TẠP CHÍ ISSN: 1859-316X, e-ISSN: 3093-3161
KHOA HỌC CÔNG NGHỆ HÀNG HẢI
-JOURNAL OF MARINE SCIENCE AND TECHNOLOGY-

PHÁT HIỆN VÀ CẢNH BÁO HÀNH VI SỬ DỤNG THIẾT BỊ DI ĐỘNG CỦA NGƯỜI LÁI XE DỰA TRÊN HỌC SÂU

DEEP LEARNING-BASED DRIVER MOBILE DEVICE USAGE DETECTION AND WARNING

HỒ THỊ HƯƠNG THƠM

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

Email liên hệ: thomhth@vimaru.edu.vn

DOI: https://doi.org/10.65154/jmst.830

Tóm tắt

Số vụ tai nạn ô tô ngày càng tăng là việc rất cần được quan tâm các hệ thống giao thông hiện nay. Theo WHO (Tổ chức Y tế Thế giới), tai nạn giao thông đường bộ là nguyên nhân gây tử vong đứng thứ tám trên toàn cầu. Hơn 80% vụ tai nạn giao thông trên đường bộ đều xuất phát từ người lái xe mất tập trung, chẳng hạn như sử dụng điện thoại di động, ăn uống, nói chuyện với hành khách hoặc hút thuốc. Đã có rất nhiều nỗ lực nhằm hạn chế lái xe mất tập trung; tuy nhiên, chưa đưa ra được biện pháp tối ưu nào. Một cách tiếp cận thực tế để giải quyết vấn đề này là áp dụng các biện pháp định lượng cho mọi hoạt động của người lái cũng như thiết kế một hệ thống nhận diện những hành động gây mất tập trung.

Trong bài báo này đã triển khai một mô hình học sâu YOLO11 có thể phân loại hiệu quả hành động sử dụng điện thoại gây mất tập trung của người lái và đưa ra khuyến nghị trong xe để giảm thiểu mức độ mất tập trung và tăng cường nhận thức trong xe nhằm cải thiện an toàn. Mô hình đã này đạt độ chính xác cao với tập dữ liệu huấn luyện 9668 ảnh. Sau đó mô hình được cài đặt chạy trên điện thoại di động và thử nghiệm cho 5 video quan sát với độ chính xác là 92,6%. Với kết quả này góp phần quan trọng vào bài toán cảnh báo kịp thời khi người lái ô tô có hành vi mất tập trung, giảm thiểu nguy cơ tai nạn giao thông từ đó nâng cao an toàn đường bộ.

Từ khóa: *Học sâu, sử dụng điện thoại khi lái xe, người lái ô tô mất tập trung.*

Abstract

The increasing number of automobile accidents is a significant problem in current traffic systems. According to the World Health Organization (WHO), road traffic accidents are the eighth leading cause of death globally. More than 80% of road traffic accidents are caused by driver distraction, such as using mobile phones, eating, talking to passengers, or smoking. Many efforts have been made to address the issue of distracted driving; however, no opti-mal solution has yet been proposed. A practical approach to tackle this prob-lem is to apply quantitative measures to driver activities and design a system to detect distracting actions.

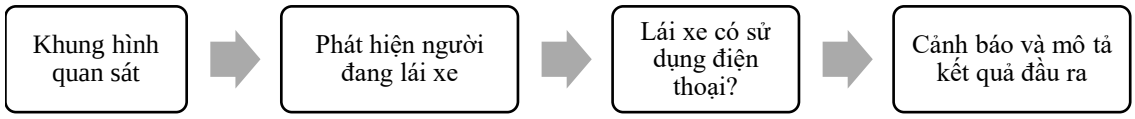
In this paper, a YOLO11 deep learning model was implemented, capable of effectively classifying distracting mobile phone usage behavior by drivers and providing in-vehicle recommendations to mitigate distraction levels and enhance in-vehicle awareness for improved safety. The model achieved high accuracy for training datasets of 9668 images. Subsequently, the model was deployed on mobile phones and tested on 5 observation videos, yielding an accuracy of 92.6%. This significantly contributes to timely warn-ings for drivers, reducing the risk of traffic accidents due to distraction, thereby enhancing road safety.

Keywords: *Deep Learning, Mobile phone usage while driving, driver distraction.*

1. Mở đầu

An toàn giao thông đường bộ đang nhận được sự quan tâm hàng đầu trên toàn thế giới. Theo các báo cáo thống kê, tai nạn giao thông vẫn là nguyên nhân gây ra hàng triệu ca tử vong và thương tích mỗi năm, trong đó một phần lớn xuất phát từ sự mất tập trung của người lái xe. Giữa các yếu tố gây xao nhãng, hành vi sử dụng điện thoại di động khi đang điều khiển phương tiện được xem là một trong những mối nguy hiểm phổ biến và nghiêm trọng nhất, làm tăng đáng kể nguy cơ xảy ra va chạm. Trước thực trạng đó, việc nghiên cứu và phát triển các hệ thống hỗ trợ lái xe tiên

TẠP CHÍ ISSN: 1859-316X, e-ISSN: 3093-3161
KHOA HỌC CÔNG NGHỆ HÀNG HẢI
JOURNAL OF MARINE SCIENCE AND TECHNOLOGY



Hình 1. Sơ đồ phát hiện hành vi sử dụng điện thoại của người lái xe

tiến (Advanced Driver-Assistance Systems - ADAS) giúp tự động phát hiện và cảnh báo hành vi này đã trở thành một hướng đi cấp thiết và đầy tiềm năng.

Thời gian gần đây, với sự phát triển vượt trội của trí tuệ nhân tạo, các phương pháp dựa trên thị giác máy tính và học sâu đã chứng tỏ hiệu quả vượt trội trong việc xử lý bài toán này. Nhiều công trình nghiên cứu đã tập trung vào việc ứng dụng Mạng Nơ-ron Tích chập (Convolutional Neural Networks - CNN) để phân tích các khung hình của video hoặc hình ảnh từ camera đặt trong cabin xe, qua đó xác định hành vi sử dụng điện thoại của tài xế. Các nghiên cứu nền tảng như của P. K. Singh và cộng sự [5] hay M. A. M. Ali và cộng sự [1] đã chứng minh tính khả thi của việc sử dụng kiến trúc CNN để phân loại và phát hiện, đặt nền móng cho các cải tiến sau này.

Để đáp ứng yêu cầu về xử lý thời gian thực và triển khai trên các thiết bị có tài nguyên hạn chế, các nhà nghiên cứu đã hướng đến việc tối ưu hóa mô hình. Điển hình là công trình của Y. Li và cộng sự [2] đã cải tiến kiến trúc YOLOv5, mô hình phát hiện đối tượng khá tốt và hiệu quả, nhằm tăng mức độ xử lý mà vẫn giữ được độ tin cậy cao. Cũng theo hướng đó, J. Wang và cộng sự [6] cũng đề xuất một phương pháp sử dụng kiến trúc CNN hạng nhẹ (lightweight), tập trung vào việc giảm độ phức tạp tính toán của mô hình. Một số nghiên cứu khác tiếp cận bài toán trong bối cảnh rộng hơn là phát hiện mất tập trung nói chung, trong đó sử dụng điện thoại là một trường hợp con quan trọng. Các hệ thống của R. Al-Tahar và cộng sự [3] hay A. K. Dubey [7] đều sử dụng học sâu để xây dựng một giải pháp toàn diện, có khả năng cảnh báo nhiều loại hành vi xao nhãng khác nhau. Để tăng cường độ chính xác, S. Mondal và cộng sự [4] đã khám phá một hướng đi mới bằng cách sử dụng mạng 3D-CNN đa luồng hạng nhẹ, cho phép mô hình nắm bắt cả thông tin không gian và thời gian từ chuỗi hình ảnh.

Mặc dù đã có nhiều kết quả đáng khích lệ, các hệ thống hiện tại vẫn tồn tại những thách thức nhất định. Việc cân bằng giữa độ chính xác và tốc độ xử lý vẫn là một bài toán khó, đặc biệt khi triển khai trên các hệ thống nhúng. Hơn nữa, hiệu suất của mô hình có thể bị ảnh hưởng bởi các điều kiện môi trường phức tạp như ánh sáng thay đổi (ngày/đêm, ngược sáng), góc quay camera, hay các trường hợp bị che khuất một phần.

Để cải thiện độ chính xác, nghiên cứu này triển khai xây dựng "mô hình hệ thống phát hiện và cảnh báo người lái xe có hành vi sử dụng thiết bị di động dựa trên thị giác máy tính" với mục tiêu tối ưu hóa cả về độ chính xác lẫn hiệu suất tính toán. Mục tiêu cuối cùng là xây dựng một mô hình đáng tin cậy, có thể tích hợp dễ dàng vào các phương tiện giao thông góp phần nâng cao an toàn và giảm thiểu tai nạn.

Nội dung chính của nghiên cứu được trình bày chi tiết trong các mục 2, 3, 4 và kết luận sẽ trình bày trong 5.

2. Mô hình phát hiện sử dụng thiết bị di động của người lái xe

Sơ đồ tổng quát của mô hình phát hiện người lái xe sử dụng điện thoại được thể hiện trong Hình 1.

Trong sơ đồ này có các thành phần cụ thể như sau:

Khung hình quan sát: Đây là bước đầu tiên, có thể hiểu là thu thập thông tin hình ảnh hoặc khung hình quan sát.

Phát hiện người đang lái xe: Từ khung hình quan sát, hệ thống sẽ tiến hành phát hiện xem có người đang điều khiển phương tiện hay không.

Phân tích hành vi có sử dụng điện thoại?: Sau khi phát hiện người lái xe, bước tiếp theo là phân tích hành vi của họ để xác định xem họ có đang sử dụng điện thoại hay không.

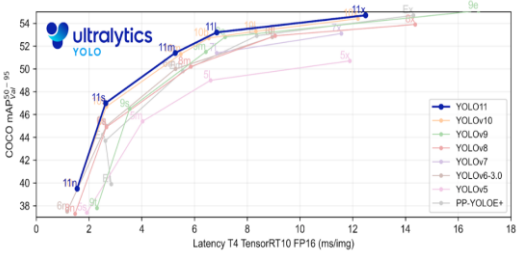
Cảnh báo và mô tả kết quả đầu ra: Bước cuối cùng là đưa ra thông tin mô tả sau khi phân tích (ví dụ: Có sử dụng điện thoại hay không) và phát ra cảnh báo cần thiết.

3. Phương pháp phát hiện người lái xe có hành vi sử dụng điện thoại di động

Trong những năm gần đây, với sự phát triển không ngừng của học sâu, nhiều nhà nghiên cứu đã đưa mạng nơ ron tích chập (CNN) vào lĩnh vực phát hiện đối tượng và đạt những kết quả được ghi nhận. Đặc biệt là mạng nơ ron YOLO trong phát hiện đối tượng. YOLO (You Only Look Once) là một họ mô hình phát hiện vật thể (object detection) nổi tiếng, được giới thiệu lần đầu vào năm 2016 bởi Joseph Redmon. Khác với các mô hình hai giai đoạn như R-CNN, YOLO xử lý toàn bộ hình ảnh chỉ một lần duy nhất thông qua

mạng nơ-ron tích chập (CNN), từ đó dự đoán trực tiếp vị trí (bounding box) và loại vật thể trong cùng một bước. Chính nhờ thiết kế này, YOLO đạt tốc độ phát hiện thời gian thực (real-time) với độ chính xác cao — rất phù hợp cho các ứng dụng như giám sát giao thông, xe tự hành hay an ninh thông minh.

Năm 2024, YOLO11 được Ultralytics giới thiệu là phiên bản mới nhất, kế thừa và nâng cấp toàn diện từ YOLOv8. Kiến trúc YOLOv11 gồm các khối C3k2, C2PSA (Parallel Spatial Attention) và SPPF giúp trích xuất đặc trưng mạnh mẽ hơn, giảm số tham số nhưng tăng mAP. Mô hình hỗ trợ nhiều tác vụ: Phát hiện vật thể, phân đoạn ảnh, ước lượng tư thế (pose), phát hiện hộp xoay (OBB) và phân loại. Ngoài ra, YOLOv11 được tối ưu để hoạt động hiệu quả trên thiết bị biên (edge device), hỗ trợ xuất sang ONNX, TensorRT, CoreML,... giúp triển khai linh hoạt [8].



Hình 2. So sánh mAP50-95 của YOLO11 với một số phiên bản trước [8]

Trong Hình 2 là kết quả so sánh giá trị trung bình mAP50-95 của YOLO11 trên dữ liệu COCO [9] với các phiên bản bản tiền nhiệm cho thấy giá trị chính xác cao hơn hẳn.

Dòng YOLO thể hiện sự phát triển liên tục từ một ý tưởng đơn giản — "nhìn một lần, phát hiện tất cả" — trở thành chuẩn mực của phát hiện vật thể thời gian thực. YOLO11 hiện là cột mốc mới nhất, đạt hiệu năng cao hơn, linh hoạt hơn và sẵn sàng cho các ứng dụng AI hiện đại như giám sát giao thông thông minh hay phân tích video tốc độ cao thời gian thực nên trong nghiên cứu này sử dụng YOLO11 để xây dựng mô hình phát hiện người đang lái xe sử dụng điện thoại.

4. Cài đặt thử nghiệm

Sử dụng ngôn ngữ lập trình Python để cài đặt mô hình phát hiện người lái xe sử dụng điện thoại và các thư viện cần thiết hỗ trợ cho YOLO11.

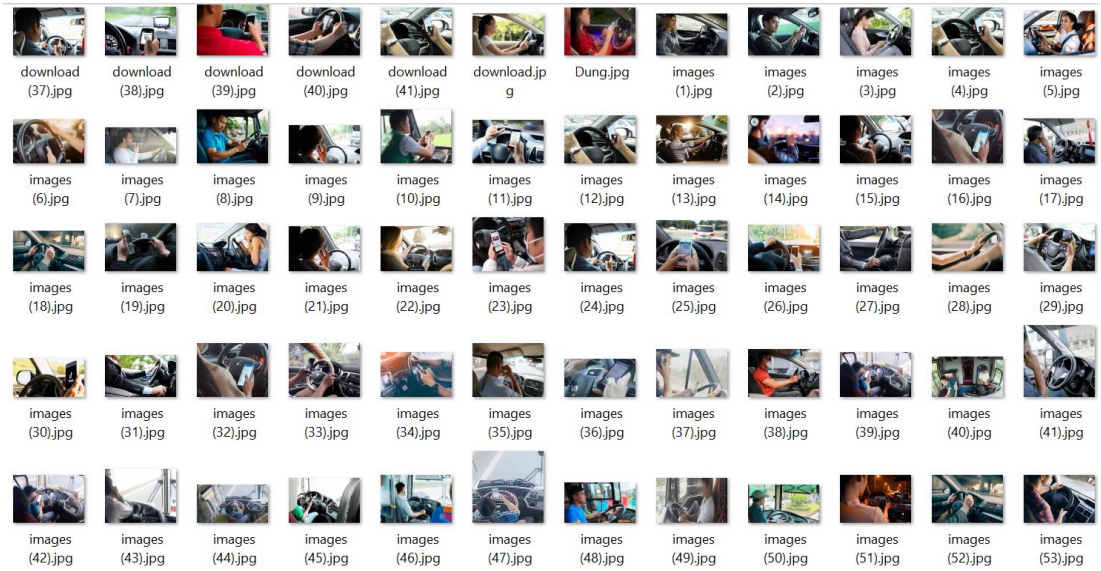
Tập dữ liệu ảnh dùng để huấn luyện cho mô hình gồm 9668 ảnh trong đó 9462 ảnh thu thập từ [10] và ảnh tự thu thập là 206 ảnh. Tập ảnh sẽ được gán nhãn, các ảnh được gán nhãn theo quy ước sau:

- su_dung: Người lái cầm điện thoại, đưa lên gần tai hoặc nhìn xuống màn hình.

- khong_sudung: Người lái không cầm điện thoại.

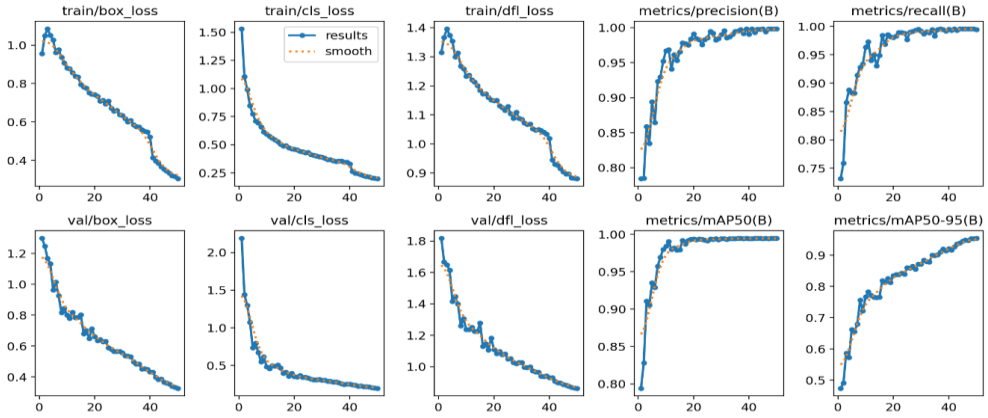
Tập ảnh sau khi được gán nhãn xong được chia làm 2 tập: Tập huấn luyện (train) có 8.218 ảnh (chiếm 85%) và tập kiểm tra (test) có 1.450 ảnh (15%)). Hình 3 là một số ảnh trong tập dữ liệu huấn luyện.

Dùng YOLO11 để tiến hành huấn luyện với chu kỳ là 50 epochs, trong thời gian 0,978 giờ được các tỷ lệ tin cậy trung bình cho các phương tiện như sau: P (Precision) là 0,989, R (Recall) là 0,985 và mAP50 = 0,985, mAP50-95 là 0,944 (mAP50-95 trên bộ dữ liệu COCO là 0,53 [8]). Cụ thể độ tin cậy của mô hình sau khi huấn luyện được thể hiện trong Hình 4.



Hình 3. Một số ảnh minh họa trong tập huấn luyện phát hiện người lái xe sử dụng điện thoại

TẠP CHÍ ISSN: 1859-316X, e-ISSN: 3093-3161
KHOA HỌC CÔNG NGHỆ HÀNG HẢI
JOURNAL OF MARINE SCIENCE AND TECHNOLOGY

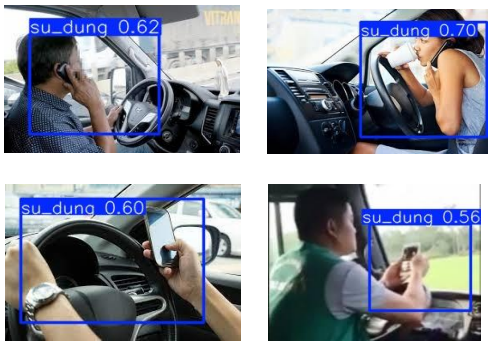


Hình 4. Thông tin về độ tin cậy của mô hình YOLO11 sau khi huấn luyện tập dữ liệu người lái xe sử dụng điện thoại

Ngoài ra nghiên cứu cũng thực hiện quy trình đánh giá k-fold ($k=5$) trên tập dữ liệu 8.218 ảnh. Tập ảnh được chia thành 5 fold không trùng lặp; mỗi lần huấn luyện sử dụng khoảng 6.574 ảnh cho train và 1.644 ảnh cho validation. Kết quả mAP@0,5 trung bình đạt 0,912 ± 0,003, trong khi mAP@0,5:0,95 đạt 0,636 ± 0,004 qua 5 fold.

Từ kết quả huấn luyện, mô hình đã đạt độ chính xác và độ bao phủ tương đối tốt, thích hợp để triển khai vào phát hiện và hỗ trợ đưa ra cảnh báo cho người lái xe tránh được rủi ro có thể xảy ra khi đang lái xe trong thời gian thực.

Vận dụng mô hình phát hiện người lái xe sử dụng điện thoại di động qua khung hình video trên một số đoạn đường ở miền bắc Việt Nam được thể hiện trong Hình 5.



Hình 5. Ảnh kết quả phát hiện người lái xe ô tô trên video bằng mô hình YOLO11 đã huấn luyện

Để xây dựng ứng dụng trên thiết bị di động, mô hình huấn luyện phải được chuyển từ định dạng .pt sang định dạng .tflite. Việc chuyển đổi được tiến hành trên Google Colab. Chuyển đổi định dạng .pt sang định dạng .tflite theo lệnh trong Hình 6.

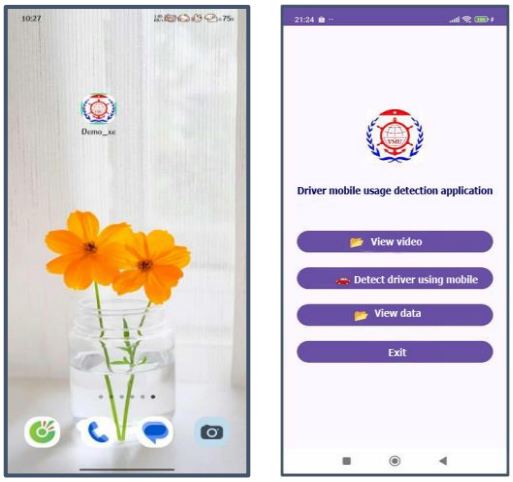
Để chạy mô hình trên thiết bị di động, ta cần Java phiên bản 11.0.16.1, Android Studio phiên bản Meerkat Feature Drop thông qua trang chính thức và các thư viện cần thiết trên máy tính. Sau khi kết nối điện thoại di động với máy tính và chạy chương trình trên Android Studio, ứng dụng được cài đặt trên điện thoại di động với giao diện như trong Hình 7.

```
from ultralytics import YOLO

# Load mô hình đã huấn luyện
model = YOLO("best.pt")

# Export ra TFLite với data.yaml đúng
model.export(format='tflite', data='data.yaml')
```

Hình 6. Chuyển đổi định dạng cho mô hình huấn luyện



Hình 7. Mô hình phát hiện lái xe sử dụng điện thoại được cài đặt trên thiết bị điện thoại

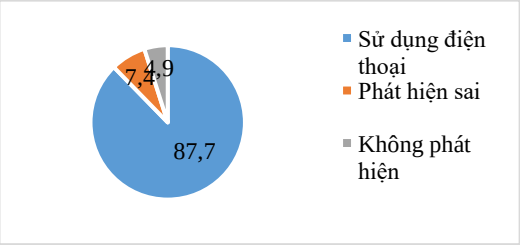
Chọn 5 video để thử nghiệm mô hình, các video này đều được lưu ý đến các yếu tố đạo đức, ẩn danh và bảo mật trong thu thập và sử dụng dữ liệu video có sự xuất hiện của người tham gia, thông tin cụ thể của video cụ thể theo Bảng 1.

Tiêu chí đánh giá, cách tính precision/recall/F1

Bảng 1. Thông tin của 5 video được sử dụng để đánh giá mô hình

Video	Độ dài (giây)	Số khung hình	Thời điểm quay	Điều kiện ánh sáng/khí hậu	Loại xe chính	Góc camera	Khoảng cách camera–đường (m)
V1	46	2875	Ban ngày	Nắng mạnh	Xe ô tô	30°	2.5
V2	48	2937	Chiều	Ánh sáng yếu	Xe bus	25°	2
V3	35	2316	Chiều muộn	Bóng đổ nhiều	Xe tải	35°	1.5
V4	44	2765	Sáng	Trời râm	Xe tải	22°	1.3
V5	45	2856	Có mưa nhẹ	Tối ưu cho stress-test	Xe ô tô	30°	2

trên video. Đối với dữ liệu video, các chỉ số cấp khung hình chưa mô tả đầy đủ hành vi kéo dài theo thời gian. Vì vậy, nghiên cứu này áp dụng đánh giá cấp sự kiện (event-level). Một sự kiện được định nghĩa là một chuỗi các khung hình liên tiếp mà mô hình dự đoán cùng một hành vi (sử dụng điện thoại). Các khung hình liên tục được ghép thành sự kiện bằng thuật toán nối chuỗi, và mỗi sự kiện dự đoán được ghép nối với sự kiện đúng (ground-truth) dựa trên độ giao nhau theo trục thời gian (Temporal IoU). Một dự đoán được xem là đúng (TP-event) nếu Temporal IoU ≥ 0,5; ngược lại được xem là FP-event. Những sự kiện đúng không được phát hiện sẽ được tính là FN-event.



Hình 8. Kết quả thử nghiệm với 5 video về lái xe sử dụng điện thoại

Từ tập TP-event, FP-event và FN-event, các chỉ số đánh giá được tính như sau:

$$Precision_{event} = \frac{TP_{event}}{TP_{event} + FP_{event}} \quad (1)$$

$$Recall_{event} = \frac{TP_{event}}{TP_{event} + FN_{event}} \quad (2)$$

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (3)$$

Sử dụng 5 video để quan sát qua ứng dụng điện thoại cho người lái xe ô tô trên đường dựa trên sự kiện, thu được kết quả như sau:

- Số sự kiện người lái xe có trong video: 162;
- Số sự kiện phát hiện đúng: 142/162;
- Số sự kiện phát hiện sai: 12/162;
- Số sự kiện không phát hiện được: 8/162.

Ta có bảng ma trận nhầm lẫn (Confusion Matrix) theo Bảng 2.

Từ công thức (1) (2) và (3) được kết quả tổng hợp trong Bảng 3.

Bảng 2. Kết quả tổng hợp trong 5 video thử nghiệm

	Dự đoán: Event	Dự đoán: Không event
Event	TP = 142	FN = 8
None	FP = 12	TN = - (không xác định)

Bảng 3. Kết quả tổng hợp trong 5 video thử nghiệm

Chỉ số	Giá trị
Precision	92.21%
Recall	94.67%
F1-score	93.39%

5. Kết luận

Nghiên cứu đã thành công trong việc xây dựng một mô hình phát hiện người lái xe có hành vi sử dụng điện thoại di động dựa trên học sâu. Mô hình này, sử dụng kiến trúc học sâu YOLO11, cho thấy khả năng phân loại hiệu quả hành vi mất tập trung của người lái xe, đồng thời đưa ra các khuyến nghị trong xe để giảm thiểu sự mất tập trung và nâng cao nhận thức, góp phần cải thiện an toàn giao thông.

Mô hình YOLO11 được huấn luyện trên bộ dữ liệu gồm 9462 ảnh người lái xe có hoạt động sử dụng điện thoại, đạt độ chính xác P (Precision) là 0,989, R (Recall) là 0,985 và mAP50=0,985. Đặc biệt, trong thử nghiệm với 5 video quan sát qua thiết bị di động, các kết quả trong Bảng 3. cho thấy mô hình đạt Precision 92,21% và Recall 94,67%, phản ánh khả năng phát hiện đúng cao và tỷ lệ bỏ sót thấp. F1-score 93,39% chứng tỏ mô hình hoạt động ổn định và đáng tin cậy trong việc nhận diện người lái xe sử dụng điện thoại từ video thực tế,… Các kết quả này mô hình thể hiện độ tin cậy cao và phù hợp cho các ứng dụng giám

TẠP CHÍ ISSN: 1859-316X, e-ISSN: 3093-3161
KHOA HỌC CÔNG NGHỆ HÀNG HẢI
JOURNAL OF MARINE SCIENCE AND TECHNOLOGY

sát hành vi người lái xe, mặc dù vẫn còn một tỷ lệ nhỏ sự kiện bị nhận dạng sai hoặc bỏ sót, có thể được cải thiện thêm ở các thử nghiệm quy mô lớn hơn.

Việc chuyển đổi mô hình từ định dạng .pt sang .tflite cho phép triển khai trên các thiết bị di động, mở ra tiềm năng tích hợp vào các hệ thống hỗ trợ lái xe tiên tiến (ADAS). Điều này góp phần quan trọng vào việc kịp thời cảnh báo cho người lái xe, nhằm hạn chế nguy cơ tai nạn giao thông do người lái xe mất tập trung, từ đó nâng cao nhận thức an toàn đường bộ.

Dựa trên thành công của mô hình YOLO11 trong việc phát hiện lái xe có hành vi sử dụng điện thoại, định hướng nghiên cứu tiếp theo có thể mở rộng để giải quyết các trường hợp mất tập trung khác như: Buồn ngủ/ngáp, ăn uống/hút thuốc, nói chuyện với khách hàng,…, góp phần toàn diện hơn vào việc nâng cao an toàn giao thông.

Lời cảm ơn

Nghiên cứu này được tài trợ bởi Trường Đại học Hàng hải Việt Nam trong đề tài mã số: **DT25-26.78**.

TÀI LIỆU THAM KHẢO

[1] M. A. M. Ali, M. A. Ab-del-Haleem, H. M. Abdel-Atty, and M. A. A. Wahab (2023), *An Efficient CNN-Based Approach for Detecting Driver Phone Usage*, in 2023 5th International Conference on Computer and Information Sciences (ICCIS), Al-Jouf, Saudi Arabia, pp.1-6.

[2] Y. Li, G. Liang, Y. Chang, and C. Huang (2022), *Real-time Detection of Drivers' Phone Use based on Improved YOLOv5*, in 2022 14th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA), Changsha, China, pp.648-652.

[3] R. Al-Tahar, O. Al-Ata, and N. M. Tahat (2023), *A Smart Real-Time Driver Distraction Detection System Using Deep Learning*, in 2023 International Conference on Information Technology (ICIT), Amman, Jordan, 2023, pp.611-616.

[4] S. Mondal, S. K. Singh, and S. K. Vipparthi (2023), *Driver's Distraction Detection using Multi-stream lightweight 3D-CNN*, in 2023 IEEE International Conference on Image Processing (ICIP), Kuala Lumpur, Malaysia, pp.2095-2099.

[5] P. K. Singh, M. P. Singh, and D. K. Singh (2023), *Driver Distraction Detection using Convolutional Neural Network*, in 2023 3rd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), Greater Noida, India, pp.1530-1534.

[6] J. Wang, W. He, J. Chen, and Y. Wang (2022), *A Driver Distraction Detection Method Based on Lightweight Convolutional Neural Network*, in 2022 IEEE 5th International Conference on Automation, Electronics and Electrical Engineering (AUTEEE), Shenyang, China, pp.660-664.

[7] A. K. Dubey (2022), *An Automated System to Detect Driver Distraction to avoid Accidents using Deep Learning*, in 2022 2nd International Conference on Innovative Practices in Technology and Management (ICIPTM), Noida, India, pp.297-302.

[8]. Asif Razzaq (2024), *YOLO11 Released by Ultralytics: Unveiling Next-Gen Features for Real-time Image Analysis and Autonomous Systems*, 2024 Gartner® Cool Vendors™ in AI Engineering.
https://www.marktechpost.com/2024/10/03/yolo11-released-by-ultralytics-unveiling-next-gen-features-for-real-time-image-analysis-and-autonomous-systems/

[9] Coco dataset: https://www.cocodataset.org.

[10] Database: https://universe.roboflow.com/final-task/phonedetectv2.

Ngày nhận bài: 30/10/2025
Ngày nhận bản sửa: 19/11/2025
Ngày duyệt đăng: 27/11/2025