

TRƯỜNG ĐẠI HỌC CÔNG NGHIỆP HÀ NỘI
KHOA CÔNG NGHỆ THÔNG TIN



BÁO CÁO BÀI TẬP LỚN
HỌC PHẦN: ĐỒ ÁN CHUYÊN NGÀNH

ĐỀ TÀI
XÂY DỰNG HỆ THỐNG NHẬN DẠNG BIỂN SỐ XE TỰ ĐỘNG SỬ DỤNG
MÔ HÌNH HỌC SÂU

GVHD:	TS. Nguyễn Mạnh Cường	
Lớp:	20231IT6052002	
Nhóm:	2	
Thành viên:	Nguyễn Văn Vũ	2020601029
	Trần Thị Khánh Linh	2020602060
	Phạm Văn Đức	2020601553

Hà Nội, 2024

MỤC LỤC

CHƯƠNG 1. GIỚI THIỆU TỔNG QUAN.....	1
1.1. Hiện trạng giao thông Việt Nam hiện nay.....	1
1.2. Học sâu và ứng dụng	1
1.2.1. Khái niệm học sâu	1
1.2.2. Ứng dụng của học sâu	2
1.3. Bài toán nhận dạng và ứng dụng của bài toán nhận dạng trong giao thông	3
1.3.1. Đặc điểm của bài toán nhận dạng	3
1.3.2. Ứng dụng của bài toán nhận dạng trong giao thông	3
1.4. Tổng quan về bài toán nhận dạng biển số xe	5
1.4.1. Đặc điểm về bài toán nhận dạng biển số xe	5
1.4.2. Phát biểu bài toán	6
1.4.3. Phạm vi đề tài	7
CHƯƠNG 2. CÁC KIẾN THỨC NGHIÊN CỨU LIÊN QUAN.....	8
2.1. Phương hướng tiếp cận bài toán.....	8
2.2. Một số kĩ thuật trong bài toán phát hiện vật thể.....	8
2.2.1. Phương pháp hai giai đoạn	8
2.2.2. Phương pháp một giai đoạn.....	15
2.3. Đề xuất giải quyết bài toán.....	17
2.4. Các phương pháp đánh giá mô hình.....	25
2.5. Nhận dạng biển số xe	28

CHƯƠNG 3. PHƯƠNG PHÁP GIẢI QUYẾT BÀI TOÁN.....	29
3.1. Lựa chọn mô hình.....	29
3.2. Gán nhãn dữ liệu	30
3.3. Phân loại kí tự.....	31
CHƯƠNG 4. CHƯƠNG TRÌNH THỰC NGHIỆM.....	33
4.1. Chuẩn bị bộ dữ liệu	33
4.2. Huấn luyện mô hình YOLOv8m	34
4.3. Thiết kế phần mềm	35
4.3.1. Thiết bị hệ thống.....	35
4.3.2. Môi trường thực hiện và công cụ	35

DANH MỤC HÌNH ẢNH

Hình 2.2.1. Tổng quan mô hình mạng R-CNN 2014 [4]	9
Hình 2.2.2. Tổng quan mô hình mạng Fast R – CNN [5]	11
Hình 2.2.3. Tổng quan về mô hình mạng Faster R-CNN 2015 [3]	13
Hình 2.2.4. Tổng quan mô hình mạng Mask R-CNN [6]	14
Hình 2.3.1. Kiến trúc mạng YOLOv1 [8]	18
Hình 2.3.2. Kiến trúc mạng YOLOv3 [8][9]	20
Hình 2.3.3. Kiến trúc mạng YOLOv8	24
Hình 2.4.1. Ma trận hỗn loạn (Confusion matrix)	26
Hình 2.4.2. Công thức tính IoU	26
Hình 2.4.3. Công thức tính độ đo precision	27
Hình 2.4.4. Công thức tính độ đo recall	27
Hình 2.4.5. Công thức tính độ đo mAP	27
Hình 3.1.1. Kết quả so sánh với các mô hình khác	29
Hình 3.1.2. So sánh các kích cỡ mô hình YOLOv8	30
Hình 3.1.3. Mô tả quá trình gán nhãn dữ liệu	31
Hình 3.1.4. Hình ảnh biên số xe gốc	32
Hình 3.1.5. Hình ảnh sau khi được chia đôi	32
Hình 4.1.1. Bộ dữ liệu GreenParking	33
Hình 4.2.1. Kết quả huấn luyện mô hình YOLOv8m	34
Hình 4.3.1. Giao diện chương trình	36
Hình 4.3.2. Kết quả khi chạy chương trình	37

LỜI CẢM ƠN

Lời đầu tiên, chúng em xin chân thành cảm ơn các thầy cô là giảng viên các bộ môn trong khoa Công nghệ thông tin – Trường Đại học Công nghiệp Hà Nội đã truyền đạt cho chúng em những kiến thức nền tảng quan trọng liên quan đến học sâu và trí tuệ nhân tạo. Đồng thời, chúng em muốn gửi lời cảm ơn sâu sắc đến giảng viên Tiến sĩ Nguyễn Mạnh Cường người đã trực tiếp giảng dạy và hỗ trợ chúng em trong suốt quá trình lên ý tưởng, nghiên cứu và hoàn thành báo cáo đề tài này. Bên cạnh đó, chúng em muốn cảm ơn các bạn là thành viên của các nhóm thực hiện những đề tài liên quan, đã giúp đỡ nhóm về mặt tinh thần để có thể thực hiện tốt nhiệm vụ cá nhân trong bài báo cáo của tập thể nhóm.

Trong quá trình nghiên cứu thực hiện đề tài, do năng lực, kiến thức, kỹ năng cũng như trình độ của bản thân các thành viên trong nhóm còn hạn hẹp, thiếu chuyên sâu nên không thể tránh khỏi những sai sót liên quan đến kỹ thuật. Vì vậy, chúng em chân thành muốn được lắng nghe những góp ý từ quý thầy cô là giảng viên các bộ môn cũng như các thành viên của các nhóm còn lại thực hiện những đề tài khác. Để từ đó rút kinh nghiệm và hoàn thiện bài báo cáo tốt hơn trong tương lai. Chúng em xin chân thành cảm ơn !

Nhóm thực hiện đề tài !

LỜI NÓI ĐẦU

Công nghệ thông tin đã trở thành một phần quan trọng trong đời sống hiện tại và đã phát triển mạnh mẽ trong thời gian gần đây. Những phần mềm thông minh có khả năng tự động xử lý các công việc khó khăn phục vụ cho lợi ích của con người. Cùng với đó không thể thiếu được các thiết bị ghi hình đang bùng nổ như điện thoại di động, hệ thống camera giám sát, ... Những năm gần đây, có nhiều các bài báo nghiên cứu khoa học về lĩnh vực trí tuệ nhân tạo, học sâu đã đưa ra rất nhiều các phương pháp, cách tiếp cận cho bài toán nhận dạng vật thể, chữ viết, khuôn mặt, giọng nói, ... cho ra được độ chính xác rất cao. Những thành tựu này hoàn toàn có khả năng để ứng dụng vào giải quyết các vấn đề về giao thông và cụ thể là nhận dạng biển số xe trong các bãi đỗ xe.

Trong những năm gần đây, số lượng các phương tiện tham gia giao thông ngày càng tăng do nhu cầu đi lại của người dân đặc biệt là học sinh, sinh viên. Lưu lượng phương tiện giao thông sử dụng ở trường học rất lớn. Hiện tại, sinh viên tham gia học tại trường đại học tương đối đông, nhà trường có bãi gửi xe cho sinh viên đến trường, mô hình quản lý gửi xe tại đây còn mang tính chất thủ công, chưa có sự quản lý chặt chẽ từ ban quản lý đến nhân viên và tự giác của sinh viên. Việc áp dụng công nghệ vào quản lý giao thông đang trở thành một giải pháp quan trọng để giảm ùn tắc giao thông và cải thiện hiệu suất hoạt động của các địa điểm công cộng và bãi đỗ xe. Các mô hình đó có thể thay thế con người làm nhiệm vụ một cách chính xác, nhanh chóng.

Đề tài “Hệ thống nhận dạng biển số xe tự động sử dụng mô hình học sâu” là một đề tài có tính thực tiễn khi có thể giúp cho bộ phận nhà xe dễ dàng hơn trong việc quản lý. Hệ thống có thể hỗ trợ bãi đỗ xe trong việc tự động hoá nhận dạng biển số xe giúp cho dễ dàng trong việc quản lý xe, kiểm soát được lưu lượng xe cộ. Từ đó giúp tiết kiệm thời gian, giảm nguy cơ xảy ra sự cố không đáng có, giảm thiểu chi phí thuê nhân lực quản lý.

Mục tiêu của đề tài tạo ra được hệ thống hỗ trợ bộ phận quản lý nhà xe có thể xuất ra được ký tự của biển số xe. Hệ thống có ứng dụng các kỹ thuật trong lĩnh vực học sâu (Deep learning) và thị giác máy tính (Computer vision) để giải quyết bài toán.

Báo cáo thực nghiệm bao gồm 4 chương như sau:

Chương 1: Khảo sát và tổng quan bài toán

Chương này trình bày tổng quan về hiện trạng giao thông hiện nay, giới thiệu tổng quan về bài toán nhận dạng, ứng dụng vào trong giao thông hiện nay và phát biểu về bài toán nhận dạng biển số xe, bên cạnh đó chỉ ra đầu ra, đầu vào của bài toán, phạm vi đề tài.

Chương 2: Các kiến thức liên quan

Trong chương này trình bày các phương pháp thường được sử dụng đối với bài toán nhận dạng biển số xe, bên cạnh đó trình bày chi tiết về phương pháp sử dụng trong đề tài.

Chương 3: Phương pháp giải quyết bài toán

Nội dung chương này là trình bày kỹ hơn về phương pháp giải quyết bài toán nhận diện đối tượng, cách gán nhãn

Chương 4: Chương trình thực nghiệm

Chương này giới thiệu về các công cụ phần mềm được sử dụng để xây dựng hệ thống, trình bày chi tiết về hệ thống đã làm được.

Qua đề tài này nhóm chúng em rút ra được nhận thức về việc áp dụng các kiến thức, công cụ kỹ thuật khoa học vào giao thông. Nhóm cũng hy vọng rằng báo cáo này không chỉ giúp hiểu rõ hơn về các công nghệ mà còn hiểu biết về các kiến thức liên quan.

CHƯƠNG 1. GIỚI THIỆU TỔNG QUAN

1.1. Hiện trạng giao thông Việt Nam hiện nay

Các trung tâm thành phố lớn là nơi tập trung nhiều cơ quan hành chính, văn hóa, lịch sử và du lịch. Lượng phương tiện tăng quá nhanh trong khi cơ sở hạ tầng chưa đáp ứng kịp, ý thức của người dân khi ra đường và doanh nghiệp vận tải chưa bao giờ cao khiến tình hình giao thông nhiều năm qua chưa được cải thiện.

Tại báo cáo mới nhất gửi Quốc hội về việc đảm bảo trật tự, an toàn giao thông năm 2023 và phương hướng năm 2024, Bộ GTVT cho biết, 9 tháng đầu năm 2023 toàn quốc xảy ra 97 vụ ùn tắc giao thông, tăng 29 vụ so với cùng kỳ năm 2022. Bộ Giao thông vận tải cho biết, tình trạng ùn tắc giao thông tại Hà Nội có xu hướng tăng trên các tuyến cao tốc, quốc lộ trọng điểm, nhất là tại cửa ngõ ra, vào các thành phố lớn. Nguyên nhân do số lượng phương tiện tham gia giao thông tăng cao [1].

Cũng giống như các bãi đỗ xe tòa nhà, chung cư. Thì lưu lượng ra vào tại các trường đại học, cao đẳng, trường học phổ thông thường có mức độ vừa phải và số lượng xe ra vào thường cố định. Tuy nhiên trong các giờ tan tầm hoặc tan học thì vẫn xảy ra tình trạng tắc đường, kẹt xe tại điểm kiểm soát. Do đó nếu áp dụng cách giữ xe thông thường sẽ gây nhiều hỗn loạn và mất nhiều thời gian hơn cho cả người gửi lẫn người kiểm soát.

1.2. Học sâu và ứng dụng

1.2.1. Khái niệm học sâu

Những năm gần đây, Trí tuệ nhân tạo (Artificial Intelligence) là một lĩnh vực khoa học nhằm mục đích đạt được trí thông minh như con người. Mục tiêu của trí tuệ nhân tạo là tạo ra các hệ thống có khả năng tư duy, học hỏi, hiểu và tương tác với môi trường xung quanh. Điều này đòi hỏi sự kết hợp của nhiều phương pháp và kỹ thuật, bao gồm học máy, xử lý ngôn ngữ tự nhiên, học sâu, ... Trong đó Học sâu (Deep learning) là một tập hợp

con của trí tuệ nhân tạo và học máy sử dụng mạng thần kinh nhân tạo nhiều lớp để mang lại độ chính xác cao nhất trong các tác vụ như phát hiện đối tượng, nhận dạng giọng nói, dịch ngôn ngữ và nhiều các vụ khác.

Học sâu (Deep learning) khác so với kỹ thuật học máy truyền thống rằng học sâu có thể tự động học các biểu diễn ngữ nghĩa từ ảnh, video hoặc tệp tài liệu mà không cần phải áp dụng các quy tắc lập trình hoặc kiến thức chuyên môn của con người. Với kỹ thuật này, học sâu đã có nhiều đột phá trong trí tuệ nhân tạo như Google DeepMind, Alpha Go, xe tự hành, trợ lý ảo sử dụng giọng nói thông minh và còn nhiều ứng dụng khác.

1.2.2. Ứng dụng của học sâu

Hiện nay, Deep Learning được ứng dụng rất nhiều vào các khía cạnh và lĩnh vực trong cuộc sống, hỗ trợ cho các quy trình nghiên cứu, giải thích và phân tích khối lượng dữ liệu một cách dễ dàng và nhanh chóng. Việc Deep Learning phát triển tạo nên sự chủ động trong công việc, con người dần có thể điều khiển cuộc sống tự động. Dưới đây là một số ứng dụng Deep Learning đem đến lợi ích cho cuộc sống của con người.

Thị giác máy tính: Học sâu đã thúc đẩy sự phát triển của hệ thống nhận dạng hình ảnh và phân loại, như nhận dạng khuôn mặt, phát hiện đối tượng, và gắn nhãn hình ảnh trong thời gian thực. Các ứng dụng bao gồm xe tự lái, nhận dạng vật thể trong hình ảnh y tế và an ninh.

Xử lý ngôn ngữ tự nhiên (NLP - Nature language processing): Học sâu đã cải thiện hiệu suất trong việc xử lý và hiểu ngôn ngữ tự nhiên, bao gồm dịch máy, xác định ngữ cảnh và tổng hợp văn bản. Ứng dụng của NLP bao gồm trợ lý ảo (như Siri, Alexa), chatbot, và phân tích cảm xúc trên mạng xã hội.

Nhận dạng giọng nói: Học sâu được sử dụng để nhận dạng giọng nói và biến giọng nói thành văn bản. Ứng dụng bao gồm hệ thống nhận dạng giọng nói trên điện thoại di động và hệ thống chuyển đổi giọng thành văn bản trong lĩnh vực y tế và pháp luật.

Tóm lại, học sâu đã có những ứng dụng đa dạng và ảnh hưởng lớn trong nhiều lĩnh vực khác nhau. Sự kết hợp giữa sức mạnh tính toán và khả năng học của học sâu đã mang lại những tiến bộ quan trọng và thúc đẩy sự phát triển của trí tuệ nhân tạo.

1.3. Bài toán nhận dạng và ứng dụng của bài toán nhận dạng trong giao thông

1.3.1. Đặc điểm của bài toán nhận dạng

Trong CNTT, nhận dạng nhấn mạnh vào việc nhận dạng các mẫu dữ liệu hoặc tính quy luật của dữ liệu trong một tình huống nhất định. Nhận dạng mẫu có thể được “giám sát”, trong đó các mẫu đã biết trước đó có thể được tìm thấy trong một dữ liệu nhất định hoặc “không được giám sát”, trong đó các mẫu hoàn toàn mới được phát hiện. Việc nhận dạng là tìm ra quy luật và các thuật toán để có thể gán đối tượng vào một lớp hay nói cách khác là gán cho đối tượng một tên.

1.3.2. Ứng dụng của bài toán nhận dạng trong giao thông

Bài toán nhận dạng trong lĩnh vực giao thông ở Việt Nam đã mở ra một loạt ứng dụng đa dạng. Không chỉ dừng lại ở việc nhận biết biển số xe, bài toán nhận dạng đã được tích hợp vào nhiều khía cạnh của hệ thống giao thông, định hình một tương lai giao thông hiệu quả và tiện lợi. Bằng cách áp dụng công nghệ học sâu và xử lý hình ảnh, bài toán nhận dạng giúp tự động nhận diện các biển báo giao thông trên đường, cung cấp thông tin quan trọng cho người lái xe và hệ thống quản lý giao thông.

Ngoài ra, bài toán nhận dạng cũng được sử dụng để giám sát hành vi lái xe. Hệ thống này có khả năng phân biệt các hành vi vi phạm như việc vượt đèn đỏ, sử dụng điện thoại khi lái xe, hoặc vi phạm tốc độ. Thông qua việc ghi nhận và báo cáo tự động, bài

toán nhận dạng giúp tăng cường tuân thủ luật lệ giao thông và giảm nguy cơ tai nạn giao thông.

Bài toán nhận dạng cũng hỗ trợ trong việc quản lý đỗ xe công cộng. Hệ thống này tự động nhận dạng xe khi đỗ và ghi nhận thời gian, giúp ngăn chặn việc lấn chiếm vỉa hè và đảm bảo trật tự đô thị. Việc này giúp tạo ra không gian đô thị sạch sẽ và an toàn, đồng thời tối ưu hóa việc sử dụng không gian đỗ xe.

Hơn thế nữa, bài toán nhận dạng được tích hợp vào các phương tiện giao thông công cộng như tàu điện ngầm và xe buýt. Bằng cách nhận diện hành khách, hệ thống quản lý giao thông có thể dự đoán lượng hành khách tại các thời điểm khác nhau và điều chỉnh lịch trình vận chuyển, giúp tối ưu hóa việc sử dụng tài nguyên và giảm ão cao điểm giao thông.

Bên cạnh đó, bài toán nhận dạng biển báo giao thông trong giao thông Việt Nam sử dụng công nghệ nhận dạng hình ảnh để tự động nhận diện và phân biệt các biểu hiện trên biển báo. Điều này giúp cảnh báo tài xế về các quy tắc và hướng dẫn giao thông một cách tự động, giảm nguy cơ vi phạm và tăng cường an toàn trên đường. Bằng cách này, hệ thống giúp tối ưu hóa hiệu suất của biển báo và đồng thời cải thiện khả năng quản lý giao thông đô thị, tạo ra một môi trường giao thông thông minh và an toàn hơn.

Bài toán nhận dạng không chỉ giúp nhận biết các yếu tố trong giao thông, mà còn đóng vai trò quan trọng trong việc cải thiện an toàn, hiệu quả và tiện lợi cho người dân khi sử dụng các dịch vụ giao thông. Sự kết hợp giữa bài toán nhận dạng và công nghệ thông tin sẽ tiếp tục định hình một tương lai giao thông thông minh và bền vững cho Việt Nam.

1.4. Tổng quan về bài toán nhận dạng biển số xe

1.4.1. Đặc điểm về bài toán nhận dạng biển số xe

Hệ thống nhận dạng biển số xe tự động, thu hút được sự quan tâm ngày càng tăng do khả năng ứng dụng trong các hệ thống giao thông thông minh đã được lắp đặt ở nhiều quốc gia cho các nhiệm vụ như thực thi luật giao thông và giám sát giao thông. Ngoài ra hệ thống này còn được sử dụng để quản lý lỗi ra vào bãi đỗ xe, thu phí cầu đường và kiểm soát các biện pháp an ninh ở các khu vực hạn như khu cắm trại quân đội và khu bảo tồn. Trước khi có các hệ thống nhận dạng biển số xe hiện đại, thực tế gặp phải nhiều khó khăn như thời gian. Việc ghi lại thông tin thường phụ thuộc vào công việc thủ công của người làm việc giao thông hoặc nhân viên kiểm soát, gây tốn thời gian và có thể sai sót. Do đó các phần mềm được sử dụng để ngăn chặn gian lận, tăng cường bảo mật và giảm thời gian xử lý.

Từ trao đổi trên, nhóm nghiên cứu đã xây dựng hệ thống nhận dạng biển số xe tự động sử dụng mô hình học sâu và có thể triển khai trên các thiết bị khi sử dụng thực tế. Hệ thống sẽ bao gồm việc kết hợp các bài toán thị giác máy tính như nhận diện vật thể (Object detection), và nhận dạng biển số xe tự động (Automatic License Plate Recognition, ALPR). Các bài toán sẽ tập trung vào việc phát hiện các ký tự trong một biển số xe. Hệ thống ALPR lấy hình ảnh hoặc luồng video làm đầu vào cho hệ thống và nếu chứa biển số xe, sẽ xuất ra nội dung của biển số xe dưới dạng văn bản. Hệ thống bao gồm một camera để ghi lại hình ảnh của phương tiện. Các kỹ thuật như phát hiện đối tượng, xử lý ảnh và nhận dạng vật thể được sử dụng để phát hiện và đọc biển số xe. Để giải quyết bài toán nhận dạng biển số xe, nhóm nghiên cứu đã thực hiện các bài toán sau đây:

- Bài toán thứ nhất: Phát hiện (Detect). Dữ liệu đầu vào là ảnh/ video với kích thước cố định tùy thuộc vào cấu trúc mô hình. Nhiệm vụ của bài toán con này là với một

ảnh đầu vào cần trích xuất được tọa độ của vùng biển số xe trong ảnh và cắt ra được vùng biển số đó.

- Bài toán thứ hai: Tiền xử lí ảnh. Dữ liệu đầu vào sẽ là một ảnh hoặc tập ảnh biển số xe đã được tách ra từ bài toán thứ nhất. Nhiệm vụ của bài toán con này là làm sạch hình ảnh biển số xe thu được từ bài toán thứ nhất (khử nhiễu ảnh viền hoặc bụi bẩn) và đưa ra được ảnh biển số xe mới dễ dàng cho việc nhận dạng. Ảnh mới là ảnh nhị phân, có góc nhìn chính diện tới biển số xe, hạn chế các chi tiết thừa và biên của các ký tự trong ảnh phải được làm rõ.
- Bài toán thứ ba: Nhận dạng ký tự (Recognize). Nhiệm vụ của bài toán con này là đưa ra được ký tự tương ứng đúng với các kí tự trong biển số xe thực.

1.4.2. Phát biểu bài toán

Biển số xe là một phần quan trọng của hệ thống giao thông và quản lý phương tiện giao thông. Bài toán nhận dạng biển số xe nhằm vào việc tự động xác định, trích xuất và hiểu thông tin từ biển số xe trên hình ảnh hoặc video. Đây là một bài toán quan trọng trong lĩnh vực thị giác máy tính và xử lý hình ảnh, và có rất nhiều ứng dụng thực tế.

Đầu vào của bài toán:

Gồm dòng video liên tục chứa biển số xe được ghi lại bởi camera gắn ở trước mỗi nhà gửi xe. Dữ liệu này được thu thập với nhiều điều kiện ánh sáng khác nhau, góc chụp cố định, nhiều loại biển số xe.

Đầu ra của bài toán:

Gồm có 2 thông tin chính đó là xác định vị trí và trích xuất thông tin biển số xe từ dòng video liên tục, Có thể có các thông tin bổ sung như: loại xe, ngày tháng,...

Kỳ vọng của bài toán:

Độ chính xác cao, một yêu cầu tối thiểu của bài toán này là đạt được mức độ chính xác cao trong việc xác định biển số xe. Điều này đặc biệt quan trọng trong các tình huống như ánh sáng yếu, biển số bị che khuất bởi vật thể khác, tính ứng dụng rộng rãi, bài toán nhận diện biển số xe có nhiều ứng dụng thực tiễn, từ giám sát giao thông để kiểm soát tốc độ và quản lý luồng xe, đến quản lý bãi đỗ xe, và kiểm soát an ninh trong các khu vực quan trọng như sân bay và trạm cảng.

Cùng với đó là khả năng thực hiện trong thời gian thực, bài toán nhận dạng biển số xe cần phải hoạt động trong thời gian thực để có thể đưa ra cảnh báo hoặc áp dụng biện pháp kiểm soát kịp thời.

1.4.3. Phạm vi đề tài

Nhận dạng biển số xe là một bài toán cụ thể, tồn tại nhiều phương pháp tiếp cận khác nhau, tùy thuộc vào điều kiện hoạt động, trang thiết bị phần cứng mà từng phương pháp phù hợp riêng. Bài toán yêu cầu phải phát hiện được biển số xe và trích xuất được các thông tin trên đó. Nhóm nghiên cứu không xét đến trường hợp ngoại lệ đó vì có một số trường hợp rất khó khăn để phát hiện ngay cả mắt thường cũng không thể phát hiện được. Với lượng kiến thức, kinh nghiệm, thời gian, phần cứng có được nên nhóm sẽ giới hạn bài toán trong khuôn khổ sau:

- Sử dụng Camera giám sát thông dụng làm đầu vào cho việc nhận diện.
- Thực hiện trong không gian hẹp, nhỏ, số lượng 1 xe cho mỗi khung hình, biển số xe không có các vật thể che khuất.
- Camera được cố định.
- Hệ thống có hỗ trợ GPU để chạy trên thời gian thực.

CHƯƠNG 2. CÁC KIẾN THỨC NGHIÊN CỨU LIÊN QUAN

2.1. Phương hướng tiếp cận bài toán

Quá trình giải quyết bài toán bắt đầu bằng việc thu thập và tiền xử lý bộ dữ liệu thực nghiệm liên quan tới biển số xe, đặc biệt tập trung vào các biển số xe Việt Nam. Tiếp đến là nghiên cứu và áp dụng kỹ thuật mạng nơ-ron, học sâu cùng với các kỹ thuật xử lý ảnh để nhận dạng biển số xe thông qua camera.

2.2. Một số kỹ thuật trong bài toán phát hiện vật thể

2.2.1. Phương pháp hai giai đoạn

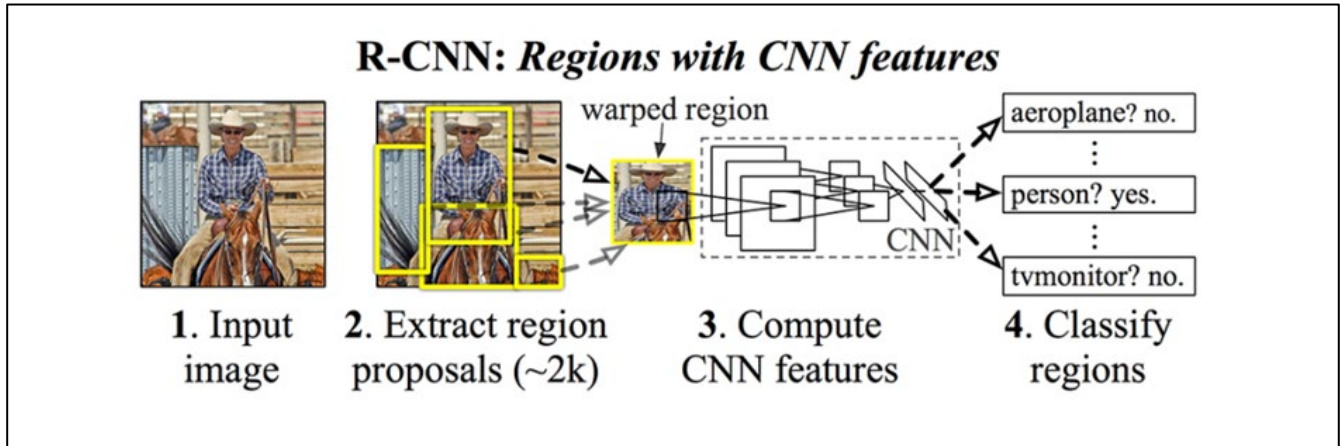
Phương pháp Two-stage detector[2] bao gồm hai công việc chính: giai đoạn định vị (region proposal) và giai đoạn phân loại. Cấu trúc cơ bản của phát hiện hai giai đoạn:

Giai đoạn 1 - Tạo các đề xuất khu vực (Region Proposal)[2]: Trong giai đoạn này, mô hình sử dụng mạng đề xuất khu vực (RPN), như trong Faster R-CNN[3], để tạo ra các vùng đề xuất trên ảnh. RPN đề xuất các khu vực có thể chứa vật thể và cũng dự đoán giá trị tọa độ của bounding box.

Giai đoạn 2 - Phân loại đối tượng cho mỗi khu vực đề xuất (Object Classification): Các vùng đề xuất từ giai đoạn trước sau đó được sử dụng để phân loại đối tượng và định vị chính xác của chúng. Các vùng này thường được trích xuất đặc trưng và đưa qua một mạng phân loại để xác định loại đối tượng và bounding box chính xác.

Mô hình tiêu biểu sử dụng kiến trúc hai giai đoạn:

R – CNN (Region-based Convolutional Neural Network 2014):[4]



Hình 2.2.1. Tổng quan mô hình mạng R-CNN 2014 [4]

Bước 1: Tìm kiếm chọn lọc trên ảnh đầu vào để lựa chọn các vùng đề xuất tiềm năng. Các vùng đề xuất thông thường sẽ có nhiều tỷ lệ với hình dạng và kích thước khác nhau. Hạng mục và khung chứa nhãn gốc sẽ được gán cho từng vùng đề xuất.

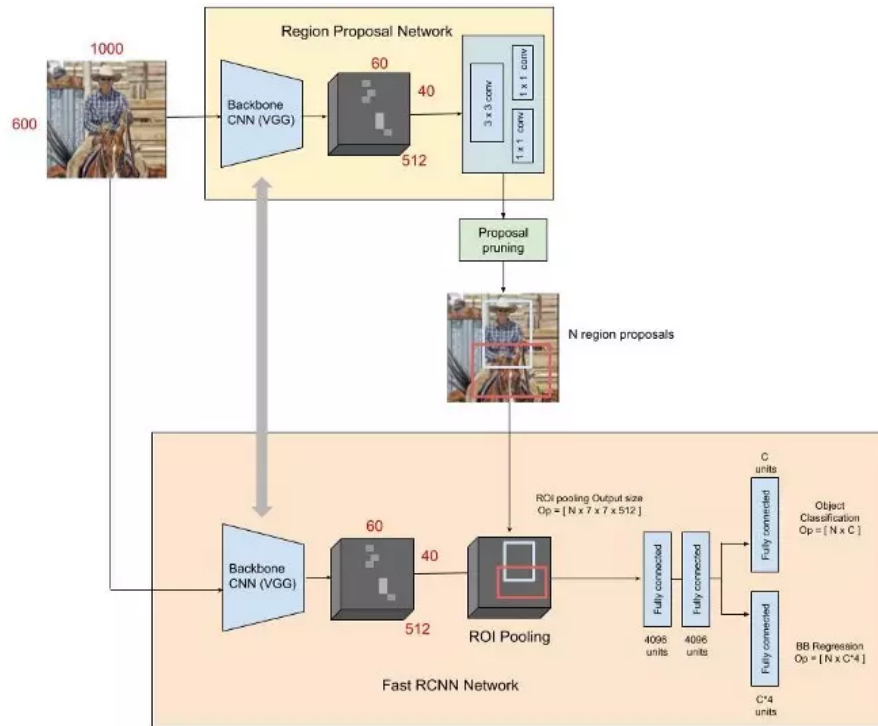
Bước 2: Sử dụng một mạng CNN đã qua tiền huấn luyện, ở dạng rút gọn, đặt trước tầng đầu ra. Mạng này biến đổi từng vùng đề xuất thành các đầu vào có chiều phù hợp với mạng và thực hiện các lượt truyền xuôi để trích xuất đặc trưng từ các vùng đề xuất tương ứng.

Bước 3: Các đặc trưng và nhãn hạng mục của từng vùng đề xuất được kết hợp thành một mẫu để huấn luyện các máy vector hỗ trợ cho phép phân loại vật thể. Ở đây, mỗi máy vector hỗ trợ được sử dụng để xác định một mẫu có thuộc về một hạng mục nào đó hay không.

Bước 4: Các đặc trưng và khung chứa được gán nhãn của mỗi vùng đề xuất được kết hợp thành một mẫu để huấn luyện mô hình hồi quy tuyến tính, để phục vụ dự đoán khung chứa nhãn gốc.

Mặc dù các mô hình R-CNN sử dụng các mạng CNN đã được tiền huấn luyện để trích xuất các đặc trưng ảnh một cách hiệu quả, điểm hạn chế chính yếu đó là tốc độ chậm. Có thể hình dung, với hàng ngàn vùng đề xuất từ một ảnh, ta cần tới hàng ngàn phép tính truyền xuôi từ mạng CNN để phát hiện vật thể. Khối lượng tính toán nặng nề khiến các mô hình R-CNN không được sử dụng rộng rãi trong các ứng dụng thực tế

Fast R-CNN [5]:



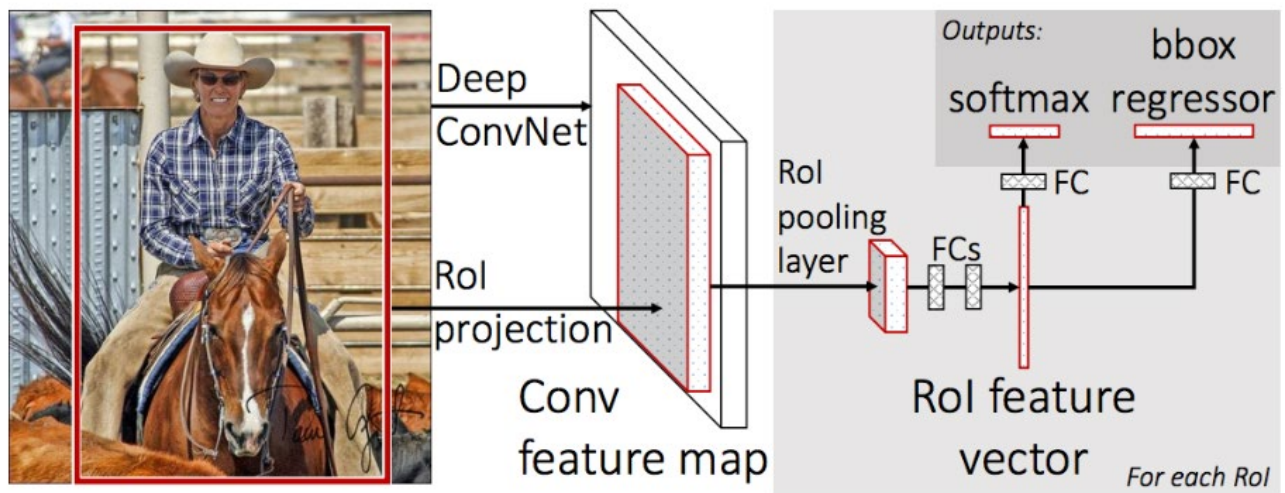
Hình 2.2.2. Tổng quan mô hình mạng Fast R – CNN [5]

Là một phiên bản cải tiến của R-CNN. Fast R-CNN kết hợp quá trình tạo đề xuất khu vực (region proposals) và quá trình phân loại thành một quy trình duy nhất, giúp tăng tốc quá trình huấn luyện so với R-CNN (2015). So với mạng R-CNN, mạng Fast R-CNN sử dụng toàn bộ ảnh làm đầu vào cho CNN để trích xuất đặc trưng thay vì từng vùng đề xuất. Hơn nữa, mạng này được huấn luyện như bình thường để cập nhật tham số mô hình. Do đầu vào là toàn bộ ảnh, đầu ra của mạng CNN có kích thước $1 \times c \times h1 \times w1$.

Giả sử thuật toán tìm kiếm chọn lọc chọn ra n vùng đề xuất, kích thước khác nhau của các vùng này chỉ ra rằng vùng quan tâm (regions of interests - RoI) tại đầu ra của CNN có kích thước khác nhau. Các đặc trưng có cùng kích thước phải được trích xuất từ các vùng quan tâm này (giả sử có chiều cao là $h2$ và chiều rộng là $w2$). Mạng Fast R-

CNN đề xuất phép gộp RoI (RoI pooling), nhận đầu ra từ CNN và các vùng quan tâm làm đầu vào rồi ghép nối các đặc trưng được trích xuất từ mỗi vùng quan tâm làm đầu ra có kích thước $n \times c \times h_2 \times w_2$.

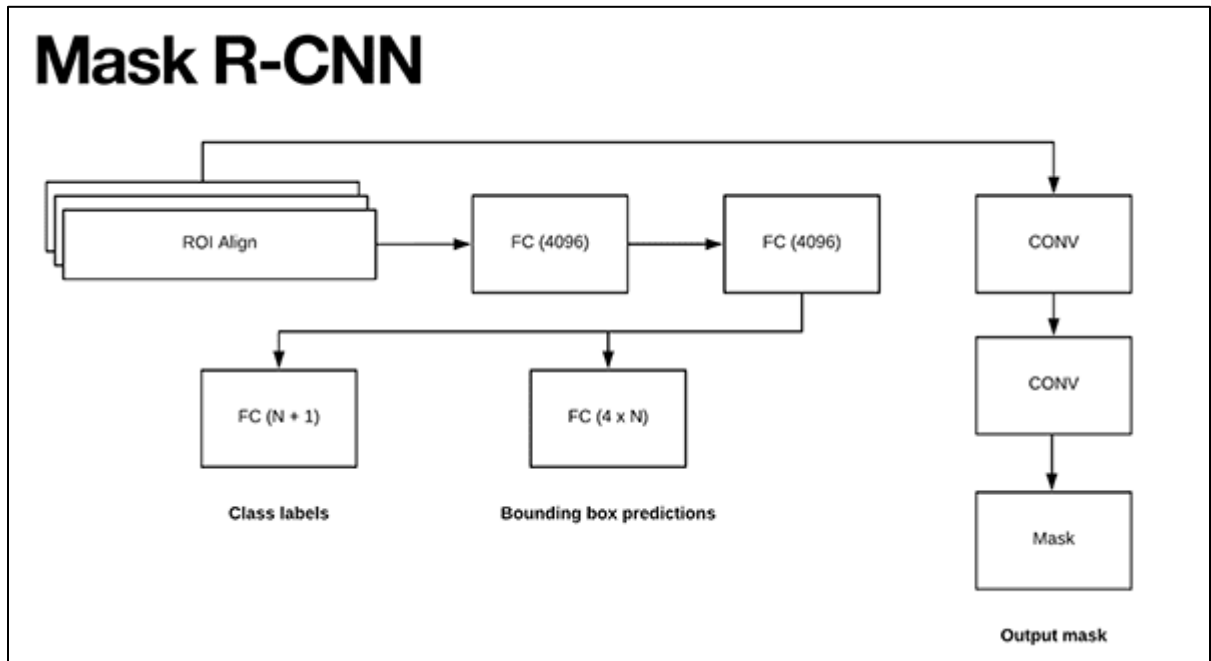
Tầng kết nối đầy đủ được sử dụng để biến đổi kích thước đầu ra thành $n \times d$, trong đó d được xác định khi thiết kế mô hình. Khi dự đoán hạng mục, kích thước đầu ra của tầng kết nối đầy đủ lại được biến đổi thành $n \times q$ và áp dụng phép hồi quy softmax (q là số lượng hạng mục). Khi dự đoán khung chứa, kích thước đầu ra của tầng đầy đủ lại được biến đổi thành $n \times 4$.

Faster R-CNN (2015) [3]:

Hình 2.2.3. Tổng quan về mô hình mạng Faster R-CNN 2015 [3]

Được thiết kế để cải thiện tốc độ và hiệu suất so với các mô hình trước đó như R-CNN và Fast R-CNN. Một trong những cải tiến quan trọng nhất của Faster R-CNN là sử dụng một mạng đặc biệt gọi là Region Proposal Network (RPN) để tạo ra các đề xuất khu vực (region proposals) một cách tự động. RPN được tích hợp vào mô hình, giúp loại bỏ quá trình tạo đề xuất trước đó và làm tăng tốc quá trình. Điều này làm giảm số lượng vùng đề xuất tạo ra, nhưng vẫn đảm bảo độ chính xác khi phát hiện đối tượng.

Mask R-CNN [6]



Hình 2.2.4. Tổng quan mô hình mạng Mask R-CNN [6]

Là một phiên bản mở rộng của Faster R-CNN, Mask R-CNN có cấu trúc cơ bản giống Faster R-CNN, nhưng có thêm một mạng nơ-ron tích chập đầy đủ giúp định vị đối tượng ở cấp điểm ảnh và cải thiện hơn nữa độ chính xác của việc phát hiện đối tượng.

Tóm lại Sự linh hoạt và chi phí tính toán của mô hình hai giai đoạn: Mô hình two-stage được mô tả là linh hoạt, phù hợp hơn cho phân loại theo vùng quan tâm, đảm bảo độ chính xác cao hơn so với mô hình một giai đoạn. Tuy nhiên, điều này đi kèm với chi phí tính toán lớn và tốc độ chậm, đặc biệt là khi so sánh với các mô hình Single Stage

2.2.2. Phương pháp một giai đoạn

Phương pháp Single-stage [7] trong bài toán phát hiện vật thể đặc trưng bởi việc thực hiện cả quá trình trích xuất đặc trưng và dự đoán bounding boxes (hộp giới hạn) và xác suất phân loại trong một lần chạy của mô hình. Nhờ vậy, tốc độ phát hiện đối tượng nhanh hơn nhưng độ chính xác thường thấp hơn so với phương pháp hai giai đoạn.

- Mô hình YOLO (You Only Look Once): Chia ảnh thành một lưới và dự đoán bounding boxes và xác suất phân loại trực tiếp tại các ô lưới. Mô hình này thực hiện dự đoán một lần duy nhất cho toàn bộ ảnh
- Mô hình SSD (Single Shot Multibox Detector): SSD cũng chia ảnh thành một lưới, nhưng nó sử dụng các lớp chuyên tiếp để dự đoán bounding box ở nhiều tỷ lệ và kích thước khác nhau.
- RetinaNet: Mô hình này sử dụng anchor boxes và một phương pháp gọi là Focal Loss để giải quyết vấn đề của class imbalance trong quá trình huấn luyện.

Ưu điểm của phương pháp Single-stage trong bài toán phát hiện vật thể:

- Tốc độ xử lý nhanh: Mô hình Single-Stage thường có thể đưa ra dự đoán nhanh chóng vì nó thực hiện cả hai bước quan trọng - định vị và phân loại - trong một lần chạy. Điều này là quan trọng đặc biệt khi áp dụng cho bài toán nhận dạng biển số xe trong thời gian thực trên hệ thống giám sát giao thông hoặc trong các ứng dụng an ninh.
- Đơn giản và linh hoạt: Mô hình Single-Stage thường đơn giản hóa quá trình nhận dạng bằng cách giảm độ phức tạp của kiến trúc. Điều này không chỉ làm cho chúng dễ triển khai mà còn tăng khả năng linh hoạt, giúp chúng hoạt động tốt trên nhiều điều kiện ánh sáng và môi trường khác nhau.

- Tính nhẹ và phù hợp cho thiết bị di động: Mô hình Single-Stage thường có số lượng tham số ít hơn so với mô hình Two-Stage, làm cho chúng nhẹ hơn và dễ triển khai trên thiết bị di động. Điều này quan trọng đối với các ứng dụng di động, như ứng dụng di động về giao thông và an ninh.
- Khả năng xử lý với biến động và chồng chéo: Do đặc tính của mình, mô hình Single-Stage có thể xử lý tốt với sự biến động và vấn đề chồng chéo trong các kịch bản thực tế, nơi nhiều vật thể có thể xuất hiện và chồng lên nhau.
- Hiệu suất đối với biến động: Các mô hình Single-Stage có thể phát hiện biên số xe hiệu quả trong các tình huống đa dạng và biến động, giúp chúng trở thành lựa chọn phổ biến trong các ứng dụng thực tế như giám sát giao thông và an ninh.

Nhược điểm của phương pháp một giai đoạn trong bài toán nhận dạng biên số xe

- Đối với biên số xe nhỏ: Phương pháp này sẽ gặp khó khăn khi định vị chính xác biên số xe nhỏ, đặc biệt khi áp dụng anchor boxes (anchor boxes được sử dụng để dự đoán bounding box và phân loại các vật thể trong ảnh) cố định. Nếu mô hình được huấn luyện trên ảnh có kích thước và tỷ lệ khác với ảnh thực tế, có thể dẫn đến giảm độ chính xác vì các anchor box không phù hợp với biên số xe trong bức ảnh thực tế.
- Biên số xe bị biến dạng hoặc hình dạng không cân đối: phương pháp Single-stage sẽ không phù hợp với các dạng biên số xe này. Nếu dữ liệu huấn luyện không cân đối, mô hình có thể không học được đủ thông tin để nhận diện các hình dạng đặc biệt này.
- Nhiều biên số xe trong một khung hình: Giảm hiệu suất phát hiện biên số xe do phương pháp này cần phải xử lý cả định vị và phân loại trong một lần.
- Vấn đề chồng chéo (overlap): Các bounding box chồng lên nhau hoặc gần nhau đến mức gây khó khăn cho mô hình phát hiện vật thể định rõ vị trí và kích thước của

từng vật thể. Các bounding box chồng lên nhau cũng làm tăng khả năng lẫn lộn giữa các vật thể, đặc biệt là khi mô hình phải phân loại các vật thể chồng lên nhau. Điều này có thể dẫn đến việc phân loại không chính xác hoặc mô hình không biết phân loại nào được áp dụng cho vật thể nào.

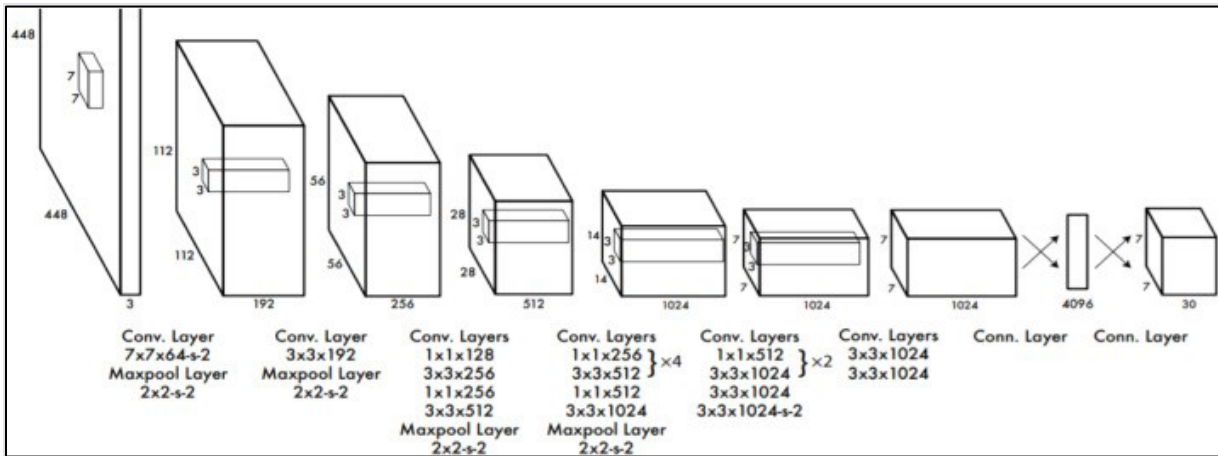
- Ảnh thiếu sáng hoặc nền có nhiều chi tiết: Khi ánh sáng yếu, vấn đề nhận dạng vật thể sẽ trở nên khó khăn hơn, mô hình có thể mất đi thông tin đặc trưng quan trọng của vật thể dẫn đến việc tổng hợp thông tin và thực hiện dự đoán không được chính xác.

2.3. Đề xuất giải quyết bài toán

Mạng Yolo[8] (You Only Look Once) được ra đời nhằm cải thiện độ chính xác của các mạng CNN đang được áp dụng hiện nay cũng như khả năng hoạt động với các bài toán nhận diện yêu cầu thời gian thực. Yolo hướng tới khả năng mạng Neural có thể được huấn luyện cũng như thực thi các tính toán thời gian thực trên các nền tảng GPU thông thường. Mạng Yolo là một thuật toán phát hiện đối tượng một giai đoạn, sử dụng một mạng CNN duy nhất để thực hiện xử lý hình ảnh và có thể trực tiếp tính toán kết quả phân loại và tọa độ vị trí của các đối tượng trong khung hình được đưa vào tính toán. Với việc sử dụng định vị và phân loại đối tượng end-to-end đã làm tăng tốc độ tính toán lên đáng kể qua các phiên bản.

Tóm lại, YOLO không phải là thuật toán tốt nhất nhưng sẽ là thuật toán nhanh nhất để đạt tốc độ gần như realtime. Nhờ việc xuất ra được vị trí của vật thể trong từng bức ảnh nên đây được coi là phương pháp tốt nhất để phát hiện ra nhiều vật thể trong cùng bức ảnh. Đầu vào của bài toán phát hiện vật là một bức ảnh. Chúng ta không chỉ phải phân loại được object trên bức ảnh mà còn phải định vị được vị trí của đối tượng đó. Đầu ra sẽ là các đối tượng, các vật thể kèm theo vị trí của các đối tượng, vật thể có trong bức ảnh đó.

Mô hình Yolo đầu tiên được giới thiệu bởi Joseph Redmon và tất cả trong bài báo năm 2015 với tiêu đề “You Only Look Once: Unified, Real-Time Object Detection” [3]. YOLO được tạo ra với mục tiêu loại bỏ thuật toán nhiều giai đoạn và thực hiện phát hiện đối tượng chỉ trong một giai đoạn duy nhất, do đó cải thiện thời gian suy luận. YOLOv1 là phiên bản đầu tiên của YOLO. Nó là một trong những phương pháp đầu tiên sử dụng mạng neural để nhận diện vật thể. YOLOv1 có tốc độ xử lý nhanh và độ chính xác tương đối cao, nhưng vẫn còn nhiều hạn chế về khả năng phát hiện các vật thể nhỏ hoặc gần nhau.



Hình 2.3.1. Kiến trúc mạng YOLOv1 [8]

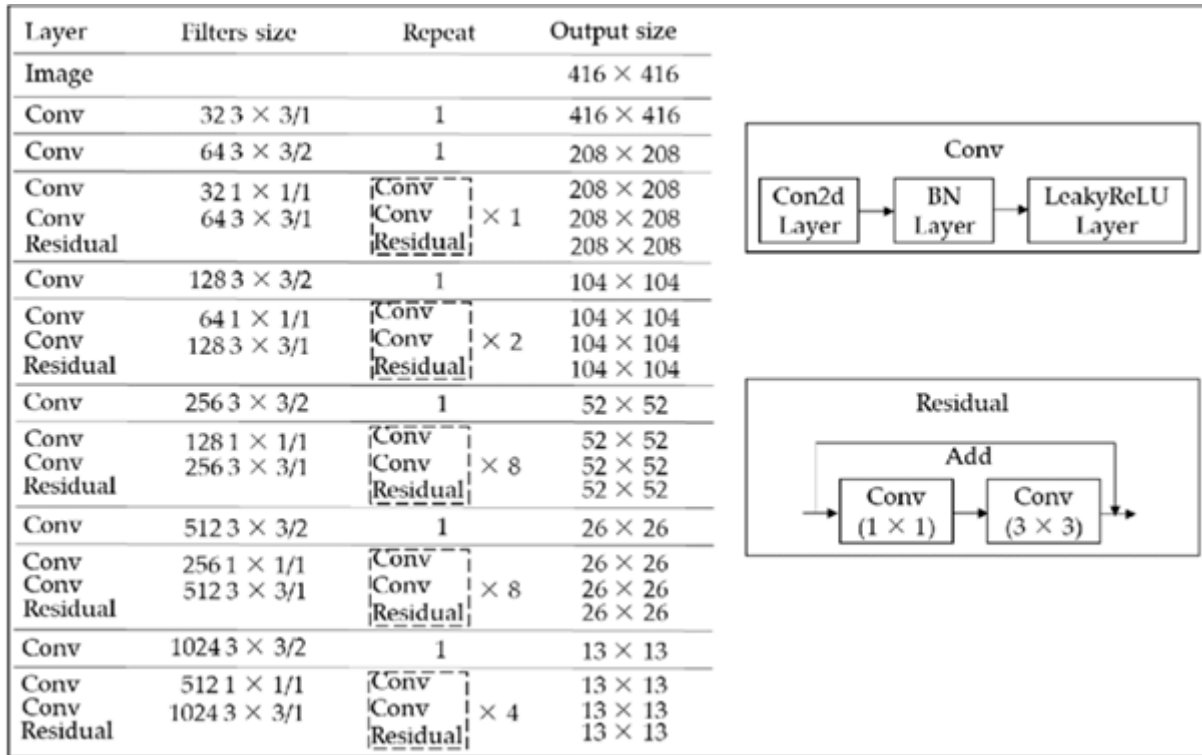
- YOLOv1 chia hình ảnh đầu vào thành một lưới ô vuông và mỗi ô sẽ dự đoán các đối tượng có khả năng xuất hiện trong ô đó.
- Sử dụng phân lớp đa nhãn (multi-label classification): YOLOv1 có thể dự đoán nhiều đối tượng khác nhau trong cùng một ô.
- Sử dụng hàm mất mát tổng quát (generalized IoU loss): YOLOv1 sử dụng hàm mất mát tổng quát để đánh giá độ chính xác của mô hình.

- Sử dụng kết hợp giữa phân lớp và dự đoán vị trí (objectness score): YOLOv1 sử dụng một điểm số (objectness score) để đánh giá khả năng xuất hiện của đối tượng trong ô.

YOLOv2 [8] được phát hành bởi Joseph Redmon và Ali Farhadi vào năm 2016 trong bài báo của họ có tiêu đề “YOLO9000: Better, Faster, Stronger” [10]. 9000 biểu thị rằng YOLOv2 có thể phát hiện hơn 9000 loại đối tượng. YOLOv2 có tốc độ xử lý nhanh hơn và độ chính xác cao hơn so với YOLOv1. Nó cũng có khả năng phát hiện các vật thể nhỏ hơn và gần nhau hơn.

- YOLOv2 sử dụng kỹ thuật batch normalization để giảm độ lệch và tăng độ chính xác của mô hình.
- YOLOv2 sử dụng một mạng nơ-ron tích chập (convolutional neural network) có cấu trúc mới để tăng tốc độ xử lý.
- YOLOv2 sử dụng kỹ thuật anchor boxes để tăng độ chính xác của việc dự đoán vị trí đối tượng.
- YOLOv2 sử dụng kỹ thuật softmax để tăng độ chính xác của việc phân loại đối tượng.

Vào năm 2018, Joseph Redmon và Ali Farhadi đã giới thiệu phiên bản thứ ba của YOLOv3 trong bài viết “YOLOv3: An Incremental Improvement” [9]. Một trình trích xuất đặc trưng tốt hơn đã được giới thiệu cùng với việc 53 lớp tích chập được đào tạo trên ImageNet. Độ chính xác của YOLOv3 tốt hơn YOLOv2 nhưng chậm hơn YOLO v2 do có nhiều lớp hơn.



Hình 2.3.2. Kiến trúc mạng YOLOv3 [8][9]

- Backbone network được cải tiến để tăng độ chính xác và tốc độ xử lý bằng mô hình Darknet-53.
- Sử dụng kỹ thuật skip connection để kết hợp thông tin từ các tầng khác nhau của mạng neural, giúp cải thiện khả năng phát hiện các vật thể nhỏ và gần nhau.
- Sử dụng kỹ thuật kết hợp đa tầng để tăng độ chính xác của việc phân loại lớp vật thể.
- YOLOv3 còn hỗ trợ phát hiện các vật thể trong các tình huống khó như các vật thể bị che khuất hoặc có độ lệch góc lớn.

YOLOv4 không phải do Joseph Redmon phát hành mà bởi Alexey Bochkovskiy, v.v. trong bài báo năm 2020 của họ “YOLOv4: Optimal Speed and Accuracy of Object Detection”[11]. YOLOv4 so với các phiên bản trước đó, có nhiều cải tiến đáng kể như:

- Sử dụng mô hình CSPDarknet sử dụng kỹ thuật kết hợp các lớp convolution giảm số lượng tham số và tăng khả năng học tập, mô hình này giúp YOLOv4 giảm hiện tượng overfitting, tăng tốc độ xử lý và độ chính xác cao hơn.
- Sử dụng kỹ thuật SPP (Spatial Pyramid Pooling) để giảm thiểu số lượng tính toán và tăng độ chính xác.
- Sử dụng kỹ thuật PAN (Path Aggregation Network) để kết hợp thông tin từ các tầng khác nhau của mạng neural, giúp cải thiện khả năng phát hiện các vật thể nhỏ và gần nhau.
- Sử dụng kỹ thuật YOLOv4 Neck để tăng độ chính xác của việc phân loại lớp vật thể.
- Hỗ trợ phát hiện các vật thể trong các tình huống khó như các vật thể bị che khuất hoặc có độ lệch góc lớn.

YOLOv5 được cho là thành viên tiếp theo của gia đình YOLO được phát hành vào năm 2020 bởi công ty Ultralytics chỉ vài ngày sau YOLOv4. YOLOv5 có nhiều điểm khác biệt so với phiên bản trước đó là YOLOv4:

- Tốc độ xử lý và độ chính xác cao hơn
- Sử dụng mô hình EfficientNet để tăng tốc độ xử lý và độ chính xác, trong khi YOLOv4 sử dụng mô hình CSPDarknet.
- YOLOv5 sử dụng kỹ thuật BiFPN (Bidirectional Feature Pyramid Network) để kết hợp thông tin từ các tầng khác nhau của mạng neural, giúp cải thiện khả năng phát hiện các vật thể nhỏ và gần nhau, trong khi YOLOv4 sử dụng kỹ thuật PAN.
- YOLOv5 có khả năng phát hiện các vật thể nhỏ và gần nhau tốt hơn so với YOLOv4.
- YOLOv5 có tính năng tự động tinh chỉnh các siêu tham số để tối ưu hóa độ chính xác và tốc độ xử lý.

Trong năm 2022, Chuyi Li, Lulu Li, Hongliang Jiang, Kaiheng Weng, Yifei Geng, Liang Li, Zaidan Ke, Qingyuan Li, Meng Cheng, Weiqiang Nie, Yiduo Li, Bo Zhang, Yufei Liang, Linyuan Zhou, Xiaoming Xu, Xiangxiang Chu, Xiaoming Wei, Xiaolin Wei đã phát hành YOLOv6 trong bài báo “YOLOv6: A Single-Stage Object Detection Framework for Industrial Applications”[12]. YOLOv6 là phiên bản cũng được đánh giá có bước tiến lớn, so với YOLOv5 thì nó vượt trội hơn hẳn ở các mô hình và kỹ thuật cải tiến dựa trên ý tưởng thiết kế mạng thần kinh nhận biết phân cứng:

- Mô hình không máy dò neo có khả năng khái quát hóa mạnh mẽ và logic mã đơn giản cải thiện hơn về tốc độ
- Chính sách gán thẻ SimOTA tự động phân bổ cải thiện độ chính xác.
- Hộp hồi quy giới hạn mất mát SIOU giám sát quá trình học của mạng cải thiện độ chính xác hồi quy.
- YOLOv6 có thể phát hiện các vật thể nhỏ hơn một cách đáng tin cậy hơn, nhưng nó lại nhạy cảm với nhiễu so với YOLOv5 và gặp khó khăn với các vật thể ở gần. So với YOLOv5, MT-YOLOv6 thiếu tính ổn định nhưng bù lại khả năng phát hiện đối tượng nhỏ trong môi trường đông đúc rất ấn tượng. Về tính linh hoạt, YOLOv5 sử dụng YAML và YOLOv6 xác định các tham số mô hình trực tiếp bằng Python. Người ta nhận thấy rằng YOLOv5 có thể tùy chỉnh nhiều hơn YOLOv6.

Cùng năm, YOLOv7[8] được phát hành bởi Chien-Yao Wang, Alexey Bochkovskiy, Hong-Yuan Mark Liao với bài báo “YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors” [13]. YOLOv7 là mô hình cơ bản được tối ưu hóa cho điện toán GPU thông thường, cũng có cải thiện về tốc độ và độ chính xác bằng việc cải cách một số kiến trúc. Tương tự như YOLOv4 được chia tỷ lệ, các xương sống YOLOv7 không sử dụng các xương sống được đào tạo trước ImageNet. Thay vào đó, các mô hình được đào tạo hoàn toàn bằng bộ dữ liệu COCO.

- Cải cách mạng E-ELAN (Mạng tổng hợp lớp hiệu quả mở rộng), nó là khối tính toán lấy cảm hứng từ nghiên cứu trước đây về hiệu quả mạng và đã được thiết kế bằng cách phân tích các yếu tố ảnh hưởng đến tốc độ, độ chính xác về chi phí truy cập bộ nhớ, tỷ lệ đầu vào và đầu ra, yếu tố hoạt động khôn ngoan, khả năng kích hoạt, đường đi đạo hàm.
- Mở rộng các mô hình dựa trên phép tìm kiếm kiến trúc mạng NAS (Network Architecture Search) có yếu tố tỉ lệ và độc lập.
- Trainable BoF (Bag of Freebies) là một kỹ thuật giám sát sâu có khuôn khổ đa đầu, có thể sử dụng huấn luyện để tích hợp tái tham số hóa theo kế hoạch và huấn luyện thô cho việc phụ trợ trong mô hình học tập và phạt nếu làm mất đầu chỉ- đầu đầu tiên trong dãy các đầu dự đoán. Bổ trợ công cụ gán nhãn mềm thay vì nhãn cứng phù hợp cho mức độ học tập mạnh của mô hình.

Vào ngày 10 tháng 1 năm 2023, Ultralytics tiếp tục cho ra mắt phiên bản mới nhất của YOLO là YOLO8 khẳng định những tiến bộ về cấu trúc và thay đổi kiến trúc với kết quả tốt hơn. YOLOv8 chưa có bài báo nào được xuất bản, vì vậy chúng tôi đã tham khảo thông tin có sẵn và ghi lại những điểm mới trong YOLOv8.

2.4. Các phương pháp đánh giá mô hình

Độ chính xác trung bình trung bình (mAP) là chỉ số được sử dụng để đánh giá các mô hình phát hiện đối tượng như Fast R-CNN, YOLO, Mask R-CNN, v.v. Giá trị trung bình của độ chính xác trung bình (AP) được tính dựa trên các giá trị recall từ 0 đến 1.

Công thức mAP dựa trên các chỉ số phụ sau:

- Ma trận hỗn loạn (Confusion Matrix)
- Giao trên hợp (IoU)
- Recall
- Độ chính xác (Precision)


Có bốn thuộc tính để tạo ma trận hỗn loạn:

- Dương tính thật (TP): Mô hình đã dự đoán nhãn và khớp với nhãn thực tế.
- Âm tính thật (TN): Mô hình không dự đoán nhãn và không phải là một phần của nhãn thực tế.
- Dương tính giả (FP): Mô hình đã dự đoán một nhãn, nhưng nó không phải là một phần của nhãn thực tế (Lỗi Loại I).
- Âm tính giả (FN): Mô hình không dự đoán nhãn, nhưng nó là một phần của nhãn thực tế (Lỗi Loại II).

		Actual Classes	
		POSITIVE	NEGATIVE
Predicted Classes	POSITIVE	TRUE POSITIVE (TP)	FALSE POSITIVE (FP)
	NEGATIVE	FALSE NEGATIVE (FN)	TRUE NEGATIVE (TN)

Hình 2.4.1. Ma trận hỗn loạn (Confusion matrix)

Giao trên hợp (IOU): là tỉ lệ giữa đo lường mức độ giao nhau giữa hai đường bao (thường là đường bao dự đoán và đường bao thực). Tỷ lệ này được tính dựa trên phần diện tích giao nhau giữa 2 đường bao với phần tổng diện tích giao nhau và không giao nhau giữa chúng.

$$\text{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}}$$


Hình 2.4.2. Công thức tính IoU

Tiêu chí theo ngưỡng tự đặt:

- IOU \geq ngưỡng: TP - True Positive
- IOU $<$ ngưỡng: FP - False positive
- Đối tượng không được nhận dạng: FN - False negative

Độ chính xác (Precision): đo lường mức độ có thể tìm thấy các kết quả dương tính thật (TP) trong số tất cả các dự đoán (TP+FP).

$$Precision = \frac{TP}{TP + FP}$$

Hình 2.4.3. Công thức tính độ đo precision

Recall: đo lường mức độ có thể tìm thấy các kết quả dương tính thật (TP) trong số tất cả các dự đoán (TP+FN).

$$Recall = \frac{TP}{TP + FN}$$

Hình 2.4.4. Công thức tính độ đo recall

Cách tính mAP: tìm Độ chính xác Trung bình (AP) cho mỗi lớp và sau đó lấy trung bình trên một số lớp.

$$mAP = \frac{1}{N} \sum_{i=1}^N AP_i$$

Hình 2.4.5. Công thức tính độ đo mAP

Chỉ số mAP kết hợp sự đánh đổi giữa độ chính xác và recall, đồng thời xem xét cả kết quả dương tính giả (FP) và âm tính giả (FN). Thuộc tính này làm cho mAP trở thành thước đo phù hợp cho hầu hết các ứng dụng phát hiện.

2.5. Nhận dạng biển số xe

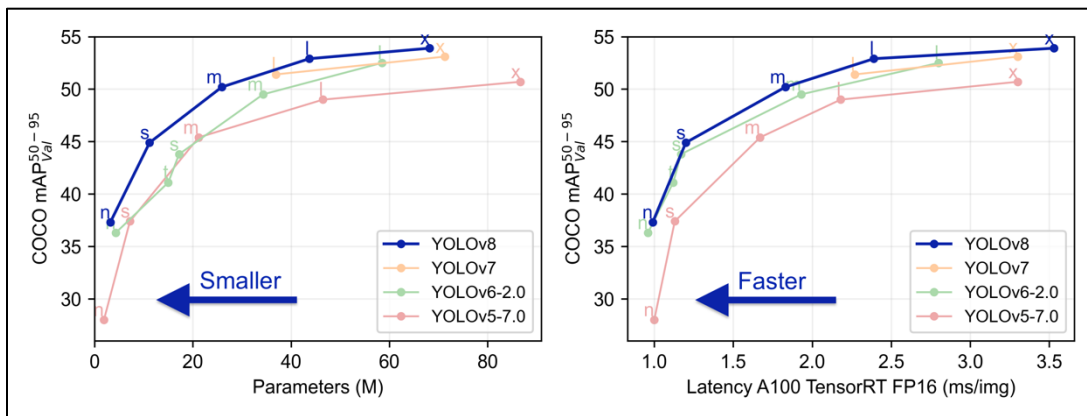
Bài toán nhận dạng biển số xe là một bài toán điển hình dựa trên các phương pháp xử lý ảnh và cả những kỹ thuật sử dụng Deep Learning, trong đó tiêu biểu là bài toán nhận diện ký tự quang học (Optical Character Recognition). Bài toán nhận dạng biển số xe sẽ tập trung vào việc phát hiện các chữ số trong một biển số xe gắn máy. Bản chất của mô hình này là một bài toán phát hiện các ký tự bao gồm các chữ cái viết hoa, viết thường, chữ số, và các ký tự đặc biệt có trên biển số xe máy. Ban đầu mô hình thu nhận biển số xe, trích xuất từng ký tự trong biển số, sau quá trình tính toán, mô hình sẽ đưa ra dự đoán xem các ký tự đó là gì, ghép chúng lại và đưa ra dự đoán hoàn chỉnh biển số xe đó.

Trong báo cáo này nhóm đã áp dụng mô hình mạng yolov5 của tác giả Lin Wang trong dự án “Chinese_license_plate_detection_recognition”[14]. Định dạng lại biển số xe bằng cách ghép các ký tự cùng hàng với nhau.

CHƯƠNG 3. PHƯƠNG PHÁP GIẢI QUYẾT BÀI TOÁN

3.1. Lựa chọn mô hình

Cách tiếp cận của nhóm nghiên cứu là sử dụng mô hình YOLOv8 để giải quyết bài toán nhận diện biển vật thể (object detection). Một số ưu điểm của mô hình này đó là có độ chính xác lớn hơn, việc triển khai mô hình vào các bài toán dễ dàng, huấn luyện mô hình nhanh hơn bất kì mô hình nhận diện vật thể hai giai đoạn khác. Tuy nhiên mô hình này không hỗ trợ việc huấn luyện với hình ảnh có độ phân giải cao, và mô hình này hoạt động tốt ở những vật có thể nhìn thấy trong tầm 10 mét rất phù hợp cho bài toán nhận diện biển số xe của nhóm. Ultralytics đã so sánh YOLOv8 với các phiên bản trước dựa trên bộ dữ liệu COCO và đạt được kết quả ấn tượng trên tất cả các kích cỡ của mô hình.



Hình 3.1.1. Kết quả so sánh với các mô hình khác

Đối với so sánh phát hiện đối tượng của 5 kích thước mô hình, mô hình YOLOv8m đạt được mAP là 50,2% trên bộ dữ liệu COCO, trong khi mô hình lớn nhất, YOLOv8x đạt được 53,9% với số lượng tham số nhiều hơn gấp đôi. Bên cạnh đó, mô hình YOLOv8n có kích thước nhỏ nhất với chỉ 3,2 triệu tham số và có độ chính xác tương đối mAP là 37,3 % trên bộ dữ liệu. Hình dưới đây thể hiện sự so sánh giữa 5 kích thước mô hình khác nhau:

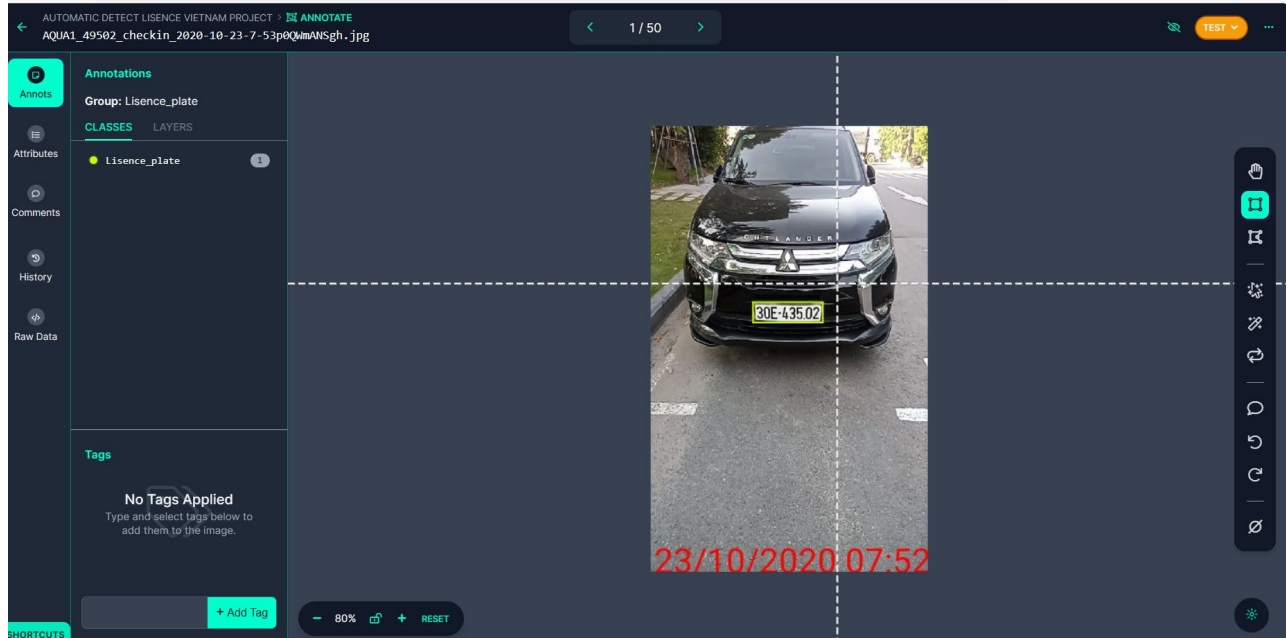
Model	size (pixels)	mAP^{val} 50-95	Speed CPU ONNX (ms)	Speed A100 TensorRT (ms)	params (M)	FLOPs (B)
YOLOv8n	640	37.3	80.4	0.99	3.2	8.7
YOLOv8s	640	44.9	128.4	1.20	11.2	28.6
YOLOv8m	640	50.2	234.7	1.83	25.9	78.9
YOLOv8l	640	52.9	375.2	2.39	43.7	165.2
YOLOv8x	640	53.9	479.1	3.53	68.2	257.8

Hình 3.1.2. So sánh các kích cỡ mô hình YOLOv8

Trong bài báo cáo này, nhóm sử dụng mô hình YOLOv8m để nhận diện biển số xe. Các tham số của mô hình này vừa phải phù hợp cho nhiều thiết bị hiện nay có hỗ trợ tính toán hiệu năng cao.

3.2. Gán nhãn dữ liệu

Trước khi huấn luyện mô hình, bộ dữ liệu cần được gán nhãn thủ công với thông tin của hộp giới hạn (bounding box) cho một lớp là biển số xe, việc gán nhãn dữ liệu này được thực hiện trên website Roboflow. Phần mềm giúp gán nhãn dữ liệu và có thể trích xuất ra nhiều định dạng dữ liệu khác nhau.



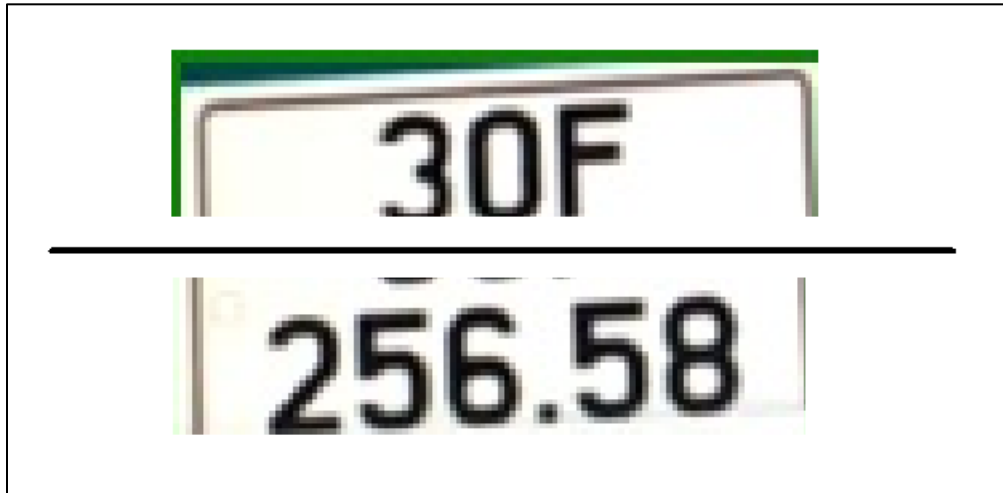
Hình 3.1.1. Mô tả quá trình gán nhãn dữ liệu

3.3. Phân loại kí tự

Hiện nay trên biển số xe ở ô tô và xe máy chấp nhận 31 kí tự bao gồm cả chữ và số. Nếu sử dụng mô hình MNIST đã có thì không thể sử dụng được do là bộ dữ liệu này đã được sử lý và trong các trường hợp đẹp nhất, thực tế bên ngoài có rất nhiều yếu tố ảnh hưởng như độ sáng, ảnh bị lệch, v.v... Trong dự án ... tác giả đã xây dựng bộ dữ liệu với các kí tự chữ cái chuẩn quốc tế alphabet nên nhóm đã chọn một phần làm bộ dữ liệu kí tự và mô hình. Tác giả xây dựng mô hình với số bộ lọc tăng dần và có sử dụng maxpooling chọn ra các đặc trưng tốt nhất của hình ảnh và đưa ra kết quả bài toán. Nhưng dự án này dự đoán các kí tự trên biển số xe mà ở đó chỉ có một dòng. Để áp dụng cho biển số xe 2 dòng, nhóm nghiên cứu đã thực hiện chia đôi bức ảnh sao cho nửa trên là dòng 1 nửa dưới là dòng 2 và đưa vào mô hình dự đoán. Dưới đây là ví dụ cụ thể



Hình 3.1.1. Hình ảnh biển số xe gốc



Hình 3.1.2. Hình ảnh sau khi được chia đôi

Mặc dù kết quả của biển số xe 2 dòng chưa được tốt, nhưng cũng đã khắc phục phần nào vấn đề chỉ nhận diện biển 1 dòng của dựa án.

CHƯƠNG 4. CHƯƠNG TRÌNH THỰC NGHIỆM

4.1. Chuẩn bị bộ dữ liệu

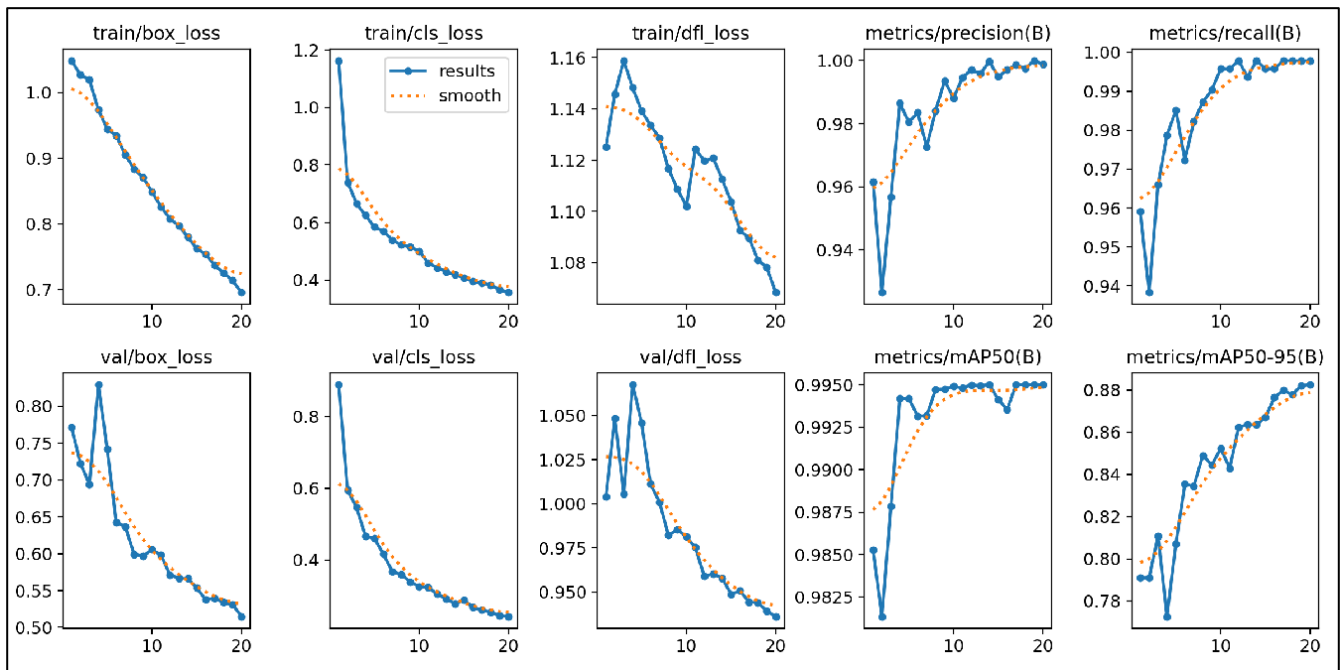
Bộ dữ liệu về biển số xe trong bài báo cáo này, nhóm đã sử dụng tổng hợp 2 bộ dữ liệu của công ty GreenParking được công khai tại ThiGiacMayTinh. Bộ dữ liệu này được chụp ở nhiều góc khác nhau và trên nhiều thiết bị khác nhau gồm 2612 bức ảnh cả xe máy và ô tô. Được gán nhãn trên Roboflows sau đó bộ dữ liệu được chia làm 70% huấn luyện, 15% kiểm tra và 15% xác thực.



Hình 4.1.1. Bộ dữ liệu GreenParking

4.2. Huấn luyện mô hình YOLOv8m

Nhóm nghiên cứu sử dụng Google Collab để huấn luyện. Google Collaboratory cung cấp môi trường biên dịch ngôn ngữ Python chạy trên nền tảng web và sử dụng các phần cứng như Tesla T4 GPU với 15 GB RAM. Chương trình được sử dụng để nhận diện biển số xe được viết bằng ngôn ngữ Python phiên bản 3.9.13. Mô hình phát hiện đối tượng được huấn luyện trên mô hình YOLOv8m đã được huấn luyện (pretrained) với 20 epochs và batch size là 16.



Hình 4.2.1. Kết quả huấn luyện mô hình YOLOv8m

Mô hình phát hiện đối tượng YOLOv8m được lựa chọn đem lại độ chính xác cao khi đánh giá trên dữ liệu thử nghiệm (test set) với các con số như: mAP với ngưỡng IoU là 50 đạt 97% và mAP với ngưỡng IoU từ 50 đến 95 đạt 65,9%.

4.3. Thiết kế phần mềm

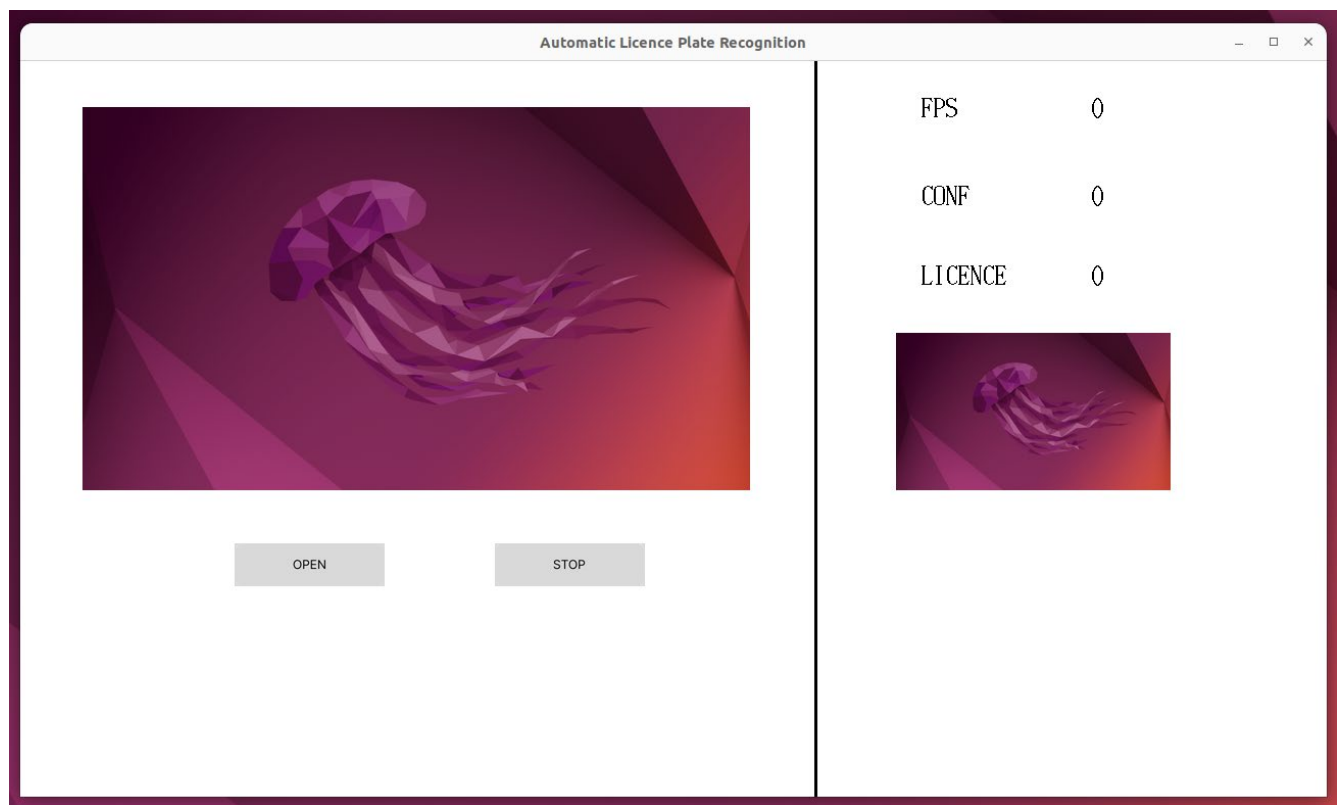
4.3.1. Thiết bị hệ thống

Thiết bị sử dụng trong báo cáo này là trên cả Ubuntu 22.04.3 và Windows. Với bộ xử lý AMD® Ryzen 7 5800hs with radeon graphics \times 16 và Đồ hoạ NVIDIA Corporation GA106M [GeForce RTX 3060 Mobile / Max-Q]. Với hệ thống như này thì với mô hình YOLOv8m có thể chạy được trong thời gian thực.

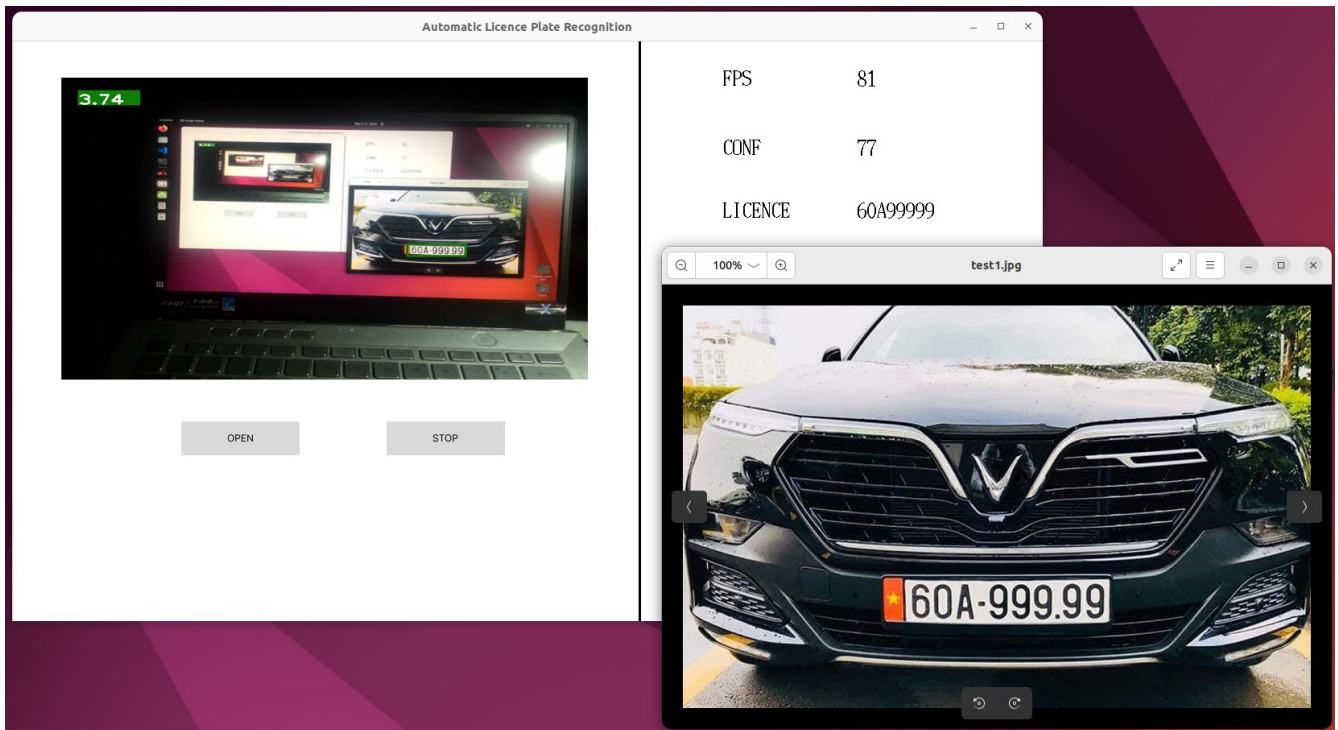
4.3.2. Môi trường thực hiện và công cụ

Môi trường thực hiện là python 3.9.13, công cụ sử dụng cho đề tài là Pycharm Community. Các thư viện sử dụng trong dự án là:

- Tkinter: giúp thiết kế giao diện, phần này nhóm đã sử dụng Figma để thiết kế và chuyển đổi giao diện đó thành giao diện hệ thống nhờ Tkinter GUI.
- Pytorch: giúp huấn luyện và chạy các mô hình học sâu như YOLO
- OpenCV: giúp xử lý ảnh đầu vào trong video



Hình 4.3.1. Giao diện chương trình



Hình 4.3.2. Kết quả khi chạy chương trình

Tất cả các mô hình, đoạn mã chương trình đều nằm ở dự án `license_plate_app[15]`.

KẾT LUẬN

Trong suốt quá trình thực hiện đồ án chuyên ngành, nhóm chúng em đã đạt được nhiều kinh nghiệm và trải nghiệm quý giá. Chúng em đã nghiên cứu và áp dụng các mô hình học sâu để nhận dạng được biển số xe. Quá trình này không chỉ là cơ hội tích lũy kiến thức chuyên môn, mà còn phát triển kỹ năng làm việc độc lập và quản lý thời gian. Kinh nghiệm này sẽ chắc chắn là hành trang quý giá cho sự nghiệp tương lai của chúng em.

Trong báo cáo này, chúng em đã đạt được các kết quả sau:

- Nghiên cứu mô hình Yolov8
- Ứng dụng mô hình Yolov8 để nhận dạng được biển số xe
- Đánh giá hiệu quả mô hình khi áp dụng
- Xây dựng được hệ thống nhận dạng biển số xe

Tuy nhiên, chúng em còn một số hạn chế:

- Mô hình chỉ tập trung vào nhận diện những xe đứng yên
- Đối với các tập dữ liệu lớn có thể dẫn đến thời gian chạy lâu và tốn nhiều tài nguyên

Trong tương lai, chúng em dự định sẽ tiếp tục nghiên cứu các vấn đề sau

- Nghiên cứu các mô hình mới trong nhận dạng biển số xe
- Nghiên cứu các phương pháp cải tiến hiệu suất cho hệ thống

Chúng em hy vọng rằng nghiên cứu của mình sẽ góp phần thúc đẩy các nghiên cứu về nhận dạng biển số xe trong tương lai

TÀI LIỆU THAM KHẢO

- [1] Báo Điện tử Chính phủ, <https://baochinhphu.vn/can-thiet-phat-trien-van-tai-khach-cong-cong-giam-un-tac-cho-do-thi-102231005180038574.htm>, truy cập gần nhất 4/1/2024.
- [2] Du, Lixuan & Zhang, Rongyu & Wang, Xiaotian, Overview of two-stage object detection algorithms. (2020), Journal of Physics: Conference Series. 1544. 012033. 10.1088/1742-6596/1544/1/012033.
- [3] Ren, S., He, K., Girshick, R., & Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks, (2015), arXiv preprint arXiv:1506.01497.
- [4] Ross Girshick and Jeff Donahue and Trevor Darrell and Jitendra Malik, Rich feature hierarchies for accurate object detection and semantic segmentation, (2014), arXiv preprint arXiv:1311.2524.
- [5] Ross Girshick, Fast R-CNN, (2015), arXiv preprint arXiv:1504.08083.
- [6] Kaiming He and Georgia Gkioxari and Piotr Dollár and Ross Girshick, Mask R-CNN, (2017), arXiv preprint arXiv:1703.06870.
- [7] Xuecheng Nie, Jianfeng Zhang, Shuicheng Yan, Jiashi Feng, Single-Stage Multi-Person Pose Machines, (2019), arXiv preprint arXiv:1908.09220.
- [8] Juan Terven, Diana Cordova-Esparza, A Comprehensive Review of YOLO: From YOLOv1 and Beyond, (2023), arXiv preprint arXiv:2304.00501.
- [9] Joseph Redmon, Ali Farhadi, YOLOv3: An Incremental Improvement, (2018), arXiv preprint arXiv:1804.02767.
- [10] Joseph Redmon, Ali Farhadi, YOLO9000: Better, Faster, Stronger, (2016), arXiv preprint arXiv:1612.08242.

- [11] Alexey Bochkovskiy, Chien-Yao Wang, Hong-Yuan Mark Liao, YOLOv4: Optimal Speed and Accuracy of Object Detection, (2020), arXiv preprint arXiv:2004.10934.
- [12] Chuyi Li, Lulu Li, Hongliang Jiang, Kaiheng Weng, Yifei Geng, Liang Li, Zaidan Ke, Qingyuan Li, Meng Cheng, Weiqiang Nie, Yiduo Li, Bo Zhang, Yufei Liang, Linyuan Zhou, Xiaoming Xu, Xiangxiang Chu, Xiaoming Wei, Xiaolin Wei, YOLOv6: A Single-Stage Object Detection Framework for Industrial Applications, (2022), arXiv preprint arXiv:2209.02976.
- [13] Chien-Yao Wang, Alexey Bochkovskiy, Hong-Yuan Mark Liao, YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors, (2022), arXiv preprint arXiv:2207.02696.
- [14] we0091234, LinWangEiGen, 2022
https://github.com/we0091234/Chinese_license_plate_detection_recognition
- [15] nguyenvanvutlv, 2023
https://github.com/nguyenvanvutlv/license_plate_app