

**ĐẠI HỌC QUỐC GIA TP. HỒ CHÍ MINH**  
**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN**  
**KHOA KHOA HỌC MÁY TÍNH**

**ĐẶNG XUÂN MAI**

**KHÓA LUẬN TỐT NGHIỆP**  
**NHẬN DẠNG NGƯỜI NỔI TIẾNG VÀ**  
**SỰ KIỆN TƯƠNG ỨNG**

**CỬ NHÂN NGÀNH KHOA HỌC MÁY TÍNH**

**TP. HỒ CHÍ MINH, 2022**

**ĐẠI HỌC QUỐC GIA TP. HỒ CHÍ MINH**  
**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN**  
**KHOA KHOA HỌC MÁY TÍNH**

**ĐẶNG XUÂN MAI - 19521820**

**KHÓA LUẬN TỐT NGHIỆP  
NHẬN DẠNG NGƯỜI NỔI TIẾNG VÀ  
SỰ KIỆN TƯƠNG ỨNG**

**CỬ NHÂN NGÀNH KHOA HỌC MÁY TÍNH**

**GIẢNG VIÊN HƯỚNG DẪN**

**TS. MAI TIẾN DŨNG**

**TP. HỒ CHÍ MINH, 2022**

## ĐĂNG KÝ ĐỀ TÀI KHÓA LUẬN TỐT NGHIỆP

**Tên đề tài: NHẬN DẠNG NGƯỜI NỔI TIẾNG VÀ SỰ KIỆN TƯƠNG ỨNG**

**Tên đề tài tiếng Anh: FAMOUS PERSON RECOGNITION AND THEIR EVENT**

**Ngôn ngữ thực hiện: Tiếng Việt**

**Cán bộ hướng dẫn: TS. Mai Tiến Dũng**

**Thời gian thực hiện:** Từ ngày 05/09/2022 đến ngày 24/12/2022.

**Sinh viên thực hiện:**

Đặng Xuân Mai – MSSV:19521820

**Lớp:** KHCL2019.2

**Email:** 19521820@gm.uit.edu.vn

**Điện thoại:** 0792108632

**Nội dung đề tài:**

- Giới thiệu đề tài:**

- Ngày nay, nhờ sự phát triển của công nghệ thông tin và các thiết bị điện tử, lượng tin tức trong một ngày được cập nhật đã tăng lên rất nhiều và sự lan truyền của những tin tức này trở nên nhanh chóng một cách đáng kinh ngạc. Theo số liệu được thống kê trong năm 2022 [1], trong vòng một phút thì có đến 275.000 tweet được gửi đi trên Twitter. Với số lượng thông tin nhiều như vậy, người dùng sẽ rất khó để kiểm chứng được tính đúng sai của các thông tin.

- Những người nổi tiếng như các vị nguyên thủ quốc gia, diễn viên, ca sĩ, ... là những người có sức ảnh hưởng rất lớn đối với công chúng. Không khó để chúng ta có thấy hình ảnh, tin tức liên quan của họ trên những phương tiện truyền thông đại chúng, mạng xã hội, ... Tuy nhiên, nếu hình ảnh của những người nổi tiếng được sử dụng trong những bài báo, thông tin, ... không chính xác thì sẽ gây ảnh hưởng lớn và phát sinh nhiều vấn đề. Lấy một ví dụ cụ thể và gần đây nhất là trong thời điểm dịch bệnh covid hoành hành, đã có không ít những tin tức sai sự thật được đưa ra dưới cái mác là những phát ngôn của những nguyên thủ quốc gia trong công tác chống dịch và kèm theo hình ảnh của những vị nguyên thủ đó trong “một buổi họp nào đó gần đây”.

- Từ đó có thể thấy, một phần mềm hệ thống có thể xác định được gương mặt của người nổi tiếng và sự kiện người đó đang tham gia là một phần mềm có thể giúp giảm được những tác hại của những thông tin không chính xác này. Hệ thống sẽ yêu cầu input là một bức ảnh của một người nổi tiếng bất kỳ, sau khi hệ thống phân tích và xử lý, sẽ cho ra output là tên người nổi tiếng đó, sự kiện và ngày diễn ra sự kiện mà người đó đã tham gia. Như vậy sẽ góp được phần nào trong việc xác định hình ảnh được sử dụng và nội dung của những thông tin đi kèm có liên quan đến nhau hay không.

- **Mục tiêu của đề tài:**

- Hiểu được các phương pháp nhận diện gương mặt, cách để phân tích những yếu tố khác trong bức hình như trang phục, phông nền, ... để xác định được thêm nhiều thông tin liên quan đến sự kiện người đó tham gia.

- Ứng dụng những phương pháp đã tìm hiểu được và dataset tự thu thập.

- Minh họa những kết quả đã thu được thông qua một chương trình phần mềm.

- **Nội dung nghiên cứu của đề tài:**

Nội dung 1: Tìm hiểu về quy trình của các phương pháp

- Tìm kiếm, khảo sát các tài liệu, công nghệ liên quan đến:

1. Nhận diện gương mặt: CNN, các mô hình như FaceNet, OpenFace in OpenCV,...
2. Phân tích những yếu tố xung quanh: dựa trên các yếu tố như phân bố màu sắc (do trang phục và phông nền phía sau), ... để tìm kiếm sự tương đồng trong những sự kiện giống nhau.

Từ đó xây dựng lên một mô hình hoàn chỉnh nhất có thể.

- Kết quả mong đợi: tài liệu kỹ thuật chi tiết các phương pháp được ứng dụng và thu được một mô hình có thể xử lý hai nhiệm vụ nêu trên một cách hoàn thiện, nhanh chóng.

Nội dung 2: Xây dựng một bộ dataset.

- Sử dụng máy tính để thu thập các hình ảnh của những người nổi tiếng trong các sự kiện đặc biệt. Những đối tượng nổi tiếng cụ thể sẽ là các chính trị gia hoạt động trong khoảng thời gian từ năm 2020 – 2022. (Thu thập thủ công hoặc viết script để thu thập).

- Kiểm tra lại các hình ảnh có được và gán nhãn.

- Training model và chạy thử trên dataset tạo được.

- Kết quả mong đợi: dataset có gán nhãn (tên người nổi tiếng – sự kiện tham gia – ngày diễn ra sự kiện) thu thập của tối thiểu là 10 nhân vật nổi tiếng và báo cáo kết quả chạy thử trên mô hình.

Nội dung 3: Xây dựng ứng dụng minh họa

Xây dựng một phần mềm có thể nhận đầu vào là một tấm ảnh, đầu ra là tên người trong ảnh – sự kiện tham gia – ngày diễn ra sự kiện.

- Kết quả mong đợi: một phần mềm minh họa

- **Tài liệu tham khảo:**

- [1] “What Happens in an Internet Minute? [2022 Statistics],” *LOCALiQ*. <https://localiq.com/blog/what-happens-in-an-internet-minute/> (accessed Sep. 07, 2022).

**Kế hoạch thực hiện:**

Ngày bắt đầu	Ngày kết thúc	Nội dung thực hiện
05/09/2022	09/09/2022	Xác định đề tài thực hiện khoá luận và viết đơn đăng ký đề tài
10/09/2022	16/10/2022	Thu thập dữ liệu và tìm kiếm các tài liệu liên quan đến phương pháp thực hiện đề tài.
17/10/2022	06/11/2022	Xây dựng mô hình nhận diện gương mặt
07/11/2022	04/12/2022	Xây dựng mô hình phân tích bối cảnh xung quanh
05/12/2022	17/12/2022	Xây dựng ứng dụng minh họa
18/12/2022	24/12/2022	Làm slide thuyết trình
10/09/2022	24/12/2022	Viết báo cáo

**Xác nhận của CBHD**  
(Ký tên và ghi rõ họ tên)

**TP. HCM, ngày 09 tháng 09 năm 2022**  
**Sinh viên**  
(Ký tên và ghi rõ họ tên)

## **DANH SÁCH HỘI ĐỒNG BẢO VỆ KHÓA LUẬN TỐT NGHIỆP**

Hội đồng chấm khóa luận tốt nghiệp, thành lập theo Quyết định số .....

ngày ..... của Hiệu trưởng Trường Đại học Công nghệ Thông tin.

1. TS. Lê Minh Hưng..... - Chủ tịch
2. ThS Nguyễn Thị Ngọc Diễm..... - Thư ký
3. ThS Đỗ Văn Tiến..... - Uỷ viên

## **LỜI CẢM ƠN**

Em xin gửi lời cảm ơn đến các Quý Thầy cô Khoa Khoa học máy tính, Trường Đại học Công nghệ thông tin – Đại học Quốc gia Thành phố Hồ Chí Minh đã tạo cơ hội cho em được học tập, rèn luyện và tích lũy kiến thức trong suốt quá trình học tập bậc Đại học để em thể hoàn thành khoá luận này.

Đặc biệt, em xin chân thành cảm ơn giảng viên hướng dẫn TS. Mai Tiến Dũng đã tận tình chỉ dẫn, theo dõi và đưa ra những lời khuyên bổ ích giúp em giải quyết được các vấn đề gặp phải trong quá trình nghiên cứu và hoàn thành khoá luận một cách tốt nhất.

Em đã cố gắng hết sức mình để hoàn thành luận. Tuy vậy, những hạn chế trong việc thiếu kiến thức chuyên môn và kinh nghiệm thực tiễn khiến cho khoá luận của em không thể tránh khỏi những thiếu sót. Em kính mong nhận được sự thông cảm cũng như góp ý, chỉ dạy thêm từ Quý Thầy cô.

Trân trọng.

# Mục lục

<b>TÓM TẮT KHOÁ LUẬN</b>	<b>xvi</b>
<b>1 TỔNG QUAN</b>	<b>1</b>
1.1 Giới thiệu bài toán . . . . .	1
1.1.1 Lý do chọn đề tài . . . . .	1
1.1.2 Phát biểu bài toán . . . . .	9
1.1.3 Giới hạn phạm vi khoá luận . . . . .	9
Giới hạn về những người nổi tiếng . . . . .	9
Giới hạn về những sự kiện . . . . .	11
1.2 Tính ứng dụng . . . . .	12
1.3 Những thách thức . . . . .	12
1.3.1 Bộ dữ liệu . . . . .	12
1.3.2 Hình ảnh của cùng một sự kiện sẽ có thể khác biệt rất lớn .	14
1.3.3 Sự kiện không có nhiều ảnh . . . . .	16
1.3.4 Các phương pháp trích xuất đặc trưng đại diện cho thông tin sự kiện . . . . .	17
1.3.5 Phương hướng giải quyết các thách thức . . . . .	17
1.4 Mục tiêu khóa luận . . . . .	18
1.5 Cấu trúc Khóa luận tốt nghiệp . . . . .	18
<b>2 NGHIÊN CỨU LIÊN QUAN</b>	<b>19</b>
2.1 Hệ thống tìm kiếm hình ảnh - Visual Search . . . . .	19
2.2 Hệ thống nhận diện PicTriev . . . . .	21
2.3 Hệ thống nhận diện StarByFace . . . . .	23
2.4 Nhận xét . . . . .	27

<b>3 HƯỚNG TIẾP CẬN TRONG KHÓA LUẬN</b>	<b>28</b>
3.1 Nhận dạng gương mặt của người nổi tiếng . . . . .	28
3.1.1 Phát hiện gương mặt trong hình ảnh . . . . .	29
3.2 Nhận diện sự kiện tương ứng . . . . .	32
3.2.1 Tiền xử lý dữ liệu - Image matting . . . . .	32
3.2.2 Thông tin từ ngữ cảnh . . . . .	37
3.2.3 Thông tin từ màu sắc . . . . .	38
Histogram . . . . .	38
Màu chủ đạo . . . . .	40
3.3 Mô hình máy học sử dụng . . . . .	41
<b>4 THỰC NGHIỆM VÀ ĐÁNH GIÁ</b>	<b>42</b>
4.1 Môi trường và ngôn ngữ cài đặt . . . . .	42
4.2 Tập dữ liệu thực nghiệm . . . . .	43
4.2.1 Tăng cường dữ liệu . . . . .	43
4.2.2 Nhận diện gương mặt người nổi tiếng . . . . .	46
4.2.3 Nhận diện sự kiện tương ứng . . . . .	50
4.3 Xây dựng ứng dụng thực tế . . . . .	55
<b>5 KẾT LUẬN</b>	<b>58</b>
5.1 Kết luận . . . . .	58
5.2 Hướng phát triển . . . . .	58
5.3 Hạn chế của khoá luận . . . . .	59
<b>Tài liệu tham khảo</b>	<b>60</b>

# Danh sách hình vẽ

1.1	Chú thích của ảnh được minh họa trong bài báo sai . . . . .	2
1.2	Ngoại trưởng Nga Sergey Lavrov . . . . .	3
1.3	Bộ trưởng Bộ quốc phòng Nga - Sergey Shoigu . . . . .	4
1.4	Ảnh minh họa cuộc Điện đàm trong bài báo không phải là ảnh minh họa đúng . . . . .	5
1.5	Ảnh minh họa cuộc Điện đàm vào ngày 30/12/2021 trên trang báo VnExpress . . . . .	6
1.6	Ảnh minh họa trong một bài báo về Tổng thống Mỹ Joe Biden . . .	7
1.7	Hình ảnh minh họa Input và Output của bài toán . . . . .	9
1.8	Trang US News - World Report - Bình chọn quốc gia quyền lực . . .	10
1.9	The 'Celebrity Together' Dataset - Bộ dữ liệu của trường Đại học Oxford . . . . .	13
1.10	Large-scale CelebFaces Attributes (CelebA) Dataset . . . . .	13
1.11	Thủ tướng Nhật Bản Kishida Fumio bắt tay Tổng thống Indonesia Joko Widodo trong Hội nghị G20 . . . . .	14
1.12	Thủ tướng Nhật Bản Kishida Fumio trong phòng họp khi tham dự Hội nghị G20 . . . . .	15
1.13	Kết quả tìm kiếm trên trang Alamy . . . . .	16
2.1	Giao diện Visual Search . . . . .	19
2.2	Kết quả Visual Search cho hình ảnh trong một bài báo [12] . . . .	20
2.3	Ảnh truy vấn của Tổng thống Mỹ - Joe Biden trong sự kiện Hội nghị Thượng đỉnh Mỹ - châu Phi . . . . .	21
2.4	Kết quả trả về của hệ thống PicTriev . . . . .	22
2.5	Tổng thống Nga Vladimir Putin trong sự kiện Diễn đàn kinh tế phương Đông . . . . .	23

2.6	Ảnh kết quả của hệ thống StarByFace - Những gương mặt nam giống nhất . . . . .	24
2.7	Ảnh kết quả của hệ thống StarByFace - Những gương mặt nữ giống nhất . . . . .	25
2.8	Ảnh kết quả của hệ thống StarByFace - Nhân vật nam và nhân vật nữ giống với người trong ảnh truy vấn nhất . . . . .	26
3.1	Quá trình học trong bài toán Nhận diện gương mặt . . . . .	28
3.2	Quá trình nhận diện trong bài toán Nhận diện gương mặt . . . . .	29
3.3	Kết quả khi dùng MTCNN (trái) và Kết quả khi dùng Insightface (phải) . . . . .	29
3.4	Kết quả khi dùng MTCNN . . . . .	30
3.5	Kết quả khi dùng Insightface . . . . .	30
3.6	Quá trình huấn luyện mô hình nhận diện sự kiện . . . . .	32
3.7	Quá trình nhận diện sự kiện . . . . .	32
3.8	Hình ảnh minh họa của PaddlePaddle . . . . .	33
3.9	Quy trình matting - Tách ảnh người và nền . . . . .	34
3.10	Quy trình matting - Chèn ảnh nền khác . . . . .	34
3.11	Quy trình trích xuất ảnh nền - Tách ảnh người và nền . . . . .	35
3.12	Loại bỏ ảnh kết quả matting đã thu được . . . . .	35
3.13	Ảnh nền trích xuất được . . . . .	37
3.14	Chia hình thành 3 * 3 vùng . . . . .	39
3.15	So sánh vùng ảnh giữa 2 ảnh trong 1 sự kiện . . . . .	40
3.16	Một minh họa cụ thể về màu chủ đạo . . . . .	40
4.1	Ảnh gốc ban đầu - trước khi thực hiện tăng cường dữ liệu . . . . .	43
4.2	Ảnh đã được tỉ lệ lại 640x640 . . . . .	44
4.3	Ảnh đã được lật theo chiều ngang . . . . .	45
4.4	Hai kết quả minh họa cho bước shear ảnh ban đầu . . . . .	46
4.5	Phân bổ dữ liệu để huấn luyện và thử nghiệm trong mô hình nhận dạng gương mặt . . . . .	47
4.6	Kết quả nhận diện gương mặt với tập dữ liệu gốc ban đầu . . . . .	48
4.7	Phân bổ tập dữ liệu đã qua tăng cường . . . . .	49
4.8	Kết quả nhận diện gương mặt trên tập dữ liệu đã qua tăng cường .	49

4.9	Phân bố số lượng dữ liệu trong tập các sự kiện ban đầu . . . . .	50
4.10	Sự khác biệt của màu chủ đạo trong cùng 1 sự kiện - Chuyến thăm Kyiv của Thủ tướng Anh . . . . .	52
4.11	Phân bố dữ liệu trong tập nhận diện sự kiện đã qua tăng cường .	53
4.12	Giao diện trang web sản phẩm . . . . .	56
4.13	Kết quả trả về của trang web . . . . .	57

# Danh sách bảng

1.1	Những nguyên thủ quốc gia trong phạm vi khoá luận . . . . .	11
1.2	Những sự kiện trong phạm vi khoá luận . . . . .	11
3.1	So sánh 2 phương pháp phát hiện gương mặt . . . . .	31
4.1	Kết quả đánh giá nhận diện sự kiện . . . . .	51
4.2	Kết quả đánh giá nhận diện sự kiện . . . . .	52
4.3	Kết quả đánh giá nhận diện sự kiện . . . . .	54
4.4	Kết quả đánh giá nhận diện sự kiện . . . . .	55

# Danh mục từ viết tắt

**KNN** K- Nearest Neighbors

**MTCNN** Multi- task Cascaded Convolutional Network

**RGB** Red Green Blue

**Acc** Accuracy

# Danh mục từ tạm dịch

Bộ dữ liệu	dataset
Đặc trưng	feature
Đặc trưng thông tin ngữ cảnh	content-based feature
Những ngôi sao nổi tiếng	celebrities
Những chính trị gia	politicians
Những nhân vật lịch sử	historical figures
Nhận diện gương mặt	face detection
Căn chỉnh gương mặt	face alignment
Thư viện hỗ trợ	framework
Đặc trưng học sâu	deep feature
Đặc trưng thủ công	hand-crafted feature
Cảnh nền	background
Bảng màu chính	palette
Pixel	điểm ảnh

# TÓM TẮT KHOÁ LUẬN

Trong thời đại công nghệ, chúng ta có thể dễ dàng cập nhật tin tức thông qua các trang báo điện tử. Các trang báo này thường sẽ sử dụng các hình ảnh để minh họa cho nội dung bài báo. Tuy nhiên, trong một số trường hợp thì các bài báo này không chú thích đúng cho các hình ảnh minh họa hoặc là chú thích sơ sài.

Từ thực trạng đó, mục tiêu của khoá luận là xây dựng được một hệ thống nhận vào một bức ảnh, kết quả đầu ra mong đợi gồm tên của những người nổi tiếng và sự kiện tương ứng trong ảnh.

Bài toán mà khoá luận giải quyết tương đương với bài toán Tạo mô tả - chú thích cho ảnh (Image captioning). Tuy nhiên, trong khoá luận này, em sẽ tập trung vào phần xây dựng một ứng dụng chứ không đi sâu vào nghiên cứu các phương pháp đã có hoặc đề xuất ra một giải pháp mới cho bài toán. Vì thế, trong khoá luận này, em sẽ thực hiện các phương pháp đã có để xây dựng hệ thống, và tìm kiếm phương pháp đạt được hiệu quả tốt nhất.

Ngoài ra, để có thể xây dựng được một hệ thống nhận diện hoàn chỉnh thì việc xây dựng một bộ dữ liệu là không thể thiếu. Đó cũng là một mục tiêu mà khoá luận này sẽ giải quyết - xây dựng một bộ dữ liệu về các sự kiện mà người nổi tiếng tham gia.

## Chương 1

# TỔNG QUAN

### 1.1 Giới thiệu bài toán

#### 1.1.1 Lý do chọn đề tài

Ngày nay, nhờ sự phát triển của công nghệ thông tin và các thiết bị điện tử, lượng tin tức trong một ngày được cập nhật đã tăng lên rất nhiều và sự lan truyền của những tin tức này trở nên nhanh chóng một cách đáng kinh ngạc. Theo số liệu được thống kê trong năm 2022 [1], trong vòng một phút thì có đến 275.000 tweet được gửi đi trên Twitter. Chúng ta có thể dễ dàng cập nhật những thông tin đó thông qua mạng xã hội, các trang báo điện tử,...

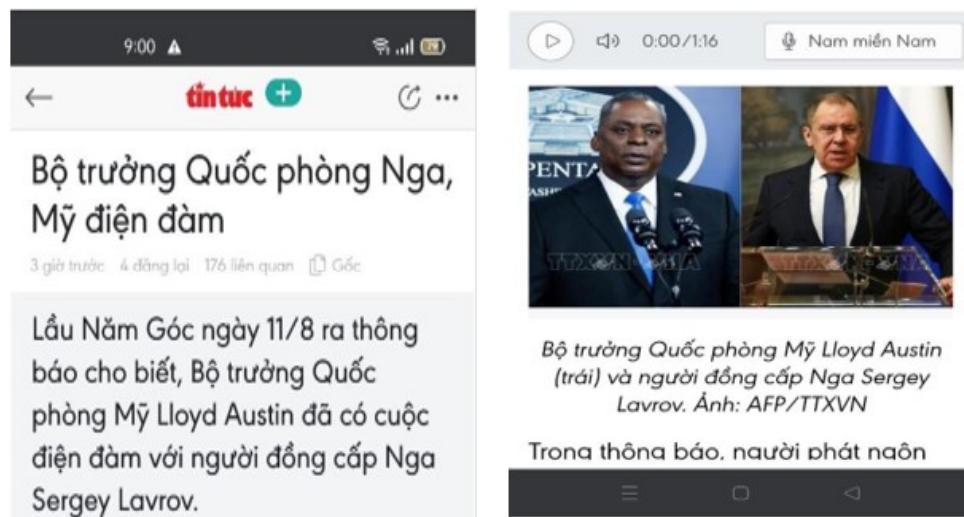
Những người nổi tiếng như các vị nguyên thủ quốc gia, diễn viên, ca sĩ,... là những người có sức ảnh hưởng rất lớn đối với công chúng. Không khó để chúng ta có thấy hình ảnh, tin tức liên quan của họ trên những phương tiện truyền thông đại chúng, mạng xã hội,...

Đối tượng cụ thể mà khoá luận muốn đề cập đến là các trang báo điện tử có sử dụng hình ảnh của người nổi tiếng để minh họa cho nội dung bài báo.

Một vài trường hợp hay gặp phải khi đọc các bài báo có thể kể đến như sau:

- Chú thích của ảnh được minh họa trong bài báo sai, từ đó dẫn đến việc người đọc hiểu sai thông tin được đề cập.

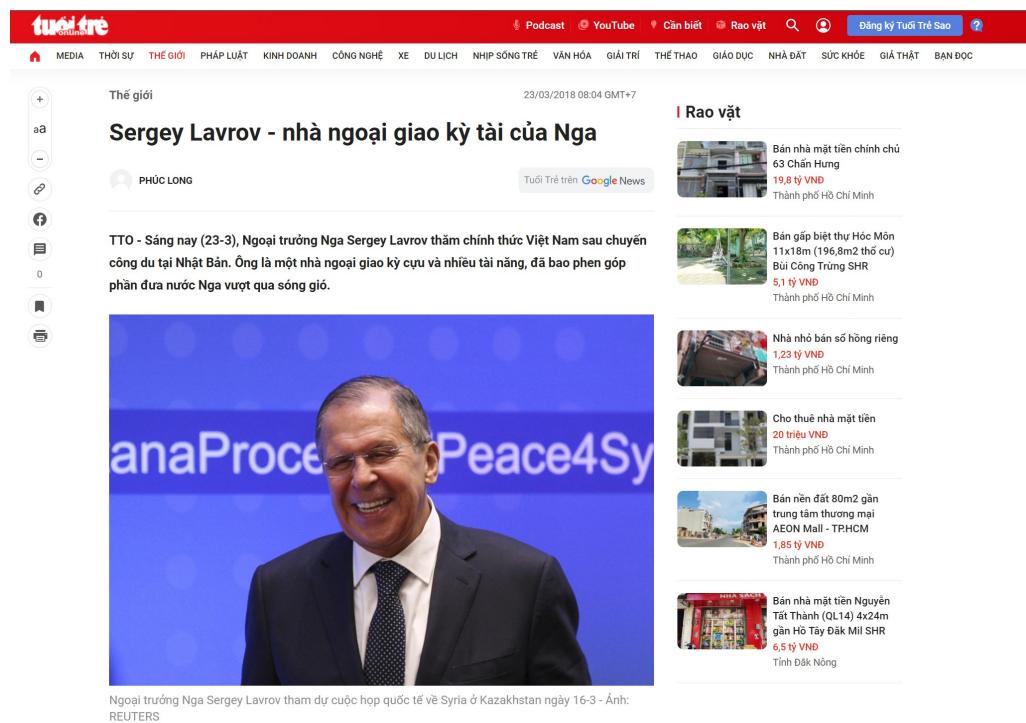
Hình 1.1 dưới đây là một ví dụ cụ thể khi chú thích của ảnh sai khiến người đọc tiếp thu thông tin bị sai lệch.



HÌNH 1.1: Chú thích của ảnh được minh họa trong bài báo sai

Trong chú thích của bài báo này, "Bộ trưởng Quốc phòng Mỹ Lloyd Austin" và cụm "người đồng cấp Nga Sergey Lavrov" khiến cho người đọc hiểu rằng Bộ trưởng Quốc phòng Nga là ông Sergey Lavrov. Điều này hoàn toàn không đúng với sự thật. Ông Sergey Lavrov là Ngoại trưởng Nga hay còn biết đến với chức danh là Bộ trưởng Bộ Ngoại giao Nga.

Hình 1.2 là trích đoạn của một bài báo khác [2] trên trang báo Tuổi Trẻ, một trang được thành lập từ năm 1975 và cũng là cơ quan ngôn luận của Đoàn Thanh niên Cộng sản Thành phố Hồ Chí Minh [3].



HÌNH 1.2: Ngoại trưởng Nga Sergey Lavrov

So sánh thông tin giữa một cơ quan ngôn luận do một tổ chức lâu đời chủ quản và một trang báo mạng không do một tổ chức đáng tin cậy nào vận hành, xét về độ đáng tin cậy của thông tin, em nghĩ mọi người cũng sẽ như em, đều lựa chọn tin tưởng vào thông tin của trang báo Tuổi Trẻ. Từ đó có thể dẫn đến kết luận là chú thích của hình 1.1 là sai.

Ngoài ra, cũng trên trang báo Tuổi Trẻ, em có thể dễ dàng tìm được một bài báo khác có thông tin của Bộ trưởng Bộ Quốc phòng Nga thật sự - Ông Sergey Shoigu [4] như sau:

The screenshot shows a news article from **Tuổi Trẻ** (Tuoi Tre) website. The headline reads: **Chào năm mới, Bộ trưởng Quốc phòng Nga tuyên bố 'chắc thắng' ở Ukraine**. The article is dated 31/12/2022 18:19 GMT+7. On the right side of the page, there is a sidebar titled **Rao vặt** (Real estate ads) showing several property listings.

**BÌNH AN** | Nghe đọc bài 2:15 | 1x | **Tuổi Trẻ trên Google News**

Trong thông điệp năm mới qua video, Bộ trưởng Quốc phòng Nga Sergei Shoigu tuyên bố chiến thắng của Nga trước Ukraine là điều 'chắc chắn xảy ra', khi ông ca ngợi 'chủ nghĩa anh hùng' của các binh sĩ Nga.

**Bộ trưởng Quốc phòng Nga Sergei Shoigu**

Bán nhà mặt tiền chính chủ  
63 Chân Hưng  
19.8 tỷ VND  
Thành phố Hồ Chí Minh

Bán gấp biệt thự Hóc Môn  
11x18m (196,8m<sup>2</sup> thổ cư)  
Bùi Công Trứ SHR  
5,1 tỷ VND  
Thành phố Hồ Chí Minh

Nhà nhỏ bán sổ hồng riêng  
1,23 tỷ VND  
Thành phố Hồ Chí Minh

Cho thuê nhà mặt tiền  
20 triệu VND  
Thành phố Hồ Chí Minh

Bán nền đất 80m<sup>2</sup> gần  
trung tâm thương mại  
AEON Mall - TP.HCM  
1,85 tỷ VND  
Thành phố Hồ Chí Minh

Bán nhà mặt tiền Nguyễn  
Tất Thành (QL14) 4x24m  
gần Hố Tây Đăk Mil SHR  
6,5 tỷ VND  
Tỉnh Đăk Nông

HÌNH 1.3: Bộ trưởng Bộ quốc phòng Nga - Sergey Shoigu

Việc đưa chú thích về hình ảnh của hai vị nguyên thủ quốc gia sai lệch như vậy, sẽ khiến người người ít cập nhật thông tin vô tình nhận diện sai mặt của hai người đó.

- Hình ảnh minh họa không đúng với nội dung trong bài báo

The screenshot shows the homepage of VietnamNet, a Vietnamese news website. The header includes the logo 'vietnamnet' and the tagline 'VÌ VIỆT NAM HƯNG CƯỜNG'. It features a navigation bar with categories like Chính trị, Thời sự, Kinh doanh, Thể giới, Giải trí, Thể thao, Sức khỏe, Đời sống, Giáo dục, Pháp luật, Xe, Công nghệ, Bất động sản, Tuần Việt Nam, Du lịch, Bạn đọc, and Video. Below the header, there is a large headline in bold: 'Điện đàm không đạt kết quả, ông Biden cảnh báo Nga về 'hậu quả nghiêm trọng''. The main text discusses the phone call between Biden and Putin, mentioning it lasted over an hour. A photograph of Joe Biden at his desk is shown, and a caption below it reads: 'Điện đàm Biden - Putin kéo dài hơn 1 giờ đồng hồ. Ảnh: Nhà Trắng'.

HÌNH 1.4: Ảnh minh họa cuộc Điện đàm trong bài báo không phải là ảnh minh họa đúng

Trong bài báo trên [5], nội dung được nêu lên rằng vào ngày 13/02/2022, Tổng thống Mỹ Joe Biden và Tổng thống Nga Vladimir Putin đã có một cuộc điện đàm và hình 1.4 là hình minh họa được cho là chụp trong cuộc điện đàm đó. Tuy nhiên, theo một trang báo khác là VnExpress, một trang báo do Bộ Khoa học Công nghệ chủ quản [6], đó thực ra là hình ảnh của một cuộc điện đàm khác đã được diễn ra trước đó, vào tháng 12 năm 2021.

VNEXPRESS

Tin theo khu vực

Các quan chức Nhà Trắng cho rằng không giống như người tiền nhiệm Donald Trump, Tổng thống Biden đã cố tình né tránh những cuộc gặp mặt đối mặt với ông Putin, điều mà ông cho là Moskva mong muốn đạt được để thể hiện vị thế ngang hàng với Mỹ.

Thay vào đó, ông để cho các lãnh đạo phương Tây khác thảo luận với Tổng thống Nga. Khi cuộc khủng hoảng Ukraine bùng nổ, Tổng thống Mỹ không tham gia các nỗ lực ngoại giao trực tiếp với Nga, mà các đợt ngoại giao còn lại do lãnh đạo ở châu Âu tiến hành.

"Điều ông Putin muốn làm là gây sức ép với Kiev", nghị sĩ Dân chủ Greg Meeks, chủ tịch Ủy ban Đối ngoại Hạ viện, nói. "Còn điều Tổng thống Biden đang làm là để cả thế giới gây sức ép với ông Putin".

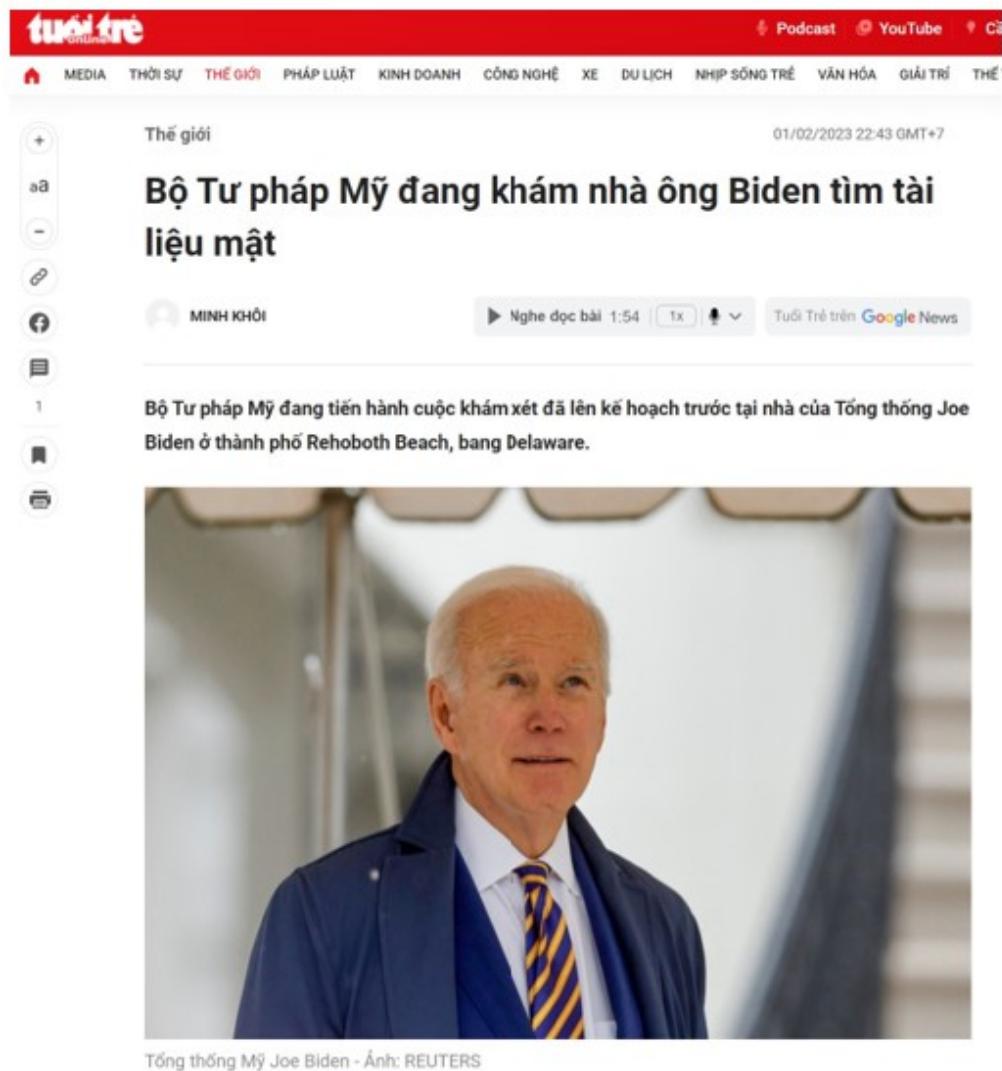


Tổng thống Mỹ Joe Biden điện đàm với người đồng cấp Nga tại Wilmington, Delaware, Mỹ hôm 30/12/2021.  
Ảnh: Nhà Trắng.

HÌNH 1.5: Ảnh minh họa cuộc Điện đàm vào ngày 30/12/2021 trên trang báo VnExpress

Tuy nội dung và hình ảnh minh họa của bài báo đúng là đều nói về sự kiện cuộc điện đàm của Tổng thống Mỹ Joe Biden và Tổng thống Nga Vladimir Putin, tuy nhiên nội dung nói về cuộc điện đàm ngày 13/2/2022 còn hình ảnh là của cuộc điện đàm ngày 30/12/2021, cách đó gần 2 tháng. Đây là tình huống mà chúng ta gặp phải rất nhiều trong lúc đọc các trang tin tức như vậy, hình ảnh và nội dung không đồng nhất với nhau, dẫn đến việc người đọc sẽ dễ bị đánh lừa. Lấy một ví dụ cụ thể và gần gũi với chúng ta là trong lúc dịch bệnh Covid-19, cũng có rất nhiều bài báo đưa tin và đăng kèm hình ảnh của các vị lãnh đạo nước ta, và nói rằng có chỉ thị mới ra gần đây. Hình ảnh của phiên họp đó có khi đã diễn ra từ rất lâu về trước, nhưng đó vẫn là hình ảnh hội họp, dẫn đến người đọc sẽ tin tưởng nội dung thông tin hơn, dù cho đó là tin không chính xác đi nữa.

- Hình ảnh minh họa thiêu thông tin



HÌNH 1.6: Ảnh minh họa trong một bài báo về Tổng thống Mỹ Joe Biden

Hình 1.6 là một đoạn được trích trong một bài báo về Tổng thống Mỹ Joe Biden của trang báo Tuổi Trẻ [7], chú thích của hình là "Tổng thống Mỹ Joe Biden - Ảnh: REUTERS" nhưng không nêu cụ thể là đó là hình ảnh của sự kiện nào. Có thể là sự kiện trong nội dung bài báo không có hình ảnh minh họa, và người viết bài báo chỉ muốn tìm một ảnh về nhân vật được đề cập để người đọc dễ hình dung, tuy nhiên vì không có thông tin về sự kiện cụ thể nào đó, người đọc sẽ dễ bị hiểu nhầm ý của bài báo, và hiểu thành hình ảnh đó chính là ảnh của sự kiện đang đề cập trong bài báo.

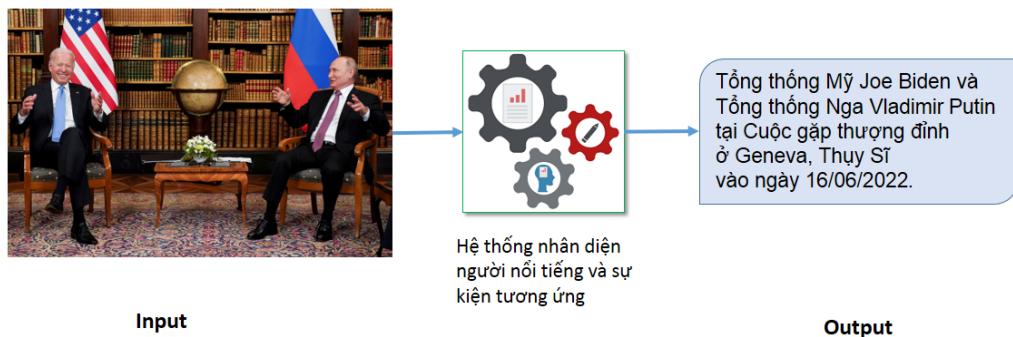
Từ đó có thể thấy, một hệ thống có thể xác định được gương mặt của người nổi tiếng và sự kiện người đó đang tham gia là giải pháp có thể góp phần giảm thiểu tình trạng người đọc tin tức bị ảnh hưởng bởi những thông tin không đủ chính xác như vậy, giúp người đọc hiểu rõ hơn về hình ảnh minh họa trong bài báo.

### 1.1.2 Phát biểu bài toán

Theo những thông tin đã được nêu ở trên, khoá luận này sẽ hướng đến mục tiêu là xây dựng một hệ thống có thể nhận diện gương mặt của người nổi tiếng và sự kiện người đó tham gia.

- Đầu vào: Một ảnh có chứa (một hoặc nhiều) người nổi tiếng.
- Đầu ra: Cho biết tên của người nổi tiếng và sự kiện tương ứng của hình ảnh đầu vào.

Hình 1.7 minh họa bài toán mà khóa luận sẽ giải quyết.



HÌNH 1.7: Hình ảnh minh họa Input và Output của bài toán

### 1.1.3 Giới hạn phạm vi khoá luận

#### Giới hạn về những người nổi tiếng

Theo định nghĩa của trang Bing Visual Search do Microsoft sáng lập, người nổi tiếng có thể là "celebrities, politicians, historical figures, and other famous faces" [8]. Trong phạm vi khoá luận này, em sẽ chỉ hướng đến nhóm "politicians" - những nguyên thủ quốc gia.

Dựa theo một bình chọn trên trang US News & World Report - một trang web đã tồn tại hơn 30 năm, trang web này bao gồm các bảng xếp hạng khác nhau thuộc các lĩnh vực như kinh tế chính trị, giáo dục, y tế,...

Trong một bình chọn thống kê những quốc gia quyền lực nhất trong năm 2022 [9], thống kê này dựa trên tầm ảnh hưởng của quốc gia đó đối với thế giới, sự áp

đảo về lượng tin tức được cập nhật của quốc gia đó so với các quốc gia khác, mối liên kết với các quốc gia đồng minh, lực lượng quân đội,...[10]. Các tiêu chí để bình chọn được nêu trong hình 1.8



Home / News / Best Countries / Most Powerful

## Power

The world's most powerful countries also are the ones that consistently dominate news headlines, preoccupy policymakers and shape global economic patterns. Their foreign policies and military budgets are tracked religiously. When they make a pledge, at least some in the international community trust they will keep it. These countries project their influence on the world stage.

---

HÌNH 1.8: Trang US News - World Report - Bình chọn quốc gia quyền lực

Kết quả gồm 85 quốc gia như sau:

1. Mỹ
2. Trung Quốc
3. Nga
4. Đức
5. Anh
6. Hàn Quốc
7. Pháp
8. Nhật Bản
9. Các Tiểu vương quốc Ả Rập Thống nhất
10. Israel
11. ...

Trong phạm vi khoá luận này, em chỉ chọn nhóm 8 những quốc gia đầu tiên, lý do dẫn đến sự lựa chọn này là vì em ưu tiên tiêu chí "có số lượng tin tức áp đảo" của các quốc gia đó. Và với mỗi quốc gia sẽ có một nguyên thủ quốc gia tương ứng. Danh sách các nguyên thủ quốc gia được thể hiện trong bảng 1.1 dưới đây:

BẢNG 1.1: Các quốc gia trong phạm vi khoá luận và nguyên thủ quốc gia tương ứng

Quốc gia	Tên nguyên thủ quốc gia	Chức vụ
Mỹ	Joe Biden	Tổng thống Mỹ
Trung Quốc	Tập Cận Bình	Chủ tịch Trung Quốc
Nga	Vladimir Putin	Tổng thống Nga
Đức	Olaf Scholz	Thủ tướng Đức
Anh	Boris Johnson	Thủ tướng Anh
Hàn Quốc	Yoon Suk Yeol	Tổng thống Hàn Quốc
Pháp	Emmanuel Macron	Tổng thống Pháp
Nhật Bản	Kishiada Fumio	Thủ tướng Nhật Bản

### Giới hạn về những sự kiện

Những sự kiện được chọn dùng để thực hiện khoá luận đều là những sự kiện có các nguyên thủ quốc gia trong bảng 1.1 tham dự, và các sự kiện này đều diễn ra vào năm 2022.

Các sự kiện và thông tin liên quan được trình bày cụ thể trong bảng 1.2 dưới đây:

BẢNG 1.2: Thông tin về các sự kiện trong phạm vi khoá luận

Tên sự kiện	Ngày diễn ra sự kiện	Thành phố diễn ra sự kiện	Quốc gia diễn ra sự kiện
Chuyến thăm Ấn Độ	19/3/2022	New Delhi	Ấn Độ
Hội nghị G7	26/6/2022	Bayern	Đức
Hội nghị thượng đỉnh NATO	29/6/2022	Madrid	Tây Ban Nha
Chuyến thăm Kyiv nhân dịp Quốc khánh Ukraine	24/8/2022	Kyiv	Ukraine
Diễn đàn Kinh tế Phương Đông	7/9/2022	Vladivostok	Nga
Khoá họp thứ 77 Đại hội đồng Liên hợp quốc	20/9/2022	New York	Mỹ
Sự kiện của Ủy ban Quốc gia Đảng Dân chủ	18/10/2022	Washington	Mỹ
Hội nghị Cấp cao ASEAN	13/11/2022	Phnom Penh	Campuchia
Hội nghị G20	15/11/2022	Bali	Indonesia
Hội nghị thượng đỉnh Mỹ - châu Phi	15/12/2022	Washington	Mỹ

Các sự kiện trên có thể chỉ có 1 nguyên thủ quốc gia tham dự hoặc nhiều nguyên thủ quốc gia cùng tham gia.

## 1.2 Tính ứng dụng

Như đã đề cập, bài toán này có thể góp phần trong việc xác định được thông tin đã được đăng tải và hình ảnh minh họa có mối liên hệ nào với nhau hay không, từ đó người đọc có thể xem xét tính đúng đắn của bài báo, thông tin mà mình đang đọc.

Ngoài ra, kết quả của bài toán này cũng có thể ứng dụng trong lĩnh vực giáo dục. Chẳng hạn như khi các bạn học sinh nhìn thấy ảnh của một vị anh hùng lịch sử, dân tộc nào đó, các bạn có thể sử dụng hệ thống này để tìm hiểu xem nhân vật lịch sử đó là ai, đang tham gia sự kiện gì, và hệ thống sẽ mở rộng thêm những thông tin về sự kiện đó. Từ đó, việc học hỏi thêm được những kiến thức mới bổ ích qua cách làm trực quan như vậy sẽ khiến các bạn hứng thú hơn với bộ môn Lịch sử, Ngữ văn,...

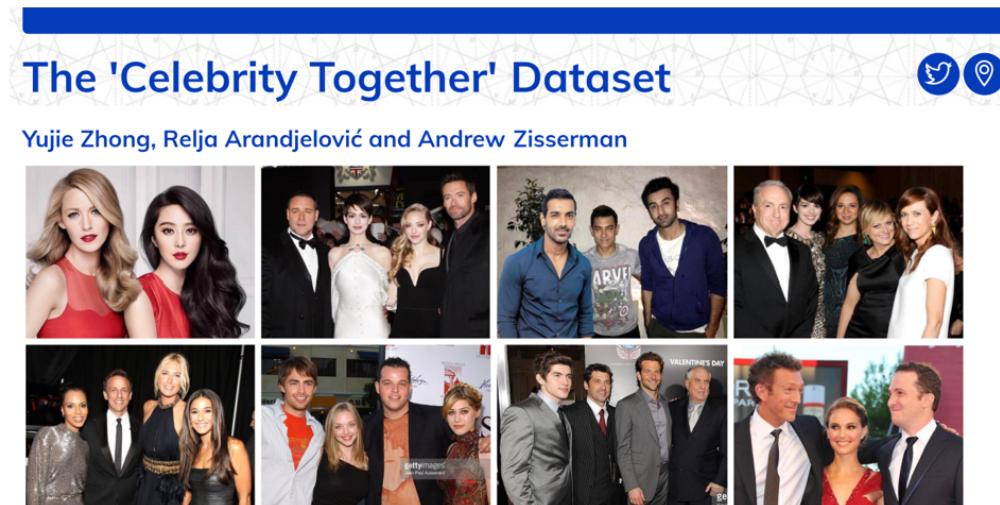
## 1.3 Những thách thức

Bên cạnh những thách thức chung giống với các bài toán nhận diện gương mặt, trích xuất đặc trưng từ hình ảnh khác như là góc chụp ảnh, độ sáng, biểu cảm của gương mặt,... Dưới đây là những thách thức riêng của bài toán này cần phải giải quyết.

### 1.3.1 Bộ dữ liệu

Đa phần các nghiên cứu trước đây đều tập trung vào việc xây dựng các bộ dữ liệu để nhận diện gương mặt của người nổi tiếng như ngôi sao, ca sỹ,...

Đây là một vài bộ dữ liệu nhận diện gương mặt người nổi tiếng đã được thu thập và chú thích. Tuy nhiên, các bộ dữ liệu đều không có thông tin gì về việc người đó đang tham gia vào sự kiện nào.



HÌNH 1.9: The 'Celebrity Together' Dataset - Bộ dữ liệu của trường  
Đại học Oxford

#### Sample Images



HÌNH 1.10: Large-scale CelebFaces Attributes (CelebA) Dataset

### 1.3.2 Hình ảnh của cùng một sự kiện sẽ có thể khác biệt rất lớn

Trong cùng một sự kiện, các vị nguyên thủ có thể di chuyển sang một địa điểm khác để chụp ảnh, hoặc thay đổi quần áo chung cho sự kiện đó,... Như trong minh họa dưới đây, hình 1.11 và hình 1.12 đều cùng là hình chụp của Thủ tướng Nhật Bản Kishida Fumio trong Hội nghị G20 tổ chức tại Indonesia, tuy nhiên, do đã thay đổi địa điểm chụp ảnh, điều đó dẫn đến việc ảnh của cùng một sự kiện có thể khác biệt nhau rất nhiều, không có nhiều sự tương đồng để có thể nhận biết.



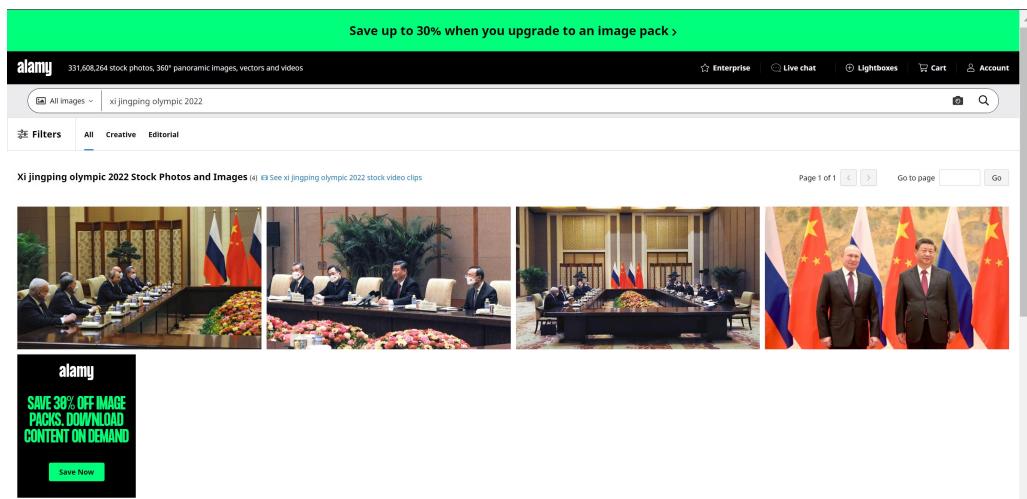
HÌNH 1.11: Thủ tướng Nhật Bản Kishida Fumio bắt tay Tổng thống Indonesia Joko Widodo trong Hội nghị G20



HÌNH 1.12: Thủ tướng Nhật Bản Kishida Fumio trong phòng họp  
khi tham dự Hội nghị G20

### 1.3.3 Sự kiện không có nhiều ảnh

Có thể do tính bảo mật của các sự kiện, nên khi có một sự kiện nào đó cần riêng tư, các nhà báo chỉ có thể chụp được một vài tấm ảnh tượng trưng chứ không thể nào theo sát được nguyên thủ quốc gia đó để chụp được nhiều ảnh hơn. Ví dụ như trong một sự kiện gặp gỡ ở Bắc Kinh trước thềm diễn ra Thế vận Hội mùa đông 2022 của Tổng thống Nga Vladimir Putin và Chủ tịch Trung Quốc Tập Cận Bình, theo trang web Alamy [11], nơi có nhiều rất nhiều nguồn ảnh của các ký giả được tổng hợp lại, thì sự kiện đó chỉ có 4 ảnh.



HÌNH 1.13: Kết quả tìm kiếm trên trang Alamy

Điều đó dẫn đến việc có rất nhiều các sự kiện có ít ảnh, hình ảnh về sự kiện đó không phong phú sẽ gây khó khăn trong quá trình xây dựng bộ dữ liệu.

### **1.3.4 Các phương pháp trích xuất đặc trưng đại diện cho thông tin sự kiện**

Ta có thể rút trích được rất nhiều đặc trưng khác nhau trong một bức hình như là đặc trưng về màu sắc, sử dụng mạng học sâu để rút trích đặc trưng thông tin ngữ cảnh,... tuy nhiên, cho đến hiện tại, vẫn chưa có bài báo khoa học nào cho biết được trong rất nhiều những đặc trưng được trích xuất đó, đặc trưng nào có thể đại diện cho thông tin về sự kiện của một bức hình.

### **1.3.5 Phương hướng giải quyết các thách thức**

Trong khoá luận này, với các thách thức về bộ dữ liệu, em sẽ giải quyết bằng cách chọn lựa những sự kiện không có biến đổi quá nhiều về cảnh nền để có thể rút trích đặc trưng tốt hơn. Hơn nữa, để xây dựng tính đa dạng cho dữ liệu, em sẽ tìm kiếm những sự kiện lớn, công khai và có nhiều ảnh được chụp nhất có thể. Thách thức này đã được giải quyết một phần khi em chọn lọc những sự kiện như trong bảng 1.2. Ngoài ra, còn có thể tăng cường dữ liệu trước khi thực hiện rút trích thông tin, điều đó cũng sẽ giúp cho dữ liệu được đa dạng hơn, từ đó tăng độ chính xác của hệ thống và tính phong phú của dữ liệu.

Về thách thức về các phương pháp, phương hướng giải quyết của em là cố gắng thử nhiều phương pháp rút trích đặc trưng khác nhau nhất có thể, không giới hạn ở đặc trưng thủ công hay đặc trưng học sâu, từ đó có thể chọn ra phương pháp rút trích thông tin sự kiện tốt nhất đối với bộ dữ liệu mà em có cho bài toán này.

## 1.4 Mục tiêu khóa luận

Mục tiêu chính của khoá luận này là tìm ra được một phương pháp rút trích đặc trưng phù hợp với bộ dữ liệu thu thập được, để có thể xây dựng lên một hệ thống nhận biết được nhân vật trong hình và xác định được đó là sự kiện nào với độ chính xác tốt nhất có thể trong khả năng.

## 1.5 Cấu trúc Khóa luận tốt nghiệp

Nội dung Khóa luận tốt nghiệp được tổ chức như sau:

- Chương 1 giới thiệu tổng quan về khóa luận.
- Chương 2 trình bày các nghiên cứu liên quan đến bài toán này.
- Chương 3 trình bày chi tiết các hướng tiếp cận của khóa luận.
- Chương 4 trình bày môi trường thực nghiệm, tập dữ liệu, phương pháp đánh giá và kết quả thực nghiệm.
- Chương 5 là phần kết luận và hướng phát triển của khóa luận.

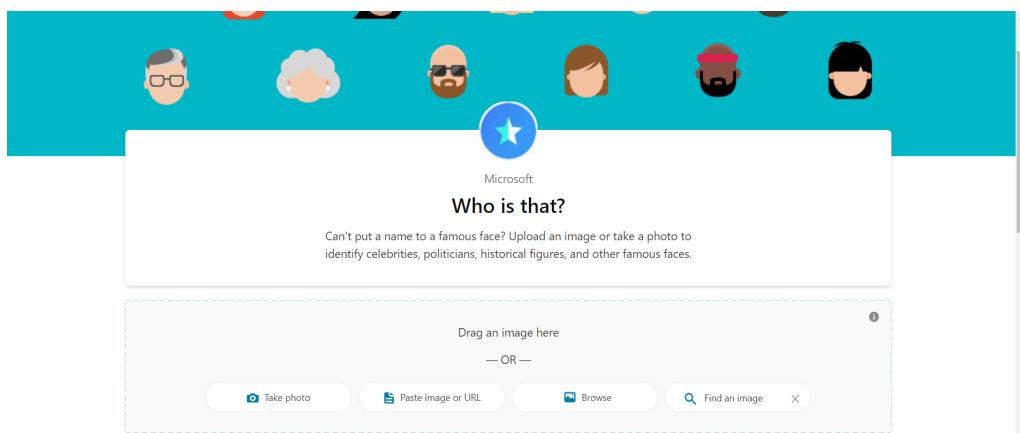
## Chương 2

# NGHIÊN CỨU LIÊN QUAN

Phương hướng chính của khoá luận là xây dựng một ứng dụng có thể nhận diện, do đó, trong phần các nghiên cứu liên quan này, em xin được phép trình bày các ứng dụng khác đã có để nhận diện người nổi tiếng.

### 2.1 Hệ thống tìm kiếm hình ảnh - Visual Search

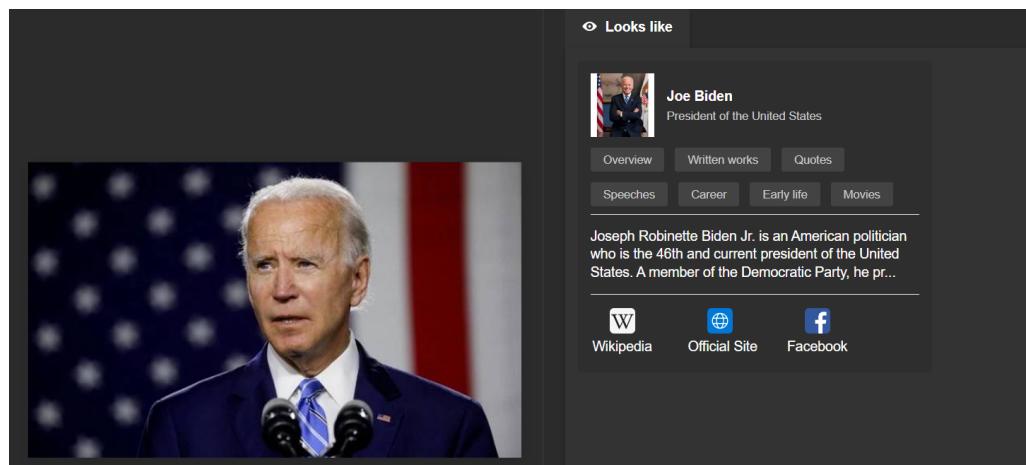
Như trong chương 1 em đã có đề cập khi nói đến định nghĩa "người nổi tiếng", công cụ tìm kiếm thông tin về một người nổi tiếng Visual Search của Microsoft Bing đã được phát triển và sử dụng rộng rãi.



HÌNH 2.1: Giao diện Visual Search

Chúng ta có thể tìm kiếm thông tin về một người nổi tiếng nào đó thông qua việc đưa url của hình ảnh hoặc tải ảnh đã có lên để hệ thống tìm kiếm và đưa ra những thông tin liên quan về người đó.

Dưới đây là minh họa kết quả tìm kiếm có được khi sử dụng hình ảnh của Tổng thống Mỹ Joe Biden được minh họa trong một bài báo [12].



HÌNH 2.2: Kết quả Visual Search cho hình ảnh trong một bài báo [12]

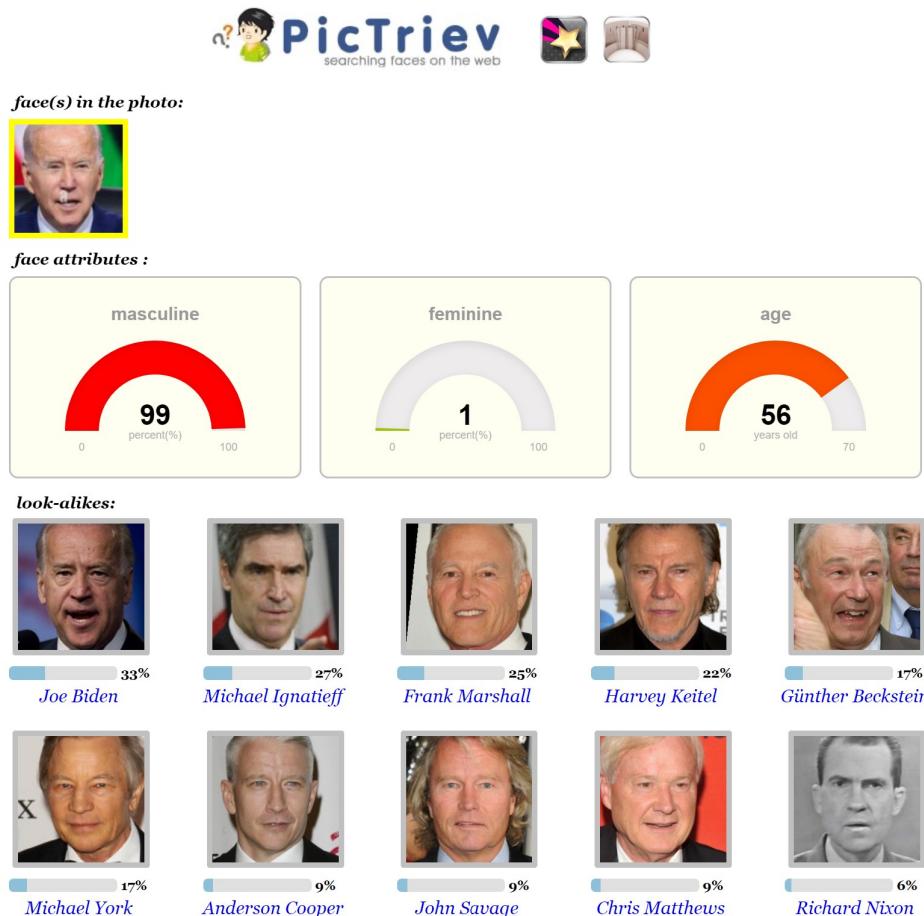
Có thể thấy, khả năng nhận diện gương mặt và xác định nhân vật đó là ai của Visual Search không cần phải bàn cãi, tuy nhiên, hạn chế của Visual Search là chỉ dừng lại ở mức đưa ra các thông tin liên quan đến nhân vật, chứ không đi sâu hơn về việc nhân vật đang ở sự kiện nào. Hạn chế này lại là mục tiêu lớn nhất mà bài toán này hướng đến. Nếu có thể cải thiện được điểm này, hệ thống tìm kiếm hình ảnh của người nổi tiếng sẽ càng giúp người khác biết thêm được nhiều thông tin hơn (dựa vào sự kiện tham gia), chứ không chỉ dựa vào những thông tin về lý lịch như vậy nữa.

## 2.2 Hệ thống nhận diện PicTriev

Tương tự với Visual Search của Microsoft, một hệ thống nhận diện người nổi tiếng khác là PicTriev [13] đã được phát triển và sử dụng. Hình 2.3 là hình ảnh truy vấn để đưa vào hệ thống, và hình 2.4 là kết quả trả về của hệ thống này.



HÌNH 2.3: Ảnh truy vấn của Tổng thống Mỹ - Joe Biden trong sự kiện Hội nghị Thượng đỉnh Mỹ - châu Phi

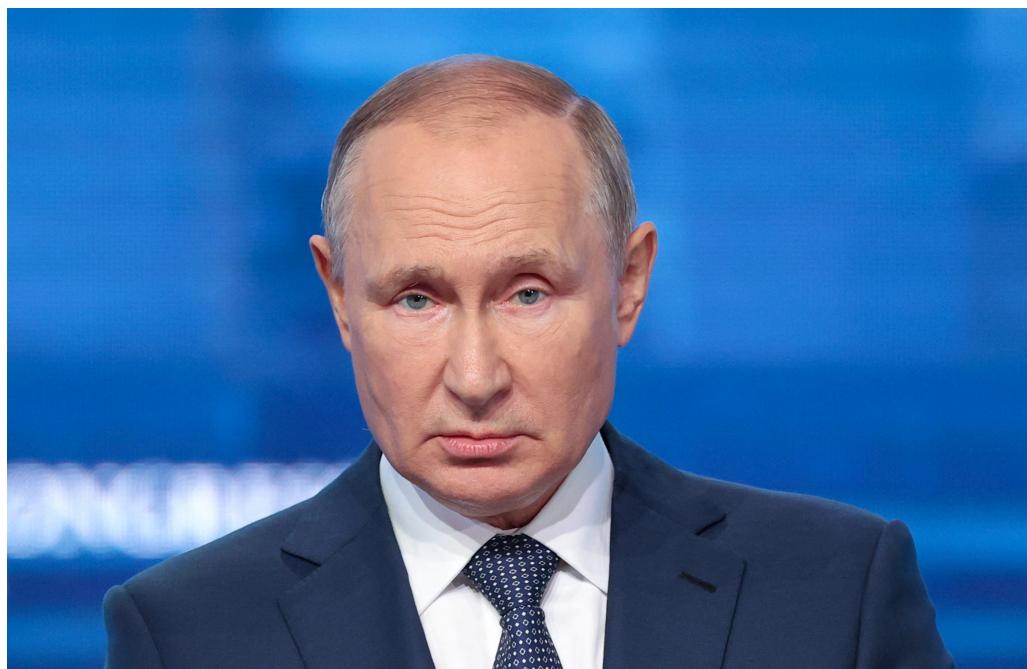


HÌNH 2.4: Kết quả trả về của hệ thống PicTriev

Khác với kết quả của hệ thống Visual Search của Microsoft, PicTriev sẽ trả về kết quả gồm phần trăm giới tính của người trong ảnh, độ tuổi xấp xỉ của người trong ảnh. Ngoài ra, PicTriev còn trả về một danh sách những gương mặt tương đương với ảnh truy vấn kèm theo phần trăm tương đương giữa hai gương mặt như trong hình 2.4 mô tả. Kết quả khi sử dụng hệ thống PicTriev nhận diện được đây là Tổng thống Mỹ Joe Biden, tuy nhiên hình ảnh minh họa của hệ thống đã khá cũ, có thể thấy trong hình trả về của hệ thống, tổng thống Mỹ lúc đó vẫn còn trẻ hơn rất nhiều so với hiện tại. Từ đó có thể kết luận được hệ thống vẫn sẽ đưa ra được kết quả chính xác, tuy nhiên hệ thống chỉ có thể hoạt động tốt nhất với bộ dữ liệu cũ, các ảnh cũ.

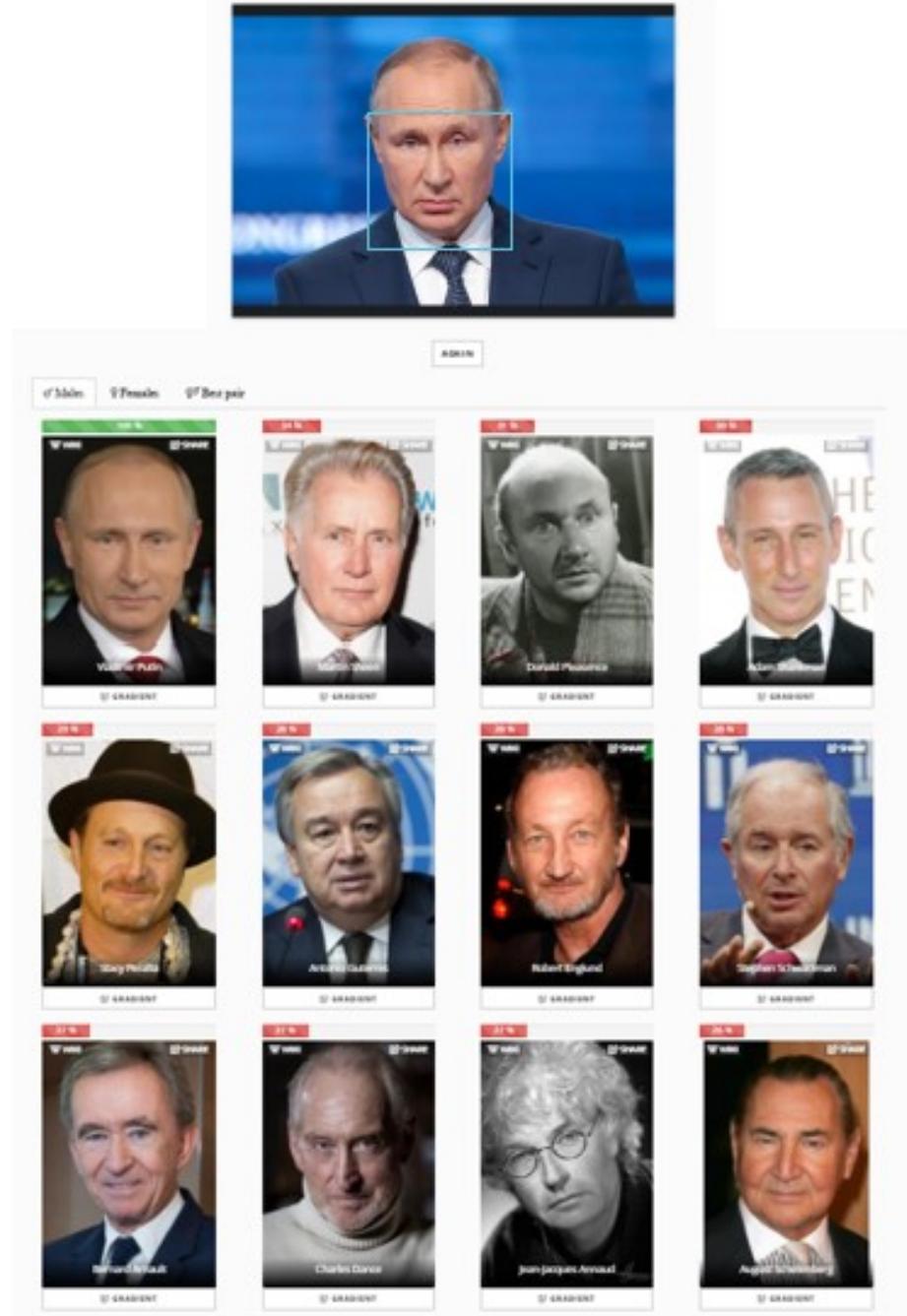
## 2.3 Hệ thống nhận diện StarByFace

Một hệ thống nhận diện gương mặt người nổi tiếng khác là StarByFace [14], em sẽ dùng hình 2.5 để làm ảnh input cho hệ thống. Kết quả trả về của hệ thống

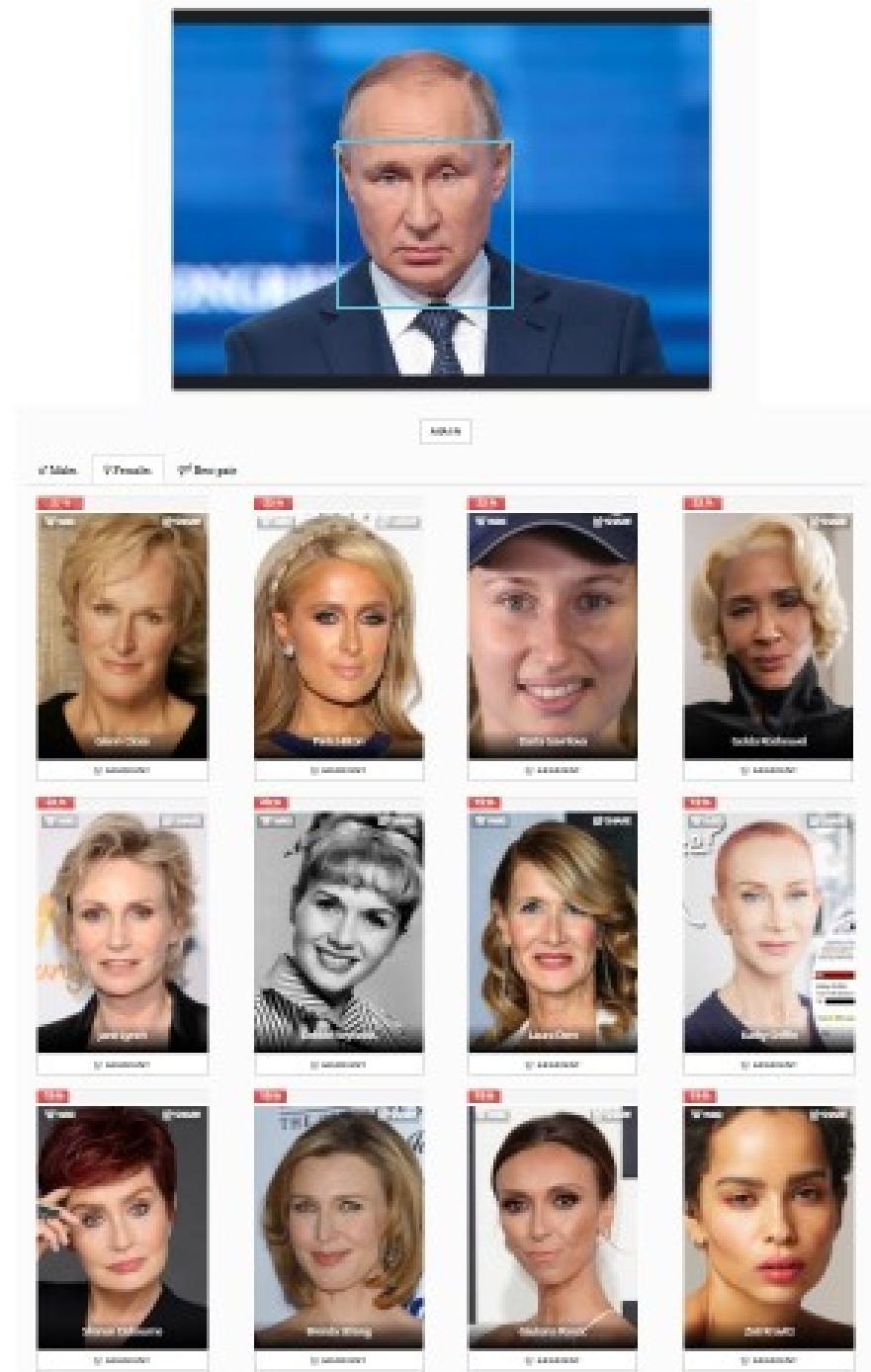


HÌNH 2.5: Tổng thống Nga Vladimir Putin trong sự kiện Diễn đàn kinh tế phương Đông

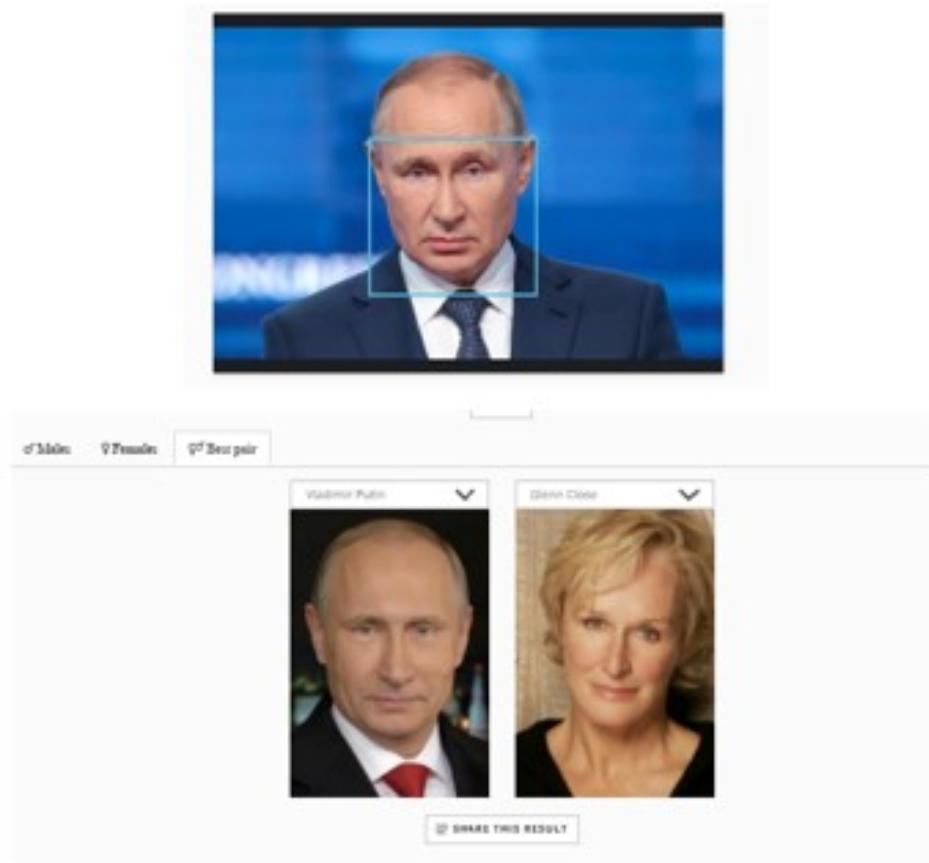
StarByFace sẽ có 2 lựa chọn, là nhóm những nhân vật nam giống với người trong ảnh nhất - hình 2.6 và nhóm những nhân vật nữ giống với người trong ảnh nhất hình - 2.7. Ngoài ra còn một lựa chọn khác đó chính là một cặp nhân vật nam - nữ giống với người trong ảnh nhất - hình 2.8. Bên cạnh đó, ngoài hình ảnh của những gương mặt tương đồng, hệ thống còn đưa thêm đường dẫn đến trang Wikipedia của gương mặt đó để người dùng có thể biết thêm được một số thông tin cá nhân cơ bản. Ta có thể nhận xét được rằng hệ thống StarByFace có thể nhận diện được đây là Tổng thống Nga Vlidimir Putin trong nhóm những gương mặt nam giống nhất, và có độ tương đồng bằng 100% với ảnh truy vấn. Đây là một kết quả rất khả quan, tuy nhiên, hệ thống này đang có một nhược điểm giống với PicTriev, đó chính là bộ dữ liệu được dùng để nhận diện đã lỗi thời.



HÌNH 2.6: Ảnh kết quả của hệ thống StarByFace - Những gương mặt nam giống nhất



HÌNH 2.7: Ảnh kết quả của hệ thống StarByFace - Những gương mặt nữ giống nhất



HÌNH 2.8: Ảnh kết quả của hệ thống StarByFace - Nhân vật nam và nhân vật nữ giống với người trong ảnh truy vấn nhất

## 2.4 Nhận xét

Những hạn chế của các ứng dụng đã nêu trên gồm:

- Chỉ có thể nhận diện được nhân vật nổi tiếng trong ảnh, đưa ra thông tin lý lịch tổng quan, chứ không thể nêu rõ rằng người đó đang tham gia sự kiện nào, diễn ra ở đâu.
- Sử dụng những bộ dữ liệu cũ, không còn phù hợp với thời điểm hiện tại do những người nổi tiếng đã thay đổi nhiều về ngoại hình; không thể nhận diện được những chính trị gia mới nhậm chức do khác biệt thời điểm thu thập dữ liệu.

## Chương 3

# HƯỚNG TIẾP CẬN TRONG KHÓA LUẬN

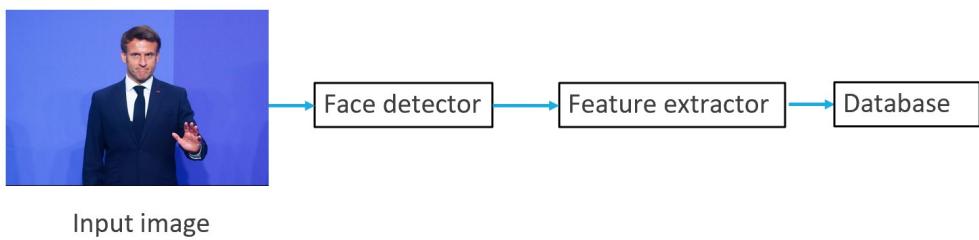
Trong bài toán mà khoá luận đang muôn giải quyết, ta có thể chia nhỏ vấn đề ra thành hai bài toán con gồm:

- Nhận diện gương mặt người nổi tiếng
- Nhận diện sự kiện

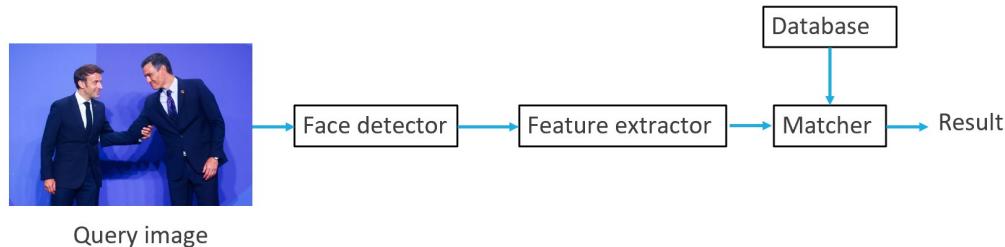
Từ đó, em sẽ tách bài toán ban đầu thành hai bài toán nhỏ hơn để giải quyết, sau đó sẽ kết hợp hai kết quả của các bài toán con, tổng hợp lại thành kết quả đầu ra của bài toán nhận diện người nổi tiếng và sự kiện tương ứng

### 3.1 Nhận dạng gương mặt của người nổi tiếng

Quá trình học - hình 3.1 và quá trình nhận diện - hình 3.2 bao gồm các bước xử lý như sau:



HÌNH 3.1: Quá trình học trong bài toán Nhận diện gương mặt



HÌNH 3.2: Quá trình nhận diện trong bài toán Nhận diện gương mặt

### 3.1.1 Phát hiện gương mặt trong hình ảnh

Trong bài toán này, em sẽ thử khảo sát hai phương pháp phát hiện gương mặt - face detector đang được sử dụng rộng rãi: MTCNN [15] và Insightface [16].

Em sẽ thực hiện phát hiện gương mặt trong một tập dữ liệu 120 ảnh thu thập để đánh giá tính hiệu quả của hai phương pháp, sau đó mới quyết định lựa chọn một phương pháp để thực hiện trong chương 4 trên bộ dữ liệu đã được chuẩn bị cho bài toán.

Dưới đây là một số minh họa cho việc đánh giá hai phương pháp phát hiện gương mặt:



HÌNH 3.3: Kết quả khi dùng MTCNN (trái) và Kết quả khi dùng Insightface (phải)

Trong hình 3.3, khi sử dụng MTCNN đã không thể phát hiện được gương mặt góc nghiêng của Tổng thống Mỹ Joe Biden, còn khi sử dụng Insightface thì có thể phát hiện được cả 2 gương mặt.



HÌNH 3.4: Kết quả khi dùng MTCNN

Trong hình 3.4, MTCNN đã có thể phát hiện được hầu hết những gương mặt trong bức ảnh, ngoài ra thì góc bàn hội nghị cũng bị detector phát hiện là một gương mặt người. Còn đối với Insightface như trong hình 3.5, tuy không thể phát hiện ra được hết những gương mặt ở phía sau, nhưng phương pháp này vẫn đảm bảo được việc phát hiện hai nhân vật chính là Chủ tịch Trung Quốc Tập Cận Bình và Tổng thống Nga Vladimir Putin, hơn nữa, kết quả của Insightface cũng không bị nhầm góc bàn hội nghị thành gương mặt người.



HÌNH 3.5: Kết quả khi dùng Insightface

Trong bảng 3.1 trên, số gương mặt của nhân vật khác sẽ là những nhân vật không nằm trong số 8 nguyên thủ quốc gia đã nêu ở chương 1, và ảnh không

BÀNG 3.1: So sánh các kết quả khi dùng MTCNN và Insightface

	MTCNN	Insightface
Tổng số gương mặt phát hiện được	474	424
Số gương mặt của các nhân vật khác	259	231
Ảnh không phải là gương mặt	54	32

phải là gương mặt sẽ là những ảnh giống như góc bàn hội nghị trong hình 3.4 đã đề cập.

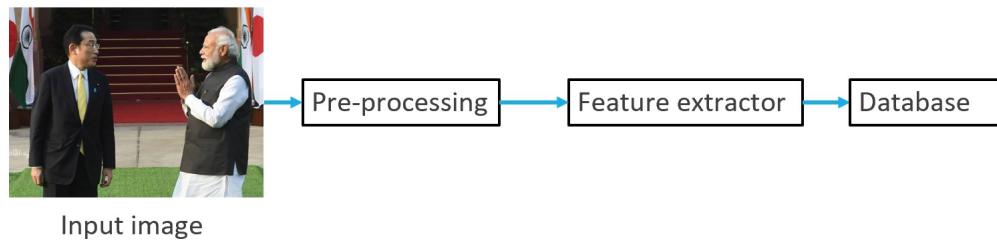
Từ thử nghiệm trên một tập dữ liệu nhỏ như vậy, em có thể tạm đưa ra được một kết luận rằng sử dụng Insightface sẽ hiệu quả hơn MTCNN trong bài toán của khoá luận đang muôn giải quyết. Nguyên nhân là vì Insightface đảm bảo được sẽ phát hiện được tất cả những ảnh của 8 nguyên thủ đã lựa chọn, không bị nhầm lẫn những họa tiết, hoa văn,... thành gương mặt người. Bài toán trong khoá luận đang hướng đến việc xây dựng hệ thống với độ chính xác cao nhất có thể.

Vì thế dẫn đến kết luận là sẽ sử dụng Insightface để làm face detector. Kéo theo đó, khi sử dụng Insightface làm phương pháp face detector, việc trích xuất đặc trưng gương mặt cũng sẽ được hỗ trợ rất dễ dàng. Nên phương pháp thực hiện cho bài toán này trong chương 4 phần Nhận diện gương mặt gồm các phương pháp:

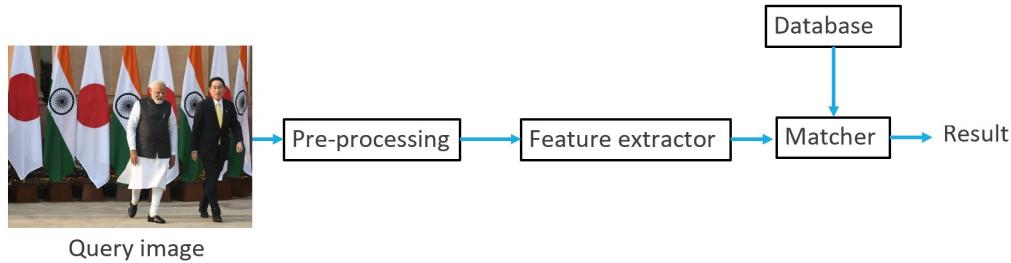
- Face detector: Insightface
- Feature extractor: Insightface

## 3.2 Nhận diện sự kiện tương ứng

Cũng tương tự với bài toán nhận diện gương mặt người nổi tiếng, phần nhận diện sự kiện này cũng được chia thành hai quá trình: quá trình huấn luyện mô hình - hình 3.6 và quá trình nhận diện sự kiện - hình 3.7.



HÌNH 3.6: Quá trình huấn luyện mô hình nhận diện sự kiện



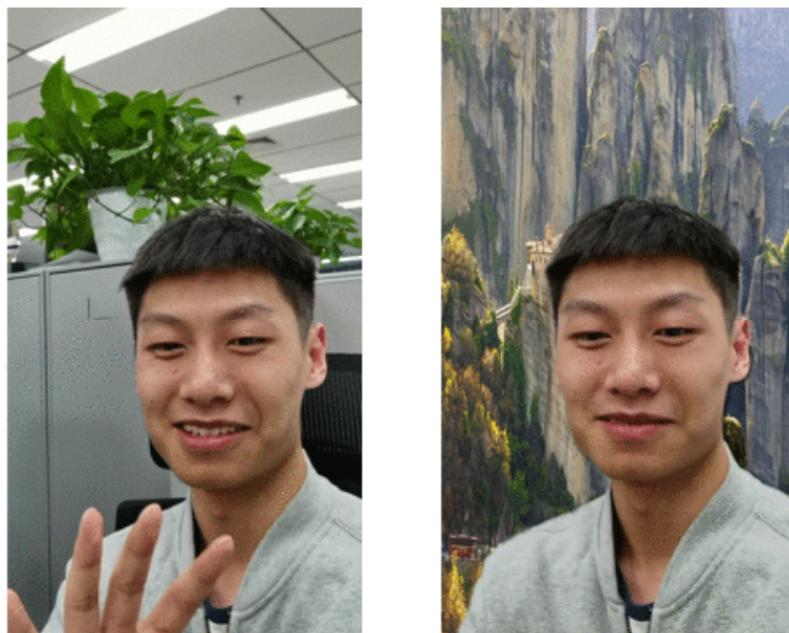
HÌNH 3.7: Quá trình nhận diện sự kiện

### 3.2.1 Tiền xử lý dữ liệu - Image matting

Từ nhu cầu của việc nhận diện sự kiện trong bài toán này, việc phân tách yếu tố cảnh nền trong ảnh ra khỏi nhân vật sẽ là một hướng tiếp cận có tiềm năng. Sử dụng một công cụ xử lý hình ảnh PaddlePaddle [17] của một nhóm kỹ sư người Trung Quốc xây dựng để phân tách nền và nhân vật là một cách xử lý hiệu quả.

Image matting là một kỹ thuật được sử dụng nhiều trong việc nền công nghiệp phim ảnh, nó được ứng dụng trong việc tách người và phông xanh trong quay phim để xây dựng lên các hiệu ứng đặc biệt [18].

Tuy nhiên, khi nhu cầu làm việc tại nhà trong thời kỳ dịch bệnh, ứng dụng của image matting được sử dụng một cách rộng rãi hơn: sử dụng tách nền để thay đổi cảnh nền đang có khi chụp hình, tham dự buổi họp,...



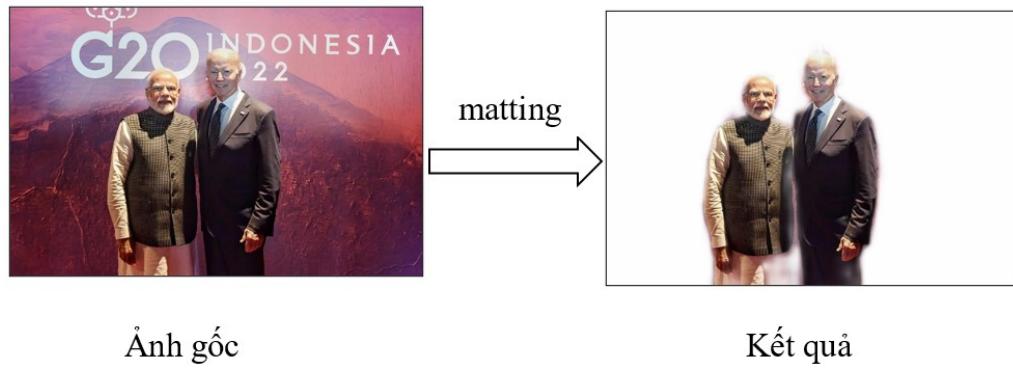
---

HÌNH 3.8: Hình ảnh minh họa của PaddlePaddle

Bên trái là hình của một người chụp trong văn phòng, còn bên phải là kết quả khi đã tách người ra nền văn phòng, và thay bằng một hình ảnh nền khác.

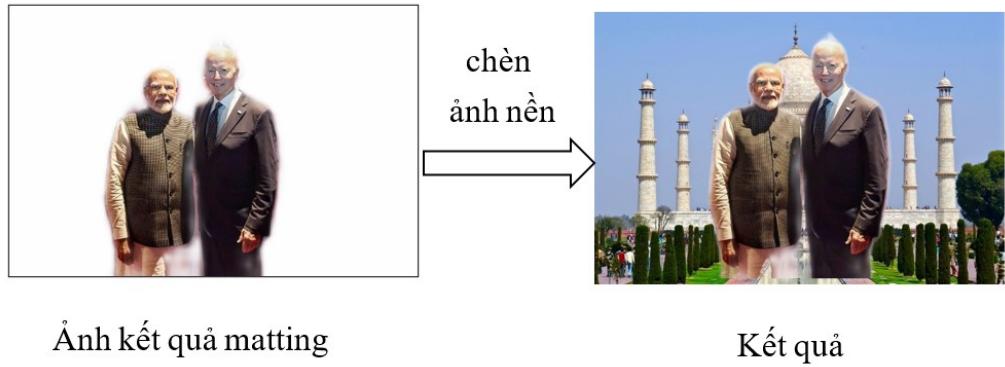
Thực chất, bài toán image matting được cài đặt trong thư viện PaddlePaddle sẽ thực hiện theo quy trình như sau:

- Bước 1: tách ảnh của người và ảnh nền.



HÌNH 3.9: Quy trình matting - Tách ảnh người và nền

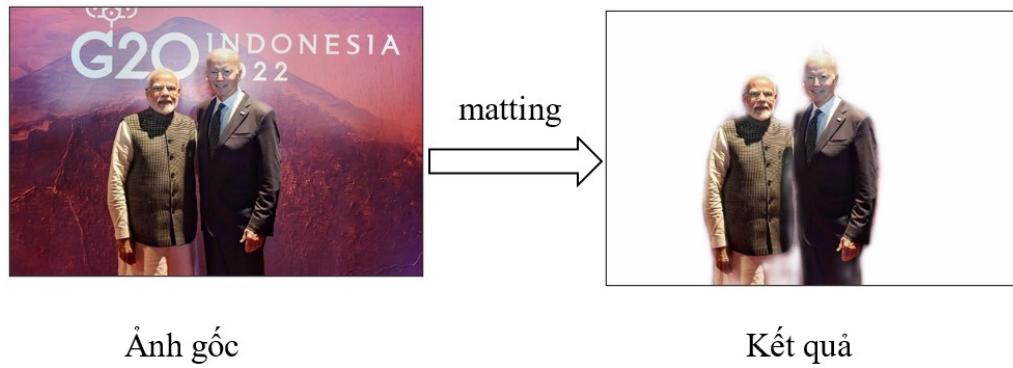
- Bước 2: chèn kết quả đã thu được lên một ảnh nền khác [19].



HÌNH 3.10: Quy trình matting - Chèn ảnh nền khác

Dựa trên kết quả của những nghiên cứu đã có, em đã điều chỉnh một phần để có thể ứng dụng vào bài toán mà khoá luận đang hướng đến như sau:

- Bước 1: tách ảnh của người và ảnh nền.



HÌNH 3.11: Quy trình trích xuất ảnh nền - Tách ảnh người và nền

- Bước 2: sử dụng Ảnh gốc ban đầu ở hình 3.11 và loại bỏ đi phần Kết quả ở hình 3.11.



HÌNH 3.12: Loại bỏ ảnh kết quả matting đã thu được

Có thể mô tả quá trình tách ảnh nền thông qua biểu diễn các ma trận như sau  
Ma trận ảnh gốc  $A_{raw}$

$$A_{raw} = \begin{bmatrix} [1, 2, 3] & [4, 5, 6] & [7, 8, 9] \\ [4, 5, 6] & [7, 8, 9] & [1, 2, 3] \\ [7, 8, 9] & [1, 2, 3] & [4, 5, 6] \end{bmatrix}$$

Ma trận kết quả sau khi matting  $A_{result}$

$$A_{result} = \begin{bmatrix} [255, 255, 255] & [255, 255, 255] & [255, 255, 255] \\ [255, 255, 255] & [7, 8, 9] & [255, 255, 255] \\ [255, 255, 255] & [1, 2, 3] & [255, 255, 255] \end{bmatrix}$$

Vì ảnh kết quả sau khi matting chỉ giữ lại những vùng có người trong ảnh, còn phần nền đều chuyển thành màu trắng, nên giá trị các điểm ảnh trong ảnh kết quả không có người đều có giá trị [255, 255, 255].

Từ hai ma trận trên, ta có thể thu được một ma trận chỉ chứa các điểm ảnh nền  $A_{background}$  theo công thức sau:

$$A_{background} = \begin{cases} A_{raw}, & \text{if } A_{raw} \neq A_{result} \\ [0, 0, 0] & \end{cases} \quad (3.1)$$

Công thức trên được diễn đạt lại như sau:

- Nếu một điểm ảnh ở ma trận ảnh gốc  $A_{raw}$  khác với ma trận ảnh kết quả sau khi matting  $A_{result}$  có nghĩa là điểm ảnh đó đã thay đổi giá trị sau khi trải qua quá trình matting  $\Rightarrow$  Điểm ảnh đó là ảnh nền, cần lấy giá trị của ảnh nền ban đầu để gán cho điểm ảnh đó trong ma trận  $A_{background}$ .
- Trong trường hợp còn lại, có nghĩa là giá trị của điểm ảnh trước và sau khi matting không thay đổi,  $A_{raw} = A_{result} \Rightarrow$  Điểm ảnh đó là điểm ảnh của người trong ảnh, cần phải tách đi, nên giá trị của điểm ảnh sẽ được gán thành giá trị của màu đen [0, 0, 0].

Như vậy, dựa vào 2 ma trận đã đề cập ở trên, ma trận chỉ chứa các điểm ảnh nền  $A_{background}$  sẽ có giá trị như sau:

$$A_{background} = \begin{bmatrix} [1, 2, 3] & [4, 5, 6] & [7, 8, 9] \\ [4, 5, 6] & [0, 0, 0] & [1, 2, 3] \\ [7, 8, 9] & [0, 0, 0] & [4, 5, 6] \end{bmatrix}$$

Thực hiện theo phương pháp trên, ảnh gốc ban đầu được biến đổi lại như hình 3.13 dưới đây.



HÌNH 3.13: Ảnh nền trích xuất được

### 3.2.2 Thông tin từ ngữ cảnh

Sử dụng các mạng học sâu để học được thông tin của hình ảnh (content-based), từ đó rút trích thông tin sự kiện.

- Restnet-18. Hình ảnh được trích xuất thành một vector đặc trưng với số chiều là 512.

- MobileNetV2. Hình ảnh được trích xuất thành một vector đặc trưng với số chiều là (5,5,1280).

### 3.2.3 Thông tin từ màu sắc

#### Histogram

Sử dụng đặc trưng thủ công histogram để làm một đặc trưng nhận diện sự kiện người nổi tiếng đang tham gia. Trong khoá luận này, các hướng sử dụng histogram chính gồm:

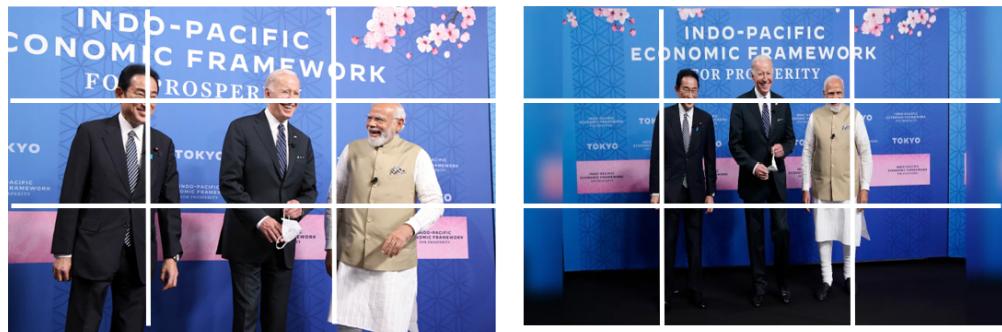
- Bin [8, 8, 8]. Với mỗi kênh màu RGB, giá trị của mỗi pixel sẽ dao động trong khoảng 0 - 255. Với phương pháp này, mỗi kênh màu sẽ được chia thành 8 phần bằng nhau 0-31, 32-63, 63-95,... sau đó lấy giá trị màu đầu tiên của mỗi phần, ta thu được các giá trị gồm [0, 32, 64, 96, 128, 160, 192, 224], và tạo nên sự kết hợp của các kênh màu. Nên vector đặc trưng của phương pháp này sẽ có số chiều gồm  $8 \times 8 \times 8 = 512$ .
- Đặc trưng histogram, xét sự kết hợp giữa 3 kênh màu chủ đạo, số chiều của vector đặc trưng này =  $255 \times 3 = 765$
- Đặc trưng chia vùng, xét sự kết hợp giữa 3 kênh màu chủ đạo trong hình ảnh được chia thành từng vùng nhỏ  $3 \times 3$  hoặc  $4 \times 4$ . Nguyên nhân dẫn đến việc chia vùng để khảo sát sự phân bố của các vùng khác nhau của các bức ảnh, với cảnh nền giống nhau thì trong các bức ảnh khác nhau, những vùng tương ứng nhau sẽ có màu tương tự nhau. Giả sử mỗi bức hình sẽ được

1	2	3
4	5	6
7	8	9

---

HÌNH 3.14: Chia hình thành  $3 * 3$  vùng

chia thành 9 vùng được đánh số như hình 3.14 thì trong ví dụ như hình 3.15, vùng 1, 2, 3 của hai bức hình sẽ có histogram tương tự nhau. Từ đó học được thông tin về sự kiện.



HÌNH 3.15: So sánh vùng ảnh giữa 2 ảnh trong 1 sự kiện

### Màu chủ đạo

Mỗi bức hình đều có thể rút trích ra bảng màu chính để làm đặc trưng, trong bài toán này, em sẽ áp dụng số màu chính là 5, và mỗi màu đặc trưng bởi 3 kênh màu khác nhau, số chiều vector đặc trưng thu được là  $3 * 5 = 15$ .



HÌNH 3.16: Một minh họa cụ thể về màu chủ đạo

### 3.3 Mô hình máy học sử dụng

Về phần mô hình máy học sẽ sử dụng trong cả hai bài toán nhận diện gương mặt và nhận diện sự kiện, trong khoá luận này em sẽ lựa chọn mô hình máy học KNN và khảo sát k trong khoảng từ 1 đến 5 để lựa chọn ra số k tốt nhất.

Nguyên nhân dẫn đến lựa chọn như vậy là vì bộ dữ liệu tư thu thập mà em xây dựng có số lượng khá ít, đối với các phương pháp máy học khác, nếu bộ dữ liệu có số lượng quá ít thì sẽ không thể huấn luyện mô hình một cách hiệu quả. Đối với KNN, số lượng dữ liệu của cả bộ dữ liệu hoặc của riêng một lớp vượt trội hơn những lớp khác sẽ không phải là vấn đề mà phương pháp này không giải quyết được.

## Chương 4

# THỰC NGHIỆM VÀ ĐÁNH GIÁ

### 4.1 Môi trường và ngôn ngữ cài đặt

Bài toán trong khoá luận được tiến hành cài đặt mô hình bằng ngôn ngữ Python3, cùng với các thư viện bổ trợ sau: opencv, numpy, scikit-learn, sklearn, pandas. Để chạy được trên GPU, cần cài đặt thêm thư viện onnxruntime-gpu. Các thư viện phụ trợ cho từng mô hình gồm:

- Phát hiện gương mặt trong hình ảnh và trích xuất đặc trưng gương mặt: insightface.
- Nhận diện sự kiện: github repository PaddlePaddle, img2vecpytorch, Pillow, scipy, tensorflow.
- Xây dựng ứng dụng: sử dụng github và thư viện streamlit - tự tạo một server để các máy khác có thể truy cập ứng dụng, chứ không chỉ riêng máy tính cài đặt; pickle - dùng để lưu lại những mô hình máy học để sử dụng trong ứng dụng.

Môi trường thực hiện chính của bài toán sẽ dùng Google Colab với các thông số như sau:

- RAM: 12.68GB
- Disk: 107.72GB

## 4.2 Tập dữ liệu thực nghiệm

Trong quá trình thực hiện khoá luận, em đã thu thập được 669 bức ảnh khác nhau, của 8 nguyên thủ quốc gia trong 10 sự kiện khác nhau như đã nêu trong chương 1.

### 4.2.1 Tăng cường dữ liệu

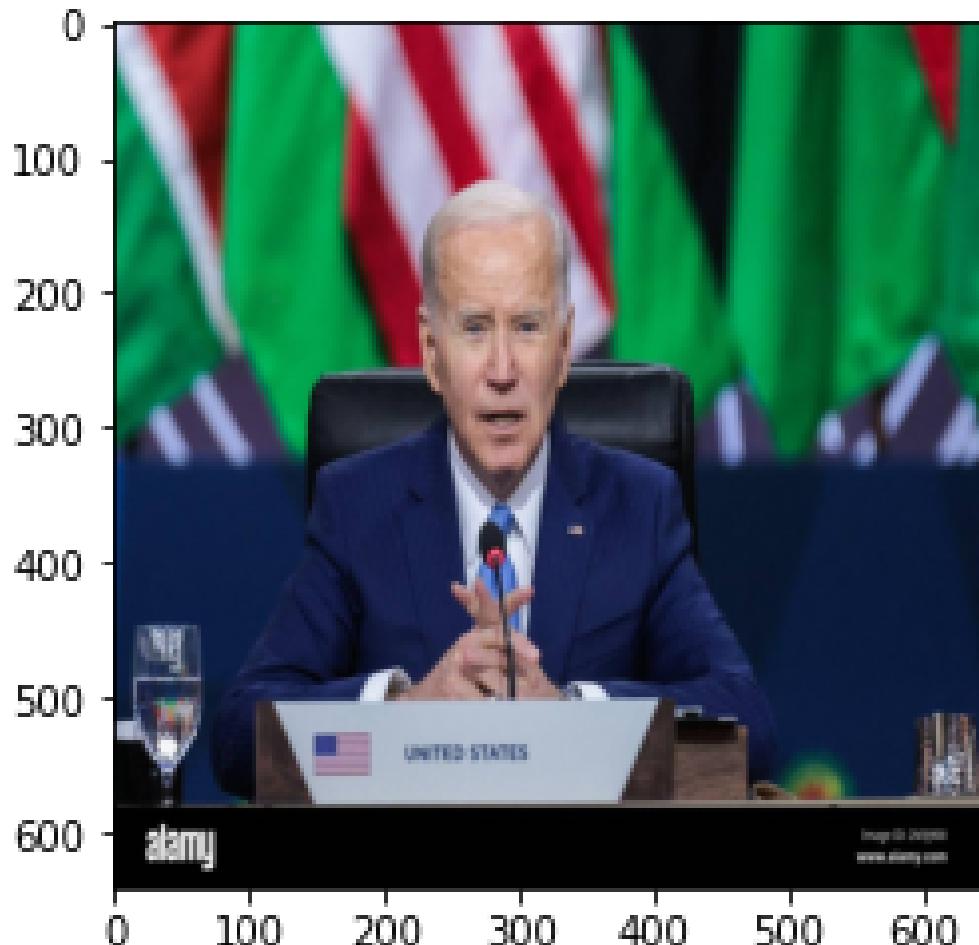
Để tăng tính đa dạng của dữ liệu, em tiến hành tăng cường dữ liệu để có thể thu được những góc độ ảnh gương mặt khác nhau, việc này sẽ tạo cho dữ liệu có tính thực tế hơn. Trong một sự kiện thì có thể sẽ có nhiều góc chụp ảnh, góc độ thay đổi sẽ khiến ảnh thu được khác nhau, việc thay đổi kích thước ban đầu của ảnh để đăng bài báo cũng sẽ khiến hình ảnh bị biến đổi,...

Với một hình gốc ban đầu như hình 4.1, em sẽ thực hiện 3 phương pháp tăng cường dữ liệu khác nhau.



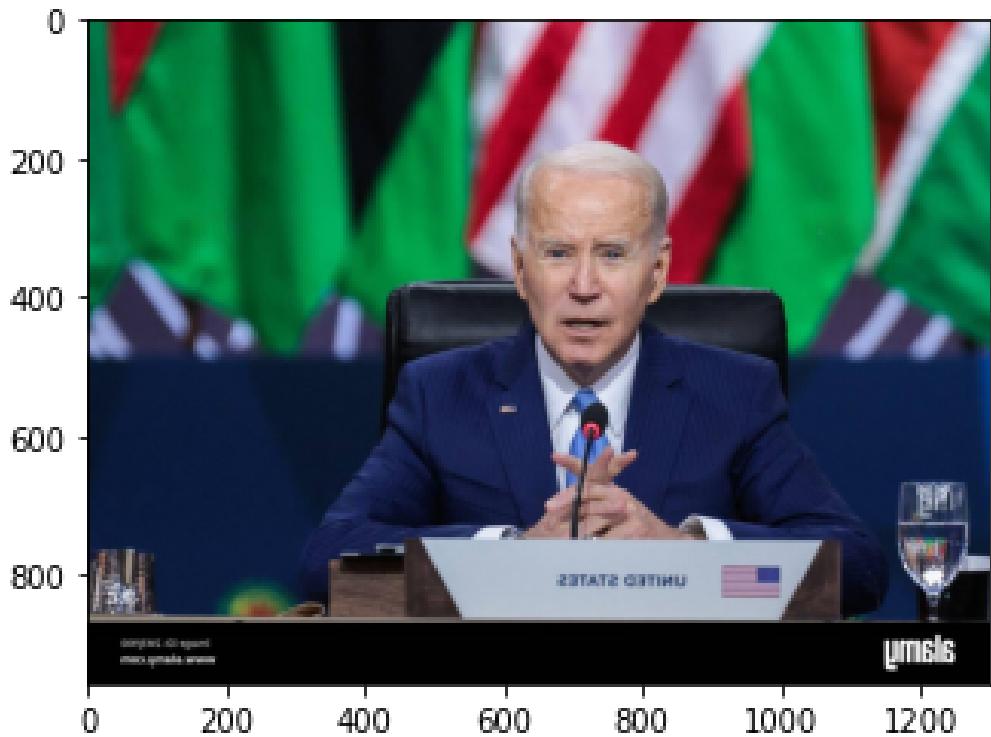
HÌNH 4.1: Ảnh gốc ban đầu - trước khi thực hiện tăng cường dữ liệu

- Thực hiện resize ảnh về 640x640 - Mô phỏng việc ảnh sẽ có thể bị thay đổi kích thước trong quá trình đăng ảnh lên bài báo.



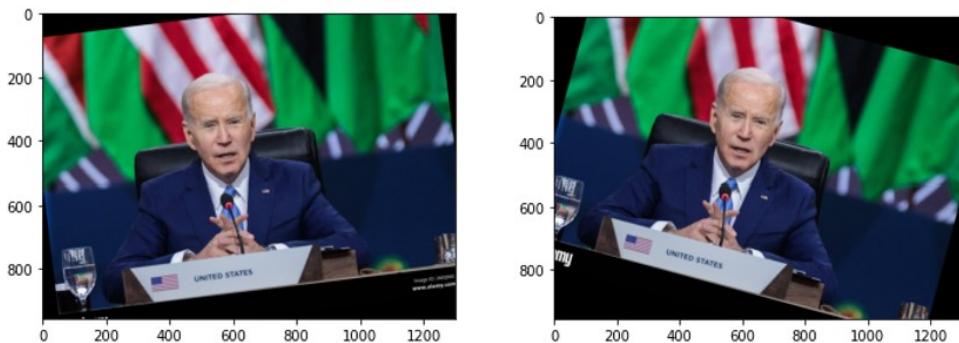
HÌNH 4.2: Ảnh đã được tỉ lệ lại 640x640

- Thực hiện lật ảnh theo chiều ngang, mô phỏng việc thiết bị chụp ảnh có thể lật hình lại, tạo tính đa dạng cho các chi tiết trong hình.



HÌNH 4.3: Ảnh đã được lật theo chiều ngang

- Thực hiện shear ảnh, với độ shear là ngẫu nhiên trong khoảng từ (-25°, 25°). Với mỗi ảnh gốc sẽ lấy 2 phiên bản ngẫu nhiên của bước shear ảnh.



HÌNH 4.4: Hai kết quả minh họa cho bước shear ảnh ban đầu

Như vậy, với mỗi ảnh trong bộ dữ liệu sẽ có 5 phiên bản khác nhau: ảnh gốc, ảnh resize, ảnh lật ngang, 2 ảnh shear ngẫu nhiên.

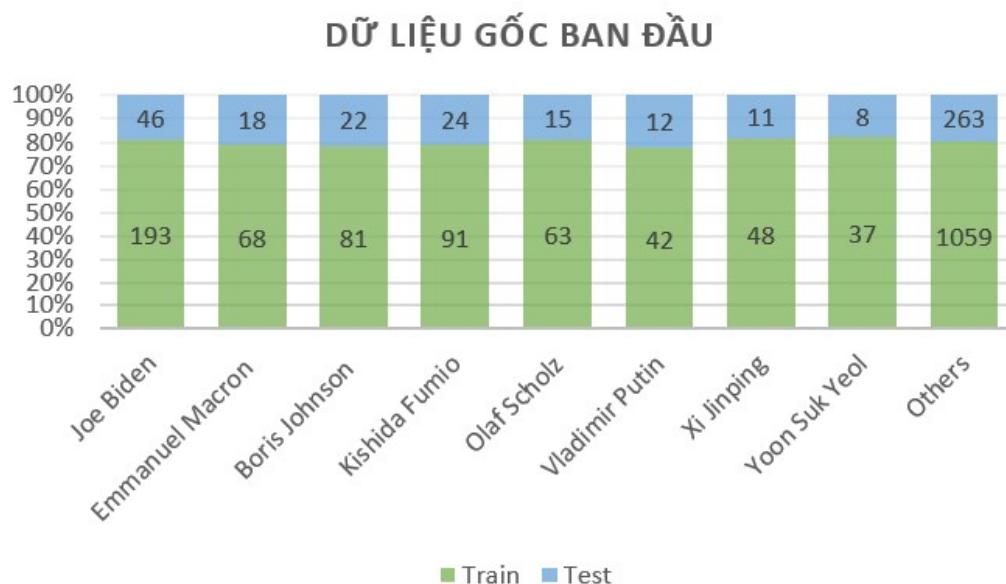
Sau đó em sẽ thực nghiệm 2 lần trên 2 bộ dữ liệu:

- Bộ dữ liệu chỉ có ảnh gốc thu thập được
- Bộ dữ liệu đã qua tăng cường - augmented data

#### 4.2.2 Nhận diện gương mặt người nổi tiếng

Sau khi thực hiện phát hiện gương mặt trên 669 ảnh ban đầu, kết quả thu được 2101 ảnh gương mặt người thuộc 9 lớp khác biệt (8 người nổi tiếng đã chọn trong chương 1 và một lớp còn lại thuộc về các nhân vật khác xuất hiện trong ảnh). Việc xây dựng bộ dữ liệu có 9 lớp như vậy sẽ khiến cho kết quả nhận diện gương mặt trực quan hơn, nếu trong một bức ảnh đầu vào không có người thuộc 8 lớp đã chọn thì kết quả sẽ trả về là không phát hiện người nổi tiếng nào trong hình, và không gây ảnh hưởng đến phần nhận diện sự kiện.

Như trong hình 4.5, phân bố tỉ lệ của các lớp không đều nhau. Nếu chỉ xét đến 8 nhân vật đã chọn thì số ảnh của Tổng thống Mỹ Joe Biden đang nhiều hơn hẳn so với các lớp còn lại, nguyên nhân là do lượng tin tức về Tổng thống Mỹ chiếm đa số và Tổng thống Mỹ cũng tham dự nhiều sự kiện được chọn trong bảng 1.2



HÌNH 4.5: Phân bổ dữ liệu để huấn luyện và thử nghiệm trong mô hình nhận dạng gương mặt

hơn các nguyên thủ quốc gia khác. Lớp Others thì có số lượng vượt trội hơn 8 lớp còn lại, vì khi có một sự kiện hội họp nào đó, các nguyên thủ quốc gia sẽ hội tụ lại rất đông, từ đó các gương mặt không thuộc 8 lớp đang hướng đến trong khoá luận là rất nhiều. Đây cũng là một lý do cung cấp thêm lựa chọn sử dụng mô hình KNN - sự chênh lệch giữa các lớp có thể được giải quyết.

Và kết quả của mô hình được thể hiện như trong hình 4.6

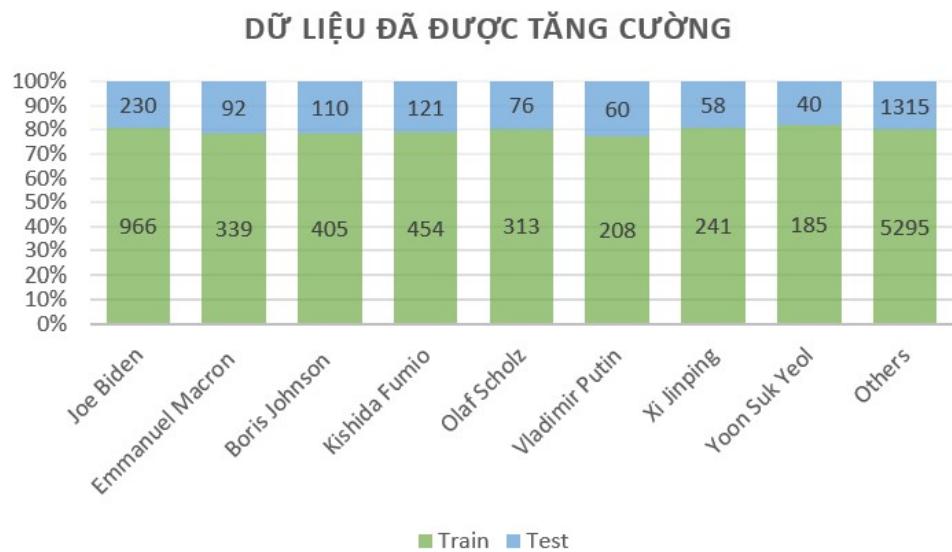
Mô hình có kết quả với độ chính xác xấp xỉ 1.0, với số k tốt nhất khảo sát được trên tập dữ liệu này = 2. Phân tích kỹ hơn những điểm chưa đạt được hiệu quả tối đa, nguyên nhân rất có thể là do sự chênh lệch của tập dữ liệu, như trong hình 4.5 đã có thể hiện, lớp 0 của kết quả ứng với lớp Tổng thống Mỹ Joe Biden, số lượng ảnh để nhận diện là nhiều hơn so với 7 nhân vật còn lại (46 ảnh), việc nhận diện nhiều ảnh sẽ dễ dẫn đến việc giảm độ chính xác xuống. Còn lớp 7 của ảnh kết quả tương ứng với Tổng thống Hàn Quốc Yoon Suk Yeol, precision của lớp này bị giảm xuống do ảnh dùng để huấn luyện cho lớp này ít hơn hẳn những lớp còn lại, nhưng hiệu quả của mô hình vẫn rất chính xác với recall = 1.0.

	precision	recall	f1-score
0	0.99	0.99	0.99
1	1.00	1.00	1.00
2	1.00	1.00	1.00
3	1.00	1.00	1.00
4	1.00	1.00	1.00
5	1.00	1.00	1.00
6	1.00	1.00	1.00
7	0.98	1.00	0.99
8	1.00	1.00	1.00
accuracy			1.00
macro avg	1.00	1.00	1.00
weighted avg	1.00	1.00	1.00

HÌNH 4.6: Kết quả nhận diện gương mặt với tập dữ liệu gốc ban đầu

Sau đó thực hiện nhận diện gương mặt trên bộ dữ liệu đã được tăng cường, với tổng cộng 10508 gương mặt đã được phát hiện, và phân bố như trong hình 4.7.

Phân bố dữ liệu sau khi đã trải qua tăng cường dữ liệu vẫn giữ tỉ lệ như cũ, nên lớp 0 của Tổng thống Mỹ vẫn là lớp có số ảnh test nhiều nhất và lớp 7 của Tổng thống Hàn Quốc vẫn sẽ là lớp có số ảnh được huấn luyện ít nhất. Tuy nhiên, kết quả hiện tại của tất cả các lớp đã được nâng lên mức tối đa = 1.0 – 100%. Từ đó có thể kết luận được rằng việc tăng cường dữ liệu trong bài toán này đã góp phần cải thiện được độ chính xác của dữ liệu, ngoài ra thì số k của mô hình KNN khi dùng dữ liệu đã tăng cường = 1, điều đó cũng góp phần trong việc dễ xác định được vector đặc trưng tương đương dễ hơn.



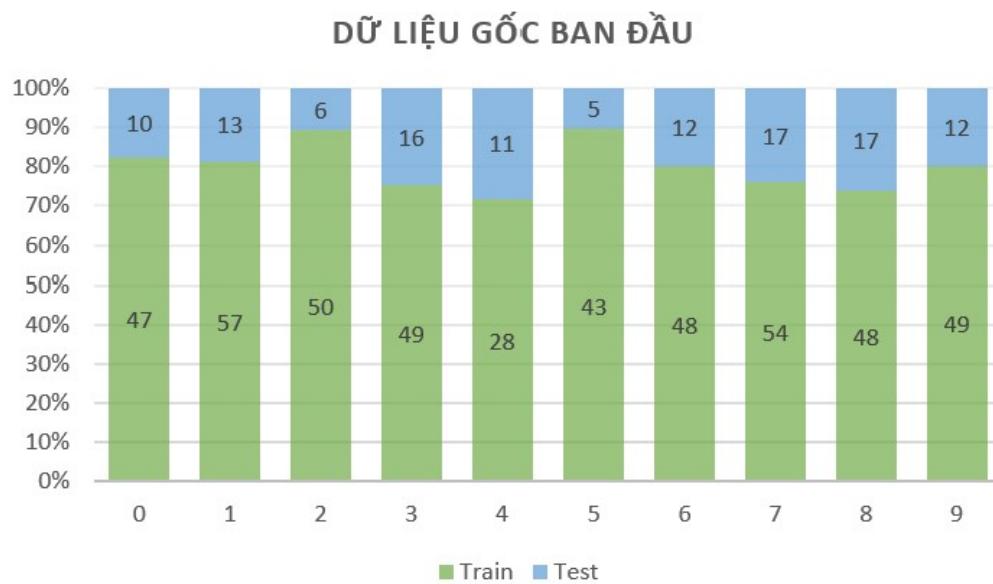
HÌNH 4.7: Phân bố tập dữ liệu đã qua tăng cường

	precision	recall	f1-score	support
0	1.00	1.00	1.00	233
1	1.00	1.00	1.00	82
2	1.00	1.00	1.00	115
3	1.00	1.00	1.00	108
4	1.00	1.00	1.00	79
5	1.00	1.00	1.00	64
6	1.00	1.00	1.00	48
7	1.00	1.00	1.00	51
accuracy			1.00	780
macro avg	1.00	1.00	1.00	780
weighted avg	1.00	1.00	1.00	780

HÌNH 4.8: Kết quả nhận diện gương mặt trên tập dữ liệu đã qua tăng cường

### 4.2.3 Nhận diện sự kiện tương ứng

Trong 669 hình đã thu thập được, có 592 ảnh là thuộc về 10 sự kiện đã đã đề cập, số ảnh còn lại là của những sự kiện rời rạc khác nhau, thu thập vì để làm bộ dữ liệu cho phần nhận diện gương mặt nên trong phần nhận diện sự kiện sẽ không dùng đến.



HÌNH 4.9: Phân bổ số lượng dữ liệu trong tập các sự kiện ban đầu

Ngoài ra, với mỗi lớp sự kiện, sẽ được đánh giá trên hai loại ảnh khác nhau: ảnh gốc - có cả nhân vật và phông nền, và ảnh nền - sử dụng matting để loại bỏ nhân vật như trong chương 3 đã nói. Từ đó nhận xét được mức độ hiệu quả trong việc trích xuất đặc trưng trên ảnh.

Đối với nhóm các phương pháp sử dụng thông tin về màu sắc, kết quả thực hiện được thể hiện trong bảng 4.3.

Từ kết quả trên có thể đưa ra được những kết luận như sau:

- Khi sử dụng histogram để làm vector mô tả đặc trưng sự kiện, nên dùng ảnh gốc ban đầu để thực hiện rút trích.

BẢNG 4.1: Kết quả nhận diện sự kiện trên bộ dữ liệu không qua tăng cường

Tên phương pháp	Ảnh sử dụng	Accuracy
Histogram bin	Ảnh gốc	0.8655
Histogram bin	Ảnh nền	0.7815
Histogram chia vùng 3x3	Ảnh gốc	0.8991
Histogram chia vùng 3x3	Ảnh nền	0.7983
Histogram chia vùng 4x4	Ảnh gốc	0.8739
Histogram chia vùng 4x4	Ảnh nền	0.7899
Histogram RGB	Ảnh nền	0.7478
Dominant color	Ảnh gốc	0.3277
Dominant color	Ảnh nền	0.3781

- Trong bộ dữ liệu hiện tại, việc chia vùng thành 3x3 sẽ có hiệu quả cao hơn là 4x4. Chỉ cần chia ảnh thành 9 vùng riêng biệt thì thông tin đặc trưng màu sắc của từng vùng cũng đủ để đại diện cho vùng ảnh đó.
- Khi sử dụng dominant color, sử dụng ảnh nền sẽ đạt được hiệu quả tốt hơn do đã loại bỏ đi màu đen thê vào, cũng loại bỏ những màu sắc như trang phục của người tham dự sự kiện. Sử dụng màu chủ đạo sẽ hoàn toàn phụ thuộc vào sự tương đồng màu sắc giữa các cảnh nền trong sự kiện.
- Kết quả sử dụng màu chủ đạo làm vector đặc trưng cho sự kiện thấp hơn hẳn so với sử dụng histogram, có thể lý giải kết quả bằng lý do: Các sự kiện giống nhau sẽ không hẳn có cùng cảnh nền với nhau, từ đó dẫn đến việc màu chủ đạo trong cảnh nền cũng sẽ khác biệt nhau. Ví dụ như trong hình 4.10.



HÌNH 4.10: Sự khác biệt của màu chủ đạo trong cùng 1 sự kiện -  
Chuyến thăm Kyiv của Thủ tướng Anh

Đối với nhóm các phương pháp sử dụng thông tin ngữ cảnh, sử dụng mạng học sâu để rút trích đặc trưng, kết quả thực hiện được thể hiện trong bảng 4.4.

BẢNG 4.2: Kết quả nhận diện sự kiện trên bộ dữ liệu không qua tăng cường - sử dụng mạng học sâu

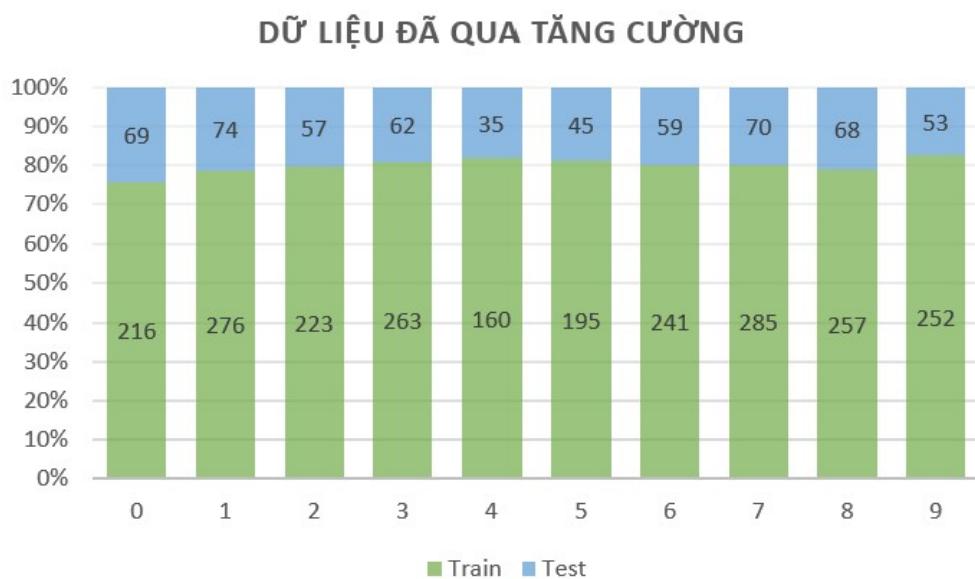
Tên phương pháp	Ảnh sử dụng	Accuracy
Resnet	Ảnh gốc	0.9327
Resnet	Ảnh nền	0.8823
MobileNet	Ảnh gốc	0.9075
MobileNet	Ảnh nền	0.8655

Từ kết quả trên ta có thể rút ra được kết luận: Khi sử dụng mạng học sâu để rút trích đặc trưng, sử dụng ảnh gốc ban đầu sẽ đạt được hiệu quả tốt hơn. Ảnh nền sau khi đã trải qua xử lý, sẽ có nhiều vết đen do phần ảnh của người đã bị thay thế đi, từ đó các mạng học sâu không thể học được thông tin gì nhiều như ảnh gốc ban đầu.

Ngoài ra thì các phương pháp sử dụng mạng học sâu đạt được kết quả tốt hơn so với các phương pháp sử dụng đặc trưng thủ công như histogram, màu chủ đạo.

Vì muốn cải thiện độ chính xác của các phương pháp, em sẽ thực hiện các phương pháp đó trên bộ dữ liệu đã được tăng cường để so sánh kết quả thu

được. Với 592 ảnh ban đầu, sau khi trải qua tăng cường dữ liệu sẽ thu được 2960 ảnh. Phân bổ của bộ dữ liệu nhận diện sự kiện đã qua tăng cường được minh họa như trong hình 4.11.



HÌNH 4.11: Phân bổ dữ liệu trong tập nhận diện sự kiện đã qua tăng cường

Đối với nhóm các phương pháp sử dụng thông tin về màu sắc, kết quả thực hiện được thể hiện trong bảng 4.3.

Sau khi thực hiện trên bộ dữ liệu đã qua tăng cường, những kết luận rút ra được cũng tương tự như bộ dữ liệu gốc ban đầu:

- Đối với histogram thì dùng ảnh gốc sẽ đạt kết quả cao hơn.
- Histogram chia vùng 3x3 hiệu quả hơn so với chia vùng 4x4.
- Khi dùng màu chủ đạo làm vector đặc trưng thì dùng ảnh nền sẽ đạt kết quả tốt hơn.
- Và một kết luận tổng kết hơn là tăng cường dữ liệu sẽ giúp cải thiện độ chính xác của các phương pháp rút trích đặc trưng.

BẢNG 4.3: Kết quả nhận diện sự kiện trên bộ dữ liệu không qua tăng cường

Tên phương pháp	Ảnh sử dụng	Accuracy
Histogram bin	Ảnh gốc	1.0000
Histogram bin	Ảnh nền	0.9881
Histogram chia vùng 3x3	Ảnh gốc	0.9983
Histogram chia vùng 3x3	Ảnh nền	0.9763
Histogram chia vùng 4x4	Ảnh gốc	0.9932
Histogram chia vùng 4x4	Ảnh nền	0.9611
Histogram RGB	Ảnh nền	0.9577
Dominant color	Ảnh gốc	0.4155
Dominant color	Ảnh nền	0.4543

Đối với nhóm các phương pháp sử dụng thông tin ngữ cảnh, sử dụng mạng học sâu để rút trích đặc trưng, kết quả thực hiện được thể hiện trong bảng 4.4.

Kết quả đạt được của các phương pháp rút trích đặc trưng sử dụng mạng học sâu đã được cải thiện đáng kể, cả 2 phương pháp Resnet và MobileNet đều đã đạt độ chính xác 100% khi dùng ảnh gốc.

BẢNG 4.4: Kết quả nhận diện sự kiện trên bộ dữ liệu đã qua tăng cường - sử dụng mạng học sâu

Tên phương pháp	Ảnh sử dụng	Accuracy
Resnet	Ảnh gốc	1.0000
Resnet	Ảnh nền	0.9864
MobileNet	Ảnh gốc	1.0000
MobileNet	Ảnh nền	0.9898

### 4.3 Xây dựng ứng dụng thực tế

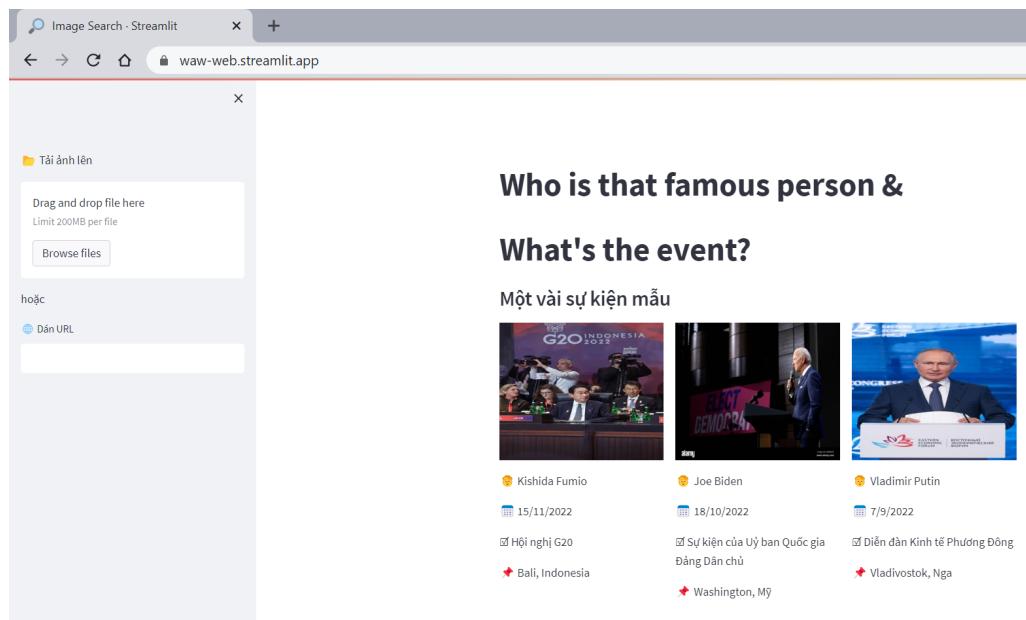
Sử dụng những phương pháp trích xuất tốt nhất để xây dựng lên một hệ thống nhận diện người nổi tiếng và sự kiện tương ứng. Từ những kết luận của phần thực nghiệm, em sẽ lựa chọn những phương pháp để xây dựng hệ thống gồm:

- Phát hiện gương mặt: detector Insightface.
- Nhận diện gương mặt: phương pháp rút trích đặc trưng Insightface.
- Nhận diện sự kiện: dùng histogram bin - không tách nền.
- Sử dụng bộ dữ liệu đã qua tăng cường để huấn luyện mô hình

Đường link truy cập vào trang web: bit.ly/waw-website

Giao diện của trang web được xây dựng như hình minh họa 4.12.

Ở khung bên trái, người dùng có thể chọn tải ảnh lên hoặc dán đường dẫn hình ảnh để hệ thống nhận diện và trả về kết quả tương ứng.



HÌNH 4.12: Giao diện trang web sản phẩm

Kết quả trả về không chỉ nhận diện người nổi tiếng và sự kiện người đó đang tham gia, mà có có bounding box và nhãn chú thích để chú thích gương mặt được phát hiện ứng với nhân vật nào trong bộ dữ liệu. Nếu gương mặt người đó không nằm trong nhóm 8 nguyên thủ quốc gia đã đề cập thì sẽ không vẽ bounding box bao quanh.

Mô phỏng kết quả khi chạy thử hệ thống được thể hiện như sau:

## Who is that famous person & What's the event?

👤 0: Kishida Fumio

📅 15/11/2022

☑ Hội nghị G20

📍 Bali, Indonesia



HÌNH 4.13: Kết quả trả về của trang web

## Chương 5

# KẾT LUẬN

### 5.1 Kết luận

Trong khoá luận này, bài toán được hướng đến để giải quyết chính là Nhận dạng người nổi tiếng và sự kiện tương ứng. Trong quá trình thực hiện khoá luận em đã thử các phương pháp khác nhau để rút trích đặc trưng một các hiệu quả nhất. Sau khi thử nghiệm trên bộ dữ liệu tự thu thập, các phương pháp rút trích đặc trưng có kết quả tốt nhất là:

- Phát hiện gương mặt trong hình ảnh: Sử dụng Insightface
- Trích xuất đặc trưng gương mặt: Sử dụng Insightface
- Trích xuất đặc trưng sự kiện: Đôi với thông tin ngữ cảnh: ResNet, MobileNet dùng ảnh gốc ban đầu Đôi với thông tin màu sắc: dùng histogram phân vùng  $3 * 3$  với ảnh gốc ban đầu. Đôi với màu chủ đạo: dùng ảnh đã tách nền.
- Mô hình máy học sử dụng: KNN.

### 5.2 Hướng phát triển

Có thể thử thêm nhiều phương pháp rút trích đặc trưng khác để tăng độ hiệu quả. Hoặc cũng có thể kết hợp những đặc trưng hiện có lại thành một đặc trưng mới, dùng nó để làm đặc trưng nhận diện.

### 5.3 Hạn chế của khoá luận

Khoá luận này vẫn còn một số hạn chế như:

- Không thể nhận diện được người ở các thời điểm khác nhau. Vì mô hình nhận diện gương mặt là riêng biệt với nhận diện sự kiện, nên nếu cùng một người nổi tiếng tham dự cùng 1 sự kiện nhưng ở 2 thời điểm khác nhau như năm 2022 và năm 2023, hệ thống sẽ không thể nhận diện được sự khác biệt trong đó.
- Không có sự đa dạng của các sự kiện trong bộ dữ liệu. Khi xây dựng bộ dữ liệu em chỉ nhắm đến việc tìm kiếm càng nhiều ảnh của cùng một sự kiện càng tốt, chú trọng số lượng ảnh của một sự kiện, từ đó dẫn đến việc số lượng sự kiện trong phạm vi khoá luận chỉ có 10 sự kiện cũng như 8 chính trị gia chứ không phải nhiều hơn.

# Tài liệu tham khảo

- [1] Stephanie Heitman. *What Happens in an Internet Minute in 2022: 90 Fascinating Online Stats*. 2022. URL: <https://localiq.com/blog/what-happens-in-an-internet-minute>.
- [2] Phúc Long. "Sergey Lavrov - nhà ngoại giao kỳ tài của Nga". In: *Báo Tuổi Trẻ* (Mar. 23, 2018). URL: <https://tuoitre.vn/sergey-lavrov-nha-ngoai-giao-ky-tai-cua-nga-20180322143510387.htm> (visited on 03/23/2018).
- [3] Báo Tuổi Trẻ. *Tuổi Trẻ: Tổ hợp truyền thông đa phương tiện hàng đầu Việt Nam. Niềm tin của bạn*. Accessed: 13-02-2023. URL: <https://quangcao.tuoitre.vn/pdf/mediakit.pdf>.
- [4] Bình An. "Chào năm mới, Bộ trưởng Quốc phòng Nga tuyên bố 'chắc thắng' ở Ukraine". In: *Báo Tuổi Trẻ* (Dec. 31, 2022). URL: <https://tuoitre.vn/chao-nam-moi-bo-truong-quoc-phong-nga-tuyen-bo-chac-thang-o-ukraine-2022123117545993.htm>.
- [5] "Điện đàm không đạt kết quả, ông Biden cảnh báo Nga về 'hậu quả nghiêm trọng'". In: *Báo VietNamNet* (Feb. 13, 2022). URL: <https://vietnamnet.vn/dien-dam-khong-dat-ketqua-ong-biden-canhbao-nga-ve-hauqua-nghiem-trong-815129.html> (visited on 02/13/2022).
- [6] *Tờ báo có nhiều độc giả nhất Việt Nam*. Accessed: 13-02-2023. URL: <https://vnexpress.net/contact.htm>.
- [7] Minh Khôi. "Bộ Tư pháp Mỹ đang khám nhà ông Biden tìm tài liệu mật". In: *Báo Tuổi Trẻ* (Feb. 1, 2023). URL: <https://tuoitre.vn/bo-tu-phap-my-dang-kham-nha-ong-biden-tim-tai-lieu-mat-20230201223615587.htm> (visited on 02/01/2023).
- [8] *Who is that? | Bing Visual Search*. URL: <https://www.bing.com/visualesearch/Microsoft/Famouspeople> (visited on 09/12/2022). (accessed: 12.09.2022).

- [9] “Những quốc gia nào quyền lực nhất hành tinh trong năm 2022?”. In: *Quỹ Hòa bình và Phát triển Việt Nam* (Jan. 1, 2023). URL: <http://vpdf.org.vn/tin-tuc-su-kien/doi-ngoai-va-hoi-nhap/nhung-quoc-gia-nao-quyen-luc-nhat-hanh-tinh-trong-nam-2022-.html> (visited on 01/01/2023).
- [10] “Power”. In: *U.S.News* (Sept. 20, 2022). URL: <https://www.usnews.com/news/best-countries/rankings/power> (visited on 09/20/2022).
- [11] *Get creative with stock photos and videos from Alamy*. URL: <https://www.alamy.com> (visited on 12/07/2022). (accessed: 07.12.2022).
- [12] Thanh Trần. “Thương chiến Mỹ - Trung sẽ ra sao nếu ông Biden đắc cử?”. In: *Tạp chí Tài Chính* (Sept. 30, 2020). URL: <https://tapchitaichinh.vn/thuong-chien-my-trung-se-ra-sao-neu-ong-biden-dac-cu.html> (visited on 09/30/2020).
- [13] *pictriev, face search engine*. URL: <http://www.pictriev.com/> (visited on 02/19/2023). (accessed: 19.02.2023).
- [14] *Celebrity look alike face-recognition app | Celebs Like Me*. URL: <https://starbyface.com/> (visited on 09/12/2022). (accessed: 12.09.2022).
- [15] Kaipeng Zhang et al. “Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks”. In: *IEEE Signal Processing Letters* 23.10 (Oct. 2016), pp. 1499–1503. DOI: 10.1109/lsp.2016.2603342. URL: <https://doi.org/10.1109/lsp.2016.2603342>.
- [16] Jia Guo et al. *Sample and Computation Redistribution for Efficient Face Detection*. 2021. DOI: 10.48550/ARXIV.2105.04714. URL: <https://arxiv.org/abs/2105.04714>.
- [17] PaddlePaddle Contributors. *PaddleSeg, End-to-end image segmentation kit based on PaddlePaddle*. <https://github.com/PaddlePaddle/PaddleSeg>. 2019.
- [18] *Image Matting | Papers With Code*. URL: <https://paperswithcode.com/task/image-matting> (visited on 02/19/2023). (accessed: 19.02.2023).
- [19] Vitasta Gurtoo. “Have You Visited These Historic Indian Buildings?”. In: *Travel Earth* (Jan. 15, 2021). URL: <https://travel.earth/have-you-visited-these-historic-indian-buildings/> (visited on 02/19/2023).