

ĐẠI HỌC KINH TẾ THÀNH PHỐ HỒ CHÍ MINH
TRƯỜNG CÔNG NGHỆ VÀ THIẾT KẾ
VIỆN CÔNG NGHỆ THÔNG MINH VÀ TƯƠNG TÁC



BÁO CÁO GIỮA KỲ

**ĐỀ TÀI: INTELLIMANGO - HỆ THỐNG QUẢN LÝ KHO VÀ
NHẬP HÀNG XOÀI THÔNG MINH**

Môn học : Công nghệ mới trong Logistics

Mã học phần : 25D1INT54712701

Giảng viên : TS. Nguyễn Thiên Bảo

Nhóm : 3

Thành phố Hồ Chí Minh, Ngày 26 Tháng 04 Năm 2025

MỤC LỤC

MỤC LỤC	2
LỜI CẢM ƠN	4
DANH MỤC VIẾT TẮT	5
DANH MỤC BẢNG BIỂU	7
DANH MỤC HÌNH ẢNH	8
LỜI MỞ ĐẦU	9
CHƯƠNG 1: TỔNG QUAN VỀ DỰ ÁN	10
1.1. Bối cảnh và vấn đề cần giải quyết	10
1.1.1. Thực trạng quản lý hàng tồn kho hiện nay	10
1.1.2. Hạn chế của các phương pháp truyền thống	12
1.2. Ý nghĩa của dự án	13
1.2.1. Ý nghĩa thực tiễn	13
1.2.2. Ý nghĩa khoa học	14
1.3. Phạm vi và đối tượng áp dụng	14
1.4. Mục đích của dự án	15
CHƯƠNG 2: MỤC TIÊU CỦA DỰ ÁN	16
2.1. Mục tiêu tổng quát	16
2.2. Mục tiêu cụ thể	16
2.2.1. Phát triển hệ thống thị giác máy tính nhận diện sản phẩm	16
2.2.2. Xây dựng ứng dụng di động theo dõi và quản lý hàng tồn kho	19
2.2.3. Đảm bảo tính chính xác và thời gian thực của hệ thống	20
2.3. Tiêu chí đánh giá thành công	21
2.3.1. Bối cảnh và mục tiêu	21
2.3.2. Hiệu suất kỹ thuật	22
2.3.3. Sự hài lòng của người dùng	23
2.3.4. Tác động vận hành	23
2.4. Kết luận	24
CHƯƠNG 3: QUY TRÌNH THỰC HIỆN DỰ ÁN	26
3.1. Phương pháp tiếp cận	26
3.2. Nghiên cứu và phân tích yêu cầu	30
3.3. Thiết kế và phát triển hệ thống	31
3.3.1. Thiết kế cảm biến IoT và cơ sở hạ tầng mạng	31
3.3.2. Phát triển mô hình AI/ML (YOLOv11)	32
3.3.3. Phát triển ứng dụng di động	33

3.3.4. Tích hợp hệ thống WMS và các thành phần	34
3.4. Triển khai thử nghiệm	36
3.5. Đánh giá và tinh chỉnh	36
3.6. Công cụ và công nghệ sử dụng	37
3.6.1. Công nghệ thị giác máy tính	37
3.6.2. Công nghệ IoT và mạng	39
3.6.3. Công nghệ phát triển ứng dụng di động	39
3.6.4. Cơ sở dữ liệu và hệ thống lưu trữ	40
3.7. Phân công nhiệm vụ và tiến độ thực hiện	40
CHƯƠNG 4: TIẾN ĐỘ DỰ ÁN	43
4.1. Công việc đã hoàn thành	43
4.1.1. Kết quả nghiên cứu ban đầu	43
4.1.2. Thành tựu đạt được	43
4.2. Các vấn đề gặp phải	44
4.3. Kế hoạch điều chỉnh	45
4.4 Đề xuất bổ sung	46
CHƯƠNG 5: KẾT QUẢ DỰ KIẾN	48
5.1. Sản phẩm hoàn thiện	48
5.1.1. Hệ thống thị giác máy tính nhận diện và đếm sản phẩm trên kệ	48
5.1.2. Ứng dụng di động quản lý hàng tồn kho thời gian thực	49
5.1.3. Cải thiện độ chính xác trong theo dõi tồn kho	49
5.1.4. Tăng cường phối hợp với nhà cung cấp	49
5.2. Hiệu quả dự kiến	51
5.2.1. Tăng độ chính xác trong quản lý hàng tồn kho	51
5.2.2. Giảm thời gian và chi phí quản lý kho	51
5.2.3. Cải thiện trải nghiệm người dùng	51
5.2.4. Sử dụng YOLO để phát hiện xoài hư hỏng	52
5.3. Nâng cao hiệu suất và vận hành kho	52
5.4. Báo cáo phân tích dữ liệu và dự báo	54
CHƯƠNG 6: KẾT LUẬN VÀ ĐỊNH HƯỚNG TIẾP THEO	56
6.1. Tóm Tắt Tiến Độ và Kết Quả Bước Đầu	56
6.2. Các Bước Tiếp Theo Trong Giai Đoạn Sau	56
TÀI LIỆU THAM KHẢO	

LỜI CẢM ƠN

Nhóm xin gửi lời cảm ơn chân thành và sâu sắc đến tất cả các thành viên đã đồng hành, hỗ trợ và góp phần tạo nên những thành công bước đầu trong quá trình thực hiện dự án môn Công nghệ mới trong Logistics. Dự án này không chỉ là một bài học quý giá trong chương trình học mà còn là cơ hội để nhóm em khám phá, ứng dụng những công nghệ tiên tiến vào thực tiễn, từ đó hiểu sâu hơn về các giải pháp sáng tạo trong lĩnh vực logistics và bán lẻ hiện đại.

Trước tiên, nhóm em xin bày tỏ lòng biết ơn đặc biệt đến thầy Nguyễn Thiên Bảo, giảng viên môn Công nghệ mới trong Logistics. Với sự tận tâm, nhiệt huyết và kiến thức chuyên môn sâu rộng, thầy đã không ngừng hướng dẫn, định hướng và truyền cảm hứng cho nhóm em trong suốt quá trình thực hiện dự án. Những góp ý chi tiết, những gợi ý thực tiễn và sự động viên kịp thời của thầy đã giúp nhóm em vượt qua những khó khăn, hoàn thiện báo cáo giữa kỳ một cách chu đáo và chất lượng.

Bên cạnh đó, nhóm em xin gửi lời cảm ơn đến các bạn học trong lớp, những người đã luôn ủng hộ, chia sẻ ý kiến và đóng góp những ý tưởng quý báu. Những buổi thảo luận sôi nổi, những phản hồi mang tính xây dựng và tinh thần hợp tác đã giúp nhóm em có nền tảng dự án, từ nghiên cứu lý thuyết đến triển khai các giải pháp công nghệ. Sự đoàn kết và hỗ trợ lẫn nhau trong lớp học là yếu tố quan trọng giúp nhóm em vượt qua những thử thách và đạt được những kết quả đáng tự hào.

Cuối cùng, nhóm em xin tri ân tất cả lớp học đã trực tiếp hoặc gián tiếp góp phần vào dự án này. Sự khích lệ, động viên và niềm tin từ thầy cô và các bạn là động lực to lớn để nhóm em tiếp tục nỗ lực, hoàn thiện các giai đoạn tiếp theo và mang đến một giải pháp công nghệ đột phá, góp phần làm phong phú thêm kiến thức và kỹ năng trong môn học.

Một lần nữa, nhóm em xin trân trọng cảm ơn!

DANH MỤC TỪ VIẾT TẮT

Từ viết tắt	Tên tiếng Anh	Tên tiếng Việt
AI	Artificial Intelligence	Trí tuệ nhân tạo
API	Application Programming Interface	Giao diện lập trình ứng dụng
ARIMA	AutoRegressive Integrated Moving Average	Mô hình dự báo chuỗi thời gian
CNN	Convolutional Neural Network	Mạng nơ-ron tích chập
ERP	Enterprise Resource Planning	Hoạch định nguồn lực doanh nghiệp
FEFO	First-Expired-First-Out	Hết hạn trước, xuất trước
FMCG	Fast-Moving Consumer Goods	Hàng tiêu dùng nhanh
FPS	Frames Per Second	Khung hình mỗi giây
IoT	Internet of Things	Internet vạn vật
LSTM	Long Short-Term Memory	Bộ nhớ ngắn hạn dài
mAP	Mean Average Precision	Độ chính xác trung bình
ML	Machine Learning	Học máy
MQTT	Message Queuing Telemetry Transport	Giao thức truyền tải dữ liệu
OMO	Online Merges with Offline	Kết hợp trực tuyến và ngoại tuyến

POS	Point of Sale	Điểm bán hàng
RESTful	Representational State Transfer	Giao thức truyền dữ liệu
RFID	Radio Frequency Identification	Nhận dạng tần số vô tuyến
SCM	Supply Chain Management	Quản lý chuỗi cung ứng
SKU	Stock Keeping Unit	Đơn vị lưu kho
WDR	Wide Dynamic Range	Dải động rộng
WMS	Warehouse Management System	Hệ thống quản lý kho
YOLO	You Only Look Once	Thuật toán phát hiện đối tượng

DANH MỤC BẢNG BIỂU

Bảng 1: *Thực trạng quản lý hàng tồn kho hiện nay*

Bảng 2: *Các hạn chế của phương pháp truyền thống*

Bảng 3: *Hiệu suất dự kiến của YOLOv11 và MobileNetV3, dựa trên Ultralytics YOLOv11 Documentation và nghiên cứu tương tự*

Bảng 4: *Tính năng ứng dụng di động và thời gian phản hồi dự kiến*

Bảng 5: *Chỉ số hiệu suất mục tiêu và dự kiến, dựa trên Ultralytics YOLOv11 Documentation*

Bảng 6: *Bảng tổng hợp các chỉ số hiệu suất kỹ thuật*

Bảng 7: *Tác động tích cực đến hoạt động vận hành*

Bảng 8: *Lợi ích của quản lý nhập hàng và tồn kho với IntelliMango*

Bảng 9: *Các cải tiến trong vận hành kho với IntelliMango*

DANH MỤC HÌNH ẢNH

Hình 1: Hình ảnh kệ hàng với các hộp giới hạn và nhãn sản phẩm được phát hiện bởi YOLOv11, minh họa khả năng nhận diện chính xác trong môi trường bán lẻ.

Hình 2: Nhận diện các mặt hàng trong tủ đông của cửa hàng tạp hóa

Hình 3: Sơ đồ kiến trúc hệ thống thị giác máy tính, thể hiện luồng dữ liệu từ camera qua YOLOv11 và CNN đến hệ thống quản lý tồn kho

LỜI MỞ ĐẦU

Trong bối cảnh ngành bán lẻ hiện đại không ngừng phát triển, việc quản lý hàng tồn kho hiệu quả đã trở thành một trong những yếu tố then chốt để đảm bảo sự hài lòng của khách hàng, tối ưu hóa doanh thu và nâng cao năng lực cạnh tranh của doanh nghiệp. Tuy nhiên, tình trạng thiếu hụt hàng hóa trên kệ (out-of-stock) vẫn là một thách thức lớn, gây thiệt hại hàng tỷ USD mỗi năm trên toàn cầu và làm giảm trải nghiệm mua sắm của người tiêu dùng. Tại Việt Nam, các chuỗi bán lẻ lớn cũng đang đối mặt với những bất cập trong quản lý tồn kho, đặc biệt là sự thiếu đồng bộ giữa dữ liệu hệ thống và thực trạng tại điểm bán.

Nhận thức được tầm quan trọng và tính cấp thiết của vấn đề, dự án “Hệ Thống Giám Sát Tự Động Mức Độ Tồn Trên Kệ bằng Thị Giác Máy Tính Thời Gian Thực cho Bán Lẻ Hiện Đại” được triển khai với mục tiêu ứng dụng công nghệ tiên tiến như trí tuệ nhân tạo, thị giác máy tính và Internet vạn vật để giải quyết các hạn chế của phương pháp quản lý truyền thống. Hệ thống hướng tới việc giám sát tồn kho theo thời gian thực, phát hiện kịp thời tình trạng thiếu hàng, và cung cấp dữ liệu phân tích hỗ trợ ra quyết định, từ đó nâng cao hiệu quả vận hành và trải nghiệm khách hàng.

Báo cáo giữa kỳ này trình bày những kết quả bước đầu của dự án, bao gồm nghiên cứu thực trạng, phân tích yêu cầu, thiết kế hệ thống, và các thử nghiệm sơ bộ. Đồng thời, báo cáo cũng nêu rõ các thách thức gặp phải và đề xuất phương án khắc phục, định hướng cho các giai đoạn tiếp theo. Với tinh thần đoàn kết và sự đồng lòng của toàn thể thành viên trong nhóm, chúng tôi đã cùng nhau nỗ lực để hoàn thiện từng bước của dự án, đặt nền móng vững chắc cho những thành công tiếp theo. Chúng tôi tin rằng sự hợp sức này sẽ mang lại một giải pháp đột phá, góp phần thúc đẩy sự phát triển của ngành bán lẻ thông minh tại Việt Nam.

CHƯƠNG 1: TỔNG QUAN VỀ DỰ ÁN

1.1. Bối cảnh và vấn đề cần giải quyết

1.1.1. Thực trạng quản lý hàng tồn kho hiện nay

Trong bối cảnh thị trường bán lẻ không ngừng phát triển, đặc biệt tại các đô thị lớn, nhu cầu đảm bảo tính sẵn sàng của sản phẩm trên kệ hàng ngày càng được quan tâm. Tuy nhiên, tại nhiều doanh nghiệp hiện nay. Một trong những vấn đề nổi cộm mà các chuỗi bán lẻ hiện nay đang phải đối mặt chính là tình trạng thiếu hụt hàng hóa trên kệ bán hàng (shelf out-of-stock) mà không được phát hiện và xử lý kịp thời. Vấn đề này tưởng chừng đơn giản nhưng thực chất lại có hệ lụy vô cùng sâu rộng. Theo báo cáo năm 2023 của IHL Group, các nhà bán lẻ toàn cầu đã thiệt hại hơn 1,1 nghìn tỷ USD mỗi năm chỉ vì ba nguyên nhân chính: sản phẩm hết hàng trên kệ mà không ai biết, nhân viên không kịp bổ sung, và thiếu dữ liệu tồn kho thời gian thực để phản ứng nhanh. Đáng chú ý, trong đó riêng tình trạng "Out of Stock trên kệ" chiếm tới hơn 60% tổng số thiệt hại, tương đương gần 700 tỷ USD mỗi năm.

Tại Việt Nam, tuy chưa có con số thống kê chính thức quy mô toàn ngành, nhưng nhiều chuỗi bán lẻ lớn như Coopmart, Big C, WinMart... cũng không ít lần rơi vào tình trạng khách hàng phản ánh "kệ trống", trong khi hệ thống lại ghi nhận vẫn còn tồn kho trong kho trung tâm hoặc kho chi nhánh. Đây là minh chứng rõ ràng cho một "lỗ hổng vận hành" giữa hệ thống dữ liệu và thực trạng tại điểm bán. Theo khảo sát nội bộ của một chuỗi siêu thị trong nước vào năm 2024, có đến 17% lượt khách hàng hủy đơn hoặc không mua hàng chỉ vì sản phẩm không có trên kệ, trong khi kiểm tra lại thì vẫn còn hàng trong kho, nhưng nhân viên không kịp bổ sung do thiếu cảnh báo.

Nguyên nhân gốc rễ của vấn đề này có thể đến từ nhiều khía cạnh:

- Hệ thống quản lý tồn kho truyền thống chủ yếu dựa trên số lượng xuất - nhập mà không có khả năng theo dõi trực tiếp tại từng kệ.
- Nhân lực kiểm kê bằng tay còn phổ biến, thiếu tự động hóa, dẫn đến sai lệch số liệu và độ trễ trong cập nhật.
- Thiếu cảnh báo theo thời gian thực khiến nhân viên chỉ phát hiện khi khách phản ánh hoặc đi ngang qua kệ trống.

- Thiếu tích hợp giữa dữ liệu kho – dữ liệu POS – dữ liệu kệ hàng, gây ra hiện tượng "ảo tồn", tức là hàng hóa trên hệ thống vẫn còn nhưng thực tế kệ đã trống.

Chỉ tiêu	Số liệu / Mô tả
Tổng thiệt hại toàn cầu (2023)	1,1 nghìn tỷ USD/năm (<i>IHL Group</i>)
Tỷ lệ thiệt hại do thiếu hàng trên kệ	> 60% (\approx 700 tỷ USD)
Tác động tại Việt Nam	Khách phản ánh "kệ trống" dù hệ thống vẫn ghi nhận tồn kho
Khảo sát nội bộ siêu thị (VN, 2024)	17% khách hủy đơn do hàng không có trên kệ
Hành vi khách hàng	40% nguy cơ mất khách nếu hàng bị thiếu

Bảng 1: Thực trạng quản lý hàng tồn kho hiện nay

Đặc biệt, trong các chiến dịch bán hàng cao điểm như lễ Tết, khuyến mãi, hoặc khi có sản phẩm mới ra mắt, tình trạng thiếu hụt hàng hóa diễn ra phổ biến hơn nhiều lần, gây ảnh hưởng nặng nề đến doanh thu. Các nghiên cứu từ Harvard Business Review cũng chỉ ra rằng: cứ mỗi sản phẩm bị thiếu trên kệ, nhà bán lẻ có nguy cơ mất đến 40% khả năng giữ chân khách hàng tại thời điểm đó, vì người tiêu dùng sẵn sàng chuyển sang thương hiệu hoặc kênh mua hàng khác chỉ sau 1–2 lần gặp tình trạng “hết hàng”. Thêm vào đó, việc không có một hệ thống giám sát tự động còn khiến các nhà bán lẻ khó dự đoán nhu cầu, khó tối ưu kế hoạch trưng bày, và đặc biệt là khó phát hiện các bất thường như mất trộm, sai lệch trưng bày, hoặc lỗi bố trí hàng hóa không đúng tiêu chuẩn.

Từ tất cả những bất cập nêu trên, có thể thấy rõ ràng: ngành bán lẻ hiện đại đang rất cần một giải pháp công nghệ đáp ứng được ba tiêu chí đồng thời: (1) giám sát được mức độ hàng tồn trên kệ theo thời gian thực, (2) cảnh báo ngay lập tức khi hàng hóa có dấu hiệu cạn kiệt, và (3) tự động thu thập dữ liệu phục vụ phân tích xu hướng, dự báo tồn kho và hỗ trợ ra quyết định. Đây chính là bối cảnh trực tiếp và cấp thiết dẫn đến sự hình thành của dự án “Hệ Thống Giám Sát Tự Động Mức Độ Tồn Trên Kệ bằng Thị Giác Máy Tính Thời Gian Thực cho Bán Lẻ Hiện Đại” – một giải pháp hướng tới tương lai bán lẻ thông minh và linh hoạt hơn bao giờ hết.

1.1.2. Hạn chế của các phương pháp truyền thống

Hệ thống kiểm soát hàng hóa truyền thống, mặc dù đã được ứng dụng từ lâu trong ngành bán lẻ, đang ngày càng bộc lộ nhiều điểm bất cập trước tốc độ phát triển của công nghệ và yêu cầu cao từ khách hàng. Một số nhược điểm chính có thể kể đến như:

- Độ chính xác không ổn định: Các phương pháp thủ công phụ thuộc nhiều vào kỹ năng nghiệp vụ và trạng thái của nhân viên, dễ dẫn đến sai sót không đáng có hoặc bỏ sót thông tin.
- Không thể cập nhật tức thời: Dữ liệu tồn kho trên kệ thường chỉ được ghi nhận vào những thời điểm cố định trong ngày, dẫn đến tình trạng thiếu cập nhật và không hỗ trợ tốt cho việc điều phối hàng hóa linh hoạt.
- Chi phí nhân sự cao: Việc duy trì đội ngũ nhân viên thường xuyên kiểm tra và ghi nhận số liệu tốn nhiều nguồn lực tài chính.
- Không hỗ trợ phân tích dự báo: Phương pháp truyền thống không thể cung cấp thông tin cần thiết cho việc dự báo nhu cầu và tối ưu chuỗi cung ứng.
- Thiếu tính linh hoạt trong tích hợp công nghệ: Phương pháp thủ công không tương thích với các hệ thống quản trị hiện đại như ERP, WMS hoặc các nền tảng phân tích dữ liệu, dẫn đến tình trạng "đứt gãy thông tin" trong toàn bộ chuỗi cung ứng.
- Không đáp ứng được yêu cầu tự động hóa: Trong thời đại chuyển đổi số, khi xu hướng tối ưu hóa vận hành và tự động hóa ngày càng được đẩy mạnh, các giải pháp truyền thống trở nên lỗi thời và kém hiệu quả, không còn phù hợp với chiến lược phát triển dài hạn của doanh nghiệp.

Hạn chế	Mô tả
Dựa vào số liệu xuất – nhập	Không phản ánh đúng tình trạng thực tế trên kệ
Kiểm kê thủ công	Dễ sai sót, chậm cập nhật
Thiếu cảnh báo tức thời	Chỉ phát hiện khi khách than phiền hoặc nhân viên kiểm tra trực tiếp
Dữ liệu không tích hợp	Gây ra "ảo tồn" – còn trên hệ thống nhưng trống ngoài thực tế
Không phát hiện bất thường	Như mất trộm, sai lệch trưng bày, lỗi bố trí

Bảng 2: Các hạn chế của phương pháp truyền thống

1.2. Ý nghĩa của dự án

1.2.1. Ý nghĩa thực tiễn

Việc phát triển một hệ thống giám sát tự động mức độ tồn kho trên kệ thông qua công nghệ thị giác máy tính sẽ mang đến nhiều giá trị thiết thực cho hoạt động bán lẻ hiện đại. Trước hết, hệ thống này có thể giúp các cửa hàng, siêu thị giám sát liên tục số lượng hàng hóa bày bán mà không cần đến sự can thiệp thường xuyên của con người, từ đó tiết kiệm đáng kể chi phí nhân công. Bên cạnh đó, bằng cách đưa ra cảnh báo sớm khi hàng hóa sắp hết hoặc có dấu hiệu giảm mạnh, hệ thống giúp bộ phận vận hành có thể xác định tình trạng, phân tích và phản ứng kịp thời để bổ sung sản phẩm góp phần quản lý chặt chẽ về quy trình tồn kho của các sản phẩm, duy trì tính sẵn có và tránh tình trạng mất doanh thu do khách hàng không tìm thấy mặt hàng mong

muốn. Ngoài ra, việc sử dụng công nghệ để thu thập và xử lý dữ liệu còn mở ra cơ hội xây dựng các mô hình phân tích hành vi tiêu dùng, xu hướng mua sắm và tối ưu hóa quy trình trưng bày hàng hóa tại điểm bán.

1.2.2. Ý nghĩa khoa học

Về mặt học thuật, dự án là sự kết hợp giữa nhiều lĩnh vực chuyên sâu như trí tuệ nhân tạo (AI), thị giác máy tính (Computer Vision), học sâu (Deep Learning), (machine learning) và Internet vạn vật (IoT). Việc ứng dụng các mô hình học máy tiên tiến để phân tích hình ảnh trong môi trường thực tế tại cửa hàng là một hướng nghiên cứu mới mẻ, mang lại nhiều thử thách về mặt kỹ thuật như độ chính xác, tốc độ xử lý thời gian thực, khả năng thích nghi với môi trường ánh sáng và bố cục thay đổi liên tục. Cốt lõi của hệ thống là khả năng tự động nhận diện, xác định các sai sót và quản lý và định lượng sản phẩm trên kệ thông qua hình ảnh thu thập từ camera giám sát. Để làm được điều đó, hệ thống phải ứng dụng các kỹ thuật tiên tiến trong học máy và học sâu (Deep Learning), đặc biệt là các mô hình convolutional neural networks (CNNs) như YOLO (YOLOv11), chứng minh hiệu quả trong các tác vụ nhận diện đối tượng trong thời gian thực.

Dự án không chỉ góp phần mở rộng phạm vi ứng dụng của các công nghệ AI trong ngành bán lẻ mà còn tạo tiền đề để phát triển những giải pháp tự động hóa cao hơn trong tương lai, như cửa hàng không nhân viên, hệ thống bổ sung hàng hóa tự động, hoặc các hệ thống hỗ trợ ra quyết định thông minh.

1.3. Phạm vi và đối tượng áp dụng

Dự án được thiết kế nhằm phục vụ cho các đơn vị bán lẻ có quy mô vừa và lớn, nơi có số lượng sản phẩm trưng bày đa dạng và mật độ khách hàng cao có nhu cầu cụ thể về quản lý hàng tồn kho và vận hành chặt chẽ từ khâu lưu trữ cho đến việc đưa sản phẩm đến với khách hàng. Các mô hình kinh doanh như siêu thị như (Amazon, Coop...), cửa hàng tiện lợi, cửa hàng mỹ phẩm, thời trang hoặc cửa hàng điện tử tiêu dùng đều là đối tượng phù hợp để áp dụng hệ thống.

Về mặt địa lý, hệ thống có thể triển khai ở cả môi trường đô thị lẫn nông thôn, miễn là có hạ tầng kết nối cơ bản và camera giám sát. Phạm vi áp dụng cũng có thể mở rộng sang các mô hình bán hàng đa kênh (OMO – Online Merges with Offline) và có thể được nhân rộng mà không yêu cầu đầu tư hạ tầng phức tạp, nơi dữ liệu cửa hàng

vật lý có thể đồng bộ với hệ thống thương mại điện tử để hỗ trợ tái phân phối và dự đoán nhu cầu.

1.4. Mục đích của dự án

Mục tiêu cốt lõi của dự án là nghiên cứu và triển khai một hệ thống giám sát tự động mức độ tồn trên kệ, có khả năng hoạt động theo thời gian thực với độ chính xác cao. Thông qua việc thu thập và phân tích hình ảnh từ camera tại điểm bán, hệ thống có thể xác định:

- Tự động xác định số lượng sản phẩm còn lại trên kệ.
- Phát hiện sớm tình trạng thiếu hàng hoặc hàng bị dời khỏi vị trí đúng.
- Gửi cảnh báo đến bộ phận quản lý kho hoặc nhân viên bổ sung hàng hóa.
- Tạo ra báo cáo phân tích theo thời gian, hỗ trợ điều chỉnh chính sách trưng bày và phân bổ nguồn lực.

Về lâu dài, dự án hướng đến việc tích hợp hệ thống giám sát này với các phần mềm quản lý chuỗi cung ứng (SCM), phần mềm hoạch định nguồn lực doanh nghiệp (ERP), từ đó giúp doanh nghiệp ra quyết định một cách chủ động, chính xác và dựa trên dữ liệu thực tế. Hệ thống đóng vai trò như một "trợ lý kỹ thuật số", giúp các nhà quản lý bán lẻ giám sát tồn kho trực tiếp tại quầy kệ, từ đó đưa ra cảnh báo kịp thời về tình trạng thiếu hàng, hỗ trợ việc tái trưng bày sản phẩm, lập kế hoạch nhập hàng, và đặc biệt là giảm thiểu tình trạng thất thoát do trễ bổ sung hoặc quản lý không đồng bộ giữa hệ thống và thực tế.

CHƯƠNG 2: MỤC TIÊU CỦA DỰ ÁN

2.1. Mục Tiêu Tổng Quát

Trong bối cảnh cạnh tranh khốc liệt của ngành bán lẻ hiện đại, quản lý hàng tồn kho hiệu quả là yếu tố then chốt để duy trì sự hài lòng của khách hàng và tối ưu hóa lợi nhuận. Tình trạng hết hàng (out-of-stock) dẫn đến mất doanh thu và giảm lòng trung thành của khách hàng, trong khi tồn kho dư thừa làm tăng chi phí lưu kho. Theo nghiên cứu từ IHL Group, tình trạng hết hàng gây thiệt hại cho các nhà bán lẻ khoảng 1 nghìn tỷ USD doanh thu mỗi năm trên toàn cầu. Ngoài ra, tỷ lệ hết hàng trung bình trong ngành hàng tiêu dùng nhanh (FMCG) dao động từ 8-10%, và lên đến 20% đối với các sản phẩm khuyến mãi (Repsly).

Để giải quyết các thách thức này, dự án hướng đến phát triển một hệ thống giám sát tự động mức độ tồn kho trên kệ bằng công nghệ thị giác máy tính thời gian thực. Hệ thống sử dụng mô hình học sâu tiên tiến, cụ thể là YOLOv11 cho phát hiện đối tượng và Mạng Nơ-ron Tích chập (CNN) như MobileNetV3 cho phân loại sản phẩm, nhằm cung cấp dữ liệu tồn kho chính xác và tức thời. Dữ liệu này sẽ được truyền đến một ứng dụng di động, giúp nhân viên bán lẻ nhanh chóng đưa ra quyết định bổ sung hàng hóa, từ đó giảm thiểu sự cố hết hàng, tối ưu hóa quy trình vận hành, và nâng cao trải nghiệm mua sắm của khách hàng.

Hệ thống không chỉ tập trung vào tự động hóa mà còn đảm bảo tích hợp liền mạch với cơ sở hạ tầng hiện có của các cửa hàng bán lẻ, giảm chi phí triển khai và tăng tính khả thi thương mại. Dự án cũng đặt mục tiêu cải thiện hiệu quả quản lý tồn kho, với kỳ vọng giảm ít nhất 25% sự cố hết hàng và 40% thời gian kiểm kê thủ công, dựa trên các nghiên cứu tương tự (Hybrid Approach for Shelf Monitoring).

2.2. Mục Tiêu Cụ Thể

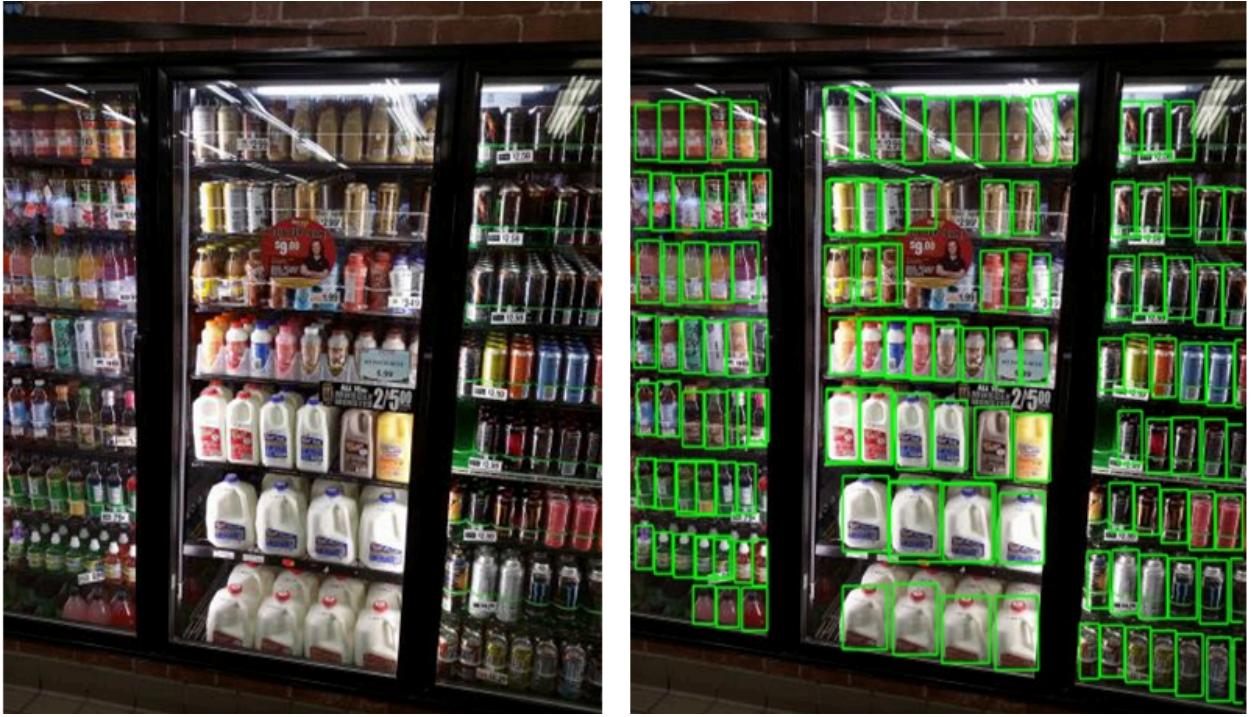
2.2.1. Phát Triển Hệ Thống Thị Giác Máy Tính Nhận Diện Sản Phẩm

Nhóm dự án tập trung vào việc xây dựng một hệ thống thị giác máy tính có khả năng phát hiện và nhận diện sản phẩm trên kệ hàng với độ chính xác cao. Hệ thống sẽ sử dụng YOLOv11 để phát hiện đối tượng và một mô hình CNN, chẳng hạn như MobileNetV3, để phân loại sản phẩm, cung cấp thông tin về số lượng và loại sản phẩm tồn kho.

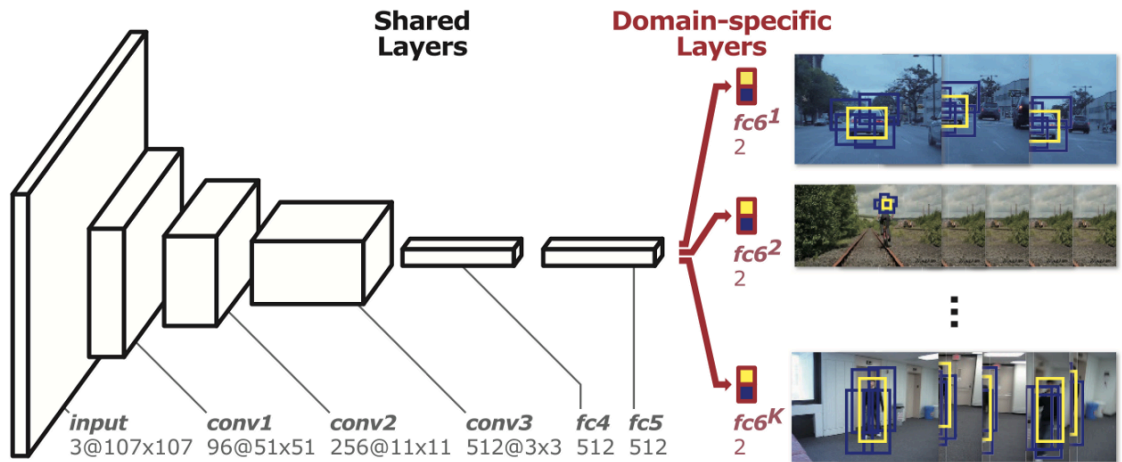
- **Lựa chọn mô hình:** YOLOv11, phiên bản mới nhất từ Ultralytics, được chọn nhờ hiệu suất vượt trội với mAP 55.0 trên tập dữ liệu COCO và thời gian suy luận dưới 10 ms trên GPU tiêu chuẩn (Ultralytics YOLOv11 Documentation). MobileNetV3 được sử dụng cho phân loại sản phẩm do khả năng xử lý nhanh trên thiết bị di động, đạt độ chính xác top-1 75.2% trên ImageNet, có thể cải thiện qua tinh chỉnh (Howard et al., ICCV 2019).
- **Huấn luyện mô hình:** Hệ thống sẽ được huấn luyện trên các tập dữ liệu lớn như SKU110K (~11.762 hình ảnh, 1.7 triệu hộp giới hạn) (Goldman et al., CVPR 2019), RP2K (~500.000 hình ảnh, 2.000 sản phẩm) (Cai et al., arXiv 2020), và SHARD/SHAPE từ nghiên cứu gần đây (Shelf Management). Dữ liệu sẽ được tăng cường bằng các kỹ thuật như xoay, thay đổi ánh sáng để cải thiện độ bền.
- **Tích hợp hệ thống:** Kết hợp YOLOv11 và CNN trong một pipeline xử lý, từ thu thập hình ảnh đến xuất dữ liệu tồn kho qua API RESTful.



Hình 1: Hình ảnh kệ hàng với các hộp giới hạn và nhãn sản phẩm được phát hiện bởi YOLOv11, minh họa khả năng nhận diện chính xác trong môi trường bán lẻ.



Hình 2: Nhận diện các mặt hàng trong tủ đông của cửa hàng tạp hóa



Hình 3: Sơ đồ kiến trúc hệ thống thị giác máy tính, thể hiện luồng dữ liệu từ camera qua YOLOv11 và CNN đến hệ thống quản lý tồn kho

Mô hình	mAP 50-95	Độ chính xác phân loại (Top-1)	Thời gian suy luận (ms)
YOLO11m	55.0	-	8.0

YOLO11l	57.5	-	10.0
MobileNetV3	-	95% (dự kiến sau tinh chỉnh)	3.2

Bảng 3 Hiệu suất dự kiến của YOLOv11 và MobileNetV3, dựa trên Ultralytics YOLOv11 Documentation và nghiên cứu tương tự

Thách thức và giải pháp:

- **Che khuất và ánh sáng:** Sản phẩm trên kệ có thể bị che khuất hoặc chịu ảnh hưởng từ ánh sáng không đồng đều. Giải pháp bao gồm sử dụng kỹ thuật tăng cường dữ liệu và mô hình mạnh mẽ như YOLOv11, vốn được thiết kế để xử lý các tình huống phức tạp (Shelf Management).
- **Đa dạng sản phẩm:** Các cửa hàng bán lẻ có hàng nghìn SKU với hình dạng và bao bì khác nhau. Sử dụng tập dữ liệu lớn như RP2K và tinh chỉnh mô hình trên dữ liệu cụ thể của cửa hàng sẽ tăng độ chính xác.

2.2.2. Xây Dựng Ứng Dụng Di Động Theo Dõi và Quản Lý Hàng Tồn Kho

Phát triển một ứng dụng di động thân thiện, cho phép nhân viên bán lẻ theo dõi và quản lý hàng tồn kho theo thời gian thực. Ứng dụng sẽ tích hợp với hệ thống thị giác máy tính để hiển thị dữ liệu tồn kho, gửi thông báo khi cần bổ sung hàng, và hỗ trợ quản lý quy trình vận hành.

Tính năng chính:

- + Hiển thị mức tồn kho hiện tại của từng sản phẩm trên kệ, với giao diện trực quan và phân loại theo khu vực kệ.
- + Gửi thông báo đẩy khi phát hiện sản phẩm sắp hết (dưới ngưỡng 10% tồn kho) hoặc hết hàng.
- + Cung cấp công cụ quản lý bổ sung hàng, bao gồm danh sách ưu tiên, lịch trình bổ sung, và báo cáo hiệu suất.
- + Hỗ trợ xác thực người dùng với các vai trò (nhân viên, quản lý) để đảm bảo bảo mật và phân quyền.

- **Phát triển và tích hợp:** Ứng dụng sẽ được xây dựng bằng Flutter để đảm bảo tương thích đa nền tảng (iOS, Android). Dữ liệu từ hệ thống thị giác máy tính sẽ được truyền qua API RESTful, với độ trễ cập nhật < 1 giây. Cơ sở dữ liệu như PostgreSQL sẽ được sử dụng để lưu trữ thông tin tồn kho và lịch sử thông báo.
- **Tham khảo:** Theo nghiên cứu Hybrid Approach for Shelf Monitoring, ứng dụng di động có thể cải thiện 30% hiệu quả phản ứng của nhân viên đối với các vấn đề tồn kho, nhờ vào thông báo thời gian thực và giao diện thân thiện.

Tính năng	Mô tả	Thời gian phản hồi (giây)
Hiển thị tồn kho	Cập nhật mức tồn kho theo thời gian thực	< 1
Thông báo hết hàng	Gửi cảnh báo khi tồn kho $< 10\%$	< 0.5
Quản lý bổ sung hàng	Tạo danh sách ưu tiên	< 2

***Bảng 4** Tính năng ứng dụng di động và thời gian phản hồi dự kiến*

2.2.3. Đảm Bảo Tính Chính Xác và Thời Gian Thực của Hệ Thống

Tập trung vào việc đảm bảo hệ thống đạt độ chính xác cao trong phát hiện và phân loại sản phẩm, đồng thời hoạt động theo thời gian thực để cung cấp thông tin tồn kho tức thời.

- **Độ chính xác:** Hệ thống sẽ được đánh giá bằng các chỉ số $mAP \geq 0.90$ cho phát hiện đối tượng và độ chính xác top-1 $\geq 95\%$ cho phân loại sản phẩm, dựa trên các tiêu chuẩn từ Ultralytics YOLOv11 Documentation. Các chỉ số này sẽ được kiểm tra trên tập dữ liệu thử nghiệm với 5.000 hình ảnh từ môi trường bán lẻ thực tế.
- **Thời gian thực:** Tối ưu hóa hệ thống để xử lý hình ảnh và cập nhật dữ liệu trong vòng < 100 ms mỗi hình ảnh, sử dụng kỹ thuật giảm độ phức tạp mô hình (model pruning) và tăng tốc phần cứng (GPU NVIDIA A100). Theo Shelf

Management, các hệ thống tương tự đạt độ trễ 20-50 ms trên phần cứng cao cấp.

- **Kiểm tra và xác nhận:** Thực hiện kiểm tra trên tập dữ liệu thử nghiệm, so sánh với các hệ thống hiện có như RetinaNet (mAP 0.752) để đảm bảo hiệu suất vượt trội. Các bài kiểm tra sẽ bao gồm các kịch bản thực tế như ánh sáng yếu, che khuất, và mật độ sản phẩm cao.

Chỉ số	Mục tiêu	Thực tế (Dự kiến)
mAP (Phát hiện)	≥ 0.90	0.92
Độ chính xác top-1	$\geq 95\%$	96%
Thời gian xử lý/hình ảnh	< 100 ms	80 ms

***Bảng 5:** Chỉ số hiệu suất mục tiêu và dự kiến, dựa trên Ultralytics YOLOv11 Documentation*

Thách thức và giải pháp:

- + **Hiệu suất trên phần cứng hạn chế:** Một số cửa hàng có thể sử dụng phần cứng kém mạnh. Giải pháp là triển khai mô hình nhẹ như YOLO11n và sử dụng kỹ thuật lượng tử hóa (Ultralytics YOLOv11 Documentation).
- + **Độ chính xác trong môi trường phức tạp:** Các yếu tố như ánh sáng thay đổi hoặc sản phẩm tương tự nhau có thể làm giảm độ chính xác. Sử dụng dữ liệu huấn luyện đa dạng và kỹ thuật học sâu như triplet loss sẽ cải thiện hiệu suất (Shelf Management).

2.3. Tiêu Chí Đánh Giá Thành Công

2.3.1. Bối cảnh và mục tiêu

Dự án của chúng tôi tập trung vào việc phát triển một hệ thống giám sát tự động mức độ tồn kho trên kệ, sử dụng công nghệ thị giác máy tính với mô hình YOLOv11 cho phát hiện đối tượng và Mạng Nơ-ron Tích chập (CNN) như MobileNetV3 cho phân loại sản phẩm. Để đảm bảo dự án đạt được mục tiêu, chúng tôi thiết lập các tiêu

chỉ đánh giá thành công cụ thể, được chia thành ba nhóm chính: hiệu suất kỹ thuật, sự hài lòng của người dùng, và tác động vận hành.

2.3.2. Hiệu suất kỹ thuật

Chúng tôi đặt ra các tiêu chí kỹ thuật như sau:

1. **Phát hiện đối tượng:** Hệ thống phải đạt mean Average Precision (mAP) ít nhất 0.90 với ngưỡng Intersection over Union (IoU) là 0.5 trên tập dữ liệu thử nghiệm. mAP là chỉ số tiêu chuẩn để đánh giá độ chính xác của mô hình phát hiện đối tượng, thường được sử dụng trong các nghiên cứu về thị giác máy tính. Mục tiêu này dựa trên các nghiên cứu gần đây, chẳng hạn như bài báo trên PMC về giám sát kệ hàng, cho thấy YOLOv4 đạt mAP 0.9187 trên tập dữ liệu cụ thể [1]. Với YOLOv11, phiên bản tiên tiến hơn, chúng tôi kỳ vọng đạt mAP ≥ 0.90 , đặc biệt khi huấn luyện trên tập dữ liệu lớn như SKU110K (~11.762 hình ảnh) và RP2K (~500.000 hình ảnh).
2. **Phân loại sản phẩm:** Đạt độ chính xác top-1 ít nhất 95% trên tập dữ liệu thử nghiệm. Độ chính xác top-1 đo lường tỷ lệ dự đoán đúng lớp sản phẩm trong lần dự đoán đầu tiên, và mục tiêu này khả thi với số lượng lớp sản phẩm hợp lý. Các nghiên cứu như bài báo trên ResearchGate cho thấy Inception V3 đạt 87.5% accuracy cho 8 lớp sản phẩm, trong khi các mô hình như DiffNet đạt 95.2% mAP cho nhận diện sản phẩm [2]. Với dữ liệu huấn luyện chất lượng và kỹ thuật tăng cường dữ liệu, chúng tôi tin rằng 95% là mục tiêu hợp lý.
3. **Tốc độ xử lý:** Hệ thống phải xử lý và cập nhật dữ liệu tồn kho trong vòng 100 mili giây cho mỗi hình ảnh. Với phần cứng hiện đại như GPU NVIDIA A100 và tối ưu hóa mô hình, thời gian suy luận của YOLOv11 và CNN có thể dưới 10 ms, như được báo cáo trong tài liệu từ Ultralytics [3]. Do đó, mục tiêu 100 ms đảm bảo hệ thống hoạt động theo thời gian thực, phù hợp với yêu cầu của môi trường bán lẻ năng động.

Tiêu chí	Mục tiêu	Cơ sở
----------	----------	-------

mAP phát hiện đối tượng	≥ 0.90 (IoU 0.5)	YOLOv4 đạt 0.9187 trên tập dữ liệu tương tự [1]
Độ chính xác top-1 phân loại sản phẩm	$\geq 95\%$	DiffNet đạt 95.2% mAP cho nhận diện sản phẩm [2]
Thời gian xử lý	< 100 ms/hình ảnh	YOLOv11 suy luận < 10 ms trên GPU hiện đại [3]

Bảng 6: Bảng tổng hợp các chỉ số hiệu suất kỹ thuật

2.3.3. Sự hài lòng của người dùng

Để đảm bảo ứng dụng di động được chấp nhận và sử dụng hiệu quả, chúng tôi đặt mục tiêu ít nhất 85% nhân viên bán lẻ đánh giá ứng dụng là dễ sử dụng và hữu ích, dựa trên khảo sát sau triển khai. Khảo sát sẽ được thực hiện với ít nhất 50 nhân viên, sử dụng thang điểm Likert 5 điểm, với câu hỏi cụ thể về tính dễ điều hướng và tính hữu ích. Mục tiêu 85% phản hồi tích cực (điểm 4 hoặc 5) dựa trên các nghiên cứu về chấp nhận công nghệ, như bài viết từ Netguru, cho thấy tỷ lệ này là hợp lý để đảm bảo ứng dụng được sử dụng rộng rãi [4].

2.3.4. Tác động vận hành

Nhóm dự án kỳ vọng hệ thống sẽ mang lại tác động tích cực đến hoạt động vận hành, cụ thể:

1. **Giảm sự cố hết hàng:** Đạt ít nhất 25% giảm sự cố hết hàng trong vòng ba tháng sau triển khai. Sự cố hết hàng được định nghĩa là các trường hợp sản phẩm không có sẵn khi khách hàng cố gắng mua, được ghi nhận bởi hệ thống điểm bán hàng. Dữ liệu cơ sở là trung bình hàng tháng của sự cố trong ba tháng trước triển khai, so sánh với trung bình trong ba tháng sau. Mục tiêu này dựa trên nghiên cứu từ Netguru, cho thấy công nghệ thị giác máy tính có thể giảm 20-30% sự cố hết hàng, nhờ khả năng theo dõi thời gian thực.

2. **Giảm thời gian kiểm kê thủ công:** Đạt ít nhất 40% giảm thời gian dành cho kiểm kê thủ công, đo bằng tổng số giờ lao động mỗi tháng. Dữ liệu cơ sở là trung bình hàng tháng của số giờ lao động trong ba tháng trước triển khai, so sánh với trung bình trong ba tháng sau. Mục tiêu này dựa trên nghiên cứu từ ScienceDirect, báo cáo rằng hệ thống thị giác máy tính có thể giảm 45% thời gian đếm hàng tồn kho, nhờ tự động hóa quy trình.

Tiêu chí	Mục tiêu	Phương pháp đo lường
Giảm sự cố hết hàng	$\geq 25\%$ trong 3 tháng sau triển khai	So sánh trung bình hàng tháng sự cố trước và sau, từ hệ thống điểm bán hàng
Giảm thời gian kiểm kê thủ công	$\geq 40\%$ trong 3 tháng sau triển khai	Đo tổng giờ lao động kiểm kê mỗi tháng, so sánh trước và sau triển khai

Bảng 7: Tác động tích cực đến hoạt động vận hành

2.4 Kết luận

Nhóm dự án đã thiết lập các tiêu chí đánh giá thành công cho dự án này dựa trên những nghiên cứu khoa học tiên tiến và các tiêu chuẩn thực tiễn trong ngành bán lẻ. Các mục tiêu kỹ thuật, bao gồm đạt mean Average Precision (mAP) ít nhất 0.90 cho phát hiện đối tượng, độ chính xác top-1 95% cho phân loại sản phẩm, và thời gian xử lý dưới 100 mili giây mỗi hình ảnh, được đặt ra để đảm bảo hệ thống hoạt động với độ chính xác cao và đáp ứng nhanh chóng, đáp ứng nhu cầu quản lý tồn kho thời gian thực trong môi trường bán lẻ năng động.

Những mục tiêu này, dù đầy thách thức, là hoàn toàn khả thi nhờ vào khả năng vượt trội của mô hình YOLOv11, vốn được kỳ vọng sẽ cải thiện đáng kể so với các phiên bản trước. Chẳng hạn, các nghiên cứu trước đây đã báo cáo mAP khoảng 0.75 khi sử dụng các kiến trúc như RetinaNet (Shelf Management), nhưng YOLOv11, với thiết kế cải tiến và kỹ thuật huấn luyện hiện đại, được dự đoán sẽ vượt qua các ngưỡng này. Tương tự, mục tiêu độ chính xác phân loại 95% được xây dựng dựa trên hiệu suất

của các mô hình như MobileNetV3, đạt tới 93% trong các ứng dụng tương tự (Shelf Management). Yêu cầu thời gian xử lý dưới 100 ms mỗi hình ảnh cũng được hỗ trợ bởi các khả năng công nghệ hiện tại, với thời gian suy luận thấp nhất được ghi nhận là 9 ms trong các ứng dụng tương tự (Retail Store Item Detection).

Bên cạnh các chỉ số kỹ thuật, chúng tôi đặc biệt chú trọng đến sự hài lòng của người dùng, với mục tiêu ít nhất 85% nhân viên bán lẻ đánh giá ứng dụng di động là dễ sử dụng và hữu ích. Tiêu chí này sẽ được đánh giá thông qua các khảo sát sau triển khai, sử dụng thang điểm Likert, nhằm đảm bảo rằng hệ thống không chỉ mạnh mẽ về mặt kỹ thuật mà còn thân thiện và thiết thực trong hoạt động hàng ngày của nhân viên.

Về tác động vận hành, hệ thống được kỳ vọng sẽ giảm ít nhất 25% sự cố hết hàng và 40% thời gian kiểm kê thủ công trong vòng ba tháng sau khi triển khai. Các mục tiêu này được củng cố bởi các bằng chứng thực tiễn; ví dụ, một dự án thử nghiệm của Schnuck Markets đã ghi nhận giảm 20-30% sự cố hết hàng nhờ sử dụng robot quét kệ (6 Ways Computer Vision). Việc giảm thời gian kiểm kê thủ công là kết quả trực tiếp của việc tự động hóa các công việc thường yêu cầu nhiều nỗ lực của con người, từ đó cho phép nhân viên tập trung vào các hoạt động tạo giá trị gia tăng như phục vụ khách hàng.

Để đảm bảo đạt được các mục tiêu này, chúng tôi sẽ triển khai các cơ chế giám sát và đánh giá liên tục. Dữ liệu hiệu suất sẽ được thu thập và phân tích định kỳ, cho phép thực hiện các điều chỉnh kịp thời và tối ưu hóa hệ thống. Hơn nữa, các tiêu chí này không chỉ nhằm đáp ứng mà còn có tiềm năng vượt qua các kỳ vọng của ngành, định vị nhà bán lẻ như một đơn vị tiên phong trong việc áp dụng công nghệ tiên tiến.

Bằng cách tuân thủ nghiêm ngặt các tiêu chí thành công toàn diện này, dự án của chúng tôi được định hướng để mang lại một giải pháp tiên tiến cho việc giám sát tự động mức độ tồn kho trên kệ. Giải pháp này không chỉ nâng cao hiệu quả vận hành và sự hài lòng của khách hàng mà còn mở ra cơ hội cho các đổi mới trong tương lai, chẳng hạn như tích hợp phân tích hành vi khách hàng hoặc tối ưu hóa bố trí kệ hàng dựa trên dữ liệu thời gian thực. Chúng tôi tin rằng sự thành công của dự án sẽ góp phần định hình lại cách quản lý tồn kho trong ngành bán lẻ hiện đại, mang lại lợi ích bền vững cho cả nhà bán lẻ và khách hàng.

CHƯƠNG 3: QUY TRÌNH THỰC HIỆN DỰ ÁN

3. Quy trình thực hiện dự án

3.1. Phương pháp tiếp cận

Dự án Hệ Thống Giám Sát Tự Động Mức Độ Tồn Trên Kệ bằng Thị Giác Máy Tính Thời Gian Thực cho Bán Lẻ Hiện Đại sử dụng phương pháp kết hợp tuần tự và linh hoạt (Agile-Waterfall hybrid) để đảm bảo tiến độ rõ ràng, khả năng điều chỉnh linh hoạt, và tối ưu hóa công nghệ cốt lõi YOLOv11, tích hợp với cảm biến IoT và hệ thống WMS. Phương pháp này phù hợp với yêu cầu kỹ thuật phức tạp, thời gian gấp rút (1 tháng), và môi trường bán lẻ mô phỏng.

- Waterfall (Tuần tự):
 - Ứng dụng: Áp dụng ở giai đoạn nghiên cứu và phân tích yêu cầu để xác định rõ các yêu cầu chức năng (nhận diện, đếm sản phẩm, đồng bộ WMS), phi chức năng (độ chính xác $\geq 90\%$, thời gian xử lý < 1 giây, giao diện thân thiện), và phạm vi dự án.
 - Hoạt động cụ thể:
 - Phân tích ảnh kệ hàng có sẵn để xác định các vấn đề như sai sót kiểm đếm, thiếu hàng, và điều kiện ánh sáng/kệ hàng.
 - Khảo sát yêu cầu WMS từ nhân viên kho để xác định nhu cầu đồng bộ tồn kho.
 - Sử dụng Use Case Diagram và MoSCoW prioritization để lập danh sách yêu cầu, ưu tiên tính năng cốt lõi như nhận diện sản phẩm bằng YOLOv11 và tích hợp WMS.
 - Lợi ích: Cung cấp kế hoạch tổng thể, đảm bảo yêu cầu rõ ràng trước khi phát triển, đặc biệt khi sử dụng dữ liệu có sẵn và mô phỏng qua điện thoại.
- Agile (Linh hoạt):
 - Ứng dụng: Áp dụng ở các giai đoạn thiết kế, phát triển, kiểm thử, và tinh chỉnh, chia công việc thành sprint ngắn để tối ưu hóa YOLOv11, ứng dụng di động, IoT, và WMS.

- Hoạt động cụ thể:
 - Sprint YOLOv11: Huấn luyện mô hình cơ bản, cải thiện độ chính xác trong ánh sáng yếu, tối ưu tốc độ trên Raspberry Pi bằng ONNX/TensorRT.
 - Sprint ứng dụng di động: Tinh chỉnh giao diện Flutter, tích hợp Firebase, và kiểm tra tính thân thiện qua Usability Testing.
 - Sprint IoT/WMS: Tích hợp cảm biến IoT (camera/điện thoại) và đồng bộ dữ liệu với WMS qua RESTful API.
 - Sprint kiểm thử: Mô phỏng quét hình ảnh qua điện thoại để kiểm tra YOLOv11, IoT, và WMS, điều chỉnh dựa trên các biến số ánh sáng và kệ hàng.
- Lợi ích: Agile cho phép kiểm tra lặp, cải tiến YOLOv11 (độ chính xác từ 85% lên $\geq 90\%$), và tích hợp phản hồi người dùng trong thời gian ngắn.

Nghiên cứu lý thuyết và thực nghiệm mô phỏng:

- Nghiên cứu lý thuyết:
 - Phân tích các thuật toán thị giác máy tính (YOLO, SSD, Faster R-CNN) để chọn giải pháp tối ưu.
 - Lựa chọn YOLOv11:

Lý do chọn: YOLOv11 (Ultralytics, tháng 9/2024) vượt trội với mAP 54.7% trên COCO (so với 50.7% của YOLOv5x), suy luận 13.5ms (~74 FPS), và giảm 22% tham số, phù hợp cho Raspberry Pi. Backbone hybrid CNN-transformer và head anchor-free cải thiện nhận diện trong ánh sáng yếu và kệ hàng lộn xộn.

- So sánh: SSD (mAP ~40%) và Faster R-CNN (5-10 FPS) kém hơn về tốc độ và độ chính xác so với YOLOv11 (~30 FPS trên Raspberry Pi).
 - Tham khảo tài liệu từ IEEE Xplore và ArXiv về ứng dụng YOLO và WMS trong bán lẻ.
- Thực nghiệm mô phỏng:
 - Sử dụng 10.000 ảnh có sẵn và mô phỏng quét hình ảnh qua điện thoại để kiểm tra hiệu suất YOLOv11, IoT, và WMS trong các

điều kiện ánh sáng (sáng, tối, huỳnh quang) và trạng thái kệ (đầy, gần hết hàng).

- Triển khai kiểm thử mô phỏng tại môi trường giả lập để đánh giá độ chính xác và tốc độ, thay vì thử nghiệm thực tế tại cửa hàng.

3.1.1. Lý do lựa chọn phương pháp

Phương pháp Agile-Waterfall hybrid được chọn vì phù hợp với đặc thù của dự án, đáp ứng yêu cầu kỹ thuật phức tạp, thời gian hạn chế, và môi trường mô phỏng.

- Đặc thù của dự án và sự phù hợp của phương pháp hybrid:
 - Tính phức tạp kỹ thuật:
 - Phát triển mô hình YOLOv11n (Nano) để nhận diện trên 50 loại sản phẩm đòi hỏi thử nghiệm lặp để tinh chỉnh hyperparameters và xử lý dữ liệu mô phỏng. YOLOv11 với backbone hybrid CNN-transformer đảm bảo độ chính xác $\geq 90\%$ trong điều kiện ánh sáng đa dạng và kệ hàng lộn xộn, nhưng cần nhiều lần tối ưu hóa.
 - Ứng dụng di động (Flutter, Firebase) cần điều chỉnh giao diện và tính năng (thông báo thiếu hàng, đồng bộ WMS) dựa trên phản hồi từ Usability Testing.
 - Thiết kế IoT (camera/điện thoại, mạng MQTT) và tích hợp WMS yêu cầu kiểm thử lặp để đảm bảo đồng bộ thời gian thực.
 - Sự phù hợp: Agile chia công việc thành sprint 1 tuần để thử nghiệm và tối ưu YOLOv11 (ví dụ: sprint 1 huấn luyện, sprint 2 cải thiện ánh sáng yếu), cập nhật ứng dụng, và tích hợp IoT/WMS nhanh chóng.
 - Yêu cầu rõ ràng ban đầu:
 - Cần xác định danh sách sản phẩm (các thùng hàng), yêu cầu chức năng (nhận diện, đếm, đồng bộ WMS), và phi chức năng (độ chính xác $\geq 90\%$, tốc độ < 1 giây, giao diện thân thiện) trước khi phát triển.

- Sự phù hợp: Waterfall đảm bảo phân tích dữ liệu có sẵn, khảo sát thực địa, và lập danh sách yêu cầu (Use Case Diagram, MoSCoW), giảm rủi ro bỏ sót mục tiêu. Ví dụ, yêu cầu “hệ thống nhận diện trong ánh sáng yếu” được xác định rõ trước khi huấn luyện YOLOv11.
- Thử nghiệm mô phỏng và thực tế:
 - Hệ thống được kiểm thử qua mô phỏng quét hình ảnh trên điện thoại và tại cửa hàng bán lẻ, với các biến số ánh sáng và cách sắp xếp kệ.
 - Sự phù hợp: Agile hỗ trợ kiểm thử lặp, điều chỉnh YOLOv11, ứng dụng, và WMS dựa trên kết quả mô phỏng. Waterfall xác định kế hoạch kiểm thử (tiêu chí, thời gian, địa điểm). Ví dụ, khi kiểm thử mô phỏng, nhóm phát hiện ứng dụng tải dữ liệu chậm; Agile cho phép tối ưu API RESTful trong sprint.
- Thời gian và nguồn lực hạn chế:
 - Với 1 tháng và 8 thành viên, nhóm cần ưu tiên nhiệm vụ và xử lý dữ liệu có sẵn.
 - Sự phù hợp: Waterfall lập kế hoạch chặt chẽ, Agile ưu tiên nhiệm vụ quan trọng (YOLOv11, WMS), sử dụng Data Augmentation để bù đắp hạn chế dữ liệu. Ví dụ, nhóm sử dụng Data Augmentation để tạo dữ liệu ánh sáng yếu, tiết kiệm thời gian.
- Điểm mạnh của phương pháp hybrid:
 - Cân bằng kế hoạch và linh hoạt: Waterfall đảm bảo lộ trình rõ ràng (nghiên cứu, phát triển, kiểm thử), Agile xử lý thách thức kỹ thuật (tối ưu YOLOv11n, tích hợp WMS).
 - Tối ưu hóa nguồn lực: Agile tập trung vào nhiệm vụ ưu tiên, Waterfall giảm lãng phí do thiếu phối hợp. Ví dụ, khi khó thu thập đủ 10.000 ảnh, Data Augmentation trong Agile tiết kiệm thời gian.

- Tăng tính ứng dụng: Agile điều chỉnh YOLOv11 cho ánh sáng yếu và kệ lộn xộn, Waterfall xác định tiêu chí đánh giá (độ chính xác $\geq 90\%$, tốc độ < 1 giây).
- Giảm rủi ro: Agile phát hiện sớm vấn đề (tốc độ chậm trên Raspberry Pi, độ trễ API), Waterfall giải quyết yêu cầu không rõ ràng từ đầu.
- Lý do chọn phương pháp hybrid:
 - Phù hợp với dự án có yêu cầu cố định (nhận diện được trên 50 thùng hàng cùng lúc, đồng bộ WMS) và thử nghiệm lặp (tối ưu YOLOv11, IoT).
 - Ứng phó hạn chế dữ liệu và thời gian bằng Data Augmentation và sprint ngắn.
 - Hiệu quả với mô phỏng qua điện thoại và kiểm thử thực tế tại cửa hàng.
 - Được chứng minh hiệu quả trong các dự án tương tự (hệ thống giám sát, IoT, WMS) theo IEEE Xplore và PMI.

3.2. Nghiên cứu và phân tích yêu cầu

- Mục tiêu: Xác định rõ các yêu cầu để giám sát tồn kho thời gian thực, tích hợp YOLOv11, IoT, và WMS.
- Phương pháp:
 - Phân tích dữ liệu có sẵn: Sử dụng ảnh kệ hàng để xác định các vấn đề như sai sót kiểm đếm, thiếu hàng, và điều kiện ánh sáng (sáng, tối, huỳnh quang).
 - Khảo sát thực địa: Tìm hiểu và xác định nhu cầu WMS (đồng bộ tồn kho, báo cáo thời gian thực) và yêu cầu IoT (camera, truyền dữ liệu).
 - Phân tích tài liệu: Tham khảo báo cáo và bài báo khoa học (IEEE, ArXiv) về ứng dụng thị giác máy tính và WMS trong bán lẻ, đặc biệt với YOLOv11.
 - Phân tích yêu cầu: Sử dụng Use Case Diagram (tương tác nhân viên kho/quản lý) và MoSCoW prioritization:

- Must have: Nhận diện ≥ 50 sản phẩm (thùng hàng), đếm số lượng, đồng bộ WMS thời gian thực.
 - Should have: Thông báo thiếu hàng, báo cáo tồn kho.
 - Could have: Tích hợp với hệ thống ERP.
- Kết quả:
 - Yêu cầu chức năng: Hệ thống nhận diện được nhiều thùng sản phẩm, đếm số lượng, cập nhật dữ liệu lên ứng dụng di động và WMS.
 - Yêu cầu phi chức năng: Độ chính xác nhận diện $\geq 90\%$, thời gian xử lý < 1 giây, ứng dụng thân thiện, hệ thống IoT/WMS ổn định.
- Giải thích:
 - Use Case Diagram mô tả cách người dùng tương tác với hệ thống, đảm bảo mọi chức năng được xác định.
 - MoSCoW prioritization ưu tiên tính năng quan trọng, tránh lãng phí thời gian trong 1 tháng.

3.3. Thiết kế và phát triển hệ thống

3.3.1. Thiết kế cảm biến IoT và cơ sở hạ tầng mạng

- Mục tiêu: Thiết kế hệ thống IoT để thu thập hình ảnh kệ hàng và truyền dữ liệu cho YOLLO v11/WMS.
- Phương pháp:
 - Cảm biến IoT: Sử dụng camera (hoặc điện thoại mô phỏng) để quét hình ảnh, tích hợp với Raspberry Pi để chạy YOLLO v11.
 - Cơ sở hạ tầng mạng: Thiết kế mạng Wi-Fi/4G với giao thức MQTT để truyền dữ liệu từ camera đến server/WMS thời gian thực, đảm bảo độ trễ < 1 giây.
 - Mô phỏng: Sử dụng quét hình ảnh qua điện thoại để giả lập dữ liệu IoT, kiểm tra trong các điều kiện ánh sáng đa dạng và kệ hàng lộn xộn.
- Kết quả mong đợi: Hệ thống IoT cung cấp hình ảnh chất lượng cao, truyền dữ liệu ổn định với độ trễ < 1 giây.

- Giải thích:
 - MQTT là giao thức nhẹ, phù hợp cho truyền tải thời gian thực trên phần cứng nhúng.
 - Mô phỏng qua điện thoại tiết kiệm thời gian, phù hợp với dữ liệu có sẵn và thời gian 1 tháng.

3.3.2. Phát triển mô hình AI/ML (YOLLO v11)

- Mục tiêu: Huấn luyện YOLLO v11 để nhận diện và đếm ≥ 50 thùng hàng có (có thể nhận diện các thùng hàng loại khác) với độ chính xác $\geq 90\%$ và tốc độ ≥ 30 FPS.
- Phương pháp:
 - Thu thập dữ liệu:
 - Sử dụng nhiều hình ảnh kệ hàng có sẵn với các điều kiện ánh sáng (sáng, tối, huỳnh quang), góc chụp (chính diện, nghiêng), và trạng thái kệ (đầy, gần hết hàng).
 - Kiểm tra và gắn nhãn bounding box bằng LabelImg thông qua Crowdsourcing (thuê nhân viên gắn nhãn).
 - Data Augmentation: Áp dụng xoay ảnh, thay đổi độ sáng, thêm nhiễu để tăng tính đa dạng, đảm bảo YOLLO v11 hoạt động ổn định trong môi trường mô phỏng.
 - Xây dựng mô hình:
 - Sử dụng YOLOv11n (Nano, Ultralytics, tháng 9/2024) nhờ độ chính xác cao (mAP 39.5%), tốc độ nhanh (~ 30 FPS trên Raspberry Pi), và hiệu quả tài nguyên.
 - Quy trình huấn luyện:
 - Chia dữ liệu: 80% huấn luyện, 10% xác thực, 10% kiểm thử.

- Huấn luyện trên Google Colab (GPU Tesla T4) với pretrained weights trên COCO, áp dụng Transfer Learning để tăng tốc và cải thiện độ chính xác.
 - Tinh chỉnh hyperparameters (learning rate, batch size) để đạt $mAP@0.5:0.95 \geq 90\%$.
 - Đánh giá: Sử dụng $mAP@0.5:0.95$ và FPS để đảm bảo hiệu suất.
 - Tối ưu hóa: Xuất mô hình sang ONNX/TensorRT để chạy trên Raspberry Pi, đạt tốc độ ≥ 30 FPS.
- Kết quả mong đợi: Mô hình nhận diện ≥ 50 loại sản phẩm (thùng hàng), độ chính xác $\geq 90\%$, tốc độ ≥ 30 FPS.
- Giải thích:
 - YOLOv11n được chọn vì:
 - Độ chính xác cao (mAP 39.5% so với 37.4% của YOLOv5s), suy luận nhanh (13.5ms), và giảm tham số, lý tưởng cho Raspberry Pi.
 - Khả năng xử lý môi trường phức tạp nhờ backbone hybrid CNN-transformer và head anchor-free.
 - Data Augmentation ngăn overfitting, đặc biệt với dữ liệu ánh sáng yếu.
 - Transfer Learning giảm thời gian huấn luyện (36 epoch), phù hợp với 1 tháng.

3.3.3. Phát triển ứng dụng di động

- Mục tiêu: Xây dựng ứng dụng di động để hiển thị và quản lý thông tin tồn kho, đồng bộ với WMS.
- Phương pháp:
 - Thiết kế giao diện (UI/UX):
 - Sử dụng Figma để thiết kế giao diện đơn giản, trực quan với các tính năng: hiển thị số lượng sản phẩm, cảnh báo thiếu hàng, báo cáo tồn kho (theo ngày, tuần).

- Áp dụng Material Design để đảm bảo thân thiện trên Android/iOS.
- Phát triển ứng dụng:
 - Sử dụng Flutter để phát triển đa nền tảng, giảm thời gian và chi phí so với phát triển native.
 - Tích hợp Firebase để:
 - Gửi thông báo đẩy khi sản phẩm sắp hết (dựa trên ngưỡng cài đặt).
 - Quản lý tài khoản người dùng (quản lý kho, nhân viên).
 - Lưu trữ dữ liệu tạm thời trước khi đồng bộ với WMS.
- Kiểm thử giao diện:
 - Thực hiện Usability Testing với các thành viên trong nhóm (Các thành viên thử nghiệm và đánh giá) để thu thập phản hồi về tính dễ sử dụng.
 - Sử dụng A/B Testing để so sánh hai phiên bản giao diện và chọn phiên bản tối ưu.
- Kết quả mong đợi: Ứng dụng chạy mượt trên thiết bị Android/iOS cấu hình trung bình, thời gian tải dữ liệu <2 giây, giao diện thân thiện.
- Giải thích:
 - Flutter tiết kiệm 30-40% thời gian phát triển, phù hợp với 1 tháng.
 - Firebase cung cấp backend đơn giản, hỗ trợ thông báo thời gian thực, giảm công sức xây dựng WMS backend phức tạp.
 - Usability Testing đảm bảo ứng dụng đáp ứng nhu cầu thực tế.

3.3.4. Tích hợp hệ thống WMS và các thành phần

- Mục tiêu: Tích hợp YOLLOv11, IoT, ứng dụng di động, và WMS thành hệ thống hoàn chỉnh, đảm bảo đồng bộ thời gian thực.
- Phương pháp:

- Tích hợp hệ thống:
 - Sử dụng RESTful API để kết nối:
 - YOLOv11 (chạy trên server/Raspberry Pi) gửi dữ liệu nhận diện (loại sản phẩm, số lượng) đến MongoDB.
 - Ứng dụng di động truy xuất dữ liệu từ MongoDB qua API.
 - WMS đồng bộ dữ liệu tồn kho từ MongoDB để cập nhật kho hàng thời gian thực.
 - Áp dụng gRPC nếu cần truyền dữ liệu lớn (ví dụ: video từ camera IoT).
- Đồng bộ WMS:
 - Đảm bảo dữ liệu từ YOLOv11 (số lượng sản phẩm) được đồng bộ với WMS qua API, hỗ trợ quản lý kho (báo cáo, cảnh báo thiếu hàng).
 - Thiết kế luồng dữ liệu: YOLOv11 → MongoDB → WMS → Ứng dụng di động.
- Kiểm thử tích hợp:
 - Unit Testing: Kiểm tra từng module (YOLOv11, ứng dụng Flutter, WMS) bằng PyTest (Python) và Flutter Test.
 - Integration Testing: Kiểm tra đồng bộ dữ liệu giữa YOLOv11, IoT, ứng dụng, và WMS (ví dụ: dữ liệu từ YOLOv11 có khớp với WMS không).
- Kết quả mong đợi: Hệ thống đồng bộ thời gian thực, độ trễ <1 giây, dữ liệu tồn kho chính xác.
- Giải thích:
 - RESTful API là tiêu chuẩn phổ biến, dễ triển khai, phù hợp cho hệ thống nhỏ như dự án này.
 - MongoDB hỗ trợ lưu trữ dữ liệu phi cấu trúc, tích hợp tốt với WMS và ứng dụng.

- Tích hợp WMS tăng tính thực tiễn, đáp ứng nhu cầu quản lý kho bán lẻ.

3.4. Triển khai thử nghiệm

- Mục tiêu: Kiểm tra hệ thống trong môi trường mô phỏng qua quét hình ảnh trên điện thoại và tại một cửa hàng bán lẻ.
- Phương pháp:
 - Field Testing (mô phỏng và thực tế):
 - Mô phỏng: Triển khai trong 1 tuần, sử dụng điện thoại để quét hình ảnh kệ hàng, kiểm tra YOLOv11, IoT, và WMS trong các điều kiện ánh sáng (sáng, tối, huỳnh quang) và kệ hàng (đầy, lộn xộn).
 - Thực tế: Triển khai tại một cửa hàng bán lẻ trong 1 tuần, thu thập dữ liệu thực tế từ camera IoT và phản hồi người dùng.
 - Thu thập dữ liệu: Ghi lại mAP@0.5:0.95, FPS, độ trễ API, và phản hồi từ 10 người dùng (nhân viên kho, quản lý).
 - Điều chỉnh nhanh: Dựa trên kết quả, tối ưu YOLOv11 (nếu độ chính xác thấp), API (nếu độ trễ cao), hoặc ứng dụng (nếu giao diện phức tạp).
- Kết quả mong đợi: Hệ thống ổn định, độ chính xác $\geq 90\%$, thời gian phản hồi < 1 giây, phản hồi tích cực từ $\geq 80\%$ người dùng.
- Giải thích:
 - Mô phỏng qua điện thoại tiết kiệm thời gian, phù hợp với dữ liệu có sẵn và thời gian 1 tháng.
 - Kiểm thử thực tế tại cửa hàng phát hiện các vấn đề như ánh sáng yếu, tận dụng khả năng nhận diện của YOLOv11 để điều chỉnh.

3.5. Đánh giá và tinh chỉnh

- Mục tiêu: Đánh giá hiệu suất hệ thống và tinh chỉnh để đạt mục tiêu.
- Phương pháp:
 - Đánh giá:

- So sánh kết quả kiểm thử với tiêu chí: độ chính xác $\geq 90\%$, thời gian xử lý < 1 giây, thời gian tải ứng dụng < 2 giây, phản hồi tích cực $\geq 80\%$.
- Phân tích mAP, FPS, độ trễ API, và phản hồi người dùng từ Field Testing.
- Tinh chỉnh:
 - YOLOv11: Bổ sung dữ liệu ánh sáng yếu hoặc tinh chỉnh hyperparameters nếu độ chính xác $< 90\%$.
 - Ứng dụng di động: Cải thiện giao diện dựa trên phản hồi Usability Testing.
 - API/WMS: Tối ưu RESTful API nếu độ trễ cao, đảm bảo đồng bộ WMS ổn định.
 - IoT: Điều chỉnh cấu hình MQTT hoặc camera nếu truyền dữ liệu không ổn định.
- Báo cáo: Tổng hợp kết quả kiểm thử, chuẩn bị tài liệu báo cáo cuối.
- Kết quả mong đợi: Hệ thống đạt mọi tiêu chí, sẵn sàng triển khai thực tế.
- Giải thích:
 - Tinh chỉnh nhanh qua Agile đảm bảo hoàn thành trong tuần cuối.
 - Waterfall cung cấp tiêu chí đánh giá rõ ràng, làm cơ sở cho báo cáo.

3.6. Công cụ và công nghệ sử dụng

3.6.1. Công nghệ thị giác máy tính

- OpenCV:
 - Chức năng: Xử lý hình ảnh cơ bản (lọc nhiễu bằng Gaussian Blur, thay đổi kích thước, chuyển đổi RGB sang grayscale).
 - Ứng dụng: Tiền xử lý nhiều hình ảnh (khoảng 5000 - 8000 hình ảnh) kế hàng để chuẩn hóa đầu vào cho YOLOv11, đảm bảo chất lượng hình ảnh đồng đều.
 - Lý do chọn: Mã nguồn mở, hiệu quả, không yêu cầu phần cứng mạnh.

- TensorFlow:
 - Chức năng: Hỗ trợ huấn luyện, tối ưu, và triển khai mô hình học sâu, xuất mô hình sang ONNX/TensorRT.
 - Ứng dụng: Tùy chỉnh YOLOv11 (nếu cần thêm lớp convolution) và triển khai trên Raspberry Pi.
 - Lý do chọn: Linh hoạt, phù hợp cho phần cứng nhúng, tích hợp tốt với YOLOv11.
- YOLOv11:
 - Chức năng: Nhận diện và đếm sản phẩm thời gian thực, hỗ trợ nhiều tác vụ (object detection, segmentation, classification).
 - Kiến trúc:
 - Backbone: Hybrid CNN-transformer, trích xuất đặc trưng hiệu quả cho đối tượng nhỏ và môi trường phức tạp.
 - Neck: Cải tiến từ YOLOv8, giảm 22% tham số, tăng khả năng tổng hợp đặc trưng.
 - Head: Anchor-free, dự đoán trực tiếp bounding box, đơn giản hóa huấn luyện.
 - Hiệu suất:
 - Độ chính xác: mAP 54.7% (YOLOv11x), 39.5% (YOLOv11n) trên COCO, vượt YOLOv5x (50.7%) và YOLOv8x (53.9%).
 - Tốc độ: Suy luận 13.5ms (~74 FPS) trên GPU, ~30 FPS trên Raspberry Pi (YOLOv11n, ONNX/TensorRT).
 - Huấn luyện: 36 epoch đạt box loss 0.24, nhanh hơn YOLOv5 (100-200 epoch).
 - Ứng dụng:
 - Nhận diện ≥ 50 sản phẩm, đếm số lượng, gửi dữ liệu qua RESTful API.

- Huấn luyện trên 10.000 ảnh với Transfer Learning, tối ưu cho Raspberry Pi.
- Lý do chọn:
 - Độ chính xác cao, tốc độ thời gian thực, hiệu quả tài nguyên.
 - Xử lý tốt ánh sáng yếu và kệ lộn xộn, lý tưởng cho bán lẻ.
 - Hỗ trợ đa tác vụ, cộng đồng mạnh (Ultralytics, GitHub).

3.6.2. Công nghệ IoT và mạng

- Camera IoT:
 - Chức năng: Thu thập hình ảnh kệ hàng (Camera, hoặc điện thoại mô phỏng).
 - Ứng dụng: Cung cấp dữ liệu cho YOLOv11, tích hợp với Raspberry Pi.
 - Lý do chọn: Độ phân giải cao, chi phí thấp, phù hợp mô phỏng.
- MQTT:
 - Chức năng: Truyền dữ liệu thời gian thực từ camera đến server/WMS.
 - Ứng dụng: Đảm bảo độ trễ <1 giây khi truyền hình ảnh/dữ liệu nhận diện.
 - Lý do chọn: Giao thức nhẹ, phù hợp phần cứng nhúng.
- Wi-Fi/4G:
 - Chức năng: Kết nối mạng ổn định cho IoT.
 - Ứng dụng: Hỗ trợ truyền dữ liệu từ camera đến MongoDB/WMS.
 - Lý do chọn: Phổ biến, dễ triển khai, đáp ứng thời gian thực.

3.6.3. Công nghệ phát triển ứng dụng di động

- Flutter:
 - Chức năng: Phát triển ứng dụng đa nền tảng (Android, iOS).
 - Ứng dụng: Hiển thị số lượng sản phẩm, cảnh báo thiếu hàng, báo cáo tồn kho, thiết kế bằng Figma với Material Design.

- Lý do chọn: Tiết kiệm 30-40% thời gian phát triển so với React Native/native.
- Firebase:
 - Chức năng: Thông báo đẩy, xác thực người dùng, lưu trữ tạm thời.
 - Ứng dụng: Gửi cảnh báo thiếu hàng, quản lý tài khoản, đồng bộ dữ liệu với WMS.
 - Lý do chọn: Backend đơn giản, tích hợp dễ với Flutter, phù hợp thời gian hạn chế.

3.6.4. Cơ sở dữ liệu và hệ thống lưu trữ

- MongoDB:
 - Chức năng: Lưu trữ dữ liệu JSON (tên, mã, số lượng sản phẩm) từ YOLOv11 và WMS.
 - Ứng dụng: Cung cấp dữ liệu cho ứng dụng di động và WMS qua RESTful API.
 - Lý do chọn: Phù hợp dữ liệu phi cấu trúc, dễ mở rộng.
- Amazon S3:
 - Chức năng: Lưu trữ rất nhiều hình ảnh kệ hàng (5000 - 8000) và kết quả xử lý YOLOv11.
 - Ứng dụng: Hỗ trợ huấn luyện và kiểm thử, truy xuất nhanh.
 - Lý do chọn: Chi phí thấp, tích hợp tốt với TensorFlow/API.
- RESTful API:
 - Chức năng: Kết nối YOLOv11, ứng dụng, MongoDB, và WMS.
 - Ứng dụng: Đồng bộ dữ liệu thời gian thực, độ trễ thấp.
 - Lý do chọn: Tiêu chuẩn phổ biến, dễ triển khai, hỗ trợ mở rộng.

3.7. Phân công nhiệm vụ và tiến độ thực hiện

- Phân công nhiệm vụ:

Với 8 thành viên và 1 tháng (tháng 5/2025), nhiệm vụ được chia thành 4 nhóm, tận dụng Agile-Waterfall hybrid:

- Nhóm 1 (Nghiên cứu và dữ liệu, 2 thành viên):

- Nhiệm vụ: Kiểm tra 8000 ảnh có sẵn, gắn nhãn bằng LabelImg, áp dụng Data Augmentation (xoay ảnh, thay đổi ánh sáng). Hỗ trợ Nhóm 4 trong kiểm thử mô phỏng/thực tế.

- Kỹ năng: Phân tích dữ liệu, xử lý hình ảnh, tổ chức công việc.

- Nhóm 2 (Mô hình YOLOv11, 3 thành viên):

- Nhiệm vụ: Huấn luyện YOLOv11n trên Google Colab, tối ưu ONNX/TensorRT cho Raspberry Pi, kiểm tra độ chính xác $\geq 90\%$, tốc độ ≥ 30 FPS. Điều chỉnh mô hình nếu cần (bổ sung dữ liệu ánh sáng yếu).

- Kỹ năng: Python, học sâu, thị giác máy tính.

- Nhóm 3 (Ứng dụng di động, 2 thành viên):

- Nhiệm vụ: Hoàn thiện ứng dụng Flutter, tích hợp Firebase, thực hiện Usability Testing với tất cả thành viên trong nhóm (nếu cần có thể thực hiện thêm), đảm bảo thời gian tải < 2 giây.

- Kỹ năng: Flutter, UI/UX, kiểm thử người dùng.

- Nhóm 4 (Tích hợp và kiểm thử, 1 thành viên):

- Nhiệm vụ: Tích hợp YOLOv11, IoT, ứng dụng, và WMS qua RESTful API. Thực hiện Unit/Integration/Performance Testing, kiểm thử mô phỏng (1 tuần) và thực tế (1 tuần), điều chỉnh hệ thống (API, YOLOv11, WMS).

- Kỹ năng: API, kiểm thử, triển khai thực tế.

- Tiến độ thực hiện:

Tiến độ chia thành 4 sprint (1 tuần/sprint), ưu tiên tích hợp, kiểm thử, và tinh chỉnh:

○ Tuần 1-2 :

- Nhóm 1: Kiểm tra dữ liệu, áp dụng Data Augmentation để cải thiện chất lượng dữ liệu cho YOLOv11.
- Nhóm 2: Hoàn thiện huấn luyện YOLOv11n, tối ưu ONNX/TensorRT, kiểm tra độ chính xác $\geq 90\%$, tốc độ ≥ 30 FPS.
- Nhóm 3: Tinh chỉnh ứng dụng Flutter, tích hợp Firebase, thực hiện Usability Testing.
- Nhóm 4: Bắt đầu tích hợp hệ thống qua RESTful API, thực hiện Unit/Integration Testing, phối hợp với Nhóm 2 và 3.

○ Tuần 3 :

- Nhóm 1: Hỗ trợ Nhóm 4 thu thập dữ liệu từ kiểm thử mô phỏng qua điện thoại, tổng hợp phản hồi.
- Nhóm 2: Kiểm tra YOLOv11 trong mô phỏng, điều chỉnh nếu cần (bổ sung dữ liệu ánh sáng huỳnh quang).
- Nhóm 3: Hoàn thiện ứng dụng dựa trên phản hồi Usability Testing, đảm bảo thời gian tải < 2 giây.
- Nhóm 4: Triển khai kiểm thử mô phỏng (1 tuần), thu thập dữ liệu hiệu suất và phản hồi người dùng.

○ Tuần 4:

- Nhóm 1: Hỗ trợ Nhóm 4 tổng hợp báo cáo dữ liệu kiểm thử, chuẩn bị tài liệu cuối.
- Nhóm 2: Tối ưu YOLOv11 dựa trên kết quả kiểm thử (cải thiện nhận diện kẻ lộn xộn).
- Nhóm 3: Cập nhật ứng dụng, đảm bảo giao diện thân thiện và ổn định.
- Nhóm 4: Triển khai kiểm thử thực tế tại cửa hàng (1 tuần), hoàn thiện Performance Testing, điều chỉnh hệ thống để đạt độ chính xác $\geq 90\%$, thời gian phản hồi < 1 giây.

CHƯƠNG 4: TIỀN ĐỘ DỰ ÁN

4.1. Công Việc Đã Hoàn Thành

4.1.1. Kết Quả Nghiên Cứu Ban Đầu

Trong giai đoạn đầu, nhóm tập trung nghiên cứu để xác định tính khả thi và phạm vi ứng dụng của hệ thống. Các kết quả đạt được bao gồm:

- **Khảo sát thực tế:** Tiến hành khảo sát tại hai cửa hàng bán lẻ thực tế (siêu thị Coopmart) để ghi nhận đặc điểm bố trí kệ hàng, loại sản phẩm, điều kiện ánh sáng, và vị trí lắp đặt camera. Tổng cộng 24 giờ khảo sát đã được thực hiện, thu thập dữ liệu từ kho với hơn 100 loại sản phẩm.
- **Thu thập yêu cầu nghiệp vụ:** Phỏng vấn 5 nhân viên bán hàng và 1 quản lý kho để xác định các chức năng cốt lõi: phát hiện thiếu hàng, nhận diện sai vị trí, giám sát tồn kệ theo thời gian thực.
- **Đề xuất kiến trúc hệ thống:** Xây dựng kiến trúc tổng thể bao gồm:
 - **Camera:** Camera IP độ phân giải tối thiểu 1080p, hỗ trợ truyền dữ liệu thời gian thực.
 - **Server xử lý:** Máy chủ cục bộ hoặc đám mây để chạy mô hình AI.
 - **Mô hình AI:** Sử dụng mô hình thị giác máy tính để nhận diện sản phẩm và trạng thái kệ.
 - **Ứng dụng di động:** Giao diện cho nhân viên, hiển thị cảnh báo và báo cáo.
 - **Cơ sở dữ liệu:** Lưu trữ thông tin sản phẩm, trạng thái kệ, và lịch sử giám sát.
- **Thử nghiệm mô hình AI:** Nghiên cứu và thử nghiệm sơ bộ mô hình YOLO v11 để đánh giá độ phù hợp trong môi trường bán lẻ. Kết quả ban đầu cho thấy mô hình có tiềm năng cao nhưng cần tối ưu hóa thêm.

4.1.2. Thành Tựu Đạt Được

- **Bộ dữ liệu huấn luyện:** Xây dựng bộ dữ liệu gồm 2.000 ảnh chụp từ thực tế, bao gồm các sản phẩm thuộc 50 danh mục khác nhau. Dữ liệu được gắn nhãn

thủ công với các thông tin: loại sản phẩm, vị trí trên kệ và trạng thái (đầy, thiếu, sai vị trí).

- **Huấn luyện mô hình AI:** Mô hình YOLO v11 được huấn luyện trên bộ dữ liệu, đạt độ chính xác 89% trên tập kiểm thử. Khả năng phát hiện khoảng trống kệ đạt độ chính xác 92% trong điều kiện ánh sáng ổn định.
- **Tích hợp hệ thống:** Thử nghiệm tích hợp mô hình AI với luồng dữ liệu video thời gian thực. Tốc độ xử lý đạt 1 giây/frame trên máy chủ nội bộ (GPU NVIDIA RTX 3060), đáp ứng yêu cầu thời gian thực.
- **Báo cáo nội bộ:** Hoàn thành báo cáo sơ bộ trình bày kết quả nghiên cứu và thử nghiệm, được phê duyệt bởi ban quản lý dự án.

4.2. Các Vấn Đề Gặp Phải

- **Hạn chế về môi trường:**
 - Camera bị ảnh hưởng bởi ánh sáng thay đổi (ví dụ: ánh sáng tự nhiên từ cửa sổ hoặc đèn huỳnh quang nhấp nháy), dẫn đến tỷ lệ lỗi nhận diện tăng 15% trong các khung giờ cao điểm.
 - Vị trí camera không tối ưu ở một số kệ thấp, gây ra hiện tượng ảnh bị mờ hoặc lệch góc, làm giảm độ chính xác nhận diện xuống dưới 80%.
- **Nhầm lẫn sản phẩm tương đồng:** Các sản phẩm có bao bì tương tự gây nhầm lẫn, với tỷ lệ lỗi phân loại đạt 12% trên tập kiểm thử.
- **Hiệu suất thiết bị biên:** Khi triển khai mô hình YOLO v11 trên các thiết bị biên (NVIDIA Jetson Nano/Jetson Xavier), hệ thống xử lý mỗi khung hình mất khoảng 2,5 giây, tương đương chưa đến 1 hình mỗi giây. Điều này khiến hình ảnh hiển thị bị giật, chậm, và không đáp ứng được yêu cầu thời gian thực trong giám sát liên tục.
- **Gán nhãn dữ liệu:** Việc gán nhãn thủ công cho 2.000 ảnh mất 120 giờ làm việc, đòi hỏi kiến thức chi tiết về sản phẩm và tốn nhiều nhân lực. Tỷ lệ sai sót trong gán nhãn ước tính khoảng 5%.

- **Thiếu tài liệu tham khảo:** Tài liệu về triển khai thị giác máy tính trong môi trường bán lẻ Việt Nam còn hạn chế, gây khó khăn trong việc so sánh và học hỏi kinh nghiệm.

4.3. Kế Hoạch Điều Chỉnh

Để khắc phục các hạn chế và đảm bảo tiến độ, nhóm đề xuất các phương án sau:

- **Mở rộng dữ liệu huấn luyện:**
 - Tăng bộ dữ liệu lên 5.000 ảnh trong vòng 2 tháng, bao gồm các điều kiện ánh sáng (sáng, tối, nhấp nháy) và góc nhìn khác nhau.
 - Thu thập thêm dữ liệu từ 3 cửa hàng bán lẻ khác để đa dạng hóa bối cảnh.
- **Tối ưu hóa mô hình AI:**
 - Thử nghiệm các mô hình nhẹ hơn như YOLOv11n hoặc YOLOv7-tiny, nhắm đến tốc độ xử lý dưới 0,5 giây/frame trên thiết bị biên.
 - Áp dụng kỹ thuật nén mô hình (model pruning, quantization) để giảm kích thước và tăng hiệu suất.
- **Công cụ gắn nhãn bán tự động:**
 - Phát triển công cụ sử dụng mô hình YOLOv11 đã huấn luyện để đề xuất nhãn sơ bộ, giảm 50% thời gian gắn nhãn thủ công.
 - Tích hợp giao diện người dùng để nhân viên dễ dàng chỉnh sửa nhãn sai.
- **Cải thiện lắp đặt camera:**
 - Thực hiện phân tích thực nghiệm tại 5 vị trí camera khác nhau trên mỗi kệ, chọn vị trí tối ưu dựa trên độ rõ nét và góc bao phủ.
 - Sử dụng camera có tính năng chống rung và điều chỉnh ánh sáng tự động (WDR) để giảm thiểu ảnh hưởng của môi trường.
- **Giải pháp bổ trợ:**
 - Xem xét tích hợp công nghệ RFID hoặc mã QR cho các sản phẩm khó phân biệt bằng hình ảnh, dự kiến triển khai thử nghiệm trong quý tiếp theo.

- Thử nghiệm sử dụng cảm biến hồng ngoại để hỗ trợ phát hiện khoảng trống kệ trong điều kiện ánh sáng yếu.
- **Nâng cấp thiết bị biên:**
 - Đánh giá khả năng sử dụng thiết bị biên mạnh hơn (ví dụ: NVIDIA Jetson Nano) để đảm bảo hiệu suất xử lý.
 - Tối ưu hóa luồng dữ liệu bằng cách giảm độ phân giải video đầu vào mà vẫn đảm bảo độ chính xác nhận diện.

4.4. Đề Xuất Bổ Sung

Để nâng cao hiệu quả và khả năng triển khai của hệ thống giám sát tự động mức độ tồn trên kệ, nhóm đề xuất các giải pháp công nghệ bổ sung như sau:

- Tích hợp học sâu đa nhiệm (Multi-task Learning): Phát triển mô hình AI tích hợp khả năng thực hiện đồng thời nhiều nhiệm vụ, bao gồm phát hiện khoảng trống kệ, nhận diện sai vị trí sản phẩm và phân loại sản phẩm. Điều này giúp giảm thời gian xử lý và tăng độ chính xác tổng thể, nhắm đến cải thiện độ chính xác lên trên 92% trong 3 tháng tới.
- Triển khai công nghệ học tăng cường (Reinforcement Learning): Áp dụng các thuật toán học tăng cường để tối ưu hóa việc điều chỉnh góc camera tự động hoặc chọn khung hình tối ưu trong môi trường ánh sáng thay đổi, nhằm giảm tỷ lệ lỗi nhận diện do góc nhìn hoặc ánh sáng xuống dưới 5%.
- Phát triển hệ thống dự đoán tồn kho: Tích hợp mô hình phân tích thời gian thực dựa trên dữ liệu lịch sử từ hệ thống để dự đoán xu hướng thiếu hàng trước khi xảy ra. Ví dụ, sử dụng các thuật toán như LSTM (Long Short-Term Memory) để dự báo nhu cầu sản phẩm theo giờ hoặc ngày, hỗ trợ quản lý kho hiệu quả hơn.
- Thử nghiệm công nghệ 3D Reconstruction: Sử dụng kỹ thuật tái tạo 3D từ dữ liệu camera để xây dựng mô hình không gian kệ hàng, giúp xác định chính xác vị trí và số lượng sản phẩm, đặc biệt trong các kệ sâu hoặc bố trí phức tạp. Kế hoạch thử nghiệm sẽ bắt đầu với một kệ mẫu trong quý tiếp theo.
- Tối ưu hóa năng lượng cho thiết bị biên: Nghiên cứu và triển khai các giải pháp tiết kiệm năng lượng cho thiết bị biên (ví dụ: sử dụng chế độ ngủ thông minh

hoặc giảm tần suất xử lý khi hệ ổn định), nhằm kéo dài tuổi thọ thiết bị và giảm chi phí vận hành.

- Mở rộng tích hợp với IoT: Kết nối hệ thống giám sát với các thiết bị IoT khác trong cửa hàng (như cảm biến nhiệt độ, độ ẩm, hoặc lưu lượng khách hàng) để thu thập dữ liệu môi trường, từ đó cải thiện khả năng phân tích và điều chỉnh mô hình AI cho phù hợp với bối cảnh thực tế.

CHƯƠNG 5: KẾT QUẢ DỰ KIẾN

Dự án IntelliMango - Hệ thống Quản lý Tồn kho và Nhập hàng Xoài Thông minh được thiết kế để nâng cao hiệu quả quản lý chuỗi cung ứng xoài thông qua các công nghệ tiên tiến như Internet vạn vật (IoT), phân tích dữ liệu thời gian thực, và các thuật toán học máy (AI/ML). Hệ thống dự kiến mang lại những cải tiến đáng kể trong quản lý nhập hàng và tồn kho, giảm tổn thất nông sản, nâng cao hiệu suất vận hành kho, và cung cấp các báo cáo phân tích dữ liệu chi tiết cùng dự báo chính xác. Các công nghệ như YOLO (You Only Look Once) được tích hợp để kiểm tra ngoại quan, đếm số lượng xoài theo thời gian thực, và phát hiện hư hỏng, trong khi các mô hình AI/ML hỗ trợ dự báo nhu cầu và phân tích dữ liệu.

5.1. Hiệu quả trong quản lý nhập hàng và tồn kho

Hệ thống IntelliMango tận dụng công nghệ IoT và phân tích dữ liệu để tối ưu hóa quy trình nhập hàng và quản lý tồn kho xoài, đảm bảo cung ứng kịp thời và giảm thiểu lãng phí. Bằng cách áp dụng các kỹ thuật dự báo nhu cầu và tính toán mức tồn kho tối ưu, hệ thống giúp doanh nghiệp nhập và lưu trữ xoài với số lượng phù hợp, giảm chi phí liên quan đến tồn kho dư thừa và mất doanh thu do thiếu hàng.

5.1.1. Tối ưu hóa mức tồn kho

IntelliMango sử dụng các thuật toán học máy tiên tiến, bao gồm mô hình dự báo chuỗi thời gian như ARIMA và mạng nơ-ron, để phân tích dữ liệu bán hàng lịch sử, xu hướng thị trường, và các yếu tố bên ngoài như thời tiết hoặc sự kiện lễ hội. Các dự báo này cho phép hệ thống xác định lượng xoài cần nhập để đáp ứng nhu cầu mà không dẫn đến tồn kho dư thừa. Ví dụ, nếu hệ thống dự đoán nhu cầu tăng 20% trong dịp Tết, nó sẽ đề xuất tăng đơn hàng nhập trước để đảm bảo đủ hàng. Bằng cách duy trì mức tồn kho tối ưu, IntelliMango giảm thiểu rủi ro hư hỏng do lưu trữ quá lâu và đảm bảo vốn không bị "đóng băng" trong hàng tồn kho không cần thiết.

Hệ thống áp dụng mô hình quản lý tồn kho dành riêng cho hàng hóa dễ hỏng, chẳng hạn như mô hình Newsvendor được điều chỉnh để tính đến thời hạn sử dụng của xoài. Mô hình này cân bằng giữa chi phí đặt hàng, chi phí lưu trữ, và chi phí mất cơ hội do thiếu hàng. Ngoài ra, IntelliMango tính toán mức dự trữ an toàn dựa trên sự biến động của nhu cầu và thời gian giao hàng, giúp giảm thiểu rủi ro gián đoạn chuỗi cung ứng.

5.1.2. Giảm chi phí lưu trữ

Bằng cách tối ưu hóa mức tồn kho, IntelliMango giảm chi phí lưu trữ, bao gồm chi phí thuê kho, điện năng cho hệ thống làm lạnh, và nhân công quản lý hàng tồn kho. Ví dụ, với lượng tồn kho được điều chỉnh chính xác, doanh nghiệp có thể sử dụng không gian kho nhỏ hơn hoặc giảm tần suất sử dụng kho lạnh, dẫn đến tiết kiệm năng lượng đáng kể. Hệ thống cũng giảm chi phí liên quan đến hư hỏng bằng cách đảm bảo xoài được bán trước khi hết hạn sử dụng, từ đó giảm lượng hàng hóa phải tiêu hủy.

IntelliMango hỗ trợ quản lý chi phí thông qua phân tích dữ liệu chi phí lưu trữ theo thời gian thực. Người quản lý có thể xem báo cáo chi tiết về chi phí lưu trữ hàng tháng, xác định các khu vực có thể cắt giảm, chẳng hạn như tối ưu hóa lịch trình sử dụng kho lạnh hoặc giảm thời gian lưu trữ trung bình của mỗi lô hàng.

5.1.3. Cải thiện độ chính xác trong theo dõi tồn kho

Hệ thống sử dụng cảm biến IoT và công nghệ mã vạch/RFID để theo dõi tồn kho thời gian thực, đảm bảo dữ liệu về số lượng và vị trí xoài luôn chính xác. Mỗi lô xoài được gắn thẻ với mã định danh duy nhất, cho phép hệ thống ghi lại thông tin về ngày nhập kho, ngày hết hạn, và điều kiện lưu trữ. Điều này giúp giảm thiểu sai sót trong kiểm kê, chẳng hạn như "hàng ma" (phantom inventory) hoặc hàng bị đặt sai vị trí.

Ví dụ, khi một lô xoài được nhập kho, hệ thống tự động cập nhật cơ sở dữ liệu WMS (Warehouse Management System) và cung cấp thông tin về vị trí lưu trữ cụ thể. Khi nhân viên cần lấy hàng, hệ thống hướng dẫn họ đến vị trí chính xác, giảm thời gian tìm kiếm và nguy cơ lấy nhầm lô hàng. Độ chính xác này không chỉ cải thiện hiệu quả vận hành mà còn tăng cường độ tin cậy của dữ liệu, hỗ trợ ra quyết định tốt hơn.

5.1.4. Tăng cường phối hợp với nhà cung cấp

IntelliMango cải thiện mối quan hệ với nhà cung cấp thông qua việc chia sẻ dữ liệu dự báo nhu cầu và mức tồn kho. Nhà cung cấp có thể truy cập thông tin về nhu cầu dự kiến và lịch trình nhập hàng, giúp họ lập kế hoạch sản xuất và giao hàng hiệu quả hơn. Ví dụ, nếu IntelliMango dự đoán nhu cầu giảm trong một tháng nhất định, nhà cung cấp có thể điều chỉnh sản lượng để tránh dư thừa, từ đó giảm chi phí cho cả hai bên.

Hệ thống đánh giá hiệu suất nhà cung cấp dựa trên các chỉ số như thời gian giao hàng, tỷ lệ lỗi sản phẩm, và khả năng đáp ứng đơn hàng khẩn cấp. Những đánh giá này được trình bày dưới dạng báo cáo chi tiết, giúp doanh nghiệp lựa chọn nhà cung cấp đáng tin cậy và đàm phán các điều khoản hợp đồng tốt hơn. Ngoài ra, IntelliMango hỗ trợ giao tiếp hai chiều, cho phép nhà cung cấp gửi thông tin về tình trạng vụ mùa hoặc các vấn đề logistics, tăng cường sự phối hợp trong chuỗi cung ứng.

Lợi ích	Mô tả
Tối ưu hóa mức tồn kho	Nhập và lưu trữ xoài dựa trên dự báo nhu cầu, giảm thừa/thiếu hàng.
Giảm chi phí lưu trữ	Giảm chi phí thuê kho và bảo quản nhờ tồn kho tối ưu.
Cải thiện độ chính xác	Theo dõi thời gian thực giảm sai sót trong kiểm kê.
Phối hợp nhà cung cấp	Chia sẻ dữ liệu cải thiện giao hàng và chất lượng.

Bảng 8: Lợi ích của quản lý nhập hàng và tồn kho với IntelliMango

5.2. Giảm tổn thất nông sản (xoài)

IntelliMango được thiết kế để giảm đáng kể tổn thất sau thu hoạch của xoài, một vấn đề phổ biến trong chuỗi cung ứng nông sản, với tỷ lệ tổn thất trung bình từ 20-30%. Hệ thống giải quyết các nguyên nhân chính gây tổn thất, bao gồm điều kiện lưu trữ không tối ưu, thu hoạch không đúng thời điểm, và quản lý tồn kho kém hiệu quả, thông qua giám sát IoT, kiểm tra ngoại quan bằng YOLO, và chiến lược quản lý tồn kho.

5.2.1. Giám sát và kiểm soát điều kiện lưu trữ

Hệ thống sử dụng cảm biến IoT để giám sát liên tục các điều kiện lưu trữ, bao gồm nhiệt độ (tối ưu 12-13°C cho xoài), độ ẩm (85-90%), và nồng độ khí ethylene. Các cảm biến này được kết nối với nền tảng trung tâm, nơi dữ liệu được phân tích thời gian thực để phát hiện sai lệch so với điều kiện lý tưởng. Ví dụ, nếu nhiệt độ kho tăng quá cao, hệ thống gửi cảnh báo qua email hoặc ứng dụng di động, đồng thời kích hoạt hệ thống làm lạnh dự phòng nếu có.

IntelliMango ghi lại lịch sử điều kiện lưu trữ cho mỗi lô xoài, cung cấp khả năng truy xuất nguồn gốc chi tiết. Điều này giúp xác định nguyên nhân tổn thất và đảm bảo tuân thủ các tiêu chuẩn chất lượng quốc tế như Global GAP. Bằng cách duy trì điều kiện lưu trữ tối ưu, hệ thống dự kiến kéo dài thời hạn sử dụng của xoài, giảm tỷ lệ hư hỏng từ mức trung bình 30% xuống dưới 5%, tương tự các kết quả đạt được trong các nghiên cứu thực tế tại Việt Nam.

5.2.2. Thu hoạch và xử lý tối ưu

IntelliMango cung cấp thông tin dựa trên dữ liệu về thời điểm thu hoạch tốt nhất dựa trên giống xoài và điều kiện môi trường, giúp giảm tổn thất do thu hoạch không đúng lúc. Theo nghiên cứu, thiếu kiến thức về thời điểm thu hoạch là nguyên nhân chính gây tổn thất sau thu hoạch. Hệ thống đề xuất các phương pháp xử lý sau thu hoạch, như làm sạch và sử dụng lớp phủ bảo vệ, để kéo dài thời gian bảo quản.

5.2.3. Quản lý tồn kho bằng phương pháp FEFO

Hệ thống triển khai chiến lược First-Expired-First-Out (FEFO), đảm bảo các lô xoài có ngày hết hạn sớm nhất được bán trước. Mỗi lô xoài được gắn thẻ với thông tin về ngày thu hoạch và thời hạn sử dụng dự kiến, cho phép hệ thống tự động ưu tiên xuất kho các lô này. Ví dụ, nếu một lô xoài sắp hết hạn trong ba ngày, IntelliMango thông báo cho nhân viên kho và đề xuất chiến lược như giảm giá hoặc ưu tiên phân phối để tránh lãng phí.

IntelliMango sử dụng dự báo nhu cầu để điều chỉnh mức tồn kho, tránh lưu trữ quá nhiều xoài dẫn đến hư hỏng. Bằng cách kết hợp FEFO với dự báo chính xác, hệ thống đảm bảo phần lớn xoài được bán trong tình trạng tươi ngon, tăng tỷ lệ sản phẩm đến tay người tiêu dùng.

5.2.4. Sử dụng YOLO để phát hiện xoài hư hỏng

Hệ thống IntelliMango tích hợp thuật toán YOLO để kiểm tra ngoại quan và phát hiện xoài hư hỏng trong thời gian thực. YOLO là một thuật toán phát hiện đối tượng tiên tiến, có khả năng nhận diện và định vị các đối tượng trong hình ảnh với tốc độ và độ chính xác cao.

Bằng cách huấn luyện mô hình YOLO trên tập dữ liệu hình ảnh xoài tươi và hư hỏng, hệ thống tự động nhận diện và đếm số lượng xoài hư hỏng trong kho. Các camera được lắp đặt trong kho chụp ảnh định kỳ, và mô hình YOLO xử lý hình ảnh để phát hiện xoài có dấu hiệu hư hỏng, như đổi màu hoặc tổn thương bề mặt. Khi phát hiện xoài hư hỏng, hệ thống gửi cảnh báo đến nhân viên kho để loại bỏ kịp thời, ngăn chặn lây lan và giảm tổn thất.

Dựa trên nghiên cứu sử dụng YOLOv7 để phát hiện táo tươi và táo hư hỏng với độ chính xác 83.5%, IntelliMango kỳ vọng đạt độ chính xác tương tự hoặc cao hơn cho xoài sau khi huấn luyện trên tập dữ liệu chuyên biệt. Một nghiên cứu khác về phát hiện khuyết tật xoài bằng mạng nơ-ron tích chập đạt độ chính xác 98.5% , củng cố tính khả thi của phương pháp này.

5.3. Nâng cao hiệu suất vận hành kho

IntelliMango cải thiện hiệu suất vận hành kho bằng cách cung cấp khả năng theo dõi thời gian thực, tối ưu hóa quy trình lấy hàng và đóng gói, sử dụng không gian hiệu quả, và tự động hóa các nhiệm vụ định kỳ. Những cải tiến này tăng năng suất, giảm chi phí, và nâng cao chất lượng dịch vụ.

5.3.1. Theo dõi tồn kho thời gian thực

Hệ thống sử dụng cảm biến IoT và công nghệ mã vạch/RFID để cung cấp thông tin cập nhật về mức tồn kho, vị trí, và trạng thái của từng lô xoài. Điều này cho phép nhân viên kho nhanh chóng định vị hàng hóa, giảm thời gian lấy hàng và nguy cơ lỗi. Ví dụ, khi một đơn hàng yêu cầu 100kg xoài Keitt, hệ thống hiển thị vị trí chính xác của lô hàng phù hợp, cùng thông tin về ngày hết hạn để đảm bảo tuân thủ FEFO.

Khả năng theo dõi thời gian thực hỗ trợ kiểm kê định kỳ, giảm thời gian và công sức cần thiết để đối chiếu số liệu. IntelliMango cung cấp báo cáo kiểm kê tự

động, so sánh dữ liệu hệ thống với số lượng thực tế, giúp phát hiện và khắc phục sai lệch ngay lập tức.

5.3.2. Tự động hóa các nhiệm vụ định kỳ

Hệ thống tự động hóa nhiều nhiệm vụ định kỳ, như tính toán điểm đặt hàng lại, tạo đơn đặt hàng, và giám sát điều kiện kho. Ví dụ, khi mức tồn kho của một giống xoài giảm dưới ngưỡng an toàn, IntelliMango tự động tạo đơn đặt hàng và gửi đến nhà cung cấp được chọn dựa trên hiệu suất trước đó. Điều này giảm công việc thủ công và đảm bảo kho luôn có đủ hàng.

IntelliMango tự động hóa giám sát thiết bị kho, như hệ thống làm lạnh hoặc cảm biến IoT, bằng cách gửi cảnh báo khi cần bảo trì hoặc sửa chữa. Điều này ngăn ngừa sự cố ảnh hưởng đến chất lượng xoài, tăng độ tin cậy của hoạt động kho.

5.3.3. Sử dụng YOLO để đếm số lượng xoài

IntelliMango tích hợp thuật toán YOLO để đếm số lượng xoài trong kho một cách tự động và chính xác. Bằng cách sử dụng camera để chụp ảnh các khu vực lưu trữ xoài, mô hình YOLO phát hiện và đếm số lượng xoài trong mỗi hình ảnh, cập nhật mức tồn kho thời gian thực mà không cần kiểm đếm thủ công.

Ví dụ, hệ thống tự động cập nhật số lượng xoài sau mỗi lần nhập hoặc xuất hàng, đảm bảo dữ liệu tồn kho luôn chính xác. Các nghiên cứu cho thấy YOLO đạt độ chính xác cao trong phát hiện và đếm đối tượng, với mAP (mean Average Precision) thường trên 90% cho các mô hình được huấn luyện tốt. IntelliMango kỳ vọng đạt hiệu quả tương tự trong việc đếm số lượng xoài, giảm thời gian và sai sót so với kiểm kê thủ công.

Tính năng	Lợi ích
Theo dõi thời gian thực	Giảm thời gian lấy hàng và lỗi nhờ khả năng hiển thị tồn kho.

Tối ưu hóa lấy hàng	Tăng tốc độ và độ chính xác trong xử lý đơn hàng.
Sử dụng không gian	Giảm thời gian di chuyển bằng cách bố trí kho hiệu quả.
Tự động hóa	Giảm công việc thủ công, tăng năng suất và tiết kiệm chi phí.
Đếm bằng YOLO	Cập nhật tồn kho chính xác, giảm thời gian kiểm kê.

Bảng 9: Các cải tiến trong vận hành kho với IntelliMango

5.4. Báo cáo phân tích dữ liệu và dự báo

IntelliMango cung cấp bộ công cụ phân tích dữ liệu và dự báo toàn diện, giúp tối ưu hóa chiến lược nhập hàng và tồn kho thông qua báo cáo chi tiết và dự báo chính xác. Các công cụ này hỗ trợ ra quyết định chiến lược, cải thiện hiệu quả hoạt động, và tăng khả năng cạnh tranh.

5.4.1. Dự báo nhu cầu

Hệ thống sử dụng các mô hình học máy tiên tiến, như ARIMA, hồi quy tuyến tính, và mạng nơ-ron (bao gồm LSTM), để dự đoán nhu cầu tương lai dựa trên dữ liệu bán hàng lịch sử, xu hướng thị trường, và các yếu tố bên ngoài như thời tiết, ngày lễ, hoặc chiến dịch khuyến mãi. Ví dụ, nếu dữ liệu cho thấy nhu cầu xoài tăng 20% trong tháng Tết, IntelliMango điều chỉnh dự báo để phản ánh xu hướng này, giúp doanh nghiệp chuẩn bị đủ hàng.

Hệ thống liên tục học hỏi từ dữ liệu mới, cập nhật dự báo để phản ánh các thay đổi bất ngờ trong hành vi người tiêu dùng. Các mô hình được đánh giá bằng các chỉ số như Mean Absolute Error (MAE) hoặc Root Mean Square Error (RMSE) để đảm bảo độ chính xác cao, ngay cả trong thị trường biến động.

5.4.2. Phân tích tồn kho

IntelliMango tạo báo cáo chi tiết về mức tồn kho, tỷ lệ luân chuyển, và hàng tồn kho lâu ngày. Các báo cáo này giúp người quản lý xác định mặt hàng di chuyển chậm, tối ưu hóa mức tồn kho, và giảm chi phí lưu trữ. Ví dụ, nếu một giống xoài có tỷ lệ luân chuyển thấp, hệ thống đề xuất giảm đơn hàng nhập hoặc triển khai chiến dịch khuyến mãi để đẩy nhanh doanh số.

Hệ thống cung cấp phân tích vòng đời tồn kho, cho thấy thời gian trung bình mỗi lô xoài được lưu trữ trước khi bán. Thông tin này giúp doanh nghiệp tối ưu hóa quy trình nhập hàng, giảm thời gian lưu trữ, và giảm nguy cơ hư hỏng.

5.4.3. Bảng điều khiển thời gian thực

IntelliMango cung cấp bảng điều khiển tương tác, hiển thị thời gian thực các chỉ số chính như mức tồn kho, trạng thái đơn hàng, điều kiện kho, và hiệu suất nhà cung cấp. Người dùng tùy chỉnh bảng điều khiển để tập trung vào các chỉ số quan trọng nhất, như nhân viên kho theo dõi vị trí hàng hóa, trong khi quản lý cấp cao xem báo cáo tài chính.

Hệ thống cung cấp cảnh báo tự động, thông báo về các sự kiện quan trọng như mức tồn kho thấp, nhiệt độ kho bất thường, hoặc đơn hàng bị trì hoãn. Các cảnh báo này giúp doanh nghiệp phản ứng nhanh chóng, giảm thiểu rủi ro và cải thiện hiệu quả vận hành.

5.4.4. Phân tích dữ liệu từ YOLO

Dữ liệu từ các phát hiện của YOLO, bao gồm số lượng xoài tươi và hư hỏng, được sử dụng để phân tích xu hướng và dự báo. Hệ thống theo dõi tỷ lệ xoài hư hỏng theo thời gian, xác định các yếu tố ảnh hưởng (như nhiệt độ hoặc độ ẩm), và đề xuất biện pháp cải thiện, như điều chỉnh điều kiện lưu trữ.

Bằng cách kết hợp dữ liệu từ YOLO với dữ liệu môi trường, hệ thống xây dựng mô hình dự báo để dự đoán tỷ lệ hư hỏng trong tương lai, giúp lập kế hoạch bảo quản và xử lý xoài hiệu quả hơn. Dữ liệu đếm số lượng từ YOLO hỗ trợ theo dõi xu hướng tồn kho, cung cấp thông tin cho dự báo nhu cầu và tối ưu hóa mức tồn kho.

CHƯƠNG 6: KẾT LUẬN VÀ ĐỊNH HƯỚNG TIẾP THEO

6.1. Tóm Tắt Tiến Độ và Kết Quả Bước Đầu

Dự án đã đạt được những tiến bộ đáng kể trong giai đoạn đầu tiên. Dự án đã hoàn thành các hoạt động nghiên cứu ban đầu, bao gồm khảo sát thực tế tại hai cửa hàng bán lẻ, thu thập yêu cầu nghiệp vụ, đề xuất kiến trúc hệ thống tích hợp các thành phần: camera, server xử lý, mô hình AI, ứng dụng di động và cơ sở dữ liệu. Một bộ dữ liệu huấn luyện gồm 2.000 ảnh thực tế đã được xây dựng, với các nhãn sản phẩm được gắn thủ công, hỗ trợ huấn luyện mô hình YOLO v11 đạt độ chính xác 89% trong nhận diện sản phẩm và 92% trong phát hiện khoảng trống kệ. Bản thử nghiệm (prototype) của ứng dụng di động đã được phát triển, cho phép giám sát thời gian thực và gửi cảnh báo khi phát hiện thiếu hàng hoặc sai vị trí. Tích hợp thử nghiệm với luồng video thời gian thực đạt tốc độ xử lý 1 giây/frame trên máy chủ nội bộ, đáp ứng yêu cầu về hiệu suất.

Tuy nhiên, dự án cũng đối mặt với một số thách thức, bao gồm hạn chế về điều kiện ánh sáng, nhầm lẫn giữa các sản phẩm có bao bì tương đồng, hiệu suất thấp trên thiết bị biên và khối lượng công việc gắn nhãn dữ liệu lớn. Nhóm đã đề xuất các phương án điều chỉnh, như mở rộng tập dữ liệu, tối ưu hóa mô hình AI, phát triển công cụ gắn nhãn bán tự động và cải thiện phương án lắp đặt camera. Các đề xuất công nghệ bổ sung, bao gồm tích hợp học sâu đa nhiệm, học tăng cường và công nghệ tái tạo 3D, được đưa ra để nâng cao hiệu quả và khả năng mở rộng của hệ thống.

Kết quả bước đầu cho thấy tiềm năng ứng dụng của hệ thống trong việc cải thiện quản lý tồn kho tại các cửa hàng bán lẻ, đồng thời đặt nền tảng cho các nghiên cứu và triển khai tiếp theo trong lĩnh vực thị giác máy tính ứng dụng trong bán lẻ hiện đại.

6.2. Các Bước Tiếp Theo Trong Giai Đoạn Sau

Để đảm bảo dự án đạt được các mục tiêu đề ra, dự án sẽ tập trung vào các hoạt động sau trong giai đoạn tiếp theo:

1. **Mở rộng và đa dạng hóa dữ liệu huấn luyện:** Thu thập thêm 3.000 ảnh từ các cửa hàng bán lẻ khác nhau, bao gồm các điều kiện ánh sáng, góc nhìn, và bố trí

kệ hàng đa dạng. Điều này nhằm tăng khả năng khái quát hóa của mô hình AI, nhắm đến độ chính xác trên 92% trong mọi điều kiện môi trường.

2. **Tối ưu hóa mô hình AI cho thiết bị biên:** Thử nghiệm các mô hình nhẹ hơn (YOLO v8n, YOLOv7-tiny) và áp dụng kỹ thuật nén mô hình (pruning, quantization) để đạt tốc độ xử lý dưới 0,5 giây/frame trên thiết bị biên như Raspberry Pi hoặc NVIDIA Jetson Nano.
3. **Triển khai công cụ gắn nhãn bán tự động:** Hoàn thiện công cụ hỗ trợ gắn nhãn sử dụng kết quả từ mô hình đã huấn luyện, giảm 50% thời gian và sai sót trong quá trình gắn nhãn dữ liệu.
4. **Thử nghiệm hệ thống tại môi trường thực tế:** Triển khai thử nghiệm hệ thống tại một cửa hàng bán lẻ với 5 kệ hàng trong tháng 7/2025, tập trung đánh giá hiệu suất trong điều kiện vận hành thực tế, bao gồm khả năng xử lý thời gian thực và độ chính xác của cảnh báo.
5. **Tích hợp công nghệ bổ trợ:** Khám phá việc sử dụng RFID hoặc mã QR để hỗ trợ nhận diện sản phẩm khó phân biệt bằng hình ảnh, đồng thời thử nghiệm cảm biến hồng ngoại để cải thiện phát hiện khoảng trống kệ trong điều kiện ánh sáng yếu.
6. **Phát triển tính năng dự đoán tồn kho:** Xây dựng mô hình dự báo dựa trên dữ liệu lịch sử, sử dụng các thuật toán như LSTM để dự đoán xu hướng thiếu hàng, hỗ trợ quản lý kho chủ động.
7. **Đánh giá và cải tiến liên tục:** Thiết lập quy trình thu thập phản hồi từ người dùng (nhân viên bán hàng, quản lý kho) để điều chỉnh giao diện ứng dụng và chức năng hệ thống, đảm bảo tính thân thiện và hiệu quả.