

Applying adaptive deep learning model for detecting traffic vehicles

Article info

Type of article:

Original research paper

DOI:

<https://doi.org/10.58845/jstt.utt.2023.vn.3.2.38-46>

*Corresponding author:

E-mail address:

nxthang@hanu.edu.vn

Received: 26/02/2023

Accepted: 08/06/2023

Published: 28/06/2023

Thang Xuan Nguyen*
Hanoi University, Ha Noi, Viet Nam

Abstract: Image analysis for detecting traffic vehicles is a problem in the field of computer vision. This problem has many useful applications in self-driving vehicle systems, traffic management, and vehicle flow measurement at important locations and routes. There are many approaches to solve this problem, such as contour representation, feature extraction, machine learning, and deep learning networks. In this paper, the author proposes a solution using an adaptive learning model based on a deep learning network to solve the problem. To evaluate the effectiveness of the solution, the author built a testing system based on the Darknet-53 deep learning network. The system was tested on both standard and self-collected datasets. The results show that the system achieves high accuracy and feasibility when applied to real-world applications.

Keywords: Vehicle detection, vehicle classification, adaptive learning, deep learning, computer vision.

Thông tin bài viết
Dạng bài viết:
Bài báo nghiên cứu

DOI:
<https://doi.org/10.58845/jstt.utt.2023.vn.3.2.38-46>

*Tác giả liên hệ:
Địa chỉ E-mail:
nxthang@hanu.edu.vn

Ngày nộp bài: 26/02/2023
Ngày chấp nhận: 08/06/2023
Ngày đăng bài: 28/06/2023

Ứng dụng mô hình học sâu thích ứng trong bài toán phát hiện phương tiện giao thông

Nguyễn Xuân Thắng*
Trường Đại học Hà Nội, Hà Nội, Việt Nam

Tóm tắt: Phân tích hình ảnh để phát hiện phương tiện giao thông là một bài toán trong lĩnh vực thị giác máy tính. Bài toán này có nhiều ứng dụng hữu ích trong các hệ thống xe tự hành, quản lý giao thông và đo lưu lượng xe tại các địa điểm, các tuyến đường quan trọng. Có nhiều cách tiếp cận để giải quyết bài toán này như biểu diễn đường viền, trích chọn đặc trưng, học máy, mạng học sâu. Trong bài báo này, tác giả đề xuất giải pháp sử dụng mô hình học thích ứng trên nền mạng học sâu để giải quyết bài toán. Để đánh giá hiệu quả của giải pháp, tác giả đã xây dựng hệ thống thử nghiệm dựa trên mạng học sâu YOLO3. Hệ thống được thử nghiệm trên bộ dữ liệu chuẩn và bộ dữ liệu tự thu thập. Kết quả cho thấy, hệ thống đạt được độ chính xác cao và khả thi khi áp dụng vào các ứng dụng thực tế.

Từ khóa: Phát hiện phương tiện giao thông, mạng học sâu, học thích ứng, thị giác máy tính.

1. Giới thiệu

Phát hiện phương tiện giao thông là một bài toán nằm trong nhóm các bài toán phát hiện đối tượng, là một lĩnh vực con của ngành thị giác máy tính và có rất nhiều ứng dụng trong thực tế. Vì vậy, vấn đề này đã thu hút sự quan tâm của các nhà khoa học và các hãng sản xuất công nghiệp lớn. Bài toán phát hiện phương tiện giao thông được ứng dụng chủ yếu trong việc phát triển các hệ thống lái tự động và kiểm soát giao thông (xác định lưu lượng giao thông tại các điểm nút để điều chỉnh, phân luồng và quy hoạch hạ tầng giao thông, phát hiện các phương tiện vi phạm trong quá trình tham gia giao thông ...).

Phát hiện phương tiện giao thông là một bài toán khó và phức tạp khi cần xác định vị trí và phân loại các phương tiện giao thông. Một hệ thống phát hiện phương tiện giao thông bao gồm ba bước xử lý chính như mô tả trong Hình 1:

Bước 1: Thu nhận ảnh từ các hệ thống camera giao thông và thực hiện tiền xử lý dữ liệu

đầu vào.

Bước 2: Sử dụng một mô hình phát hiện đã được huấn luyện để phát hiện và trả về kết quả bao gồm hộp giới hạn và đối tượng xuất hiện trong các hộp giới hạn đó.

Bước 3: Thực hiện hậu xử lý để loại bỏ các hộp chồng chéo, theo dõi các phương tiện...

Có thể chia các phương pháp xây dựng mô hình nhận dạng phương tiện giao thông thành 4 nhóm [1]: các phương pháp dựa vào kỹ thuật trích chọn đặc trưng (màu sắc, hình dạng, cạnh và đường viền ...) [2], các phương pháp dựa vào hiệu giữa các frame liên tiếp trong dữ liệu video [3], các phương pháp dựa vào luồng quang học (optical flow) [4] và các phương pháp dựa vào các mô hình mạng học sâu phát hiện đối tượng như YOLO [5], Retinanet [6], SSD [7] hay Fast R-CNN [8]. Các phương pháp thuộc các cách tiếp cận [2], [3] và [4] có ưu điểm là không cần nhiều dữ liệu huấn luyện hay năng lực xử lý mạnh mẽ của các hệ thống phần cứng, tốc độ nhanh nhưng hạn chế về độ

chính xác. Các mô hình học sâu ([5], [6], [7] và [8]) có điểm chung là cần dữ liệu huấn luyện lớn, các phần cứng triển khai phải có năng lực xử lý mạnh

mẽ (các card đồ họa GPU chuyên dụng) và tốc độ thực thi chậm hơn, nhưng lại có độ chính xác cao hơn.



Hình 1. Bài toán phát hiện phương tiện giao thông

Mô hình học thích ứng (Adaptive Learning) [9] là một giải pháp nhằm khắc phục các hạn chế của các mạng học sâu, khai phá những vấn đề mà các mạng học sâu chưa làm được. Một mô hình Adaptive Learning hoàn chỉnh sẽ cho phép hệ thống nhận dạng có khả năng tự học hỏi, tự thông minh mô phỏng theo hoạt động của bộ não con người. Hệ thống có thể tự động chọn lọc các dữ liệu được cho là phù hợp, tự huấn luyện lại mô hình và tự cập nhật thay thế mô hình cũ. Từ đó giảm bớt chi phí trong thu thập, xử lý dữ liệu và huấn luyện hệ thống nhận dạng.

ở điều kiện thời tiết xấu tốt hơn so với mô hình YOLO3 thông thường (giá trị của tham số mAP cao hơn khoảng 5%).

Các phần tiếp theo của bài báo được cấu trúc như sau: phần 2 trình bày về mô hình mạng học sâu YOLO3 và mô hình học thích ứng trên nền mạng học sâu này. Hệ thống phát hiện phương tiện giao thông và các kết quả thử nghiệm hệ thống được trình bày cụ thể trong phần 3. Cuối cùng là kết luận và một số đề xuất.

2. Mô hình học thích ứng trên nền mạng học sâu YOLO3

2.1. Mô hình nhận dạng đối tượng YOLO3

Trong phạm vi của bài báo này, tác giả đề xuất giải pháp sử dụng mô hình Adaptive Learning trên nền mạng học sâu YOLO3 [10] để xây dựng một hệ thống phát hiện các phương tiện giao thông phổ biến như xe đạp, xe máy, xe con, xe tải và xe buýt. Hệ thống được xây dựng dựa trên nền tảng Darknet-53 [11] và ngôn ngữ lập trình Python 3. Để huấn luyện và thử nghiệm hệ thống, tác giả sử dụng bộ dữ liệu chuẩn MS COCO [12] và dữ liệu ảnh thực tế gồm hơn 8500 hình ảnh phương tiện do tác giả thu thập.

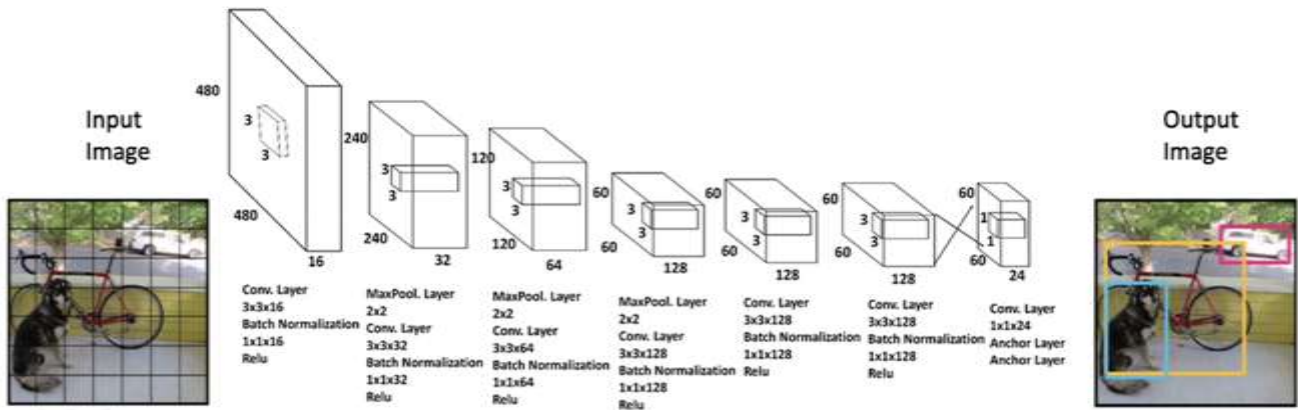
Mô hình mạng học sâu YOLO (You Only Look Once) [5] là một mô hình phát hiện các đối tượng trong ảnh được giới thiệu vào năm 2016. So với bài toán phân lớp đối tượng, bài toán phát hiện đối tượng phức tạp hơn vì phải trả lời hai câu hỏi: có loại đối tượng nào và vị trí của chúng ở đâu trong bức ảnh đầu vào. Mô hình YOLO có thể xác định vị trí và lớp của các đối tượng có trong ảnh đầu vào chỉ với một lần chạy qua mô hình. Thử nghiệm cho thấy YOLO có tốc độ xử lý đạt 28,2ms/1 ảnh có độ phân giải 320 x 320 điểm ảnh, tương đương với tốc độ khung hình khoảng 35,5 FPS.

Kết quả thử nghiệm cho thấy, so với các giải pháp hiện tại, hệ thống do tác giả đề xuất có thể phát hiện tốt 5 loại phương tiện giao thông là xe đạp, xe máy, ô tô con, xe tải, xe buýt (xe khách) với tỷ lệ chính xác cao (trong khoảng 95% - 97%). Bên cạnh đó hệ thống cũng có khả năng phát hiện các phương tiện giao thông có trong các ảnh chụp

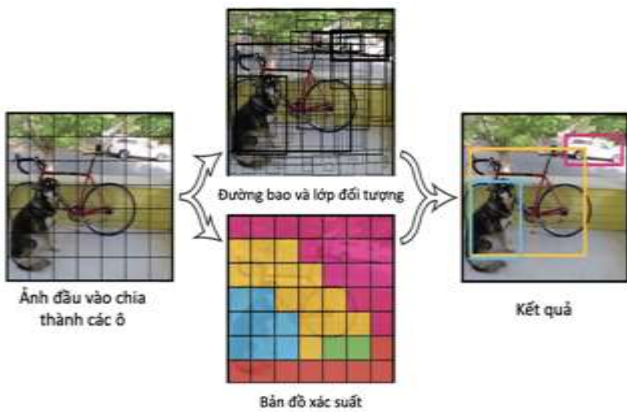
Hình 2 mô tả kiến trúc của mô hình mạng học sâu YOLO. Mô hình sử dụng một mạng neuron tích chập (convolutional neural network - CNN) để phân tích hình ảnh và tìm kiếm các đối tượng. Ảnh đầu vào được chia thành các ô nhỏ để xử lý, sau đó mô

hình sẽ nhận dạng và dự đoán xác suất cho các lớp tương ứng với các đối tượng trong các đường bao (bounding boxes) xung quanh mỗi ô nhỏ này.

Những đường bao có xác suất cao sẽ được giữ lại và sử dụng cho nhiệm vụ xác định vị trí của đối tượng trong ảnh như mô tả trong Hình 3.



Hình 2. Kiến trúc của mô hình mạng học sâu YOLO [5]



Hình 3. Phát hiện đối tượng trong mô hình YOLO [5]

Mô hình YOLO3 là phiên bản nâng cấp của YOLO, được công bố vào năm 2018 [10]. Mô hình YOLO3 sử dụng một kiến trúc mạng CNN mới và áp dụng những kỹ thuật mới để cải thiện hiệu suất của mô hình. Các cải tiến của YOLOv3 bao gồm: i) sử dụng một kiến trúc mạng neuron tích chập mới gọi là Darknet-53 [11] để tạo ra các đặc trưng của ảnh đầu vào; ii) sử dụng mạng trích xuất đặc trưng kim tự tháp (Feature Pyramid Network – FPN) để tăng độ chính xác của việc phát hiện đối tượng trong các kích cỡ khác nhau của hình ảnh; lii) thực hiện dự đoán trên 3 lớp khác nhau để đạt được độ chính xác cao hơn; iv) sử dụng thuật toán Non-maximum suppression - NMS để loại bỏ các đường bao trùng lặp và giảm số lượng dự đoán sai.

Mô hình YOLO3 đã đạt được kết quả tốt trên nhiều bài toán phát hiện đối tượng khác nhau trong

thị giác máy tính, bao gồm cả bài toán phát hiện đối tượng trong thời gian thực. Do đó, YOLO3 đã trở thành một trong những mô hình phát hiện đối tượng phổ biến nhất hiện nay.

2.2. Mô hình học thích ứng

Mặc dù những mô hình mạng học sâu như YOLO3 đã giải quyết được các bài toán phát hiện đối tượng và cho kết quả tốt nhưng những mô hình này vẫn tồn tại một số hạn chế cần phải khắc phục và nghiên cứu thêm: thứ nhất, để tạo ra một hệ thống có khả năng phát hiện đa dạng các đối tượng, cần có một lượng dữ liệu đầu vào khổng lồ để huấn luyện các mô hình học sâu. Quy trình huấn luyện này mất nhiều thời gian và cần tài nguyên xử lý cực kỳ lớn. Thứ hai, các mô hình học sâu vẫn chưa thể nhận dạng được những đối tượng phức tạp, và gặp khó khăn khi cần nhận dạng những đối tượng tương tự nhau. Lý do là vì hiện chưa có kỹ thuật nào đủ tốt để các mô hình học sâu có thể rút ra những kết luận một cách logic. Bên cạnh đó, vẫn còn nhiều thách thức trong việc tích hợp kiến thức trừu tượng vào các hệ thống phát hiện đối tượng sử dụng mô hình học sâu, ví dụ như thông tin về đối tượng đó là gì, dùng để làm gì, dùng như thế nào ...

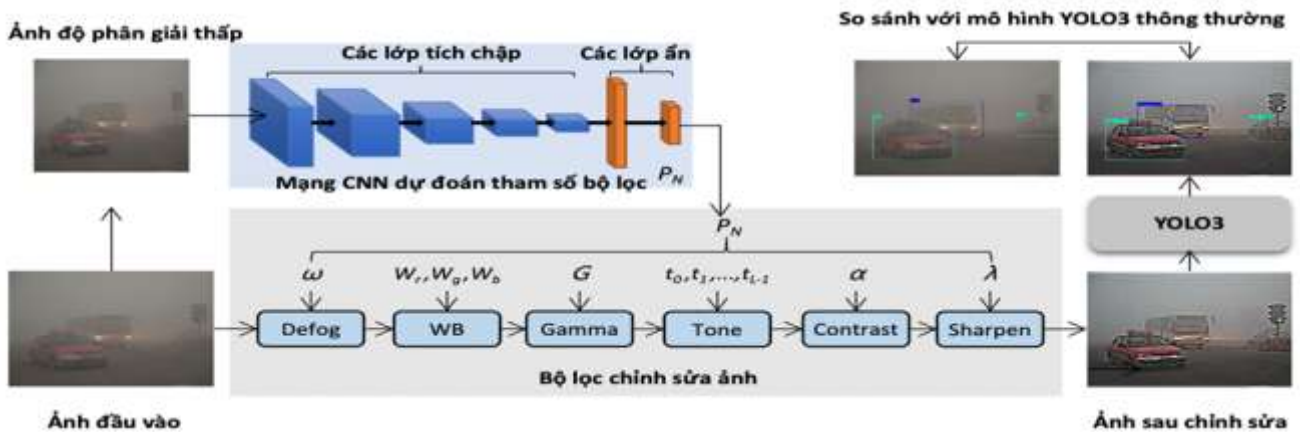
Mô hình học thích ứng (Adaptive Learning) [9] là một giải pháp nhằm khắc phục các hạn chế nêu trên. Trong mô hình này các mạng học sâu được thiết kế để thích nghi với sự thay đổi của dữ

liệu đầu vào. Mô hình cho phép tự động tối ưu hóa các tham số và cấu trúc của mạng nơ-ron để đảm bảo rằng nó có thể hoạt động hiệu quả trên các dữ liệu mới và không được huấn luyện trước. Các mô hình học thích ứng được sử dụng trong nhiều ứng dụng như nhận dạng khuôn mặt, xử lý ngôn ngữ tự nhiên, điều khiển robot, xe tự lái, và các ứng dụng y sinh học. Với sự phát triển của các mô hình học thích ứng, ứng dụng học sâu có khả năng thích nghi tốt hơn và đạt được độ chính xác cao hơn đối với các dữ liệu mới và phức tạp.

Mô hình học thích ứng có thể được áp dụng đối với bài toán phát hiện phương tiện giao thông nhằm cải thiện hiệu quả của hệ thống. Với mô hình học thích ứng, các mạng nơ-ron có khả năng tự động thích nghi với các thay đổi về ánh sáng, thời tiết, hình dạng và vị trí của các phương tiện giao

thông, giúp cải thiện độ chính xác và hiệu suất của hệ thống. Cụ thể, khi gặp phải dữ liệu mới về phương tiện giao thông mà hệ thống chưa từng gặp trước đó, mô hình học thích ứng sẽ tự động điều chỉnh các tham số và cấu trúc của mạng để có thể nhận dạng được các đặc trưng đó.

Ví dụ, nếu phương tiện giao thông có kích thước, màu sắc khác so với các dữ liệu mà mô hình đã học, mô hình sẽ tự động thích nghi bằng cách điều chỉnh kích thước nhân của một số lớp mạng. Mô hình học thích ứng cũng có thể được sử dụng để giảm thiểu tác động của các yếu tố ngoại cảnh như thời tiết và ánh sáng. Nếu ánh sáng môi trường thay đổi, mô hình sẽ cập nhật lại các tham số và cấu trúc của mạng để phù hợp với các biến đổi này và đưa ra kết quả nhận dạng chính xác hơn.



Hình 4. Mô hình học thích ứng nhận dạng phương tiện giao thông

Trong bài báo này, tác giả đề xuất mô hình học thích ứng trên nền mạng học sâu YOLO3 để cải thiện độ chính xác của hệ thống phát hiện phương tiện giao thông khi làm việc với những ảnh đầu vào được chụp trong điều kiện thời tiết xấu (trời mưa, hoặc có nhiều sương mù, ngược sáng ...). Mô hình YOLO3 thông thường hoạt động kém chính xác khi phải phát hiện các phương tiện giao thông có trong ảnh đầu vào được chụp trong điều kiện thời tiết bất lợi, bởi vì: i) chưa được huấn luyện trước do dữ liệu ảnh dùng để huấn luyện mô hình YOLO3 thường là các ảnh chụp trong điều kiện thông thường; ii) chất lượng ảnh đầu vào kém (mờ, nhòe ...) do ảnh hưởng của thời tiết. Mô hình học thích ứng được trình bày trong Hình 4, bao gồm

một bộ lọc dùng để chỉnh sửa ảnh theo 06 loại tham số đầu vào, một mạng học sâu CNN dùng để dự đoán tham số của bộ lọc, một mạng học sâu YOLO3 dùng để phát hiện phương tiện giao thông có trong ảnh. Hoạt động của hệ thống như sau:

Bước 1: Hệ thống tạo ra một bản sao có độ phân giải thấp (256 x 256 điểm ảnh) của ảnh đầu vào, bản sao này được sử dụng để làm đầu vào của mô hình học sâu CNN dự đoán tham số cho bộ lọc chỉnh sửa ảnh.

Bước 2: Ảnh đầu vào được cho qua bộ lọc chỉnh sửa ảnh để chỉnh sửa, các tham số của bộ lọc chỉnh sửa ảnh được tạo ra từ mạng CNN mô tả trong bước một. Bộ lọc chỉnh sửa ảnh dựa trên giá trị của 06 nhóm tham số: i) Lọc sương mù (defog);

ii) Cân bằng trắng (WB); iii) Hiệu chỉnh Gamma; iv) Chỉnh tông màu (Tone); v) Chỉnh độ tương phản (Contract); vi) Chỉnh độ nét (Sharpen).

Bước 3: Ảnh sau chỉnh sửa được đưa qua mô hình YOLO3 để phát hiện phương tiện giao thông có trong ảnh.

3. Thử nghiệm và đánh giá hệ thống

3.1. Cài đặt

Hệ thống phát hiện phương tiện giao thông ứng dụng mô hình học thích ứng dựa trên nền tảng YOLO3 được cài đặt trên máy tính có cấu hình phần cứng như sau: CPU Intel Core i9-10900X, GPU RTX 3080 TI 12GB, RAM 64GB, ổ cứng 1TB SSD. Về nền tảng phần mềm, hệ thống sử dụng hệ điều hành Ubuntu 22.04.1 LTS, phiên bản Python 3.10, Tensorflow 2.10 và OpenCV 4.6.

3.2. Dữ liệu

Dữ liệu để huấn luyện và kiểm thử hệ thống được trích ra từ bộ dữ liệu MS COCO [12] với năm nhóm (loại phương tiện giao thông): xe đạp, xe máy, ô tô con, xe tải, xe buýt (xe khách). Bên cạnh đó tác giả cũng sử dụng bộ dữ liệu ảnh tự thu thập gồm các ảnh chụp các tình huống giao thông thực tế. Hệ thống camera được sử dụng thu thập dữ liệu phía trước hoặc phía sau các phương tiện tham gia giao thông và được thu thập từ các bối cảnh thực tế khác nhau trên các tuyến đường giao thông khác nhau. Các phương tiện giao thông xuất hiện trong bộ dữ liệu bao gồm: xe đạp, xe máy, ô tô con, xe tải, xe buýt (xe khách), với 8688 hình ảnh phương tiện giao thông.

Để huấn luyện và thử nghiệm mô hình học thích ứng tác giả sử dụng bộ dữ liệu RTTS [13] gồm 4270 ảnh chụp người và phương tiện đang tham gia giao thông trong điều kiện thời tiết mưa, sương mù, thiếu sáng, ngược sáng ... Các đối tượng có trong bộ dữ liệu này bao gồm: người, xe đạp, ô tô con, xe buýt (xe khách)

Các ảnh có trong hai bộ dữ liệu được chia thành 02 nhóm. Nhóm ảnh dùng để huấn luyện chiếm 60% và nhóm ảnh dùng để đánh giá chiếm 40%. Bảng 1 mô tả các thông tin về bộ dữ liệu thử nghiệm.

Bảng 1. Mô tả dữ liệu thử nghiệm

Loại phương tiện	Số lượng mẫu			Độ phân giải
	Tổng	Huấn luyện	Đánh giá	
Xe đạp	1026	616	410	416 x 416
Xe máy	2458	1475	983	416 x 416
Ô tô con	2521	1513	1008	416 x 416
Xe tải	1245	747	498	416 x 416
Xe buýt (xe khách)	1438	863	575	416 x 416
<i>Ảnh chụp trong điều kiện thời tiết xấu</i>				
Người	1120	762	448	416 x 416
Xe đạp	980	588	392	416 x 416
Ô tô con	1260	756	504	416 x 416
Xe buýt (xe khách)	910	546	364	416 x 416

3.3. Kết quả

Bảng 2. Ma trận lỗi (Confusion Matrix)

Lớp thực tế	Lớp dự đoán					
		Xe đạp	Xe máy	Ô tô con	Xe tải	Xe buýt
	Xe đạp	384	26	0	0	0
Xe máy	75	908	0	0	0	
Ô tô con	0	0	958	22	28	
Xe tải	0	0	36	377	85	
Xe buýt	0	0	7	45	523	

Bảng 3. Các độ đo đánh giá hệ thống

Lớp	Các độ đo					
	Preci.	Recall	F1	Acc	Avg IOU	ms/img.
Xe đạp	0,84	0,94	0,88	0,97	0,79	48,5
Xe máy	0,97	0,92	0,95	0,97	0,81	47,4
Ô tô con	0,96	0,95	0,95	0,97	0,86	46,2
Xe tải	0,85	0,76	0,80	0,95	0,83	48,8
Xe buýt	0,82	0,91	0,86	0,95	0,84	47,2

Kết quả thử nghiệm hệ thống với các ảnh chụp ở điều kiện thông thường được mô tả tại

Bảng 2 và Bảng 3.

Để đánh giá hiệu quả của hệ thống học thích ứng đối với nhiệm vụ phát hiện phương tiện giao thông từ các ảnh chụp trong điều kiện thời tiết xấu, tác giả thử nghiệm và so sánh với mô hình phát hiện sử dụng mạng học sâu YOLO3 thông thường. Hai hệ thống được chạy thử trên cùng bộ dữ liệu RTTS [13] và tính toán giá trị của độ đo mAP (mean Average Percision) là độ đo được sử dụng phổ biến hiện nay để đánh giá hiệu quả của các hệ thống nhận dạng đối tượng [14]. Kết quả so sánh được trình bày trong Bảng 4.

Bảng 4. So sánh giá trị mAP giữa hai hệ thống

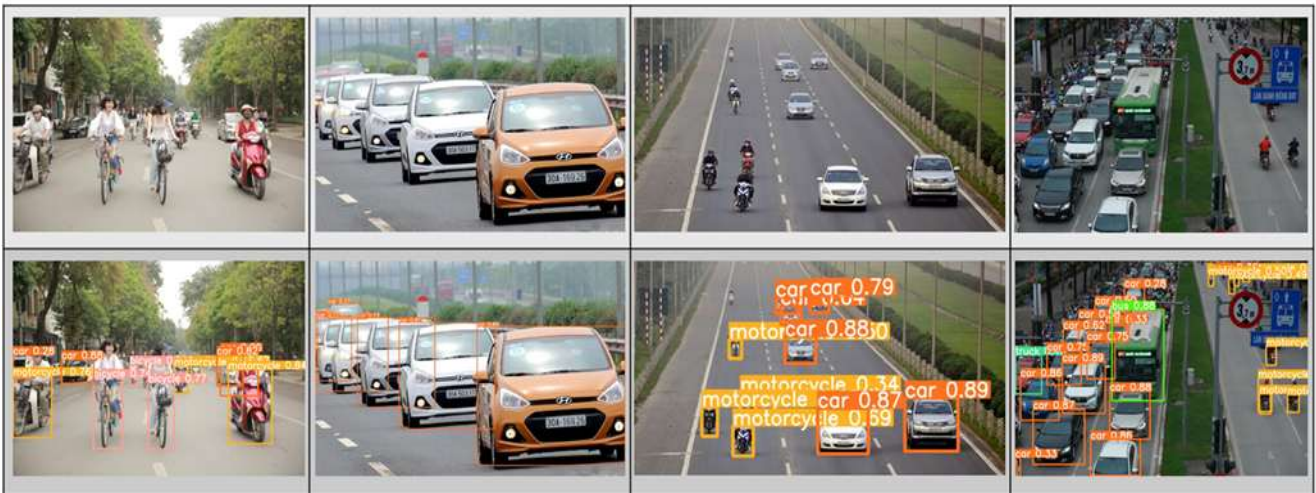
Mô hình	Lớp đối tượng			
	Người	Xe đạp	Ô tô con	Xe buýt
YOLO3	65,33	64,08	69,13	68,97
Adaptive YOLO3	72,03	72,23	73,23	73,23

Hình vẽ 5, 6, 7, trình bày một số ví dụ về kết quả phát hiện phương tiện giao thông của hệ thống

khi dữ liệu đầu vào là các ảnh được chụp trong điều kiện thời tiết bình thường và ảnh được chụp trong điều kiện thời tiết xấu.

Các kết quả thử nghiệm cho thấy những điểm đáng chú ý sau:

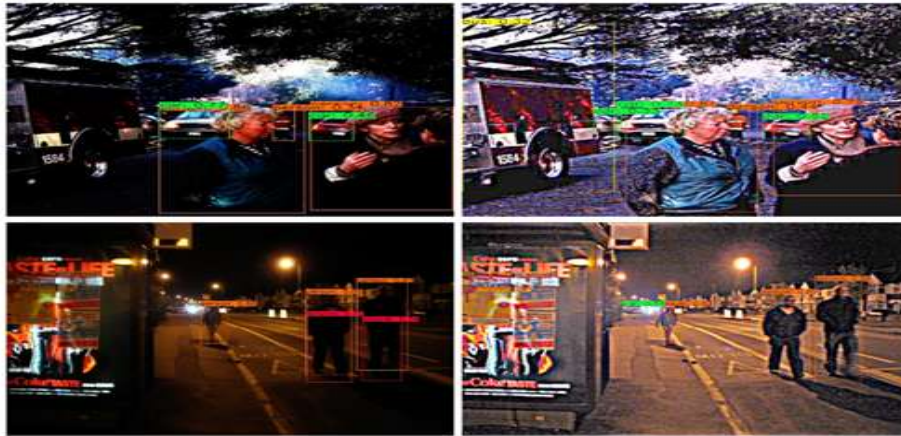
1. Hệ thống có thể phát hiện tốt 5 loại phương tiện giao thông là xe đạp, xe máy, ô tô con, xe tải, xe buýt (xe khách) với tỷ lệ chính xác cao (trong khoảng 95% - 97%).
2. Hệ thống có khả năng phát hiện các phương tiện giao thông có trong các ảnh chụp ở điều kiện thời tiết xấu tốt hơn so với mô hình YOLO3 thông thường (giá trị của tham số mAP cao hơn khoảng 5%).
3. Hệ thống có nhầm lẫn giữa xe tải và xe buýt khi số lượng đối tượng trong ảnh lớn.
4. Hệ thống có khả năng chạy trong thời gian thực (tốc độ xử lý trung bình khoảng 48ms/1 hình ảnh, tương đương với tốc độ khung hình 20 FPS).



Hình 5. Phát hiện phương tiện giao thông từ ảnh chụp trong điều kiện thường



Hình 6. Phát hiện phương tiện giao thông từ ảnh chụp trong điều kiện thời tiết xấu



Hình 7. Ví dụ phát hiện đối tượng từ ảnh chụp thiếu sáng

4. Kết luận

Phát hiện phương tiện giao thông là một bài toán khó và phức tạp. Trong bài báo này, tác giả xây dựng hệ thống phát hiện phương tiện giao thông sử dụng mô hình học thích ứng trên nền mạng học sâu YOLO3. Hệ thống có khả năng phát hiện các phương tiện giao thông phổ biến như xe đạp, xe máy, xe con, xe tải và xe buýt. Thêm vào đó nhờ ứng dụng mô hình học thích ứng, hệ thống có khả năng phát hiện phương tiện giao thông từ ảnh chụp trong điều kiện thời tiết xấu với hiệu quả cao hơn so với hệ thống dùng mạng YOLO3 thông thường. Hệ thống được thử nghiệm trên bộ dữ liệu chuẩn và bộ dữ liệu tự thu thập. Kết quả thử nghiệm cho thấy hệ thống đạt được độ chính xác cao và có khả năng áp dụng vào các dự án thực tế trong lĩnh vực xe tự lái hoặc kiểm soát giao thông.

Lời cảm ơn

Bài báo là sản phẩm của đề tài nghiên cứu cấp Trường, tên đề tài “Ứng dụng mạng học sâu trong tự động mô tả hình ảnh”, mã số “CS-2021”, thời gian thực hiện năm học 2021 – 2022.

Tài liệu tham khảo

- [1] Yang, Z., & Pun-Cheng, L. S. (2018). Vehicle detection in intelligent transportation systems and its applications under varying environments: A review. *Image and Vision Computing*, 69, 143-154.
- [2] Jazayeri, A., Cai, H., Zheng, J. Y., & Tuceryan, M. (2011). Vehicle detection and tracking in car video based on motion model. *IEEE*

Transactions on Intelligent Transportation Systems, 12(2), 583-595.

- [3] Ge, D. Y., Yao, X. F., Xiang, W. J., & Chen, Y. P. (2022). Vehicle detection and tracking based on video image processing in intelligent transportation system. *Neural Computing and Applications*, 1-13.
- [4] Aslani, S., & Mahdavi-Nasab, H. (2013). Optical flow based moving object detection and tracking for traffic surveillance. *International Journal of Electrical, Computer, Energetic, Electronic and Communication Engineering*, 7(9), 1252-1256.
- [5] Han, X., Chang, J., & Wang, K. (2021). Real-time object detection based on YOLO-v2 for tiny vehicle object. *Procedia Computer Science*, 183, 61-72.
- [6] Lin, T. Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2017). Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision* (pp. 2980-2988).
- [7] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. (2016). Ssd: Single shot multibox detector. In *Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14* (pp. 21-37). Springer International Publishing.
- [8] Fan, Q., Brown, L., & Smith, J. (2016, June). A closer look at Faster R-CNN for vehicle detection. In *2016 IEEE intelligent vehicles*

- symposium (IV)* (pp. 124-129).
- [9] Tran, D. P., & Hoang, V. D. (2019). Adaptive learning based on tracking and Reldentifying objects using convolutional neural network. *Neural Processing Letters*, 50(1), 263-282.
- [10] Huang, Y. Q., Zheng, J. C., Sun, S. D., Yang, C. F., & Liu, J. (2020). Optimized YOLOv3 algorithm and its application in traffic flow detections. *Applied Sciences*, 10(9), 3079.
- [11] Zhang, X., Dong, X., Wei, Q., & Zhou, K. (2019). Real-time object detection algorithm based on improved YOLOv3. *Journal of electronic imaging*, 28(5), 053022-053022.
- [12] Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D.,... & Zitnick, C. L. (2014). Microsoft coco: Common objects in context. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13* (pp. 740-755). Springer International Publishing.
- [13] Li, B., Ren, W., Fu, D., Tao, D., Feng, D., Zeng, W., & Wang, Z. (2018). Benchmarking single-image dehazing and beyond. *IEEE Transactions on Image Processing*, 28(1), 492-505.
- [14] Padilla, R., Netto, S. L., & Da Silva, E. A. (2020, July). A survey on performance metrics for object-detection algorithms. In *2020 international conference on systems, signals and image processing (IWSSIP)* (pp. 237-242). IEEE.