

Phát hiện vật thể lạ trên băng tải bằng mô hình học máy

Phạm Thị Thanh Loan¹, Doãn Văn Hoàn²

¹ Bộ môn Tự động hoá, Trường Đại học Mở-Địa chất (phamthithanhloan@humg.edu.vn);

THÔNG TIN BÀI BÁO

TÓM TẮT

Quá trình:
Nhận bài
Chấp nhận
Đăng online

Từ khóa: Vật thể lạ; băng tải; Yolo.

Trong các ngành công nghiệp hiện đại, đặc biệt là khai thác khoáng sản, băng tải đóng vai trò thiết yếu trong việc vận chuyển vật liệu liên tục và hiệu quả. Tuy nhiên, sự xuất hiện của các vật thể lạ trên băng tải than là một thách thức nghiêm trọng, là nguyên nhân chính dẫn đến hư hỏng thiết bị, gián đoạn sản xuất và gây ra rủi ro an toàn lao động. Để giải quyết vấn đề này, bài báo trình bày một phương pháp tiếp cận dựa trên học sâu. Dữ liệu được thu thập từ hệ thống băng tải quy mô phòng thí nghiệm, được tiền xử lý và tăng cường nhằm cải thiện chất lượng của tập dữ liệu. Nghiên cứu tập trung vào việc đánh giá và so sánh hiệu suất của các kiến trúc YOLO khác nhau, bao gồm YOLOv8, YOLOv9, YOLOv10 và YOLOv11, để tìm ra mô hình tối ưu cho việc nhận diện vật thể lạ theo thời gian thực. Kết quả thực nghiệm cho thấy mô hình YOLOv9 đạt độ chính xác và tốc độ xử lý vượt trội, đáp ứng yêu cầu khắt khe của một hệ thống tự động hóa. Việc tích hợp các mô hình nhận diện đối tượng tiên tiến là một giải pháp khả thi để nâng cao an toàn và hiệu quả sản xuất trong môi trường công nghiệp. Nghiên cứu này mở ra hướng tích hợp hệ thống với các cánh tay robot, tạo tiền đề cho việc xây dựng các nhà máy khai thác thông minh và tự động hóa hoàn toàn trong bối cảnh của cuộc cách mạng công nghiệp 4.0.

© 2020 Trường Đại học Mở - Địa chất. Tất cả các quyền được bảo đảm.

1. Mở đầu

Trong bối cảnh công nghiệp hóa, hiện đại hóa ngày càng phát triển mạnh mẽ, hệ thống vận tải băng chuyền đóng vai trò thiết yếu trong nhiều ngành công nghiệp, từ sản xuất đến khai thác và logistics. Đặc biệt, trong lĩnh vực khai thác than và khoáng sản, băng tải là thành phần trung tâm, đảm bảo vận chuyển vật liệu liên tục và nâng cao năng suất.

Tuy nhiên, hoạt động của băng tải đối mặt với nhiều thách thức, trong đó đáng kể nhất là sự xuất hiện của các vật thể lạ như đá to, kim loại, cành cây hoặc vật liệu không mong muốn. Những vật thể này có thể gây ra hậu quả nghiêm trọng bao gồm hư hỏng thiết bị, gián đoạn sản xuất, tiềm ẩn nguy cơ mất an toàn lao động và làm gia tăng chi phí vận

hành. Để giải quyết vấn đề này, xu thế tích hợp các công nghệ tiên tiến như trí tuệ nhân tạo, Internet vạn vật (IoT) và thị giác máy tính vào hệ thống băng tải đang định hình lại cách thức sản xuất hiện đại, hướng tới các nhà máy thông minh trong kỷ nguyên công nghiệp 4.0. Nhiều nghiên cứu đã tập trung vào việc áp dụng các mô hình học sâu để phát hiện vật thể lạ trên băng tải. Phương pháp này cho thấy tiềm năng to lớn trong quá trình giám sát và xử lý vật liệu.

Tập dữ liệu hình ảnh và video chất lượng cao về quặng sắt trên băng tải mô hình phòng thí nghiệm, nhằm phát hiện vật thể lạ như gỗ, nhựa, đá gangue, jaspilite đã được thiết lập [1]. Dữ liệu được thu thập bằng camera NVIDIA Jetson TX2 ở độ phân giải cao và tốc độ 120 FPS, sau đó được gán nhãn thủ công và tiền xử lý bằng mô hình YOLO để tập trung vào vùng quan tâm. Ưu điểm chính của tập dữ liệu là cung cấp một môi trường

* Tác giả liên hệ

E - mail: phamthithanhloan@humg.edu.vn

DOI:

được kiểm soát để đánh giá các phương pháp phát hiện dị vật bằng thị giác máy tính cho quặng sắt.

Tuy nhiên, nghiên cứu còn một số hạn chế: dữ liệu được thu thập trên băng tải trong điều kiện môi trường ổn định, chưa phản ánh được sự phức tạp của yếu tố khai thác thực tế (như độ ẩm, bụi bẩn, biến đổi ánh sáng). Ngoài ra, số lượng và sự đa dạng của các lớp vật thể lạ còn giới hạn.

Vấn đề độ chính xác thấp và hiệu suất thời gian thực kém trong việc phát hiện vật thể lạ trên băng tải, nguyên nhân chính gây rách băng tải đã được đề xuất cải tiến bằng phương pháp mới dựa trên mô hình Nanode [2]. Hệ thống phần cứng được thiết kế với bộ xử lý ARM và phần mềm dựa trên Android. Các cải tiến bao gồm: thay thế hàm mất mát vị trí gốc bằng Siou để tăng độ chính xác và tốc độ hội tụ; sửa đổi mạng xương sống bằng cách sử dụng tích chập 5x5 để mở rộng trường tiếp nhận và giảm tham số mô hình. Kết quả thử nghiệm cho thấy hệ thống đạt độ chính xác phát hiện trung bình 94.3% và tốc độ 30 khung hình/giây, đáp ứng yêu cầu thời gian thực và không đòi hỏi phần cứng cao cấp.

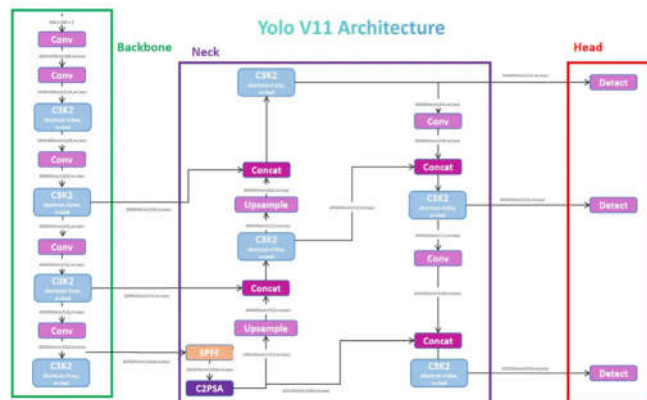
Thuật toán Yolov8-SEML, một phiên bản cải tiến của Yolov8n đã được đề xuất, nhằm phát hiện vật thể lạ nhẹ và hiệu quả trên băng tải mỏ than, giải quyết các vấn đề về tốc độ chậm, tham số quá lớn và yêu cầu tính toán cao. Mạng xương sống được thay thế bằng StarNet, tích hợp module C2f-EIEM để tăng cường học đặc trưng, cải tiến SPPF với cơ chế chú ý LSKA, thay thế cơ chế chú ý kênh bằng C2f_MLCA, thêm đầu dò Detect-LSDECD và sử dụng hàm mất mát MPDIoU [3].

Kết quả thử nghiệm cho thấy mô hình đề xuất đã giảm 80% số lượng tham số, giảm 60.49% tải tính toán và 69.35% kích thước lưu trữ, đồng thời tăng độ chính xác lên 1.9% so với mô hình gốc, trong khi vẫn duy trì tỷ lệ recall. Mô hình cũng đạt tốc độ phát hiện nhanh (5.58 it/s so với 4.74 it/s của bản gốc) và hiệu quả hơn trong các môi trường phức tạp như ánh sáng yếu hay vật thể mờ. Tuy nhiên, khả năng tổng quát hóa của mô hình còn hạn chế vì chỉ tập trung vào các vật thể lạ cụ thể như đá và neo.

Tiếp nối xu hướng này, nhóm tác giả đặt ra mục tiêu xây dựng một hệ thống phát hiện vật thể lạ trên băng tải trong hệ thống khai thác than bằng cách ứng dụng thị giác máy tính và các kỹ thuật

điều khiển thông minh. Hệ thống sử dụng mô hình nhận dạng đối tượng là các cấu trúc khác nhau của YOLO để giám sát băng tải trong thời gian thực. Dữ liệu được thu thập, tăng cường nhằm cải thiện chất lượng của các đặc trưng. Hệ thống có khả năng mở rộng để tích hợp với robot công nghiệp, tạo nên dây chuyền khai thác và xử lý vật liệu tự động hóa cao.

2. Các mô hình nhận dạng đối tượng YOLO



Hình 1. Kiến trúc tổng quát của YOLO

YOLO (You Only Look Once) là một mô hình mạng CNN cho việc phát hiện, nhận dạng, phân loại đối tượng. YOLO được tạo ra từ việc kết hợp giữa các convolutional layers và connected layers. Trong đó các convolutional layers có nhiệm vụ trích xuất các feature của ảnh, full-connected layers dự đoán xác suất và tọa độ của đối tượng. Thay vì sử dụng nhiều bước như các mô hình truyền thống, YOLO chia hình ảnh thành một lưới, trong đó mỗi ô lưới chịu trách nhiệm dự đoán các đối tượng xuất hiện trong khu vực đó. Mô hình đồng thời dự đoán vị trí và xác suất của từng lớp đối tượng, giúp đạt được tốc độ xử lý vượt trội. Được phát triển bởi Ultralytics, YOLO nổi bật với hiệu năng cao, khả năng mở rộng và tính linh hoạt trên nhiều nền tảng. Sự xuất hiện của YOLO đánh dấu một bước tiến quan trọng trong thị giác máy tính (Computer Vision), đặc biệt trong các ứng dụng thời gian thực như giám sát an ninh, xe tự hành, robot công nghiệp, phân tích video thông minh...[4]-[8].

Kiến trúc YOLO gồm ba thành phần chính: Backbone; Neck và Head.

+) Backbone: chịu trách nhiệm trích xuất đặc trưng từ ảnh đầu vào ở nhiều tỷ lệ. Quá trình được

thực hiện thông qua việc xếp chồng các lớp tích chập và các khối chuyên biệt, tạo ra các bản đồ đặc trưng đa độ phân giải để phục vụ cho các bước xử lý sau.

+) Các lớp tích chập: Cấu trúc các lớp tích chập truyền thống được duy trì nhằm giảm kích thước không gian ảnh đầu vào và tăng số lượng kênh đặc trưng.

+) Neck: đóng vai trò tổng hợp đặc trưng từ nhiều tầng khác nhau và truyền chúng đến Head để thực hiện dự đoán. Neck thực hiện phép nối và lấy mẫu từ bản đồ đặc trưng ở các độ phân giải khác nhau, giúp mô hình xử lý tốt đối tượng đa tỷ lệ.

+) Head: Cơ chế dự đoán, xuất ra tọa độ hộp giới hạn và xác suất phân loại đối tượng.

Đầu ra của mô hình YOLO là một vector bao gồm các thành phần:

$$y^T = [p_0, (t_x, t_y, t_w, t_h), (p_1, p_2, \dots, p_c)] \quad (1)$$

Trong đó:

p_0 : xác suất dự báo vật thể xuất hiện trong bounding box.

(t_x, t_y, t_w, t_h) : giúp xác định bounding box.

t_x, t_y : tọa độ tâm; t_w, t_h : kích thước rộng, dài của bounding box.

$(p_1, p_2, \dots, p_c)$: véc tơ phân phối xác suất dự báo của các lớp.

*.) Thông số đánh giá hiệu suất của mô hình YOLO

+) mAP50 (mean Average Precision) với IOU threshold 0.5: trung bình của AP được tính cho tất cả các lớp.

+) P(Precision): đo lường độ chính xác mô hình.

$$P = \frac{TP}{TP + FP} \quad (2)$$

TP là số lượng True Positive (dự đoán đúng) và FP là số lượng False Positive (dự đoán sai).

+) R (Recall): đo lường khả năng tìm thấy tất cả các kết quả tích cực. Recall được tính bằng công thức:

$$R = \frac{TP}{TP + FN} \quad (3)$$

FN là số lượng False Negative (không dự đoán được)

+) F1 score: trung bình điều hòa (Harmonic Mean) của precision và recall. F1 score cao cho thấy cả Precision và Recall đều cao.

Cho đến nay, nhiều cấu trúc YOLO đã được phát triển và cải tiến. Trong đó, YOLOv8 là một bước tiến đáng kể, nổi bật với khả năng nhận diện đối tượng hiệu quả và cân bằng tốt giữa tốc độ và độ chính xác. YOLOv8 dễ dàng sử dụng, hỗ trợ nhiều tác vụ, cùng với hiệu suất mạnh mẽ trên các thiết bị phổ thông. Tuy nhiên, nhược điểm là độ chính xác không cao, đặc biệt đối với các đối tượng nhỏ.

YOLOv9 cải thiện độ chính xác nhờ áp dụng khái niệm Giga-Grid và các khối kiến trúc mới như GSC-block. Ưu điểm chính là khả năng xử lý tốt hơn, cải thiện hiệu suất trên các tập dữ liệu phức tạp. Nhược điểm của v9 là tốc độ xử lý chậm hơn do sự phức tạp của kiến trúc.

YOLOv10 tập trung vào tối ưu hóa tốc độ và hiệu quả cho việc triển khai trên thiết bị nhúng. Ưu điểm nổi bật là tốc độ suy luận nhanh với độ chính xác cao, nhờ việc loại bỏ các thao tác hậu xử lý tốn kém. Tuy nhiên, YOLOv10 yêu cầu việc tinh chỉnh và tối ưu hóa phức tạp hơn.

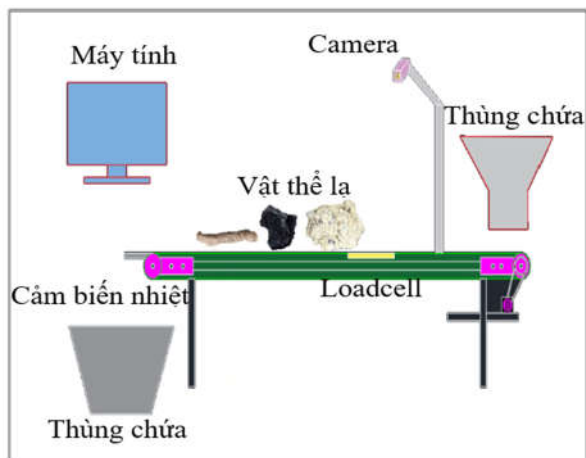
YOLOv11 là phiên bản mới, kết hợp các điểm mạnh từ các phiên bản trước. YOLOv11 đạt được độ chính xác cao, xử lý hiệu quả, đặc biệt trong các tác vụ phân đoạn và nhận dạng đối tượng đa tỷ lệ. Đồng lại, chúng yêu cầu tài nguyên tính toán cao hơn cho quá trình huấn luyện. Do đó, bài báo này sẽ so sánh và đánh giá hiệu suất của bốn kiến trúc YOLO (YOLOv8, YOLOv9, YOLOv10 và YOLOv11) để xác định mô hình tối ưu cho việc nhận diện vật thể lạ trên băng tải.

3. Triển khai hệ thống dò vật thể lạ trên băng tải

3.1. Thiết lập mô hình băng tải

Mô hình được thiết kế nhằm nghiên cứu và đánh giá hệ thống phát hiện vật thể lạ trên băng tải trong khai thác than bằng thị giác máy tính. Các chức năng cốt lõi bao gồm: Giám sát băng tải trong thời gian thực bằng camera và hệ thống xử lý ảnh;

Phát hiện vật thể bất thường (đá lớn, kim loại, gỗ, tạp chất không mong muốn) trên băng tải.



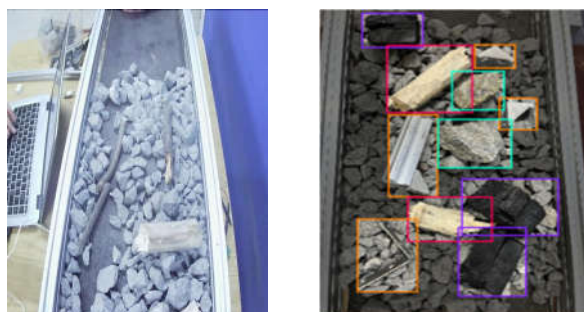
Hình 2. Sơ đồ hệ thống băng tải

Bảng 1. Thông số kỹ thuật

TT	Thiết bị	Thông số
1	Băng tải	Dài: 1.2m; rộng 0.2m Tải trọng vận chuyển: 2-5kg Tốc độ: 0.1-0.5 m/s Động cơ: DC
2	Camera	Webcam 2K
3	Loadcell	Dải đo: tối đa 5kg
4	CB nhiệt	LM75A: 55°C-125°C
5	Vật thể	Branch; Coal; Iron; Stick; Stone
6	Màn hình	LCD

3.2. Thu thập dữ liệu

1697 ảnh được lấy từ camera độ phân giải cao trong điều kiện môi trường ánh sáng với các góc chụp khác nhau.



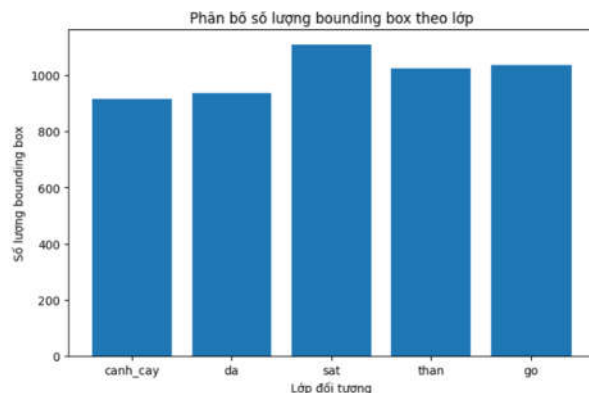
Hình 3. Thu thập dữ liệu và gán nhãn

Ảnh được resize về 640x640 và được gán nhãn bằng Roboflow. Dữ liệu được chia tỷ lệ train_val_test theo tỷ lệ 70-20-10.

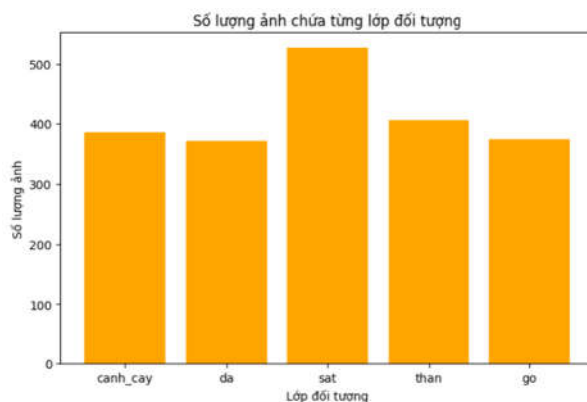
3.3. Tiền xử lý dữ liệu

Tiền xử lý dữ liệu là một giai đoạn nền tảng tối quan trọng, quyết định trực tiếp đến hiệu suất của mô hình học sâu. Việc làm sạch, chuẩn hóa, và tăng cường dữ liệu một cách chiến lược đảm bảo ổn định quá trình hội tụ, tránh hiện tượng quá khớp, từ đó đạt được kết quả chính xác và đáng tin cậy trong các kịch bản thực tế.

Dữ liệu được trực quan hóa để kiểm tra tính chính xác của nhãn gán và phát hiện các lỗi có thể xảy ra trong quá trình gán nhãn. Quá trình tăng cường dữ liệu giúp mô hình học được nhiều đặc trưng hơn, giảm hiện tượng overfitting và nâng cao khả năng khái quát hóa. Đồng thời, ảnh được chuẩn hóa, đưa giá trị pixel về [0,1] nhằm thống nhất kích thước và phạm vi dữ liệu đầu vào, giúp quá trình huấn luyện nhanh và ổn định.



Hình 4. Phân bố số lượng bounding box theo lớp



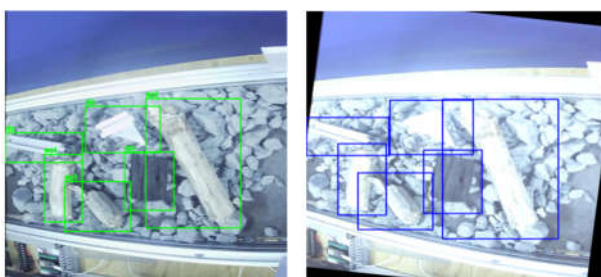
Hình 5. Số lượng ảnh chứa từng lớp đối tượng

Biểu đồ thống kê cho thấy sự phân bố không đồng đều giữa các lớp đối tượng trong tập dữ liệu.

Lớp như sắt có tỷ lệ cao hơn đáng kể so với các lớp khác. Sự mất cân bằng này có thể dẫn đến hiện tượng mô hình thiên lệch, có xu hướng dự đoán tốt đối với lớp chiếm ưu thế nhưng kém chính xác đối với các lớp hiếm.

Bảng 2. Số lượng ảnh được tăng cường

	Cảnh cây	Sắt	Than	Gỗ	Đá
Số lượng	390	520	405	375	370
Tăng cường	130		115	145	150



Hình 6. Dữ liệu được trực quan hoá và tăng cường chọn lọc

3.4. Huấn luyện mô hình

Huấn luyện mô hình được tiến hành sau khi hoàn tất quá trình tiền xử lý dữ liệu. Trong nghiên cứu này, các phiên bản YOLO từ v8 đến v12 được lựa chọn nhằm so sánh hiệu quả giữa các thể hệ mô hình hiện đại. Tập dữ liệu sau khi tăng cường được chia thành ba phần: huấn luyện, kiểm định và kiểm thử với tỷ lệ 70:20:10.

Tham số mô hình:

+ batch size: 16;

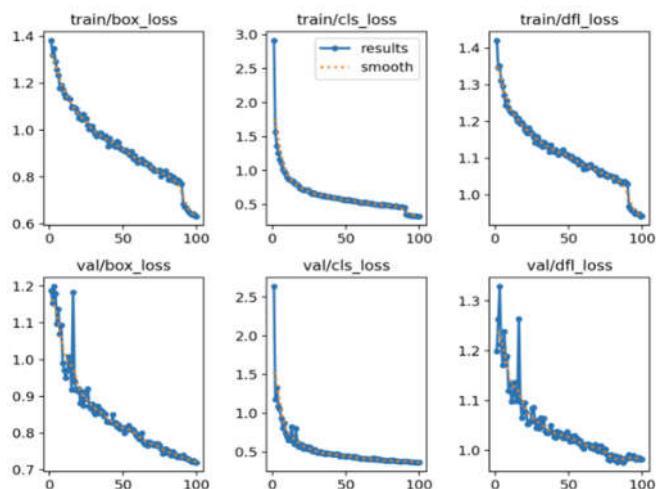
+ learning rate: 0.01

+ số epoch: 100

+ Thuật toán tối ưu: AdamW

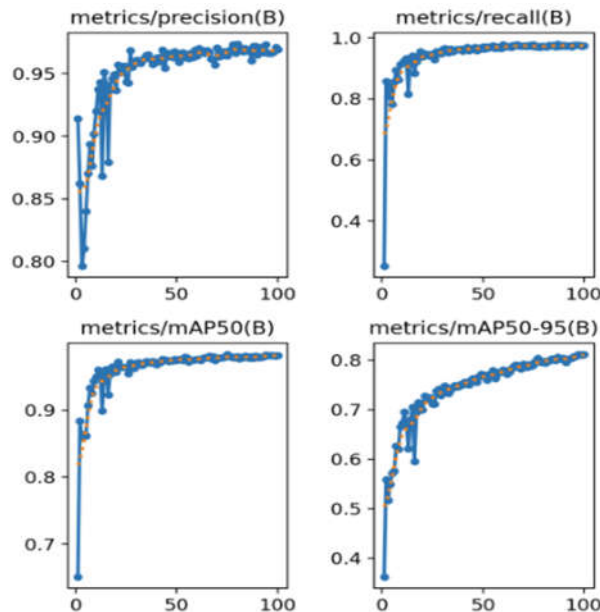
Quá trình huấn luyện được triển khai trên môi trường Google Colab với GPU T4 nhằm rút ngắn thời gian xử lý. Sau mỗi epoch, mô hình được đánh giá dựa trên các chỉ số chuẩn gồm precision, recall, mAP@50 và mAP@50-95.

Biểu đồ loss curves cho thấy các thành phần hàm mất mát (box_loss, cls_loss, dfl_loss) trên cả tập huấn luyện và tập kiểm định đều có xu hướng giảm dần ổn định theo số epoch.



Hình 7. Biểu đồ loss curves

Ở giai đoạn đầu, các giá trị loss giảm nhanh, phản ánh khả năng mô hình học được đặc trưng từ dữ liệu. Sau khoảng 50 epoch, tốc độ giảm loss chậm lại và tiến tới trạng thái hội tụ. Sự tương đồng giữa các đường loss của tập huấn luyện và tập kiểm định cho thấy mô hình không bị overfitting, quá trình huấn luyện đạt hiệu quả và mô hình đã học được các đặc trưng hình ảnh một cách ổn định.



Hình 8. Chỉ số đánh giá hiệu suất mô hình

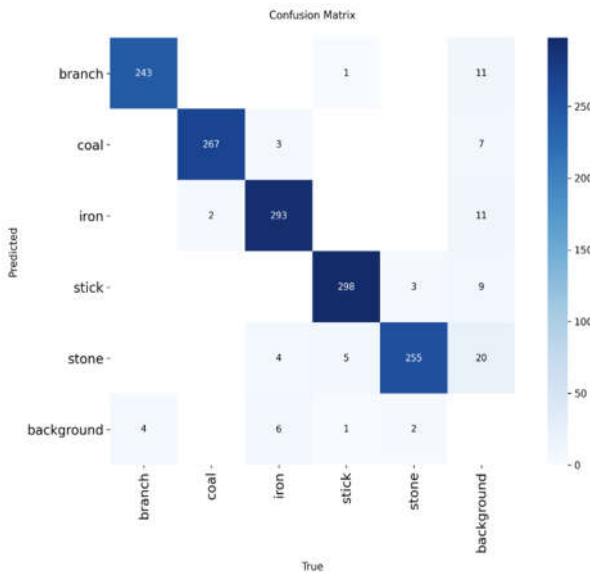
Các chỉ số đánh giá mô hình Precision, Recall, mAP50, mAP50-95 đều có xu hướng tăng nhanh trong giai đoạn đầu huấn luyện và dần đạt trạng thái ổn định sau khoảng 50 epoch.

+) Precision đạt giá trị tiệm cận ~ 0.97 , phản ánh tỷ lệ dự đoán đúng trong số các dự đoán của mô hình là rất cao.

+) Recall đạt gần 0.98: mô hình hầu như không bỏ sót đối tượng trong ảnh.

+) mAP50 tiến gần đến 0.98, khẳng định hiệu suất phát hiện đối tượng ở ngưỡng IoU 0.5 là rất tốt.

+) mAP50-95 đạt ~ 0.81 , thể hiện khả năng tổng quát hóa của mô hình trên nhiều mức IoU khác nhau vẫn duy trì ở mức cao.



Hình 9. Hiệu suất phân loại của mô hình

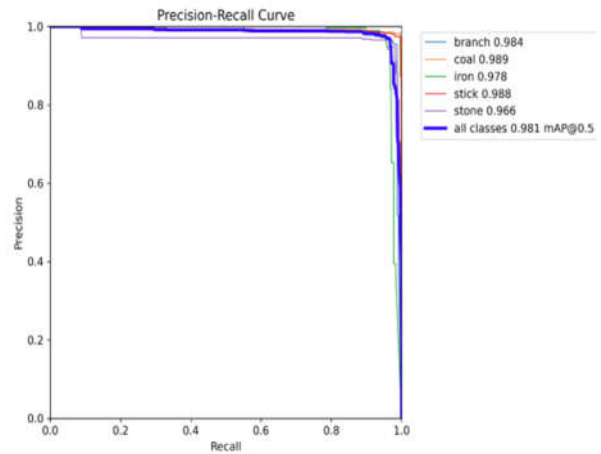
Ma trận nhầm lẫn thể hiện độ chính xác cao của mô hình.

+) Đường chéo chính (True Positive) có giá trị lớn, branch = 243, coal = 267, iron = 293, stick = 298, stone = 255, chứng tỏ phần lớn đối tượng được phân loại đúng.

+) Sai nhầm ít xảy ra: một số đối tượng bị nhầm lẫn với background hoặc nhầm lẫn lẫn với nhau.

+) Lớp background cũng có sai số nhỏ, khi một số đối tượng bị mô hình nhầm là background.

Biểu đồ Precision-Recall (Hình 10) cho thấy mô hình đạt hiệu suất rất cao ở tất cả các lớp.

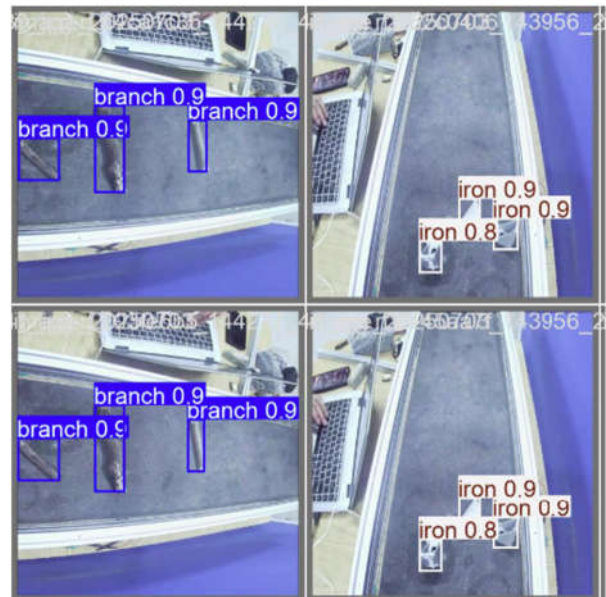


Hình 10. Đường cong Precision-Recall (PR Curve)

+) Giá trị Precision và Recall đều xấp xỉ 1.0 với đường cong áp sát góc trên bên phải của đồ thị, thể hiện mô hình duy trì được sự cân bằng tốt giữa khả năng phát hiện đầy đủ (Recall) và độ chính xác trong dự đoán (Precision).

+) Coal đạt hiệu suất cao nhất (Precision ≈ 0.989), tiếp theo là stick (0.988) và branch (0.984), iron có Precision thấp hơn một chút (0.978), stone là lớp có độ chính xác thấp nhất (0.966), phản ánh xu hướng nhầm lẫn với các đối tượng khác.

+) mAP@0.5 toàn bộ các lớp đạt 0.981, thể hiện độ bao quát và khả năng khái quát tốt của mô hình trên nhiều loại đối tượng.



Hình 11. Kết quả phát hiện vật thể

Kết quả phân loại trên hình 11 đã chứng minh khả năng phân loại và định vị các vật thể mục tiêu với độ tin cậy cao. Mô hình đã phát hiện ba vật thể thuộc lớp "branch" với các điểm tin cậy lần lượt là 0.9, 0.9, và 0.9. Sự đồng nhất trong các giá trị này phản ánh độ ổn định và tin cậy cao của mô hình đối với lớp vật thể "branch". Tương tự, ba vật thể thuộc lớp "iron" cũng được phát hiện với các điểm tin cậy 0.9, 0.9, và 0.8. Mặc dù có một chút biến thiên nhỏ ở một trong các điểm tin cậy, tất cả các giá trị này đều vượt xa ngưỡng chấp nhận thông thường (thường là 0.5), cho thấy khả năng phân biệt mạnh mẽ của mô hình.

4. Kết quả và thảo luận

Các cấu trúc YOLOv8s; YOLOv9s; YOLOv10s; YOLOv11s được huấn luyện với cùng tập dữ liệu

của dự án. Hiệu năng của các cấu trúc YOLO khác nhau được so sánh nhằm lựa chọn cấu trúc tối ưu cho bài toán nhận dạng. Các chỉ số đánh giá sử dụng bao gồm độ chính xác (Precision), độ bao phủ (Recall), chỉ số mAP ở ngưỡng IoU 0.5 và mAP trung bình trên các ngưỡng IoU từ 0.5 đến 0.95 (mAP@0.5:0.95). Bên cạnh đó, các thông số về độ phức tạp mô hình như số lượng tham số (Parameters), phép tính suy luận (FLOPs), kích thước mô hình và tốc độ suy luận (FPS) cũng được xem xét. Việc so sánh toàn diện này không chỉ giúp đánh giá chất lượng phát hiện đối tượng mà còn cân nhắc đến hiệu quả triển khai trong thực tế, nơi yêu cầu mô hình vừa có độ chính xác cao vừa đảm bảo tốc độ xử lý thời gian thực.

Bảng 3. So sánh hiệu suất các cấu trúc YOLO.

TT	Mô hình	Precision	Recall	mAP0.5	mAP@0.5:0.95	Params(M)	FLOPs(G)	FPS(GPU)
1	YOLOv8s	0.969	0.980	0.982	0.83	11.12	28.4	8.5
2	YOLOv9c	0.969	0.982	0.987	0.835	25.323	102,3	17.7
3	YOLOv10n	0.965	0.964	0.982	0.818	7.2	21.4	6.9
4	YOLOv11s	0.966	0.975	0.981	0.811	2.58	6.3	5.5
5	YOLOv12s	0.973	0.976	0.98	0.818	9.233	21.2	11.3

Bảng 3 cho thấy sự khác biệt rõ rệt giữa các thế hệ mô hình. Ở mức tổng thể, YOLOv9s đạt được độ chính xác cao nhất với Precision = 0.969, Recall = 0.982, mAP@0.5 = 0.987 và mAP@0.5:0.95 = 0.835, vượt trội hơn so với YOLOv8s và YOLOv11s. Tuy nhiên, để đạt được hiệu năng này, YOLOv9s có số lượng tham số lớn hơn đáng kể (25.3M tham số, 102.3 GFLOPs) so với YOLOv8s (11.1M tham số, 28.4 GFLOPs) và đặc biệt so với YOLOv11s (chỉ 2.58M tham số, 6.3 GFLOPs). Điều này phản ánh sự đánh đổi giữa độ chính xác và độ phức tạp mô hình.

Về tốc độ xử lý, YOLOv9s có tốc độ 17.7 FPS, nhanh gấp đôi YOLOv8s (8.5 FPS) và gấp ba lần YOLOv11s (5.5 FPS). Mặc dù YOLOv9s có độ phức tạp cao, nhưng khi chạy trên GPU, mô hình vẫn duy trì khả năng suy luận nhanh, phù hợp cho các ứng dụng yêu cầu cả độ chính xác lẫn tốc độ. Trong khi đó, YOLOv11s tuy nhẹ nhất về tham số và FLOPs, nhưng lại cho kết quả mAP thấp hơn

(mAP@0.5:0.95 = 0.811), hạn chế về khả năng khái quát hóa so với các phiên bản trước.

Như vậy, YOLOv9s là lựa chọn cân bằng nhất giữa độ chính xác và tốc độ, thích hợp cho các ứng dụng yêu cầu phát hiện chính xác trong thời gian thực. Ngược lại, YOLOv11s có thể phù hợp hơn với các hệ thống nhúng hoặc môi trường tính toán hạn chế, khi sự gọn nhẹ của mô hình được ưu tiên hơn hiệu năng phát hiện.

5. Kết luận

Nghiên cứu này nhằm giải quyết thách thức phát hiện vật thể lạ trên băng tải than bằng phương pháp học sâu. Bộ dữ liệu hình ảnh đa dạng gồm 1697 ảnh đã được thu thập cho mô hình băng tải quy mô phòng thí nghiệm. Dữ liệu sau được tiền xử lý và tăng cường chọn lọc, chuẩn hóa kích thước ảnh về 640x640 pixel để cải thiện chất lượng, đảm bảo khả năng tổng quát hóa và chống quá khớp cho mô hình học sâu. Tập dữ liệu được

chia theo tỷ lệ 70% huấn luyện, 20% kiểm định và 10% kiểm thử.

Các kiến trúc YOLO tiên tiến, bao gồm YOLOv8, YOLOv9, YOLOv10 và YOLOv11, YOLOv12 đã được huấn luyện và đánh giá với các tham số nhất quán (batch size 16, learning rate 0.01, 100 epoch với optimizer AdamW). Kết quả huấn luyện cho thấy các đường cong hàm mất mát hội tụ ổn định, không có dấu hiệu quá khớp đáng kể, và các chỉ số hiệu suất như Precision, Recall, mAP đều tăng nhanh và đạt mức cao.

Nghiên cứu này khẳng định tính khả thi và hiệu quả của việc tích hợp các mô hình YOLO tiên tiến vào hệ thống phát hiện vật thể lạ, góp phần nâng cao an toàn và năng suất trong khai thác, đồng thời mở ra tiềm năng tích hợp robot để tự động hóa hoàn toàn các nhà máy thông minh.

Lời cảm ơn

Tác giả bài báo trân trọng cảm ơn Bộ môn Tự động hoá – Trường Đại học Mở địa chất đã hỗ trợ cung cấp thiết bị để xây dựng mô hình.

Đóng góp của các tác giả

Phạm Thị Thanh Loan: Tổng hợp tài liệu, viết báo cáo, phân tích kết quả.

Doãn Văn Hoàn: Xây dựng mô hình bằng tải; xây dựng bộ dữ liệu; huấn luyện mô hình.

Tài liệu tham khảo

- [1] Frederico L.M. de Sousa , Thiago E.A. de Oliveira , Saul E. Delabrida Silva , Bruno Nazário Coelho (2025). Image Dataset for Foreign Object Detection in Iron Ore Conveyor Belt Systems. *Elsevier Inc.* <https://doi.org/10.1016/j.dib.2025.111537>
- [2] Di miao, Yimin Wang, Ling yang, and Shihao wei (2023). Foreign Object Detection Method of Conveyor Belt Based on Improved Nanodet. *IEEE Access*.
- [3]. Jierui ling, Zhibo fu & Xinpeng Yuan (2025). Lightweight coal mine conveyor belt foreign object detection based on improved Yolov8n. *Scientific Reports*, tập 15, art. no. 10361.
- [4] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2015). You Only Look Once: Unified, Real-Time Object Detection. *arXiv preprint arXiv:1506.02640*.
- [5] Howse, J., & Minichino, J. (2020). Learning OpenCV 4 Computer Vision with Python 3 (Third Edition). *Packt Publishing*.
- [6] Prakash, I. V., & Palanivelan, M. (2024). A Study of YOLO (You Only Look Once) to YOLOv8. Trong: Algorithms in Advanced Artificial Intelligence. *CRC Press / Taylor & Francis Group*.
- [7] Haque, M. (2023). Mastering YOLO: Build an Automatic Number Plate Recognition System. *Independently Published*.
- [8] Omar, A. (2023). YOLO Object Detection Explained: Definitive Reference. *Independently Published*.