

**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

## TÓM TẮT KHOÁ LUẬN

xv

<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 . . . . .	10
Giới hạn về những người nổi tiếng . . . . .	10
Giới hạn về những sự kiện . . . . .	12
1.2 Tính ứng dụng . . . . .	13
1.3 Những thách thức . . . . .	14
1.3.1 Bộ dữ liệu . . . . .	14
1.3.2 Hình ảnh của cùng một sự kiện sẽ có thể khác biệt rất lớn .	16
1.3.3 Sự kiện không có nhiều ảnh . . . . .	18
1.3.4 Các phương pháp trích xuất đặc trưng đại diện cho thông tin sự kiện . . . . .	19
1.3.5 Phương hướng giải quyết các thách thức . . . . .	19
1.4 Mục tiêu khoá luận . . . . .	20
1.5 Cấu trúc Khóa luận tốt nghiệp . . . . .	20
<b>2 NGHIÊN CỨU LIÊN QUAN</b>	<b>21</b>
2.1 Nhận dạng gương mặt . . . . .	21
2.1.1 Phát hiện gương mặt trong ảnh . . . . .	21
Multi-task Cascaded Convolutional Networks - MTCNN .	21
2.1.2 Hệ thống tìm kiếm hình ảnh - Visual Search . . . . .	23
2.2 Nhận dạng sự kiện . . . . .	25
2.2.1 Image matting . . . . .	25

<b>3 HƯỚNG TIẾP CẬN TRONG KHÓA LUẬN</b>	<b>30</b>
3.1 Hướng tiếp cận chung . . . . .	31
3.2 Phát hiện gương mặt trong hình ảnh . . . . .	31
3.3 Nhận dạng gương mặt người nổi tiếng . . . . .	32
3.4 Nhận diện sự kiện tương ứng . . . . .	33
3.4.1 Thông tin từ ngữ cảnh . . . . .	33
3.4.2 Thông tin từ màu sắc . . . . .	33
Histogram . . . . .	33
Màu chủ đạo . . . . .	35
<b>4 THỰC NGHIỆM VÀ ĐÁNH GIÁ</b>	<b>37</b>
4.1 Môi trường và ngôn ngữ cài đặt . . . . .	37
4.2 Phương pháp đánh giá . . . . .	38
4.3 Tập dữ liệu thực nghiệm . . . . .	38
4.3.1 Đối với bài toán phát hiện và nhận diện gương mặt . . . . .	38
4.3.2 Đối với bài toán nhận diện sự kiện tương ứng . . . . .	39
4.4 Xây dựng ứng dụng thực tế . . . . .	42
<b>5 KẾT LUẬN</b>	<b>46</b>
5.1 Kết luận . . . . .	46
5.2 Hướng phát triển . . . . .	46
<b>Tài liệu tham khảo</b>	<b>47</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	The 'Celebrity Together' Dataset - Bộ dữ liệu của trường Đại học Oxford . . . . .	14
1.9	Large-scale CelebFaces Attributes (CelebA) Dataset . . . . .	15
1.10	Thủ tướng Nhật Bản Kishida Fumio bắt tay Tổng thống Indonesia Joko Widodo trong Hội nghị G20 . . . . .	16
1.11	Thủ tướng Nhật Bản Kishida Fumio trong phòng họp khi tham dự Hội nghị G20 . . . . .	17
1.12	Kết quả tìm kiếm trên trang Alamy . . . . .	18
2.1	Ảnh minh họa về 3 giai đoạn của phương pháp MTCNN [12] . . .	22
2.2	Kết quả khi sử dụng MTCNN . . . . .	22
2.3	Giao diện Visual Search . . . . .	23
2.4	Kết quả Visual Search cho hình ảnh trong một bài báo [13] . . .	23
2.5	Hình ảnh minh họa của PaddlePaddle . . . . .	25
2.6	Quy trình matting - Tách ảnh người và nền . . . . .	26
2.7	Quy trình matting - Chèn ảnh nền khác . . . . .	26
2.8	Quy trình trích xuất ảnh nền - Tách ảnh người và nền . . . . .	27
2.9	Loại bỏ ảnh kết quả matting đã thu được . . . . .	27

2.10	Ảnh nền trích xuất được . . . . .	29
3.1	Quy trình giải bài toán . . . . .	31
3.2	Chia hình thành $3 * 3$ vùng . . . . .	34
3.3	So sánh vùng ảnh giữa 2 ảnh trong 1 sự kiện . . . . .	35
3.4	Một minh họa cụ thể về màu chủ đạo . . . . .	36
4.1	Phân bố dữ liệu để huấn luyện và thử nghiệm trong mô hình nhận dạng gương mặt . . . . .	39
4.2	Phân bố số lượng dữ liệu trong tập các sự kiện . . . . .	40
4.3	Kết quả rút trích đặc trưng sự kiện . . . . .	44
4.4	Giao diện trang web sản phẩm . . . . .	45
4.5	Kết quả trả về của trang web . . . . .	45

# 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 . . . . .	12
4.1	Kết quả đánh giá . . . . .	39

# 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

Nội dung chính của Khoá luận này sẽ tập trung vào bài toán Nhận diện người nổi tiếng và sự kiện tương ứng. Trình bày quá trình cũng như các phương pháp thực hiện để xây dựng một hệ thống có thể phát hiện được gương mặt trong một bức ảnh, sau đó nhận diện gương mặt đó là của người nổi tiếng nào; đồng thời cũng trích xuất các thông tin trong bức ảnh để xác định người đó đang ở sự kiện nào.

Khoá luận cũng sẽ đề cập đến những phương pháp liên quan đã có để giải quyết bài toán, phân tích, đánh giá những phương pháp khác nhau, từ đó chọn ra phương pháp tốt nhất để sử dụng trên tập dữ liệu đã có.

Từ những kết quả có trên sẽ có thể xây dựng được một sản phẩm ứng dụng hoàn chỉnh để giải quyết bài toán

## Chương 1

# TỔNG QUAN

### Tóm tắt

Chương này sẽ trình bày định nghĩa về bài toán Nhận dạng người nổi tiếng và các sự kiện tương ứng, các ứng dụng trong thực tế và các thách thức mà bài toán đang gặp phải. Chương này cũng sẽ trình bày các phương pháp đã có để giải quyết bài toán, đồng thời giới thiệu phương pháp được tập trung tìm hiểu. Các mục tiêu, nội dung và phương pháp thực hiện sẽ được nêu cụ thể. Cấu trúc của Khóa luận tốt nghiệp cũng sẽ được đề cập trong chương này.

### 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. Với số lượng thông tin nhiều như vậy, người dùng sẽ rất khó để kiểm chứng được tính đúng sai của các thông tin.

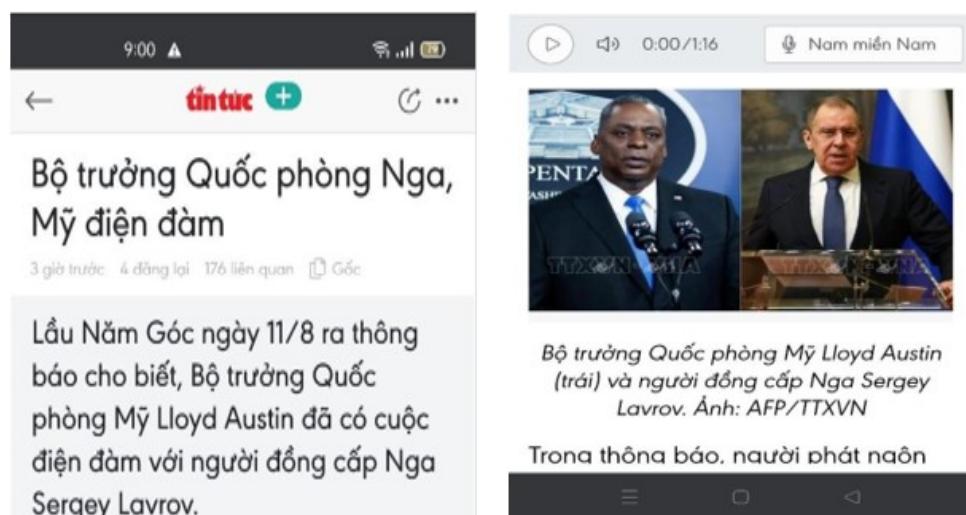
Những người nổi tiếng như các vị nguyên thủ quốc gia, diễn viên, ca sĩ,... là những người có sức ảnh hưởng rất lớn đối với công chúng. Không khó để chúng ta có thấy hình ảnh, tin tức liên quan của họ trên những phương tiện truyền thông đại chúng, mạng xã hội,... Tuy nhiên, nếu hình ảnh của những người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

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 liên tưởng và 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 discusses the Russian Defense Minister's statement during the New Year's address. On the right side of the page, there is a sidebar titled **Rao vặt** (Real estate ads) displaying several real estate listings.

**Báo chí** | **Podcast** | **YouTube** | **Cần biết** | **Rao vặt** | **Đăng ký Tuổi Trẻ Sao**

MEDIA THỜI SỰ THẾ GIỚI PHÁP LUẬT KINH DOANH CÔNG NGHỆ XE DU LỊCH NHỊP SỐNG TRẺ VĂN HÓA GIẢI TRÍ THỂ THAO GIÁO DỤC NHÀ ĐẤT SỨC KHỎE GIÁ THẬT BẢN ĐỌC

Thế giới | 31/12/2022 18:19 GMT+7

**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 VNĐ  
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 VNĐ  
Thành phố Hồ Chí Minh

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

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

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

Bán nhà mặt tiền Nguyễn  
Tất Thành (QL14) 4x24m  
gần Hố Tây Đăk Mil SHR  
6,5 tỷ VNĐ  
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, từ đó dẫn đến sự thiếu hiểu biết sẽ dễ bị dẫn dắt bởi những luồng tin không chính thống.

- 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 is a main title in bold: 'Điện đàm không đạt kết quả, ông Biden cảnh báo Nga về 'hậu quả nghiêm trọng''. The date and time '13/02/2022 06:52 (GMT+07:00)' are also visible. The article summary states: 'Tổng thống Mỹ Joe Biden và người đồng cấp Nga Vladimir Putin đã kết thúc cuộc nói chuyện qua điện thoại trong bối cảnh căng thẳng liên quan đến Ukraine vẫn leo thang không ngừng.' Below the summary, it says: 'Đây là cuộc điện đàm đầu tiên giữa hai nhà lãnh đạo kể từ cuối tháng 12. Cuộc gọi được thực hiện theo yêu cầu của Nhà Trắng, một ngày sau khi giới chức Mỹ cảnh báo Nga sắp tấn công Ukraine.' At the bottom of the article section is a photograph of Joe Biden sitting at his desk in the Oval Office, talking on the phone. The desk has several framed photos and a lamp.

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".

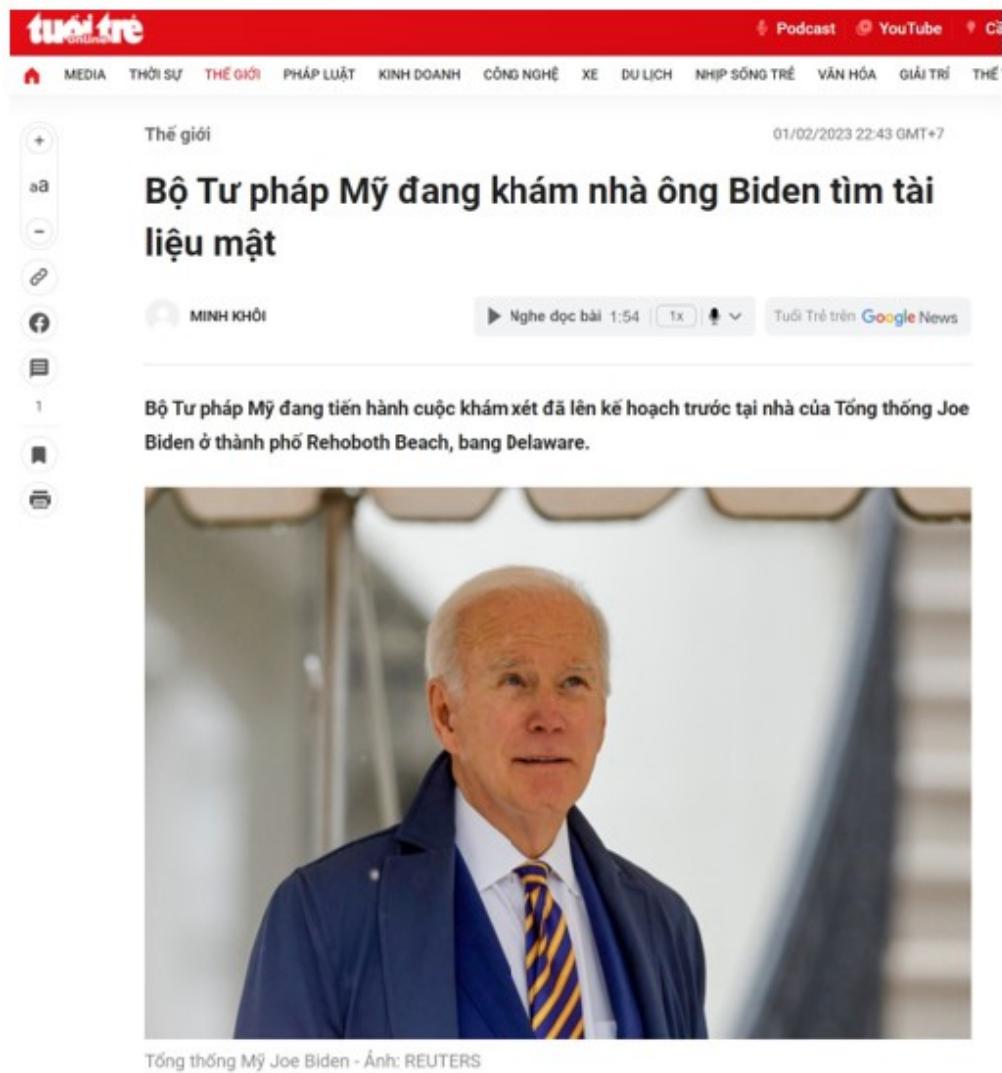


Tổng thống Mỹ Joe Biden điện đàm với người đồng cấp Nga tại Wilmington, Delaware, Mỹ hôm 30/12/2021.  
Ảnh: Nhà Trắng.

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, 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 để 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.

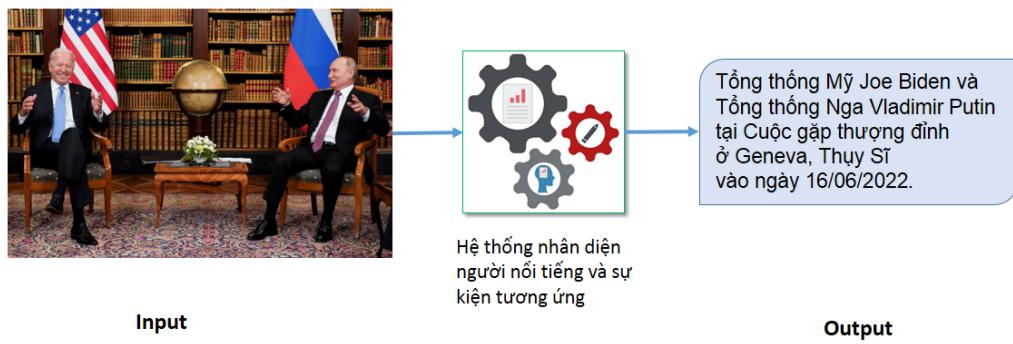
Từ đó có thể thấy, một hệ thống có thể xác định được gương mặt của người nổi tiếng và sự kiện người đó đang tham gia là giải pháp có thể góp phần giảm thiểu tình trạng người đọc tin tức bị ảnh hưởng bởi những thông tin không đủ độ chính xác như vậy.

### 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.

Trong một bình chọn thống kê những quốc gia quyền lực nhất trong năm 2022 do US News & World Report công bố [9], thống kê này dựa trên tầm ảnh hưởng của quốc gia đó đối với thế giới, sự liên kết với các quốc gia đồng minh, lực lượng quân đội,...[10]. Kết quả gồm 10 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

Trong phạm vi khoá luận này, em chỉ chọn nhóm 8 những quốc gia đầu tiên, 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
Điểm đầu Kinh tế Phương Đông	7/9/2022	Vladivostok	Nga
Khoa học kỹ thuật 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, Văn học,...

## 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 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 1.8: The 'Celebrity Together' Dataset - Bộ dữ liệu của trường  
Đại học Oxford

**Sample Images**



HÌNH 1.9: 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 1.10 và hình 1.11 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.



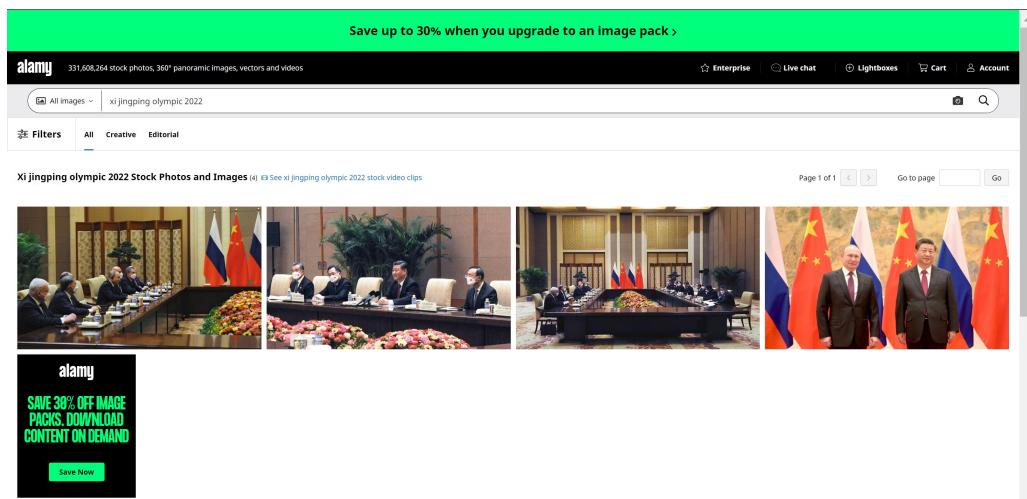
HÌNH 1.10: Thủ tướng Nhật Bản Kishida Fumio bắt tay Tổng thống Indonesia Joko Widodo trong Hội nghị G20



HÌNH 1.11: 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 1.12: 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, những sự kiện đó đã được em chọn lọc kỹ trong quá trình thu thập dữ liệu, đảm bảo được những ảnh trong cùng 1 sự kiện sẽ có độ tương đồng với nhau. 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

### Tóm tắt

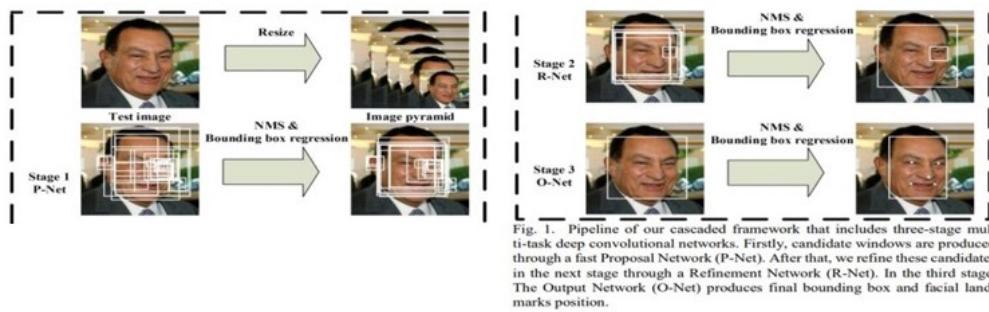
Ở chương này em xin trình bày tổng quan về các phương pháp, ứng dụng đã có để giải quyết một phần nào đó trong bài toán Nhận dạng người nổi tiếng và sự kiện tương ứng

### 2.1 Nhận dạng gương mặt

#### 2.1.1 Phát hiện gương mặt trong ảnh

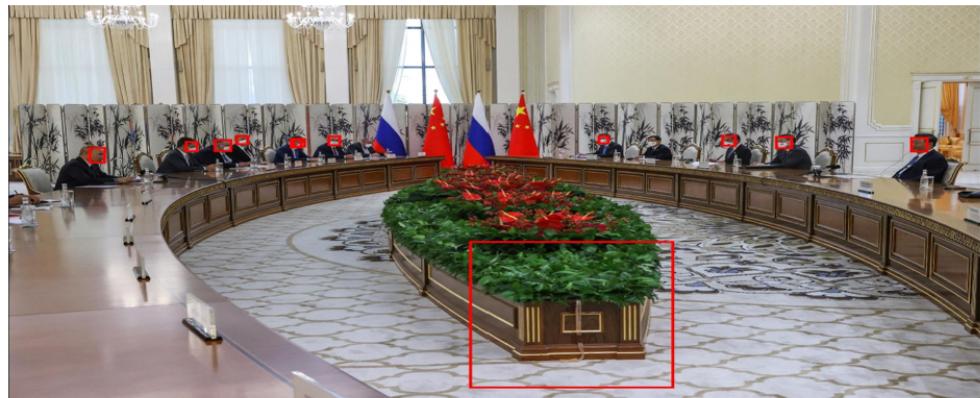
##### Multi-task Cascaded Convolutional Networks - MTCNN

Đây là một thư viện có sẵn được giới thiệu vào năm 2016 [12] và được ứng dụng nhiều trong quá trình phát hiện gương mặt và căn chỉnh gương mặt. Quy trình thực hiện của MTCNN gồm có 3 giai đoạn chính và được minh họa như trong ảnh 2.1



HÌNH 2.1: Ảnh minh họa về 3 giai đoạn của phương pháp MTCNN [12]

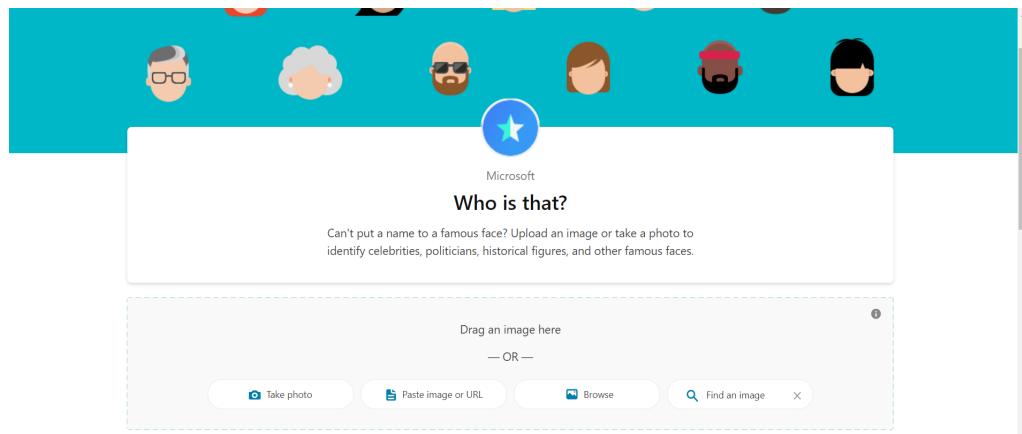
Việc resize bức ảnh nhiều lần như vậy sẽ giúp cho phương pháp này phát hiện nhiều gương mặt nhất có thể - dù cho gương mặt trong ảnh có bị nhỏ hoặc bị che khuất cũng sẽ không khó để phát hiện. Tuy nhiên, từ đó cũng sẽ phát sinh một vấn đề là các chi tiết nhỏ sẽ dễ bị nhầm lẫn thành hình gương mặt, ví dụ như trong hình dưới đây,



HÌNH 2.2: Kết quả khi sử dụng MTCNN

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

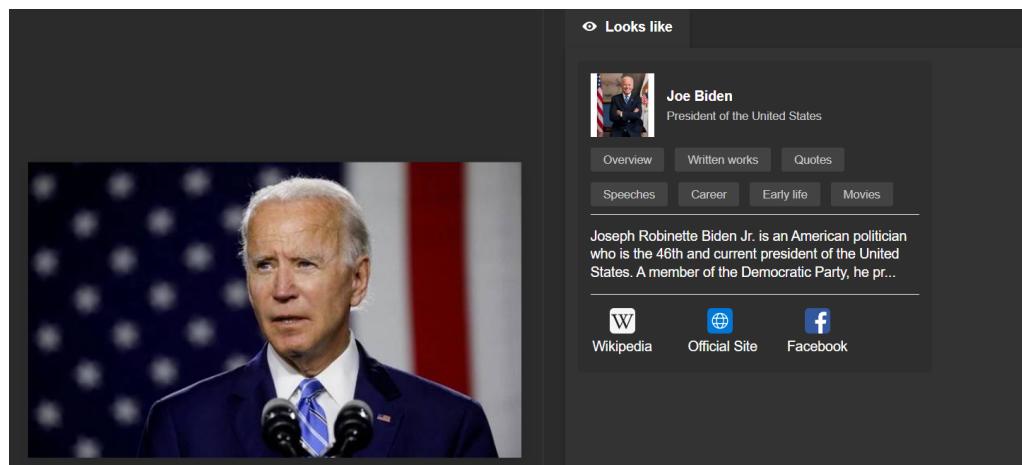
Công cụ tìm kiếm thông tin về một người nổi tiếng Visual Search của Microsoft Bing.



HÌNH 2.3: 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 [13].



HÌNH 2.4: Kết quả Visual Search cho hình ảnh trong một bài báo [13]

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.

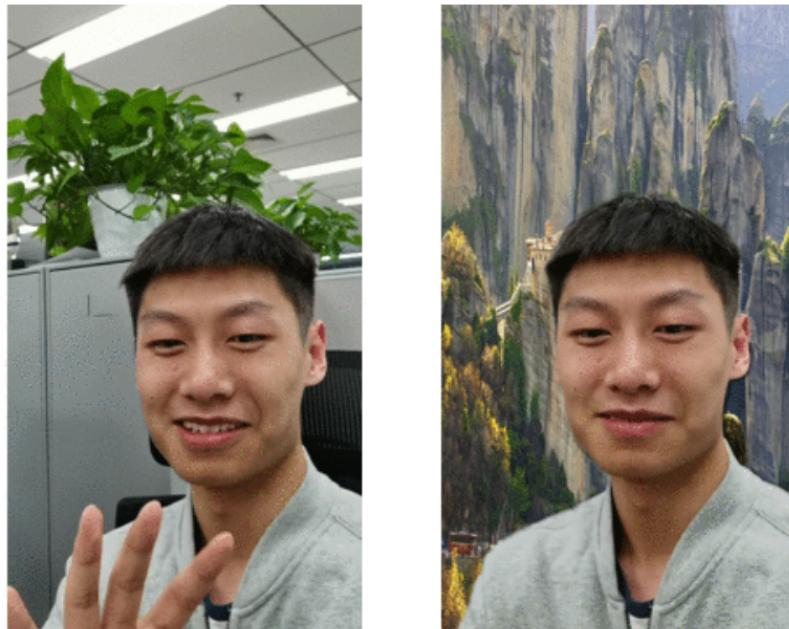
## 2.2 Nhận dạng sự kiện

### 2.2.1 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 [14] 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 [15].

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,...

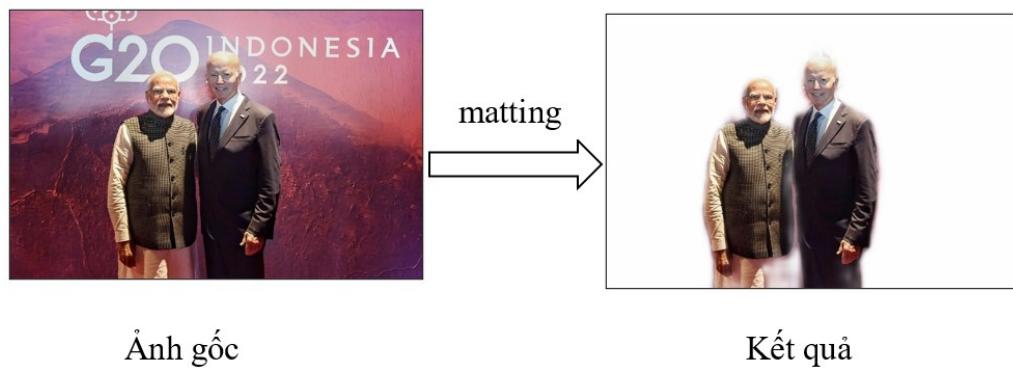


HÌNH 2.5: Hình ảnh minh họa của PaddlePaddle

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.

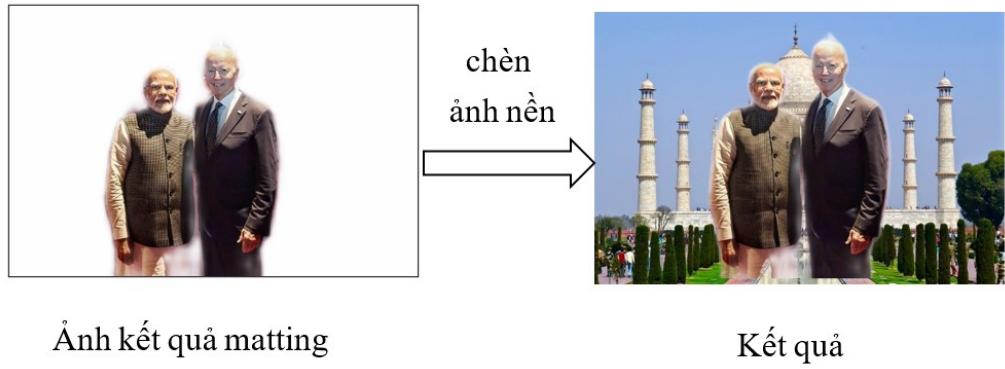
Thực chất, bài toán image matting được cài đặt trong thư viện PaddlePaddle sẽ thực hiện theo quy trình như sau:

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



HÌNH 2.6: Quy trình matting - Tách ảnh người và nền

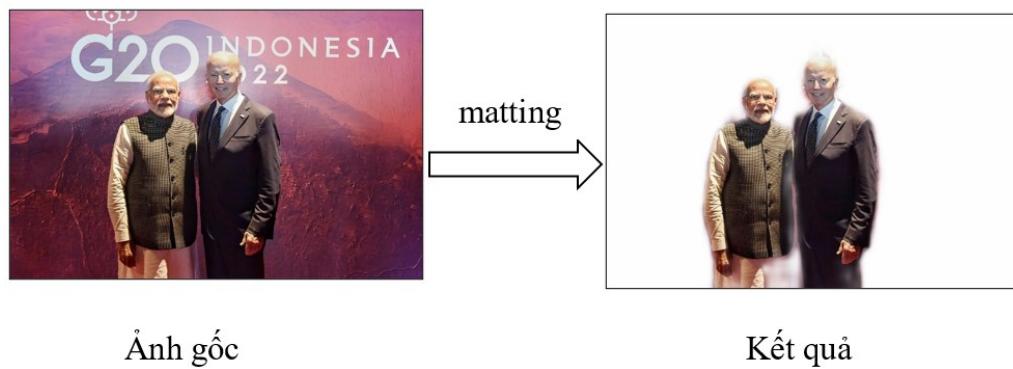
- Bước 2: chèn kết quả đã thu được lên một ảnh nền khác.



HÌNH 2.7: 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 2.8: Quy trình trích xuất ảnh nền - Tách ảnh người và nền

- Bước 2: sử dụng Ảnh gốc ban đầu ở hình 2.8 và loại bỏ phần Kết quả ở hình 2.8.



HÌNH 2.9: 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:

$$A_{background} = \begin{cases} A_{raw}, & \text{if } A_{raw} \neq A_{result} \\ [0, 0, 0] & \end{cases} \quad (2.1)$$

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 2.10 dưới đây.



---

HÌNH 2.10: Ảnh nền trích xuất được

## Chương 3

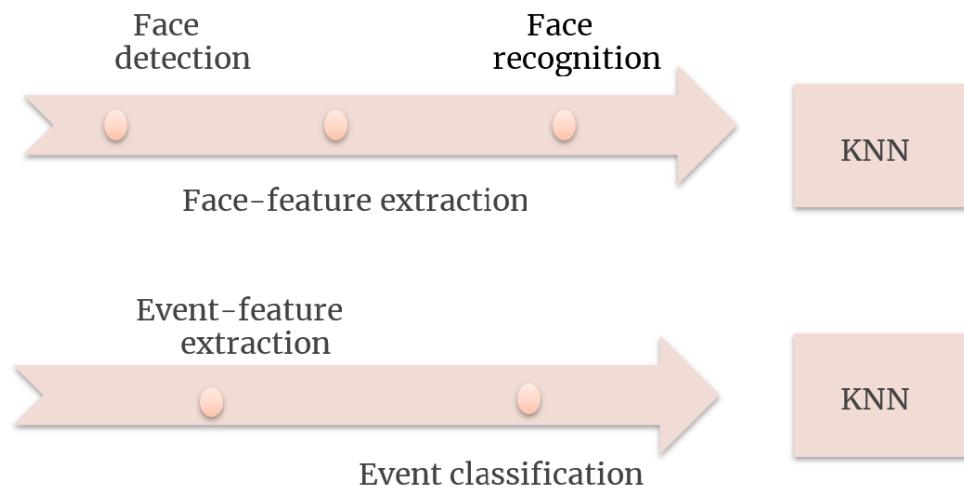
# HƯỚNG TIẾP CẬN TRONG KHÓA LUẬN

### Tóm tắt

Trong chương này, em sẽ trình bày quy trình giải quyết bài toán và các hướng tiếp cận chính được áp dụng. Từ bài toán Nhận dạng người nổi tiếng và sự kiện tương ứng, em xin được phép chia thành 3 bài toán nhỏ hơn để trình bày như sau:

- Phát hiện gương mặt trong hình ảnh
- Nhận diện mặt người nổi tiếng
- Nhận diện sự kiện tương ứng

### 3.1 Hướng tiếp cận chung



HÌNH 3.1: Quy trình giải bài toán

Trong bài toán này, bài toán phát hiện gương mặt và nhận diện gương mặt sẽ dùng chung một mô hình máy học KNN. Các thông tin về sự kiện dùng cho bài toán nhận diện sự kiện sẽ dùng một mô hình máy học KNN khác.

Nguyên nhân dẫn đến lựa chọn sử dụng mô hình máy học KNN là vì mục tiêu của bài toán đang hướng đến việc trích xuất đặc trưng hiệu quả nhất, chỉ cần phương pháp trích xuất đủ tốt thì các đặc trưng giống nhau sẽ được gom cụm, từ đó không cần thêm bất kỳ một mô hình có sự tính toán phức tạp nào khác nữa.

### 3.2 Phát hiện gương mặt trong hình ảnh

- Insightface: Là một thư viện hỗ trợ trong ngôn ngữ lập trình Python. Insightface sẽ sử dụng cách tiếp cận SCRFD [16] để nhận diện gương mặt một cách hiệu quả, lấy cấu trúc mạng Basic Res là cơ sở phát triển.

### 3.3 Nhận dạng gương mặt người nổi tiếng

Sau khi đã trích xuất được mặt người trong những bức hình, ta sẽ dùng các phương pháp rút trích đặc trưng để chuẩn bị cho bài toán nhận diện mặt người nổi tiếng.

Dưới đây là cấu trúc mạng chính mà các thư viện hỗ trợ cho việc trích xuất đặc trưng gương mặt này sử dụng.

- Facenet Pytorch: Sử dụng InceptionResnetV1 với mô hình đã được học với bộ dữ liệu Vggface2 để trích xuất đặc trưng. Kết hợp với thư viện Torch của Python.
- Facenet Keras: Sử dụng InceptionResnetV1 với mô hình đã được học với bộ dữ liệu Vggface2 để trích xuất đặc trưng. Kết hợp với thư viện Keras của Python.
- Insightface: Sử dụng mô hình mạng ResNet50 và bộ dữ liệu WebFace600K để trích xuất đặc trưng gương mặt.

Cả 3 phương pháp rút trích đặc trưng trên đều sẽ tạo ra một vector đặc trưng gồm 512 chiều với một bức ảnh mặt người tương ứng.

## 3.4 Nhận diện sự kiện tương ứng

### 3.4.1 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).
- Sử dụng Bag of Visual word với K = 128, extractor SIFT.

### 3.4.2 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:

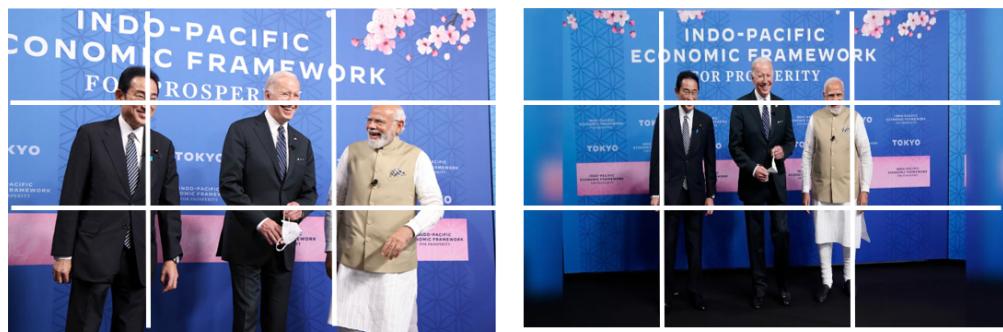
- Bin [8, 8, 8]. Với mỗi kênh màu RGB, giá trị của mỗi pixel sẽ dao động trong khoảng 0 - 255. Với phương pháp này, mỗi kênh màu sẽ được chia thành 8 phần bằng nhau 0-31, 32-63, 63-95,... sau đó lấy giá trị màu đầu tiên của mỗi phần, ta thu được các giá trị gồm [0, 32, 64, 96, 128, 160, 192, 224], và tạo nên sự kết hợp của của các kênh màu. Nên vector đặc trưng của phương pháp này sẽ có số chiều gồm 8 (R) \* 8 (G) \* 8 (B) = 512.

- Đặc trưng histogram, xét sự kết hợp giữa 3 kênh màu chủ đạo, số chiều của vector đặc trưng này =  $255 * 3 = 765$
- Đặc trưng chia vùng, xét sự kết hợp giữa 3 kênh màu chủ đạo trong hình ảnh được chia thành từng vùng nhỏ  $3 * 3$  hoặc  $4 * 4$ . 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 3.2: Chia hình thành  $3 * 3$  vùng

chia thành 9 vùng được đánh số như hình 3.2 thì trong ví dụ như hình 3.3, vùng 1, 2, 3 của hai bức hình sẽ có histogram tương tự nhau. Từ đó học được thông tin về sự kiện.



HÌNH 3.3: 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  $3 * 5 = 15$ .



[164, 175, 86]



[43, 63, 59]



[85, 105, 59]



[226, 226, 224]



[157, 169, 19]

HÌNH 3.4: Một minh họa cụ thể về màu chủ đạo

## Chương 4

# THỰC NGHIỆM VÀ ĐÁNH GIÁ

### Tóm tắt

Chương này sẽ trình bày kết quả thực nghiệm của các hướng tiếp cận trong khóa luận trên tập dữ liệu tự thu thập được. Đồng thời cài đặt một ứng dụng hoàn thiện dựa trên kết quả tốt nhất thu được.

### 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 hỗ 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: insightface.
- Nhận diện gương mặt: torch, torchvision, facenet-pytorch, keras-facenet.
- Nhận diện sự kiện: github repository PaddlePaddle, img2vecpytorch, Pillow, scipy, joblib, 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.

## 4.2 Phương pháp đánh giá

Để đo tính hiệu quả của các phương pháp, sử dụng độ đo Accuracy để đánh giá tính hiệu quả của các đặc trưng trên mô hình máy học KNN.

## 4.3 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 1.

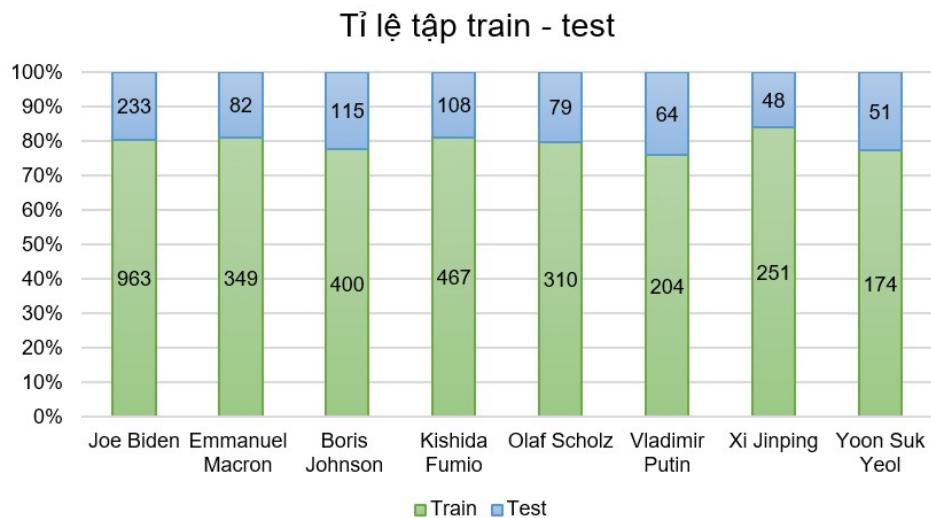
Sau đó với mỗi bức ảnh, em sẽ thực hiện tăng cường dữ liệu bằng cách tạo ra thêm bốn phiên bản khác của bức ảnh như sau:

- Lật hình theo chiều ngang.
- Thay đổi kích thước hình thành 640 x 640
- Shear 15 độ
- Shear 25 độ

Từ đó, mỗi bức ảnh sẽ có 5 phiên bản khác nhau - tính cả ảnh gốc ban đầu không thay đổi, bộ dữ liệu lúc này có tổng cộng 3345 ảnh.

### 4.3.1 Đối với bài toán phát hiện và nhận diện gương mặt

Với riêng bài toán con này, sau khi đã trích xuất gương mặt bằng insightface, có tất cả 3898 gương mặt được trích xuất và được phân bố như sau:



HÌNH 4.1: Phân bổ dữ liệu để huấn luyện và thử nghiệm trong mô hình nhận dạng gương mặt

Thực hiện đánh giá với 3 phương pháp rút trích đặc trưng khác nhau, và kết quả được thể hiện trong bảng 4.1 dưới đây:

Số liệu của pytorch và keras vẫn chưa có, đang chạy model

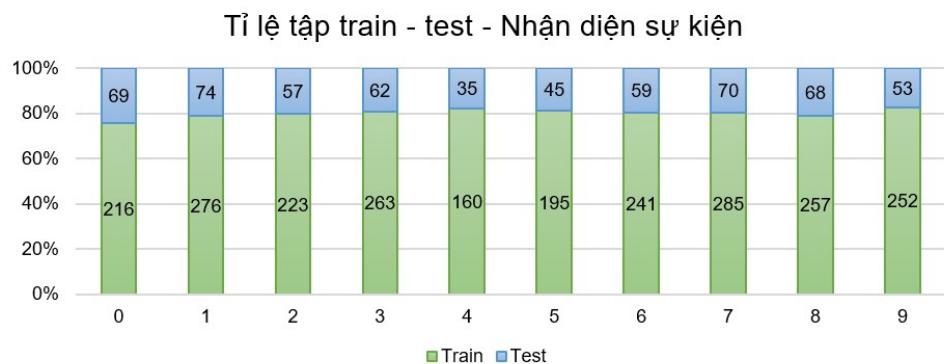
BẢNG 4.1: Kết quả đánh giá của các phương pháp rút trích đặc trưng

Tên phương pháp rút trích đặc trưng	Số k tốt nhất trong mô hình KNN	Accuracy
Facet Pytorch	4	
Facet Keras	2	
Insightface	1	1.0

Theo như kết quả trên, thì sử dụng Insightface là một phương pháp rút trích đặc trưng sẽ là phương pháp có hiệu quả tốt nhất trên bộ dữ liệu này.

### 4.3.2 Đối với bài toán 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 4.2: Phân bổ số lượng dữ liệu trong tập các sự kiện

Ngoài ra, với mỗi lớp sự kiện, sẽ được đánh giá trên hai loại ảnh khác nhau: ảnh nguyên bản - có cả nhân vật và phông nền, và ảnh phông nền - sử dụng matting để loại bỏ nhân vật như trong 2 đã nói. Từ đó nhận xét được mức độ hiệu quả trong việc trích xuất đặc trưng trên ảnh.

Kết quả đánh giá được thể hiện trong ảnh sau:

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

- khả năng rút trích thông tin của hướng tiếp cận sử dụng thông tin ngữ cảnh content based, dựa trên các đặc trưng học sâu như Resnet, Mobilenet là tốt nhất trong các phương pháp rút trích đặc trưng.
- Đối với hướng tiếp cận sử dụng thông tin ngữ cảnh nói chung (gồm đặc trưng học sâu, bag of words thì khả năng rút trích sẽ hiệu quả hơn đối với ảnh gốc, không phân tách nền.
- Đối với hướng tiếp cận về màu sắc (histogram, dominant colors), khả năng rút trích thông tin của ảnh gốc vẫn tốt hơn ảnh chỉ có phông nền. Và đối với nhóm rút trích thông tin dựa trên histogram thì chia ảnh thành nhiều vùng nhỏ hơn thì đặc trưng rút trích càng thể hiện rõ hơn thông tin về sự kiện.

## 4.4 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.

- Phát hiện gương mặt: detector Insightface.
- Nhận diện gương mặt: phương pháp rút trích đặc trưng Insightface.
- Nhận diện sự kiện: dùng đặc trưng học sâu Resnet18 cho ảnh đầu vào, không cần tách nền.

Đường link truy cập vào trang web: [bit.ly/waw-web](http://bit.ly/waw-web)

Giao diện của trang web được xây dựng như hình minh họa 4.4.

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.

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

		Method	Accuracy
0	Resnet, background image	0.986486	
1	Resnet, original image	1.000000	
2	Histogram bin, background image	0.988176	
3	Histogram bin, original image	1.000000	
4	Histogram region 3 x 3, background image	0.976351	
5	Histogram region 3 x 3, original image	0.998311	
6	Histogram region 4 x 4, background image	0.961149	
7	Histogram region 4 x 4, original image	0.993243	
8	MobileNet, background image	0.989865	
9	MobileNet, original image	1.000000	
10	Histogram RGB, background image	0.957770	
11	Bag of visual words, k=128, background image	0.837838	
12	Dominant colors, background image	0.454392	
13	Dominant colors, origin image	0.415541	

HÌNH 4.3: Kết quả rút trích đặc trưng sự kiện

The screenshot shows a user interface for identifying famous people and events. On the left, there is a file upload area with options to "Tải ảnh lên" (Upload image), "Drag and drop file here" (Limit 200MB per file), "Browse files", and "hoặc" (or) "Dán URL". On the right, the title "Who is that famous person & What's the event?" is displayed. Below it, the text "Một vài sự kiện mẫu" (Some sample events) is shown. Three images are presented, each with a small profile icon and a caption:

- Vladimir Putin, Tập Cận Bình (4/2/2022)
- Donald Trump (3/9/2022)
- Joe Biden (25/8/2022)

Below each image is a link to more information: "Lễ khai mạc Olympic", "Một sự kiện ở Pennsylvania", and "Một sự kiện vận động Nước Mỹ an toàn hơn ở Maryland".

HÌNH 4.4: Giao diện trang web sản phẩm

This screenshot shows the same web application interface as Figure 4.4. On the left, there is a file upload area with options to "Tải ảnh lên" (Upload image), "Drag and drop file here" (Limit 200MB per file), "Browse files", and "hoặc" (or) "Dán URL". On the right, the title "Who is that famous person & What's the event?" is displayed. Below it, the text "Một vài sự kiện mẫu" (Some sample events) is shown. One image is presented, showing King Charles III speaking at a podium. A red box highlights his face. The file upload section on the left shows a file named "0\_Proclamation-Of-King-C... 46.8KB". The search results on the right are identical to Figure 4.4.

HÌNH 4.5: 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 với backbone là Basic Net, kết hợp với phương pháp SCRFD.
- Trích xuất đặc trưng gương mặt: Sử dụng Insightface với cấu trúc ResNet50 để trích xuất đặc trưng hiệu quả nhất.
- Trích xuất đặc trưng sự kiện: Đối với thông tin ngữ cảnh: dùng ResNet18; Đối với thông tin màu sắc: dùng histogram phân vùng  $3 * 3$ .
- 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.

# 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] 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>.
- [13] 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).
- [14] PaddlePaddle Contributors. *PaddleSeg, End-to-end image segmentation kit based on PaddlePaddle*. <https://github.com/PaddlePaddle/PaddleSeg>. 2019.
- [15] *Image Matting | Papers With Code*. URL: <https://paperswithcode.com/task/image-matting> (visited on 02/19/2023). (accessed: 19.02.2023).
- [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>.