

# Tự động hóa quy trình đếm cốt thép và lập báo cáo hiện trường dựa trên công nghệ thị giác máy tính và học sâu

Lê Thái Hòa<sup>1\*</sup>, Lê Đình Tiến<sup>1</sup>, Trần Quang Dũng<sup>1</sup>, Vũ Anh Tuấn<sup>1</sup>, Nguyễn Ngọc Toàn<sup>1</sup>

<sup>1</sup> Nhóm chuyên môn Công nghệ và Quản lý Xây dựng, Khoa Xây dựng Dân dụng và Công nghiệp, Trường Đại học Xây dựng Hà Nội

## TỪ KHOÁ

Kiểm soát chất lượng  
Học sâu  
Thị giác máy tính  
Đếm cốt thép  
Tự động  
Thời gian thực

## TÓM TẮT

Hiện nay, công tác kiểm soát chất lượng tại công trường xây dựng vẫn còn phức tạp và yêu cầu nhiều thao tác thủ công, đặc biệt đối với công việc kiểm đếm số lượng cốt thép, vốn tiêu tốn nhiều thời gian và dễ xảy ra sai sót. Trong quá trình xây dựng, mặc dù số lượng lớn hình ảnh vẫn được ghi lại hàng ngày, dữ liệu này chủ yếu được sử dụng cho mục đích lưu trữ, chưa được tận dụng hiệu quả trong việc kiểm soát chất lượng tự động. Sự phát triển mạnh mẽ gần đây của công nghệ học sâu, đặc biệt trong lĩnh vực thị giác máy tính, đã mở ra nhiều cơ hội ứng dụng tiềm năng nhằm tự động hóa quy trình kiểm đếm cốt thép trực tiếp tại công trường một cách hiệu quả. Bài báo này đề xuất một khung giải pháp dựa trên công nghệ học sâu để tự động hóa công tác kiểm đếm cốt thép, góp phần nâng cao hiệu quả và độ chính xác trong kiểm soát chất lượng xây dựng. Khung giải pháp gồm bốn bước chính: (1) Thu thập dữ liệu, bao gồm việc chụp ảnh cốt thép, và truyền tải dữ liệu về máy chủ để xử lý; (2) Nhận diện cốt thép, sử dụng các mô hình thị giác máy tính tiên tiến để nhận diện cốt thép; (3) Đếm cốt thép tự động; và (4) Lập báo cáo theo thời gian thực. Nhằm đạt được hiệu quả nhận diện tốt nhất, các mô hình phát hiện đối tượng phổ biến như Faster-RCNN, YOLOv8, YOLOv11, YOLOv12 và RF-DETR được huấn luyện và so sánh. Kết quả kiểm chứng cho thấy mô hình YOLOv8 đạt hiệu suất vượt trội với độ chính xác trung bình (mAP@50) lên đến 98%, xử lý tác vụ trong thời gian thực. Giải pháp đề xuất giúp giảm đáng kể khối lượng công việc thủ công, hạn chế sai sót trong quá trình kiểm đếm cốt thép, góp phần nâng cao hiệu quả tổng thể công tác kiểm soát chất lượng xây dựng.

## KEYWORDS

Quality control  
Deep learning  
Computer vision  
Rebar counting  
Automation

## ABSTRACT

Quality control on construction sites remains complex and predominantly relies on manual effort, particularly for tasks such as rebar counting, which are time-consuming and prone to errors. Although a large number of images are captured daily during the construction process, these visual data are mainly archived and not yet effectively utilized for automated quality control. Recent advances in deep learning, especially in computer vision, offer promising opportunities to automate the rebar counting process directly onsite. This paper proposes a comprehensive framework based on deep learning techniques to automate rebar counting, aiming to enhance both the efficiency and accuracy of construction quality control. The proposed framework consists of four main stages: (1) Data Collection, involving the capture and organization of rebar images; (2) Rebar Detection, applying advanced computer vision models to accurately identify rebars; (3) Automated Rebar Counting; and (4) Real-Time Report Generation. To ensure optimal detection performance, several popular object detection models, including Faster-RCNN, YOLOv8, YOLOv11, YOLOv12, and RF-DETR, were trained and compared. Evaluation results show that YOLOv8 outperformed the others, achieving a mean Average Precision (mAP@50) of up to 98% while operating in real-time. The proposed solution significantly reduces manual labor, minimizes counting errors, and improves the overall effectiveness of quality control procedures in construction projects.

## 1. Giới thiệu

Trong quá trình xây dựng, công tác cốt thép giữ vai trò quan trọng trong việc đảm bảo chất lượng cho kết cấu công trình. Việc kiểm soát chính xác số lượng và chất lượng của thép không chỉ ảnh hưởng trực tiếp đến độ bền công trình mà còn tác động lớn đến tiến độ và chi

phí thi công. Trước khi thép được nhập vào kho bãi hoặc chuyển đến khu vực gia công, khâu kiểm đếm số lượng tại công trường là bước bắt buộc nhằm đảm bảo việc bàn giao chính xác giữa nhà cung cấp và đơn vị thi công. Đây cũng là cơ sở để kiểm soát chi phí, tránh thất thoát vật liệu và duy trì sự minh bạch trong quá trình quản lý. Tại Việt Nam, công tác kiểm tra chất lượng cốt thép được thực hiện theo các tiêu chuẩn

\*Liên hệ tác giả: hoalt2@huce.edu.vn

Nhận ngày 04/09/2025, sửa xong ngày 17/10/2025, chấp nhận đăng ngày 20/10/2025

Link DOI: <https://doi.org/10.54772/jomc.05.2025.1153>

quốc gia và quy chuẩn kỹ thuật (ví dụ, TCVN 1651-1:2018 Thép cốt bê tông – Phần 1: Thép thanh tròn trơn, TCVN 1651-2:2018 Thép cốt bê tông – Phần 2: Thép thanh vằn [1,2]) khi nhập vật tư vào công trường. Tuy nhiên, quá trình đếm và kiểm tra thép trên công trường vẫn còn nhiều hạn chế. Các phương pháp truyền thống, vốn phụ thuộc chủ yếu vào quy trình thủ công của công nhân và cán bộ giám sát, thường tốn thời gian và dễ xảy ra sai sót, đặc biệt khi số lượng thanh thép nhiều, bó thép dày, hoặc thép bị chồng chéo. Trong điều kiện công trường thường xuyên chịu tác động bởi bụi bẩn, ánh sáng thay đổi, và nhịp độ làm việc gấp rút, việc kiểm đếm thủ công càng khó đảm bảo tính chính xác và hiệu quả. Điều này tạo ra điểm nghẽn trong khâu tiếp nhận vật liệu, làm chậm tiến độ và gia tăng chi phí quản lý.

Trước những hạn chế này, nhiều nghiên cứu đã hướng đến việc ứng dụng công nghệ thị giác máy tính và xử lý ảnh để tự động hóa quá trình kiểm đếm cốt thép tại công trường. Các phương pháp phổ biến như phát hiện cạnh (edge detection), biến đổi Hough (Hough transformation), hoặc dùng ảnh tham chiếu (template matching) đã được sử dụng để nhận diện và tách các thanh thép trong ảnh chụp, từ đó hỗ trợ công tác đếm nhanh hơn [3-5]. Tuy nhiên, các kỹ thuật này vẫn dễ gặp sai sót khi xử lý các tình huống thực tế như thép bị gỉ, có vết bẩn, bị chồng lấn hoặc khi ánh sáng công trường thay đổi liên tục. Trong bối cảnh đó, các giải pháp dựa trên trí tuệ nhân tạo và học sâu (deep learning) đã nổi lên như một hướng đi đầy tiềm năng. Các mô hình sử dụng mạng nơ-ron tích chập (Convolutional Neural Network – CNN), Faster R-CNN, hay YOLO với các phiên bản cải tiến cho phép phát hiện và kiểm đếm thép trong thời gian thực với độ chính xác cao [6-8], kể cả trong điều kiện phức tạp. Mô hình CNN đã được áp dụng cho hệ thống đếm thép khi thu hồi sản phẩm trong nhà máy tại Việt Nam [9]. Song, hệ thống mới chỉ được nghiên cứu cho môi trường trong nhà máy, yêu cầu sử dụng camera tốc độ cao. Một số nghiên cứu khác còn kết hợp mô hình học sâu với kỹ thuật xử lý ảnh cổ điển nhằm nâng cao khả năng phân biệt thanh thép khi bị chồng chéo hoặc bó thép có hình dạng không đồng đều [5,10]. Hơn nữa, việc tích hợp các hệ thống camera đặt tại công trường hoặc sử dụng thiết bị di động cầm tay để ghi nhận hình ảnh khi xe chở thép đến bãi đã mở ra giải pháp khả thi cho bài toán kiểm đếm. Nhờ đó, toàn bộ quá trình giao nhận vật liệu có thể được số hóa, giảm thiểu sự phụ thuộc vào tay nghề nhân công, nâng cao tính chính xác và tốc độ xử lý, đồng thời tăng cường tính minh bạch trong công tác quản lý vật liệu xây dựng.

Tóm lại, việc phát triển các phương pháp kiểm đếm cốt thép tự động dựa trên thị giác máy tính và học sâu ngay tại công trường, trong bối cảnh giao nhận vật liệu, là một yêu cầu cấp thiết. Đây không chỉ là bước tiến quan trọng trong quản lý vật liệu đầu vào mà còn góp phần nâng cao hiệu quả và tính chuyên nghiệp của ngành xây dựng hiện đại.

## 2. Tổng quan về công nghệ hỗ trợ phát hiện và kiểm đếm cốt thép

Trong những năm gần đây, nhu cầu tự động hóa việc phát hiện và kiểm đếm cốt thép ngày càng trở nên cấp thiết nhằm thay thế cho các phương pháp thủ công vốn tốn kém nhân lực và dễ sai sót. Nhiều

công trình nghiên cứu đã đưa ra các giải pháp khác nhau, có thể chia thành hai hướng chính: (1) sử dụng phương pháp xử lý ảnh cổ điển dựa trên các đặc trưng được tạo thủ công (hand-crafted features) và (2) sử dụng phương pháp học sâu tận dụng ưu thế của mạng nơ-ron tích chập. Sự phát triển của hai hướng tiếp cận này phản ánh quá trình chuyển dịch từ các kỹ thuật truyền thống sang các giải pháp hiện đại hơn, phù hợp với điều kiện công trường thực tế.

### 2.1. Nhận diện cốt thép dựa trên phương pháp xử lý ảnh cổ điển

Trước đây, các nghiên cứu về nhận diện và xác định số lượng vật thể, bao gồm cả kiểm đếm cốt thép, chủ yếu dựa trên các phương pháp xử lý ảnh cổ điển, trích xuất các đặc trưng được tạo thủ công (hand-crafted features). Những phương pháp này thường bắt đầu từ bước tiền xử lý ảnh như khử nhiễu, cân bằng sáng, rồi áp dụng các kỹ thuật nhị phân hóa để tách nền – vật thể, trong đó các thuật toán phân chia vùng ảnh (segmentation threshold) được sử dụng rộng rãi nhằm phân biệt thanh thép với môi trường xung quanh [11]. Về sau, các thuật toán phát hiện biên như Sobel hoặc Canny được sử dụng để trích xuất đường bao [12,13], giúp làm nổi bật tiết diện thép và phục vụ bước đếm số lượng. Nhiều hệ thống sau đó tận dụng phép biến đổi Hough (Hough transformation), đặc biệt là Circle Hough Transform (CHT), để xác định tâm và bán kính của tiết diện, từ đó tính ra số lượng cốt thanh thép [14]. Cách tiếp cận này đã cho thấy hiệu quả cao trong môi trường nhà máy hoặc trên dây chuyền sản xuất, nơi điều kiện chiếu sáng và góc ảnh chụp tương đối ổn định. Kết quả thu được có tỷ lệ sai số thấp và đảm bảo khả năng nhận dạng nhanh chóng [12,15].

Ngoài biến đổi Hough, phương pháp sử dụng ảnh tham chiếu (template matching) cũng được dùng để tìm kiếm và nhận diện tiết diện thép trong ảnh. Phương pháp này so sánh một mẫu chuẩn đã biết với từng vùng ảnh để tìm ra vị trí trùng khớp, cho kết quả chính xác trong trường hợp các thanh thép ít bị biến dạng hay bị che khuất [5]. Bên cạnh đó, các kỹ thuật phân tích hình thái học và phân tích đường bao cũng được khai thác nhằm xử lý tình huống các thanh thép dính chùm. Thông qua việc xác định các điểm lõm (concave points) trên đường bao, hệ thống có thể tách bó thép thành từng đơn vị riêng biệt để đếm, đồng thời bổ sung cơ chế dung sai lỗi nhiều cấp (multi-level fault tolerance) để hạn chế sai số trong phân tách [16].

Nhìn chung, các phương pháp truyền thống đã đạt độ chính xác rất cao trong môi trường kiểm soát như kho lưu trữ hoặc dây chuyền sản xuất, nơi nền ảnh đơn giản và ít biến đổi. Tuy nhiên, khi áp dụng ngoài công trường, hiệu quả lại suy giảm rõ rệt do nhiều yếu tố phức tạp như ánh sáng không đồng nhất, thanh thép bị rỉ sét, bụi bẩn, khung cảnh công trường lộn xộn, cũng như sự đa dạng về kích thước và chồng lấn của cốt thép. Các nghiên cứu về đếm thép từ hình ảnh UAV tại công trường cũng chỉ ra rằng các đặc trưng thủ công chưa đủ mạnh để khái quát trong các kịch bản thực tế có nhiều biến thiên [5]. Do đó, mặc dù các phương pháp truyền thống đóng vai trò quan trọng trong việc đặt nền móng cho lĩnh vực này, những hạn chế về tính bền vững và khả năng khái quát trong điều kiện phức tạp đã trở thành động lực thúc đẩy sự

chuyển dịch sang các phương pháp học máy và học sâu hiện đại, nhằm đáp ứng yêu cầu khắt khe về độ chính xác và tính ứng dụng thực tiễn.

## 2.2. Nhận diện cốt thép thông qua phương pháp học sâu

Trong những năm gần đây, học sâu, đặc biệt là phương pháp mạng nơ-ron tích chập CNN và mạng Transformer (kiến trúc mạng học sâu dựa trên cơ chế Attention), đã trở thành công cụ chủ đạo trong việc giải quyết các bài toán thị giác máy tính như phát hiện và phân loại đối tượng. Trong lĩnh vực xây dựng, nhiều nghiên cứu đã ứng dụng phương pháp học sâu để nhận diện và đếm cốt thép từ ảnh chụp hiện trường, giúp khắc phục các nhược điểm của phương pháp xử lý ảnh truyền thống vốn.

Faster R-CNN là một trong những phương pháp được đề xuất trong nghiên cứu của Wang, Kim [5], để phát hiện cốt thép nhờ khả năng nhận diện đối tượng chính xác trong các điều kiện đa dạng. Phương pháp này cho phép trích xuất đặc trưng và xác định các vùng chứa đối tượng thông qua vùng đề xuất (region proposal network), sau đó phân loại và xác định vị trí đối tượng cần phát hiện. Theo Fang, Li [17], mô hình Faster R-CNN có thể đạt độ chính xác 95,7 % khi áp dụng tại công trường thực tế với các chi tiết phức tạp như đồ bảo hộ lao động.

Chuỗi mô hình YOLO được phát triển sau Faster R-CNN, là mô hình phát hiện đối tượng trong một giai đoạn (one-stage detector), có quy trình xử lý đơn giản hơn và phù hợp cho các ứng dụng yêu cầu tốc độ xử lý nhanh trong thời gian thực. Phiên bản YOLOv5 và YOLOv6 đã được ứng dụng để nhận diện cốt thép khi có nhiều đối tượng nhỏ, chồng chéo. Đây là một đặc điểm phổ biến ở các bộ thép trên công trường. Mô hình YOLO được đánh giá cao về tốc độ và khả năng triển khai trên các thiết bị có tài nguyên giới hạn, trong khi vẫn duy trì độ chính xác rất cao trên 90 % [5,8].

Thời gian gần đây, một số mô hình cải tiến như DSSD-YOLO hoặc RF-DETR cũng được đề xuất để nâng cao độ chính xác trong điều kiện đối tượng cần nhận diện nhỏ và xuất hiện dày đặc. DSSD-YOLO cải thiện độ chính xác khi phát hiện các đối tượng nhỏ và xếp chồng bằng cách sử dụng cơ chế Attention theo tọa độ (Coordinate Attention - CA) và mạng kim tự tháp đặc trưng nhẹ (Lightweight Feature Pyramid Network - LFPN), cho khả năng nhận diện tốt hơn về mặt hình học và vị trí không gian của cốt thép [18]. RF-DETR, được cải tiến từ mô hình DETR (Detection Transformer), kết hợp giữa cấu trúc CNN và cơ chế Transformer để cải thiện độ ổn định trong phân loại và định vị chính

xác vật thể nhỏ [19,20].

Ngoài ra, một số nghiên cứu còn tích hợp các kỹ thuật học máy như K-Nearest Neighbor (KNN) và Support Vector Machine (SVM) với CNN để phân loại các vùng ảnh và trả về kết quả có hoặc không có cốt thép. Phương pháp này rất khả thi cho điều kiện môi trường ánh sáng ổn định, mô hình được huấn luyện có thể đạt độ chính xác trung bình trên 95 % [21,22].

Mặc dù nhiều mô hình học sâu được chứng minh có hiệu quả khi nhận diện cốt thép, hầu hết các nghiên cứu hiện tại tập trung vào điều kiện trong phòng thí nghiệm hoặc nhà máy sản xuất, là môi trường có tính kiểm soát cao. Trong khi đó, công trường xây dựng tại Việt Nam thường có điều kiện ánh sáng phức tạp, mặt bằng bố trí vật tư vẫn khiến cốt thép bị che khuất và chồng chéo, và dữ liệu ảnh không đồng nhất giữa các thiết bị chụp. Vì vậy, việc phát triển một quy trình áp dụng thuật toán học sâu có khả năng hoạt động hiệu quả trong điều kiện thực tế của Việt Nam là điều cần thiết. Bài báo này không chỉ đề xuất quy trình hoàn thiện – từ thu thập dữ liệu, huấn luyện mô hình nhận diện thép, kiểm đếm, và tạo báo cáo kết quả – mà còn so sánh các mô hình phát hiện đối tượng hiện đại như Faster-RCNN, YOLOv8, YOLOv11, YOLOv12 và RF-DETR. Đây là điểm nổi bật giúp chuyển hóa mô hình từ lý thuyết sang ứng dụng thực tiễn, tạo tiền đề cho việc số hóa các quy trình quản lý chất lượng trong ngành xây dựng.

## 3. Thiết kế hệ thống tự động kiểm đếm cốt thép trên hiện trường

Trong bối cảnh yêu cầu kiểm soát chất lượng ngày càng khắt khe và khối lượng dữ liệu hình ảnh tại công trường ngày một tăng, nhu cầu cấp thiết đặt ra là phải có một hệ thống tự động, chính xác và vận hành thời gian thực. Nghiên cứu này đề xuất một khung giải pháp tích hợp công nghệ học sâu với các thành phần từ thu thập, xử lý đến báo cáo kết quả, nhằm xây dựng một hệ thống hoàn chỉnh cho việc kiểm đếm cốt thép tại hiện trường.

### 3.1. Khung giải pháp

Hệ thống được thiết kế thành bốn bước chính: (1) Thu thập dữ liệu, (2) Nhận diện cốt thép, (3) Đếm cốt thép tự động và (4) Lập báo cáo theo thời gian thực, với mỗi bước được tối ưu hóa để phù hợp với điều kiện thi công tại Việt Nam và hướng đến khả năng triển khai thực tế (Hình 1).



Hình 1. Khung giải pháp đề xuất cho công tác kiểm đếm cốt thép tự động.

### 3.1.1. Thu thập dữ liệu

Giai đoạn đầu tiên trong khung giải pháp là thu thập dữ liệu hình ảnh cốt thép. Dữ liệu ảnh đầu vào có thể được chụp từ các thiết bị phổ biến như điện thoại thông minh, máy ảnh, hoặc từ camera giám sát cố định. Trong nghiên cứu này, nhằm đạt độ chính xác tốt nhất cho bước huấn luyện mô hình nhận diện cốt thép, điện thoại thông minh được đề xuất sử dụng. Trên thực tế, điện thoại di động có chức năng chụp ảnh là thiết bị phù hợp vì cán bộ kỹ thuật có thể chủ động điều chỉnh chế độ chụp trong nhiều điều kiện thực tế công trường khác nhau như những thay đổi về ánh sáng, góc chụp, hay gặp vật cản. Sau khi chụp, ảnh sẽ được truyền về máy tính xử lý ngay lập tức thông qua mạng nội bộ để thực hiện xử lý phân tích và nhận diện cốt thép.

Trong quá trình chụp ảnh, cán bộ kỹ thuật cần đảm bảo phương chụp vuông góc mặt cắt ngang của bó thép (Hình 2). Góc chụp này mô phỏng đúng theo tư thế đếm thép thông thường trong quá trình kiểm tra vật tư, đồng thời tránh việc cây thép che khuất lẫn nhau.



Hình 2. Ảnh cốt thép chụp đúng yêu cầu.

### 3.1.2. Nhận diện cốt thép

Ở bước thứ hai, mô hình học sâu được sử dụng để phát hiện cốt thép trong ảnh đầu vào. Các mô hình thị giác máy tính hiện đại được huấn luyện và so sánh nhằm tìm ra mô hình hoạt động tốt nhất, bao gồm Faster R-CNN, YOLOv8, YOLOv11, YOLOv12 và RF-DETR. Trong đó, Faster R-CNN đại diện cho nhóm phát hiện hai giai đoạn (two-stage detector), cho kết quả chính xác cao nhưng tốc độ chậm hơn. YOLOv8, YOLOv11, và YOLOv12 là các mô hình phát hiện một giai đoạn (one-stage detector), tối ưu cho tác vụ thời gian thực và có thể được huấn luyện để nhận diện cốt thép một cách nhanh chóng. RF-DETR là phiên bản cải tiến của mô hình DETR (Detection Transformer), kết hợp giữa mạng CNN và cơ chế Attention trong Transformer, giúp tăng cường độ ổn định khi phát hiện các vật thể nhỏ như tiết diện cốt thép. Nhược điểm của mô hình này là thời gian huấn luyện lâu, độ phức tạp cao, và

đòi hỏi tài nguyên lớn (về cả số lượng dữ liệu và cấu hình máy tính huấn luyện mô hình).

Nhằm có sự so sánh tương đương nhất, các mô hình được huấn luyện với cùng bộ dữ liệu cốt thép. Bộ dữ liệu dùng để huấn luyện gồm 788 ảnh, trong đó có 638 ảnh dùng cho tập huấn luyện (train set), 75 ảnh cho mỗi tập kiểm tra (test set) và tập xác thực (validation set), tương ứng tỷ lệ 80 % – 10 % – 10 %. Toàn bộ dữ liệu được gán nhãn thủ công với một lớp duy nhất là “rebar”, tạo tính nhất quán trong nhận dạng. Mặc dù không lớn, bộ dữ liệu này vẫn thích hợp cho nghiên cứu do ảnh tập trung vào thanh thép (chụp tiết diện ngang). Ngoài ra, việc sử dụng mô hình có nền tảng (backbone) đã được tiền huấn luyện (pre-trained) và các kỹ thuật tăng cường dữ liệu sẽ nâng cao hiệu quả của mô hình. Kỹ thuật tăng cường dữ liệu được áp dụng bao gồm các phép biến đổi xoay, lật, thay đổi độ sáng – độ tương phản, biến đổi góc nhìn ( $\pm 15^\circ$ ). Nhờ đó, mô hình có thể học được các đặc trưng ổn định hơn và tăng khả năng khái quát hóa. Kết quả so sánh giữa các mô hình được thể hiện ở Bảng 1 dưới đây.

Chỉ số mAP@50 (mean Average Precision ở ngưỡng 0,5) được sử dụng để đánh giá hiệu năng mô hình nhận diện đối tượng. Trong đó, một dự đoán được xem là đúng nếu hình bao dự đoán và hình bao nhãn chồng lấn nhau với IoU (Intersection over Union)  $\geq 0,5$ . Giá trị mAP@50 càng cao cho thấy mô hình phát hiện đối tượng càng chính xác và ổn định.

### 3.1.3. Đếm cốt thép tự động

Sau khi có mô hình nhận diện cốt thép, bước tiếp theo là đưa mô hình này vào hệ thống đếm số lượng cốt thép trong ảnh. Quá trình này được thực hiện bằng cách đếm số lượng hình bao được xác định từ mô hình YOLOv8 có độ tin cậy (confidence score) lớn hơn 0,25. Việc lựa chọn ngưỡng tin cậy bằng 0,25 cho bài toán kiểm đếm cốt thép được cân nhắc giữa độ chính xác (precision) và khả năng phát hiện đầy đủ (recall). Đối với thực tế công trường, yêu cầu nhận dạng sẽ thiên về việc tăng khả năng phát hiện (recall) nhằm ít bỏ sót các thanh thép kể cả khi hình ảnh bị mờ do các thanh thép chồng sát nhau, thiếu sáng hoặc bụi bẩn. Ngoài ra, ngưỡng này không quá thấp để tạo ra nhiều nhận dạng giả. Giá trị 0,25 cũng được khuyến nghị bởi tác giả của mô hình YOLO khi tiến hành phát hiện các vật thể nhỏ.

### 3.1.4. Lập báo cáo theo thời gian thực

Trong giai đoạn cuối cùng này, hệ thống sẽ tự động tạo báo cáo kết quả trong thời gian thực. Giao diện người dùng (Graphical User Interface - GUI) được xây dựng bằng thư viện PyQt5, cho phép cán bộ kỹ thuật chọn ảnh được tải lên từ thiết bị di động. Hình 3 thể hiện giao diện của hệ thống. Tại giao diện này, cán bộ cần nhập loại thanh thép và đường kính của bó thép đang cần đếm vào hai ô “Nhập tên vật liệu” và “Nhập đường kính”. Kết quả hiển thị sẽ gồm ảnh đã gán nhãn và số lượng thanh cốt thép đã được nhận diện.

#### 4. Kết quả thực nghiệm và thảo luận

Hệ thống kiểm đếm cốt thép tự động được triển khai dựa trên khung giải pháp gồm bốn bước chính đã giới thiệu ở phần trước. Mô hình YOLOv8 đã được huấn luyện để phát hiện tiết diện tròn của các thanh thép, với ngưỡng độ tin cậy được đặt ở 0,25. Trong ví dụ minh họa tại Hình 4, sau khi nhận diện, hệ thống đưa ra kết quả kiểm đếm có 60 thanh thép D18. Số lượng này được tự động trích xuất sang biểu mẫu “Báo cáo nhập xuất vật liệu xây dựng hàng ngày” cùng thông tin tương ứng trong cột “Tên vật liệu” (Hình 5). Tổng thời gian cho cả quá trình xử lý ảnh, đếm thép và hiển thị thông tin trong báo cáo là nhỏ hơn 5 giây. Cán bộ kỹ thuật có thể tiếp tục chọn ảnh trong thư mục chứa ảnh được truyền từ thiết bị di động chụp ảnh trên hiện trường để kiểm đếm.

Mặc dù hệ thống đếm cốt thép này đã cho thấy hiệu quả, vẫn tồn tại một số tình huống đặc thù trên công trường gây ảnh hưởng đến độ chính xác của kết quả. Trường hợp đầu tiên (Hình 6), các thanh thép đã qua gia công (cắt, mài) khiến mặt cắt không còn tiết diện tròn đều như lúc xuất xưởng. Mô hình học sâu vốn được huấn luyện trên các

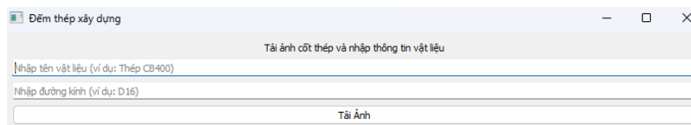
mặt cắt nguyên vẹn và có tính đối xứng cao, lúc này gặp khó khăn khi nhận diện tiết diện đã bị biến dạng. Điều này dẫn đến việc bỏ sót các đối tượng. Đây là trường hợp điển hình cho việc cần bổ sung thêm lượng mẫu ảnh thực tế của cốt thép sau khi gia công vào bộ dữ liệu huấn luyện.

Ở trường hợp tiếp theo, mô hình đã nhận diện các vị trí vị trí gân thép hoặc đường viền cạnh là các mặt cắt ngang (Hình 7 – vùng khoanh đỏ). Điều này xảy ra do các chi tiết gân thép có hình tròn hoặc bo góc nhẹ, khiến mô hình bị nhầm lẫn. Trường hợp này đặc biệt quan trọng vì số lượng đếm có thể bị tăng lên nhiều lần. Giải pháp tiềm năng là tăng cường dữ liệu âm (negative samples) trong tập huấn luyện để mô hình học được cách phân biệt rõ giữa gân và mặt cắt.

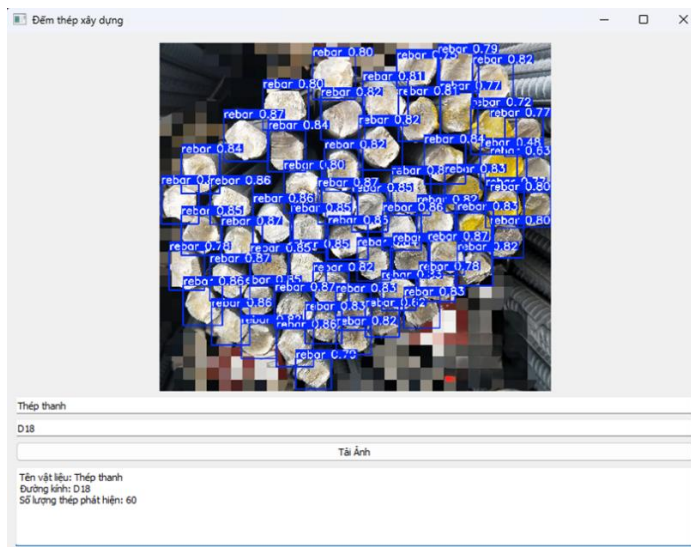
Cuối cùng, cán bộ kỹ thuật cần chú ý tới tình huống sai sót vì lý do ảnh chụp không bao phủ toàn bộ bó thép, dẫn đến việc thiếu đối tượng cần đếm (Hình 8 – phần khoanh đỏ). Đây là sai lệch đến từ quy trình chụp ảnh, không phải từ mô hình nhận diện đã được huấn luyện. Vấn đề này nhấn mạnh sự cần thiết của việc xây dựng quy trình chuẩn khi thu thập dữ liệu đầu vào, chẳng hạn như đảm bảo góc chụp chính diện, đủ ánh sáng, và bao trùm toàn bộ bề mặt mặt cắt bó thép.

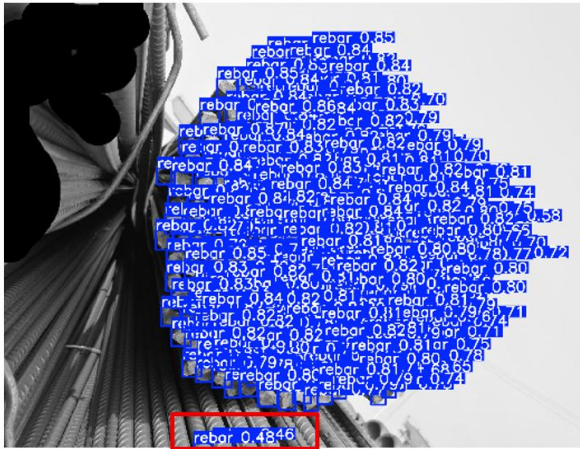
**Bảng 1.** So sánh độ chính xác trung bình giữa các mô hình nhận diện cốt thép.

| Mô hình                          | Faster R-CNN | YOLOv8 | YOLOv11 | YOLOv12 | RF-DETR |
|----------------------------------|--------------|--------|---------|---------|---------|
| Độ chính xác trung bình (mAP@50) | 71,4 %       | 98,0 % | 97,8 %  | 94,8 %  | 62,4 %  |

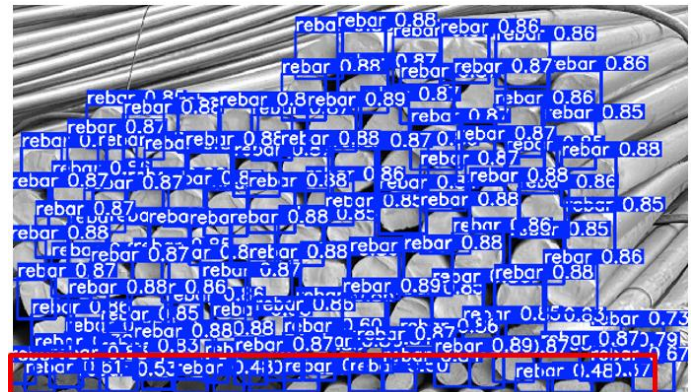


**Hình 3.** Giao diện của hệ thống đếm cốt thép tự động.





Hình 7. Mô hình nhận diện nhằm gán thép thành mặt cắt thép.



Hình 8. Thép đếm bị thiếu do ảnh chụp không đúng yêu cầu của hệ thống.

## 5. Kết luận

Bài báo đã đề xuất một khung giải pháp ứng dụng công nghệ học sâu để tự động hóa quá trình kiểm đếm cốt thép trong xây dựng, giúp nâng cao hiệu quả và độ chính xác của công tác kiểm soát chất lượng vật tư. Khung giải pháp gồm bốn bước chính: thu thập dữ liệu, nhận diện cốt thép, đếm tự động và lập báo cáo theo thời gian thực.

Các mô hình học sâu hiện đại như Faster R-CNN, YOLOv8, YOLOv11, YOLOv12 và RF-DETR đã được huấn luyện và đánh giá trên bộ dữ liệu ảnh thực tế chụp tại hiện trường. Kết quả thực nghiệm cho thấy mô hình YOLOv8 đạt độ chính xác mAP@50 cao nhất (98,0 %) và được lựa chọn để tích hợp vào hệ thống kiểm đếm tự động. Hệ thống giao diện thân thiện cho phép người dùng nhanh chóng phân tích dữ liệu từ ảnh chụp thực địa, xử lý và trả về báo cáo trong thời gian ngắn, đáp ứng yêu cầu thực tiễn tại công trường.

Tuy nhiên, kết quả cũng chỉ ra một số hạn chế trong việc nhận diện chính xác trong các tình huống ảnh không đảm bảo yêu cầu ban đầu (thép đã gia công, sai góc chụp, v.v.). Do đó, để nâng cao độ tin cậy và khả năng tổng quát của hệ thống, cần mở rộng tập dữ liệu huấn luyện bao gồm cả các trường hợp tiêu cực (negative samples) và xây dựng quy trình chụp ảnh tiêu chuẩn. Ngoài ra, cần bổ sung thêm vào bộ dữ liệu các trường hợp thanh thép có tiết diện đặc biệt như thép hộp, thép ống, thép hình, v.v... Hướng phát triển trong tương lai sẽ tập trung vào việc tích hợp hệ thống lên thiết bị di động, cải tiến khả năng nhận diện theo thời gian thực và mở rộng sang các loại vật liệu xây dựng khác, góp phần thúc đẩy chuyển đổi số trong ngành xây dựng tại Việt Nam.

## Tài liệu tham khảo

- [1]. TCVN (2018), TCVN 1651-1:2018 *Thép cốt bê tông – Phần 1: Thép thanh tròn trơn*, Bộ Khoa học và Công nghệ, Việt Nam.
- [2]. TCVN (2018), TCVN 1651-2:2018 *Thép cốt bê tông – Phần 2: Thép thanh vằn*, Bộ Khoa học và Công nghệ, Việt Nam.

- [3]. Gibb S., La H. M. (2016), "Automated rebar detection for ground-penetrating radar", *International Symposium on Visual Computing*, Springer, 815-824.
- [4]. Ghazali M. F., Wong L.-K., See J. (2016), "Automatic detection and counting of circular and rectangular steel bars", *9th International Conference on Robotic, Vision, Signal Processing and Power Applications: Empowering Research and Innovation*, Springer, 199-207.
- [5]. Wang S., Kim M., Hae H., Cao M., Kim J. (2023), "The development of a rebar-counting model for reinforced concrete columns: Using an unmanned aerial vehicle and deep-learning approach", *Journal of Construction Engineering and Management*, 149 (11): 04023111.
- [6]. Chen M., Shi X., Zhang Y., Wu D., Guizani M. (2017), "Deep feature learning for medical image analysis with convolutional autoencoder neural network", *IEEE transactions on big data*, 7 (4): 750-758.
- [7]. He K., Zhang X., Ren S., Sun J. (2016), "Deep residual learning for image recognition", *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770-778.
- [8]. Li Y., Lu Y., Chen J. (2021), "A deep learning approach for real-time rebar counting on the construction site based on YOLOv3 detector", *Automation in Construction*, 124: 103602.
- [9]. Long P. Đ. (2014), "Đếm thép cây tự động bằng xử lý ảnh", *Tạp chí KHOA HỌC & CÔNG NGHỆ*, 118 (4): 119-124.
- [10]. Liu C., Zhu L., Zhang X. (2019), "Bundled round bars counting based on iteratively trained SVM", *International Conference on Intelligent Computing*, Springer, 156-165.
- [11]. Su Z., Fang K., Peng Z., Feng Z. (2010), "Rebar automatically counting on the product line", *2010 IEEE International Conference on Progress in Informatics and Computing*, IEEE, 756-760.
- [12]. Ying X., Wei X., Pei-xin Y., Qing-da H., Chang-hai C. (2010), "Research on an automatic counting method for steel bars' image", *2010 International Conference on Electrical and Control Engineering*, IEEE, 1644-1647.
- [13]. Xu X., Xia T., Venkatachalam A., Huston D. (2013), "Development of high-speed ultrawideband ground-penetrating radar for rebar detection", *Journal of Engineering Mechanics*, 139 (3): 272-285.
- [14]. Wang H., Polden J., Jirgens J., Yu Z., Pan Z. (2019), "Automatic rebar counting using image processing and machine learning", *2019 IEEE 9th Annual International Conference on CYBER Technology in Automation, Control, and Intelligent Systems (CYBER)*, IEEE, 900-904.
- [15]. Zhang D., Xie Z., Wang C. (2008), "Bar section image enhancement and positioning method in on-line steel bar counting and automatic separating system", *2008 Congress on Image and Signal Processing*, IEEE, 319-323.

- [16]. Wu Y., Zhou X., Zhang Y. (2015), "Steel bars counting and splitting method based on machine vision", *2015 IEEE International Conference on Cyber Technology in Automation, Control, and Intelligent Systems (CYBER)*, IEEE, 420-425.
- [17]. Fang Q., et al. (2018), "Detecting non-hardhat-use by a deep learning method from far-field surveillance videos", *Automation in construction*, 85: 1-9.
- [18]. Zhang Z., et al. (2024), "DSSO-YOLO: A fast detection model for densely stacked small object", *Displays*, 82: 102659.
- [19]. Zhao Y., et al. (2024), "Detrs beat yolos on real-time object detection", *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 16965-16974.
- [20]. Wang G., Sun D., Li H., Cheng J., Yan P., Li H. (2025), "Helmet Detection in Underground Coal Mines via Dynamic Background Perception with Limited Valid Samples", *Machine Learning and Knowledge Extraction*, 7 (3): 64.
- [21]. Hernández-Ruiz A. C., Martínez-Nieto J. A., Buldain-Pérez J. D. (2021), "Steel bar counting from images with machine learning", *Electronics*, 10 (4): 402.
- [22]. Fan W., et al. (2021), "Machine learning applied to the design and inspection of reinforced concrete bridges: Resilient methods and emerging applications", *Structures*, Elsevier, 3954-3963.