BỘ GIÁO DỤC VÀ ĐÀO TẠO
ĐẠI HỌC PHENIKAA

BÁO CÁO BÀI TẬP LỚN
HỌC PHẦN: THỊ GIÁC MÁY TÍNH

PHÁT HIỆN LỖI CHAI NƯỚC TRÊN BĂNG CHUYỀN

Họ và Tên	Mã sinh viên	Ngành
Nguyễn Ngọc Trâm	22010506	Công nghệ thông tin Việt Nhật
Nguyễn Thị Thu Thảo	22010496	Công nghệ thông tin Việt Nhật
Đỗ Thị Ngân	22010105	Công nghệ thông tin Việt Nhật

Giảng viên hướng dẫn: TS. Nguyễn Văn Tới

Hà Nội – 2025

LỜI CẢM ƠN

Lời đầu tiên, em xin được gửi lời cảm ơn chân thành và sâu sắc nhất đến thầy **Nguyễn Văn Tới** giảng viên bộ môn Thị giác máy tính**.** Thầy đã tận tình giảng dạy, truyền đạt những kiến thức chuyên môn cốt lõi về thị giác máy tính và học sâu, đồng thời đưa ra những định hướng và góp ý học thuật quý báu, giúp em có nền tảng vững chắc để triển khai và hoàn thiện đề tài này.

Tiếp theo, em xin gửi lời cảm ơn đặc biệt đến bạn **Nguyễn Xuân Phước**, người đã đảm nhiệm vai trò nhóm trưởng. Bạn không chỉ xuất sắc trong việc điều phối công việc chung, lập kế hoạch và bám sát tiến độ, mà còn là người trực tiếp cùng em thảo luận, giải quyết nhiều vấn đề kỹ thuật phức tạp trong quá trình huấn luyện mô hình YOLO. Sự chủ động, tinh thần trách nhiệm và khả năng lãnh đạo của bạn là yếu tố then chốt giúp nhóm vượt qua khó khăn và hoàn thành tốt mục tiêu đề ra.

Em cũng xin trân trọng cảm ơn Ban Giám hiệu Đại học Phenikaa cùng toàn thể quý thầy cô đã tạo mọi điều kiện học tập, nghiên cứu tốt nhất và trang bị cho em những nền tảng tri thức quý giá trong suốt những năm tháng học tập tại trường.

Nhóm 7

Hà Nội, 15/11/2025

BẢNG PHÂN CÔNG NHIỆM VỤ

Họ và Tên	Nhiệm vụ
Đỗ Thị Ngân	Thu thập data, xử lý dữ liệu
Nguyễn Ngọc Trâm	Thu thập data, xây dựng mô hình
Nguyễn Thị Thu Thảo	Xử lý dữ liệu, train model

MỤC LỤC

DANH MỤC HÌNH ẢNH

DANH MỤC BẢNG BIỂU

CHƯƠNG 1. GIỚI THIỆU

1.1 Giới thiệu

Trong ngành công nghiệp sản xuất đồ uống đóng chai, bao gồm các sản phẩm như nước uống tinh khiết, nước giải khát hay đồ uống thể thao, công đoạn kiểm tra chất lượng cuối dây chuyền đóng một vai trò tối quan trọng. Đây được xem là hàng rào bảo vệ cuối cùng, đảm bảo rằng mỗi sản phẩm thành phẩm trước khi được đóng thùng và xuất kho phải đáp ứng đầy đủ các tiêu chí nghiêm ngặt về cả ngoại quan lẫn chất lượng. Các tiêu chí này bao gồm việc chai không bị móp méo hay biến dạng, mức dung dịch bên trong đạt chuẩn quy định, nắp chai được vặn chặt và còn nguyên vẹn, đồng thời nhãn mác phải được dán ngay ngắn, không bị rách, lệch hoặc in sai. Bất kỳ một sản phẩm lỗi nào nếu vô tình lọt ra thị trường đều có thể dẫn đến những hậu quả nghiêm trọng, không chỉ là các chi phí trực tiếp như thu hồi sản phẩm, xử lý khiếu nại của khách hàng, mà còn là nguy cơ ảnh hưởng sâu sắc đến uy tín và hình ảnh thương hiệu mà doanh nghiệp đã dày công xây dựng.

Trong bối cảnh đó, việc tự động hóa khâu kiểm định chất lượng bằng cách ứng dụng các công nghệ tiên tiến như thị giác máy tính và học sâu đang trở thành một xu hướng tất yếu và là yêu cầu cấp thiết, nhằm thay thế hoặc hỗ trợ phương pháp quan sát thủ công truyền thống vốn còn nhiều hạn chế.

1.2 Bối cảnh

Hiện nay, các dây chuyền sản xuất nước uống đóng chai tại nhiều nhà máy lớn đã đạt đến mức độ tự động hóa rất cao ở hầu hết các công đoạn cơ khí. Quy trình từ chiết rót, đóng nắp, dán nhãn, in hạn sử dụng cho đến đóng lốc và đóng thùng... phần lớn đều được cơ khí hóa và điều khiển bằng các hệ thống tự động. Tuy nhiên, nghịch lý là một trong những khâu quan trọng nhất, ảnh hưởng trực tiếp đến cảm nhận của khách hàng là khâu kiểm tra ngoại quan thành phẩm, thì lại thường vẫn phụ thuộc chủ yếu vào sức người. Trong khâu này, các công nhân QC phải đứng giám sát liên tục để phát hiện và loại bỏ bằng tay các chai không đạt chuẩn khỏi băng chuyền. Các lỗi phổ biến thường gặp rất đa dạng, bao gồm: lỗi về

thân chai (móp, biến dạng), lỗi về nắp chai (vặn lệch ren, hở) và các lỗi về nhãn mác (lệch, nhăn, rách). Vấn đề mấu chốt và cũng là thách thức lớn nhất của công đoạn này chính là tốc độ. Trong sản xuất công nghiệp, tốc độ băng chuyền có thể rất cao, lên đến hàng chục, thậm chí hàng trăm chai mỗi phút. Điều này đặt ra một yêu cầu kép cho hệ thống kiểm tra, nó không chỉ phải phát hiện lỗi một cách chính xác mà còn phải phản ứng kịp thời trong thời gian thực, nếu không chai lỗi sẽ đi qua và lọt vào khâu đóng gói.

1.3 Lý do chọn đề tài

Phương pháp kiểm tra thủ công bằng mắt thường, vốn đang được áp dụng phổ biến, bộc lộ nhiều hạn chế nghiêm trọng không thể khắc phục. Thứ nhất, hiệu suất kiểm tra thấp và hoàn toàn thiếu tính ổn định. Con người không thể duy trì sự tập trung cao độ trong thời gian dài, dẫn đến mệt mỏi, lơ là và bỏ sót lỗi, đặc biệt là vào cuối ca làm việc. Chất lượng kiểm định vì thế mà không đồng nhất giữa các công nhân và các thời điểm khác nhau. Thứ hai, phương pháp thủ công không có khả năng mở rộng quy mô. Khi nhà máy muốn tăng tốc dây chuyền để nâng cao năng suất, con người đơn giản là không thể đáp ứng kịp tốc độ quan sát. Thứ ba, việc kiểm tra bằng mắt không để lại dữ liệu hình ảnh, khiến doanh nghiệp thiếu khả năng truy vết, thống kê và phân tích nguyên nhân gốc rễ của lỗi để cải tiến quy trình sản xuất. Mặc dù các hệ thống thị giác máy tính truyền thống (dựa trên các thuật toán xử lý ảnh cổ điển như lọc biên, phân ngưỡng màu sắc, hay so khớp mẫu) đã được nghiên cứu, chúng lại tỏ ra cứng nhắc, phức tạp khi cài đặt và rất nhạy cảm với sự thay đổi của môi trường như ánh sáng phản xạ, bóng lóa trên bề mặt chai nhựa, hoặc khi nhà máy thay đổi mẫu mã sản phẩm. Sự phát triển vượt bậc của Trí tuệ Nhân tạo, đặc biệt là các mô hình phát hiện đối tượng dựa trên học sâu, đã mở ra một hướng tiếp cận mới ưu việt hơn hẳn. Trong đó, mô hình YOLO nổi bật lên như một giải pháp lý tưởng nhờ kiến trúc độc đáo, cho phép vừa xác định vị trí vừa phân loại đối tượng chỉ trong một lần xử lý duy nhất, đạt tốc độ suy luận cực cao. Chính đặc tính này làm cho YOLO trở thành ứng cử viên hàng đầu cho các bài toán đòi hỏi xử lý thời gian thực trong môi trường công nghiệp khắc nghiệt. Xuất phát từ nhu cầu thực tế cấp bách của sản xuất và tiềm năng ứng dụng mạnh

mẽ của công nghệ, đề tài "Phát hiện lỗi chai nước trên băng chuyền" môn thị giác máy tính dùng YOLO Detection được lựa chọn, nhằm giải quyết trực tiếp bài toán tự động hóa kiểm định chất lượng, góp phần hiện thực hóa mô hình nhà máy thông minh.

1.4 Mục tiêu

Đề tài này được thực hiện nhằm hướng tới ba mục tiêu nghiên cứu chính, rõ ràng và cụ thể. Mục tiêu thứ nhất, đó là xây dựng một hệ thống phát hiện lỗi chai nước hoàn chỉnh dựa trên hình ảnh. Hệ thống này có khả năng nhận đầu vào là luồng hình ảnh hoặc video thu được trực tiếp từ camera giám sát băng chuyền, và tạo ra đầu ra là thông tin nhận dạng dưới dạng các hộp bao cho từng chai xuất hiện trong khung hình. Quan trọng hơn, hệ thống phải đồng thời phân loại được từng chai đó là "đạt" hay "lỗi", và lý tưởng nhất là có thể chỉ rõ được loại lỗi cụ thể ví dụ: móp thân, lệch nắp. Mục tiêu thứ hai, đề tài sẽ tập trung vào việc ứng dụng và kiểm chứng khả năng của mô hình YOLO để giải quyết bài toán đặc thù này. Quá trình này nhằm đánh giá xem YOLO có thực sự đáp ứng được yêu cầu về tốc độ xử lý và độ chính xác trong điều kiện thực tế của dây chuyền sản xuất công nghiệp hay không, nơi các chai di chuyển liên tục với tốc độ cao và yêu cầu độ trễ xử lý thấp. Mục tiêu thứ ba, đề tài sẽ xây dựng và thực thi một quy trình huấn luyện, đánh giá mô hình một cách bài bản. Quy trình này bao gồm toàn bộ các bước: từ thu thập và xây dựng bộ dữ liệu hình ảnh thực tế, thực hiện gán nhãn cho từng đối tượng, tiến hành huấn luyện mô hình, cho đến đánh giá hiệu năng cuối cùng. Việc đánh giá sẽ dựa trên cả các chỉ số học máy khách quan như Precision, Recall, mAP lẫn các chỉ số hiệu năng kỹ thuật như tốc độ xử lý FPS - Frames Per Second, qua đó cung cấp một cái nhìn toàn diện để chứng minh tính khả thi của giải pháp trước khi xem xét triển khai thực tế.

CHƯƠNG 2. PHƯƠNG PHÁP PHÁT HIỆN LỖI CHAI NƯỚC TRÊN BĂNG CHUYỀN SỬ DỤNG MÔ HÌNH YOLOv11

2.1 Mô tả hệ thống

Hệ thống phát hiện lỗi chai nước trên băng chuyền sử dụng mô hình YOLO (You Only Look Once) là một giải pháp ứng dụng thị giác máy kết hợp trí tuệ nhân tạo nhằm tự động hóa quy trình kiểm tra chất lượng trong sản xuất.

Toàn bộ hệ thống bao gồm các thành phần phần cứng và phần mềm được tích hợp thành một chuỗi xử lý (pipeline) khép kín, hoạt động theo ba giai đoạn chính:

Giai đoạn 1: Thu nhận hình ảnh

- Một camera điện thoại được lắp đặt cố định tại một vị trí chiến lược phía trên băng chuyền thử nghiệm. Vị trí này được tính toán để đảm bảo camera có thể quan sát toàn bộ đối tượng chai nước khi nó đi qua vùng quan sát.
- Hệ thống chiếu sáng được bố trí để tối ưu hóa độ tương phản và làm nổi bật các khuyết điểm.

Giai đoạn 2: Phân tích và suy luận

- Hình ảnh thu được từ camera được đưa trực tiếp vào bộ xử lý trung tâm.
- Tại đây, mô hình YOLOv11 đã được huấn luyện sẽ thực hiện suy luận trên hình ảnh. Mô hình nhận một khung hình đầu vào và trả về một danh sách các dự đoán.
- Mỗi dự đoán bao gồm: Tọa độ khung chứa bounding box của lỗi, nhãn lỗi, và điểm tự tin.

Giai đoạn 3: Đánh dấu và xử lý

- Hệ thống phần mềm sẽ phân tích kết quả suy luận. Nếu một sản phẩm được phát hiện là lỗi và có điểm tự tin vượt qua một ngưỡng, hệ thống sẽ coi đó là một phát hiện hợp lệ.
- Ghi nhận thông tin: số lượng chai lỗi, loại lỗi.

2.2 Cơ sở lý thuyết và mô hình YOLO

2.2.1 Tổng quan về phát hiện đối tượng

Phát hiện đối tượng là một bài toán cốt lõi trong thị giác máy tính, yêu cầu mô hình không chỉ phân loại đối tượng trong ảnh là gì, mà còn phải định vị chính xác vị trí của nó bằng một khung chứa. Các hướng tiếp cận chính bao gồm:

- Hệ thống hai giai đoạn: R-CNN, Fast R-CNN, Faster R-CNN.
- Hệ thống một giai đoạn: SSD (Single Shot Detector), YOLO.

Bảng 2.1 Bảng so sánh hướng tiếp cận đối tượng

Tiêu chí	**Hai giai đoạn** **R-CNN, Faster R-CNN**	**Một giai đoạn** **YOLO, SSD, YOLOv11**
Cách hoạt động	Giai đoạn 1: Tìm vùng nghi ngờ. Giai đoạn 2: Phân loại từng vùng.	Dự đoán trực tiếp vị trí và phân loại chỉ trong 1 lần.
Đại diện	R-CNN, Fast R-CNN, Faster R-CNN	YOLOv11, YOLOv8, SSD
Độ chính xác	Rất cao (mAP >50%)	Cao (YOLOv11: ~51.5% mAP)
Tốc độ	Chậm (5-30 FPS)	Siêu nhanh (100-200 FPS)
Thời gian thực	Không phù hợp	Rất phù hợp với băng chuyền, camera công nghiệp
Tài nguyên	Cần GPU mạnh	Chạy được CPU, Jetson, Raspberry Pi
Phát hiện lỗi nhỏ	Tốt	Tốt

Dễ huấn luyện	Phức tạp	Dễ, end-to-end
Ứng dụng	Y tế, nghiên cứu	Công nghiệp, băng chuyền, giám sát

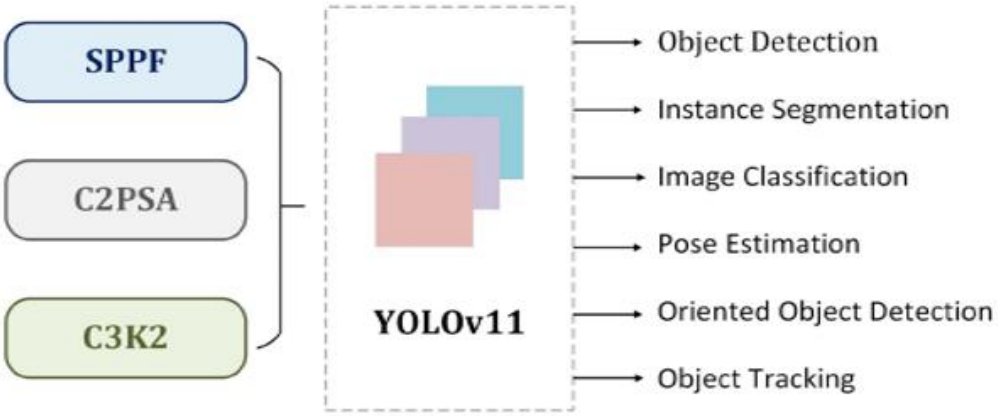
Đối với bài toán giám sát băng chuyền, tốc độ là yếu tố tiên quyết. Do đó, phương pháp tiếp cận một giai đoạn sử dụng YOLO là lựa chọn tối ưu.

2.2.2 Kiến trúc của mô hình YOLOv11

Khung YOLO đã cách mạng hóa việc phát hiện đối tượng bằng cách giới thiệu một kiến trúc mạng nơ-ron thống nhất xử lý đồnng thời giới hạn và các tác vụ phân loại đối tượng. Phương pháp tiếp cận tích hợp này đánh dấu sự thay đổi đáng kể so với các phương pháp phát hiện hai giai đoạn truyền thống, cung cấp khả năng đào tạo toàn diện thông qua thiết kế có thể phân biệt hoàn toàn tăng độ chính xác của học máy.

Về cốt lõi, kiến trúc YOLO bao gồm ba thành phần cơ bản. Đầu tiên, backbone đóng vai trò là trình trích xuất tính năng chính, sử dụng mạng nơ-ron tích chập để chuyển đổi dữ liệu hình ảnh thô thành bản đồ tính năng đa tỷ lệ. Thứ hai, thành phần neck hoạt động như một giai đoạn xử lý trung gian, sử dụng các lớp chuyên biệt để tổng hợp và tăng cường các biểu diễn tính năng trên các tỷ lệ khác nhau. Thứ ba, thành phần head hoạt động như cơ chế dự đoán, tạo ra các đầu ra cuối cùng để định vị và phân loại đối tượng dựa trên các bản đồ tính năng đã tinh chỉnh và xây dựng trước đó.

Dựa trên kiến trúc đã được thiết lập này, YOLO11 mở rộng và nâng cao nền tảng được đặt ra bởi YOLOv8, giới thiệu những cải tiến về kiến trúc và tối ưu hóa tham số để đạt được hiệu suất phát hiện vượt trội như minh họa trong hình dưới đây. Các phần sau đây trình bày chi tiết các sửa đổi kiến trúc chính được triển khai trong YOLO11:



Hình 2.1 Các mô-đun kiến trúc chính trong YOLO11

2.2.2.1 BackBone

BackBone là một thành phần quan trọng của kiến trúc YOLO, chịu trách nhiệm trích xuất các đặc điểm từ hình ảnh đầu vào ở nhiều tỷ lệ. Quá trình này bao gồm việc xếp chồng các lớp tích chập và các khối chuyên biệt để tạo ra các bản đồ đặc trưng ở nhiều độ phân giải khác nhau.

2.2.2.2 Các lớp tích chập ban đầu

YOLOv11 vẫn sử dụng các lớp tích chập ban đầu nhằm giảm kích thước không gian của ảnh trong khi tăng số lượng kênh đặc trưng. Các lớp này tạo nên nền tảng cho quá trình trích xuất đặc trưng sâu hơn.

Một cải tiến quan trọng trong YOLOv11 là việc giới thiệu khối C3k2, thay thế cho khối C2f của YOLOv8 và các phiên bản trước: C3k2 là phiên bản rút gọn và hiệu quả hơn của khối CSP (Cross Stage Partial). Khối này sử dụng hai phép tích chập nhỏ thay vì một phép tích chập lớn, giúp giảm chi phí tính toán và tăng tốc độ xử lý. “K2” biểu thị việc sử dụng hạt nhân tích chập nhỏ (kernel size nhỏ), giúp tăng hiệu năng mà vẫn duy trì độ chính xác.

2.2.2.3 Khối SPPF và C2PSA

YOLOv11 tiếp tục sử dụng SPPF (Spatial Pyramid Pooling – Fast) nhằm tăng khả năng trích xuất đặc trưng theo không gian với chi phí tính toán thấp.

Bên cạnh đó, YOLOv11 giới thiệu một khối mới: C2PSA (Cross Stage Partial with Spatial Attention). Khối này bổ sung cơ chế chú ý không gian (spatial attention), giúp mô hình tập trung vào các vùng quan trọng trong ảnh. Điều này giúp cải thiện hiệu suất phát hiện các đối tượng nhỏ, đối tượng chồng lấp hoặc nằm trong môi trường phức tạp.

2.2.2.4 Neck

Neck là bộ phận tổng hợp đặc trưng từ nhiều tầng của backbone. Nhiệm vụ chính của nó là:

+ Gộp thông tin đa tỷ lệ (multi-scale)
+ Lấy mẫu lên – lấy mẫu xuống
+ Nối (concatenate) các bản đồ đặc trưng

Trong YOLOv11, neck có một thay đổi quan trọng, khối C3k2 được sử dụng thay cho C2f, giúp tăng tốc độ tổng hợp đặc trưng, đồng thời giảm số lượng tham số. Cải tiến này giúp neck xử lý nhanh hơn và hiệu quả hơn mà không ảnh hưởng đến chất lượng trích xuất đặc trưng.

2.2.2.5 Head

Head là phần chịu trách nhiệm tạo ra các dự đoán cuối cùng, bao gồm: Tọa độ hộp giới hạn (bounding box), điểm đối tượng (objectness score), điểm lớp (class score)

Trong YOLOv11, head sử dụng nhiều khối C3k2 để tinh chỉnh đặc trưng đầu vào, đảm bảo mô hình có khả năng học tốt các đặc trưng sâu và phức tạp trước khi đưa ra dự đoán. Các khối này được bố trí ở nhiều nhánh khác nhau, tương ứng với các mức độ trích xuất đặc trưng đa tỷ lệ.

Khối C3k2 trong phần head có hai cấu hình hoạt động:

Giá trị c3k	**Cấu trúc tương ứng**	**Chức năng**
False	Cấu trúc tương tự C2f	Trích xuất đặc trưng nhanh, ít tham số
True	Cấu trúc dạng C3 sâu hơn	Học đặc trưng phức tạp, tăng độ chính xác

2.2.2.6 Lớp tích chập cuối và lớp phát hiện

Mỗi nhánh phát hiện kết thúc bằng một tập hợp các lớp Conv2D nhằm giảm số chiều đặc trưng xuống số lượng cần thiết cho đầu ra. Lớp phát hiện cuối cùng hợp nhất các thông tin và sinh ra: tọa độ hộp giới hạn, điểm objectness và điểm phân loại. Đây là các giá trị đầu ra chính phục vụ nhiệm vụ phát hiện đối tượng của YOLOv11.

2.2.3 Phiên bản YOLO sử dụng trong đề tài

Trong phạm vi đề tài, nhóm sử dụng YOLOv11s nhờ khả năng cân bằng tốt giữa tốc độ suy luận và độ chính xác, phù hợp với yêu cầu hệ thống và cấu hình phần cứng hiện có. Bộ mô hình YOLOv11 nói chung mang lại hiệu suất vượt trội trong các tác vụ nhận dạng đối tượng, phân đoạn và ước lượng tư thế. Các ưu điểm nổi bật bao gồm:

- Mô hình đã được huấn luyện trước: Giúp rút ngắn thời gian phát triển và đạt hiệu quả nhanh chóng.
- Độ chính xác cao: mAP cải thiện đáng kể so với các thế hệ trước.
- Tốc độ xử lý nhanh: Tối ưu hóa cho nhận dạng thời gian thực.
- Tính linh hoạt cao: Hỗ trợ xuất sang nhiều định dạng như ONNX, TensorRT, CoreML,…

Bảng 2.2 Bảng so sánh các phiên bản yolo11

Tiêu chí	**YOLOv11n (nano)**	**YOLOv11s (small)**	**YOLOv11m (medium)**
Số tham số	Thấp	Trung bình	Cao
Độ chính xác (mAP)	Thấp	Cao hơn n	Cao
Tốc độ suy luận	Nhanh nhất	Nhanh, ổn định	Chậm hơn s

Yêu cầu phần cứng	Rất thấp	Trung bình	Cao
Khả năng phù hợp với real-time	Tốt nhưng độ chính xác thấp	Cân bằng tốt giữa tốc độ và mAP	Phụ thuộc vào GPU, có thể không ổn định real-time
Ưu điểm	Nhẹ, dễ triển khai trên thiết bị yếu	Cân bằng giữa tốc độ và độ chính xác	Độ chính xác cao
Mức độ phù hợp với đề tài	Không phù hợp	Phù hợp nhất	Không tối ưu

Vì vậy, nhóm quyết định sử dụng YOLOv11s vì:

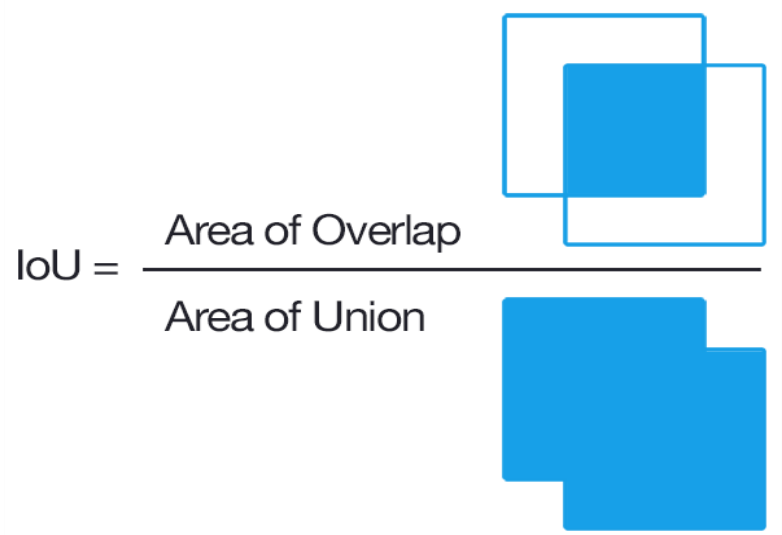
- Độ chính xác cao hơn YOLOv11n, đáp ứng tốt yêu cầu nhận dạng đối tượng trong đề tài.
- Tốc độ suy luận nhanh và ổn định, phù hợp cho xử lý thời gian thực trên thiết bị phần cứng của nhóm.
- Nhẹ hơn và yêu cầu tài nguyên thấp hơn YOLOv11m, giúp triển khai dễ dàng mà vẫn đảm bảo hiệu suất.

YOLOv11s đạt điểm cân bằng tối ưu giữa độ chính xác, tốc độ và khả năng triển khai, do đó phù hợp nhất cho hệ thống của đề tài.

2.3 Các chỉ số đánh giá hiệu năng

2.3.1 Chỉ số IoU

IoU (Intersection over Union - Giao trên Hợp) là chỉ số dùng để đo mức độ trùng khớp giữa hộp dự đoán của mô hình và hộp thực tế của đối tượng. IoU được tính bằng tỉ lệ giữa diện tích phần hai hộp chồng lên nhau và diện tích hợp lại của cả hai hộp. Chỉ số này phản ánh trực tiếp độ chính xác của quá trình xác định vị trí trong bài toán phát hiện đối tượng.



Hình 2.2 Công thức IoU

Ý nghĩa của các giá trị IoU:

- IoU = 1.0: Hai hộp trùng khớp hoàn toàn với nhau.
- IoU = 0.0: Hai hộp không chồng lấp, mô hình dự đoán sai hoàn toàn.
- IoU ≥ 0.5: Thường được xem là mức đủ tốt để chấp nhận dự đoán đúng.

Khi IoU thấp, điều đó cho thấy mô hình đang gặp khó khăn trong việc xác định chính xác vị trí của đối tượng. Trong trường hợp này, có thể cần xem xét lại các phương pháp dự đoán bounding box hoặc cải tiến dữ liệu và kiến trúc mô hình.

2.3.2 Chỉ số Precision và Recall

Độ chính xác (Precision) và Độ phủ (Recall) là hai chỉ số quan trọng dùng để đánh giá hiệu quả của mô hình trong bài toán phát hiện đối tượng.

Precision đo tỷ lệ các trường hợp mô hình phát hiện đúng trong số tất cả các mục mà mô hình cho là có đối tượng. Nói cách khác, chỉ số này thể hiện khả năng mô hình tránh nhận nhầm những vùng không phải đối tượng.

Recall đo tỷ lệ các trường hợp mô hình phát hiện đúng trong số toàn bộ các đối tượng thực sự xuất hiện trong dữ liệu. Chỉ số này phản ánh mức độ đầy đủ trong việc tìm ra tất cả các đối tượng cần phát hiện.

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$\frac{2}{F1} = \frac{1}{\text{Precision}} + \frac{1}{\text{Recall}}$$

Hình 2.3 Công thức Precision và Recall

Trong đó: TP – True Positive, TN – True Negative, FP – Flase Positve, FN – False Negative.

Độ Chính Xác Thấp: Mô hình có thể đang phát hiện quá nhiều đối tượng không tồn tại. Điều chỉnh ngưỡng tin cậy có thể giảm thiểu điều này.

Độ Phủ Thấp: Mô hình có thể bỏ sót các đối tượng thực. Cải thiện trích xuất đặc trưng hoặc sử dụng nhiều dữ liệu hơn có thể giúp ích.

2.3.3 Chỉ số F1 Score

F1 Score là trung bình điều hòa của precision (độ chính xác) và recall (độ phủ), cung cấp một đánh giá cân bằng về hiệu suất của mô hình trong khi xem xét cả false positives (dương tính giả) và false negatives (âm tính giả).

$$F1 = 2\frac{Precision * Recall}{Precision + Recall}$$

Hình 2.4 Công thức F1 Score

F1 Score: Hữu ích khi cần sự cân bằng giữa precision (độ chính xác) và recall (độ phủ).

2.3.4 Chỉ số mAP

Mean Average Precision (mAP): mAP mở rộng khái niệm AP bằng cách tính giá trị AP trung bình trên nhiều lớp đối tượng. Điều này hữu ích trong các tình huống phát hiện đối tượng đa lớp để cung cấp đánh giá toàn diện về hiệu suất của mô hình.

Có hai loại mAP:

+ mAP50: Độ chính xác trung bình được tính toán tại ngưỡng giao nhau trên hợp (IoU) là 0.50. Đây là thước đo độ chính xác của mô hình chỉ xét đến các phát hiện "dễ".
+ mAP50-95: Giá trị trung bình của độ chính xác trung bình được tính toán ở các ngưỡng IoU khác nhau, từ 0.50 đến 0.95. Nó cung cấp một cái nhìn toàn diện về hiệu suất của mô hình trên các mức độ khó phát hiện khác nhau.

2.3.5 Chỉ số đánh giá hiệu suất công nghiệp

FPS (Frames Per Second): Tốc độ xử lý khung hình. Chỉ số này phải cao hơn tốc độ của băng chuyền để đảm bảo không bỏ sót sản phẩm.

Latency (Độ trễ): Thời gian (tính bằng mili giây - ms) từ lúc camera chụp ảnh đến khi mô hình đưa ra kết quả. Đây là chỉ số quan trọng hơn FPS trong các hệ thống điều khiển thời gian thực.

CHƯƠNG 3. THỰC NGHIỆM

3.1 Môi trường thực nghiệm

3.1.1 Cấu hình phần cứng

Máy thử nghiệm sử dụng GPU NVIDIA GeForce RTX 5060 (8GB VRAM). Băng chuyền thử nghiệm được đặt tại phòng lab A4.609 - Đại học Phenikaa, có các thông số:

Bảng 3.1 Bảng giá trị thông số đặt băng chuyền

Thông số	Giá trị
Chiều dài băng chuyền	115cm
Thời gian chạy hết băng	8.6s
Vận tốc trung bình	13.37 cm/s
Màu nền ảnh	Trắng
Màu băng chuyền	Xanh lá
Góc chiếu ánh sáng đèn 1 so với băng chuyền	90 độ
Góc chiếu ánh sáng đèn 2 so với băng chuyền	60 độ

Ánh sáng được bố trí đều để giảm nhiễu và đảm bảo độ rõ bề mặt sản phẩm. Đèn 1 đặt góc 90° phía trước giúp giảm bóng đổ, làm rõ các nếp hoặc biến dạng nhỏ. Đèn 2 đặt góc 60° phía sau tạo viền sáng nhẹ, giúp phát hiện thủng vỏ hoặc mép nắp cong qua hiệu ứng rò sáng.

3.1.2 Cấu hình phần mềm

Dự án được phát triển và huấn luyện trong môi trường cục bộ, sử dụng Visual Studio Code làm công cụ lập trình chính. Quá trình thực thi diễn ra trên hệ

điều hành Windows 11 với ngôn ngữ Python, đáp ứng đầy đủ yêu cầu linh hoạt khi xây dựng và kiểm thử mô hình.

Về nền tảng học sâu, PyTorch được sử dụng làm framework cốt lõi, kết hợp với CUDA của NVIDIA để tăng tốc tính toán song song trên GPU. Thư viện Ultralytics giữ vai trò trung tâm khi cung cấp triển khai của mô hình YOLOv11 và hỗ trợ các thao tác tăng cường dữ liệu trong quá trình huấn luyện. Song song đó, OpenCV được dùng cho các tác vụ xử lý ảnh như đọc, ghi và biến đổi hình ảnh, trong khi NumPy đóng vai trò thư viện tính toán khoa học hỗ trợ các phép toán nền tảng.

Roboflow được sử dụng trong giai đoạn xử lý dữ liệu, bao gồm quản lý tập dữ liệu, gán nhãn hình ảnh, tăng cường dữ liệu và tiền xử lý trước khi đưa vào môi trường huấn luyện. Các công cụ và thư viện trên kết hợp tạo nên một môi trường phát triển đầy đủ và ổn định cho toàn bộ quy trình của dự án.

3.2 Dữ liệu thực nghiệm

3.2.1 Nguồn dữ liệu và cách thu thập

3.2.1.1 Thông tin chung về bộ dữ liệu

Đối tượng được sử dụng trong nghiên cứu là chai nước 500 ml trà BUPNON TEA 365 do Masan Consumer sản xuất.

Dữ liệu phục vụ huấn luyện và đánh giá mô hình được thu thập hoàn toàn (100%) tại hệ thống băng chuyền thử nghiệm của phòng lab A4-606. Hình ảnh được ghi lại trong quá trình sản phẩm di chuyển trên băng chuyền nhằm mô phỏng điều kiện vận hành thực tế.

Mục tiêu của bài toán là ứng dụng Object Detection để phát hiện và khoanh vùng các lỗi xuất hiện trên sản phẩm khi nó đang chuyển động trên băng chuyền.

Hệ thống phân loại lỗi được xây dựng thủ công, gồm ba lớp: Không lỗi, móp méo, rách vỏ và nắp vênh.

Sau khi thu thập và thực hiện bước lọc dữ liệu nhằm loại bỏ các ảnh không đạt yêu cầu, tổng số ảnh còn lại là 418 ảnh, được sử dụng làm tập dữ liệu cho mô hình.

3.2.1.2 Quy trình thu thập

Quy trình thu thập dữ liệu được thực hiện bằng cách gắn một camera cố định ở vị trí phía trên băng chuyền để đảm bảo góc nhìn bao quát và ổn định. Camera ghi lại toàn bộ quá trình các chai bao gồm cả chai bình thường và chai có lỗi di chuyển trên băng chuyền. Video được thu ở tốc độ 30 FPS nhằm bảo đảm số lượng khung hình đủ lớn cho việc trích xuất dữ liệu. Sau khi ghi hình, toàn bộ video được tách thành các khung hình riêng lẻ, tạo thành nguồn dữ liệu đầu vào cho các bước tiền xử lý và gán nhãn.

3.2.1.3 Khó khăn và biện pháp xử lý

Trong quá trình thu thập, nhóm gặp phải một số khó khăn đáng chú ý. Trước hết là dữ liệu trùng lặp, do vận tốc băng chuyền khá chậm (13.37 cm/s) trong khi camera ghi hình ở 30 FPS, nhiều khung hình liên tiếp gần như không có sự thay đổi, dẫn đến lượng lớn dữ liệu lặp lại và không mang thêm thông tin hữu ích. Tiếp theo là dữ liệu nhiễu, xuất hiện nhiều khung hình trống không có sản phẩm, khiến việc lưu trữ và xử lý trở nên kém hiệu quả. Ngoài ra, ánh sáng không ổn định cũng gây ảnh hưởng. Do băng chuyền và vỏ chai có độ phản xạ cao, hiện tượng bóng gắt và phản chiếu xuất hiện trên nền trắng và bề mặt chai, tạo nhiễu cho quá trình trích xuất hình ảnh.

Để khắc phục các vấn đề trên, nhóm đã áp dụng một quy trình lọc tự động thay vì sử dụng toàn bộ dữ liệu gốc.

Đối với dữ liệu trùng lặp, nhóm không giữ lại toàn bộ 30 khung hình trong mỗi giây. Thay vào đó, một bộ lọc thời gian được áp dụng, chỉ trích xuất 1 khung hình sau mỗi 15 khung hình, tương đương 0,5 giây/khung hình. Cách làm này giúp giảm đáng kể sự trùng lặp, đảm bảo các khung hình được chọn có sự thay đổi rõ rệt và mang thông tin cần thiết cho mô hình.

Đối với khung hình trống, nhóm sử dụng phương pháp phân tích màu sắc dựa trên đặc điểm nổi bật của sản phẩm là vỏ chai có màu đỏ đặc trưng, khác biệt hoàn toàn so với nền. Mỗi khung hình được kiểm tra số lượng điểm ảnh màu đỏ. Chỉ những khung hình có tổng số điểm ảnh vượt ngưỡng 30.000 mới được giữ lại. Điều này cho thấy khung hình có sự xuất hiện của sản phẩm. Các khung hình

không đạt ngưỡng sẽ bị loại bỏ ngay lập tức để tránh nhiễu và giảm tải cho quá trình xử lý tiếp theo.

3.2.2 Phương pháp tiền xử lý

Từ 418 ảnh gốc thu được, toàn bộ dữ liệu được tải lên nền tảng Roboflow để tiến hành xử lý và chuẩn hóa trước khi huấn luyện. Dữ liệu ban đầu được phân chia theo tỉ lệ 80% cho tập huấn luyện và 20% cho tập kiểm định, tương ứng với 335 ảnh thuộc tập huấn luyện và 83 ảnh thuộc tập kiểm định.

Do số lượng dữ liệu gốc còn hạn chế, nhóm đã áp dụng chiến lược tăng cường dữ liệu hệ số 3x trên Roboflow dành riêng cho tập huấn luyện. Quá trình tăng cường này đã mở rộng số lượng mẫu, nâng tổng số ảnh trong tập huấn luyện lên 1005 ảnh. Các phép tăng cường được lựa chọn đều là các kỹ thuật an toàn, không làm thay đổi hình học của đối tượng nhằm giữ nguyên cấu trúc và vị trí của sản phẩm trong ảnh. Các phép biến đổi bao gồm điều chỉnh độ sáng (15%), thay đổi độ phơi sáng (±15%), làm mờ ở mức 2px, bổ sung nhiễu lên đến 10% và thay đổi độ bão hòa (±25%).

Sau khi hoàn tất quá trình xử lý và tăng cường, bộ dữ liệu cuối cùng bao gồm 1005 ảnh huấn luyện và 83 ảnh kiểm định. Bộ dữ liệu này được tải về và sử dụng làm nguồn dữ liệu chính cho quá trình huấn luyện mô hình.

3.3 Cài đặt và huấn luyện mô hình

3.3.1 Mô tả mô hình sử dụng

Trong nghiên cứu này, mô hình được lựa chọn là YOLOv11, cụ thể là phiên bản yolo11s.pt (Small). Đây là biến thể cho tỷ lệ giữa tốc độ xử lý và độ chính xác ở mức tối ưu, phù hợp với yêu cầu phát hiện lỗi theo thời gian thực trong môi trường băng chuyền công nghiệp.

Kiến trúc của YOLOv11s được xây dựng trên nền tảng PyTorch và tích hợp các thành phần cải tiến như C3k2 và C2PSA (Cross-Stage Partial with Parallel Spatial Attention). Các khối này giúp mô hình tăng cường khả năng trích xuất đặc trưng, cải thiện hiệu suất trong việc phát hiện các vật thể có hình dạng thay đổi nhỏ hoặc xuất hiện trong điều kiện ánh sáng phức tạp.

Để phù hợp với bài toán của dự án, đầu ra của mô hình được tùy chỉnh lại nhằm nhận diện ba lớp lỗi cụ thể:

+ Class 0 – Không lỗi
+ Class 1 – Lỗi rách vỏ và nắp vênh
+ Class 2 – Lỗi méo chai

Cấu hình này đảm bảo mô hình tập trung vào đúng các trường hợp lỗi cần phát hiện, đồng thời giảm nhiễu và tránh sự dư thừa trong quá trình huấn luyện.

3.3.2 Các siêu tham số huấn luyện

Bảng 3.2 Bảng giá trị tham số huấn luyện

Tham số	Giá trị	Lý do chọn
Mô hình	yolo11s.pt	Phiên bản "Small" được chọn để cân bằng giữa độ chính xác và tốc độ, đồng thời giảm nguy cơ overfitting trên bộ dữ liệu huấn luyện nhỏ (1.011 ảnh), phù hợp với triển khai thời gian thực.
Kích thước ảnh	640x640	Đây là kích thước đầu vào tiêu chuẩn cho các mô hình YOLO, cung cấp sự cân bằng tốt giữa khả năng phát hiện chi tiết lỗi và yêu cầu tài nguyên VRAM.
Số epochs	100	Cung cấp đủ thời gian để mô hình hội tụ.
Batch size	16	Giá trị 16 là lớn nhất có thể nhét vừa 8GB VRAM của RTX 5060, giúp tăng tốc độ huấn luyện.
Trình tối ưu	AdamW (tự động)	Được optimizer tự động lựa chọn. AdamW là một optimizer mạnh mẽ, cải tiến từ Adam, có khả năng quản lý weight decay hiệu quả hơn, thường dẫn đến hội tụ tốt hơn.
Seed	0 (mặc định)	Giá trị này được thư viện ultralytics tự động đặt. Việc giữ cố định seed đảm bảo rằng kết quả huấn luyện có thể được tái lập chính xác trong những lần chạy sau.

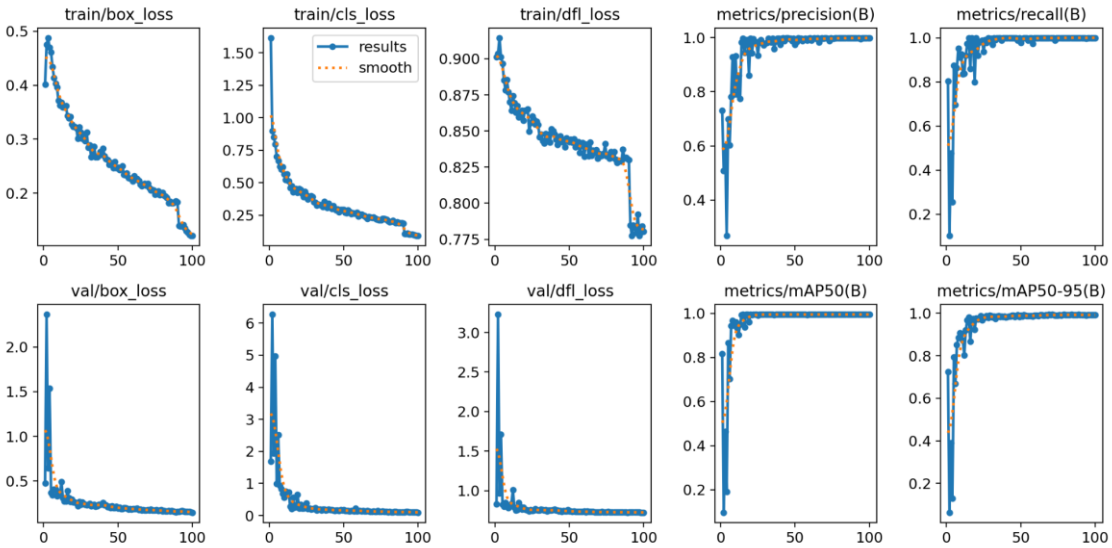
Tăng cường	mosaic=1.0	Ghép 4 ảnh làm 1, buộc mô hình học cách phát hiện lỗi ở các kích thước nhỏ và bị che khuất.
	copy_paste=0.5	Tự động dán các mẫu "lỗi" lên ảnh nền "không lỗi", giải quyết vấn đề mất cân bằng dữ liệu và tăng số lượng mẫu lỗi hiếm.
Cấu hình	amp=False	Tắt tính năng Mixed Precision để tránh gây ra lỗi tràn số và lỗi mất ổn định trong mô hình. Từ đó đảm bảo quá trình huấn luyện ổn định.
	workers=0	Tắt chế độ đa luồng trên windows để tránh gây ra lỗi treo hoặc lỗi sập liên quan đến bộ tải dữ liệu.

Quá trình huấn luyện được cấu hình với các siêu tham số được lựa chọn và tối ưu hóa dựa trên môi trường và đặc tính của bộ dữ liệu.

3.3.3 Quy trình huấn luyện

Khởi chạy huấn luyện: Quá trình huấn luyện được khởi chạy trong môi trường Python bằng cách sử dụng hàm train() từ thư viện ultralytics. Đối tượng mô hình YOLO("yolo11s.pt") được tải, sau đó hàm train() được gọi với đầy đủ các siêu tham số đã được thiết lập.

Giám sát quá trình huấn luyện: Quá trình huấn luyện được giám sát qua log tại mỗi epoch. Log này cung cấp thông tin về các chỉ số mất mát và độ chính xác (mAP) trên tập kiểm định. Các biểu đồ trực quan sau được tự động tạo ra từ log này để dễ dàng theo dõi.



Hình 3.1 Biểu đồ các chỉ số trong quá trình huấn luyện 100 epoch.

Phân tích biểu đồ huấn luyện:

+ Các biểu đồ train/box_loss, train/cls_loss (hàng trên): Giảm đều và nhanh chóng, cho thấy mô hình học tốt trên tập huấn luyện.
+ Các biểu đồ val/box_loss, val/cls_loss (hàng dưới): Cũng giảm song song với tập train và giữ ở mức thấp. Quan trọng nhất là không có dấu hiệu tăng ngược trở lại, cho thấy mô hình không bị Overfitting dù chạy 100 epoch.
+ Các biểu đồ metrics/precision, metrics/recall (hàng trên, bên phải): Tăng vọt lên gần 1.0 và duy trì ổn định, cho thấy mô hình nhanh chóng đạt được độ chính xác và độ phủ cao.
+ Các biểu đồ metrics/mAP50(B) và metrics/mAP50-95(B) (hàng dưới, bên phải): Đây là các chỉ số quan trọng nhất. Cả hai đều tăng nhanh và đạt mức gần 1.0 chỉ sau khoảng 40-50 epoch, cho thấy hiệu suất của mô hình đã đạt đỉnh và rất ổn định.

3.4 Kết quả

3.4.1 Kết quả định lượng

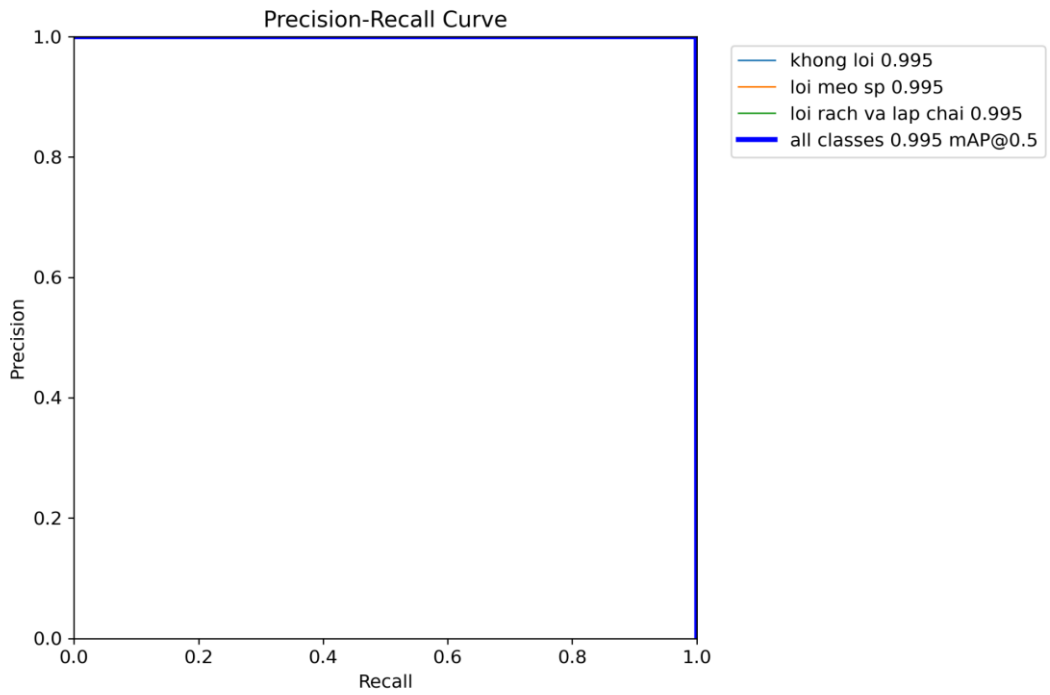
Quá trình huấn luyện hoàn thành sau 2.937 giờ (100 epochs). Mô hình đạt được độ chính xác cao nhất đã được lưu lại tự động trong file best.pt. Dưới đây là kết quả thực tế được ghi nhận sau huấn luyện mô hình:

```
                 Class     Images  Instances      Box(P          R      mAP50  mAP50-95): 1
                   all         83         83      0.998          1      0.995      0.991
             khong loi         29         29      0.997          1      0.995      0.995
             loi meo sp         28         28      0.998          1      0.995      0.995
  loi rach va lap chai         26         26      0.998          1      0.995      0.983
Speed: 0.1ms preprocess, 6.3ms inference, 0.0ms loss, 0.8ms postprocess per image
```

Hình 3.2 Kết quả thực tế sau huấn luyện mô hình.

Kết quả tóm tắt này cho thấy mô hình YOLOv11s đã đạt được hiệu suất vượt trội và hoàn toàn đáp ứng yêu cầu thời gian thực của hệ thống. Độ chính xác tổng thể mAP0.5-0.95 đạt 99.1%, khẳng định khả năng định vị chính xác đối tượng lỗi với độ tin cậy rất cao. Cụ thể, mô hình đạt Precision 99.8% và Recall tuyệt đối 100%, chỉ ra rằng mô hình không bỏ sót bất kỳ lỗi nào trong 83 ảnh kiểm tra, đây là yêu cầu tối quan trọng trong sản xuất. Phân tích từng lớp cho thấy lớp khó nhất “loi rach va lap chai” cũng đạt mAP rất cao 98.3%, chứng minh tính hiệu quả của chiến lược tăng cường dữ liệu (copy_paste). Về tốc độ, mô hình xử lý một khung hình chỉ mất 6.3ms, đạt tốc độ 158 FPS, đáp ứng yêu cầu phát hiện lỗi theo thời gian thực trên băng chuyền.

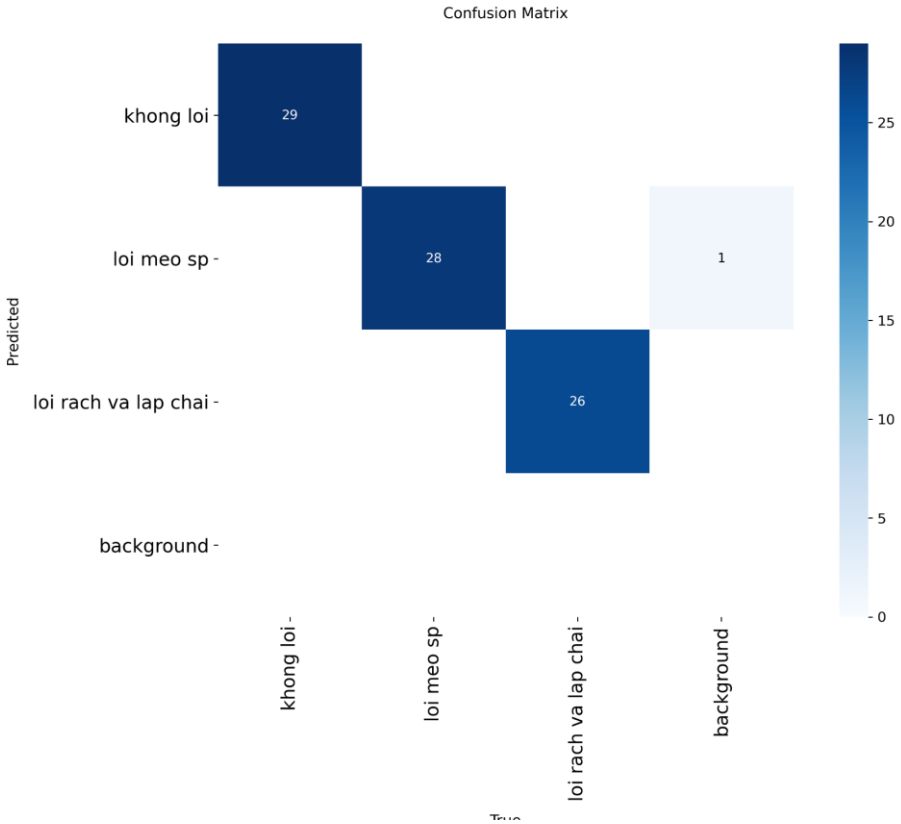
Phân tích đường cong Precision-Recall: Đường cong PR thể hiện sự cân bằng giữa Precision và Recall. Một mô hình lý tưởng sẽ có đường cong tiệm cận góc trên bên phải (Precision=1.0, Recall=1.0).



Hình 3.3 Đường cong Precision-Recall cho tất cả các lớp.

Như trong Hình 3.3, đường cong PR của mô hình gần như là một hình vuông hoàn hảo ở góc trên bên phải. Mô hình đạt mAP 0.5 - 0.995 (99.5%) cho tất cả các lớp. Đây là một kết quả gần như hoàn hảo, cho thấy mô hình có khả năng phát hiện chính xác (Precision cao) mà không bỏ sót (Recall cao) ở ngưỡng IoU=0.5.

Phân tích ma trận nhầm lẫn: Ma trận nhầm lẫn giúp đánh giá chi tiết các trường hợp mô hình dự đoán đúng và các trường hợp dự đoán sai.



Hình 3.4 Ma trận nhầm lẫn trên tập validation

Qua quan sát đường chéo chính ta thấy rõ, về sản phẩm khong loi có 29 trường hợp được dự đoán chính xác. Sản phẩm méo có 28 trường hợp được dự đoán chính xác. Và sản phẩm có lỗi rách và lắp chai hở có 26 trường hợp được dự đoán chính xác. Ngoài đường chéo chính có 1 trường hợp mô hình dự đoán là loi meo sp trong khi nhãn thực tế là background. Cho thấy đây là một kết quả cực kỳ tốt. Mô hình không hề bỏ sót bất kỳ lỗi nào, và cũng không nhầm lẫn giữa các loại lỗi với nhau.

3.4.2 Kết quả trực quan

Mô hình cho thấy khả năng khoanh vùng chính xác các chai nước trà BUPNON TEA 365, minh chứng qua các mẫu sau:

Hình 3.5 Kết quả trực quan khi mô hình chạy với sản phẩm thực nghiệm

3.5 Nhận xét và thảo luận

3.5.1 Phân tích kết quả đạt được

Hiệu suất của mô hình YOLOv11s được chứng minh bằng các đường cong sau.

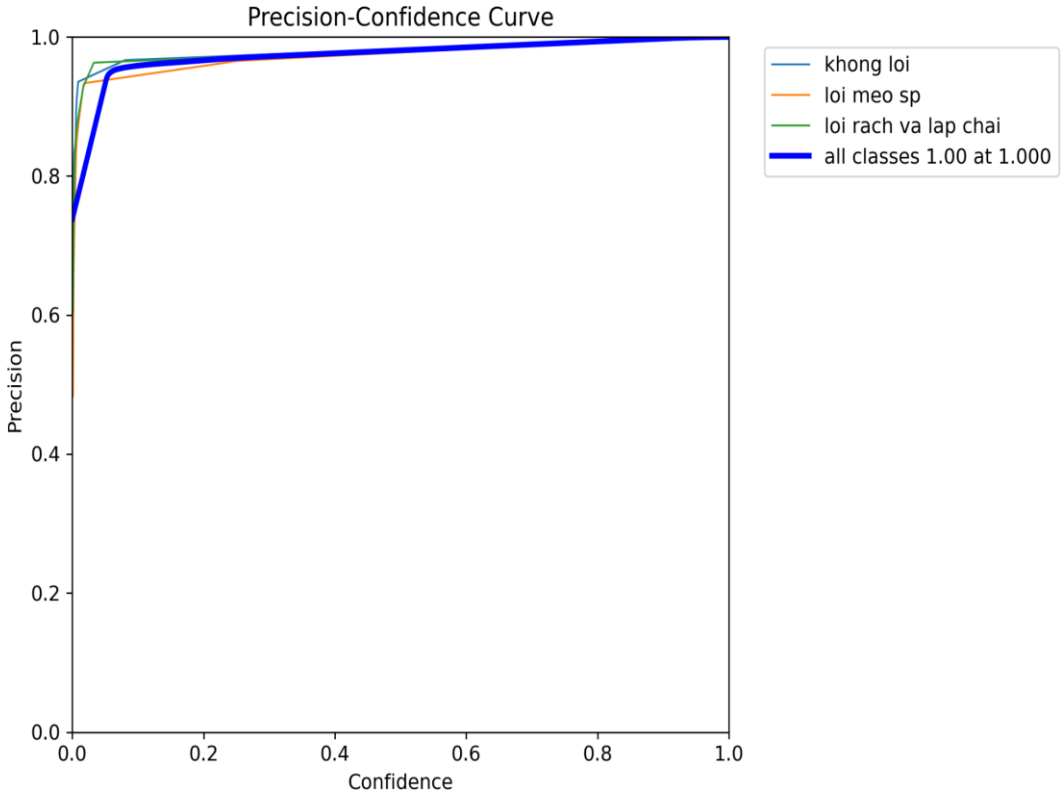
Về hiệu năng tổng thể, mô hình đạt mAP 0.5 là 99.5% và mAP 0.5-0.95 là 99.1%, đây là kết quả gần như hoàn hảo. Phân tích ma trận nhiễu cho thấy mô hình chỉ mắc 1 lỗi dự đoán sai trên toàn bộ 83 ảnh kiểm định và không bỏ sót bất kỳ lỗi có thật nào.

Về tốc độ suy luận đạt 6.3ms/khung hình (tương đương 158 FPS) trên GPU RTX 5060, hoàn toàn đáp ứng vượt trội yêu cầu thời gian thực của băng chuyền (chỉ cần khoảng 30 FPS).

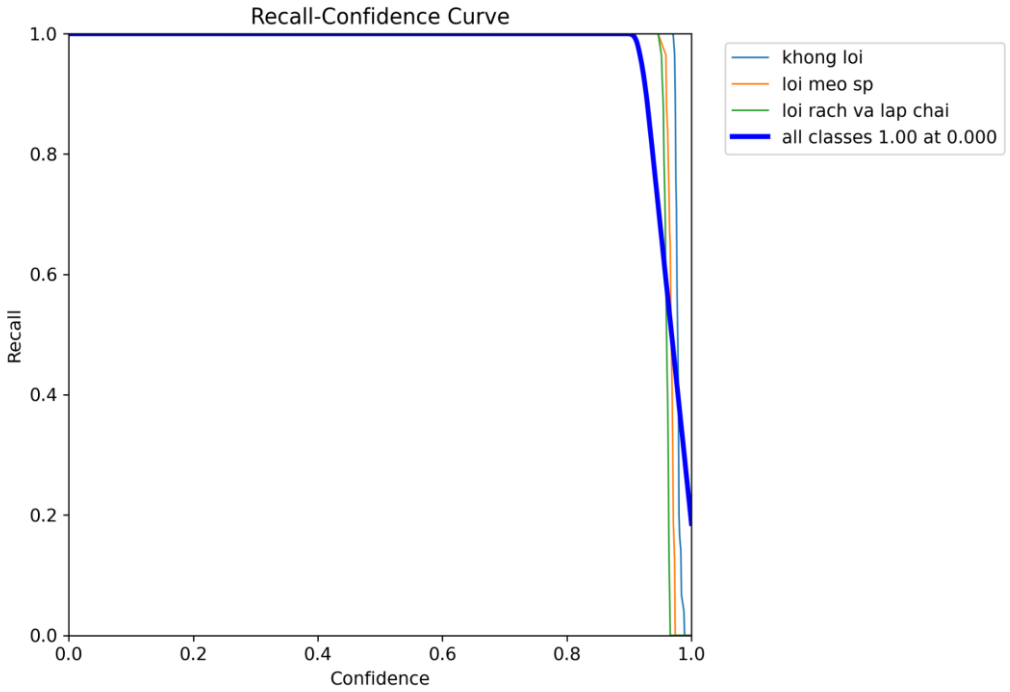
Phân tích chi tiết (P, R, F1 Curves):

+ Precision-Confidence : Đường Precision (độ chính xác) duy trì ở mức gần 1.0 ở hầu hết các ngưỡng tin cậy. Điều này khẳng định rằng hầu hết mọi hộp lỗi mô hình dự đoán đều là chính xác.
+ Recall-Confidence : Đường Recall (Độ phủ) giữ ở mức 1.000 (tìm thấy mọi lỗi) cho đến ngưỡng tin cậy ~0.9. Điều này chứng tỏ mô hình có khả năng tìm thấy tất cả các lỗi có thật với độ tin cậy rất cao.

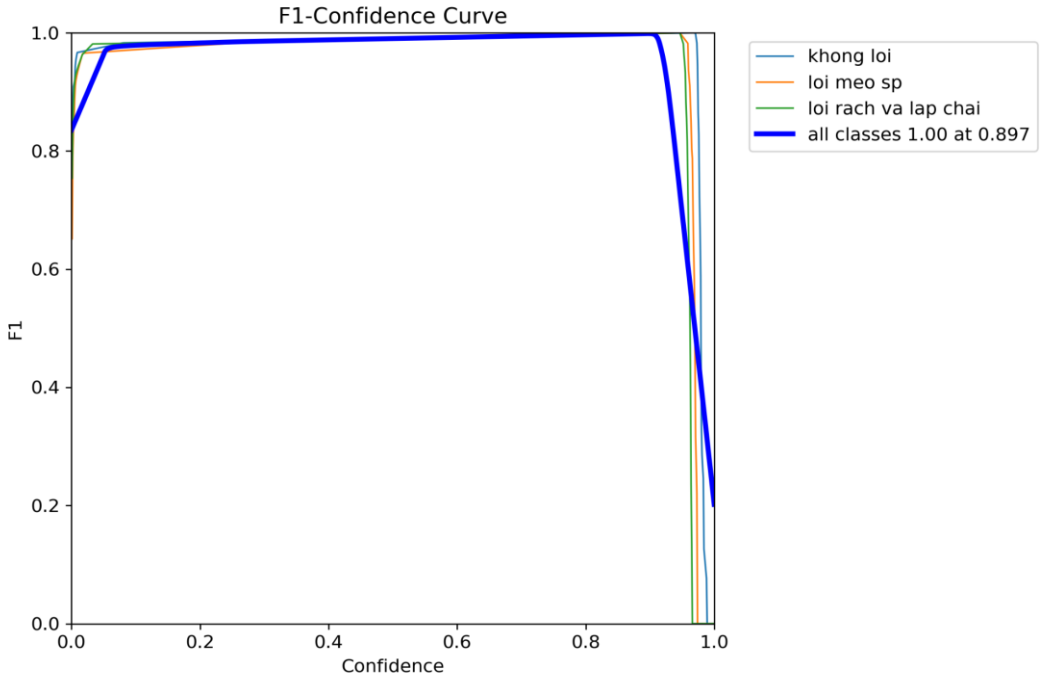
+ F1-Confidence : Biểu đồ này cho thấy điểm F1 (chỉ số cân bằng giữa P và R) đạt giá trị tối ưu. Đây chính là ngưỡng tin cậy tốt nhất để triển khai mô hình trên thực tế.



Hình 3.6 Biểu đồ Precision-Confidence



Hình 3.7 Biểu đồ Recall-Confidence.



Hình 3.8 Biểu đồ F1-Confidence

3.5.2 Hạn chế của dự án

Mặc dù kết quả mAP rất cao, mô hình này có một hạn chế nghiêm trọng cần phải được làm rõ.

Thứ nhất, Overfitting với môi trường, toàn bộ 418 ảnh gốc và 85 ảnh kiểm định đều được thu thập trong một môi trường duy nhất (phòng lab A4-606), với các đặc điểm: nền tường trắng và bề mặt băng chuyền xanh lá. Kết quả 99.1% mAP chỉ chứng minh rằng mô hình hoạt động hoàn hảo trong chính môi trường đó. Tuy nhiên, mô hình có khả năng đã học "thuộc lòng" bối cảnh này. Nếu mô hình được triển khai trên một dây chuyền thực tế, nó có khả năng thất bại vì bối cảnh đã thay đổi.

Hơn nữa tập kiểm định cũng nhỏ, độ chính xác được đo trên 85 ảnh. Con số này có thể chưa đủ lớn để đại diện cho mọi trường hợp lỗi.

3.5.3 Đề xuất hướng cải thiện

Dựa trên hạn chế lớn nhất đã nêu, hướng cải thiện rõ ràng nhất là phá vỡ sự phụ thuộc vào bối cảnh.

Giai đoạn thu thập dữ liệu cần được mở rộng. Cần thu thập thêm ảnh của chai trên nhiều bối cảnh khác nhau (ví dụ: băng chuyền kim loại, nền xi măng, bàn gỗ) và dưới các điều kiện ánh sáng khác nhau.

Sử dụng ảnh nền, đây là một giải pháp hiệu quả mà không cần thu thập thêm chai lỗi.

+ Cách làm: Thêm hàng trăm ảnh ảnh nền (ví dụ: ảnh các băng chuyền nhà máy thực tế, không có chai) vào thư mục huấn luyện.
+ Kết quả: Khi huấn luyện, tham số copy_paste=0.5 sẽ tự động lấy các "lỗi méo", "lỗi rách" từ dữ liệu gốc và dán chúng ngẫu nhiên lên các nền nhà máy thực tế này. Điều này buộc mô hình phải học cách nhận diện cái lỗi thay vì học cái bàn xanh lá.

KẾT LUẬN

Báo cáo đã trình bày chi tiết quá trình nghiên cứu, xây dựng và đánh giá hệ thống phát hiện lỗi chai nước trên băng chuyền, ứng dụng mô hình học sâu YOLO, qua đó hoàn thành các mục tiêu ban đầu đề ra là xây dựng thành công quy trình hệ thống, kiểm chứng hiệu năng mô hình YOLOv11s và đạt được các kết quả định lượng cụ thể. Về mặt thực nghiệm, mô hình đã đạt hiệu suất gần như hoàn hảo trong môi trường lab, với các chỉ số mAP 0.5-95 đạt 99.1% và mAP 0.5 đạt 99.5%. Đặc biệt, việc mô hình đạt Recall 100% (không bỏ sót lỗi) và tốc độ suy luận 6.3ms/khung hình (158 FPS) đã khẳng định khả năng đáp ứng vượt trội yêu cầu kiểm định công nghiệp theo thời gian thực. Tuy nhiên, nhóm nghiên cứu nhận thức rõ hạn chế nghiêm trọng của mô hình rằng toàn bộ dữ liệu được thu thập trong một môi trường duy nhất , dẫn đến nguy cơ mô hình bị học thuộc lòng bối cảnh thay vì học đặc điểm của lỗi. Kết quả 99.1% mAP chỉ chứng minh mô hình hoạt động tốt trong môi trường đó và có khả năng thất bại khi triển khai thực tế. Tóm lại, đề tài đã xây dựng thành công một mô hình nguyên mẫu với độ chính xác và tốc độ rất cao trong môi trường thử nghiệm, khẳng định tính khả thi của YOLOv11, đồng thời chỉ ra hạn chế về tính tổng quát và đề xuất giải pháp cải thiện cụ thể để đưa mô hình từ phòng thí nghiệm đến ứng dụng thực tế.

TÀI LIỆU THAM KHẢO

[1] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," in *Proc. IEEE CVPR*, 2016, pp. 779–788.

[2] Ultralytics, "YOLO – Where to Start," 2024. [Online]. Available: https://docs.ultralytics.com/vi/#where-to-start. [Truy cập: 01-11-2025].

[3] Ultralytics, "YOLOv11 Release Notes," 2024. [Online]. Available: https://docs.ultralytics.com/models/yolo11. [Truy cập: 05-11-2025].

[4] A. Paszke *et al.*, "PyTorch: An Imperative Style, High-Performance Deep Learning Library," *Advances in Neural Information Processing Systems*, vol. 32, pp. 8024–8035, 2019.

[5] Roboflow, "Roboflow Dataset Management and Augmentation," 2024. [Online]. Available: https://roboflow.com. [Truy cập: 05-11-2025].

[6] Viblo, "Tìm hiểu về YOLO trong bài toán Real-time Object Detection," 2020. [Online]. Available: https://viblo.asia/p/tim-hieu-ve-yolo-trong-bai-toan-real-time-object-detection-yMnKMdvr57P. [Truy cập: 05-11-2025].

[7] R. Szeliski, *Computer Vision: Algorithms and Applications*, 2nd ed., Springer, 2022. [Online]. Available: https://dlib.phenikaa-uni.edu.vn/handle/PNK/6042. [Truy cập: 05-11-2025].