

RESEARCH YOLOv8 AND YOLO-NAS VERSIONS IN LICENSE PLATE DETECTION

Dang Thi Dung*, Ha Le Ngoc Dung, Truong Le Chuong, Thai Chi Hao, Tran Van Phuc

Can Tho University of Engineering - Technology

ARTICLE INFO		ABSTRACT
Received:	09/5/2024	Recently, license plate recognition systems have been an important part of many traffic management and security systems such as automatic speed control, stolen vehicle tracking, automatic toll management, and control of vehicles entering and exiting bus station areas, schools, hospitals, etc. During the research process, we compared versions of YOLOv8 and YOLO-NAS based on the criteria of Accuracy, Precision, Recall, and F1 score to evaluate the most suitable models for vehicle license plate recognition in Vietnam under different environmental conditions. This review provides perspective for developers or last users to choose the most suitable technique for their application. The results show that for applications with good infrastructure and high accuracy requirements, YOLO-NAS-S is a suitable model with an Accuracy of 83.92%, Precision of 0.9125; Recall is 0.9125, and F1 score is 0.9125. For less developed infrastructure and speed requirements, YOLOv8n can be used with a smaller number of parameters but the accuracy is acceptable, Accuracy is 81.4%; Precision is 0.9625; Recall is 0.8415, and F1 score is 0.8979.
Revised:	10/6/2024	
Published:	11/6/2024	
KEYWORDS		
YOLOv8		
YOLO-NAS		
Vehicle License Plate Detection		
Machine Learning		
Deep Learning		

NGHIÊN CỨU CÁC PHIÊN BẢN YOLOv8 VÀ YOLO-NAS TRONG PHÁT HIỆN BIỂN SỐ XE

Đặng Thị Dung*, Hà Lê Ngọc Dung, Trương Lê Chương, Thái Chí Hào, Trần Văn Phúc

Trường Đại học Kỹ thuật – Công nghệ Cần Thơ

THÔNG TIN BÀI BÁO		TÓM TẮT
Ngày nhận bài:	09/5/2024	Trong thời gian gần đây, hệ thống nhận dạng biển số xe là một phần quan trọng trong nhiều hệ thống quản lý giao thông và an ninh như kiểm soát tốc độ tự động, theo dõi xe bị đánh cắp, quản lý phí tự động và kiểm soát xe ra vào các khu vực bến xe, trường học, bệnh viện,... Trong quá trình nghiên cứu, chúng tôi tiến hành so sánh các phiên bản của YOLOv8 và YOLO-NAS theo các tiêu chí về độ Accuracy, Precision, Recall và F1 score để đánh giá các mô hình phù hợp nhất đối với việc nhận diện biển số xe ở Việt Nam trong các điều kiện môi trường khác nhau. Đánh giá này đưa ra quan điểm để các nhà phát triển hoặc người dùng cuối lựa chọn kỹ thuật phù hợp nhất cho ứng dụng của họ. Kết quả cho thấy đối với các ứng dụng có cơ sở hạ tầng tốt và yêu cầu có độ chính xác cao thì YOLO-NAS-S là một mô hình phù hợp với Accuracy 83,92%, Precision 0,9125; Recall 0,9125 và F1 score 0,9125. Đối với cơ sở hạ tầng kém phát triển hơn và yêu cầu về tốc độ thì có thể sử dụng YOLOv8n với số lượng tham số ít hơn nhưng độ chính xác lại khá ổn Accuracy 81,4%; Precision 0,9625; Recall 0,8415 và F1 score 0,8979.
Ngày hoàn thiện:	10/6/2024	
Ngày đăng:	11/6/2024	
TỪ KHÓA		
YOLOv8		
YOLO-NAS		
Phát hiện biển số xe		
Học máy		
Học sâu		

DOI: <https://doi.org/10.34238/tnu-jst.10336>

* Corresponding author. Email: dtdung@ctu.edu.vn

1. Giới thiệu

Phát hiện đối tượng là một trong những bài toán quan trọng của thị giác máy tính với các ứng dụng trải nghiệm rộng rãi trong nhiều lĩnh vực khác nhau như: Robot công nghệ, xử lý ảnh y khoa, hệ thống giám sát, hệ thống tương tác người máy, giao thông thông minh,... Phát hiện đối tượng có hai hướng tiếp cận: Hướng tiếp cận truyền thống sử dụng đặc trưng tự thiết kế như Haar-like [1], HOG [2], DPM [3] và hướng tiếp cận hiện đại dựa trên mạng học sâu. Khi đề cập đến sử dụng phương pháp Deep Learning được chia thành hai nhóm chính: Máy dò hai giai đoạn R-CNN [4], Fast R-CNN [5], Faster R-CNN [6], Mask R-CNN [7] và máy dò một giai đoạn YOLO [8], SSD [9], Retinanet [10].

Máy dò một giai đoạn sử dụng mạng thần kinh riêng biệt để thực hiện một chuyển tiếp giữa việc tạo ra các khung (Bounding Box) và xác định đối tượng trong thời gian thực. Mô hình YOLOv8 [11] đã được phát triển và cải tiến liên tục để nâng cao khả năng nhận diện đặc biệt là biển số xe. YOLO-NAS [12] sử dụng kỹ thuật tìm kiếm Nơ-ron để tối ưu hóa mô hình, cải thiện hiệu suất và sử dụng tài nguyên hiệu quả. Đánh giá hai mô hình YOLOv8 và YOLO-NAS dựa trên việc so sánh hình ảnh trong các điều kiện khác nhau như vị trí, ánh sáng và khoảng cách.

1.1. Tổng quan tình hình nghiên cứu

Nghiên cứu sử dụng các mô hình máy học trong đời sống được ứng dụng rộng rãi trong nhiều lĩnh vực khác nhau, đặc biệt là các nghiên cứu về nhận diện biển số xe: Hệ thống giám sát tốc độ xe, hệ thống bãi giữ xe thông minh... Phương pháp so khớp mẫu là phương pháp dựa trên sự tương quan giữa mẫu đưa vào và mẫu có sẵn trong cơ sở dữ liệu. Một trong những phương pháp được sử dụng trong bài nghiên cứu [13], sau 3 lần thử nghiệm với mỗi lần sử dụng 100 ảnh khác nhau nhóm nghiên cứu thu được độ chính xác ở mức 80%. Đối với nghiên cứu [14], sử dụng ba góc nhìn khác nhau để thu thập dữ liệu biển số xe. Sau đó, dữ liệu được sử dụng để huấn luyện một mô hình học sâu để nhận dạng các ký tự trên biển số xe. Phương pháp đạt được độ chính xác cao trong việc nhận dạng các ký tự trên biển số xe, vượt trội so với các phương pháp trước đây. Sử dụng tập dữ liệu hơn 12.500 hình ảnh được chia tập huấn luyện 10.000 hình ảnh và tập kiểm thử 2.500 hình ảnh. Kết quả thử nghiệm cho thấy mô hình đạt độ chính xác cao 96,9% được thể hiện trong nghiên cứu [15]. Trong bài nghiên cứu [16], mô hình đã đạt được tỷ lệ nhận dạng 96,9% trên tất cả các tập dữ liệu, vượt trội so với các công trình trước đó và các hệ thống thương mại. Với tập dữ liệu hơn 10.000 ảnh gồm hai tập huấn luyện 8.000 ảnh và tập kiểm thử 2.000 ảnh cùng với việc đa dạng các biển số xe gồm biển số xe cá nhân, biển số xe thương mại, biển số xe ngoại giao. Mô hình đem lại độ chính xác nhận dạng trung bình là 99,8% và có thể hoạt động hiệu quả trong các điều kiện môi trường phức tạp được nhóm tác giả thực hiện trong nghiên cứu [17]. Trong bài nghiên cứu [18], bài báo sử dụng tập dữ liệu địa phương về biển số xe Saudi được thu thập và chú thích bởi tác giả. Kết quả hệ thống đạt được Accuracy là 0,97; Recall 0,985; F1-score 0,982. Đặc biệt, trong bài nghiên cứu [19], bài báo sử dụng bộ dữ liệu Stanford Cars được lấy từ Kaggle gồm 16200 hình ảnh để đào tạo và kiểm tra, gồm một tập hợp các hình ảnh ô tô đa dạng chụp từ nhiều góc độ và điều kiện chiếu sáng khác nhau. Mô hình YOLOv8 cho Accuracy 93% trong cả tập dữ liệu đào tạo và xác thực, giảm nhẹ xuống 90% trong tập dữ liệu thử nghiệm, đối với Faster R-CNN lần lượt là 71%, 71% và 74%.

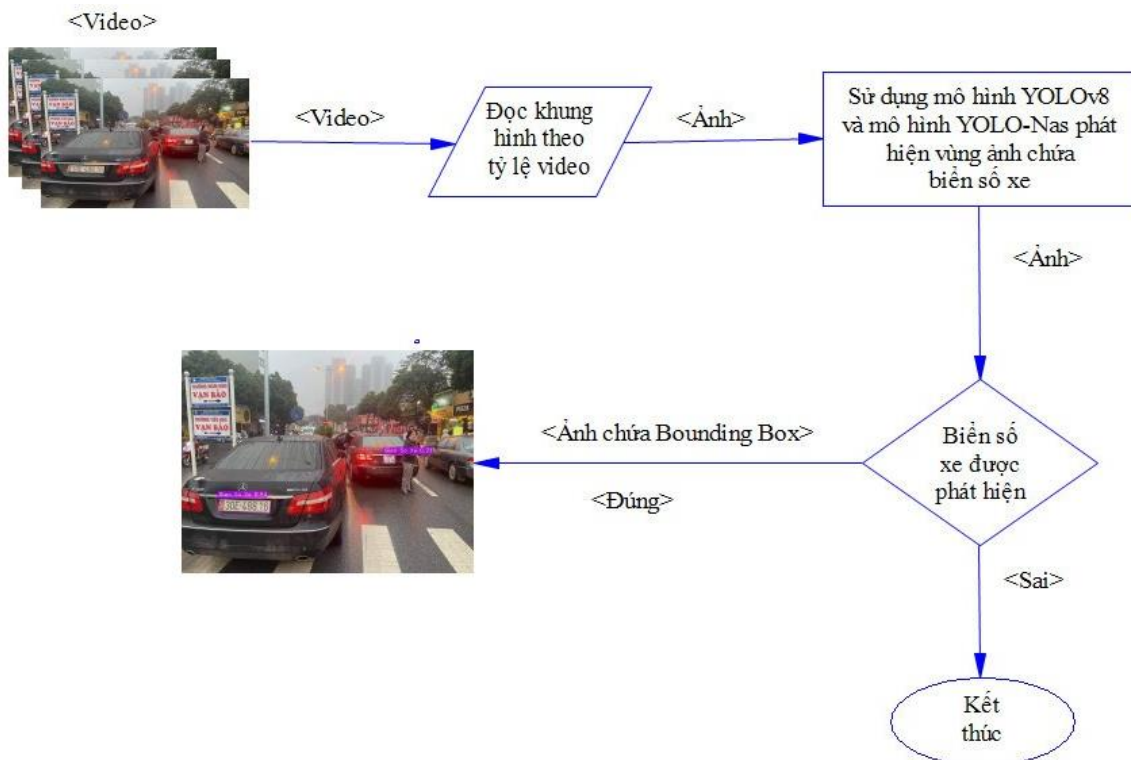
1.2. Quá trình đào tạo và nhận diện của mô hình YOLO

YOLO là một mô hình phát hiện đối tượng theo thời gian thực được thiết kế dựa trên kiến trúc mạng CNN có khả năng phát hiện phân loại đối tượng trong hình ảnh hoặc video. Quá trình đào tạo đòi hỏi phải chuẩn bị một tập dữ liệu gồm các hình ảnh biển số xe, mỗi hình ảnh phải được gắn nhãn với thông tin về vị trí và loại đối tượng. Lựa chọn phiên bản YOLO và xây dựng mạng Nơ-ron CNN phù hợp. Tiếp theo sử dụng dữ liệu đã được chuẩn bị sẵn (Dataset) và mô hình

YOLO để thực hiện quá trình huấn luyện. Quá trình này sử dụng nhiều thuật toán để điều chỉnh trọng số mô hình sao cho sai số dự đoán giảm xuống thấp nhất. Bước đầu của giai đoạn nhận diện, YOLO sẽ chia hình ảnh ban đầu thành mạng lưới các ô $S \times S$, mỗi ô lưới chứa B hộp giới hạn. Mỗi hộp giới hạn được đặc trưng bởi năm giá trị: x, y, w, h và độ tin cậy. Tọa độ (x, y) thể hiện tâm của hộp so với lưới, chiều rộng (w) và chiều cao (h) tương ứng với toàn bộ hình ảnh. Cuối cùng, là độ tin cậy cho biết khả năng chứa đối tượng thực tế và mức độ chính xác của dự đoán.

2. Phương pháp đề xuất

2.1. Mô hình bài toán



Hình 1. Mô hình bài toán

Hình 1 thể hiện mô hình bài toán thực hiện việc nhận diện biển số xe của các mô hình YOLOv8 và YOLO-NAS. Dữ liệu cần kiểm tra sẽ được thu thập từ nhiều nguồn khác nhau có thể là ảnh hoặc video. Dữ liệu sẽ được đưa vào mô hình, đối với video sẽ thực hiện việc đọc khung hình để đưa ra ảnh có pixel phù hợp được quy định. Sử dụng mô hình YOLOv8 hoặc YOLO-NAS để nhận diện ảnh có chứa biển số xe, nếu ảnh có chứa biển số xe mô hình sẽ khoanh vùng bằng Bounding Box và đưa ra hình ảnh hoặc video tùy thuộc theo dữ liệu đầu vào; nếu ảnh không chứa biển số xe mô hình sẽ kết thúc.

2.2. Cấu hình bộ dữ liệu

Trong phạm vi nghiên cứu này, chúng tôi tập trung vào một lớp duy nhất: Biển số xe ở Việt Nam. Hình ảnh được thu thập thông qua việc tìm kiếm trên Web, từ các Camera giao thông và chụp bằng máy ảnh, các thiết bị di động. Tổng cộng, chúng tôi thu thập hơn 1500 hình ảnh, bao gồm ảnh biển số xe máy và biển số xe ô tô mặt trước và mặt sau. Tập dữ liệu này chúng tôi chỉ tập trung chủ yếu vào biển số xe của các phương tiện trên đường, trên bãi đỗ xe, trên lề đường,... Chúng tôi loại bỏ những hình ảnh không liên quan hoặc mờ, không rõ ràng, hoặc biển số bị biến dạng quá nhiều. Tại Hình 2, một số hình ảnh mô phỏng được trích xuất nhằm phục vụ nghiên cứu.

Chúng tôi sử dụng nền tảng Make để tiền xử lý hình ảnh. Các hộp giới hạn được thực hiện thủ công trong các khu vực liên quan đến biển số xe trên mỗi hình ảnh và được gán lớp tương ứng. Trong tổng số 1567 hình ảnh, có 1334 hình dành cho tập huấn luyện (Train), 133 cho tập xác nhận (Val) và 100 cho tập thử nghiệm (Test).



Hình 2. Một số hình ảnh được trích xuất từ tập dữ liệu được sử dụng cho nghiên cứu này: (a) Hình ảnh xe ô tô chạy trên đường (mặt trước), (b) Hình ảnh xe ô tô chạy trên đường (mặt sau), (c) Hình ảnh xe ô tô và xe máy chạy trên đường, (d) Hình ảnh xe máy ở phía sau được chụp từ camera của một bãi đỗ xe từ một công viên, (e) Hình ảnh xe máy chạy trên đường, (f) Hình ảnh một số xe máy đang đỗ

2.3. Nền tảng thử nghiệm

Nền tảng chúng tôi sử dụng cho các thử nghiệm trong bài viết này là Google Colab với phần cứng của hệ thống:

- CPU: Intel Xeon E5-2676 v4
- GPU: Nvidia Tesla V100-SXM2-16GB với 16 GB bộ nhớ GDDR6X.
- RAM: 51GB
- Ổ cứng: SSD 166 GB

2.4. Bản thông số chính của hai mô hình YOLO

Bảng 1. Thông số đào tạo mô hình

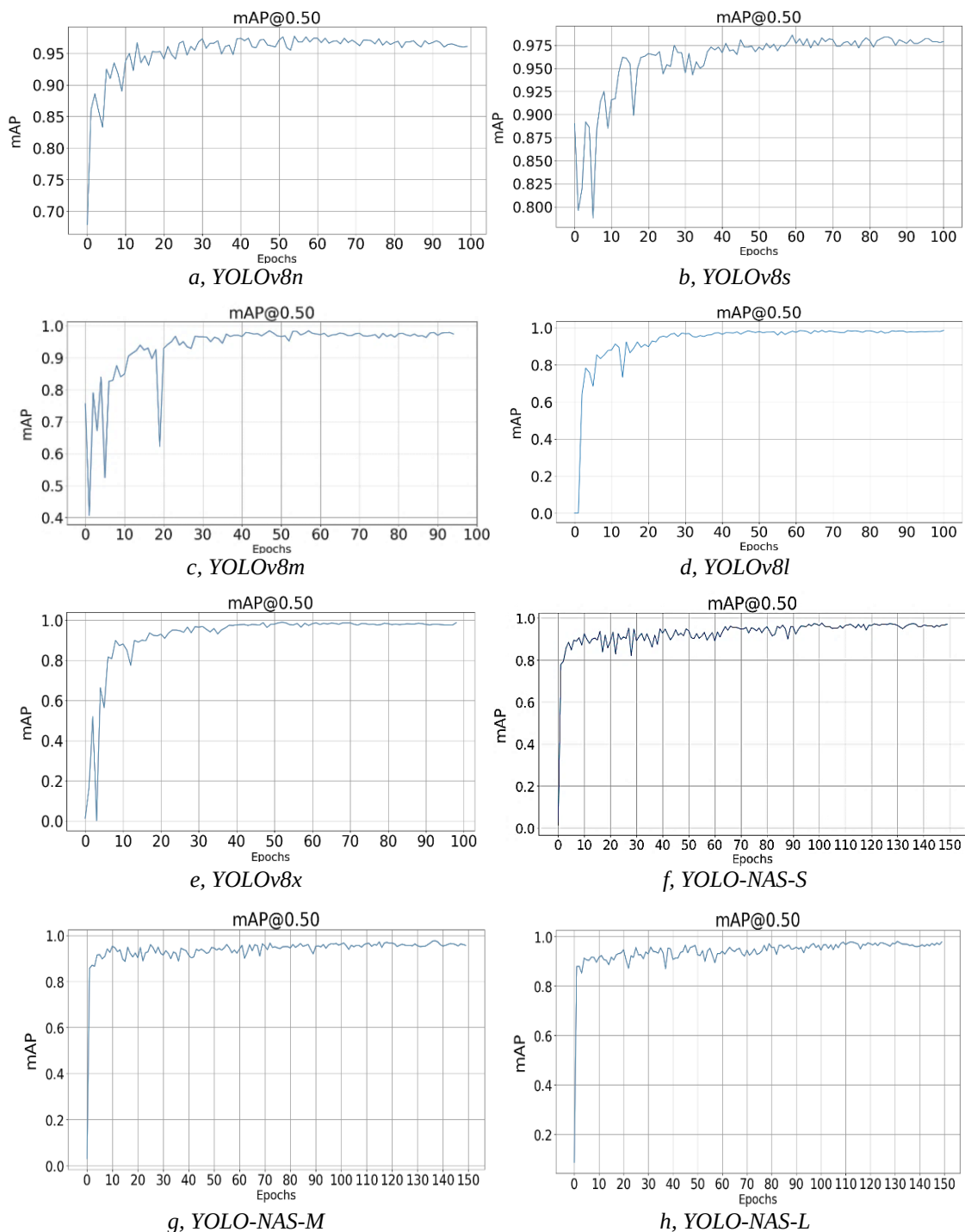
	YOLOv8n	YOLOv8s	YOLOv8m	YOLOv8l	YOLOv8x	YOLO-NAS-s	YOLO-NAS-m	YOLO-NAS-l
Epochs	100	100	100	100	100	150	150	150

Bảng 1 hiển thị các thông số đào tạo mô hình YOLOv8 và YOLO-NAS. Đối với YOLOv8, sau nhiều lần thử nghiệm với số Epochs ban đầu là 50, các chỉ số vẫn chưa đạt đủ yêu cầu chúng tôi đã tăng/giảm số lượng Epochs để có kết quả tốt nhất và kết quả trung bình chúng tôi chọn để Training với các Model V8 Epochs là 100, imgsz= 640 (kích thước ảnh của tập dữ liệu).

Với mô hình YOLO-NAS thời gian trung bình để Training trên cùng một tập dữ liệu tương đối lâu hơn so với mô hình YOLOv8. Ban đầu số Epochs mặc định của chúng tôi vẫn là 50 với Batchsize = 16 và sau nhiều lần điều chỉnh để có kết quả tốt nhất, chỉ số Epochs cuối cùng là 150, Batchsize = 16 cùng với các thông số mặc định khác.

3. Kết quả và thảo luận

3.1. Kết quả sau khi đào tạo mô hình



Hình 3. Đánh giá huấn luyện các mô hình

Đối với một bài toán, có nhiều cách để đánh giá một mô hình nhận diện như các chỉ số Accuracy, Precision, Recall và F1 cores nhưng để xem xét rõ ràng nhất quá trình đào tạo một tập

dữ liệu có tính chính xác hay không chủ yếu dựa vào chỉ số AP hoặc mAP. Đối với mAP là tính trung bình của chỉ số AP được đánh giá trong khoảng (0; 1), với mAP càng lớn thì mô hình càng được đánh giá cao. Do vậy, Hình 3 thể hiện các biểu đồ mAP của các mô hình khi huấn luyện, đối với mô hình YOLOv8 chỉ cần 100 Epoch đã đem lại mAP > 0,9, nhờ vậy quá trình đào tạo YOLOv8 tiêu tốn ít thời gian hơn đào tạo một mô hình YOLO-NAS rất nhiều. YOLO-NAS cần số Epoch khoảng 100 đã ổn định tuy nhiên độ Loss của mô hình cao nên số Epoch đã tăng lên khoảng 150 để giảm các chỉ số Loss.

3.2. So sánh thực tế các phiên bản YOLOv8 và YOLO-NAS



Ảnh đầu vào



(a) Ảnh nhận diện bởi YOLOv8n

(b) Ảnh nhận diện bởi YOLOv8s

(c) Ảnh nhận diện bởi YOLOv8m



(d) Ảnh nhận diện bởi YOLOv8l



(e) Ảnh nhận diện bởi YOLOv8x



(f) Ảnh nhận diện bởi YOLO-NAS-S



(g) Ảnh nhận diện bởi YOLO-NAS-M



(h) Ảnh nhận diện bởi YOLO-NAS-L

Hình 4. Hình ảnh xe trong nhà được nhận diện

Để tìm ra đối với các phiên bản của YOLOv8 và chứng minh khả năng phát hiện biển số xe, chúng tôi đã tiến hành thí nghiệm so sánh năm mô hình YOLOv8 cơ bản: YOLOv8n, YOLOv8s, YOLOv8m, YOLOv8l và YOLOv8x.

Đối với hình ảnh đầu tiên là các xe máy được đặt trong nhà: Điều kiện ánh sáng đầy đủ cũng như độ sắc nét hoàn chỉnh, gồm 6 hình ảnh biển số xe không trực diện, được chụp xéo thể hiện tại Hình 4. Về cơ bản các mô hình đều nhận diện được ít nhất 2 biển số xe và cao nhất 6 biển số, không nhận diện thêm đối tượng nào khác ngoài biển số xe. Mô hình YOLOv8n nhận diện được ít biển số xe nhất chỉ 2 biển số, YOLOv8m và YOLOv8s có thể nhận diện đúng 3 biển số xe. Trong khi đó các mô hình YOLOv8l và YOLOv8x lại vượt trội hơn nhận diện được 4 biển số xe. Đối với các phiên bản YOLO-NAS lần lượt đối với phiên bản YOLO-NAS-S nhận diện được 5 biển số xe, với YOLO-NAS-M và YOLO-NAS-L nhận diện được toàn bộ biển số xe cụ thể là 6 biển số.

Đối với hình ảnh thứ hai là hình ảnh ngoài trời: Điều kiện trời râm, không có đầy đủ về điều kiện ánh sáng, gồm 3 hình ảnh biển số xe trực diện và không trực diện thể hiện tại Hình 5. Các mô hình đều đồng thời nhận diện được 2 biển số xe chỉ có một mô hình nhận diện được 1 biển số đó là YOLO-NAS-S, nhưng có hiện tượng nhận diện sai nhiều đối tượng đối với mô hình YOLOv8l.



Ảnh đầu vào



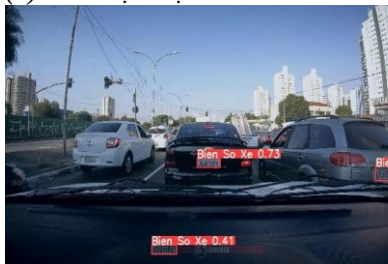
(a) Ảnh nhận diện bởi YOLOv8n



(b) Ảnh nhận diện bởi YOLOv8s



(c) Ảnh nhận diện bởi YOLOv8m



(d) Ảnh nhận diện bởi YOLOv8l



(e) Ảnh nhận diện bởi YOLOv8x



(f) Ảnh nhận diện bởi YOLO-NAS-S



(g) Ảnh nhận diện bởi YOLO-NAS-M



(h) Ảnh nhận diện bởi YOLO-NAS-L

Hình 5. Hình ảnh xe ngoài trời không đầy đủ ánh sáng được nhận diện

Đối với hình ảnh thứ ba là hình ảnh ngoài trời: Điều kiện trời có ánh sáng tốt, hình ảnh có độ sắc nét, gồm 4 biển số xe gồm cả trực diện, che khuất và xéo được thể hiện tại Hình 6. Các mô hình nhận diện thấp nhất 1 biển số và cao nhất 3 biển số xe và có xuất hiện việc nhận diện sai đối tượng đối với mô hình YOLOv8l nhận diện sai 1 đối tượng. Cả hai mô hình YOLOv8m và YOLOv8s có khả năng nhận diện khá kém chỉ đưa ra kết quả 1 biển số xe, mô hình YOLOv8n và YOLO-NAS-S nhận diện được 2 biển số xe. Các mô hình YOLOv8x, YOLO-NAS-M và YOLO-NAS-L đều nhận diện được 3 biển số xe có hình ảnh rõ nét.



Ảnh đầu vào



(a) Ảnh nhận diện bởi YOLOv8n



(b) Ảnh nhận diện bởi YOLOv8s



(c) Ảnh nhận diện bởi YOLOv8m



(d) Ảnh nhận diện bởi YOLOv8l



(e) Ảnh nhận diện bởi YOLOv8x



(f) Ảnh nhận diện bởi YOLO-NAS-S



(g) Ảnh nhận diện bởi YOLO-NAS-M



(h) Ảnh nhận diện bởi YOLO-NAS-L

Hình 6. Hình ảnh xe ngoài trời đầy đủ ánh sáng được nhận diện

Nhận diện của các phiên bản YOLO chủ yếu sẽ phụ thuộc vào các việc tách màu và nhận diện các vật thể thông qua hình ảnh trắng đen. Biển số xe Việt Nam sẽ có hai phần chính là hình dạng (bao gồm hình chữ nhật và hình vuông) và các ký tự do đó các phiên bản YOLO có nhiều tham số sẽ dễ bị nhầm lẫn đối với các biển hiệu hoặc các vật thể có hình dạng tương tự. Điều này sẽ dẫn đến việc độ chính xác của các mô hình trong việc nhận diện sẽ giảm đáng kể so với khả năng thực tế của chúng.

3.3. So sánh thực tế các phiên bản YOLOv8

Việc nhận xét trên, chúng tôi chỉ đưa ra cách xem xét khả năng nhận diện của các mô hình trong một phạm vi nhỏ các hình ảnh phải nhận diện. Vì vậy, chúng tôi đã sử dụng 100 hình ảnh có các điều kiện về ánh sáng, môi trường, độ sắc nét của ảnh, hình ảnh biển số xe trực diện hay không trực diện để làm tiền đề tính các giá trị Accuracy, Precision, Recall và F1score. Thông qua các thông số, có thể đánh giá trực quan nhất khả năng nhận diện của một mô hình được chúng tôi thể hiện ở Bảng 3.

Bảng 3. Các thông số so sánh YOLOv8

Tiêu chí	YOLOv8n	YOLOv8s	YOLOv8m	YOLOv8l	YOLOv8x
<i>Yêu cầu về CPU</i>	Thấp	Trung bình	Trung bình	Cao	Cao nhất
<i>Tốc độ CPU ONNX</i>	80,4 ms	128,4 ms	234,7 ms	375,2 ms	479,1 ms
<i>Accuracy</i>	81,4%	81%	77,27%	79,48%	81,58%
<i>Precision</i>	0,9625	0,9562	0,9107	0,9337	0,9567
<i>Recall</i>	0,8415	0,8361	0,836	0,8424	0,847
<i>F1 score</i>	0,8979	0,8921	0,8718	0,8857	0,8989

Yêu cầu về CPU của mỗi mô hình khác nhau, phụ thuộc nhiều vào kích thước của mô hình. Dựa theo bảng 3, yêu cầu về CPU của các mô hình được sắp xếp theo YOLOv8n, YOLOv8s, YOLOv8m, YOLOv8l và YOLOv8x. Có nghĩa là để có thể chạy mô hình một cách tốt nhất thì YOLOv8n có thể chạy ở các thiết bị CPU phổ biến như Core i5 hoặc AMD Ryzen5. YOLOv8x là mô hình có kích thước lớn nhất trong các thể hệ YOLOv8, do đó có yêu cầu khá cao chủ yếu là các CPU thế hệ mới, tuy các CPU thế hệ cũ có thể chạy được nhưng tốc độ không thể tối ưu.

Về tốc độ: Do các mô hình phụ thuộc nhiều vào tốc độ xung nhịp của CPU và kích thước của mô hình, ở đây chúng tôi xét tốc độ của các mô hình dự đoán trên CPU ONNX. YOLOv8n có tốc độ nhanh nhất 80,4 ms, tăng dần YOLOv8s, YOLOv8m, YOLOv8l và mô hình YOLOv8x có tốc độ chậm nhất. Điều này có nghĩa là YOLOv8n có thể xử lý nhiều ảnh trong một khoảng thời gian nhất định, trong khi YOLOv8x chậm hơn nhưng có thể xử lý ảnh có kích thước lớn và phức tạp hơn.

Về độ chính xác: Xét theo nhà phát triển thì độ chính xác của các mô hình YOLOv8 sẽ giảm dần theo kích thước mô hình. Tuy nhiên ở ngoài thực tế, đối với biển số xe Việt Nam sẽ có thể bị nhầm lẫn với một số biển báo có kí tự tương tự như biển số xe. Do đó, mô hình YOLOv8n (81,4%) và YOLOv8x (81,58%) là hai mô hình có độ chính xác tốt nhất khi so sánh với các phiên bản còn lại.

Xét trên tiêu chí Precision: Các mô hình đều đạt tỉ lệ cao trên 0,9, điều này cho thấy đối với các đối tượng biển số xe các mô hình đều nhận diện đúng nhưng vẫn sẽ có khả năng bỏ sót một số đối tượng.

Xét theo tiêu chí Recall: Các mô hình đều đạt tỉ lệ trên 0,8 các đối tượng biển số xe nhận diện mang tỉ lệ cao nhưng sẽ phụ thuộc vào Precision để đánh giá mô hình có nhận diện đúng các đối tượng phải là biển số xe hay không.

Do đó, để đánh giá tốt nhất ta có thể sử dụng tiêu chí F1 score để có thể đánh giá các mô hình trong việc có cân bằng tốt giữa các chỉ số Precision và Recall.

Như vậy, để có độ chính xác cao thì Precision là yêu cầu quan trọng nhất và trên các mô hình YOLOv8 thì YOLOv8n đạt 0,9625. Tuy nhiên Recall của mô hình lại khá thấp (0,8415) có thể sẽ không thể nhận diện được tất cả các biển số xe xuất hiện. Nếu yêu cầu về độ che phủ cao thì YOLOv8x là ứng cử viên sáng giá nhất, đồng thời đây cũng là mô hình có chỉ số F1 score cao nhất (0,8989).

3.4. So sánh thực tế các phiên bản YOLO-NAS

Sau khi so sánh các phiên bản YOLOv8, chúng tôi sẽ tiếp tục so sánh một số phiên bản YOLO-NAS. Các mô hình sử dụng kiến trúc CNN theo mô hình Darknet53 - NAS phân rõ các nhu cầu về độ chính xác và thời gian nhận diện thông qua việc thể hiện các khối trong kiến trúc

CNN. Thông qua bảng sau, có thể so sánh cụ thể các thông số của mô hình trong việc nhận diện biển số xe được thể hiện ở Bảng 4.

Bảng 4. Các thông số so sánh YOLO-NAS

Tiêu chí	YOLO-NAS-S	YOLO-NAS-M	YOLO-NAS-L
Yêu cầu về CPU	Thấp	Trung bình	Cao
Tốc độ	Nhanh	Trung bình	Chậm
Accuracy	83,92%	83,42%	85,8%
Precision	0,9125	0,9425	0,97
Recall	0,9125	0,8798	0,89
F1 score	0,9125	0,91	0,9287

Các mô hình YOLO-NAS tuy có sự khác biệt về thông số và yêu cầu về CPU tuy nhiên các mô hình YOLO-NAS đều yêu cầu phải có CPU cao mới có thể đáp ứng mô hình chạy tốt nhất.

Về tốc độ: Do các mô hình phụ thuộc nhiều vào tốc độ xung nhịp của CPU và kích thước của mô hình, ở đây chúng tôi xét tốc độ của các mô hình dựa trên CPU ONNX. YOLO-NAS-S có tốc độ nhanh nhất, tăng dần YOLO-NAS-M và mô hình YOLO-NAS-L có tốc độ chậm nhất. Điều này có nghĩa là YOLO-NAS-S có thể xử lý nhiều ảnh trong một khoảng thời gian nhất định, trong khi YOLO-NAS-L chậm hơn nhưng có thể xử lý ảnh có kích thước lớn và phức tạp hơn.

Về độ chính xác: Xét theo nhà phát triển thì độ chính xác của các mô hình YOLO-NAS sẽ giảm dần theo kích thước mô hình. Thực tế, biển số xe Việt Nam có thể bị nhầm lẫn với một số biển báo có kí tự tương tự như biển số xe. Do đó, mô hình YOLO-NAS-L là mô hình có độ chính xác tốt nhất khi so sánh với các phiên bản còn lại.

Để có độ chính xác cao thì Precision là yêu cầu quan trọng nhất và trên các mô hình YOLO-NAS và YOLO-NAS-M đúng nhất đạt (0,9425), tuy nhiên Recall của mô hình lại khá thấp (0,8798) có thể sẽ không thể nhận diện được tất cả các biển số xe xuất hiện. Nếu yêu cầu về độ che phủ cao thì YOLO-NAS-S (0,9125) là ứng cử viên sáng giá nhất và mô YOLO-NAS-L là mô hình có khả năng cân bằng các chỉ số Recall và Precision trên các mô hình.

3.5. Tổng quan so sánh

Trải qua hai lần so sánh và chọn phiên bản có thể phù hợp với việc nhận diện biển xe ở Việt Nam, chúng tôi chọn ra một số phiên bản nổi bật nhất để có thể bình chọn và đưa ra những cách nhìn trực quan nhất trong việc lựa chọn ra phiên bản YOLOv8 và YOLO-NAS. Đối với phiên bản YOLO-NAS-M, chúng tôi không đưa vào bảng so sánh bởi vì tuy YOLO-NAS-M có độ che phủ cao nhất trong các phiên bản YOLO-NAS nhưng các tiêu chí còn lại thì thấp hơn khá nhiều thể hiện rõ ở Bảng 5.

Bảng 5. Tổng quan so sánh các phiên bản tốt nhất

Tiêu chí	YOLOv8n	YOLOv8x	YOLO-NAS-S	YOLO-NAS-L
Accuracy	81,4%	81,58%	83,92%	85,8%
Precision	0,9625	0,9567	0,9125	0,97
Recall	0,8415	0,847	0,9125	0,89
F1 score	0,8979	0,8989	0,9125	0,9287

Đối với các mô hình khác nhau, sẽ có tốc độ đào tạo và nhận diện khác nhau. Tuy nhiên ở thực tế đối với việc đào tạo, các mô hình YOLO-NAS chậm hơn rất nhiều so với các mô hình YOLOv8.

Xét về độ chính xác Accuracy, YOLO-NAS-L có độ chính xác cao nhất, tiếp theo là YOLO-NAS-S. Hai mô hình YOLOv8n và YOLOv8x độ chính xác khá tương đương nhau và thấp nhất trong các phiên bản so sánh.

Xét về các giá trị Precision, Recall, F1 score thì:

- YOLOv8n và YOLOv8x có chỉ số Recall tương đương nhau và độ Precision đều cao hơn 0,95 điều này cho thấy mô hình có thể dự đoán đúng các biển số xe, nhưng sẽ dự đoán thiếu.
- YOLO-NAS-L có chỉ số Precision và Recall khá tương đương nhau nên giá trị F1 score sẽ cao điều này cho thấy mô hình cân bằng tốt khả năng nhận diện biển số xe đúng.
- YOLO-NAS-S có các chỉ số bằng nhau và đều trên 0,9. Mô hình có thể cân bằng gần như tốt tất các chỉ số.

Như vậy nhìn về mặt tổng quan, các mô hình được phù hợp với mỗi yêu cầu về cơ sở hạ tầng khác nhau. Nếu cơ sở hạ tầng tốt và ứng dụng có yêu cầu về độ chính xác cao thì YOLO-NAS-L là một mô hình đáng để tham khảo. Nếu ứng dụng yêu cầu về tốc độ đào tạo và nhận diện cao, thì YOLOv8n là lựa chọn phù hợp.

Đối với việc nhận diện biển số xe Việt Nam thì mô hình YOLO-NAS-L không hẳn là một mô hình phù hợp, bởi vì địa hình Việt Nam có nhiều vật che khuất khác nhau nên độ che phủ là một chỉ số ưu tiên cao hơn nên mô hình YOLO-NAS-S là một mô hình phù hợp nếu yêu cầu về mặt chính xác cao cũng như việc cung cấp cơ sở hạ tầng tốt. Đối với các ứng dụng được sử dụng phổ biến thì việc phát triển dựa trên mô hình YOLOv8n là lựa chọn phù hợp, mô hình cung cấp đủ tốc độ nhận diện cũng như các chỉ số về độ chính xác và độ che phủ đều ở mức khá phù hợp.

4. Kết luận

Trong bài viết này, chúng tôi tiến hành đánh giá hiệu suất toàn diện của các kiến trúc YOLOv8 và YOLO-NAS để phát hiện biển số xe. Qua các số liệu phân tích, chúng tôi thấy rằng hiện tại hai phiên bản YOLO-NAS-S và YOLOv8n là hai phiên bản phù hợp nhất trong việc nhận diện biển số xe, mỗi phiên bản phù hợp cho các nhu cầu về độ chính xác, tốc độ và cơ sở hạ tầng. Đối với các ứng dụng có cơ sở hạ tầng tốt và yêu cầu có độ chính xác cao thì YOLO-NAS-S là một mô hình phù hợp với độ chính xác là 83,92%, Precision là 0,9125; Recall là 0,9125; và F1 score là 0,9125. Đối với cơ sở hạ tầng kém phát triển hơn và yêu cầu về tốc độ thì có thể sử dụng YOLOv8n với số lượng tham số ít hơn nhưng độ chính xác lại khá ổn 81,4%; Precision là 0,9625; Recall 0,8415 và F1 score là 0,8979.

Sau khi huấn luyện và kiểm thử hệ thống, chúng tôi nhận thấy được một số hạn chế: Tốc độ xử lý video và thời gian thực còn chậm do hạn chế về phần cứng, chưa tiếp cận được nguồn dữ liệu video từ các camera bãi đỗ xe ngoài đường để kiểm tra hiệu quả thực tế của mô hình. Hướng phát triển cần cài đặt ứng dụng trên máy tính có phần cứng mạnh hơn để cải thiện khả năng xử lý, sử dụng thêm các kỹ thuật tiền xử lý ảnh để cải thiện khả năng chống nhiễu ở các mô hình, kết nối với các thiết bị quản lý giao thông thực tiễn để tăng tính ứng dụng của hệ thống nhằm phát triển và sử dụng có hiệu quả các mô hình.

TÀI LIỆU THAM KHẢO/ REFERENCES

- [1] P. Viola and M. J. Jones, "Robust Real-Time Face Detection," *International Journal of Computer Vision*, vol. 57, pp. 137–154, 2004, doi.org/10.1023/B:VISI.0000013087.49260.fb.
- [2] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, San Diego, CA, USA, 2005, vol. 1, pp. 886-893, doi: 10.1109/CVPR.2005.177.
- [3] P. Felzenszwalb, D. McAllester, and D. Ramanan, "A discriminatively trained, multiscale, deformable part model," *2008 IEEE Conference on Computer Vision and Pattern Recognition*, Anchorage, AK, USA, 2008, pp. 1-8, doi: 10.1109/CVPR.2008.4587597.
- [4] J. Chen, Z. Liu, H. Wang, A. Núñez, and Z. Han, "Automatic Defect Detection of Fasteners on the Catenary Support Device Using Deep Convolutional Neural Network," *IEEE Transactions on Instrumentation and Measurement*, vol. 67, no. 2, pp. 257-269, Feb. 2018, doi: 10.1109/TIM.2017.2775345.

-
- [5] Q. Guo, L. Liu, W. Xu, Y. Gong, X. Zhang, and W. Jing, "An Improved Faster R-CNN for High-Speed Railway Dropper Detection," *IEEE Access*, vol. 8, pp. 105622-105633, 2020, doi: 10.1109/ACCESS.2020.3000506.
- [6] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137-1149, June 2017, doi: 10.1109/TPAMI.2016.2577031.
- [7] V. L. Trinh, T. L. T. Dao, X. T. Le, and E. Castelli, "Emotional Speech Recognition Using Deep Neural Networks," *Sensors*, vol. 22, 2022, Art. no. 1414, doi: 10.3390/s22041414.
- [8] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016, pp. 779-788, doi: 10.1109/CVPR.2016.91.
- [9] W. Liu, et al., "SSD: Single Shot MultiBox Detector," in *Computer Vision – ECCV 2016. ECCV 2016, Lecture Notes in Computer Science*, vol. 9905, B. Leibe, J. Matas, N. Sebe, and M. Welling, (eds). Springer, Cham, 2016, doi.org/10.1007/978-3-319-46448-0_2.
- [10] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal Loss for Dense Object Detection," *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 2980-2988, doi: 10.48550/arXiv.1708.02002.
- [11] S. S. Patil et al., "Vehicle Number Plate Detection using YoloV8 and EasyOCR," *2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, Delhi, India, 2023, pp. 1-4, doi: 10.1109/ICCCNT56998.2023.10307420.
- [12] D. Mane, P. Kumbharkar, S. Sangve, N. Earan, K. Patil, and S. Bonde, "A Metaphor Analysis on Vehicle License Plate Detection using Yolo-NAS and Yolov8," *Journal of Electrical Systems*, vol. 20, no. 1s, pp. 152-164, 2024, doi:10.52783/jes.761.
- [13] D. L. Nguyen, V. N. Nguyen, and V. D. Dang, "Building a License Plate Recognition Application with OpenCV Open Source" (in Vietnamese), *Quang Binh Journal for Science and Technology Information*, vol. 1, pp. 35-40, 2018.
- [14] T.-A. Dat, K. L. Tran, and H.-N. Vu, "License Plate Recognition Based on MULTI-ANGLE View Model," *arXiv:2309.12972*, 2023, doi: 10.48550/arXiv.2309.12972.
- [15] H. Shi and D. Zhao, "License Plate Recognition System Based on Improved YOLOv5 and GRU," *IEEE Access*, vol. 11, pp. 10429-10439, 2023, doi: 10.1109/ACCESS.2023.3240439.
- [16] R. Laroca, L. A. Zanlorensi, G. R. Gonçalves, E. Todt, W. R. Schwartz, and D. Menotti, "Anefficient and layout-independent automatic license platerecognition system based on the YOLO detector," *IETIntell Transp Syst.*, vol. 15, pp. 483–503, 2021, doi: 10.1049/itr2.12030.
- [17] S. Pan, J. Liu, and D. Chen, "Research on License Plate Detection and Recognition System based on YOLOv7 and LPRNet," *Academic Journal of Science and Technology*, vol. 4, no. 2, pp. 62-68, 2023, doi: 10.54097/ajst.v4i2.3971
- [18] S. Mejdil, A. Abdulmalik, and A. Sultan, "Efficient Multistage License Plate Detection and Recognition Using YOLOv8 and CNN for Smart Parking Systems," *Journal of sensors*, 2024, doi: 10.1155/2024/4917097.
- [19] D. A. Mulia, S. Safitri, and G. P. K. Negara, "YOLOv8 and Faster R-CNN Performance Evaluation with Super-resolution in License Plate Recognition," *International Journal of Computing and Digital Systems*, 2024, doi: 10.12785/ijcds/160129.