

## BÁO CÁO THỰC TẬP

CHỦ ĐỀ:

# NHẬN DIỆN KHUÔN MẶT BẰNG PHƯƠNG PHÁP HỌC MÁY

Sinh viên thực hiện

**Vũ Mạnh Đức – 20002046**

**Vũ Trọng Quân – 20002087**

**Phan Tấn Sơn - 20002090**

# Nội dung

- 1 Giới thiệu
- 2 Công việc triển khai
- 3 Kết quả đạt được

# Nội dung

- 1 Giới thiệu
- 2 Công việc triển khai
- 3 Kết quả đạt được

# Giới thiệu

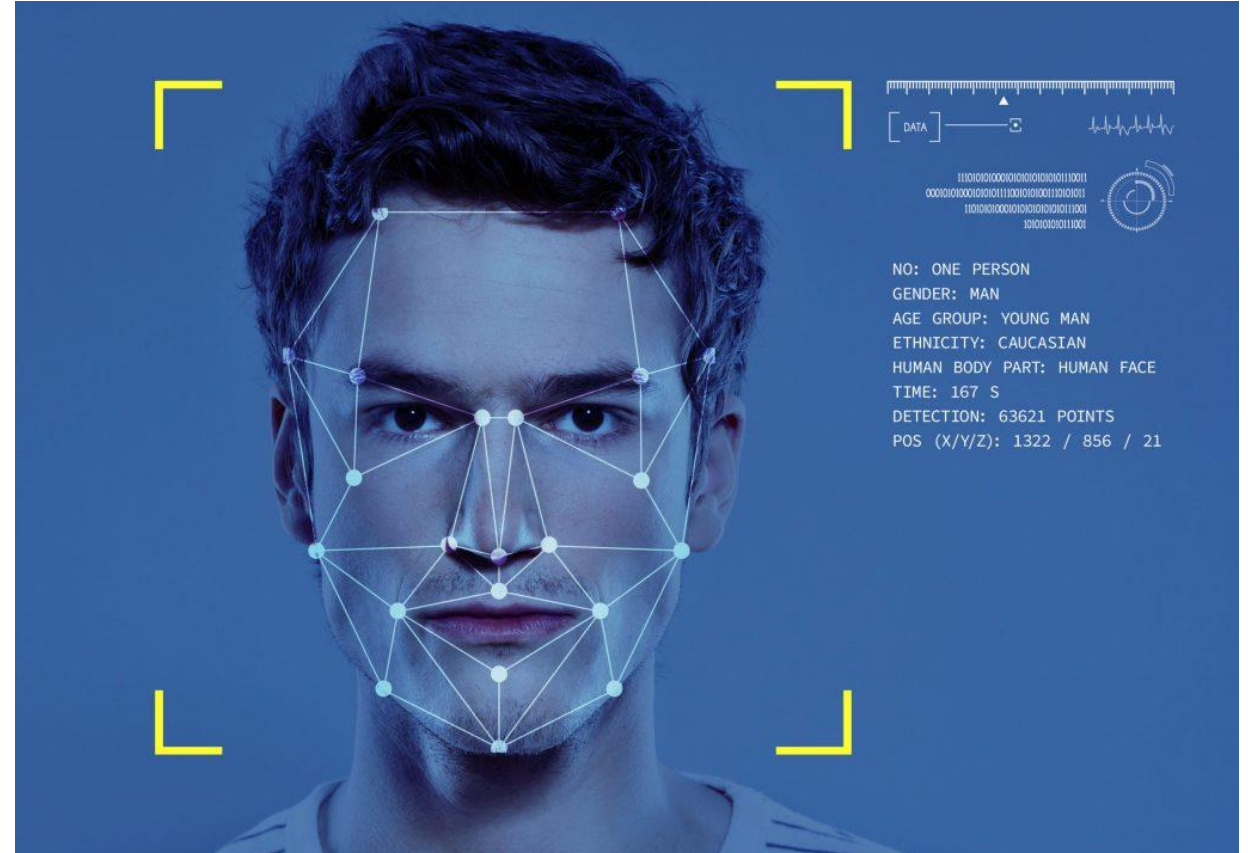
- 1 Phát biểu bài toán
- 2 Giới thiệu các công nghệ sử dụng

# Giới thiệu

- 1 Phát biểu bài toán
- 2 Giới thiệu các công nghệ sử dụng

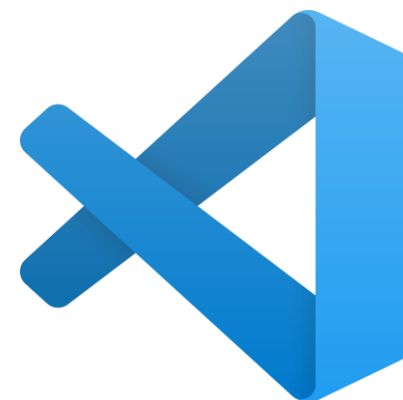
# Phát biểu bài toán

- Nhận diện khuôn mặt (Face recognition) là bài toán phổ biến trong AI.
- Nhận diện khuôn mặt được ứng dụng trong nhiều lĩnh vực: điểm danh, phát hiện tội phạm, tìm kiếm thông tin,...



# Giới thiệu

- 1 Phát biểu bài toán
- 2 Giới thiệu các công nghệ sử dụng





# Nội dung

- 1 Giới thiệu
- 2 Công việc triển khai
- 3 Kết quả đạt được

# Công việc triển khai

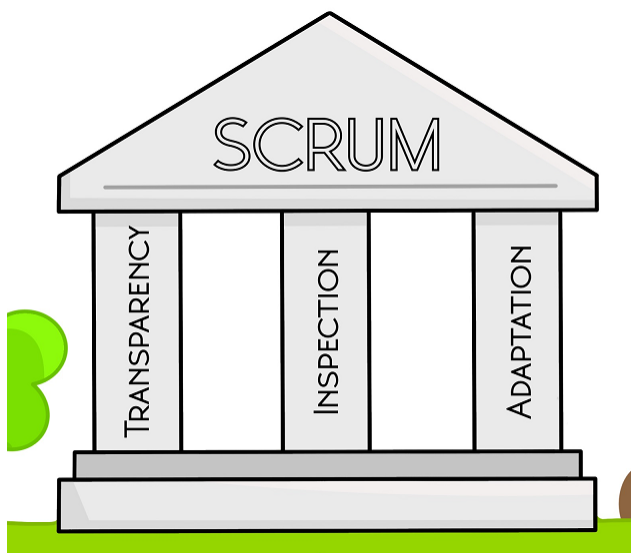
- 1 Học agile scrum course
- 2 Phân tích bài toán
- 3 Phát hiện khuôn mặt
- 4 YOLOv7
- 5 Facial Landmarks
- 6 MTCNN
- 7 Face Classification

# Công việc triển khai<sup>2</sup>

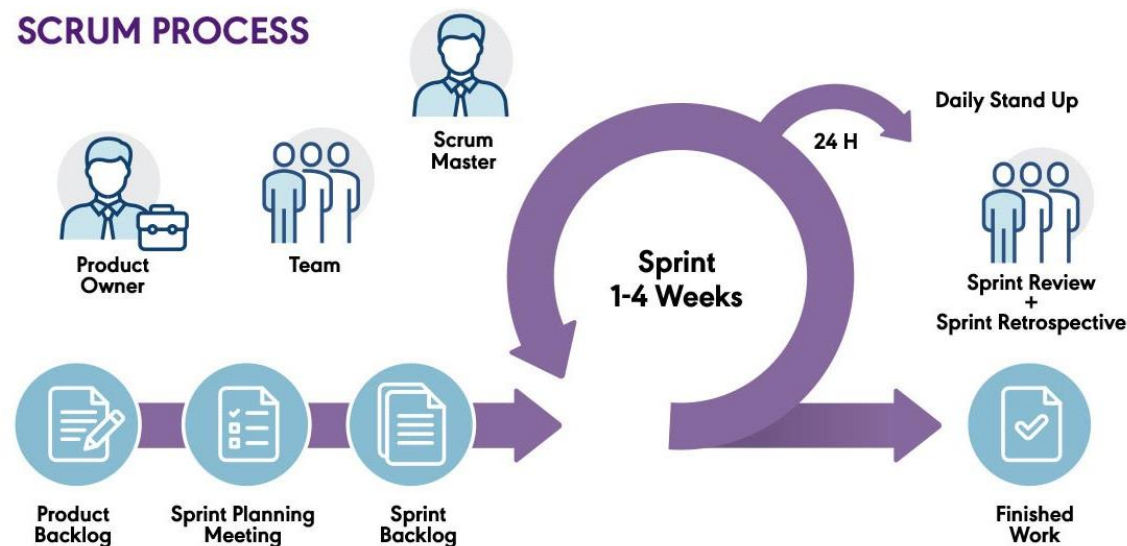
- 1 Học agile scrum course
- 2 Phân tích bài toán
- 3 Phát hiện khuôn mặt
- 4 YOLOv7
- 5 Facial Landmarks
- 6 MTCNN
- 7 Face Classification

# Agile scrum

Scrum là một khung tổ chức công việc (framework) dùng trong các dự án phát triển phần mềm với mục tiêu là chuyển giao các sản phẩm mới đều đặn, sau từ 1-4 tuần.



## SCRUM PROCESS



Ba giá trị cốt lõi của Scrum:

- Minh bạch.
- Thanh tra.
- Thích nghi.

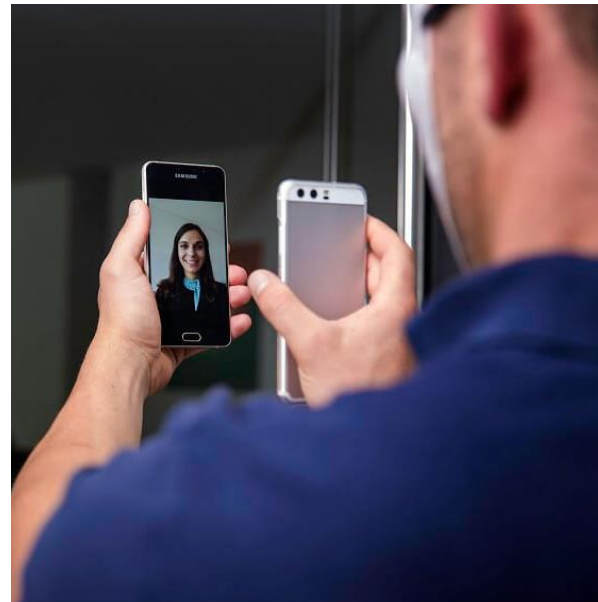
# Công việc triển khai<sup>2</sup>

- 1 Học agile scrum course
- 2 Phân tích bài toán**
- 3 Phát hiện khuôn mặt
- 4 YOLOv7
- 5 Facial Landmarks
- 6 MTCNN
- 7 Face Classification

# Phân tích bài toán

Khó khăn của bài toán nhận diện khuôn mặt:

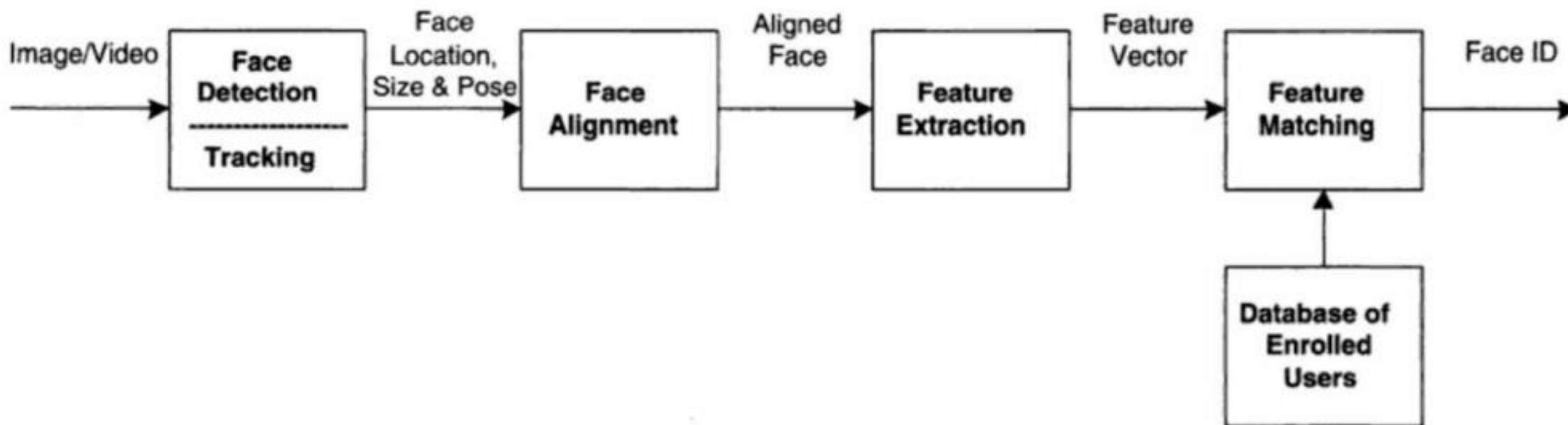
- Vấn đề chống giả mạo khuôn mặt
- Góc chụp
- Khuôn mặt thay đổi theo thời gian
- Điều kiện ánh sáng



# Phân tích bài toán

Các vấn đề cần giải quyết trong bài toán nhận diện khuôn mặt:

- Face Detection – Phát hiện khuôn mặt
- Face Extraction (Face Embedding) – Trích xuất đặc trưng
- Face Classification



# Công việc triển khai<sup>2</sup>

- 1 Học agile scrum course
- 2 Phân tích bài toán
- 3 Phát hiện khuôn mặt**
- 4 YOLOv7
- 5 Facial Landmarks
- 6 MTCNN
- 7 Face Classification



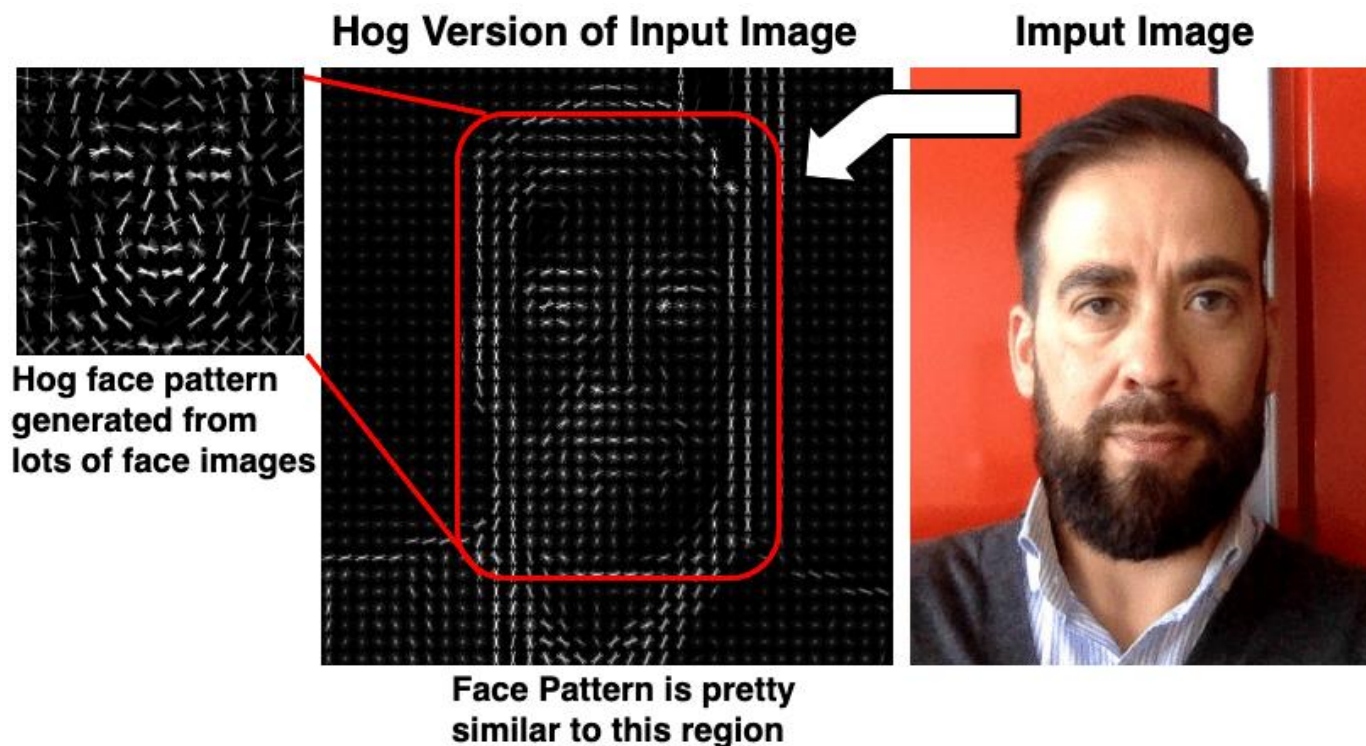
# Phát hiện khuôn mặt

Phát hiện khuôn mặt là quá trình tự động định vị khuôn mặt người trong phương tiện trực quan (hình ảnh kỹ thuật số hoặc video).



# Phát hiện khuôn mặt

Feature-based: Sử dụng các bộ lọc thủ công (hand-crafted filters) với phương pháp HOG.



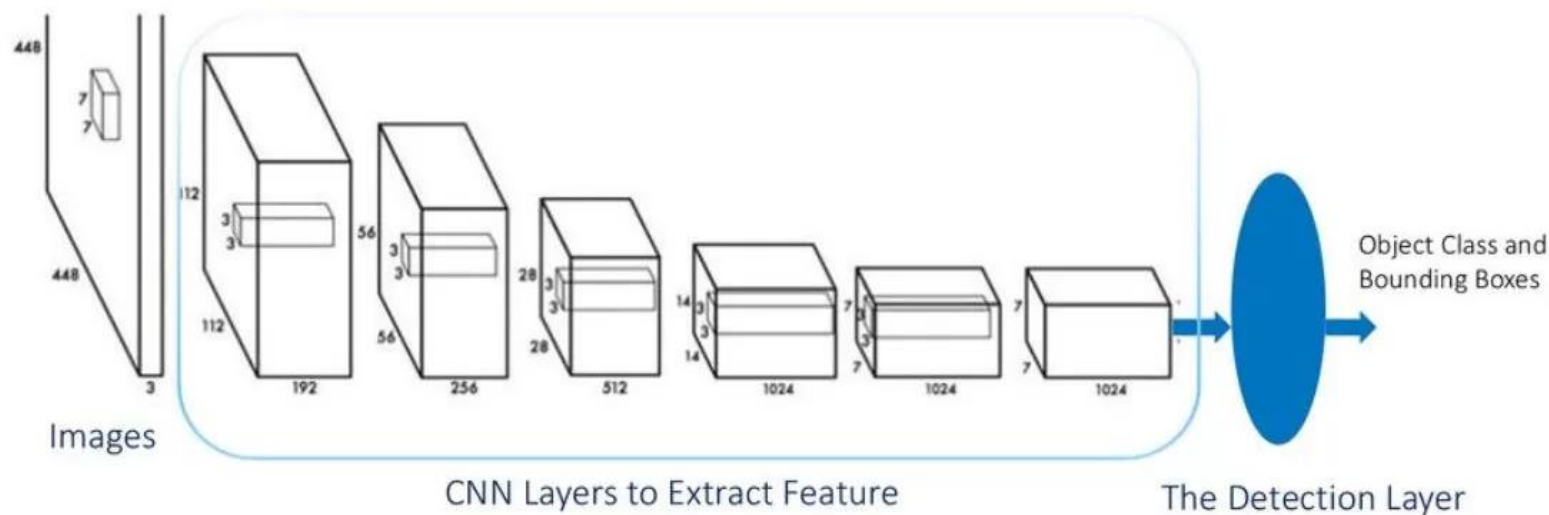
# Công việc triển khai<sup>2</sup>

- 1 Học agile scrum course
- 2 Phân tích bài toán
- 3 Phát hiện khuôn mặt
- 4 YOLOv7**
- 5 Facial Landmarks
- 6 MTCNN
- 7 Face Classification

# YOLOv7

YOLO (You only look once) là một mô hình mạng CNN cho việc phát hiện, nhận dạng, phân loại đối tượng được tạo ra từ việc kết hợp giữa các convolutional layers và fully connected layers.

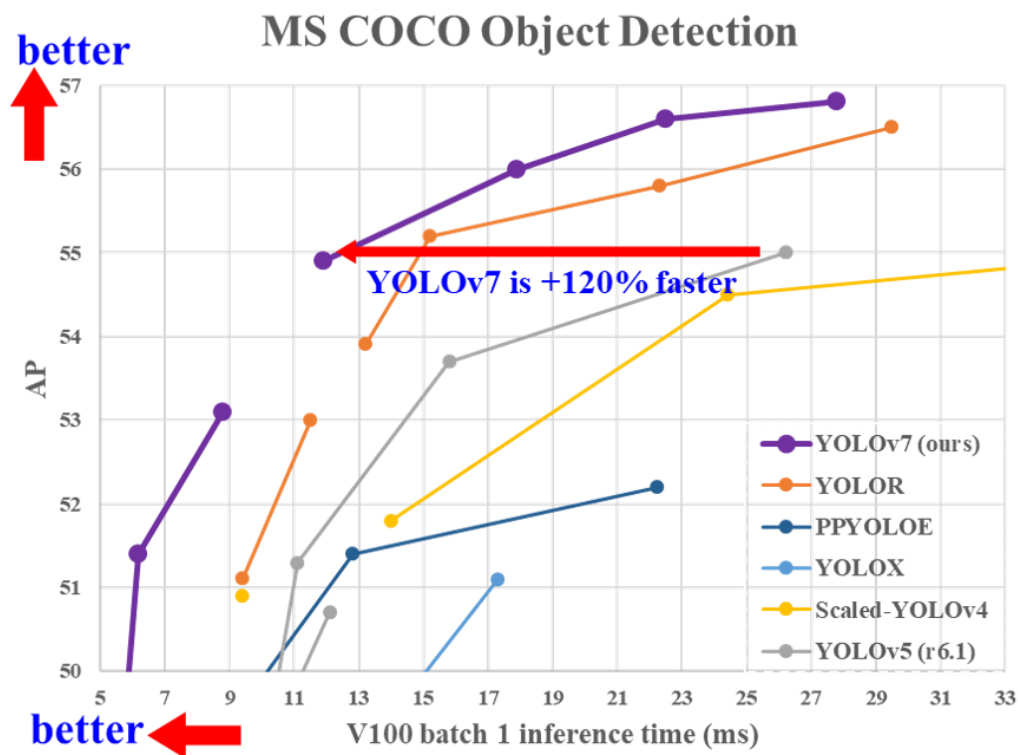
YOLO đạt tốc độ gần như real-time và đạt độ chính xác cao.



# YOLOv7

Trên bộ dữ liệu COCO:

- Tốc độ của YOLOv7 nhanh hơn 120% so với YOLOv5.
- Độ chính xác đạt 73,7% trên ngưỡng IoU 0,5



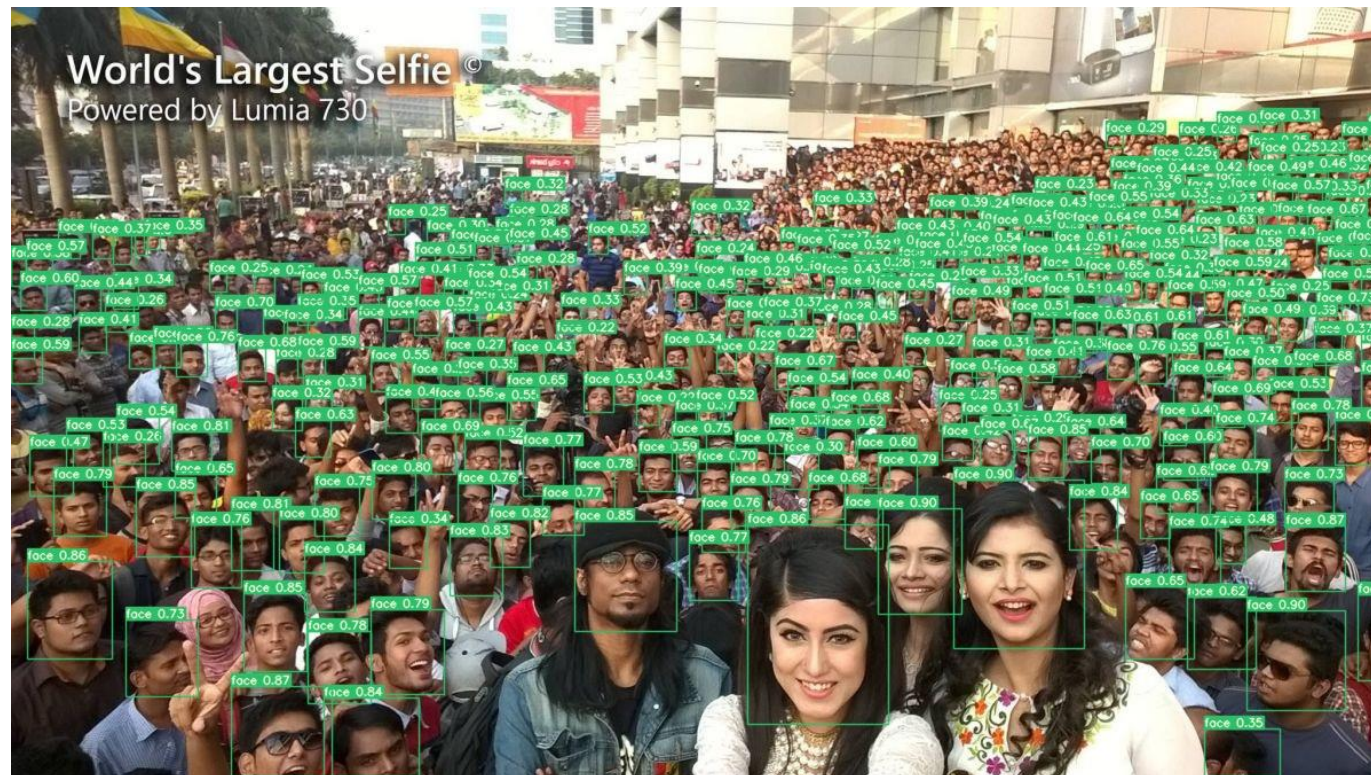
Model	#Param.	FLOPs	Size	AP <sup>val</sup>	AP <sup>val</sup> <sub>50</sub>	AP <sup>val</sup> <sub>75</sub>	AP <sup>val</sup> <sub>S</sub>	AP <sup>val</sup> <sub>M</sub>	AP <sup>val</sup> <sub>L</sub>
YOLOv4 [3]	64.4M	142.8G	640	49.7%	68.2%	54.3%	32.9%	54.8%	63.7%
YOLOR-u5 (r6.1) [81]	46.5M	109.1G	640	50.2%	68.7%	54.6%	33.2%	55.5%	63.7%
YOLOv4-CSP [79]	52.9M	120.4G	640	50.3%	68.6%	54.9%	34.2%	55.6%	65.1%
YOLOR-CSP [81]	52.9M	120.4G	640	50.8%	69.5%	55.3%	33.7%	56.0%	65.4%
YOLOv7	36.9M	104.7G	640	<b>51.2%</b>	<b>69.7%</b>	<b>55.5%</b>	<b>35.2%</b>	<b>56.0%</b>	<b>66.7%</b>
improvement	-43%	-15%	-	+0.4	+0.2	+0.2	+1.5	=	+1.3
YOLOR-CSP-X [81]	96.9M	226.8G	640	52.7%	<b>71.3%</b>	57.4%	36.3%	57.5%	68.3%
YOLOv7-X	71.3M	189.9G	640	<b>52.9%</b>	71.1%	<b>57.5%</b>	<b>36.9%</b>	<b>57.7%</b>	<b>68.6%</b>
improvement	-36%	-19%	-	+0.2	-0.2	+0.1	+0.6	+0.2	+0.3
YOLOv4-tiny [79]	6.1	6.9	416	24.9%	42.1%	25.7%	8.7%	28.4%	39.2%
YOLOv7-tiny	6.2	5.8	416	<b>35.2%</b>	<b>52.8%</b>	<b>37.3%</b>	<b>15.7%</b>	<b>38.0%</b>	<b>53.4%</b>
improvement	+2%	-19%	-	+10.3	+10.7	+11.6	+7.0	+9.6	+14.2
YOLOv4-tiny-3l [79]	8.7	5.2	320	30.8%	47.3%	32.2%	<b>10.9%</b>	31.9%	51.5%
YOLOv7-tiny	6.2	3.5	320	<b>30.8%</b>	<b>47.3%</b>	<b>32.2%</b>	10.0%	<b>31.9%</b>	<b>52.2%</b>
improvement	-39%	-49%	-	=	=	=	-0.9	=	+0.7
YOLOR-E6 [81]	115.8M	683.2G	1280	55.7%	73.2%	60.7%	40.1%	<b>60.4%</b>	69.2%
YOLOv7-E6	97.2M	515.2G	1280	<b>55.9%</b>	<b>73.5%</b>	<b>61.1%</b>	<b>40.6%</b>	60.3%	<b>70.0%</b>
improvement	-19%	-33%	-	+0.2	+0.3	+0.4	+0.5	-0.1	+0.8
YOLOR-D6 [81]	151.7M	935.6G	1280	56.1%	73.9%	61.2%	<b>42.4%</b>	60.5%	69.9%
YOLOv7-D6	154.7M	806.8G	1280	56.3%	73.8%	61.4%	41.3%	60.6%	70.1%
YOLOv7-E6E	151.7M	843.2G	1280	<b>56.8%</b>	<b>74.4%</b>	<b>62.1%</b>	40.8%	<b>62.1%</b>	<b>70.6%</b>
improvement	=	-11%	-	+0.7	+0.5	+0.9	-1.6	+1.6	+0.7



# YOLOv7

Training YOLOv7 trên bộ dữ liệu WIDER FACE với 55 epoch:

- Giá trị Precision đạt 0.98.
- Giá trị Recall đạt 0.96.
- Giá trị trung bình chính xác (Average Precision) tại ngưỡng IoU 0.5 đạt 0.98.



Epoch	gpu_mem	box	obj	cls	total	labels	img_size
54/54	13.9G	0.01839	0.009006	0	0.02739	42	640: 100% 139/139 [02:21<00:00, 1.02s/it]
	Class	Images	Labels		P	R	mAP@.5 mAP@.5:.95: 100% 70/70 [00:46<00:00, 1.51it/s]
	all	2211	11321		0.98	0.962	0.985 0.831

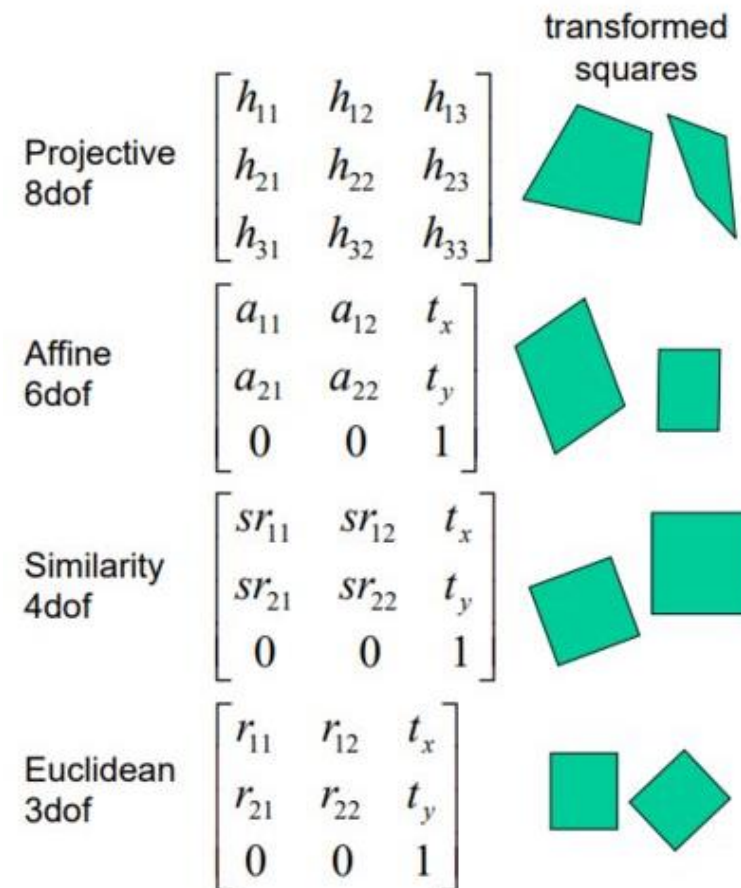
# Công việc triển khai<sup>2</sup>

- 1 Học agile scrum course
- 2 Phân tích bài toán
- 3 Phát hiện khuôn mặt
- 4 YOLOv7
- 5 Facial Landmarks**
- 6 MTCNN
- 7 Face Classification

# Facial Landmarks

Sau khi xác định được vị trí khuôn mặt, sẽ xuất hiện những khuôn mặt bị chéo, bị lệch. Do đó, cần căn chỉnh khuôn mặt để xác định được cấu trúc khuôn mặt (mắt, mũi, miệng,...).

Có thể áp dụng các phương pháp biến đổi hình học để căn chỉnh khuôn mặt về theo quy chuẩn.

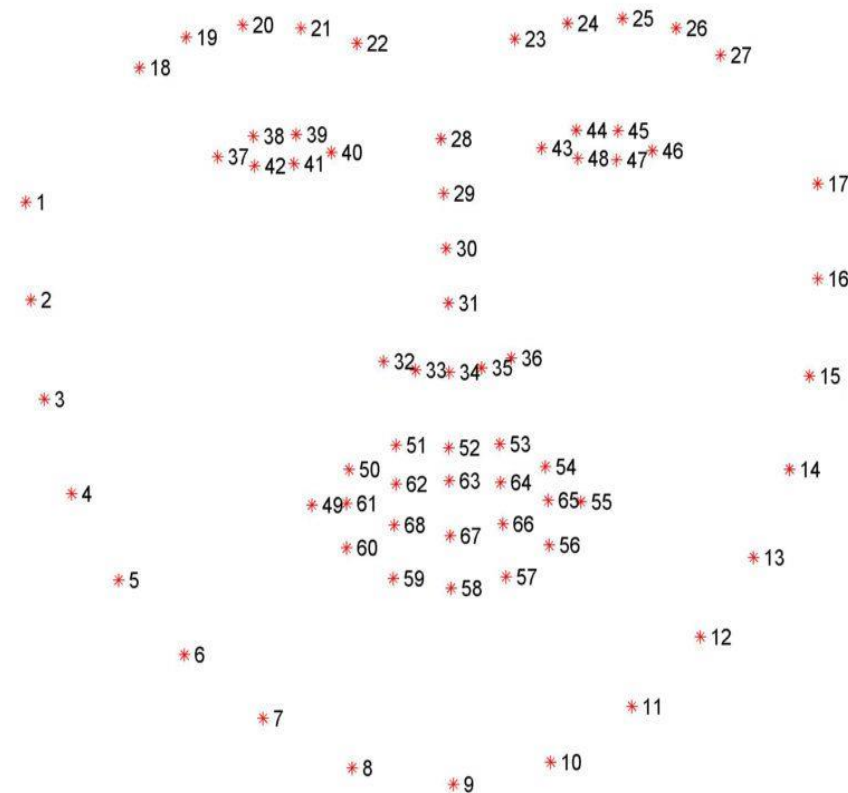




# Facial Landmarks

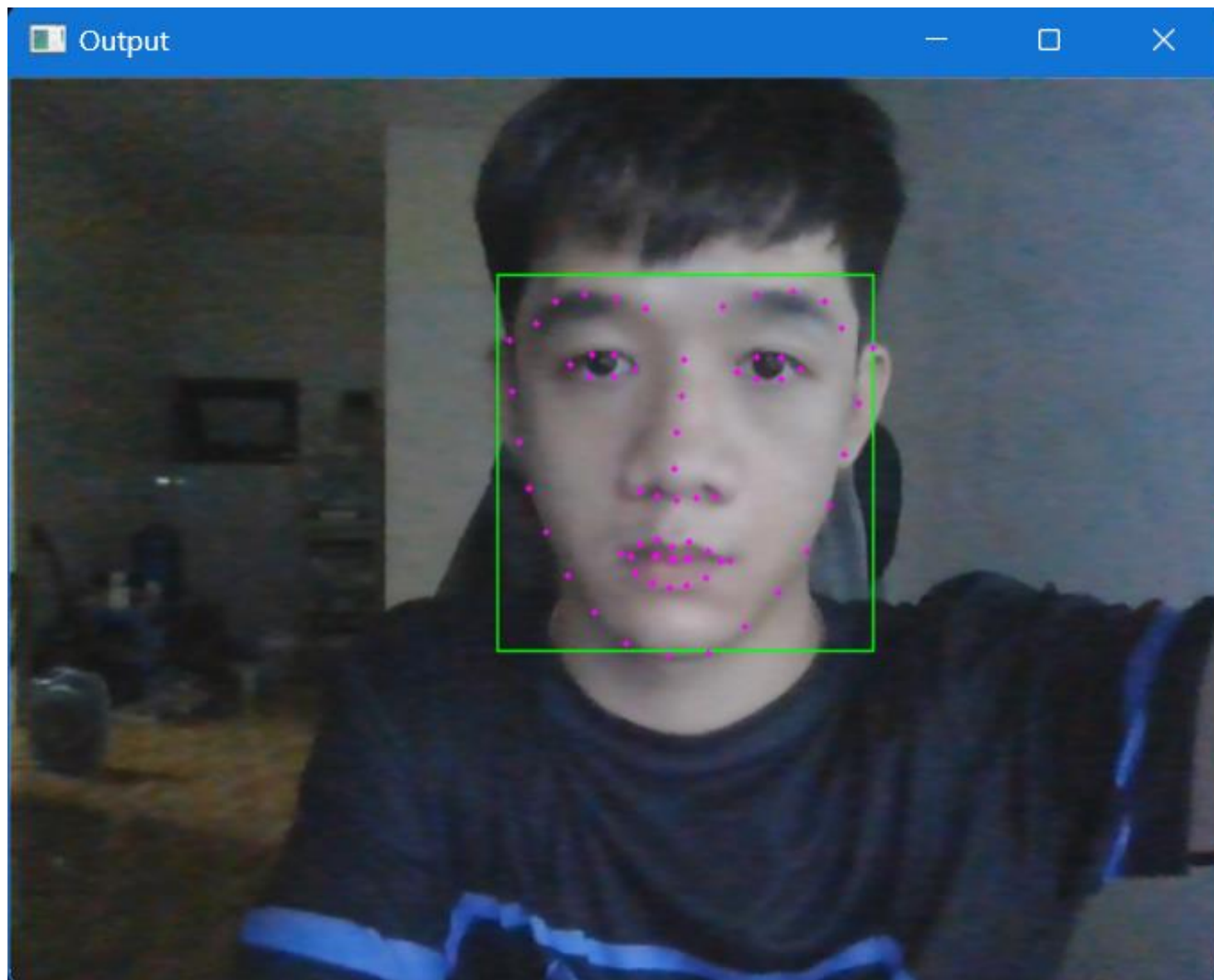
68 điểm cụ thể (được gọi là mốc) tồn tại trên mỗi khuôn mặt - đỉnh cằm, cạnh ngoài của mỗi mắt, cạnh trong của mỗi lông mày, v.v

Sau đó, chúng ta sẽ huấn luyện một máy học thuật toán để có thể tìm thấy 68 điểm cụ thể này trên mọi mặt:



# Facial Landmarks

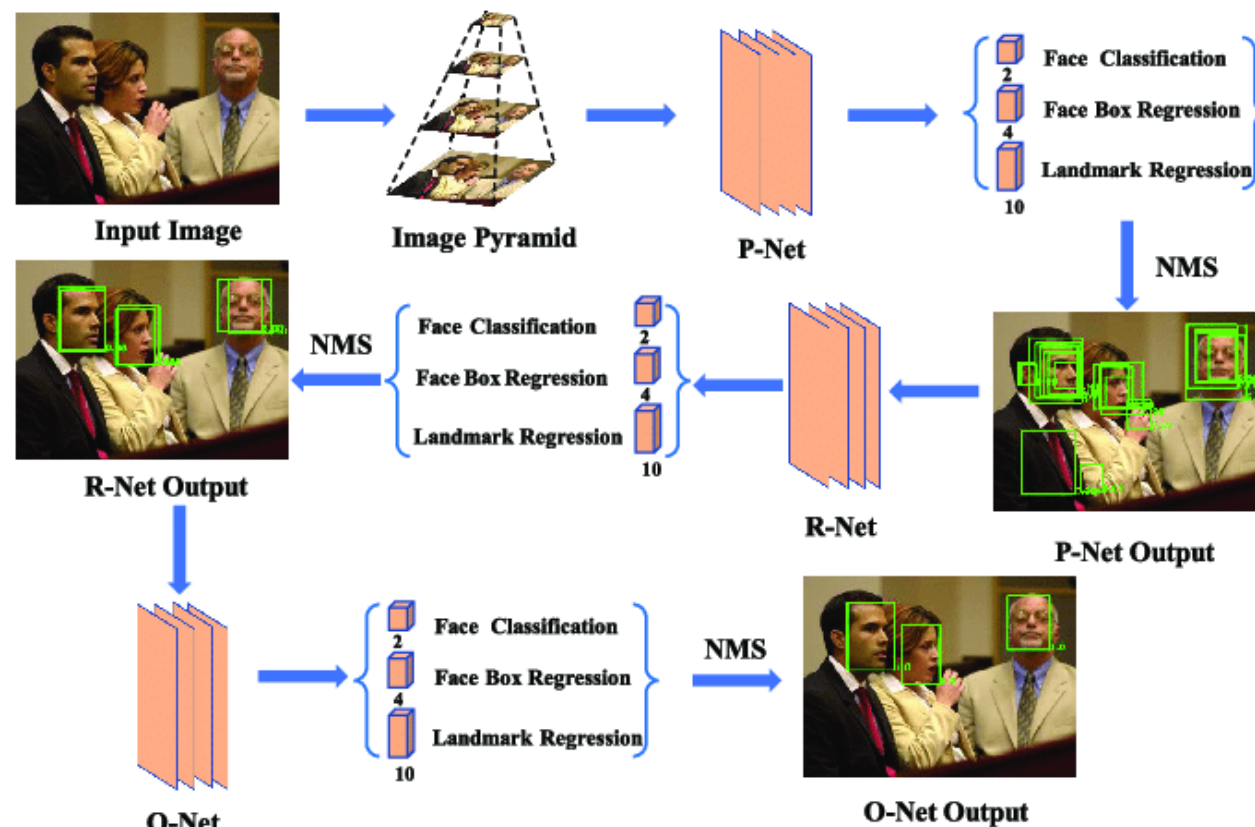
The pre-trained facial landmark detector trong dlib được sử dụng để ước lượng 68 (x, y)-coordinates tương ứng với tọa độ các facial landmarks trên khuôn mặt.

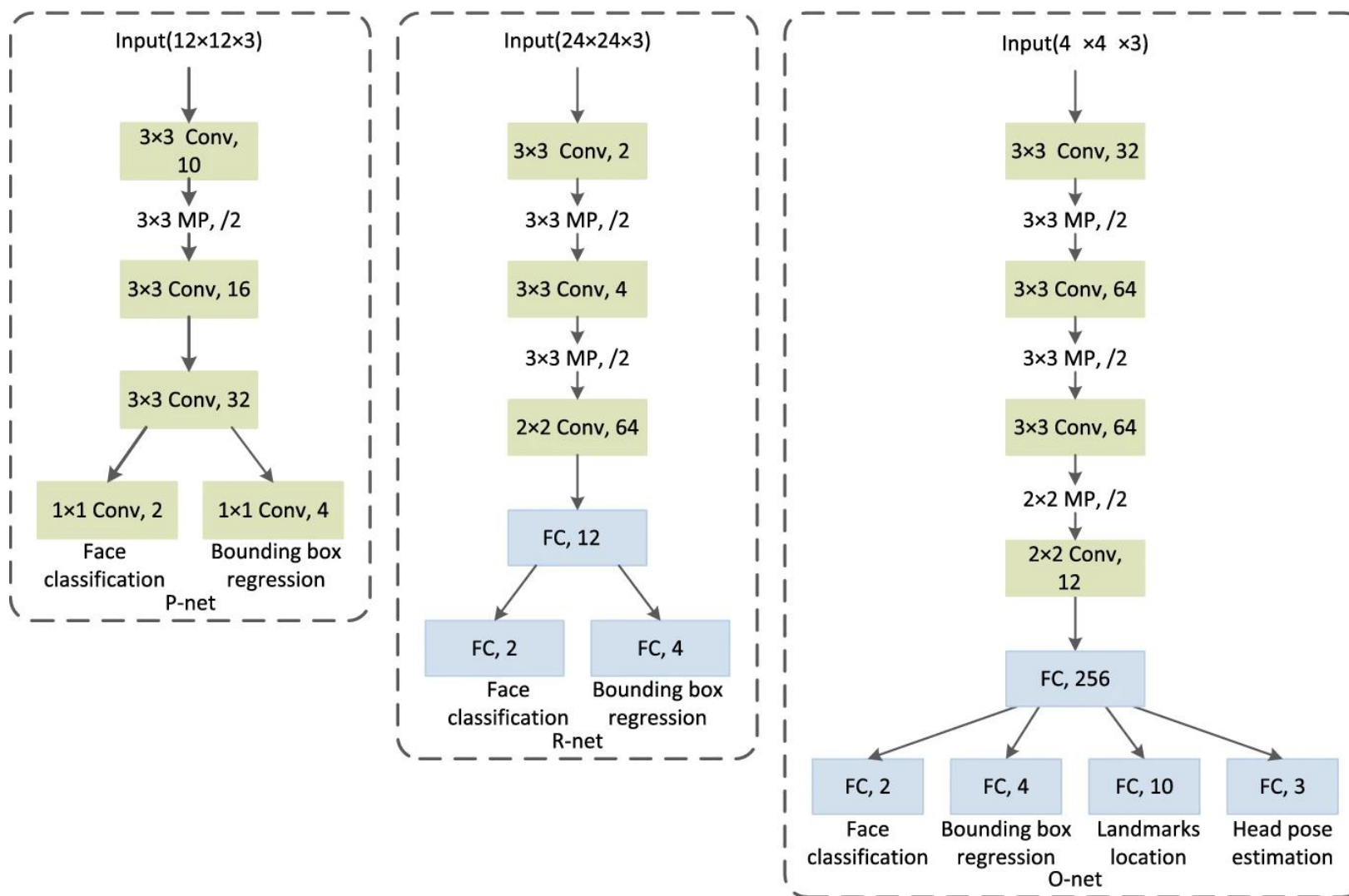


# Công việc triển khai<sup>2</sup>

- 1 Học agile scrum course
- 2 Phân tích bài toán
- 3 Phát hiện khuôn mặt
- 4 YOLOv7
- 5 Facial Landmarks
- 6 MTCNN**
- 7 Face Classification

MTCNN (Multi – Cascaded Convolution Networks) được xây dựng trên cơ sở của CNN và kết hợp nhiều tác vụ như phát hiện khuôn mặt (face detection), xác định các điểm đặc trưng trên khuôn mặt (landmark localization) và xác định vùng khuôn mặt (face alignment).





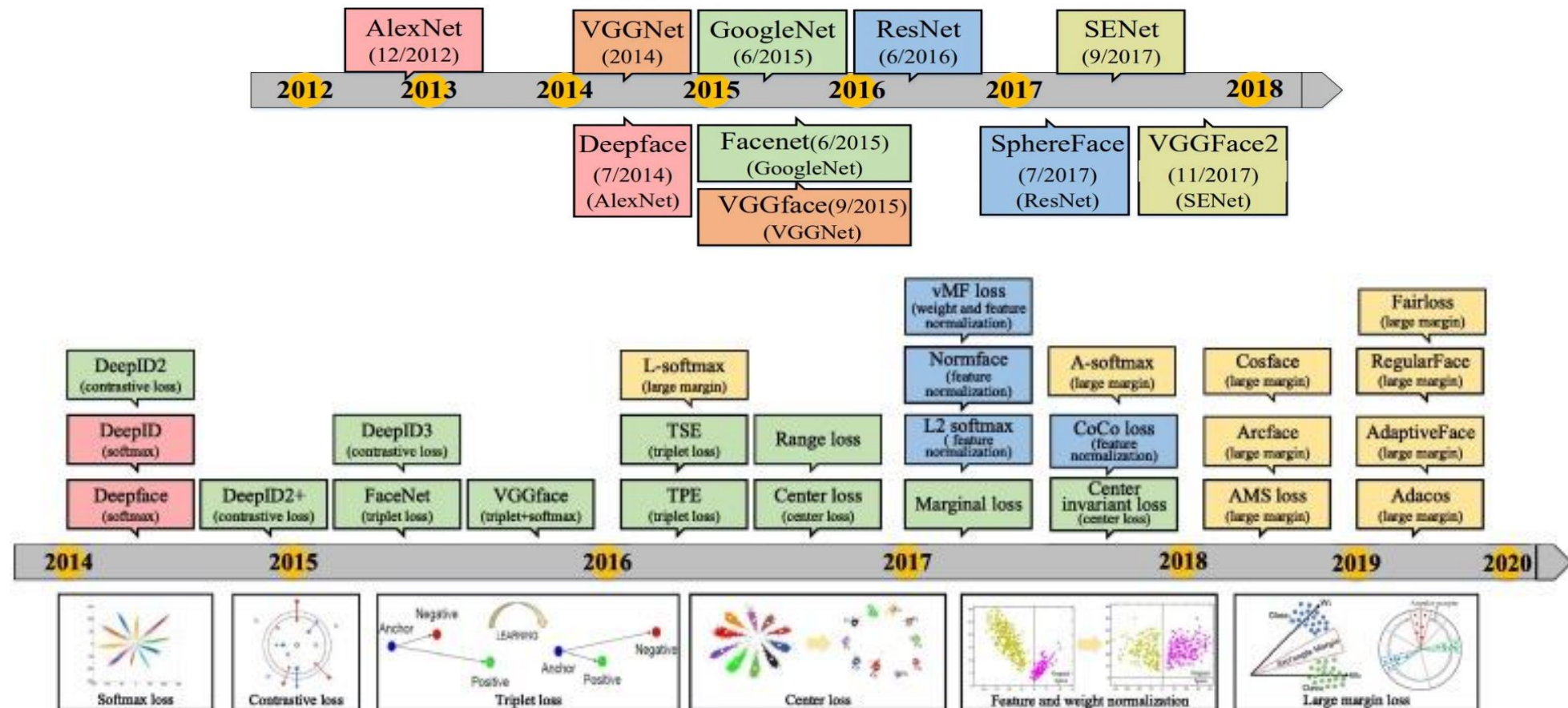
# Công việc triển khai<sup>2</sup>

- 1 Học agile scrum course
- 2 Phân tích bài toán
- 3 Phát hiện khuôn mặt
- 4 YOLOv7
- 5 Facial Landmarks
- 6 MTCNN
- 7 Face Classification**

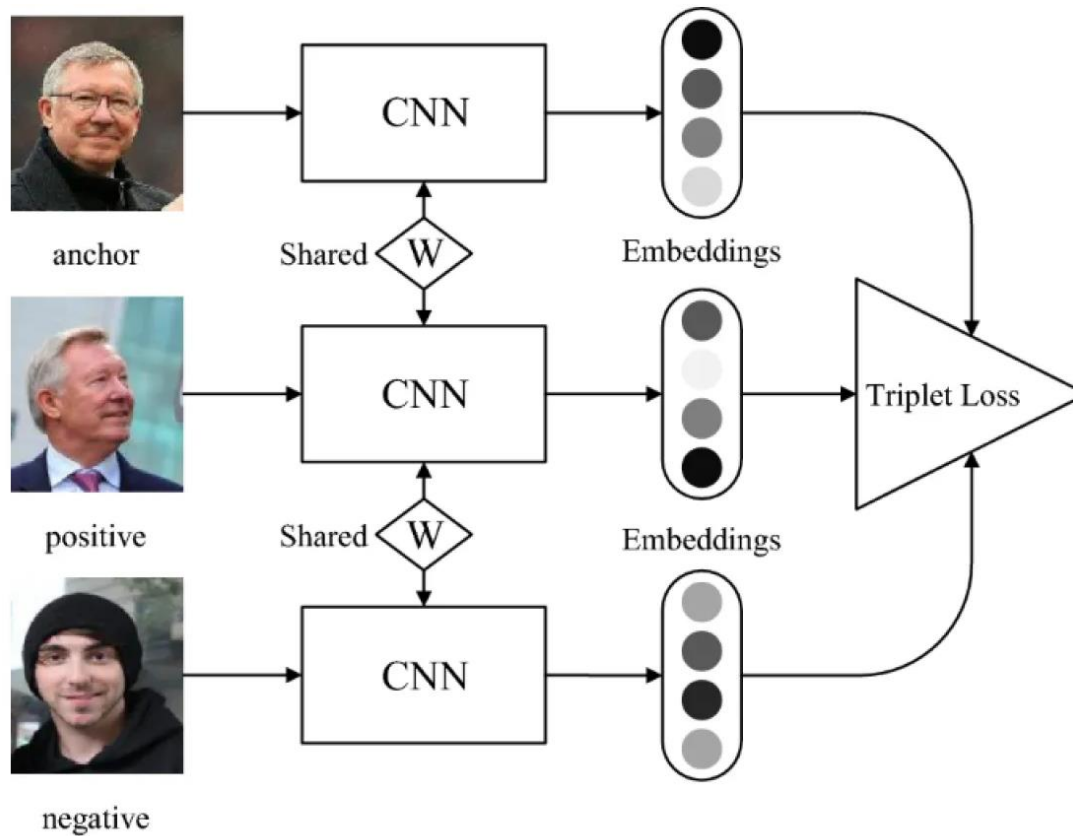


# Face Classification

Ảnh khuôn mặt sẽ được đi qua một bộ trích chọn đặc trưng mà đầu ra ở đây là một vector biểu thị cho khuôn mặt.



# Face Classification



Công thức hàm Triple loss

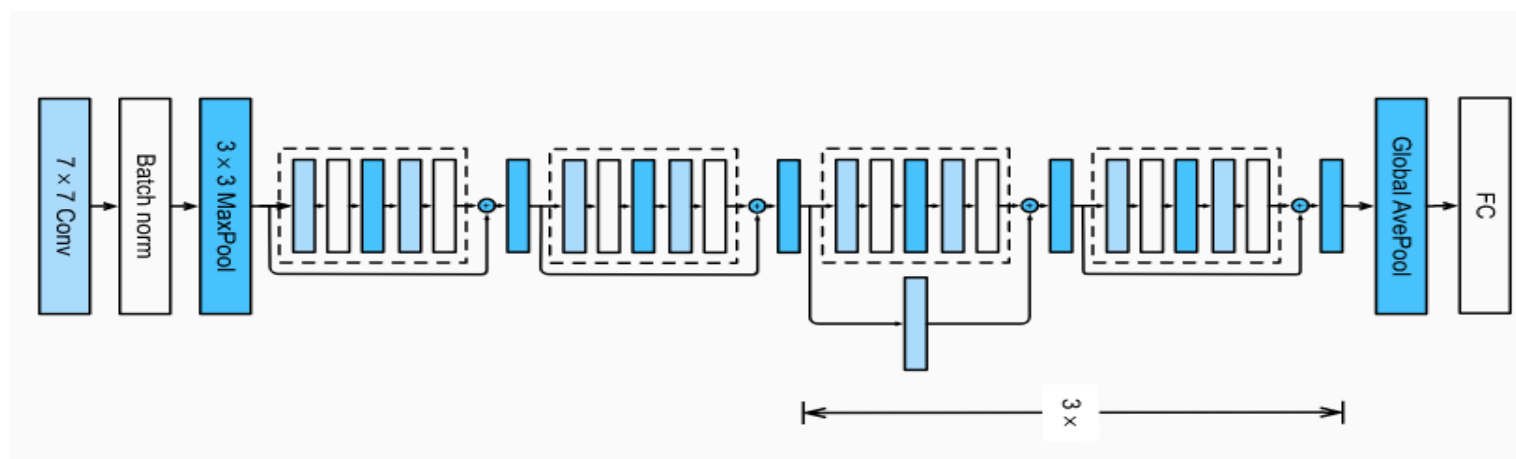
$$\sum_i^N \left[ \|f(x_i^a) - f(x_i^p)\|_2^2 - \|f(x_i^a) - f(x_i^n)\|_2^2 + \alpha \right]$$



# Face Classification

Dlib extract feature sử dụng mô hình mạng ResNet-29 conv layers và triplet loss. Mạng được train trên bộ dữ liệu khoảng 3 triệu khuôn mặt The face scrub dataset, the VGG dataset, và các ảnh trên internet (Tổng số người là 7485). Độ chính xác lên tới 99,38%.

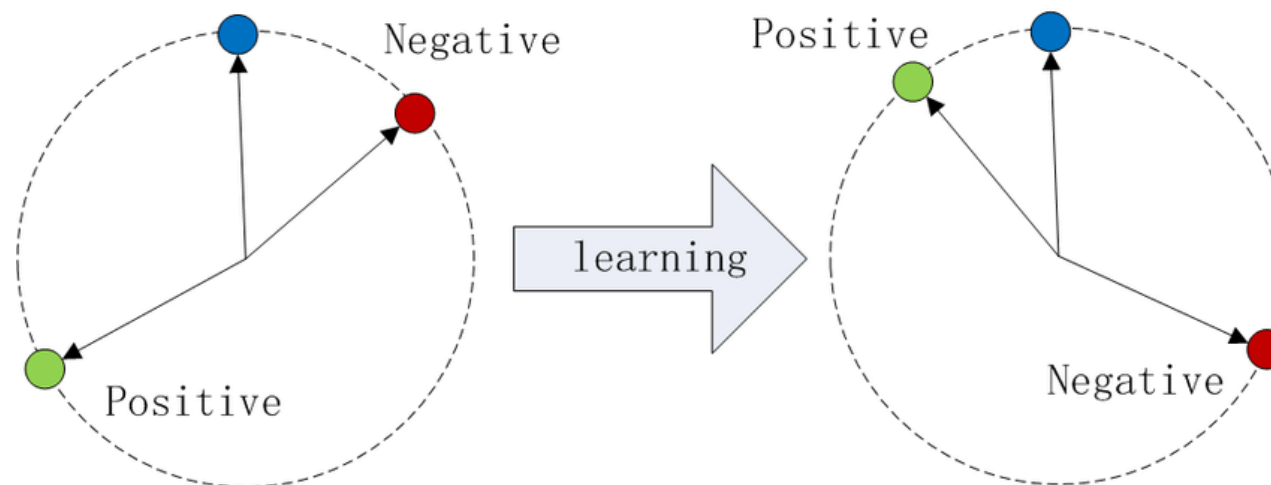
Mạng ResNet (R) - 29 là một mạng CNN được thiết kế để làm việc với 29 lớp chập. ResNet gồm có các lớp như convolution, pooling, activation và fully connected.



# Face Classification

Quá trình encoding của mạng CNN đã giúp ta mã hóa bức ảnh về 128 chiều. Sau đó những vectơ này sẽ làm đầu vào cho hàm loss function đánh giá khoảng cách giữa các véc tơ.

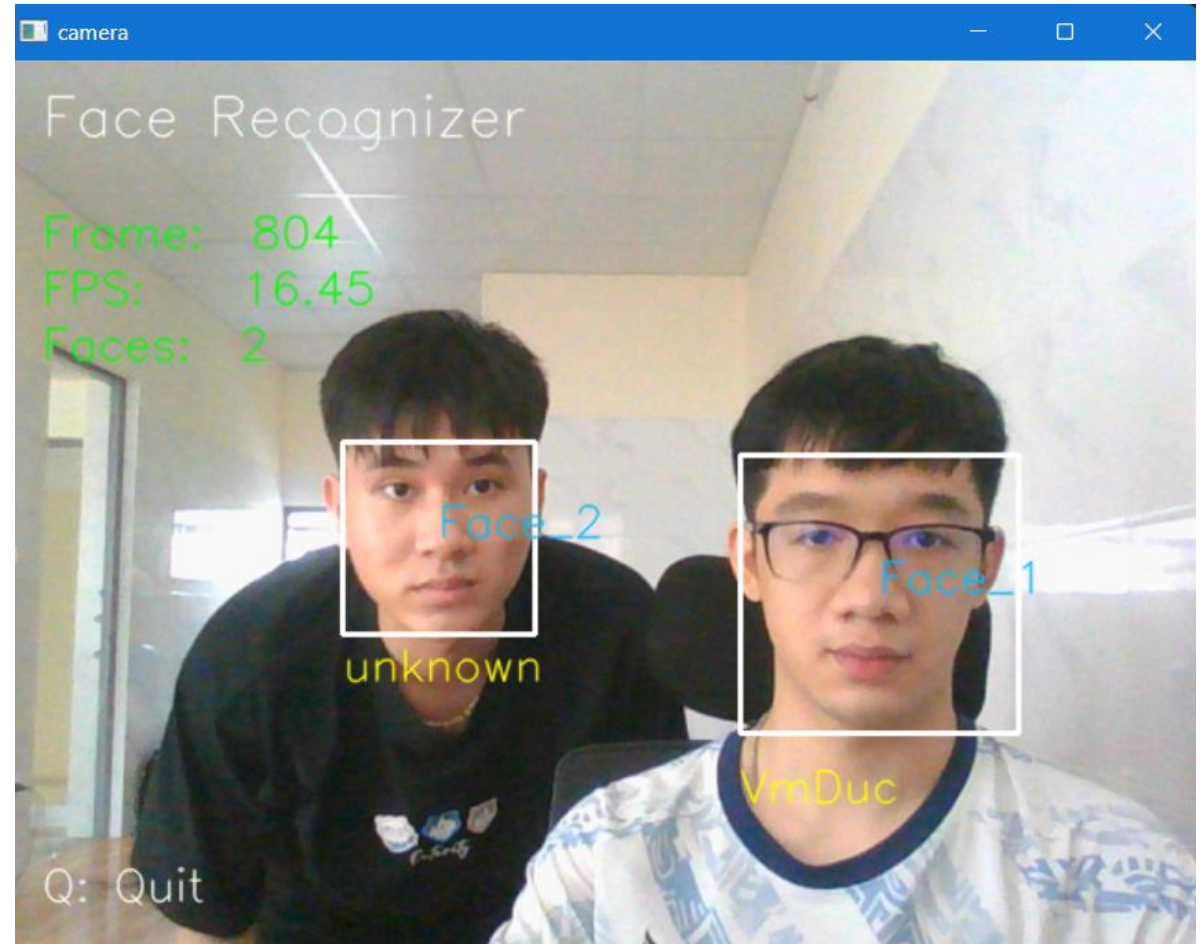
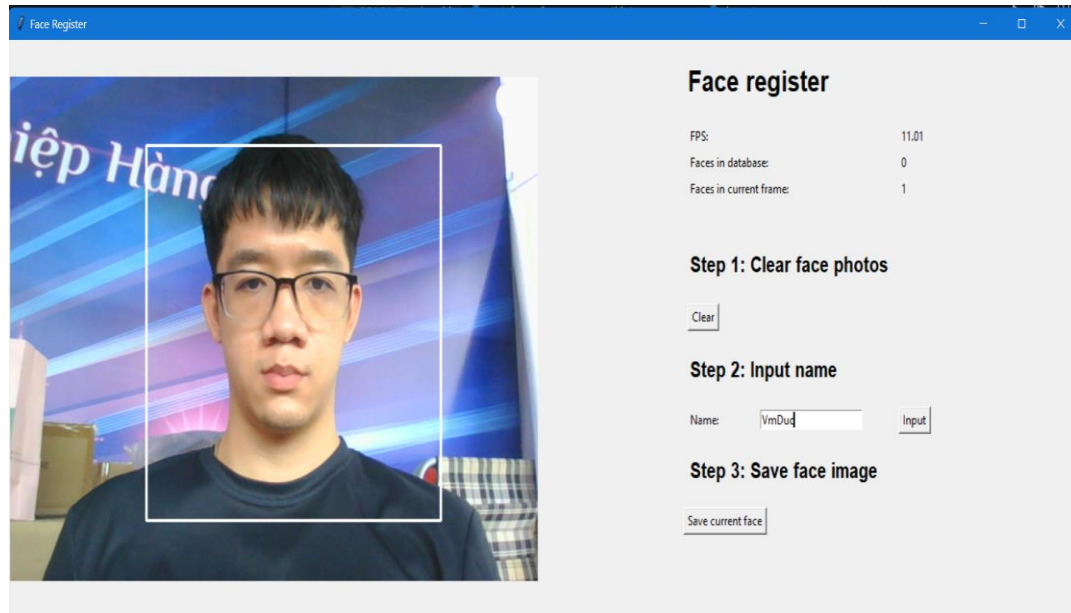
$$\mathcal{L}(A, P, N) = \sum_i^n \max(\|f(A_i) - f(p_i)\|_2^2 - \|f(A_i) - f(N_i)\|_2^2 + \alpha, 0)$$



# Nội dung

- 1 Giới thiệu
- 2 Công việc triển khai
- 3 Kết quả đạt được

# Kết quả đạt được



CẢM ƠN THẦY CÔ VÀ CÁC BẠN  
ĐÃ LẮNG NGHE