

ĐẠI HỌC ĐÀ NẴNG  
TRƯỜNG ĐẠI HỌC BÁCH KHOA  
KHOA ĐIỆN TỬ - VIỄN THÔNG

**ĐỒ ÁN TỐT NGHIỆP**  
**NGÀNH: KỸ THUẬT ĐIỆN TỬ VÀ VIỄN THÔNG**  
**CHUYÊN NGÀNH: KỸ THUẬT MÁY TÍNH**

**ĐỀ TÀI:**  
**HỆ THỐNG QUẢNG CÁO**  
**THÔNG MINH**

|                     |                      |
|---------------------|----------------------|
| Người hướng dẫn     | : TS. NGÔ MINH TRÍ   |
| Sinh viên thực hiện | : NGUYỄN ĐÌNH NGHĨA  |
| Số thẻ sinh viên    | : 106190119          |
| Sinh viên thực hiện | : HUỖNH THỊ LỆ TRINH |
| Số thẻ sinh viên    | : 106190134          |
| Lớp                 | : 19DTCLC3           |

Đà Nẵng, 07/2023

ĐẠI HỌC ĐÀ NẴNG  
TRƯỜNG ĐẠI HỌC BÁCH KHOA  
KHOA ĐIỆN TỬ - VIỄN THÔNG

**ĐỒ ÁN TỐT NGHIỆP**  
**NGÀNH: KỸ THUẬT ĐIỆN TỬ VÀ VIỄN THÔNG**  
**CHUYÊN NGÀNH: KỸ THUẬT MÁY TÍNH**

**ĐỀ TÀI:**  
**HỆ THỐNG QUẢNG CÁO**  
**THÔNG MINH**

|                     |                      |
|---------------------|----------------------|
| Người hướng dẫn     | : TS. NGÔ MINH TRÍ   |
| Sinh viên thực hiện | : NGUYỄN ĐÌNH NGHĨA  |
| Số thẻ sinh viên    | : 106190119          |
| Sinh viên thực hiện | : HUỖNH THỊ LỆ TRINH |
| Số thẻ sinh viên    | : 106190134          |
| Lớp                 | : 19DTCLC3           |

Đà Nẵng, 07/2023

## NHẬN XÉT ĐỒ ÁN TỐT NGHIỆP

### I. Thông tin chung:

| STT | Họ và tên sinh viên | Mã số sinh viên | Lớp      | Ngành              |
|-----|---------------------|-----------------|----------|--------------------|
| 1   | Nguyễn Đình Nghĩa   | 106190119       | 19DTCLC3 | Điện tử-viễn thông |
| 2   | Huỳnh Thị Lệ Trinh  | 106190134       | 19DTCLC3 | Điện tử-viễn thông |

Người hướng dẫn: Ngô Minh Trí

Học hàm/ học vị: Tiến sĩ

### II. Nhận xét, đánh giá đồ án tốt nghiệp:

1. Về tính cấp thiết, tính mới, khả năng ứng dụng của đề tài: (điểm tối đa là 2đ)

.....  
.....

2. Về kết quả giải quyết các nội dung nhiệm vụ yêu cầu của đồ án: (điểm tối đa là 4đ)

.....  
.....

3. Về hình thức, cấu trúc, bố cục của đồ án tốt nghiệp: (điểm tối đa là 2đ)

.....  
.....

4. Đề tài có giá trị khoa học/ có bài báo/ giải quyết vấn đề đặt ra của doanh nghiệp hoặc nhà trường: (điểm tối đa là 1đ)

.....  
.....

5. Các tồn tại, thiếu sót cần bổ sung, chỉnh sửa:

.....  
.....

### III. Tinh thần, thái độ làm việc của sinh viên: (điểm tối đa 1đ)

.....

### IV. Đánh giá:

1. Điểm đánh giá: ...../10 (lấy đến 1 số lẻ thập phân)

2. Đề nghị: ☐ Được bảo vệ đồ án    ☐ Bổ sung để bảo vệ    ☐ Không được bảo vệ

Đà Nẵng, ngày      tháng      năm 2023

**Người hướng dẫn**

## NHẬN XÉT PHẢN BIỆN ĐỒ ÁN TỐT NGHIỆP

### I. Thông tin chung:

| STT | Họ và tên sinh viên | Mã số sinh viên | Lớp      | Ngành              |
|-----|---------------------|-----------------|----------|--------------------|
| 1   | Nguyễn Đình Nghĩa   | 106190119       | 19DTCLC3 | Điện tử-viễn thông |
| 2   | Huỳnh Thị Lệ Trinh  | 106190134       | 19DTCLC3 | Điện tử-viễn thông |

Người hướng dẫn: Ngô Minh Trí

Học hàm/ học vị: Tiến sĩ

### II. Nhận xét, đánh giá đồ án tốt nghiệp:

| TT | Các tiêu chí đánh giá                                                                                                                                                                                     | Điểm tối đa | Điểm đánh giá |
|----|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------|---------------|
| 1  | <b>Sinh viên có phương pháp nghiên cứu phù hợp, giải quyết đủ nhiệm vụ đồ án được giao</b>                                                                                                                | 80          |               |
| 1a | - Tính mới (nội dung chính của ĐATN có những phần mới so với các ĐATN trước đây).<br>- Đề tài có giá trị khoa học, công nghệ; có thể ứng dụng thực tiễn.                                                  | 15          |               |
| 1b | - Kỹ năng giải quyết vấn đề; hiểu, vận dụng được kiến thức cơ bản, cơ sở, chuyên ngành trong vấn đề nghiên cứu.<br>- Chất lượng nội dung ĐATN (thuyết minh, bản vẽ, chương trình, mô hình,...).           | 50          |               |
| 1c | - Có kỹ năng vận dụng thành thạo các phần mềm ứng dụng trong vấn đề nghiên cứu;<br>- Có kỹ năng đọc, hiểu tài liệu bằng tiếng nước ngoài ứng dụng trong vấn đề nghiên cứu;<br>- Có kỹ năng làm việc nhóm; | 15          |               |
| 2  | <b>Kỹ năng viết:</b>                                                                                                                                                                                      | 20          |               |
| 2a | - Bố cục hợp lý, lập luận rõ ràng, chặt chẽ, lời văn súc tích                                                                                                                                             | 15          |               |
| 2b | - Thuyết minh đồ án không có lỗi chính tả, in ấn, định dạng                                                                                                                                               | 5           |               |
| 3  | <b>Tổng điểm đánh giá theo thang 100:</b>                                                                                                                                                                 |             |               |
|    | <b>Quy về thang 10 (lấy đến 1 số lẻ)</b>                                                                                                                                                                  |             |               |

- Các tồn tại, thiếu sót cần bổ sung, chỉnh sửa:

.....

- Câu hỏi đề nghị sinh viên trả lời trong buổi bảo vệ:

.....

- Đề nghị: ☐ Được bảo vệ đồ án    ☐ Bổ sung đề bảo vệ    ☐ Không được bảo vệ

Đà Nẵng, ngày      tháng      năm 2023

**Người phản biện**

## TÓM TẮT

Tên đề tài: “HỆ THỐNG QUẢNG CÁO THÔNG MINH”.

Sinh viên thực hiện:

Nguyễn Đình Nghĩa Lớp: 19DTCLC3 Số thẻ sinh viên: 106190119

Huỳnh Thị Lệ Trinh Lớp: 19DTCLC3 Số thẻ sinh viên: 106190134

Lĩnh vực quảng cáo hiện nay đang chuyển đổi theo xu hướng kỹ thuật số và tận dụng sự phát triển công nghệ để tiếp cận khách hàng mục tiêu và tạo ra sự tương tác. Xuất hiện như một tất yếu khách quan, quảng cáo thông minh ứng dụng AI là xu hướng của các doanh nghiệp trong thời buổi hiện đại với một số công dụng nổi bật sau đây:

**Phân tích khách hàng:** Biểu quảng cáo thông minh AI có khả năng phân tích dữ liệu và nhận biết các đặc điểm của khách hàng, bao gồm độ tuổi, giới tính, sở thích và hành vi tiêu dùng. Dựa trên thông tin này, nó có thể cung cấp quảng cáo được cá nhân hóa và phù hợp với nhóm khách hàng cụ thể, tăng cường hiệu quả của chiến dịch quảng cáo.

**Tăng cường tương tác:** Biểu quảng cáo thông minh có khả năng tương tác với khách hàng một cách thông minh. Với sự hỗ trợ của trí tuệ nhân tạo, nó có thể nhận dạng và phản hồi, đồng thời cung cấp thông tin một cách tự động, điều này tạo ra trải nghiệm tương tác tích cực và thu hút sự chú ý của khách hàng

**Tối ưu hóa quảng cáo:** Với trí tuệ nhân tạo, biểu quảng cáo thông minh có thể tự động tối ưu hóa chiến dịch quảng cáo. Nó có khả năng phân tích dữ liệu liên tục và điều chỉnh nội dung quảng cáo, vị trí hiển thị và thời gian phát để đạt được hiệu quả tốt nhất. Điều này giúp tiết kiệm thời gian và nguồn nhân lực, đồng thời tăng cường hiệu suất quảng cáo.

**Nhận dạng khuôn mặt và nhận diện đối tượng:** Biểu quảng cáo thông minh được trang bị công nghệ nhận dạng khuôn mặt và nhận diện đối tượng. Điều này cho phép nó nhận biết và phân loại khách hàng dựa trên đặc điểm giới tính và cung cấp nội dung quảng cáo phù hợp và tăng cường tương tác.

**Đo lường hiệu quả:** Biểu quảng cáo thông minh có khả năng đo lường hiệu quả của chiến dịch quảng cáo. Nó có thể thu thập dữ liệu về tương tác khách hàng, số lần chuyển đổi quảng cáo giúp nhà quảng cáo hiểu rõ hơn về khách hàng và điều chỉnh chiến dịch để đạt được kết quả tốt hơn.

Chính vì vậy, với đề án này chúng em thực hiện “Hệ thống quảng cáo thông minh”, với khả năng nhận dạng và tính toán để phát ra một quảng cáo thích hợp nhất.

Nội dung của Đề án tốt nghiệp bao gồm 4 chương:

Chương 1: Giới thiệu tổng quan đề tài

Chương 2: Cơ sở lý thuyết hệ thống quảng cáo thông minh

Chương 3: Thiết kế và đánh giá hệ thống

Chương 4: Kết quả nghiên cứu và đánh giá kết quả

**NHIỆM VỤ ĐỒ ÁN TỐT NGHIỆP**

| STT | Họ và tên sinh viên | Mã số sinh viên | Lớp      | Ngành              |
|-----|---------------------|-----------------|----------|--------------------|
| 1   | Nguyễn Đình Nghĩa   | 106190119       | 19DTCLC3 | Điện tử-viễn thông |
| 2   | Huỳnh Thị Lệ Trinh  | 106190134       | 19DTCLC3 | Điện tử-viễn thông |

- Tên đề tài đồ án: “Hệ thống quảng cáo thông minh”*
- Đề tài thuộc diện: ☐ Có ký kết thỏa thuận sở hữu trí tuệ đối với kết quả thực hiện*
- Các số liệu và dữ liệu ban đầu:*  
*Sử dụng bộ dữ liệu được thu thập bằng cách chụp từ camera điện thoại, máy tính; các trang web chuyên cung cấp cơ sở dữ liệu.*
- Nội dung các phần thuyết minh và tính toán:*

*Phần chung:*

| STT | Họ và tên          | Nội dung                                                                                                                                                                                                                                                          |
|-----|--------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1   | Nguyễn Đình Nghĩa  | <ul style="list-style-type: none"><li>- Tìm hiểu về mô hình huấn luyện YOLOv5</li><li>- Tìm hiểu và tiến hành thực hiện huấn luyện hệ thống với mô hình YOLOv5</li><li>- Tìm hiểu về thuật toán DeepSORT</li><li>- Xây dựng các thuật toán trong đồ án.</li></ul> |
| 2   | Huỳnh Thị Lệ Trinh |                                                                                                                                                                                                                                                                   |

*Phần riêng:*

| STT | Họ và tên         | Nội dung                                                                                                                                                                                                                  |
|-----|-------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1   | Nguyễn Đình Nghĩa | <ul style="list-style-type: none"> <li>- Thu thập cơ sở dữ liệu khuôn mặt nam giới</li> <li>- Áp dụng mô hình huấn luyện YOLOv5 vào hệ thống</li> <li>- Xây dựng các thuật toán thuộc phần nhận dạng đối tượng</li> </ul> |

| STT | Họ và tên          | Nội dung                                                                                                                                                                                                                   |
|-----|--------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 2   | Huỳnh Thị Lệ Trinh | <ul style="list-style-type: none"> <li>- Thu thập cơ sở dữ liệu khuôn mặt nữ giới</li> <li>- Áp dụng thuật toán DeepSORT vào hệ thống</li> <li>- Xây dựng các thuật toán để hiển thị quảng cáo một cách phù hợp</li> </ul> |

5. Các bản vẽ, đồ thị ( ghi rõ các loại và kích thước bản vẽ ):

| STT | Họ và tên          | Nội dung |
|-----|--------------------|----------|
| 1   | Nguyễn Đình Nghĩa  |          |
| 2   | Huỳnh Thị Lệ Trinh |          |

6. Họ và tên người hướng dẫn: TS. Ngô Minh Trí

7. Ngày giao nhiệm vụ đồ án: ...../...../2023

8. Ngày hoàn thành đồ án: ...../...../2023

Đà Nẵng, ngày      tháng      năm 2023

**Trưởng Bộ môn .....**

**Người hướng dẫn**

## LỜI NÓI ĐẦU

Trong suốt thời gian gần 4 năm học tập và rèn luyện tại Trường Đại Học Bách Khoa – Đại học Đà Nẵng, nhóm em đã nhận được rất nhiều sự quan tâm, giúp đỡ của quý thầy cô và bạn bè. Với lòng biết ơn sâu sắc và chân thành nhất, chúng em xin gửi đến quý thầy cô ở Khoa Điện tử - Viễn Thông, trường Đại học Bách Khoa Đà Nẵng đã cùng với tri thức và tâm huyết của mình để truyền đạt vốn kiến thức quý báu cho chúng em trong suốt thời gian học tập tại trường, tạo tiền đề cho nhóm em được thực hiện đồ án này.

Và đặc biệt, những bài học mà thầy cô truyền đạt đã giúp cho chúng em tiếp cận những sự phát triển của khoa học kỹ thuật, biết cách vận dụng lý thuyết vào thực tiễn, có thể tự mình thiết kế ra những hệ thống riêng mang tính cá nhân, nhìn nhận một cách thiết thực và rõ ràng hơn về những gì mình cần phải làm sau khi ra trường. Và bằng tất cả những gì nhóm em học hỏi được cùng với định hướng của thầy Ngô Minh Trí, nhóm em đã thực hiện đề tài “ Hệ thống quảng cáo thông minh”. Nhóm chúng em xin chân thành cảm ơn thầy giáo Ngô Minh Trí đã tận tâm hướng dẫn chúng em từng buổi học trên lớp cũng như những buổi nói chuyện, thảo luận, góp ý, tạo điều kiện tốt nhất cho chúng em để thực hiện đồ án lần này.

Chúng em cũng xin bày tỏ lòng biết ơn đến ban lãnh đạo của Trường Đại học Bách Khoa Đà Nẵng và các khoa phòng ban chức năng đã trực tiếp và gián tiếp giúp đỡ chúng em trong suốt quá trình học tập và nghiên cứu đề tài này. Với điều kiện thời gian cũng như kinh nghiệm còn hạn chế của một sinh viên chuẩn bị bước ra đời, đồ án này không thể tránh khỏi được những thiếu sót. Chúng em rất mong nhận được sự chỉ bảo, đóng góp ý kiến từ các thầy cô giáo để nhóm kịp thời bổ sung, sửa chữa những hạn chế về kiến thức và đặc biệt trang bị hành trang tốt nhất sau khi tốt nghiệp để bước vào một chặng đường mới.

Chúng em xin chân thành cảm ơn.

Đà Nẵng, ngày ... tháng...năm 2023

Sinh viên thực hiện 1

Sinh viên thực hiện 2

**Nguyễn Đình Nghĩa**

**Huỳnh Thị Lệ Trinh**

## **CAM ĐOAN**

Chúng tôi xin cam đoan:

1. Nội dung trong đồ án này là do chúng tôi thực hiện dưới sự hướng dẫn trực tiếp của thầy giáo Ngô Minh Trí.
2. Các tham khảo trong đồ án này đều được trích dẫn rõ ràng tên tác giả, tên công trình, địa điểm công bố.
3. Mọi hình ảnh và video demo trong báo cáo đều do nhóm chúng tôi tự làm, không sao chép toàn bộ từ bất kì nguồn nào. Nếu có những sao chép không hợp lệ, vi phạm liên chính học thuật, nhóm chúng tôi xin chịu hoàn toàn trách nhiệm.

Sinh viên thực hiện 1

Sinh viên thực hiện 2

**Nguyễn Đình Nghĩa**

**Huỳnh Thị Lệ Trinh**

## MỤC LỤC

|                                                     |      |
|-----------------------------------------------------|------|
| TÓM TẮT .....                                       | i    |
| NHIỆM VỤ ĐỒ ÁN TỐT NGHIỆP .....                     | iii  |
| LỜI NÓI ĐẦU .....                                   | v    |
| CAM ĐOAN .....                                      | vi   |
| MỤC LỤC.....                                        | vii  |
| DANH MỤC BẢNG.....                                  | ix   |
| DANH MỤC HÌNH ẢNH .....                             | x    |
| DANH SÁCH CÁC KÝ HIỆU, CHỮ VIẾT TẮT .....           | xiii |
| MỞ ĐẦU.....                                         | 1    |
| CHƯƠNG 1: GIỚI THIỆU TỔNG QUAN ĐỀ TÀI.....          | 2    |
| 1.1. Giới thiệu chương .....                        | 2    |
| 1.2. Tình hình thực tế.....                         | 2    |
| 1.3. Nhiệm vụ đề tài.....                           | 4    |
| 1.4. Thực tiễn.....                                 | 4    |
| 1.5. Ứng dụng của hệ thống.....                     | 4    |
| 1.6. Kết luận chương .....                          | 4    |
| CHƯƠNG 2: CƠ SỞ LÝ THUYẾT HỆ THỐNG.....             | 5    |
| 2.1. Giới thiệu chương .....                        | 5    |
| 2.2. Bài toán phát hiện và nhận diện đối tượng..... | 5    |
| 2.2.1. Khái quát bài toán.....                      | 5    |
| 2.2.2. Các khái niệm trong bài toán.....            | 6    |
| 2.3. Các phương pháp ứng dụng học sâu.....          | 12   |
| 2.3.1. Phương pháp một pha.....                     | 13   |
| 2.3.2. Phương pháp hai pha .....                    | 13   |
| 2.4. Mô hình YOLO .....                             | 14   |
| 2.5. Thuật toán Deepsort.....                       | 18   |
| 2.6. Kết luận chương .....                          | 19   |

|                                                                       |    |
|-----------------------------------------------------------------------|----|
| CHƯƠNG 3: THIẾT KẾ VÀ ĐÁNH GIÁ HỆ THỐNG.....                          | 20 |
| 3.1. Giới thiệu chương .....                                          | 20 |
| 3.2. Quy trình thiết kế hệ thống .....                                | 20 |
| 3.2.1. Khối tiền xử lý.....                                           | 22 |
| 3.2.2. Khối nhận dạng và phân loại đối tượng .....                    | 29 |
| 3.2.3. Khối theo dõi đối tượng .....                                  | 32 |
| 3.2.4. Khối hiển thị .....                                            | 35 |
| 3.3. Kết luận chương .....                                            | 36 |
| CHƯƠNG 4: KẾT QUẢ NGHIÊN CỨU VÀ ĐÁNH GIÁ KẾT QUẢ.....                 | 37 |
| 4.1. Giới thiệu chương .....                                          | 37 |
| 4.2. Kết quả nghiên cứu đạt được.....                                 | 37 |
| 4.2.1. Kết quả mô hình học máy nhận dạng nam nữ.....                  | 37 |
| 4.2.2. Kết quả hiển thị hệ thống quảng cáo thông minh .....           | 42 |
| 4.3. Đánh giá kết quả .....                                           | 46 |
| 4.3.1. Đánh giá kết quả mô hình học máy nhận dạng khuôn mặt nam nữ .. | 46 |
| 4.3.2. Đánh giá kết quả hệ thống quảng cáo thông minh .....           | 47 |
| 4.4. Kết luận chương .....                                            | 47 |
| KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN ĐỀ TÀI .....                             | 48 |
| TÀI LIỆU THAM KHẢO.....                                               | 49 |

## DANH MỤC BẢNG

|                                                 |    |
|-------------------------------------------------|----|
| Bảng 3.1: Số lượng nhãn đã thu thập.....        | 27 |
| Bảng 3.2: Bảng mô hình tóm tắt của YOLOv5s..... | 29 |
| Bảng 4.1: Hiệu suất kết quả mô hình.....        | 40 |

## DANH MỤC HÌNH ẢNH

|                                                                                                              |    |
|--------------------------------------------------------------------------------------------------------------|----|
| Hình 1.1: Tổng chỉ tiêu quảng cáo của thị trường Việt Nam từ 2018 đến 2027 (từ 2023 trở đi là dự đoán) ..... | 2  |
| Hình 1.2: Tổng số người dùng Internet ở Việt Nam vào tháng 01/2023 .....                                     | 3  |
| Hình 2.1: Mối liên hệ giữa các bài toán .....                                                                | 5  |
| Hình 2.2: Bài toán phát hiện đối tượng .....                                                                 | 6  |
| Hình 2.3: Mô phỏng hộp giới hạn .....                                                                        | 7  |
| Hình 2.4: Hộp giới hạn xác định vật thể .....                                                                | 7  |
| Hình 2.5: Các chỉ số IoU tương ứng .....                                                                     | 8  |
| Hình 2.6: Hộp Hộp mỏ neo (Anchor box) .....                                                                  | 9  |
| Hình 2.7: Kiến trúc của một lớp mạng nơ ron tích chập .....                                                  | 11 |
| Hình 2.8: Lớp gộp (pooling layer) .....                                                                      | 12 |
| Hình 2.9: Hình ảnh minh họa một trình tự CNN .....                                                           | 12 |
| Hình 2.10: Mô phỏng phương pháp một pha trong Object Detection .....                                         | 13 |
| Hình 2.11: Mô phỏng phương pháp hai pha trong Object Detection .....                                         | 14 |
| Hình 2.12: Kiến trúc mô hình YOLO .....                                                                      | 15 |
| Hình 2.13: Mô phỏng quá trình xử lý của YOLO .....                                                           | 15 |
| Hình 2.14: Mô tả quy trình phát hiện đối tượng YOLO .....                                                    | 16 |
| Hình 2.15: Nhận diện hình ảnh trên nhiều feature map của YOLO .....                                          | 16 |
| Hình 2.16: Mô phỏng các hộp neo (anchor box) .....                                                           | 17 |
| Hình 3.1: Sơ đồ khối của bộ xử lý trung tâm .....                                                            | 20 |
| Hình 3.2: Lưu đồ thuật toán của hệ thống .....                                                               | 21 |
| Hình 3.3: Tổng quan sơ đồ khối tiền xử lý .....                                                              | 22 |
| Hình 3.4: Hình ảnh tập dữ liệu dùng để huấn luyện .....                                                      | 23 |
| Hình 3.5: Hình ảnh tập dữ liệu dùng để kiểm tra .....                                                        | 23 |
| Hình 3.6: Nội dung nhãn cho một hình ảnh .....                                                               | 24 |
| Hình 3.7: Mô tả thông số cho nhãn dữ liệu .....                                                              | 24 |
| Hình 3.8: Hình ảnh sử dụng labelImg để gán nhãn .....                                                        | 25 |

|                                                                                                                                                                                                    |    |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| Hình 3.9: Hình ảnh thông số của nhân sau khi gán.....                                                                                                                                              | 26 |
| Hình 3.10: Cách tính các thông số của hộp biên.....                                                                                                                                                | 27 |
| Hình 3.11: Số lượng và vị trí thể hiện mỗi lớp trong tập huấn luyện .....                                                                                                                          | 28 |
| Hình 3.12: Biểu đồ mô tả vị trí và kích thước các lớp trong tập huấn luyện.....                                                                                                                    | 28 |
| Hình 3.13: Tổng quan sơ đồ khối nhận dạng và phân loại đối tượng.....                                                                                                                              | 29 |
| Hình 3.14: Tổng quan khối theo dõi đối tượng.....                                                                                                                                                  | 32 |
| Hình 3.15: Hệ thống Real-time Multi-person Tracking với YOLO và DeepSORT.....                                                                                                                      | 33 |
| Hình 3.16: Luồng xử lý của DeepSORT.....                                                                                                                                                           | 33 |
| Hình 3.17: Quá trình theo dõi đối tượng của thuật toán DeepSORT.....                                                                                                                               | 34 |
| Hình 3.18: Tổng quan khối hiển thị .....                                                                                                                                                           | 35 |
| Hình 4.1: Các đồ thị Loss Function của mô hình YOLOv5 trên tập dữ liệu dùng để huấn luyện với 6691 bức ảnh (Train Set).....                                                                        | 37 |
| Hình 4.2: Các đồ thị Loss Function của mô hình YOLOv5 trên tập dữ liệu dùng để kiểm tra với 1400 bức ảnh (Validation Set).....                                                                     | 38 |
| Hình 4.3: Đồ thị tỷ lệ giữa số lượng các đối tượng dự đoán đúng (true positives) và tổng số lượng các đối tượng dự đoán là positive (bao gồm cả true positives và false positives) .....           | 38 |
| Hình 4.4: Đồ thị tỷ lệ giữa số lượng các đối tượng dự đoán đúng (true positives) và tổng số lượng các đối tượng đối tượng thực tế là positive (bao gồm cả true positives và false negatives) ..... | 39 |
| Hình 4.5: Biểu đồ đánh giá hiệu suất tổng thể của mô hình phát hiện đối tượng với ngưỡng IoU là 0.5 .....                                                                                          | 39 |
| Hình 4.6: Biểu đồ đánh giá hiệu suất tổng thể của mô hình phát hiện đối tượng trong khoảng ngưỡng IoU từ 0.5 đến 0.95 .....                                                                        | 40 |
| Hình 4.7: Kết quả nhận dạng của mô hình YOLOv5s .....                                                                                                                                              | 41 |
| Hình 4.8: Kết quả hiển thị quảng cáo dành cho nữ giới .....                                                                                                                                        | 42 |
| Hình 4.9: Kết quả hiển thị quảng cáo dành cho nam giới.....                                                                                                                                        | 42 |
| Hình 4.10: Kết quả hiển thị quảng cáo ngẫu nhiên khi số lượng nam nữ bằng nhau .....                                                                                                               | 43 |
| Hình 4.11: Kết quả hiển thị quảng cáo ngẫu nhiên khi số lượng nam nữ bằng nhau .....                                                                                                               | 43 |

|                                                                                   |    |
|-----------------------------------------------------------------------------------|----|
| Hình 4.12: Kết quả hiển thị quảng cáo khi có nhiều đối tượng trong khung hình.... | 44 |
| Hình 4.13: Kết quả hiển thị quảng cáo dành cho nam giới trong thực tế .....       | 44 |
| Hình 4.14: Kết quả hiển thị quảng cáo dành cho nữ giới trong thực tế .....        | 45 |
| Hình 4.15: Kết quả hiển thị quảng cáo ngẫu nhiên trong thực tế .....              | 45 |
| Hình 4.16: Ma trận nhầm lẫn mô hình YOLOv5s .....                                 | 46 |

## DANH SÁCH CÁC KÝ HIỆU, CHỮ VIẾT TẮT

| STT | Ký hiệu chữ viết tắt | Chữ viết đầy đủ                                                                                                    |
|-----|----------------------|--------------------------------------------------------------------------------------------------------------------|
| 1   | CNN                  | Convolutional Neural Network hay là mạng nơ ron tích chập.                                                         |
| 2   | AI                   | Artificial Intelligence hay là trí tuệ nhân tạo.                                                                   |
| 3   | YOLO                 | “You only look once” là mô hình nhận dạng thuộc phương pháp một pha                                                |
| 4   | IOU                  | INTERSECTION OVER UNION là hàm đánh giá độ chính xác của object detector trên tập dữ liệu cụ thể                   |
| 5   | DeepSORT             | Deep Learning for Real-Time Object Tracking                                                                        |
| 6   | SSD                  | Single Shot Multibox Detector là một mô hình nhận dạng vật thể trong hình ảnh                                      |
| 7   | GFLOPs               | Giga Floating-point Operations Per second là số lượng phép tính lượng tử động thực hiện bởi mô hình trong mỗi giây |
| 8   | FMCG                 | Fast-Moving Consumer Goods là nhóm hàng tiêu dùng nhanh                                                            |

## MỞ ĐẦU

Trí tuệ nhân tạo (Artificial Intelligence - AI) đã và đang thay đổi cách chúng ta sống và làm việc. Từ việc hỗ trợ tự động hóa các nhiệm vụ phức tạp đến việc cải thiện trải nghiệm người dùng, AI đã trở thành một công cụ quan trọng trong nhiều lĩnh vực, bao gồm cả lĩnh vực quảng cáo. Quảng cáo là một phần không thể thiếu trong xã hội hiện đại, nơi thông tin được truyền tải đến các khách hàng tiềm năng. Tuy nhiên, quảng cáo truyền thống thường gặp phải nhiều thách thức, bao gồm sự phức tạp trong việc xác định mục tiêu khách hàng, tối ưu hóa chiến dịch quảng cáo và tăng cường tương tác với khách hàng.

Vì vậy, chúng em đã quyết định lựa chọn xây dựng hệ thống “Quảng cáo thông minh” bằng cách sử dụng các thuật toán và mô hình AI, chúng ta có thể phân tích dữ liệu khách hàng, dự đoán hành vi tiêu dùng và tạo ra những chiến dịch quảng cáo tối ưu hóa hơn, tăng cường tương tác với khách hàng và tạo ra trải nghiệm người dùng tốt hơn.

Có rất nhiều phương pháp để thực hiện nên hệ thống, tuy nhiên, việc sử dụng kiến trúc YOLOv5, đáp ứng được mục tiêu khả năng nhận dạng đa đối tượng nhanh chóng và chính xác trên các hình ảnh và video mà đề tài đặt ra. Chương 2 sẽ trình bày cơ sở lý thuyết mô hình, thuật toán sử dụng trong đồ án. Sau đó mô hình sẽ được xây dựng và đánh giá định lượng và định tính trong chương 3,4.

Nội dung của Đồ án tốt nghiệp bao gồm 4 chương:

Chương 1: Giới thiệu tổng quan đề tài

Chương 2: Cơ sở lý thuyết hệ thống

Chương 3: Thiết kế và đánh giá hệ thống

Chương 4: Kết quả nghiên cứu và đánh giá kết quả

## CHƯƠNG 1: GIỚI THIỆU TỔNG QUAN ĐỀ TÀI

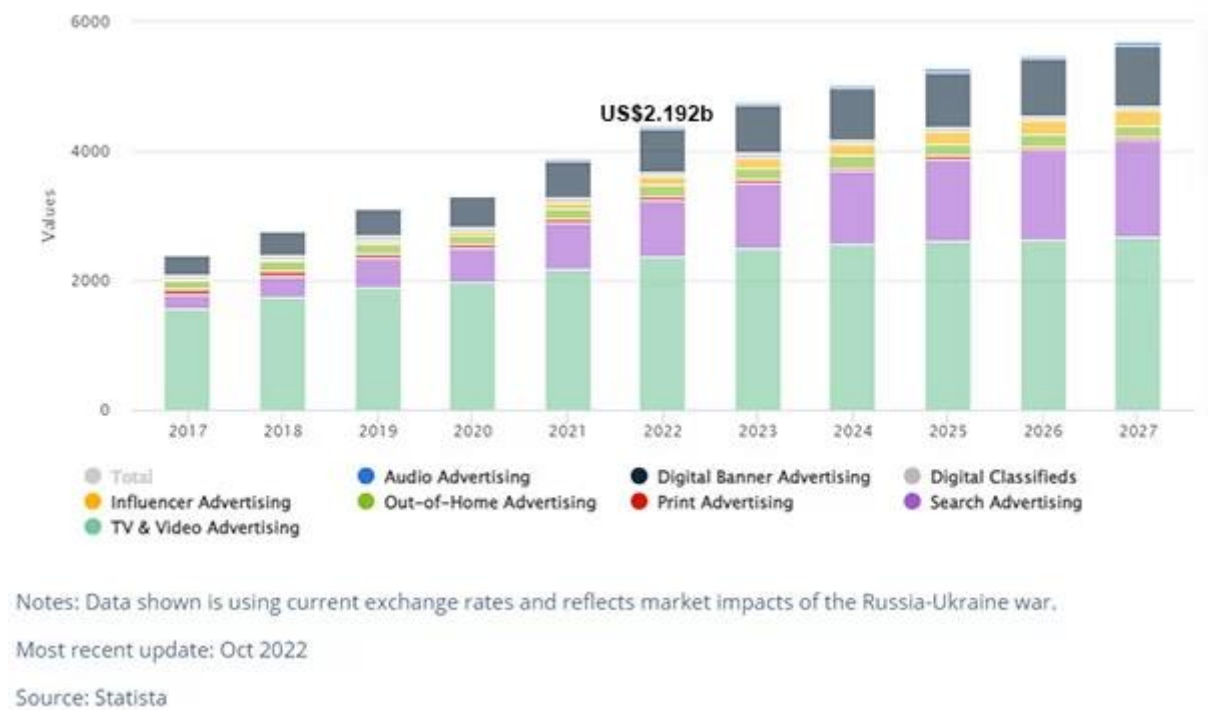
### 1.1. Giới thiệu chương

Trong chương này, mục tiêu đề tài sẽ được nêu ra để thấy được tầm quan trọng của sản phẩm mà đề án này nghiên cứu. Từ mục tiêu ấy, đặt ra được nhiệm vụ mà nhóm chúng em cần phải thực hiện tại đề án này. Bên cạnh đó, chương này cũng đưa ra một số ví dụ về ứng dụng của đề án cũng như là thực tiễn mà đề án mang lại.

### 1.2. Tình hình thực tế

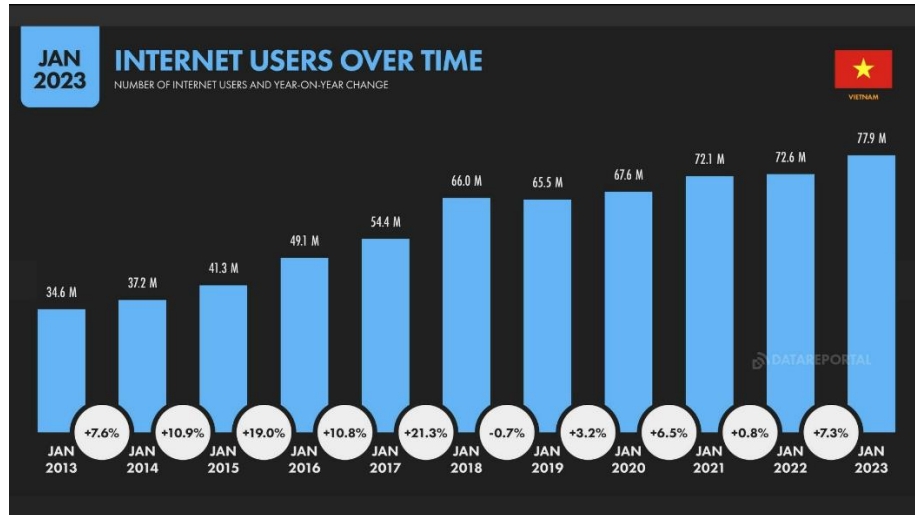
Trong những năm gần đây, dịch vụ quảng cáo tại Việt Nam đã có sự chuyển dịch từ quảng cáo truyền thống sang quảng cáo trực tuyến. Với sự phát triển của công nghệ thông tin và internet, người dân Việt Nam đã trở nên ngày càng kết nối mạnh mẽ với thế giới số. Điều này đã tạo ra một cơ hội lớn cho các công ty quảng cáo để tiếp cận khách hàng tiềm năng thông qua các nền tảng trực tuyến.

Theo số liệu của Statista tính đến tháng 10/2022, tổng doanh thu thị trường quảng cáo Việt Nam dự kiến đạt 2.192 tỷ USD vào năm 2022, tăng 12.7% so với năm 2021, trong đó, phân khúc lớn nhất của thị trường là Quảng cáo TV & Video với khối lượng thị trường dự kiến là 1.19 tỷ USD [Hình 1.1].



Hình 1.1: Tổng chi tiêu quảng cáo của thị trường Việt Nam từ 2018 đến 2027  
(từ 2023 trở đi là dự đoán)

Trong lĩnh vực quảng cáo trực tuyến, các nền tảng quảng cáo như Google Ads, Facebook Ads, và Zalo Ads đã trở thành những công cụ phổ biến để tiếp cận khách hàng. Theo thống kê, Việt Nam có hơn 77 triệu người dùng internet [Hình 1.2] và hơn 76 triệu người dùng mạng xã hội, tạo điều kiện thuận lợi cho các công ty quảng cáo định hình chiến lược quảng cáo trực tuyến.



Hình 1.2: Tổng số người dùng Internet ở Việt Nam vào tháng 01/2023

Các ngành công nghiệp chủ chốt đầu tư nhiều vào quảng cáo bao gồm FMCG (Fast-Moving Consumer Goods), ngân hàng, bất động sản và du lịch. Đây là các ngành có nhu cầu tiếp cận khách hàng lớn và cạnh tranh cao, do đó việc đầu tư vào quảng cáo giúp tăng cường nhận diện thương hiệu và tạo điểm khác biệt trong thị trường cạnh tranh.

Tóm lại, dịch vụ quảng cáo tại Việt Nam đang trên đà phát triển mạnh mẽ. Sự chuyển dịch từ quảng cáo truyền thống sang quảng cáo trực tuyến cùng với tăng trưởng kinh tế và sự phổ biến của internet đã tạo ra một môi trường thuận lợi cho việc tiếp cận và tương tác với khách hàng tiềm năng. Số liệu thực tế chỉ ra rằng các công ty và thương hiệu đang đầu tư ngày càng nhiều vào quảng cáo để tăng cường sự hiện diện và tạo ấn tượng trong lòng người tiêu dùng tại Việt Nam.

Chính vì vậy, trong thời đại 4.0 hiện nay cùng với sự phát triển mạnh của mọi lĩnh vực, ngành nghề đều chuyển dịch theo hướng khai thác và sử dụng các ứng dụng công nghệ thông tin. Xuất hiện như một tất yếu khách quan, quảng cáo thông minh ứng dụng AI sẽ là xu hướng tiếp theo của các doanh nghiệp trong thời buổi hiện đại. Điều này đã thúc đẩy chúng em thực hiện đề tài “Hệ thống quảng cáo thông minh”.

### **1.3. Nhiệm vụ đề tài**

Để thực hiện được mục tiêu trên, đề tài này sẽ nghiên cứu về phương pháp phát hiện và nhận dạng khuôn mặt của đối tượng, phương pháp để phân loại dựa trên giới tính của đối tượng từ đó đưa ra quảng cáo sản phẩm thích hợp với đối tượng. Phát triển bằng ngôn ngữ lập trình Python tích hợp trên các nền tảng thư viện đi kèm để xây dựng hệ thống.

Hệ thống được hoàn thành phải đảm bảo được các yêu cầu:

- Nhận dạng được nam và nữ trong khung hình
- So sánh và hiển thị số lượng nam nữ
- Hệ thống lấy dữ liệu và trả kết quả để hiển thị quảng cáo trong thời gian nhanh nhất

### **1.4. Thực tiễn**

Nghiên cứu và xây dựng một chương trình có thể phát hiện được giới tính và số lượng của đối tượng đứng trước camera, từ đó hiển thị quảng cáo phù hợp với các đối tượng. Điều này đem lại sự tối ưu về việc hiển thị quảng cáo cho từng đối tượng hay một nhóm đối tượng thay vì chiếu thủ công một cách ngẫu nhiên trên màn hình.

### **1.5. Ứng dụng của hệ thống**

Hệ thống được xây dựng nhằm mục đích đưa ra các sản phẩm thích hợp với người sử dụng và có thể ứng dụng trong việc phát triển, tăng số lượng khách hàng mua sản phẩm.

Một số nơi mà hệ thống có thể được đặt:

- Sân bay
- Trung tâm thương mại
- Rạp chiếu phim
- Khu phố đi bộ

### **1.6. Kết luận chương**

Kết thúc chương 1 đã trình bày được mục tiêu cũng như nhiệm vụ mà đồ án này hướng đến. Chúng ta đã biết được ứng dụng mà hệ thống này sẽ đem lại sau khi được hoàn thành. Để hiểu hơn về lý thuyết cũng như quá trình xây dựng hệ thống, chúng ta sẽ đến với chương tiếp theo.

## CHƯƠNG 2: CƠ SỞ LÝ THUYẾT HỆ THỐNG

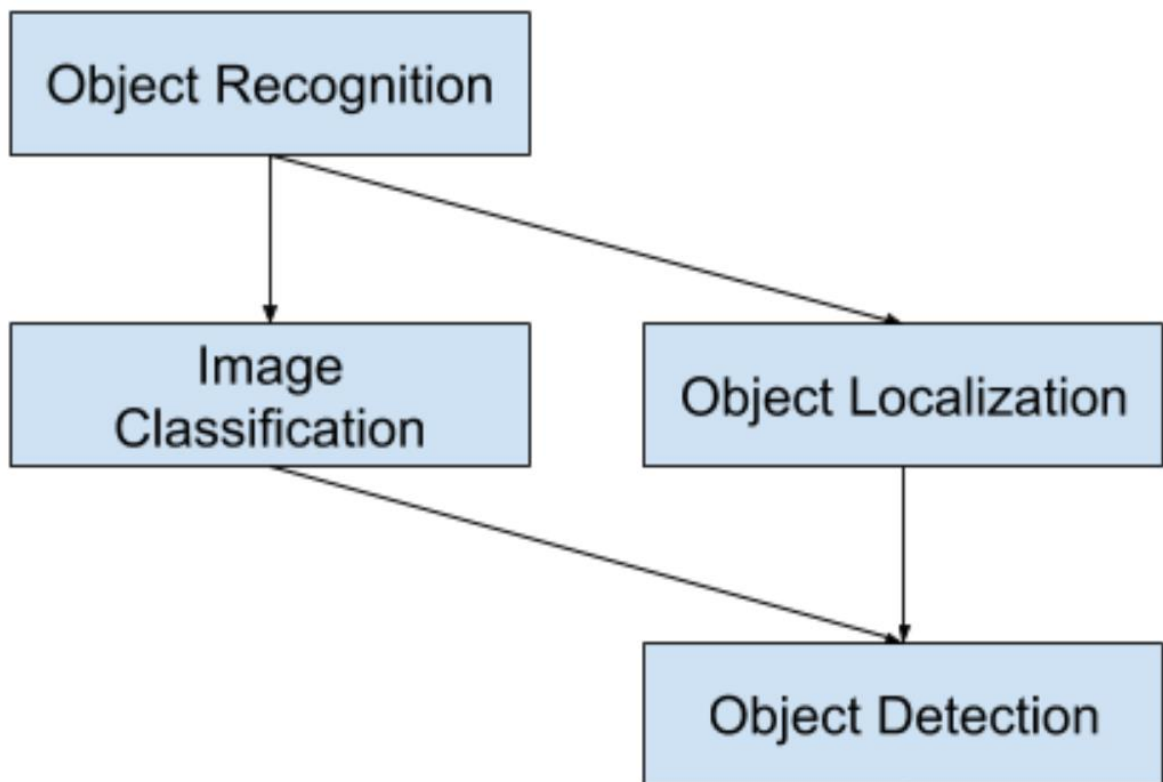
### 2.1. Giới thiệu chương

Chương này trình bày về lý thuyết phát hiện giới tính bằng cách nhận dạng khuôn mặt, mạng nơ ron tích chập và việc lựa chọn kiến trúc mạng nơ ron tích chập phù hợp cho hệ thống được đề cập. Chương này cũng trình bày được những ưu điểm nổi trội từ việc lựa chọn mô hình YOLOv5 và thuật toán Deepsort được sử dụng trong đề án này.

### 2.2. Bài toán phát hiện và nhận diện đối tượng

#### 2.2.1. Khái quát bài toán

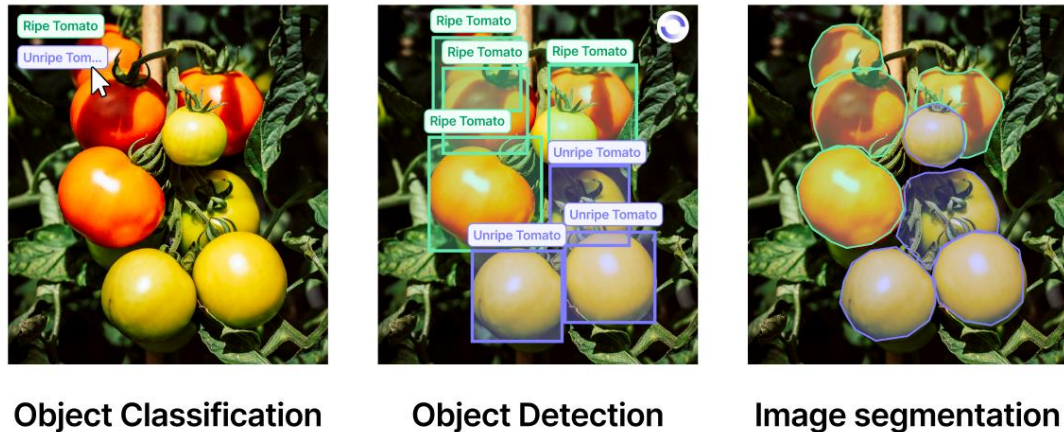
Để tìm hiểu về bài toán phát hiện và nhận diện đối tượng (object detection) trước tiên ta cần hiểu về 2 bài toán nhỏ hơn là:



Hình 2.1: Mối liên hệ giữa các bài toán

- Phân loại hình ảnh (image classification): là một bài toán học có giám sát, xác định một tập hợp các lớp mục tiêu (các đối tượng để xác định trong hình ảnh) và đào tạo một mô hình để nhận ra chúng bằng cách sử dụng các ảnh mẫu được gắn nhãn [Hình 2.1].

- Định vị vật thể (object localization): là bài toán định vị một thể hiện của một loại đối tượng cụ thể trong một hình ảnh, thường bằng cách chỉ định một hộp giới hạn (bounding box) xung quanh thể nhằm khoanh vùng đối tượng.



ENCORD

Hình 2.2: Bài toán phát hiện đối tượng

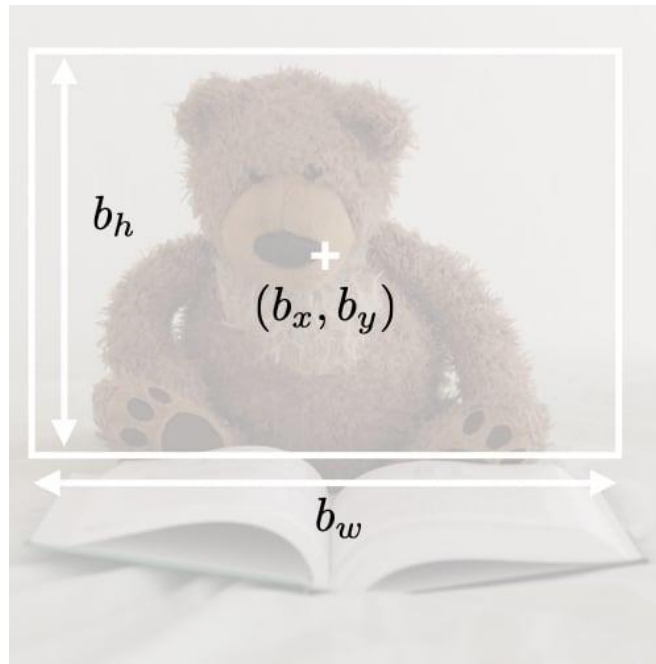
Bài toán phát hiện đối tượng (Object detection) là: quá trình phức tạp hơn, bao gồm cả việc định vị chúng trong ảnh bằng cách vẽ một hộp giới hạn (bounding box) xung quanh từng đối tượng và phân loại đối tượng ấy [Hình 2.2].

### 2.2.2. Các khái niệm trong bài toán

#### a. Hộp giới hạn:

Hộp giới hạn (Bounding box) [Hình 2.3] là một khái niệm trong lĩnh vực thị giác máy tính và trí tuệ nhân tạo. Nó được sử dụng để xác định vị trí và định hình của một đối tượng trong hình ảnh hoặc video. Hộp giới hạn là một hình chữ nhật hay hình hộp bao quanh đối tượng và được xác định bằng cách xác định tọa độ của hai điểm đối diện, thường là tọa độ góc trái trên và tọa độ góc phải dưới.

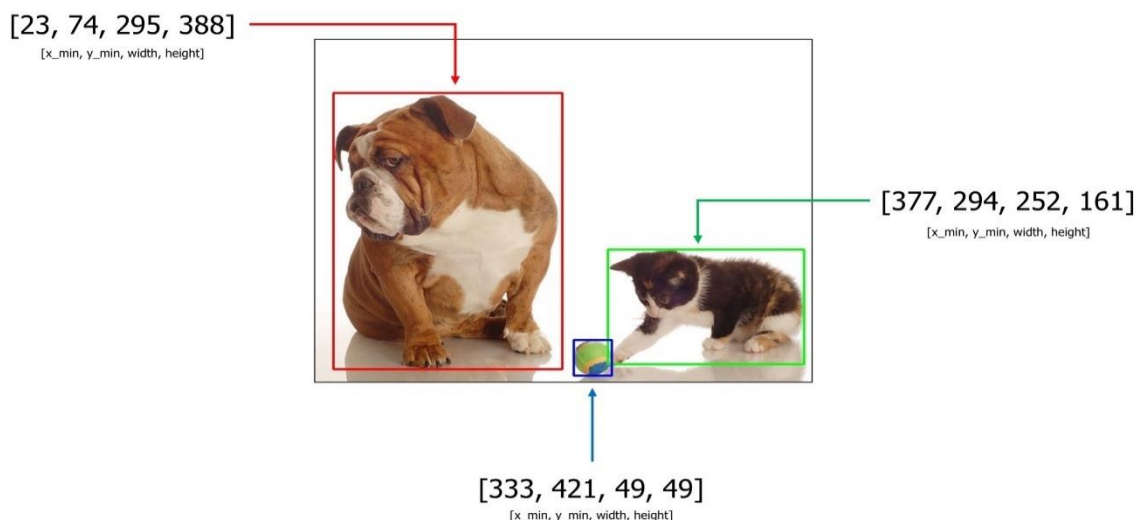
Hộp giới hạn cung cấp thông tin quan trọng về vị trí và kích thước của đối tượng trong hình ảnh. Nó thường được sử dụng trong các nhiệm vụ như phát hiện đối tượng, nhận dạng đối tượng, và theo dõi đối tượng trong thời gian thực. Đối với mỗi đối tượng trong hình ảnh, có thể có nhiều hộp giới hạn để mô tả các đặc trưng khác nhau của đối tượng.



Hình 2.3: Mô phỏng hộp giới hạn

Hộp giới hạn thường được biểu diễn bằng cách sử dụng tọa độ (x, y) của góc trái trên và (width, height) của hình chữ nhật hoặc tọa độ (x, y) của hai góc đối diện và kích thước (width, height) của hình hộp. Thông tin từ hộp giới hạn có thể được sử dụng để xác định vị trí, phân loại và tương tác với các đối tượng trong hình ảnh hoặc video.

Trong một hình ảnh, bounding box là tọa độ (x, y) của một hình chữ nhật xung quanh một khu vực quan tâm, chẳng hạn như con chó trong hình ảnh bên dưới [Hình 2.4].



Hình 2.4: Hộp giới hạn xác định vật thể

b. IoU (Intersection over Union):

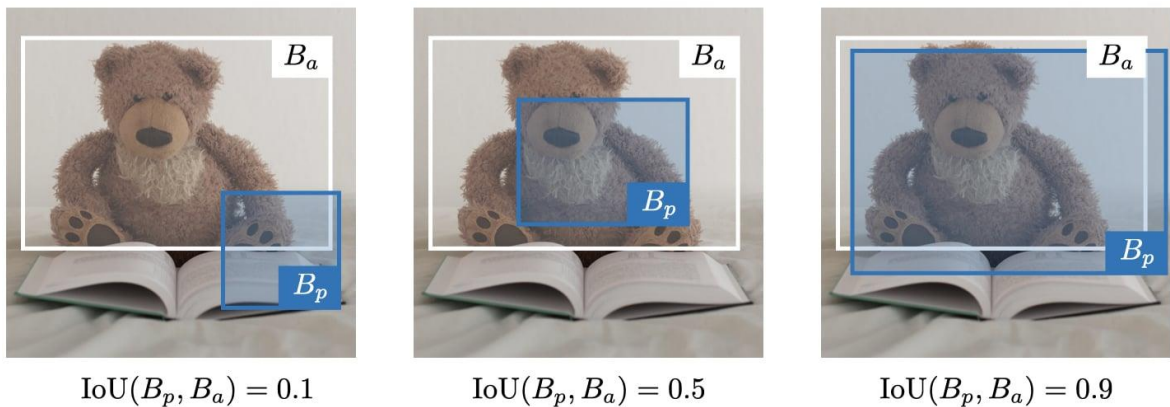
IoU (Intersection over Union) [Hình 2.5] là một chỉ số định lượng sự trùng lặp giữa hai hộp giới hạn, thường được sử dụng để đánh giá hiệu suất của các thuật toán phát hiện đối tượng trong lĩnh vực thị giác máy tính và trí tuệ nhân tạo.

IoU được tính bằng cách lấy diện tích của phần giao giữa hai hộp giới hạn và chia cho diện tích của phần hợp của chúng.

Một hàm định lượng vị trí  $B_p$  của hộp giới hạn dự đoán được định vị đúng như thế nào so với hộp giới hạn thực tế  $B_a$ . Nó được định nghĩa:

$$\text{IoU}(B_p, B_a) = \frac{B_p \cap B_a}{B_p \cup B_a} \quad (1)$$

Giá trị IoU nằm trong khoảng từ 0 đến 1. Khi IoU gần bằng 1, tức là hai hộp giới hạn trùng lặp hoàn toàn. Ngược lại, khi IoU gần bằng 0, tức là không có sự trùng lặp giữa hai hộp giới hạn.

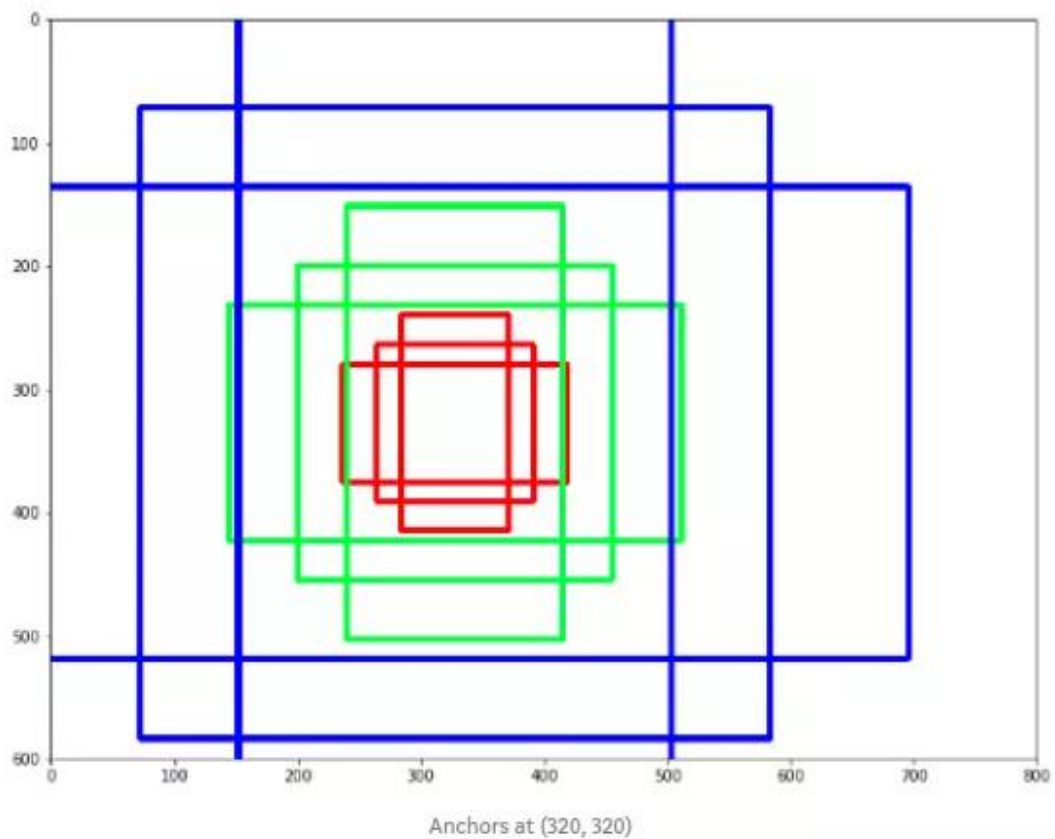


Hình 2.5: Các chỉ số IoU tương ứng

IoU thường được sử dụng để đánh giá chất lượng của việc phát hiện đối tượng, trong đó một giá trị IoU ngưỡng được sử dụng để quyết định xem một hộp giới hạn dự đoán có được coi là chính xác hay không.

c. Hộp mỏ neo (Anchor boxes)

Hộp mỏ neo (Anchor box) [Hình 2.6] là một phần tử trong kỹ thuật mạng nơ-ron tích chập (Convolutional Neural Network - CNN) được sử dụng trong các mô hình phát hiện đối tượng, như Faster R-CNN hoặc YOLO (You Only Look Once).



Hình 2.6: Hộp Hộp mỏ neo (Anchor box)

Ý tưởng của hộp mỏ neo là tạo ra một tập các hộp giới hạn có kích thước và tỷ lệ khác nhau trên không gian đầu vào. Mỗi hộp mỏ neo đại diện cho một kỳ vọng về hình dạng và kích thước của các đối tượng trong hình ảnh. Thay vì dự đoán trực tiếp các hộp giới hạn tùy ý, mô hình chỉ cần dự đoán sự thay đổi và điều chỉnh so với các hộp mỏ neo.

Bằng cách sử dụng hộp mỏ neo, mạng có thể dự đoán nhiều hộp giới hạn cùng một lúc, mỗi hộp giới hạn được ràng buộc bởi các tính chất hình học đã được xác định trước. Ví dụ, mạng có thể được huấn luyện để dự đoán đầu ra cho mỗi hộp mỏ neo gồm các thông tin như tọa độ x, y, chiều rộng, chiều cao và xác suất của một đối tượng xuất hiện trong hộp đó.

Qua đó, hộp mỏ neo giúp tăng tính linh hoạt và đa dạng trong việc dự đoán các hộp giới hạn, cho phép mô hình có khả năng phát hiện đối tượng ở các kích thước và tỷ lệ khác nhau trong một hình ảnh, cũng như xử lý các trường hợp khi các đối tượng chồng lấn lên nhau.

d. Bản đồ đặc trưng (Feature Map):

Bản đồ đặc trưng (Feature map) là kết quả của quá trình trích xuất đặc trưng từ ảnh bằng các mạng nơ-ron tích chập (Convolutional Neural Networks - CNN) trong lĩnh vực thị giác máy tính và trí tuệ nhân tạo.

Feature map là một tập hợp các ma trận hai chiều, nơi mỗi giá trị trong ma trận đại diện cho một đặc trưng cụ thể trong ảnh. Mỗi feature map tương ứng với một kênh hoặc một bộ lọc (filter) được áp dụng trên ảnh đầu vào. Kết quả của việc áp dụng bộ lọc này là một feature map.

Quá trình trích xuất đặc trưng bằng CNN bao gồm các bước liên tiếp của các lớp tích chập và lớp pooling. Các lớp tích chập áp dụng các bộ lọc để tìm kiếm các đặc trưng cục bộ trong ảnh, trong khi các lớp pooling giúp giảm kích thước không gian của feature map và làm nổi bật các đặc trưng quan trọng.

Feature map ở các lớp sâu của CNN thường bao gồm các đặc trưng trừu tượng và ngày càng trừu tượng hơn so với các feature map ở các lớp đầu vào. Điều này cho phép mô hình học các đặc trưng phức tạp và trừu tượng từ ảnh.

Feature map cuối cùng của mạng CNN thường được sử dụng để dự đoán và phân loại các đối tượng trong ảnh, trong khi các feature map ở các lớp trước có thể chứa các đặc trưng cụ thể như cạnh, góc, hoặc textural của ảnh.

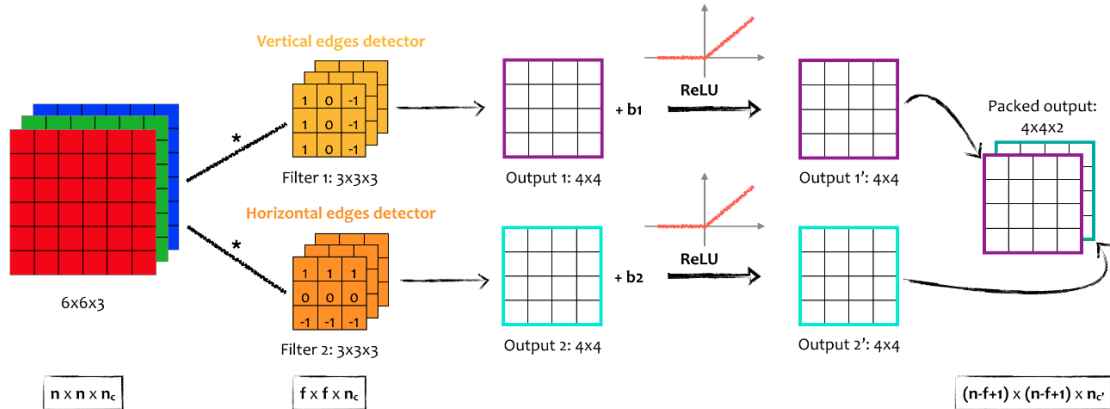
Tổng quan, feature map là kết quả của quá trình trích xuất đặc trưng từ ảnh thông qua mạng CNN và chứa thông tin về các đặc trưng quan trọng trong ảnh được sử dụng cho các tác vụ như phân loại, phát hiện, và nhận dạng đối tượng.

e. Mạng nơ-ron tích chập

Mạng nơ-ron tích chập (Convolutional Neural Network - CNN)[1] là một trong những mô hình mạng nơ-ron phổ biến nhất hiện nay trong lĩnh vực xử lý ảnh và âm thanh. Điều này bởi vì mạng CNN có thể tự động học các đặc trưng cấp thấp và cấp cao từ dữ liệu đầu vào, giúp cho việc xử lý dữ liệu trở nên nhanh chóng và chính xác hơn. Với kiến trúc đặc biệt được thiết kế dựa trên các lớp tích chập, CNN đã đạt được rất nhiều thành công trong các lĩnh vực như phân loại đối tượng, nhận dạng khuôn mặt, xử lý ngôn ngữ tự nhiên và rất nhiều ứng dụng khác.

Các lớp tích chập trong mạng CNN có nhiệm vụ trích xuất các đặc trưng cấp thấp và cấp cao từ dữ liệu đầu vào. Với ảnh làm đầu vào, các lớp tích chập được sử dụng để quét qua toàn bộ ảnh và tìm kiếm các đặc trưng như các cạnh, đường nét và

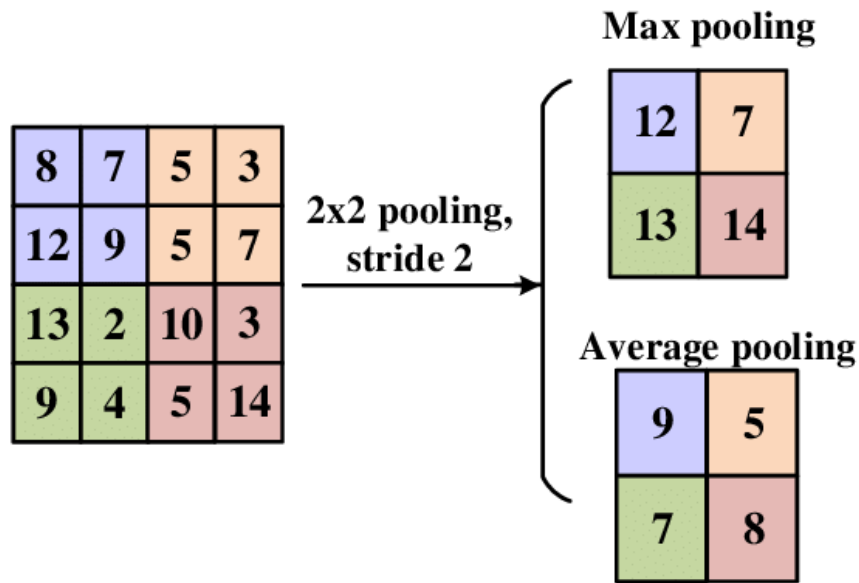
hình dạng đối tượng. Mỗi lớp tích chập [Hình 2.7] sử dụng một tập hợp các bộ lọc (filter) để quét qua ảnh đầu vào và tìm kiếm các đặc trưng đặc biệt. Các bộ lọc này có thể có kích thước và hình dạng khác nhau, tùy thuộc vào mục đích của mô hình.



Hình 2.7: Kiến trúc của một lớp mạng nơ ron tích chập

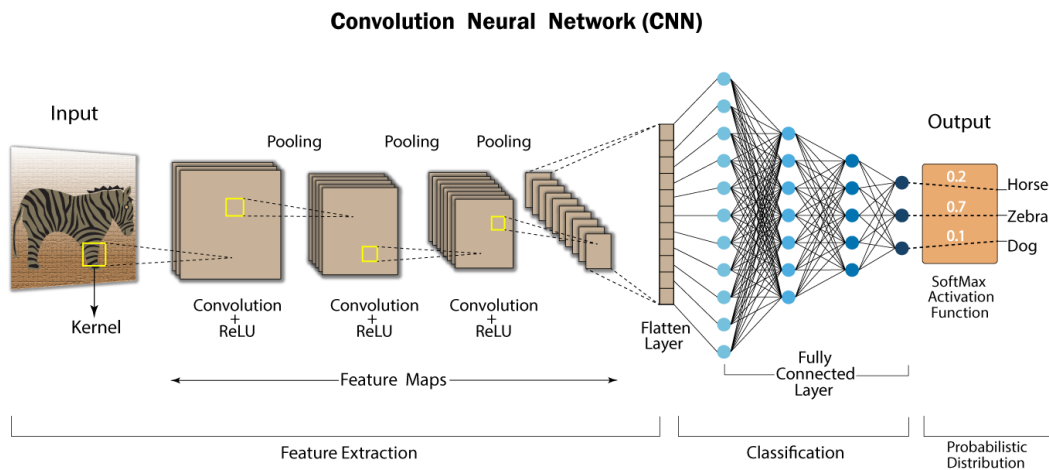
Khi quét qua ảnh đầu vào, các bộ lọc này sẽ tạo ra một ma trận kết quả gọi là feature map [Hình 2.9], trong đó mỗi phần tử của feature map là kết quả của phép tính tích chập giữa bộ lọc và vùng của ảnh được quét qua. Sau đó, các feature map này được đưa vào các lớp kích hoạt (activation layer) để tạo ra các đặc trưng phi tuyến tính.

Các lớp tích chập có thể được xếp chồng lên nhau để tăng độ sâu của mô hình và cũng có thể được kết hợp với các lớp khác như lớp gộp (pooling layer) [Hình 2.8] hoặc lớp chuẩn hóa (normalization layer) để giảm chiều dữ liệu đầu vào và cải thiện độ chính xác của mô hình.



Hình 2.8: Lớp gộp (pooling layer)

Sau đó, các đặc trưng được đưa vào các lớp kết nối đầy đủ (fully connected layer) để phân loại hoặc dự đoán đầu ra. Các lớp kết nối đầy đủ này tương tự như trong mạng nơ-ron thần kinh thông thường, với mỗi nơ-ron kết nối với tất cả các đặc trưng từ các lớp tích chập trước đó. Sau đó, một lớp đầu ra được sử dụng để tính toán xác suất của các lớp phân loại khác nhau, hoặc giá trị của đầu ra cần dự đoán.



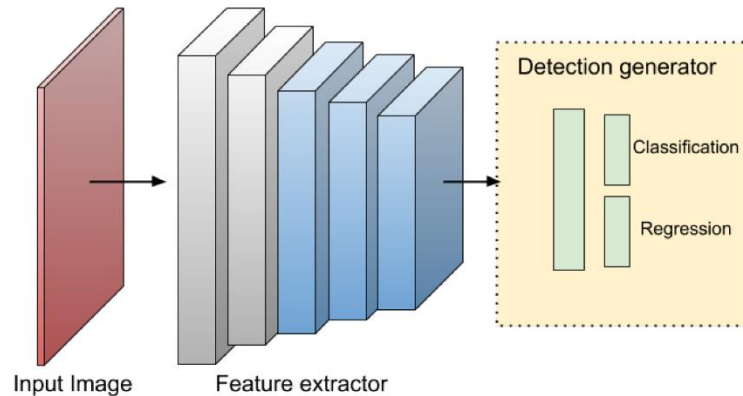
Hình 2.9: Hình ảnh minh họa một trình tự CNN

### 2.3. Các phương pháp ứng dụng học sâu

Hệ thống quảng cáo thông minh có thể được xem là một vấn đề thuộc bài toán phát hiện đối tượng (Object Detection). Các mô hình học sâu có khả năng trích xuất các đặc trưng mạnh mẽ, có lợi thế đáng kể so với các phương pháp truyền thống trước đây.

Nhìn chung đặc điểm cấu trúc của các thuật toán phát hiện đối tượng (Object Detection) hiện nay được chia làm hai nhóm chính là: phương pháp một pha và phương pháp hai pha.

### **2.3.1. Phương pháp một pha**



Hình 2.10: Mô phỏng phương pháp một pha trong Object Detection

Phương pháp một pha (one-stage) [Hình 2.10] là một phương pháp tiên tiến trong phát hiện đối tượng, khác với phương pháp hai pha (two-stage) trong đó các vùng đối tượng được đề xuất trước đó trước khi thực hiện việc phân loại và vị trí hóa đối tượng. Khác với phương pháp hai pha thay vì có một bước tìm ra các vùng có khả năng tồn tại đối tượng thì trong phương pháp một pha, mô hình phát hiện đối tượng chỉ sử dụng một lần đưa toàn bộ ảnh qua mạng nơ-ron tích chập. Mô hình sẽ cùng lúc thực hiện việc phân loại và xác định vị trí các đối tượng trên ảnh.

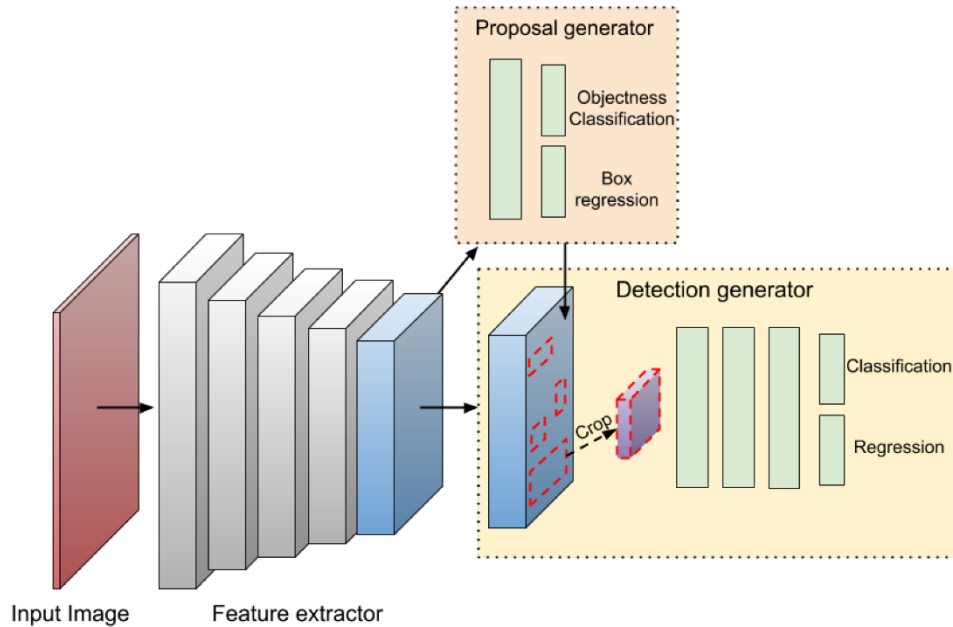
Các mô hình one-stage thường được thiết kế để có tốc độ xử lý nhanh và độ chính xác cao, cho phép phát hiện đối tượng trong thời gian thực. Một số mô hình one-stage phổ biến trong phát hiện đối tượng là: SSD (Single Shot MultiBox Detector), RetinaNet và điển hình là YOLO. Đây cũng là phương pháp được sử dụng trong việc nhận dạng khuôn mặt để phân loại giới tính của hệ thống quảng cáo thông minh.

### **2.3.2. Phương pháp hai pha**

Phương pháp hai pha (two-stage) [Hình 2.11] là phương pháp giải quyết bài toán phát hiện đối tượng qua hai giai đoạn (hai pha):

- Pha đầu tiên dùng để xác định các vùng có tiềm năng là đối tượng.

- Pha thứ hai dùng để phân loại và cải tiến hộp giới hạn (bounding box) cho đối tượng.



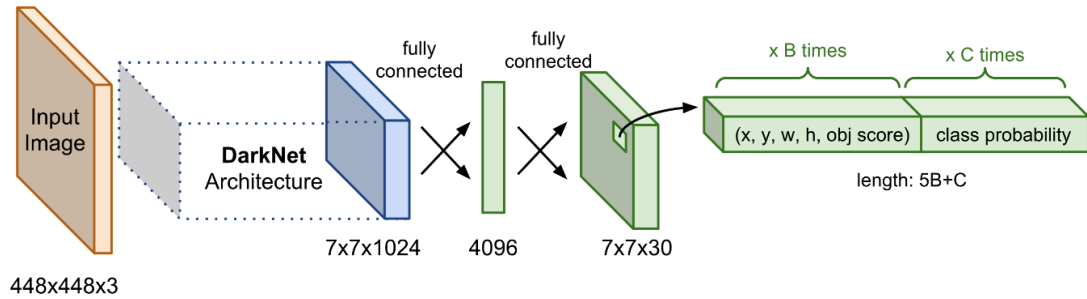
Hình 2.11: Mô phỏng phương pháp hai pha trong Object Detection

Hình ảnh dữ liệu đầu vào sẽ được trích xuất đặc trưng, xác định các vùng tiềm năng trong hình ảnh có thể chứa đối tượng. Mục tiêu là tìm ra những vùng quan trọng có khả năng chứa đối tượng dựa trên các đặc trưng như cạnh, góc, màu sắc, hoặc kết hợp các đặc trưng khác.

Sau khi có được các vùng đề xuất, sẽ tiến hành việc phân loại và định vị đối tượng trong từng vùng đề xuất. Một mô hình học máy hoặc mạng nơ-ron sẽ được sử dụng để phân loại đối tượng và dự đoán vị trí của chúng trong vùng đề xuất.

## 2.4. Mô hình YOLO

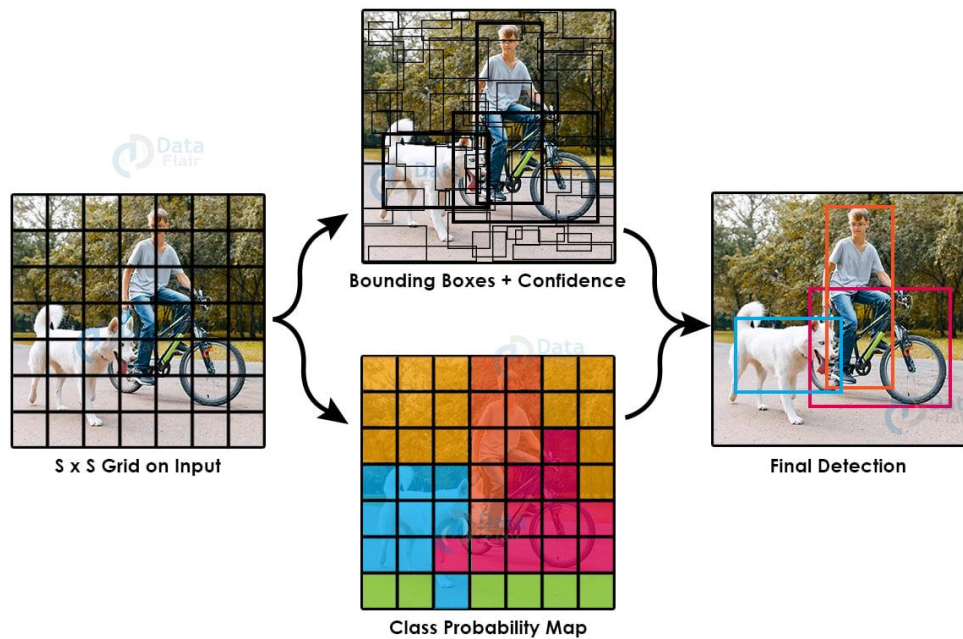
YOLO (You Only Look Once) [Hình 2.12] là một mô hình mạng nơ-ron tích chập (CNN) được sử dụng trong phát hiện đối tượng trên ảnh. Mô hình YOLO được thiết kế để đạt được tốc độ xử lý nhanh và độ chính xác cao, giúp phát hiện đối tượng trong thời gian thực.



Hình 2.12: Kiến trúc mô hình YOLO

Kiến trúc của mạng Yolo bao gồm[1]:

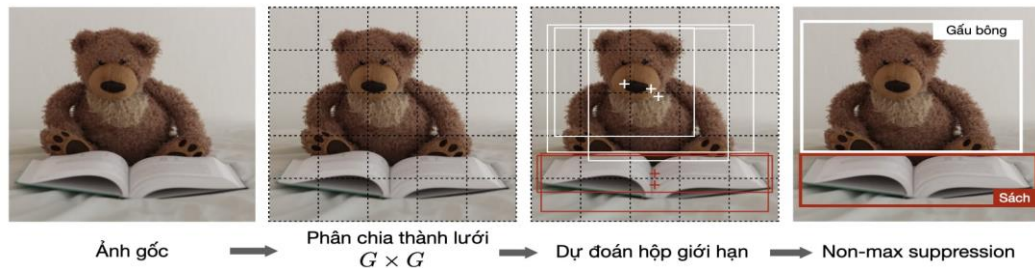
- Base Network (Darknet Architecture): Gồm các mạng nơ ron tích chập được dùng để trích xuất đặc trưng.
- Extra Layer: gồm các lớp fully connected được dùng để phát hiện đối tượng trên feature map của Base Network.



Hình 2.13: Mô phỏng quá trình xử lý của YOLO

Mô hình YOLO sử dụng một lần duy nhất đưa toàn bộ ảnh qua mạng CNN và đưa ra kết quả phát hiện đối tượng. YOLO sử dụng một lớp grid [Hình 2.13] để chia nhỏ ảnh đầu vào thành các ô vuông và dự đoán các bounding box cho các đối tượng có trong mỗi ô vuông[1]. Mỗi bounding box sẽ được dự đoán vị trí, độ lớn và xác suất có chứa đối tượng. Sau đó, YOLO sử dụng một thuật toán [Hình 2.14] để loại bỏ

các bounding box không cần thiết và chỉ giữ lại các bounding box tốt nhất để đưa ra kết quả phát hiện cuối cùng.



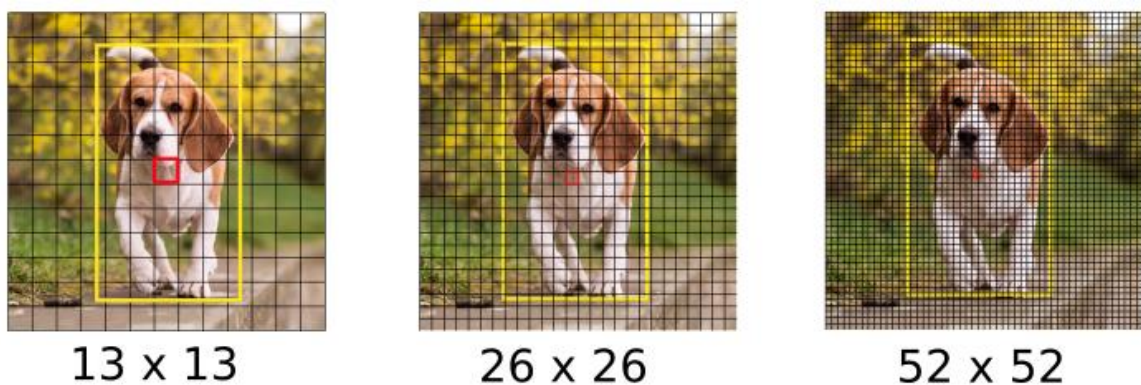
Hình 2.14: Mô tả quy trình phát hiện đối tượng YOLO

Kết quả của YOLO là một vector bao gồm:  $[p_0, (tx, ty, tw, th), (p_1, p_2, p_3, \dots, p_n)]$ .

Trong đó:

- $P_0$  là xác suất dự báo vật thể xuất hiện trong bounding box.
- $(tx, ty, tw, th)$ : Lần lượt là kích thước tọa độ tâm, kích thước rộng và dài của bounding box.
- $(p_1, p_2, p_3, \dots, p_n)$ : Lần lượt là xác suất dự đoán của  $n$  lớp vật thể.

YOLO sẽ dựa đoán trên nhiều feature map [Hình 2.15], những feature map có kích thước nhỏ giúp dự đoán được các vật thể có kích thước lớn. Những feature map có kích thước lớn giúp ta dự đoán được các vật thể nhỏ một cách tốt hơn.

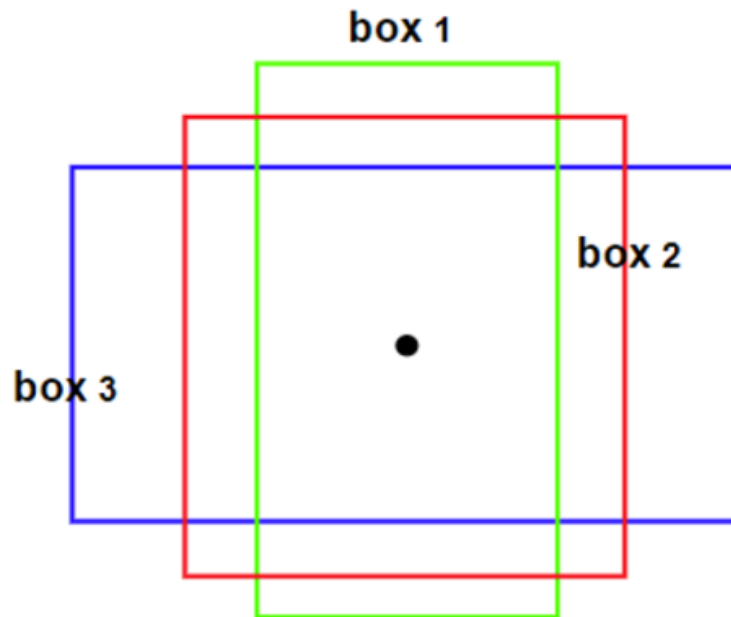


Hình 2.15: Nhận diện hình ảnh trên nhiều feature map của YOLO

Trên mỗi một cell của feature map sẽ áp dụng một số lượng anchor boxes cố định với kích thước khác nhau sao cho tâm của các anchor boxes trùng với cell dùng để dự đoán bounding boxes của vật thể. Anchor box là các hình chữ nhật có kích thước và tỷ lệ khác nhau, được đặt trước và dùng để ước lượng bounding box của đối

tượng trong ảnh. Mô hình YOLO sẽ dự đoán xác suất và hệ số tọa độ cho mỗi anchor box. Xác suất này đại diện cho khả năng của anchor box chứa một đối tượng, và hệ số tọa độ dùng để dự đoán tọa độ của bounding box.

Ở hình ảnh trên, ta có 3 feature map có kích thước lần lượt là 13x13, 26x26, 52x52. Trên mỗi một cell của feature này, áp dụng 3 anchor box [Hình 2.16] có kích thước khác nhau như hình bên dưới.



Hình 2.16: Mô phỏng các hộp neo (anchor box)

Như vậy số lượng anchor box của bức ảnh đầu vào sẽ là:

$$(13 \times 13 + 26 \times 26 + 52 \times 52) \times 3 \text{ anchor box} = 10.647 \text{ (anchor box)}$$

Đây là một số lượng rất lớn và cũng là nguyên nhân khiến quá trình huấn luyện của mô hình YOLO vô cùng chậm bởi chúng ta cần dự đoán đồng thời nhãn và bounding box của 10.647 bounding box.

Khi huấn luyện YOLO, có một số lưu ý quan trọng:

- Cần có RAM với dung lượng lớn hơn để lưu trữ 10647 bounding box trong quá trình huấn luyện YOLO.
- Không thể đặt batch\_size quá lớn như trong các mô hình phân loại (classification) vì có nguy cơ gặp lỗi "Out of Memory".
- Quá trình xử lý của một bước trên YOLO mất thời gian nhiều lần so với các mô hình phân loại. Do đó, cần giới hạn số lượng bước huấn luyện cho YOLO.

Một kinh nghiệm khi thực hiện huấn luyện mô hình YOLO là không nên chọn số lượng epoch lớn, chọn số lượng epoch vừa đủ để không mất quá nhiều thời gian để huấn luyện giảm thiểu rủi ro nguy cơ từ phần cứng (máy tự động reset khi quá nóng, tràn bộ nhớ, ...) gây mất mát dữ liệu huấn luyện.

## **2.5. Thuật toán Deepsort**

Object tracking là một kỹ thuật thị giác máy tính (computer vision technique) liên quan đến việc tự động theo dõi chuyển động và vị trí của một đối tượng trong chuỗi video theo thời gian. Mục tiêu của theo dõi đối tượng là duy trì danh tính của đối tượng, ngay cả khi đối tượng di chuyển và thay đổi hình thức, đồng thời định vị chính xác đối tượng trong từng khung hình của video. Deep SORT (Deep Simple Online Realtime Tracking) là thuật toán theo dõi đối tượng tiên tiến kết hợp trình phát hiện đối tượng dựa trên học sâu với thuật toán theo dõi để đạt được độ chính xác cao và mạnh mẽ trong môi trường có nhiều đối tượng.

Deep SORT dựa trên thuật toán SORT (Theo dõi trực tuyến và thời gian thực đơn giản), sử dụng bộ lọc Kalman và thuật toán Hungary để liên kết phát hiện đối tượng trên các khung. Deep SORT mở rộng SORT bằng cách kết hợp một bộ phát hiện đối tượng dựa trên deep neural network-based, điển hình là convolutional neural network (CNN), để cải thiện độ chính xác của việc phát hiện đối tượng.

DeepSORT (Deep Learning-based SORT) là một phần mở rộng của thuật toán theo dõi đối tượng phổ biến có tên là SORT (Simple Online and Realtime Tracking). DeepSORT bổ sung tính năng trích xuất tính năng xuất hiện sâu và quy trình khớp với thuật toán SORT ban đầu để cải thiện độ chính xác và độ mạnh theo dõi của nó.

DeepSORT (Deep Learning-based SORT) có một số ưu điểm giúp nó tốt hơn các thuật toán theo dõi đối tượng khác:

- Độ chính xác theo dõi được cải thiện: Bằng cách kết hợp các tính năng xuất hiện sâu bên cạnh thông tin vị trí và chuyển động được sử dụng bởi thuật toán SORT ban đầu, DeepSORT có thể xử lý các tình huống trong đó các đối tượng có thể tạm thời biến mất hoặc che khuất lẫn nhau, giúp theo dõi đối tượng chính xác và mạnh mẽ hơn
- Xử lý nhiều đối tượng: DeepSORT có thể theo dõi đồng thời nhiều đối tượng và duy trì danh tính của chúng theo thời gian, ngay cả trong những cảnh đông đúc
- Hiệu suất thời gian thực: DeepSORT được thiết kế để hoạt động trong thời gian thực, giúp nó phù hợp với nhiều ứng dụng như giám sát, người máy và ô tô tự lái

- Chi phí tính toán thấp: Việc sử dụng trình trích xuất tính năng sâu trong DeepSORT là hiệu quả về mặt tính toán và không làm tăng đáng kể chi phí tính toán tổng thể của thuật toán
- Dễ dàng tích hợp với các trình phát hiện đối tượng hiện có: DeepSORT có thể dễ dàng tích hợp với các trình phát hiện đối tượng hiện có, chẳng hạn như YOLO, để cung cấp một quy trình theo dõi và phát hiện đối tượng từ đầu đến cuối.

## **2.6. Kết luận chương**

Kết thúc chương 2, nhóm chúng em đã trình bày cơ sở lý thuyết về phát hiện và nhận dạng khuôn mặt, cũng như quy trình nhận dạng khuôn mặt nhằm xác định và phân loại giới tính. Bên cạnh đó, nắm được lý thuyết về mạng CNN cũng như là kiến trúc của mô hình YOLO phục vụ cho việc huấn luyện mô hình. Từ cơ sở lý thuyết, chúng em sẽ bắt đầu thiết kế chương trình. Cụ thể quy trình thiết kế sẽ được trình bày tại chương 3.

## CHƯƠNG 3: THIẾT KẾ VÀ ĐÁNH GIÁ HỆ THỐNG

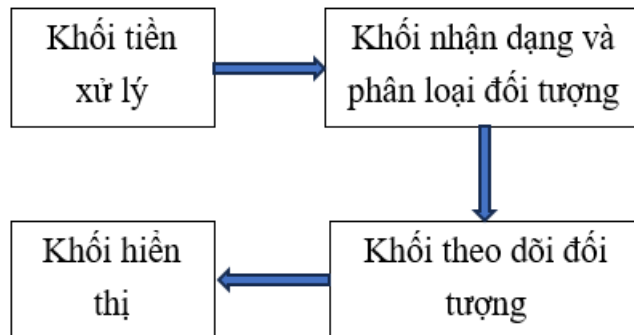
### 3.1. Giới thiệu chương

Chương này nhóm chúng em sẽ trình bày về những vấn đề liên quan đến quá trình thiết kế và đánh giá mô hình huấn luyện. Nắm được cấu trúc để xây dựng nên hệ thống. Thiết kế bắt đầu với yêu cầu đã đặt ra tại chương 1, đó là việc xây dựng được hệ thống nhận dạng khuôn mặt để phân loại giới tính và đếm số lượng nhằm mục đích hiển thị quảng cáo phù hợp với đối tượng.

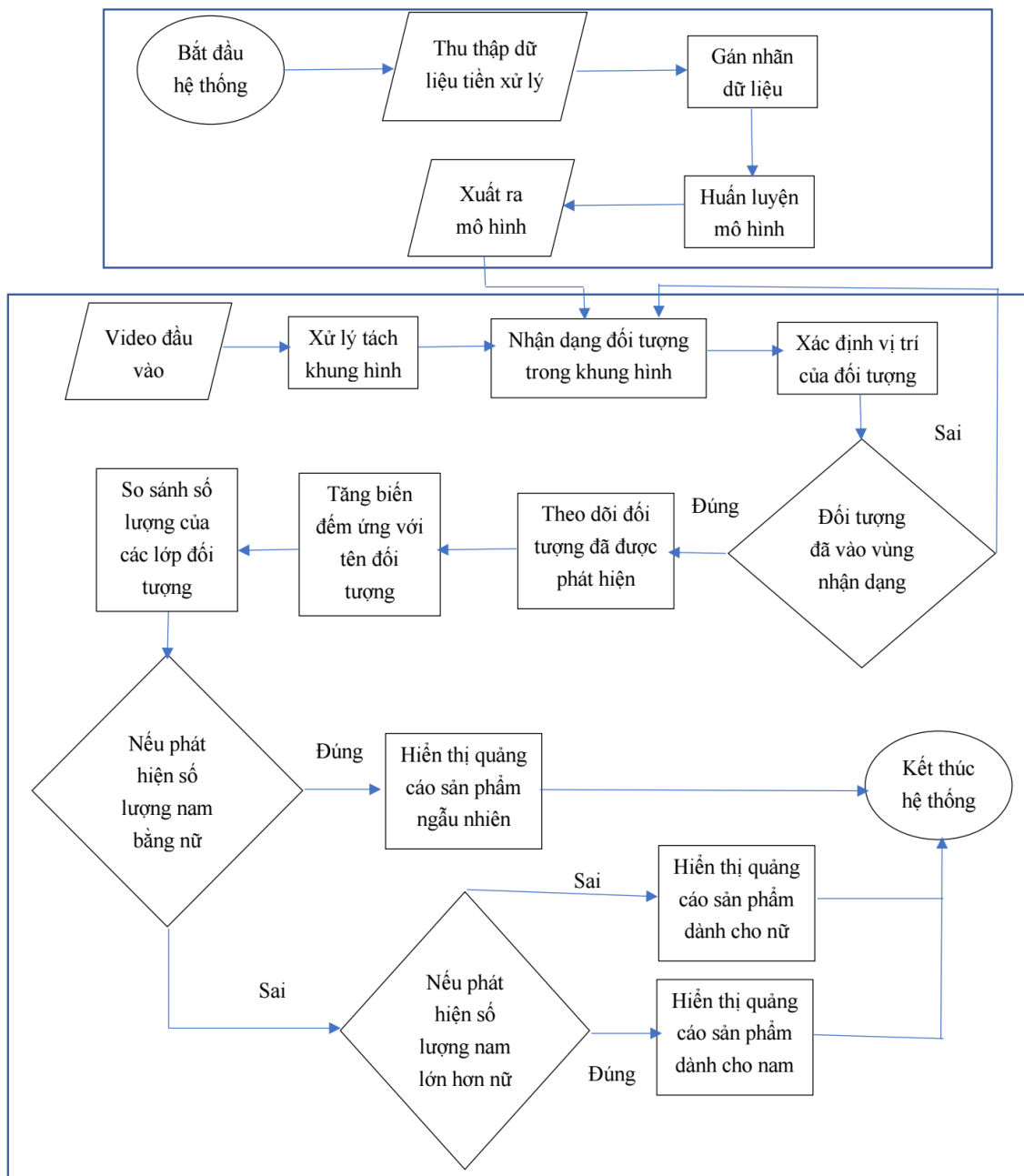
### 3.2. Quy trình thiết kế hệ thống

Để xây dựng được một hệ thống hoàn chỉnh thì ta cần một thiết bị thu nhận dữ liệu (ở đây chúng em sẽ sử dụng webcam máy tính) và một thiết bị xử lý trung tâm (máy tính). Vì thiết bị phần cứng đã có sẵn nên chúng em sẽ tập trung xây dựng mô hình cho bộ xử lý trung tâm [Hình 3.1] gồm các khối như sau:

1. Khối tiền xử lý
2. Khối nhận dạng và phân loại đối tượng
3. Khối theo dõi đối tượng
4. Khối hiển thị



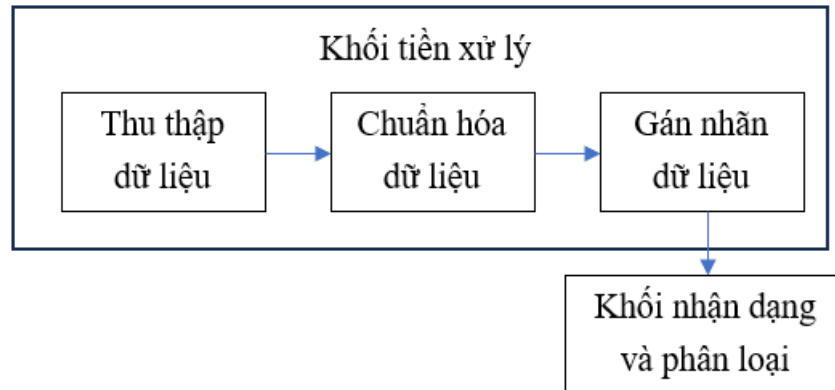
Hình 3.1: Sơ đồ khối của bộ xử lý trung tâm



Hình 3.2: Lưu đồ thuật toán của hệ thống

### 3.2.1. Khối tiền xử lý

Sơ đồ tổng quan:



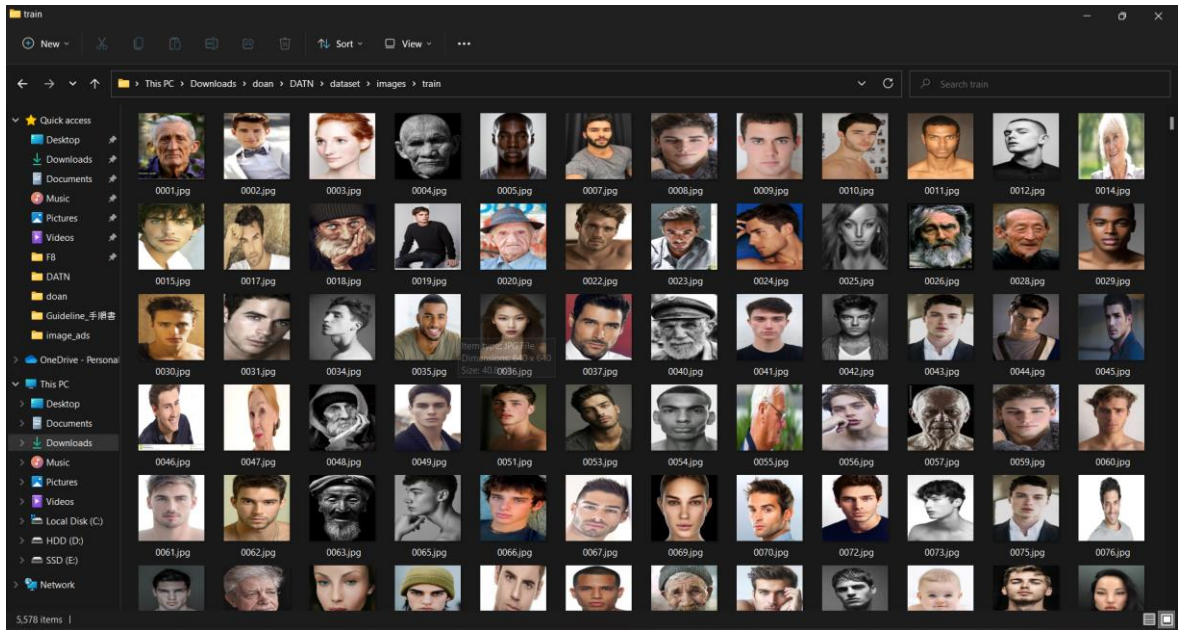
Hình 3.3: Tổng quan sơ đồ khối tiền xử lý

Mô tả:

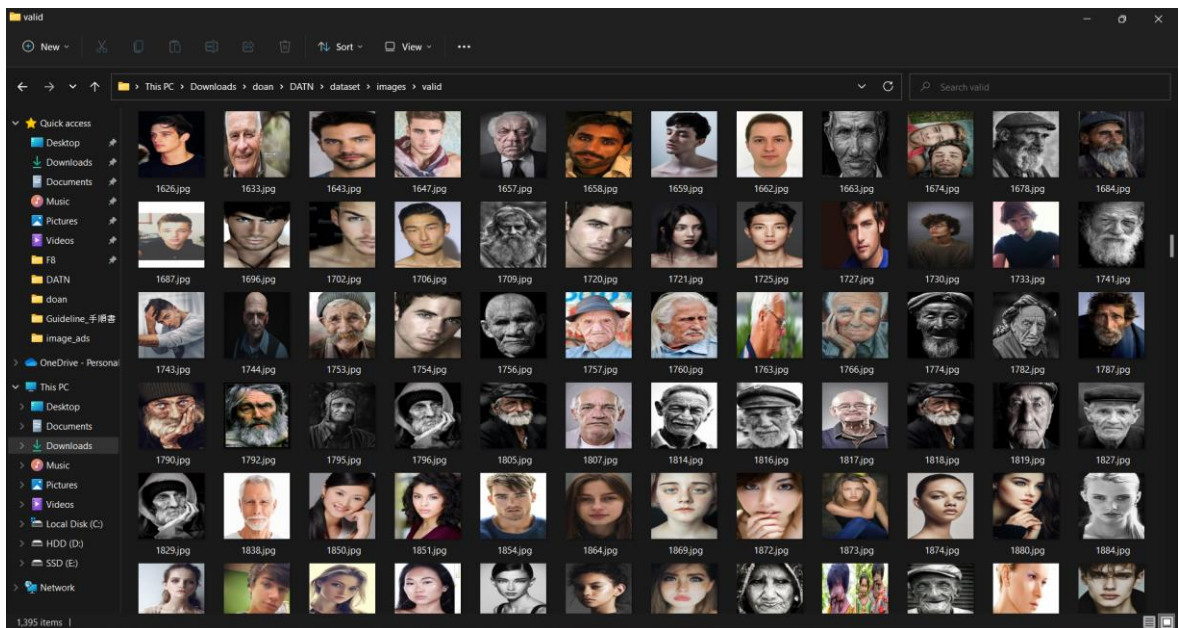
Khối tiền xử lý trong mô hình học máy là quá trình chuẩn bị và biến đổi dữ liệu đầu vào trước khi đưa vào mô hình để đạt được hiệu suất tốt hơn và dễ dàng xử lý dữ liệu.

Chức năng:

- Thu thập dữ liệu: Thu thập dữ liệu từ các nguồn khác nhau để tạo thành tập dữ liệu huấn luyện.
- Xử lý dữ liệu bị thiếu: Kiểm tra và xử lý các giá trị thiếu trong dữ liệu, thông qua việc điền giá trị thiếu hoặc loại bỏ các mẫu dữ liệu bị thiếu.
- Chuẩn hóa dữ liệu: Đưa các đặc trưng về cùng phạm vi hoặc dạng dữ liệu chuẩn.
- Chia tập dữ liệu: Phân chia tập dữ liệu thành tập huấn luyện và tập kiểm tra để đánh giá hiệu suất của mô hình.

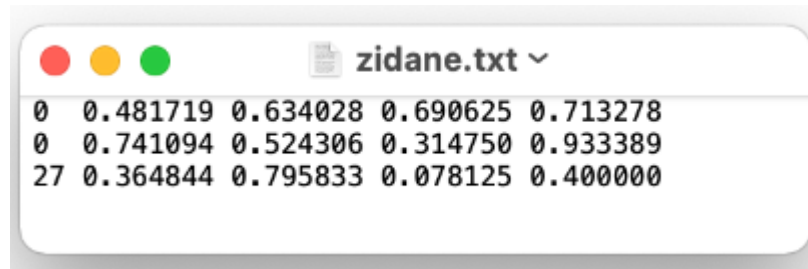


Hình 3.4: Hình ảnh tập dữ liệu dùng để huấn luyện



Hình 3.5: Hình ảnh tập dữ liệu dùng để kiểm tra

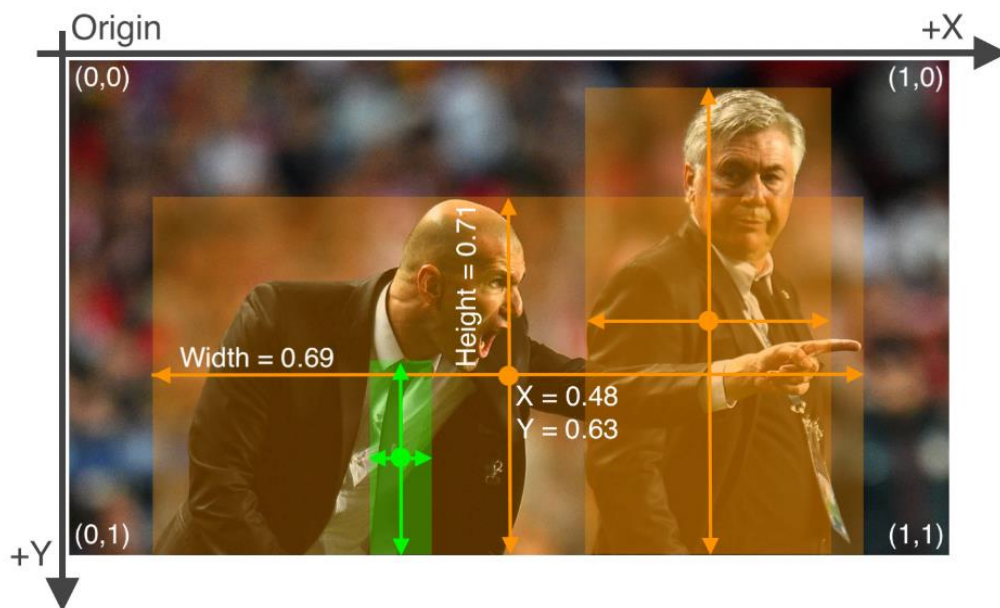
Để đào tạo mô hình YOLOv5, tập dữ liệu huấn luyện cần phải tuân thủ định dạng yêu cầu của YOLOv5, mỗi hình ảnh trong tập dữ liệu sẽ phải có một tệp văn bản dạng text \*.txt (nếu không có đối tượng trong hình ảnh thì không cần) chứa thông tin về nhãn đối tượng nhận dạng của ảnh [Hình 3.6].



Hình 3.6: Nội dung nhãn cho một hình ảnh

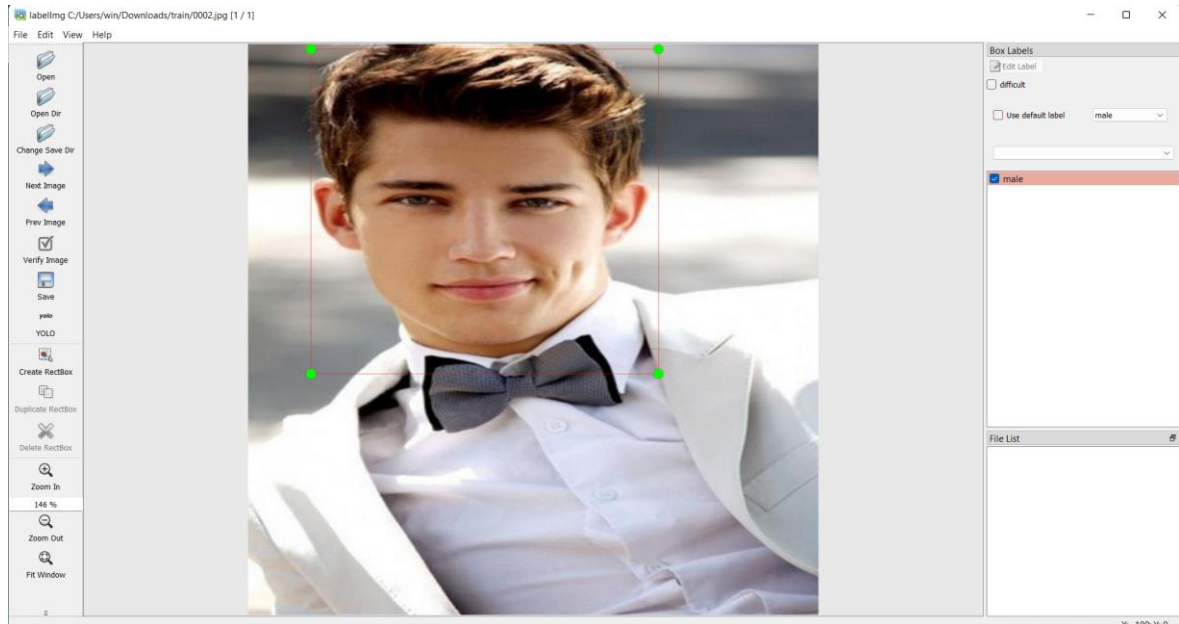
Thông số kỹ thuật [Hình 3.7] của tệp \*.txt là:

- Một hàng cho mỗi đối tượng
- Mỗi hàng bao gồm: lớp, tọa độ điểm trung tâm  $x\_center$  và  $y\_center$ , chiều rộng và chiều cao của đối tượng (được chuẩn hoá về 0-1)
- Số lớp không được lập chỉ mục (bắt đầu từ 0).



Hình 3.7: Mô tả thông số cho nhãn dữ liệu

Tập dữ liệu cho việc huấn luyện [Hình 3.4] phân loại nam nữ : khoảng 3000 ảnh khuôn mặt nam giới và 3000 ảnh khuôn mặt nữ giới, được chia thành 2 class. Các hình ảnh được chụp từ camera điện thoại kết hợp với bộ dữ liệu lấy ở các trang web trên internet. Dữ liệu bộ dữ liệu huấn luyện hầu hết có kích thước là 640px x 640px, tiến hành resize các hình ảnh có kích thước rời rạc về cùng kích thước này để có thể thực hiện huấn luyện mô hình.



Hình 3.8: Hình ảnh sử dụng labellng để gắn nhãn

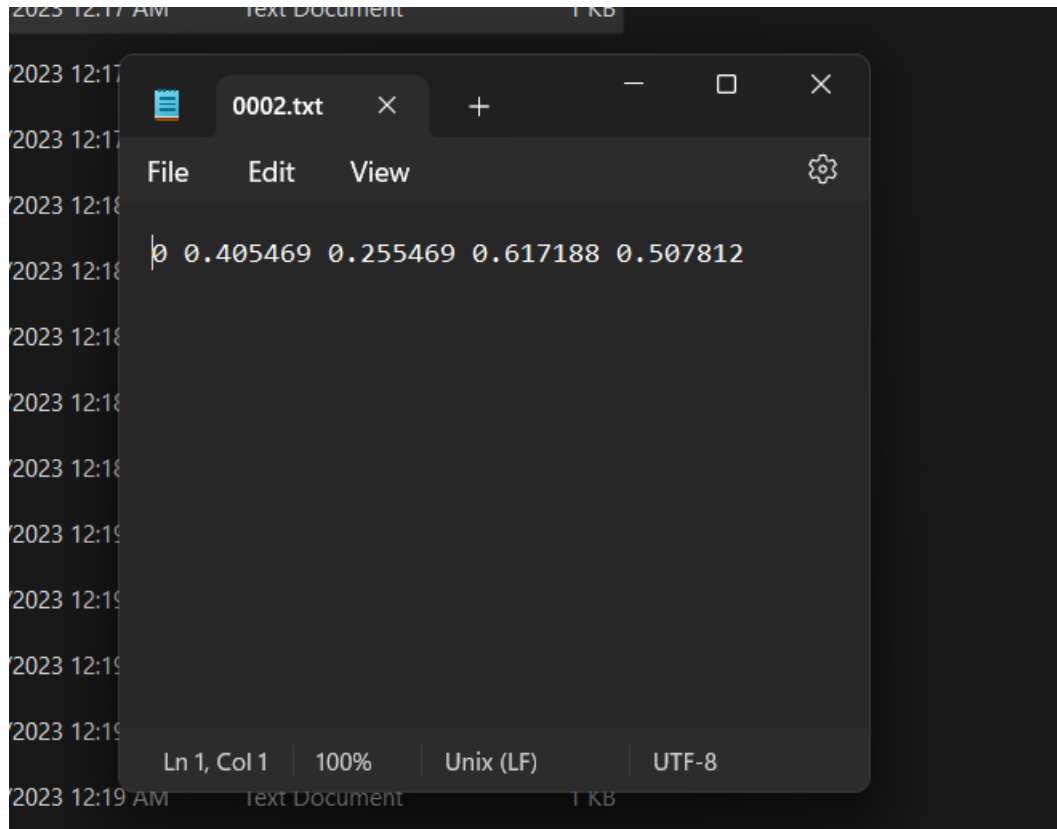
Các bước thực hiện gắn nhãn [Hình 3.8]:

Bước 1: Chọn hình ảnh cần để gắn nhãn.

Bước 2: Tạo các khung hình chữ nhật và kéo giữ chuột sao cho vừa với phương tiện. Sau đó có hộp thoại xuất hiện thì điền tên phương tiện. Tiếp tục thực hiện bước này cho đến khi đã gắn hết các nhãn cho các phương tiện cần quan tâm.

Bước 3: Lưu lại hình ảnh đã gắn nhãn. Phần mềm sẽ tạo ra các file “.txt” tương ứng với các hình. File này chứa dữ liệu đã gắn nhãn của ảnh đó.

File label được lưu dưới dạng file “.txt” với các thông số như sau:



Hình 3.9: Hình ảnh thông số của nhãn sau khi gán

Phần mềm sẽ tạo các file txt [Hình 3.9] cho mỗi ảnh trong cùng đường dẫn và cùng tên. Nhưng trong file có đuôi mở rộng txt sẽ chứa các thông tin như số của vật thể và tọa độ của vật thể trong hình. Với mỗi vật thể sẽ có các thông số <lớp> <x> <y> <chiều rộng> <chiều cao>

Trong đó:

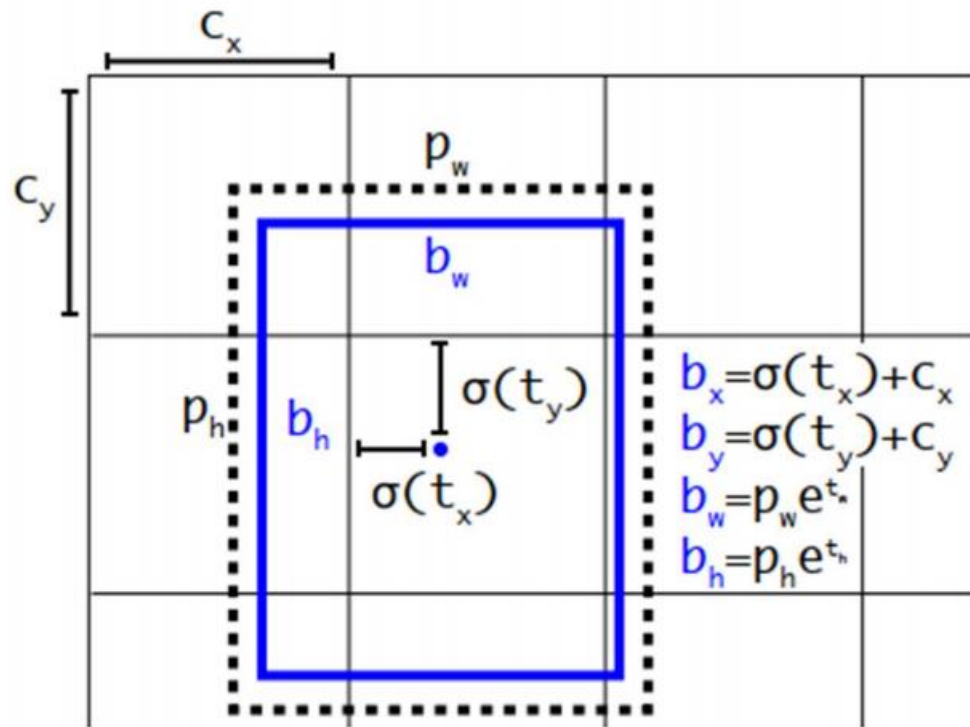
- <lớp> là một số tự nhiên của vật thể từ 0 đến số lớp – 1. Với số lớp của chúng em là 2 lớp giới tính thì sẽ tương ứng là nam là 0 và nữ là 1.
- <x> <y> <chiều rộng> <chiều cao> là các giá trị thập phân liên quan đến vị trí của hình và nằm trong khoảng 0 đến 1 và được tính toán như sau:

$$+ <x> = \text{<tọa độ thực của x> / <chiều rộng của hình>}$$

$$+ <y> = \text{<tọa độ thực của y> / <chiều cao của hình>}$$

$$+ <chiều rộng> = \text{<chiều rộng thực của khung> / <chiều rộng của hình>}$$

$$+ <chiều cao> = \text{<chiều cao thực của khung> / <chiều cao của hình>}$$



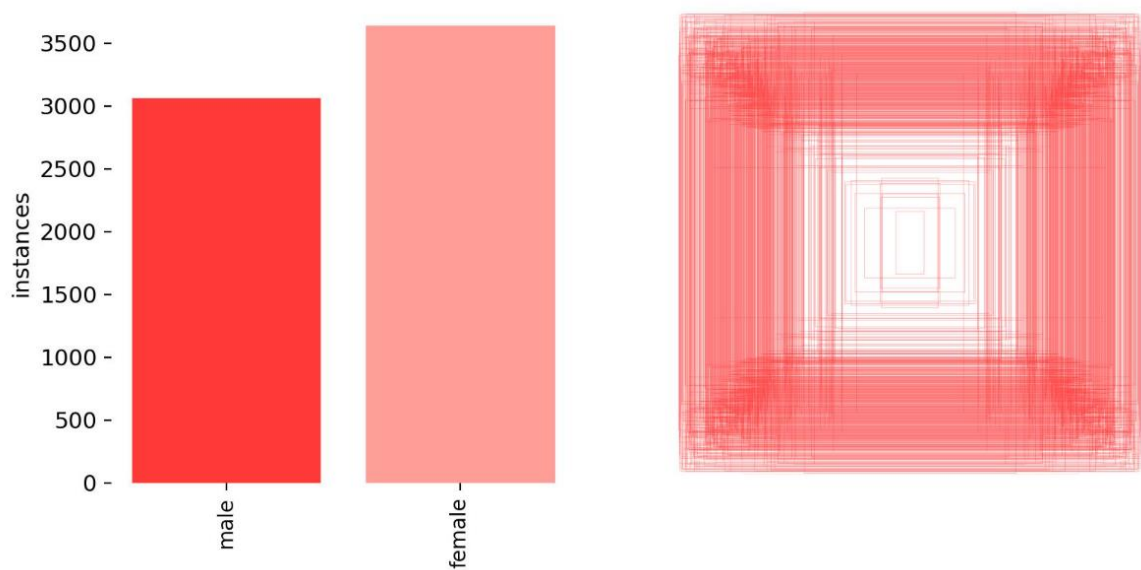
Hình 3.10: Cách tính các thông số của hộp biên

Điều này thuận lợi cho việc xác minh vị trí của vật khi hình bị thay đổi với các kích thước khác nhau.

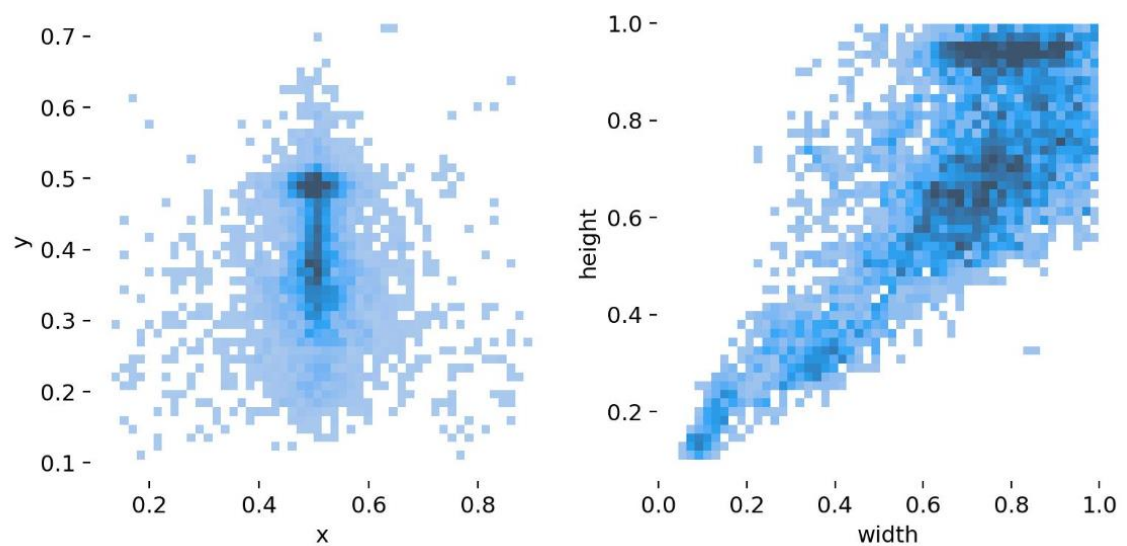
Quá trình gán nhãn thu được tổng cộng :

Bảng 3.1: Số lượng nhãn đã thu thập

|       | Male | Female |
|-------|------|--------|
| Train | 3057 | 3634   |
| Valid | 687  | 713    |
| Total | 8091 |        |



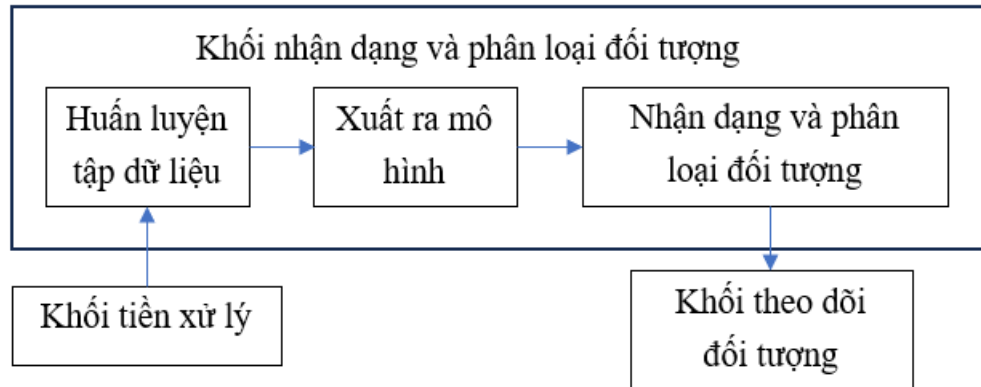
Hình 3.11: Số lượng và vị trí thể hiện mỗi lớp trong tập huấn luyện



Hình 3.12: Biểu đồ mô tả vị trí và kích thước các lớp trong tập huấn luyện

### 3.2.2. Khối nhận dạng và phân loại đối tượng

Sơ đồ tổng quan:



Hình 3.13: Tổng quan sơ đồ khối nhận dạng và phân loại đối tượng

Mô tả:

Sử dụng mô hình YOLOv5, đây là một mô hình học sâu (deep learning) được sử dụng để phát hiện và nhận dạng đối tượng trong ảnh và video. Nó là phiên bản tiếp theo của YOLO (You Only Look Once), một kiến trúc mô hình phát hiện đối tượng nhanh chóng và chính xác. Đầu vào là ảnh thu thập được sau khi qua khối tiền xử lý. Đầu ra là tập tin sau khi được huấn luyện qua mô hình YOLOv5.

Chức năng:

Chức năng chính của YOLOv5 là phát hiện và nhận dạng đối tượng trong ảnh và video. Mô hình này có khả năng phát hiện nhanh chóng các đối tượng và đưa ra các bounding box (hình chữ nhật) xác định vị trí của chúng trong ảnh hoặc video. Ngoài ra, YOLOv5 cũng có khả năng phân loại các đối tượng đó vào các lớp (class) khác nhau. Đặc điểm và chức năng của YOLOv5 bao gồm:

- Kiến trúc mạng nơ-ron: YOLOv5 sử dụng một kiến trúc mạng nơ-ron tích chập (CNN) để trích xuất đặc trưng từ ảnh. Mô hình bao gồm các layer convolution, pooling và các kỹ thuật như skip connections để xây dựng một mạng nơ-ron sâu có khả năng học biểu diễn đa dạng và phát hiện các đối tượng.

Bảng 3.2: Bảng mô hình tóm tắt của YOLOv5s

|   | from | n | params | module             | arguments        |
|---|------|---|--------|--------------------|------------------|
| 0 | -1   | 1 | 3520   | models.common.Conv | [3, 32, 6, 2, 2] |
| 1 | -1   | 1 | 18560  | models.common.Conv | [32, 64, 3, 2]   |

|    |          |   |         |                                      |                      |
|----|----------|---|---------|--------------------------------------|----------------------|
| 2  | -1       | 1 | 18816   | models.common.C3                     | [64, 64, 1]          |
| 3  | -1       | 1 | 73984   | models.common.Conv                   | [64, 128, 3, 2]      |
| 4  | -1       | 2 | 115712  | models.common.C3                     | [128, 128, 2]        |
| 5  | -1       | 1 | 295424  | models.common.Conv                   | [128, 256, 3, 2]     |
| 6  | -1       | 3 | 625152  | models.common.C3                     | [256, 256, 3]        |
| 7  | -1       | 1 | 1180672 | models.common.Conv                   | [256, 512, 3, 2]     |
| 8  | -1       | 1 | 1182720 | models.common.C3                     | [512, 512, 1]        |
| 9  | -1       | 1 | 656896  | models.common.SPPF                   | [512, 512, 5]        |
| 10 | -1       | 1 | 131584  | models.common.Conv                   | [512, 256, 1, 1]     |
| 11 | -1       | 1 | 0       | torch.nn.modules.upsampling.Upsample | [None, 2, 'nearest'] |
| 12 | [-1, 6]  | 1 | 0       | models.common.Concat                 | [1]                  |
| 13 | -1       | 1 | 361984  | models.common.C3                     | [512, 256, 1, False] |
| 14 | -1       | 1 | 33024   | models.common.Conv                   | [256, 128, 1, 1]     |
| 15 | -1       | 1 | 0       | torch.nn.modules.upsampling.Upsample | [None, 2, 'nearest'] |
| 16 | [-1, 4]  | 1 | 0       | models.common.Concat                 | [1]                  |
| 17 | -1       | 1 | 90880   | models.common.C3                     | [256, 128, 1, False] |
| 18 | -1       | 1 | 147712  | models.common.Conv                   | [128, 128, 3, 2]     |
| 19 | [-1, 14] | 1 | 0       | models.common.Concat                 | [1]                  |
| 20 | -1       | 1 | 296448  | models.common.C3                     | [256, 256, 1, False] |
| 21 | -1       | 1 | 590336  | models.common.Conv                   | [256, 256, 3, 2]     |
| 22 | [-1, 10] | 1 | 0       | models.common.Concat                 | [1]                  |
| 23 | -1       | 1 | 1182720 | models.common.C3                     | [512, 512, 1, False] |

|                                                                                      |              |   |       |                    |                                                                                                            |
|--------------------------------------------------------------------------------------|--------------|---|-------|--------------------|------------------------------------------------------------------------------------------------------------|
|                                                                                      | [17, 20, 23] |   |       |                    | [2, [[10, 13, 16, 30, 33, 23], [30, 61, 62, 45, 59, 119], [116, 90, 156, 198, 373, 326]], [128, 256, 512]] |
| 24                                                                                   |              | 1 | 18879 | models.yolo.Detect |                                                                                                            |
| <b>Model summary: 214 layers, 7025023 parameters, 7025023 gradients, 16.0 GFLOPs</b> |              |   |       |                    |                                                                                                            |

- Mô tả chi tiết bảng trên [Bảng 3.2]:

+ "214 layers" (214 lớp): Đây là số lượng lớp trong mô hình YOLOv5s. Mỗi lớp thực hiện các phép tính như tích chập, kích hoạt, pooling, và normalization để trích xuất đặc trưng từ dữ liệu đầu vào.

+ "7025023 parameters" (7025023 tham số): Đây là số lượng tham số trong mô hình YOLOv5s. Các tham số bao gồm các ma trận trọng số và các tham số dịch chuyển (bias) của các lớp. Các tham số này được tối ưu hóa trong quá trình huấn luyện để mô hình có thể dự đoán chính xác vị trí và loại đối tượng trong hình ảnh.

+ "7025023 gradients" (7025023 gradient): Đây là số lượng gradient tương ứng với mỗi tham số trong mô hình YOLOv5s. Gradient đại diện cho độ dốc của hàm mất mát theo từng tham số, cho biết cần điều chỉnh tham số như thế nào để giảm thiểu sai số. Gradient được tính toán thông qua quá trình lan truyền ngược (backpropagation) trong quá trình huấn luyện.

+ "16.0 GFLOPs": Đây là số lượng phép tính lượng tử động thực hiện bởi mô hình trong mỗi giây. GFLOPs (giga floating-point operations per second) đo lường khả năng tính toán của mô hình và thể hiện số lượng phép tính số học (nhân, cộng, trừ) được thực hiện trên các số dấu chấm động (floating-point) trong mỗi giây. Điều này cho thấy mức độ phức tạp tính toán của mô hình và ảnh hưởng đến thời gian xử lý hình ảnh.

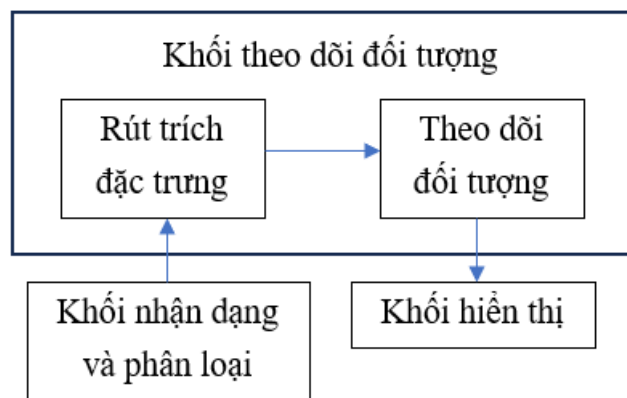
- Phát hiện và nhận dạng đối tượng: YOLOv5 sử dụng anchor boxes và lưới ô (grid) để phát hiện và nhận dạng đối tượng trong ảnh và video. Mô hình dự đoán các bounding box và xác suất của các lớp cho mỗi ô trong lưới, sau đó sử dụng các ngưỡng (threshold) để lọc các đối tượng phát hiện.

- Độ chính xác và tốc độ: YOLOv5 được thiết kế để cân bằng giữa độ chính xác và tốc độ. Mô hình này có khả năng phát hiện và nhận dạng đối tượng nhanh chóng với độ chính xác tương đối cao, giúp xử lý ảnh và video trong thời gian thực.

- Huấn luyện và triển khai: YOLOv5 có thể được huấn luyện trên các bộ dữ liệu chứa đối tượng được gắn nhãn. Sau khi huấn luyện, mô hình có thể được triển khai trên nhiều nền tảng khác nhau, bao gồm máy tính cá nhân, hệ thống nhúng và thiết bị di động.

### **3.2.3. Khối theo dõi đối tượng**

Sơ đồ tổng quan:



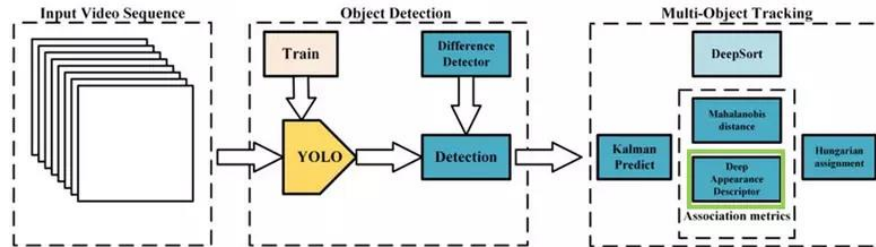
Hình 3.14: Tổng quan khối theo dõi đối tượng

Mô tả:

DeepSORT (Deep Learning for Real-Time Object Tracking) là một thuật toán theo dõi đối tượng thời gian thực sử dụng học sâu (deep learning). DeepSORT kết hợp các mô hình deep learning để xác định và theo dõi các đối tượng trong các khung hình video liên tục [Hình 3.15].

# DeepSORT

Where is the Deep Learning in all of this?

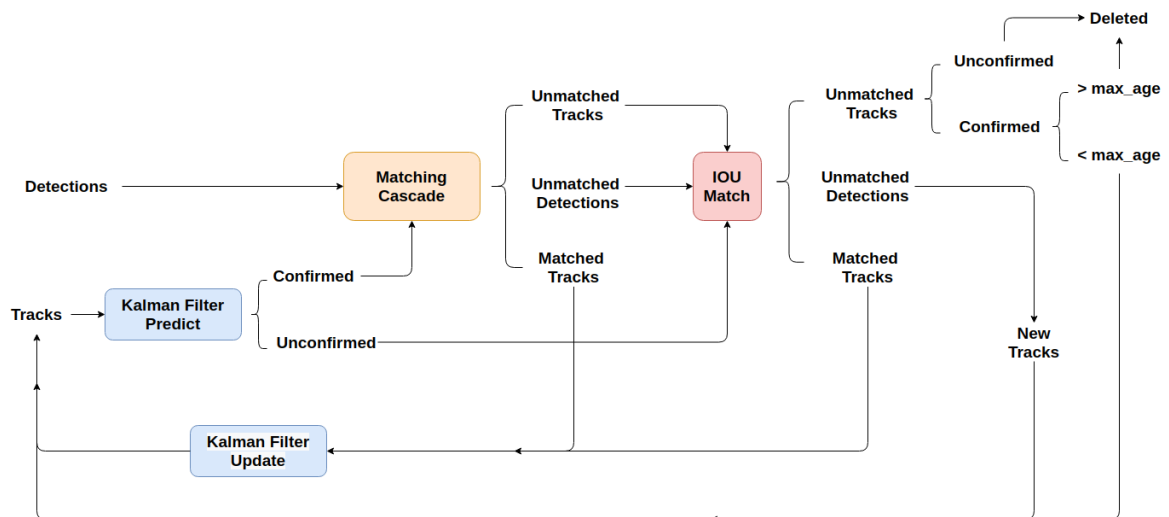


Hình 3.15: Hệ thống Real-time Multi-person Tracking với YOLO và DeepSORT

Chức năng:

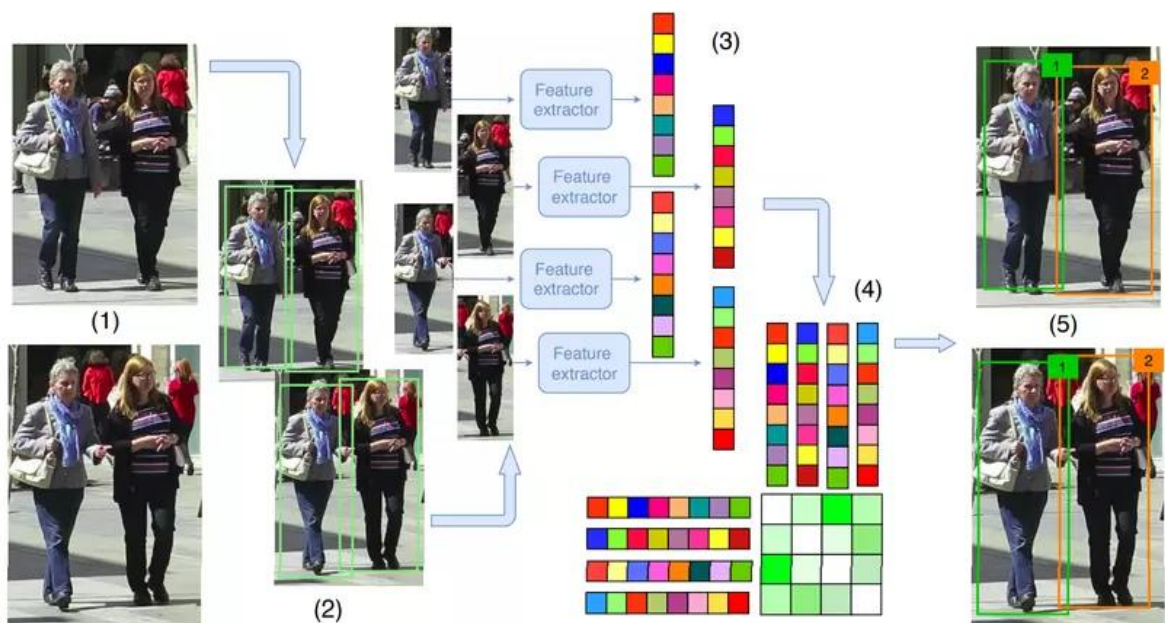
Dưới đây là quá trình hoạt động chính của DeepSORT trong hệ thống [Hình 3.17]:

- Nhận diện đối tượng: Sử dụng mô hình nhận diện đối tượng YOLOv5 để xác định các đối tượng trong khung hình ban đầu. Mô hình này sẽ tạo ra các hình chữ nhật bao quanh các đối tượng và đưa ra các điểm đặc trưng của chúng.
- Rút trích đặc trưng: Sử dụng một mô hình deep learning CNN để rút trích các đặc trưng từ các hình chữ nhật xác định ở bước trước. Mô hình này chuyển đổi hình chữ nhật thành một vector đặc trưng có kích thước cố định.



Hình 3.16: Luồng xử lý của DeepSORT

- Xác định đối tượng gần nhất: Sử dụng thuật toán Nearest Neighbor (k-nearest neighbors) để xác định các đối tượng gần nhất dựa trên các vector đặc trưng. Các đối tượng có vector đặc trưng tương tự sẽ được gom nhóm lại và xem như là cùng một đối tượng.
- Theo dõi đối tượng: Sử dụng bộ lọc Kalman để dự đoán vị trí tiếp theo của các đối tượng theo thời gian. Bộ lọc Kalman tính toán các ước lượng về tốc độ và gia tốc của đối tượng để dự đoán vị trí tiếp theo.
- Gắn kết các đối tượng: Sử dụng các thông tin từ bước trước và kỹ thuật gắn kết đối tượng để xác định xem các đối tượng được theo dõi có liên quan với nhau hay không. Các đối tượng liên quan sẽ được gắn kết lại thành các quỹ đạo theo dõi liền mạch.

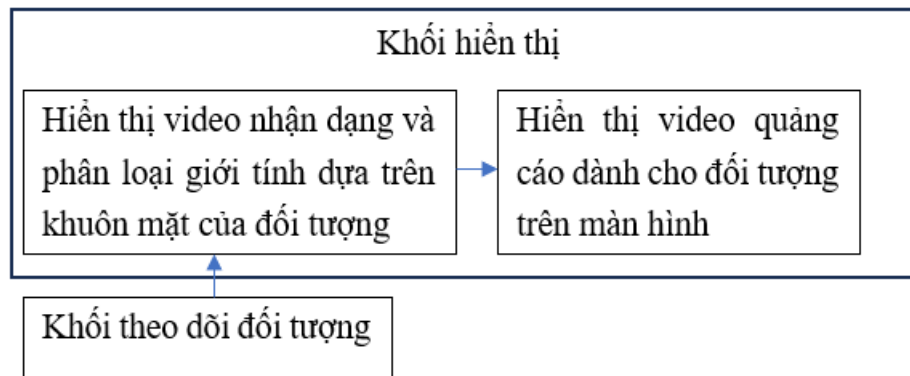


Hình 3.17: Quá trình theo dõi đối tượng của thuật toán DeepSORT

Thuật toán DeepSORT kết hợp các mô hình deep learning và các kỹ thuật theo dõi đối tượng truyền thống như bộ lọc Kalman để cải thiện khả năng theo dõi đối tượng trong thời gian thực. Nó được sử dụng rộng rãi trong các ứng dụng như theo dõi đối tượng trong video giám sát, đếm lượng người và nhận dạng hành vi trong phân tích video.

### 3.2.4. Khối hiển thị

Sơ đồ tổng quan:



Hình 3.18: Tổng quan khối hiển thị

Mô tả:

Sử dụng mô hình nhận dạng đối tượng YOLOv5 để phát hiện và nhận dạng các đối tượng kết hợp với thuật toán DeppSORT để theo dõi đối tượng trong video nhận dạng. Sau đó, ta có thể áp dụng mô hình này để xác định vị trí, số lượng và lớp của đối tượng có trong video nhận dạng từ đó hiển thị quảng cáo lên màn hình quảng cáo (ở đây chúng em sử dụng màn hình máy tính thay cho màn hình quảng cáo).

Chức năng:

- Xác định khối hiển thị: Dựa trên vị trí và loại đối tượng đã nhận dạng, ta có thể xác định khối hiển thị trong video nhận dạng. Điều này có thể được thực hiện bằng cách xác định vùng bao quanh các đối tượng đã nhận dạng và xác định khối hiển thị là khu vực chứa tất cả các đối tượng này.
- Phân tích video quảng cáo: Trong cùng khung hình hoặc các khung hình liên tiếp, xác định và phân tích video quảng cáo. Ta sử dụng các thuật toán để dựa vào điều kiện nhận được từ video nhận dạng đối tượng để xác định nội dung và thông tin quảng cáo cần phải hiển thị trong video quảng cáo.
- Đồng bộ hóa và phân tích cùng lúc: Chúng em đã xác định được khối hiển thị video nhận dạng và phân tích video quảng cáo. Từ đó, nhóm chúng em có thể đồng bộ hóa và phân tích cả hai cùng lúc. Điều này cho phép chúng em kiểm tra sự tương tác giữa video nhận dạng và video quảng cáo, ví dụ như xác định xem liệu video quảng cáo có liên quan đến nội dung video nhận dạng hay không.

### **3.3. Kết luận chương**

Chương 3 đã trình bày được quy trình thiết kế hệ thống, từ quy trình đó, phát triển ra sản phẩm demo dựa trên các mô hình được huấn luyện được đánh giá từ bộ cơ sở dữ liệu của chúng em thu thập được. Trình bày được cấu trúc của hệ thống và thiết bị sử dụng để xây dựng hệ thống. Kết quả nghiên cứu và kết luận sẽ được trình bày ở chương tiếp theo.

## CHƯƠNG 4: KẾT QUẢ NGHIÊN CỨU VÀ ĐÁNH GIÁ KẾT QUẢ

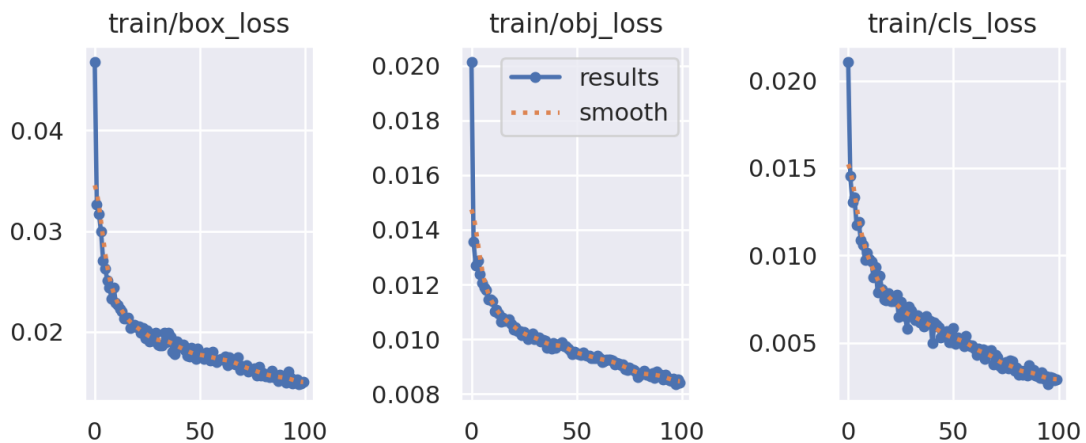
### 4.1. Giới thiệu chương

Chương này trình bày kết quả nghiên cứu từ cơ sở lý thuyết cho đến thiết kế chương trình. Ngoài ra, thử nghiệm hệ thống với các trường hợp khác nhau về số lượng người trong ảnh để kiểm tra độ hiệu quả của hệ thống.

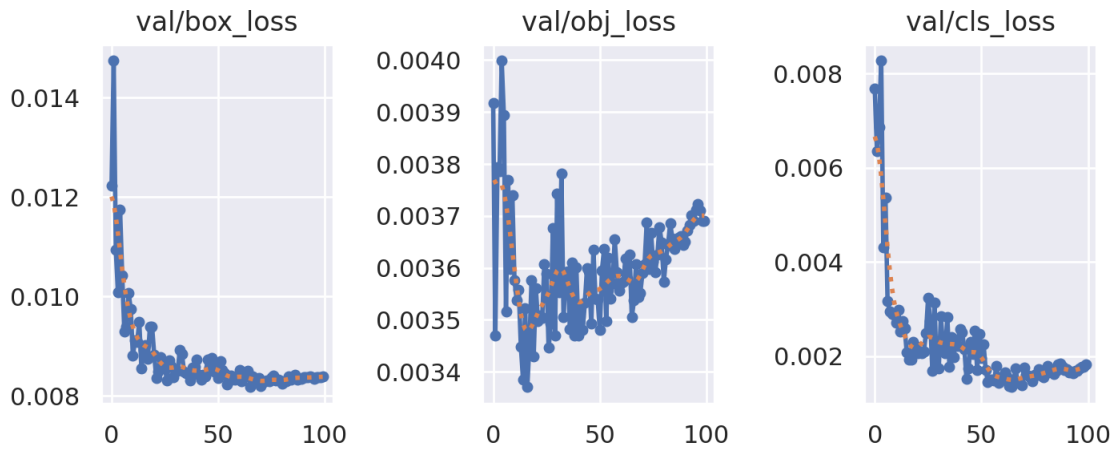
### 4.2. Kết quả nghiên cứu đạt được

#### 4.2.1. Kết quả mô hình học máy nhận dạng nam nữ

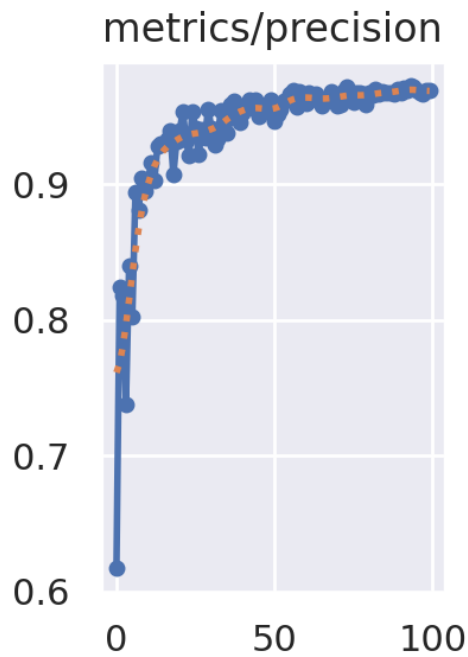
Mô hình được triển khai huấn luyện trên Google Colab với phần cứng RAM hệ thống 12.7 GB, GPU là Nvidia Tesla T4 15 GB. Sau khi huấn luyện ta được các đồ thị hàm loss function bên dưới.



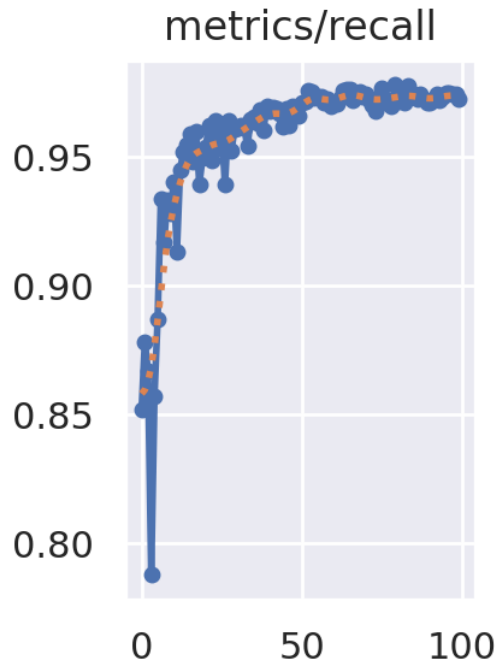
Hình 4.1: Các đồ thị Loss Function của mô hình YOLOv5 trên tập dữ liệu dùng để huấn luyện với 6691 bức ảnh (Train Set)



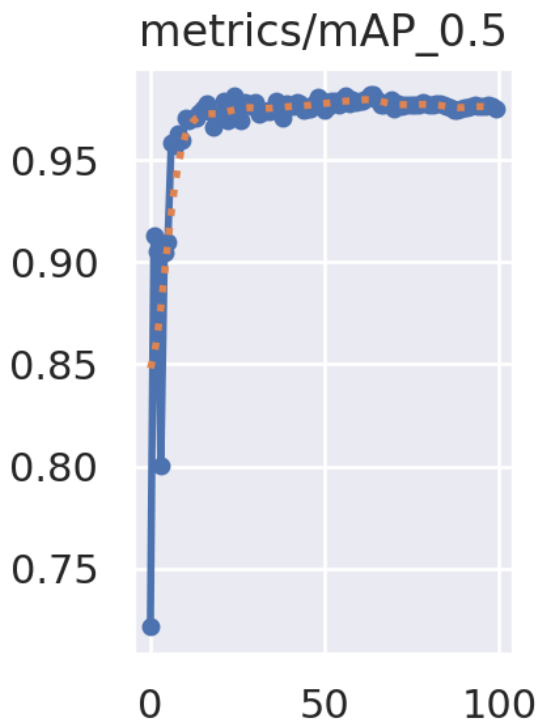
Hình 4.2: Các đồ thị Loss Function của mô hình YOLOv5 trên tập dữ liệu dùng để kiểm tra với 1400 bức ảnh (Validation Set)



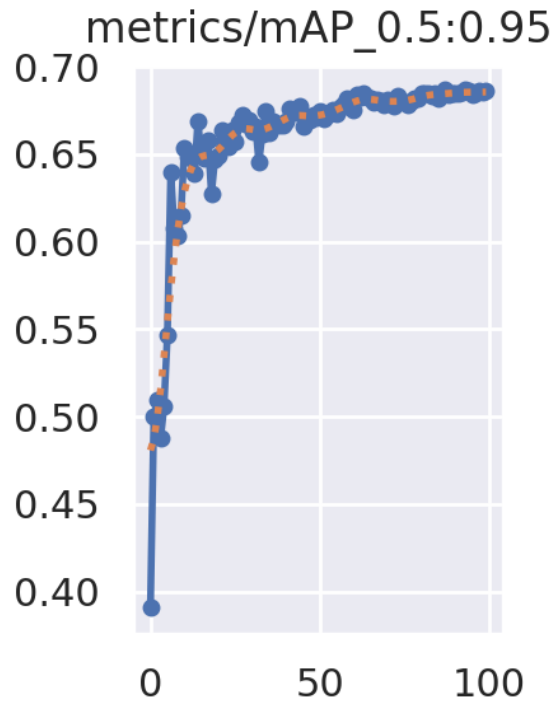
Hình 4.3: Đồ thị tỷ lệ giữa số lượng các đối tượng dự đoán đúng (true positives) và tổng số lượng các đối tượng dự đoán là positive (bao gồm cả true positives và false positives)



Hình 4.4: Đồ thị tỷ lệ giữa số lượng các đối tượng dự đoán đúng (true positives) và tổng số lượng các đối tượng đối tượng thực tế là positive (bao gồm cả true positives và false negatives)



Hình 4.5: Biểu đồ đánh giá hiệu suất tổng thể của mô hình phát hiện đối tượng với ngưỡng IoU là 0.5



Hình 4.6: Biểu đồ đánh giá hiệu suất tổng thể của mô hình phát hiện đối tượng trong khoảng ngưỡng IoU từ 0.5 đến 0.95

Bảng 4.1: Hiệu suất kết quả mô hình

| Đánh giá         | YOLOv5s      |
|------------------|--------------|
| Số lượng tham số | 7,015,519    |
| Kiểm thử         | Tập xác thực |
| mAP50            | 97%          |
| mAP50:95         | 68%          |
| Độ chính xác     | 96%          |
| Độ bao phủ       | 97%          |

Nhận xét:

- Mô hình YOLO vẫn chứng minh được là một trong những mô hình có nhận diện hình ảnh theo thời gian thực với độ chính xác cao.
- Mô hình YOLOv5s với số lượng tham số khoảng 7 triệu đã cho thấy hiệu suất kết quả nhận diện rất tốt và tốc độ xử lý nhận dạng khuôn mặt đối tượng một cách nhanh chóng với tốc độ trung bình khoảng 0.02s cho một hình ảnh có chứa đối tượng.
- Chỉ số mAP\_0.5 (mean Average Precision ứng với ngưỡng IoU > 0.5) và chỉ số mAP\_0.5:0.95 (giá trị của mean Average Precision ứng với 10 ngưỡng: IoU > 0.5, IoU>0.55, IoU>0.60...IoU>0.95) cho thấy mô hình YOLOv5s đã cho ra độ chính xác cao.

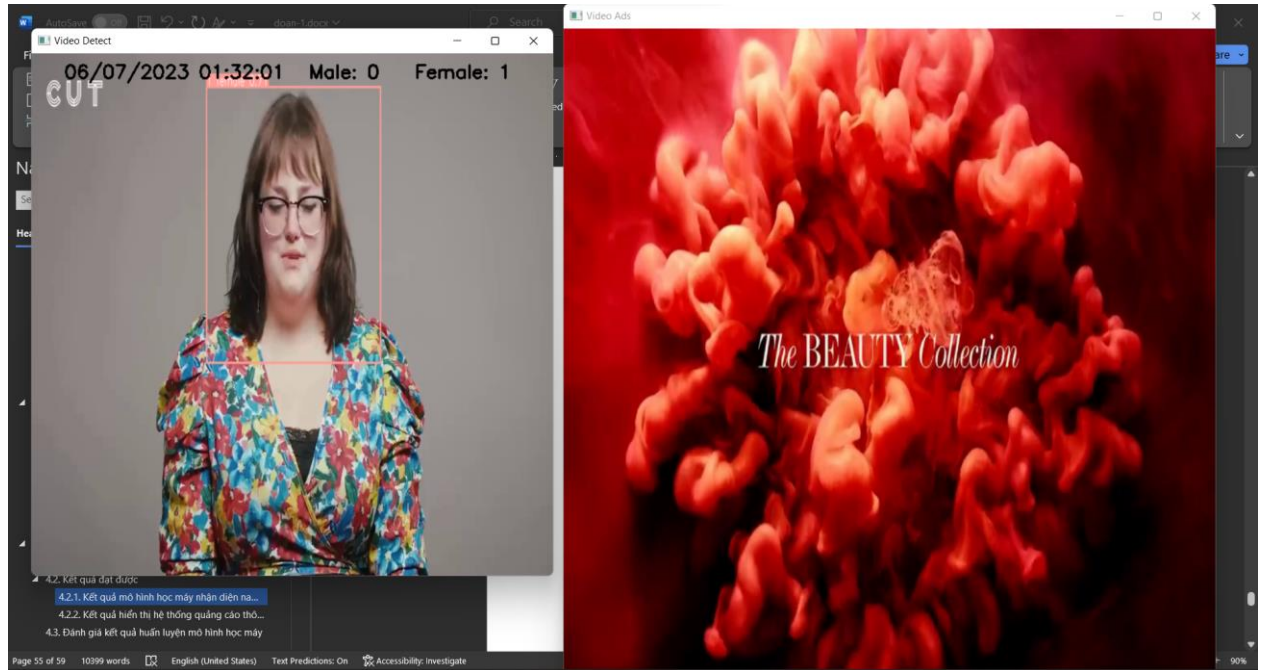
Dưới đây là một số hình ảnh nhận diện trên mỗi lớp tương ứng của các mô hình:



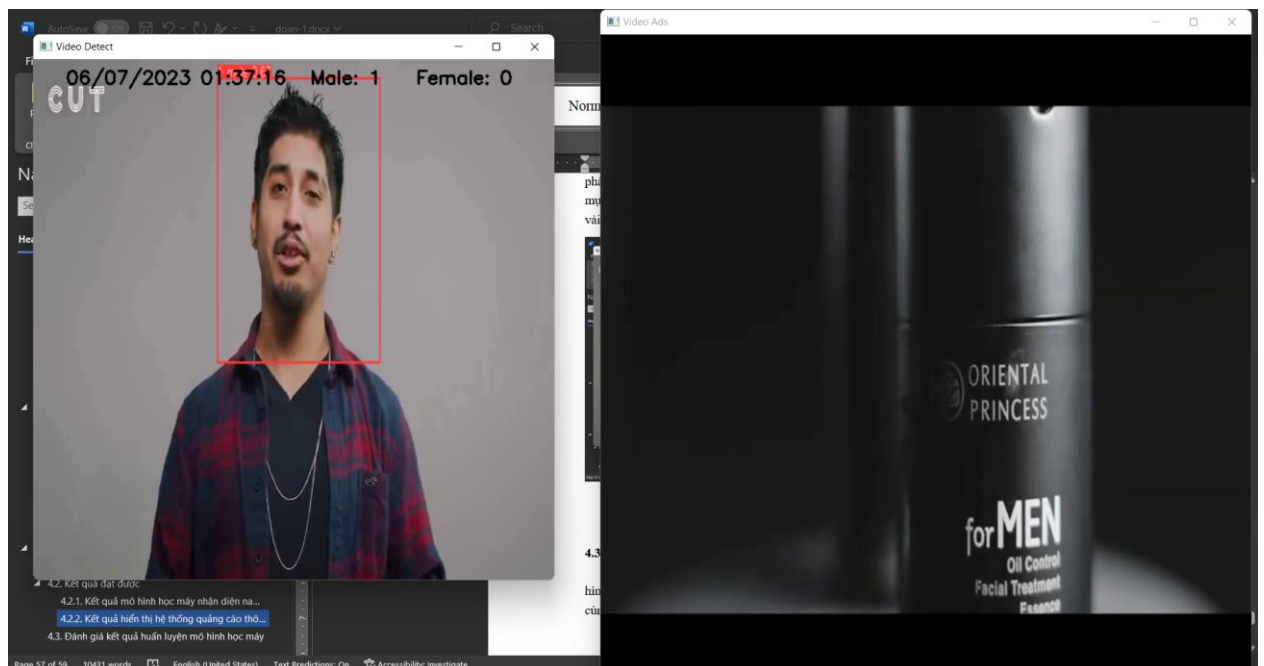
Hình 4.7: Kết quả nhận dạng của mô hình YOLOv5s

#### 4.2.2. Kết quả hiển thị hệ thống quảng cáo thông minh

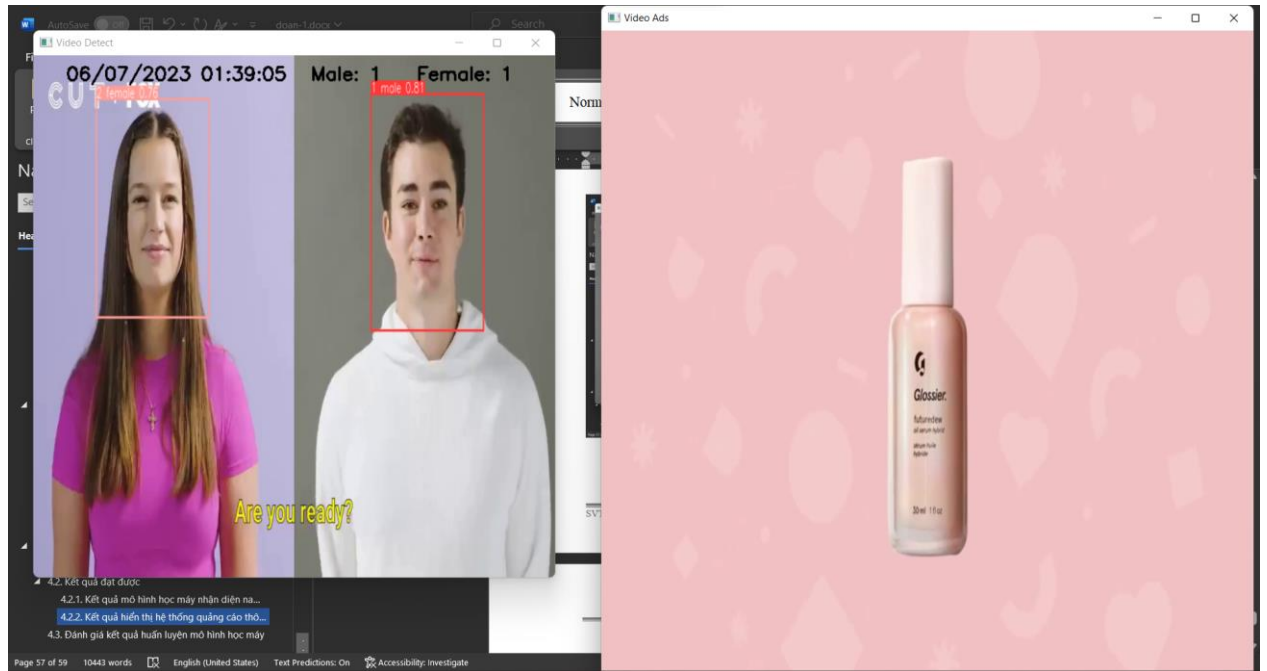
Sau khi huấn luyện mô hình nhận dạng khuôn mặt, chúng em tiếp tục dựa vào đó để phân loại theo giới tính và so sánh số lượng của đối tượng nam nữ giữa hai lớp nhằm mục đích hiển thị quảng cáo phù hợp với đối tượng nhận dạng được. Sau đây là một vài hình ảnh hiển thị kết quả đạt được của hệ thống quảng cáo thông minh.



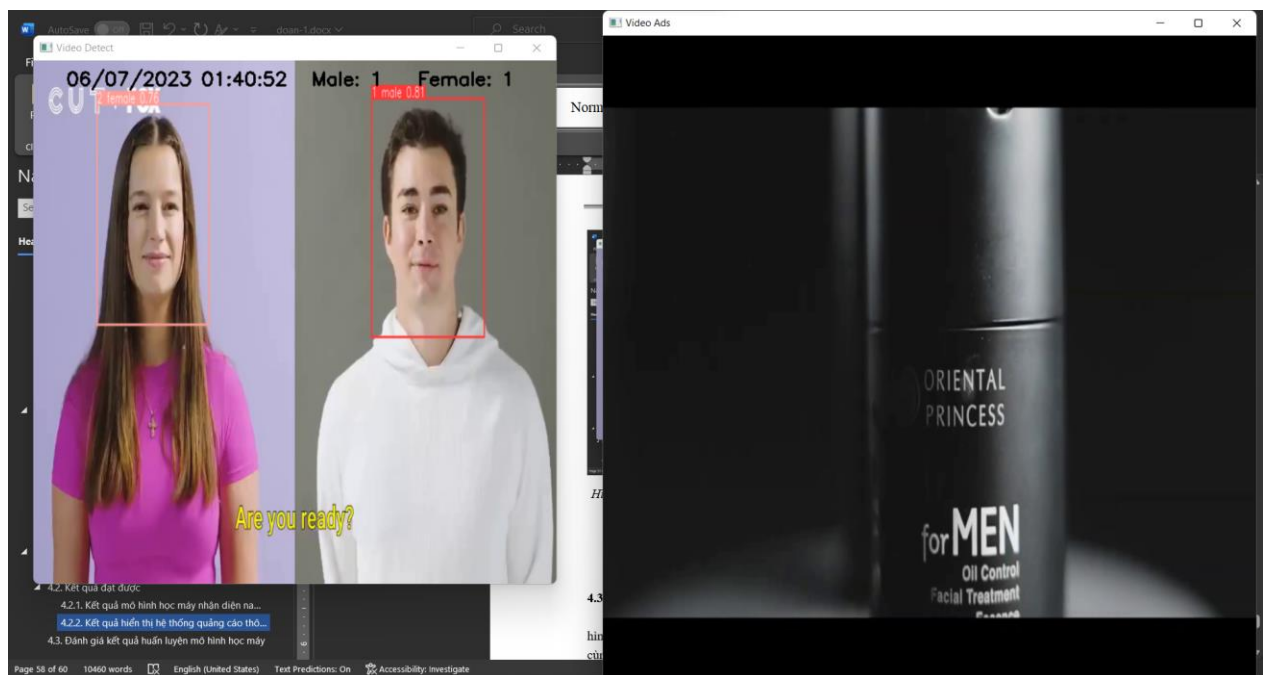
Hình 4.8: Kết quả hiển thị quảng cáo dành cho nữ giới



Hình 4.9: Kết quả hiển thị quảng cáo dành cho nam giới



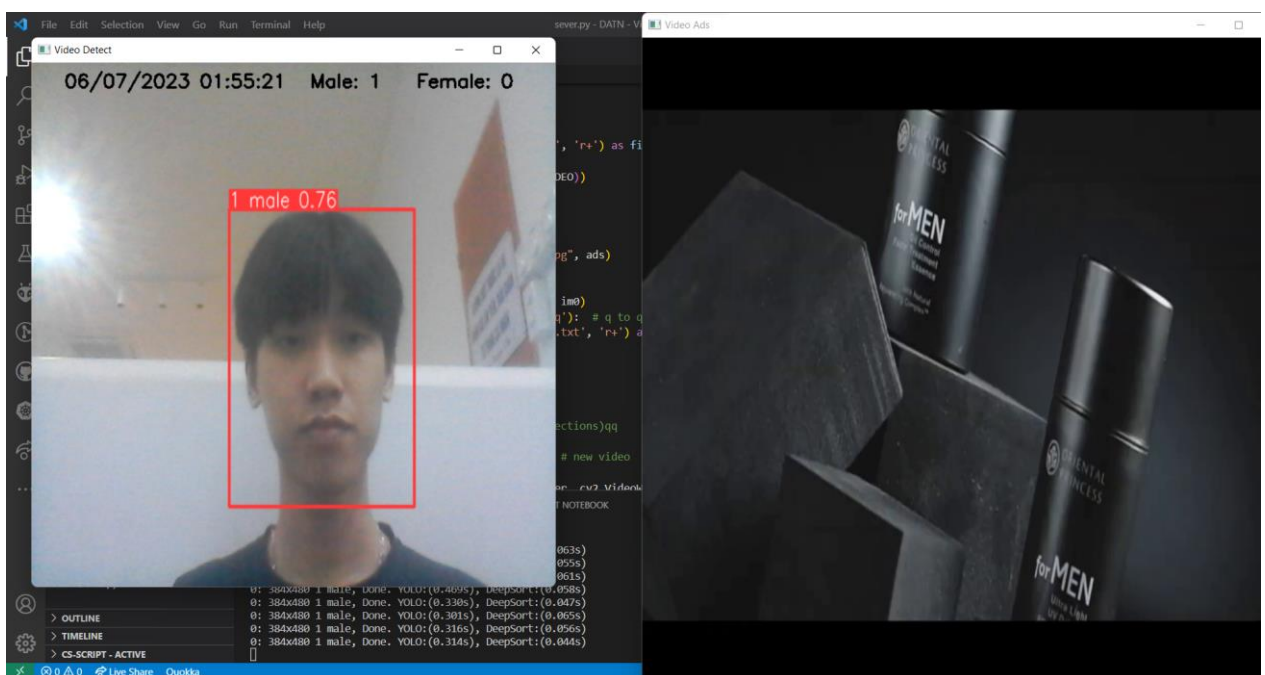
Hình 4.10: Kết quả hiển thị quảng cáo ngẫu nhiên khi số lượng nam nữ bằng nhau



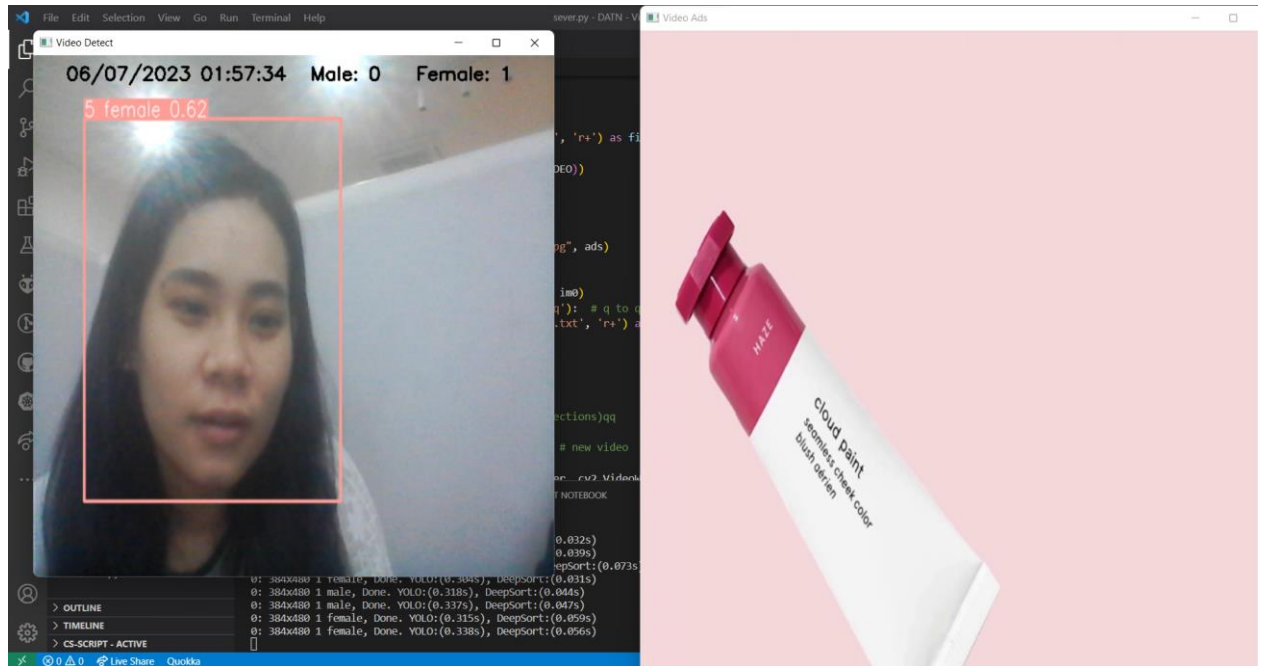
Hình 4.11: Kết quả hiển thị quảng cáo ngẫu nhiên khi số lượng nam nữ bằng nhau



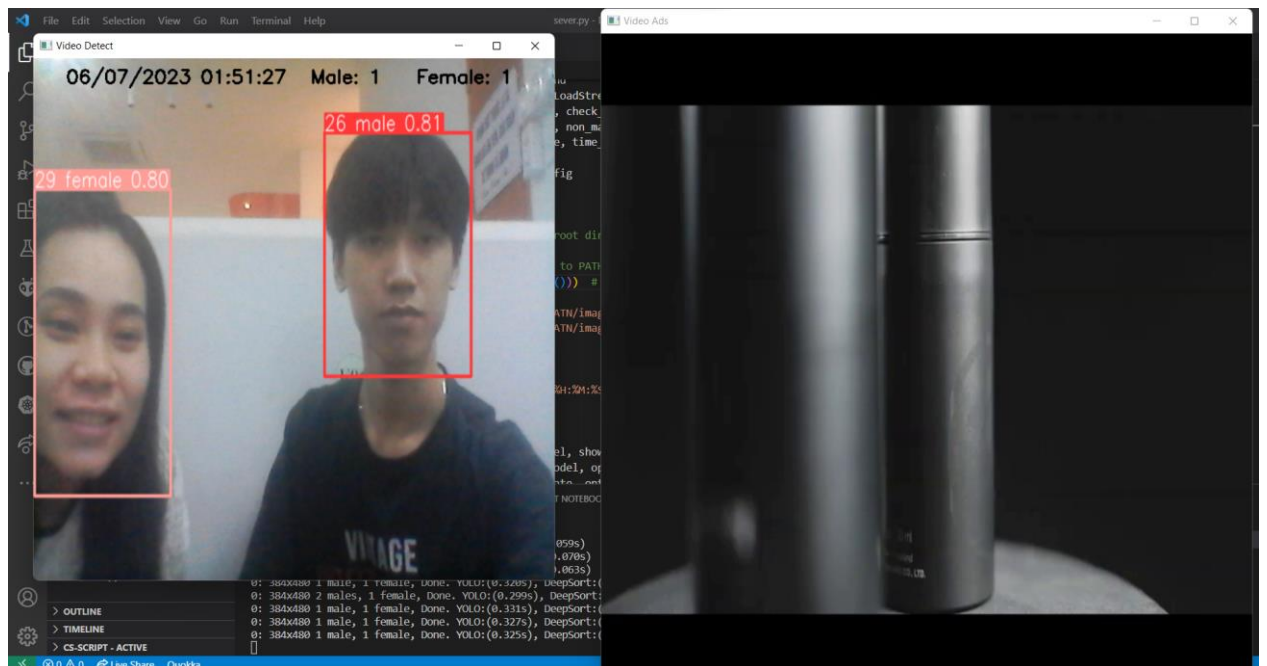
Hình 4.12: Kết quả hiển thị quảng cáo khi có nhiều đối tượng trong khung hình



Hình 4.13: Kết quả hiển thị quảng cáo dành cho nam giới trong thực tế



Hình 4.14: Kết quả hiển thị quảng cáo dành cho nữ giới trong thực tế

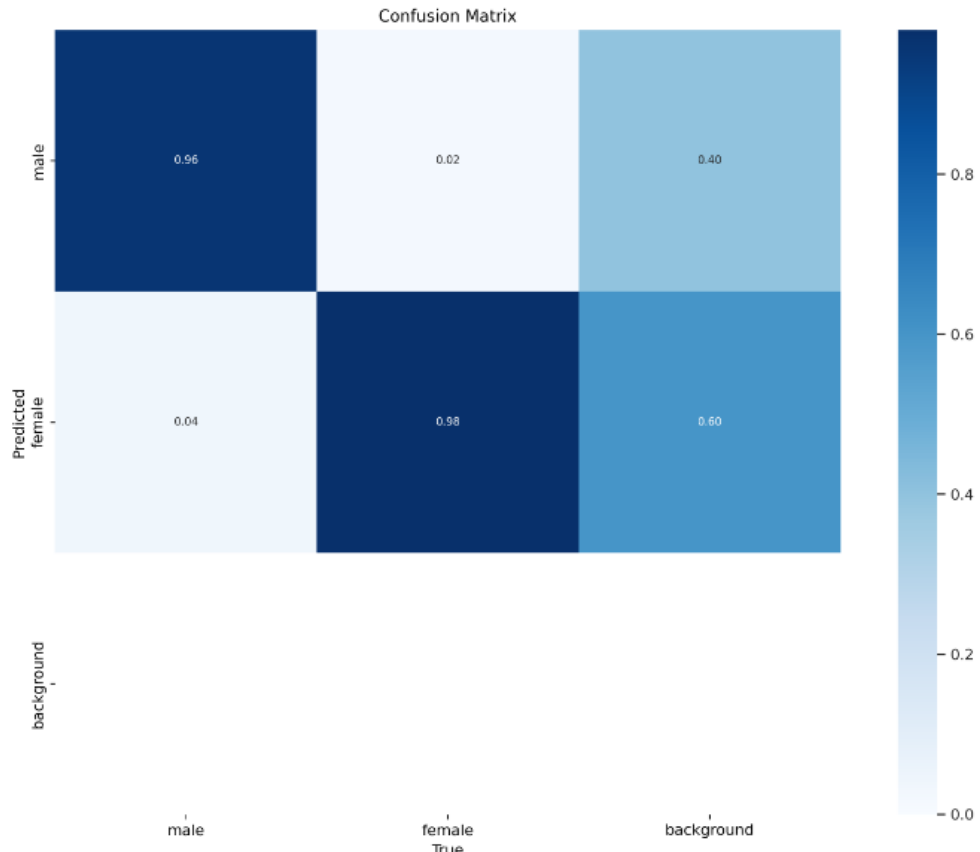


Hình 4.15: Kết quả hiển thị quảng cáo ngẫu nhiên trong thực tế

### 4.3. Đánh giá kết quả

#### 4.3.1. Đánh giá kết quả mô hình học máy nhận dạng khuôn mặt nam nữ

Mô hình YOLOv5s có thông số huấn luyện rất tốt, phù hợp để chọn làm mô hình triển khai cho hệ thống. Để rõ hơn về chi tiết các đánh giá về mô hình YOLO cùng tìm hiểu ma trận nhầm lẫn của chúng.



Hình 4.16: Ma trận nhầm lẫn mô hình YOLOv5s

Ma trận nhầm lẫn cho thấy mô hình có kết quả dự đoán chính xác rất cao trên các lớp. Tỷ lệ các đối tượng bị bỏ sót giữa các lớp là khá cân bằng (mô hình không nhận ra được) khoảng 4% cho mỗi lớp. Việc giảm tỷ lệ các đối tượng bị bỏ sót có tầm quan trọng rất lớn trong quá trình hoạt động thực tế vì nó sẽ giảm thiểu khả năng không nhận dạng được giới tính của đối tượng từ đó đưa ra quảng cáo không phù hợp với đối tượng.

#### **4.3.2. Đánh giá kết quả hệ thống quảng cáo thông minh**

Dựa vào kết quả đạt được ở trên, ta có thể thấy hệ thống hoạt động một cách trơn tru với độ chính xác cao. Điều này có nghĩa là mô hình YOLOv5s đã cho kết quả nhận dạng và phát hiện đối tượng chính xác và đáng tin cậy.

Ngoài ra, nhóm chúng em đã sử dụng truyền dữ liệu từ quá trình nhận dạng sang một tập tin trung gian. Điều này giúp tối ưu hóa việc nhận dữ liệu và tốc độ xử lý của hệ thống. Thay vì xử lý dữ liệu trực tiếp từ video nhận diện nam nữ, việc sử dụng tập tin trung gian giúp giảm thời gian xử lý và đảm bảo rằng việc hiển thị video quảng cáo được thực hiện nhanh chóng.

Điều này giúp tối ưu hóa quá trình phát quảng cáo một cách nhanh chóng và phù hợp với đối tượng được nhắm đến. Thời gian xử lý nhanh cho phép phát quảng cáo ngay sau khi nhận diện, mang lại hiệu quả và tương tác tốt với đối tượng mục tiêu.

#### **4.4. Kết luận chương**

Qua chương này, nhóm chúng em đã trình bày được kết quả nghiên cứu cũng như là các kết luận, bàn luận về mô hình được xây dựng trong đồ án này. Dựa trên kết quả đạt được và mục tiêu đã đề ra, chúng em nhận định được kết quả đạt được là tương đối tốt và có thể áp dụng hệ thống vào trong thực tế.

## KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN ĐỀ TÀI

Trong thời gian thực hiện đồ án, nhóm chúng em đã thiết kế “Hệ thống quảng cáo thông minh” khá hoàn chỉnh và đầy đủ đúng với những yêu cầu đề ra của đề tài. Tuy hệ thống chỉ dừng lại dưới dạng demo sản phẩm nhưng cũng đã thực hiện được đầy đủ các chức năng mà hệ thống đặt ra.

### **Ưu điểm:**

- Hệ thống có khả năng phân loại được nam và nữ, so sánh số lượng và đưa ra quảng cáo dựa vào số liệu được xử lý.
- Tốc độ xử lý sau khi nhận tín hiệu từ video nhận dạng đến lúc hiển thị quảng cáo gần như không có độ trễ.
- Có thể áp dụng vào thực tế.

### **Hạn chế:**

- Hệ thống bị ảnh hưởng bởi một số tác nhân bên ngoài như là ánh sáng, góc để mặt, khoảng cách với camera, camera độ phân giải thấp.
- Kết quả phân loại với độ chính xác lên đến 96% nhưng vẫn chưa phải là tuyệt đối.

### **Hướng phát triển đề tài:**

- Để cải thiện đề tài này, chúng em tiếp tục tìm hiểu và kết hợp các phương pháp khác nhằm tăng độ chính xác trong quá trình phân loại và nhận diện nam nữ không chỉ trên giới tính mà còn ở độ tuổi.
- Thêm chức năng cảnh tổng hợp dữ liệu và phân tích hành vi người dùng để tối ưu hóa quảng cáo.
- Có thể phát triển sản phẩm thêm tính năng giao tiếp và tương tác : Xây dựng chatbot hoặc trợ lý ảo có khả năng tương tác và giao tiếp với khách hàng. Điều này giúp nâng cao trải nghiệm của khách hàng và cung cấp thông tin quảng cáo theo cách dễ dàng và thuận tiện.

## TÀI LIỆU THAM KHẢO

- [1] Amidi, A., & Amidi, S. (2018). VIP Cheatsheet: Convolutional Neural Networks. *Deep Learn.*: [CS 230 - Convolutional Neural Networks Cheatsheet \(stanford.edu\)](#)
- [2] Vaibhaw Singh Chandel, *Selective Search for Object Detection (C++ / Python)*, 2017. Available: [Selective Search for Object Detection \(C++ / Python\) | LearnOpenCV](#) [Accessed: June 15, 2023]
- [3] Kaggle Dataset Human Faces: [Human Faces | Kaggle](#)
- [4] Model YOLOv5: [ultralytics/yolov5: YOLOv5 🚀 in PyTorch > ONNX > CoreML > TFLite \(github.com\)](#)
- [5] Object Tracking with DeepSORT |Authors: Shoeb AhmadAccenture: [\(PDF\) Object Tracking with DeepSORT | \(researchgate.net\)](#)
- [6] Glenn Jocher, *Train Custom Data*, 2023. Available: <https://github.com/ultralytics/yolov5/wiki/Train-Custom-Data> [Accessed: April 28, 2023]
- [7] Benchmarking YOLOv5 and YOLOv7 models with DeepSORT for droplet tracking applications: [Benchmarking YOLOv5 and YOLOv7 models with DeepSORT for droplet tracking applications | SpringerLink](#)