

CẢI THIỆN ĐỘ CHÍNH XÁC TRUNG BÌNH KHI SỬ DỤNG MẠNG YOLOV11 ĐỂ XÂY DỰNG MÔ HÌNH PHÁT HIỆN BIỂN BÁO GIAO THÔNG ĐƯỜNG BỘ

IMPROVING MEAN AVERAGE PRECISION USING YOLOV11 TO BUILD A ROAD SIGN DETECTION MODEL

NGUYỄN ĐÌNH KHIÊM*, NGUYỄN ĐỨC SANG

Viện Cơ khí, Trường Đại học Hàng hải Việt Nam

**Email liên hệ: khiemnd.vck@vimaru.edu.vn*

Tóm tắt

Trong bài báo này, trình bày về những điểm mới của mạng YOLOV11 so với các phiên bản trước. Đồng thời sử dụng mạng học sâu YOLOv11 để xây dựng mô hình nhận diện biển báo giao thông trên một tập dữ liệu gồm 9.635 ảnh của 56 loại biển báo giao thông đường bộ tại Việt Nam. Mô hình đã huấn luyện đạt được độ chính xác trung bình - mean Average Precision là 97,06% với mô hình YOLOv11x sau 100 epoch huấn luyện. Mặt khác việc sử dụng transfer learning trong huấn luyện mạng YOLOv11 thì độ chính xác trung bình đạt 99,09% ở mô hình YOLOv11x, tăng 2,03%. Với kết quả này, nhóm tác giả đã xây dựng mô hình thực nghiệm nhận dạng biển báo giao thông.

Từ khóa: YOLOv11, độ chính xác trung bình, phát hiện và nhận dạng biển báo giao thông.

Abstract

In this article, we present the new features of the YOLOV11 network compared to previous versions. At the same time, we use the YOLOv11 deep learning network to build a traffic sign recognition model on a data set of 9.635 images of 56 types of road traffic signs in Vietnam. The trained model achieved an average accuracy of 97,06% with the YOLOv11x model after 100 training epochs. On the other hand, using transfer learning in training the YOLOv11 network, the average accuracy reached 99,09% in the YOLOv11x model, an increase of 2,03%. With this result, the authors have built an experimental model for traffic sign recognition.

Keywords: YOLOv11, mean average precision, detect and recognize traffic signs.

1. Mở đầu

Xe tự lái là một trong những lĩnh vực được quan tâm nghiên cứu nhiều nhất trong những năm gần đây. Một trong những thách thức quan trọng đối với xe tự lái khi tham gia giao thông là khả năng nhận diện và phản ứng chính xác với biển báo giao thông, những yếu tố có vai trò quyết định trong việc điều hướng và đảm bảo an toàn. Chính vì thế, việc nghiên cứu các hệ thống nhận diện biển báo giao thông nhanh, chính xác và hiệu quả là một yếu tố không thể thiếu trong quá trình phát triển công nghệ xe tự hành. Đây cũng là tiền đề quan trọng để đưa xe tự hành vào thực tiễn một cách an toàn và hiệu quả.

Đã có nhiều công trình nghiên cứu để giải quyết bài toán phát hiện và nhận diện biển báo giao thông. Nhóm tác giả Bùi Quốc Tú đã sử dụng mô hình YOLOv4 [1] huấn luyện trên 1500 ảnh với 32 loại biển báo và tín hiệu đèn giao thông, đạt được độ chính xác trung bình 91% [2]. Mô hình YOLOv5 [3] đã được Yanzhao Zhu, Wei Qi Yan sử dụng để huấn luyện trên 2182 ảnh với 8 loại biển báo, đạt độ chính xác trung bình ấn tượng là 97,7% [4]. Một mô hình khác cũng đạt được độ chính xác cao 94,4% khi sử dụng YOLOv7 [5] để huấn luyện trên 6945 tấm ảnh với 9 loại biển báo do Trần Thịnh Mạnh Đức, Đỗ Trí Nhựt đề xuất [6]. X. Dai và cộng sự [7] đã phát triển một chiến lược mới để cải thiện độ chính xác của việc nhận dạng biển báo giao thông trong các điều kiện ánh sáng khác nhau. Họ đạt được điều này bằng cách tận dụng các tính năng màu sắc, mang lại độ chính xác là 78% và tốc độ xử lý là 11 khung hình/giây. C. Dewi và cộng sự [8] đã sử dụng GAN (Generative Adversarial Network) để tạo ra bộ sưu tập hình ảnh đào tạo lớn hơn và đa dạng hơn, từ đó nâng cao tập dữ liệu. Bằng cách tích hợp hình ảnh tổng hợp với hình ảnh gốc, họ nhằm mục đích nâng cao chất lượng tổng thể của tập dữ liệu. Kết quả nghiên cứu của họ cho thấy độ chính xác là 84,9% khi sử dụng YOLOv3 và cải thiện hơn nữa lên 89,33% với YOLOv4. Có thể thấy các cách tiếp cận được giới thiệu ở trên là sử dụng các mô hình phát hiện đối tượng (chủ yếu là họ các mô hình

YOLO), số lượng ảnh chưa nhiều nên độ chính xác mặc dù cao nhưng chưa thực sự tin cậy, số loại biển báo nhận diện được còn ít. Ngoài ra, các cách tiếp cận này cũng chưa đề xuất hướng xây dựng mô hình vật lý để có thể áp dụng mô hình nhận diện biển báo giao thông vào thực tế.

Gần đây, phiên bản kế tiếp của các mô hình YOLO là YOLOv11 [9] đã được giới thiệu và đạt được độ chính xác và tốc độ cao hơn so với các mô hình YOLO trước đó. Trong bài báo này, chúng tôi đề xuất áp dụng mô hình phát hiện đối tượng YOLOv11 trên tập dữ liệu hình ảnh lớn với ngữ cảnh phức tạp để xây dựng mô hình nhận dạng 56 (trên hơn 382) loại biển báo giao thông đường bộ ở Việt Nam. Nội dung tiếp theo của bài báo sẽ được trình bày theo bố cục sau: Phần 2 sẽ trình bày những về những điểm mới trong mô hình kiến trúc của mạng YOLOv11; Phần 3 là ứng dụng mạng YOLOv11 để huấn luyện mô hình phát hiện biển báo giao thông; Phần 4 là triển khai mô hình phát hiện biển báo giao thông trên một mô hình vật lý và phần cuối cùng là kết luận.

2. Mạng YOLOv11

Mạng YOLO, viết tắt của cụm từ You Only Look Once (bạn chỉ nhìn một lần) [10] nổi bật vì sự cân bằng đáng kinh ngạc giữa tốc độ và độ chính xác. Mạng YOLO cho phép nhận dạng nhanh chóng và đáng tin cậy các vật thể trong hình ảnh. Kể từ khi ra đời, mạng YOLO đã phát triển qua nhiều phiên bản, mỗi phiên bản đều dựa trên các phiên bản trước để giải quyết các hạn chế và nâng cao hiệu suất.

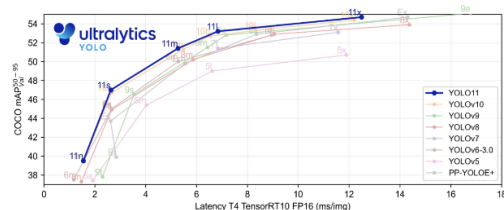
YOLOv11 là phiên bản mới được giới thiệu vào tháng 9 năm 2024, định nghĩa lại những gì có thể về độ chính xác, tốc độ và hiệu quả, đại diện cho một bước tiến đáng kể trong công nghệ phát hiện đối tượng theo thời gian thực. Dựa trên những tiến bộ ấn tượng của các phiên bản YOLO trước đó, YOLOv11 giới thiệu những cải tiến đáng kể về kiến trúc và phương pháp đào tạo, biến nó thành lựa chọn linh hoạt cho nhiều tác vụ thị giác máy tính. YOLO11 mở rộng và nâng cao nền tảng do YOLOv8 [11] đặt ra. Các phần sau đây nêu chi tiết các sửa đổi kiến trúc quan trọng được triển khai trong YOLOv11:

Đổi mới về kiến trúc: Mô hình kết hợp các yếu tố kiến trúc mới giúp tăng cường khả năng trích xuất và xử lý tính năng. Việc kết hợp các yếu tố mới như khối C3k2 (Cross Stage Partial with kernel size 2) thay thế cho khối C2f, SPFF (Spatial Pyramid Pooling - Fast) và C2PSA (Convolutional block with Parallel Spatial Attention) góp phần trích xuất và xử lý tính năng hiệu quả hơn. Những cải tiến này cho phép mô hình phân

tích và diễn giải thông tin hình ảnh phức tạp tốt hơn, có khả năng dẫn đến cải thiện độ chính xác phát hiện trong nhiều tình huống khác nhau. Kiến trúc của mô hình mạng YOLOv11 có thể xem ở [12].

Cơ chế chú ý nâng cao: Một tiến bộ quan trọng trong YOLOv11 là tích hợp các cơ chế chú ý không gian tinh vi, đặc biệt là thành phần C2PSA. Việc đưa C2PSA vào giúp YOLOv11 khác biệt so với người tiền nhiệm của nó, YOLOv8, không có cơ chế chú ý cụ thể này. Tính năng này cho phép mô hình tập trung hiệu quả hơn vào các vùng quan trọng trong hình ảnh, tăng cường khả năng phát hiện và phân tích các đối tượng. Khả năng chú ý được cải thiện đặc biệt có lợi cho việc xác định các đối tượng phức tạp hoặc bị che khuất một phần, giải quyết một thách thức phổ biến trong các tác vụ phát hiện đối tượng. Sự tinh chỉnh này trong nhận thức không gian góp phần vào những cải tiến hiệu suất tổng thể của YOLOv11, đặc biệt là trong các môi trường thị giác đầy thách thức.

Biểu đồ so sánh hiệu suất được mô tả trong Hình 1 bao gồm một số thông tin chi tiết quan trọng.

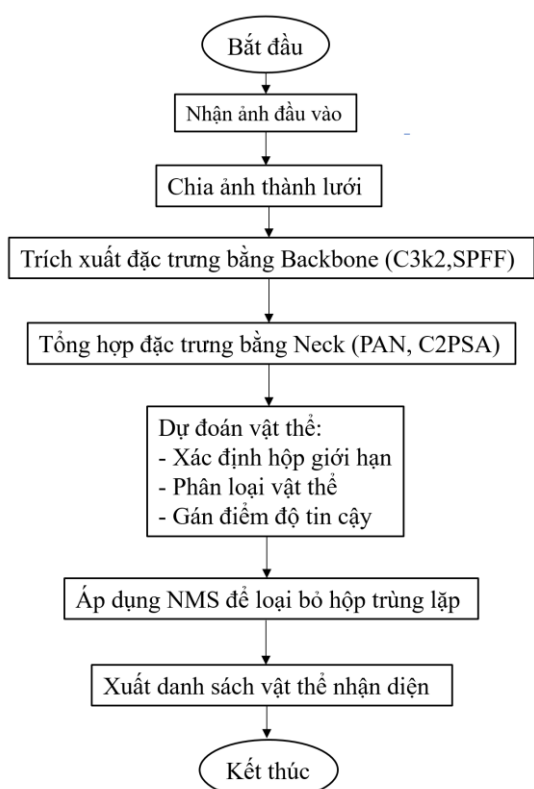


Hình 1. So sánh YOLOv11 với các phiên bản trước [9]

Các biến thể YOLOv11 (11n, 11s, 11m, 11l và 11x) tạo thành một ranh giới hiệu suất riêng biệt, với mỗi mô hình đạt điểm COCO mAP 50-95 cao hơn tại các điểm độ trễ tương ứng của chúng. Đáng chú ý, YOLOv11x đạt được khoảng 54,5% mAP 50-95 ở độ trễ 13ms, vượt qua tất cả các phiên bản YOLO trước đó. Các biến thể trung gian, đặc biệt là YOLOv11m, chứng minh hiệu quả đặc biệt bằng cách đạt được độ chính xác tương đương với các mô hình lớn hơn từ các thế hệ trước trong khi yêu cầu thời gian xử lý ít hơn đáng kể. Cụ thể, YOLO11m đạt độ chính xác trung bình (mAP) cao hơn trên tập dữ liệu COCO với số tham số ít hơn 22% so với YOLOv8m [9], giúp tăng hiệu quả tính toán mà không ảnh hưởng đến độ chính xác.

Thuật toán YOLOv11 được trình bày ở Hình 2, còn ở Hình 3 là lưu đồ huấn luyện mạng YOLOv11.

Trong đó PAN (Path Aggregation Network) là một kiến trúc giúp tăng cường thông tin từ các tầng sâu và các tầng nông của mạng Backbone. Thuật toán



Hình 2. Lưu đồ thuật toán YOLOv11

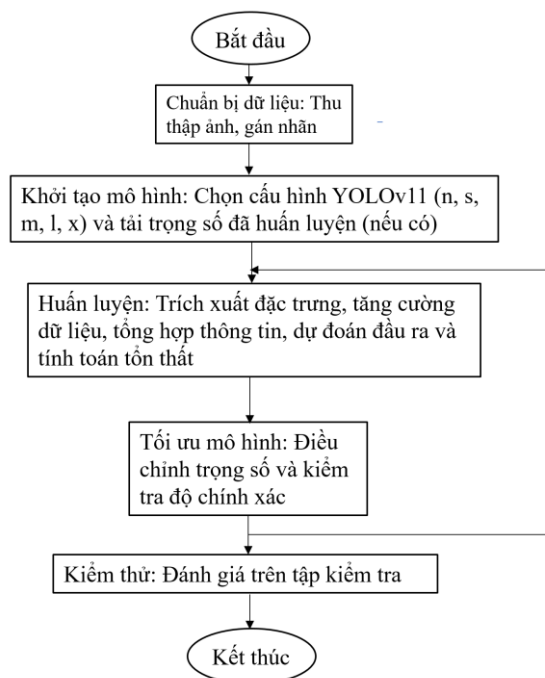
YOLOv11 (Hình 2) vẫn áp dụng NMS (Non-Maximum Suppression), giúp loại bỏ các hộp giới hạn trùng lặp để chỉ giữ lại hộp có độ tin cậy cao nhất cho mỗi vật thể.

Dựa trên lưu đồ huấn luyện ở Hình 3 nhóm tác giả đã tiến hành xây dựng mô hình phát hiện biển báo giao thông ở Phần 3.

3. Cải thiện độ chính xác trung bình khi sử dụng mạng YOLOv11 để xây dựng mô hình nhận diện biển báo giao thông

Để xây dựng mô hình nhận diện biển báo giao thông, nhóm tác giả đã lựa chọn tập dữ liệu công cộng gồm hình ảnh các biển báo giao thông đường bộ ở Việt Nam, cụ thể tập huấn luyện gồm 6750 ảnh (minh họa ở Hình 4), tập đánh giá gồm 1922 ảnh và tập kiểm tra gồm 963 ảnh với kích thước 640x640 (<https://universe.roboflow.com/giang-yp9gl/vietnam-traffic-sign-altsi>).

YOLOv11 có 5 phiên bản khác nhau là Nano (n), Small (s), Medium (m), Large (l) và Extra Large (x). Các phiên bản này khác nhau ba tham số chính: Depth (độ sâu của mạng), Width (độ rộng của mạng), Max Channels (giới hạn số lượng kênh tối đa). Trong đó mô hình Nano có kích thước nhỏ nhất, độ chính xác thấp



Hình 3. Lưu đồ huấn luyện mạng YOLOv11

nhất, tốc độ thực hiện nhanh nhất còn mô hình Extra Large có kích thước lớn nhất, độ chính xác cao nhất và thời gian thực hiện lâu nhất. Trong bài báo này nhóm tác giả đã thực hiện huấn luyện trên cả 5 mô hình với YOLOv11, sau 100 epoch thì thu được kết quả về độ chính xác trung bình trên tập huấn luyện và tập đánh giá như ở Bảng 1. Phương pháp huấn luyện có sử dụng transfer learning, các kỹ thuật tăng cường dữ liệu và được thực hiện trên Google Colab với GPU A100.



Hình 4. Một số biển báo giao thông trong tập huấn luyện

Trong Bảng 1 thông số mAP(50), là độ chính xác trung bình được tính ở ngưỡng giao nhau trên hợp nhau (IoU) là 0,50. IoU là một số liệu quan trọng trong phát hiện đối tượng, cung cấp thước đo định lượng về mức độ chồng chéo giữa hộp giới hạn thực tế cơ bản (ground-truth bounding box) và hộp giới hạn dự đoán (predicted bounding box) do mô hình phát hiện đối tượng tạo ra. Thông số mAP (50-95) là trung bình của

độ chính xác trung bình được tính ở các ngưỡng IoU khác nhau, dao động từ 0,50 đến 0,95.

Từ Bảng 1 cho thấy độ chính xác trung bình (mean Average Precision) mAP(50) của cả 4 phiên bản (trừ phiên bản YOLOv11no) đều cao, thấp nhất là 92,88% ở mô hình YOLOv11so và cao nhất là 99,09% ở mô hình YOLOv11x. Các kết quả với giá trị đánh giá mAP(50-90) trên tập huấn luyện cũng đạt được 78,2% với mô hình YOLOv11so và tăng đến 92,79% ở YOLOv11x. Đồng thời trên tập đánh giá thì độ chính xác của mô hình cũng tương tự như tập huấn luyện, cho thấy mô hình đã huấn luyện là tin cậy.

Bảng 1. So sánh độ chính xác trung bình

STT	YOLO	Tập huấn luyện (%)		Tập đánh giá (%)	
		Ảnh 640x640			
		mAP (50)	mAP (50-90)	mAP (50)	mAP (50-90)
1	v11n	94,99	84,70	94,80	84,70
2	v11s	98,74	91,35	98,30	91,30
3	v11m	98,50	91,13	98,10	91,10
4	v11l	98,54	91,93	98,40	91,90
5	v11x	99,09	92,79	99,10	92,80
6	v11no	51,40	41,49	51,60	41,70
7	v11so	92,88	78,20	92,80	78,20
8	v11mo	96,81	87,85	96,80	87,80
9	v11lo	97,09	89,13	97,10	89,10
10	v11xo	97,06	89,62	97,00	89,60

Ở Bảng 1, các mô hình YOLOv11no đến YOLOv11xo là các ký hiệu thể hiện không sử dụng transfer learning trong quá trình huấn luyện để phân biệt với các mô hình YOLOv11n đến YOLOv11x là có sử dụng transfer learning (học chuyển giao) trong khi huấn luyện. Transfer learning được sử dụng ở đây là việc tải trọng số từ mô hình đã được huấn luyện trước với dữ liệu lớn (Tải trọng số từ các mô hình YOLOv11n.pt đến YOLOv11x.pt một cách tương ứng). Việc tải trọng số không những tránh phải huấn luyện từ đầu mà còn giúp tận dụng được các kiến thức về đặc trưng của ảnh, kiến thức về phát hiện vật thể, một số đối tượng tham gia giao thông đã biết,... Điều này góp phần nâng cao độ chính xác của mô hình khi huấn luyện. Kết quả ở Bảng 2 cho thấy, khi sử dụng transfer learning thì độ chính xác của mô hình tăng từ 1,45% (YOLOv11l so với YOLOv11lo) đến 13,15% (YOLOv11s so với YOLOv11so) ở tập huấn luyện và còn trên tập đánh giá thì tăng từ 1,3% (YOLOv11m so

với YOLOv11mo) đến 13,1% ở (YOLOv11s so với YOLOv11so). Đặc biệt, ở YOLOv11n đã tăng 43,59% ở tập huấn luyện và 43,2% trên tập đánh giá.

4. Xây dựng mô hình thực nghiệm

Trước khi xây dựng mô hình thực nghiệm, nhóm tác giả đã đánh giá các mô hình đã huấn luyện ở Bảng 1 trên một laptop có GPU Quadro P4000 qua một số video thu thập từ Internet. Việc đánh giá cũng được thực hiện khi mô hình huấn luyện ở định dạng file của thư viện Pytorch và dạng triển khai thực tế TensorRT-FP16. Kết quả cho thấy mô hình có độ chính xác cao hơn thì tốc độ phát hiện (khung hình trên giây) chậm hơn. Với laptop có GPU Quadro P4000 ở trên thì mô hình thực nghiệm có thể đạt tới 25 FPS (khung hình trên giây) và độ chính xác đạt 71% ở một số loại biển báo, minh họa ở Hình 5.

Bảng 2. Mức tăng mAP khi sử dụng transfer learning

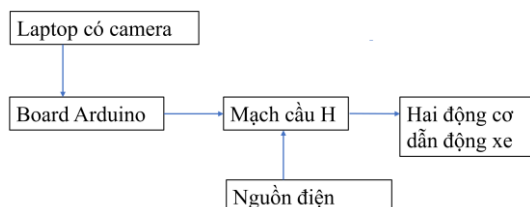
STT	YOLO	Tập huấn luyện (%)		Tập đánh giá (%)	
		Ảnh 640x640			
		mAP (50)	mAP (50-90)	mAP (50)	mAP (50-90)
1	v11n	43,59	43,21	43,2	43,00
2	v11s	5,86	13,15	5,50	13,10
3	v11m	1,69	3,28	1,3	3,3
4	v11l	1,45	2,8	1,3	2,8
5	v11x	2,03	3,17	2,10	3,20



Hình 5. Kết quả nhận diện biển báo với dữ liệu từ Internet

Dựa trên kết quả kiểm thử, nhóm tác giả đã xây dựng một mô hình thực nghiệm theo sơ đồ khối ở Hình 6, được minh họa ở Hình 7. Trong Hình 6, laptop có camera sẽ triển khai mô hình phát hiện biển báo giao thông đã được lựa chọn dựa trên tốc độ (khung hình trên giây) và độ chính xác qua ngôn ngữ Python. Laptop sẽ gửi tín hiệu đến board arduino để điều khiển hai động cơ dẫn động xe thông qua mạch cầu H. Nguồn điện được sử dụng ở đây là một ắc quy nhỏ 12V. Trong quá trình xe di chuyển, camera trên laptop

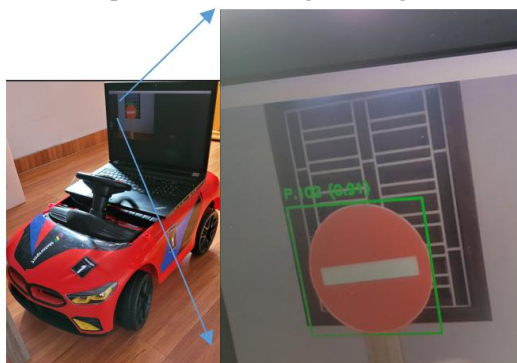
sẽ chụp ảnh, nếu gặp biển báo giao thông ví dụ biển P.102 (biển cấm đi ngược chiều) thì dừng xe. Hình 8 minh họa mô hình phát hiện biển báo giao thông được áp dụng vào mô hình thực nghiệm cho kết quả nhận diện biển P.102 là 91%.



Hình 6. Sơ đồ khối mô hình thực nghiệm



Hình 7. Các thành phần thực tế của mô hình phát hiện biển báo giao thông



Hình 8. Mô hình thực nghiệm

5. Kết luận

Để cải thiện độ chính xác trung bình khi ứng dụng mạng YOLOv11 để xây dựng mô hình phát hiện biển báo giao thông đường bộ ở Việt Nam, chúng tôi đã huấn luyện các phiên bản YOLOv11 khác nhau trên tập dữ liệu gồm 9635 ảnh. Chúng tôi cũng đưa ra kết quả so sánh giữa việc huấn luyện có sử dụng transfer learning và không sử dụng transfer learning. Kết quả cho thấy việc sử dụng transfer learning đã giúp tăng độ chính xác trung bình của mô hình lên 2,03%, từ 97,06 lên 99,09%, vượt trội so với các công trình đã giới thiệu trong bài báo. Ngoài ra, mô hình thực nghiệm đã xây dựng có thể đạt được tốc độ 25 FPS và độ chính xác đạt 71% ở một số loại biển báo.

Lời cảm ơn

Nghiên cứu này được tài trợ bởi Trường Đại học Hàng hải Việt Nam trong đề tài mã số: **DT24-25.36**.

TÀI LIỆU THAM KHẢO

- [1] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao (2020), *YOLOv4: Optimal Speed and Accuracy of Object Detection*. arXiv, Apr.22.
doi: 10.48550/arXiv.2004.10934.
- [2] Bùi Quốc Tú, Nguyễn Huy Hoàng, Trương Quang Phúc, Lê Quang Bình, Hồ Nhật Minh (2022), *Nhận diện biển báo và tín hiệu đèn giao thông sử dụng YOLOv4 trên phần cứng Jetson TX2*, Tạp chí Khoa học Công nghệ và Thực phẩm, tr.132-142.
- [3] Ultralytics (2021), *YOLOv5: A state-of-the-art real-time object detection system*.
[Online]. Available: <https://docs.ultralytics.com>
- [4] Yanzhao Zhu, Wei Qi Yan (2022), *Traffic sign recognition based on deep learning*, Multimedia Tools and Application, pp.17779-17791.
- [5] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao (2022), *YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors*, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp.7464-7475.
doi: 10.48550/arXiv.2207.02696.
- [6] Trần Thịnh Mạnh Đức, Đỗ Trí Nhật (2024), *Thiết kế hệ thống nhận dạng biển báo giao thông với ứng dụng YOLO*, Tạp chí Đại học Thủ Dầu Một, Số 2(69).
- [7] X. Dai, X. Yuan, G. Le and L. Zhang (2018), *Detection method of traffic signs based on color pair and MSER in the complex environment*, Journal of Beijing Jiaotong University, Vol.42, pp.107-115.
- [8] C. Dewi, R.-C. Chen, Y.-T. Liu, X. Jiang and K. D. Hartomo (2021), *YOLO v4 for advanced traffic sign recognition with synthetic training data generated by various GAN*, IEEE Access, Vol.9, pp.97228-97242.
- [9] [Online]. Available:
<https://docs.ultralytics.com/models/yolo11/>
- [10] Nguyễn Hữu Tuấn, Nguyễn Duy Trường Giang (2024), *Ứng dụng mạng YOLOv8 trong phát hiện hư hỏng trên mặt đường bộ*, Tạp chí Khoa học Công nghệ Hàng hải, Số 78, tr.74-78.

- [11] J. Glenn, C. Ayush, and Q. Jing (2023),
Ultralytics YOLOv8.

[Online]. Available:

<https://github.com/ultralytics/ultralytics>.

- [12] Zhou Wang, Yuting Su, Feng Kang, Lijin Wang,
Yaohua Lin, Qingshou Wu, Huicheng Li, Zhiling
Cai (2025), *PC-YOLO11s: A Lightweight and
Effective Feature Extraction Method for Small
Target Image Detection*, Sensors, Vol.25(2), 348p.

<https://doi.org/10.3390/s25020348>.

Ngày nhận bài:	11/03/2025
Ngày nhận bản sửa:	25/03/2025
Ngày duyệt đăng:	25/03/2025