

ỨNG DỤNG CÁC MÔ HÌNH MẠNG HỌC SÂU NHẬN DẠNG BIỂN SỐ XE Ô TÔ TỰ ĐỘNG TOWARDS AUTOMATIC CAR LICENSE PLATE NUMBER RECOGNITION USING DEEP LEARNING MODELS

NGUYỄN HỮU TUÂN*, NGUYỄN DUY TRƯỜNG GIANG

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

*Email liên hệ: huu-tuan.nguyen@vimaru.edu.vn

DOI: <https://doi.org/10.65154/jmst.1049>

Tóm tắt

Nhận dạng biển số xe ô tô tự động là một bài toán của lĩnh vực thị giác máy tính có tính khoa học và thực tiễn cao. Giải quyết tốt bài toán này là nền tảng cho việc chuyển đổi số trong lĩnh vực giao thông đường bộ và nhiều bài toán về quản lý giao thông đô thị khác như giám sát phương tiện, phạt nguội, truy vết phương tiện. Bài báo này nghiên cứu và đề xuất ứng dụng mô hình mạng YOLOv12 và PaddleOCR v3.3.2 để nhận dạng tự động biển số xe ô tô Việt Nam. Ở bước đầu tiên, mô hình YOLOv12, một phiên bản cải tiến từ các phiên bản trước của họ các mô hình YOLO với cơ chế chú ý sẽ được sử dụng để phát hiện biển số xe ô tô. Sau đó phần biển số xe sẽ được nhận dạng bằng PaddleOCR v3.3.2. Kết quả huấn luyện và kiểm thử trên tập dữ liệu biển số xe ô tô Việt Nam gồm 25.364 bức ảnh cho phần phát hiện biển số và 6.340 bức ảnh cho phần nhận dạng biển số cho thấy phương pháp đề xuất có độ chính xác trong phát hiện và nhận dạng rất tốt khi chỉ số phát hiện biển số $mAP@0.5$ là 99,3% và độ chính xác của nhận dạng biển số lên đến 89,21%. Kết quả khi đánh giá với các video thử nghiệm thu thập được từ thực tế cũng cho kết quả phát hiện và nhận dạng tốt và hoàn toàn có thể áp dụng vào các tình huống thực tế.

Từ khóa: Nhận dạng biển số xe ô tô, phát hiện biển số xe, nhận dạng ký tự, YOLOv12, OCR.

Abstract

Automated License Plate Recognition (ALPR) constitutes a significant problem within the field of computer vision, possessing both high scientific and practical relevance. Successfully addressing this challenge forms a foundation for digital transformation within the road transport sector and enables further advancements in urban traffic management applications such as vehicle

monitoring, automated enforcement (e.g., photo-based ticketing), and vehicle tracking. This paper investigates and proposes an application of the YOLOv12 and PaddleOCR v3.3.2 models for automated Vietnamese license plate recognition. Firstly, a YOLOv12 model—an improved version leveraging attention mechanisms over previous YOLO architectures—is employed to detect license plates within images. Then, the detected license plate regions are recognized using PaddleOCR v3.3.2. Training and testing results on a dataset comprising 25,364 images for license plate detection and 6,340 images for license plate recognition demonstrate excellent performance in both detection and recognition tasks. The mean Average Precision (mAP) at IoU threshold of 0.5 for license plate detection reaches 99.3%, while the license plate recognition accuracy achieves a precision of 89.21%. Evaluation using real-world video sequences further validates robust performance, indicating suitability for practical deployment in diverse operational scenarios.

Keywords: Car license plate recognition, license plate detection, YOLOv12, OCR.

1. Mở đầu

Nhận dạng biển số xe ô tô tự động là một bài toán thực tế và là cơ sở cho việc giải quyết nhiều vấn đề thực tế liên quan. Xét trong bối cảnh tốc độ đô thị hóa tăng nhanh, dân cư tại các đô thị lớn ngày càng đông đúc và tỉ lệ người dân sở hữu ô tô ngày càng tăng thì những vấn đề cần giải quyết liên quan đến việc nhận dạng chính xác biển số xe ô tô là một nhu cầu cấp thiết. Các bài toán thực tế cần giải quyết dựa trên nhận dạng biển số xe ô tô bao gồm: quản lý ra vào bãi xe công cộng hay trong các khu dân cư, phạt nguội các chủ xe có hành vi vi phạm luật Giao thông đường bộ như vượt đèn đỏ, đi sai làn, gây tai nạn giao thông, dừng đỗ không đúng vị trí hoặc xác định thông tin xe và chủ

xe trong các tình huống giao thông dựa trên hình ảnh thu được từ camera giám sát. Hiện nay tại các đô thị ở Việt Nam, các camera giám sát được lắp đặt khá nhiều tại các điểm nút giao thông, các tuyến đường đô thị và các khu dân cư. Đây là hệ thống thu thập dữ liệu hình ảnh có thể sử dụng làm nền tảng cho nhiều bài toán dựa trên các mô hình học sâu của thị giác máy tính, trong đó có bài toán nhận dạng biển số xe ô tô tự động. Việc nhận dạng biển số xe ô tô tự động cũng có thể giúp lực lượng công an giao thông, các đơn vị quản lý hành chính nhà nước ở cấp Phường, Xã, Thành phố thu thập dữ liệu để làm cơ sở cho việc giải quyết các vấn đề phát sinh liên quan đến chủ xe sau này một cách hiệu quả và chủ động thay cho việc rà soát thủ công hoặc kiểm soát thủ công bằng con người qua việc quan sát bằng mắt thường tại các điểm giám sát.

Bài báo này tập trung nghiên cứu ứng dụng mô hình mạng học sâu YOLOv12 và công cụ PaddleOCR v3.3.2 để phát hiện biển số xe và nhận dạng biển số xe ô tô tự động từ các hình ảnh thu được qua các camera giám sát có thể đặt tại các điểm trên đường giao thông hoặc tại các ngã tư giao cắt hay các khu vực cần quan sát khác như bãi xe, khu vực cổng vào ra. Cách tiếp cận này không đòi hỏi phải trang bị các camera giám sát chuyên dụng mà có thể tận dụng các camera thông thường đã được lắp đặt sẵn kết hợp với các thiết bị biên hoặc máy tính để triển khai thực tế mà không làm phát sinh chi phí đầu tư quá lớn.

Trong [1], Dong và các cộng sự đã sử dụng một mô hình mạng đề xuất vùng (region proposal network - RPN) kết hợp với mô hình mạng R-CNN truyền thống [2] để phát hiện biển số xe và sau đó áp dụng đồng thời nhiều mô hình mạng CNN nhỏ để nhận dạng biển số xe và đạt được độ chính xác 89.1% trên tập ảnh test có 5.999 ảnh. Laroca và các cộng sự đã áp dụng mô hình mạng YOLOv2 để xây dựng một hệ thống nhận dạng biển số xe theo thời gian thực với độ chính xác 78,33% và tốc độ 35 ảnh trong 1 giây (frames per second - FPS) trong [3]. Một cách tiếp cận khác dựa trên các thuật toán xử lý ảnh để phát hiện biển số xe và thuật toán phân lớp rừng ngẫu nhiên (Random Forest) để nhận dạng biển số đã được đề xuất bởi Akhtar và các cộng sự [4]. Mặc dù có độ chính xác tới 90.9% nhưng cách tiếp cận này sẽ không đảm bảo độ ổn định và tính chính xác với các ảnh chụp ở các điều kiện ánh sáng và góc chụp khác nhau. Một mô hình nhẹ của kiến trúc mạng CNN [5] và Tesseract [6] đã được đề xuất cho phát hiện và nhận dạng biển số xe trong [7] với độ chính xác 90% trên tập test gồm 1000 ảnh. Mô hình YOLOv5 [8] và một mạng CNN sửa đổi đã được áp dụng bởi Khan và các cộng sự

trong [9] và đạt được độ chính xác lên đến 92,8% trên một tập dữ liệu ảnh thu thập qua camera. Một cách tiếp cận khác [10] sử dụng mô hình Faster R-CNN [11] để phát hiện biển số và MobileNetv3 [12] để nhận dạng biển số với độ chính xác 92%. Gần đây, một phiên bản khác của họ các mô hình YOLO là YOLOv8 [13] cũng được đề xuất để phát hiện và nhận dạng biển số xe ô tô bởi Vargoorani và các cộng sự [14] và đạt được độ chính xác khá cao (94%).

Có thể thấy, điểm chung trong các cách tiếp cận để giải quyết bài toán nhận dạng biển số xe ô tô tự động là sử dụng một mô hình mạng học sâu để phát hiện biển số (có thể là mô hình hai bước dựa trên mô hình R-CNN hoặc các biến thể của nó, hoặc mô hình một bước là các biến thể của YOLO), sau đó sử dụng các mô hình CNN hoặc các thuật toán học máy để nhận dạng ký tự trong vùng biển số phát hiện được. So với các mô hình hai bước dựa trên R-CNN và các biến thể của nó, cách tiếp cận sử dụng các mô hình một bước dựa trên các biến thể của mạng YOLO có ưu điểm về tốc độ huấn luyện và triển khai trong khi vẫn đạt được độ chính xác phát hiện tốt.

YOLOv12, một phiên bản mới của họ các mô hình YOLO được đề xuất mới đây với nhiều điểm mới, đặc biệt là cơ chế chú ý theo vùng (area attention), đã đạt được các kết quả tốt hơn so với các phiên bản trước của YOLO với bài toán phát hiện đối tượng trong khi vẫn giữ được tốc độ cao để có thể áp dụng vào các bài toán đòi hỏi thời gian thực. Bên cạnh đó PaddleOCR v3.3.2, một phiên bản thế hệ mới của công cụ PaddleOCR với mô hình nhận dạng ký tự PP-OCRv5, cũng cho các kết quả vượt trội đối với các bài toán nhận dạng ký tự quang học từ dữ liệu hình ảnh. Trong bài báo này, chúng tôi đề xuất áp dụng các mô hình mạng học sâu YOLOv12 và PP-OCRv5 cho bài toán nhận dạng biển số xe ô tô tự động. Cấu trúc của bài báo được tổ chức thành các phần như sau: Phần 2 sẽ trình bày về mô hình YOLOv12 và công cụ PaddleOCR v3.3.2, phần 3 trình bày việc áp dụng hai công cụ trên vào bài toán nhận dạng biển số xe ô tô tự động với các kết quả huấn luyện mô hình, phát hiện biển số và phân tích các kết quả thực nghiệm, phần 4 là Kết luận.

2. Mô hình mạng YOLOv12 và công cụ PaddleOCR v3.3.2

2.1. Mạng học sâu YOLOv12

Các mô hình YOLO (You Only Look Once) [15] là một cách tiếp cận phổ biến và hiệu quả hàng đầu cho bài toán phát hiện đối tượng kể từ khi được Redmon và các cộng sự giới thiệu vào năm 2016. Các

mô hình YOLO hoạt động theo cơ chế một bước (Look Once) bằng cách sử dụng duy nhất 1 mô hình mạng CNN xương sống cho việc dự đoán bounding box và xác suất của đối tượng phát hiện nhằm đảm bảo thời gian phát hiện nhanh, sử dụng ít tài nguyên mà vẫn đảm bảo độ chính xác cao. Các mô hình mạng YOLO đã phát triển từ phiên bản YOLOv1 ban đầu đến YOLOv11 với mục tiêu cân bằng giữa độ chính xác và hiệu quả sử dụng tài nguyên. Trong khi các mô hình ban đầu (từ v1 đến v6) sử dụng các mạng xương sống mạnh cho việc trích chọn đặc trưng đa tỉ lệ, thì các mô hình sau này (từ v7 đến v11) đã đi theo hướng tích hợp các kiến trúc mới, cơ chế dự đoán không cần mỏ neo (anchor-free), và các thành phần dựa trên cơ chế transformer để nâng cao khả năng ứng dụng của chúng trong các bài toán có độ phức tạp ngày càng cao.

YOLOv12 [16] đánh dấu một bước phát triển đột phá của họ các mô hình YOLO. Các phiên bản trước của YOLO chủ yếu dựa trên các mạng CNN làm xương sống mà chưa có sự ứng dụng của cơ chế chú ý (attention) một cách hiệu quả. Lý do là cơ chế chú ý mặc dù có khả năng mô hình hóa cao trong việc trích chọn đặc trưng nhưng lại có rào cản là độ phức tạp tính toán cao và kém hiệu quả trong việc truy cập bộ nhớ dẫn tới quá chậm trong các tác vụ đòi hỏi tốc độ thời gian thực. YOLOv12 đã giải quyết vấn đề này bằng cách chuyển sang một thiết kế lấy sự chú ý làm trung tâm (attention-centric design) giúp hệ thống phát hiện đối tượng đạt được tốc độ như các mô hình dựa trên mạng CNN trong khi lại khai thác được lợi thế của cơ chế chú ý trong việc trích chọn đặc trưng

hiệu quả hơn.

Ba điểm mới của YOLOv12 so với các phiên bản trước đó là: 1) mô đun chú ý theo vùng (area attention), 2) mạng tổng hợp đặc trưng hiệu quả có kết nối tắt R-ELAN và 3) tối ưu hóa kiến trúc với các phép nhân chập tách biệt kích thước 7x7.

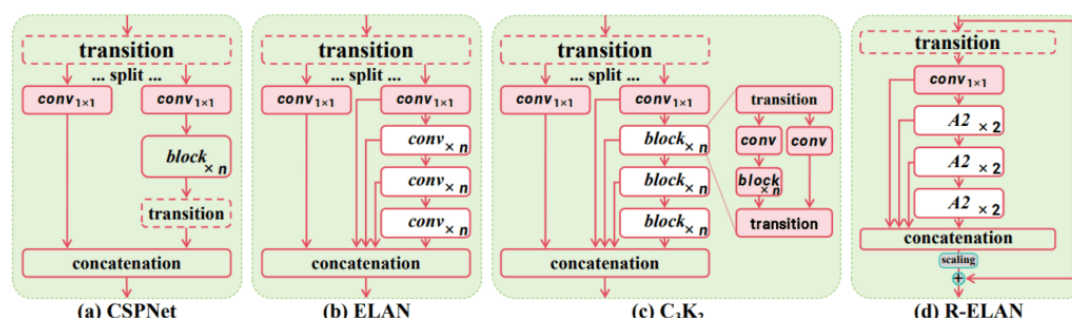
Mô đun chú ý theo vùng (area attention) (xem chi tiết hơn ở Hình 1): Thay vì tính toán attention toàn cục, mô đun này chia các bản đồ đặc trưng thành các phân đoạn, cho phép mô hình tập trung hiệu quả vào các vùng quan trọng nhằm giảm đáng kể độ phức tạp tính toán trong khi vẫn duy trì được hiệu quả trong việc trích chọn đặc trưng từ ảnh.

Mạng tổng hợp đặc trưng hiệu quả có kết nối tắt R-ELAN (xem chi tiết hơn ở Hình 2): Một bản thiết kế mới (cải tiến mô hình ELAN trong các phiên bản YOLO trước đây) của mạng xương sống sử dụng các kết nối tắt sâu (deep residual connection) để giải quyết các vấn đề về nghẽn gradient và cải thiện khả năng dung hợp các đặc trưng, một vấn đề thường gặp khi áp dụng cơ chế attention. Ở đây YOLOv12 sử dụng các kết nối tắt từ đầu vào đến đầu ra để cải thiện độ ổn định và thay vì chia nhỏ đầu ra sau lớp chuyển tiếp thì điều chỉnh kích thước kênh và tạo ra một bản đồ đặc trưng duy nhất.

Tối ưu hóa kiến trúc: YOLOv12 tích hợp cơ chế Flash Attention [17] để vượt qua vấn đề nút thắt truy cập bộ nhớ, sử dụng các phép nhân chập tách biệt (separable convolution) với kích thước nhân 7x7 để bảo toàn ngữ cảnh không gian mà không cần dựa vào các phương pháp mã hóa vị trí truyền thống.



Hình 1. Mô đun chú ý theo vùng của YOLOv12



Hình 2. Kiến trúc R-ELAN của YOLOv12

2.2. PaddleOCR v3.3.2 nhận dạng ký tự quang học

PaddleOCR v3.3.2 [18] là một thư viện nhận dạng ký tự thể hệ tiếp theo của nền tảng Paddle với tốc độ nhanh và độ chính xác cao trong các bài toán có liên quan đến nhận dạng ký tự quang học (gọi chung là OCR - Optical Character Recognition) trong đó có bài toán nhận dạng biển số xe ô tô. Ở phiên bản v3.3.2, PaddleOCR giới thiệu mô hình kiến trúc PP-OCR v5 có cấu trúc siêu nhẹ (khoảng 0,07 tỉ tham số) với các ưu điểm chính sau: Nhận dạng đa ngôn ngữ trong một mô hình duy nhất; cung cấp các mô hình khác nhau phù hợp với các nền tảng phần cứng khác nhau từ các hệ thống di động tới các hệ thống được trang bị các card đồ họa GPU mạnh mẽ nhằm đạt được hiệu năng cao nhất; độ chính xác rất cao so với các phiên bản trước đó và thậm chí cả các mô hình ngôn ngữ thị giác VLM [19].

3. Nhận dạng biển số xe ô tô tự động sử dụng YOLOv12 và Paddle OCR v3.3.2

Để xây dựng một hệ thống nhận dạng biển số xe có độ chính xác cao và tốc độ nhanh nhằm ứng dụng

và các tình huống đòi hỏi xử lý theo thời gian thực, chúng tôi đề xuất ứng dụng mô hình YOLOv12 để phát hiện biển số và PaddleOCR v3.3.2 để nhận dạng biển số xe sau khi phát hiện. Khi đó hệ thống nhận dạng biển số xe ô tô tự động sẽ gồm có 2 bước chính: Phát hiện biển số bằng YOLOv12 và nhận dạng biển số bằng PaddleOCR v3.3.2.

Để huấn luyện mô hình YOLOv12 cho bước phát hiện biển số xe ô tô, nhóm tác giả đã thu thập 1 tập dữ liệu ảnh huấn luyện gồm các bức ảnh chụp (qua thiết bị di động, camera tại các bãi gửi xe,...) ô tô Việt Nam với 25.364 ảnh (xem Hình 3 để biết thêm chi tiết), và 1 tập ảnh kiểm thử gồm 6.340 ảnh. Tất cả các ảnh đều được huấn luyện với độ phân giải 640x640.

Với mục đích hướng tới ứng dụng hệ thống vào thực tế nên chúng tôi chỉ huấn luyện các phiên bản siêu nhỏ (nano), nhỏ (small) và vừa (medium) của YOLOv12 trên một máy trạm được trang bị 32 GB RAM, 1 card GPU RTX 5060 16 GB và phiên bản Pytorch 2.9 trên nền tảng CUDA 13.0. Các mô hình đều được huấn luyện với 200 epoch. Tham số học learning_rate được thiết lập bằng 0,001, sử dụng



Hình 3. Một số ảnh biển số xe ô tô trong tập huấn luyện

Bảng 1. Kết quả huấn luyện mô hình YOLOv12 phát hiện biển số xe ô tô

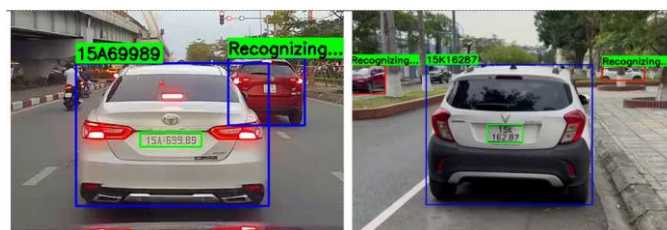
TT	Phiên bản	Kết quả huấn luyện		Kết quả với dữ liệu kiểm tra	
		mAP@0.50	mAP@0.50-0.95	mAP@0.50	mAP@0.5-0.95
1	YOLOv12n	99.5%	83.3%	99.3%	83.1%
2	YOLOv12s	99.5%	83.4%	99.3%	83.2%
3	YOLOv12m	99.5%	83.5%	99.3%	83.2%

Bảng 2. Kết quả nhận dạng biển số sau phát hiện

TT	Công cụ OCR	Độ chính xác (%)
1	PaddleOCR	89.21% (5630/6311)
2	EasyOCR	25.97% (1639/6311)
3	TesseractOCR	12.77% (806/6311)



Hình 4. Một số ảnh có kết quả nhận dạng biển số sai



Hình 5. Một số kết quả nhận dạng biển số xe với video thực tế

phương pháp tối ưu Adam và giá trị batch_size lần lượt là 8, 32, 48 cho các mô hình phiên bản Nano, Small và Medium.

Kết quả phát hiện biển số: Kết quả huấn luyện YOLOv12 phát hiện biển số được trình bày trong Bảng 1 (theo chỉ số độ chính xác trung bình mAP - mean Average Precision) cho thấy các mô hình YOLOv12 đều có khả năng phát hiện biển số tốt khi đạt được mAP@0.5-0.95 cao nhất là 83,2% và mAP@0.5 cao nhất là 99,3% trên tập kiểm thử với mô hình YOLOv12m. Các mô hình siêu nhỏ và nhỏ có kém hơn 1 chút nhưng không đáng kể khi chỉ suy giảm khoảng 0,1% so với mô hình trung bình với cả hai chỉ số mAP@0.5-0.95 và mAP@0.5.

Kết quả nhận dạng biển số: Mô hình YOLOv12m sau khi huấn luyện được sử dụng với một tập test gồm 6.340 bức ảnh (không trùng với tập huấn luyện) để thực hiện phát hiện biển số và kết quả là YOLOv12m phát hiện được 6.311/6.340 ảnh (tỉ lệ phát hiện chính xác biển số đạt 99,5%). Biển số xe (6.311 ảnh) sau khi phát hiện được đưa qua công cụ OCR để nhận dạng với 3 công cụ là EasyOCR, TesseractOCR và PaddleOCR v3.3.2. Bảng 2 là kết quả nhận dạng biển số xe. Nhìn vào bảng 2, có thể thấy PaddleOCR v3.3.2 cho kết quả tốt hơn nhiều so với các công cụ TesseractOCR và EasyOCR khi đạt được độ chính xác cao nhất là 89,21% trong khi EasyOCR là 25,97% còn TesseractOCR là 12,77%. Với độ chính xác 89,21% (89,21% là độ chính xác khớp hoàn toàn toàn bộ biển số, tức là phải nhận diện đúng 100% tất cả các ký tự, sai 1 ký tự coi như sai toàn bộ), hệ thống nhận dạng biển số xe ô tô đề xuất trong bài báo này hoàn toàn có khả năng triển khai vào các ứng dụng thực tế. Trong

Hình 4 làm một số ảnh bị nhận dạng sai, có thể thấy các ảnh bị nhận dạng sai có các đặc điểm chung sau: 1) Phần biển số có kích thước bé và mờ, 2) Góc chụp quá nghiêng và 3) Biển bị che khuất 1 phần.

Để đánh giá với dữ liệu thực, 4 video quay từ môi trường thực tế đã được thử nghiệm và kết quả hệ thống đã phát hiện và nhận dạng biển số khá chính xác (xem chi tiết trong Hình 5). Với các biển số được phát hiện ở góc nghiêng không quá lớn và không bị che khuất thì kết quả chính xác, tuy nhiên với các góc nghiêng lớn và bị che khuất 1 phần biển số thì hệ thống vẫn cho kết quả chưa chính xác (tương tự như các trường hợp sai của tập ảnh đã test). Tốc độ vừa phát hiện và nhận dạng với mô hình YOLOv12m là 21 frame trong 1 giây, tốc độ này đảm bảo cách tiếp cận bài báo đề xuất có khả năng đáp ứng việc triển khai theo thời gian thực.

4. Kết luận

Hướng tới mục tiêu ứng dụng các mô hình mạng học sâu để giải quyết bài toán nhận dạng biển số xe ô tô tự động tại Việt Nam, bài báo này đề xuất áp dụng mô hình học sâu YOLOv12 để phát hiện biển số và PaddleOCR v3.3.2 để nhận dạng biển số sau khi phát hiện. Ba phiên bản khác nhau (siêu nhỏ, nhỏ và trung bình) của YOLOv12 đã được huấn luyện trên một tập dữ liệu hình ảnh gồm 25.364 ảnh và đạt được kết quả phát hiện rất tốt (tính theo chỉ số mAP@0.5 là 99,3% và theo mAP@0.5-0.95 là 83,2% trên tập ảnh test gồm 6.340 ảnh). Tỉ lệ nhận dạng đúng các ảnh biển số sau khi phát hiện với PaddleOCR v3.3.2 là 89,21%. Các kết quả này chứng tỏ hệ thống nhận dạng biển số xe ô tô tự động đề xuất trong bài báo này là đúng đắn và có khả năng áp dụng được vào các bài toán thực tế.

Trong thời gian tới, hướng phát triển tiếp theo của bài báo sẽ là cải thiện độ chính xác của việc nhận dạng biển số sau khi phát hiện bằng YOLOv12 với các mô hình VLM hiện đại và đưa hệ thống vào các ứng dụng thực tế.

Lời cảm ơn

Nghiên cứu này được tài trợ bởi Trường Đại học Hàng hải Việt Nam trong đề tài mã số: **DT25-26.81**.

TÀI LIỆU THAM KHẢO

- [1] Dong, M., He, D., Luo, C., Liu, D. & Zeng, W. (2017), *A CNN-Based Approach for Automatic License Plate Recognition in the Wild*, *Proceedings of the British Machine Vision Conference 2017*, London, UK: British Machine Vision Association, p.175.
 Doi: 10.5244/C.31.175.
- [2] Girshick, R., Donahue, J., Darrell, T. & Malik, J. (2014), *Rich feature hierarchies for accurate object detection and semantic segmentation*, arXiv:1311.2524.
 Doi: 10.48550/arXiv.1311.2524.
- [3] Laroca, R. et al. (2018), *A Robust Real-Time Automatic License Plate Recognition Based on the YOLO Detector*, *2018 International Joint Conference on Neural Networks (IJCNN)*, pp. 1-10.
 Doi: 10.1109/IJCNN.2018.8489629.
- [4] Akhtar, Z. & Ali, R. (2020), *Automatic Number Plate Recognition Using Random Forest Classifier*, *SN Computer Science*, Vol. 1(3), p. 120.
 Doi: 10.1007/s42979-020-00145-8.
- [5] LeCun, Y., Bottou, L., Bengio, Y. & Ha, P. (1998), *Gradient-Based Learning Applied to Document Recognition*.
- [6] Tesseract-ocr/tesseract (2026), *C++. tesseract-ocr*, Retrieved from <https://github.com/tesseract-ocr/tesseract>.
- [7] Velarde, M. D. C. & Velarde, G. (2022), *Benchmarking Algorithms for Automatic License Plate Recognition*, arXiv:2203.14298.
 Doi: 10.48550/arXiv.2203.14298.
- [8] Ultralytics (2024), *YOLOv5: A state-of-the-art real-time object detection system*, Retrieved from <https://github.com/ultralytics/yolov5>.
- [9] Khan, I. R. et al. (2022), *Automatic License Plate Recognition in Real-World Traffic Videos Captured in Unconstrained Environment by a Mobile Camera*, *Electronics*, Vol. 11(9), p.1408.
 Doi: 10.3390/electronics11091408.
- [10] Vargoorani, Z. E. & Suen, C. Y. (2024), *License Plate Detection and Character Recognition Using Deep Learning and Font Evaluation*, Vol. 15154, pp. 231-242.
 Doi: 10.1007/978-3-031-71602-7_20.
- [11] Ren, S., He, K., Girshick, R. & Sun, J. (2016), *Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks*, arXiv:1506.01497.
 Doi: 10.48550/arXiv.1506.01497.
- [12] Howard, A. et al. (2019), *Searching for MobileNetV3*, arXiv:1905.02244.
 Doi: 10.48550/arXiv.1905.02244.
- [13] Jocher, G., Chaurasia, A. & Qiu, J. (2023), *Ultralytics YOLOv8*. Python, Retrieved from <https://github.com/ultralytics/ultralytics>.
- [14] Vargoorani, Z. E., Ghoreyshi, A. M. & Suen, C. Y. (2025), *Efficient License Plate Recognition via Pseudo-Labeled Supervision with Grounding DINO and YOLOv8*, *2025 IEEE 35th International Workshop on Machine Learning for Signal Processing (MLSP)*, pp. 1-6.
 Doi: 10.1109/MLSP62443.2025.11204315.
- [15] Redmon, J., Divvala, S., Girshick, R. & Farhadi, A. (2016), *You Only Look Once: Unified, Real-Time Object Detection*, arXiv:1506.02640.
 Doi: 10.48550/arXiv.1506.02640.
- [16] Tian, Y., Ye, Q. & Doermann, D. (2025), *YOLOv12: Attention-Centric Real-Time Object Detectors*, arXiv:2502.12524.
 Doi: 10.48550/arXiv.2502.12524.
- [17] Dao, T., Fu, D. Y., Ermon, S., Rudra, A. & Ré, C. (2022), *FlashAttention: Fast and Memory-Efficient Exact Attention with IO-Awareness*, arXiv:2205.14135.
 Doi: 10.48550/arXiv.2205.14135.
- [18] Cui, C. et al. (2025), *PaddleOCR v3.3.2.0 Technical Report*, arXiv:2507.05595.
 Doi: 10.48550/arXiv.2507.05595.
- [19] Bordes, F. et al. (2024), *An Introduction to Vision-Language Modeling*, arXiv:2405.17247.
 Doi: 10.48550/arXiv.2405.17247.

Ngày nhận bài:	11/03/2026
Ngày nhận bản sửa:	25/03/2026
Ngày duyệt đăng:	07/04/2026