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

ỨNG DỤNG AI TRONG PHÁT HIỆN NGƯỜI ĐI XE MÁY KHÔNG ĐỘI MŨ BẢO HIỂM

AI APPLICATION IN DETECTING MOTORCYCLISTS WITHOUT HELMETS

NGUYỄN HỮU TUÂN*, VŨ THANH HƯƠNG
Khoa Công nghệ Thông tin, Trường Đại học Hàng hải Việt Nam
**Email liên hệ: huu-tuan.nguyen@vimaru.edu.vn*

Tóm tắt

Điều khiển xe máy không đội mũ bảo hiểm là một hành vi vi phạm Luật Trật tự, an toàn giao thông đường bộ năm 2024. Khi tai nạn xảy ra, hậu quả đối với những người đi xe máy không đội mũ bảo hiểm thường nguy hiểm và có nguy cơ ảnh hưởng tới tính mạng. Mặc dù Luật đã quy định nhưng vẫn có những người lái xe máy không đội mũ khi tham gia giao thông. Trong bài báo này, chúng tôi đề xuất ứng dụng trí tuệ nhân tạo sử dụng mô hình mạng YOLO11 đối với các hình ảnh thu nhận được từ các camera giám sát tại các điểm nút giao thông để giải quyết bài toán phát hiện người điều khiển xe máy không đội mũ bảo hiểm. Kết quả sau khi huấn luyện trên tập dữ liệu ảnh có 4267 ảnh chứng tỏ cách tiếp cận này có độ chính xác khá tốt (độ chính xác trung bình - mean Average Precision mAP@0,5 là 89,8% đối với phiên bản YOLO11n với quá trình huấn luyện 200 epoch). Kết quả thử nghiệm với các dữ liệu test tự thu thập sử dụng mô hình sau huấn luyện cũng cho thấy hệ thống ứng dụng mô hình YOLO11 có độ chính xác khá tốt trong phát hiện người điều khiển xe máy không đội mũ bảo hiểm và có tiềm năng ứng dụng trên thực tế.

Từ khóa: *YOLO11, phát hiện người đi xe máy không đội mũ bảo hiểm.*

Abstract

Riding a motorcycle without wearing a helmet is a violation of the 2024 Road Traffic Order and Safety Law. In the event of an accident, the consequences for motorcyclists not wearing helmets are often severe and can pose risks to their lives. Despite the law's stipulations, some motorcyclists continue to ride without helmets. In this article, we propose applying artificial intelligence with the YOLO11 network model to images obtained from surveillance cameras at traffic intersections to detect motorcyclists not wearing helmets. Training results on a dataset comprising 4,267 images demonstrate a promising training accuracy (mean Average Precision mAP@0.5 is 89.8% with the YOLO11n model after 200 epochs of training). Testing results using self-collected test data and the trained model also show that the YOLO11-based system achieves commendable accuracy in detecting motorcyclists without helmets and is practically applicable.

Keywords: *YOLO11, motorcyclist helmet detection.*

1. Mở đầu

Trong bối cảnh an toàn giao thông đường bộ ngày càng trở thành vấn đề cấp thiết, việc phát hiện, xử lý người tham gia giao thông vi phạm luật, như không đội mũ bảo hiểm khi đi xe máy, đóng vai trò thiết yếu trong việc giảm thiểu tai nạn giao thông và bảo vệ tính mạng người tham gia giao thông. Theo cách tiếp cận truyền thống, các phương pháp giám sát và phát hiện tình huống không đội mũ bảo hiểm của người đi xe máy thường dựa vào lực lượng cảnh sát giao thông có nhiệm vụ tuần tra, lập chốt kiểm tra và xử phạt trực tiếp. Tuy nhiên, cách tiếp cận này có nhiều hạn chế, thứ nhất là đòi hỏi một lực lượng nhân lực tham gia lớn nhưng vẫn không bao quát được hết các vị trí cần kiểm tra, phát hiện và xử lý vi phạm. Thứ hai là người vi phạm sẽ tìm cách tránh, né các điểm kiểm tra, giám sát của lực lượng chức năng nên dẫn tới không hiệu quả. Thứ ba là nguy cơ xảy ra sai sót do yếu tố con người. Với sự phát triển của công nghệ hiện đại, đặc biệt trong lĩnh vực trí tuệ nhân tạo, các giải pháp tự động hóa đã ra đời nhằm cải thiện hiệu quả phát hiện và xử lý vi phạm giao thông từ việc phân tích các hình ảnh thu nhận được từ các camera giám sát trên đường và tại các điểm nút giao thông. Trong bài báo này, chúng tôi nghiên cứu ứng dụng kiến trúc mạng YOLO11 nhằm phát hiện hành vi không đội mũ bảo hiểm qua hình ảnh từ các camera giám sát tại các giao lộ. Phương pháp này không chỉ giúp tự động hóa quá trình giám sát mà còn nâng cao độ chính xác, tiết kiệm nhân lực và tối ưu hóa chi phí. So với các cách tiếp

cận truyền thống, việc áp dụng công nghệ hiện đại như YOLO11 đánh dấu bước tiến đáng kể trong lĩnh vực quản lý giao thông, mang đến những lợi ích thiết thực và tiềm năng ứng dụng rộng rãi trong thực tế.

Để xây dựng một hệ thống thị giác máy tính nhằm phát hiện người đi xe máy không đội mũ bảo hiểm, Dahiya và các cộng sự [1] đã sử dụng các phương pháp trích chọn đặc trưng cục bộ như LBP (Local Binary Patterns) [2], SIFT (Scale Invariant Feature Transformation) [3] và HOG [4] trên một tập dữ liệu video quay trong 2 tiếng và đạt được kết quả tốt nhất với HOG (độ chính xác 93,8%) và thời gian xử lý khoảng 11,58ms cho 1 frame ảnh. Cách tiếp cận sử dụng các mô hình mạng học sâu như VGG [5], MobileNet [6], SSD [7] được Boonsirisumpun và các cộng sự sử dụng [8] với độ chính xác 85,4%. Trong [9], các tác giả đã sử dụng phương pháp học đa nhiệm dựa trên mạng CNN và đạt được kết quả với F-measure là 67,3% và tốc độ 8 khung hình 1 giây (8 FPS). Một mô hình phát hiện đối tượng dựa trên các mạng học sâu là Faster R-CNN [10] được áp dụng trong [11] với độ chính xác 92%. Các cách tiếp cận được đề cập trên đây có hai hạn chế: Một là đa số thực hiện trên các tập dữ liệu khá nhỏ với số lượng ảnh ít nên hạn chế về độ tin cậy và khả năng áp dụng thực tế, thứ hai là tốc độ xử lý bị hạn chế do sử dụng các mô hình mạng học sâu có kích thước khá lớn không đảm bảo khi đưa vào áp dụng thực tế đòi hỏi tốc độ cao theo thời gian thực.

Gần đây, một kiến trúc mới trong họ các mô hình mạng học sâu YOLO là YOLO11 [12] đã được đề xuất với khả năng phát hiện chính xác và tốc độ tốt hơn các phiên bản trước đó. Trong bài viết này, nhóm tác giả nghiên cứu ứng dụng mạng YOLO11 để xây dựng hệ thống phát hiện người điều khiển xe máy không đội mũ bảo hiểm. Nội dung bài báo được chia thành các phần như sau: Phần 2 bao gồm các chi tiết về mô hình mạng học sâu YOLO11, phần 3 trình bày về ứng dụng của YOLO11 cho bài toán phát hiện người điều khiển xe máy không đội mũ bảo hiểm và các kết quả huấn luyện, thực nghiệm, cuối cùng là phần Kết luận.

2. Mạng YOLO11 phát hiện đối tượng

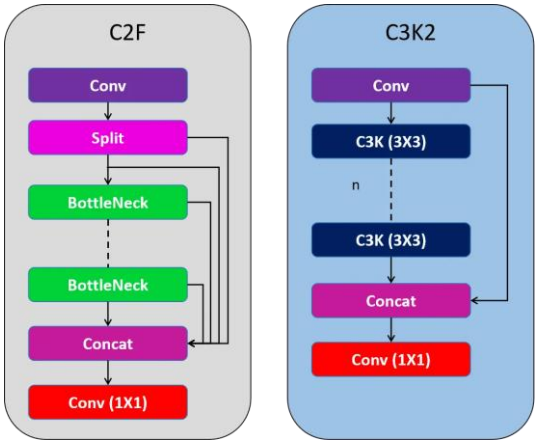
YOLO (thuật ngữ thành lập từ các chữ cái đầu của các từ You Only Look Once) [13] là một họ các mô hình mạng học sâu được thiết kế để thực hiện các thao tác cơ bản trong các bài toán thị giác máy tính là phát hiện đối tượng (Object Detection), Phân lớp (Classification), Phân đoạn ảnh (Segmentation). Một trong các lý do các mô hình YOLO thường được sử dụng giải quyết các bài toán phân tích nội dung ảnh sử dụng hướng tiếp cận phát hiện đối tượng là do tính chính xác và tốc độ thực hiện của chúng khá tốt và khả thi trên thực tế, kể cả với các hệ thống máy tính chỉ có CPU. Các phiên bản sau của YOLO như YOLOv8, YOLOv10, YOLO11 đều được xây dựng dựa trên việc kế thừa kiến trúc và các ưu điểm của phiên bản trước đo và bổ sung thêm các kỹ thuật mới và cải tiến trong phần kiến trúc của mạng. Xét về mặt kiến trúc mạng, phiên bản YOLO11 (xem chi tiết tại địa chỉ: [14]) dựa trên các phiên bản trước đó là YOLOv8, YOLOv9. YOLO11 (You Only Look Once version 11) là một cải tiến mới trong các mô hình phát hiện đối tượng thời gian thực dựa trên deep learning. YOLO11 có nhiều điểm ưu điểm so với các phiên bản trước nhờ sử dụng kiến trúc backbone hiệu quả hơn, được tối ưu hóa để giảm chi phí tính toán mà vẫn đạt được độ chính xác cao. Một trong những điểm nổi bật của YOLO11 là việc áp dụng cơ chế chú ý (attention mechanism) để tăng khả năng phát hiện các đối tượng ckích thước nhỏ hoặc các đối tượng trong điều kiện quan sát phức tạp. Kiến trúc của YOLO11 được hình thành dựa trên các lớp convolutional cải tiến, kết hợp với các khối residual để tăng cường khả năng huấn luyện mạng sâu. Mô hình này sử dụng head prediction đa cấp độ, cho phép phát hiện đối tượng ở nhiều kích thước khác nhau. YOLO11 cũng được trang bị thuật toán non-maximum suppression (NMS) cải tiến để giảm các lỗi chồng lặp vùng dự đoán. Mô hình hỗ trợ tích hợp các hàm mất mát (loss functions) tiên tiến, như CIOU và Focal Loss, để cải thiện chất lượng dự đoán các bounding box. Với khả năng xử lý thời gian thực, YOLO11 phù hợp với các tình huống yêu cầu tốc độ và độ chính xác cao.

Về cơ bản sự khác biệt và cải tiến của YOLO11 so với các mô hình YOLOv8, YOLOv9 và YOLOv10 thể hiện ở 3 yếu tố: Các khối kiến trúc C3K2 ở phần mạng xương sống (backbone), kỹ thuật SPFF (Spatial Pyramid Pooling Fast) ở phần cổ của mạng (neck) và cơ chế chú ý C2PSA (Cross Stage Partial with Spatial Attention).

Kiến trúc khối nhân chập mới C3K2: YOLOv11 sử dụng các khối C3K2 (xem chi tiết trong Hình 1) để xử lý trích xuất đặc trưng ở các giai đoạn khác nhau của phần backbone. Các nhân kích thước 3x3 nhỏ hơn cho phép tính toán hiệu quả hơn trong khi vẫn giữ nguyên khả năng của mô hình trong việc nắm bắt các tính năng cần thiết trong hình ảnh. Trọng tâm của phần backbone YOLO11 là khối C3K2, đây là sự phát triển của khối CSP (Cross Stage Partial) được giới thiệu trong các phiên bản trước đó. Khối C3K2 tối ưu hóa luồng thông tin qua mạng bằng cách chia tách bản đồ

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

đặc trưng và áp dụng một loạt các phép tích chập nhân nhỏ hơn (3x3), nhanh hơn và rẻ hơn về mặt tính toán so với các phép tích chập nhân kích thước lớn. Bằng cách xử lý các bản đồ đặc trưng nhỏ hơn, riêng biệt và hợp nhất chúng sau một số phép tích chập, khối C3K2 cải thiện biểu diễn tính năng với ít tham số hơn so với các khối C2f của YOLOv8. Khối C3K có cấu trúc tương tự như khối C2F nhưng không có sự phân tách nào được thực hiện ở đây, đầu vào được truyền qua một khối Conv theo sau là một loạt các lớp Bottle Neck có nối và kết thúc bằng khối Conv cuối cùng.



Hình 1. Khối nhân chập trong YOLO11 so với YOLOv8

Kỹ thuật SPFF: YOLOv11 vẫn giữ lại mô-đun SPFF (Spatial Pyramid Pooling Fast), được thiết kế để gộp các đặc điểm từ các vùng khác nhau của một hình ảnh ở nhiều tỷ lệ khác nhau. Điều này tăng sức mạnh của mạng trong việc phát hiệm các đối tượng có kích thước thay đổi, đặc biệt là các đối tượng nhỏ, vốn là một thách thức đối với các phiên bản YOLO trước đó. SPFF gộp các đặc điểm bằng nhiều hoạt động gộp tối đa (với các kích thước bộ lọc khác nhau) để tổng hợp thông tin theo ngữ cảnh đa tỷ lệ. Mô-đun này đảm bảo rằng ngay cả các đối tượng nhỏ cũng được mô hình nhận dạng, vì nó kết hợp hiệu quả thông tin trên các độ phân giải khác nhau. Việc đưa SPFF vào đảm bảo rằng YOLOv11 có thể đạt được tốc độ thời gian thực đồng thời tăng cường khả năng phát hiện các đối tượng trên nhiều tỷ lệ. Tuy nhiên điểm khác biệt ở đây là khối C2PSA được sử dụng ngay sau khối SPFF.

Khối CSPSA: Khối C2PSA sử dụng hai mô-đun PSA (Partial Spatial Attention) hoạt động trên các nhánh riêng biệt của bản đồ đặc trưng và sau đó được nối lại, tương tự như cấu trúc khối C2F. Thiết lập này đảm bảo mô hình tập trung vào thông tin không gian trong khi vẫn duy trì sự cân bằng giữa chi phí tính toán và độ chính xác phát hiện. Khối C2PSA tinh chỉnh khả năng tập trung có chọn lọc vào các vùng quan tâm của mô hình bằng cách áp dụng sự chú ý không gian vào các đặc điểm được trích xuất. Điều này cho phép

Hình 2. Ảnh người điều khiển xe máy, mũ bảo hiểm trong tập huấn luyện

YOLOv11 hoạt động tốt hơn các phiên bản trước như YOLOv8 trong các tình huống cần có chi tiết đối tượng tốt để phát hiện chính xác.

Khả năng phát hiện đối tượng của YOLO11: Theo [12] và các số liệu được công bố tại website của Ultralytics [15] thì YOLO11 có khả năng phát hiện chính xác hơn và tốc độ nhanh hơn khi đối sánh với các phiên bản YOLO trước.

3. Ứng dụng YOLO11 phát hiện người điều khiển xe máy không đội mũ bảo hiểm

Với mục tiêu ứng dụng mô hình YOLO11 xây dựng hệ thống phát hiện người điều khiển xe máy không đội mũ bảo hiểm, chúng tôi đã lựa chọn kiến trúc mạng học sâu YOLO11 do độ chính xác cao và tốc độ phát hiện nhanh chóng mà nó mang lại.

Để xây dựng hệ thống phát hiện người điều khiển xe máy không đội mũ bảo hiểm, nhóm tác giả đã tìm kiếm, đánh giá và sử dụng 1 tập dữ liệu công cộng gồm 4267 ảnh (https://universe.roboflow.com/abdullah-kqdi3/no-helmet-no-ride) với tập huấn luyện có 3912 ảnh và tập kiểm thử trong quá trình huấn luyện có 355 ảnh (xem ảnh minh họa trong Hình 2). Hai tập ảnh này gồm các ảnh chụp có kích thước tối đa 640x640 được chọn lọc từ các camera giám sát và ảnh chụp đường phố của các nước trong khu vực Châu Á như Thái Lan, Ấn Độ, Myanmar, Việt Nam nên khá tương đồng với Việt Nam. Quá trình huấn luyện đã sử dụng phương pháp học chuyển tiếp các mô hình YOLO11 huấn luyện trên tập ảnh COCO và các kỹ thuật tăng cường dữ liệu bằng các thuật toán xử lý ảnh như biến đổi HSV, xoay ảnh, lật ảnh theo chiều ngang, dịch ảnh, thay đổi tỉ lệ, tạo ảnh khảm (mosaic) nên trên thực tế số ảnh sử dụng để huấn luyện là 23472 ảnh (6x3912).

Cũng như các phiên bản trước, YOLOv11 có 5 cấu hình là siêu nhỏ - Nano (n), nhỏ - Small (s), trung bình - Medium (m), lớn - Large (l) và siêu lớn - Extra Large (x). Các cấu hình này khác nhau về số lượng tham số, do đó sẽ khác nhau về độ chính xác và tốc độ huấn luyện, thực hiện. Cấu hình Nano là nhỏ nhất với số tham số ít nhất và cũng phù hợp nhất với các đối tượng kích thước nhỏ và ảnh huấn luyện có độ phân giải hạn chế trong khi cấu hình Extra Large có số tham số nhiều nhất, độ chính xác theo lý thuyết là cao nhất và tốc độ huấn luyện và phát hiện đối tượng chậm nhất. Trong bài báo này, nhóm tác giả đã huấn luyện và thử nghiệm cả 5 cấu hình của YOLO11 với thư viện Pytorch 2.6 trên nền tảng CUDA 12.4 với 1 GPU RTX 3060 12 GB RAM. Toàn bộ các mô hình được huấn luyện với độ phân giải ảnh là 640x640 trong 200 epoch. Kết quả trong bảng 1 chứng tỏ cách tiếp cận ứng dụng mô hình mạng YOLO11 cho bài toán phát hiện người đi xe máy không đội mũ bảo hiểm là đúng và có hiệu quả với độ chính xác mAP@0,5 của toàn bộ 5 cấu hình YOLOv11 đều khá cao: Cấu hình

Bảng 1. Kết quả huấn luyện YOLO11 phát hiện người điều khiển xe máy không đội mũ bảo hiểm

TT	Phiên bản	Kết quả huấn luyện mAP (%)		Kết quả với dữ liệu kiểm tra mAP(%)	
		@0.50	@0.50-0.95	@0.50	@0.5-0.95
1	YOLO11n	89.8%	60.4%	89.7%	60.3%
2	YOLO11s	88.6%	61.4%	88.7%	61.3%
3	YOLO11m	87.6%	62.5%	87.6%	62.4%
4	YOLO11l	87.5%	61.9%	87.6%	61.9%
5	YOLO11x	88.6%	60.9%	89.8%	60.8%
6	YOLOv8m	87.7%	61.5%	87.9%	61.5%

Hình 3. Kết quả phát hiện người đi xe máy không đội mũ bảo hiểm với dữ liệu test

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

YOLO11l và YOLO11x đạt kết quả thấp nhất là khoảng 87% và cấu hìn đạt độ chính xác cao nhất là 89,8% với cấu hình YOLO11n. Các kết quả đối mAP@0,5-0,95 trên dữ liệu đào tạo cũng khá tốt khi có giá trị tối thiểu 60,3% với cấu hình Nano và tăng đôi chút với các cấu hình còn lại. Các kết quả mAP@0,5 và mAP@0,5-0,95 trên cơ sở dữ liệu ảnh kiểm tra cũng tương đương (khác biệt không quá lớn) với các số liệu trên tập ảnh huấn luyện. Để kiểm tra các kết quả của cách tiếp cận sử dụng YOLO11 là đúng đắn, chúng tôi đã huấn luyện thêm các mô hình của YOLOv8 trên cùng tập dữ liệu và chọn kết quả tốt nhất đạt được (với mô hình YOLOv8m) và so sánh trong Bảng 1. Có thể thấy kết quả tốt nhất của YOLOv8 thấp hơn YOLO11 (phiên bản YOLO11m) khoảng 1% độ chính xác @mAP0,5-0,95 trên cả tập huấn luyện và tập kiểm tra.

Để kiểm tra tính chính xác các mô hình đã đào tạo, nhóm tác giả đã thu thập một số video thực tế (5 video) từ mạng Internet và kết quả cho thấy các mô hình sau huấn luyện có khả năng phát hiện được người điều khiển xe máy không đeo mũ bảo hiểm khá tốt (xem minh họa trong Hình 3). Để đánh giá tốc độ hướng tới việc triển khai thực tế, chúng tôi đã chuyển đổi format các mô hình từ Pytorch (.pt) sang TensorRT (.engine) và thử nghiệm. Kết quả thực hiện với video thì tốc độ xử lý đạt được là 51,8 khung hình 1 giây. Tốc độ này cho thấy cách tiếp cận của bài báo có khả năng ứng dụng vào thực tế với yêu cầu xử lý dữ liệu theo tốc độ thời gian thực.

4. Kết luận

Với mục đích ứng dụng các mô hình AI để nhằm xây dựng một hệ thống phát hiện người điều khiển xe máy không đội mũ bảo hiểm, nhóm tác giả đã đề xuất sử dụng kiến trúc mạng YOLO11. Chúng tôi đã huấn luyện các cấu hình khác nhau của YOLO11 trên một cơ sở dữ liệu ảnh và kiểm tra trên tập dữ liệu ảnh test cũng như dữ liệu thực tế. Các kết quả nhận được (tỉ lệ phát hiện chính xác tính theo mAP@0,5 và mAP@0,5-0,95, tốc độ thực hiện) cho thấy cách tiếp cận của chúng tôi là đúng và kết quả có thể triển khai thực tế. Sắp tới, nhóm tác giả sẽ bổ sung thêm dữ liệu thực tế, đánh nhãn, huấn luyện, đánh giá nhằm nâng cao độ chính xác cũng như kết hợp với các mô hình khác để có thể ứng dụng vào thực tế.

Lời cảm ơn

Nghiên cứu này được tài trợ bởi Trường Đại học Hàng hải Việt Nam trong đề tài mã số: **DT24-25.75**.

TÀI LIỆU THAM KHẢO

[1] K. Dahiya, D. Singh, and C. K. Mohan (2016), *Automatic detection of bike-riders without helmet using surveillance videos in real-time*, in 2016 International Joint Conference on Neural Networks (IJCNN), Vancouver, BC, Canada: IEEE, Jul. 2016, pp.3046-3051.

[2] T. Ojala, M. Pietikainen, and D. Harwood (1994), *Performance evaluation of texture measures with classification based on Kullback discrimination of distributions*, in Proceedings of 12th International Conference on Pattern Recognition, Vol.1, pp.582-585.

[3] D. G. Lowe (2004), *Distinctive Image Features from Scale-Invariant Keypoints*, Int. J. Comput. Vis., Vol.60, No.2, pp.91-110.

[4] N. Dalal and B. Triggs (2005), *Histograms of oriented gradients for human detection*, in 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), Vol.1, pp.886-893.

[5] K. Simonyan and A. Zisserman (2015), *Very Deep Convolutional Networks for Large-Scale Image Recognition*, arXiv:1409.1556.

[6] A. G. Howard et al. (2017), *MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications*, arXiv:1704.04861.

[7] W. Liu et al. (2016), *SSD: Single Shot MultiBox Detector*, Vol.9905, pp.21-37.

[8] N. Boonsirisumpun, W. Puarungroj, and P. Wairotchanaphuttha (2018), *Automatic Detector for Bikers with no Helmet using Deep Learning*, in 2018 22nd International Computer Science and Engineering Conference (ICSEC), Chiang Mai, Thailand: IEEE, pp.1-4.

[9] H. Lin, J. D. Deng, D. Albers, and F. W. Siebert (2020), *Helmet Use Detection of Tracked Motorcycles Using CNN-Based Multi-Task Learning*, IEEE Access, Vol.8, pp.162073-162084.

[10] S. Ren, K. He, R. Girshick, and J. Sun (2016), *Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks*, arXiv:1506.01497.

[11] S. Raju, S. P. Paul, S. Sajeev, and A. Johny (2025), *Detection of Helmetless Riders Using Faster R-CNN*, Accessed: Mar. 24, 2025.

[12] R. Khanam and M. Hussain (2024), *YOLOv11: An Overview of the Key Architectural Enhancements*, arXiv:2410.17725.

[13] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi (2016), *You Only Look Once: Unified, Real-Time Object Detection*, arXiv:1506.02640.

[14] S. N. Rao (2025), *YOLOv11 Explained: Next-Level Object Detection with Enhanced Speed and Accuracy*, Medium.

[15] Ultralytics (2025), *YOLO11 NEW*. Accessed: Mar. 24, 2025. [Online-Available] https://docs.ultralytics.com/models/yolo11.

Ngày nhận bài:	25/03/2025
Ngày nhận bản sửa:	10/04/2025
Ngày duyệt đăng:	11/04/2025