VẬT LIỆU & XÂY DỰNG
Tạp chí Vật liệu và Xây dựng
ISSN 1859 - 381X
Website: www. jomc.vn

Ứng dụng mô hình học sâu YOLOv8 để kiểm soát hành vi sử dụng phương tiện bảo vệ cá nhân trên công trường xây dựng

Nguyễn Ngọc Thoan[1], Nguyễn Anh Đức [1]*, Đào Chí Hiếu[1]

[1]Bộ môn Công nghệ và Quản lý xây dựng, Khoa Xây dựng Dân dụng và Công nghiệp, Trường Đại học Xây dựng Hà Nội, số 55 đường Giải Phóng, quận Hai Bà Trưng, Hà Nội, Việt Nam

TỪ KHOÁ

Học sâu
YOLOv8
Quản lý an toàn
Phát hiện đối tượng
Phương tiện bảo vệ cá nhân

TÓM TẮT

Trong quá trình thi công tại các công trường xây dựng, việc giám sát các hành vi sử dụng phương tiện bảo vệ cá nhân của công nhân luôn cần sự quan tâm kỹ càng, nhằm tránh xảy ra tai nạn đáng tiếc. Cùng với sự phát triển của khoa học kỹ thuật, công nghệ hình ảnh và các thuật toán trí tuệ nhân tạo liên tục được nghiên cứu và áp dụng vào các công tác ngoài hiện trường. Nghiên cứu này đề xuất một phương pháp ứng dụng mô hình học sâu YOLOv8 – một trong những mô hình tiên tiến nhất – để tự động xác định hành vi sử dụng phương tiện bảo vệ cá nhân của công nhân trên công trường, bao gồm các hành vi sử dụng và không sử dụng phương tiện bảo vệ cá nhân. Kết quả của nghiên cứu cho thấy các phát hiện được cải thiện đáng kể so với những mô hình trước đó, cụ thể độ chính xác với phát hiện mũ bảo hộ, găng tay và khẩu trang lần lượt lên tới 96,6 %; 99,6 % và 98,2 %. Không chỉ vậy, mô hình còn có ý nghĩa rất lớn khi phát hiện được các hành vi không sử dụng phương tiện bảo vệ cá nhân, những trường hợp vốn cần được quan tâm hơn trong công tác giám sát an toàn lao động. Do đó, mô hình học sâu đề xuất có khả năng ứng dụng thực tiễn vào các công trường xây dựng, tăng cường khả năng quản lý, kiểm soát việc sử dụng phương tiện bảo vệ cá nhân, đảm bảo an toàn lao động cho công nhân khi thi công tại các công trường.

KEYWORDS

YOLO
Deep learning
Safety management
Object detection
Personal protective equipment (PPE)

ABSTRACT

The supervision of workers' use of PPE is of utmost importance to prevent worksite accidents. The development of image processing, computer vision, and machine learning algorithms has allowed achievements in such safety procedures. This study proposes a method that applies the YOLOv8 deep learning model – one of the most advanced algorithms – to automatically detect the use of safety protective equipment by workers on construction sites, including behaviors of both using and not using safety protective equipment. The research results show significant improvements compared to previous models, specifically in accuracy rates for detecting helmets, gloves, and masks, which reach 96.6%, 99.6%, and 98.2% respectively. Moreover, the model also holds considerable significance in detecting instances of non-compliance with PPE, cases that require increased attention in occupational safety monitoring. Therefore, the proposed deep learning model is capable of practical application on construction sites, enhancing management capabilities and control over PPE usage, thereby ensuring occupational safety for workers during construction activities.

1. Giới thiệu

Trong lĩnh vực xây dựng, người lao động luôn phải đối mặt với những nguy cơ thương tích cao hơn nhiều khi so với những ngành nghề khác. Tỷ lệ tử vong hay việc xảy ra các tai nạn lao động trong quá trình thi công công trình luôn cao hơn mặt bằng chung ở nhiều quốc gia, trong đó có cả ở Việt Nam. Tuy vậy, công tác đảm bảo an toàn lao động tại nhiều công trường đôi khi không được chú trọng, hoặc không thể kiểm soát được hoàn toàn do các vấn đề đặc thù trong môi trường lao động xây dựng. Thống kê từ nghiên cứu của Bhole [1] cho thấy, ngã từ trên cao chiếm 28 %, trượt ngã chiếm 15 %, điện giật chiếm 9 %, va chạm bởi vật di chuyển chiếm 8% các ca tử vong trong tai nạn trên công trường xây dựng. Các dạng tai nạn này đều có thể được giảm thiểu cả về số lượng và mức độ nghiêm trọng thông qua giám sát an toàn lao động. Điều này cho thấy tầm quan trọng của việc cải thiện quá trình giám sát an toàn lao động trong các công trường xây dựng.

Các phương tiện bảo vệ cá nhân được kể đến như quần áo bảo hộ, thiết bị hoặc dụng cụ chuyên dụng được thiết kế để bảo vệ người lao động khỏi các mối nguy hiểm khác nhau tại nơi làm việc. Việc nghiên cứu ứng dụng các công nghệ kỹ thuật để giám sát hành vi sử dụng các phương tiện bảo vệ cá nhân trở thành một yếu tố vô cùng quan trọng, cần được triển khai vào các hệ thống công trường nhanh

*Liên hệ tác giả: ducna@huce.edu.vn
Nhận ngày 28/06/2024, sửa xong ngày 04/07/2024, chấp nhận đăng ngày 15/07/2024
Link DOI: https://doi.org/10.54772/jomc.04.2024.755

chóng để giảm thiểu các rủi ro. Những năm gần đây, các nghiên cứu cứu tập trung vào việc phát hiện hành vi sử dụng phương tiện bảo vệ cá nhân (Personal Protective Equipment - PPE) đã thu hút được sự quan tâm đáng kể trong lĩnh vực xây dựng [2, 3]. Công trường thi công luôn tồn tại những nguy hiểm khó có thể kiểm soát được hoàn toàn, vì vậy, vấn đề an toàn cá nhân cho người lao động làm việc tại hiện trường luôn cần được giám sát sát sao.

2. Các nghiên cứu liên quan

Hiện nay đã có hai loại kỹ thuật giám sát được nghiên cứu áp dụng, bao gồm các phương pháp phát hiện dựa trên cảm biến (sensor-based) và các phương pháp dựa trên công nghệ thị giác (vision-based) [4]. Các phương pháp liên quan đến việc sử dụng cảm biến sẽ thu thập dữ liệu hoặc thông tin từ môi trường xung quanh, dữ liệu được thu thập sau đó sẽ được phân tích và xem xét ứng dụng ngoài thực tế [5-7]. Tuy nhiên, các phương pháp dựa trên cảm biến bắt buộc phải gắn thẻ thủ công hoặc cảm biến tích hợp vào phương tiện bảo vệ cá nhân, dẫn đến việc thiết lập phức tạp, làm tăng các chi phí thiết bị, ảnh hưởng đến thao tác của công nhân, do đó gây trở ngại cho việc áp dụng vào các công tác ngoài thực tế.

Các phương pháp tiếp cận dựa trên thị giác máy tính, khác với các phương pháp dựa trên cảm biến, được ứng dụng trong nhiều ngành khác nhau và sử dụng dữ liệu hình ảnh bằng máy quay phim hoặc các cảm biến hình ảnh. Phương pháp này toàn diện, chính xác hơn và có thể dễ áp dụng trong các công trường xây dựng phức tạp chỉ với các camera giám sát phổ thông [8]. Park và cộng sự [9] đã giới thiệu một kỹ thuật cải tiến dựa trên thị giác có khả năng tự động phát hiện cả cơ thể người và mũ bảo hiểm một cách hiệu quả và đồng thời trong mỗi khung hình đoạn phim quay được. Phương pháp này được xây dựng thông qua phép trừ nền (background substraction) và sử dụng biểu đồ của các tính năng gradient định hướng (Histogram of oriented gradient - HOG). Các tác giả này sử dụng tập dữ liệu từ các đoạn phim quay tại công trường xây dựng, ban đầu xác định các đối tượng thông qua ma trận độ lệch chuẩn và sau đó áp dụng bộ mô tả HOG để xác định hành vi sử dụng mũ bảo hộ.

Các kỹ thuật phát hiện sử dụng chuỗi hình ảnh bao gồm việc xác định chuyển động của khuôn mặt và đặc điểm của chúng, cùng với việc thu thập và phân tích dữ liệu liên quan đến màu sắc [10], cũng như phát hiện góc cạnh bằng kỹ thuật dựa trên độ dốc trong hình ảnh [11] đã được sử dụng để theo dõi hành vi sử dụng phương tiện bảo vệ cá nhân của công nhân ngoài công trường. Ngoài ra, các phương pháp học máy dựa trên HOG được sử dụng trong nhiều ứng dụng khác nhau, chẳng hạn như xác định các hành vi không an toàn trên công trường, trong đó chúng liên quan đến các quy trình nhiều giai đoạn tạo ra các tính năng tùy chỉnh để đánh giá. Tuy nhiên, những phương pháp này gặp phải thách thức lớn trong việc phát hiện chính xác hành vi tuân thủ sử dụng phương tiện bảo vệ cá nhân do điều kiện thời tiết thay đổi, các góc nhìn khác nhau và các trường hợp có vật cản trong khung hình.

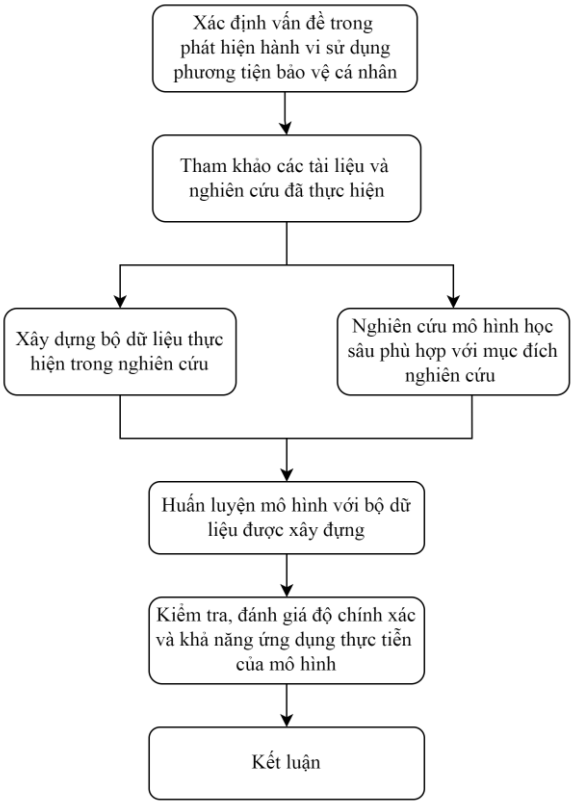
Với nhiều ưu điểm vượt trội, các kỹ thuật học sâu được giới thiệu trong khoảng thời gian gần đây đã thu hút được sự quan tâm đáng kể trong lĩnh vực phát hiện vật thể vì tính mạnh mẽ của chúng trong việc xử lý các đặc điểm dữ liệu đa quy mô. Fang và các cộng sự [12] đã sử dụng faster R-CNN để tự động giám sát việc sử dụng mũ bảo hộ trên các công trường xây dựng, bằng cách huấn luyện với bộ dữ liệu hình ảnh gồm 81.000 ảnh để đào tạo và thử nghiệm. Tuy nhiên, mô hình này phải đối mặt với những thách thức trong việc xác định chính xác màu sắc của mũ bảo hộ lao động. Kolar và cộng sự [13] đã tạo ra một mô hình phát hiện lan can bảo vệ an toàn bằng cách sử dụng mạng nơ-ron tích chập (Convolutional neural network - CNN) với các dữ liệu hình ảnh. Ding và cộng sự [14] đã xây dựng một mô hình học sâu kết hợp giữa CNN và mô hình trí nhớ ngắn – dài hạn (Long short-term memory - LSTM) có thể trích xuất và phân loại chính xác các hành vi không an toàn từ dữ liệu hình ảnh thu được trong các máy quay phổ thông. Tuy nhiên, mô hình này chỉ dành riêng cho một thiết bị duy nhất và cần nhiều nghiên cứu hơn cho các tình huống liên quan đến nhiều thiết bị hoặc công nhân đồng thời làm việc trên công trường. Siddula và cộng sự [15] tập trung vào việc đánh giá hiệu suất của CNN bằng cách kết hợp mô hình hỗn hợp Gaussian để xác định các mục tiêu di động trên mái nhà của các công trường xây dựng. Nath và cộng sự [16] gần đây đã thiết kế các mô hình dựa trên CNN cho việc phân loại một hoặc nhiều nhãn với các đối tượng trong xây dựng. Các phương pháp này tuy có độ chính xác tương đối cao, nhưng lại tốn kém về mặt tốc độ xử lý, thường không đáp ứng được yêu cầu từ các tình huống theo thời gian thực.

Trong các nghiên cứu đã được tiến hành, các mô hình đã được nghiên cứu chủ yếu tập trung vào việc phát hiện phương tiện bảo vệ cá nhân của công nhân, có thể kể đến như mũ bảo hộ, áo, khẩu trang, găng tay. Tuy nhiên, bên cạnh việc xác định các hành vi có sử dụng phương tiện bảo vệ cá nhân, hành vi không sử dụng các phương tiện bảo vệ cá nhân đa phần lại không được quan tâm nghiên cứu phát hiện, mặc dù các hành vi này lại nguy hiểm hơn và cần được phát hiện kịp thời ngoài công trường. Trên thực tế, các hành vi không sử dụng phương tiện bảo vệ cá nhân, hay là việc phát hiện sự thiếu hụt của các loại (nhãn) phương tiện bảo hộ phổ biến mới là vấn đề cần được quan tâm hơn. Vì vậy, trong nghiên cứu này, một bộ dữ liệu bao gồm các nhãn hành vi không sử dụng phương tiện bảo vệ cá nhân đã được nhóm nghiên cứu tiến hành xây dựng, sau đó dữ liệu được huấn luyện với mô hình YOLOv8 để kiểm tra khả năng làm việc của mô hình đối với các dữ liệu hành vi này. Các kết quả đạt được trong nghiên cứu sẽ làm cơ sở để ứng dụng phương pháp đề xuất vào thực tế công trường, khi các hành vi không sử dụng phương tiện cá nhân này luôn cần được phát hiện kịp thời để tránh xảy ra các rủi ro về an toàn lao động.

3. Phương pháp nghiên cứu

Hình 1 thể hiện các bước phát triển và triển khai mô hình học sâu tự động phát hiện hành vi sử dụng phương tiện bảo vệ cá nhân của công nhân trên công trường. Bắt đầu bằng việc xác định các vấn đề

trong việc kiểm soát hành vi sử dụng trang thiết bị bảo vệ cá nhân của công nhân hoạt động trên công trường nói chung. Với việc nghiên cứu các tài liệu về các nghiên cứu đã được thực hiện, một bộ dữ liệu hình ảnh chứa các nhãn hành vi sử dụng và không sử dụng phương tiện bảo vệ cá nhâncùng với một mô hình học sâu YOLOv8 được lựa chọn để thực hiện nghiên cứu. Sau khi đã xây dựng dược bộ dữ liệu cùng mô hình học sâu, việc huấn luyện sẽ được tiến hành và kiểm tra các kết quả. Cuối cùng, mô hình sẽ được kiểm tra với các hình ảnh ngoài công trường thực tế để đưa ra các kết luận về tính hiệu quả khi ứng dụng vào quá trình theo dõi hành vi sử dụng phương tiện bảo vệ cá nhân trên công trường xây dựng.



Hình 1. Trình tự các bước thực hiện nghiên cứu.

4. Cơ sở lý thuyết

4.1. Mô hình học sâu YOLOv8

Với những ưu điểm vượt trội khi là mô hình mới nhất được phát triển, đặc biệt là khả năng phát hiện nhanh và chính xác với các tác vụ thời gian thực, YOLOv8 rất phù hợp với mục đích của nghiên cứu. YOLOv8 là mô hình YOLO mới nhất được giới thiệu bởi Glen Jocher, người sáng lập và là giám đốc điều hành của Ultralytics – một công ty hàng đầu trong việc phát triển các thuật toán tiên tiến thuộc lĩnh vực trí tuệ nhân tạo. Kế thừa các thế hệ YOLO trước, vốn đã được ứng dụng rất thành công trong các thao tác nhận dạng trong giao thông, y học, thể thao [17], kiến trúc của mô hình YOLOv8 được thay đổi, bao gồm 3 thành phần chính, trong đó bao gồm phần xương sống (backbone), phần cổ (Neck) và phần đầu (Head).

4.1.1. Phần xương

YOLOv8 sử dụng cấu trúc CSPDarknet53 đã được chỉnh sửa làm mạng xương sống, mỗi dữ liệu hình ảnh đầu vào được lấy các đặc trưng hình ảnh ở năm cấp độ, tương ứng với 5 lần giảm kích thước dữ liệu, lần lượt được ký hiệu là B1 – B5. Cấu trúc của mạng xương sống được thể hiện trong Hình 2. Mô-đun Cross Stage Partial (CSP) trong mạng xương sống trước đó được thay thế bằng mô-đun C2f (Hình 2f). Mô-đun C2f sử dụng kết nối shunt gradient để làm phong phú luồng thông tin của mạng trích xuất đặc trưng trong khi vẫn duy trì được trọng lượng nhẹ. Mô-đun CBS thực hiện thao tác tích chập trên thông tin đầu vào, sau đó là chuẩn hóa hàng loạt và cuối cùng kích hoạt luồng thông tin bằng SiLU để thu được kết quả đầu ra, như trong Hình 2g. Mạng xương sống cuối cùng sử dụng mô-đun tổng hợp nhanh hình kim tự tháp không gian (spatial pyramid pooling fast - SPPF) để gộp các bản đồ tính năng đầu vào thành một bản đồ có kích thước cố định cho đầu ra có kích thước tương ứng. So với cấu trúc tổng hợp kim tự tháp không gian (spatial pyramid pooling - SPP), SPPF giảm khối lượng tính toán và thời gian xử lý nhanh hơn bằng cách kết nối tuần tự ba lớp tổng hợp tối đa, như trong Hình 2d.

4.1.2. Phần cổ

Lấy cảm hứng từ cấu trúc PANet, YOLOv8 được thiết kế với cấu trúc PAN-FPN ở phần cổ, như trong Hình 2b. So với cấu trúc cổ của các mẫu YOLOv5 và YOLOv7, YOLOv8 loại bỏ thao tác tích chập sau khi lấy mẫu lên trong cấu trúc PAN, giúp duy trì hiệu suất ban đầu trong khi đạt được mô hình nhẹ. Cấu trúc FPN (Feature Pyramid Network – mạng lưới kim tự tháp đặc trưng) thông thường sử dụng cách tiếp cận từ trên xuống để truyền tải các thông tin ngữ nghĩa sâu sắc. FPN tăng cường thông tin ngữ nghĩa của các tính năng bằng cách kết hợp B4-P4 và B3-P3, nhưng một số thông tin bản địa hóa đối tượng sẽ bị mất. Để giảm bớt vấn đề này, cấu trúc PAN-FPN thêm một mạng tổng hợp đường dẫn (PAN – Path Aggregation Networks) vào trong FPN. Cấu trúc PAN này tăng cường việc tìm hiểu thông tin vị trí bằng cách kết hợp các lớp P4-N4 và B5-N5 để nhận ra việc nâng cao đường dẫn ở dạng từ trên xuống. PAN-FPN xây dựng cấu trúc mạng từ trên xuống và từ dưới lên, giúp nhận ra sự bổ sung của thông tin vị trí nông và thông tin ngữ nghĩa sâu thông qua phản ứng tổng hợp tính năng, dẫn đến tính đa dạng và đầy đủ của tính năng.

4.1.3. Phần đầu

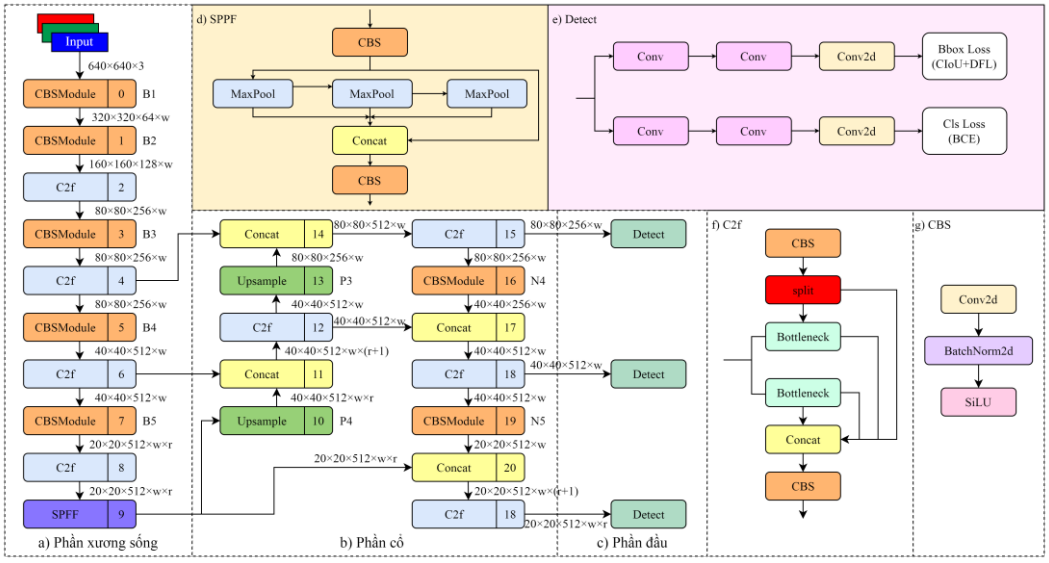
Phần phát hiện đối tượng của mô hình YOLOv8 sử dụng cấu trúc đầu phân nhánh, như trong Hình 2e. Cấu trúc phần đầu sử dụng hai nhánh riêng biệt để phân loại đối tượng và đưa ra giá trị dự đoán nhãn với mỗi phát hiện, đồng thời sử dụng các hàm mất mát khác nhau cho

hai loại nhiệm vụ này. Đối với nhiệm vụ phân loại, mất mát entropy chéo nhị phân (binary cross-entropy loss – BCE loss) được sử dụng. Đối với tác vụ hồi quy giới hạn hộp được dự đoán, tổn thất tiêu điểm phân phối (distribution focal loss - DFL) [19] và CIoU [20] được sử dụng. Cấu trúc phát hiện này có thể cải thiện độ chính xác và tăng tốc độ hội tụ mô hình, làm tăng hiệu quả khi huấn luyện.

5. Thử nghiệm, kết quả và phân tích

5.1. Xây dựng bộ dữ liệu sử dụng phương tiện bảo vệ cá nhân trên công trường xây dựng

Các dữ liệu hình ảnh hành vi sử dụng phương tiện bảo vệ cá nhân của công nhân trong nghiên cứu được thu thập tại nhiều công trường tại Việt Nam, kết hợp cùng với các dữ liệu được chia sẻ trên các nền tảng mã nguồn mở. Tổng cộng 960 hình ảnh chứa các hành vi sử dụng phương tiện bảo vệ cá nhân của công nhân đã được thu thập tại nhiều công trường trong nước. Những hình ảnh được thu thập sau đó được tổng hợp lại và gán nhãn trên nền tảng Roboflow, một nền tảng mở rộng phục vụ cho quá trình xử lý dữ liệu cho các nghiên cứu ứng dụng mô hình YOLO. Trong nghiên cứu này, các hành vi được theo dõi bao gồm mang các phương tiện bảo vệ cá nhân của công nhân như các nhãn mũ bảo hộ (*helmet*), áo phản quang (*safety vest*), khẩu trang (*mask*), găng tay (*gloves*) và giày (*shoes*), cùng với đó các hành vi không mang phương tiện bảo hộ cũng được nhóm nghiên cứu ghi lại (*NO_helmet, NO_safety vest, NO_mask, NO_gloves, NO_shoes*). Như vậy, tổng cộng có 5 nhãn phát hiện hành vi có sử dụng phương tiện bảo hộ với 5 nhãn không sử dụng. Thử nghiệm được tiến hành trên ba bộ dữ liệu, trong đó, bộ dữ liệu *Use* được gán các nhãn nhằm xác định phương tiện bảo vệ cá nhân có được mặc trên người công nhân. Bộ dữ liệu *NotUse* đánh dấu phát hiện các hành vi không tuân thủ quy định trang bị phương tiện bảo vệ cá nhân, như việc không đội mũ bảo hộ (*NO_helmet*) hay không mặc áo bảo hộ (*NO_safety vest*), … Cuối cùng, bộ dữ liệu *All* là tổng hợp của cả hai bộ dữ liệu *Use* và *NotUse* để kiểm tra sự làm việc đồng thời khi mô hình được huấn luyện với cả hai loại nhãn hành vi có sử dụng và không sử dụng phương tiện bảo vệ cá nhân. Hình 3 thể hiện hình ảnh các nhãn dữ liệu được gán trong các bộ dữ liệu nghiên cứu và Bảng 1 thể hiện các nhãn có trong các bộ dữ liệu và số lượng tương ứng của chúng. Trong khi đó, Bảng 2thể hiện các nhãn dữ liệu tạo thành các bộ dữ liệu con để phục vụ cho việc đào tạo mô hình.



Hình 2. Cấu trúc của mô hình YOLOv8 [18].

a) Bộ dữ liệu *Use* b) Bộ dữ liệu *NotUse* c) Bộ dữ liệu *All*

Hình 3. Hình ảnh trong các bộ dữ liệu.

Bảng 1. Các nhãn phân loại và số lượng.

	gloves	*helmet*	*mask*	*safety vest*	*shoes*
Số lượng	535	802	305	1092	974
	NO_gloves	*NO_helmet*	*NO_mask*	*NO_safety vest*	*NO_shoes*
Số lượng	1299	925	1105	721	114

Bảng 2. Các bộ dữ liệu được tạo và các nhãn trong bộ dữ liệu đó.

	Bộ dữ liệu *Use*	Bộ dữ liệu *NotUse*	Bộ dữ liệu *All*
gloves	×		×
helmet	×		×
mask	×		×
safety vest	×		×
shoes	×		×
NO_gloves		×	×
NO_helmet		×	×
NO_mask		×	×
NO_safety vest		×	×
NO_shoes		×	×

5.2. Huấn luyện mô hình và các tham số đánh giá

Quá trình huấn luyện các thử nghiệm và đánh giá mô hình được tiến hành trên hệ thống của Ultralytics và Google Colab, một nền tảng miễn phí cung cấp hệ thống tính toán đám mây cho các nghiên cứu ứng dụng học sâu. GPU NVIDIA Tesla T4 được cung cấp sẵn có trong Google Colab đặc biệt phù hợp trong những bài toán cần huấn luyện với dữ liệu hình ảnh lớn. GPU Tesla T4 được xây dựng trên kiến trúc Turing và được thiết kế đặc biệt để tăng tốc suy luận mô hình học sâu. CPU Intel(R) Core (TM) i7-6820 @ 2.70GHz đóng vai trò là bộ xử lý trung tâm (CPU). Ngôn ngữ lập trình được sử dụng là Python 3.11.2, hoạt động trong môi trường hệ thống Ubuntu 16.04. CUDA Toolkit 9.0 đang được sử dụng để tăng tốc tính toán.

Việc đánh giá tính hiệu quả của các bộ dữ liệu hành vi sử dụng phương tiện bảo vệ cá nhân với mô hình YOLOv8 được đề xuất sử dụng các số liệu đánh giá thường được sử dụng trong phân loại và phát hiện đối tượng, bao gồm *Precision, Recall và mAP*. Trong đó *Precision* thể hiện mức độ chính xác trong các phát hiện của mô hình, *Recall* thể hiện cho khả năng mô hình không bỏ sót dữ liệu cần được phát hiện còn *mAP0.5* là độ chính xác trung bình của mô hình khi ngưỡng phát hiện được thiết lập ở mốc 0,5. Công thức cho ba tham số này như sau:

$$Precision\ (Pr) = \frac{TP}{TP + FP} \quad (1)$$

$$Recall\ (Rc) = \frac{TP}{TP + FN} \quad (2)$$

$$Mean\ Avarage\ Precision\ (mAP0.5) = \frac{1}{N}\sum_{i=1}^{N} AP_i \quad (3)$$

5.3. Kết quả và phân tích

5.3.1. Kết quả so sánh với các mô hình tương tự

Để kiểm tra khả năng làm việc của mô hình YOLOv8, các mô hình học sâu khác đã được đưa vào huấn luyện và so sánh độ chính xác của chính với nhau. Bảng 3 thể hiện kết quả huấn luyện của các mô hình với bộ dữ liệu *All*. Có thể thấy bộ dữ liệu được huấn luyện với cấu trúc mô hình YOLOv8 đem lại hiệu quả cải thiện đáng kể khi so sánh với các mô hình có chức năng tương tự, kể cả với mô hình YOLOv5, phiên bản trước của YOLOv8. Mô hình cũng đạt hiệu suất tốt hơn về mặt thời gian, khi mất ít thời gian huấn luyện nhất so với các mô hình còn lại.

Bảng 3. Kết quả so sánh mô hình YOLOv8 với các mô hình trước đó.

Phương pháp	Chỉ số đánh giá				
	Precision	*Recall*	*mAP*	*Cải thiện (%)*	*Thời gian thực hiện (s)*
YOLOv8	0,725	0,719	0,705		1295
YOLOv5	0,683	0,634	0,661	8,74	1354
Faster-RCNN MobilenetV3	0,594	0,643	0,618	15,98	1825
Faster-RCNN Resnet50	0,653	0,578	0,605	17,32	2147

Cụ thể, với mô hình YOLOv8 đề xuất các giá trị *Precision*, *Recall* và *mAP* đều cải thiện đáng kể, tăng trung bình *8,74 %* so với mô hình cao thứ hai là YOLOv5. Việc các tham số đánh giá đều được cải thiện thể hiện khả năng phát hiện đầy đủ và dự đoán chính xác các hành vi sử dụng phương tiện bảo vệ cá nhân cần quan tâm của mô hình đề xuất hiệu quả hơn so với các mô hình có chức năng tương tự. Thời gian huấn luyện mô hình với cùng một bộ dữ liệu cũng giảm đi, cho thấy hiệu năng của mô hình đề xuất tốt hơn rất nhiều khi không cần phải tiêu tốn quá nhiều thời gian vào quá trình huấn luyện.

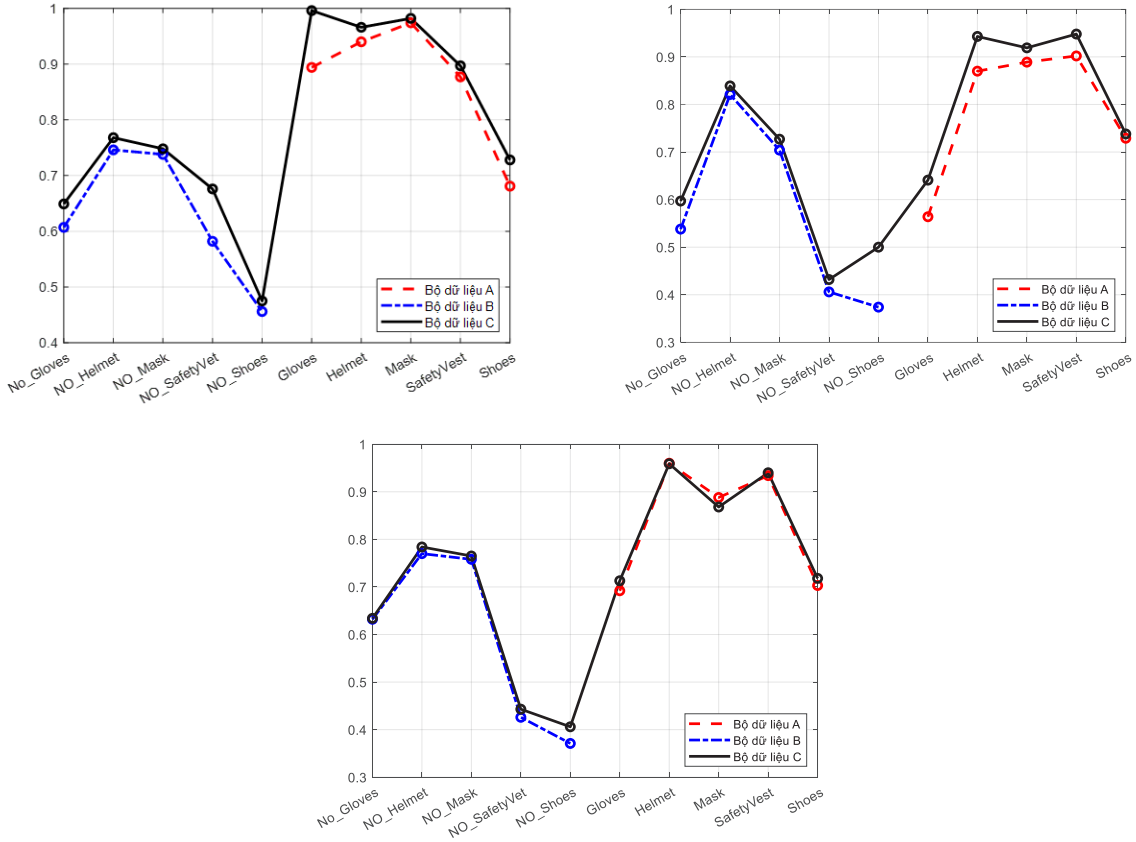
5.3.2. Kết quả huấn luyện các bộ dữ liệu nghiên cứu

Bảng 4 thể hiện các kết quả huấn luyện của mô hình với các bộ dữ liệu và Hình 4 là biểu đồ thể hiện các tham số của các bộ dữ liệu đó ứng với từng nhãn phân loại. Với các tham số trung bình cho toàn bộ các nhãn, độ chính xác của mô hình khi huấn luyện với bộ dữ liệu *All* chứa chung các nhãn hành vi sử dụng và không sử dụng trang bị bảo hộ thể hiện kết quả không quá khác biệt khi so với kết quả từ bộ dữ liệu *Use*. Điều này có thể giải thích do độ chính xác trong bộ dữ liệu này bị ảnh hưởng bởi các nhãn không mang phương tiện bảo vệ cá nhân với độ chính xác thấp hơn, khiến kết quả phát hiện trung bình bị giảm đi. Từ Hình 4 và Bảng 5 thể hiện các kết quả tham số đánh giá với từng loại nhãn, có thể thấy, các nhãn phát hiện hành vi công nhân sử dụng phương tiện bảo vệ cá nhân có độ chính xác cao hơn, do các phương tiện vệ cá nhân đã có các đặc điểm dễ nhận dạng và số lượng dữ liệu đủ để mô hình đủ khả năng phát hiện. Ngược lại, các nhãn phát hiện hành vi không mang phương tiện bảo vệ cá nhân thể hiện khả năng phát hiện kém hơn hẳn, đặc biệt trong trường hợp hành vi không

mang giày (*No_Shoes*) do không có đủ dữ liệu. Bộ dữ liệu khi huấn luyện với cả hai loại nhãn mang lại kết quả tốt hơn khi chúng được huấn luyện riêng lẻ. Với nhãn hành vi không mang giày bảo hộ được huấn luyện trong bộ dữ liệu *All*, chỉ số *recall* của nhãn này lại được cải thiện đáng kể (từ *0,374* tăng lên *0,5*). Điều đó cho thấy, việc huấn luyện với các nhãn hành vi sử dụng trái ngược nhau có ý nghĩa lớn trong việc cải thiện các kết quả phát hiện đối với các dữ liệu không đủ số lượng.

Bảng 4. Kết quả các tham số trung bình với các bộ dữ liệu đã được huấn luyện.

	Precision	*Recall*	*mAP*
Bộ dữ liệu *Use*	0,873	0,791	0,835
Bộ dữ liệu *NotUse*	0,626	0,579	0,591
Bộ dữ liệu *All*	0,725	0,719	0,705



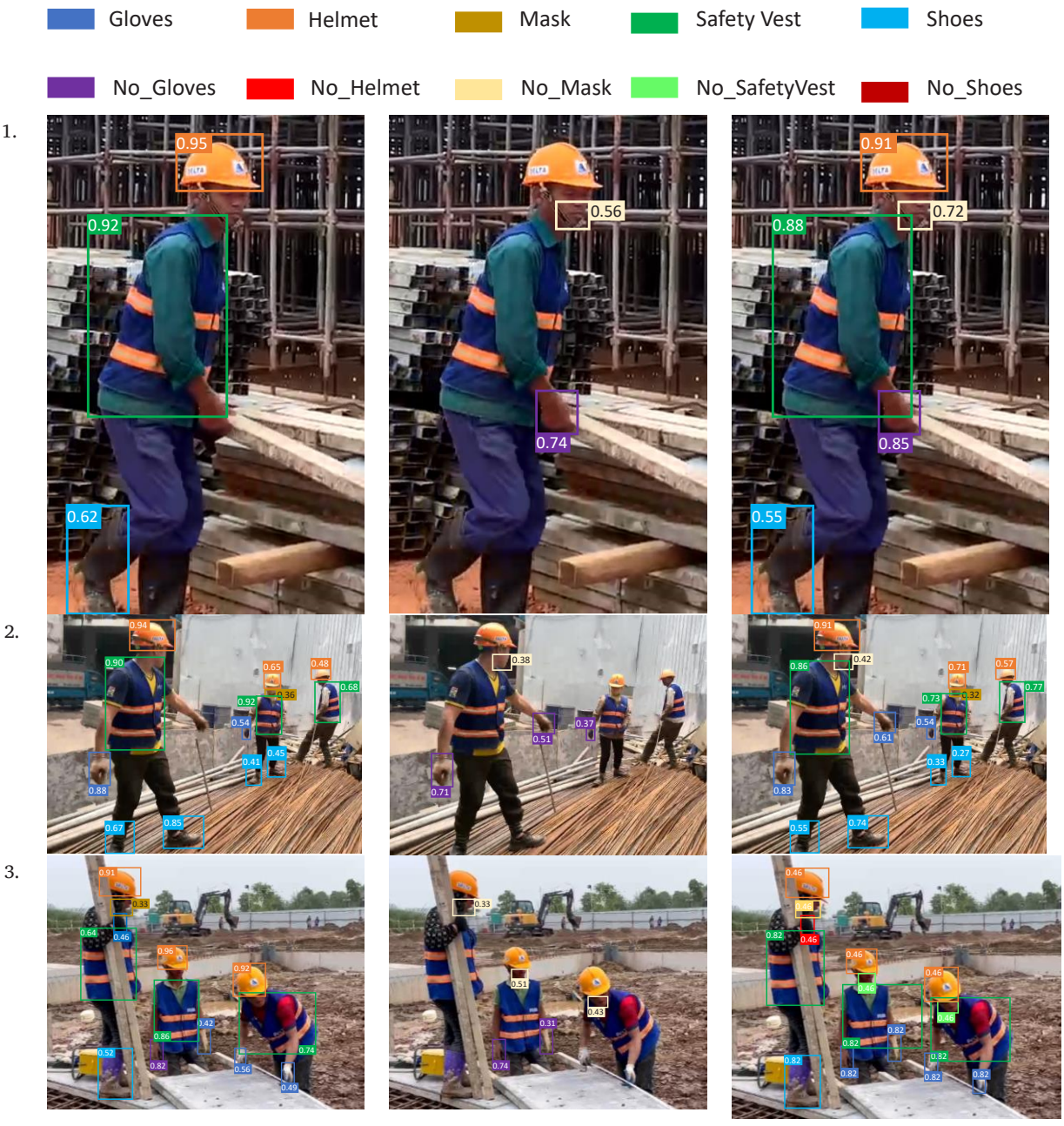
Hình 4. Biểu đồ các tham số kết quả khi huấn luyện mô hình với các bộ dữ liệu.

Bảng 5. Kết quả các tham số trung bình với các bộ dữ liệu đã được huấn luyện.

		Glove	*Helmet*	*Mask*	*Safety vest*	*Shoes*	*No_ Gloves*	*No_ Helmet*	*No_ Mask*	*No_ Safety vest*	*No_ Shoes*
Dataset *Use*	*Precision*	0,894	0,940	0,974	0,877	0,681					
	Recall	0,564	0,870	0,889	0,902	0,729					
	mAP	0,692	0,960	0,888	0,934	0,703					
Dataset *NotUse*	*Precision*						0,607	0,746	0,738	0,582	0,456
	Recall						0,538	0,821	0,704	0,406	0,374
	mAP						0,632	0,770	0,758	0,426	0,371
Dataset *All*	*Precision*	0,996	0,966	0,982	0,897	0,728	0,649	0,768	0,748	0,676	0,475
	Recall	0,641	0,943	0,919	0,948	0,738	0,597	0,839	0,727	0,432	0,500
	mAP	0,713	0,959	0,868	0,940	0,718	0,634	0,784	0,765	0,443	0,406

5.3.3. Kết quả phát hiện của mô hình đã phát hiện

Hình 5 thể hiện kết quả phát hiện của các mô hình, được so sánh. Có thể thấy việc huấn luyện với bộ dữ liệu chứa nhiều nhãn sẽ mang lại hiệu quả phát hiện hơn so với các bộ dữ liệu khác. Với việc huấn luyện khi tách riêng từng trường hợp với bộ dữ liệu *Use* và *NotUse*, một vài trường hợp mô hình có thể nhận nhầm do sự giống nhau của một vài nhãn, có thể kể đến như đeo găng tay (*gloves)* và không đeo găng tay (*NO_gloves)* (Hình 5.2). Kết quả thể hiện cũng cho thấy được những nhãn do mô hình đào tạo bởi bộ dữ liệu *All* cho được các kết quả đầy đủ và ý nghĩa hơn, độ chính xác trong các trường hợp phát hiện hành vi không sử dụng cũng được cải thiện đáng kể. Do đó, chứng minh độ tin cậy của phương pháp đề xuất khi áp dụng vào công tác thực tế trong việc giám sát và cảnh báo hành vi sử dụng phương tiện bảo vệ cá nhân của công nhân tại các công trường.



a) Bộ dữ liệu *Use* b) Bộ dữ liệu *NotUse* c) Bộ dữ liệu *All*

Hình 5. Kết quả phát hiện của mô hình YOLOv8 với dữ liệu thực tế.

6. Kết luận và kiến nghị

Trong nghiên cứu này, một phương pháp nhằm tự động phát hiện các hành vi sử dụng phương tiện bảo vệ cá nhân của công nhân ngoài công trường đã được nhóm nghiên cứu đề xuất. Các bộ dữ liệu phát hiện hành vi sử dụng phương tiện bảo vệ cá nhân của công nhân ngoài công trường đã được xây dựng. Trong đó, ngoài các hành vi có sử dụng phương tiện bảo vệ cá nhân, các nhãn công nhân không mang các trang thiết bị đó cũng được nhóm nghiên cứu quan tâm và bổ sung vào bộ dữ liệu, đây chính là một trong những cách tiếp cận mới của nghiên cứu này. Các dữ liệu bao gồm 960 hình ảnh trong đó chứa tổng cộng 7872 nhãn hành vi, bao gồm các hành vi sử dụng và không sử dụng các phương tiện bảo vệ cá nhân như găng tay, khẩu trang, mũ, áo bảo hộ và giày. Các kết quả sau khi huấn luyện mô hình cho thấy khi được huấn luyện với bộ dữ liệu chứa cả hai loại nhãn, độ chính xác của mô hình có được cải thiện, đặc biệt với các dữ liệu có số lượng ít. Việc huấn luyện với các nhãn hành vi đối lập như vậy cũng giảm bớt trường hợp mô hình nhầm lẫn với các nhãn có giống nhau về màu sắc hoặc hình dạng. Mô hình học sâu là kết quả nghiên cứu chính xác và hiệu quả trong việc sử dụng tài nguyên tính toán, do đó có khả năng áp dụng thời gian thực. Qua đó, phương pháp đề xuất trong nghiên cứu có thể tăng độ an toàn, kịp thời cảnh báo trên công trường nhằm làm giảm các nguy cơ gây rủi ro do các hành vi thiếu sót không sử dụng các phương tiện bảo vệ cá nhân trên công trường xây dựng.

Lời cảm ơn

Nghiên cứu này được tài trợ bởi Bộ Xây dựng trong đề tài mã số RD 31-24.

Tài liệu tham khảo

[1]. S. Bhole, "Safety problems and injuries on construction site: a review," *International Journal of Engineering and Techniques,* vol. 2, no. 4, pp. 24-35, 2016.

[2]. S. Ammad *et al.*, "Personal protective equipment in construction, accidents involved in construction infrastructure projects," *Solid State Technology,* vol. 63, no. 6, pp. 4147-4159, 2020.

[3]. V. S. K. Delhi, R. Sankarlal, and A. Thomas, "Detection of personal protective equipment (PPE) compliance on construction site using computer vision based deep learning techniques," *Frontiers in Built Environment,* vol. 6, p. 136, 2020.

[4]. N. D. Nath, A. H. Behzadan, and S. G. Paal, "Deep learning for site safety: Real-time detection of personal protective equipment," *Automation in Construction,* vol. 112, p. 103085, 2020/04/01/ 2020, doi: https://doi.org/10.1016/j.autcon.2020.103085.

[5]. S. Dong, Q. He, H. Li, and Q. Yin, "Automated PPE Misuse Identification and Assessment for Safety Performance Enhancement," in *ICCREM 2015*, 2015, pp. 204-214.

[6]. H. Zhang, X. Yan, H. Li, R. Jin, and H. Fu, "Real-Time Alarming, Monitoring, and Locating for Non-Hard-Hat Use in Construction," *Journal of Construction Engineering and Management,* vol. 145, no. 3, p. 04019006, 2019, doi: doi:10.1061/(ASCE)CO.1943-7862.0001629.

[7]. A. Kelm *et al.*, "Mobile passive Radio Frequency Identification (RFID) portal for automated and rapid control of Personal Protective Equipment (PPE) on construction sites," *Automation in Construction,* vol. 36, pp. 38-52, 2013/12/01/ 2013, doi: https://doi.org/10.1016/j.autcon.2013.08.009.

[8]. J. Seo, S. Han, S. Lee, and H. Kim, "Computer vision techniques for construction safety and health monitoring," *Advanced Engineering Informatics,* vol. 29, no. 2, pp. 239-251, 2015/04/01/ 2015, doi: https://doi.org/10.1016/j.aei.2015.02.001.

[9]. M.-W. Park, N. Elsafty, and Z. Zhu, "Hardhat-Wearing Detection for Enhancing On-Site Safety of Construction Workers," *Journal of Construction Engineering and Management,* vol. 141, no. 9, p. 04015024, 2015/09/01 2015, doi: 10.1061/(ASCE)CO.1943-7862.0000974.

[10]. D. Shan, M. Shehata, and W. Badawy, "Hard hat detection in video sequences based on face features, motion and color information," in *2011 3rd International Conference on Computer Research and Development*, 11-13 March 2011 2011, vol. 4, pp. 25-29, doi: 10.1109/ICCRD.2011.5763846.

[11]. K. Shrestha, P. P. Shrestha, D. Bajracharya, and E. A. Yfantis, "Hard-Hat Detection for Construction Safety Visualization," *Journal of Construction Engineering,* vol. 2015, p. 721380, 2015/02/01 2015, doi: 10.1155/2015/721380.

[12]. W. Fang, L. Ding, B. Zhong, P. E. D. Love, and H. Luo, "Automated detection of workers and heavy equipment on construction sites: A convolutional neural network approach," *Advanced Engineering Informatics,* vol. 37, pp. 139-149, 2018/08/01/ 2018, doi: https://doi.org/10.1016/j.aei.2018.05.003.

[13]. Z. Kolar, H. Chen, and X. Luo, "Transfer learning and deep convolutional neural networks for safety guardrail detection in 2D images," *Automation in Construction,* vol. 89, pp. 58-70, 2018/05/01/ 2018, doi: https://doi.org/10.1016/j.autcon.2018.01.003.

[14]. L. Ding, W. Fang, H. Luo, P. E. D. Love, B. Zhong, and X. Ouyang, "A deep hybrid learning model to detect unsafe behavior: Integrating convolution neural networks and long short-term memory," *Automation in Construction,* vol. 86, pp. 118-124, 2018/02/01/ 2018, doi: https://doi.org/10.1016/j.autcon.2017.11.002.

[15]. M. Siddula, F. Dai, Y. Ye, and J. Fan, "Unsupervised Feature Learning for Objects of Interest Detection in Cluttered Construction Roof Site Images," *Procedia Engineering,* vol. 145, pp. 428-435, 2016/01/01/ 2016, doi: https://doi.org/10.1016/j.proeng.2016.04.010.

[16]. N. D. Nath, T. Chaspari, and A. H. Behzadan, "Single- and multi-label classification of construction objects using deep transfer learning methods," *Journal of Information Technology in Construction,* vol. 24, no. Special issue Virtual, Augmented and Mixed: New Realities in Construction, pp. 511-526, 2019.

[17]. T. Diwan, G. Anirudh, and J. V. Tembhurne, "Object detection using YOLO: Challenges, architectural successors, datasets and applications," *multimedia Tools and Applications,* vol. 82, no. 6, pp. 9243-9275, 2023.

[18]. G. Wang, Y. Chen, P. An, H. Hong, J. Hu, and T. Huang, "UAV-YOLOv8: a small-object-detection model based on improved YOLOv8 for UAV aerial photography scenarios," *Sensors,* vol. 23, no. 16, p. 7190, 2023.

[19]. X. Li *et al.*, "Generalized focal loss: Learning qualified and distributed bounding boxes for dense object detection," *Advances in Neural Information Processing Systems,* vol. 33, pp. 21002-21012, 2020.

[20]. Z. Zheng, P. Wang, W. Liu, J. Li, R. Ye, and D. Ren, "Distance-IoU loss: Faster and better learning for bounding box regression," in *Proceedings of the AAAI conference on artificial intelligence*, 2020, vol. 34, no. 07, pp. 12993-13000.