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)*

**NHẬN XÉT CỦA THÀNH VIÊN HỘI ĐỒNG**

*Vĩnh Long, ngày ….. tháng …… năm ……*

**Thành viên hội đồng**

*(Ký tên và ghi rõ họ tên)*

**LỜI CẢM ƠN**

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](#_Toc217859578)

[1. Lý do chọn đề tài 9](#_Toc217859579)

[2. Mục tiêu nghiên cứu 9](#_Toc217859580)

[3. Đối tượng và pham vi nghiên cứu 9](#_Toc217859581)

[CHƯƠNG 1 : TỔNG QUAN 10](#_Toc217859582)

[1.1 Giới thiệu tổng quan về đề tài 10](#_Toc217859583)

[1.2 Mục tiêu của đề tài 10](#_Toc217859584)

[1.3 Bài toán cần giải quyết 10](#_Toc217859585)

[1.4 Giải quyết bài toán 11](#_Toc217859586)

[CHƯƠNG 2 : CƠ SỞ LÝ THUYẾT 12](#_Toc217859587)

[2.1 Tổng quan về thị giác máy tính 12](#_Toc217859588)

[2.1.1 Khái niệm về thị giác máy tính 12](#_Toc217859589)

[2.1.2 Nguyên lý hoạt động cơ bản của thị giác máy tính 12](#_Toc217859590)

[2.1.3 Ứng dụng của thị giác máy tính 12](#_Toc217859591)

[2.2 Tổng quan về học sâu (Deep Learning) 13](#_Toc217859592)

[2.2.1 Khái niệm học sâu 13](#_Toc217859593)

[2.2.2 Mạng nơ-ron tính chập (Convolutional Neural Network - CNN) 14](#_Toc217859594)

[2.3 Bài toán phát hiện đối tượng (Object Detection) 15](#_Toc217859595)

[2.3.1 Khái niệm 15](#_Toc217859596)

[2.3.2 Các hướng tiếp cận 15](#_Toc217859597)

[2.4 Giới thiệu mô hình YOLO (You Only Look Once) 15](#_Toc217859598)

[2.4.1 Tổng quan về mô hình YOLO 15](#_Toc217859599)

[2.4.2 Các phiên bản của YOLO 16](#_Toc217859600)

[2.5 Giới thiệu mô hình YOLOv11 17](#_Toc217859601)

[2.5.1 Tổng quan về mô hình 17](#_Toc217859602)

[2.5.2 Tổng quan về kiến trúc 18](#_Toc217859603)

[2.5.3 Những tiến bộ và tính năng của YOLOv11 22](#_Toc217859604)

[CHƯƠNG 3 : HIỆN THỰC HÓA NGHIÊN CỨU 24](#_Toc217859605)

[3.1 Mô tả bài toán 24](#_Toc217859606)

[3.2 Huấn Luyện mô hình phát hiện và đếm số lượng phương tiện 24](#_Toc217859607)

[3.2.1 Nguồn dữ liệu 24](#_Toc217859608)

[3.2.2 Gán nhãn và khung bao 25](#_Toc217859609)

[3.2.3 Huấn luyện mô hình YOLOv11 26](#_Toc217859610)

[3.2.4 Kiểm tra mô hình 26](#_Toc217859611)

[CHƯƠNG 4 : KẾT QUẢ NGHIÊN CỨU 28](#_Toc217859612)

[4.1 Kết quả huấn luyện mô hình YOLOv11 28](#_Toc217859613)

[4.2 Đánh giá độ chính xác 28](#_Toc217859614)

[4.3 Nhận xét chung về mô hình 29](#_Toc217859615)

[4.4 Sơ đồ mô hình YOLOv11 30](#_Toc217859616)

[CHƯƠNG 5 : KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN 30](#_Toc217859617)

[5.1 Kết luận 30](#_Toc217859618)

[5.2 Hướng phát triển 31](#_Toc217859619)

[DANH MỤC TÀI LIỆU THAM KHẢO 32](#_Toc217859620)

[PHỤ LỤC 33](#_Toc217859621)

**DANH MỤC HÌNH ẢNH**

[Hình 2.1 Sơ đồ tổng quát lịch sử YOLO 17](#_Toc217860018)

[Hình 2.2 Bảng tác vụ và chế độ hỗ trợ 18](#_Toc217860019)

[Hình 2.3 Kiến trúc của mô hình YOLOv11 19](#_Toc217860020)

[Hình 2.4 So sánh giữa khối C2f và C3K2 20](#_Toc217860021)

[Hình 2.5 Khối C2PSA 21](#_Toc217860022)

[Hình 2.6 So sánh YOLOv11 và các phiên bản trước 23](#_Toc217860023)

[Hình 3.1 Tập ảnh về phương tiện 24](#_Toc217860024)

[Hình 3.2 Tạo khung bao và gán nhãn cho đối tượng 25](#_Toc217860025)

[Hình 3.3 Label sau khi tạo khung bao và gán nhãn 25](#_Toc217860026)

[Hình 3.4 File data.yaml 26](#_Toc217860027)

[Hình 3.5 Huấn luyện mô hình 26](#_Toc217860028)

[Hình 3.6 Kết quả chạy thử nghiệm của mô hình sau huấn luyện 27](#_Toc217860029)

[Hình 4.1 Kết quả của mô hình 28](#_Toc217860030)

[Hình 4.2 Sơ đồ quy trình làm việc của YOLOv11 30](#_Toc217860031)

**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.[1]

### 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.[2]

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 phân 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)

### Tổng quan về mô hình YOLO

YOLO (You Only Look Once) là mô hình mạng CNN cho việc phát hiện, nhận dạng, phân loại đối tượng. YOLO được tạo ra từ việc kết hợp giữa convolutional layers và connected layers.

### Các phiên bản của YOLO

Nhờ vào tốc độ xử lý và độ chính xác cao YOLO đã dần trở nên phổ biến qua nhiều phiên bản như:

YOLOv1: Ra mắt năm 2015 đây là mô hình ban đầu của YOLO có tốc độ xử lý nhanh nhưng độ chính xác còn hạn đối với các đối tượng nhỏ và bị che khuất.

YOLOv2: Được phát năm 2016, cải thiện nhiều về tốc độ xử lý và độ chính xác hơn so với phiên bản YOLOv1, nó cung cấp khả năng phát hiện nhiều đối tượng hơn nhờ vào các kỹ thuật batch normalization, anchor boxes và huấn luyện trên tập dữ liệu lớn (khoảng 9000 lớp).

YOLOv3: Ra mắt năm 2018, tiếp tục nâng cao về mặt hiệu suất của mô hình nhờ sử dụng mạng xương sống (backbone), anchor boxes được điều chỉnh để phù hợp với kích thước và hình dạng của đối tượng để đạt hiệu quả cao. YOLOv3 cũng giới thiện Feature Pyramid Networks (FPN) để cải thiện phát hiện ở nhiều tỷ lệ khung hình. Cải thiện hiệu suất phát hiện đối tượng nhỏ.

YOLOv4: Phát hành vào năm 2020, giới thiệu nhiều cải tiến mới như tăng cường dữ liệu Mosaic, sử dụng kiến trúc CNN mới và sử dụng anchor boxes với k-means clustering để tập trung cải thiện tốc độ xử lý và độ chính xác. Nó giới thiệu một Loss Function mới và cải thiện lại cấu trúc FPN so với phiên bản trước.

YOLOv5: Sử dụng kiến trúc phức tạp hơn gọi là EfficientDet và đào tạp ttreen tập dữ liệu lớn bao gồm nhiều loại đối tượng. Nó sử dụng dynamic anchor boxes, spatial pyramid pooling (SPP) và CIoU loss để cải thiện hiệu suất đối tượng phát hiện trên nhiều loại đối tượng. Đây là phiên bản không được phát triển từ nhóm ban đầu của YOLO nhưng được phát triển rất phổ biến từ cộng nhờ vào tính dễ dàng triển khai và sử dụng.

YOLOv6: Phiên bản này mang lại nhiều cải tiến đáng kể trong việc phát hiện đối tượng với việc sử dụng các kiến thức mới, kiến trúc mạng nâng cấp và phương pháp huấn luyện tốt hơn.

YOLOv7: Được cải tiến rất nhiều so với các phiên bản trước, sử dụng anchor box cho phép phát hiện phạm vi hình dạng và kích thước đối tượng rộng hơn, sử dụng loss function mới gọi là “focal loss” giải quyết vấn đề kém hiệu quả với đối tượng nhỏ bằng cách giảm trọng số mất mát của các đối tượng dễ phát hiện và tập trung vào đối tượng khó phát hiện, cải thiện về độ phân giải từ 416 x 416 pixel lên 608 x 608 pixel cho ra độ chính xác cao hơn.

YOLOv8: Phiên bản này mang lại nhiều ưu điểm so với các phiên bản trước như phát hiện không dùng anchor, giới thiệu lớp tích chập C3 và tăng cường mosaic.

YOLOv9: Phiên bản này được cải tiến từ các phiên bản trước về tốc độ xử lý, độ chính xác và phát hiện tốt hơn các đối tượng nhỏ hoặc bị che khuất mà vẫn duy trì hiệu suất thời gian thực nhờ vào kiến trúc được tối ưu hóa và cơ chế lan truyền gradient hiệu quả hơn.

YOLOv10: Được phát triển bởi nhóm nghiên cứu của Đại học Thanh Hoa, phiên bản này tập trung vào tối ưu kiến trúc để đạt được khả năng phát hiện đối tượng thời gian thực hiệu quả hơn mà không cần các bước hậu xử lý phức tạp.

YOLOv11: Do Ultralytics phát triển, có độ chính xác và hiệu quả tiên tiến. Dựa trên những tiến bộ ấn tượng của các phiên bản YOLO trước đó, nó cải thiện đáng kể về kiến trúc và phương pháp huấn luyện biến nó trở thành lựa chọn phổ biến và ứng dụng rộng rãi trong lĩnh vực thị giác máy tính.

A timeline with blue circles and white text

AI-generated content may be incorrect.

Hình 2.1 Sơ đồ tổng quát lịch sử YOLO

## Giới thiệu mô hình YOLOv11

### Tổng quan về mô hình

YOLOv11 là phiên bản mới nhất của Ultralytics, cải thiện đáng kể về độ chính xác, tốc độ xử lý và hiệu quả cao. Nhờ vào các yếu tố đó mà nó trở nên là lựa chọn phổ biến trong các tác vụ cho nhiều lĩnh vực của thị giác máy tính.[3]

YOLOv11 được xây dựng dựa trên phạm vi mô hình đa dạng được Ultralytics phát hành, cung cấp và hỗ trợ nhiều tác vụ thị giác máy tính. A screenshot of a computer

AI-generated content may be incorrect.

Hình 2.2 Bảng tác vụ và chế độ hỗ trợ

### Tổng quan về kiến trúc

Kiến trúc YOLOv11 bao gồm ba phần chính. Thứ nhất, Backbone (Xương sống) đóng vai trò là bộ trích xuất đặc trưng chính, sử dụng mạng nơ-ron tích chập (CNN) để chuyển đổi dữ liệu hình ảnh thô thành các bản đồ đặc trưng đa tỷ lệ. Thứ hai, Neck (Cổ mạng) hoạt động như một giai đoạn xử lý trung gian, sử dụng các lớp chuyên biệt để tổng hợp và tăng cường các biểu diễn đặc trưng qua nhiều tỷ lệ khác nhau. Thứ ba, Head (Đầu mạng) là cơ ché dự đoán, tạo ra các đầu ra cuối cùng cho việc định vị và phân loại đối tượng dựa trên các bản đồ đặc trưng được tinh chỉnh.

Dựa trên kiến trúc đã được thiết lập này, YOLOv11 mở rộng và cải tiến nền tảng được tạo ra bởi YOLOv8, giới thiệu các đổi mới về kiến trúc và tối ưu hóa tham số để đạt được hiệu suất phát hiện vượt trội.

A diagram of a computer program

AI-generated content may be incorrect.

Hình 2.3 Kiến trúc của mô hình YOLOv11

Dưới đây là chi tiết từng phần của kiến trúc mạng YOLOv11:

Backbone (Xương sống):

Đây là một thành phần quan trọng của kiến trúc YOLOv11, chịu trách nhiệm trích xuất các đặc điểm từ hình ảnh đầu vào ở nhiều tỷ lệ. Quá trình này bao gồm việc xếp chồng các lớp tích chập và khối chuyên biệt để tạo ra các bản đồ đặc trưng ở nhiều độ phân giải khác nhau.

- Khối C3K2: Một cải tiến trong YOLOv11 là sự ra đời của khối C3K2, thay thế khối C2f được sử dụng ở các phiên bản trước. Khối C3K2 là một triển khai hiệu quả hơn về mặt tính toán của CSP (Cross Stage Partial). Sử dụng các phép tnhs kernel 3x3 nhỏ hơn để tối ưu hóa luồng thông tin và giảm chi phí tính toán, trong khi vẫn giữ được khả năng nắm bắt các đặc trưng cần thiết.

A diagram of a diagram

AI-generated content may be incorrect.

Hình 2.4 So sánh giữa khối C2f và C3K2

Neck (Cổ mạng):

Đây là phần trung gian giữa Backbone và Head, có nhiệm vụ kết hợp các đặc trưng từ nhiều tầng khác nhau để giúp mô hình phát hiện được các đối tượng ở kích thước khác nhau.

Kiến trúc Neck được cải tiến bao gồm:

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

- Khối C2PSA Blocks (Cross Stage Partial with Spatial Attention): Đây là 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 vào các vùng quan trong trong hình ảnh, nâng cao độ chính xác, đặc iệt là đối với các đối tượng nhỏ hoặc bị che khuất.

A diagram of a system

AI-generated content may be incorrect.

Hình 2.5 Khối C2PSA

Head (Đầu mạng):

Đây là phần chịu trách nhiệm tạo ra các dự đoán cuối cùng về phát hiện và phân loại đối tượng. Nó xử lý các bản đồ đặc trưng được truyền từ phần neck, cuối cùng xuất ra các hộp chứa đối tượng (Bounding boxes) và nhãn lớp (Class labels) cho các đối tượng trong ảnh.

Trong phần head, khối C3K2 được sử dụng nhiều để xử lý và tinh chỉnh hiệu quả các bản đồ đặc trưng. Các khối này được đặt trong nhiều đường dẫn khác nhau trong phần head, giúp xử lý các đặc trưng đa tỷ lệ ở các độ sâu khác nhau.

Đặc điểm chính của khối C3K2 trong phần này:

- Xử lý nhanh hơn: Việc sử dụng hai tích chập nhỏ thay cho một tích chập lớn làm giảm chi phí tính toán, giúp quá trình trích xuất đặc trưng nhanh hơn.

- Hiệu quả về tham số: C3K2 là phiên bản gọn nhẹ hơn của bottleneck CSP, là giảm số lượng tham số cần huấn luyện.

### Những tiến bộ và tính năng của YOLOv11

YOLOv11 là một bước tiến quan trọng trong công nghệ phát hiện đối tượng, được xây dựng trên nền tảng của YOLOv9 và YOLOv10. Phiên bản này được Ultralytics phát triển mang đến thiết kế kiến trúc nâng cao, kỹ thuật trích xuất đặc trưng tinh vi và phương pháp huấn luyện cải tiến, kết hợp giữa tốc độ xử lý nhanh, độ chính xác cao và hiệu quả tính toán. YOLOv11 có khả năng phát hiện chi tiết trong các điều kiện không thuận lợi và xử lý dữ liệu phức tạp. Nó cũng đạt được nhiều cải tiến so với các phiên bản trước, giúp nâng cao hiệu quả và khả năng phát hiện.

- Độ chính xác được cải thiện với độ phức tạp được giảm bớt: Biến thể YOLOv11m đạt điểm Độ chính xác trung bình (mAP) vượt trội trong khi sử dụng ít hơn 22% tham số so với biến thể YOLOv8m, chứng minh hiệu quả tính toán được cải thiện mà không ảnh hưởng đến độ chính xác.

- Tính linh hoạt trong các tác vụ: YOLOv11 thể hiện sự thành thạo trong nhiều ứng dụng CV khác nhau, bao gồm ước tính tư thế, nhận diện đối tượng, phân loại hình ảnh, phân đoạn thể hiện đối tượng và phát hiện giới hạn đối tượng.

- Tốc độ và hiệu suất được tối ưu: thông qua các thiết kế kiến trúc tinh chỉnh và quy trình huấn luyện hợp lý, YOLOv11 đạt tốc độ xử lý nhanh hơn trong khi vẫn duy trì sự cân bằng giữa độ chính xác và hiệu quả tính toán.

- Số lượng tham số được sắp xếp hợp lý: Việc giảm các tham số góp phần làm cho hiệu suất mô hình nhanh hơn mà không ảnh hưởng đáng kể đến độ chính xác tổng thể của YOLOv11.

- Trích xuất tính năng nâng cao: YOLOv11 kết hợp các cải tiến trong kiến trúc backbone và neck, mang lại khả năng trích xuất tính năng nâng cao và do đó phát hiện đối tượng chính xác hơn.

- Khả năng thích ứng theo ngữ cảnh: YOLOv11 chứng minh tính linh hoạt trong nhiều tình huống triển khai khác nhau, bao gồm nền tảng đám mây, thiết bị biên và các hệ thống tối ưu cho GPU NVIDIA.[4]

A graph of different colored lines

AI-generated content may be incorrect.

Hình 2.6 So sánh YOLOv11 và các phiên bản trước

YOLOv11 mang lại nhiều cải tiến về độ chính xác, tốc độ xử lý dữ liệu và tính linh hoạt trong lĩnh vực thị giác máy tính. Với kiến trúc tối ưu và khả năng thích trích xuất đặc trưng tiên tiến, YOLOv11 trở thành một công cụ mạnh mẽ cho các nhiệm vụ từ phát hiện đối tượng cho đến phân đoạn và dự đoán tư thế, đồng thời có thể triển khai trên nhiều nền tảng.

YOLO11 Nano là một biến thể siêu nhỏ gọn của mô hình YOLOv11, được thiết kế đặc biệt để tối ưu hóa tốc độ và hiệu quả tính toán mà vẫn duy trì hiệu suất tốt trong các nhiệm vụ phát hiện đối tượng. Nhờ vào ưu điểm trên em sẽ sử dụng mô hình YOLO11 Nano cho bài toán phát hiện và đếm số lượng phương tiện giao thông.

# : 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 sát cho mô hình.

## Huấn Luyện mô hình phát hiện và đếm số lượng phương tiện

### Nguồn dữ liệu

Dữ liệu sử dụng trong bài được thu thập từ google

A collage of images of cars and motorcycles

AI-generated content may be incorrect.

Hình 3.1 Tập ảnh về phương tiện

### 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 3.2 Tạo khung bao và gán nhãn cho đối tượng

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>

Sau khi gán nhãn và tạo khung bao các file .txt sẽ được lưu vào thư mục label để thực hiện huấn luyện.

A screenshot of a computer

AI-generated content may be incorrect.

Hình 3.3 Label sau khi tạo khung bao và gán nhãn

File cấu hình data.yaml chứa đường dẫn đến thư mục hình ảnh và thông tin đối tượng. Trong đó nc là số class và names là tên của các class.

A black screen with orange text

AI-generated content may be incorrect.

Hình 3.4 File data.yaml

### 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 3.5 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 3.6 Kết quả chạy thử nghiệm của mô hình sau huấn luyện

# : 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.

A screenshot of a computer

AI-generated content may be incorrect.

Hình 4.1 Kết quả của mô hình

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 mô hình YOLOv11

A diagram of a program

AI-generated content may be incorrect.

Hình 4.2 Sơ đồ quy trình làm việc của YOLOv11

# : 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

[1] AI V., “Computer Vision là gì? Lợi ích và ứng dụng của thị giác máy tính.” Accessed: Dec. 30, 2025. [Online]. Available: https://vnptai.io/vi/blog/detail/computer-vision-la-gi

[2] AI V., “Deep Learning là gì? Ứng dụng và xu hướng tương lai của học sâu.” Accessed: Dec. 30, 2025. [Online]. Available: https://vnptai.io/vi/blog/detail/deep-learning-la-gi

[3] Ultralytics, “Ultralytics YOLO11.” Accessed: Dec. 30, 2025. [Online]. Available: https://docs.ultralytics.com/vi/models/yolo11/

[4] “Báo cáo BTL 17226: Xây dựng Hệ Thống Đếm Người Trên Camera,” Studocu. Accessed: Dec. 30, 2025. [Online]. Available: https://www.studocu.vn/vn/document/dai-hoc-hang-hai-viet-nam/sql-server/bao-cao-copy-adasaaaaaaaaaaaaa/118170069

# PHỤ LỤC