

ỨNG DỤNG MẠNG NƠ RON TÍCH CHẬP LAI GHEP ĐỂ XỬ LÝ ẢNH TRONG HỆ THỐNG BÁO ĐỘNG TRỰC CA HÀNG HẢI

APPLYING HYBRID CONVOLUTIONAL NEURAL NETWORKS FOR IMAGE PROCESSING TO BRIDGE NAVIGATIONAL WATCH AND ALARM SYSTEM

Đặng Xuân Kiên, Nguyễn Việt Chính, Phan Thanh Minh
 Trường Đại học Giao thông vận tải Tp. Hồ Chí Minh

Tóm tắt: Hiện nay, mạng nơ ron tích chập được coi như một công cụ hữu hiệu nhất trong kỹ thuật chuẩn đoán và xử lý hình ảnh. Trong bài báo này, các tác giả đã nghiên cứu và áp dụng giải thuật lai ghép SSDLite_MobileNetV2 cải tiến vào hệ thống báo động trực ca hàng hải GTS.VI. Phân cứng thiết kế dựa trên Raspberry Pi-3, một board máy tính nhúng với CPU cấp điện thoại thông minh, RAM giới hạn và không có GPU CUDA. Kết quả thực nghiệm cho thấy hoàn toàn có thể đưa một mô hình học sâu áp dụng lên các máy tính nhúng với độ chính xác cao như các kết quả trong nghiên cứu này.

Từ khóa: Mạng nơ ron tích chập, phát hiện đối tượng, BNWAS.

Chỉ số phân loại: 2.2

Abstract: Currently, Convolutional Neural Networks are considered as the most effective tool in image diagnosis and processing techniques. In this paper, the authors have studied and applied the modified SSDLite_MobileNetV2 hybrid algorithm to Bridge Navigational Watch & Alarm System - GTS.VI(BNWAS). The hardware was designed based on Raspberry Pi-3, a embedded single board computer with CPU smartphone level, limited RAM without CUDA GPU. Experimental results showed that we can deploy a deep learning model on an embedded single board computer with high effective.

Keywords: Convolutional neural networks, object detection, BNWAS.

Classification number: 2.2

1. Giới thiệu

Phát hiện đối tượng có thể được coi là một thách thức lớn đối với thị giác máy tính, vì nó liên quan đến sự kết hợp giữa phân loại đối tượng và định vị đối tượng trong khung hình. Sự ra đời của những tiến bộ hiện đại trong học sâu [1], [2] đã dẫn đến những tiến bộ đáng kể trong phát hiện đối tượng, với phần lớn nghiên cứu tập trung vào việc thiết kế các mạng phát hiện đối tượng ngày càng phức tạp hơn để cải thiện độ chính xác như SSD [2, 3], Faster R-CNN [4], YOLO [5]. Hiệu suất phát hiện đối tượng trên ImageNet và PASCAL VOC đã được cải thiện đáng kể với sự phát triển của các giải thuật xử lý ảnh dựa trên mạng nơ ron tích chập (CNNs). Tuy vậy các giải thuật đều hình đòi hỏi lượng tài nguyên hệ thống lớn và để đưa chúng lên các thiết bị có phần cứng hạn chế như Hệ thống Trực ca hàng hải buồng lái (BNWAS) cần tinh gọn và biên dịch lên phần cứng hạn chế của hệ thống.

Để đáp ứng được yêu cầu về hạn chế phần cứng, nhóm nghiên cứu sử dụng cấu trúc SSDLite MobileNetV2 cải tiến bởi nó có tốc độ nhanh và chính xác. Không những đáp ứng được các yêu cầu về xử lý hình ảnh, phát hiện và phân loại đối tượng, hệ thống còn đáp ứng

được các quy định của IMO [6], [7], [9] và IEC [8] để có thể đưa vào thử nghiệm và vận hành trực tiếp trên tàu.

2. Xử lý hình ảnh và video dựa trên SSDLite-MobileNet

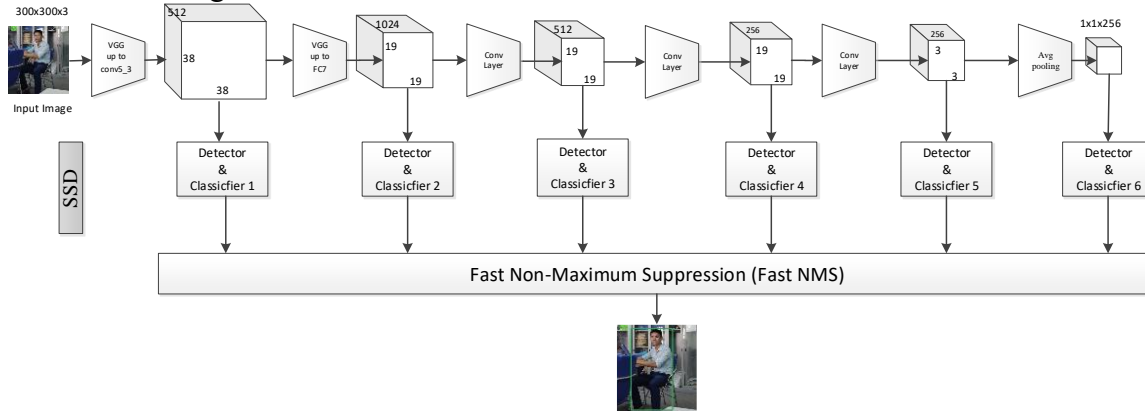
Các mô hình CNN tuy có độ chính xác cao, nhưng chúng đều có một điểm hạn chế chung đó là không phù hợp với các ứng dụng trên mobile hay các hệ thống nhúng có khả năng tính toán thấp.

Nếu muốn phát triển các mô hình trên cho các ứng dụng trong thời gian thực ta cần phải có cấu hình cực kì mạnh mẽ (GPU / CPU), còn đối với các hệ thống nhúng (Raspberry Pi, Nano pc) hay các ứng dụng chạy trên smartphone, ta cần có một mô hình "nhẹ" như SSDLite-MobileNet lai ghép.

Yếu tố chính giúp SSDLite-MobileNet có được độ chính xác cao trong khi thời gian tính toán thấp nằm ở cấu trúc được lai ghép từ SSD và cấu trúc MobileNet.

- SSD (Single Shot MultiBox Detector) là một bộ dò đối tượng (hình 1) thực hiện hai công đoạn chính: Trích xuất bản đồ đặc trưng của các đối tượng (feature maps) và áp dụng các bộ lọc tích chập (convolution filters) để

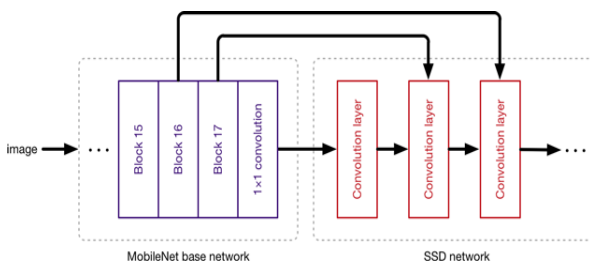
phát hiện đối tượng.



Hình 1. Cấu trúc SSD (Single Shot MultiBox Detector) [2].

- Trong khi cấu trúc của MobileNetV2 chứa lớp tích chập hoàn toàn ban đầu với 32 bộ lọc, sau đó là 19 lớp bottleneck. Cấu trúc chi tiết MobileNetV2 được M. Sandler mô tả cụ thể trong tài liệu [11].

Để lai ghép SSD và MobileNetV2, các tác giả đã thay thế tất cả các lớp tích chập thông thường (regular convolutions) bằng các lớp tích chập phân tách (separable convolutions) trong các lớp dự đoán của mạng SSD [2] nhằm giúp mô hình giảm số lượng tham số so với các mạng tích chập thông thường, ngoài ra giúp mô hình giảm được một số bước tính toán. Cấu trúc mạng lai ghép SSD-MobileNet V2 có thể biểu diễn như trên hình 2 [2].



Hình 2. Cấu trúc SSDLite-MobileNetV2 lai ghép.

Các thử nghiệm của tác giả M. Sandler [11] chứng minh kết quả SSDLite-MobileNetV2 hiệu quả hơn 20 lần và dung lượng nhỏ hơn 10 lần so với phương pháp YOLOv2 [10] trên tập dữ liệu COCO. Trong đó, kết quả đầu ra được gán nhãn đối tượng và mức độ tin cậy theo phần trăm. Trong các thực nghiệm của bài báo này, phương pháp SSD-MobileNet V2 cải tiến còn cho thấy hiệu quả cao hơn phương pháp của [11] đặc biệt khi áp dụng cho hệ thống báo động trực ca hàng hải buồng lái (BNWAS).

3. Ứng dụng CNN trong hệ thống báo động trực ca hàng hải

3.1. Các yêu cầu IMO theo SOLAS về BNWAS

Bridge Navigational Watch & Alarm System (BNWAS) là hệ thống giám sát và báo động, thông báo cho các nhân viên điều hướng hoặc thuyền trưởng khác nếu sỹ quan trực ca không trả lời hay không thể thực hiện các nhiệm vụ theo dõi hiệu quả, có thể dẫn đến tai nạn hàng hải. Các yêu cầu IMO MSC. 128(75) và SOLAS Chapter V, Reg.19 MSC. 282(86) [9] sửa đổi ngày 5 tháng 6 năm 2009 [6] có hiệu lực đối với tàu phân loại theo kích thước:

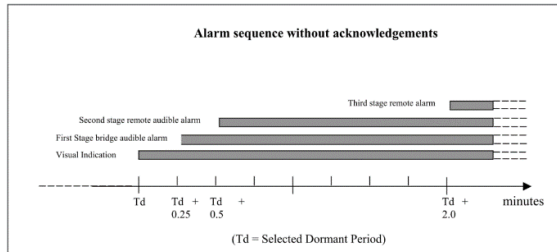
- Từ tháng 7 năm 2011: Tàu mới vượt quá 150 tấn;
- Từ tháng 7 năm 2011: Tất cả các tàu khách;
- Từ tháng 7 năm 2012: Tất cả các tàu vượt quá 3.000 tấn;
- Từ tháng 7 năm 2013: Tất cả các tàu từ 500 đến 3.000 tấn;
- Từ tháng 7 năm 2014: Tất cả các tàu từ 150 đến 500 tấn.

Ứng dụng phát hiện đối tượng (cụ thể là Sỹ quan trực ca) vào hệ thống BNWAS giúp cho hệ thống tự nhận dạng sỹ quan trực ca có mặt và hoạt động bình thường trong buồng lái, nhằm đảm bảo an toàn hàng hải và phòng ngừa rủi ro.

3.2. Thiết kế phần cứng hệ thống BNWAS

Yêu cầu tối thiểu đối với một BNWAS

theo tiêu chuẩn Tổ chức Hàng hải Quốc tế là có một giai đoạn không hoạt động và ba giai đoạn báo động, ngoại trừ trên một tàu không phải tàu chở hành khách, giai đoạn thứ hai có thể được bỏ qua [6]. Cụ thể các giai đoạn như sau:



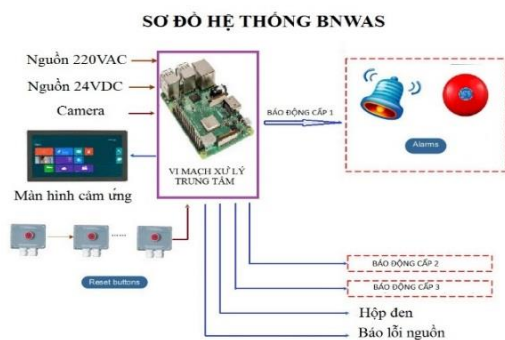
Hình 3. Thời gian các cấp báo động của BNWAS.

Giai đoạn 1: Khi máy lái tự động tham gia, sỹ quan trực ca phải báo hiệu sự hiện diện của anh ta với hệ thống BNWAS cứ sau 3 đến 12 phút (hình 3) để phản ứng với ánh sáng nhấp nháy, bằng cách di chuyển trước một cảm biến chuyển động, nhấn một nút xác nhận hoặc trực tiếp áp dụng ngay trên trung tâm BNWAS.

Giai đoạn 2: Khi không có tín hiệu xác nhận trong vòng 15 giây trong giai đoạn 1, chuông báo động sẽ kêu trên buồng lái và nếu vẫn không có tín hiệu xác nhận sau khi thêm 15 giây, chuông sẽ kêu trong cabin thuyền trưởng và cabin của các sỹ quan chỉ huy. Một trong số họ sau đó phải đi đến buồng lái và hủy bỏ báo động.

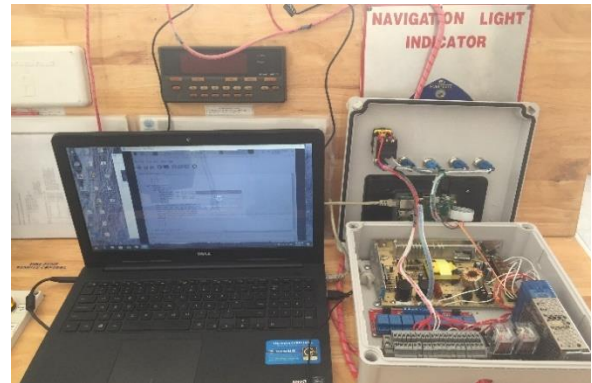
Giai đoạn 3: Nếu không có thuyền trưởng hay các sỹ quan chỉ huy hủy báo động trong một khoảng thời gian nhất định (từ 90 giây đến 3 phút tùy thuộc vào kích cỡ của tàu), một báo động sẽ phát ra ở những vị trí mà các nhân viên khác thường có mặt.

Gọi khẩn cấp: chức năng này cho phép thuyền viên có thể kích hoạt báo động Giai đoạn 2 hoặc giai đoạn 3 để yêu cầu trợ giúp.



Hình 4. Sơ đồ nguyên lý hệ thống BNWAS thiết kế.

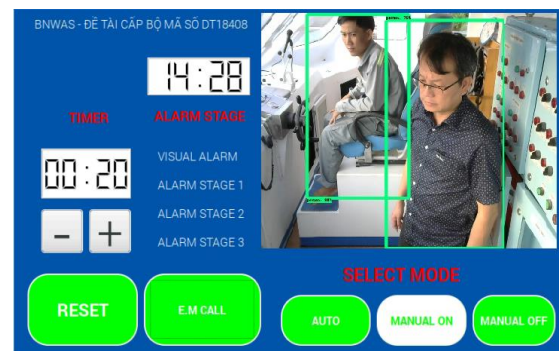
Sơ đồ nguyên lý thiết kế như trên hình 4 và được áp dụng để thiết kế hệ thống thực trên hình 5. Máy tính được kết nối làm việc song song với Raspberry Pi-3 (đóng vai trò của vi mạch xử lý trung tâm trên hình 4) để thu thập dữ liệu vào ra của quá trình thử nghiệm. Phần cứng được thiết kế thực hiện các chức năng báo động thỏa mãn các quy định về hàng hải.



Hình 5. Hệ thống BNWAS – GTS.VI thử nghiệm tại Đại học GTVT TP.HCM

3.3. Thử nghiệm BNWAS trên tàu Saigon Millennium

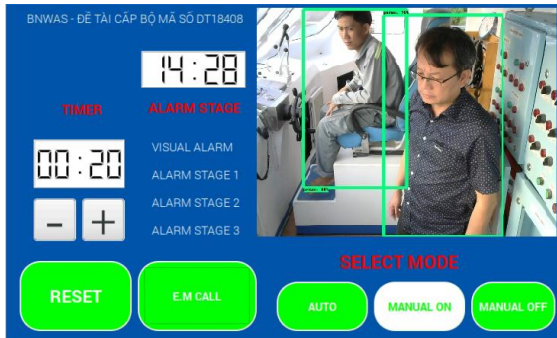
Hình ảnh được ghi lại từ tàu Saigon Millennium tại Nhà máy đóng tàu Sơn Hải, Phú Xuân, Nhà Bè.



Hình 6. Hình ảnh trực tiếp từ buồng lái được ghi lại bởi Hệ thống Trực ca hàng hải buồng lái.

Hình ảnh được ghi lại thông qua camera Logitech C270, sau đó được xử lý bằng các giải thuật xác định đối tượng dựa trên mạng lai ghép SSD-MobileNet V2 cải tiến. Đầu ra là hình ảnh đã được xử lý, trích xuất ra khung bao quanh đối tượng được phát hiện và độ tin cậy tính theo phần trăm. Với kỹ thuật sử dụng trong bài báo này, hệ thống có thể xác định nhiều sỹ quan trong buồng lái và số lượng tối đa có thể truy xuất khung phát hiện là 20 người trong cùng một thời điểm. Khi nhận

diện được sỹ quan trong buồng lái, hệ thống cho phép tùy chỉnh các chức năng thông qua màn hình cảm ứng hoặc nút nhấn trên tủ điều khiển.



Hình 7. Thử nghiệm xác định thời gian báo động sơ cấp.

Sau khi xử lý hình ảnh, nếu hệ thống xác định không có sỹ quan trực ca trong buồng lái, một timer sẽ được bật lên và thời gian đếm ngược chờ sự xuất hiện của sỹ quan. Trong khoảng thời gian timer hoạt động, các chức năng chuyên chế độ, tăng giảm timer đều bị vô hiệu hóa. Nếu trong khoảng thời gian đếm ngược này có sự xuất hiện của sỹ quan trực ca trong buồng lái (không cần sự tác động vật lý lên hệ thống) thì timer được reset và hệ thống trở lại trạng thái bình thường, các sỹ quan có thể thao tác và sử dụng các phím chức năng của hệ thống.

Trường hợp không có sỹ quan quay lại và timer đã đếm về không (hết thời gian chờ), một tín hiệu cảnh báo bằng đèn chớp sẽ được kích hoạt trong buồng lái, giai đoạn này gọi là giai đoạn báo động sơ cấp. Tín hiệu này có thể được nhìn thấy tại bất cứ đâu trong buồng lái và phù hợp với các tiêu chuẩn của IMO. Trên màn hình điều khiển sẽ hiển thị cấp báo động, đồng thời mọi thông số hệ thống sẽ được lưu vào file lịch sử, sau đó một timer tiếp theo được khởi động để chuyển tiếp qua giai đoạn báo động tiếp theo.



Hình 8. Ảnh chụp màn hình sau khi hệ thống trải qua thử nghiệm các cấp báo động.

Các thử nghiệm về cấp báo động tiếp theo được thử nghiệm và cho kết quả cuối cùng phù hợp với yêu cầu của IMO. Không những nhận diện sự có mặt của sỹ quan trong buồng lái, hệ thống còn phân tích hành động của sỹ quan và đưa ra cảnh báo khi nhận thấy sỹ quan đứng yên quá lâu hoặc nằm ngửa trong quá trình làm việc.



Hình 9. Thử nghiệm phát hiện sỹ quan trực ca ngồi im quá lâu hoặc có dấu hiệu ngủ gật.

Thử nghiệm được ghi lại khi chúng tôi yêu cầu một sỹ quan ngồi im trên ghế lái (tối thiểu 20 giây), điều này tương trưng cho việc sỹ quan đứng im quá lâu hoặc đang có hiện tượng ngủ gật. Lúc này một bộ đếm ngầm chạy sẽ phân tích vị trí tương đối của sỹ quan và đưa ra sai số tương đối. Dựa vào kết quả phân tích từng khung hình, nếu sau 20 giây liên tiếp sai số vị trí tương đối không quá 10%, báo động sơ cấp được đặt lên và timer cho báo động tiếp theo sẽ bắt đầu đếm ngược.

3.4. Đánh giá và so sánh các kết quả thử nghiệm

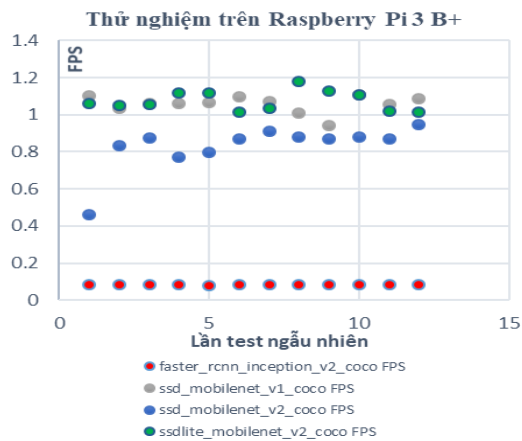
Để so sánh hiệu quả của giải pháp đưa ra với các tác giả khác, dựa trên phần cứng và điều kiện thực tế tại buồng lái của tàu Saigon Millennium, nhóm nghiên cứu đã triển khai đồng bộ bốn giải pháp để lấy kết quả. Lựa chọn hai yếu tố là tốc độ xử lý và độ tin cậy đầu ra khi áp dụng các bộ dò đối tượng trên hệ thống đã thiết kế.

Với cấu hình cao, các mô hình chạy trên GPU TITAN X cho ra kết quả tốc độ xử lý từ 17 đến 37 khung hình trên giây. Tuy vậy khi thử nghiệm trên bộ dữ liệu COCO và tính toán mAP trên tất cả các lớp đối tượng, kết quả chỉ đạt 21 - 28%. Kết quả thử nghiệm về tốc độ xử lý trên BNWAS – GTS.VI, thử nghiệm trực tiếp trên buồng lái với điều kiện làm việc bình thường khi chỉ

cần xác định một lớp đối tượng là con người cho ra độ tin cậy cao từ 76 - 97% (bảng 1).

Bảng 1: Kết quả thử nghiệm.

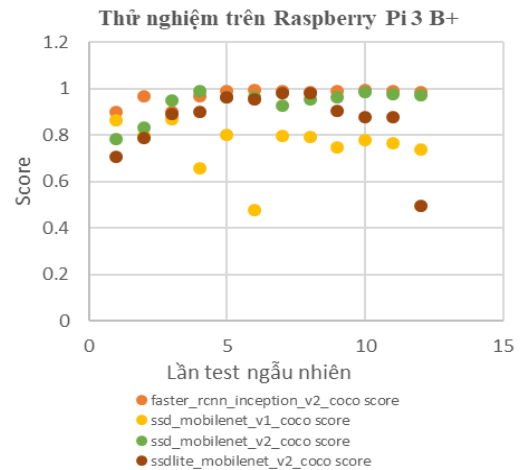
Tên mô hình	Thử nghiệm trên GPU TITAN X		Thử nghiệm trên Raspberry Pi 3B+	
	Tốc độ (ms)	Thử nghiệm trong tập COCO (mAP)	Tốc độ (FPS)	Thử nghiệm trong buồng lái (mAP)
ssd_mobilenet_v1_coco	30	21	1.05	0.76
ssd_mobilenet_v2_coco	31	22	0.83	0.94
ssdlite_mobilenet_v2_coco	27	22	1.08	0.86
faster_rcnn_inception_v2_coco	58	28	0.08	0.97



Hình 10. So sánh tốc độ xử lý của các bộ dò đối tượng trên phân cứng của BNWAS.

Kết quả ấn tượng này có được khi lắp đặt camera trong buồng lái ở vị trí thuận lợi trong khi với phần cứng là thiết bị di động chỉ có CPU ARM và không có GPU tích hợp, tốc độ xử lý cao nhất chỉ đạt xấp xỉ 1 FPS. Tốc độ FPS của các phương pháp thử nghiệm trên hình 10, đây là tốc độ đáp ứng tốt đối với một hệ thống giám sát.

Độ tin cậy đầu ra cao nhất khi thử nghiệm với bộ dò Faster RCNN, tuy vậy với 0.08FPS (khoảng 12.5 giây để xử lý một khung hình) là không thể đáp ứng được trên một hệ thống giám sát. Các bộ dò đối tượng dựa trên cấu trúc SSD_MobileNet (màu nâu) cho ra kết quả tin cậy cao và đáp ứng được yêu cầu về tốc độ xử lý.



Hình 11. So sánh độ tin cậy đầu ra của các bộ dò đối tượng trên phân cứng của BNWAS.

Trong khi đó, thực tế cho thấy (hình 11), kết quả của bộ SSD_MobileNetV1 (vàng) và SSD_MobileNetV2 (xanh lá) gần như tương đương, nhưng thời gian nạp mô hình chậm do dung lượng lớn, kết quả đầu ra thực tế vẫn có sự sai lệch nhất định. Như vậy, giải pháp SSDLite_MobileNetV2 cải tiến cho ra kết quả tốt và cân bằng giữa các yếu tố: chất lượng, tốc độ xử lý, thời gian nạp mô hình nhanh (chạy ổn định trên Raspberry PI-3), điều này khẳng định bằng kết quả thử nghiệm theo thực tế có độ chính xác cao hơn các giải pháp còn lại.

4. Kết luận

Bài báo đã nghiên cứu và cải tiến giải thuật lai ghép SSDLite_MobileNetV2 và ứng dụng thiết kế cho hệ thống báo động

trực ca hàng hải *BNWAS – GTS.V1*, kết quả thực nghiệm tốt khi so sánh với các giải thuật khác trên cùng một phần cứng. Hệ thống *BNWAS – GTS.V1* được kiểm nghiệm trong môi trường phòng thí nghiệm trước khi lắp đặt thử trên tàu Saigon Millennium để thực nghiệm và thu thập dữ liệu để đánh giá, phân tích. Hệ thống không chỉ áp dụng cho việc xác định sự hiện diện của sỹ quan trực ca trên buồng lái, hệ thống còn có thêm các tính năng nhận dạng từng sỹ quan khác nhau, nhận dạng hành vi và cảnh báo các sự cố khác, đồng thời có thể truy xuất dữ liệu hình ảnh vào bộ nhớ đen □

Tài liệu tham khảo

- [1] Krizhevsky, I. Sutskever, and G. E. Hinton. “Imagenet classification with deep convolutional neural networks”. In NIPS, 2012, pp 1097-1105.
- [2] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. E. Reed, C.-Y. Fu, and A. C. Berg, “Ssd: Single shot multibox detector”. In ECCV, 2016.
- [3] Chengcheng Ning, Huajun Zhou, Yan Song, Linhui Tang, “Inception Single Shot Multibox Detector for Object Detection” In ICME, 2017.
- [4] S. Ren, K. He, R. Girshick, J. Sun. “Faster r-cnn: To-wards real-time object detection with region proposal networks” IEEE Volume: 39, Issue: 6, June 1 2017, pp 1137 – 1149.
- [5] Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi, “You Only Look Once: Unified, Real-Time Object Detection”, IEEE(CVPR) June 2016, pp 1063-6919.
- [6] IMO MSC.128(75) Performance Standards for a Bridge Navigational Watch Alarm System (BNWAS).
- [7] IMO A.1021(26) Code on Alerts and Indicators 2009.
- [8] IEC 62616:2010 Maritime Navigation and Radio Communication Equipment and Systems- Bridge Navigational Watch Alarm System (BNWAS).
- [9] IMO MSC.282(86) Chapter V, Reg.19, adopted on 5 June 2009.
- [10] Joseph Redmon and Ali Farhadi, “Yolo9000: Better, faster, stronger”, arXiv: 1612.08242, 2016.
- [11] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, Liang-Chieh Chen, “MobileNetV2: Inverted Residuals and Linear Bottlenecks”, arXiv:1801.04381.

Ngày nhận bài: 27/3/2019

Ngày chuyển phản biện: 4/4/2019

Ngày hoàn thành sửa bài: 25/4/2019

Ngày chấp nhận đăng: 2/5/2019