**TRƯỜNG ĐẠI HỌC NGUYỄN TẤT THÀNH**

**KHOA CÔNG NGHỆ THÔNG TIN**



**TIỂU LUẬN NHẬN DẠNG ĐỐI TƯỢNG**

Ứng Dụng YOLOv8 Phân Loại Xe

Để Giám Sát Trong Trạm Thu Phí

**Giảng viên hướng dẫn: ThS.Trần Châu Thanh Thiện**

**Sinh viên thực hiện: Võ Nguyễn Bảo Phước**

**MSSV: 2200008333**

**Khoá: 2022 - 2026**

**Ngành/ chuyên ngành: Trí Tuệ Nhân Tạo**

Tp HCM, 14 tháng 5 năm 2025

**TRƯỜNG ĐẠI HỌC NGUYỄN TẤT THÀNH**

**KHOA CÔNG NGHỆ THÔNG TIN**



**TIỂU LUẬN NHẬN DẠNG ĐỐI TƯỢNG**

Ứng Dụng YOLOv8 Phân Loại Xe

Để Giám Sát Trong Trạm Thu Phí

**Giảng viên hướng dẫn: ThS.Trần Châu Thanh Thiện**

**Sinh viên thực hiện: Võ Nguyễn Bảo Phước**

**MSSV: 2200008333**

**Khoá: 2022 - 2026**

**Ngành/ chuyên ngành: Trí Tuệ Nhân Tạo**

Tp HCM, 14 tháng 5 năm 2025

LỜI MỞ ĐẦU

Trong bối cảnh giao thông ngày càng phát triển và nhu cầu tự động hóa các hệ thống quản lý phương tiện ngày càng cao, việc ứng dụng trí tuệ nhân tạo vào giám sát và phân loại xe tại các trạm thu phí trở thành một hướng đi thiết thực và đầy tiềm năng. Các phương pháp truyền thống như nhận dạng thủ công hoặc dựa vào cảm biến vật lý thường tồn tại nhiều hạn chế về chi phí, độ chính xác và khả năng mở rộng.

Với sự tiến bộ mạnh mẽ của thị giác máy tính (Computer Vision), các mô hình học sâu như YOLO (You Only Look Once) đã chứng minh được hiệu quả vượt trội trong việc phát hiện và phân loại đối tượng theo thời gian thực. Trong đó, YOLOv8, phiên bản mới nhất của dòng mô hình YOLO, mang đến nhiều cải tiến về tốc độ và độ chính xác, đặc biệt phù hợp với các ứng dụng yêu cầu xử lý nhanh như giám sát giao thông.

Đề tài này tập trung nghiên cứu và triển khai ứng dụng YOLOv8 vào việc phát hiện và phân loại các loại xe (xe máy, ô tô con, xe tải, xe buýt, v.v...) tại trạm thu phí, nhằm phục vụ cho mục đích giám sát, thống kê và hỗ trợ thu phí tự động. Việc áp dụng mô hình này không chỉ nâng cao hiệu quả quản lý mà còn góp phần thúc đẩy quá trình chuyển đổi số trong lĩnh vực giao thông vận tải.

LỜI CẢM ƠN

Trước tiên, em xin gửi lời cảm ơn chân thành đến Ban Giám Hiệu và các thầy cô trường Đại học Nguyễn Tất Thành, đặc biệt là Khoa Công nghệ thông tin, đã tạo điều kiện và cung cấp kiến thức trong suốt quá trình học tập và thực hiện đề tài.

Em xin gửi lời biết ơn sâu sắc đến thầy ThS.Trần Châu Thanh Thiện đã hết lòng hướng dẫn, góp ý và định hướng để em có thể hoàn thành tốt nhất đề tài này.

Em xin chân thành cảm ơn gia đình và bạn bè đã động viên, hỗ trợ em trong quá trình thực hiện đề tài. Đặc biệt, em xin cảm ơn những ý kiến đóng góp quý báu từ các thầy cô và đồng nghiệp trong suốt quá trình nghiên cứu.

Cuối cùng, em xin gửi lời cảm ơn đến tất cả những ai đã gián tiếp hoặc trực tiếp đồng hành cùng em trong suốt hành trình thực hiện đề tài này.

TPHCM, 14 tháng 5 năm 2025

Họ tên sinh viên thực hiện

(ký, ghi rõ họ tên)

|  |  |
| --- | --- |
| TRƯỜNG ĐẠI HỌC NGUYỄN TẤT THÀNH  **TRUNG TÂM KHẢO THÍ** | **KỲ THI KẾT THÚC HỌC PHẦN**  **HỌC KỲ 2 NĂM HỌC 2024 – 2025** |

Text Box 8, Textbox**PHIẾU CHẤM THI TIỂU LUẬN**

Môn thi: **Nhận dạng đối tượng** Lớp học phần:**22DTH3C**

Nhóm sinh viên thực hiện: Nhóm 1

1. **Nguyễn Đoàn Thanh Thảo** Tham gia đóng góp: ALL

2. **Võ Nguyễn Bảo Phước** Tham gia đóng góp: ALL

3. **Võ Trần Duy Tân** Tham gia đóng góp: ALL

Ngày thi:**14/5/2025** Phòng thi: L.610

Đề tài tiểu luận: **Ứng Dụng YOLOv8 Phân Loại Xe Để Giám Sát Trong Trạm Thu Phí**

Phần đánh giá của giảng viên (căn cứ trên thang rubrics của môn học):

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Tiêu chí (theo CĐR HP)** | **Đánh giá của GV** | | **Điểm tối đa** | **Điểm đạt được** |  |
| Cấu trúc của báo cáo |  | |  |  |  |
| Nội dung |  | |  |  |  |
| Các nội dung thành phần |  | |  |  |  |
| Lập luận |  | |  |  |  |
| Kết luận |  | |  |  |  |
| Trình bày |  | |  |  |  |
| **TỔNG ĐIỂM** |  | |  |  |  |
|  | | **Giảng viên chấm thi**  *(ký, ghi rõ họ tên)* | | | |

MỤC LỤC

[LỜI MỞ ĐẦU 1](#_Toc196323557)

[LỜI CẢM ƠN 2](#_Toc196323558)

[Chương 1: Giới Thiệu Đề Tài 1](#_Toc196323560)

[1.1 Đặt vấn đề 1](#_Toc196323561)

[1.2 Vấn đề nghiên cứu 2](#_Toc196323563)

[1.3 Đối tượng nghiên cứu 2](#_Toc196323564)

[1.4 Phạm vi nghiên cứu 2](#_Toc196323565)

[1.5 Lý do chọn đề tài 2](#_Toc196323566)

[Chương 2: Cơ Sở Lý Thuyết 4](#_Toc196323567)

[2.1. Phát hiện đối tượng là gì? 4](#_Toc196323568)

[2.2 Các loại mô hình phát hiện đối tượng 4](#_Toc196323570)

[2.2.1 Các mô hình sử dụng một giai đoạn 5](#_Toc196323572)

[2.2.2 Các mô hình sử dụng hai giai đoạn 5](#_Toc196323573)

[2.3 Mạng YOLO (You Only Look Once): 6](#_Toc196323576)

[2.3.1 Giới thiệu về YOLO 6](#_Toc196323577)

[2.3.2 Các phiên bản YOLO 7](#_Toc196323579)

[2.3.3 Ứng dụng YOLO 8](#_Toc196323580)

[2.3.4 YOLOv1 9](#_Toc196323581)

[2.3.5 YOLOv8 16](#_Toc196323597)

[2.4 Tổng quan về bài toán phân loại xe trong giám sát trạm thu phí 19](#_Toc196323600)

[2.4.1 Ứng dụng phát hiện đối tượng trong giám sát và phân loại xe 19](#_Toc196323601)

[2.4.2 Thách thức và định hướng nghiên cứu 19](#_Toc196323602)

[Chương 3: Cơ Sở Thực Tiễn 21](#_Toc196323603)

[3.1. Thực trạng và nhu cầu giám sát tại trạm thu phí 21](#_Toc196323604)

[3.2. Xu hướng ứng dụng trí tuệ nhân tạo trong giám sát giao thông 21](#_Toc196323605)

[3.3. Tính khả thi và lợi ích của đề tài 22](#_Toc196323606)

[3.4. Giải pháp triển khai đề tài 22](#_Toc196323607)

[Chương 4: Thực Nghiệm 25](#_Toc196323611)

[4.1 Mô hình hệ thống đề xuất 25](#_Toc196323612)

[4.2 Môi trường và công cụ thực hiện 25](#_Toc196323614)

[4.3 Quy trình xử lý và thuật toán chính: 26](#_Toc196323615)

[4.4 Kết quả và đánh giá: 30](#_Toc196323625)

[Chương 5: Kết Luận 33](#_Toc196323628)

[5.1 Nội dung đã thực hiện 33](#_Toc196323629)

[5.2 Hạn chế 33](#_Toc196323630)

[5.3 Hướng phát triển 33](#_Toc196323631)

[TÀI LIỆU THAM KHẢO 36](#_Toc196323632)

DANH MỤC HÌNH

[Hình 1.1 1](#_Toc196324089)

[Hình 2.1 4](#_Toc196324090)

[Hình 2.2 5](#_Toc196324091)

[Hinh 2.3 6](#_Toc196324092)

[Hình 2.4 6](#_Toc196324093)

[Hình 2.5 7](#_Toc196324094)

[Hình 2.6 10](#_Toc196324095)

[Hình 2.7 10](#_Toc196324096)

[Hình 2.8 11](#_Toc196324097)

[Hình 2.9 12](#_Toc196324098)

[Hình 2.10 12](#_Toc196324099)

[Hình 2.11 13](#_Toc196324100)

[Hình 2.12 13](#_Toc196324101)

[Hình 2.13 14](#_Toc196324102)

[Hình 2.14 14](#_Toc196324103)

[Hình 2.15 15](#_Toc196324104)

[Hình 2.16 15](#_Toc196324105)

[Hình 2.17 15](#_Toc196324106)

[Hình 2.18 16](#_Toc196324107)

[Hình 2.19 17](#_Toc196324108)

[Hình 2.20 17](#_Toc196324109)

[Hình 2.21 18](#_Toc196324110)

[Hình 2.22 20](#_Toc196324111)

[Hình 3.1 24](#_Toc196324112)

[Hình 3.2 24](#_Toc196324113)

[Hình 3.3 25](#_Toc196324114)

[Hình 4.1 26](#_Toc196324115)

[Hình 3.2 28](#_Toc196324116)

[Hình 4.3 28](#_Toc196324117)

[Hình 4.4 28](#_Toc196324118)

[Hình 4.5 29](#_Toc196324119)

[Hình 4.6 29](#_Toc196324120)

[Hình 4.7 29](#_Toc196324121)

[Hình 4.8 30](#_Toc196324122)

[Hình 4.9 31](#_Toc196324123)

[Hình 4.10 31](#_Toc196324124)

[Hình 4.11 32](#_Toc196324125)

[Hình 4.12 32](#_Toc196324126)

Chương 1: Giới Thiệu Đề Tài

1.1 Đặt vấn đề

**Làm thế nào để giám sát và phân loại phương tiện tại trạm thu phí một cách chính xác, hiệu quả và tiết kiệm?**

Đây là một trong những thách thức lớn đặt ra trong bối cảnh ngành giao thông vận tải Việt Nam đang chuyển mình mạnh mẽ theo hướng hiện đại hóa và số hóa.

Việt Nam hiện là một trong những quốc gia có tốc độ phát triển kinh tế nhanh ở khu vực Đông Nam Á. Tính đến năm 2024, tăng trưởng GDP đạt 7,09% – mức cao so với khu vực. Trong tiến trình phát triển đó, hạ tầng giao thông đóng vai trò thiết yếu, đặc biệt là hệ thống đường cao tốc và các trạm thu phí, vốn đang ngày càng mở rộng để đáp ứng nhu cầu lưu thông hàng hóa và hành khách.

Tính đến năm 2024, Việt Nam có hơn 2.000 km đường cao tốc và hàng trăm trạm thu phí BOT hoạt động. Dự kiến đến năm 2030, con số này sẽ còn tăng mạnh nhằm đáp ứng chiến lược phát triển giao thông quốc gia. Tuy nhiên, bên cạnh việc mở rộng, công tác giám sát, phân loại và kiểm soát phương tiện tại trạm thu phí cũng đặt ra những thách thức mới về chi phí vận hành, độ chính xác và khả năng xử lý thời gian thực.

1.1 Hình đường cao tốc


Hình 1.1

Hiện nay, một số trạm thu phí vẫn còn sử dụng các phương pháp kiểm soát truyền thống như kiểm tra thủ công, quét RFID đơn lẻ hoặc camera thông thường chưa tích hợp AI, dẫn đến các vấn đề như:

Xếp hàng kéo dài, gây ùn tắc cục bộ.

Sai sót trong phân loại phương tiện dẫn đến thu phí không chính xác.

Khó giám sát hành vi vi phạm hoặc xe cố tình tránh phí.

Do đó, nhu cầu áp dụng công nghệ thông minh, đặc biệt là thị giác máy tính và trí tuệ nhân tạo, vào hệ thống giám sát tại trạm thu phí đang trở nên cấp thiết.

1.2 Vấn đề nghiên cứu

Hệ thống giám sát tại trạm thu phí có thể trở nên hiệu quả hơn khi được tích hợp với camera thông minh và mô hình AI, đặc biệt là các mô hình học sâu (Deep Learning) có khả năng phân tích hình ảnh theo thời gian thực. Trong đó, mô hình YOLOv8 – phiên bản mới nhất trong chuỗi mô hình phát hiện đối tượng "You Only Look Once" – nổi bật với khả năng:

Nhận diện phương tiện (xe máy, ô tô con, xe tải, xe buýt...) với độ chính xác cao.

Xử lý thời gian thực, phù hợp cho các trạm thu phí nơi xe liên tục lưu thông.

Khả năng tích hợp vào các hệ thống camera giám sát sẵn có.

Hệ thống có thể phục vụ cho các mục tiêu sau:

Phân loại xe chính xác để tính phí đúng loại phương tiện.

Giám sát hành vi vi phạm như vượt trạm, xe không đủ điều kiện, hoặc chạy sai làn.

Lưu trữ dữ liệu cho việc phân tích lưu lượng, lên kế hoạch bảo trì, và hỗ trợ quản lý giao thông.

1.3 Đối tượng nghiên cứu

Mô hình YOLOv8 trong việc phát hiện và phân loại phương tiện giao thông.

Ứng dụng công nghệ thị giác máy tính (Computer Vision) và mạng nơ-ron tích chập (CNN). Phân tích độ chính xác của hệ thống trong môi trường trạm thu phí thực tế.

1.4 Phạm vi nghiên cứu

Về mặt kỹ thuật: Tìm hiểu, đánh giá và thử nghiệm mô hình YOLOv8 trong phân loại các phương tiện qua hình ảnh từ camera trạm thu phí.

Về thực tiễn ứng dụng: Mô phỏng hoạt động của hệ thống tại trạm thu phí tại Việt Nam, so sánh với hệ thống các nước tiên tiến như Hàn Quốc, Nhật Bản.

Về hiệu suất hệ thống: Đo lường tốc độ xử lý, độ chính xác nhận diện trong điều kiện ánh sáng, thời tiết và mật độ xe khác nhau.

Về thách thức triển khai: Phân tích khó khăn về chi phí, tích hợp hạ tầng, bảo mật dữ liệu và xử lý thời gian thực.

1.5 Lý do chọn đề tài

Trong bối cảnh chuyển đổi số toàn diện trong ngành giao thông vận tải, các trạm thu phí truyền thống đang dần được thay thế bằng hệ thống thu phí không dừng (ETC). Tuy nhiên, để đảm bảo hiệu quả vận hành, các hệ thống giám sát cần được nâng cấp với khả năng tự động phát hiện và phân loại xe chính xác.

Phương pháp truyền thống như kiểm tra bằng mắt người, quét thẻ thủ công hay cảm biến đơn lẻ không còn đáp ứng được yêu cầu về tốc độ và độ tin cậy trong thời đại mới. Bên cạnh đó, các hành vi gian lận như tráo đổi biển số, xe đi sai làn, hoặc tránh trạm vẫn là vấn đề nhức nhối.

Việc ứng dụng mô hình YOLOv8 vào phân loại xe tại trạm thu phí là một giải pháp hiện đại, có khả năng nhận diện chính xác trong thời gian thực, phù hợp với điều kiện hạ tầng hiện nay và dễ dàng mở rộng quy mô.

**Vì vậy, em lựa chọn đề tài:**

“Ứng Dụng YOLOv8 Trong Phân Loại Xe Để Giám Sát Trong Trạm Thu Phí”

với các mục tiêu:

Tìm hiểu chuyên sâu về ứng dụng AI và thị giác máy tính trong giao thông.

Triển khai thử nghiệm mô hình trong tình huống mô phỏng thực tế tại trạm thu phí.

Đề xuất hướng cải tiến, phát triển hệ thống giám sát thông minh có tính ứng dụng cao tại Việt Nam.

Đây không chỉ là một đề tài mang tính học thuật mà còn có tiềm năng áp dụng thực tiễn cao, góp phần vào mục tiêu phát triển giao thông thông minh, an toàn và hiệu quả trong thời đại chuyển đổi số.

Chương 2: Cơ Sở Lý Thuyết

2.1. Phát hiện đối tượng là gì?

Phát hiện đối tượng là một nhiệm vụ quan trọng trong thị giác máy tính, cho phép máy móc xác định và định vị các đối tượng cụ thể trong hình ảnh hoặc video.

Không giống như phân loại hình ảnh, chỉ xác định sự hiện diện của một đối tượng trong hình ảnh, phát hiện đối tượng vẽ các hộp giới hạn xung quanh mỗi đối tượng được phát hiện, chỉ định vị trí của nó.

Công nghệ này thu hẹp khoảng cách giữa cách máy móc nhận thức dữ liệu hình ảnh và cách con người hiểu môi trường xung quanh.

Về bản chất, phát hiện đối tượng kết hợp hai quy trình chính: phân loại và định vị.

Phân loại xác định những đối tượng nào có mặt (ví dụ: ô tô, người, cây), trong khi định vị xác định vị trí của những đối tượng này trong hình ảnh, thường bằng cách vẽ một hộp giới hạn xung quanh chúng.

Điều này thường đạt được bằng cách sử dụng các thuật toán phức tạp, thường dựa trên Mạng nơ-ron tích chập (CNN) , học cách nhận dạng các mẫu và đặc điểm đặc trưng cho các đối tượng khác nhau.

Độ chính xác của các mô hình phát hiện đối tượng thường được đánh giá bằng các số liệu như Giao điểm trên hợp (IoU) và Độ chính xác trung bình trung bình (mAP).

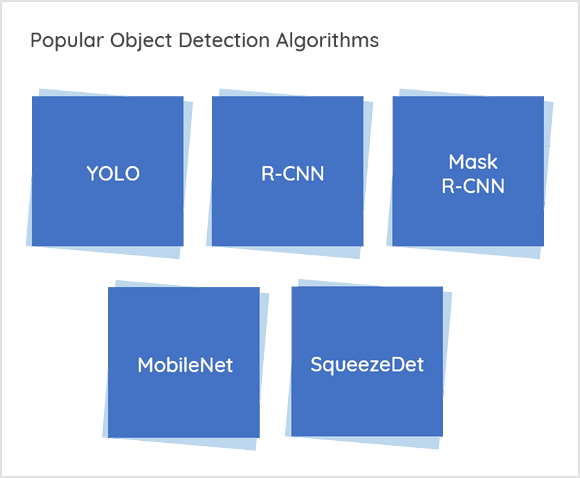
Hình 2.1

2.2 Các loại mô hình phát hiện đối tượng

Các mô hình phát hiện đối tượng có thể được phân loại thành hai loại chính:

Các mô hình sử dụng một giai đoạn.

Các mô hình sử dụng hai giai đoạn.



Hình 2.2

2.2.1 Các mô hình sử dụng một giai đoạn

**R-CNN (Region-based ConvNet)**

Đề xuất một số vùng tiềm năng bằng thuật toán thô khác, chẳng hạn selective search.

Dùng mạng CNN trích xuất đặc trưng từng vùng rồi phân loại bằng SVM.

**Fast-RCNN**

Đẩy tất cả các vùng (khoảng 2000) qua mạng trích xuất CNN cùng một lúc.

Crop thông tin ở lớp đầu ra của CNN thay vì crop vùng trên ảnh gốc như R-CNN.

Đẩy qua nhánh phân loại và nhánh hiệu chỉnh tọa độ box.

2.2.2 Các mô hình sử dụng hai giai đoạn

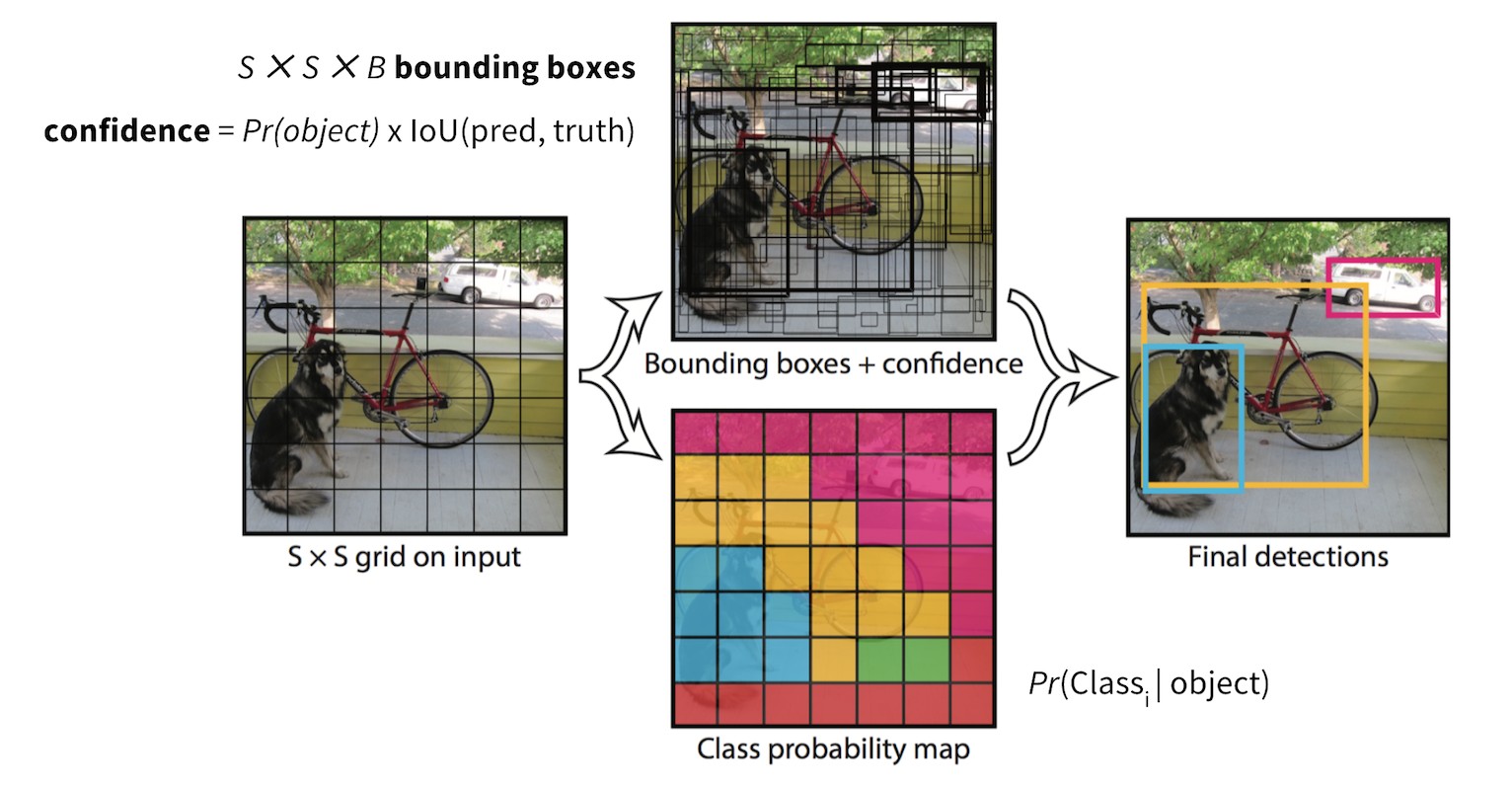
Ngược lại, mô hình một giai đoạn, như Ultralytics YOLO, SSD, cung cấp hiệu suất nhanh hơn bằng cách dự đoán trực tiếp các hộp giới hạn và xác suất lớp trong một lần chạy.

Các mô hình này thường đề xuất một lưới box dày đặc trên ảnh ban đầu, thường có bước nhảy đều (stride).

Từng box này sẽ được phân loại và hiệu chỉnh tọa độ (nếu box chứa đối tượng) bằng mạng CNN.

Các mô hình một giai đoạn thường nhanh hơn và đơn giản hơn các mạng hai giai đoạn.

**YOLO- You Only Look Once**



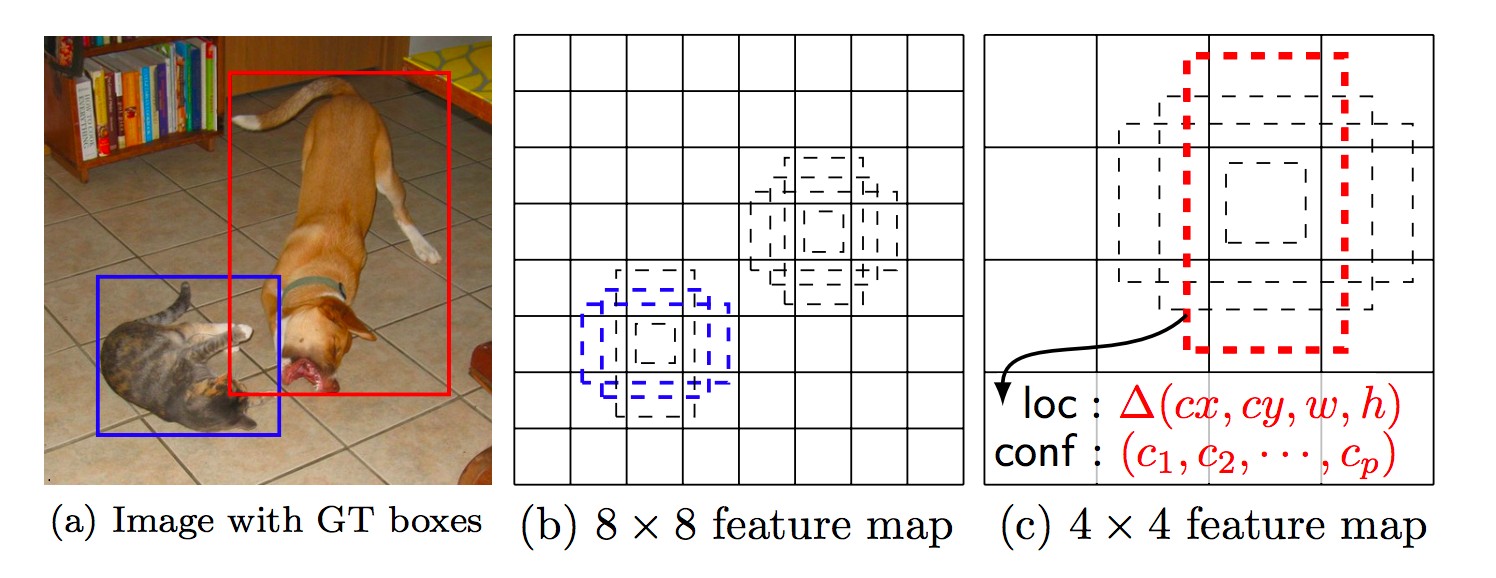
Hinh 2.3

**SSD: Single Shot Detector**

Tương tự YOLO nhưng lưới box dày đặc hơn, có nhiều lưới với các kích thước box khác nhau.

Kiến trúc mạng backbone khác với YOLO.

Data augmentation + Hard negative mining.



Hình 2.4

2.3 Mạng YOLO (You Only Look Once):

2.3.1 Giới thiệu về YOLO

YOLO là một mô hình mạng CNN cho việc phát hiện, nhận dạng, phân loại đối tượng. YOLO được tạo ra từ việc kết hợp giữa các convolutional layers (Conv) và connected layers. Trong đó các convolutional layers sẽ trích xuất ra các feature của ảnh, còn full-connected layers (FC) sẽ dự đoán ra xác suất đó và tọa độ của đối tượng.

Đầu vào của mô hình là một ảnh, mô hình sẽ nhận dạng ảnh đó có đối tượng nào hay không, sau đó sẽ xác định tọa độ của đối tượng trong bức ảnh.

Mô hình YOLO được giới thiệu lần đầu tiên vào năm 2016 bởi Joseph Redmon và các cộng sự trong bài báo "You Only Look Once: Unified, Real-Time Object Detection".



Hình 2.5

2.3.2 Các phiên bản YOLO

**Bảng 2.1**

|  |  |  |
| --- | --- | --- |
| **Phiên bản** | **Năm phát hành** | **Điểm khác biệt** |
| YOLOv1 | 2016 | Phiên bản đầu tiên, tốc độ nhanh nhưng độ chính xác thấp |
| YOLOv2 | 2017 | Cải thiện tốc độ và độ chính xác so với YOLOv1, sử dụng Darknet-19 làm backbone |
| YOLOv3 | 2018 | Cải thiện hiệu suất và khả năng khái quát hóa so với YOLOv2, sử dụng Darknet-53 làm backbone |
| YOLOv4 | 2020 | Cải thiện độ chính xác và tốc độ so với YOLOv3, hỗ trợ nhiều mô hình backbone hơn |
| YOLOv5 | 2021 | Cải thiện tốc độ và hiệu quả so với YOLOv4, đơn giản hóa mã nguồn và dễ sử dụng hơn, sử dụng PyTorch |
| YOLOv6 | 2022 | Cải thiện hiệu suất và khả năng khái quát hóa so với YOLOv5, sử dụng mô hình Transformer và Swin Transformer |
| YOLOv7 | 2022 | Một trong những ưu điểm chính của YOLO v7 là tốc độ. Nó có thể xử lý hình ảnh với tốc độ 155 khung hình mỗi giây, nhanh hơn nhiều so với các thuật toán phát hiện đối tượng hiện đại khác. |
| YOLOv8 | 2023 | Tối ưu thời gian xử lý |

2.3.3 Ứng dụng YOLO

**Giám sát an ninh:**

YOLO được sử dụng để phát hiện các đối tượng nghi ngờ trong các hệ thống giám sát an ninh, chẳng hạn như camera giám sát.

Ví dụ: YOLO có thể được sử dụng để phát hiện người xâm nhập, hành vi trộm cắp hoặc các hoạt động bất thường khác.

**Nhận diện khuôn mặt:**

YOLO được sử dụng để phát hiện khuôn mặt trong ảnh và video, giúp cho việc nhận diện khuôn mặt trở nên hiệu quả hơn.

YOLO có thể được sử dụng trong các hệ thống kiểm soát ra vào, thanh toán di động hoặc các ứng dụng xác thực danh tính khác.

**Phân loại ảnh:**

YOLO được sử dụng để phân loại ảnh tự động, giúp xác định các đối tượng trong ảnh một cách nhanh chóng và chính xác.

Ví dụ: YOLO có thể được sử dụng để phân loại ảnh sản phẩm, ảnh y tế hoặc ảnh phong cảnh.

**Xe tự lái:**

YOLO được sử dụng để phát hiện các đối tượng khác trên đường, chẳng hạn như xe cộ, người đi bộ và biển báo giao thông, giúp xe tự lái di chuyển an toàn hơn.

YOLO có thể được tích hợp vào hệ thống cảm biến của xe tự lái để nhận biết môi trường xung quanh và đưa ra quyết định lái xe phù hợp.

**Robot:**

YOLO được sử dụng để giúp robot nhận biết môi trường xung quanh và thực hiện các nhiệm vụ một cách hiệu quả hơn.

Ví dụ: YOLO có thể được sử dụng để giúp robot chọn và sắp xếp các đối tượng, hoặc để điều hướng robot trong môi trường phức tạp.

**Kiểm tra chất lượng sản phẩm:**

YOLO có thể được sử dụng để phát hiện các lỗi sản phẩm trong quá trình sản xuất.

**Giám sát y tế:**

YOLO có thể được sử dụng để phát hiện các dấu hiệu bất thường trong hình ảnh y tế.

**Nâng cao trải nghiệm người dùng:**

YOLO có thể được sử dụng để tạo ra các ứng dụng tương tác với người dùng dựa trên cử chỉ hoặc hành động.

**Ưu điểm của YOLO:**

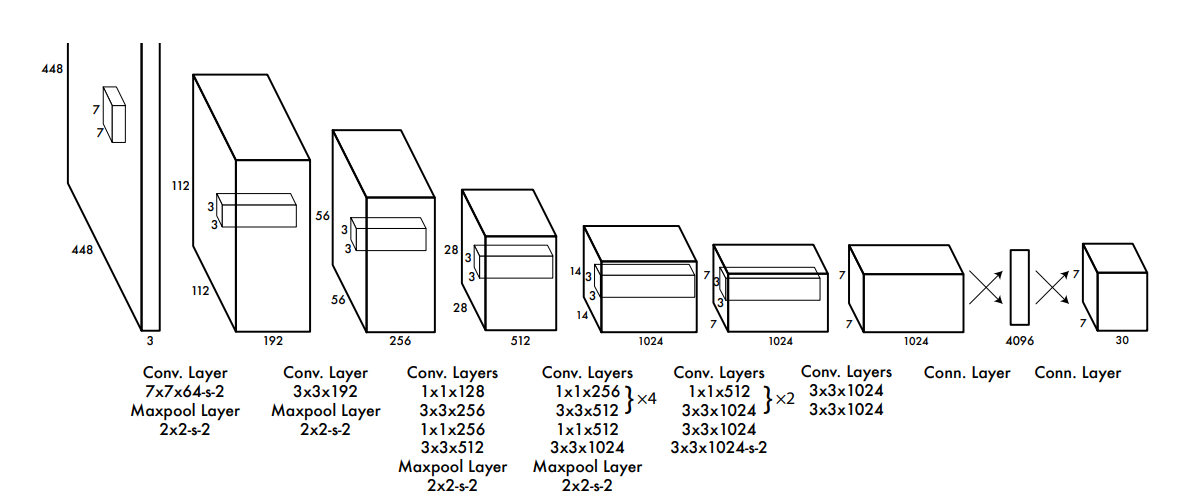
Tốc độ xử lý nhanh

Độ chính xác cao

Khả năng khái quát hóa tốt

Dễ dàng sử dụng.

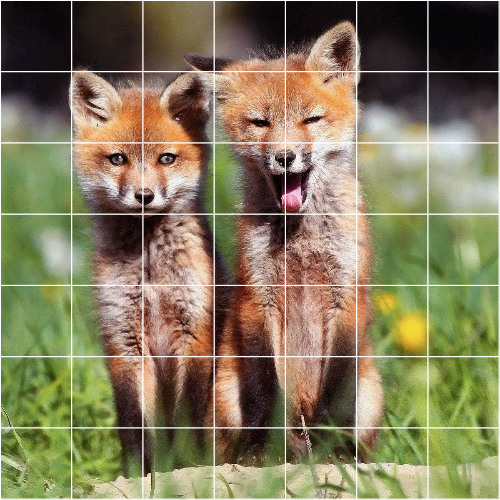
2.3.4 YOLOv1



Hình 2.6

24 lớp tích chập và 2 lớp FCL với input image 448\* 448, ta có được output là một vector với 4096 tham số được resize về 7\* 7\* 30

**Cách thức hoạt động**



Hình 2.7

Ý tưởng chia ảnh thành các grid cell để học vị trí của vật thể, nếu tâm của vật thể nằm tại ô nào thì ô đó sẽ đảm nhiệm việc detect ra bounding box cho vật thể ấy.

Mỗi grid cell sẽ dự đoán 2 bounding boxes và confidence score cho mỗi box đó.

Tổng bounding boxes cần dự đoán: 7x7x2 = 98 bounding boxes/ ảnh.

**Bounding boxes**

Bounding box là các hình chữ nhật bao xung quanh vật thể mang những thông tin về tọa độ và chiều dài chiều rộng của bounding box. Một bouding box sẽ có 5 giá trị thành viên, cụ thể:

Δx, Δy : Độ lệch giữa tọa độ tâm của bounding box dự đoán và tọa độ điểm bên trái trên cùng của grid cell mà nó thuộc về, giá trị của Δx,Δy vì vậy mà nằm trong khoảng [0,1].

Δw, Δh : Chiều rộng, chiều dài của bbox được chuẩn hóa theo kích thước của toàn bộ bức ảnh, giá trị của Δw, Δh nhờ vậy cũng bị ràng buộc trong khoảng [0,1]

c : Xác suất bên trong bouding box có vật thể là bao nhiêu.

Lý do độ lệch Δx, Δy được tính theo grid cell còn Δw, Δh lại được chuẩn hóa độ lớn của cả ảnh?

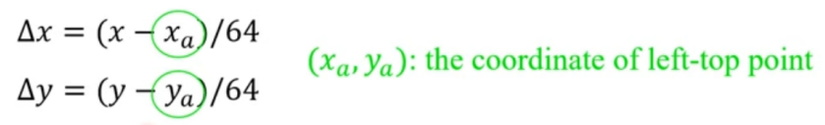
Vật thể sẽ được xác định vị trí theo grid cell, nghĩa là mô hình chỉ quan tâm việc dự đoán giá trị độ lệch ấy mà không cần quan tâm đến chính xác vị trí tâm vật thể ở đâu (tọa độ cụ thể).

Vị trí tâm của vật thể luôn có ràng buộc đến grid cell mà nó thuộc về, nhưng điều này cũng là một yếu điểm lớn của YOLOv1 vì khi một grid cell xác định một tâm vật thể và chỉ dự đoán xác suất một lần duy nhất dù cho có bao nhiêu bounding box trong grid cell => một grid cell tối đa chỉ dự đoán được một vật thể, cả ảnh tối đa chỉ dự đoán được 7\* 7 = 49 vật thể.

w, h được chuẩn hóa theo độ lớn của toàn ảnh vì khi đó giá trị của w, h nhỏ và việc dự đoán chiều rộng và chiều dài sẽ không bị ảnh hưởng bởi độ phân giải của ảnh.

Ví dụ ảnh đầu vào có là bao nhiêu thì model cũng chỉ việc dự đoán độ lệch chiều dài, chiều rộng trong khoảng [0,1] rồi về sau scale lại. Ngoài ra giá trị w,h trong khoảng [0,1] còn khiến việc tính loss và cập nhật tham số dễ dàng, nhanh chóng hơn.

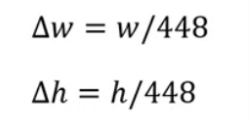
Công thức chuẩn hóa x, y center về [0,1]:



Hình 2.8

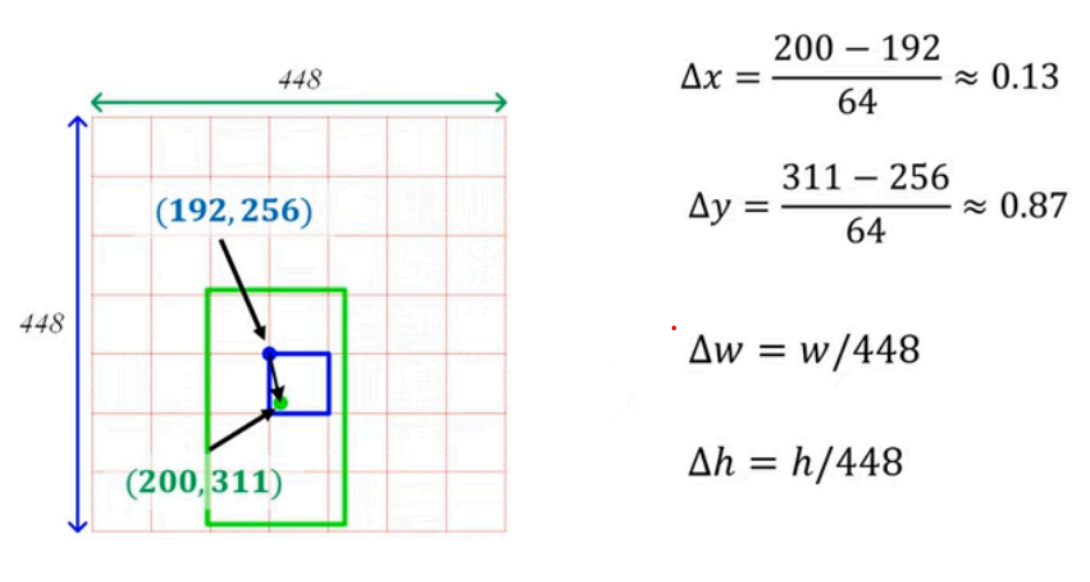
Tọa độ tâm của bounding box (x) trừ đi tọa độ của điểm góc trên cùng bên trái của grid cell, sau đó chia cho chiều rộng của một grid cell (trong ví dụ này là 64) với kích thước toàn ảnh là 448 \* 448.

Chuẩn hóa chiều dài chiều rộng của Bounding boxes:



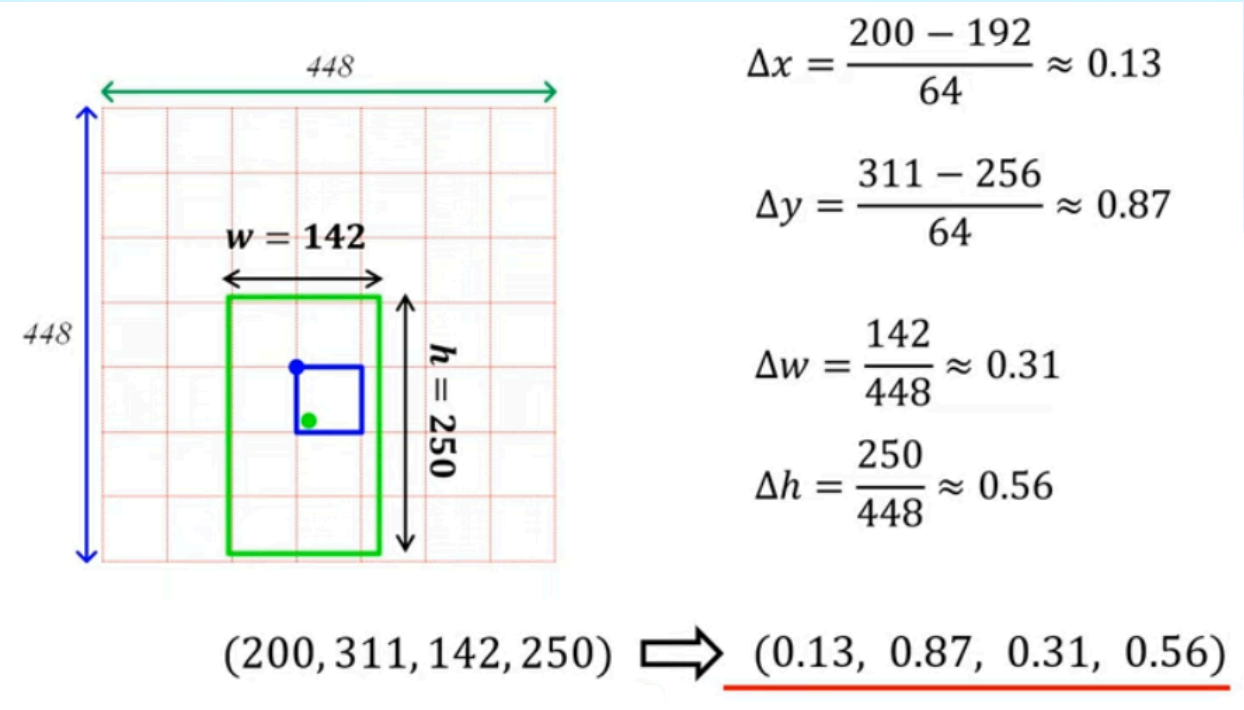
Hình 2.9

Ví dụ khi chuẩn hóa bằng công thức trên, với tọa độ tâm của vật thể là (200,311), Tọa độ góc trên cùng bên trái của grid cell chứa vật thể là (192,256), ta có 4 tọa độ của bounding box:



Hình 2.10

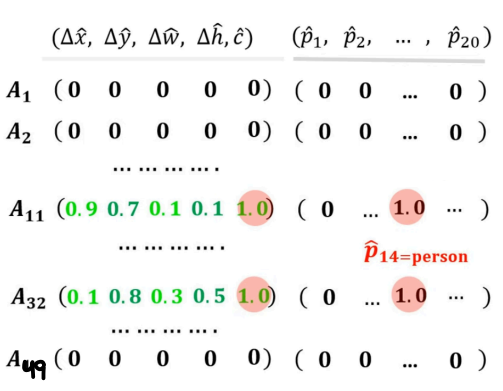
Ví dụ khi chuẩn hóa bằng công thức trên, với tọa độ tâm của vật thể là (200,311), Tọa độ góc trên cùng bên trái của grid cell chứa vật thể là (192,256), ta có 4 tọa độ của bounding box:



Hình 2.11

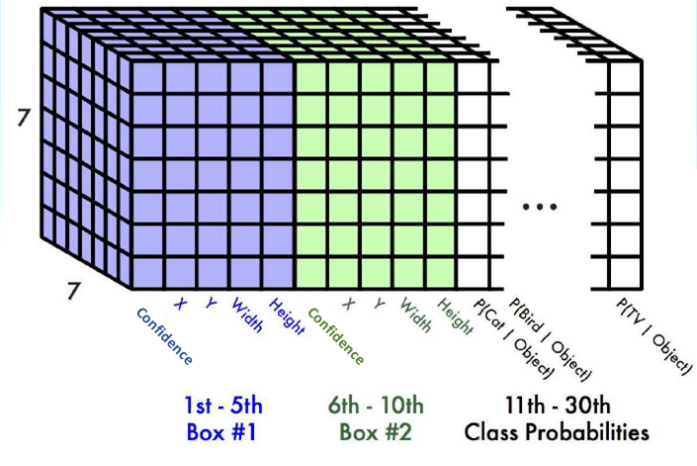
**Label encoding ground truth box**

Sau khi có các giá trị tọa độ bounding box thì tiến hành label encoding ground truth box chuẩn bị cho quá trình training.



Hình 2.12

**Predicted bounding box**

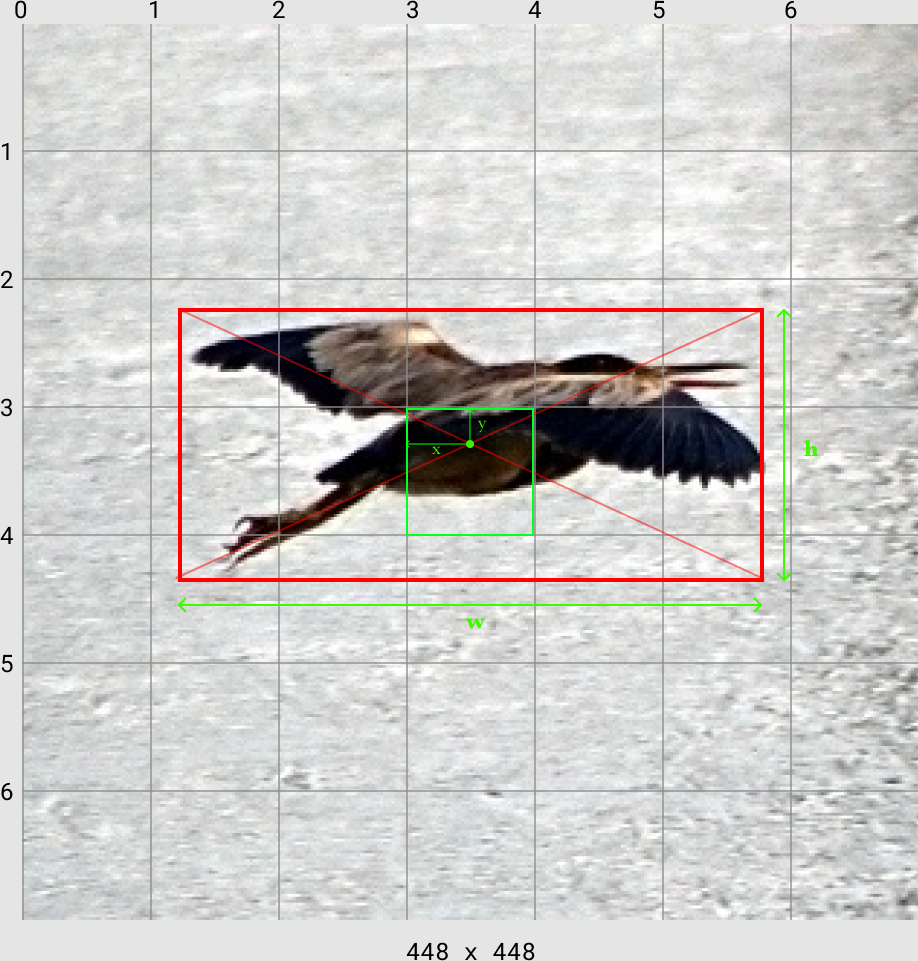


Hình 2.13

Thường một grid cell trong YOLOv1 sẽ dự đoán ra 2 bouding box, mỗi box có 5 giá trị ( x1, y1, w1, h1, c1), (x2, y2, w2, h2, c2).

Khác với labeled ground truth bbox, cho dù có 2 hay nhiều bounding box dự đoán trong một grid cell thì các nhãn label cũng chỉ được dự đoán một lần => tối đa một grid cell chỉ dự đoán được một object.

Như vậy, giả sử ta có 20 class trong bộ dữ liệu, cộng thêm 10 giá trị đến từ cả hai bounding box => Một grid cell mã hóa ra được một vector có độ lớn 20 + 10 = 30.



Hình 2.14

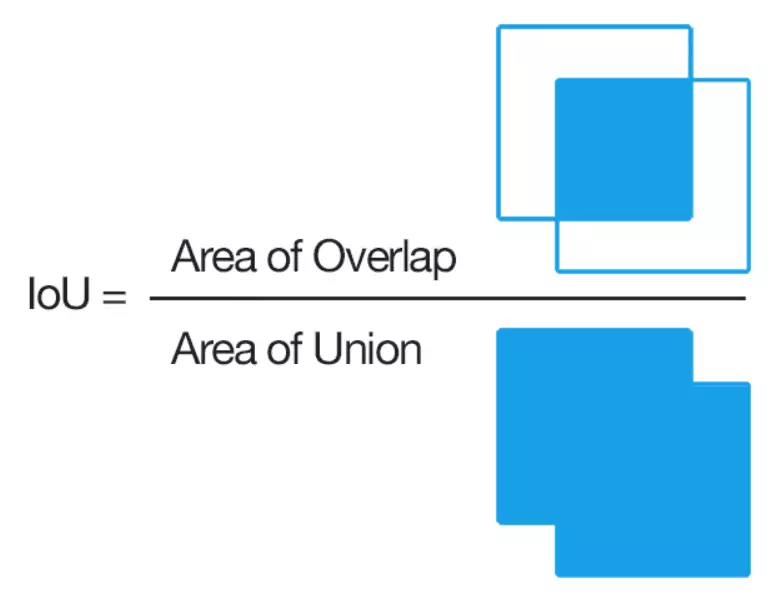
**Class confidence (Độ tự tin grid cell có chứa vật thể hay không)**



Hình 2.15

Pr(Object) là xác suất có vật thể ( 0 hoặc 1)

IoU(intersection over union) là công thức tính xem mức độ trùng nhau của bounding box dự đoán và ground truth box.



Hình 2.16

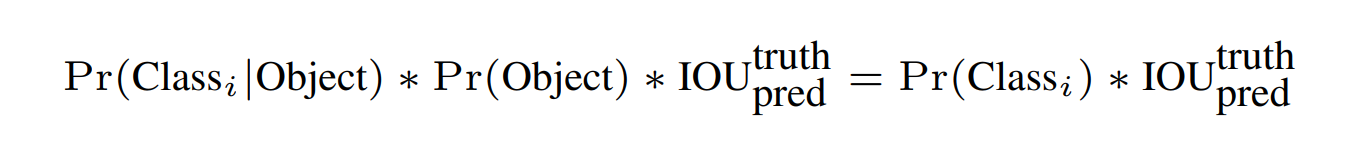
Phần tử là diện tích phần giao nhau của predicted bounding box và ground truth box, phần mẫu là tổng diện tích của 2 box trên hợp lại.

Người ta sẽ quy định một Ngưỡng, thường thì IoU score >= 0.5 có nghĩa là predicted bouding box có độ tương đồng cao với ground truth box.

Từ công thức tính trên ta thấy objectness confidence chính là IoU của box đó với bất kỳ ground truth box nào rồi từ đó chọn ra kết quả có IoU cao nhất để làm objectness confidence cho bbox.

Do YOLOv1 thường sẽ dự doạn 2 box cho một ô, nghĩa là không quan tâm ô đó có vật thể hay không thì model vẫn sẽ dự đoán ra 2 bouding box, khi ấy một ngưỡng IoU được người dùng đặt ra ( thường là 0.5), nếu ngưỡng IoU của predicted bounding box nào nhỏ hơn 0.5 thì sẽ bị loại bỏ.

**Confidence score (xác suất box đó chứa loại vật thể gì)**



Hình 2.17

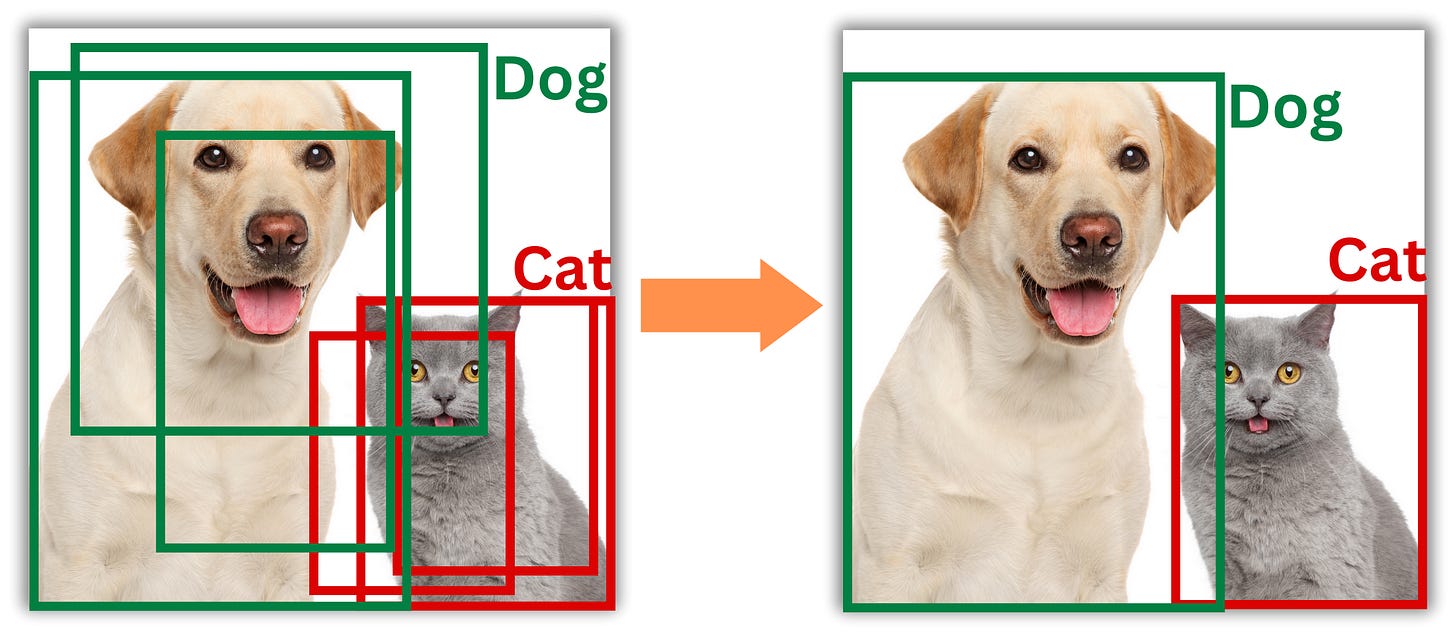
p(0,...,i,...n) được tính bằng xác suất có điều kiện giữa việc có vật thể thì nhãn của vật thể đó là gì, nhân với objectness confidence vừa tính ở trên. Khi có object, Pr(Object) = 1, Pr(Class(i)|Object) = Pr(Class(i)) nên ta có công thức tính xác suất của từng nhãn ở phía bên phải.

Confident score vừa thể hiện mức độ tương đồng của predicted bounding box và xác suất của nhãn được gán cho box là bao nhiêu.

**Non-max suppression**

Sau quá trình predicted bounding box của model, vẫn còn nhiều bounding box trùng lặp lên nhau với ngưỡng IoU với ground truth box cao.

Vậy làm thế nào để ta có thể loại bỏ đi những box dự đoán cùng một vật thể, chỉ giữ lại 1 box với độ chính xác cao nhất cho một vật thể. Câu trả lời là sử dụng thuật toán Non maximum Suppression.

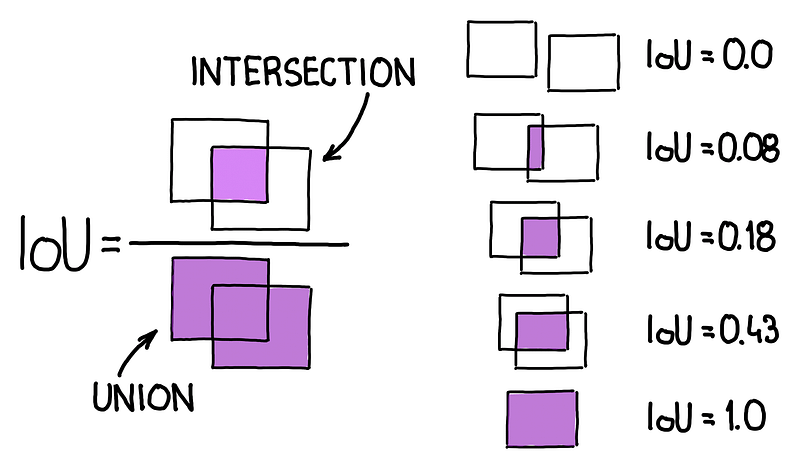


Hình 2.18

**Các bước:**

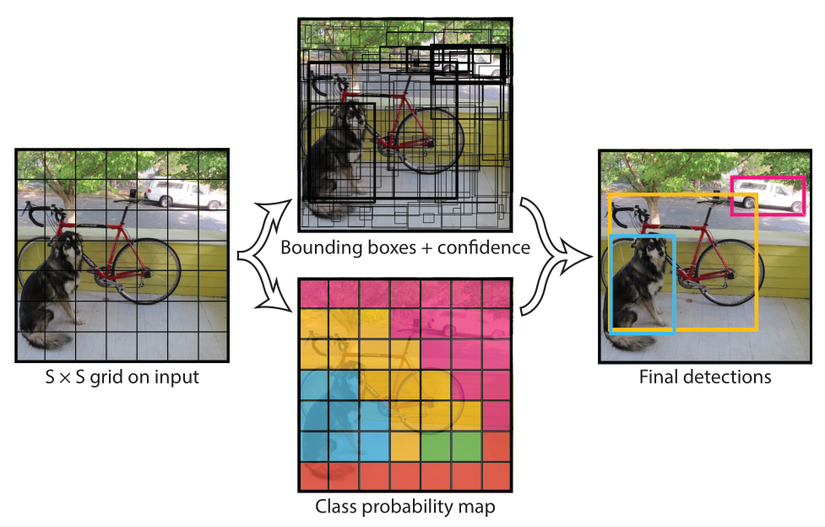
Bước 1: Ta có một list các bounding box còn lại gọi là P(box1, box2, box3,....) được sắp xếp theo thứ tự giảm dần về objeccness score. Những box đầu có độ tự tin cao sẽ được ưu tiên trước. Tiếp theo ta tạo một danh sách S gồm các box chuẩn đầu ra cuối cùng.

Bước 2: Lấy box đầu tiên từ tập P bỏ vào tập S, xóa box đó khỏi tập P sau đó tính tỷ lệ IoU với từng box còn lại trong tập P, nếu box nào trong tập P có tỷ lệ IoU so với box đang lấy làm chuẩn nhỏ hơn ngưỡng đặt ra (thường là 0,5) sẽ bị xóa khỏi tập P.



Hình 2.19

Bước 3: Nếu tập P vẫn còn phần tử thì tiếp tục lặp lại bước 2 cho đến khi P không còn phần tử nào nữa. Kết quả ta được tập S gồm các box phân biệt nhau cho từng vật thể.



Hình 2.20

2.3.5 YOLOv8

**Tổng quan về kiến trúc YOLOv8**

YOLOv8 là một mô hình anchor-free, nghĩa là nó không còn phụ thuộc vào các anchor box cố định như các phiên bản trước. Điều này giúp đơn giản hóa quá trình huấn luyện và cải thiện hiệu suất phát hiện.

YOLOv8 là một mô hình nhận dạng đối tượng dựa trên mạng convolutional neural network (CNN) được phát triển bởi Joseph Redmon và nhóm nghiên cứu của ông tại Đại học Washington.

YOLOv8 là phiên bản nâng cấp của YOLOv7, với khả năng nhận diện đối tượng nhanh hơn và chính xác hơn. Điều này được đạt được thông qua một số cải tiến, bao gồm mạng kim tự tháp đặc trưng, các mô-đun chú ý không gian và các kỹ thuật tăng cường dữ liệu tiên tiến.

Mô hình YOLOv8 sử dụng một mạng neural kiến trúc darknet-53 để trích xuất đặc trưng của hình ảnh và áp dụng thuật toán nhận dạng đối tượng YOLOv8 trên các đặc trưng đó.

**Key features:**

Tăng cường mô hình bằng cách thêm các kênh phân tán để tăng tốc độ tính toán.

Sử dụng kỹ thuật Attention để cải thiện khả năng nhận dạng đối tượng của mô hình.

Áp dụng phương pháp đào tạo mới để tăng tốc độ hội tụ.

Sử dụng kiến trúc mạng nơ-ron mới: Sử dụng kiến trúc YOLOv4 làm cơ sở để tăng hiệu suất và độ chính xác của mô hình.

Tích hợp cơ chế tự động điều chỉnh tỷ lệ tăng kích thước của hình ảnh đầu vào (AutoScale).

Hỗ trợ giám sát bằng video (Video Supervision): Mô hình có khả năng phát hiện và giám sát vật thể trong các video và đưa ra dự đoán liên tục trên toàn bộ video.

Tích hợp công nghệ Ensemble.

Tính năng điều chỉnh tỷ lệ tự động (AutoAnchor): Cải thiện việc phát hiện đối tượng với nhiều tỷ lệ khác nhau.

**Progression Timeline**



Hình 2.21

**YOLOv8 có thể được chia thành 3 thành phần chính:**

**A. Backbone (Xương sống – Trích xuất đặc trưng)**

Backbone trong YOLOv8 được xây dựng trên kiến trúc CSP (Cross Stage Partial) cải tiến, giúp:

Trích xuất đặc trưng hiệu quả hơn.

Giảm độ phức tạp mô hình (số tham số).

Tăng tốc độ suy luận.

Trong YOLOv8, backbone thường được gọi là C2f module, là phiên bản nâng cấp của C3 (từ YOLOv5). C2f tối ưu hóa tốc độ tính toán và độ chính xác.

**B. Neck (Truyền tải và tổng hợp đặc trưng)**

Phần Neck sử dụng FPN + PAN (Feature Pyramid Network + Path Aggregation Network) để:

Kết hợp đặc trưng từ nhiều tầng (đa cấp).

Phát hiện vật thể ở nhiều kích thước khác nhau (nhỏ – vừa – lớn).

Cải thiện khả năng định vị chính xác.

YOLOv8 vẫn giữ nguyên triết lý sử dụng Neck nhẹ và hiệu quả, nhấn mạnh khả năng suy luận thời gian thực.

**C. Head (Dự đoán kết quả)**

YOLOv8 dự đoán đầu ra theo dạng [x\_center, y\_center, width, height, confidence, class\_scores...] cho mỗi ô lưới, không còn dựa vào preset anchor.

Output kích thước: [batch, num\_predictions, 4 + 1 + num\_classes]

Trong đó:

4: tọa độ hộp (x, y, w, h)

1: độ tin cậy (objectness)

num\_classes: xác suất cho mỗi lớp

**Mô hình đầu ra (Output)**

YOLOv8 dự đoán đầu ra theo dạng [x\_center, y\_center, width, height, confidence, class\_scores...] cho mỗi ô lưới, không còn dựa vào preset anchor.

Output kích thước: [batch, num\_predictions, 4 + 1 + num\_classes]

Trong đó:

4: tọa độ hộp (x, y, w, h)

1: độ tin cậy (objectness)

num\_classes: xác suất cho mỗi lớp.

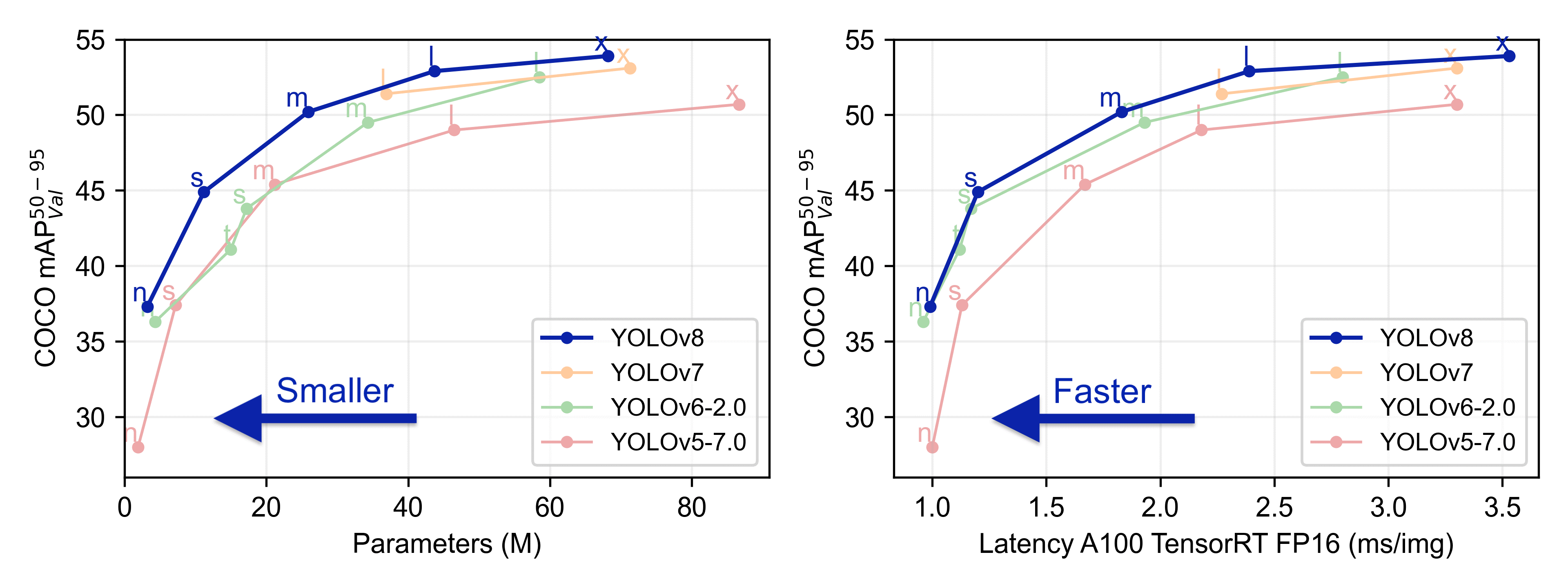
**Ưu điểm và hạn chế của YOLOv8**

**Ưu điểm**

Tốc độ: YOLOv8 được đánh giá là nhanh chóng và thời gian phản hồi thấp, giúp xử lý các tác vụ nhận diện đối tượng và phân-segment ảnh trong thời gian thực.

Độ chính xác: YOLOv8 được xây dựng trên các tiến bộ về học sâu và thị giác máy tính, đảm bảo độ chính xác cao trong việc nhận diện đối tượng.

Sự linh hoạt: YOLOv8 hỗ trợ việc nhận diện đối tượng và phân-segment trên cả GPU và CPU, tận dụng các công nghệ như TensorRT của Nvidia và OpenVino của Intel.



Hình 2.22

**Nhược điểm**

Để sử dụng YOLOv8 hiệu quả cần phải:

Có kiến thức chuyên sâu về Machine Learning, Deep Learning và các thuật toán liên quan.

Cần phải được huấn luyện trên một tập dữ liệu đủ lớn và đa dạng để đạt được hiệu quả cao nhất.

Yêu cầu các tài nguyên tính toán cao để đạt được tốc độ xử lý nhanh và chính xác.

Thuật toán YOLOv8 không phải là mã nguồn mở và chỉ có sẵn thông qua các thỏa thuận cấp phép với người tạo ra nó, Joseph Redmon.

YOLOv8 có thể không hoạt động tốt trong tất cả các môi trường và có thể cần thêm điều chỉnh hoặc tối ưu hóa để đạt được hiệu suất tối ưu.

2.4 Tổng quan về bài toán phân loại xe trong giám sát trạm thu phí

2.4.1 Ứng dụng phát hiện đối tượng trong giám sát và phân loại xe

Trong hệ thống giao thông hiện đại, trạm thu phí không chỉ đóng vai trò thu ngân sách mà còn là nơi quản lý và điều phối luồng giao thông. Một trong những yêu cầu quan trọng tại các trạm thu phí là phân loại phương tiện một cách chính xác để áp dụng mức phí phù hợp, đồng thời hỗ trợ công tác giám sát, chống gian lận và nâng cao hiệu quả vận hành hệ thống thu phí tự động (ETC).

Mỗi loại xe (xe máy, ô tô con, xe tải nhẹ, xe tải nặng, container, xe buýt...) đều có mức phí và điều kiện lưu thông khác nhau, do đó việc phân loại sai có thể dẫn đến thất thoát doanh thu hoặc khiếu nại từ người dân. Trong thực tế, việc kiểm tra bằng mắt thường hoặc các cảm biến đơn giản dễ xảy ra nhầm lẫn, đặc biệt trong điều kiện lưu lượng cao hoặc xe bị thay đổi hình dạng (lắp mui, chở quá tải...).

Với sự phát triển của thị giác máy tính (Computer Vision), đặc biệt là các mô hình phát hiện đối tượng (Object Detection) như YOLOv8, việc tự động nhận diện và phân loại phương tiện tại trạm thu phí đã trở nên khả thi, chính xác và có thể hoạt động trong thời gian thực.

Quy trình ứng dụng mô hình phân loại xe tại trạm thu phí gồm các bước:

Ghi nhận video/ảnh từ camera trạm thu phí.

Phát hiện và phân loại phương tiện trong từng khung hình thông qua mô hình YOLOv8.

Gắn nhãn loại xe tương ứng với cơ sở dữ liệu thu phí.

Gửi thông tin đến hệ thống quản lý thu phí để áp dụng mức phí chính xác hoặc kiểm tra vi phạm (xe vượt trạm, không dán thẻ, v.v.).

Ứng dụng này không chỉ hỗ trợ tăng tốc độ xử lý tại trạm, mà còn giúp giảm thiểu sai sót trong phân loại, tăng cường khả năng phát hiện các hành vi bất thường hoặc gian lận thu phí, từ đó góp phần nâng cao hiệu quả và minh bạch trong quản lý giao thông.

2.4.2 Thách thức và định hướng nghiên cứu

**Thách thức:**

Việc ứng dụng mô hình phát hiện đối tượng như YOLOv8 trong giám sát trạm thu phí thực tế gặp phải một số thách thức lớn, bao gồm:

Môi trường vận hành phức tạp: Các phương tiện tại trạm có thể bị che khuất, chạy gần nhau hoặc thay đổi hình dạng (lắp mái, chở hàng...).

Đa dạng chủng loại phương tiện: Khó phân biệt giữa các loại xe tương tự như xe tải nhẹ và bán tải, hoặc SUV với xe khách nhỏ.

Điều kiện ánh sáng thay đổi: Hoạt động vào ban đêm hoặc trong điều kiện thời tiết xấu (mưa, sương mù) làm giảm chất lượng ảnh đầu vào.

Yêu cầu xử lý thời gian thực: Hệ thống cần phát hiện và phản hồi nhanh chóng để không gây cản trở lưu thông.

Thiếu bộ dữ liệu chuyên biệt: Phần lớn các mô hình quốc tế chưa được huấn luyện với hình ảnh phương tiện đặc thù tại Việt Nam.

**Định hướng nghiên cứu:**

Việc ứng dụng mô hình phát hiện đối tượng như YOLOv8 trong giám sát trạm thu phí thực tế gặp phải một số thách thức lớn, bao gồm:

Môi trường vận hành phức tạp: Các phương tiện tại trạm có thể bị che khuất, chạy gần nhau hoặc thay đổi hình dạng (lắp mái, chở hàng...).

Đa dạng chủng loại phương tiện: Khó phân biệt giữa các loại xe tương tự như xe tải nhẹ và bán tải, hoặc SUV với xe khách nhỏ.

Điều kiện ánh sáng thay đổi: Hoạt động vào ban đêm hoặc trong điều kiện thời tiết xấu (mưa, sương mù) làm giảm chất lượng ảnh đầu vào.

Yêu cầu xử lý thời gian thực: Hệ thống cần phát hiện và phản hồi nhanh chóng để không gây cản trở lưu thông.

Thiếu bộ dữ liệu chuyên biệt: Phần lớn các mô hình quốc tế chưa được huấn luyện với hình ảnh phương tiện đặc thù tại Việt Nam.

Chương 3: Cơ Sở Thực Tiễn

3.1. Thực trạng và nhu cầu giám sát tại trạm thu phí

Hiện nay, các trạm thu phí tại Việt Nam và nhiều quốc gia khác đang áp dụng hai hình thức thu phí chính: thu phí thủ công và thu phí tự động không dừng (ETC – Electronic Toll Collection). Hình thức thu phí thủ công đòi hỏi sự tham gia của nhân viên thu phí để kiểm tra, phân loại phương tiện và thu tiền trực tiếp. Trong khi đó, hệ thống ETC hiện đại hơn, cho phép phương tiện được gắn thẻ định danh và qua trạm mà không cần dừng lại, giúp giảm thời gian chờ đợi.

Tuy nhiên, cả hai hình thức này đều đang đối mặt với nhiều vấn đề bất cập trong thực tế. Với thu phí thủ công, tình trạng ùn tắc cục bộ vào giờ cao điểm là điều thường xuyên xảy ra do thời gian xử lý lâu. Hệ thống ETC dù hiện đại hơn nhưng vẫn tồn tại rủi ro sai lệch trong việc nhận dạng loại xe, đặc biệt với những xe có cấu trúc không phổ biến hoặc bị che khuất. Ngoài ra, còn có nguy cơ gian lận trong việc phân loại xe, cố tình dùng thẻ sai loại để trốn phí hoặc gây nhầm lẫn trong hệ thống. Cả hai hình thức đều cần đến một lực lượng nhân sự đáng kể để giám sát và xử lý các tình huống phát sinh, làm gia tăng chi phí vận hành.

Chính vì vậy, việc phát triển một hệ thống giám sát thông minh, có khả năng **phân loại phương tiện một cách tự động, chính xác và không phụ thuộc vào yếu tố con người** là vô cùng cấp thiết. Giải pháp này không chỉ giúp **giảm chi phí nhân công**, mà còn nâng cao tính **minh bạch, hiệu quả và độ tin cậy** trong quá trình thu phí giao thông hiện đại

3.2. Xu hướng ứng dụng trí tuệ nhân tạo trong giám sát giao thông

Trong những năm gần đây, trí tuệ nhân tạo (AI) và học sâu (Deep Learning) đã trở thành xu hướng nổi bật và được ứng dụng mạnh mẽ trong nhiều lĩnh vực, đặc biệt là giao thông thông minh. Nhiều quốc gia phát triển như Trung Quốc, Hàn Quốc và Hoa Kỳ đã triển khai thành công các hệ thống giám sát tại trạm thu phí dựa trên nền tảng AI. Tại Trung Quốc, các camera được kết hợp với thuật toán học sâu giúp nhận dạng phương tiện, phân loại xe, ghi nhận biển số và xử lý thanh toán tự động, tất cả đều diễn ra trong thời gian thực mà không cần phương tiện phải dừng lại. Hàn Quốc đã triển khai hệ thống giám sát xe tự động kết hợp giữa AI và cơ sở dữ liệu quốc gia để đảm bảo độ chính xác tuyệt đối trong việc thu phí. Trong khi đó, Hoa Kỳ ứng dụng AI không chỉ trong nhận dạng xe mà còn trong phân tích luồng giao thông, tối ưu hóa vị trí đặt trạm thu phí và dự đoán thời gian cao điểm.

Bên cạnh các hệ thống ứng dụng thực tế, nhiều nghiên cứu học thuật cũng đã chứng minh hiệu quả của các mô hình học sâu như CNN, R-CNN và đặc biệt là YOLO (You Only Look Once) trong việc phát hiện và phân loại phương tiện giao thông với độ chính xác cao và tốc độ xử lý nhanh. Trong đó, mô hình YOLOv8 là một trong những phiên bản cải tiến hiện đại, giúp hệ thống có thể hoạt động ổn định trong môi trường thực tế với yêu cầu tài nguyên thấp.

Tại Việt Nam, việc triển khai các giải pháp giám sát thông minh ứng dụng AI tại các tuyến đường cao tốc, cầu BOT hay trạm thu phí đang dần được quan tâm. Với hạ tầng công nghệ ngày càng phát triển, việc ứng dụng các mô hình như YOLOv8 để nhận dạng phương tiện và giám sát tự động hoàn toàn khả thi. Điều này mở ra cơ hội hiện đại hóa giao thông, giảm tải nhân lực và nâng cao hiệu quả vận hành tại các điểm thu phí trong cả nước.

3.3. Tính khả thi và lợi ích của đề tài

Đề tài “Ứng dụng YOLOv8 phân loại xe để giám sát trong trạm thu phí” mang tính khả thi cao cả về mặt kỹ thuật lẫn điều kiện triển khai thực tế. Trước hết, mô hình YOLOv8 là một trong những kiến trúc mạng nơ-ron tiên tiến trong lĩnh vực nhận dạng đối tượng, với khả năng phát hiện và phân loại phương tiện theo thời gian thực với độ chính xác đáng tin cậy. Đặc biệt, phiên bản nhẹ YOLOv8n (nano) giúp giảm thiểu yêu cầu phần cứng, phù hợp để triển khai trên các hệ thống giám sát chi phí thấp như camera tại các trạm thu phí hiện nay.

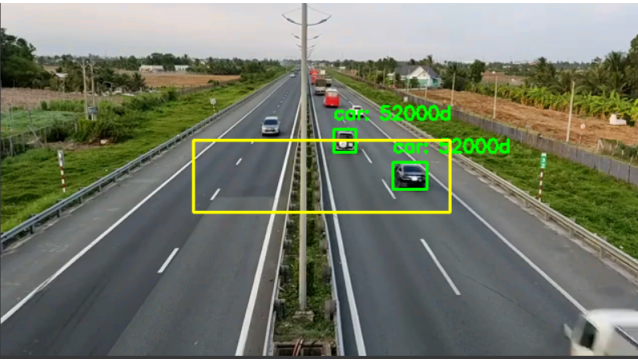
Bên cạnh đó, hệ thống không yêu cầu các thiết bị chuyên dụng như cảm biến từ trường hay hệ thống nhận dạng biển số phức tạp, mà chỉ cần camera quan sát thông thường kết hợp với phần mềm xử lý hình ảnh. Điều này giúp tiết kiệm chi phí đầu tư ban đầu và dễ dàng tích hợp vào hạ tầng có sẵn. Đặc biệt, việc sử dụng thuật toán vùng ảnh hưởng để xác định xe đi qua giúp đơn giản hóa logic kiểm soát, đồng thời có thể tùy biến để áp dụng cho từng loại trạm thu phí khác nhau.

Lợi ích của đề tài không chỉ dừng lại ở việc tự động hóa quá trình phân loại xe và tính phí, mà còn ở khả năng thống kê, giám sát lưu lượng xe theo thời gian và loại hình phương tiện. Điều này mang lại sự minh bạch trong quá trình thu phí, đồng thời hỗ trợ cơ quan quản lý trong việc phân tích dữ liệu, lập kế hoạch bảo trì đường bộ và dự báo doanh thu một cách hiệu quả hơn. Trong tương lai, hệ thống có thể dễ dàng mở rộng để tích hợp thêm chức năng nhận dạng biển số, phát hiện xe quá tải hoặc truy vết phương tiện vi phạm, góp phần xây dựng nền giao thông thông minh và hiện đại.

3.4. Giải pháp triển khai đề tài

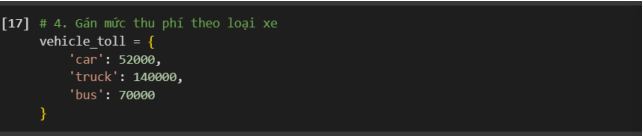
Để triển khai hệ thống giám sát và phân loại phương tiện tại trạm thu phí sử dụng mô hình YOLOv8, đề tài đề xuất giải pháp bao gồm các bước cụ thể như sau. Trước tiên, hệ thống sẽ được lắp đặt tại vị trí thuận lợi trong khu vực trạm thu phí, với một camera giám sát hướng về phía làn xe đang lưu thông. Camera này có nhiệm vụ ghi lại các đoạn video theo thời gian thực.

Tiếp theo, phần mềm xử lý sẽ thực hiện trích xuất khung hình (frames) từ video đầu vào và sử dụng mô hình YOLOv8 để nhận dạng các phương tiện trong từng khung hình. Để tăng tính chính xác và tránh đếm trùng, một vùng ảnh hưởng sẽ được xác định tại trung tâm khung hình – đây là vùng quyết định xe nào đã đi qua vùng ảnh hưởng. Chỉ khi phương tiện đi qua vùng ảnh hưởng này và không trùng lặp với phương tiện đã nhận dạng trước đó (được kiểm tra bằng thuật toán IOU), hệ thống mới tiến hành tính phí.



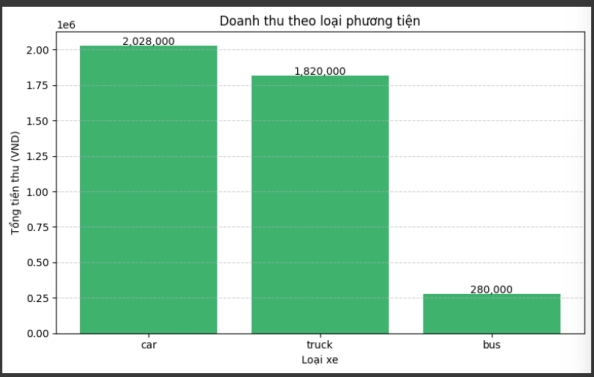
Hình 3.1

Mỗi loại phương tiện được gán một mức phí cố định dựa trên quy định hiện hành, và hệ thống sẽ cộng dồn giá trị thu được theo thời gian. Đồng thời, dữ liệu thu phí cũng được phân loại theo loại xe để phục vụ công tác thống kê và phân tích sau này. Kết quả có thể được hiển thị bằng biểu đồ trực quan, hoặc tích hợp vào hệ thống quản lý của đơn vị vận hành.



Hình 3.2

Giải pháp này hoàn toàn có thể triển khai tại các trạm thu phí hiện có mà không cần thay đổi nhiều về cơ sở hạ tầng. Việc sử dụng mô hình YOLOv8n nhẹ và hiệu quả cho phép vận hành hệ thống trên các thiết bị tầm trung hoặc chạy trên nền tảng đám mây như Google Colab, từ đó giảm chi phí đầu tư ban đầu. Hơn nữa, hệ thống có khả năng mở rộng linh hoạt để tích hợp thêm các chức năng nâng cao như nhận dạng biển số, phân tích lưu lượng hoặc phát hiện bất thường.



Hình 3.3

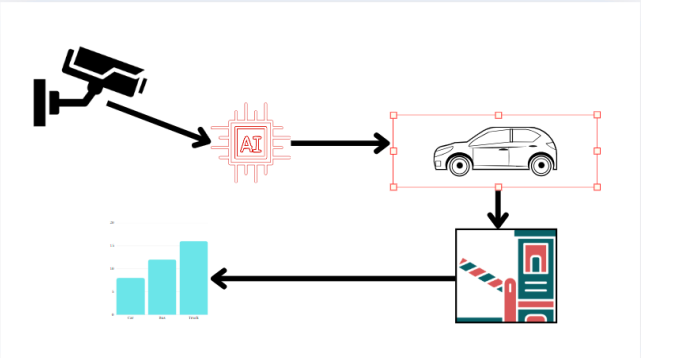
Chương 4: Thực Nghiệm

4.1 Mô hình hệ thống đề xuất

Hệ thống giám sát và phân loại phương tiện tại trạm thu phí được thiết kế theo một quy trình xử lý logic, gắn liền với ứng dụng của mô hình học sâu YOLOv8. Đầu tiên, một camera giám sát được đặt tại vị trí trạm thu phí nhằm ghi lại video phương tiện di chuyển qua trạm. Dữ liệu video này sau đó được truyền vào hệ thống xử lý trung tâm, nơi sử dụng mô hình YOLOv8 để nhận dạng và phân loại các đối tượng là phương tiện giao thông trong từng khung hình.

Tiếp theo, hệ thống sẽ xác định xem phương tiện có thực sự đi qua trạm thu phí hay không bằng cách kiểm tra vị trí trung tâm của xe có nằm trong “vùng ảnh hưởng” đã định trước giữa khung hình hay không. Chỉ những xe vượt qua vùng ảnh hưởng này và chưa được ghi nhận trước đó (dựa vào chỉ số IOU so với các xe đã tính phí) mới được xem là hợp lệ để tiến hành tính phí.

Mỗi phương tiện hợp lệ sẽ được đối chiếu với bảng giá thu phí tương ứng theo từng loại xe (car, truck, bus). Cuối cùng, hệ thống tự động cộng dồn số tiền thu được và đồng thời thực hiện thống kê, vẽ biểu đồ doanh thu theo loại phương tiện. Toàn bộ quá trình trên được tự động hóa, không yêu cầu can thiệp thủ công, đảm bảo tính minh bạch và hiệu quả trong công tác quản lý tại trạm thu phí.



Hình 4.1

4.2 Môi trường và công cụ thực hiện

Để đảm bảo tính hiệu quả và khả năng tái hiện trong môi trường học thuật cũng như thực tiễn, hệ thống được xây dựng và triển khai trên nền tảng điện toán đám mây **Google Colab**. Sử dụng ngôn ngữ lập trình **Python** - một trong những ngôn ngữ phổ biến và mạnh mẽ nhất trong lĩnh vực trí tuệ nhân tạo (**AI - Artificial Intelligence**)

Về mô hình nhận dạng, hệ thống sử dụng **YOLOv8n** ( phiên bản nano nhẹ của **YOLOv8**) do thư viện **Ultralytics** phát triển với khả năng nhận diện nhanh, chính xác và yêu cầu tài nguyên tính toán thấp. Việc sử dụng **YOLOv8** giúp phát hiện các đối tượng trong khung hình theo thời gian thực, phù hợp với mục tiêu giám sát giao thông ở trạm thu phí.

Ngoài ra, còn sử dụng thêm các thư viện khác hỗ trợ thêm như **OpenCV** được sử dụng để xử lý hình ảnh đầu vào và trích xuất thông tin từ video, trong khi đó **Matplotlib** được tích hợp để trực quan hoá dữ thiệu thông qua biểu đồ doanh thu theo từng loại phương tiện. Dữ liệu đầu vào là các video mô phỏng hoặc thực tế ghi lại quá trình lưu thông qua trạm thu phí. Hệ thống tập trung nhận diện các phương tiện chính trên cao tốc sau: xe hơi (**car**), xe tải (**truck**), và xe buýt (**bus**), mỗi loại được gán một mức phí cố định tương ứng.

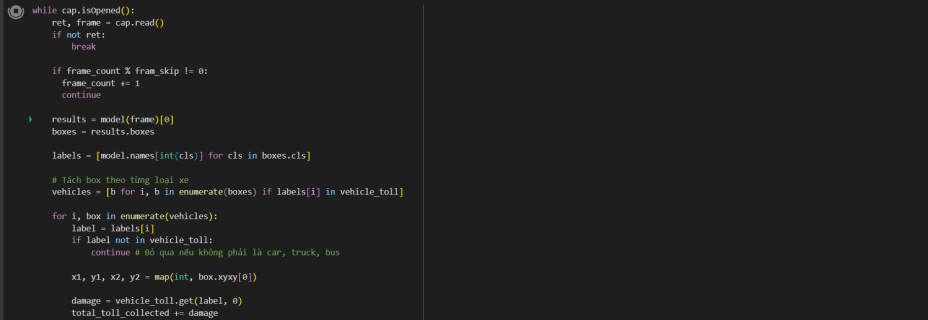
|  |  |
| --- | --- |
| Vehicle (Phương tiện) | Vehicle toll (Phí xe) |
| xe hơi (car) | 52000 |

|  |  |
| --- | --- |
| xe buýt (bus) | 70000 |
| xe tải | 140000 |

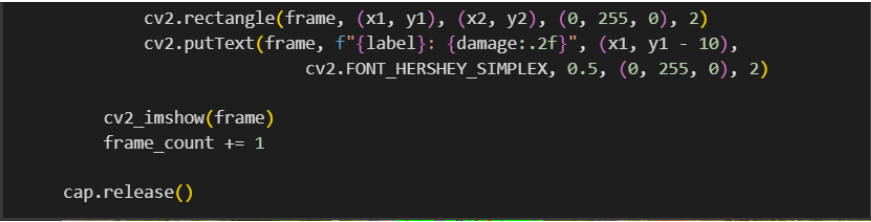
Việc triển khai trên Google Colab giúp sinh viên chúng em, nhà nghiên cứu hoặc tổ chức dễ dàng thử nghiệm, nâng cấp mô hình mà cần đầu tư hạ tầng máy chủ đắt đỏ, đồng thời thuận tiện chia sẻ và tái sử dụng mã nguồn.

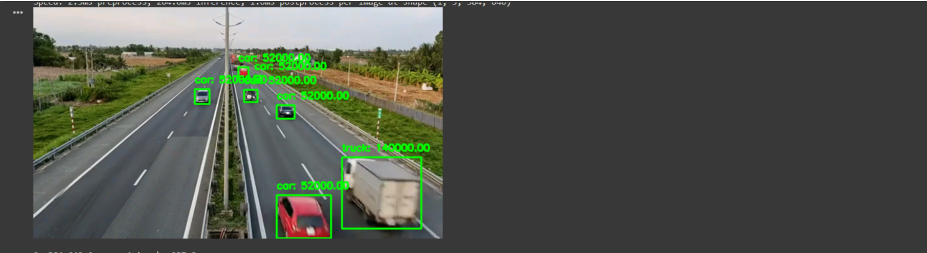
4.3 Quy trình xử lý và thuật toán chính:

Quy trình xử lý trong hệ thống giám sát thu phí được tổ chức theo chuỗi các bước rõ ràng và có tính tự động cao. Đầu tiên, hệ thống trích xuất từng khung hình (frame) từ video đầu vào. Với mỗi khung hình, mô hình YOLOv8 sẽ được sử dụng để nhận dạng và phân loại các phương tiện xuất hiện như xe hơi, xe tải hay xe buýt.



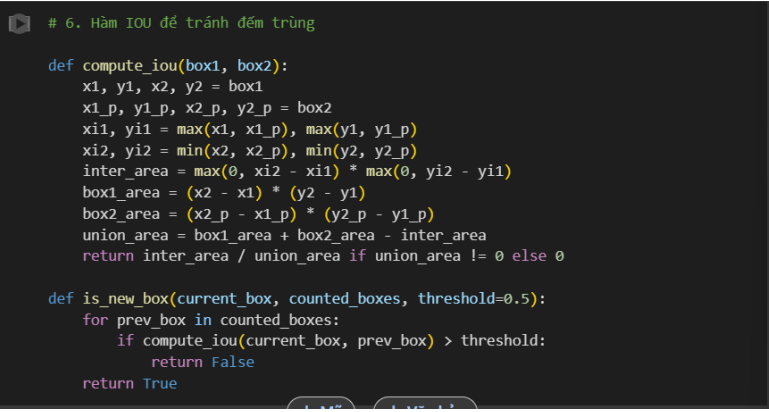
Hình 3.2

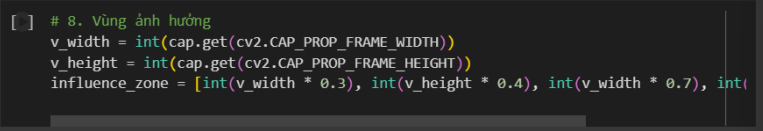


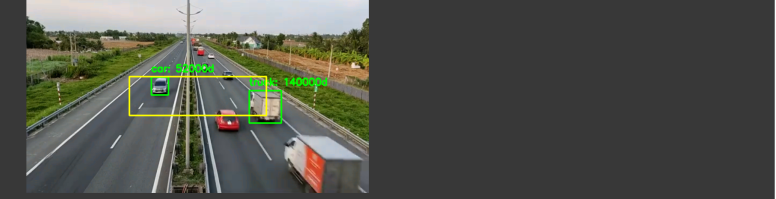
Hình 4.3

Hình 4.4

Sau khi phát hiện phương tiện, hệ thống tính toán vị trí trung tâm của bounding box để kiểm tra xem phương tiện có đi qua “vùng ảnh hưởng” giữa khung hình hay chưa. Đây là vùng đại diện cho vị trí xe đi qua trạm thu phí thực tế. Nếu xe đi qua vùng này và không bị trùng với các phương tiện đã được tính trước đó (kiểm tra bằng chỉ số giao nhau trên hợp – IOU), hệ thống sẽ gán mức phí tương ứng cho loại xe đó và cộng dồn vào tổng doanh thu.

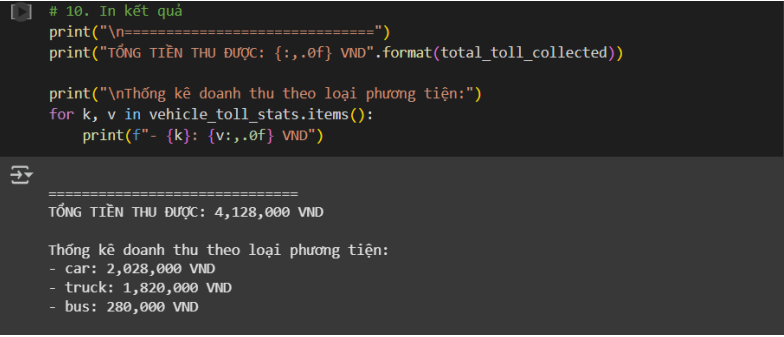


Hình 4.5

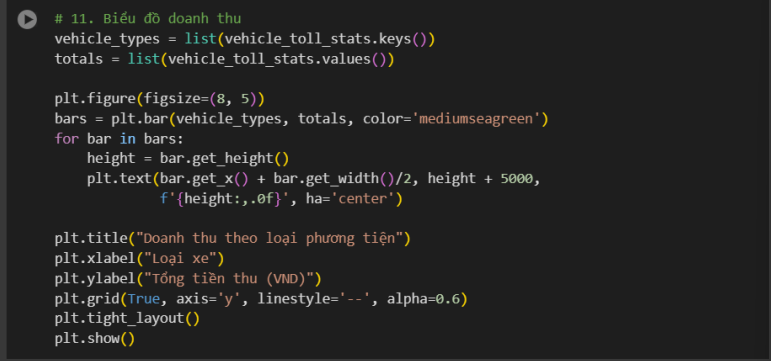
Hình 4.6

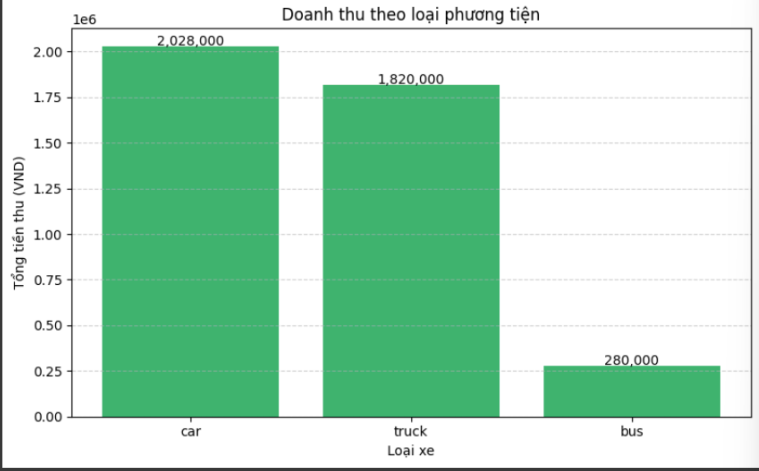
Hình 4.7

Toàn bộ thông tin về số lần phương tiện đi qua và tổng tiền thu được sẽ được lưu trữ để phục vụ thống kê, đồng thời được hiển thị dưới dạng biểu đồ giúp đánh giá hiệu quả vận hành trạm thu phí.



Hình 4.8

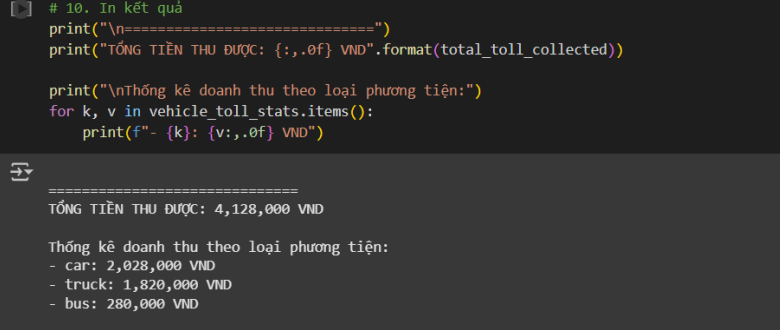
Hình 4.9



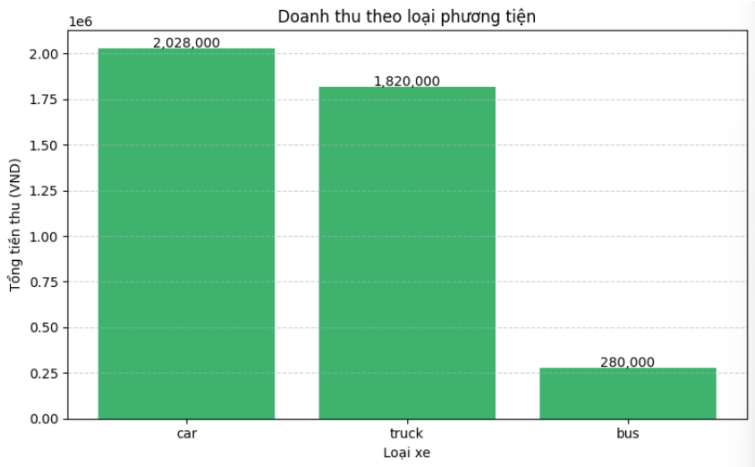
Hình 4.10

4.4 Kết quả và đánh giá:

Sau khi hệ thống được triển khai và thực nghiệm trên tập video mô phỏng hoạt động của trạm thu phí, kết quả cho thấy hệ thống hoạt động ổn định, khả năng nhận dạng phương tiện đạt độ chính xác cao. Các phương tiện được phân loại đúng theo từng loại: xe hơi, xe máy, xe tải và xe buýt, đồng thời được gán đúng mức phí theo cấu hình ban đầu.



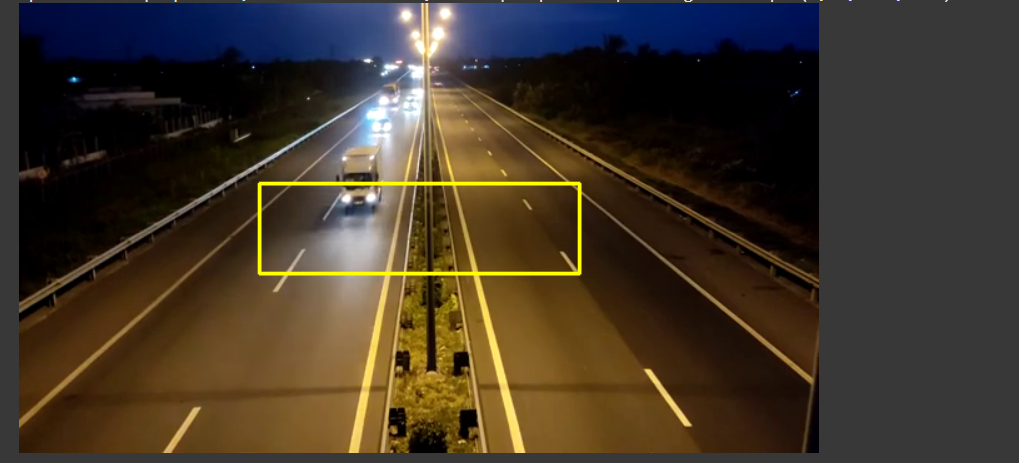
Hình 4.11

Hệ thống đã ghi nhận tổng số tiền thu được, đồng thời thống kê chi tiết số tiền tương ứng cho từng loại phương tiện. Kết quả này không chỉ được in trực tiếp mà còn được biểu diễn dưới dạng biểu đồ cột bằng thư viện **Matplotlib**, giúp việc đánh giá doanh thu trở nên trực quan và dễ hiểu hơn.

Hình 4.12

Về mặt hiệu suất, việc sử dụng mô hình YOLOv8n giúp quá trình xử lý diễn ra nhanh chóng ngay cả khi thực hiện trên Google Colab – một nền tảng điện toán đám mây miễn phí. Kỹ thuật bỏ qua frame (frame skip) được áp dụng cũng góp phần cải thiện thời gian xử lý mà vẫn đảm bảo độ chính xác cần thiết trong quá trình nhận dạng.

Tuy nhiên, hệ thống vẫn còn một số hạn chế như phụ thuộc vào góc quay của camera và điều kiện ánh sáng môi trường. Bên cạnh đó, việc chưa tích hợp tính năng nhận diện biển số xe khiến hệ thống chưa thể xác định danh tính cụ thể của từng phương tiện. Đây là những điểm có thể được cải tiến trong các phiên bản tiếp theo nhằm tăng độ tin cậy và mở rộng ứng dụng thực tế.

Hình minh hoạ khi ở trong môi trường loé sáng nhận diện của YOLOv8n còn bị hạn chế kém

Hình 4.13

Chương 5: Kết Luận

5.1 Nội dung đã thực hiện

Trong quá trình nghiên cứu và thực hiện ứng dụng YOLOv8 để phân loại xe tại trạm thu phí, chúng tôi đã hoàn thành các nội dung chính sau:

Tổng quan công nghệ YOLOv8: Đã nghiên cứu và áp dụng phiên bản YOLOv8, một trong những mô hình nhận diện đối tượng tiên tiến nhất, vào việc phân loại xe. Mô hình này được lựa chọn nhờ khả năng nhận dạng nhanh và độ chính xác cao trong các bài toán thị giác máy tính.

Thu thập và xử lý dữ liệu: Đã xây dựng một bộ dữ liệu chuyên dụng bao gồm các hình ảnh xe từ camera tại các trạm thu phí, với sự đa dạng về loại xe (xe tải, xe buýt, v.v.). Quá trình này bao gồm gắn nhãn dữ liệu để mô hình có thể học hiệu quả.

Sử dụng thuật toán IoU: Để tránh đếm trùng lặp các phương tiện khi đi qua tâm của video nên chúng em áp dụng thuật toán để nhằm tăng độ chính xác và khả năng áp dụng thực tế cao hơn.

Đánh giá kết quả: Hệ thống đạt được độ chính xác cao trong việc nhận dạng các loại xe phổ biến tại trạm thu phí, đảm bảo đáp ứng yêu cầu giám sát trong môi trường thực tế.

5.2 Hạn chế

Mặc dù đạt được một số kết quả khả quan, quá trình thực hiện vẫn gặp phải các hạn chế sau:

Dữ liệu không đủ đa dạng: Một số loại xe ít xuất hiện trong bộ dữ liệu, dẫn đến khả năng nhận dạng chưa tối ưu.

Hạn chế về hiệu năng: Khi áp dụng trong thời gian thực, mô hình yêu cầu phần cứng mạnh, gây khó khăn trong việc triển khai tại các trạm thu phí quy mô lớn hoặc vùng xa.

Vấn đề môi trường: Điều kiện ánh sáng kém hoặc thời tiết xấu (mưa, sương mù) làm giảm độ chính xác của mô hình.

Chi phí triển khai cao: Việc tích hợp mô hình vào hệ thống thực tế đòi hỏi đầu tư vào phần cứng và phần mềm.

5.3 Hướng phát triển

Để khắc phục các hạn chế và nâng cao hiệu quả ứng dụng, các hướng phát triển sau đây sẽ được cân nhắc:

Mở rộng bộ dữ liệu: Tiến hành thu thập thêm hình ảnh từ nhiều trạm thu phí khác nhau, bao gồm cả các điều kiện môi trường khắc nghiệt để tăng tính đa dạng và khả năng tổng quát hóa của mô hình.

Cải tiến mô hình: Tích hợp các phương pháp xử lý ảnh tiên tiến như làm sạch hình ảnh hoặc tăng cường dữ liệu (data augmentation). Đồng thời, áp dụng các thuật toán tối ưu hóa để giảm thiểu tài nguyên cần thiết.

Ứng dụng công nghệ bổ trợ: Kết hợp YOLOv8 với công nghệ IoT (Internet of Things) để truyền dữ liệu thời gian thực hoặc sử dụng blockchain để đảm bảo tính minh bạch trong hệ thống thanh toán.

Nghiên cứu mở rộng: Tiếp tục ứng dụng mô hình vào các lĩnh vực giám sát khác như phân tích hành vi lái xe, phát hiện vi phạm giao thông hoặc tự động hóa quy trình thu phí.

TÀI LIỆU THAM KHẢO

Ultralytics. (2023). *YOLOv8 - Ultralytics Documentation.* Retrieved from <https://docs.ultralytics.com>

Ultralytics. (2023). *YOLOv8 GitHub Repository.* GitHub. <https://github.com/ultralytics/ultralytics>

Source code tiểu luận :

https://colab.research.google.com/drive/1Nj9VewvpLde-NoYtT\_x0Wzhpi5d30kH-?usp=sharing