

# NHẬN DIỆN KHUÔN MẶT

Lê Tiến Quyết<sup>1, 2, 3, 4</sup>, Nguyễn Công Nguyên<sup>1, 2, 3, 5</sup>

<sup>1</sup>Đại học Quốc gia Thành phố Hồ Chí Minh

<sup>2</sup>Khoa Khoa học máy tính, Trường Đại học Công nghệ Thông tin

<sup>3</sup>Thị giác máy tính nâng cao - CS331.P11

{<sup>4</sup>21520428, <sup>5</sup>21521200}@gm.uit.edu.vn

## 1 Thông tin thành viên

Bảng 1: Thông tin thành viên và đóng góp vào đồ án

STT	Họ và tên	MSSV	Đóng góp	Đánh giá
1	Lê Tiến Quyết	21520428	100%	10/10
2	Nguyễn Công Nguyên	21521200	100%	10/10

## 2 Lý Do Chọn Bài Toán

### 2.1 Tầm Quan Trọng

Nhận diện khuôn mặt là một trong những lĩnh vực cốt lõi của trí tuệ nhân tạo (AI) và đang ngày càng đóng vai trò thiết yếu trong nhiều ứng dụng thực tiễn. Công nghệ này không chỉ góp phần nâng cao mức độ bảo mật mà còn cải thiện đáng kể trải nghiệm người dùng. Trong bối cảnh an ninh ngày càng trở thành ưu tiên hàng đầu, nhận diện khuôn mặt cung cấp giải pháp xác thực danh tính, giúp tăng cường an toàn và hiệu quả.

Bên cạnh đó, công nghệ nhận diện khuôn mặt còn mở ra những cơ hội đột phá trong lĩnh vực thương mại. Từ việc cá nhân hóa các chiến lược marketing để tiếp cận khách hàng hiệu quả hơn, đến tối ưu hóa trải nghiệm dịch vụ, công nghệ này đang định hình tương lai của các ngành công nghiệp và tạo ra giá trị vượt trội cho doanh nghiệp.

### 2.2 Ứng Dụng

Công nghệ nhận diện khuôn mặt hiện nay đang được ứng dụng rộng rãi trong nhiều lĩnh vực, với những tiềm năng vượt trội như sau:

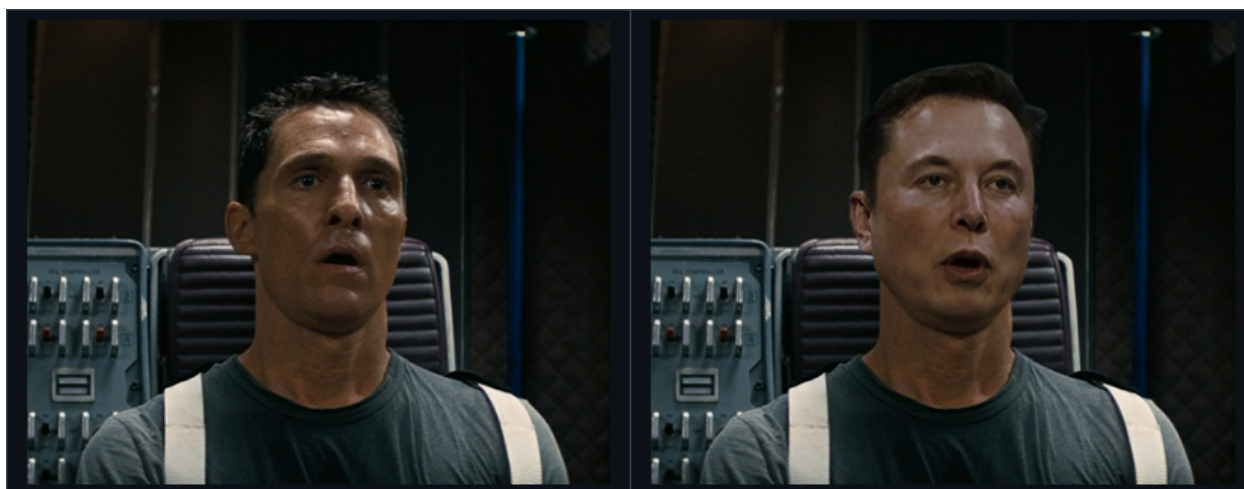
- Kiểm soát truy cập:** Nhận diện khuôn mặt được sử dụng để xác thực danh tính tại các tòa nhà, khu vực hạn chế, thiết bị cá nhân (như điện thoại thông minh) và các hệ thống quản lý doanh nghiệp. Đây là một trong những ứng dụng phổ biến nhất, đặc biệt trong các hệ thống bảo mật và an ninh, giúp giảm thiểu nguy cơ truy cập trái phép và tăng cường an toàn.
- Giám sát và an ninh công cộng:** Nhận diện khuôn mặt được tích hợp vào các hệ thống camera giám sát để phát hiện và ngăn chặn các hành vi đáng ngờ hoặc nguy hiểm trong thời gian thực. Công nghệ này đặc biệt hữu ích trong việc truy vết tội phạm, quản lý giao thông và đảm bảo an ninh tại các sự kiện đông người.
- Thương mại và bán lẻ:** Các cửa hàng sử dụng nhận diện khuôn mặt để nhận biết khách hàng thân thiết, từ đó đề xuất các sản phẩm hoặc dịch vụ phù hợp. Ngoài ra, kết hợp nhận diện khuôn mặt với hệ thống thanh toán tự động, giúp cải thiện tốc độ và sự tiện lợi trong giao dịch.

Những ứng dụng đa dạng này không chỉ chứng minh vai trò thiết yếu của công nghệ nhận diện khuôn mặt trong cuộc sống hiện đại mà còn mở ra tiềm năng phát triển mạnh mẽ trong tương lai. Công nghệ này hứa hẹn tiếp tục đóng góp quan trọng vào các lĩnh vực như bảo mật, tiện ích cá nhân, và quản lý xã hội.

## 2.3 Thách Thức

Mặc dù công nghệ nhận diện khuôn mặt mang lại nhiều lợi ích, nhưng việc triển khai nó vẫn phải đối mặt với những thách thức lớn:

- **Sự đa dạng về góc nhìn và điều kiện ánh sáng:** Hệ thống nhận diện khuôn mặt thường gặp khó khăn trong việc xử lý các tình huống thực tế, nơi góc nhìn khuôn mặt không đồng nhất hoặc điều kiện ánh sáng không ổn định, chẳng hạn như ánh sáng yếu, ánh sáng gắt hoặc bóng tối. Sự thay đổi này có thể làm giảm độ chính xác của hệ thống, đặc biệt trong các môi trường ngoài trời hoặc nơi công cộng. Giải quyết vấn đề này đòi hỏi sự cải tiến trong các thuật toán và khả năng học sâu của mô hình AI.
- **Giả mạo khuôn mặt:** Một trong những rủi ro lớn nhất đối với nhận diện khuôn mặt là khả năng bị lừa bởi các hình ảnh, video hoặc công nghệ giả lập như deepfake. Các hành vi như sử dụng ảnh chụp, video hoặc mặt nạ 3D có thể đánh bại hệ thống nhận diện, gây nguy cơ mất an toàn trong các ứng dụng bảo mật. Điều này đòi hỏi sự phát triển của các cơ chế phát hiện giả mạo tiên tiến để bảo vệ hệ thống.



Hình 1: Sử dụng Deep Fake để thay đổi khuôn mặt (Ảnh thật bên trái, ảnh giả bên phải) [1]

Những thách thức này không chỉ là trở ngại cần vượt qua mà còn là động lực thúc đẩy các nghiên cứu và phát triển trong lĩnh vực nhận diện khuôn mặt, nhằm mang lại những giải pháp hiệu quả và an toàn hơn trong tương lai.

## 2.4 Mục Tiêu

Mục tiêu của đồ án là xây dựng và phát triển một hệ thống nhận diện khuôn mặt có độ chính xác cao, đáp ứng hiệu quả các yêu cầu trong điều kiện thực tế. Hệ thống cần có khả năng hoạt động ổn định trong nhiều môi trường khác nhau, bao gồm cả những tình huống ánh sáng yếu, ánh sáng mạnh không đồng đều hoặc góc nhìn đa dạng.

Ngoài ra, hệ thống sẽ được thiết kế với tính ứng dụng cao, dễ dàng tích hợp vào các lĩnh vực như kiểm soát truy cập, giám sát an ninh hoặc hỗ trợ trải nghiệm người dùng trong các dịch vụ thông minh.

## 3 Phát biểu bài toán

### Input:

- Một bộ dữ liệu  $D$  gồm  $N$  ảnh, mỗi ảnh chỉ chứa một khuôn mặt duy nhất đã được xác định danh tính:  $D = \{(I_1, y_1), (I_2, y_2), \dots, (I_N, y_N)\}$ . Với  $y_i$  là danh tính của mặt người trong ảnh  $I_i$  ( $1 \leq i \leq N$ ).
- Một ảnh  $I_x$  chứa một khuôn mặt duy nhất cần được xác định danh tính.

### Output:

- $\hat{y}_x$ : Danh tính dự đoán của khuôn mặt trong ảnh  $I_x$  được xác định bằng cách chọn ảnh tham chiếu có độ tương đồng lớn nhất với  $I_x$  trong bộ dữ liệu tham chiếu. Nếu độ tương đồng lớn nhất này vượt ngưỡng cho trước, trả về danh tính của ảnh tham chiếu đó; ngược lại trả về "Unknown".

## 4 Phương pháp thực hiện

### 4.1 Phương pháp

Mỗi cặp ảnh đầu vào, sau khi được tiền xử lý sẽ được đưa qua Backbone để trích xuất đặc trưng (gồm 512 giá trị cho mỗi đặc trưng), *Feature 1* ( $\mathbf{X}_1$ ) và *Feature 2* ( $\mathbf{X}_2$ ) như trong Hình [2, 3], .

Tuy nhiên, tùy vào hướng tiếp cận, cách khai thác và sử dụng các đặc trưng này sẽ có sự khác biệt đôi chút.

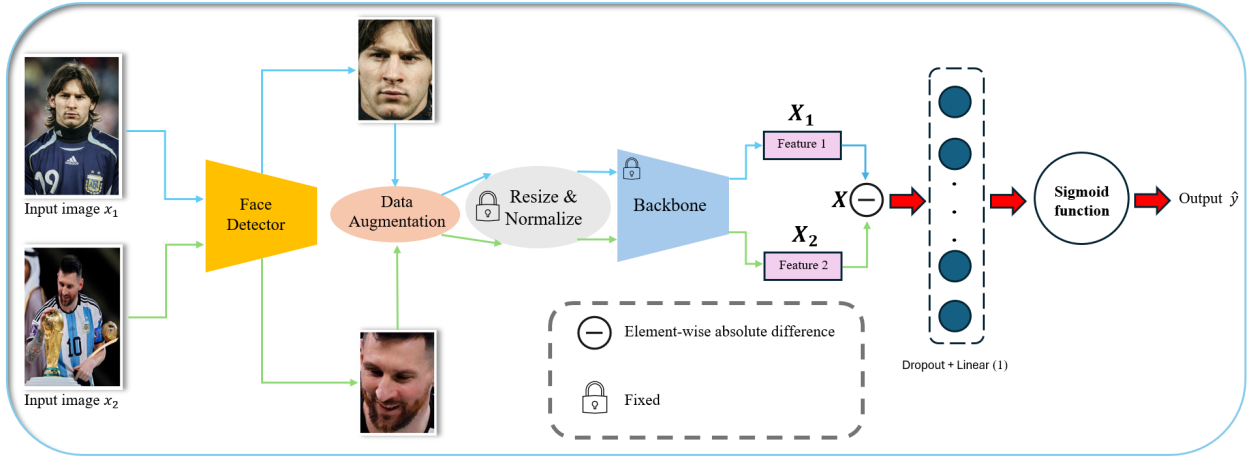
#### 4.1.1 Hướng tiếp cận (1)

**Hướng tiếp cận (1):** Tính xác suất hai ảnh đầu vào cùng thuộc một người [2].

**Element-wise absolute difference:**

Phép toán được áp dụng trên hai mảng hoặc ma trận có cùng kích thước, trong đó mỗi phần tử của kết quả được tính bằng giá trị tuyệt đối của hiệu số giữa hai phần tử tương ứng từ hai mảng đầu vào. Công thức được biểu diễn như sau:

$$\mathbf{X} = |\mathbf{X}_1 - \mathbf{X}_2| = |X_{11} - X_{21}, \dots, X_{1i} - X_{2i}|, \text{ với } 1 \leq i \leq 512 \quad (1)$$



Hình 2: Tổng quan Hướng tiếp cận (1)

Sau khi tính toán chênh lệch  $\mathbf{X}$  giữa hai đặc trưng  $\mathbf{X}_1$  và  $\mathbf{X}_2$ , kết quả được đưa qua một lớp Dropout và tiếp theo qua một lớp Fully Connected có kích thước đầu ra là một giá trị duy nhất  $z$ , được tính theo công thức:

$$z = \mathbf{W}^T \times \mathbf{X} \quad (2)$$

Trong đó:

- $\mathbf{W}$  đại diện trọng số của lớp Fully Connected, có kích thước bằng kích thước của  $X$  (bằng 512).

**Dự đoán:**

Đầu ra  $z$  từ lớp Fully Connected được chuyển đổi thành một giá trị xác suất bằng cách sử dụng hàm Sigmoid:

$$\hat{y} = \text{Sigmoid}(z) = \frac{1}{1 + e^{-z}} \quad (3)$$

Hàm Sigmoid nén giá trị  $z$  vào khoảng  $[0, 1]$ , giúp biểu diễn xác suất mà cặp ảnh đầu vào thuộc cùng một người.

**Hàm mất mát (Loss Function):**

Để tối ưu hóa mô hình, sử dụng **Binary Cross-Entropy Loss (BCE Loss)** để đo lường sự khác biệt giữa xác suất dự đoán  $\hat{y}$  và nhãn thực tế  $y$  (với  $y = 1$  biểu thị cùng một người và  $y = 0$  biểu thị khác người). Hàm mất mát được định nghĩa như sau:

$$\mathbb{L}_{BCE}(y, \hat{y}) = -\frac{1}{N} \sum_{i=1}^N [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)] \quad (4)$$

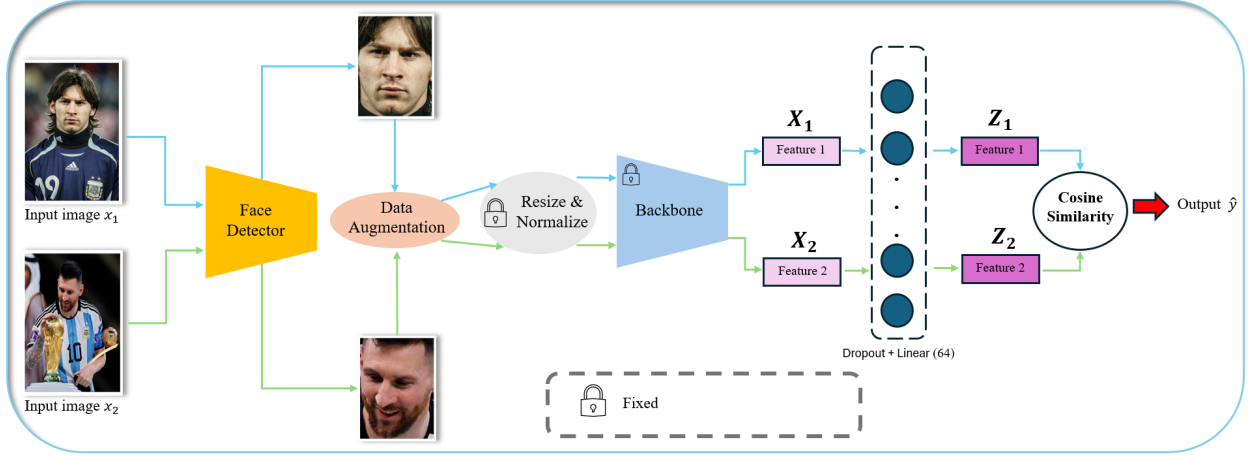
Trong đó:

- $N$ : Số lượng mẫu trong tập dữ liệu.
- $y_i$ : Nhân thực tế của mẫu thứ  $i$
- $\hat{y}_i$ : Xác suất dự đoán của mẫu thứ  $i$

#### 4.1.2 Hướng tiếp cận (2)

**Hướng tiếp cận (2):** Tính mức độ tương đồng giữa hai ảnh đầu vào.

Thay vì áp dụng **Element-wise absolute difference** như trong Hướng tiếp cận (1), các đặc trưng  $X_1$  và  $X_2$  sẽ được xử lý qua một lớp Dropout và một lớp Fully Connect với đầu ra có kích thước 64, kết quả thu được các đặc trưng mới tương ứng lần lượt là  $Z_1$  và  $Z_2$ .



Hình 3: Tổng quan Hướng tiếp cận (2)

##### Dự đoán:

Để dự đoán mức độ tương đồng giữa hai ảnh đầu vào, các đặc trưng  $Z_1$  và  $Z_2$  được sử dụng để tính giá trị dự đoán  $\hat{y}$  bằng cách sử dụng Cosine Similarity. Phép đo này sẽ phản ánh mức độ liên quan giữa hai đặc trưng, giúp xác định mối quan hệ giữa hai ảnh:

$$\hat{y} = \text{Cosine\_Similarity}(Z_1, Z_2) = \frac{Z_1 \cdot Z_2}{\|Z_1\| \|Z_2\|} = \frac{\sum_{i=1}^{64} Z_{1i} \cdot Z_{2i}}{\sqrt{\sum_{i=1}^{64} Z_{1i}^2} \sqrt{\sum_{i=1}^{64} Z_{2i}^2}} \quad (5)$$

##### Hàm mất mát (Loss Function):

Để tối ưu hóa mô hình, sử dụng **Cosine Loss** để đo lường sự khác biệt giữa giá trị tương đồng dự đoán  $\hat{y}$  và nhân thực  $y$  (với  $y = 1$  biểu thị cùng một người và  $y = -1$  biểu thị khác người).

Phương trình [6] thể hiện Cosine Loss:

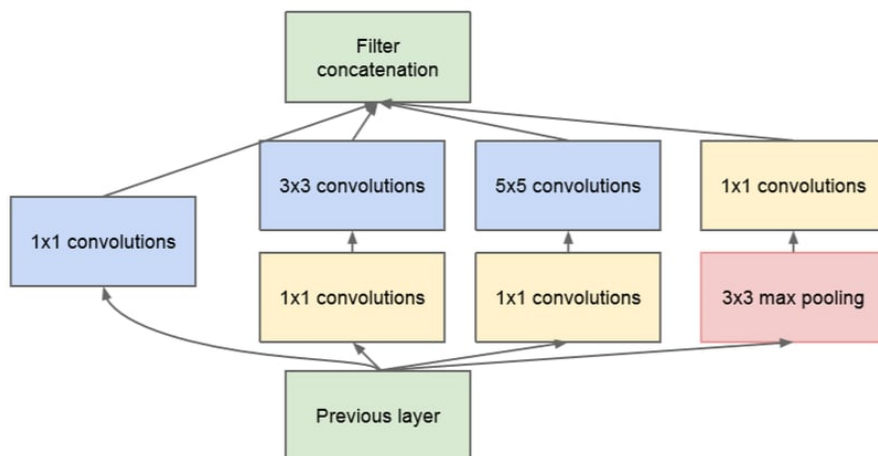
$$\mathbb{L}_{\text{Cosine}} = \sum_1^N \left[ \frac{1+y}{2} \cdot (1-\hat{y}) + \frac{1-y}{2} \cdot \max(0, \hat{y}) \right] \quad (6)$$

## 4.2 Backbone

Backbone được sử dụng trong đồ án là mô hình mạng học sâu Inception-ResNet (v1) [3] đã được pretrained trên bộ dữ liệu VGGFace2 [4].

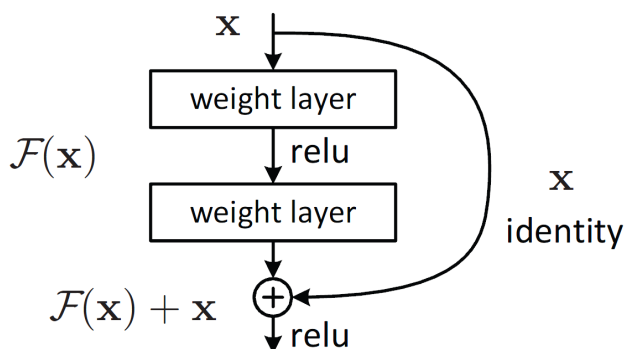
Inception-Resnet là một kiến trúc mạng kết hợp các đặc điểm nổi bật của InceptionNet [5] và ResNet [6], nhằm cải thiện hiệu suất và khả năng học sâu của mô hình.

- InceptionNet tận dụng phương pháp xử lý song song nhiều kernel có kích thước khác nhau ( $1 \times 1$ ,  $3 \times 3$ ,  $5 \times 5$ ), sau đó tổng hợp các đầu ra nhằm khai thác thông tin ở nhiều phạm vi không gian khác nhau. Ngoài ra, mô hình còn sử dụng các lớp tích chập  $1 \times 1$  để giảm số lượng kênh trước khi thực hiện các phép tích chập kích thước lớn hơn. Điều này giúp giảm đáng kể số lượng tham số của mạng.



Hình 4: Inception Module [5]

- ResNet giải quyết vấn đề vanishing gradient, một trở ngại phổ biến khi huấn luyện các mạng học sâu, bằng cách sử dụng các kết nối phân dư (residual connections).



Hình 5: Residual learning: a building block. [6]

Kiến trúc Inception-ResNet kế thừa ưu điểm của cả hai mô hình trên [5, 6], vừa duy trì khả năng xử lý linh hoạt của InceptionNet, vừa đảm bảo độ sâu và tính ổn định trong huấn luyện nhờ ResNet.

### 4.3 Tiền xử lý

Trước khi đưa vào Backbone để trích xuất đặc trưng, các hình ảnh đầu vào sẽ được tiền xử lý nhằm tối ưu hóa khả năng học của mô hình.

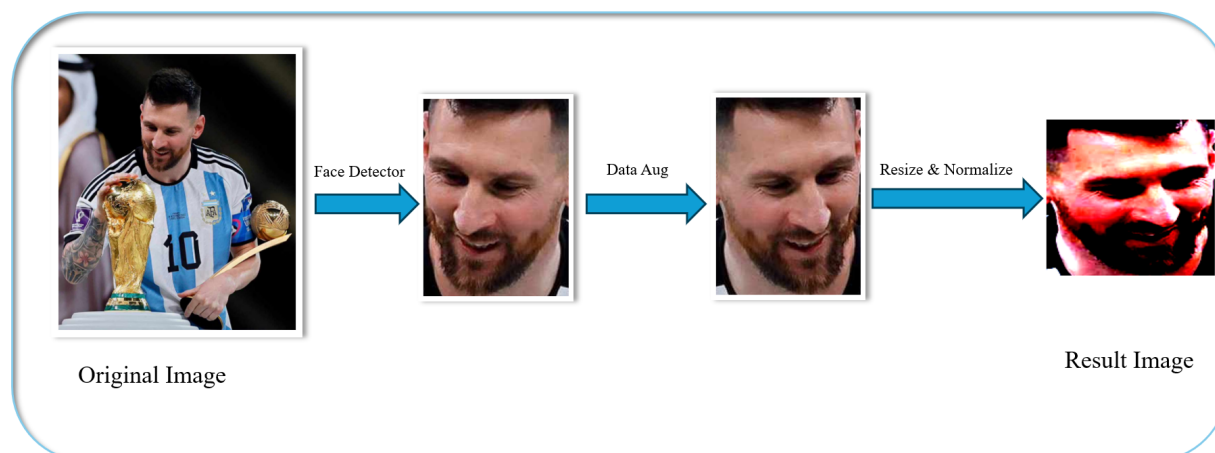
Cụ thể, Face Detector sẽ được sử dụng để xác định và cắt vùng chứa khuôn mặt trong ảnh. Đối với giai đoạn huấn luyện, vùng khuôn mặt được cắt ra sẽ được áp dụng một số phương pháp Data Augmentation nhằm cải thiện tính đa dạng của tập dữ liệu và khả năng tổng quát hóa của mô hình. Cuối cùng, ảnh sẽ được resize và chuẩn hóa về phân phối chuẩn để đảm bảo ổn định trong quá trình huấn luyện.

#### 4.3.1 Face Detector

Face Detector trong đồ án được sử dụng là YOLOv5 (cụ thể là YOLOv5n). Mục tiêu của việc áp dụng YOLOv5n là xác định chính xác vị trí khuôn mặt xuất hiện trong ảnh, giúp mô hình tập trung học các đặc trưng của khuôn mặt thay vì bị phân tán bởi các yếu tố không gian xung quanh.

YOLOv5n được lựa chọn không chỉ vì khả năng phát hiện khuôn mặt chính xác ổn định, mà còn nhờ tốc độ xử lý nhanh, phù hợp với yêu cầu xử lý thời gian thực mà đồ án hướng đến.

Để có khả năng phát hiện khuôn mặt, YOLOv5n đã được fine-tuning trên bộ dữ liệu Face-Detection-Dataset [7]. Bộ



Hình 6: Toàn bộ quá trình tiền xử lý ảnh

dữ liệu này gồm 13400 mẫu dành cho huấn luyện và 3347 mẫu dùng để đánh giá.

Về cài đặt huấn luyện cho YOLOv5n:  $epoch = 30$ ,  $batch\ size = 256$  và các ảnh đều được resize về kích thước (640, 640). Kết quả sau khi huấn luyện  $mAP50 = 0.875$  đủ để thực hiện việc phát hiện khuôn mặt trong ảnh.

### 4.3.2 Data Augmentation

Các phương pháp tăng cường dữ liệu được sử dụng trong đồ án:

- **Tăng, giảm sáng:** Ảnh sẽ được điều chỉnh độ sáng dựa vào việc sinh một giá trị ngẫu nhiên  $factor \sim \text{Uniform}(0.5, 1.5)$ . Nếu  $factor = 1$  thì độ sáng ảnh giữ nguyên,  $factor > 1$  thì ảnh sẽ sáng lên và ảnh sẽ tối đi nếu  $factor < 1$ .
- **Lật ảnh:** Ảnh sẽ được lật theo chiều ngang với xác suất 0.5.

Các kỹ thuật trên giúp tăng khả năng tổng quát hóa của mô hình, đặc biệt là giúp mô hình linh hoạt trước các biến đổi hình ảnh trong môi trường thực tế.

## 5 Kết quả thực nghiệm

### 5.1 Bộ dữ liệu

#### 5.1.1 LFW Dataset

**Labeled Faces in the Wild (LFW)** [8] là một bộ dữ liệu hình ảnh khuôn mặt nổi tiếng, được sử dụng rộng rãi trong nghiên cứu về nhận diện và xác thực danh tính. Bộ dữ liệu này bao gồm các bức ảnh của nhiều cá nhân nổi tiếng, được thu thập từ các nguồn trực tuyến, phản ánh sự đa dạng trong điều kiện môi trường và góc chụp.

LFW chứa hơn 13000 bức ảnh màu đại diện cho 5749 cá nhân khác nhau với nhiều góc nhìn, điều kiện ánh sáng và biểu cảm khuôn mặt.

LFW được thiết kế nhằm mô phỏng các tình huống thực tế, nơi hình ảnh khuôn mặt không được chuẩn bị sẵn trong phòng thu mà phản ánh đặc điểm của môi trường tự nhiên.

**Mục đích sử dụng:** Bộ dữ liệu LFW được thiết kế chủ yếu để đánh giá và thử nghiệm các phương pháp nhận diện khuôn mặt.

#### 5.1.2 Celebrity Face Image Dataset

**Celebrity Face Image Dataset** [9] là một bộ dữ liệu được cung cấp bởi tác giả Vishesh Thakur trên nền tảng Kaggle. Bộ dữ liệu này bao gồm:

- 1800 ảnh khuôn mặt của 17 diễn viên Hollywood nổi tiếng, chẳng hạn như Leonardo DiCaprio, Johnny Depp, Tom Cruise, ...



Hình 7: Hình ảnh trong Celebirty Face Image Dataset

**Mục đích sử dụng:** Celebrity Face Image Dataset được sử dụng để đánh giá hiệu suất của các mô hình nhận diện khuôn mặt trên các ảnh của những người nổi tiếng, cung cấp một bối cảnh thực tế để kiểm tra các thuật toán.

## 5.2 Độ đo đánh giá

### 5.2.1 Accuracy

**Accuracy:** Đo lường tỷ lệ cặp ảnh được phân loại chính xác (cùng danh tính hoặc khác danh tính) trên tổng số cặp ảnh đầu vào.

$$Accuracy = \frac{\sum_{i=1}^N \mathbb{K}(\hat{y}_i = y_i)}{N} \quad (7)$$

Trong đó:

- $N$ : Tổng số cặp ảnh đầu vào  $(x_{i1}, x_{i2})$ .
- $\hat{y}_i$ : Nhân dự đoán của mô hình cho cặp ảnh  $(x_{i1}, x_{i2})$ .
- $\mathbb{K}(\hat{y}_i = y_i)$ : Hàm chỉ thị, trả về 1 nếu  $\hat{y}_i = y_i$ , ngược lại trả về 0.

### 5.2.2 1-shot Accuracy

Với mỗi ảnh truy vấn  $q_i$  thì danh tính dự đoán của ảnh thỏa mãn:

$$\hat{y}_i = \arg \max_{c \in \{1, 2, \dots, C\}} \text{simi}(q_i, s_c) \quad (8)$$

**1-shot Accuracy:** Đo lường tỷ lệ dự đoán đúng của mô hình khi chỉ có một ảnh tham chiếu cho mỗi danh tính.

$$1 - \text{shot Accuracy} = \frac{\sum_{i=1}^N \mathbb{K}(\hat{y}_i = y_i)}{N} \quad (9)$$

Trong đó:

- $S = \{s_1, s_2, \dots, s_C\}$ : Tập các ảnh tham chiếu, mỗi ảnh  $s_c$  đại diện duy nhất cho danh tính  $c$ .
- $Q = \{q_1, q_2, \dots, q_N\}$ : Tập các ảnh cần truy vấn, gồm  $N$  ảnh.
- $\text{simi}(q_i, s_c)$ : Hàm mức độ tương đồng giữa ảnh truy vấn  $q_i$  và ảnh tham chiếu  $s_c$ .

## 5.3 Cài đặt & Kết quả

Mô hình được huấn luyện trên tập dữ liệu LFW với số lượng *epoch* tối đa là 100, đồng thời áp dụng kỹ thuật Early Stopping với giá trị *patience* = 10 để ngăn chặn hiện tượng overfitting.

Kích thước ảnh trước khi qua Backbone sẽ được resize về (160,160). Quá trình huấn luyện sử dụng *batch size* là 256



và tối ưu các tham số bằng thuật toán Adam với *learning rate* được thiết lập là 0.1.

Nhóm thực hiện các thực nghiệm chi tiết để đánh giá hiệu quả của YOLOv5n cũng như các kỹ thuật Data Augmentation. Ngoài ra, điều chỉnh tham số Dropout Rate  $\in \{0.5, 0.75\}$ . Tổng cộng, 8 thực nghiệm khác nhau đã được tiến hành cho mỗi hướng tiếp cận, hai thực nghiệm có kết quả tốt nhất trên tập kiểm thử sẽ được lựa chọn để đánh giá cuối cùng.

Kết quả đạt được được thể hiện ở Bảng [2, 3].

Bảng 2: Accuracy trên bộ dữ liệu LFW (%)

	Method	Dropout Rate	Data Augmentation	Accuracy
Hướng tiếp cận (1)	Inception-Resnet (v1)	0.75	No	92.17
	YOLOv5n + Inception-Resnet (v1)	0.5	No	<b>93.10</b>
Hướng tiếp cận (2)	YOLOv5n + Inception-Resnet (v1)	0.5	No	91.33
	YOLOv5n + Inception-Resnet (v1)	0.75	No	91.60

Đối với tiêu chí đánh giá bằng 1-shot Accuracy, mỗi danh tính yêu cầu một ảnh tham chiếu duy nhất. Tuy nhiên, trong phạm vi đồ án này, các ảnh tham chiếu được chọn ngẫu nhiên, điều này có thể chưa phản ánh đầy đủ tiềm năng thực sự của mô hình.

Bảng 3: 1-shot Accuracy trên bộ dữ liệu Celebrity Face Image Dataset

	Method	Dropout Rate	Data Augmentation	1-shot Accuracy
Hướng tiếp cận (1)	Inception-Resnet (v1)	0.75	No	0.4375
	YOLOv5n + Inception-Resnet (v1)	0.5	No	<b>0.9798</b>
Hướng tiếp cận (2)	YOLOv5n + Inception-Resnet (v1)	0.5	No	0.9568
	YOLOv5n + Inception-Resnet (v1)	0.75	No	0.9209

## 6 Nhận xét

### 6.1 Ưu điểm

- **Hiệu quả cao của phương pháp YOLOv5n kết hợp với Inception-Resnet (v1):** Việc tích hợp mô hình YOLOv5n với Inception-Resnet (v1) theo Hướng tiếp cận (1) đã mang lại kết quả ấn tượng. Cụ thể, đối với bài toán nhận diện khuôn mặt 1-shot, mô hình đạt độ chính xác lên đến 97.98%. Trong khi đó, đối với bài toán phân loại khuôn mặt, độ chính xác đạt mức 93.10%..
- **Tiềm năng của phương pháp trong các ứng dụng thực tế:** Kết quả này chứng tỏ rằng phương pháp này có tiềm năng lớn trong các ứng dụng nhận diện khuôn mặt thực tế, đặc biệt trong các hệ thống yêu cầu độ chính xác cao khi chỉ có một mẫu ảnh tham chiếu.

### 6.2 Hạn chế

- **Chưa thử nghiệm trên bộ dữ liệu lớn và phức tạp hơn:** Một trong những hạn chế của nghiên cứu này là chỉ sử dụng bộ dữ liệu LFW và Celebrity Face Image Dataset, những bộ dữ liệu có quy mô tương đối nhỏ. Việc thử nghiệm trên các bộ dữ liệu lớn và phức tạp hơn sẽ giúp đánh giá chính xác hơn về hiệu suất của mô hình khi đối mặt với các tình huống thực tế.
- **Phụ thuộc vào chất lượng phát hiện khuôn mặt và quy trình tiền xử lý:** Hiệu suất của mô hình vẫn còn phụ thuộc nhiều vào chất lượng của việc phát hiện khuôn mặt bằng YOLOv5n và quy trình tiền xử lý dữ liệu. Nếu dữ liệu đầu vào chứa các khuôn mặt bị mờ hoặc không rõ ràng, hiệu quả nhận diện có thể bị suy giảm.

## 7 Hướng phát triển

- **Tích hợp Liveness Detection:** Nâng cao bảo mật bằng cách thêm tính năng phát hiện gian lận, giảm thiểu rủi ro từ các hình thức tấn công như sử dụng ảnh hoặc video giả mạo.
- **Cải tiến YOLO và pipeline nhận diện:** Tối ưu hóa YOLO để nhận diện khuôn mặt trong các điều kiện phức tạp. Kết hợp các phương pháp mới trong pipeline xử lý để tăng độ chính xác.



- **Thử nghiệm trên các bộ dữ liệu lớn hơn:** Mở rộng quy mô thí nghiệm trên các bộ dữ liệu đa dạng hơn, như MegaFace hoặc MS-Celeb-1M, để kiểm tra khả năng ứng dụng thực tế.
- **Ứng dụng thực tiễn:** Triển khai thử nghiệm trong các hệ thống như chấm công, kiểm tra danh tính hoặc kiểm tra an ninh.

## Tài liệu

- [1] DeepFakeCreator. Rogozin replied to elon musk [deepfake]. <https://www.youtube.com/watch?v=RTjgkhMugVw>, 2020.
- [2] Gregory Koch, Richard Zemel, and Ruslan Salakhutdinov. Siamese neural networks for one-shot image recognition. In *ICML Deep Learning Workshop*, 2015.
- [3] Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, and Alex Alemi. Inception-v4, inception-resnet and the impact of residual connections on learning, 2016.
- [4] Qiong Cao, Li Shen, Weidi Xie, Omkar M. Parkhi, and Andrew Zisserman. Vggface2: A dataset for recognising faces across pose and age, 2018.
- [5] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions, 2014.
- [6] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition, 2015.
- [7] Fares Elmenhawii. Face-detection-dataset. <https://www.kaggle.com/datasets/fareselmenhawii/face-detection-dataset>, 2023.
- [8] Gary B. Huang, Manu Ramesh, Tamara Berg, and Erik Learned-Miller. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical Report 07-49, University of Massachusetts, Amherst, October 2007.
- [9] Vishesh Thakur. Celebrity face image dataset. <https://www.kaggle.com/datasets/vishesh1412/celebrity-face-image-dataset>, 2022.