PHÁT TRIỂN HỆ THÔNG NHẬN DIỆN SẢN PHẨM THƯƠNG MẠI DỰA TRÊN CƠ CHẾ PHÁT HIỆN VÀ PHÂN LOẠI HAI GIAI ĐOẠN SỬ DỤNG YOLO VÀ EFFICIENTNET

PHẠM NGỌC CƯỜNG[1], LÊ PHÁT HIỆN[1], TÁI THANH TUẤN[2], NGUYỄN TRỌNG ĐĂNG KHOA[2], LỮ THỊ CẨM TÚ[2,]*

[1] Bộ môn Công nghệ Thông tin, Trường Cao đẳng FPT Polytechnic, Trường Đại học FPT

[2] Khoa Công nghệ Thông tin, Trường Đại học Công Thương TP.Hồ Chí Minh

[*]email: cuongpn11@fpt.edu.vn, hienlp2@fpt.edu.vn, 1001250022@huit.edu.vn, 1001250012@huit.edu.vn, tultc@huit.edu.vn

TÓM TẮT

Nghiên cứu này đề xuất một hệ thống nhận diện sản phẩm thương mại tự động nhằm khắc phục những hạn chế của phương pháp quét mã vạch truyền thống trong môi trường bán lẻ tại Việt Nam. Thách thức chính xuất phát từ sự tương đồng cao về hình dạng giữa các loại đồ uống nội địa, hiện tượng che khuất sản phẩm và sự thiếu hụt dữ liệu gán nhãn chuẩn hóa. Để giải quyết, nhóm nghiên cứu xây dựng pipeline hai giai đoạn kết hợp mô hình YOLOv8n cho nhiệm vụ phát hiện đa đối tượng và EfficientNet-B0 làm backbone để phân loại chi tiết. Kỹ thuật transfer learning cùng các chiến lược tối ưu hóa như Automatic Mixed Precision (AMP) và label smoothing được áp dụng nhằm nâng cao hiệu quả huấn luyện trong điều kiện dữ liệu và tài nguyên tính toán hạn chế. Kết quả thực nghiệm trên bộ dữ liệu gồm 136 nhãn sản phẩm cho thấy YOLOv8n + EfficientNet-B0 đạt độ chính xác 91.00%, cao hơn 1.77% so với mô hình phân loại đơn lẻ. Kết quả này khẳng định tính khả thi của cơ chế phát hiện và phân loại hai giai đoạn trong việc nâng cao hiệu năng nhận diện sản phẩm có hình thái tương đồng, đồng thời mở ra tiềm năng ứng dụng cho các hệ thống thanh toán thông minh tự chủ công nghệ tại Việt Nam.

Từ khóa: Thị giác máy tính, EfficientNet-B0, Transfer Learning, Nhận diện sản phẩm, Bán lẻ thông minh

1. GIỚI THIỆU

Thị giác máy tính hiện đóng vai trò trọng yếu trong sự phát triển của khoa học dữ liệu và trí tuệ nhân tạo, thu hút sự quan tâm lớn từ cộng đồng nghiên cứu học thuật [1], [2],

[3]. Các thuật toán trong lĩnh vực này đã được triển khai rộng rãi và chứng minh hiệu quả trong nhiều bài toán thực tiễn như xe tự hành, nhận diện biển số xe, theo dõi đối tượng và nhiều bài toán khác. Sự phát triển mạnh mẽ của thị giác máy tính còn thúc đẩy các nghiên cứu liên ngành, tạo ra những bước tiến quan trọng trong y tế, kỹ thuật cơ điện tử và thương mại. Tuy nhiên, dù có tốc độ phát triển mạnh mẽ, việc áp dụng thị giác máy tính vào các lĩnh vực chuyên sâu phục vụ các tác vụ cụ thể vẫn còn tồn tại những khoảng trống nghiên cứu nhất định [1].

Trong lĩnh vực thương mại tại Việt Nam, đặc biệt là tại Thành phố Hồ Chí Minh, quy trình thanh toán thủ công dựa trên mã vạch truyền thống thường gặp nhiều trở ngại do các yếu tố vật lý như nhãn dán bị biến dạng, rách hoặc nhăn. Thách thức này càng trở nên phức tạp đối với nhóm mặt hàng đồ uống do tính nội địa hóa cao, dẫn đến sự tương đồng lớn về hình dáng giữa các loại chai và lon, cùng hiện tượng các sản phẩm che khuất lẫn nhau trong môi trường bán lẻ thực tế. Bên cạnh đó, việc thiếu hụt các tập dữ liệu được gán nhãn chuẩn cho các mặt hàng nội địa cũng gây khó khăn cho việc triển khai các mô hình thị giác máy tính hiện có [1].

Nhằm giải quyết các thách thức trên, nghiên cứu này đề xuất một pipeline nhận diện sản phẩm thương mại tự động, đề xuất sử dụng mô hình YOLOv8n để nhận diện các đối tượng trong điều kiện thực tế và EfficientNet-B0 để phân loại từng đối tượng do YOLOv8n đã nhận diện [4], [5]. Do hạn chế về số lượng dữ liệu gán nhãn cho các sản phẩm tại thị trường Việt Nam, kỹ thuật transfer learning[6] được áp dụng để tối ưu hóa khả năng trích xuất đặc trưng và nhận diện của mô hình. Cách tiếp cận này cho phép hệ thống duy trì độ chính xác cần thiết trong việc phân loại hàng hóa mà không đòi hỏi một tập dữ liệu huấn luyện quá lớn ngay từ đầu [6].

Nghiên cứu không chỉ đóng góp một giải pháp có khả năng mở rộng cho các hệ thống thanh toán thông minh mà còn hỗ trợ các doanh nghiệp nội địa trong việc tự chủ công nghệ, giảm thiểu sự phụ thuộc vào các nguồn dữ liệu và giải pháp từ nước ngoài. Việc ứng dụng thành công mô hình này sẽ góp phần nâng cao trải nghiệm người dùng và mang lại giá trị thực tiễn cho lộ trình phát triển hạ tầng bán hàng tự động tại Việt Nam.

2. PHƯƠNG PHÁP NGHIÊN CỨU

2.1. Tập dữ liệu

Bộ dữ liệu được xây dựng từ các loại đồ uống phổ biến như Coca-Cola, Pepsi, nước suối, sữa… thu thập tại Trung tâm thương mại Aeon Mall Bình Tân (Thành phố Hồ Chí Minh) bằng điện thoại Xiaomi 14T Pro, từ tháng 11/2025 đến tháng 01/2026. Các sản phẩm được đặt trên mặt phẳng nghiêng màu trắng, thả lăn tự do để ghi lại hình ảnh từ nhiều góc nhìn khác nhau, kể cả trong điều kiện ánh sáng đa dạng và tình huống khó như sản phẩm bị móp nhẹ.

Từ 136 video thu thập, nhóm nghiên cứu tách frame để tạo ảnh tĩnh, chuẩn hóa theo hệ màu RGB nhằm giữ đặc trưng bao bì. Các ảnh được xử lý loại bỏ nền và cắt xén để đảm

bảo nhất quán cho huấn luyện. Kết quả cuối cùng là tạo ra 16,752 ảnh thuộc 136 nhãn sản phẩm khác nhau, với trung bình khoảng 120 - 125 ảnh cho mỗi lớp. Việc phân bố số lượng ảnh tương đối đồng đều giữa các lớp giúp bộ dữ liệu duy trì độ cân bằng, hạn chế tình trạng mô hình bị thiên lệch trong quá trình huấn luyện. Các nhãn được phân biệt theo kiểu dáng (chai, lon), thể tích, hãng sản xuất và đặc tính (có đường, không đường…), tạo nên bộ dữ liệu phong phú và sát thực tế.

Bảng 1. Danh sách 10 nhãn sản phẩm trong bộ dữ liệu 136 nhãn sản phẩm

				
7up - Chanh - Chai 390	C2 - Freeze Anh Dao - Chai 455	Fanta - Cam - Chai 390	Boncha - O Long Viet Quat - Chai 450	String - Dau Tay Do - Chai 330
				
CocaCola - Nguyen Ban - Lon 320	CocaCola - Light - Lon 320	CocaCola - Plus - Lon 320	Pepsi - Khong Duong - Lon 320	Pepsi - Nguyen Ban - Lon 320

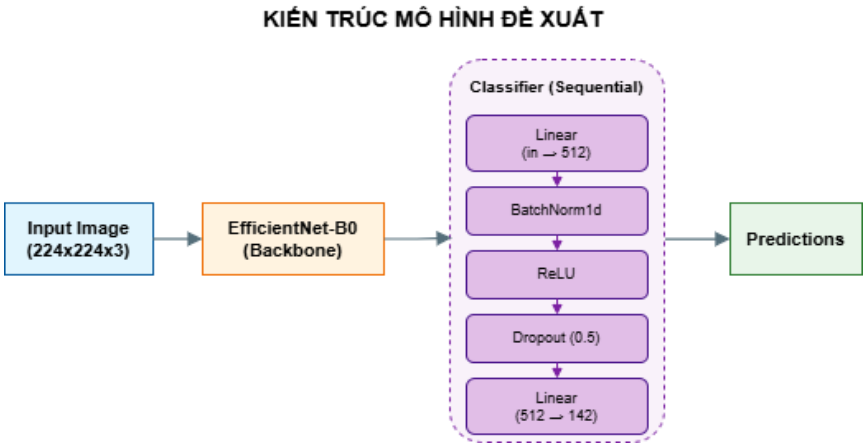
Bảng 2. Tỉ lệ phân chia và số lượng ảnh được sử dụng của bộ dữ liệu

	Train	***Validation***	***Test***
Tỉ lệ phân chia	80%	10%	10%
Tổng số lượng ảnh	13,482	1,636	1,634
Số lượng ảnh trong mỗi lớp	100	12	12

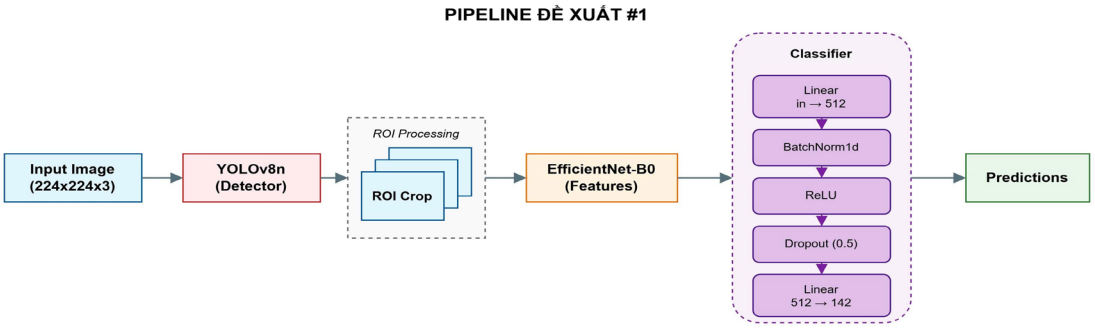
2.2. Mô hình đề xuất

Nghiên cứu áp dụng kỹ thuật transfer learning nhằm khắc phục hạn chế về số lượng mẫu trong tập dữ liệu đồ uống nội địa tại Việt Nam. Bằng cách kế thừa các trọng số đã được huấn luyện sẵn, phương pháp này không chỉ rút ngắn thời gian huấn luyện mà còn giúp hệ thống đạt được độ chính xác cao ngay cả khi điều kiện dữ liệu gán nhãn thực tế còn hạn chế [1].

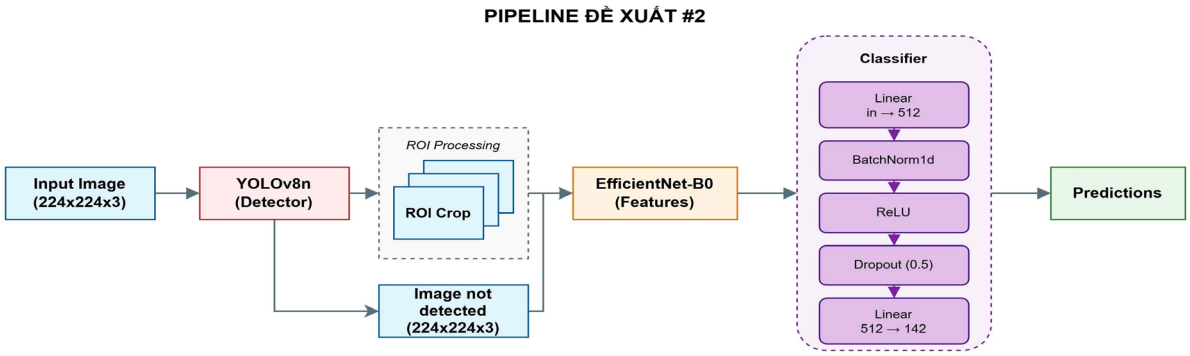
Kiến trúc EfficientNet-B0 được lựa chọn làm mạng backbone nhờ sự tối ưu hóa thông qua cơ chế compound scaling, giúp cân bằng giữa độ sâu, độ rộng của mạng và độ phân giải của hình ảnh đầu vào. Việc sử dụng các khối tích chập tách biệt theo chiều sâu giúp giảm thiểu đáng kể số lượng tham số và khối lượng tính toán nhưng vẫn duy trì được khả năng biểu diễn đặc trưng mạnh mẽ. Điều này đảm bảo mô hình hoạt động hiệu quả trên các thiết bị có tài nguyên phần cứng hạn chế tại các điểm bán hàng trong khi vẫn phân biệt tốt các loại bao bì sản phẩm có thiết kế tương đồng [6].



Hình 1. Kiến trúc mô hình đề xuất sử dụng để phân lớp



Hình 2. Pipeline #1 đề xuất kết hợp YOLOv8n và EfficientNet-B0 để nhận diện sản phẩm



Hình 3. Pipeline #2 đề xuất kết hợp linh hoạt YOLOv8n và EfficientNet-B0 để nhận diện sản phẩm kể cả khi YOLO không phát hiện sản phẩm

Sự kết hợp giữa EfficientNet-B0 và YOLOv8n tạo nên một hệ thống phát hiện và phân loại sản phẩm nhất quán, hoạt động theo cơ chế một giai đoạn như minh họa ở **Hình 2**. Trong cấu trúc này, hình ảnh đầu vào chứa nhiều sản phẩm đồ uống sẽ được YOLOv8n xử lý để xác định vị trí đối tượng. Các vùng chứa sản phẩm sau đó được cắt (ROI Crop) và đưa vào EfficientNet-B0 nhằm thực hiện phân loại chi tiết. Cách tiếp cận này giúp giải quyết hiệu quả tình huống sản phẩm bị che khuất hoặc sắp xếp chồng chéo trong môi trường bán lẻ thực tế, đồng thời đáp ứng yêu cầu xử lý thời gian thực cho hệ thống thanh toán tự động.

Pipeline #1 phát huy hiệu quả khi YOLOv8n nhận diện chính xác đối tượng, trong khi Pipeline #2 (**Hình 3**) được thiết kế nhằm duy trì khả năng phân loại ngay cả khi YOLO không phát hiện được sản phẩm. Cách thiết kế này góp phần nâng cao độ tin cậy của hệ thống trong môi trường bán lẻ phức tạp, nơi dữ liệu hình ảnh thường bị nhiễu hoặc thiếu hụt, đồng thời mở ra tiềm năng ứng dụng cho các hệ thống thanh toán tự động trong tương lai.

Đặc biệt, Pipeline #2 tích hợp cơ chế xử lý ngoại lệ (fallback) để đảm bảo tính ổn định. Trong trường hợp YOLOv8n không phát hiện được đối tượng (không có vùng bao đạt ngưỡng tin cậy), hệ thống sẽ tự động kích hoạt luồng dự phòng: bỏ qua bước ROI Crop và đưa trực tiếp ảnh gốc (224×224×3) vào EfficientNet-B0 để tiến hành phân loại toàn cục. Cơ chế fallback này đóng vai trò quan trọng trong việc xử lý các trường hợp sản phẩm bị biến dạng, che khuất hoặc điều kiện ánh sáng bất lợi khiến YOLO bỏ sót, qua đó tăng cường độ bao phủ và tính ổn định của hệ thống trong môi trường bán lẻ thực tế.

2.3. Môi trường nghiên cứu

Quá trình huấn luyện được cấu hình với các tham số tối ưu nhằm đảm bảo khả năng hội tụ và tính tổng quát hóa của mô hình. Mô hình thực hiện huấn luyện trong 50 epochs với batch size là 32 và learning rate khởi tạo ở mức 1e-3. Hàm mất mát Cross-Entropy Loss kết hợp kỹ thuật label smoothing (0.1) được sử dụng để giảm thiểu sự tự tin thái quá vào các nhãn nhiễu [7]. Thuật toán AdamW với hệ số weight decay 1e-4 và cơ chế cắt bỏ gradient tại ngưỡng 1.0 được lựa chọn để duy trì sự ổn định, tránh hiện tượng bùng nổ gradient và tối ưu hóa trọng số hiệu quả hơn [8].

Để kiểm soát hiệu năng trong suốt quá trình huấn luyện, hai chiến lược điều chỉnh động đã được triển khai. Cơ chế ReduceLROnPlateau thực hiện giảm một nửa learning rate nếu độ chính xác không cải thiện sau 5 epochs liên tiếp, với ngưỡng tối thiểu là 1e-6. Chiến lược này cho phép mô hình tinh chỉnh trọng số ở các vùng không gian hẹp trước khi đạt đến giới hạn hội tụ. Bên cạnh đó, kỹ thuật Early Stopping với tham số patience bằng 5 được sử dụng để tự động chấm dứt quá trình huấn luyện ngay khi mô hình có dấu hiệu đạt ngưỡng bão hòa trên tập kiểm thử, giúp tối ưu hóa tài nguyên tính toán và ngăn chặn hiện tượng học vẹt dữ liệu.

Thêm vào đó nhằm tối ưu hóa tài nguyên phần cứng, nghiên cứu tích hợp kỹ thuật tính toán số thực dấu phẩy động hỗn hợp (Automatic Mixed Precision - AMP), giúp tăng tốc độ

xử lý và tiết kiệm bộ nhớ VRAM nhưng vẫn duy trì độ chính xác của các trọng số. Ngoài ra, phương pháp tích lũy gradient được áp dụng để đảm bảo sự ổn định trong quá trình cập nhật tham số, giúp mô hình đạt được hiệu quả tương đương khi huấn luyện với batch size lớn trong điều kiện tài nguyên giới hạn.

3. KẾT QUẢ THỰC NGHIỆM

3.1. Đánh giá giai đoạn phát hiện đối tượng (YOLOv8n)

Trong thiết kế của pipeline hai giai đoạn, mô hình YOLOv8n được giữ nguyên trọng số huấn luyện trước (pre-trained) nhằm đảm nhiệm vai trò mạng đề xuất vùng (Region Proposal Network). Về mặt độ chính xác phát hiện, nghiên cứu kế thừa các chỉ số nền tảng của kiến trúc YOLOv8n được đánh giá trên tập dữ liệu chuẩn MS COCO (kiểm thử ở kích thước ảnh 640x640). Cụ thể, mô hình đạt độ chính xác trung bình mAP@0.5 là 52.5% và mAP@0.5:0.95 là 37.3% .

Để tối ưu hóa việc trích xuất và cắt (crop) đối tượng, tham số của mô hình được thiết lập với ngưỡng tin cậy (Confidence) là 0.25 và ngưỡng giao nhau (IoU threshold) là 0.60 nhằm phục vụ thuật toán loại bỏ hộp giới hạn trùng lặp (Non-Maximum Suppression). Khi triển khai thực tế trên bộ dữ liệu đồ uống nội địa, hệ thống đạt tốc độ xử lý 12.42 FPS. Các chỉ số này chứng minh YOLOv8n hoàn toàn đáp ứng tốt nhiệm vụ quét và trích xuất vùng ứng viên theo thời gian thực.

3.2. Đánh giá giai đoạn phân loại và phân tích lỗi

Hiệu năng của hệ thống phân loại được đánh giá thông qua các chỉ số Accuracy, Precision, Recall và F1-score. Kết quả so sánh giữa mô hình EfficientNet-B0 độc lập và hai pipeline được trình bày trong **Bảng *3***.

Bảng 3. Kết quả thực nghiệm mô hình EfficientNet-B0 khi không có và khi có sử dụng YOLOv8n để nhận diện sản phẩm

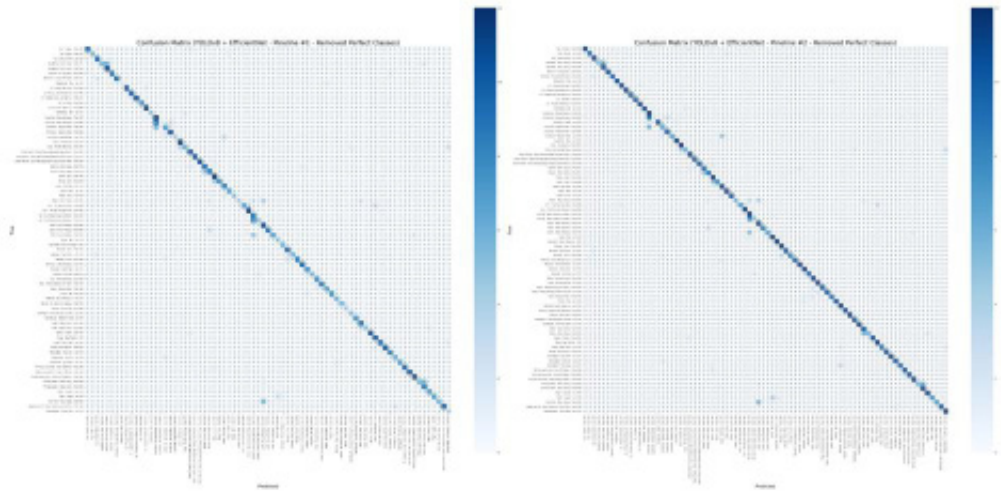
Mô hình / Chỉ số	EfficientNet-B0	YOLOv8n + EfficientNet-B0 (Pipeline #1)	YOLOv8n + EfficientNet-B0 (Pipeline #2)
Accuracy	89.23%	88.75%	91.00%
Macro Avg Precision	93.54%	93.16%	93.84%
Macro Avg Recall	89.26%	89.94%	91.02%
Macro Avg F1-score	88.76%	89.51%	90.66%
Weighted Avg Precision	93.54%	92.61%	93.83%
Weighted Avg Recall	89.23%	88.75%	91.00%
Weighted Avg F1-score	88.73%	88.28%	90.66%

Kết quả cho thấy mô hình EfficientNet-B0 (**Hình 1**) độc lập đạt độ chính xác 89.23%. Khi tích hợp YOLOv8n theo Pipeline #1 (**Hình 2**), độ chính xác giảm nhẹ xuống 88.75% do phụ thuộc hoàn toàn vào bước phát hiện. Ngược lại, Pipeline #2 với cơ chế fallback (**Hình 3**) linh hoạt đạt kết quả tốt nhất với độ chính xác 91.00% và Macro F1-score 90.66%

Ma trận nhầm lẫn rút gọn (**Hình 4**), đã loại bỏ các lớp đạt độ chính xác tuyệt đối, được sử dụng để đánh giá hiệu năng của Pipeline #1 và Pipeline #2. Phân tích ma trận cho thấy các lỗi nhận diện chủ yếu tập trung ở nhóm sản phẩm có mức độ tương đồng cao về màu sắc và kiểu dáng, điển hình là sự nhầm lẫn giữa các loại nước suối đóng chai trong suốt hoặc các dòng bia lon có cùng tông màu chủ đạo. Điều này phản ánh rằng trong một số trường hợp, đặc trưng màu sắc tổng thể có thể lấn át các chi tiết nhỏ trên nhãn mác.

Kết quả thực nghiệm cho thấy phần lớn dự đoán đúng tập trung dọc theo đường chéo chính của ma trận, với số lượng phổ biến từ 8 đến 12 mẫu cho mỗi lớp, chứng tỏ mô hình đạt hiệu suất phân loại khá ấn tượng. Tuy nhiên, vẫn tồn tại những khó khăn trong việc phân biệt các biến thể sản phẩm có đặc trưng thị giác gần nhau. Sự nhầm lẫn xuất hiện chủ yếu ở hai nhóm: (i) nhóm có nhiều biến thể về dung tích: Đây là sai số phổ biến nhất, điển hình ở thương hiệu Aquafina và Dasani. Cụ thể, sản phẩm Dasani - Chai 500ml và Chai 350ml ghi nhận sự phân loại chồng lấn (lần lượt đạt 4 và 5 mẫu đúng, còn lại bị nhầm lẫn với nhau). Điều này cho thấy khi hình dáng chai và nhãn hiệu hoàn toàn tương đồng, việc phân biệt dựa trên kích thước tương đối trong ảnh là một thách thức lớn đối với bộ trích xuất đặc trưng; (ii) nhóm cùng dòng sản phẩm nhưng khác biệt về thành phần/hương vị. Ví dụ, dòng Coca-Cola Nguyên bản - Lon 330ml (11 mẫu đúng) bị nhầm lẫn với Coca-Cola Light và Coca-Cola Zero. Tương tự, các loại Mirinda (Cam, Xá xị) cũng ghi nhận sự phân tán nhẹ giữa các lớp. Nguyên nhân được xác định là do sự thống nhất trong ngôn ngữ thiết kế của hãng, nơi các đặc trưng về logo và màu sắc chủ đạo chiếm diện tích lớn, làm mờ nhạt các chi tiết nhỏ về văn bản định danh hương vị.

Nhìn chung, mặc dù vẫn tồn tại sự nhầm lẫn ở các lớp có quan hệ gần gũi, mô hình vẫn thể hiện khả năng phân biệt rõ ràng giữa các thương hiệu khác nhau. Điều này cho thấy bộ trích xuất đặc trưng đã học được những dấu hiệu nhận dạng cốt lõi của từng thương hiệu, đồng thời gợi mở hướng cải thiện bằng cách tập trung nhiều hơn vào các vùng chứa thông tin về dung tích và nhãn phụ của sản phẩm.



Hình 4. Ma trận nhầm lẫn rút gọn của Pipeline #1 và Pipeline #2

3.3. Kiểm định thống kê

Nhằm tăng độ tin cậy và đáp ứng tính chặt chẽ trong thực nghiệm khoa học, cấu hình tốt nhất (Pipeline #2) được tiến hành huấn luyện và kiểm thử lặp lại 5 lần với các thiết lập khởi tạo ngẫu nhiên (random seed) khác nhau. Kết quả kiểm định thống kê cho thấy độ chính xác của hệ thống đạt trung bình . Biên độ dao động thấp khẳng định tính ổn định cao của cơ chế hai giai đoạn trên bộ dữ liệu thực tế, loại trừ yếu tố sai số ngẫu nhiên trong quá trình hội tụ của mô hình.

4. KẾT LUẬN

Trong nghiên cứu này, chúng tôi đã đề xuất và triển khai hệ thống nhận diện sản phẩm thương mại dựa trên pipeline hai giai đoạn kết hợp YOLOv8n và EfficientNet-B0. Kết quả thực nghiệm cho thấy pipeline #2 đạt độ chính xác 91.00%, cao hơn so với mô hình phân loại đơn lẻ, khẳng định tính khả thi của việc kết hợp phát hiện và phân loại trong môi trường bán lẻ phức tạp. Đây là minh chứng cho hiệu quả của transfer learning và các kỹ thuật tối ưu hóa nhẹ trong điều kiện dữ liệu hạn chế tại Việt Nam. Tuy nhiên, nghiên cứu vẫn tồn tại một số hạn chế. Hệ thống mới chỉ thử nghiệm với EfficientNet-B0 mà chưa mở rộng sang các mô hình tiên tiến hơn như EfficientNetV2, ConvNeXt hay transformer-based. Ngoài ra, các kỹ thuật ensemble hoặc data augmentation nâng cao chưa được áp dụng, khiến khả năng tổng quát hóa của mô hình còn hạn chế. Trong tương lai, việc mở rộng sang các kiến trúc mạnh hơn và tích hợp thêm các kỹ thuật huấn luyện tiên tiến có thể giúp nâng cao độ chính xác và tốc độ xử lý. Nhìn chung, nghiên cứu này đặt nền tảng cho việc phát triển hệ thống thanh toán tự động thông minh, đồng thời hỗ trợ doanh nghiệp nội địa trong quá trình tự chủ công nghệ và nâng cao trải nghiệm người dùng tại thị trường bán lẻ Việt Nam.

5. TÀI LIỆU THAM KHẢO

[1] Y. Wei, S. Tran, S. Xu, B. Kang, and M. Springer, “Deep Learning for Retail Product Recognition: Challenges and Techniques,” *Computational Intelligence and Neuroscience*, vol. 2020, pp. 1–23, Nov. 2020, doi: 10.1155/2020/8875910.

[2] F. Atban, S. E. Guleryuz, Y. E. Kocaoglu, and H. O. Ilhan, “Deep learning based automated non-barcoded product identification system for in-person shopping,” *The European Physical Journal Special Topics*, vol. 234, no. 15, pp. 4269–4284, Oct. 2025, doi: 10.1140/epjs/s11734-025-01775-w.

[3] J. Y. Jeon, S. W. Kang, H. J. Lee, and J. S. Kim, “A Retail Object Classification Method Using Multiple Cameras for Vision-Based Unmanned Kiosks,” *IEEE Sensors Journal*, vol. 22, no. 22, pp. 22200–22209, Nov. 2022, doi: 10.1109/JSEN.2022.3210699.

[4] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You Only Look Once: Unified, Real-Time Object Detection,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016, pp. 779–788. doi: 10.1109/CVPR.2016.91.

[5] M. Tan and Q. V. Le, “EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks,” in *Proceedings of the 36th International Conference on Machine Learning (ICML)*, Long Beach, CA, USA, 2019, vol. 97, pp. 6105-6114.

[6] F. Zhuang *et al.*, "A Comprehensive Survey on Transfer Learning," *Proceedings of the IEEE*, vol. 109, no. 1, pp. 43–76, Jan. 2021, doi: 10.1109/JPROC.2020.3004555.

[7] R. Müller, S. Kornblith, and G. Hinton, "When Does Label Smoothing Help?," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2019, vol. 32.

[8] I. Loshchilov and F. Hutter, "Decoupled Weight Decay Regularization," in *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*, OpenReview.net, 2019. [Online]. Available: https://openreview.net/forum?id=Bkg6RiCqY7

[9] G. Jocher, A. Chaurasia, and J. Qiu, "YOLO by Ultralytics (Version 8.0.0)," 2023. [Online]. Available: https://github.com/ultralytics/ultralytics.