

# THÔNG TIN CHUNG CỦA NHÓM

- Link YouTube video của báo cáo:

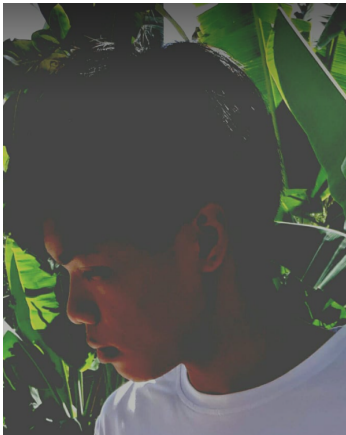
<https://www.youtube.com/watch?v=5lyQuIMeMcg>

- Link slides:

<https://github.com/SoulOfWindTGN/CS519.M11/blob/main/%C4%90%E1%BB%93%20%C3%A1n%20cu%E1%BB%91i%20k%E1%BB%B3/CS519.Slide.pdf>

- Họ và Tên: Trần Gia Nghĩa

- MSSV: 19521901



- Lớp: CS519.M11

- Tự đánh giá (điểm tổng kết môn): 8.5 /10

- Số buổi vắng: 0

- Số câu hỏi QT cá nhân: 11



- Số câu hỏi QT của cả nhóm: 4

- Link Github:

<https://github.com/SoulOfWindTGN/CS519.M11>

- Mô tả công việc và đóng góp của cá nhân cho kết quả của nhóm:

- Đề xuất và lên ý tưởng bài toán
- Khảo sát các nghiên cứu liên quan
- Viết nội dung báo cáo
- Thiết kế phần Description của poster

<ul style="list-style-type: none"> <li>• Họ và Tên: Nguyễn Thị Thúy An</li> <li>• MSSV: 19521183</li> </ul> 	<ul style="list-style-type: none"> <li>• Lớp: CS519.M11</li> <li>• Tự đánh giá (điểm tổng kết môn): 7.5/10</li> <li>• Số buổi vắng: 0</li> <li>• Số câu hỏi QT cá nhân: 11</li> <li>• Số câu hỏi QT của cả nhóm: 4</li> <li>• Link Github: <a href="https://github.com/UIT-19521183/CS519.M11">https://github.com/UIT-19521183/CS519.M11</a></li> <li>• Mô tả công việc và đóng góp của cá nhân cho kết quả của nhóm: <ul style="list-style-type: none"> <li>- Chỉnh sửa nội dung báo cáo</li> <li>- Tham gia làm một phần poster</li> <li>- Làm slide</li> </ul> </li> </ul>
<ul style="list-style-type: none"> <li>• Họ và Tên: Nguyễn Lâm Thảo Vy</li> <li>• MSSV: 19522547</li> </ul> 	<ul style="list-style-type: none"> <li>• Lớp: CS519.M11</li> <li>• Tự đánh giá (điểm tổng kết môn): 7/10</li> <li>• Số buổi vắng: 0</li> <li>• Số câu hỏi QT cá nhân: 11</li> <li>• Số câu hỏi QT của cả nhóm: 4</li> <li>• Link <span style="float: right;">Github:</span> <a href="https://github.com/UIT19522547/CS519.M11">https://github.com/UIT19522547/CS519.M11</a></li> <li>• Mô tả công việc và đóng góp của cá nhân cho kết quả của nhóm: <ul style="list-style-type: none"> <li>- Tham gia làm 1 phần poster</li> <li>- Làm video</li> <li>- Đóng góp ý kiến về nội dung báo cáo</li> </ul> </li> </ul>

# ĐỀ CƯƠNG NGHIÊN CỨU

## TÊN ĐỀ TÀI (IN HOA)

THIẾT KẾ HỆ THỐNG PHÂN TÍCH THÔNG TIN VÀ TÌM KIẾM CON NGƯỜI TỪ CAMERA GIÁM SÁT NGOÀI TRỜI

## TÊN ĐỀ TÀI TIẾNG ANH (IN HOA)

DESIGN SYSTEM FOR INFORMATION ANALYSIS AND PERSON SEARCH FROM SURVEILLANCE CAMERAS

## TÓM TẮT (*Tối đa 400 từ*)

Person re-identification, hay person retrieval, là một trong những bài toán khó khăn trong lĩnh vực giám sát an ninh. Với đề tài này, dựa trên các nghiên cứu của tiền nhân, nhóm chúng tôi tập trung nghiên cứu các vấn đề sau:

- **Thứ nhất:** Nghiên cứu và cải tiến mô hình khai thác và lập luận thuộc tính con người được nhóm tác giả của bài báo [1] và [2] đề xuất. Trong đó, chú trọng giải quyết các bài toán con gồm: person detection, retrieval và classification. Ngoài ra, chúng tôi sẽ nghiên cứu các phương pháp cho trường hợp người bị khuất - trường hợp rất phổ biến trong tập dữ liệu camera giao thông ở Việt Nam trên cơ sở bài báo [4].
- **Thứ hai:** Đối tượng con người trong các tập dữ liệu sẵn có phần lớn là người đi bộ. Trong thực tế, truy vấn người cần được triển khai ở các camera giao thông chứa cả người đi bộ và người lái phương tiện giao thông. Vì vậy, chúng tôi sẽ tiến hành thu thập và gán nhãn cho tập dữ liệu được thu thập từ camera giao thông ở Việt Nam và các camera an ninh ở một số khu vực công cộng (phục vụ tìm người thất lạc, tội phạm...).
- **Thứ ba:** Thiết kế và xây dựng ứng dụng nhằm mục đích trực quan hóa các kết quả nghiên cứu cũng như ứng dụng cho các bài toán thực tế.

## **GIỚI THIỆU (Tối đa 1 trang A4)**

Cách mạng công nghiệp 4.0 đã kéo theo sự phát triển rất nhanh của lĩnh vực khoa học và công nghệ. Trong những thập kỷ gần đây, nhân loại đã chứng kiến sự phát triển mạnh mẽ của lĩnh vực công nghệ thông tin cũng như công nghệ kỹ thuật số. Công nghệ thông tin nói chung và lĩnh vực trí tuệ nhân tạo nói riêng được ứng dụng rất nhiều trong đời sống của chúng ta, giúp nâng cao và thay đổi đáng kể lối sống của con người. Trong đó, sự phát triển mạnh mẽ của thị giác máy tính và công nghệ chip xử lý đã góp phần vào việc xây dựng các mô hình nhận dạng với tốc độ xử lý cao và chính xác. Điều đó đã mở ra nhiều hướng đi tiềm năng cho các bài toán lớn hơn, trong đó có nhận diện và truy xuất thông tin con người từ các đặc điểm sinh trắc học - bài toán mang tính ứng dụng cao, đóng vai trò là cơ sở xây dựng các thành phố thông minh.

Ngày nay, hệ thống camera giám sát được triển khai ở nhiều nơi, giúp con người dễ dàng quản lý và kiểm soát an ninh ở một khu vực cụ thể. Tuy nhiên, việc tìm kiếm và truy xuất thông tin đa phần vẫn phải thực hiện thủ công thông qua video, khiến cho năng suất tìm kiếm và hiệu quả không cao. Do đó, việc xây dựng một hệ thống tự động truy xuất và thu thập dữ liệu ngoại hình con người từ camera giám sát có thể ứng dụng cho nhiều mục đích như:

- Rút ngắn phạm vi và thời gian tìm kiếm con người khi biết mô tả ngoại hình, thời gian và địa điểm xuất hiện. Ứng dụng cho công việc của các điều tra viên nhằm tìm kiếm tội phạm hoặc F0 mắc Covid-19 qua mô tả địa điểm và ngoại hình
- Xây dựng hệ thống an ninh nhằm xác định đúng người có quyền ra vào trụ sở, phòng ban.
- Dữ liệu đã truy xuất và thu thập có thể chia sẻ để phục vụ cho các nghiên cứu về hành vi con người.
- Phục vụ việc tìm kiếm tội phạm, người thất lạc, người có hành vi bất thường nhanh chóng bằng camera an ninh.

Với bài toán cùng đề tài của các nhóm nghiên cứu trước được thực hiện trên tập dữ liệu người đi bộ. Trong thực tế, truy vấn người cần được triển khai ở các camera giao thông chứa cả người đi bộ và người lái phương tiện giao thông với các góc quay khác nhau và chất lượng hình ảnh không đồng đều. Vì những hạn chế trên, nhóm chúng tôi sẽ nghiên cứu cải tiến mô

hình truy xuất thông tin và tìm kiếm con người trong ngữ cảnh camera ở đường phố Việt Nam.

**INPUT:** Một video trích xuất từ camera giám sát và một ảnh chứa người cần tìm kiếm (ảnh truy vấn)

**OUTPUT:** Video chứa bounding box của người tìm được.

### **MỤC TIÊU** (*Viết trong vòng 3 mục tiêu*)

- Xây dựng một tập dữ liệu từ camera giám sát ở Việt Nam
- Đề xuất mô hình khai thác và lập luận thuộc tính con người bao gồm: giới tính, nhóm tuổi, màu sắc trang phục để giải quyết thách thức của tập dữ liệu đã thu thập.
- Xây dựng được một ứng dụng phân tích thông tin con người đã thu thập và tìm kiếm con người qua hình ảnh ngoại hình.

### **NỘI DUNG VÀ PHƯƠNG PHÁP**

#### **Nội dung 1: Nghiên cứu tổng quan và xây dựng giải thuật**

##### **Phương pháp thực hiện:**

- Tìm kiếm và nghiên cứu các công trình nghiên cứu liên quan trong và ngoài nước về bài toán person re-identification và person search. Đặc biệt, nghiên cứu phương pháp giải quyết trường hợp người bị khuất một phần để phù hợp với ngữ cảnh camera đường phố ở Việt Nam.
- Nghiên cứu các mô hình và phương pháp giải quyết 3 bài toán con sau:
  - + Person detection: nghiên cứu và so sánh các mô hình với hai kiến trúc mạng CNN (YOLO, Mask R-CNN) và Vision Transformer (Swin Transformer)
  - + Person retrieval: tập trung nghiên cứu vào các mô hình khai thác và lập luận thuộc tính (Attribute Mining and Reasoning - AMR) [2]/[3] và mô hình phân vùng và cục bộ hóa thuộc tính (Attribute CAM for Pedestrian Body Part Localization) [4]
  - + Person classification: Tập trung nghiên cứu kiến trúc Graph Convolution Network [1]
- Tổng hợp các phương pháp đã được đề xuất, phân tích ưu điểm và khuyết

điểm của các phương pháp đó. Tìm cách kết hợp 1 số phương pháp với nhau để giảm tác động của khuyết điểm.

- Tìm 5 bộ dữ liệu sẵn có: Market1501, DukeMTMC, CUHK01, CUHK02, CUHK03. Thu thập và tự gắn nhãn thêm 45 video từ camera giao thông ở Việt Nam và một số khu vực công cộng ở Việt Nam để phục vụ cho thực tiễn nước ta, mỗi video có thời lượng là 20 phút.
- Tìm hiểu và thử nghiệm các phương pháp tăng cường dữ liệu.
- Xây dựng giải thuật cho bài toán dựa trên các phương pháp được chọn.

### **Kết quả dự kiến:**

Báo cáo nghiên cứu tổng quan và chi tiết giải thuật đã đề xuất.

## **Nội dung 2: Hiện thực giải thuật trên máy tính**

### **Phương pháp thực hiện:**

- Xây dựng mô hình dựa trên giải thuật đã nghiên cứu.
- Huấn luyện và tinh chỉnh tham số cho mô hình đã nghiên cứu trên tập dữ liệu đã thu thập.
- So sánh và đánh giá các mô hình đã xây dựng, phân tích ưu khuyết điểm của từng mô hình dựa trên: độ chính xác, thời gian thực thi, dung lượng bộ nhớ chính sử dụng và một số tiêu chí khác.
- Nghiên cứu cách thức tương tác giữa người dùng và hệ thống sẽ xây dựng: hệ thống có chỗ nhập video và ảnh input và xuất output video, có hiển thị bảng thống kê thông tin của tất cả người truy xuất được từ video và biểu đồ thống kê các thuộc tính của tất cả những người này.
- Đóng gói mô hình đã xây dựng để triển khai thành hệ thống hoàn chỉnh: đóng gói thành một ứng dụng web sử dụng ngôn ngữ Python và framework streamlit. Sau khi người dùng nhập ảnh input, người dùng có thể tải về video kết quả và một file pdf chứa thống kê thuộc tính của những người người truy xuất được trong video.

### **Kết quả dự kiến:**

- Hệ thống có khả năng truy xuất, thống kê thông tin con người và tìm kiếm con người trong video bằng mô tả ngoại hình dựa trên hình ảnh.
- Ứng dụng web có giao diện đơn giản, thân thiện với người dùng.
- Báo cáo đánh giá kết quả thử nghiệm hệ thống.

### **Nội dung 3: Thực nghiệm và đánh giá kết quả nghiên cứu**

#### **Phương pháp thực hiện:**

- Thử nghiệm hệ thống với dữ liệu thực tế.
- Ghi nhận kết quả và phân tích ưu điểm và hạn chế so với các phương pháp trước đó.
- Viết báo cáo kết quả nghiên cứu.

#### **Kết quả dự kiến:**

Báo cáo kết quả của đề tài nghiên cứu.

### **KẾT QUẢ MONG ĐỢI**

- Hệ thống có thể phát hiện người và truy xuất thông tin của người trong video giám sát.
- Hệ thống có chức năng thống kê thông tin và tìm kiếm con người dựa vào ảnh truy vấn (chứa thông tin ngoại hình con người) một cách chính xác và trong thời gian giới hạn có thể chấp nhận (nếu hơn thời gian con người tìm kiếm thì không còn ý nghĩa).
- Báo cáo tổng kết gồm các bước thực hiện của giải thuật, các mô hình đề xuất và kết quả thực nghiệm.

### **KẾ HOẠCH THỰC HIỆN**

<b>Thời gian thực hiện</b>	Tháng thứ nhất - tháng thứ 2  (2 tháng)	Tháng thứ 3 - tháng thứ 5  (3 tháng)	Tháng thứ 6  (1 tháng)
<b>Nội dung</b>	Nội dung 1: Nghiên cứu tổng quan và xây dựng giải thuật	Nội dung 2: Hiện thực giải thuật trên máy tính	Nội dung 3: Thực nghiệm và đánh giá kết quả nghiên cứu

## **TÀI LIỆU THAM KHẢO (Định dạng DBLP)**

[1] Thomas N. Kipf, Max Welling:

*Semi-Supervised Classification with Graph Convolutional Networks.*

*ICLR, 2017*

[2] Hiren Galiyawala, Kenil Shah, Vandit Gajjar, Mehul S. Raval:

*Person Retrieval in Surveillance Video using Height, Color and Gender.*

*AVSS, 2018*

[3] Yuxuan Shi, Zhen Wei, Hefei Ling, Ziyang Wang, Jialie Shen, Ping Li:

*Person Retrieval in Surveillance Videos via Deep Attribute Mining and Reasoning.*

*IEEE T MULTIMEDIA, 2020*

[4] Hanyang Jin, Shenqi Lai, and Xueming Qian:

*Occlusion-sensitive Person Re-identification via Attribute-based Shift Attention .*

*IEEE Transactions on Circuits and Systems for Video Technology 2021*