

NGHIÊN CỨU MÔ HÌNH YOLOv8 ĐỂ NHẬN DẠNG BIỂN BÁO GIAO THÔNG ĐƯỜNG BỘ TẠI VIỆT NAM

● PHẠM TUẤN ĐẠT

Khoa Công nghệ Thông tin, Trường Đại học Hàng hải Việt Nam

Email: datpt@vimaru.edu.vn

TÓM TẮT:

Nhận dạng biển báo giao thông tự động đã trở thành một tiêu chuẩn bắt buộc và yêu cầu công nghệ đối với các hệ thống thông minh hiện đại cũng như xe tự lái nói riêng. Phương pháp học máy truyền thống bộc lộ những hạn chế trong vấn đề khái quát đặc trưng và hiệu quả tính toán. Bài viết khám phá những ưu điểm riêng biệt của một mô hình YOLO hiện có và áp dụng vào nhiệm vụ phát hiện biển báo giao thông. Thực nghiệm phát hiện và phân loại trên tập dữ liệu huấn luyện cho kết quả rất ấn tượng và tương đối chính xác. Hơn nữa, thời gian suy luận để dự đoán trên tập dữ liệu kiểm thử đáp ứng yêu cầu thời gian thực khi phương tiện di chuyển trong khu vực đô thị. Do đó, một mô hình YOLO tối ưu có thể mang tới một lựa chọn đáng tin cậy giải quyết các vấn đề an toàn giao thông trong tương lai.

Từ khóa: độ chính xác trung bình, khung bao, vùng giao trên vùng hợp.

1. Đặt vấn đề

Chủ đề an toàn giao thông hay an ninh giám sát bao gồm nhận dạng tự động biển báo và đèn tín hiệu giao thông, truy vết đối tượng tại những khu vực công cộng cũng như phát hiện đám cháy tại các kho bãi là những bài toán có thể triển khai ứng dụng trong lĩnh vực thương mại và công nghiệp.

Học máy áp dụng kỹ thuật phổ biến là Haar-like features [1] và Hog [2] để phát hiện đối tượng trong hình ảnh trước khi giải thuật phân loại thực hiện chức năng nhận dạng nhãn lớp đối tượng. Tuy nhiên, không dễ dàng mô tả khái quát tập đặc trưng của các loại biển báo giao thông cho 2 phương pháp này. Trái lại, mô hình YOLO xây dựng bản đồ đặc trưng biểu diễn đối

tượng thông qua cấu trúc mạng nơ ron nhân chấp. Việc dự đoán đối tượng có thể có trong một ô lưới điểm dựa vào thông tin trên toàn bộ hình ảnh giúp mô hình giảm lỗi nhận dạng nhằm vùng nền trông giống đối tượng. Nhờ những cải tiến về kiến trúc mạng cùng với các giải thuật huấn luyện tối ưu, mô hình YOLO cho kết quả nhanh và chính xác vượt trội [3,4] so với một số kiến trúc học sâu như LRF hay RFBNet [4].

Xuất phát từ những ưu thế trên, mô hình YOLOv8-Small được lựa chọn để thử nghiệm cho bài toán nhận dạng biển báo giao thông đường bộ tại Việt Nam. Mặc dù tập dữ liệu có số lượng khung hình ít ỏi nhưng ứng dụng đã cho kết quả tốt trên dữ liệu đào tạo và kiểm thử.

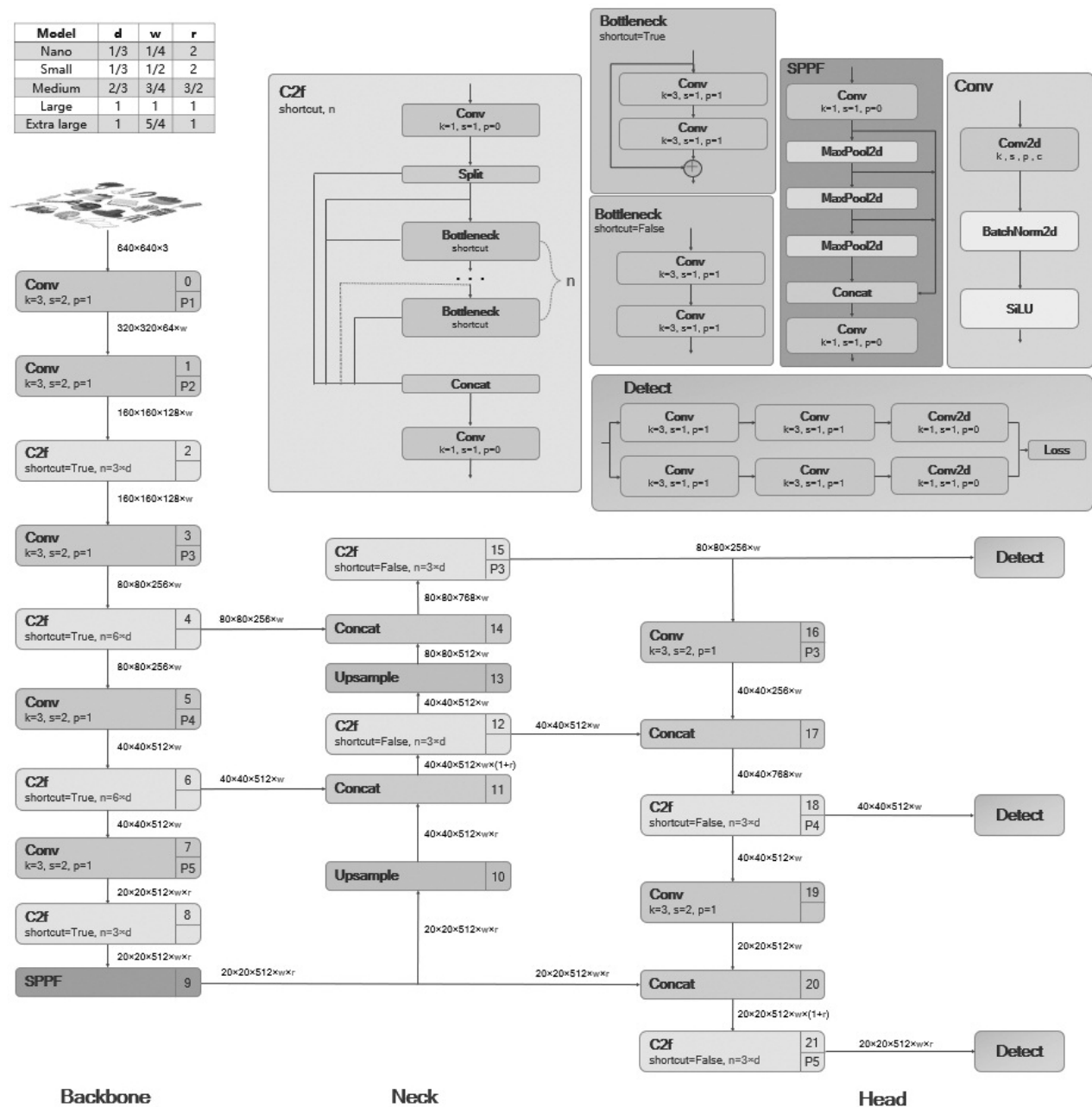
2. Kiến trúc YOLOv8

Phiên bản YOLO đầu tiên được công bố trong một bài viết khoa học [3]. Từ đó, những cải tiến sau này [5,6,7,8] được nghiên cứu bởi những nhóm tác giả khác nhau. Ultralytics phát triển YOLO phiên bản 8 và 11 nhưng nhiều phiên bản khác cũng được Ultralytics triển khai mô hình và giải thuật dưới dạng mã nguồn mở. Mỗi phiên bản YOLO của Ultralytics thường có nhiều mô hình đáp ứng một loại tác vụ.

YOLOv8 [7] được ứng dụng trong tác vụ phát hiện đối tượng, 5 mô hình (Nano, Small, Medium, Large và Extra-large) chia sẻ một khung kiến trúc nhưng khác nhau về số lượng tầng và số lượng kênh. Hình 1 minh họa khung kiến trúc chứa 3 khối chính gồm có phần xương sống mạng (Back-bone), phần cổ mạng (Neck) và phần đầu mạng (Head).

Back-bone chứa cấu trúc C2F cho phép truyền rẽ nhánh và gộp thông tin để biểu diễn đặc trưng đầy đủ.

Hình 1: Tổng quan kiến trúc mô hình YOLOv8



Mặt khác, khối trục xương sống kế thừa cấu trúc SPPF [5], nó chứa các tầng Pooling nối tiếp có cửa sổ trượt kích thước 5x5 và giảm bớt so với thành phần tương ứng của cấu trúc SPP [4]. Neck đóng vai trò cầu nối, kết hợp các đặc trưng nhận từ các tầng của cấu trúc xương sống để chuẩn bị cho việc dự báo. Khối này cũng kế thừa một số nguyên tắc thiết kế từ các kiến trúc YOLO trước đó, đồng thời mang đến những cải tiến trong việc trích xuất và phát hiện đặc trưng. Back-bone và Neck đều có những cấu trúc C2F nhưng C2F của Back-bone chứa những kết nối tắt còn C2F của Neck thì không.

Head phân chia 2 chức năng là dự đoán khung bao chứa đối tượng và phân loại lớp. Mô hình quét toàn bộ hình ảnh để dự đoán khung bao chứa đối tượng theo 3 thang phân giải. Chức năng dự đoán khung bao có thêm đánh giá tiêu điểm phân phối [9] khắc phục tình huống đối tượng mờ hoặc ranh giới đối tượng không rõ ràng bằng cách ước lượng giá trị xác suất. Sau khi đã định vị được khung bao thì chức năng phân loại sẽ ước lượng xác suất nhãn lớp đối tượng. Mỗi khung bao được đánh giá độ tin cậy bằng phép nhân giữa IoU [3,5] và xác suất phân loại lớp đối tượng. Trường hợp có nhiều hơn một khung bao với IoU vượt ngưỡng và chúng cùng chứa một loại đối tượng thì giải thuật NMS [5] giữ khung bao có điểm tin cậy cao nhất và loại bỏ các khung chồng lấn. Nhưng cách làm của NMS có thể xóa bỏ một khung bao có sự trùng khớp tọa độ gần đúng nhất với khung bao thực tế. Vì thế, những phiên bản mới đây như YOLOv10 [8] không áp dụng phương pháp này trong quá trình suy luận.

3. Thử nghiệm

Tập dữ liệu theo quy định báo hiệu đường bộ (QCVN41-2024) được sưu tầm từ Vietnam Traffic Signs Dataset của MaiTam trên Kaggle bao gồm 3216 hình ảnh thuộc 52 thể loại biển báo từ nhiều địa điểm khác nhau tại Việt Nam. Mỗi hình ảnh có thông tin các nhãn và vị trí tương đối xác định các khung chứa biển báo trong hình ảnh đó. Tập dữ liệu được trộn ngẫu nhiên (không có trùng lặp dữ

liệu giữa 2 pha đào tạo) và chia thành 3 tập con gồm dữ liệu huấn luyện, xác thực và kiểm thử theo tỷ lệ tương ứng 0.7:0.1:0.2.

Mô hình YOLOv8-Small tiếp nhận hình ảnh theo thang phân giải 640x640 trên 3 kênh RGB. Ứng dụng đào tạo mô hình có sự hỗ trợ GPU P100, tích hợp giải thuật Mosaic tăng cường sự đa dạng ngữ cảnh trên tập huấn luyện và giải thuật tối ưu AdamW [10] điều chỉnh tập trọng số nhằm tối thiểu lỗi dự đoán nhầm nhãn lớp. Trong 80 vòng lặp, lỗi dự đoán khung bao trên tập huấn luyện có xu hướng giảm đều thì lỗi tương ứng trên tập xác thực giảm chậm hơn và chưa thực sự ổn định. Tại Bảng 1, tổn thất dự đoán khung bao trên tập huấn luyện và xác thực còn cao, song mục tiêu của việc huấn luyện là nắm bắt được đặc trưng chính của từng loại biển báo chứ không ghi nhớ tọa độ. Kết thúc tiến trình đào tạo, mAP₅₀₋₉₅ trên tập huấn luyện xấp xỉ 0.81 còn tập xác thực chỉ cỡ 0.77. Nếu đặt ngưỡng thấp cho mức trùng khớp (IoU) giữa khung chứa được dự đoán và khung bao thực tế thì mô hình phân loại rất chính xác, nhưng chất lượng suy giảm dần khi nâng mức ngưỡng. Cụ thể, mAP₅₀ luôn vượt 0.94 nhưng mAP₉₅ thường dưới 0.1 trên dữ liệu kiểm thử. Nhìn chung, độ chính xác trung bình tại một dải ngưỡng trên tập xác thực và tập kiểm thử chênh lệch không quá 2%.

Chỉ số mAP (trên một ngưỡng hay dải ngưỡng) cho biết khả năng nhận dạng của mô hình này nhưng không thể hiện kết quả cụ thể và khuyết điểm tồn tại. Nhằm làm rõ vấn đề này, bên cạnh kết quả của mAP₇₀, thử nghiệm tiến hành ước lượng Precision và Recall [5] trên 52 lớp. Kết quả trong Bảng 2 dựa trên số liệu thống kê từ ma trận nhầm lẫn thỏa mãn ngưỡng (0.7) cho điểm tin cậy và độ trùng khớp. Chỉ số Precision cho thấy mô hình có tỷ lệ rất thấp hoặc

Bảng 1. Kết quả huấn luyện

	Tập huấn luyện	Tập xác thực
Tổn thất dự đoán khung bao	0.70	0.75
Tổn thất phân loại	0.36	0.39
mAP ₇₀	0.92 ± 0.02	0.88 ± 0.02
mAP ₅₀₋₉₅	0.81	0.77

thấp đối với dự đoán sai giữa các cặp biển báo khác lớp cũng như đối với dự đoán nhầm vùng ảnh nền thành biển báo. Mặt khác, chỉ số Recall cho biết chỉ có một tỷ lệ nhỏ số lớp khiến mô hình không thể phát hiện được biển báo. Tổng thể, những lớp chứa lực lượng không đáng kể thường gây cho mô hình khó khăn trong việc phát hiện hoặc phân loại đúng.

Bảng 2. Precision và Recall

	Tập huấn luyện	Tập kiểm thử
Precision	0.90±0.02	0.87±0.03
Recall	0.90±0.02	0.87±0.03

Hình 2: Minh họa kết quả nhận dạng



Tập dữ liệu được sưu tầm trong điều kiện thời tiết bình thường nên các hình ảnh đều rõ nét. Đối với tình huống phát hiện đúng, mô hình cần không quá 15 ms để xử lý và phân loại mỗi biển báo giao thông. Hình 2

thể hiện kết quả phát hiện và phân loại 2 biển báo giao thông trong một khung hình của tập kiểm thử. Mô hình suy luận chỉ đặt ngưỡng cho điểm tin cậy chứ không quy định ngưỡng IoU như trong ma trận nhầm lẫn, do vậy có thể nảy sinh trường hợp khung bao chứa đối tượng có điểm tin cậy cao vượt ngưỡng nhưng độ trùng khớp không cao. Trong thực tế, những biển báo có kích thước khác biệt nhiều so với giới hạn thang phân giải (20x20, 40x40, hay 80x80) hoặc hình ảnh biển báo mờ hoặc có nhiễu, thì mô hình này cũng có thể không phát hiện được hay nhầm lẫn biển báo.

4. Kết luận

Bài viết trình bày kết quả thử nghiệm mô hình YOLOv8-Small nhận dạng biển báo giao thông có độ chính xác nhận dạng ấn tượng trên dữ liệu đào tạo cũng như kiểm thử với chi phí thời gian thấp. Dựa trên tập dữ liệu tuy còn ít ỏi, nhưng bài viết đã phân tích và ước lượng khả năng nhận dạng, độ chính xác và bao phủ thực tế của mô hình để thấy được những ưu thế và khiếm khuyết. Việc thu thập dữ liệu từ các bối cảnh địa điểm không phong phú cũng như số lượng hình ảnh chưa bao quát đối với nhiều nhãn biển báo giao thông là một nguyên nhân dẫn tới những kết quả trên tập dữ liệu thấp hơn kỳ vọng. Với sự cải tiến trong kiến trúc và giải thuật dự đoán của những phiên bản sau này, nếu được huấn luyện trên tập dữ liệu bổ sung sẽ tăng cường độ chính xác cho ứng dụng phát hiện và phân loại biển báo giao thông ■

TÀI LIỆU TRÍCH DẪN VÀ THAM KHẢO:

[1] Paul Viola and Michael Jones (2001). Rapid Object Detection using a Boosted Cascade of Simple Features, Conference on Computer Vision and Pattern Recognition.

[2] Navneet Dalal and Bill Triggs (2005). Histograms of Oriented Gradients for Human Detection. Conference on Computer Vision and Pattern Recognition.

[3] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi (2016). You Only Look Once: Unified, Real-Time Object Detection, Conference on Computer Vision and Pattern Recognition.

[4] Alexey Bochkovskiy, Chien-Yao Wang, and Hong-Yuan Mark Liao (2020). YOLOv4: Optimal Speed and Accuracy of Object Detection, <https://doi.org/10.48550/arXiv.2004.10934>.

[5] Hao Chen, Zhan Chen and HangYu (2023). Enhanced YOLOv5: An efficient Road Object Detection Method, Sensors (MDPI).

[6] Chien-Yao Wang, Alexey Bochkovskiy, and Hong-Yuan Mark Liao (2023). YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors, Conference on Computer Vision and Pattern Recognition.

- [7] Glenn Jocher, Ayush Chaurasia, Jing Qiu (2023). YOLOv8, <https://github.com/ultralytics/ultralytics>.
- [8] Ao Wang, Hui Chen, Lihao Liu, Kai Chen, Zijia Lin, Jungong Han, and Guiguang Ding (2024). YOLOv10: Real-Time End-to-End Object Detection, <https://arxiv.org/pdf/2405.14458>.
- [9] Xiang Li, Wenhai Wang, Lijun Wu, Shuo Chen, Xiaolin Hu, Jun Li, Jinhui Tang, and Jian Yang (2020). Generalized Focal Loss: Learning Qualified and Distributed Bounding Boxes for Dense Object Detection, Conference on Neural Information Processing Systems.
- [10] Ilya Loshchilov and Frank Hutter (2019). Decoupled Weight Decay Regularization, International Conference on Learning Representations.

Ngày nhận bài: 15/03/2026

Ngày phản biện đánh giá và sửa chữa: 01/04/2026

Ngày chấp nhận đăng bài: 15/04/2026

A STUDY ON THE APPLICATION OF THE YOLOv8 MODEL FOR ROAD TRAFFIC SIGN RECOGNITION IN VIETNAM

● PHAM TUAN DAT

Faculty of Information Technology,
Vietnam Maritime University

ABSTRACT:

Automatic traffic sign recognition has become an essential technological requirement for modern intelligent transportation systems, particularly in the development of autonomous vehicles and advanced driver-assistance systems. Conventional machine-learning approaches often exhibit limitations in feature generalization, detection accuracy, and computational efficiency when applied to complex traffic environments. This study investigates the application of an existing YOLO-based model for traffic sign detection and classification, emphasizing its advantages in real-time object recognition tasks. Experimental results obtained from training and testing datasets demonstrate high detection accuracy and reliable classification performance. In addition, the model achieved inference speeds that satisfy real-time processing requirements for vehicles operating in urban traffic conditions. The findings indicate that an optimized YOLO framework provides an effective and robust solution for traffic sign recognition, thereby contributing to improved road safety and the advancement of intelligent transportation technologies.

Keywords: average accuracy, bounding box, intersection region over union region.