báo cáo được trình bày trong ba chương, cụ thể như sau:

Chương 1. *Tổng quan về ….* Trong chương này, báo cáo trình bày các khái niệm, đặc điểm,

Chương 2. *Phân tích thiết kế …* . Nội dung chương bao gồm các

Chương 3. *Xây dựng …* Chương này đề xuất…

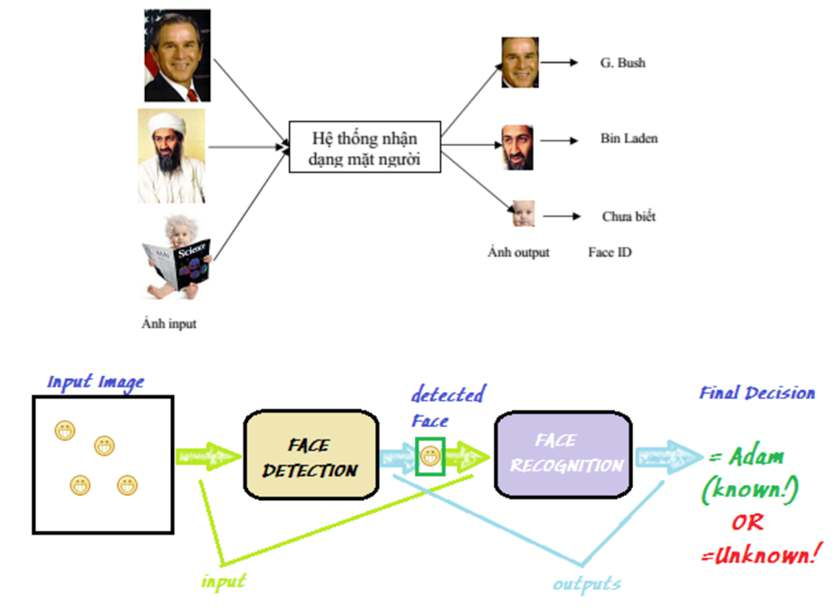
Cuối cùng là *Kết luận*, *Tài liệu tham khảo* và *Phụ lục* liên quan đến đề tài.

# CƠ SỞ LÝ THUYẾT

## Tổng quan bài toán nhận diện khuôn mặt

### Bài toán nhận diện khuôn mặt

Hệ thống nhận dạng mặt người là một hệ thống nhận vào là một ảnh hoặc một đoạn video (một dòng các hình ảnh liên tục). Qua xử lý, tính toán hệ thống xác định được vị trí mặt người (nếu có) trong ảnh và xác định là người nào trong số những người mà hệ thống đã được biết (qua quá trình học) hoặc là người lạ.



Hình . Mô tả hệ thống nhận diện khuôn mặt

### Một số bài toán nhận dạng khuôn mặt

Bảng . Bảng so sánh các phương pháp

|  |  |  |
| --- | --- | --- |
| Xác thực khuôn mặt (face verification) | Xác thực khuôn mặt (face verification) | Tìm kiếm khuôn mặt đại diện (face clustering) |
| Đây là bài toán match oneone. Bài toán này trả lời cho câu hỏi “có phải 2 ảnh đầu vào là cùng một người không?” Kết quả đầu ra sẽ là yes hoặc no. Bài toán thường được dùng trong các hệ thống bảo mật. Xác thực khuôn mặt trên điện thoại là một bài toán như vậy. | Đây là bài toán match oneone. Bài toán này trả lời cho câu hỏi “có phải 2 ảnh đầu vào là cùng một người không?” Kết quả đầu ra sẽ là yes hoặc no. Bài toán thường được dùng trong các hệ thống bảo mật. Xác thực khuôn mặt trên điện thoại là một bài toán như vậy. | Đơn giản chỉ cần tính ra trung bình của các ảnh khuôn mặt để thu được centroid image. Tính điểm tương đồng giữa trọng tâm với toàn bộ khuôn mặt còn lại để thu được khuôn mặt đặc trưng nhất giống nhất với gốc. Tương tự như vậy cũng có thể tìm ra khuôn mặt đặc trưng nhất của nam, nữ các quốc gia. |

### Tổng quan kiến trúc

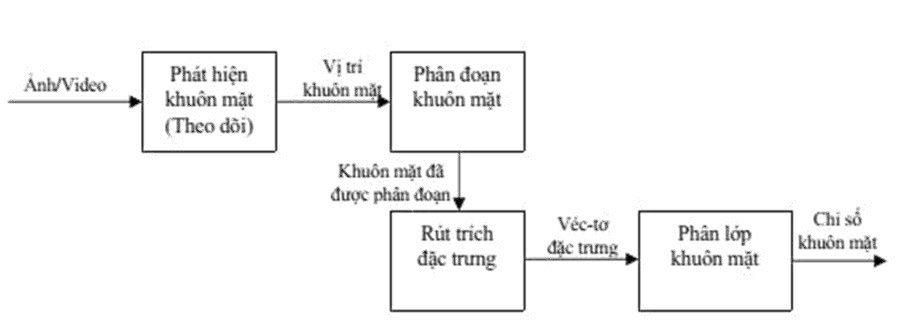
Một hệ thống nhận dạng mặt người thông thường bao gồm bốn bước xử lý sau:

- Phát hiện khuôn mặt (Face Detection).

- Phân đoạn khuôn mặt (Face Alignment hay Segmentation).

- Trích chọn đặc trưng (Feature Extraction).

- Nhận dạng (Recognition) hay Phân lớp khuôn mặt (Face Classification).



Hình .. Các bước chính trong hệ thống nhận dạng khuôn mặt

Phát hiện khuôn mặt dò tìm, định vị những vùng (vị trí) có thể là khuôn mặt xuất hiện trong ảnh hoặc các frame video. Các vùng này sẽ được tách riêng để xử lý. Phân đoạn khuôn mặt sẽ xác định vị trí mắt mũi, miệng và các thành phần khác của khuôn mặt và chuyển kết quả này cho bước trích chọn đặc trưng. Ở bước trích chọn đặc trưng, bằng một phương pháp trích chọn đặc điểm nào đó (mẫu nhị phân cục bộ-Local Binary Pattern-LBP, Gabor wavelets…) sẽ được sử dụng với ảnh mặt để trích xuất các thông tin đặc trưng cho ảnh từ các thông tin về các thành phần trên khuôn mặt, kết quả là mỗi ảnh sẽ được biểu diễn dưới dạng một vector đặc trưng (feature vector).

Những vectơ đặc trưng này sẽ là dữ liệu đầu vào cho một mô hình đã được huấn luyện trước để nhận dạng khuôn mặt (Face Recognition) hay phân lớp khuôn mặt (Face Lớpifition), tức là xác định danh tính (identity) hay nhãn của ảnh-đó là ảnh của ai. Ở bước nhận dạng khuôn mặt (Face Recognition), thường thì phương pháp k-láng giềng gần (k-nearest neighbor: kNN) sẽ được sử dụng. Bên cạnh những bước chính nêu trên, chúng ta còn có thể áp dụng thêm một số bước khác như tiền xử lý, hậu xử lý nhằm làm tăng độ chính xác cho hệ thống. Ví dụ, sau bước phát hiện khuôn mặt, ta có thể thực hiện bước tiền xử lý (Preprocessing) bao gồm các bước căn chỉnh ảnh (face image alignment) và chuẩn hóa ánh sáng (illumination normalization).

Do một số thông số như: tư thế khuôn mặt, độ sáng, điều kiện ánh sáng, v.v… phát hiện khuôn mặt được đánh giá là bước khó khăn và quan trọng nhất so với các bước còn lại của hệ thống. Tuy nhiên, trong phạm vi đồ án này, không tập trung tìm hiểu bước phát hiện khuôn mặt mà chỉ tập trung chủ yếu vào bước nhận dạng khuôn mặt.

Dữ liệu cho một hệ thống nhận dạng mặt được chia làm 3 tập: tập huấn luyện (training set), tập tham chiếu (reference set haygallery set) và tập để nhận dạng (probe set hay query set, đôi khi còn gọi là test set). Trong nhiều hệ thống, tập training trùng với tập reference. Tập training gồm các ảnh được dùng để huấn luyện (hay học-learning), thông thường tập này được dùng để sinh ra một không gian con (projection subspace) là một ma trận và phương pháp hay được sử dụng là PCA (Principal Component Analysis), WPCA (Whitened PCA), LDA (Linear Discriminant Analysis),

KPCA (Kernel PCA). Tập reference gồm các ảnh đã biết danh tính được chiếu (projected) vào không gian con ở bước training. Bước training nhằm 2 mục đích: giảm số chiều (dimension reduction) của các vector đặc điểm (feature vector) vì các vector này thường có độ dài khá lớn (vài nghìn tới vài trăm nghìn) nên nếu để nguyên thì việc tính toán sẽ rất rất lâu, thứ hai là làm tăng tính phân biệt (discriminative) giữa các ảnh khác lớp (định danh khác nhau), ngoài ra có thể làm giảm tính phân biệt giữa các ảnh thuộc về một lớp (tùy theo phương pháp, ví dụ như Linear Discriminant Analysis LDA- còn gọi là Fisher Linear Discriminant Analysis-Fisherface là một phương pháp làm việc với tập training mà mỗi đối tượng có nhiều ảnh mặt ở các điều kiện khác nhau). Sau khi thực hiện chiếu tập reference vào không gian con, hệ thống lưu lại kết quả là một ma trận với mỗi cột của ma trận là một vectơ tương ứng với ảnh (định danh đã biết) để thực hiện nhận dạng (hay phân lớp). Nhận dạng (hay phân lớp) được thực hiện với tập các ảnh probe, sau khi tiền xử lý xong, mỗi ảnh sẽ được áp dụng phương pháp trích chọn đặc điểm (như với các ảnh thuộc tập training và reference) và được chiếu vào không gian con. Tiếp đến việc phân lớp sẽ dựa trên phương pháp k-NN, định danh của một ảnh cần xác định sẽ được gán là định danh của ảnh có khoảng cách (distance) gần với nó nhất. Ở đây cần lưu ý là mỗi ảnh là một vector nên có thể dùng khái niệm hàm khoảng cách giữa hai vector để đo sự khác biệt giữa các ảnh.

## Một số lĩnh vực ứng dụng nhận diện khuôn mặt

Xác định khuôn mặt đã được ứng dụng nhiều trong các lĩnh vực:

* Hệ thống tương tác giữa người và máy (điều kiển máy tính qua các cử động của khuôn mặt).
* Hệ thống nhận dạng người (giúp cho các cơ quan an ninh quản lý tốt con người).
* Hệ thống quan sát, bảo vệ, quản lý việc ra vào cho các cơ quan và công ty.
* Hệ thống trợ giúp lái xe, hệ thống phân tích cảm xúc trên khuôn mặt, và hệ thống nhận dạng khuôn mặt cho các máy ảnh kỹ thuật số…

## Phương pháp nhận diện khuôn mặt

### Bài toán xác định mặt người:

Nhận dạng khuôn mặt người (Face Detection) là một kỹ thuật máy tính để xác định các vị trí và các kích thước của các khuôn mặt người trong các ảnh bất kỳ (ảnh kỹ thuật số) nếu có. Kỹ thuật này nhận biết các đặc trưng của khuôn mặt và bỏ qua những thứ khác, như: tòa nhà, cây cối, cơ thể.

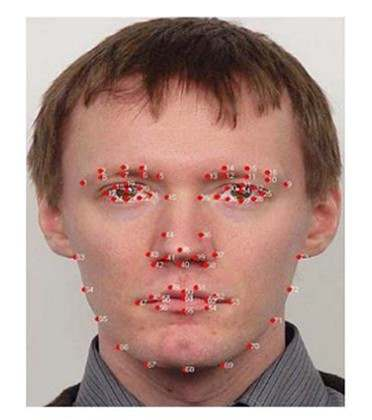
Có nhiều nghiên cứu tìm phương pháp xác định khuôn mặt người, từ ảnh xám đến ảnh màu. Dựa vào tính chất của các phương pháp xác định mặt người trên ảnh, các phương pháp này được chia thành bốn loại chính, tương ứng với bốn hướng tiếp cận khác nhau.

* Hướng tiếp cận dựa trên tri thức (knowledge-based).
* Hướng tiếp cận dựa trên các đặc trưng không đổi (feature invariant).
* Hướng tiếp cận dựa trên phương pháp đối sánh mẫu (Template matching).
* Hướng tiếp cận dựa trên diện mạo (appearance-based).

Ngoài ra cũng có rất nhiều nghiên cứu mà phương pháp xác định mặt người không chỉ dựa vào một hướng mà có liên quan đến nhiều hướng. Trong phạm vi đồ án môn học này, chỉ giới thiệu tổng quan các hướng tiếp cận để xác định khuôn mặt người trong ảnh.

### Hướng tiếp cận dựa trên tri thức (knowledge-based):

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.



Hình 1.3 Nhận dạng dựa trên tri thức

Một vấn đề khá phức tạp khi dùng hướng tiếp cận này là làm sao chuyển từ tri thức con người sang các luật một cách hiệu quả. Nếu các luật này quá chi tiết thì khi xác định có thể xác định thiếu các khuôn mặt có trong ảnh, vì những khuôn mặt này không thể thỏa mãn tất cả các luật đưa ra. Nhưng các luật tổng quát quá thì có thể chúng ta sẽ xác định nhầm một vùng nào đó trên thực tế không phải là khuôn mặt nhưng lại xác định là khuôn mặt. Và cũng khó khăn mở rộng yêu cầu từ bài toán để xác định các khuôn mặt có nhiều tư thế khác nhau.

### Hướng tiếp cận dựa trên các đặc trưng không đổi (feature invariant)

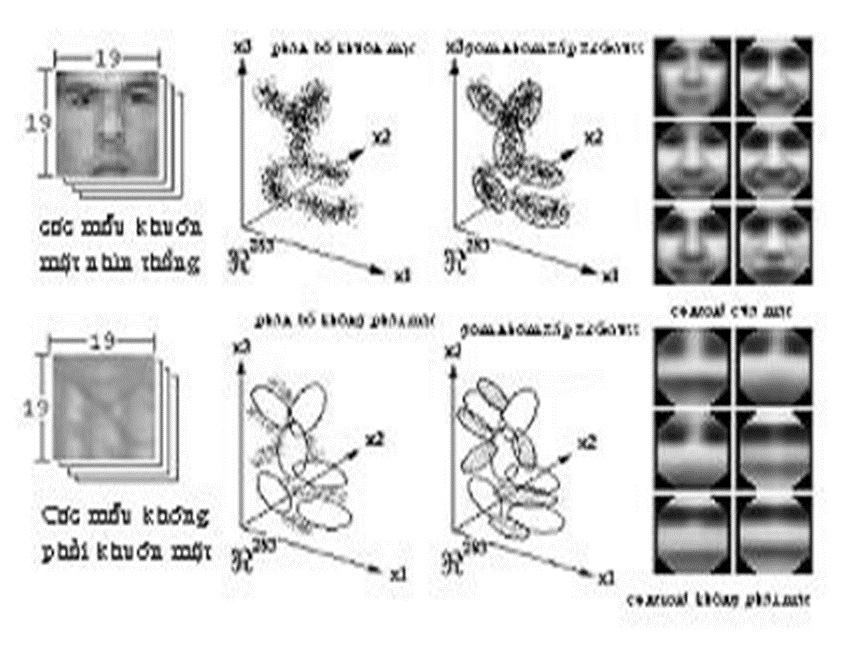
Đây là hướng tiếp cập theo kiểu bottom-up. Các tác giả cố gắng tìm các đặc trưng không thay đổi của khuôn mặt người để xác định khuôn mặt người. Dựa trên nhận xét thực tế, con người dễ dàng nhận biết các khuôn mặt và các đối tượng trong các tư thế khác nhau và điều kiện ánh sáng khác nhau, thì phải tồn 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 thuật toán theo hướng tiếp cận này đó là cần phải điều chỉnh cho phù hợp điều kiện ánh sáng, nhiễu, và bị che khuất. Đôi khi bóng của một khuôn mặt sẽ tạo thêm cạnh mới, mà cạnh này lại rõ ràng hơn cạnh thật sự của khuôn mặt, gây ra sự nhầm lẫn khi xác định các khuôn mặt.

### Hướng tiếp cận dựa trên phương pháp đối sánh mẫu (Template matching)

Trong phương pháp đối sánh 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 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.

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. I.Craw 1992 đã áp dụng một mẫu cứng trong khi A. Lanitis 1995 sử dụng một mẫu có thể biến dạng trong bước phát hiện khuôn mặt.



Hình .. Nhận diện dựa trên phương thức đối sánh mẫu

### Hướng tiếp cận dựa trên diện mạo (appearance-based)

Trái ngược hẳn với hướng tiếp cận dựa trên đối sánh mẫu, các mô hình (hay các mẫu) được học từ một tập ảnh huấn luyện trước đó. Sau đó hệ thống (mô hình) sẽ xác định khuôn mặt người. Do phương pháp này thường dùng một mô hình máy học nên còn được gọi là phương pháp dựa trên máy học (machine learning-based).



Hình .. Tập ảnh dùng để huấn luyện mặt người

Sau đó hệ thống sẽ tổng hợp tất cả các đặc tinh của khuôn mặt con người như: mắt, mũi, miệng ... thành một vector riêng. Có nhiều mô hình máy học được áp dụng trong hướng tiếp cận này:

Eigenface (M. Turk và A. Pentland 1991), Mô hình dựa trên phân phối (K. Sung and T. Poggio 1998). Mojang Neural (H. Rowley 1998), Support Vector Machine (E. Osuna et al 1997). Phân lớp Bayes (H. Schneiderman và T. Kanade 1998). Mô hình Markov ẩn (A. Rajagopalan et al 1998) và các mô hình tăng cường (Adaboost của P. Viola và M. Jones 2001; FloatBoost do Stan Z. Li và Zhen Qiu Zhang 2004).

Trên đây là giới thiệu tổng quan bốn hướng tiếp cận để xác định khuôn mặt người trong ảnh và trong đồ án này, hướng tiếp cận dựa trên diện mạo (appearance-based) được lựa chọn để thực hiện phát hiện khuôn mặt người trong ảnh.

## Hướng tiếp cận

### Mạng Nơ-ron nhân tạo

Là một hệ thống bao gồm nhiều phần tử xử lý đơn giản(hay còn gọi là nơ-ron) hoạt động song song và được nối với nhau bằng các liên kết nơ-ron. Mỗi liên kết kèm theo một trọng số nào đó, đặc trưng cho tính kích hoạt hay ức chế của các nơron.

### Ứng dụng của mạng Nơ-ron trong bài toán xác định khuôn mặt

Mạng Nơ-ron được áp dụng rộng rãi và tương đối thành công trong các bài toán nhận dạng mẫu như: Nhận diện kí tự (nhận diện chữ viết), đối tượng (object), điều khiển robot …Một trong những ưu điểm của mạng nơ-ron là tính khả thi của hệ thống học là cơ sở dữ liệu mẫu đa dạng và phức tạp vì vậy khả năng phát hiện khuôn mặt rất lớn.Tuy nhiên điều trở ngại là các kiến trúc mạng đều quá lớn ,vì vậy khi áp dụng vào trường hợp cụ thể phải xác định rõ ràng số lượng tầng ,số lượng node , tỉ lệ … cho từng trường hợp.

Theo đánh giá của các chuyên gia thì phương pháp xác định mặt người bằng mang nơ-ron của Rowel đối với ảnh xám là tốt nhất. Một mạng đa tầng được dùng để học các mẫu khuôn mặt và không phải là khuôn mặt từ các ảnh tương ứng (dựa trên quan hệ về cường độ và mặt không gian của các điểm ảnh). Đây cũng là phương pháp mạng nơ-ron mà em sẽ trình bày cụ thể trong các chương sau.

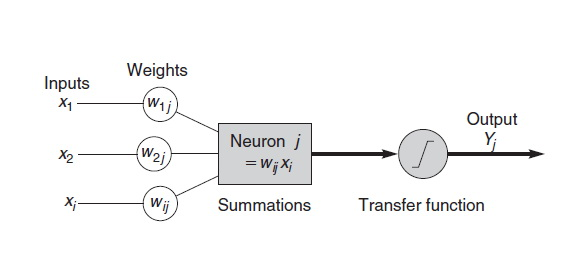
### Các bước xác định khuôn mặt dựa trên mạng Nơ-ron

Hệ thống xác định khuôn mặt thực hiện các bước chính sau:

* Ước lượng vị trí: việc dùng tiếp cận máy học, cụ thể là mạng Nơ-ron, đòi hỏi việc huấn luyện mẫu. Để giảm số lượng biến đổi trong ảnh huấn luyện dương, ảnh được canh biên với các ảnh khác để cực tiểu hoá các biến đổi vị trí đặc trưng khuôn mặt. Khi thi hành chương trình, ta không biết chính xác các vị trí đặc trưng khuôn mặt, do đó không thể dùng chúng để định vị các ứng viên khuôn mặt tiềm năng. Thay vậy, ta dò tìm toàn diện ở mọi vị trí và tỷ lệ để tìm mọi vị trí ứng viên. Các cải tiến dò tìm toàn diện làm cho thuật toán nhanh hơn, với tỷ lệ dò tìm giảm 10% đến 30%.
* Tiền xử lý: để giảm các biến đổi gây ra do chiếu sáng hay camera, ảnh được tiền xử lý với các thuật toán chuẩn như cân bằng lược đồ để cải thiện độ sáng và độ tương phản trong ảnh. Sau đó ta tiến hành phân mảnh ảnh và rút trích các đặc trưng quan trọng của khuôn mặt, đồng thời tiến hành phân tích độ nhạy để lọc bỏ các thông tin dư thừa.
* Dò tìm: các khuôn mặt tiềm năng đã chuẩn hoá về vị trí, tư thế, và độ sáng trong hai bước đầu tiên được khảo sát để xác định chúng có thực sự là khuôn mặt hay không. Quyết định này được thực hiện bằng mạng Nơ-ron đã huấn luyện với nhiều ảnh mẫu khuôn mặt và không khuôn mặt.
* Quyết định: Kết hợp nhiều mạng để có được một quyết định khách quan nhất. Mỗi mạng học những điều khác nhau từ dữ liệu huấn luyện, và đưa ra các lỗi khác nhau. Các quyết định của chúng có thể kết hợp dùng một số phương thức đơn giản khác, làm tăng độ chính xác dò tìm khuôn mặt và ngăn chặn lỗi.

## Giới thiệu mạng Neural

Mạng Neural là một mô hình xử lý thông tin mô phỏng theo cách thức xử lý thông tin của các hệ Neural sinh học của sinh vật. Mạng Neural được xây dựng nên từ một số lượng lớn các phần tử (Neural) kết nối với nhau thông qua các liên kết (trọng số liên kết) làm việc như một thể thống nhất để giải quyết một vấn đề cụ thể nào đó. Mô hình mạng Neural giống như bộ não con người, được học bởi kinh nghiệm (thông qua huấn luyện), có khả năng lưu giữ những kinh nghiệm hiểu biết (tri thức) và sử dụng những tri thức đó trong việc dự đoán các dữ liệu chưa biết.



Hình .. Cấu trúc mạng Neural

Các thành phần cơ bản của một Neural nhân tạo bao gồm:

- Input : Các tín hiệu đầu vào của Neural thể hiện thuộc tính của dữ liệu, các tín hiệu này thường được đưa vào dưới dạng một vector.

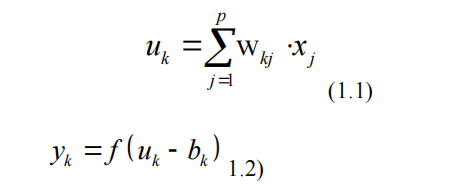
- Trọng số liên kết (Connection Weights): Mỗi liên kết được thể hiện bởi một trọng số liên kết. Trọng số liên kết giữa tín hiệu vào thứ j với Neural k thường được kí hiệu là wkj. Thông thường, các trọng số này được khởi tạo một cách ngẫu nhiên ở thời điểm khởi tạo mạng và được cập nhật liên tục trong quá trình học mạng.

- Hàm tổng (Summation) : Tính tổng trọng số của tất cả các tín hiệu đầu vào được đưa vào mỗi Neural.

- Hàm truyền (Transfer function): Hàm này được dùng để giới hạn phạm vi đầu ra của mỗi Neural. Nó nhận đầu vào là kết quả của hàm tổng và ngưỡng.

- Output: Là tín hiệu đầu ra của một Neural, với mỗi Neural sẽ có tối đa là một đầu ra.

Xét về mặt toán học, cấu trúc của một Neural k, được mô tả bằng cặp biểu thức sau:



Hình .. Công thức toán học của mạng Neural K

Trong đó:

- x1, x2, ..., xp: là các tín hiệu vào.

- (wk1, wk2,..., wkp) là các trọng số liên kết của Neural thứ k.

- uk là hàm tổng.

- bk là độ lệch trong quá trình truyền.

- f là hàm truyền.

- yk là tín hiệu đầu ra của Neural.

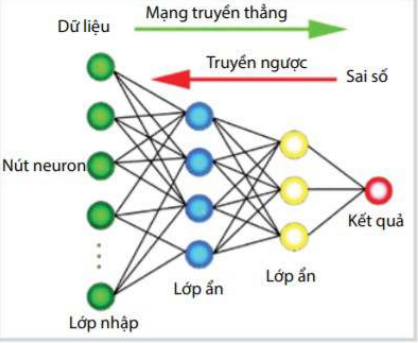
Như vậy Neural nhân tạo nhận các tín hiệu đầu vào, xử lý (nhân các tín hiệu này với trọng số liên kết, tính tổng các tích thu được rồi gửi kết quả tới hàm truyền), và cho một tín hiệu đầu ra ( là kết quả của hàm truyền).

### Phân loại

a) Mạng thần kinh nhân tạo (ANN)

ANN ra đời xuất phát từ ý tưởng mô phỏng bộ não con người. Mạng thần kinh nhân tạo (ANN), là một nhóm gồm nhiều perceptron hoặc neural ở mỗi lớp. ANN còn được gọi là mạng Neural Feed-Forward vì các đầu vào chỉ được xử lí theo hướng chuyển tiếp.

Mạng thần kinh nhân tạo là một trong những biến thể đơn giản nhất của mạng neural. Chúng truyền thông tin theo một hướng, qua các nút đầu vào khác nhau, cho đến khi nó đến được nút đầu ra.



Hình .. Hoạt động của ANN

ANN bao gồm 3 lớp - Đầu vào, Ẩn và Đầu ra. Lớp đầu vào chấp nhận các tín hiệu đầu vào, lớp ẩn xử lý các tín hiệu đầu vào và lớp đầu ra tạo ra kết quả. Về cơ bản, mỗi lớp cố gắng tìm hiểu các trọng lượng nhất định.

ANN có thể được sử dụng để giải quyết các vấn đề liên quan đến:

- Dữ liệu dạng bảng

-Dữ liệu hình ảnh

-Dữ liệu văn bản

Ưu điểm:

- Mạng nơron nhân tạo có khả năng học bất kỳ hàm phi tuyến nào. Do đó, những mạng này thường được biết đến với cái tên Phương pháp xấp xỉ hàm phổ quát . ANN có khả năng tìm hiểu các trọng số ánh xạ bất kỳ đầu vào nào với đầu ra.

- Lưu trữ thông tin trên toàn bộ mạng.

- Khả năng làm việc với kiến thức chưa hoàn thiện.

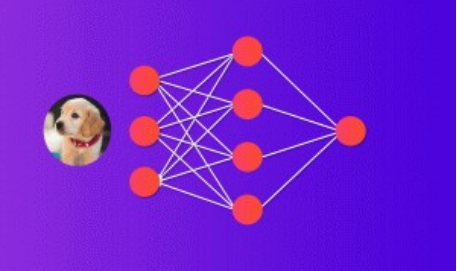
- Có khả năng chịu lỗi.

- Có bộ nhớ phân tán.

Nhược điểm:

Với bài toán phân loại ảnh bằng ANN, bước đầu tiên là chuyển ảnh 2 chiều thành vectơ 1 chiều trước khi huấn luyện mô hình. Điều này có ba nhược điểm:

- Số lượng các thông số có thể huấn luyện tăng lên đáng kể với sự gia tăng kích thước của hình ảnh.



Hình .. Phân loại hình ảnh

Trong trường hợp trên, nếu kích thước của hình ảnh là 224 x 224, thì số lượng tham số có thể tra được ở lớp ẩn đầu tiên chỉ với 4 nơ-ron là 602.112.

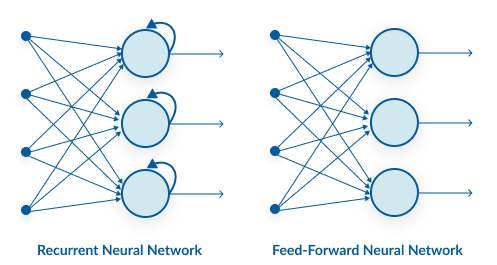
- ANN làm mất các tính năng không gian của hình ảnh. Các tính năng không gian đề cập đến sự sắp xếp của các pixel trong một hình ảnh.

- ANN không thể nắm bắt thông tin tuần tự trong dữ liệu đầu vào được yêu cầu để xử lý dữ liệu trình tự.

b) Mạng thần kinh tái tạo (RNN)

Sự khác biệt giữa RNN và ANN từ quan điểm kiến trúc:

- Một ràng buộc lặp trên lớp ẩn của ANN chuyển thành RNN.



Hình .. RNN và ANN

Mạng thần kinh tái tạo (RNN) có một kết nối lặp lại ở trạng thái ẩn. Ràng buộc lặp này đảm bảo rằng thông tin tuần tự được ghi lại trong dữ liệu đầu vào. RNN lưu đầu ra của các nút xử lý và đưa kết quả trở lại mô hình (không chuyển thông tin chỉ theo một hướng). Đây là cách mô hình được cho là học cách dự đoán kết quả của một lớp. Mỗi nút trong mô hình RNN hoạt động như một ô nhớ, tiếp tục tính toán và thực hiện các hoạt động. Nếu dự đoán của mạng không chính xác, thì hệ thống sẽ tự học và tiếp tục làm việc để hướng tới dự đoán chính xác trong quá trình nhân giống ngược.

Chúng ta có thể sử dụng mạng thần kinh tái tạo để giải quyết các vấn đề liên quan đến:

- Dữ liệu chuỗi thời gian

- Dữ liệu văn bản

- Dữ liệu âm thanh

Ưu điểm:

- RNN ghi nhớ từng thông tin theo thời gian. RNN chỉ hữu ích trong dự đoán chuỗi thời gian vì tính năng ghi nhớ các đầu vào trước đó. Đây được gọi là trí nhớ ngắn hạn dài hạn.

- RNN còn được sử dụng với các lớp phức hợp để mở rộng vùng lân cận pixel hiệu quả.

Nhược điểm:

- Các RNN sâu (RNN với một số lượng lớn các bước thời gian) cũng gặp phải vấn đề Gradient biến mất và các vấn đề bùng nổ.

- Huấn luyện một RNN là một nhiệm vụ rất khó khăn.

- RNN không thể xử lý các chuỗi rất dài nếu sử dụng tanh hoặc relu làm hàm kích hoạt.

c) Mạng thần kinh tích chập (Convolutional Neural Network - CNN)

Mạng thần kinh tích chập (CNN) là một trong những mô hình phổ biến nhất được sử dụng hiện nay. Mạng thần kinh tích chập có kiến trúc khác với Mạng thần kinh thông thường. Mạng thần kinh bình thường chuyển đổi đầu vào thông qua hàng loạt các tầng ẩn. Mỗi tầng là một tập các nơ-ron và các tầng được liên kết đầy đủ với các nơ-ron ở tầng trước đó. Và ở tầng cuối cùng sẽ là tầng kết quả đại diện cho dự đoán của mạng. Mạng thần kinh tích chập được chia thành 3 chiều: rộng, cao, và sâu. Các nơ-ron trong mạng không liên kết hoàn toàn với toàn bộ nơ-ron tầng trước chỉ liên kết tới một vùng nhỏ. Cuối cùng, một tầng đầu ra được tối giản thành véc-tơ của giá trị xác suất.

CNN gồm hai thành phần:

- Phần tầng ẩn hay phần rút trích đặc trưng: trong phần này, mạng sẽ tiến hành tính toán hàng loạt phép tích chập và phép hợp nhất (pooling) để phát hiện các đặc trưng. Ví dụ: nếu ta có hình ảnh con ngựa vằn, thì trong phần này mạng sẽ nhận diện các sọc vằn, hai tai, và bốn chân của nó.

- Phần phân lớp: tại phần này, một lớp với các liên kết đầy đủ sẽ đóng vai trò như một bộ phân lớp các đặc trưng đã rút trích được trước đó. Tầng này sẽ đưa ra xác suất của một đối tượng trong hình.

Ưu điểm:

- Độ chính xác rất cao trong các vấn đề nhận dạng hình ảnh.

- Tự động phát hiện các tính năng quan trọng mà không cần bất kỳ sự giám sát của con người.

- Chia sẻ trọng lượng.

- Dễ thiết kế mô hình nhận dạng nhanh

Nhược điểm:

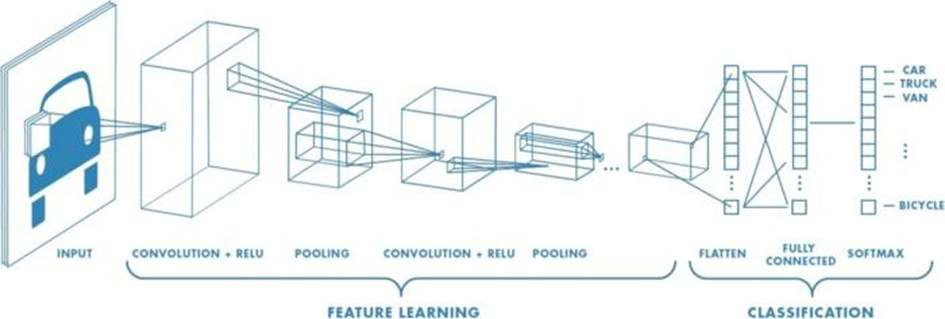
- CNN không mã hóa vị trí và hướng của đối tượng.

- Thiếu khả năng bất biến trong không gian đối với dữ liệu đầu vào.

Qua hoạt động của 3 loại mô hình mạng neraul ta thấy được các ưu nhược điểm của từng loại mô hình. Với mỗi loại có mỗi đặc điểm riêng và trong đó nổi trội về mảng nhận diện hình ảnh là mô hình CNN với khả năng dễ thiết kế nhận dạng nhanh, độ chính xác cao là lựa chọn phù hợp cho bài toán nhận diện mà chúng ta cần giải quyết. Để thấy rõ hơn về sự hoạt động của CNN ta đi phân tích chi tiết hơn về cấu trúc và hoạt động để thấy rõ sự tối ưu của mô hình mạng này.

### Cấu trúc mạng thần kinh tích chập (CNN )

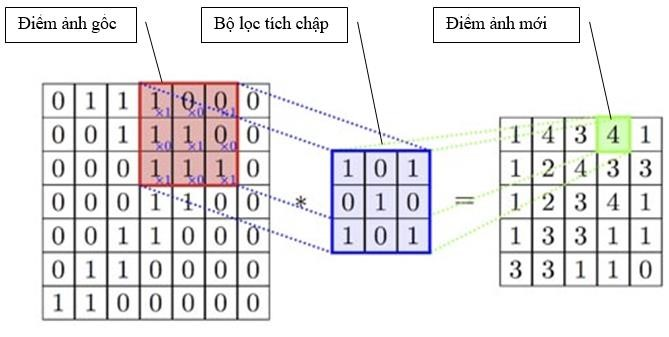
Convolutional Neural Network (CNN) là một trong những thuật toán học sâu chính để nhận dạng vật thể, nhận dạng hình ảnh,… Để nhận diện được vật thể, mô hình CNN sẽ học dữ liệu và kiểm tra liên tục để đạt được kết quả nhận diện tốt nhất. Mỗi ảnh input đầu vào sẽ được máy tính coi như 1 mảng các pixel phụ thuộc vào độ phân giải của chúng và cũng tùy vào ảnh màu hay ảnh đen trắng mà số lớp của ảnh cũng khác nhau, cách xử lý cũng khác nhau. Quá trình CNN xử lý và học dữ liệu sẽ bao gồm 1 số lớp như sau: Lớp tích chập (Convolutional Layer), Lớp kích hoạt phi tuyến ReLU (Rectified Linear Unit), Lớp lấy mẫu (Pooling Layer) và Lớp kết nối đầy đủ (Fully-connected Layer), 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.



Hình .. Cấu trúc mạng tích chập CNN

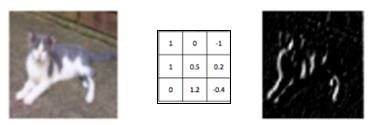
a) Lớp tích chập (Convolutional layer)

Đây là thành phần quan trọng nhất trong mạng CNN, cũng là nơi thể hiện tư tưởng xây dựng sự liên kết cục bộ thay vì kết nối toàn bộ các điểm ảnh. Các liên kết cục bộ này 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 .. Nguyên lý bộ lọc tích chập

Bộ lọc được sử dụng là một ma trận có kích thước 3x3. Bộ lọc này được 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.

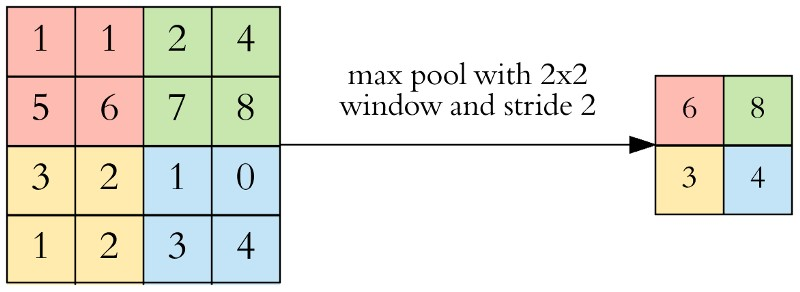


Hình .. Kết quả xử lý sau lớp tích chập

b) Lớp lấy mẫu( Pooling Layer)

Mục đích của pooling rất đơn giản, nó làm giảm số hyperparameter mà ta cần phải tính toán, từ đó giảm thời gian tính toán, tránh overfitting. Loại pooling ta thường gặp nhất là max pooling, lấy giá trị lớn nhất trong một pooling window.

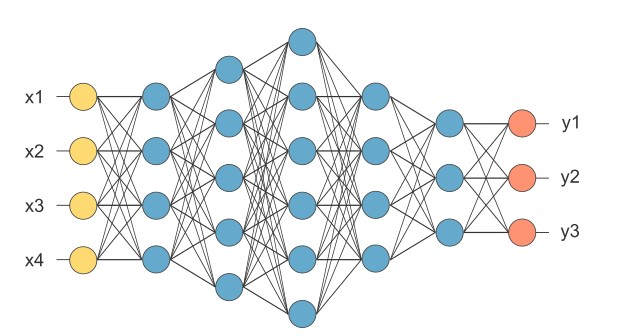
Pooling hoạt động gần giống với convolution, nó cũng có 1 cửa sổ trượt gọi là pooling window, cửa sổ này trượt qua từng giá trị của ma trận dữ liệu đầu vào (thường là các feature map trong convolutional layer), chọn ra một giá trị từ các giá trị nằm trong cửa sổ trượt (với max pooling ta sẽ lấy giá trị lớn nhất).



Hình .. Xử lý Max Pooling

c) Lớp kết nối đầy đủ( Fully Connected module)

Trong mạng Neural, các Neural input sẽ được kết nối với các Neural khác trong mạng với 1 trọng số cụ thể. Từ input đầu vào, qua các Neural trung gian sẽ tìm được kết quả đầu ra phù hợp. Nếu kết quả chưa đạt tối ưu, mạng Neural sẽ lan truyền ngược để tìm lại trọng số weight ở mỗi Neural sao cho kết quả đạt được là tối ưu.



Hình .. Fully Connected Layer

Sự ra đời của mạng CNN đã giải quyết các vấn đề còn lại của các mạng Neural nhân tạo truyền thống học thông tin trong ảnh. Do sử dụng các liên kết đầy đủ giữa các điểm ảnh vào node, các mạng Neural nhân tạo truyền thẳng (Feedforward Neural Network) bị hạn chế rất nhiều bởi kích thước của ảnh, ảnh càng lớn thì số lượng liên kết càng tăng nhanh và kéo theo sự bùng nổ khối lượng tính toán. Ngoài ra sự liên kết đầy đủ này cũng là sự dư thừa khi 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 ra đời 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.

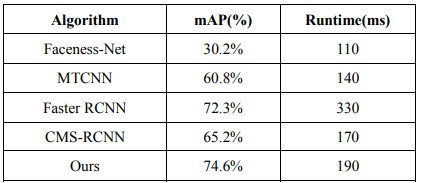
## Phát hiện khuôn mặt bằng MTCNN

Mạng nơ-ron tích chập CNN (Convolutional Neural Networks) là mô hình học sâu (Deep Learning) hiệu quả, được dùng trong nhiều bài toán phát hiện và nhận diện khuôn mặt, phân tích video, ảnh MRI,….Với những ưu điểm trong hoạt động phân tích hình ảnh của mạng CNN nhận thấy được những điều trên nên phát triển thêm mô hình “Mạng thần kinh kết nối đa tác vụ được xếp tầng” gọi tắt MTCNN. Mạng MTCNN được phát triển từ CNN. Về mặt cấu trúc MTCNN bao gồm 3 mạng CNN (Convolutional Neural Networks) xếp chồng và đồng thời hoạt động khi phát hiện và xác định khuôn mặt. Mỗi mạng CNN trong MTCNN có cấu trúc và vai trò khác nhau trong việc phát hiện khuôn mặt. Kết quả dữ liệu đầu ra của MTCNN là véc-tơ đặc trưng biểu diễn cho vị trí khuôn mặt được xác định trong bức ảnh (mắt, mũi, miệng, …)

MTCNN hoạt động theo 3 bước với 3 mạng nơ-ron riêng cho mỗi bước (P-Net, R-Net và O-Net). Khi sử dụng, MTCNN sẽ cho phép tạo ra nhiều bản sao của hình ảnh đầu vào, với các kích thước khác nhau để làm dữ liệu đầu vào.

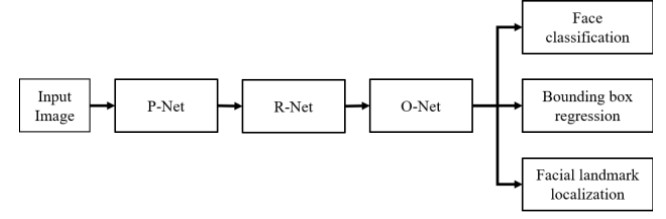
Mô hình này cho phép xác định khuôn mặt ở nhiều góc độ khác nhau, ít bị ảnh hưởng bởi ánh sáng của môi trường xung quanh và nhận diện ngay cả trong trường hợp một phần khuôn khuôn mặt bị che khuất. Phát hiện khuôn mặt và các điểm trên mặt như mắt, mũi,…Phát hiện khuôn mặt và căn chỉnh trong là một thách thức lớn do nhiều tư thế, ánh sáng khác nhau. Với Deep Learning trên nền tảng MTCNN, chúng ta hoàn toàn có thể nhận dạng, định vị chính xác khuôn mặt và các điểm trên khuôn mặt như mắt, mũi,…

Thuật toán tốc độ tính toán nhanh, hiệu quả do MTCNN đã được rất nhiều tác giả nghiên cứu và nâng cao độ chính xác. Điển hình với công bố của mình, nhóm tác giả Rong Xie, Qingyu Zhang đã so sánh đánh giá phương pháp MTCNN với các phương pháp khác để thấy được sự ưu việt với công bố “A Method of Small Face Detection Based on CNN”



Hình .. So sánh phương pháp MTCNN

Thuật toán MTCNN sử dụng mạng hiệu chuẩn để hiệu chỉnh khuôn mặt sau khi phát hiện khuôn mặt. Nó bao gồm nhiều mô hình với số lượng tính toán lớn, bỏ qua các liên kết nội tại giữa các nội giới hạn hồi quy của khuôn mặt và vị trí của khuôn mặt. Thuật toán MTCNN là một thuật toán nhận diện khuôn mặt đa tác vụ, thực hiện nhận diện khuôn mặt và điểm đặc trưng khuôn mặt. Qua phân tích so sánh của tác giả, ta có thể nhận ra tốc độ xử lý dữ liệu của thuật toán là khá cao, có thể đáp ứng yêu cầu hệ thống của mình. Chúng ta có thể biểu diễn nguyên lý hoạt động MTCNN như hình dưới :



Hình .. Nguyên lý hoạt động MTCNN

MTCNN hoạt động theo 3 bước và mỗi bước có một mạng Neural (CNN) riêng biệt gồm: P-Net, R-Net, O-Net. Tại mỗi bước sẽ thực hiện quá trình xử lý riêng với kết quả là có thể phát hiện khuôn mặt trong ảnh đầu vào nhanh và chính xác. Với mỗi bức ảnh đầu vào, nó sẽ tạo ra nhiều bản sao của hình ảnh đó với các kích thước khác nhau.Sau đây chúng ta sẽlần lượt đi sâu và phân tích cấu tạo, nguyên lý hoạt động của từng mạng xử lý, từ đó có cái nhìn tổng quan nhất về thuật toán MTCNN.

## Nhận diện khuôn mặt bằng thuật toán FACENET

Facenet 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. Hầu hết các thuật toán nhận diện khuôn mặt trước facenet đều tìm cách biểu diễn khuôn mặt bằng một véc tơ nhúng thông qua một lớp cổ chai có tác dụng giảm chiều dữ liệu. Tuy nhiên hạn chế của các thuật toán này đó là số lượng chiều nhúng tương đối lớn (thường >= 1000) và ảnh hưởng tới tốc độ của thuật toán. Thường chúng ta phải áp dụng thêm thuật toán PCA để giảm chiều dữ liệu để giảm tốc độ tính toán.

Hàm loss function chỉ đo lường khoảng cách giữa 2 bức ảnh. Như vậy trong một đầu vào huấn luyện chỉ học được một trong hai khả năng là sự giống nhau nếu chúng cùng 1 lớp hoặc sự khác nhau nếu chúng khác lớp mà không học được cùng lúc sự giống nhau và khác nhau trên cùng một lượt huấn luyện.

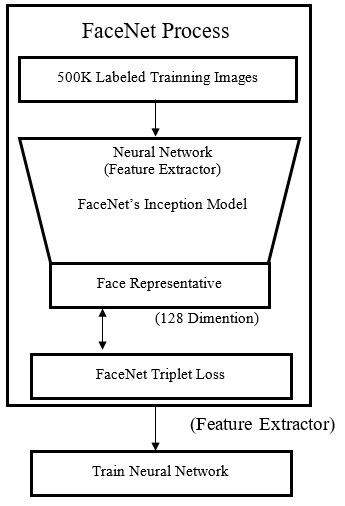
Facenet đã giải quyết cả 2 vấn đề trên bằng các hiệu chỉnh nhỏ nhưng mang lại hiệu quả lớn:

Mạng cơ sở áp dụng một mạng nơ-ron tích chập (CNN) và giảm chiều dữ liệu xuống chỉ còn 128 chiều. Do đó quá trình suy diễn và dự báo nhanh hơn và đồng thời độ chính xác vẫn được đảm bảo.

Phân tích và so sánh, đánh giá bởi các nhà khoa học cho thấy thuật toán Facenet độ chính xác khá cao, ứng dụng vào nhiều ứng dụng. Công bố ”Face Recognition Algorithm Bias: Performance Differences on Images of Children and Adults”[7] của nhóm tác giả Nisha Srinivas, Karl Ricanek, Dana Michalski đã cho thấy ứng dụng của thuật toán vào phân tích sự khácnhau giữa trẻ em và người lớn. Thuật toán có thể được training từ nhiều nguồn dataset nâng cao độ chính xác có thể thấy tính tiện lợi của thuật toán. Để thấy được sự vượt trội của thuật toán, nhóm tác giả Ivan William, De Rosal Ignatius Moses Setiadi , Eko Hari Rachmawanto đã so sánh, đánh giá với các phương pháp khác với đề tài “Face Recognition using FaceNet (Survey, Performance Test, and Comparison)”[8] để thấy tính vượt trội về độ chính xác so với phương pháp PCA,…

FaceNet cung cấp khả năng nhúng thống nhất cho các tác vụ nhận dạng, xác minh và phân cụm khuôn mặt. Nó ánh xạ mỗi hình ảnh khuôn mặt vào một không gian euclide sao cho khoảng cách trong không gian đó tương ứng với khuôn mặt, tức là hình ảnh của một người sẽ được đặt gần hơn với tất cả các hình ảnh khác của người đó so với hình ảnh của bất kỳ người nào khác có trong tập dữ liệu.

Đầu tiên ta sẽ tìm hiểu cách biểu diễn khuôn mặt khi xử lý hình ảnh. Khi xử lý, khuôn mặt sẽ được biểu diễn dưới dạng vector để quá trình nhận dạng, so sánh toán học trở nên dễ dàng. Nói đến việc nhận dạng, xác định khuôn mặt này là của ai, cần tính độ giống, khác nhau giữa các khuôn mặt chúng ta lấy được. Và nói về độ giống, khác nhau, để đơn giản, nên quy về bài toán tính khoảng cách giữa các vector.



Hình .. FaceNet process

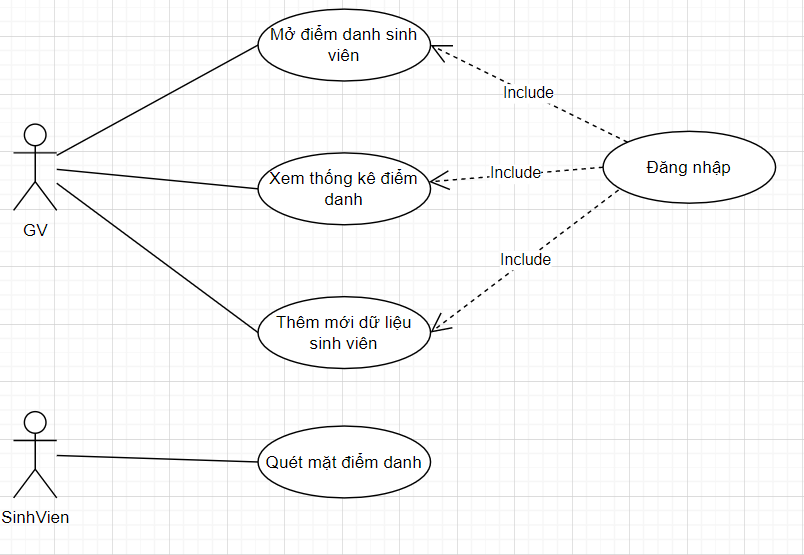
Bài toán đặt ra là tìm cách chuyển từ khung hình khuôn mặt sang vector sao cho ảnh hai khuôn mặt gần nhau thì 2 vector tương ứng cũng phải có khoảng cách gần nhau. Ảnh 2 khuôn mặt khác nhau thì 2 vector tương ứng cũng phải xa nhau hơn. Như sơ đồ trên, mô hình sử dụng Triplet Loss.

# PHÂN TÍCH THIẾT KẾ

## Biểu đồ use case

Hệ thống gồm 2 tác nhân:

* Giáo viên quản lý điểm danh sinh viên (thêm dữ liệu sinh viên, mở hệ thống điểm danh, xem thống kế sinh viên vắng)
* Sinh viên thực hiện quét mặt để điểm danh



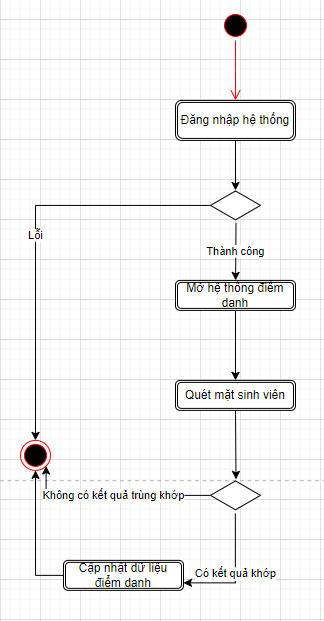
Hình .. Biểu đồ use-case

### Đặc tả usecase

|  |  |  |
| --- | --- | --- |
| Tên usecase | Mở điểm danh sinh viên | |
| Điều kiện bắt đầu | Đã đăng nhập với tài khoản giáo viên | |
| Tác nhân | Giáo viên | |
| Xử lý | Tác nhân | Hệ thống |
| Nhấn vào nút “Điểm danh” trên giao diện ứng dụng |  |
|  | Chuyển đến trang điểm danh |
| Nhập thời gian chờ để đóng hệ thống |  |
| Điều kiện kết thúc | Hệ thống đống điểm danh sau khi hết thời gian chờ | |

|  |  |  |
| --- | --- | --- |
| Tên usecase | Quét mặt điểm danh | |
| Điều kiện bắt đầu | Giáo viên đã mở hệ thống điểm danh | |
| Tác nhân | Sinh viên | |
| Xử lý | Tác nhân | Hệ thống |
| Đưa mặt đến vị trí điểm danh |  |
|  | Hệ thống xác nhận và đưa ra kết quả Cập nhật dữ liệu lên hệ thống khi xác thực danh tính sinh viên.  Báo lỗi khi không xác thực được danh tính sinh viên |
| Điều kiện kết thúc |  | |

### Biểu đồ hoạt động

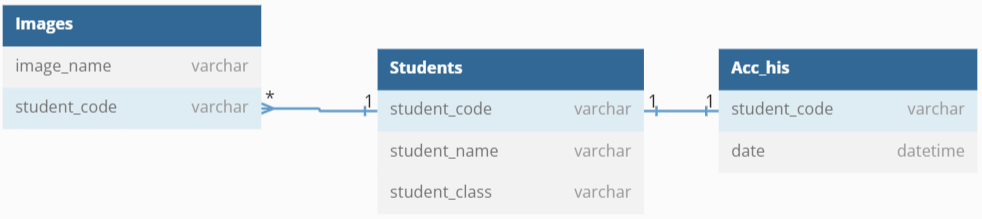


Hình .. Biểu đồ hoạt động

## Cơ sở dữ liệu

### SQL

Ngôn ngữ truy vấn mang tính cấu trúc, là một loại ngôn ngữ máy tính phổ biến để tạo, sửa, và lấy dữ liệu từ một hệ quản trị cơ sở dữ liệu quan hệ. Ngôn ngữ này phát triển vượt xa so với mục đích ban đầu là để phục vụ các hệ quản trị cơ sở dữ liệu đối tượng-quan hệ.



Hình .. Cơ sở dữ liệu hệ thống

Hệ thống ghi nhận sinh viên, ảnh nhận diện sinh viên và thời gian vào lớp

### Dataset

Tập dữ liệu là một tập hợp dữ liệu. Trong trường hợp dữ liệu dạng bảng, tập dữ liệu tương ứng với một hoặc nhiều bảng cơ sở dữ liệu, trong đó mỗi cột của bảng đại diện cho một biến cụ thể và mỗi hàng tương ứng với một bản ghi nhất định của tập dữ liệu được đề cập.

Ứng dụng lưu trữ hình ảnh của sinh viên theo mã sinh viên để training hệ thống nhận diện.

## Ngôn ngữ và công cụ hỗ trợ

### Ngôn ngữ

* Ứng dụng phát triển trên ngôn ngữ python:
* Xây dựng và ứng dụng thuật toán training model với thư viện face-detection
* Giao diện người dùng xây dựng bằng html, css
* Web-server sử dụng fask của python

### Công cụ

* Visual Studio Code
* Xampp

# XÂY DỰNG ỨNG DỤNG

KẾT LUẬN

<Kết luận: Trong thời gian làm đồ án về mặt lý thuyết tìm hiểu được gì, về mặt thực nghiệm làm được gì. Và nếu hướng phát triển của đề tài này>

TÀI LIỆU THAM KHẢO

1. Awesome Python Repositories | PythonRepo. Retrieved December 4, 2021, from <https://pythonrepo.com/>
2. Stack Overflow - Where Developers Learn, Share, & Build Careers. Retrieved December 4, 2021, from <https://stackoverflow.com/>
3. PyPI · The Python Package Index. Retrieved December 4, 2021, from <https://pypi.org/>
4. <https://www.youtube.com/watch?v=LnvxObLYO-o&ab_channel=Finxter-CreateYourCodingBusiness>
5. 3.10.0 Documentation. Retrieved December 4, 2021, from <https://docs.python.org/3/>
6. Chương 19: CƠ CHẾ MULTICAST (Phần 1) -. (2020, May 24). VnPro. Retrieved December 4, 2021, from <https://vnpro.vn/thu-vien/chuong-19-co-che-multicast-phan-1-3447.html>
7. Giao thức UDP là gì? UDP hoạt động như thế nào Giao thức UDP là gì? (2021, November 5). VinaHost. Retrieved December 4, 2021, from <https://blog.vinahost.vn/giao-thuc-udp-la-gi>
8. Python Tutorial. W3Schools. Retrieved December 4, 2021, from <https://www.w3schools.com/python/>
9. Trường Đại học Công nghệ Thông tin và Truyền thông Việt Hàn - Đại học Đà Nẵng. Retrieved December 4, 2021, from <https://vku.udn.vn/tai-nguyen/phong-dao-tao/17-mau-bao-cao-do-an-de-anthuc-tap.html>
10. Trường Đại học Công nghệ Thông tin và Truyền thông Việt Hàn - Đại học Đà Nẵng. . Retrieved December 4, 2021, from <https://vku.udn.vn/tai-nguyen/phong-dao-tao/15-huong-dan-trinh-bay-do-an-co-sothuc-tap.html>
11. Python: nhận dạng khuôn mặt: <https://v1study.com/python-tham-khao-nhan-dang-khuon-mat-voi-duoi-25-dong-ma-python.html>
12. Studocu: https://www.studocu.com/vn/document/dai-hoc-mo-dia-chat/giao-duc-the-chat/nhan-dien-khuon-mat-voi-python-va-opencv/29464701