

## Mục lục

<b>MỞ ĐẦU</b> .....	<b>3</b>
0.1 Lý do chọn đề tài .....	3
0.2 Mục đích của đề tài .....	4
0.3 Đối tượng và phạm vi nghiên cứu của đề tài .....	4
0.3.1 Đối tượng .....	4
0.3.2 Phạm vi nghiên cứu .....	4
0.4 Kết cấu của đồ án .....	5
0.5 Lời cảm ơn .....	5
0.6 Bảng các ký hiệu .....	6
<b>1 Nghiên cứu tổng quan</b> .....	<b>7</b>
1.1 Các phương pháp nghiên cứu .....	7
1.2 Ưu nhược điểm của các phương pháp .....	7
1.2.1 Nhận dạng dựa trên các đặc trưng của các phần tử trên khuôn mặt: .....	7
1.2.2 Nhận dạng dựa trên xét tổng thể khuôn mặt: .....	8
1.2.3 Kết luận .....	8
<b>2 Cơ sở lý thuyết</b> .....	<b>9</b>

2.1	Tổng quan về nhận diện khuôn mặt .....	9
2.2	Tìm hiểu về OpenCV và ngôn ngữ lập trình Python .....	9
2.3	Mô hình mạng neural tích chập (CNN - Convolutional neural network) .....	11
2.3.1	Lớp tích chập .....	11
2.4	Máy dò khuôn mặt (Face detector) .....	12
2.5	Các kĩ thuật làm giàu dữ liệu (Data agumentation) .....	12
2.6	Mô hình học sâu được huấn luyện trước (Pre-train model) .....	12
2.6.1	Sử dụng mô hình được huấn luyện trước .....	12
2.6.2	Giới thiệu Facenet .....	12
2.6.3	Giới thiệu Mạng InceptionResnetV1 .....	12
2.6.4	Kĩ thuật đánh giá bộ ba (Triplet loss) .....	12
<b>3</b>	<b>Thiết kế và xây dựng hệ thống .....</b>	<b>13</b>
3.1	Phân tích .....	13
3.2	Xây dựng .....	13
<b>4</b>	<b>Kết luận và hướng phát triển .....</b>	<b>14</b>
4.1	Kết luận .....	14
	<b>Tài liệu tham khảo .....</b>	<b>15</b>
	<b>Index .....</b>	<b>16</b>

# MỞ ĐẦU

## 0.1. Lý do chọn đề tài

Với sự phát triển không ngừng của khoa học và công nghệ, đặc biệt là với những chiếc điện thoại thông minh (smartphone) ngày càng hiện đại và được sử dụng phổ biến trong đời sống con người đã làm cho lượng thông tin thu được bằng hình ảnh ngày càng tăng. Theo đó, lĩnh vực xử lý ảnh cũng được chú trọng phát triển, ứng dụng rộng rãi trong đời sống xã hội hiện đại. Không chỉ dừng lại ở việc chỉnh sửa, tăng chất lượng hình ảnh mà với công nghệ xử lý ảnh hiện nay chúng ta có thể giải quyết các bài toán nhận dạng chữ viết, nhận dạng dấu vân tay, nhận dạng khuôn mặt... Một trong những bài toán được nhiều người quan tâm nhất của lĩnh vực xử lý ảnh hiện nay đó là nhận dạng khuôn mặt (Face Recognition). Như chúng ta đã biết, khuôn mặt đóng vai trò quan trọng trong quá trình giao tiếp giữa người với người, nó mang một lượng thông tin giàu có, chẳng hạn như từ khuôn mặt chúng ta có thể xác định giới tính, tuổi tác, chủng tộc, trạng thái cảm xúc, đặc biệt là xác định mối quan hệ với đối tượng (có quen biết hay không).

Do đó, bài toán nhận dạng khuôn mặt đóng vai trò quan trọng trong nhiều lĩnh vực đời sống hằng ngày của con người như các hệ thống giám sát, quản lý vào ra, tìm kiếm thông tin một người nổi tiếng,...đặc biệt là an ninh, bảo mật. Có rất nhiều phương pháp nhận dạng khuôn mặt để nâng cao hiệu suất tuy nhiên dù ít hay nhiều những phương pháp này đang vấp phải những thử thách về độ sáng, hướng nghiêng, kích thước ảnh, hay ảnh hưởng của tham số môi trường. Bài toán Nhận diện khuôn mặt (Face Recognition) bao gồm nhiều bài toán khác nhau như: phát hiện mặt người (face detection), đánh dấu (facial landmarking), trích chọn(rút) đặc trưng (feature extration), gán nhãn, phân lớp (classification). Trong thực tế, nhận dạng khuôn mặt người (Face Recognition) là một hướng nghiên cứu được nhiều nhà khoa học quan tâm, nghiên cứu để ứng dụng trong thực tiễn. Vì thế có những cải tiến nghiên cứu về bài toán phát hiện khuôn mặt người trong những môi trường phức tạp hơn, có nhiều khuôn mặt người trong ảnh hơn, và có nhiều tư thế thay đổi trong ảnh... Trong bài này tôi sẽ tìm hiểu

về trích rút đặc trưng sử dụng học sâu (Deep learning) và áp dụng vào bài toán nhận diện khuôn mặt .

## 0.2. Mục đích của đề tài

- Xây dựng, tìm kiếm các mô hình học sâu để trích xuất thông tin khuôn mặt người dùng 1 cách chính xác, hiệu quả.
- Tìm hiểu sử dụng thành thạo các phương pháp làm giàu dữ liệu, ứng dụng cho bài toán nhận diện với dữ liệu là các ảnh khuôn mặt
- Xây dựng các phương pháp thu thập dữ liệu
- Tìm hiểu ứng dụng các thuật toán phân loại trong học máy để áp dụng vào bài toán nhận diện khuôn mặt

## 0.3. Đối tượng và phạm vi nghiên cứu của đề tài

### 0.3.1. Đối tượng

- Các mô hình học sâu nổi tiếng về nhận dạng khuôn mặt được các nhà khoa học huấn luyện trước với những bộ dữ liệu cực lớn và chuẩn xác.
- Các mô hình, phương pháp phát hiện khuôn mặt trong ảnh với độ chính xác cao
- Các phương pháp làm giàu dữ liệu, đặc biệt là với dữ liệu khuôn mặt.
- Các phương pháp phân loại dữ liệu trong học máy và học sâu
- Các phương pháp đánh giá mất mát thương dùng để huấn luyện các mô hình học máy, học sâu

### 0.3.2. Phạm vi nghiên cứu

- Tập trung sử dụng bộ dữ liệu tự tạo của các sinh viên trong trường đại học Công nghiệp Hà Nội
- Huấn luyện mô hình theo phương pháp đánh giá bộ ba (triplet loss)
- Sử dụng các đánh giá cơ bản để đánh giá tính chính xác của mô hình

## **0.4. Kết cấu của đề án**

Gồm 4 chương:

Chương 1 : Nghiên cứu tổng quan

Chương 2 : Cơ sở lý thuyết

Chương 3 : Thiết kế và xây dựng hệ thống

Chương 4 : Kết luận và hướng phát triển

## **0.5. Lời cảm ơn**

Lời đầu tiên, em xin bày tỏ sự cảm ơn chân thành đối với Cô giáo, ThS. Nguyễn Thị Mỹ Bình – giáo viên hướng dẫn trực tiếp em.

Em cũng xin gửi lời cảm ơn tới các thầy cô trong khoa Công nghệ thông tin, trường Đại học Công Nghiệp Hà Nội đã hướng dẫn, chỉ bảo và tạo điều kiện cho em học tập cũng như nghiên cứu trong thời gian qua.







Cảm ơn Câu lạc bộ HIT, Đội Olympic Tin học khoa Công nghệ thông tin đã đồng hành cùng em trong suốt quãng thời gian học tập, làm việc tại trường.

Mặc dù đã cố gắng hoàn thành Báo cáo thực tập tốt nghiệp nhưng chắc chắn sẽ không tránh khỏi những sai sót, em kính mong nhận được sự thông cảm và chỉ bảo của các thầy cô và các bạn.

## 0.6. Bảng các ký hiệu

Các ký hiệu sử dụng trong sách được liệt kê trong Bảng 0.1.

Bảng 0.1: Các quy ước ký hiệu và tên gọi được sử dụng trong báo cáo

Ký hiệu	Ý nghĩa
$x, y, N, k$	in nghiêng, thường hoặc hoa, là các số vô hướng
$\mathbf{x}, \mathbf{y}$	in đậm, chữ thường, là các vector
$\mathbf{X}, \mathbf{Y}$	in đậm, chữ hoa, là các ma trận
$\mathbb{R}$	tập hợp các số thực
$\mathbb{N}$	tập hợp các số tự nhiên
$\mathbb{C}$	tập hợp các số phức
$\mathbb{R}^m$	tập hợp các vector thực có $m$ phần tử
$\mathbb{R}^{m \times n}$	tập hợp các ma trận thực có $m$ hàng, $n$ cột
$\mathbb{S}^n$	tập hợp các ma trận vuông đối xứng bậc $n$
$\mathbb{S}_+^n$	tập hợp các ma trận nửa xác định dương bậc $n$
$\mathbb{S}_{++}^n$	tập hợp các ma trận xác định dương bậc $n$
$\in$	phần tử thuộc tập hợp
$\exists$	tồn tại
$\forall$	mọi
$\triangleq$	ký hiệu là/bởi. Ví dụ $a \triangleq f(x)$ nghĩa là “ký hiệu $f(x)$ bởi $a$ ”.
$x_i$	phần tử thứ $i$ (tính từ 1) của vector $\mathbf{x}$
$\text{sgn}(x)$	hàm xác định dấu. Bằng 1 nếu $x \geq 0$ , bằng -1 nếu $x < 0$ .
$\exp(x)$	$e^x$
$\log(x)$	logarit <i>tự nhiên</i> của số thực dương $x$
$\underset{x}{\text{argmin}} f(x)$	giá trị của $x$ để hàm $f(x)$ đạt giá trị nhỏ nhất
$\underset{x}{\text{argmax}} f(x)$	giá trị của $x$ để hàm $f(x)$ đạt giá trị lớn nhất
o.w	<i>otherwise</i> – trong các trường hợp còn lại
$\frac{\partial f}{\partial x}$	đạo hàm của hàm số $f$ theo $x \in \mathbb{R}$
$\nabla_{\mathbf{x}} f$	gradient của hàm số $f$ theo $\mathbf{x}$ ( $\mathbf{x}$ là vector hoặc ma trận)
$\nabla_{\mathbf{x}}^2 f$	gradient bậc hai của hàm số $f$ theo $\mathbf{x}$ , còn được gọi là <i>Hesse</i>
$\odot$	Hadamard product (elementwise product). Phép nhân từng phần tử của hai vector hoặc ma trận cùng kích thước.
$\propto$	tỉ lệ với
	đường nét liền
	đường nét đứt
	đường nét chấm (đường chấm chấm)
	đường chấm gạch
	nền chấm
	nền sọc chéo

---

*Chương 1*

---

# Nghiên cứu tổng quan

---

## 1.1. Các phương pháp nghiên cứu

- Hiện nay có 2 phương pháp nhận diện khuôn mặt được sử dụng rộng rãi nhất là:
  - Nhận dạng dựa trên các đặc trưng của các phần tử trên khuôn mặt (Feature base face recognition)
  - Nhận dạng dựa trên xét tổng thể khuôn mặt (Appearance based face recognition)
- Ngoài ra còn một số phương pháp về loại sử dụng mô hình về khuôn mặt :
  - Nhận dạng 2D :Elastics Bunch Graph, Active Appearance Model.
  - Nhận dạng 3D :3D Morphale Model

## 1.2. Ưu nhược điểm của các phương pháp

### 1.2.1. Nhận dạng dựa trên các đặc trưng của các phần tử trên khuôn mặt:

Đây là phương pháp nhận dạng khuôn mặt dựa trên việc xác định các đặc trưng hình học của các chi tiết trên một khuôn mặt (vị trí, diện tích, hình dạng của mắt, mũi, miệng, ...) và mối quan hệ giữa chúng (khoảng cách của hai mắt, khoảng cách của hai lông mày, ...).

Ưu điểm của phương pháp này là nó gần với cách mà con người sử dụng để nhận biết khuôn mặt. Hơn nữa với việc xác định đặc tính cả mối quan hệ, phương pháp này có thể cho kết quả tốt trong các trường hợp ảnh có nhiều nhiễu như bị nghiêng, bị xoay hoặc ánh sáng thay đổi.

Nhược điểm của phương pháp này là cài đặt thuật toán phức tạp do việc xác định mối quan hệ giữa các đặc tính sẽ khó phân biệt. Mặt khác, với các ảnh kích thước bé thì các đặc tính sẽ khó phân biệt.

### **1.2.2. Nhận dạng dựa trên xét tổng thể khuôn mặt:**

Đây là phương pháp xem mỗi ảnh có kích thước  $R \times C$  là một vector trong không gian  $R \times C$  chiều. Ta sẽ xây dựng một không gian mới có chiều nhỏ hơn sao cho khi biểu diễn trong không gian có các đặc điểm chính của một khuôn mặt không bị mất đi. Trong không gian đó, các ảnh cùng một người sẽ được tập trung lại một nhóm gần nhau và cách xa các nhóm khác.

Ưu điểm của phương pháp này là tìm được các đặc tính tiêu biểu của đối tượng cần nhận dạng mà không cần phải xác định các thành phần và mối quan hệ giữa các thành phần đó. Phương pháp sử dụng thuật toán có thể thực hiện tốt với các ảnh có độ phân giải cao, thu gọn ảnh thành một ảnh có kích thước nhỏ hơn. Có thể kết hợp các phương pháp khác như mạng Nơ-ron, Support Vector Machine.

Nhược điểm của phương pháp này phân loại theo chiều phân bố lớn nhất của vector. Tuy nhiên, chiều phân bố lớn nhất không phải lúc nào cũng mang lại hiệu quả tốt nhất cho bài toán nhận dạng và đặc biệt là phương pháp này rất nhạy với nhiễu.

### **1.2.3. Kết luận**

Vì kết quả nghiên cứu cuối cùng là ứng dụng với yêu cầu về độ chính xác cao, khả năng thích ứng linh hoạt, hoạt động ổn định trong môi trường thực tế và hoạt động với các camera với độ phân giải thấp. Tôi quyết định chọn phương pháp nhận dạng dựa trên xét tổng thể khuôn mặt (Appearance based face recognition).



---

*Chương 2*

---

# Cơ sở lý thuyết

---

## 2.1. Tổng quan về nhận diện khuôn mặt

Nhận diện khuôn mặt (Face recognition) đang được ứng dụng trong nhiều lĩnh vực. Hệ thống nhận dạng khuôn mặt là một ứng dụng cho phép máy tính tự động xác định hoặc nhận dạng một người nào đó từ một bức hình ảnh kỹ thuật số hoặc một khung hình.

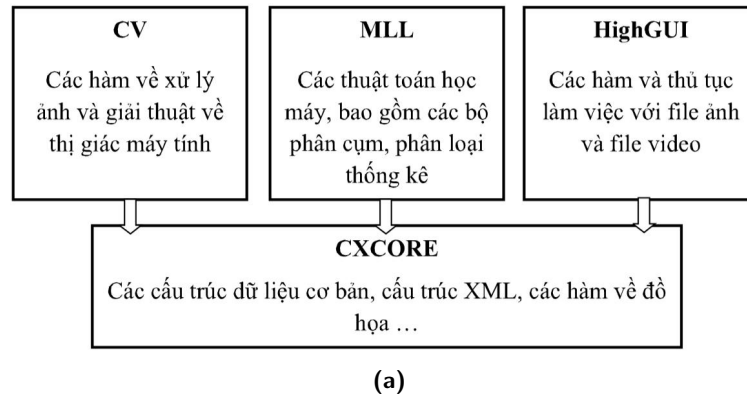
Nhận diện khuôn mặt là một bài toán phức tạp, đòi hỏi cần phải xử lý một loạt các vấn đề .

Mỗi khuôn mặt đều có những điểm mốc, những phần lồi lõm, hình dáng của các bộ phận trên khuôn mặt như mắt, mũi, miệng,... Các hệ thống nhận diện định nghĩa những điểm này là những điểm nút, và mỗi khuôn mặt có khoảng 80 nút như thế

## 2.2. Tìm hiểu về OpenCV và ngôn ngữ lập trình Python

OpenCV (Open Source Computer Vision Library) là một thư viện các chức năng lập trình chủ yếu nhằm vào tầm nhìn máy tính thời gian thực. OpenCV hỗ trợ nhiều ngôn ngữ lập trình như C++, Python, Java,... và có sẵn trên các nền tảng khác nhau bao gồm Windows, Linux, Mac OS, Android và iOS. Các giao diện cho các hoạt động GPU tốc độ cao dựa trên CUDA và OpenCL cũng đang được phát triển tích cực.

Cấu trúc tổng quan của OpenCV bao gồm 5 phần chính. 4 trong 5 phần đó được chỉ ra trong hình vẽ dưới.



**Hình 2.1.** Cấu trúc các phần của OpenCV.

Phần CV bao gồm các thư viện cơ bản về xử lý ảnh và các giải thuật về xử lý ảnh. MLL là bộ thư viện về các thuật toán học máy, bao gồm rất nhiều bộ phân cụm và phân loại thống kê. HighGUI chứa đựng những thủ tục vào ra, các chức năng về lưu trữ cũng như đọc các file ảnh và video. Phần thứ 4, Cxcore chứa đựng các cấu trúc dữ liệu cơ bản (ví dụ như cấu trúc XML, các cây dữ liệu ...). Phần cuối cùng là CvAux, phần này bao gồm các thư viện cho việc phát hiện, theo dõi và nhận dạng đối tượng (khuôn mặt, mắt ...).

OpenCV - Python là một thư viện các ràng buộc Python được thiết kế để giải quyết các vấn đề về xử lý ảnh và thị giác máy tính.

Python là ngôn ngữ lập trình có mục đích chung được bắt đầu bởi Guido van Rossum, nó trở nên rất phổ biến rất nhanh trong thời gian gần đây, chủ yếu vì tính đơn giản và khả năng đọc mã của nó. Nó cho phép lập trình viên thể hiện ý tưởng trong ít dòng mã hơn mà không làm giảm khả năng đọc.

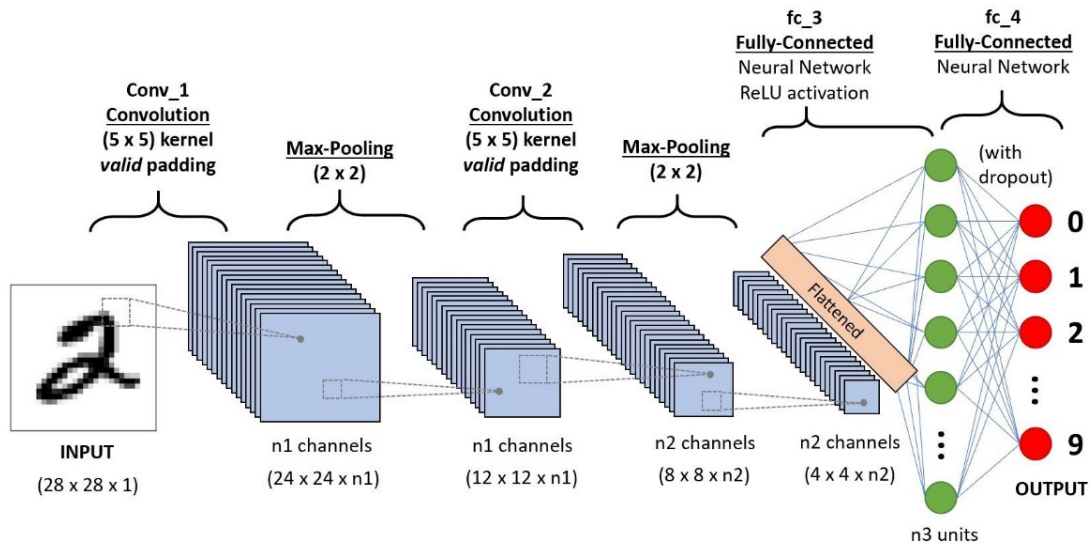
So với các ngôn ngữ như C/C++, Python chậm hơn. Điều đó nói rằng, Python có thể dễ dàng được mở rộng với C/C++, cho phép chúng ta viết mã chuyên sâu tính toán trong C/C++ và tạo các trình bao bọc Python có thể được sử dụng làm mô-đun Python. Điều này mang lại cho chúng ta hai lợi thế: thứ nhất, mã nhanh như mã C/C++ gốc (vì đây là mã C++ thực tế hoạt động ở chế độ nền) và thứ hai, mã dễ dàng hơn trong Python so với C/C++. OpenCV - Python là một trình bao bọc Python để thực hiện OpenCV C++ ban đầu.

OpenCV - Python sử dụng Numpy, một thư viện được tối ưu hóa cao cho các hoạt động số với cú pháp kiểu MATLAB. Tất cả các cấu trúc mảng OpenCV được chuyển đổi sang và từ các mảng Numpy. Điều này cũng giúp tích hợp dễ dàng hơn với các thư viện khác sử dụng Numpy như SciPy và Matplotlib.

## 2.3. Mô hình mạng neural tích chập (CNN - Convolutional neural network)

Mạng neural tích chập (CNN) là một trong những mô hình học sâu tiên tiến phổ biến nhất và có ảnh hưởng nhất với cộng đồng thị giác máy tính (Computer vision). CNN thường được dùng trong các bài toán nhận dạng ảnh, phân tích ảnh, xử lý ngôn ngữ tự nhiên dưới dạng ảnh các bước sóng. Và hầu hết đều cho hiệu quả tốt đến rất tốt.

CNN là một kiến trúc mạng neural sinh ra để xử lý các dữ liệu phi cấu trúc dạng ảnh. Có 2 loại lớp chính trong CNN : lớp tích chập (Convolutional layer) và lớp gộp (Pooling layer)



(a)

Hình 2.2. CNN cho bài toán nhận diện chữ số.

### 2.3.1. Lớp tích chập

Lớp tích chập là lớp quan trọng nhất và thường cũng là lớp đầu tiên của của mô hình CNN. Lớp này có chức năng chính là phát hiện các đặc trưng có tính không gian hiệu quả. Trong tầng này có 4 đối tượng chính là: ma trận đầu vào, bộ lọc (filters) và trường thụ cảm, bản đồ đặc trưng (feature map). Lớp tích chập nhận đầu vào là một ma trận 3 chiều và một bộ lọc cần phải học. Bộ lọc này sẽ trượt qua từng vị trí trên bức ảnh để tính tích chập (convolution) giữa bộ lọc và phần tương ứng trên bức ảnh. Phần tương ứng này trên bức ảnh gọi là trường thực cảm (receptive field), tức là vùng mà một neuron có thể nhìn thấy để đưa

ra quyết định, và mà trận cho ra bởi quá trình này được gọi là bản đồ đặc trưng (feature map).

Để hình dung, có thể tưởng tượng, bộ filters giống như các tháp canh trong nhà tù quét lần lượt qua không gian xung quanh để tìm kiếm tên tù nhân bỏ trốn. Khi phát hiện tên tù nhân bỏ trốn, thì chuông báo động sẽ reo lên, giống như các bộ lọc tìm kiếm được đặc trưng nhất định thì tích chập đó sẽ cho giá trị tương ứng.

- a. Lớp tích chập được coi như xác định đặc trưng
- b. Các tham số :

## **2.4. Máy dò khuôn mặt (Face detector)**

## **2.5. Các kĩ thuật làm giàu dữ liệu (Data agumentation)**

## **2.6. Mô hình học sâu được huấn luyện trước (Pre-train model)**

### **2.6.1. Sử dụng mô hình được huấn luyện trước**

### **2.6.2. Giới thiệu Facenet**

### **2.6.3. Giới thiệu Mạng InceptionResnetV1**

### **2.6.4. Kĩ thuật đánh giá bộ ba (Triplet loss)**

---

*Chương 3*

---

# Thiết kế và xây dựng hệ thống

---

## 3.1. Phân tích

Về cơ bản một hệ thống điểm danh bằng khuôn mặt gồm các bước sau:

- Thu thập dữ liệu khuôn mặt
- Phát hiện khuôn mặt dựa trên ảnh đầu vào và gán nhãn dữ liệu
- Làm giàu dữ liệu
- Trích xuất các đặc trưng (sử dụng học sâu)
- Đưa các đặc trưng đã được gán nhãn vào thuật toán phân loại
- Lưu trữ các thông tin và kết quả phân loại đã được học
- Nhận dạng khuôn mặt

## 3.2. Xây dựng

---

*Chương 4*

---

# Kết luận và hướng phát triển

---

## 4.1. Kết luận

Trên cơ sở tìm hiểu về bài toán nhận diện mặt người trong ảnh, sử dụng pre-trained model FaceNet được huấn luyện trước trên base-net là InceptionResnetV1 do tiến sĩ khoa học máy tính David Sandberg cung cấp, tôi đã xây dựng thành công hệ thống điểm danh thông qua hình ảnh khuôn mặt.

Về khả năng phát hiện khuôn mặt, kết quả phát hiện khá tốt hầu hết các trường hợp, kể cả trong điều kiện thiếu sáng, góc nghiêng, hay có vật che khuất như kính mắt,...

Về khả năng nhận dạng, hệ thống đạt kết quả từ 80-95% đối với các khuôn mặt thẳng và điều kiện ánh sáng thích hợp, đạt 70-85% đối với các khuôn mặt nghiêng hoặc thiếu sáng.

---

## Tài liệu tham khảo

---

# Index

dimensionality reduction – giảm chiều dữ liệu, 2	giảm chiều dữ liệu – dimensionality reduction, 2
feature extraction – trích chọn đặc trưng, 2	lựa chọn đặc trưng – feature selection, 2
feature selection – lựa chọn đặc trưng, 2	trích chọn đặc trưng – feature extraction, 2