



Final Project

BÁO CÁO ĐỒ ÁN MÔN HỌC NHẬN DẠNG

NHÓM 07



DANH SÁCH THÀNH VIÊN

Thành viên 1

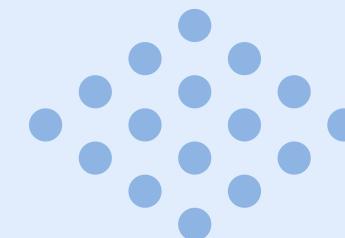
22120099

Trần Gia Hào

Thành viên 2

22120123

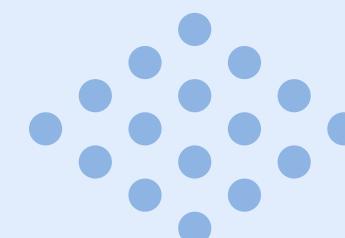
Nguyễn Minh Hưng



Thành viên 3

22120126

Nguyễn Tấn Hưng

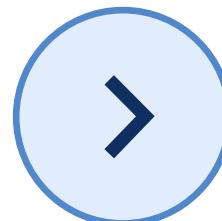


Thành viên 4

22120133

Hà Đức Huy

NỘI DUNG



NỘI DUNG CHƯƠNG SÁCH



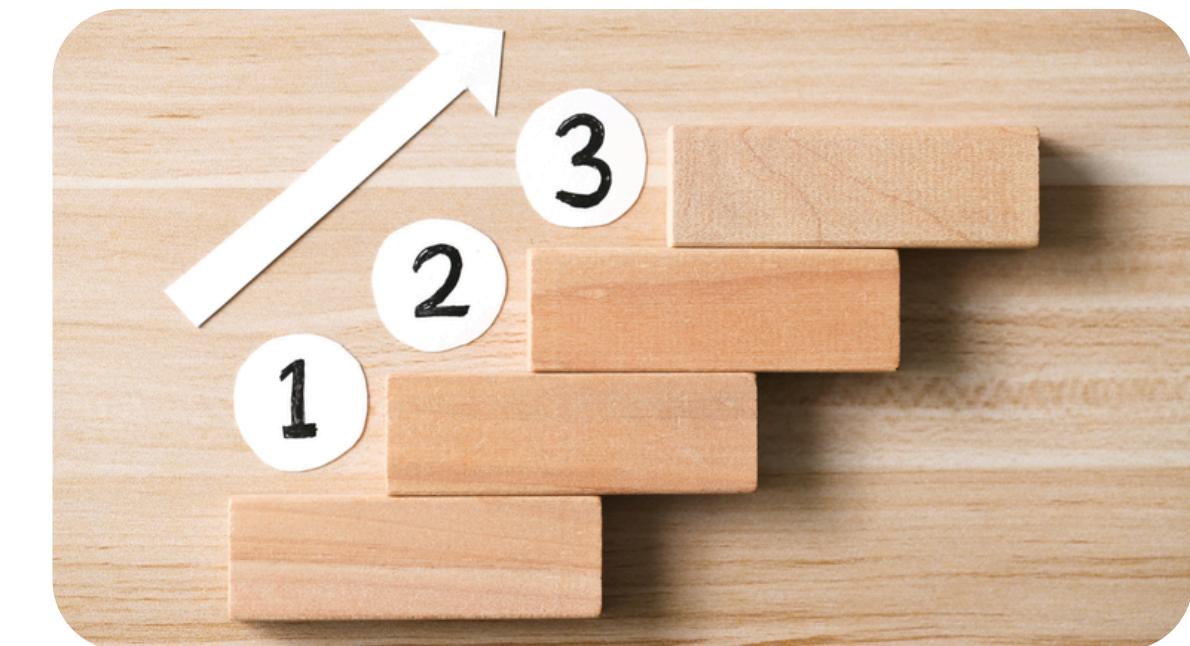
PHƯƠNG PHÁP TIÊN TIẾN

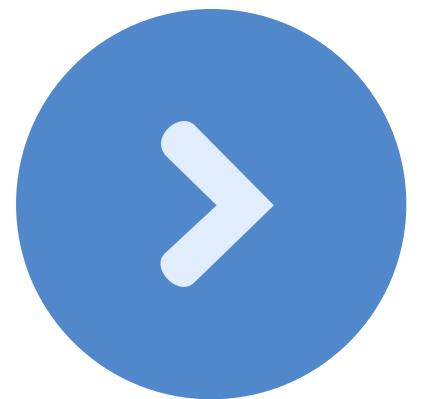


THỰC NGHIỆM

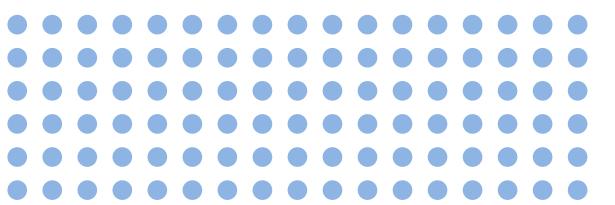


DEMO



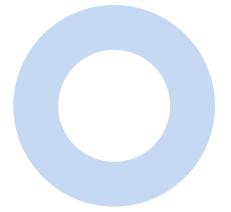


NỘI DUNG CHƯƠNG SÁCH



Introduction

Nhận dạng khuôn mặt & ánh sáng



- Nên dựa vào đặc điểm nội tại: Hình dạng 3D, độ phản xạ bề mặt khuôn mặt
- Ánh sáng, kính, tóc... là yếu tố ngoại tại cần giảm thiểu
- Thay đổi ánh sáng là thách thức chính

Hạn chế của hệ thống VIS (Hệ thống dựa trên hình ảnh ánh sáng nhìn thấy)

- Dễ bị ảnh hưởng bởi ánh sáng, nhất là góc chiếu sáng
- Ảnh cùng cùng 1 người khác ánh sáng có thể khác biệt nhiều hơn hai người khác nhau
- Bộ lọc cục bộ không khắc phục được thay đổi ánh sáng
- Hiệu suất giảm mạnh trong đánh giá thực tế (VD: FRVT - Face Recognition Vendor Test)

Giới thiệu

>>>

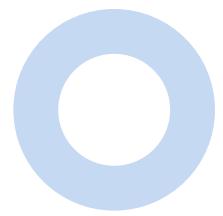


Giải pháp: Nhận dạng khuôn mặt dựa trên ánh sáng hồng ngoại gần (NIR)

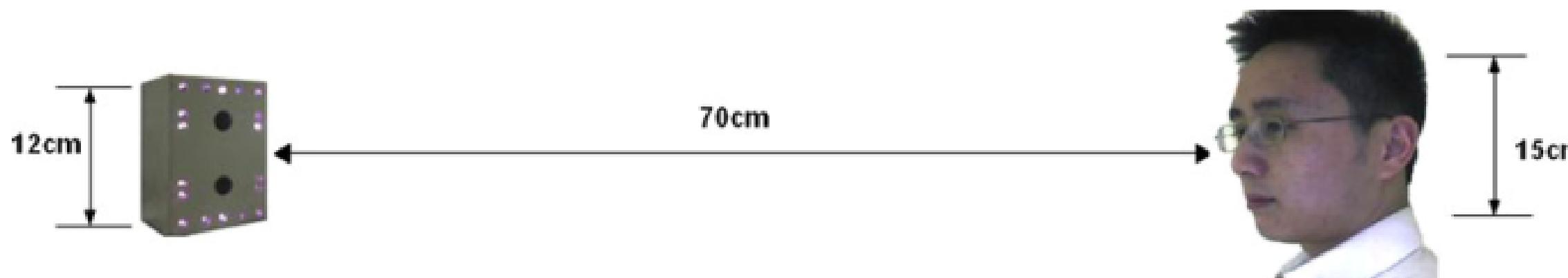
- NIR (Near Infrared) giúp chuẩn hóa chiếu sáng → Ổn định ảnh khuôn mặt.
- Dùng camera chuyên dụng, thu nhận ảnh khuôn mặt NIR chiếu sáng từ phía trước.
- Kết hợp biểu diễn đặc trưng như Local Binary Pattern (LBP) để loại bỏ ảnh hưởng ánh sáng.
- Biểu diễn khuôn mặt gần như không phụ thuộc ánh sáng.
- Hiệu quả hơn VIS trong môi trường ánh sáng không kiểm soát

Active NIR Imaging System

Hệ thống chụp ảnh NIR chủ động



- **Mục tiêu:** Loại bỏ ảnh hưởng ánh sáng môi trường, thu được ảnh khuôn mặt ổn định
- **Yêu cầu chính:**
 - Chiếu sáng chính diện, cường độ điểm ảnh phù hợp
 - Đèn không gây chói, hướng chiếu cố định
 - Ánh sáng NIR phải lấn át ánh sáng môi trường
- **Giải pháp:**
 - Dùng LED NIR 850–940 nm, gắn quanh ống kính
 - Sử dụng bộ lọc quang học để loại bỏ ánh sáng nhìn thấy

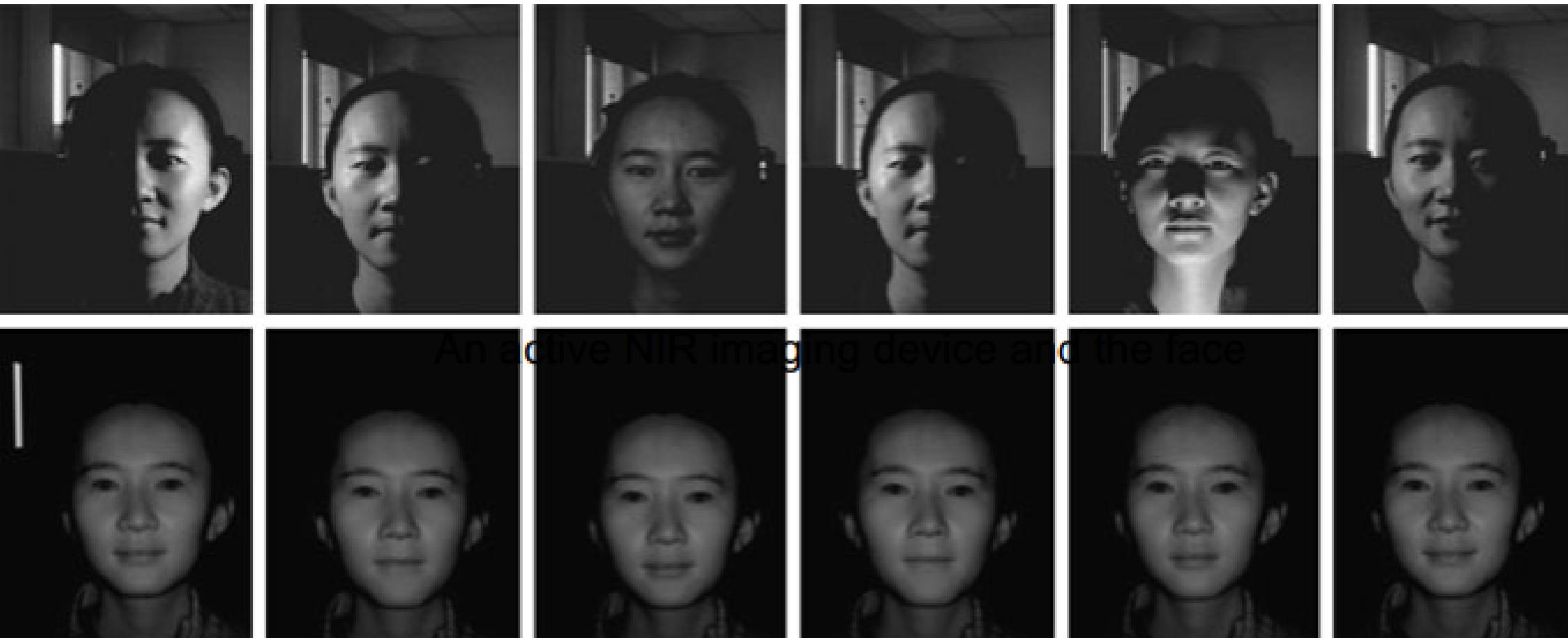


Active NIR Imaging System

Vấn đề dưới ánh sáng mặt trời

>>>>

- Ánh sáng mặt trời chứa nhiều thành phần NIR → Gây nhiễu mạnh
- Hệ thống NIR thông thường khó lẩn át ánh sáng mặt trời



Active NIR Imaging System

Hệ thống ENIR (NIR nâng cao)



- Giải quyết nhiễu từ ánh sáng mặt trời

- **Sử dụng:**

- Flash NIR công suất lớn
- Phơi sáng cực ngắn ($\sim 50 \mu\text{s}$)
- Đồng bộ flash và cảm biến
- Bộ lọc dải hẹp phù hợp bước sóng

- **Kỹ thuật xử lý:**

- Chụp 2 ảnh (có/không có NIR)
- Lấy hiệu hai ảnh \rightarrow khử nhiễu ánh sáng môi trường
- Cho ra ảnh chỉ với ánh sáng NIR chủ động



Hình 3: Từ trái sang phải: a. Ảnh khuôn mặt dưới sự kết hợp của ánh sáng mặt trời và ánh sáng hồng ngoại gần chủ động; b. Ảnh khuôn mặt chỉ có ánh sáng mặt trời; c. Ảnh hiệu số (a) – (b)

Biểu diễn khuôn mặt bất biến với ánh sáng

Mô hình ảnh NIR theo Lambertian



- Mô hình Lambertian:

$$I(x, y) = \rho(x, y) \cdot n(x, y) \cdot s$$

- Trong đó:
 - $\rho(x, y)$: Hệ số phản xạ (albedo) của vật liệu bề mặt khuôn mặt tại điểm (x, y)
 - $n(x, y)$: Vector đơn vị pháp tuyến của bề mặt tại điểm (x, y)
 - $s = (s_x, s_y, s_z)$: Vector cột biểu diễn hướng chiếu sáng, có độ lớn.
- Biểu thức có thể viết lại dưới dạng:

$$I(x, y) = \kappa \rho(x, y) \cos \theta(x, y)$$

- Tách biệt đặc tính nội tại và ngoại tại để tăng độ chính xác nhận diện
- Kiểm soát ánh sáng bằng NIR giúp ổn định nhận diện trong môi trường thay đổi

Biểu diễn khuôn mặt bất biến với ánh sáng

Bù trừ cho phép biến đổi đơn điệu (κ)



- **Vấn đề:**

- Biến đổi đơn điệu do hằng số κ làm thay đổi cường độ ảnh khuôn mặt
- Ảnh hưởng đến độ chính xác khi ánh sáng thay đổi

- **Giải pháp:** Dùng LBP (Local Binary Pattern) để biểu diễn khuôn mặt bất biến với ánh sáng

- **Đặc trưng LBP:**

- So sánh điểm ảnh trung tâm với 8 điểm xung quanh → Tạo mã nhị phân
- Tổng cộng có 256 mẫu LBP, với $LBP_{8,1}^{u^2}$ là dạng phổ biến và đồng nhất
- LBP có thể mở rộng thành LBP (P,R) với P điểm ảnh và bán kính R

Phân loại khuôn mặt NIR

Mục tiêu



- Lọc chọn đặc trưng LBP hữu ích nhất cho nhận diện.
- Giảm nhiễu, giảm chiều, tăng hiệu quả phân loại.

AdaBoost - Lựa chọn đặc trưng

- Biến bài toán nhiều lớp → bài toán hai lớp (intra-class vs. inter-class).
- AdaBoost chọn các đặc trưng histogram LBP phân biệt tốt nhất.
- Xây dựng bộ phân loại mạnh từ nhiều bộ phân loại yếu:

$$H(x) = \text{sign}\left(\sum_{t=1}^T \alpha_t h_t(x)\right)$$

- Ưu điểm:
 - Giảm số đặc trưng cần thiết → Tăng tốc tính toán.
 - Tăng độ chính xác nhờ tập trung vào đặc trưng phân biệt cao.

Phân loại khuôn mặt NIR

LDA Classifier



- LDA chiếu dữ liệu xuống không gian con để tối ưu phân biệt giữa các lớp.
- Kết hợp PCA + LDA → Tạo ra Fisherfaces, giúp giảm chiều và khử nhiễu.
- Cosine similarity được dùng để so sánh vector chiếu của ảnh mới và ảnh đăng ký sẵn.
- So sánh với ngưỡng → Quyết định có cùng người hay không.

Ưu điểm trong ảnh hồng ngoại (NIR):

- Tăng độ phân biệt giữa khuôn mặt.
- Giảm nhiễu và tăng tốc xử lý.
- Cải thiện độ chính xác nhận diện trong điều kiện ánh sáng phức tạp.

Đánh giá

Mục tiêu: Đánh giá hiệu suất nhận diện khuôn mặt NIR trong nhiều điều kiện: ánh sáng yếu, đeo kính, thời gian trôi qua, và môi trường ngoài trời.



Đánh giá cơ bản:

- **Phương pháp tốt nhất:** LBP + AdaBoost.
- **VR @ FAR=0.1%:**
 - LBP + AdaBoost: 91.8%
 - LBP + LDA: 69.9%
 - PCA (NIR): 32.0%
- **Ý nghĩa:** LBP + AdaBoost cho hiệu suất cao nhất.

Ánh sáng yếu:

- NIR phân biệt cặp trong/ngoài lớp rõ ràng hơn VIS.
- **Kết luận:** NIR ổn định hơn VIS trong điều kiện ánh sáng yếu.

Đánh giá

Khi đeo kính:

- Hiệu suất giảm nhẹ so với không đeo kính ($VR = 87.1\%$).
- **Kết luận:** Hệ thống vẫn tách biệt tốt trong lớp và ngoài lớp.



Time Lapse:

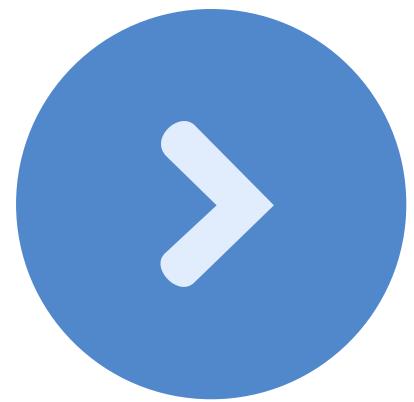
- VR giảm còn **83.24%** sau 1 năm.
- **Kết luận:** Hệ thống NIR vẫn duy trì độ chính xác cao dù có thay đổi theo thời gian.

Môi trường ngoài trời:

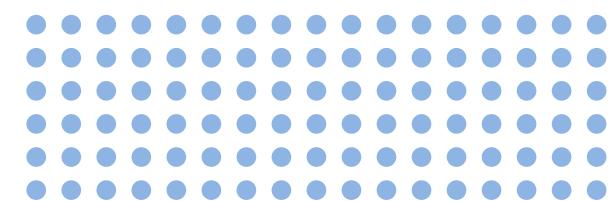
- ENIR vượt trội so với VIS/NIR trong ánh sáng ngoài trời.
- VR @ FAR=0.01:
 - ENIR: 69% (trực diện), 85% (trong nhà).
- **Kết luận:** ENIR hoạt động tốt trong môi trường ánh sáng phức tạp.

Kết luận

- **Giải pháp:** Kết hợp hình ảnh NIR chủ động với thuật toán LBP + AdaBoost + LDA để [>>>>](#) tạo hệ thống nhận diện khuôn mặt bất biến ánh sáng.
- **Kết quả:** Hệ thống đạt độ chính xác cao, tốc độ nhanh, hoạt động ổn định dưới ánh sáng thay đổi, hiệu quả với nhiều nhóm dân tộc, vượt trội hơn các phương pháp cũ.
- **Ứng dụng:** Phù hợp cho các hệ thống 1-đối-nhiều, đặc biệt trong điều kiện người dùng hợp tác và môi trường ánh sáng phức tạp.



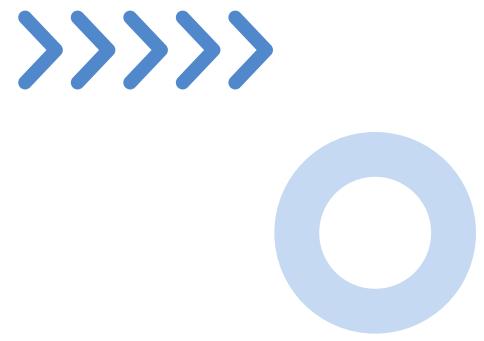
PHƯƠNG PHÁP TIÊN TIẾN





Tổng quan

- Hệ thống nhận diện khuôn mặt sử dụng ảnh NIR đã được đề xuất bởi **Athreya V. Shet et al. (CSITSS 2022)**.
- Nghiên cứu sử dụng camera IR thu dữ liệu từ CCTV thực tế, kết hợp với các kỹ thuật như:
 - **Phát hiện khuôn mặt:** Haar-Cascade, YOLO-Face, MTCNN.
 - **Nhận diện khuôn mặt:** Embedding từ mô hình VGG-Face, phân loại bằng Softmax Regression.



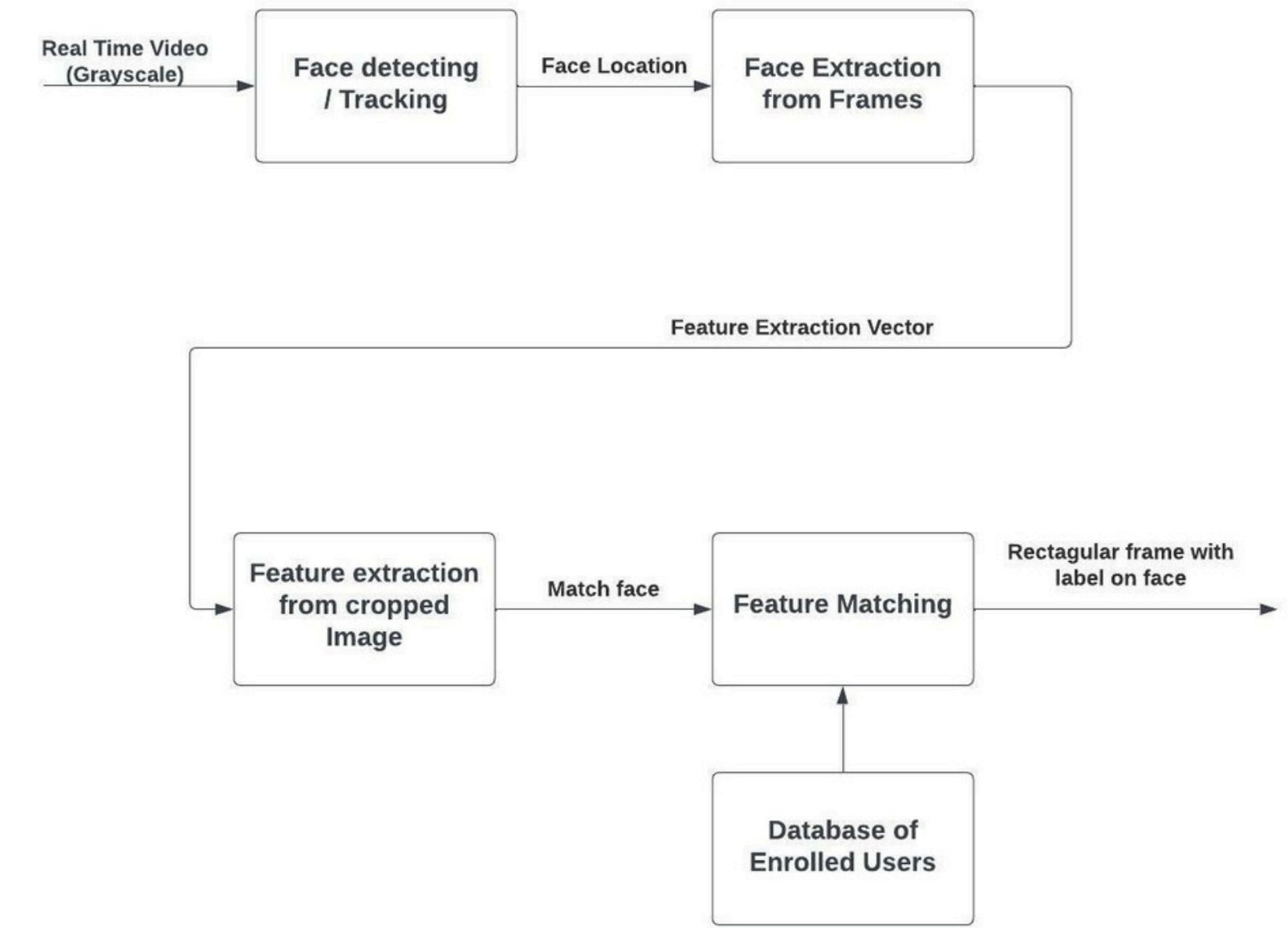
Phương pháp thực hiện

Thu thập dữ liệu: Tập dữ liệu huấn luyện được thu từ camera giám sát IR (CCTV có hồng ngoại) trong điều kiện ánh sáng yếu. Mỗi người trong tập dữ liệu có từ 40-50 ảnh NIR chụp nhiều góc khác nhau.

Phát hiện khuôn mặt: Khuôn mặt được phát hiện bằng 3 kỹ thuật khác nhau. Sau khi phát hiện, bounding box được vẽ và ảnh được cắt từ vùng này để đưa vào bước nhận diện.

Nhận diện khuôn mặt

- Sử dụng mô hình VGG-Face (CNN) để trích xuất embedding vector cho mỗi khuôn mặt.
- Các vector này được đưa vào Softmax Regression đã huấn luyện để so sánh và gán nhãn khuôn mặt.



Phát hiện khuôn mặt

Nhằm tăng độ chính xác và khả năng bao phủ trong các điều kiện ánh sáng khác nhau, nhóm tác giả sử dụng ba phương pháp phát hiện khuôn mặt:

- **Haar-Cascade (OpenCV)**: Dựa trên đặc trưng Haar và AdaBoost, nhanh nhưng kém chính xác trong môi trường phức tạp.
- **YOLO-Face**: Cải tiến từ YOLOv3, tối ưu anchor boxes cho khuôn mặt, giúp phát hiện nhanh và chính xác.
- **MTCNN**: Gồm 3 mạng nơ-ron (P-Net, R-Net, O-Net), có độ chính xác cao nhất, phát hiện ổn định kể cả trong điều kiện khó.

Face Detection Techniques Comparison

Dataset No	Haar-Cascade		YOLO-Face		MTCNN	
	Total frames detected	Actual frames with faces	Total frames detected	Actual frames with faces	Total frames detected	Actual frames with faces
1	259	259	282	282	280	280
2	62	62	57	57	63	63
3	226	226	187	185	240	240
4	193	174	197	196	234	234
5	302	284	297	297	300	300
6	225	220	232	232	233	233
7	270	239	280	280	282	282
8	292	288	292	292	300	300
9	88	83	111	111	135	135
10	110	90	157	157	160	160
11	264	263	270	268	273	273
12	281	277	279	279	296	296
13	249	208	254	254	278	266
14	66	46	94	94	101	99
15	187	132	289	289	300	297

Table 2: Detection algorithms result on the dataset created

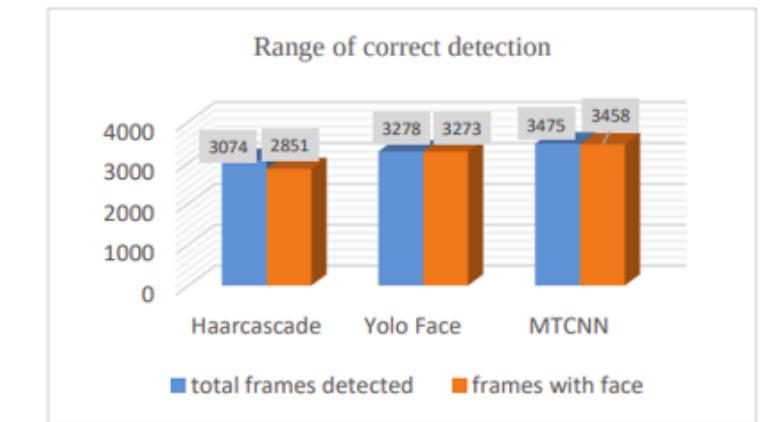


Figure 7: Detector capability of detecting faces on our dataset

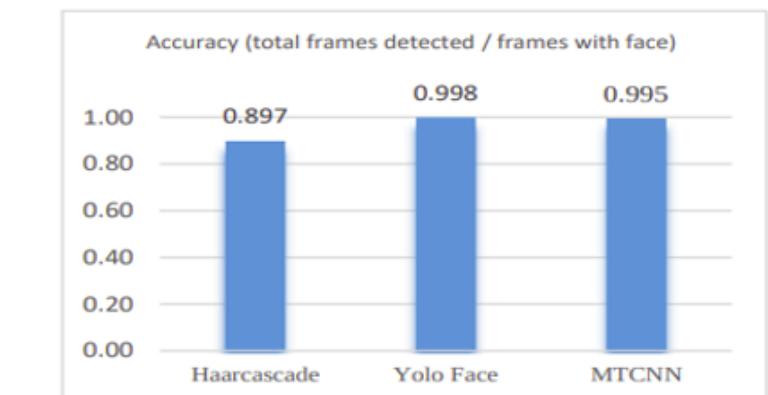
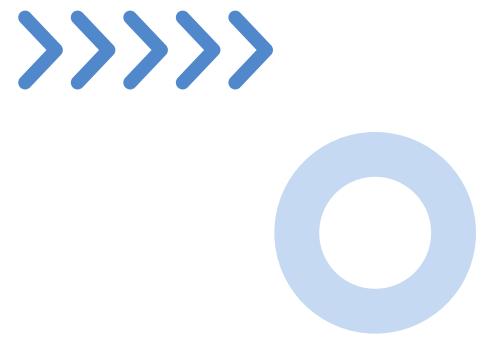


Figure 6: Accuracy of different detection techniques on our dataset

Có thể thấy MTCNN phát hiện được nhiều khuôn mặt hơn, YOLO-Face nhanh hơn; cả hai đều vượt trội hơn Haar-Cascade.



Nhận diện khuôn mặt

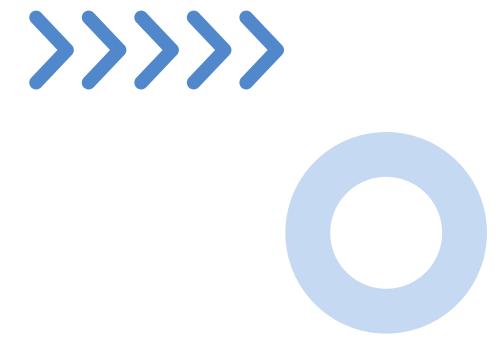
Mục tiêu:

- Huấn luyện một mô hình phân loại nhằm gán nhãn chính xác cho các khuôn mặt mới, dựa trên các khuôn mặt đã biết được lưu trữ trong tập dữ liệu.
- Quá trình nhận diện được thực hiện sau khi khuôn mặt đã được cắt từ ảnh bởi các thuật toán phát hiện.

Phương pháp thử nghiệm ban đầu:

- **FaceNet** (Google, dựa trên kiến trúc Inception, công bố năm 2015)
- **DeepFace Framework**

Tuy nhiên, khi áp dụng lên ảnh hồng ngoại gần (NIR) thu được từ hệ thống camera giám sát, các mô hình này không đạt độ chính xác như kỳ vọng do đặc tính dữ liệu khác biệt (ánh sáng yếu, không màu, nhiều nhiễu...).

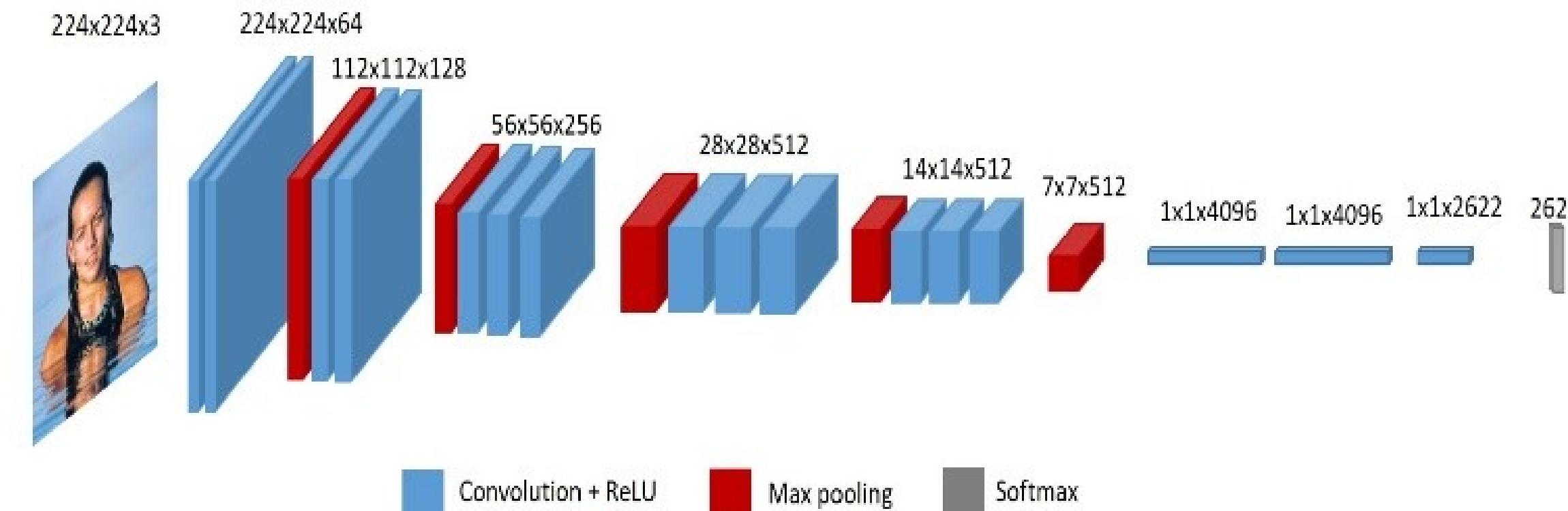


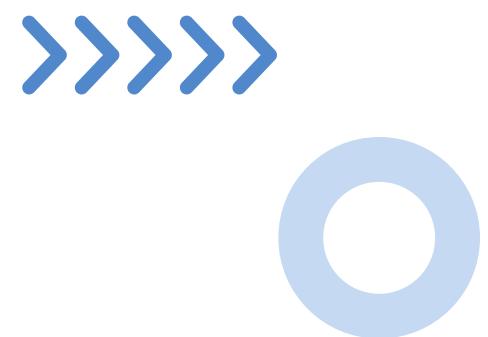
Nhận diện khuôn mặt

Giải pháp: Nhóm tác giả đã thiết kế một mô hình nhận diện mới, với cấu trúc như sau:

- **Trích đặc trưng khuôn mặt bằng mô hình VGG-Face:**

- VGG-Face được huấn luyện trên 2.6 triệu ảnh của 2.622 người.
- Mỗi ảnh đầu vào được ánh xạ thành **một vector đặc trưng 2.622 chiều (embeddings)**.
- Những embeddings này được sử dụng làm đầu vào cho bước phân loại tiếp theo.





Nhận diện khuôn mặt

Giải pháp: Nhóm tác giả đã thiết kế một mô hình nhận diện mới, với cấu trúc như sau:

- **Bộ phân loại Softmax tùy chỉnh, gồm:**

- **Lớp 1:** 100 đơn vị ẩn, hàm kích hoạt tanh.
- **Lớp 2:** 10 đơn vị ẩn, kích hoạt tanh.
- **Lớp 3:** 6-15 đơn vị (dựa vào số người trong tập dữ liệu), theo sau là hàm softmax để sinh phân phối xác suất.

Ưu điểm giải pháp:

- Nhận diện hiệu quả trong điều kiện ánh sáng yếu hoặc môi trường ngoài trời.
- Độ chính xác cao và tốc độ xử lý nhanh.
- Khả năng phân biệt rõ ràng giữa các cá nhân nhờ biểu diễn đặc trưng VGG-Face.

Kết quả mô hình

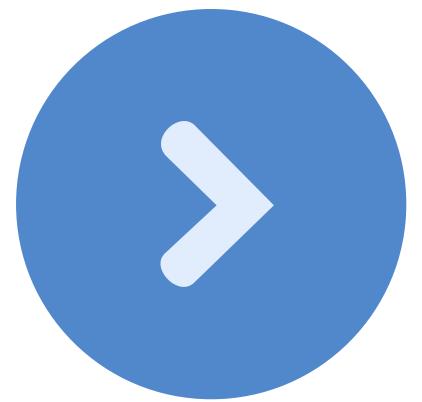
- Mô hình cuối cùng tích hợp cả hai bước phát hiện và nhận diện khuôn mặt trong một lần xử lý duy nhất, hoạt động thực tế trên video hồng ngoại (NIR) theo thời gian thực.
- Sử dụng mô hình nhận diện tùy chỉnh dựa trên VGG-Face embeddings, các khuôn mặt trong khung hình được gắn nhãn chính xác theo tập dữ liệu đã huấn luyện.



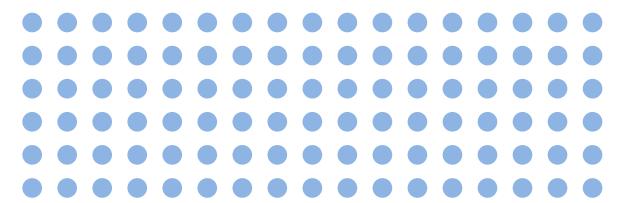
Figure 11: Real time testing of the model (sample image)

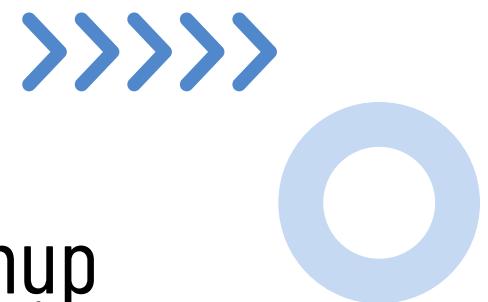
Frame rate of detection techniques along with vggface recognition. (fps, performed in 1.80Ghz cpu and Nvidia MX250)		
Detection Technique	Without recognition	With recognition
Haar-cascade	9.46	5.03
Yolo-face	1.16	0.80
Mtcnn	0.79	0.59

Table 2: Speed comparison of different approaches.



THỰC NGHIỆM





Chuẩn bị dữ liệu

- **Bộ dữ liệu sử dụng:** TD-NIR-A-Set1 - Gồm 815 ảnh NIR từ 25 người, với nhiều góc chụp (chính diện, nghiêng, trên xuống, dưới lên...).
- **Chia tập dữ liệu:** Chia theo tỷ lệ Train : Validation : Test = 7 : 2 : 1
- **Tăng cường dữ liệu trên tập huấn luyện:**
 - Xoay, lật, dịch chuyển ảnh
 - Thay đổi độ sáng, thêm nhiễu Gaussian
- **Kết quả thu được:**

Tập dữ liệu	Số lượng ảnh
Tập huấn luyện (Train)	2,794 ảnh
Tập xác thực (Validation)	165 ảnh
Tập kiểm tra (Test)	91 ảnh

Phát hiện khuôn mặt

Nhóm thực hiện các bước trong quy trình phát hiện khuôn mặt như sau:



- **Bước 1 - Tích hợp YOLOFace**

- Sử dụng **YOLOFace** - một biến thể của YOLOv3 được huấn luyện chuyên biệt cho phát hiện khuôn mặt.
- Được huấn luyện trên tập dữ liệu **WIDER FACE**, có khả năng xử lý tốt nhiều điều kiện như ánh sáng yếu, góc nghiêng, ảnh mờ.

- **Bước 2 - Phát hiện và cắt khuôn mặt**

- YOLOFace tự động xác định khuôn mặt và vẽ bounding box.
- Ảnh trong vùng này được cắt và chuẩn hóa để làm đầu vào cho bước nhận diện.

- **Bước 3 - Chia tập dữ liệu (train/val/test):** Sau khi cắt ảnh khuôn mặt bằng YOLOFace, chia dữ liệu thành 3 tập train/val/test

Nhận diện khuôn mặt

>>>



Nhóm thực hiện các bước trong quy trình nhận diện khuôn mặt như sau:

- **Bước 1 - Nạp mô hình VGG-Face:** Sử dụng mô hình VGG-Face được huấn luyện sẵn, loại bỏ lớp Softmax để sử dụng như một bộ trích xuất đặc trưng khuôn mặt (embedding extractor).
- **Bước 2 - Chuẩn bị dữ liệu huấn luyện (train):** Trích xuất vector đặc trưng từ ảnh khuôn mặt bằng VGG-Face.
 - Resize ảnh về 224×224 .
 - Tiền xử lý ảnh (preprocess_input).
 - Trích đặc trưng bằng VGG-Face (bỏ lớp softmax).
 - Lưu **x_train** (vector đặc trưng), **y_train** (nhãn), **person_rep** (tập ảnh xạ nhãn).
- **Bước 3 - Chuẩn bị dữ liệu kiểm thử (val) và kiểm tra (test):** Xử lý tương tự bước 2 cho tập validation và tập test → Thu được **x_val**, **x_test**, **y_val**, **y_test**.

Nhận diện khuôn mặt



Nhóm thực hiện các bước trong quy trình nhận diện khuôn mặt như sau:

- **Bước 4 - Lưu trữ dữ liệu đặc trưng:** Lưu các vector đặc trưng và nhãn thành file .npy để để có thể tái sử dụng sau này.
- **Bước 5 - Xây dựng mô hình phân loại Softmax:**
 - **Mô hình:** Keras Sequential gồm 2 lớp ẩn và 1 lớp đầu ra:
 - Dense(100) + BatchNorm + tanh + Dropout(0.3)
 - Dense(6) + BatchNorm + tanh + Dropout(0.2)
 - Dense(25) + softmax (cho 25 người)
 - **Loss:** SparseCategoricalCrossentropy
 - **Optimizer:** Nadam

Nhận diện khuôn mặt

>>>



Nhóm thực hiện các bước trong quy trình nhận diện khuôn mặt như sau:

- **Bước 6 - Huấn luyện và đánh giá mô hình:**
 - **Huấn luyện:** Dùng tập **x_train, y_train**, chạy 80 epoch.
 - **Đánh giá:** Trên **x_val, y_val** và **x_test, y_test**.
 - **Kết quả:** train_accuracy = 0.8285, train_loss = 0.4171
- **Bước 7 - Lưu mô hình đã huấn luyện:** Mô hình được lưu thành file .h5 để tái sử dụng sau này

Thử nghiệm

>>>>

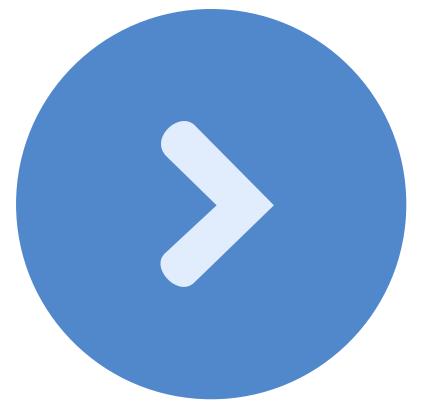


- Nhóm kiểm thử thực tế bằng cách nhận diện một số ảnh bất kỳ từ tập test.
- Mô hình dự đoán ảnh thuộc lớp 18 với độ tin cậy khoảng 95%.
- Kết quả cho thấy hệ thống nhận diện chính xác và nhanh trên dữ liệu thực tế.

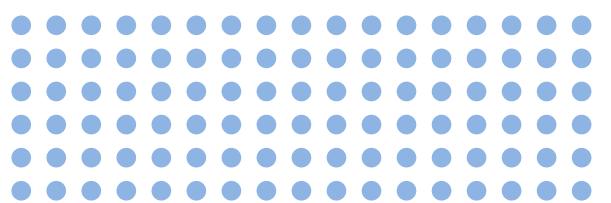
```
# Sử dụng hàm
os.makedirs(r'/content/output', exist_ok=True)
detect_faces("/content/TD_NIR_A_Set/test/18/TD_NIR_A_1_6.jpg", "/content/output", face, vgg_face, classifier_model, person_rep)

1/1 ━━━━━━━━ 1s 595ms/step
Phát hiện: 18 (score = 0.95)
```





DEMO



THỰC HIỆN TRÌNH BÀY



Datasets



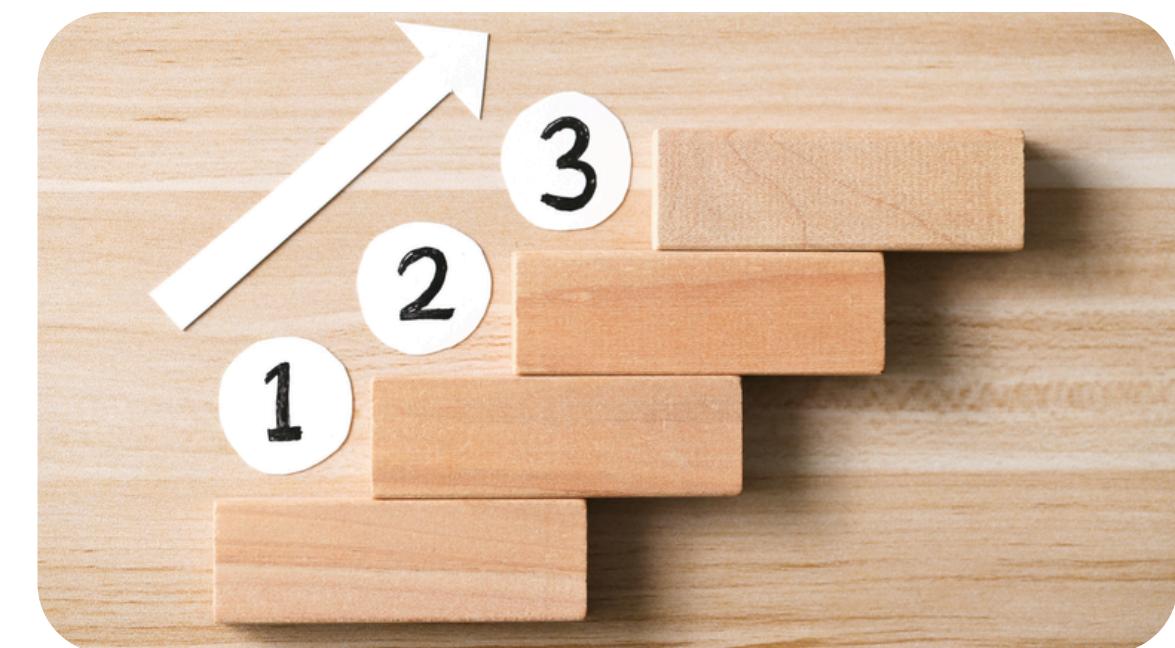
Face Detection

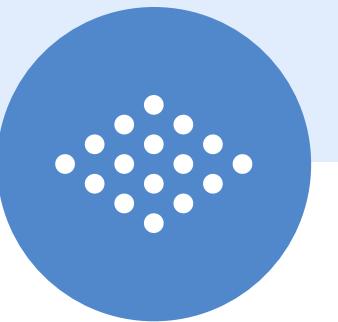


Face Recognition



Kết quả trả về





THANK YOU!

