TRƯỜNG KỸ THUẬT VÀ CÔNG NGHỆ

**KHOA CÔNG NGHỆ THÔNG TIN**

****

**THỰC TẬP ĐỒ ÁN CƠ SƠ NGÀNH**

**HỌC KỲ I, NĂM HỌC 2025-2026**

**TÌM HIỂU VỀ YOLOV11 VÀ ỨNG DỤNG PHÁT HIỆN VÀ ĐẾM SỐ LƯỢNG**

**PHƯƠNG TIỆN**

*Giảng viên hướng dẫn:*

ThS. Đoàn Phước Miền

*Sinh viên thực hiện:*

Họ tên: Phan Khánh Duy

MSSV: 110123082

Lớp: DA23TTB

***Vĩnh Long, tháng 12 năm 2025***

TRƯỜNG KỸ THUẬT VÀ CÔNG NGHỆ

**KHOA CÔNG NGHỆ THÔNG TIN**

**A blue and white logo

AI-generated content may be incorrect.**

**THỰC TẬP ĐỒ ÁN CƠ SƠ NGÀNH**

**HỌC KỲ I, NĂM HỌC 2025-2026**

**TÌM HIỂU VỀ YOLOV11 VÀ ỨNG DỤNG PHÁT HIỆN VÀ ĐẾM SỐ LƯỢNG**

**PHƯƠNG TIỆN**

*Giảng viên hướng dẫn:*

ThS. Đoàn Phước Miền

*Sinh viên thực hiện:*

Họ tên: Phan Khánh Duy

MSSV: 110123082

Lớp: DA23TTB

***Vĩnh Long, tháng 12 năm 2025***

**NHẬN XÉT CỦA GIẢNG VIÊN HƯỚNG DẪN**

*Vĩnh Long, ngày ….. tháng …… năm ……*

**Giảng viên hướng dẫn**

*(Ký tên và ghi rõ họ tên)*

*Trà Vinh, ngày ….. tháng …… năm ……*

**Giáo viên hướng dẫn**

*(Ký tên và ghi rõ họ tên)*

**LỜI CẢM ƠN**

**NHẬN XÉT CỦA THÀNH VIÊN HỘI ĐỒNG**

*Vĩnh Long, ngày ….. tháng …… năm ……*

**Thành viên hội đồng**

*(Ký tên và ghi rõ họ tên)*

Em xin chân thành gửi lời cám ơn đến thầy cô trường Đại học Trà Vinh nói chung và các thầy cô bộ môn trong khoa kỹ thuật và công nghệ nói riêng đã tạo điều kiện cho chúng em cơ hội thực hành, tiếp xúc để chúng em có thể tránh được những vướng mắc và bỡ ngỡ trong môi trường công việc thời gian tới.

Em xin chân thành cảm ơn thầy Đoàn Phước Miền. Nhờ sự giúp đỡ tận tình và những chỉ bảo của Thầy từu lúc bắt đầu cho tới lúc kết thúc đồ án mà em đã hoàn thành đúng thời hạn quy định và tích lũy được cho mình một lượng nền tảng kiến thức quý báu.

Mặc dù đã cố gắng hoàn thành đề tài tốt nhất nhưng do thời gian và kiến thức chuyên ngành còn hạn chế nên em vẫn còn nhiều thiếu sót khi tìm hiểu, đánh giá và trình về đề tài. Rất mong được sự quan tâm, góp ý của các thầy cô giảng việc bộ môn để đề tài của em được hoàn chỉnh và đầy đủ hơn.

Em xin chân thành cảm ơn!

**MỤC LỤC**

[MỞ ĐẦU 9](#_Toc217549219)

[1. Lý do chọn đề tài 9](#_Toc217549220)

[2. Mục tiêu nghiên cứu 9](#_Toc217549221)

[3. Đối tượng và pham vi nghiên cứu 9](#_Toc217549222)

[CHƯƠNG 1 : TỔNG QUAN 10](#_Toc217549223)

[1.1 Giới thiệu tổng quan về đề tài 10](#_Toc217549224)

[1.2 Mục tiêu của đề tài 10](#_Toc217549225)

[1.3 Bài toán cần giải quyết 10](#_Toc217549226)

[1.4 Giải quyết bài toán 10](#_Toc217549227)

[CHƯƠNG 2 : CƠ SỞ LÝ THUYẾT 11](#_Toc217549228)

[2.1 Tổng quan về thị giác máy tính 11](#_Toc217549229)

[2.1.1 Khái niệm về thị giác máy tính 11](#_Toc217549230)

[2.1.2 Nguyên lý hoạt động cơ bản của thị giác máy tính 11](#_Toc217549231)

[2.1.3 Ứng dụng của thị giác máy tính 11](#_Toc217549232)

[2.2 Tổng quan về học sâu (Deep Learning) 12](#_Toc217549233)

[2.2.1 Khái niệm học sâu 12](#_Toc217549234)

[2.2.2 Mạng nơ-ron tính chập (Convolutional Neural Network - CNN) 12](#_Toc217549235)

[2.3 Bài toán phát hiện đối tượng (Object Detection) 13](#_Toc217549236)

[2.3.1 Khái niệm 13](#_Toc217549237)

[2.3.2 Các hướng tiếp cận 13](#_Toc217549238)

[2.4 Giới thiệu mô hình YOLO (You Only Look Once) 13](#_Toc217549239)

[2.4.1 Nguyên lý hoạt động của YOLO 13](#_Toc217549240)

[2.4.2 Sự phát triển của các phiên bản YOLO 13](#_Toc217549241)

[2.5 Cấu trúc và đặc điểm của YOLOv11 14](#_Toc217549242)

[2.5.1 Kiến trúc tổng thể của YOLOv11 14](#_Toc217549243)

[2.5.2 Cải tiến nổi bật của YOLOv11 15](#_Toc217549244)

[2.5.3 Ứng dụng YOLOv11 trong thực tế 15](#_Toc217549245)

[CHƯƠNG 3 : HIỆN THỰC HÓA NGHIÊN CỨU 16](#_Toc217549246)

[3.1 Mô tả bài toán 16](#_Toc217549247)

[3.2 Xây dựng mô hình 16](#_Toc217549248)

[3.2.1 Nguồn dữ liệu 16](#_Toc217549249)

[3.2.2 Gán nhãn và khung bao 16](#_Toc217549250)

[3.2.3 Huấn luyện mô hình YOLOv11 16](#_Toc217549251)

[3.2.4 Kiểm tra mô hình 17](#_Toc217549252)

[CHƯƠNG 4 : KẾT QUẢ NGHIÊN CỨU 18](#_Toc217549253)

[4.1 Kết quả huấn luyện mô hình YOLOv11 18](#_Toc217549254)

[4.2 Đánh giá độ chính xác 18](#_Toc217549255)

[4.3 Nhận xét chung về mô hình 18](#_Toc217549256)

[4.4 Sơ đồ 19](#_Toc217549257)

[CHƯƠNG 5: KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN 20](#_Toc217549258)

[DANH MỤC TÀI LIỆU THAM KHẢO 21](#_Toc217549259)

[PHỤ LỤC 22](#_Toc217549260)

**DANH MỤC HÌNH ẢNH**

**DANH MỤC BẢNG BIỂU**

**TÓM TẮT ĐỒ ÁN CƠ SỞ NGÀNH**

Đồ án cơ sở ngành với đề tài “Tìm hiểu về YOLOv11 và ứng dụng phát hiện và đếm số lượng phương tiện giao thông” được thực hiện nhằm nghiên cứu và tiếp cận các kiến thức về thị giác máy tính và học sâu, đồng thời tìm hiểu khả năng ứng dụng mô hình YOLOv11 vào giao thông thông minh.

Nội dung của đồ án trình bày tổng quan về mô hình YOLOv11, cơ sở lý thuyết liên quan đến mạng nơ-ron tính chập và phương pháp phát hiện đối tượng, từ đó xây dựng, huấn luyện mô hình phát hiện và đếm số lượng phương tiện từ dữ liệu hình ảnh hoặc video mô phỏng giao thông.

Kết quả đạt được cho thấy mô hình YOLOv11 có khả năng phát hiện chính xác các loại phương tiện giao thông từ dữ liệu đưa vào huấn luyện với tốc độ xử lý nhanh và đáp ứng yêu cầu của các ứng dụng giám sát giao thông. Đồ án góp phần giúp hiểu rõ hơn về quy trình nghiên cứu, triển khai và ứng dụng mô hình học sau trong lĩnh vực công nghệ thông tin.

# MỞ ĐẦU

## 1. Lý do chọn đề tài

Trong bối cảnh đô thị hóa nhanh chóng, mật độ tham gia giao thông của các phương tiên ngày càng tăng khiến việc quản lý và giám sát trở nên khó khăn và là một vấn đề cấp thiết. Các hệ thống giám sát cảm biến vật lý và các hệ thống thủ công thường có chi phí cao, thiếu linh hoạt và khó triển khai diện rộng. Trong khi đó việc triển khai một hệ thống giám sát trí tuệ nhân tạo (AI) đặc biệt là hệ thống thị giác máy tính (Computer Vision) giúp tự động hóa việc giám sát, nhận diện và phân tích trong thời gian thực. YOLOv11 là phiên bản mới nhất của mô hình YOLO nổi bật nhờ tốc độ xử lý và độ chính xác rất cao. Tuy nhiên YOLOv11 khá phát hiện các đối tượng nhỏ nhưng nhờ vào độ linh hoạt, tốc độ xử lý cao và giám sát theo thời gian thực nên em chọn đề “Tìm hiểu về YOLOv11 và ứng dụng phát hiện và đếm số lượng phương tiện giao thông”.

## 2. Mục tiêu nghiên cứu

Tìm hiểu mô hình YOLOv11 và khả năng ứng dụng vào phát hiện và đếm số lượng phương tiện giao thông.

## 3. Đối tượng và pham vi nghiên cứu

Đối tượng: mô hình phát hiên đối tượng YOLOv11, nguyên lý hoạt động, đặc điểm kỹ thuật và hiệu quả của mô hình trong việc giám sát thời gian thực.

Phạm vi nghiên cứu: sử dụng mô hình YOLOv11 tự huấn luyện để phát hiện và đếm số lượng phương tiện qua hình ảnh hoặc video mô phỏng.

# : TỔNG QUAN

## Giới thiệu tổng quan về đề tài

Sự phát triển mạnh mẽ của giao thông đô thị trong những năm gần đây đã đặt ra yêu cầu ngày càng cao về công tác giám sát và quản lý lưu lượng phương tiện. Việc nắm bắt chính xác sô lượng và loại hình phương tiện đóng vai trò quan trọng. Tuy nhiên các phương pháp truyền thống gặp nhiều hạn chế về hiêu quả và khó thích ứng với điều kiện giao thông phức tạp.

Cùng với sự phát triển của trí tuệ nhân tạo, đặc biệt lĩnh vực thị giác máy tính, nhiều giải pháp phân tích hình ảnh và video đã được ứng dụng rộng rãi. Trong đó, mô hình YOLO khá nổi bật với khả năng phát hiện đối tượng nhanh và chính xác, phù hợp cho hệ thống giám sát thời gian thực.

Đề tài “Tìm hiểu về YOLOv11 và ứng dụng phát hiện và đếm số lượng phương tiện giao thông” tập trung tự huấn luyện mô hình thử nghiệm có khả năng phát hiện và đếm số lượng phương tiện giao thông từ ảnh và video mô phỏng.

## Mục tiêu của đề tài

Mục tiêu của đề tài là tự huấn luyện mô hình phát hiện và đếm số lượng phương tiện bằng cách sử dụng YOLOv11, các đáp ứng cơ bản của mô hình bao gồm:

- Củng cố kiến thức về thị giác máy tính và học sâu.

- Nâng cao khả năng lập trình và triển khai mô hình.

- Làm quen với việc xử lý và phân tích dữ liệu giao thông

- Nhận diện chính xác được các loại phương tiện giao thông trong cùng khung hình, trong nhiều điều kiện khác nhau.

- Tự động đếm số lượng xe và thống kê theo thời gian thực nhất định.

## Bài toán cần giải quyết

Bài toán được đặt ra là huấn luyện mô hình với các yếu tố cơ bản:

- Phát hiện phương tiện giao thông.

- Đếm số lượng phương tiện giao thông.

- Đảm bảo tính chính xác thời gian thực.

- Khả năng đáp ứng yêu cầu trong môi trường thực tế

## Giải quyết bài toán

Đề tài sử dụng mô hình phát hiện đối tượng YOLOv11 làm nền tảng chính để giải quyết bài toán trên, bao gồm:

- Thu thập và xử lý dữ liệu: Dữ liệu đầu vào của hệ thống và hình ảnh và video mô phỏng giao thông được thu thập từ nhiều nguồn. Ảnh và video được chia thành các khung hình, sau đó thực hiện bước tiền xử lý để thay đổi kích thước phù hợp với yêu cầu của mô hình YOLOv11.

- Phát hiện phương tiện giao thông: Các khung hình được xử lí đưa vào mô hình YOLOv11 để thực hiện phát hiện đối tượng. Mô hình xác định vị trí và đánh dấu lại bằng các khung bao quanh phương tiện và đánh giá mức độ chính xác ttreen khung bao.

- Đếm số lượng phương tiện giao thông: Hệ thống sẽ tổng hợp số lượng từng loại phương tiện xuất hiện trong ảnh hoặc video.

- Hiển thị và đánh giá kết quả: Kết quả đầu ra là hình ảnh hoặc video được hiển thị kèm khung bao, nhãn phân loại từng phương tiện giao thông khác nhau.

# : CƠ SỞ LÝ THUYẾT

## Tổng quan về thị giác máy tính

### Khái niệm về thị giác máy tính

Thị giác máy tính (Computer Vision) là một lĩnh vực trong trí tuệ nhân tạo (AI) nhằm giúp máy tính có khả năng “nhìn thấy” và “phân tích” hình ảnh hoặc video tương tự như con người. Nói cách khác, Computer Vision cho phép máy tính trích xuất thông tin hữu ích từ dữ liệu hình ảnh hoặc video, phục vụ cho các mục tiêu nhưu nhận dạng, phân loại, định vị hoặc theo dõi đối tượng.

### Nguyên lý hoạt động cơ bản của thị giác máy tính

Hệ thống thị giác máy tính sử dụng các thuật toán phức tạp, đặc biệt là mô hình Học Sâu (Deep Learning) như mạng nơ-ron tích chập (Convolution Neural Networks-CNN), để xử lý dữ liệu hình ảnh hoặc video qua các bước:

Bước 1 Thu hập dữ liệu: Hình ảnh hoặc video được thu thập từ camera, cảm biến hoặc các nguồn dữ liệu có sẳn.

Bước 2 Tiền xử lý dữ liệu: Hình ảnh hoặc video có thể được chuẩn hóa, giảm nhiễu, điều chỉnh độ sáng/độ tương phản để cải thiện chất lượng.

Bước 3 Trích xuất đặc trưng: Hệ thống nhận diện và tách ra yếu tố cơ bản trong ảnh như đường viền, góc cạn, hình dạng.

Bước 4 Phân tích và nhận dạng: Dựa trên các đặc trưng đã trích xuẩ, mô hình sẽ phân tích để xác định, phân loại đối tượng, hoặc theo dõi chuyển động của chúng.

Bước 5 Đưa ra quyết định: Dựa trên kết quả phân tích, hệ thống sẽ thực hiện hành động đã được lập trình.

### Ứng dụng của thị giác máy tính

Thị giác máy tính được ứng dụng nhiều và đa dạng trong đời sống công nghiệp:

Giám sát an ninh (Security & Sủveillance):

- Kiểm sát các truy cập bằng khôn mặt tại các khu vực cần an ninh cao như: sân bay, phòng thí nghiệm…

- Phân tích và đưa ra cảnh báo với các tình huống bất thường từ camera.

- Theo dõi đường đi của đối tượng cụ thể trong một khu vực lớn.

Y tế (Healthcare):

- Phân tích và phát hiện các bất thường trong cơ thể con người hoặc động vật như: khối u, tổn thương hoặc các bệnh lý từ các hình ảnh X-quang…

- Cung cấp các hướng dẫn trực quan trong thời gian thực cho bác sĩ hoặc công cụ phẫu thuật.

Công nghiệp và sản xuất (Manufacturing & Industrial):

- Phát hiện các khuyết điểm nhỏ trên các sản phẩm với tốc độ và độ chính xác cao.

- Phân tích và phát hiện các hành vi vi phạm qui định an toàn hoặc hành vi xâm nhập bất thường.

Giao thông vận tải (Transportation):

- Phát triển xe tự hành nhờ vào thị giác máy tính giúp xe có thể tự động phát hiện vât cản trong môi trường xung quanh.

- Phát hiện các hành vi vi phạm an toàn giao thông, đếm số lượng phương tiện.

## Tổng quan về học sâu (Deep Learning)

### Khái niệm học sâu

Học sâu (Deep Learning) là một nhánh của học máy (Machine Learning), dựa trên việc sử dụng mạng nơ-ron nhân tạo nhiều tầng (Aritificial Neural Network) để tự động trích xuất đặc trưng từu dữ liệu đầu vào và thực hiện các tác vụ nhận dạng, dự đoán.

Khác với các phương pháp truyện thống đòi hỏi người lập trình phải chọn thủ công các đặc trưng, Deep Learning có khả năng tự học đặc trưng (Feature Learing) thông qua quá trình huấn luyện với dữ liệu lớn.

Trong lĩnh vực thị giác máy tính, học sâu đã tạo ra bước đột phá lớn, cho phép máy tính không chỉ nhận dạng ảnh với độ chính xác cao mà còn phát hiện, theo dõi, phân đoạn và hiểu ngữ cảnh hình ảnh.

### Mạng nơ-ron tính chập (Convolutional Neural Network - CNN)

Mạng nơ-ron tính chập (CNN) là nền tảng của hầu hết các hệ thống thị giác máy tính hiện nay. CNN mô phỏng cách con người nhìn và xử lý hình ảnh bằng các trích xuất đặc trưng từ mức độ thấp đến cao.

Một mạng CNN điển hình được cáu tạo từ các lớp cơ bản xếp chồng lên nhau:

1. Lớp tính chập (Convolutional Layer)

Chức năng: Thực hiện phép toán chập giữa một bộ lọc (filter).

Nguyên lý: Bộ lọc sẽ quét qua toàn bộ đầu vào, tính roán tích vô hướng giữa các trọng số của nó và các pixel đầu vào, tạo ra một bản đồ đặc trưng (feature map). Mỗi bộ lọc được thiết kế để phát hiện một đặc trưng cụ thể nào đó.

Đầu ra của lớp tính chập là bản đồ dặc trưng (feature map).

2. Lớp hoạt hóa (Activation Layer)

Chức năng: Áp dụng một hàm phi tuyến (non-linear function) cho đầu ra của lớp tính chập, giúp mô hình học các mối quna hệ phức tạp và phi tính trong dữ liệu.

Phổ biến: Hàm ReLU (Rectified Linearr Unit) là hàm hoạt hóa được sử dụng rộng rãi nhất do hiệu quả tính toán và khả năng giảm thiểu vấn đề gradient vanishing.

3. Lớp tổng hợp (Pooling Layer / Subsampling Layer)

Chức năng: Giảm kích thước không gian của bản đồ đặc trứng, qua đó giảm số lượng tham số và tính toán trong mạng, dồng thời kiểm soát học mức

Phổ biến:

- Max Pooling: lấy giá trị lớn nhất trong số cửa sổ trượt.

- Average Pooling: Lấy giá trị trung bình một cửa sổ trượt.

4. Các lớp cuối (Fully-Connected Layers và Output)

Sau khi trích xuất đặc trưng qua các lớp tính chập và tổng hợp, dữ liệu được làm phẳng (flatten) thành một vector.

Vector sau đó được đưa vào một hoặc nhiều lớp kết nối đầy đủ (Fully-Connected-FC) để học mối quan hệ giauwx các đặc trưng và cuối cùng là đưa ra dự đoán.

Lớp đầu ra thường sử dụng hàm Softmax để đưa ra đề xuất từng lớp.

## Bài toán phát hiện đối tượng (Object Detection)

### Khái niệm

Phát hiện đối tượng (Object Detection) là một bài toán then chốt trong thị giác máy tính (Computer Vision) nhằm mục đích xác định vị trí loại đối tượng xuất hiện trong ảnh hoặc video.

Khác với phan loại ảnh (Image Classification) chỉ xác định ảnh, thì phát hiện đối tượng (Object Detection) yêu cầu mô hình phải xác định cả vị trí và nhãn của đối tượng bằng cách vẽ khung bao (Bounding Box) hình tứ giác xung quanh đối tượng.

### Các hướng tiếp cận

Có hai nhóm phương pháp tiếp cận chính: nhóm hai giai đoạn và nhóm một giai đoạn.

- Nhóm hai giai đoạn (Tow-Stage Detectors):

Giai đoạn 1: Xác định các vùng có khả năng chứa đối tượng (Region Proposal).

Giai đoạn 2: Phân loại và tinh chỉnh các vùng được đề xuất sau đó đưa vào mạng thứ 2 để phân loại đối tượng từng vùng, tinh chỉnh các khung bao để chính xác hóa vị trí.

Ví dụ: R-CNN, Fast R-CNN,…

- Nhóm một giai đoạn (One-Stage Detectors):

Mô hình trực tiếp dự đoán vị trí và nhãn đối tượng trong một lần duy nhất.

Ví dụ: SSD (Single Shot MultiBox Detector), RetinaNet và đặc biệt nhất là YOLO (You Only Look Once).

## Giới thiệu mô hình YOLO (You Only Look Once)

### Nguyên lý hoạt động của YOLO

YOLO là một mô hình đột phá trong lĩnh vực phát hiện đối tượng (Object Detection), nổi tiếng vì đã định nghĩa lại khái niệm phát hiện thời gian thực (real-time detection). Ra mắt lần đầu năm 2016 bởi Jose Redmon và cộng sự, YOLO được xem là đại diện tiêu biểu cho nhóm một giai đoạn nhờ vào tốc độ phát hiện nhanh và độ chính xác khá cao.

Thay vì tách biệt bước đề xuất vùng như R-CNN, YOLO chia sẻ ảnh đầu vào thành lưới (grid), mỗi ô trong lưới chịu trách nhiệm dự đoán khung bao (bounding box) và xác suất đối tượng nếu có vật trong đó.

Quy trình hoạt động cơ bản của YOLO:

- Ảnh đầu vào được chia thành các ô lưới.

- Mỗi ô dự đoán nhiều khung bao cùng với độ tin cậy.

- Các khung có xác xuất cao nhất được giữu lại.

- Kết quả cuối cùng là danh sách các đối tượng với nhãn và vị trí.

Ưu điểm chính của YOLO là tốc độ nhanh (có thể xử lý nhiều khung hình trên giây) đảm bảo tính chính xác và phát hiện thời gian thực.

### Sự phát triển của các phiên bản YOLO

|  |  |  |
| --- | --- | --- |
| Phiên bản | Năm | Đặc điểm nổi bật |
| YOLOv1 (2016) | 2016 | Mô hình đầu tiên, xử lý ảnh toàn cục trong một lần nhìn (single forward pass). |
| YOLOv2 (2017) | 2017 | Cải thiện độ chính xác, dùng Batch Normalization và Anchor Boxes. |
| YOLOv3 (2018) | 2018 | Hỗ trợ phát hiện đa tỉ lệ, sử dụng mạng Darknet-53. |
| YOLOv4 (2020) | 2020 | Tối ưu huấn luyện, tăng tốc bằng các kỹ thuật như Mosaic, DropBlock. |
| YOLOv5 (2020) | 2020 | Viết lại bằng PyTorch, dễ dùng, linh hoạt, tốc độ cao. |
| YOLOv7 (2022) | 2022 | Cải tiến kiến trúc, đạt kỷ lục về tốc độ và độ chính xác. |
| YOLOv8 (2023) | 2023 | Được Ultralytics phát triển, hỗ trợ nhiều tác vụ (detection, segmentation, pose). |
| YOLOv10 (2024) | 2024 | Tăng hiệu năng inference, giảm tính toán, cải thiện kiến trúc backbone. |
| YOLOv11 (2025) | 2025 | Phiên bản mới nhất, tốc độ nhanh hơn YOLOv8–10, độ chính xác cao hơn trên COCO dataset. |

## Cấu trúc và đặc điểm của YOLOv11

Mô hình YOLOv11 là phiên bản hiện đại nhất, được phát triển bởi Ultralytics. Kế thừa các ưu điểm của YOLOv8 và YOLOv10 để nâng cao hiệu suất phát hiện đối tượng thời gian thực băng cách cân bằng tối ưu tốc độ và độ chính xác.

### Kiến trúc tổng thể của YOLOv11

Backbone (Xương sống)

Chức năng: Trích xuất các đặc trưng cơ abnr và cao cấp từu ảnh hoặc video đầu vào.

Đặc điểm: YOLOv11 sử dụng các C3K2 block để xử lý đcặ trưng ở giai đoạn khác nhau.

(C3K2 block là sự cải tiến từ CSP (Cross Stage Partial) và C2f blocks, sử dụng các phép tính chập kernel 3x3 nhỏ hơn để tối ưu luồng thông tin và giảm chi phí tính toán, trong khi vẫn giữu được khả năng nắm bắt các đặc trưng thiết yếu.)

Neck (Cổ mạng):

- Chức năng: Kết hợp các đặc trưng từ cấp độ khác nhau của backbone để chuẩn bị cho việc dự đoán

- Đặc điểm: Kiến trúc neck được cải tiến bao gồm:

+ SPFF (Spatial Pyramid Pooling Fast): Module này cho phép mạng tông hợp thông tin bối cảnh từ các vùng có kích thước khác nhau, đặc biệt quan trọng để cải thiện khả năng phát hiện các đối tượng với nhiều kích cỡ khác nhau.

+ C2PSA Blocks (Cross Stage Partial with Spatial Attention): Đây là một đổi mới quan trọng. C2PSA tích hợp các cơ chế chú ý không gian (Spatial Attention) vào cấu trúc CSP. Điều này giúp mô hình tập trung hơn vào các vùng quan trọng trong hình ảnh, nâng cao dộ chính xác, đặc biệt là đối với các đối tượng nhỏ hoặc bị che khuất một phần.

Head (Đầu mạng)

- Chức năng: Sử dụng các đặc trưng từ Neck để đưa ra dự đoán cuối cùng về khung bao, điểm tin cậy và phân loại.

- Đặc điểm: YOLOv11 tiếp tục sử dụng phương pháp Anchor-free (không cần anchor boxes) và Decoupled Head (tách riêng các nhánh định vị và phân loại) để tối ưu hóa hiệu xuất.

### Cải tiến nổi bật của YOLOv11

Tích hợp cơ chế chú ý không gian (Spatial Attention): Đây được xemlaf cải tiến về kiến trúc của YOLOv11.

- Khối C2SPA: YOLOv11 tích hợp cơ chế chú ý không gian vào xử lý đặc trưng trong phần Neck.

- Lợi ích: giúp mô hình tập trung tài nguyên vào các khu vực chứa đối tượng, tăng khẳ năng định vị và phân loại.

Tối ưu hóa hiệu quả tính toán:

- Khối C3K2: được tinh chỉnh để tối ưu hóa luồng dữ liệu và giảm số lượng phép toán so với YOLOv8 sử dụng khối C2f.

- Lợi ích: Giúp giảm số lượng tham số, tăng tốc độ suy luận và đạt điểm mAP cao hơn so với YOLOv8 và YOLOv10 với ít tham số hơn, điều này mang lại hiệu quả cao mà không giảm độ chính xác.

Tăng cường phát hiện đa tỷ lệ

- Việc tích hợp và tối ưu module SFFF trong YOLOv11 đảm bảo mạng có thể tổng hợp thông tin bối cảnh ở nhiều tỷ lệ khác nhau một cách hiệu quả.

- Lợi ích: Cải thiện khả năng tổng quát hóa của mô hình giúp phát hiện đối tượng có kích thước đa dạng một cách nhất quán.

### Ứng dụng YOLOv11 trong thực tế

Nhờ vào tốc độ và độ chính xác cao, YOLOv11 được ứng dụng vào nhiều lĩnh vực vì nó cũng là phần của thị giác máy tính:

- Phát hiện và đếm số lượng phương tiện giao thông.

- Giám sát an ninh, nhân diện khuôn mặt.

- Nhận dạng động vật, cây trồng trong công nghiệp.

- Kiểm tra lỗi sản phẩm trong công nghiệp

# : HIỆN THỰC HÓA NGHIÊN CỨU

## Mô tả bài toán

Trong bối cảnh hiện nay nhu cầu tự động hóa ngày càng tăng, các hệ thống giám sát trở nên ngày càng quan trọng. Các hệ thống dựa trên cảm biến vật lý thường có chi phí rất cao, thiếu sự llinh hoạt và khó triển khai rộng rãi. Nhờ sự phát triển mạnh của trí tuệ nhân tạo và thị giác máy tính, các mô hình học sâu đã trở thành giải pháp rất hiệu quả trong việc giám sát giao thông theo thời gian thực.

Trong các mô hình phát hiện đối tượng (Object Detection), mô hình YOLO (You Only Look Once) nổi bật nhờ tốc độ xử lý và độ chính xác cao. Phiên bản YOLOv11 do Ultralytics phát triển đem đến nhiều cải tiến về kiến trúc, tối ưu và hiệu năng cao so với các phiên bản khác.

Việc sử dụng YOLOv11 vào phát hiện và đếm số lượng phương tiện nhờ khả năng nhận dạng tốt trong các môi trường phức tạp, đặc biệt đây là mô hình AI hiện đại có thể dễ thực hành lập trình huấn luyện và xử lý dữ liệu thực tế, từ đó nâng cao khả năng giám sat cho mô hình.

## Xây dựng mô hình

### Nguồn dữ liệu

Dữ liệu được sử dụng trong đề tài được thu thập từ các kho dữ liệu trên google. Các dữ liệu thu thập được chia làm hai phần lưu vào file train và file val để thực hiện huấn luyện phát hiện phương tiện cho mô hình.

### Gán nhãn và khung bao

Sau khi dữ liệu được thu thập, các hình ảnh này sẽ được đưa lên công cụ LabelImg để gán nhãn và tạo khung bao phù hợp với từng loại phương tiện

A group of people on motorcycles

AI-generated content may be incorrect.

Hình 1: Tạo khung bao và gán nhãn

Các khung bao này sẽ được tạo dưới dạng file .txt và có cấu trúc như sau:

<class\_id> <x\_center> <y\_center> <width> <height>

### Huấn luyện mô hình YOLOv11

Sau khi dữ liệu được gán nhãn và tạo khung bao quá trình huấn luyện được thực hiện trên môi trường Python với framework Ultralytics YOLOv11, sử dụng GPU để tăng tốc độ xử lý.



Hình 2: Huấn luyện mô hình

Các tham quan trong như epoch, kích thước ảnh đầu vào, batch size và thuật toán được tối ưu lựa chọn và điều chỉnh để được được hiệu quả tốt nhất. Mô hình được huấn luyện qua nhiều vòng lặp, sau khi huấn luyện mô hình tốt nhất sẽ được lưu lại để sử dụng trong bài toán phát hiện và đếm số lượng phương tiện giao thông.

### Kiểm tra mô hình

Sau khi mô hình đã được huấn luyện, sử dụng mô hình để phát hiện phương tiện trong hình ảnh. Đối với ảnh số lượng phương tiện phát hiện được xác định dựa trên số lượng đối tượng mà mô hình phát hiện được. Kết quả đếm có thể được tổng hợp lại theo từng loại phương tiện xuất hiện trong ảnh.

A group of people riding scooters

AI-generated content may be incorrect.

Hình 3: Kiểm tra mô hình

# : KẾT QUẢ NGHIÊN CỨU

## Kết quả huấn luyện mô hình YOLOv11

Sau khi hoàn tất huấn luyện mô hình YOLOv11 sẽ phát hiện được các loại phương tiện khác nhau trong ảnh và đưa ra kết quả dự đoán với các độ chính xác khác nhau theo các góc quan sát, điều kiện ánh sáng và mật độ phương tiện.



Hình 4: Kết quả huấn luyện

Trong hình minh họa kết quả, hệ thống cho thấy:

- Phát hiện được phương tiện xe máy thông qua khung bao quanh đối tượng

- Gán nhãn tương ứng với đối tượng trong dữ liệu huấn luyện.

- Đưa ra độ tin cậy cho thấy mô hình có khả năng phát hiện đối tượng.

Điều này cho thấy YOLOv11 phù hợp để triển khai trong các hệ thống giám sát giao thông thực tế.

## Đánh giá độ chính xác

Độ chính xác của mô hình YOLOv11 được đánh giá thông qua các chỉ số phổ biến trong bài toán phát hiện đối tượng bao gồm Precision, Recall và mAP. Kết quả đánh giá cho thấy mô hình đạt độ chính xác cao trên hầu hết các lớp phương tiện.

Mô hình cho kết quả đúng hầu hết trong các trường hợp thử nghiệm với số lượng mẫu huấn luyện còn hạn chế. Tuy nhiên, độ tin cậy có thể giảm với các trường hợp:

- Phương tiện bị che khuất lẫn nhau.

- Phương tiện có kích thước nhỏ trong khung hình.

- Điều kiện ánh sáng không thuận lợi.

## Nhận xét chung về mô hình

Từ kết quả thử nghiệm, có thể rút ra các kết luận:

- YOLOv11 là mô hình phù hợp và cho hiệu quả cao cho bài toán phát hiện và đếm số lượng phương tiện giao thông.

- Mô hình có tiềm năng ứng dụng cao trông hệ thống giám sát giao thông thông minh và xử lý dữ liệu thời gian thực.

- Độ chính xác của mô hình phụ thuộc vào chất lượng dữ liệu huấn luyện và bước gán nhãn tạo khung bao cho đối tượng.

## Sơ đồ

# : KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

## Kết luận

Đề tài đã nghiên cứu và ứng dụng thành công mô hình YOLOv11 trong bài toán phát hiện và đếm số lượng phương tiện giao thông từ ảnh và video.

Qua quá trình tìm hiểu lý thuyết và thực nghiệm, mô hình YOLOv11 cho thầy khả năng nhận diện chính xác, tốc độ xử lý nhanh và phù hợp với các ứng dụng thời gian thực.

Hệ thống được xây dựng có thể phát hiện các loại phương tiện khác nhau như xe máy, ô tô, xe buýt và xe tải, đồng thời hiển thị khung bao, nhãn đối tượng và số lượng phương tiện một cách trực quan.

Kết quả thử nghiệm trên tập dữ liệu kiểm tra cho thấy mô hình hoạt động ổn định trong nhiều điều kiện khác nhau.

Thông qua đề tài, em đã nâng cao kiến thức về thị giác máy tính, học sâu và bài toán phát hiện đối tượng, đồng thời rèn luyện kỹ năng xây dựng và triển khai mô hình AI thực tế.

## Hướng phát triển

Mở rộng tập dữ liệu huấn luyện: Thu thập thêm nhiều dữ liệu giao thông với nhiều điều kiện khác nhau để nâng cao độ chính xác cho mô hình.

Kết hợp thuật toán theo dõi (Tracking): Áp dụng các phương pháp DeepSORT hoặc ByteTrack để theo dỗi đối tượng giúp đếm phương tiện chính xác hơn tránh bị trùng lặp.

Triển khai hệ thống trên nền tảng web hoặc IoT: Xây dựng giao diện web, tích hợp với camera giao thông hoặc các thiết bị IoT để dễ dàng giám sát và quản lý giao thông.

Tối ưu hiệu năng mô hình: Thực hiện tinh chỉnh các tham số và tối ưu mô hình để chạy trên các thiết bị có tài nguyên hạn chế.

Mở rộng bài toán ứng dụng: Phân tích mật độ giao thông, dự báo ùn tắc, ứng dụng quản lý bãi đỗ xe, giám sát an toàn giao thông và điều tiết giao thông một cách thông minh.

# DANH MỤC TÀI LIỆU THAM KHẢO

# PHỤ LỤC