

**ĐẠI HỌC QUỐC GIA TP.HỒ CHÍ MINH**  
**TRƯỜNG ĐẠI HỌC BÁCH KHOA**  
**KHOA ĐIỆN-ĐIỆN TỬ**  
**BỘ MÔN VIỄN THÔNG**



**BÁO CÁO LUẬN VĂN**  
**HỆ THỐNG NHẬN DIỆN KHUÔN MẶT VÀ**  
**CẢM XÚC NHÂN VIÊN QUA CAMERA IP**  
**SỬ DỤNG HỌC MÁY**

**GVHD: PGS. TS. Hà Hoàng Kha**

**SVTH : Trương Minh Trí**

**MSSV : 1814475**

- 1. Giới thiệu tổng quan**
- 2. Nhiệm vụ đề tài**
- 3. Nhận diện khuôn mặt**
- 4. Ứng dụng của hệ thống**
- 5. Kết luận và hướng phát triển**

# 1. Giới thiệu tổng quan

## Đặt vấn đề



**Sự phát triển rất nhanh của các chuỗi cửa hàng tiêu biểu:**

- The Coffee House: 100 cửa hàng (4 năm)
- Circle K: 400 cửa hàng (15 năm)
- MWG: 5497 cửa hàng (10 năm)
- ...

**Các thách thức đặt ra:**

1. Nhân sự (quản lí, chăm công, ...)
2. Chất lượng dịch vụ (cảm xúc khách hàng, cảm xúc của nhân viên...)

# 1. Giới thiệu tổng quan



## Mục tiêu của đề tài:

1. Chấm công
2. Nhận diện cảm xúc
3. Theo dõi từ xa

## Yêu cầu đặt ra:

- Nhận diện thời gian thực và độ chính xác cao
- Tận dụng được những thiết bị ghi hình có sẵn
- Xây dựng ứng dụng, dễ dàng triển khai trong thực tế

## 2. Nhiệm vụ đề tài

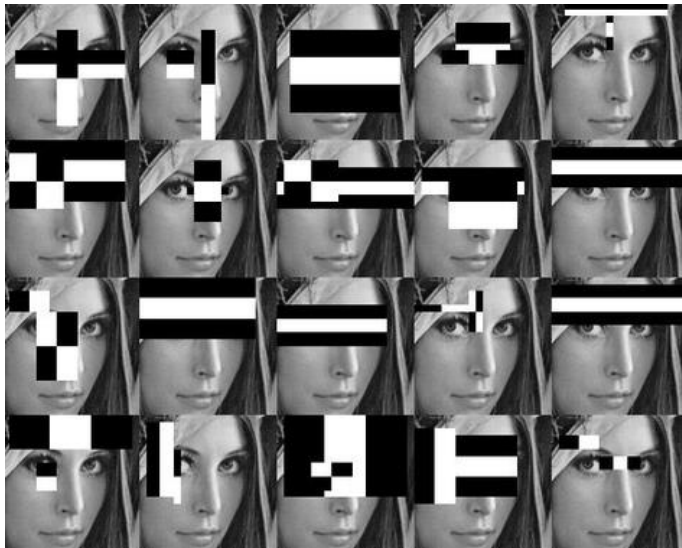
### **Xây dựng ứng dụng “HỆ THỐNG NHẬN DIỆN KHUÔN MẶT VÀ CẢM XÚC NHÂN VIÊN QUA CAMERA IP SỬ DỤNG HỌC MÁY”**

1. Thu thập dữ liệu
2. Lựa chọn thuật toán phát hiện khuôn mặt
3. Xử lý bộ dữ liệu ảnh
4. Huấn luyện mô hình trích xuất đặc trưng
5. Huấn luyện mô hình phân loại các đặc trưng
6. Tích hợp hệ thống nhận dạng vào ứng dụng Web

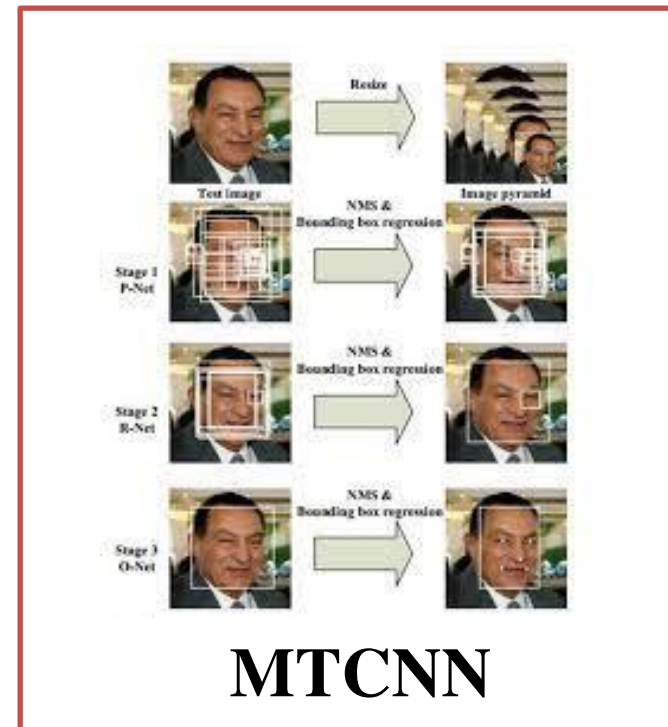
# 3. Nhận diện khuôn mặt



## Phát hiện khuôn mặt



Haar Cascade

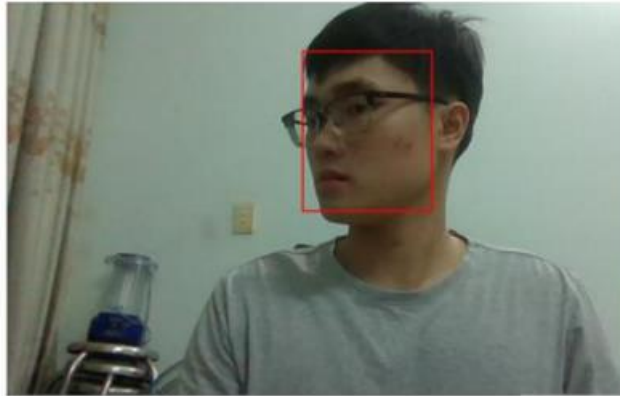




# Đánh giá và kết quả

## Kết quả phát hiện khuôn mặt

Khả năng phát hiện

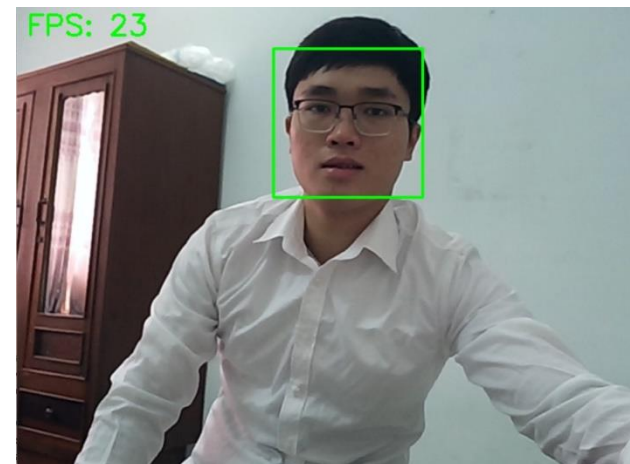
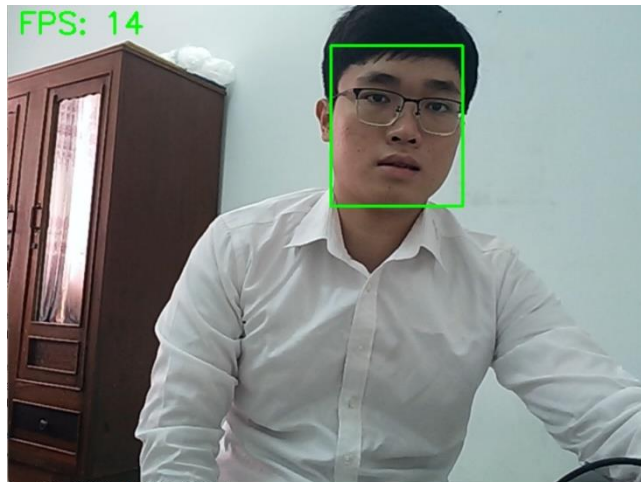


**MTCNN**



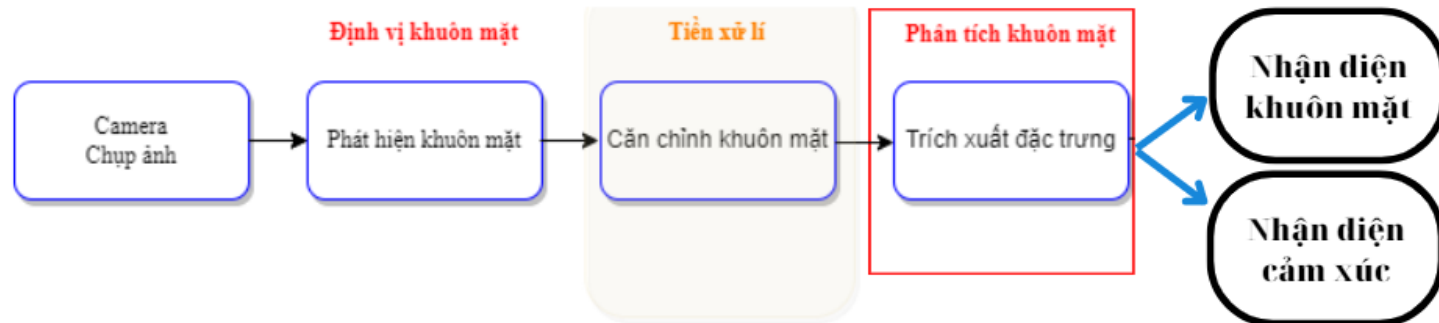
**Haar Cascade**

Tốc độ xử lý (FPS)



# 3. Nhận diện khuôn mặt

## Trích xuất đặc trưng



1. Tốc độ trích xuất đáp ứng yêu cầu thời gian thực <50ms
2. Hạn chế phụ thuộc vào điều kiện khách quan như môi trường, ánh sáng, độ nghiêng, độ tuổi, bị che khuất...
3. Có thể nhận diện khuôn mặt dựa vào khoảng cách giữa các vector đặc trưng

=> Lựa chọn hàm mất mát TripletLoss:

$$L(A, P, N) = \sum_{i=0}^n \max(\|f(A_i) - f(P_i)\| - \|f(A_i) - f(N_i)\| + \alpha, 0)$$

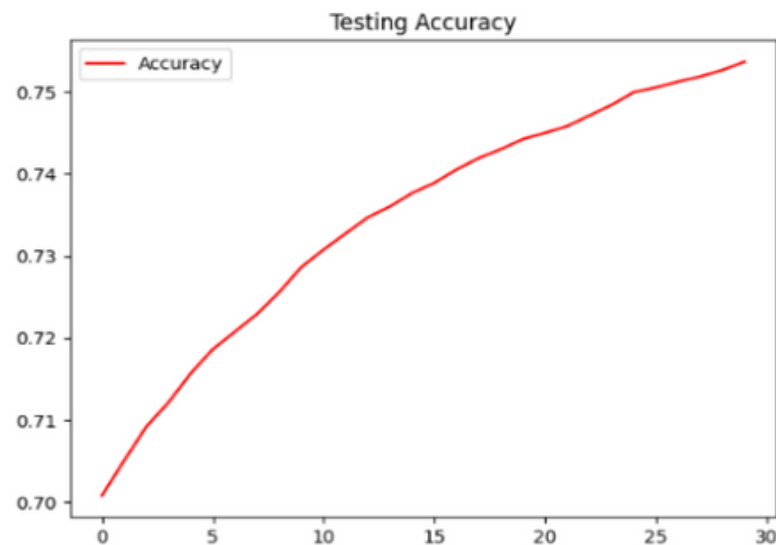
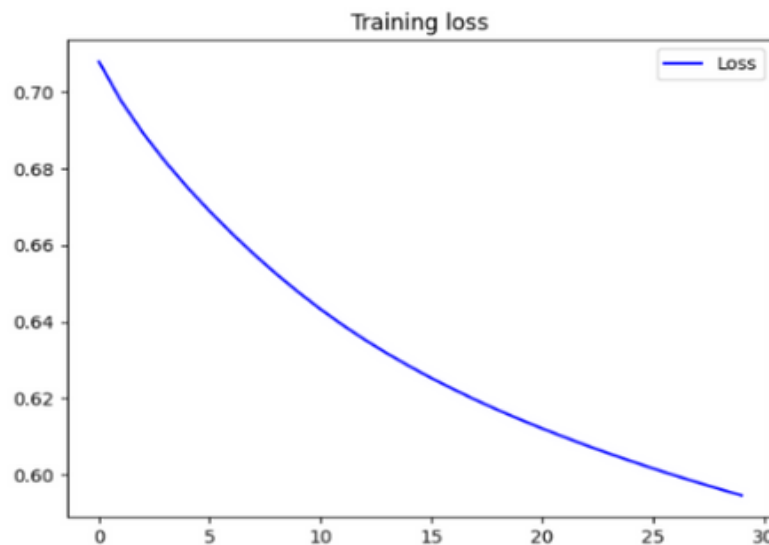


# Đánh giá và kết quả

## Transfer Learning với model Inception-ResnetV1

Bộ dữ liệu VN-Celeb đã làm sạch:

- Bao gồm 960 người
- Mỗi người có ít nhất 5 ảnh



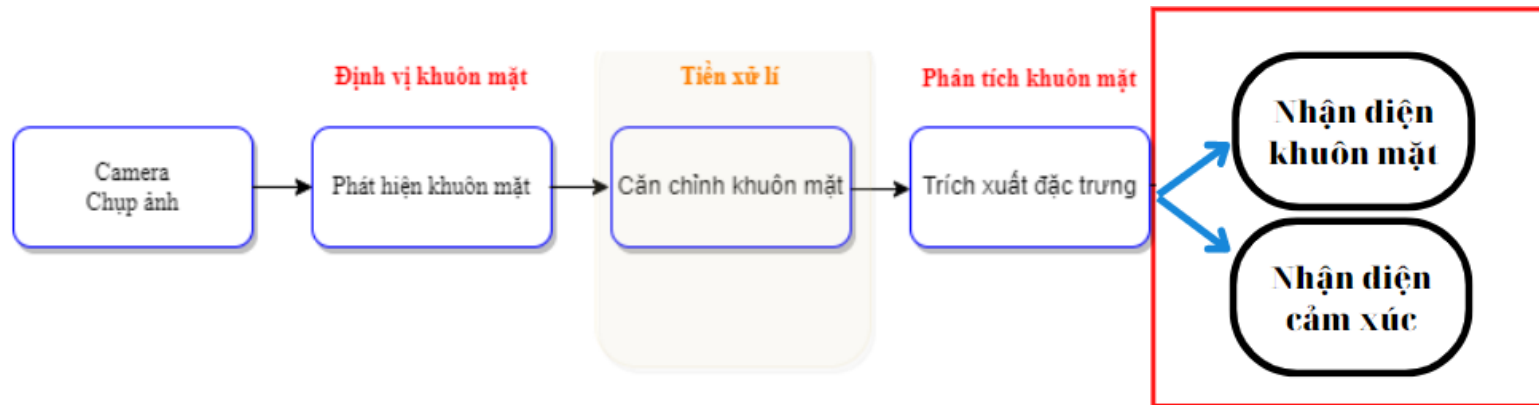
### Kết quả huấn luyện chuyển tiếp (Transfer learning)

- Quá trình huấn luyện không liền mạch
- Tốn rất nhiều thời gian (hơn 30 ngày liên tục)

**Không đáp ứng yêu cầu 3 ( Giải pháp: Thay TripletLoss thành ArcFace)**

# 3. Nhận diện khuôn mặt

## Nhận diện khuôn mặt và cảm xúc



### Nhận diện khuôn mặt:

- Bài toán phân loại nhiều lớp - với mỗi lớp là đặc trưng của mỗi người
- Thuật toán ít nhạy cảm với nhiễu
- Đường bao giữa các lớp có thể không tuyến tính

### Nhận diện cảm xúc:

- Bài toán phân loại 2 lớp "Tích cực" và "Tiêu cực"
- Giá trị Recall của mô hình cao

# Đánh giá và kết quả

## Nhận diện khuôn mặt

Trên tập dữ liệu: Độ chính xác = 100%

Trong điều kiện thực tế: Độ chính xác = 94.8% với một tỉ lệ nhỏ gán nhãn sai

		Thực tế							
Dự đoán		0	1	2	3	4	5	6	Unknown
	0	0	0	0	0	12	0	0	15
	1	0	0	0	0	0	0	0	3
	2	0	0	726	5	0	0	1	52
	3	0	0	22	853	0	0	4	5
	4	0	0	0	0	603	4	0	21
	5	0	0	0	0	0	713	0	15
	6	0	0	0	24	0	0	756	2
	Unknown	0	0	53	65	15	15	22	6535

**Kết quả đánh giá thực tế với ngưỡng khuôn mặt "Unknown" là 70%**

- Không đánh giá với người có ID = 0 và 1
- Vẫn có một tỉ lệ nhỏ nhận diện sai người

# Đánh giá và kết quả

## Nhận diện cảm xúc

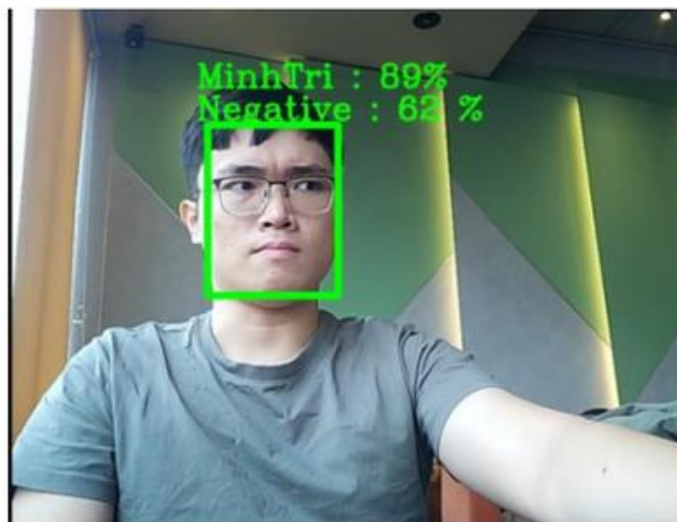
Trên tập dữ liệu: Độ chính xác = 72.8%

Trong điều kiện thực tế: Độ chính xác = 79.9%, Recall = 79.5%

Dự đoán	Thực tế	
	Positive	Negative
	Positive	Negative
Positive	6553	28
Negative	1690	270

## Kết quả đánh giá thực tế với ngưỡng "Unknown" là 70%

Tuy nhiên, bài đánh giá thực tế không mang tính khách quan khi sử dụng khuôn mặt cố ý "Tiêu cực" để đánh giá.



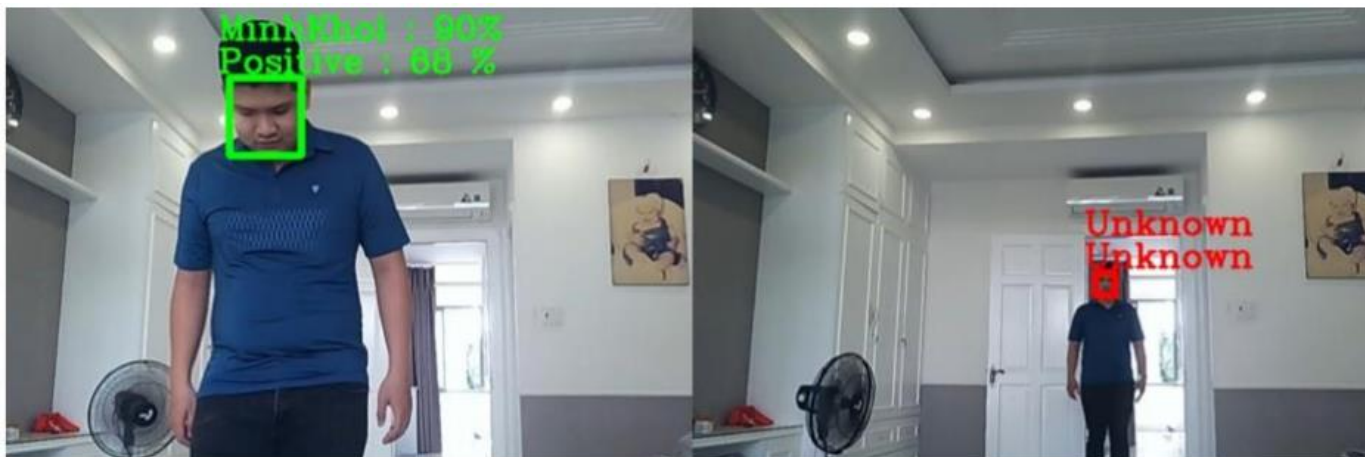
# Đánh giá và kết quả

## Điều kiện nhận diện

Khoảng cách nhận diện tin cậy: 1-6m, ngưỡng khuôn mặt nhỏ nhất: 160x160p



Khuôn mặt quá nhỏ có xu hướng gán nhãn “Tiêu cực”



Khoảng cách nhận diện với 1m (bên trái), với 6m (bên phải)

# 4. Ứng dụng của hệ thống

## Frontend

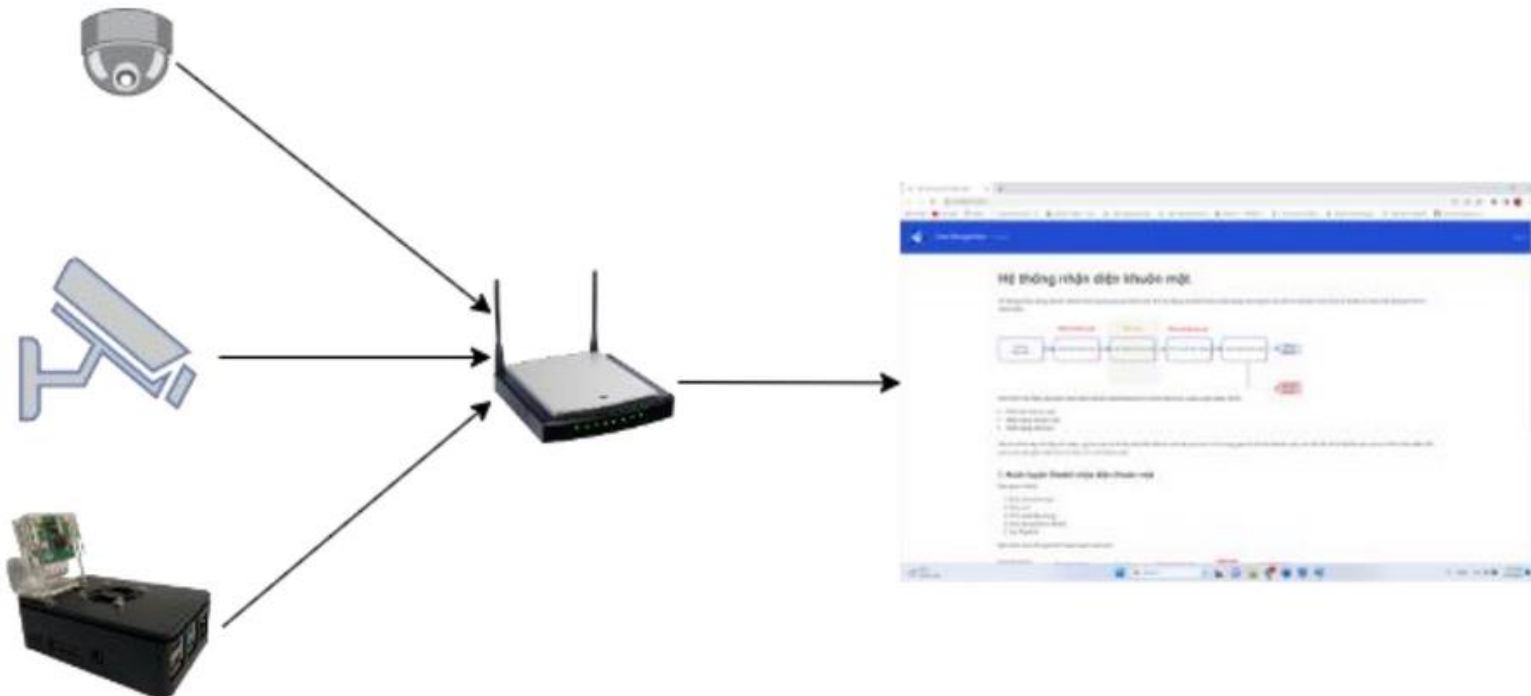
HTML CSS: Bootstrap v5.2.3

Javascript: Bootstrap v5.2.3

## Backend

Python MySQL

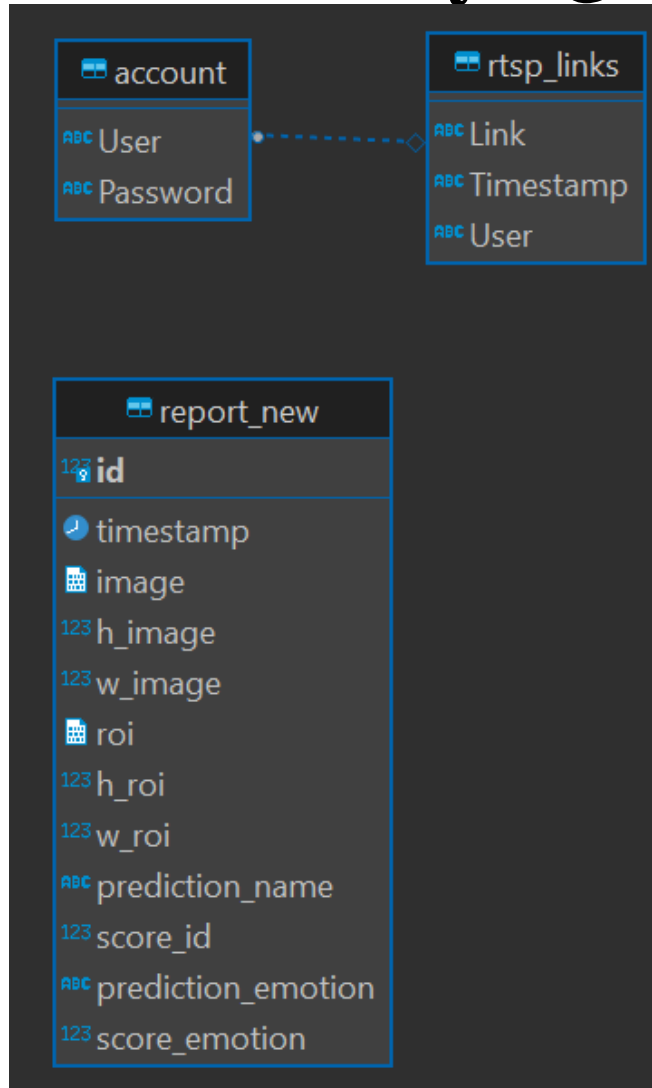
Flask Framework





## 4. Ứng dụng của hệ thống

### Database MySQL:



Bao gồm 3 bảng:

1. account: Thông tin người dùng
2. rtsp\_links: Đường dẫn RTSP
3. report\_new: Kết quả nhận diện

Sử dụng ngôn ngữ SQL để truy vấn dữ liệu

```
cursor = connection.cursor()
query = "SELECT * FROM report_new WHERE id = %s"
cursor.execute(query, id_num)
data = cursor.fetchone()
```



## 4. Ứng dụng của hệ thống

**DEMO**

## **Hệ thống nhận diện khuôn mặt và cảm xúc**

### **Mô hình nhận diện**

- Phát hiện khuôn mặt: MTCNN
- Trích xuất đặc trưng: Inception-ResnetV1
- Phân loại đặc trưng: Support Vector Machine

### **Ứng dụng Web**

- Tính năng 1: Nhận diện với camera IP
- Tính năng 2: Quản lý các đường dẫn RTSP
- Tính năng 3: Quản lý dữ liệu kết quả nhận diện trong quá khứ

## Mô hình nhận diện

- Nhận diện dựa trên việc tính toán khoảng cách giữa các đặc trưng
- Cải thiện bộ dữ liệu nhận diện cảm xúc
- Tích hợp thêm nhận diện hành vi
- Tăng khả năng phần cứng nhằm tăng tốc độ nhận diện

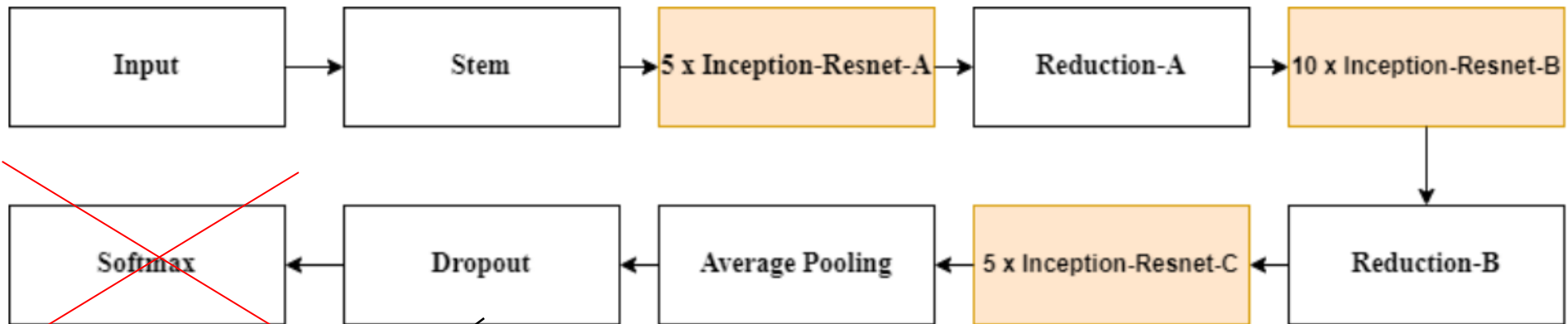
## Ứng dụng Web

- Cải thiện tính bảo mật của ứng dụng
- Nhận diện nhiều Camera cùng một thời điểm
- Sử dụng linh hoạt các loại Camera
- Xây dựng app trên điện thoại
- Tạo API kết nối giữa ứng dụng và mô hình nhận diện



**CẢM ƠN CÁC THẦY VÀ CÁC BẠN  
ĐÃ LẮNG NGHE**

## Transfer Learning model Inception-ResnetV1



**TripletLoss:**

$$L(A, P, N) = \sum_{i=0}^n \max(|f(A_i) - f(P_i)| - |f(A_i) - f(N_i)| + \alpha, 0)$$