**ĐẠI HỌC CÔNG NGHIỆP HÀ NỘI**

**TRƯỜNG CÔNG NGHỆ THÔNG TIN & TRUYỀN THÔNG**

*--------------------------------*

****

**BÁO CÁO BÀI TẬP LỚN**

**HỌC PHẦN TRÍ TUỆ NHÂN TẠO**

*ĐỀ TÀI*

**TÌM HIỂU THUẬT TOÁN FACENET VÀ ỨNG DỤNG TRONG NHẬN DẠNG KHUÔN MẶT**

GVHD: TS. Trần Thanh Huân

Nhóm Thực Hiện: Nhóm 18

|  |  |  |
| --- | --- | --- |
| 1. Nguyễn Trọng Đại | Mã SV: 2023603399 | Lớp: KHMT02-K18 |
| 2. Ngô Dương Huy | Mã SV: 2023602031 | Lớp: KHMT02-K18 |
| 4. Phạm Trọng Nghĩa | Mã SV: 2023603858 | Lớp: KHMT02-K18 |
| 5. Lê Duy Linh | Mã SV: 2023607754 | Lớp: KHMT02-K18 |

# LỜI CẢM ƠN

Trước tiên, nhóm chúng em xin chân thành cảm ơn thầy Trần Thanh Huân đã tận tình hướng dẫn và định hướng để nhóm hoàn thành đề tài “Tìm hiểu mô hình FaceNet và ứng dụng trong bài toán nhận dạng khuôn mặt”. Bên cạnh đó, chúng em cũng xin gửi lời cảm ơn đến các bạn trong nhóm vì sự nỗ lực, tinh thần làm việc nhóm và trách nhiệm cao trong suốt thời gian nghiên cứu và hoàn thiện đề tài. Mặc dù đã cố gắng hoàn thành tốt nhất trong khả năng, nhưng do thời gian và kiến thức còn hạn chế, báo cáo không thể tránh khỏi thiếu sót. Nhóm rất mong nhận được sự góp ý từ thầy Trần Thanh Huân và các bạn để cải thiện trong những lần sau. Xin chân thành cảm ơn!

***Chúng em xin chân thành cảm ơn!***

MỤC LỤC

[LỜI CẢM ƠN i](#_Toc200573643)

[DANH MỤC HÌNH VẼ iv](#_Toc200573644)

[MỞ ĐẦU 1](#_Toc200573645)

[CHƯƠNG 1: 2](#_Toc200573646)

[TỔNG QUAN VỀ NHẬN DẠNG KHUÔN MẶT 2](#_Toc200573647)

[1.1. Tổng quan về nhận dạng 2](#_Toc200573648)

[1.1.1. Khái niệm nhận dạng 2](#_Toc200573649)

[1.1.2. Bản chất của quá trình nhận dạng 2](#_Toc200573650)

[1.2.3. Ứng dụng thực tiễn tại Việt Nam 3](#_Toc200573651)

[1.2. Tổng quan về nhận dạng khuôn mặt 4](#_Toc200573652)

[1.2.1. Khái niệm 4](#_Toc200573653)

[1.2.2. Những khó khăn của hệ thống nhận dạng khuôn mặt 5](#_Toc200573654)

[1.2.3. Tổng quan kiến trúc của một hệ thống nhận dạng khuôn mặt 6](#_Toc200573655)

[1.3. Tổng kết chương 1 8](#_Toc200573656)

[CHƯƠNG 2: 9](#_Toc200573657)

[MÔ HÌNH FACENET VÀ ỨNG DỤNG TRONG NHẬN DẠNG KHUÔN MẶT 9](#_Toc200573658)

[2.1. FaceNet và đặc trưng ánh xạ không gian 9](#_Toc200573659)

[2.2. Cấu trúc mạng FaceNet 10](#_Toc200573660)

[2.2.1. Mạng nơ-ron tích chập (CNN) để trích xuất đặc trưng 10](#_Toc200573661)

[2.2.2. Lớp Fully Connected (FC) để ánh xạ đặc trưng 11](#_Toc200573662)

[2.2.3. Hàm mất mát Triplet Loss 11](#_Toc200573663)

[2.2.4. Kỹ thuật tăng cường (Data Augmentation) 13](#_Toc200573664)

[2.3. Tổng kết chương 2. 13](#_Toc200573665)

[CHƯƠNG 3: 14](#_Toc200573666)

[GIẢI QUYẾT BÀI TOÁN NHẬN DẠNG KHUÔN MẶT 14](#_Toc200573667)

[3.1. Phát biểu bài toán 14](#_Toc200573668)

[3.1.1. Đầu vào 14](#_Toc200573669)

[3.1.2. Đầu ra 14](#_Toc200573670)

[3.2. Công cụ và môi trường phát triển 15](#_Toc200573671)

[3.2.1. Anaconda 15](#_Toc200573672)

[3.2.2. Visual Studio Code (VSCode) 15](#_Toc200573673)

[3.3.3. Google Colab 15](#_Toc200573674)

[3.3. Chuẩn bị dữ liệu 16](#_Toc200573675)

[3.4. Giải quyết bài toán 16](#_Toc200573676)

[TÀI LIỆU THAM KHẢO 19](#_Toc200573677)

# DANH MỤC HÌNH VẼ

[Hình 1.1 Hệ thống nhận dạng mặt người 5](#_Toc200572920)

[Hình 1.2. Các bước chính trong hệ thống nhận dạng khuôn mặt 7](#_Toc200572921)

[Hình 2.1. Mạng nơ-ron tích chập và Triplet Loss trong FaceNet 9](#_Toc200572922)

[Hình 2.2. Cấu trúc mạng facenet 10](#_Toc200572923)

# MỞ ĐẦU

Công nghệ thông tin ngày càng phát triển mạnh mẽ và đóng vai trò then chốt trong nhiều lĩnh vực của đời sống hiện đại. Con người đang từng bước tạo ra những hệ thống thông minh có khả năng học hỏi, phân tích và xử lý dữ liệu một cách tự động, hỗ trợ hiệu quả cho các hoạt động trong thực tế. Trong những năm gần đây, một trong những bài toán nổi bật và thu hút nhiều sự quan tâm từ cộng đồng nghiên cứu đó là “Nhận diện khuôn mặt bằng xử lý ảnh”.

Ngay từ khi xuất hiện, bài toán nhận diện khuôn mặt đã nhận được nhiều sự chú ý nhờ tiềm năng to lớn trong việc thay thế các phương pháp xác thực truyền thống. Mặc dù đã có nhiều hướng tiếp cận được đề xuất, nhưng việc đạt được độ chính xác tuyệt đối trong mọi điều kiện vẫn còn là một thách thức. Tuy vậy, các kết quả hiện tại cũng đã đủ để triển khai vào thực tế và mang lại hiệu quả cao trong việc hỗ trợ quản lý và nâng cao tính tự động hóa.

Ngày nay, với sự phát triển nhanh chóng của trí tuệ nhân tạo và học sâu, việc xây dựng các hệ thống nhận diện khuôn mặt chính xác và tối ưu đã trở nên khả thi hơn. Trong phạm vi bài báo cáo này, chúng em lựa chọn sử dụng FaceNet để nhận diện khuôn mặt thông qua việc mã hóa đặc trưng – đây là mô hình đang được đánh giá cao về hiệu suất và độ chính xác.

Vì vậy, mục tiêu chính của chúng em trong bài viết này là tìm hiểu và xây dựng một ứng dụng nhỏ có khả năng nhận diện khuôn mặt dựa trên các kỹ thuật học sâu hiện đại.

Dưới đây là bố cục của phần báo cáo:

* Chương 1: Tổng quan về nhận dạng khuôn mặt
* Chương 2: Mô hình facenet và ứng dụng trong nhận dạng khuôn mặt
* Chương 3: Giải quyết bài toán nhận dạng khuôn mặt

# CHƯƠNG 1:

# TỔNG QUAN VỀ NHẬN DẠNG KHUÔN MẶT

Chương này trình bày những cơ sở lý thuyết liên quan đến bài toán nhận dạng khuôn mặt, một lĩnh vực quan trọng trong thị giác máy tính và trí tuệ nhân tạo. Nội dung chương sẽ đi sâu vào khái niệm, bản chất và quy trình của quá trình nhận dạng, đồng thời phân tích các yếu tố ảnh hưởng đến độ chính xác của hệ thống. Bên cạnh đó, chương cũng giới thiệu các ứng dụng thực tiễn của công nghệ nhận dạng khuôn mặt, đặc biệt trong bối cảnh tại Việt Nam, nhằm làm rõ tính cấp thiết và tiềm năng ứng dụng của đề tài trong đời sống hiện nay.

## 1.1. Tổng quan về nhận dạng

### 1.1.1. Khái niệm nhận dạng

Nhận dạng là quá trình mô phỏng khả năng cảm nhận, phân tích và phân biệt các đối tượng của con người bằng hệ thống máy tính. Các đối tượng được nhận dạng có thể là âm thanh, hình ảnh, tín hiệu hoặc bất kỳ dạng dữ liệu nào mà con người có thể cảm nhận và phân biệt được thông qua các đặc trưng riêng. Quá trình này tái hiện lại cách con người cảm nhận và đưa ra quyết định nhận dạng, giúp máy tính có thể tự động nhận biết các mẫu dữ liệu trong thế giới thực.

### 1.1.2. Bản chất của quá trình nhận dạng

Bản chất của nhận dạng mẫu nằm ở việc phân tích và phân loại các đối tượng trong dữ liệu dựa trên các đặc trưng có ý nghĩa. Hệ thống sẽ thực hiện việc trích xuất đặc trưng từ dữ liệu, ví dụ như hình dạng hoặc màu sắc trong ảnh, sau đó áp dụng các thuật toán học máy hoặc thống kê để tìm ra mẫu tương ứng hoặc nhóm phù hợp. Đây là một quá trình phức tạp, liên quan đến việc phát hiện và lý giải các cấu trúc hoặc quy luật trong dữ liệu đầu vào.

Một hệ thống nhận dạng hiện đại thường sử dụng các phương pháp học máy như mạng nơ-ron, máy vector hỗ trợ, cây quyết định… nhằm xây dựng khả năng học và dự đoán dựa trên dữ liệu huấn luyện. Quan trọng hơn cả là hệ thống cần có khả năng tổng quát hóa, tức là áp dụng kiến thức học được từ dữ liệu cũ vào dữ liệu mới chưa từng thấy. Quá trình nhận dạng không chỉ có ý nghĩa lý thuyết mà còn rất thực tiễn, được ứng dụng trong nhiều lĩnh vực như xử lý hình ảnh, giọng nói, y học, tài chính, và cả bảo mật.

Để xây dựng được hệ thống nhận dạng, ba giai đoạn chính cần được thực hiện: xác định mô hình biểu diễn đối tượng, lựa chọn quy tắc ra quyết định, và cuối cùng là tiến hành quá trình học nhận dạng. Khi đã xác định cách biểu diễn đối tượng phù hợp, hệ thống sẽ bước vào giai đoạn học nhằm tối ưu hóa việc phân lớp các đối tượng và gán nhãn phù hợp cho chúng. Trong đó, học có giám sát là phương pháp phổ biến, dựa trên các mẫu chuẩn có sẵn để huấn luyện hệ thống phân biệt đối tượng mới. Ví dụ như trong phân tích ảnh viễn thám, mẫu chuẩn về đồng lúa, rừng hoặc đất hoang sẽ giúp hệ thống xác định được đối tượng trong ảnh đầu vào. Hệ thống cần có khả năng đối sánh và ra quyết định dựa trên các hàm phân lớp hoặc hàm đánh giá cụ thể.

Ngược lại, học không giám sát là kỹ thuật phức tạp hơn vì không có dữ liệu gán nhãn sẵn. Hệ thống sẽ tự động tìm cách phân nhóm dữ liệu dựa trên sự tương đồng giữa các đặc trưng. Quá trình này bao gồm nhiều bước thử nghiệm và tinh chỉnh để tìm ra cấu trúc nhóm tối ưu. Dữ liệu sẽ được phân tích và phân loại dần theo hướng cải tiến, nhằm đạt được kết quả phân nhóm tốt nhất mà không cần đến sự hướng dẫn từ trước.

### 1.2.3. Ứng dụng thực tiễn tại Việt Nam

Tại Việt Nam, công nghệ nhận dạng khuôn mặt đã được đưa vào ứng dụng trong nhiều lĩnh vực khác nhau, góp phần tự động hóa, tối ưu hoá và nâng cao hiệu quả của các hệ thống quản lý. Trong lĩnh vực tài chính – ngân hàng, nhận dạng khuôn mặt giúp xác thực người dùng khi đăng nhập tài khoản, thực hiện giao dịch hay mở tài khoản trực tuyến, từ đó gia tăng tính bảo mật và hạn chế các hành vi giả mạo. Bên cạnh đó, tại các cơ quan, doanh nghiệp, công nghệ này hỗ trợ trong việc kiểm soát ra vào, điểm danh, ghi nhận thời gian làm việc của nhân viên một cách tự động và chính xác.

Trong lĩnh vực an ninh, hệ thống nhận dạng khuôn mặt đóng vai trò quan trọng tại các cửa khẩu, trạm kiểm soát, giúp phát hiện các đối tượng khả nghi hoặc ngăn chặn hành vi vượt biên trái phép. Tại các khu dân cư, chung cư, hay tòa nhà thông minh, việc tích hợp hệ thống nhận diện khuôn mặt cho phép kiểm soát truy cập, mở cửa tự động, đảm bảo an toàn và hạn chế xâm nhập trái phép. Ngoài ra, công nghệ này cũng đang được thử nghiệm trong giám sát hành khách tại sân bay, nhà ga, và phục vụ công tác tìm kiếm người mất tích.

## 1.2. Tổng quan về nhận dạng khuôn mặt

### 1.2.1. Khái niệm

Nhận dạng khuôn mặt 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.

A diagram of a face detection system

AI-generated content may be incorrect.

Hình 1.1 Hệ thống nhận dạng mặt người

### 1.2.2. Những khó khăn của hệ thống nhận dạng khuôn mặt

Bài toán nhận dạng mặt người là bài toán đã được nghiên cứu từ những năm 70. Tuy nhiên, đây là một bài toán khó nên những nghiên cứu hiện tại vẫn chưa đạt được những kết quả mong muốn. Chính vì thế, vấn đề này vẫn đang được nhiều nhóm trên thế giới quan tâm nghiên cứu. Khó khăn của bài toán nhận dạng mặt người có thể kể đến như sau:

a) Tư thế chụp, góc chụp: Ảnh chụp khuôn mặt có thể thay đổi rất nhiều bởi vì góc chụp giữa camera và khuôn mặt. Chẳng hạn như: chụp thẳng, chụp chéo bên trái 45o hay chụp chéo bên phải 45o, chụp từ trên xuống, chụp từ dưới lên, v.v… Với các tư thế khác nhau, các thành phần trên khuôn mặt như mắt, mũi, miệng có thể bị khuất một phần hoặc thậm chí khuất hết.

b) Sự xuất hiện hoặc thiếu một số thành phần của khuôn mặt: Các Đặc trưng như: râu mép, râu hàm, mắt kính, v.v… có thể xuất hiện hoặc không. Vấn đề này làm cho bài toán càng trở nên khó hơn rất nhiều.

c) Sự biểu cảm của khuôn mặt: Biểu cảm của khuôn mặt con người có thể làm ảnh hưởng đáng kể lên các thông số của khuôn mặt. Chẳng hạn, cùng một khuôn mặt một người, nhưng có thể sẽ rất khác khi họ cười hoặc sợ hãi, v.v…

d) Sự che khuất: Khuôn mặt có thể bị che khuất bởi các đối tượng

khác hoặc các khuôn mặt khác.

e) Hướng của ảnh (pose variations): Các ảnh khuôn mặt có thể biếnđổi rất nhiều với các góc quay khác nhau của trục camera. Chẳng hạn chụp với trục máy ảnh nghiêng làm cho khuôn mặt bị nghiêng so với trục của ảnh.

f) Điều kiện của ảnh: Ảnh được chụp trong các điều kiện khác nhauvề: chiếu sáng, về tính chất camera (máy kỹ thuật số, máy hồng ngoại, v.v…), ảnh có chất lượng thấp ảnh hưởng rất nhiều đến chất lượng ảnh khuôn mặt.

g) Aging condition: Việc nhận dạng ảnh mặt thay đổi theo thời gian còn là một vấn đề khó khăn, ngay cả đối với khả năng nhận dạng của con người.

h) Các hệ thống cực lớn (very large scale systems): Các CSDL ảnh mặt được test bởi các nhà nghiên cứu còn khá nhỏ (vài trăm tới vài chục nghìn ảnh mặt), tuy nhiên trên thực tế các CSDL có thể rất lớn, ví dụ CSDL ảnh mặt của cảnh sát của một đất nước có thể chứa từ hàng triệu tới hơn 1 tỷ ảnh…

### 1.2.3. Tổng quan kiến trúc của một hệ thống nhận dạng khuôn mặt

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).

A diagram of a person's face

AI-generated content may be incorrect.

Hình 1.2. 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. 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 classification), 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 bài tập này, không tập trung tìm hiểu bước phát hiện khuôn mặt mà chỉ tập trung chủ yếu vào bước nhận dạng khuôn mặt.

## 1.3. Tổng kết chương 1

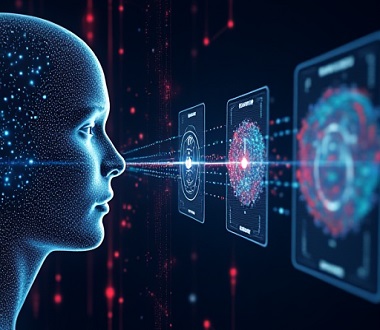
Chương 1 đã trình bày tổng quan về bài toán nhận dạng khuôn mặt, bao gồm khái niệm, vai trò và các bước chính trong hệ thống nhận dạng khuôn mặt như phát hiện, tiền xử lý, trích xuất đặc trưng và so sánh. Bên cạnh đó, chương này cũng đã điểm qua một số phương pháp tiếp cận truyền thống và hiện đại, từ các kỹ thuật dựa trên đặc trưng thủ công đến các mô hình học sâu. Những kiến thức này cung cấp cái nhìn toàn diện và là cơ sở lý thuyết quan trọng cho việc tìm hiểu sâu hơn về mô hình FaceNet trong các chương tiếp theo.

# CHƯƠNG 2:

# MÔ HÌNH FACENET VÀ ỨNG DỤNG TRONG NHẬN DẠNG KHUÔN MẶT

## 2.1. FaceNet và đặc trưng ánh xạ không gian

FaceNet là một mô hình học sâu (Deep Learning) được phát triển bởi Google vào năm 2015. Điểm nổi bật của FaceNet là nó không chỉ đơn giản là nhận diện khuôn mặt mà còn ánh xạ mỗi khuôn mặt vào một không gian đặc trưng (embedding space) sao cho khoảng cách Euclidean giữa hai vector đặc trưng đại diện cho hai khuôn mặt tương tự nhau sẽ nhỏ, trong khi khoảng cách giữa hai vector đại diện cho hai khuôn mặt khác nhau sẽ lớn.

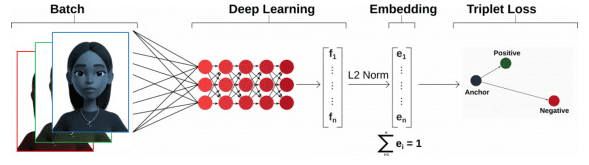


Hình 2.1. Mạng nơ-ron tích chập và Triplet Loss trong FaceNet

FaceNet sử dụng một mạng nơ-ron tích chập (Convolutional Neural Network – CNN) để học các đặc trưng của khuôn mặt. Sau khi qua mạng, mỗi khuôn mặt sẽ được ánh xạ thành một vector có độ dài cố định, thường là 128 chiều, gọi là face embedding. Vector này có thể được sử dụng cho các nhiệm vụ như xác thực khuôn mặt, tìm kiếm khuôn mặt, và nhóm khuôn mặt.

Một trong những đặc điểm chính của FaceNet là khả năng tạo ra các vector đặc trưng có tính phân biệt cao. Điều này đạt được nhờ việc sử dụng một hàm mất mát đặc biệt gọi là Triplet Loss, cùng với một kỹ thuật huấn luyện tinh vi.

## 2.2. Cấu trúc mạng FaceNet



Hình 2.2. Cấu trúc mạng facenet

Mô hình FaceNet có một cấu trúc phức tạp với nhiều thành phần khác nhau, mỗi phần đóng một vai trò quan trọng trong việc trích xuất và ánh xạ đặc trưng khuôn mặt. Cấu trúc chính của FaceNet bao gồm các phần sau:

### 2.2.1. Mạng nơ-ron tích chập (CNN) để trích xuất đặc trưng

Phần đầu tiên của FaceNet là một mạng nơ-ron tích chập (CNN) được thiết kế để trích xuất các đặc trưng từ hình ảnh khuôn mặt. Mạng này bao gồm nhiều lớp tích chập (convolutional layers) và lớp gộp (pooling layers) để học các đặc trưng hình học và phong cách của khuôn mặt.

Cấu trúc điển hình của mạng CNN trong FaceNet có thể bao gồm:

* Lớp tích chập (Convolutional Layer): Các lớp này giúp phát hiện các đặc trưng cơ bản của hình ảnh, như cạnh và kết cấu. Ví dụ, lớp tích chập đầu tiên có thể có 64 kênh đầu ra với kích thước kernel là 7×7.
* Lớp Gộp (Pooling Layer): Các lớp này giảm kích thước không gian của hình ảnh đầu vào, giúp giảm số lượng tham số và tính toán, đồng thời giữ lại các đặc trưng quan trọng. Lớp gộp thường sử dụng phương pháp gộp tối đa (max pooling) với kích thước kernel là 3×3.
* Lớp Batch Normalization: Lớp này giúp chuẩn hóa đầu ra của các lớp tích chập, giảm thiểu sự thay đổi phân phối trong quá trình huấn luyện.
* Lớp kích hoạt (Activation Layer): Lớp ReLU (Rectified Linear Unit) thường được sử dụng để thêm tính phi tuyến vào mô hình, giúp mô hình học các đặc trưng phức tạp hơn.

Ví dụ về cấu trúc mạng CNN như sau:

## 2.2.2. Lớp Fully Connected (FC) để ánh xạ đặc trưng

Sau khi các đặc trưng được trích xuất bởi các lớp CNN, chúng được đưa qua các lớp fully connected (FC) để ánh xạ các đặc trưng này thành vector embedding có độ dài cố định. Các lớp FC thường được thiết kế để giảm dần số lượng kích thước đặc trưng từ các lớp tích chập xuống một kích thước nhỏ hơn, thường là 128 hoặc 512 chiều cho vector embedding.

Cấu trúc của các lớp FC có thể như sau:

Flatten→FC1 (1024 units) →ReLU→FC2 (Embedding Dim)

Trong đó, *Embedding Dim* là kích thước của vector đặc trưng, thường là 128.

## 2.2.3. Hàm mất mát Triplet Loss

Triplet Loss là một trong những hàm mất mát quan trọng nhất trong mô hình FaceNet. Mục tiêu của Triplet Loss là đảm bảo rằng khoảng cách giữa vector đặc trưng của một khuôn mặt (gọi là *Anchor*) và một khuôn mặt khác của cùng một người (gọi là *Positive*) phải nhỏ hơn khoảng cách giữa vector đặc trưng của khuôn mặt đó và khuôn mặt của một người khác (gọi là *Negative*).

Định nghĩa:

Trong đó:

* : Hình ảnh Anchor (mỏ neo)
* : Hình ảnh Positive (tích cực, cùng một người với Anchor)
* : Hình ảnh Negative (tiêu cực, người khác với Anchor)
* : Hàm ánh xạ hình ảnh  thành vector đặc trưng (embedding)
* : Một siêu tham số gọi là margin, giúp kiểm soát khoảng cách tối thiểu giữa Positive và Negative

Hàm mất mát này yêu cầu rằng khoảng cách giữa Anchor và Positive nhỏ hơn khoảng cách giữa Anchor và Negative ít nhất là một lượng bằng . Điều này giúp mô hình học cách phân biệt rõ ràng giữa các khuôn mặt khác nhau.

Hoạt động:

Hàm Triplet Loss dựa trên việc điều chỉnh ba hình ảnh: Anchor (A), Positive (P), và Negative (N). Mục tiêu của mô hình là học sao cho:

Trong đó,  là khoảng cách Euclidean giữa các vector đặc trưng. Điều này có nghĩa là khoảng cách giữa Anchor và Positive sẽ được thu nhỏ, trong khi khoảng cách giữa Anchor và Negative sẽ được mở rộng.

Ví dụ, nếu

* là vector đặc trưng cho một khuôn mặt,
* là vector đặc trưng cho cùng một khuôn mặt nhưng với biểu cảm khác, và
* là vector đặc trưng cho một khuôn mặt khác, thì hàm Triplet Loss sẽ cố gắng đảm bảo rằng:

Nếu điều kiện này không được thoả mãn, mô hình sẽ điều chỉnh các trọng số để giảm giá trị mất mát, làm cho các vector đặc trưng di chuyển gần hoặc xa nhau tùy theo yêu cầu.

## 2.2.4. Kỹ thuật tăng cường (Data Augmentation)

Để cải thiện khả năng tổng quát của mô hình và giảm hiện tượng overfitting, các kỹ thuật tăng cường dữ liệu thường được sử dụng.

Các kỹ thuật này có thể bao gồm:

* **Chuyển đổi màu sắc:** Thay đổi độ sáng, độ tương phản, và bão hòa của hình ảnh.
* **Biến dạng hình học:** Xoay, cắt, và thay đổi kích thước của hình ảnh.
* **Làm mờ và nhiễu:** Thêm nhiễu hoặc làm mờ hình ảnh để làm cho mô hình mạnh mẽ hơn với các điều kiện không giống như trong dữ liệu huấn luyện.

Việc áp dụng các kỹ thuật tăng cường giúp mô hình FaceNet có khả năng xử lý các biến thể trong điều kiện thực tế và cải thiện hiệu suất nhận diện khuôn mặt.

## 2.3. Tổng kết chương 2.

Chương 2 đã trình bày chi tiết về mô hình **FaceNet**, bao gồm kiến trúc mạng nơ-ron sâu, cơ chế học biểu diễn đặc trưng khuôn mặt thông qua hàm mất mát *Triplet Loss*, cũng như quá trình ánh xạ ảnh khuôn mặt vào không gian vector có tính phân biệt cao. Mô hình cho phép thực hiện hiệu quả các tác vụ như **xác thực**, **nhận dạng** và **phân biệt khuôn mặt** với độ chính xác cao. Những nội dung được trình bày trong chương này đóng vai trò nền tảng lý thuyết cho việc triển khai và đánh giá mô hình trong các chương tiếp theo.

# CHƯƠNG 3:

# GIẢI QUYẾT BÀI TOÁN NHẬN DẠNG KHUÔN MẶT

Chương 3 trình bày quá trình giải quyết bài toán nhận dạng khuôn mặt từ việc phát biểu bài toán, xác định đầu vào – đầu ra cho đến các bước triển khai hệ thống. Trong chương này, nhóm sử dụng ngôn ngữ lập trình Python cùng các thư viện hỗ trợ mạnh mẽ như OpenCV, TensorFlow, Keras… để xử lý ảnh, xây dựng và vận hành mô hình học sâu. Mô hình được áp dụng là FaceNet, có khả năng ánh xạ khuôn mặt thành vector đặc trưng, từ đó thực hiện việc nhận dạng thông qua so sánh các vector. Bên cạnh đó, chương này cũng giới thiệu về công cụ và môi trường phát triển sử dụng trong quá trình xây dựng hệ thống, đồng thời trình bày các bước xử lý dữ liệu, huấn luyện mô hình và đánh giá kết quả đầu ra.

## 3.1. Phát biểu bài toán

### 3.1.1. Đầu vào

Với bài toán nhận dạng khuôn mặt, đầu vào là các bức ảnh hoặc khung hình chứa khuôn mặt người. Các ảnh này có thể thu thập từ camera giám sát, webcam, hoặc ảnh tĩnh. Sau khi được đưa vào hệ thống, ảnh sẽ trải qua bước tiền xử lý như: phát hiện khuôn mặt, cắt ảnh khuôn mặt, điều chỉnh kích thước và chuẩn hoá dữ liệu đầu vào.

Tiếp theo, ảnh khuôn mặt sau khi tiền xử lý sẽ được đưa qua mô hình FaceNet để ánh xạ thành vector đặc trưng. Vector này sẽ được so sánh với tập vector khuôn mặt đã lưu trong cơ sở dữ liệu để xác định danh tính người trong ảnh.

#ảnh

### 3.1.2. Đầu ra

Đầu ra của hệ thống là **tên của người xuất hiện trong ảnh**, nếu hệ thống tìm thấy khuôn mặt trùng khớp trong dữ liệu. Trong trường hợp không tìm được người phù hợp, hệ thống có thể trả về kết quả “Không xác định”.

## 3.2. Công cụ và môi trường phát triển

Trong quá trình triển khai hệ thống nhận dạng khuôn mặt, nhóm sử dụng nhiều công cụ và môi trường phát triển nhằm đảm bảo hiệu quả trong việc lập trình, xử lý dữ liệu và huấn luyện mô hình học sâu.

Dưới đây là các công cụ chính được sử dụng:

### 3.2.1. Anaconda

Anaconda là một nền tảng phân phối Python phổ biến, hỗ trợ quản lý môi trường ảo và các thư viện liên quan đến khoa học dữ liệu, trí tuệ nhân tạo. Việc sử dụng Anaconda giúp nhóm dễ dàng cài đặt và quản lý các gói thư viện như NumPy, OpenCV, TensorFlow, Keras… mà không gặp phải xung đột về phiên bản. Ngoài ra, Anaconda còn hỗ trợ giao diện dòng lệnh (Anaconda Prompt) và các công cụ tích hợp như Jupyter Notebook.

### 3.2.2. Visual Studio Code (VSCode)

Visual Studio Code là trình soạn thảo mã nguồn nhẹ, đa nền tảng và mạnh mẽ, được sử dụng để viết và quản lý mã nguồn Python. Với sự hỗ trợ từ các tiện ích mở rộng (extension) như Python Extension, Jupyter, Pylance, VSCode cung cấp môi trường lập trình trực quan, hỗ trợ tự động hoàn thành mã, gợi ý cú pháp và gỡ lỗi hiệu quả. Nhờ đó, nhóm có thể tổ chức cấu trúc dự án rõ ràng và thuận tiện trong quá trình phát triển hệ thống.

### 3.3.3. Google Colab

Google Colaboratory (Google Colab) là nền tảng lập trình trên nền web do Google cung cấp, cho phép người dùng chạy mã Python trực tiếp trên trình duyệt mà không cần cấu hình hệ thống cục bộ. Colab hỗ trợ truy cập GPU/TPU miễn phí, rất hữu ích trong việc huấn luyện mô hình deep learning như FaceNet. Ngoài ra, Colab cũng tích hợp tốt với Google Drive, giúp lưu trữ dữ liệu và chia sẻ tài nguyên giữa các thành viên nhóm một cách dễ dàng.

## 3.3. Chuẩn bị dữ liệu

Dữ liệu đầu vào đóng vai trò quan trọng trong việc huấn luyện và đánh giá hiệu quả của mô hình nhận dạng khuôn mặt. Trong phạm vi đề tài này, nhóm sử dụng một tập dữ liệu gồm **khoảng 200 hình ảnh khuôn mặt riêng lẻ**, mỗi ảnh chứa duy nhất một khuôn mặt và gắn liền với nhãn danh tính tương ứng.

A screenshot of a computer

AI-generated content may be incorrect.

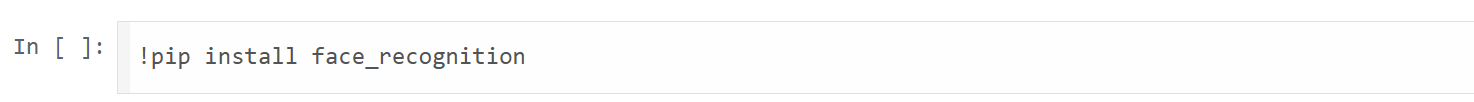
## 3.4. Giải quyết bài toán

- Kết nối Colab với Google Drive

A white card with green text

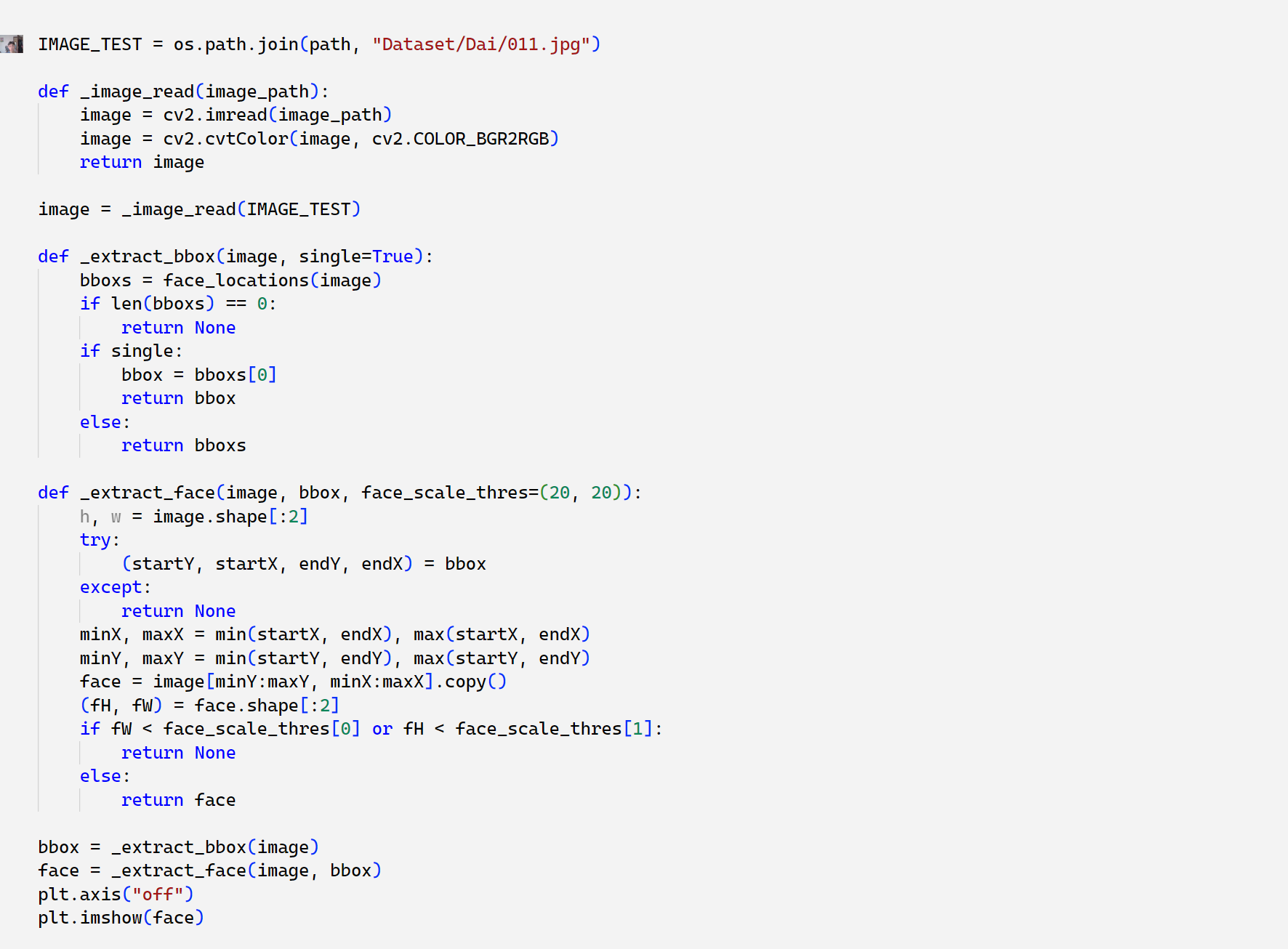
AI-generated content may be incorrect.

Cài đặt thư viện face\_recognition



Tiền xử lí dữ liệu:

* Hàm phát hiện và trích xuất khuôn mặt.



* Tải và xử lí dữ liệu

A screen shot of a computer code

AI-generated content may be incorrect.

A screen shot of a computer code

AI-generated content may be incorrect.

# TÀI LIỆU THAM KHẢO

Trần Hùng Cường, Nguyễn Phương Nga, Giáo trình Trí tuệ nhân tạo, 2014

[2] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "DeepFace: Closing the gap to human-level performance in face verification," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2014, pp. 1701-1708, doi: 10.1109/CVPR.2014.220.

[3] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," in Proc. Brit. Mach. Vis. Conf., 2015, pp. 41.1-41.12, doi: 10.5244/C.29.41.

[4] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "Labeled Faces in the Wild: A database for studying face recognition in unconstrained environments," Univ. Massachusetts, Amherst, Tech. Rep. 07-49, 2007.

[5] M. A. Turk and A. P. Pentland, "Face recognition using eigenfaces," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 1991, pp. 586-591, doi: 10.1109/CVPR.1991.139758.

[6] N. V. Thắng and Đ. Q. Minh, "Ứng dụng học sâu trong nhận dạng khuôn mặt: Nghiên cứu và triển khai," Tạp chí Khoa học và Công nghệ Việt Nam, vol. 62, no. 3, pp. 45-52, 2020.

[7] T. H. Anh, "Nghiên cứu mô hình FaceNet trong bài toán nhận diện khuôn mặt tại Việt Nam," in Proc. Hội nghị Khoa học Công nghệ Thông tin, ĐH Công nghệ, ĐHQGHN, 2021, pp. 123-130.

[8] P. V. Long and L. T. Hiền, "Xây dựng hệ thống nhận dạng khuôn mặt sử dụng mạng nơ-ron sâu," Tạp chí Nghiên cứu Khoa học và Phát triển Công nghệ, vol. 15, no. 2, pp. 33-40, 2022.