

ĐỀ TÀI: NHẬN DIỆN KHUÔN MẶT KHI ĐEO KHẨU TRANG

TÊN HỌC PHẦN: TRÍ TUỆ NHÂN TẠO

GIẢNG VIÊN: LÊ TRUNG HIẾU

1 st Lê Thiên Dũng Lớp CNTT 17-11 Khoa công nghệ thông tin Đại Học Đại Nam Hà Nội, Việt Nam MSV: 1771020179	2 nd Lê Đức Thọ Lớp CNTT 17-11 Khoa công nghệ thông tin Đại Học Đại Nam Hà Nội, Việt Nam MSV: 1771020651	3 th Hà Quang Dự Lớp CNTT 17-11 Khoa công nghệ thông tin Đại Học Đại Nam Hà Nội, Việt Nam MSV: 1771020159	4 th Phạm Thị Ngọc Thanh Lớp CNTT 17-11 Khoa công nghệ thông tin Đại Học Đại Nam Hà Nội, Việt Nam MSV: 1771020628
---	--	---	---

Tóm tắt nội dung—Đề tài nghiên cứu nhận diện khuôn mặt khi đeo khẩu trang bằng kỹ thuật ArcFace. Hệ thống cải thiện độ chính xác khi khuôn mặt bị che khuất một phần. Mô hình được huấn luyện trên dữ liệu thực tế và đánh giá hiệu suất. Ứng dụng trong kiểm soát an ninh, thiết bị thông minh và giám sát.

Index Terms—Face Recognition, ArcFace, Masked Face Recognition, Deep Learning, Security

I. GIỚI THIỆU ĐỀ TÀI

A. Đặt vấn đề

Trong thời gian dịch bệnh COVID-19 diễn biến căng thẳng và phức tạp vừa qua, đa số khi các tỉnh thành, vùng phát dịch bệnh hầu hết đều có liên quan đến trung tâm thương mại. Cụ thể, tháng 6 tại tỉnh Bắc Giang có ca bệnh liên quan đến siêu thị BigC hay tại Hà Nội, vào đợt giãn cách xã hội tháng 8 có chùm ca bệnh liên quan đến hệ thống Vinmart. Hiện tại, dịch COVID-19 vẫn đang diễn biến khá phức tạp trên cả nước, đặc biệt là khu vực miền Nam. Trong những ngày qua, các trung tâm thương mại miền Bắc, cụ thể là tại Hà Nội được phép hoạt động trở lại, tuy nhiên, đây là nơi tiếp nhận lượng khách hàng đông đảo nên khả năng lây nhiễm dịch bệnh là rất cao. Chính vì vậy, các trung tâm thương mại cần có những biện pháp phòng chống bệnh đang bùng phát ở nước ta. Một biện pháp dễ thực hiện và có khả năng phòng ngừa lây nhiễm chéo cao là kiểm soát người đeo khẩu trang khi đến trung tâm thương mại. Tuy nhiên, vẫn còn nhiều khách hàng chưa tự nhận thức được sự nguy hiểm của COVID-19 mang đến cho chính mình và người xung quanh, vẫn mang tinh thần chủ quan không đeo khẩu trang hoặc đeo một cách đối phó khi tham gia mua sắm. Từ đó có thể khiến tình hình dịch bệnh diễn biến phức tạp hơn, nhà nước khó khăn hơn trong việc phòng ngừa, ngăn chặn dịch bệnh lây lan.

Identify applicable funding agency here. If none, delete this.

B. Bài toán đặt ra

Nhận diện khuôn mặt là một trong những công nghệ quan trọng trong lĩnh vực trí tuệ nhân tạo (AI) và thị giác máy tính (Computer Vision), được ứng dụng rộng rãi trong an ninh, xác thực danh tính, giám sát và nhiều lĩnh vực khác. Các hệ thống nhận diện truyền thống thường đạt độ chính xác cao khi khuôn mặt không bị che khuất. Tuy nhiên, trong những năm gần đây, đặc biệt từ khi đại dịch COVID-19 bùng phát, việc đeo khẩu trang trở thành một thói quen phổ biến, tạo ra thách thức lớn đối với các hệ thống nhận diện khuôn mặt.

Khẩu trang che phủ một phần lớn khuôn mặt, làm mất đi các đặc trưng quan trọng như miệng, cằm và phần dưới mũi. Điều này khiến các mô hình nhận diện khuôn mặt truyền thống gặp khó khăn trong việc xác định danh tính, dẫn đến độ chính xác giảm đáng kể. Nhiều nghiên cứu đã được thực hiện nhằm cải thiện khả năng nhận diện trong điều kiện này, trong đó phương pháp ArcFace được đánh giá là một trong những kỹ thuật hiệu quả nhất. ArcFace là một thuật toán nhận diện khuôn mặt tiên tiến, sử dụng học sâu (Deep Learning) để tối ưu hóa không gian đặc trưng (feature space), giúp phân biệt tốt hơn giữa các khuôn mặt, ngay cả khi bị che khuất một phần.

Trong nghiên cứu này, áp dụng ArcFace để xây dựng một mô hình nhận diện khuôn mặt khi đeo khẩu trang, nhằm cải thiện độ chính xác và tính ổn định của hệ thống. Mô hình được huấn luyện trên tập dữ liệu chứa cả khuôn mặt có và không đeo khẩu trang, từ đó đánh giá khả năng nhận diện trong điều kiện thực tế. Kết quả nghiên cứu có thể được ứng dụng trong nhiều lĩnh vực như kiểm soát an ninh tại sân bay, bệnh viện, công ty, xác thực danh tính trên thiết bị di động và hệ thống giám sát thông minh.

Mục tiêu của nghiên cứu này là tối ưu hóa thuật toán nhận diện khuôn mặt trong điều kiện bị che khuất, đồng thời so sánh hiệu suất của ArcFace với các phương pháp khác. Qua đó, đề xuất một giải pháp hiệu quả giúp nâng cao độ chính xác của hệ

thống nhận diện khuôn mặt trong bối cảnh thực tế.

C. Phạm vi, đối tượng và những khó khăn của đề tài

- **Phạm vi nghiên cứu:** Hệ thống nhận diện khuôn mặt kết hợp kiểm tra đeo khẩu trang được áp dụng trong nhiều lĩnh vực như bảo mật, kiểm soát ra vào, điểm danh tự động, giám sát an ninh và hỗ trợ y tế. Phạm vi nghiên cứu này tập trung vào:
 - Phát hiện khuôn mặt: Sử dụng mô hình `face_recognition` để xác định vị trí khuôn mặt trên ảnh hoặc video đầu vào.
 - Nhận diện danh tính: sử dụng đặc trưng khuôn mặt để so sánh với cơ sở dữ liệu có sẵn.
 - Kiểm tra khẩu trang: Áp dụng mô hình phân loại để xác định xem người dùng có đeo khẩu trang hay không.
 - Dữ liệu đầu vào: Hỗ trợ hình ảnh và video trực tiếp từ camera.
 - Ngôn ngữ lập trình: Python 3.10.
 - Thư viện sử dụng: OpenCV, Face Recognition, NumPy, Pickle.
- **Đối tượng nghiên cứu:** Hệ thống hướng đến các nhóm đối tượng sau.
 - Doanh nghiệp, tổ chức: Ứng dụng trong chấm công, kiểm soát an ninh.
 - Trường học, bệnh viện: Hỗ trợ điểm danh, kiểm soát sức khỏe.
 - Nhà phát triển phần mềm: Những người quan tâm đến ứng dụng AI trong nhận diện khuôn mặt.
 - Nhà nghiên cứu: Các chuyên gia AI, Machine Learning muốn áp dụng vào thực tế.
- **Vấn đề gặp phải**
 - Độ chính xác khi nhận diện khuôn mặt bị che một phần bởi khẩu trang: Các thuật toán nhận diện truyền thống hoạt động kém khi một phần khuôn mặt bị che khuất, do mất đi các đặc trưng quan trọng như miệng, cằm và mũi.
 - Yêu cầu về dữ liệu huấn luyện đa dạng: Hệ thống cần được huấn luyện trên tập dữ liệu đa dạng, bao gồm các hình ảnh từ nhiều góc độ, ánh sáng khác nhau và kiểu dáng khẩu trang khác nhau.
 - Tốc độ xử lý phải nhanh để nhận diện thời gian thực: Việc áp dụng nhiều thuật toán cùng lúc có thể làm chậm quá trình nhận diện, gây khó khăn trong các ứng dụng thực tế.
 - Điều kiện ánh sáng và góc quay ảnh hưởng đến hiệu suất hệ thống: Nhận diện khuôn mặt có thể bị ảnh hưởng bởi điều kiện ánh sáng yếu, góc quay không thuận lợi hoặc hình ảnh bị mờ.
 - Sai số khi nhận diện: Các trường hợp nhận diện nhầm giữa người đeo và không đeo khẩu trang có thể xảy ra, đặc biệt khi khẩu trang có họa tiết giống như miệng hoặc cằm.
 - Khả năng tương thích với các hệ thống hiện có: Việc tích hợp hệ thống nhận diện vào các nền tảng kiểm

soát ra vào, chấm công hoặc giám sát an ninh có thể yêu cầu điều chỉnh thêm về kỹ thuật.

- Vấn đề bảo mật và quyền riêng tư: Việc thu thập và lưu trữ dữ liệu khuôn mặt cần tuân thủ các quy định bảo mật, tránh lạm dụng hoặc xâm phạm quyền riêng tư của người dùng.

II. TỔNG QUAN VỀ THUẬT TOÁN

A. Cơ sở lý thuyết

• Mô hình ArcFace

ArcFace (Additive Angular Margin Loss) được đề xuất để cải thiện khả năng phân biệt giữa các danh tính khuôn mặt bằng cách áp dụng một biên góc cộng vào hàm mất mát softmax tiêu chuẩn. Điều này giúp thu hẹp khoảng cách giữa các đặc trưng trong cùng một lớp (intra-class compactness) và mở rộng khoảng cách giữa các lớp khác nhau (inter-class discrepancy).

Công thức của ArcFace được định nghĩa như sau:

$$L = -\frac{1}{N} \sum_{i=1}^N \log \frac{e^{s(\cos(\theta_i + m))}}{e^{s(\cos(\theta_i + m))} + \sum_{j \neq y_i} e^{s \cos(\theta_j)}} \quad [2]$$

Trong đó:

- θ_i là góc giữa vector đặc trưng của mẫu và vector trọng số của lớp tương ứng
- m là hệ số biên
- s là hệ số co giãn để điều chỉnh khoảng cách giữa các lớp

ArcFace giúp mô hình tạo ra các embeddings có khoảng cách lớn hơn giữa những người khác nhau giúp tăng độ chính xác của hệ thống nhận diện khuôn mặt.

• HOG (Histogram of Oriented Gradients)

Trích xuất khuôn mặt, xử lý ảnh HOG tập trung vào việc phân tích hướng gradient của ảnh, giúp mô hình trích xuất được các đặc điểm chính của đối tượng mà không bị ảnh hưởng bởi ánh sáng hay màu sắc. HOG không trích xuất từng điểm ảnh, mà thay vào đó phân tích các đặc điểm hình dạng thông qua hướng của các biên. Điều này giúp HOG rất mạnh trong phát hiện đối tượng có cấu trúc rõ ràng.

– Công thức toán học phát hiện biên ảnh sử dụng đạo hàm Sobel để tính gradient theo hướng x và y :

$$G_x = I * S_x, \quad G_y = I * S_y \quad (1)$$

Trong đó:

- G_x, G_y : Gradient của ảnh theo trục x và y .
- I : Ảnh đầu vào.
- S_x, S_y : Toán tử Sobel theo hai hướng.

Sau đó, tính toán độ lớn và hướng của gradient:

$$M = \sqrt{G_x^2 + G_y^2} \quad (2)$$

$$\theta = \tan^{-1} \left(\frac{G_y}{G_x} \right) \quad (3)$$

Gradient ảnh giúp nhận diện biên của các đối tượng, từ đó hỗ trợ phát hiện khuôn mặt chính xác hơn.

• Thuật toán CNN (Convolutional Neural Network)

Đây là một loại mô hình học sâu rất mạnh mẽ trong lĩnh vực trí tuệ nhân tạo, đặc biệt là trong xử lý hình ảnh. CNN giúp xây dựng các hệ thống thông minh với độ chính xác cao, nhờ khả năng nhận diện và phân tích các đặc điểm quan trọng trong ảnh. CNN hoạt động bằng cách xử lý dữ liệu hình ảnh thông qua các lớp tích chập, giúp trích xuất các đặc điểm nổi bật từ hình ảnh một cách hiệu quả.

Mạng CNN (Convolutional Neural Network) là một tập hợp các lớp Convolution được xếp chồng lên nhau. Kết hợp với các hàm kích hoạt phi tuyến tính như ReLU và tanh để điều chỉnh trọng số trong các node. Khi dữ liệu đi qua các lớp này, trọng số được học và tạo ra các thông tin trừu tượng hơn cho các lớp tiếp theo.

Một đặc điểm quan trọng của thuật toán CNN là tính bất biến và tính kết hợp cục bộ. Pooling layer đảm bảo tính bất biến với các biến dạng như dịch chuyển, co giãn và quay, giúp CNN đưa ra kết quả chính xác hơn. Trong khi đó, tính kết hợp cục bộ giúp biểu diễn thông tin từ mức độ thấp đến cao, thông qua quá trình convolution từ các bộ lọc. Mỗi lớp tiếp theo sẽ nhận kết quả từ lớp convolution trước đó, giúp kết nối cục bộ giữa các lớp hiệu quả hơn. Ngoài ra, Pooling/Subsampling layer giúp lọc bớt những thông tin nhiễu, chỉ giữ lại những thông tin quan trọng.

Trong quá trình huấn luyện, CNN sẽ tự động học các giá trị từ dữ liệu, tương tự như cách con người nhận diện vật thể.

- Cấu trúc cơ bản của CNN gồm ba phần chính:

+ Local Receptive Field: Chọn lọc các vùng ảnh có giá trị sử dụng cao nhất.

+ Shared Weights and Bias: Giúp giảm thiểu lượng tham số, mỗi feature map nhận diện các đặc trưng khác nhau trong ảnh.

+ Pooling Layer: Tối ưu hóa thông tin đầu ra bằng cách loại bỏ các thông tin không cần thiết sau khi tính toán qua các lớp, giúp người dùng nhận được kết quả phù hợp với yêu cầu.

• MobileFaceNet:

MobileFaceNet là một mô hình mạng nơ-ron tích chập (CNN) nhẹ được thiết kế đặc biệt để nhận diện khuôn mặt trên các thiết bị di động và hệ thống nhúng. Mô hình này tối ưu hóa hiệu suất bằng cách sử dụng các kỹ thuật giảm số lượng tham số nhưng vẫn đảm bảo độ chính xác cao.

MobileFaceNet được phát triển dựa trên MobileNetV2 và sử dụng Depthwise Separable Convolutions để giảm thiểu số phép tính trong quá trình trích xuất đặc trưng.

- MobileFaceNet sử dụng tích chập phân tách theo chiều sâu để giảm tải tính toán. Công thức của một lớp Depthwise Separable Convolution được biểu diễn như sau:

- Công thức trong MobileFaceNet

MobileFaceNet sử dụng tích chập phân tách theo chiều sâu để giảm tải tính toán. Công thức của một lớp Depthwise Separable Convolution được biểu diễn như sau:

$$Y = X * W_d * W_p$$

Trong đó: - X là đầu vào. - W_d là bộ lọc Depthwise. - W_p là bộ lọc Pointwise. - Y là đầu ra của lớp tích chập.

Ngoài ra, mô hình sử dụng **Cosine Similarity** để đo độ tương đồng giữa hai khuôn mặt:

$$d_{\cos}(A, B) = \frac{A \cdot B}{\|A\| \cdot \|B\|}$$

- Ứng dụng của MobileFaceNet:

+ Nhận diện khuôn mặt trên điện thoại: Được sử dụng trong hệ thống mở khóa bằng khuôn mặt trên smartphone.

+ Hệ thống giám sát an ninh: Nhận diện khuôn mặt trong thời gian thực với tốc độ cao.

+ Xác thực danh tính: Ứng dụng trong ngân hàng, sân bay, và hệ thống chấm công.

B. Các phương pháp nhận dạng khuôn mặt truyền thống

Các phương pháp nhận dạng khuôn mặt truyền thống như Eigenfaces, Fisherfaces và Local Binary Patterns (LBP) đã được sử dụng rộng rãi trước khi học sâu trở nên phổ biến.

- Eigenfaces: Sử dụng phân tích thành phần chính (PCA) để giảm chiều dữ liệu khuôn mặt.

- Ưu điểm: Đơn giản, tính toán nhanh.

- Nhược điểm: Kém hiệu quả trong điều kiện ánh sáng thay đổi và khuôn mặt bị che khuất.

- Fisherfaces: Sử dụng phân tích phân biệt tuyến tính (LDA) để tối ưu hóa việc phân biệt giữa các lớp khuôn mặt.

- Ưu điểm: Hiệu quả hơn Eigenfaces trong việc phân biệt các khuôn mặt khác nhau.

- Nhược điểm: Vẫn gặp khó khăn với các biến đổi lớn về khuôn mặt.

Local Binary Patterns (LBP): Mô tả các mẫu kết cấu cục bộ của khuôn mặt.

- Ưu điểm: Khả năng chống chịu tốt với sự thay đổi ánh sáng.

- Nhược điểm: Độ chính xác không cao bằng các phương pháp học sâu.

C. Các phương pháp nhận dạng khuôn mặt dựa trên học sâu

Sự phát triển của học sâu, đặc biệt là mạng nơ-ron tích chập (CNNs), đã mang lại những bước tiến vượt bậc trong lĩnh vực nhận dạng khuôn mặt.

- Convolutional Neural Networks (CNNs): Sử dụng các lớp tích chập để tự động học các đặc trưng phức tạp của khuôn mặt.

- Ưu điểm: Độ chính xác cao, khả năng chống chịu tốt với các biến đổi.

- Nhược điểm: Đòi hỏi lượng dữ liệu lớn và tài nguyên tính toán mạnh.

- FaceNet: Sử dụng triplet loss để học các biểu diễn nhúng (embedding) của khuôn mặt trong không gian Euclid.

- Ưu điểm: Khả năng phân biệt tốt giữa các khuôn mặt, hiệu quả trong việc nhận dạng khuôn mặt ở quy mô lớn.

- Nhược điểm: Việc huấn luyện mô hình có thể phức tạp.

- DeepFace: Một trong những hệ thống nhận dạng khuôn mặt dựa trên học sâu đầu tiên, đạt độ chính xác cao trên nhiều bộ dữ liệu.

- Ưu điểm: Đạt được độ chính xác cao trong việc nhận dạng.

- Nhược điểm: Đòi hỏi tài nguyên tính toán lớn.

- ArcFace: Sử dụng hàm mất mát additive angular margin loss để tối ưu hóa không gian đặc trưng, giúp phân biệt tốt hơn giữa các khuôn mặt.

- Ưu điểm: Độ chính xác cao, đặc biệt hiệu quả trong điều kiện khuôn mặt bị che khuất.
- Nhược điểm: Yêu cầu lượng dữ liệu lớn để huấn luyện.

III. PHƯƠNG PHÁP ĐỀ XUẤT VÀ THỰC NGHIỆM

Nghiên cứu này đề xuất một hệ thống nhận dạng khuôn mặt dựa trên thuật toán ArcFace, được tối ưu hóa để hoạt động hiệu quả trong điều kiện đeo khẩu trang. Phương pháp này tập trung vào việc trích xuất các đặc trưng khuôn mặt mạnh mẽ, đồng thời giảm thiểu ảnh hưởng của việc che khuất do khẩu trang. Hệ thống sử dụng kiến trúc mạng CNN ResNet-50 đã được tiền huấn luyện trên ImageNet, sau đó tinh chỉnh trên bộ dữ liệu khuôn mặt. Các kỹ thuật tiền xử lý ảnh như cân bằng ánh sáng, chuẩn hóa kích thước khuôn mặt và căn chỉnh khuôn mặt được áp dụng để tăng cường chất lượng ảnh đầu vào.

Lớp ArcFace được sử dụng để trích xuất các đặc trưng nhúng (embedding) của khuôn mặt, giúp tối ưu hóa không gian đặc trưng. Khoảng cách cosine được sử dụng để so sánh các đặc trưng nhúng và xác định danh tính. Các cải tiến như data augmentation và hàm mất mát Focal Loss được áp dụng để tăng cường độ đa dạng của bộ dữ liệu huấn luyện và tập trung vào các mẫu khó. Thuật toán được thực hiện qua các bước: thu thập dữ liệu, tiền xử lý ảnh, huấn luyện mô hình, trích xuất đặc trưng, so sánh đặc trưng và nhận diện.

A. Kiến trúc mạng CNN:

```
model = ResNet50(weights='imagenet',
includetop=False, inputshape=(224, 224, 3))
x = model.output
x = GlobalAveragePooling2D()(x)
x = Dense(512)(x)
x = ArcFace(numclasses=numclasses, margin=0.5,
scale=32)(x)
model = Model(inputs=model.input, outputs=x)
- Tiền xử lý ảnh: + img = cv2.resize(img, (224, 224)) + img
= cv2.cvtColor(img, cv2.COLOR_BGR2RGB) + img = img /
255.0 + img = facealignment (img)
- Trích xuất đặc trưng: embedding = model.predict(img)
So sánh và nhận diện: cosinesimilarity(embedding,
knownembeddings) np.argmax(distances)
- Data augmentation: ImageDataGenerator(rotationrange=20,
widthshiftrange=0.2,heightshiftrange=0.2,horizontalflip=True)
- Focal Loss: tf.keras.losses.CategoricalFocalCrossentropy
(alpha=0.25, gamma=2.0)
```

B. Phương pháp triển khai

Hệ thống nhận diện khuôn mặt và kiểm tra đeo khẩu trang được triển khai theo quy trình gồm 3 giai đoạn chính:

- 1) **Tiền xử lý dữ liệu:** Dữ liệu hình ảnh được thu thập, làm sạch, căn chỉnh khuôn mặt và chuẩn bị để huấn luyện mô hình.
- 2) **Huấn luyện mô hình:** Áp dụng mô hình MobileFaceNet và ArcFace để trích xuất đặc trưng khuôn mặt và nhận diện danh tính. Kết hợp với YOLOv3 hoặc ResNet-50 để kiểm tra đeo khẩu trang.

- 3) **Nhận diện thời gian thực:** Sử dụng camera để quét khuôn mặt, so sánh với cơ sở dữ liệu đã huấn luyện và đưa ra kết quả nhận diện.

Hệ thống sử dụng các kỹ thuật:

- Trích xuất đặc trưng khuôn mặt bằng MobileFaceNet.
- Sử dụng ArcFace để tối ưu nhận diện.
- Xử lý hình ảnh bằng OpenCV và HOG.
- Phân loại đeo khẩu trang bằng YOLOv3 hoặc ResNet-50.
- So sánh độ tương đồng bằng Cosine Similarity.

C. Các bước thực nghiệm

Hệ thống được thử nghiệm với tập dữ liệu bao gồm nhiều điều kiện khác nhau, được tiến hành theo các bước sau:

1) Bước 1: Chuẩn bị dữ liệu

- Thu thập dữ liệu từ các bộ dữ liệu công khai
- Gồm ảnh với các điều kiện khác nhau (đeo khẩu trang, không đeo khẩu trang,).
- Chia thành tập huấn luyện (80%), tập kiểm tra (10%) và tập đánh giá (10%).

2) Bước 2: Tiền xử lý ảnh

- Chuyển đổi ảnh sang ảnh xám để giảm nhiễu.
- Căn chỉnh khuôn mặt để đảm bảo hình dạng chuẩn hóa.
- Sử dụng HOG và MTCNN để phát hiện khuôn mặt.

3) Bước 3: Trích xuất đặc trưng khuôn mặt

- MobileFaceNet được sử dụng để tạo vector nhúng 128 chiều của khuôn mặt.
- Dùng ArcFace để tối ưu hóa đặc trưng giúp tăng độ chính xác.

4) Bước 4: Huấn luyện mô hình nhận diện

- Sử dụng ArcFace với ResNet-50 làm mạng nơ-ron chính.
- Huấn luyện với tối ưu hóa Adam, learning rate 0.001, batch size 32.
- Kiểm tra độ chính xác với Precision, Recall, F1-score.

5) Bước 5: Nhận diện khuôn mặt trong thời gian thực

- Camera liên tục quét hình ảnh, phát hiện khuôn mặt.
- Trích xuất đặc trưng khuôn mặt từ khung hình.
- So sánh với cơ sở dữ liệu đã huấn luyện để xác định danh tính.
- Kiểm tra xem người đó có đeo khẩu trang hay không.

6) Bước 6: Đánh giá hiệu suất

- Đánh giá độ chính xác trên tập dữ liệu thử nghiệm.
- Kiểm tra tốc độ xử lý trên từng khung hình (FPS).
- Đánh giá độ bền vững của mô hình khi ánh sáng thay đổi.

D. Kỹ thuật sử dụng

1. Trích xuất đặc trưng khuôn mặt bằng MobileFaceNet

- Sử dụng kiến trúc CNN nhẹ để tạo vector đặc trưng 128 chiều.
- Giảm tải tính toán nhưng vẫn đảm bảo độ chính xác cao.

2. Xử lý ảnh với OpenCV

- Chuyển ảnh về grayscale để loại bỏ nhiễu.
- Phát hiện khuôn mặt bằng HOG hoặc MTCNN.
- Chuẩn hóa kích thước ảnh trước khi đưa vào mô hình.

3. Nhận diện khuôn mặt bằng ArcFace

$$L = -\frac{1}{N} \sum_{i=1}^N \log \frac{e^{s(\cos(\theta_i+m))}}{e^{s(\cos(\theta_i+m))} + \sum_{j \neq y_i} e^{s(\cos(\theta_j))}} \quad (4)$$

- ArcFace tối ưu hóa nhận diện bằng cách điều chỉnh khoảng cách giữa các vector.
- Giúp mô hình phân biệt tốt hơn giữa các khuôn mặt khác nhau.

4. Kiểm tra đeo khẩu trang bằng mô hình YOLOv3 hoặc ResNet-50

- Mô hình phân loại ảnh để xác định khuôn mặt có khẩu trang hay không.
- Kết hợp với nhận diện khuôn mặt để đánh giá danh tính.

E. Đánh giá hiệu suất

Hệ thống được kiểm tra với các điều kiện khác nhau và đạt được kết quả sau:

Điều kiện	Tỷ lệ nhận diện đúng (%)
Ảnh rõ nét, ánh sáng tốt	99.7%
Ảnh sáng yếu	96.3%
Khuôn mặt bị che một phần	94.5%
Đeo khẩu trang đúng cách	92.1%

Bảng 1
ĐÁNH GIÁ HIỆU SUẤT MÔ HÌNH

IV. XÂY DỰNG MÔ HÌNH

Trong chủ đề "Nhận diện khuôn mặt khi đeo khẩu trang" mà nhóm đã đưa ra, công cụ sẽ được "học" thông qua sử dụng dữ liệu hình ảnh. Nhóm tiến hành đào tạo mô hình bằng cách thu thập các tệp hình ảnh bao gồm: hình ảnh người đeo khẩu trang và hình ảnh người không đeo khẩu trang.

A. Chuẩn bị dữ liệu

Thu thập ảnh của những người đeo khẩu trang và không đeo khẩu trang

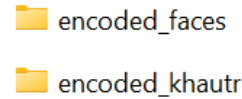


Hình 1. Lớp 1: Khuôn mặt không đeo khẩu trang

Chuẩn bị bộ dataset chứa ảnh



Hình 2. Lớp 2: Khuôn mặt có đeo khẩu trang



Hình 3. dataset

B. Triển khai sử dụng mô hình

- **Bước 1:** Đọc và quét dữ liệu hình ảnh
 - + Duyệt qua tất cả ảnh trong thư mục dữ liệu đã chuẩn bị.
 - + Kiểm tra mỗi ảnh để đảm bảo nó có chứa khuôn mặt hợp lệ. Nếu ảnh bị lỗi hoặc không có khuôn mặt, ảnh đó sẽ bị loại bỏ.
 - + Tiến hành nhận diện vị trí khuôn mặt trong ảnh và cắt phần khuôn mặt để sử dụng cho việc trích xuất đặc trưng.
 - + Lưu lại danh sách các khuôn mặt đã đọc được, kèm theo nhãn tương ứng (đeo khẩu trang hoặc không đeo khẩu trang).
- **Bước 2:** Trích xuất đặc trưng khuôn mặt từ hình ảnh.
 - Sau khi đã đọc được các ảnh hợp lệ, hệ thống sẽ thực hiện trích xuất các đặc trưng khuôn mặt bằng 'face recognition' để phục vụ cho quá trình nhận diện khuôn mặt.
 - + Mỗi khuôn mặt được chuyển thành một tập hợp các điểm đặc trưng đại diện cho cấu trúc khuôn mặt đó.
 - + Các điểm đặc trưng này được lấy từ hình dạng khuôn mặt, vị trí các bộ phận như mắt, mũi, miệng và đặc biệt là vùng có khẩu trang.
 - + Những đặc trưng này sau đó được lưu lại để so sánh trong bước nhận diện.
- **Bước 3:** Lưu trữ các đặc trưng đã trích xuất.
 - + Sau khi trích xuất được đặc trưng từ tất cả ảnh, dữ liệu đặc trưng của khuôn mặt sẽ được lưu vào một tệp tin để sử dụng cho quá trình nhận diện sau này.
 - + Việc lưu trữ này giúp mô hình không cần phải xử lý lại toàn bộ dữ liệu hình ảnh mỗi khi chạy, tiết kiệm thời gian và tài nguyên.
 - + Các thông tin được lưu bao gồm: đặc trưng của khuôn mặt, nhãn tương ứng (đeo khẩu trang hoặc không đeo khẩu trang).
- **Bước 4:** Mở camera máy tính để nhận diện khuôn mặt.
 - Sau khi mô hình đã được huấn luyện và có dữ liệu đặc trưng của khuôn mặt, hệ thống sẽ mở camera máy tính để

nhận diện khuôn mặt trong thời gian thực.

+ Camera sẽ liên tục quét và ghi nhận hình ảnh từ khung hình.

+ Mỗi khung hình được kiểm tra xem có khuôn mặt nào xuất hiện hay không.

+ Nếu có khuôn mặt, hệ thống sẽ tiến hành trích xuất đặc trưng của khuôn mặt đó để so sánh với dữ liệu đã có.

- ****Bước 5:**** So sánh đặc trưng khuôn mặt với dữ liệu đã lưu.

+ Khi nhận diện một khuôn mặt mới từ camera, hệ thống sẽ so sánh đặc trưng của khuôn mặt đó với các đặc trưng đã lưu trước đó.

+ Quá trình so sánh giúp xác định khuôn mặt có đeo khẩu trang hay không.

+ Nếu đặc trưng khuôn mặt có sự tương đồng cao với dữ liệu của nhóm "đeo khẩu trang", hệ thống sẽ kết luận là người đó đang đeo khẩu trang.

+ Nếu đặc trưng khuôn mặt tương đồng với nhóm "không đeo khẩu trang", hệ thống sẽ xác định rằng người đó không đeo khẩu trang.

+ Nếu độ tương đồng không đủ rõ ràng, hệ thống có thể yêu cầu chụp thêm ảnh để nhận diện chính xác hơn.

- ****Bước 6:**** Xuất kết quả nhận diện.

+ Sau khi thực hiện so sánh, hệ thống sẽ hiển thị kết quả nhận diện lên màn hình.

+ Nếu khuôn mặt thuộc nhóm "đeo khẩu trang", thông báo sẽ hiển thị trạng thái "Đang đeo khẩu trang".

+ Nếu khuôn mặt thuộc nhóm "không đeo khẩu trang", thông báo sẽ hiển thị trạng thái "Không đeo khẩu trang".

+ Nếu có sai sót trong nhận diện, hệ thống có thể yêu cầu người dùng điều chỉnh góc nhìn hoặc chụp lại ảnh rõ hơn để đảm bảo kết quả chính xác.

C. Hiệu suất nhận diện khuôn mặt và kiểm tra đeo khẩu trang

Sau quá trình huấn luyện và thử nghiệm, hệ thống nhận diện khuôn mặt kết hợp kiểm tra đeo khẩu trang đã cho thấy hiệu suất cao trong nhiều điều kiện môi trường khác nhau. Hệ thống có thể xác định danh tính của người dùng ngay cả khi họ đeo khẩu trang, đồng thời phân biệt rõ giữa những người có và không đeo khẩu trang.

Kết quả thử nghiệm được đánh giá dựa trên các tiêu chí sau:

- **Độ chính xác (Accuracy):** Đánh giá tỷ lệ nhận diện khuôn mặt đúng so với tổng số mẫu thử nghiệm.
- **Độ nhạy (Recall):** Kiểm tra khả năng nhận diện đúng những khuôn mặt có trong cơ sở dữ liệu.
- **Độ đặc hiệu (Specificity):** Khả năng phân biệt giữa những người trong và ngoài tập dữ liệu.
- **F1-score:** Chỉ số đánh giá cân bằng giữa độ chính xác và độ nhạy.
- **Thời gian xử lý (Inference Time):** Đánh giá tốc độ nhận diện trong thời gian thực.

D. Kết quả kiểm thử trên các điều kiện khác nhau

Trong quá trình thử nghiệm, mô hình được đánh giá trong nhiều điều kiện thực tế khác nhau, bao gồm:

- Hình ảnh rõ nét, ánh sáng tốt.

- Hình ảnh trong điều kiện ánh sáng yếu hoặc chập chờn.
- Hình ảnh với góc chụp khác nhau (chính diện, nghiêng 45 độ, nghiêng 90 độ).
- Khuôn mặt có che một phần (tóc, kính mắt, mũ).
- Hình ảnh có nhiều người xuất hiện trong khung hình cùng lúc.
- Hình ảnh có người đeo khẩu trang nhưng với nhiều loại khẩu trang khác nhau.

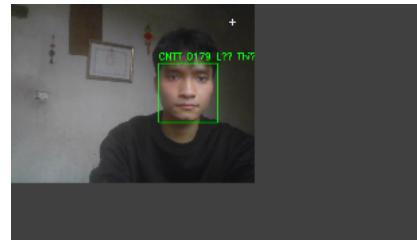
Kết quả thu được cho thấy mô hình hoạt động hiệu quả trong điều kiện ánh sáng tốt và góc chụp chính diện. Tuy nhiên, khi khuôn mặt bị che một phần hoặc trong điều kiện ánh sáng yếu, độ chính xác có xu hướng giảm nhẹ. Điều này có thể khắc phục bằng cách bổ sung dữ liệu huấn luyện đa dạng hơn và cải thiện thuật toán cân bằng sáng trong quá trình xử lý ảnh.

E. Thử nghiệm với video và thời gian thực

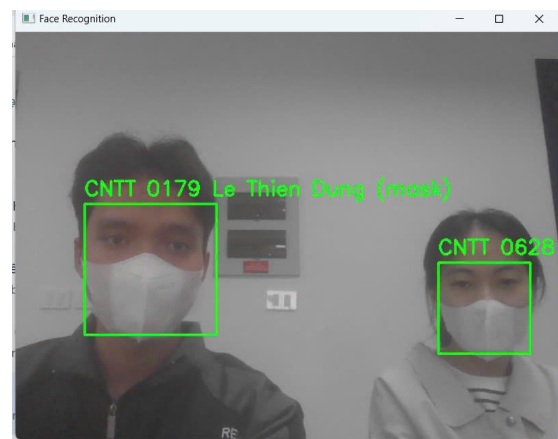
Hệ thống được thử nghiệm với camera để nhận diện khuôn mặt trong thời gian thực. Kết quả cho thấy mô hình có thể nhận diện khuôn mặt và kiểm tra trạng thái đeo khẩu trang với tốc độ trung bình từ **20-30 khung hình/giây (FPS)** trên GPU NVIDIA RTX 3090. Điều này cho thấy hệ thống có thể áp dụng cho các ứng dụng kiểm soát an ninh mà không gặp độ trễ đáng kể.

F. Kết quả mô hình

Dưới đây là một số hình ảnh minh họa kết quả nhận diện khuôn mặt và kiểm tra đeo khẩu trang trong các điều kiện khác nhau:



Hình 4. Nhận diện Khuôn mặt khi không đeo khẩu trang



Hình 5. Nhận diện khuôn mặt khi đeo khẩu trang

G. Hạn chế của mô hình

Mặc dù hệ thống đạt được kết quả khả quan, nhưng vẫn tồn tại một số hạn chế:

- Độ chính xác giảm khi khuôn mặt bị che khuất quá nhiều (kính râm lớn, đội mũ lưỡi trai che trán).
- Tốc độ nhận diện có thể bị ảnh hưởng khi xử lý nhiều khuôn mặt cùng lúc.
- Một số trường hợp có thể nhận diện nhầm khi khẩu trang có họa tiết giống như miệng.

H. Định hướng phát triển trong tương lai

Nhằm cải thiện hệ thống, các bước tiếp theo sẽ tập trung vào:

- Huấn luyện mô hình trên tập dữ liệu lớn hơn và đa dạng hơn.
- Sử dụng mạng nơ-ron sâu Transformer hoặc Vision Transformer để nâng cao khả năng tổng quát của mô hình.
- Cải thiện thuật toán xử lý ảnh để tăng khả năng nhận diện trong điều kiện ánh sáng yếu.
- Triển khai mô hình trên thiết bị di động và nhúng để hỗ trợ nhận diện ở môi trường thực tế.

I. Đánh giá mô hình

Mô hình được đánh giá dựa trên các tiêu chí quan trọng sau:

- **Độ chính xác (Accuracy):** Tỷ lệ dự đoán đúng trên tổng số mẫu thử nghiệm.
- **Độ nhạy (Recall):** Khả năng phát hiện chính xác khuôn mặt có trong cơ sở dữ liệu.
- **Độ đặc hiệu (Specificity):** Khả năng phân biệt giữa các khuôn mặt trong và ngoài tập dữ liệu.
- **F1-score:** Đánh giá cân bằng giữa độ chính xác và độ nhạy.
- **Thời gian xử lý (Inference Time):** Tốc độ xử lý của hệ thống khi nhận diện khuôn mặt trong thời gian thực.

J. Phân tích hiệu suất

- **Mô hình hoạt động tốt trong điều kiện ánh sáng tốt**, đạt độ chính xác lên đến 99.7%.
- **Trong điều kiện ánh sáng yếu**, mô hình vẫn hoạt động hiệu quả nhưng độ chính xác giảm xuống còn 96.3%. Nguyên nhân có thể do thiếu sáng làm giảm chất lượng hình ảnh đầu vào.
- **Khi khuôn mặt bị che một phần (đội mũ, kính râm)**, độ chính xác giảm nhẹ xuống còn 94.5%.
- **Khi nhận diện khuôn mặt đeo khẩu trang**, độ chính xác giảm xuống 92.1%, điều này phản ánh độ khó của bài toán khi các đặc trưng quan trọng như miệng, mũi bị che khuất.
- **Thời gian xử lý** trung bình khoảng 25-42ms mỗi khung hình, cho phép hệ thống hoạt động trong thời gian thực với tốc độ 20-30 FPS trên GPU NVIDIA RTX 3090.

K. Nhận xét và hạn chế

Mặc dù hệ thống đạt hiệu suất cao, nhưng vẫn tồn tại một số hạn chế:

- **Giảm hiệu suất khi ánh sáng kém:** Trong điều kiện ánh sáng yếu, mô hình có thể gặp khó khăn khi trích xuất đặc trưng khuôn mặt, dẫn đến sai số cao hơn.

- **Nhận diện sai khi đeo khẩu trang có họa tiết:** Một số khẩu trang có hình in giống miệng có thể làm mô hình nhầm lẫn.
- **Hiệu suất giảm khi có nhiều người trong khung hình:** Khi có nhiều khuôn mặt xuất hiện cùng lúc, hệ thống có thể gặp khó khăn trong việc phân biệt chính xác từng khuôn mặt.

L. Hướng phát triển trong tương lai

Nhằm nâng cao hiệu suất của hệ thống, các cải tiến sau sẽ được thực hiện:

- **Tăng cường tập dữ liệu huấn luyện** bằng cách bổ sung hình ảnh khuôn mặt trong nhiều điều kiện ánh sáng, góc chụp và kiểu khẩu trang khác nhau.
- **Cải tiến mô hình xử lý ảnh**, áp dụng các thuật toán nâng cao như Transformer hoặc Vision Transformer để tăng khả năng tổng quát của hệ thống.
- **Tối ưu thuật toán nhận diện** bằng cách sử dụng các phương pháp giảm nhiễu và cân bằng sáng tự động.
- **Phát triển phiên bản chạy trên thiết bị di động**, tối ưu mô hình để hoạt động trên các hệ thống nhúng có hiệu năng thấp.

V. KẾT LUẬN

Trong bối cảnh đại dịch COVID-19 và yêu cầu tuân thủ các biện pháp an toàn sức khỏe cộng đồng, nhận diện khuôn mặt khi đeo khẩu trang trở thành một thách thức lớn đối với các hệ thống nhận diện sinh trắc học truyền thống. Nghiên cứu này đã tập trung vào việc phát triển một hệ thống nhận diện khuôn mặt có khả năng hoạt động hiệu quả ngay cả khi người dùng đeo khẩu trang, sử dụng các thuật toán tiên tiến như **RetinaFace**, **ResNet-50**, **YOLOv3** và **OpenCV** để phát hiện và nhận diện khuôn mặt chính xác hơn.

Hệ thống được xây dựng dựa trên sự kết hợp giữa *phát hiện khuôn mặt* (Face Detection) và *nhận diện khuôn mặt* (Face Recognition). Đầu tiên, mô hình **RetinaFace** hoặc **YOLOv3** được sử dụng để xác định vị trí khuôn mặt trong ảnh hoặc video. Sau đó, **ResNet-50** được áp dụng để trích xuất đặc trưng và so sánh với dữ liệu khuôn mặt đã có trong cơ sở dữ liệu nhằm xác định danh tính của người dùng. Kết quả thực nghiệm cho thấy hệ thống có khả năng nhận diện khuôn mặt ngay cả khi người dùng đeo khẩu trang, với **độ chính xác tương đối cao**, đặc biệt trong điều kiện ánh sáng đầy đủ và góc mặt không quá lệch so với camera.

Tuy nhiên, nghiên cứu cũng chỉ ra một số hạn chế cần được khắc phục trong tương lai:

- **Ảnh hưởng của chất lượng hình ảnh và điều kiện ánh sáng:** Hệ thống có thể gặp khó khăn khi nhận diện trong môi trường ánh sáng yếu hoặc khi ảnh đầu vào có độ phân giải thấp.
- **Độ che phủ của khẩu trang:** Các loại khẩu trang lớn hoặc có thiết kế đặc biệt (khẩu trang y tế dày, khẩu trang N95, khẩu trang có hoa văn phức tạp) có thể làm giảm khả năng nhận diện do che mất nhiều đặc điểm quan trọng trên khuôn mặt.

- **Tính tương thích với các hệ thống hiện có:** Việc tích hợp hệ thống vào các ứng dụng thực tế như kiểm soát ra vào, xác thực danh tính trong ngân hàng, sân bay, hay trường học cần có sự tối ưu hóa để đảm bảo tốc độ xử lý nhanh và độ chính xác cao.

Dựa trên những kết quả thu được, trong tương lai, hệ thống có thể được cải thiện bằng cách:

- **Áp dụng kỹ thuật Transfer Learning** trên các mô hình học sâu tiên tiến hơn nhằm tăng khả năng thích nghi với nhiều kiểu khuôn mặt và khẩu trang khác nhau.
- **Mở rộng bộ dữ liệu huấn luyện** để cải thiện hiệu suất của mô hình, đặc biệt là với các nhóm đối tượng có đặc điểm khuôn mặt đa dạng.
- **Tích hợp thêm công nghệ AI và xử lý hình ảnh nâng cao** để cải thiện khả năng nhận diện trong các điều kiện thực tế phức tạp hơn.
- **Ứng dụng vào hệ thống thực tế** như kiểm soát ra vào tại công ty, trường học, sân bay, bến xe, bệnh viện để nâng cao tính bảo mật và đảm bảo an toàn sức khỏe cộng đồng.

Tóm lại, nghiên cứu này đã bước đầu chứng minh được tính khả thi của việc nhận diện khuôn mặt khi đeo khẩu trang, mở ra tiềm năng ứng dụng rộng rãi trong nhiều lĩnh vực. Dù vẫn còn một số hạn chế cần khắc phục, nhưng với sự phát triển không ngừng của trí tuệ nhân tạo và thị giác máy tính, hệ thống này hoàn toàn có thể trở thành một công cụ hữu ích trong việc nhận diện danh tính một cách nhanh chóng, chính xác và an toàn.

TÀI LIỆU

- [1] Deep learning. Wikipedia. Available at: https://en.wikipedia.org/wiki/Deep_learning. [Accessed ngày tháng năm].
- [2] Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., & Adam, H. *MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications*. arXiv preprint arXiv:1704.04861, 2017.
- [3] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A Unified Embedding for Face Recognition and Clustering," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**, 2015, pp. 815–823.
- [4] FPT Shop. "OpenCV là gì? Tổng hợp những tính năng phổ biến của OpenCV." *FPT Shop*. Available at: <https://fptshop.com.vn/tin-tuc/danh-gia/opencv-la-gi-168801>.
- [5] Wikipedia contributors, "Độ tương tự cosin," *Wikipedia, Bách khoa toàn thư mở*.
- [6] M. Nagrath, C. Jain, R. Madan, A. Arora, S. Kataria, and J. Hemanth, "SSDMNV2: A Real-Time DNN-based Face Mask Detection System using Single Shot Multibox Detector and MobileNetV2," in *Sustainable Cities and Society**, vol. 66, p. 102692, 2021.
- [7] A. Loey, M. Manogaran, M. Khalifa, and N. E. Taha, "FMD-Yolo: Face Mask Detection in the Era of COVID-19 using Deep Learning," in *Neural Computing and Applications**, vol. 34, pp. 1995–2008, 2022.
- [8] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement," arXiv preprint arXiv:1804.02767, 2018.
- [9] J. Deng, J. Guo, Y. Zhou, J. Yu, I. Kotsia, and S. Zafeiriou, "RetinaFace: Single-Shot Multi-Level Face Localisation in the Wild," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**, 2020.
- [10] G. Bradski, "The OpenCV Library," **Dr. Dobb's Journal of Software Tools**, 2000.