



XÁC ĐỊNH PHƯƠNG TIỆN VI PHẠM TỐC ĐỘ

Nguyễn Thị Vân¹, Nguyễn Thị Duyên¹

¹Khoa học dữ liệu K65,

Trường Đại học Khoa học Tự nhiên - Đại học Quốc gia Hà Nội.

TÓM TẮT

Việc thiết lập một hệ thống quản lý giao thông hiệu quả là rất cần thiết, đặc biệt là trên các loại đường như đường nội thành và đường cao tốc. Đã có một số luật và các biện pháp kiểm soát tốc độ được thực hiện ở mọi nơi với những quan điểm khác nhau. Ngoài ra, giới hạn tốc độ có thể khác nhau tùy theo từng loại đường. Trong bài nghiên cứu này, chúng tôi đề xuất một chiến lược theo dõi tích hợp với việc tính vận tốc và trích xuất thông tin của các phương tiện bằng cách sử dụng thuật toán "Thị giác máy tính" và học máy. Đầu tiên, chúng tôi sử dụng kết hợp giữa YOLOv8 và thuật toán SORT để nhận dạng và theo dõi các đối tượng. Sau đó, chúng tôi đo lường tốc độ của chúng. Cuối cùng, để nhận diện biển số xe, chúng tôi đã làm nét ảnh và sử dụng mô hình easyocr để đọc ký tự. Kết quả thực nghiệm cho thấy rằng phương pháp đề xuất đạt được hiệu suất cao trong việc theo dõi tất cả các đối tượng và trích xuất thông tin xe đạt hiệu suất ổn định.

Từ khóa: Nhận diện và theo dõi phương tiện, xử lý ảnh, YOLO, Sort, easyocr.

1. ĐẶT VẤN ĐỀ

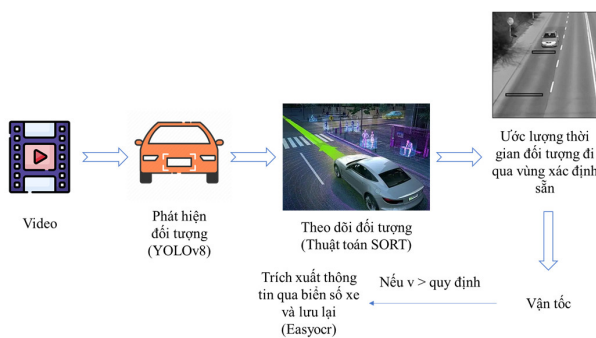
Theo số liệu của Liên hợp quốc, tại các quốc gia có thu nhập cao, tốc độ đóng vai trò là nguyên nhân của khoảng 1/3 các vụ tai nạn giao thông, trong khi ở các quốc gia có thu nhập trung bình và thấp, con số này lên tới một nửa. Hành vi chạy quá tốc độ không chỉ đe dọa tính mạng của người lái xe mà còn gây nguy hiểm cho người tham gia giao thông khác. Ở Việt Nam, từ đầu năm 2023 đến nay, đã xảy ra hơn 9.900 vụ tai nạn giao thông, khiến hơn 5.500 người thiệt mạng và khoảng 6.000 người

bị thương. Tổng thiệt hại về tài sản ước tính lên tới 518 tỷ USD.[3] Mục tiêu chính là xác định các phương tiện chạy quá tốc độ và tự động phát hiện biển số, sau đó áp dụng biện pháp xử lý đối với người lái xe. Hiện nay, chúng ta đang sử dụng các thiết bị radar cần can thiệp thủ công để phát hiện các phương tiện chạy quá tốc độ. Tuy nhiên, với sự phát triển của công nghệ, chúng ta có thể xác định các phương tiện chạy quá tốc độ mà không cần can thiệp thủ công, thông qua việc áp dụng công nghệ thị giác máy tính và học máy [1] [2]. Mặc dù đã

có nhiều phương pháp và kỹ thuật được đề xuất để ước tính tốc độ của các phương tiện, nhưng vẫn còn một số hạn chế, do đó, các nghiên cứu về phát hiện biển số xe và ước tính tốc độ vẫn đang được tiếp tục. Mặt khác, quay video từ camera giám sát không chỉ dễ dàng hơn mà còn tiết kiệm chi phí. Do đó, máy quay video đóng vai trò quan trọng trong việc phát hiện các phương tiện chạy quá tốc độ và tự động phát hiện biển số xe. Phương pháp cảnh báo quan sát nhân tạo thông thường có ưu điểm là ít can thiệp thủ công hơn, phản ứng nhanh, tỷ lệ phát hiện cao, và giúp xác định trách nhiệm sau tai nạn. Việc sử dụng công nghệ đang phát triển nhanh trong lĩnh vực này đang được đánh giá cao. Trong bài báo này, các phương tiện được xác định bằng cách sử dụng kỹ thuật thị giác máy tính và thuật toán học máy để tự động phát hiện biển số và tốc độ của phương tiện và đạt được hiệu quả, có khả năng ứng dụng thực tế.

2. PHƯƠNG PHÁP NGHIÊN CỨU

2.1. Kiến trúc hệ thống



Hình 1: Nguyên lý hoạt động của hệ thống xác định vi phạm tốc độ

Trong hệ thống chính thức, ước tính tốc độ của xe có thể được lấy từ các cảm biến khác nhau như cảm biến camera, cảm biến âm thanh, máy dò vòng lặp, cảm biến siêu âm, v.v. Trong bài báo này, tốc độ của xe và biển số của xe tương ứng có thể được

phát hiện bằng cách sử dụng video được lấy từ camera giám sát. Hầu hết các tuyến đường chính đều có camera giám sát và nó có thể được hiệu chỉnh theo khoảng cách nhất định để có thể ghi lại các phương tiện đang di chuyển, những video đó có thể được lấy làm đầu vào.

Từ đó, sử dụng YOLO để phân loại đối tượng và lựa chọn đối tượng cần xác định. Mục tiêu chính ở đây là theo dõi nhiều phương tiện cùng lúc bằng cách gán ID theo dõi của từng phương tiện và tiếp tục theo dõi cách nó di chuyển từ khung này sang khung khác trong mỗi khung. Tốc độ được ước tính bằng khoảng cách đã mặc định và tính khoảng thời gian mà đối tượng đi trong khoảng cách đó. Để xác định được khoảng thời gian chúng tôi đã tiến hành tính toán tốc độ khung hình fps của video và số frame kể từ khi đối tượng đi vào vùng đề xuất. Sau đó áp dụng công dưới đây để tính toán tốc độ bằng km/h:

$$\text{Tốc độ} = 3.6 \frac{\text{Khoảng cách}}{\text{Thời gian}}$$

với 3.6 là cách quy đổi tiêu chuẩn từ m/s sang km/h.

Đồng thời, khi ra khỏi vùng mặc định thì ID theo dõi sẽ bị xóa vì vậy có thể tránh được việc theo dõi nhiều lần cùng một đối tượng. Do đó, giai đoạn này chủ yếu nhằm mục đích gán ID cho đối tượng, theo dõi đối tượng trong mỗi khung hình, tính toán quãng đường di chuyển trong mỗi khung hình.

Sau khi đã có vận tốc các đối tượng chúng ta tiến hành so sánh ngưỡng tối đa mà nó có thể được di chuyển trên loại đường đó nếu xe nào vi phạm tốc độ chúng ta tiến hành lưu lại thông tin xe lấy từ việc lưu trữ frame, ID xe, tọa độ box của xe. Sau khi có những thông tin đó, tiếp tục áp dụng mô hình pretrain với cấu trúc YOLO và bộ dữ liệu

xác định biển số xe từ đó ta sẽ biết những biển số nào của xe nào. Do ảnh từ video cắt ra khá mờ nên chúng tôi tiến hành sử dụng mạng RRDBNet để cải thiện chất lượng hình ảnh. Cuối cùng việc trích xuất ký tự chúng ta sẽ sử dụng mô hình eas-yocr - nó có khả năng nhận diện văn bản từ ảnh với nhiều ngôn ngữ khác nhau và hỗ trợ trên nhiều nền tảng. EasyOCR sử dụng các mô hình học sâu để nhận diện văn bản và có thể được tích hợp dễ dàng vào các ứng dụng Python.

2.2. YOLO (You Only Look Once)


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 và connected layers. Trong đó các convolutional layers sẽ trích xuất ra các feature của ảnh, còn full-connected layers 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. YOLO phân chia hình ảnh thành một mạng lưới $S \times S$ ô và thương là 3×3 , 7×7 và 9×9 ,... Đầu ra của mô hình là một ma trận 3 chiều có kích thước $S \times S \times (5 \times N + M)$, trong đó S là số ô chia ảnh và N , M lần lượt là số lượng bounding box và class mà mỗi ô cần dự đoán.

Mỗi ô trong ma trận đầu ra sẽ có $(5 \times N + M)$ tham số. Trong đó, 5 tham số đầu tiên sẽ chỉ ra thông tin về bounding box, bao gồm tọa độ tâm của bounding box (x, y), chiều rộng (w), chiều cao (h) và một tham số dự đoán (prediction).

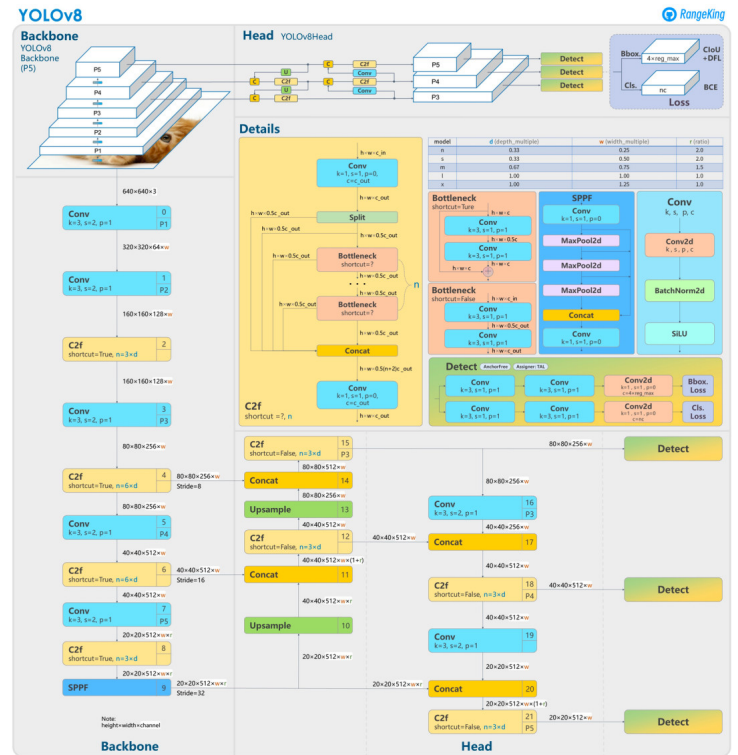
Tham số prediction được tính toán bằng công thức $P(\text{Object}) \times \text{IOU}(\text{pred}, \text{truth})$, trong đó $P(\text{Object})$ là xác suất rằng ô đó chứa một đối tượng, và $\text{IOU}(\text{pred}, \text{truth})$ là chỉ số IOU (Intersection over Union) giữa bounding box dự đoán và

bounding box thực tế.

IOU(intersection over union) là hàm đánh giá độ chính xác của object detector trên tập dữ liệu cụ thể với công thức:

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


Ở đây chúng tôi sử dụng YOLOv8 phiên bản mới nhất của YOLO và đã cải thiện được khó khăn từ phát hiện các đối tượng nhỏ ở các tỷ lệ và môi trường khác nhau của YOLO phiên bản trước đây.

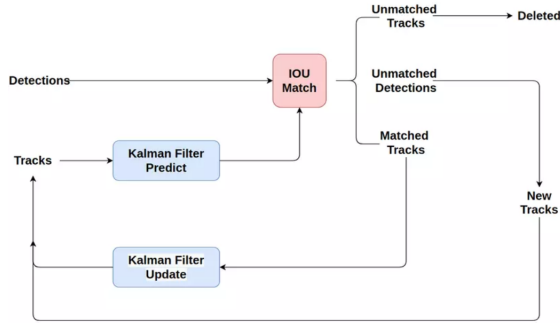


Hình 2: Cấu trúc mạng YOLOv8

2.3. SORT (Simple Online and Realtime Tracking)

[4] SORT là một thuật toán theo dõi đơn giản và thời gian thực được sử dụng để theo dõi các đối tượng trong các tập dữ liệu video. Thuật toán SORT kết hợp các bước phát hiện đối tượng và theo dõi

theo cách tách biệt, cho phép nó hoạt động trực tuyến và xử lý video theo thời gian thực.



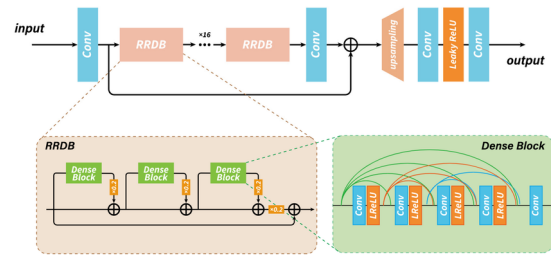
Thuật toán SORT hoạt động theo các bước chính sau:

- Bước phát hiện đối tượng: Sử dụng một mô hình phát hiện đối tượng (như YOLO, SSD, hoặc Faster R-CNN) để phát hiện và xác định các đối tượng trong mỗi khung hình của video.
- Sử dụng thuật toán Kalman Filter để dự đoán các trạng thái track mới dựa trên các track trong quá khứ.
- Sử dụng những track vừa dự đoán được, kết hợp với các detection thu được từ detector, xây dựng ma trận chi phí cho Assignment Problem. Chi phí được sử dụng để đánh giá ở đây là giá trị IOU giữa các bounding box của track và detection.
- Sử dụng giải thuật Hungary giải bài toán Assignment Problem với ma trận chi phí vừa lập.
- Xử lý, phân loại các detection.
- Sử dụng Kalman filter để update những detection đã được liên kết với track.

2.4. RRDBNET (Residual in Residual Dense Block Network)

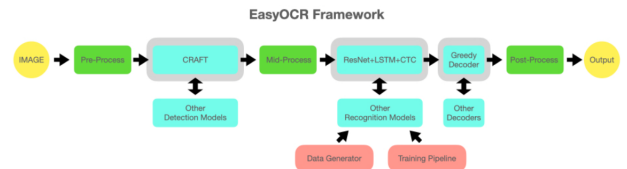
Cơ chế của RRDB Net dựa trên việc sử dụng các khối residual dense block để trích xuất đặc trưng từ ảnh. Mỗi khối này bao gồm nhiều lớp tích chập và kết nối dày đặc (dense connections) giữa các

lớp. Điều này cho phép mô hình học được các đặc trưng phức tạp từ ảnh một cách hiệu quả. Các residual dense block trong RRDB Net cho phép mô hình học được các mức độ biểu diễn đa tầng của đặc trưng, từ đó cải thiện khả năng tái tạo ảnh và giảm nhiễu. Việc sử dụng các khối này giúp mô hình học được các đặc trưng chi tiết và phức tạp từ ảnh đầu vào, từ đó tạo ra các phiên bản siêu phân giải của ảnh có chất lượng cao.



2.5. EasyOCR

EasyOCR giúp cho hệ thống OCR xử lý nhiều kiểu văn bản, phong chữ nâng cao hiệu quả của hệ thống.



Ý tưởng chính của EasyOCR là kết hợp các kỹ thuật như xử lý ảnh, mạng nơ-ron tích chập (Convolutional Neural Networks - CNN) và mô hình biến đổi ngôn ngữ tự nhiên (Transformer) để nhận dạng văn bản trong hình ảnh một cách dễ dàng và hiệu quả. Dưới đây là một tóm tắt về thuật toán của EasyOCR:

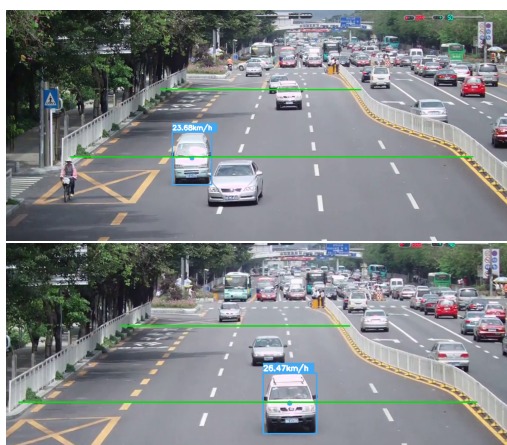
- Tiền xử lý hình ảnh: Ảnh đầu vào được tiền xử lý để cải thiện chất lượng và tách riêng văn bản với nền ảnh. Điều này có thể bao gồm các bước như làm mờ, làm rõ, chuyển đổi sang ảnh xám và phân ngưỡng.

- Xác định vùng chứa văn bản (Text Detection): EasyOCR sử dụng một mạng nơ-ron tích chập để phát hiện vùng chứa văn bản trong hình ảnh. Điều này giúp xác định các vùng chứa văn bản có thể có trong ảnh.
- Nhận dạng ký tự (Text Recognition): Sau khi xác định được vùng chứa văn bản, EasyOCR sử dụng một mô hình biến đổi ngôn ngữ tự nhiên (Transformer) để nhận dạng và trích xuất văn bản từ các vùng đó. Mô hình Transformer được huấn luyện trên dữ liệu lớn để nhận dạng nhiều ngôn ngữ và kiểu chữ khác nhau.
- Kết hợp kết quả: Cuối cùng, kết quả từ các vùng chứa văn bản được kết hợp lại để tạo ra văn bản đầy đủ và chính xác từ hình ảnh ban đầu.

3. Kết quả và thảo luận

Để thuận tiện cho việc thực hiện, nhóm nghiên cứu sử dụng các video ở các thành phố khác nhau thu thập từ các dự án về giao thông.

Đối với việc theo dõi đối tượng và tính vận tốc:



Có thể thấy việc xác định vận tốc của mô hình hoạt động hiệu quả với việc tính toán vận tốc ở trong khoảng mặc định. Đồng thời ở đây vẫn tồn tại hạn chế đó là với mỗi bất kỳ con đường nào thì

chúng tôi lại phải tiến hành chọn thủ công khoảng tính mặc định dẫn đến việc mất mát thời gian.

Đối với việc trích xuất biển số xe có thể thấy việc trích xuất biển số xe khá hiệu quả ở trong trường hợp này.



Qua những thực nghiệm với những video khác nhau chúng tôi nhận thấy rằng có những hạn chế: Không phải xe nào cũng phát hiện được biển số xe vì có xảy ra sự chồng lấp giữa các xe phía trước/phía sau. Mô hình xác định ký tự còn chưa hẳn chuẩn xác khi gặp ở góc độ quá cao, biển số xe bị quá mờ. Vì vậy chúng tôi cần có sự cải tiến hơn ở mô hình làm nét ảnh biển số xe.

4. Tài liệu tham khảo

- [1] Automated License Plate Detection and Speed Estimation of Vehicle using Machine Learning - Haar Classifier Algorithm
- [2] Real-time speed estimation of vehicles from uncalibrated view-independent traffic cameras
- [3] Hậu quả tai nạn giao thông
- [4] Bui Tien Dung, SORT - Deep SORT: Một góc nhìn về Object Tracking, VIBLO, December 2020, Available from
- [5] Nhận diện biển số xe tự động
- [6] Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network