

Phát hiện và Đếm Đối tượng trong Ảnh(Người, xe, vật thể) sử dụng YOLOv5

SVTH1: Phạm Thế Hùng - 2374802010164

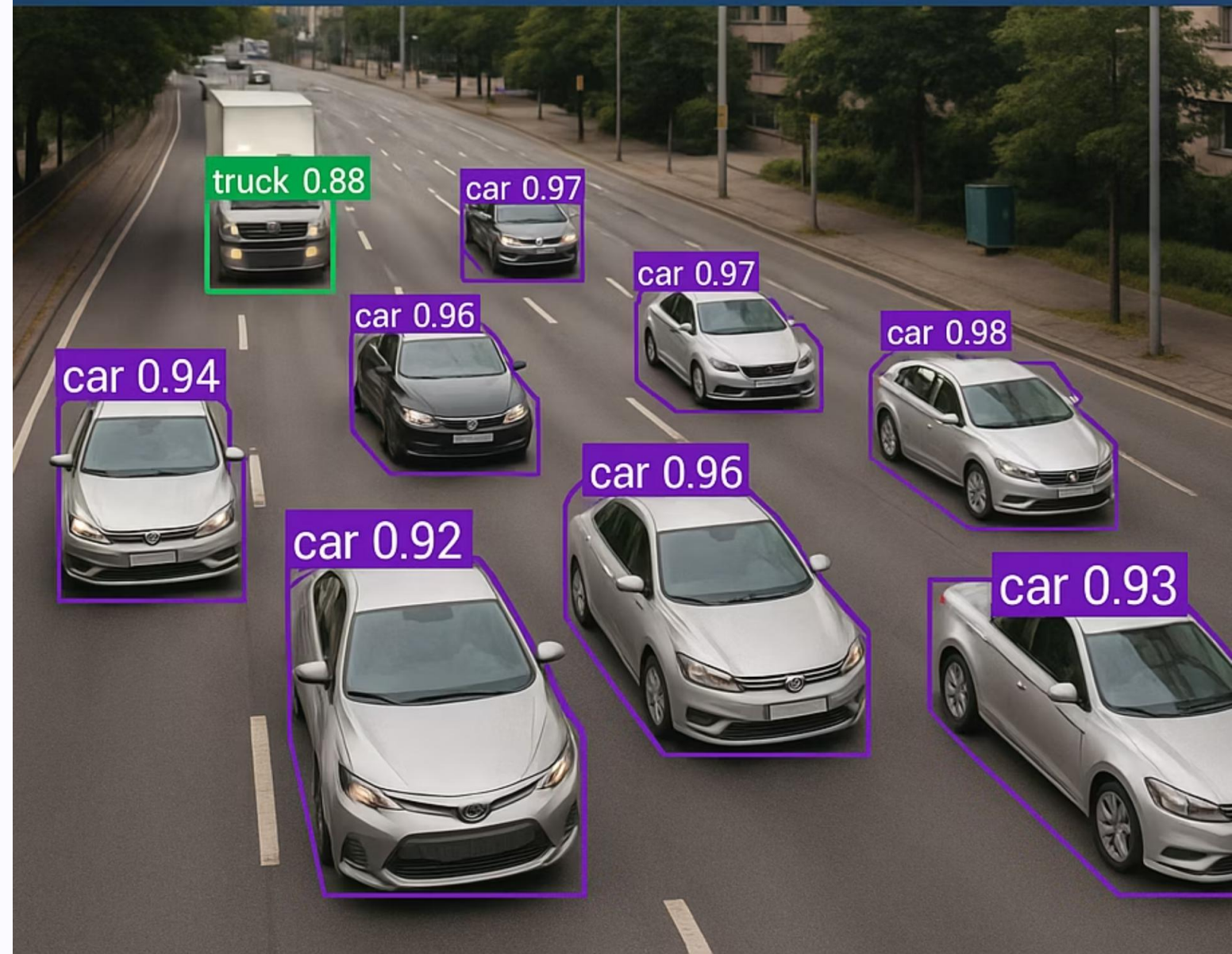
SVTH2:Phan Trần Quang Thái- 2174802010116

SVTH3: Đoàn Hoàng Quân - 2374802010412

SVTH4: Nguyễn Chí Dũng - 2374802013469

GVHD: T.S Đỗ Hữu Quân

ĐẾM XE BẰNG YOLOv5



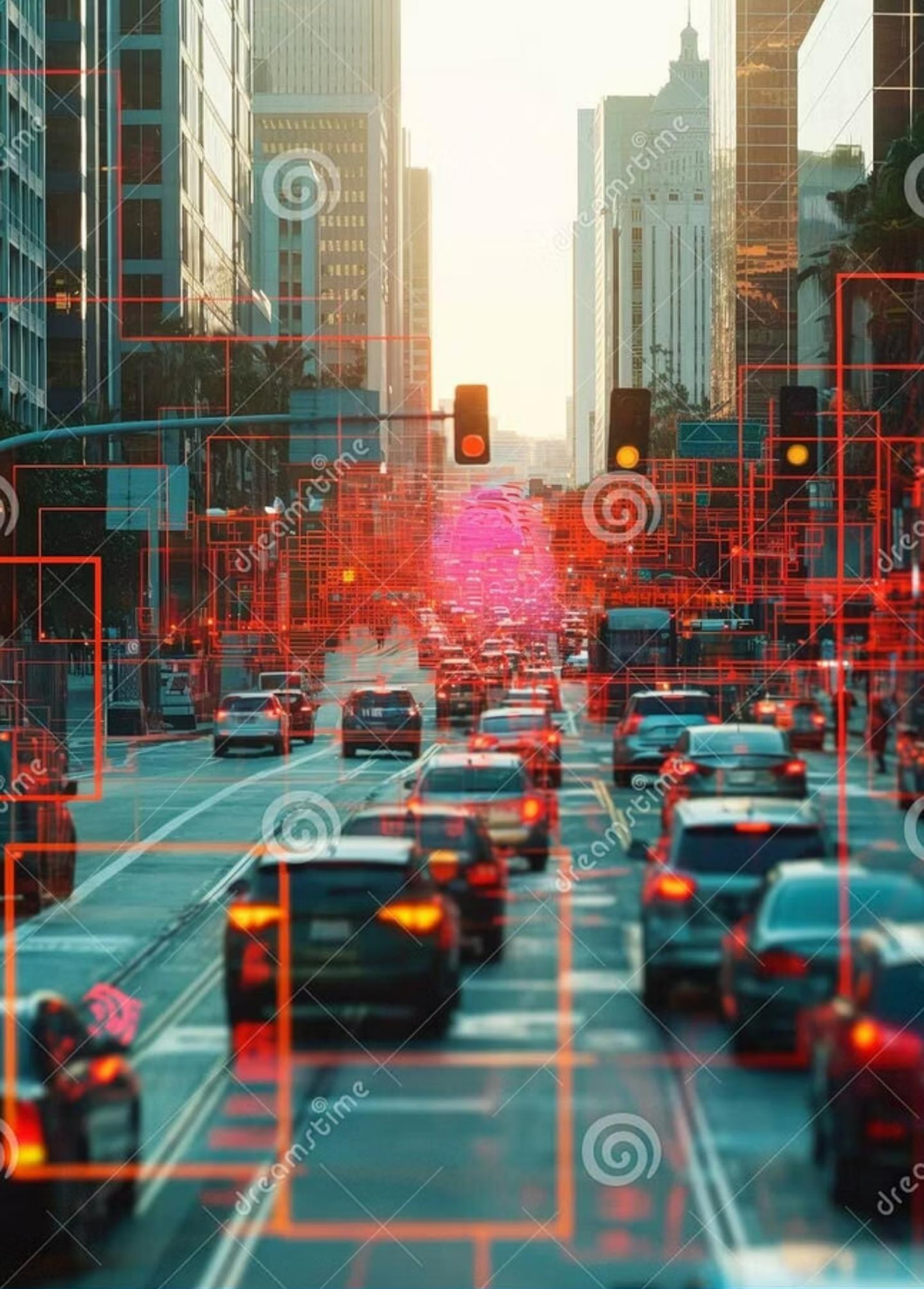
Mục Lục

- Lời cảm ơn
- Chương 1: Mở Đầu
- Chương 2: Cơ sở lý thuyết
- Chương 3: Phân tích và thiết kế hệ thống
- Chương 4: Cài đặt và thử nghiệm
- Chương 5: Kết luận và hướng phát triển
- Tài liệu tham khảo

Lời Cảm Ơn

Nhóm chúng em, nhóm 21, xin chân thành cảm ơn thầy Đỗ Hữu Quân, người đã tận tình dìu dắt chúng em từ những kiến thức cơ bản đến những buổi học cuối cùng. Công lao và tình thương của thầy là hành trang vững chắc cho sự nghiệp của chúng em trong tương lai. Chúng em mãi yêu thầy.

TP. Hồ Chí Minh, ngày 14 tháng 7 năm 2025



Chương 1: Mở Đầu

1.1. Lý do chọn đề tài nghiên cứu

Trong thời đại công nghệ 4.0, thị giác máy tính ngày càng quan trọng. Đề tài "Phát hiện và đếm số lượng đối tượng trong ảnh sử dụng YOLOv5" được chọn vì tính ứng dụng cao trong giám sát, giao thông, và robot thông minh. YOLOv5 là mô hình hiện đại, tối ưu cho nhận diện. Đề tài giúp nhóm áp dụng lý thuyết xử lý ảnh vào thực tiễn, rèn luyện kỹ năng lập trình và mở rộng sang các hệ thống thông minh khác.

1.2. Công Cụ và Thư Viện



Ngôn ngữ

Python



Thư viện chính

PyTorch, OpenCV, NumPy, matplotlib



Framework

YOLOv5 từ Ultralytics



Môi trường

Google Colab hoặc chạy local với GPU

1.3. Đối tượng và Phạm vi nghiên cứu

Đối tượng nghiên cứu

- Kỹ thuật phát hiện đối tượng (object detection) trong ảnh số.
- Mô hình YOLOv5 (You Only Look Once version 5).
- Các đối tượng cần phát hiện và đếm: người, xe máy, ô tô, xe đạp, và vật thể phổ biến khác trong ảnh tĩnh.
- Ảnh đầu vào từ tập dữ liệu có sẵn hoặc do người dùng cung cấp (JPG, PNG).

Phạm vi nghiên cứu

- Tập trung vào phát hiện và đếm đối tượng trong ảnh tĩnh, không xử lý video hay ảnh thời gian thực.
- Sử dụng mô hình YOLOv5 pre-trained trên tập dữ liệu COCO (80 lớp đối tượng).
- Không huấn luyện lại hay tinh chỉnh mô hình trên tập dữ liệu riêng.
- Chỉ áp dụng cho ảnh rõ ràng, chất lượng tương đối, không xử lý ảnh mờ, thiếu sáng, nhiễu cao.
- Không đi sâu vào tối ưu hiệu năng hay so sánh chi tiết với các mô hình khác.

1.4. Phương pháp nghiên cứu

Thu thập và Xử lý Dữ liệu

Xây dựng/sử dụng bộ dữ liệu, gán nhãn bằng công cụ chuyên dụng (LabelImg), tiền xử lý dữ liệu với các kỹ thuật tăng cường (Mosaic Augmentation).

Thiết kế và Cấu hình Mô hình

Lựa chọn kiến trúc YOLOv5 (n, s, m, l, x) phù hợp với yêu cầu tốc độ/độ chính xác, tùy chỉnh tham số (learning rate, batch size, epoch).

Huấn luyện Mô hình

Thiết lập môi trường (PyTorch, Ultralytics YOLOv5, driver GPU), tiến hành huấn luyện, sử dụng Transfer Learning từ mô hình đã huấn luyện sẵn.

Đánh giá và Phân tích Kết quả

Đánh giá bằng mAP, IoU, FPS, Recall, Precision. Phân tích hiệu suất, so sánh với các phiên bản YOLOv5 khác, đề xuất cải thiện.

Triển khai và Ứng dụng

Triển khai mô hình trên các nền tảng (web, mobile, thiết bị nhúng), tối ưu hóa mô hình (quantization, pruning) cho tài nguyên hạn chế.

Chương 2: Cơ sở lý thuyết

2.1. Tổng quan xử lý ảnh số

Xử lý ảnh số là lĩnh vực sử dụng máy tính để thao tác trên ảnh số, nhằm thu được thông tin giá trị hoặc cải thiện chất lượng ảnh. Quá trình này gồm ba giai đoạn chính:

- **Nhập ảnh:** Số hóa ảnh từ camera, máy quét, hoặc thiết bị y tế thành ma trận điểm ảnh.
- **Xử lý ảnh:** Áp dụng thuật toán để biến đổi ma trận điểm ảnh (lọc, biến đổi Fourier, nén).
- **Phân tích ảnh:** Trích xuất thông tin có ý nghĩa (xác định đối tượng, nhận dạng ký tự, đo lường đặc trưng).

2.2. Các khái niệm và thuật toán liên quan

Bộ lọc ảnh và Phát hiện cạnh

Các lớp tích chập trong YOLOv5 hoạt động như bộ lọc học được, tự động trích xuất đặc trưng như cạnh, đường nét, kết cấu. Các lớp đầu tiên phát hiện đặc trưng đơn giản, lớp sâu hơn nhận biết thành phần phức tạp.

Phân đoạn ảnh

YOLOv5 ban đầu dự đoán khung bao và tên lớp. Các phiên bản gần đây hỗ trợ phân đoạn thể hiện, xác định chính xác từng pixel thuộc về đối tượng (tạo mặt nạ).

Các thuật toán cốt lõi trong YOLOv5

Sử dụng CSPDarknet (Backbone) trích xuất đặc trưng, FPN và PAN (Neck) kết hợp đặc trưng đa tầng, và Head dự đoán vị trí, độ tin cậy, lớp đối tượng với anchor boxes.

2.3. Mô hình áp dụng YOLOv5

YOLOv5 là mô hình phát hiện đối tượng hàng đầu, được thiết kế để cân bằng tối ưu giữa tốc độ và độ chính xác. Nó có khả năng xử lý hình ảnh theo thời gian thực với hiệu suất cao.

- **Tốc độ và Hiệu quả:** Xử lý real-time, lý tưởng cho giám sát giao thông và hệ thống hỗ trợ lái xe.
- **Độ chính xác cao:** Phát hiện đa dạng đối tượng, kể cả trong điều kiện môi trường phức tạp.
- **Dễ sử dụng và Tối ưu:** Mã nguồn mở hỗ trợ tốt, dễ cài đặt, huấn luyện và triển khai với nhiều phiên bản.
- **Hỗ trợ đa nhiệm:** Ngoài phát hiện, còn hỗ trợ phân đoạn thể hiện (instance segmentation).



YOLOv5 vượt trội hơn các phương pháp truyền thống như Haar Cascade hay Otsu, và là phiên bản tối ưu của mô hình CNN cơ bản cho các ứng dụng thực tế.

Chương 3: Phân tích và thiết kế hệ thống

3.1. Quy trình xử lý tổng thể

Chuẩn bị Đầu vào

Hệ thống tiếp nhận và tiền xử lý hình ảnh (thay đổi kích thước, chuẩn hóa) để phù hợp với mô hình YOLOv5.

Tải và Áp dụng Mô hình

Tải mô hình YOLOv5 đã huấn luyện, sau đó mô hình sẽ quét và dự đoán các đối tượng tiềm năng trên ảnh.

Hiển thị Kết quả

Dựa trên dự đoán, hệ thống vẽ các hộp giới hạn (bounding boxes) và đếm tổng số lượng đối tượng, hiển thị trực quan cho người dùng.

Mở Rộng

Quy trình có thể được mở rộng để xử lý hàng loạt hình ảnh hoặc video, cho phép phát hiện và đếm đối tượng liên tục trong thời gian thực.

3.2. Kiến trúc hệ thống

AB



Khối Đầu vào

Thu thập hình ảnh hoặc video từ nhiều nguồn.

- **Camera:** Thu nhận video trực tiếp.
- **Tệp tin:** Đọc video (.mp4, .avi) hoặc ảnh (.jpg, .png).

Khối Xử lý Trung tâm

Tiền xử lý dữ liệu và thực hiện suy luận.

- **Tiền xử lý:** Cắt, thay đổi kích thước ảnh.
- **Mô hình YOLOv5:** Trích xuất đặc trưng, dự đoán đối tượng.
- **Hậu xử lý:** Áp dụng NMS loại bỏ khung bao trùng lặp.

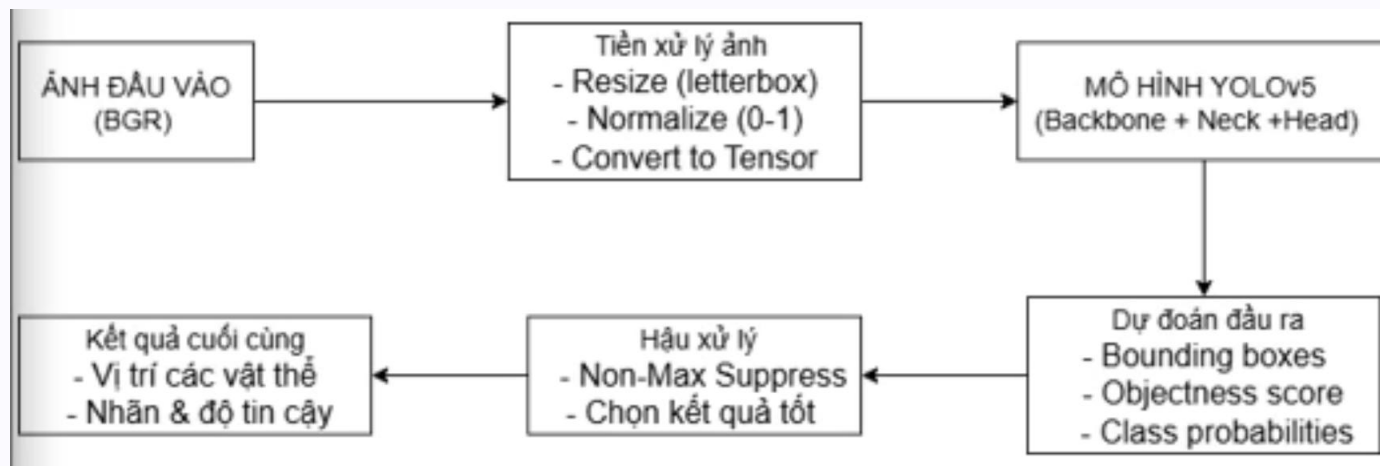
Khối Đầu ra

Trình bày kết quả nhận diện và lưu trữ.

- **Hiển thị:** Giao diện người dùng trực quan.
- **Lưu trữ:** Lưu lại video hoặc hình ảnh kết quả.

Kiến trúc hệ thống được chia thành ba khối chức năng chính, đảm bảo quy trình xử lý và hiển thị đối tượng hiệu quả.

Sơ đồ luồng xử lý:



3.3. Thiết kế các bước xử lý ảnh cụ thể

1. Đọc và Tiền xử lý Ảnh

Chuyển đổi dữ liệu ảnh hoặc video thô thành định dạng phù hợp cho mô hình YOLOv5.

- **Đọc dữ liệu:** Sử dụng OpenCV để đọc ảnh hoặc khung hình từ video.
- **Thay đổi kích thước:** Chuẩn hóa về 640x640 pixels.
- **Chuyển đổi màu:** Chuyển đổi từ BGR sang RGB nếu cần.
- **Chuẩn hóa:** Đưa giá trị pixel về khoảng 0-1.
- **Định dạng lại:** Chuyển định dạng tensor từ (H, W, C) sang (C, H, W).

2. Suy luận (Inference)

Áp dụng mô hình YOLOv5 để xác định vị trí và phân loại đối tượng trong ảnh đã tiền xử lý.

- **Tải mô hình:** Nạp mô hình YOLOv5 đã huấn luyện sẵn.
- **Dự đoán:** Đưa ảnh vào mô hình để nhận kết quả.
- **Giải mã kết quả:** Trích xuất khung bao, điểm tin cậy và lớp đối tượng.

3. Hiển thị và Lưu trữ Kết quả

Trực quan hóa kết quả nhận diện và lưu trữ nếu cần.

- **Vẽ khung bao và nhãn:** Sử dụng OpenCV vẽ hộp giới hạn và thông tin lớp, độ tin cậy.
- **Hiển thị:** Trình chiếu khung hình đã xử lý trên màn hình.
- **Lưu trữ:** Ghi lại ảnh hoặc video đã được nhận diện.

3.4. Mô tả dữ liệu đầu vào / đầu ra

Dữ liệu đầu vào

Hệ thống chấp nhận nhiều định dạng dữ liệu hình ảnh và video từ các nguồn khác nhau.

- **Định dạng:** Ảnh (.jpg, .png), Video (.mp4, .avi), Luồng thời gian thực từ camera.
- **Cấu trúc thô:** Ảnh được biểu diễn dưới dạng tensor 3D [Chiều cao, Chiều rộng, Kênh màu].
- **Sau tiền xử lý:** Chuyển đổi thành tensor [Batch size, Kênh màu, Chiều cao, Chiều rộng] với giá trị pixel chuẩn hóa [0, 1].

Dữ liệu đầu ra

Kết quả xử lý từ mô hình YOLOv5, thể hiện các đối tượng được nhận diện.

- **Từ mô hình:** Tensor chứa thông tin dự đoán thô.
- **Đã xử lý:** Danh sách các đối tượng nhận diện, mỗi đối tượng gồm:
 - **Khung bao:** [x_min, y_min, x_max, y_max] hoặc [x_center, y_center, width, height].
 - **Điểm tin cậy:** Giá trị 0-1.
 - **Lớp đối tượng:** Tên đối tượng (ví dụ: "xe", "người").
- **Kết quả trực quan:** Ảnh/video với khung bao và nhãn được vẽ trực tiếp lên hình ảnh gốc.

Chương 4: Cài đặt và thử nghiệm

4.1. Môi trường cài đặt & 4.2. Quá trình cài đặt

Môi trường đề xuất

Để triển khai YOLOv5 hiệu quả, hệ thống cần đáp ứng các yêu cầu sau:

- **Hệ điều hành:** Windows 10/11 (khuyến dùng Google Colab nếu không có GPU).
- **Ngôn ngữ:** Python 3.8+ (với pip/conda).
- **Công cụ:** CUDA Toolkit & cuDNN (cho NVIDIA GPU), môi trường ảo (venv/conda env).
- **Thư viện:** PyTorch, Ultralytics, OpenCV, NumPy, Pillow, Matplotlib.

Các bước cài đặt chính



Cài đặt CUDA & PyTorch

Cài đặt driver NVIDIA, CUDA Toolkit và PyTorch hỗ trợ GPU.



Thiết lập Python & VS Code

Cài đặt Python, VS Code, extensions, và tạo môi trường ảo.



Tải YOLOv5 & Thư viện

Clone mã nguồn YOLOv5 và cài đặt các thư viện yêu cầu qua `requirements.txt`.



Huấn luyện & Sử dụng Mô hình

Chuẩn bị dữ liệu, cấu hình `custom_data.yaml`, huấn luyện và kiểm tra mô hình.

4.3. Kết quả thực nghiệm

Sau khi cài đặt YOLOv5, nhóm đã tiến hành thử nghiệm trên các ảnh tĩnh và video MP4 rõ nét, tập trung vào việc phát hiện và đếm đối tượng như xe máy, ô tô.

Phương pháp đếm

YOLOv5 cung cấp khung bao và nhãn. Hệ thống kết hợp YOLOv5 với **DeepSORT** để gán ID và theo dõi từng đối tượng, đảm bảo đếm chính xác mà không trùng lặp.

Thiết kế vùng đếm

Để đếm xe qua, một **đường ảo** được thiết lập trên video. Xe sẽ được đếm khi ID của nó lần đầu tiên đi qua đường này.

Kết quả & Độ chính xác

Độ chính xác phụ thuộc vào góc quay camera, mật độ giao thông và khả năng nhận diện của mô hình. Trong điều kiện lý tưởng, hệ thống đạt độ chính xác trên **95%**.

4.4. Đánh giá và So sánh Hiệu suất

Dựa trên các chỉ số hoạt động, mô hình YOLOv5 thể hiện những ưu điểm rõ rệt trong việc nhận diện và đếm đối tượng.

Hiệu suất vượt trội

YOLOv5 đạt sự cân bằng hoàn hảo giữa tốc độ và độ chính xác, với mAP cao (trên 85%) và tốc độ xử lý nhanh (>20 FPS), lý tưởng cho ứng dụng thời gian thực.

Khả năng thích ứng cao

Mô hình hoạt động ổn định trong nhiều điều kiện khác nhau: ngày/đêm, mưa, hay giao thông đông đúc, nhờ khả năng học mạnh mẽ từ dữ liệu đa dạng.

Hiệu quả trọng số huấn luyện trước

Việc sử dụng trọng số từ mô hình tiền huấn luyện giúp tinh chỉnh nhanh chóng, giảm yêu cầu dữ liệu và đảm bảo độ chính xác cao ngay từ đầu.

So sánh với Haar Cascade

YOLOv5 vượt trội hoàn toàn so với Haar Cascade, vốn lỗi thời và chỉ phù hợp cho các bài toán đơn giản, không xử lý được các tình huống phức tạp như vật thể bị che khuất hoặc thay đổi điều kiện môi trường.

So sánh với YOLOv3

YOLOv5 cải tiến đáng kể với kiến trúc tinh gọn hơn, kỹ thuật tăng cường dữ liệu tiên tiến (Mosaic Augmentation), và các thành phần tối ưu hóa, mang lại độ chính xác cao hơn mà vẫn duy trì hoặc cải thiện tốc độ xử lý.

Chương 5: Kết luận và hướng phát triển

Kết luận

Hệ thống phát hiện và đếm phương tiện giao thông bằng YOLOv5 và DeepSORT đã được xây dựng thành công. Hệ thống có khả năng phân tích video, phát hiện, theo dõi, gán ID, hiển thị bounding box, và thống kê số lượng phương tiện qua vạch đếm ảo. Hoạt động ổn định, tốc độ cao, độ chính xác tốt trong điều kiện phù hợp.

Hạn chế

- Chưa tinh chỉnh mô hình theo tập dữ liệu chuyên biệt.
- Chưa xử lý tình huống phức tạp (che khuất, ánh sáng yếu, tốc độ nhanh).
- Chỉ đếm một chiều, chưa hỗ trợ đếm hai chiều hoặc phân luồng.

Hướng phát triển

- Huấn luyện lại YOLOv5 trên dữ liệu giao thông Việt Nam.
- Tăng cường xử lý ảnh đầu vào (độ sáng, làm rõ).
- Mở rộng thành ứng dụng thời gian thực với camera giám sát.
- Thêm chức năng đếm hai chiều, phân tích hướng di chuyển, mật độ phương tiện.
- Kết hợp phân tích dữ liệu, thống kê lưu lượng theo giờ.

Tài liệu tham khảo

1. T. Dang, "Sử dụng Colab train YOLOv5 với custom dataset phát hiện các đối tượng đặc thù," Viblo, Apr. 25, 2021. [Online]. Available: <https://viblo.asia/p/su-dung-colab-train-yolov5-voi-custom-dataset-phat-hien-cac-doi-tuong-dac-thu-Az45bqv6lxY>
2. Ultralytics, "YOLOv5 Quickstart Tutorial," Ultralytics Documentation. [Online]. Available: https://docs.ultralytics.com/vi/yolov5/quickstart_tutorial/
3. D. Khuong, "Yolov5AnimalCamera - videomp4main.py," GitHub, 2023. [Online]. Available: <https://github.com/KhuongDuy25/Yolov5AnimalCamera/blob/main/yolov5-master/videomp4main.py>
4. aivlogs, "Phát hiện & đếm phương tiện giao thông dùng YOLOv5 + DeepSORT - Python code," YouTube, Jun. 25, 2022. [Online]. Available: <https://www.youtube.com/watch?v=63frZYsAlVk>
5. Ghiền AI, "Hướng dẫn phát hiện & theo dõi đối tượng bằng YOLOv5 + DeepSORT," YouTube, Jul. 10, 2022. [Online]. Available: <https://www.youtube.com/watch?si=kdAKVpcevljFoupd&v=plgWCxj--tM&feature=youtu.be>

Tiến độ hoàn thành công việc

Dự án đã được hoàn thành với 100% tiến độ cho tất cả các hạng mục chính.

Hạng mục	Tiến độ
• SVTH1: Phạm Thế Hùng - 2374802010164	100%
• SVTH2: Phan Trần Quang Thái - 2174802010116	100%
• SVTH3: Đoàn Hoàng Quân - 2374802010412	100%
• SVTH4: Nguyễn Chí Dũng - 2374802013469	100%

Chân thành cảm ơn!

Chúng em xin chân thành cảm ơn quý thầy cô và các bạn đã dành thời gian quý báu để lắng nghe bài thuyết trình của nhóm. Mọi ý kiến đóng góp từ quý vị sẽ là nguồn động lực quý giá, giúp bài nghiên cứu của chúng em được hoàn thiện hơn.