

ỨNG DỤNG TRÍ TUỆ NHÂN TẠO (AI) VÀO DÂY CHUYỀN PHÂN LOẠI HOA QUẢ

ThS. Lê Minh Đức

Trường Đại học Hòa Bình

Tác giả liên hệ: lmduc@daihochoabinh.edu.vn

Ngày nhận: 15/12/2024

Ngày nhận bản sửa: 20/12/2024

Ngày duyệt đăng: 24/12/2024

Tóm tắt

Dây chuyền phân loại hoa quả được sử dụng nhiều ở Việt Nam vẫn còn thủ công, tốn nhiều thời gian và cần nhân công. Bài báo trình bày tổng quan về thiết kế dây chuyền có khả năng phân loại hoa quả tự động bằng việc sử dụng trí tuệ nhân tạo, đáp ứng được tăng tốc độ phân loại quả và có độ chính xác cao, giảm chi phí và cải thiện chất lượng sản phẩm nông nghiệp, đáp ứng các yêu cầu về chi phí đầu tư, tính linh hoạt khi thay đổi loại nông sản theo mùa vụ tại Việt Nam.

Từ khóa: Phân loại nông sản hoa quả, trí tuệ nhân tạo, dây chuyền.

Applying artificial intelligence (ai) in fruit sorting line

MA. Le Minh Duc

Hoa Binh University

Corresponding Author: lmduc@daihochoabinh.edu.vn

Abstract

Fruit sorting lines widely used in Vietnam are still manual, time-consuming, and requires labor. This paper presents an overview of the design of modern lines and equipment capable of automatic fruit sorting by using image processing technology combined with an artificial intelligence network to increase speed, fruit classification and high accuracy. Meet the requirements of investment costs, flexibility in changing the type of agricultural crops according to the season in Vietnam.

Keywords: Fruit and agricultural product sorting, artificial intelligence, production line.

1. Đặt vấn đề

Các dây chuyền phân loại hoa quả phổ biến hiện nay thường được sàng lọc dựa trên kích thước, trọng lượng, màu sắc hoặc bằng nhân công, v.v.. Các dây chuyền này, tuy đã khắc phục được các nhược điểm của phương pháp phân loại sản phẩm truyền thống, nhưng vẫn tồn tại các vấn đề như: Không gian làm việc lớn,

cần nhân công tham gia phân loại, thời gian làm việc lớn, và có sai sót. Việc áp dụng trí tuệ nhân tạo (AI) vào hỗ trợ phân loại trái cây bằng hình ảnh sẽ là bước tiến đưa kinh tế nông nghiệp phát triển, hoàn chỉnh dây chuyền phân loại nông sản tự động hoàn toàn.

Nội dung chính của bài báo sẽ làm rõ những vấn đề sau: Ứng dụng công nghệ xử lý ảnh và

mạng trí tuệ nhân tạo trong kỹ thuật nhận dạng hoa quả, tập trung phân loại các loại hoa quả họ nhà cam quýt. Cùng với đó, thiết kế hệ thống dây chuyền phân loại hoa quả theo kích thước, màu sắc (chín, xanh, thối,...), hình dạng (bị dập, bề mặt bị trầy xước). Dây chuyền phù hợp áp dụng rộng rãi với giá thành rẻ và đơn giản trong quá trình vận hành, thao tác.

2. Thu thập và tiền xử lý dữ liệu

Để có cơ sở dữ liệu cho phần mềm học và phát triển thu thập và tiền xử lý là bước đầu quan trọng không thể thiếu. Dữ liệu ảnh thu chụp từ các loại hoa quả họ nhà cam quýt trên các nguồn dữ liệu mở thường có số lượng ảnh không nhiều,

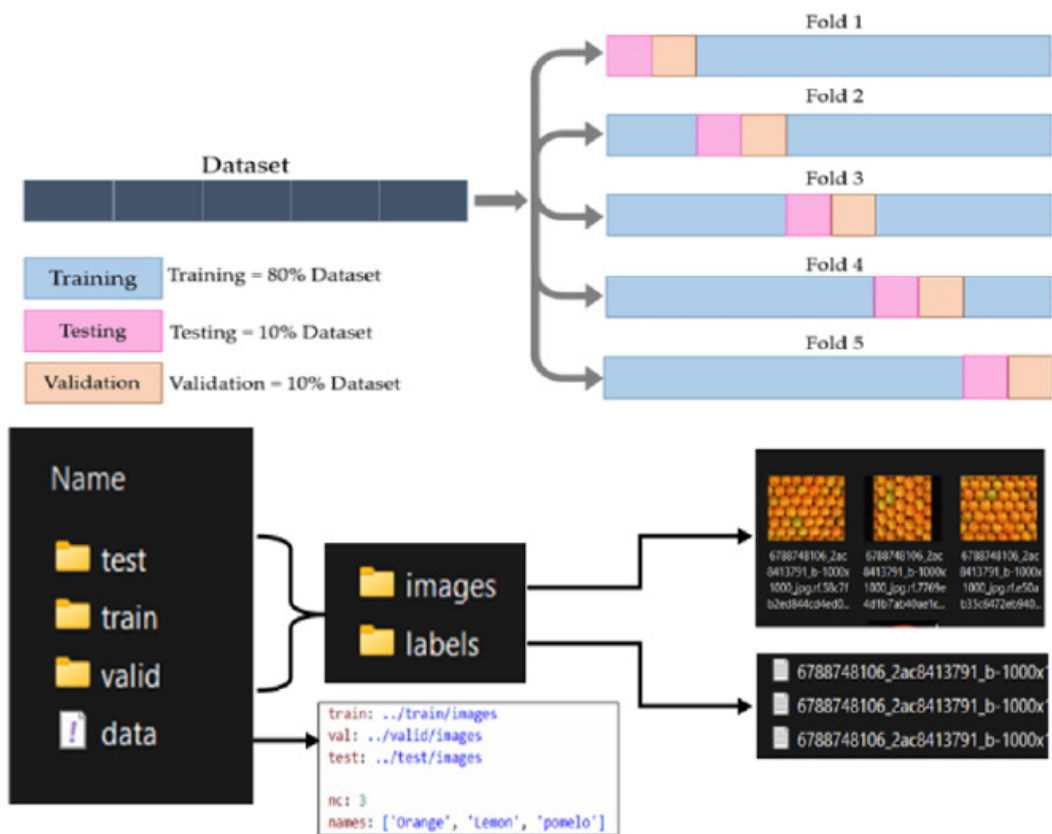
đặc biệt các loại trái cây ở Việt Nam là tương đối hiếm và không được phát hành công khai. Để kiểm nghiệm kết quả nghiên cứu, đề tài tiếp cận thu thập dữ liệu theo hai cách:

Thứ nhất, thực hiện thu thập một số mẫu trên thực tế khi chụp ảnh mẫu vật thực nghiệm tại địa điểm vườn trồng, các nhà máy chế biến và trái cây đã đóng hộp chế biến tại các địa điểm bán.

Thứ hai, để tăng tính đa dạng thực hiện biến đổi dữ liệu qua các phép xoay ảnh, lật ảnh, thay đổi tỉ lệ đối tượng chiếm chỗ, thay đổi nền, bổ sung nhiễu v.v..

Cấu trúc tập dữ liệu được mô tả như sau:

Hình 1. Cấu trúc tập dữ liệu sử dụng cho mô hình



Quá trình tiền xử lý tập dữ liệu bao gồm sắp xếp hình ảnh theo trái cây, loại trừ hình ảnh có độ phân giải thấp và bị hỏng, nhóm hình ảnh thành các lớp, thay đổi kích thước hình ảnh để vừa nền đen với khung vẽ 640×640 pixel và chuẩn hóa. Mạng nơ-ron tích chập (Convolutional Neural Network) yêu cầu hình ảnh đầu vào phải có kích thước đồng nhất, điều này cần thiết để mạng

xử lý chúng một cách hiệu quả và giảm gánh nặng về mặt tính toán. Sự thay đổi lớn về kích cỡ hình ảnh có thể làm tăng tải tính toán trên mạng, làm chậm quá trình đào tạo và suy luận. Sau khi thực hiện thay đổi thành các ảnh có kích thước đồng nhất, ta thực hiện chuẩn hoá hình ảnh. Việc chuẩn hóa các giá trị pixel thành một phạm vi chung (ví dụ: $[0, 1]$ hoặc $[-1, 1]$) giúp

ổn định và tăng tốc độ trả kết quả. Điều này đảm bảo rằng trọng số của mạng được cập nhật thống nhất, ngăn chặn sự bão hòa của các hàm kích hoạt. Cụ thể, chuẩn hóa cũng giúp giảm thiểu tác động của sự khác biệt về ánh sáng và độ tương phản giữa các hình ảnh, làm cho mạng trở nên mạnh mẽ hơn, trước những thay đổi trong dữ liệu đầu vào. Áp đặt các ràng buộc về trọng số

và kích hoạt của mạng, làm cho mô hình có khả năng chống nhiễu tốt hơn.

Sau cùng, quá trình tổng hợp các dữ liệu miêu tả cụ thể tại Bảng 1, các nguồn ảnh tham khảo tại dữ liệu mở nêu trên. Trong đó, số lượng các mẫu có sự tương đồng về số lượng để tránh hiện tượng mất cân bằng dữ liệu, để gây hiện tượng over/under fitting.

Bảng 1. Bộ dữ liệu khi đào tạo

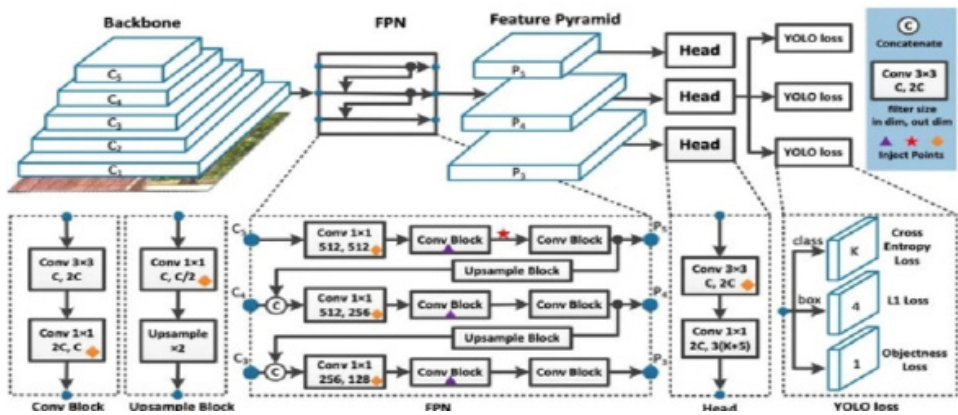
Loại trái cây	Số lượng ảnh trên toàn bộ tập dữ liệu	Tổng
Cam mật Ôn Châu	1226	4927
Cam Canh	1356	
Cam Vinh	2345	
Chanh ta	1344	4024
Chanh không hạt	2045	
Chanh đào	635	
Bưởi năm roi	965	4821
Bưởi da xanh	2645	
Bưởi diễm	1211	
Tổng số lượng ảnh toàn bộ tập dữ liệu		13772

3. Xây dựng chương trình

Trước khi thực hiện học chuyển giao (Transfer Learning), đó là mô hình ban đầu nhận diện họ YOLO (You Only Live Once). Đây là một mô hình học sâu (Deep Learning) phổ biến cho các nhiệm vụ phân loại đối tượng với ưu điểm chính xác và hoạt động

trong thời gian thực. Trong đó, YOLOv8 là phiên bản mới nhất của thuật toán YOLO cải thiện so với các phiên bản trước bằng cách giới thiệu nhiều cải tiến khác nhau về tính năng và ngữ nghĩa môi trường nhằm phát hiện chính xác hơn so với các phiên bản trước đây.

Hình 2. Cấu trúc mô hình YOLOv8



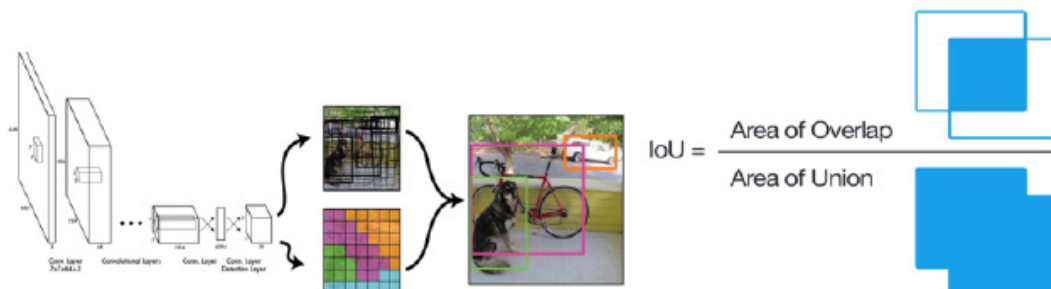
Tương tự các mô hình nhận diện khác như SSD (Single Shot MultiBox Detector), YOLOv8 cũng thực hiện dự đoán xác suất đối

tượng phát hiện trên nhiều bản đồ đặc trưng (Feature map). Những bản đồ đặc trưng ban đầu có kích thước nhỏ giúp dự báo được các

vật thể kích thước lớn. Những bản đồ đặc trưng sau có kích thước lớn hơn trong khi hộp neo (Anchor box) được giữ cố định kích thước nên sẽ giúp dự báo các vật thể kích thước nhỏ. Hộp neo chính là cơ sở để ước lượng các hộp giới hạn (Bounding box) chính xác cho các vật thể. Mặt khác, mô hình cũng dự báo ra rất

nhều hộp giới hạn trên một bức ảnh nên đối với những khung dự đoán có vị trí gần nhau, khả năng bị chồng lấp là rất cao. Trong trường hợp đó, YOLO sẽ cần đến ức chế không tối đa (Non-max suppression) để giảm bớt số lượng các khung hình được sinh ra một cách đáng kể với sơ đồ thuật toán như Hình 3.

Hình 3. Thuật toán Non-max suppression [1]



Hàm mất mát được sử dụng [2] chia thành 2 thành phần: Lloc đo lường sai số của hộp giới hạn và Lcls đo lường sai số của phân phối xác suất các lớp phân loại theo công thức sau:

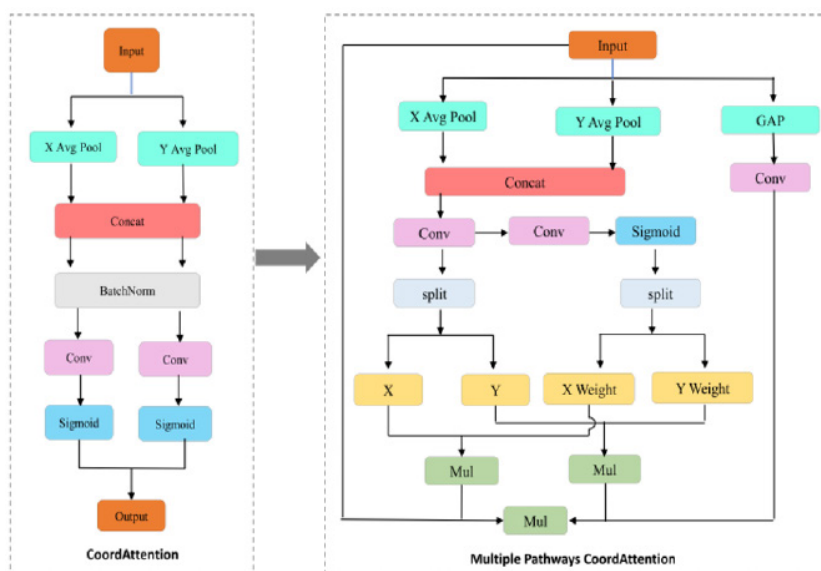
$$L_{loc} = \lambda_{ood} \sum_{i=0}^{S^2} \sum_{j=0}^B K_{ij}^{obj} \left[(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 + (\sqrt{w_i} - \sqrt{\hat{w}_i})^2 + (\sqrt{h_i} - \sqrt{\hat{h}_i})^2 \right]$$

$$L_{cls} = \sum_{i=0}^{S^2} \sum_{j=0}^B \left(K_{ij}^{obj} + \lambda_{noobj} (1 - K_{ij}^{obj}) \right) (C_{ij} - \hat{C}_{ij})^2 + \sum_{i=0}^{S^2} \sum_{c \in C} K_{ij}^{obj} (p_i(c) - \hat{p}_i(c))^2$$

$$L = L_{loc} + L_{cls}$$

Trong quá trình phát hiện trái cây, nhận thấy rằng, bên cạnh các loại quả dễ phát hiện như cam và bưởi thì chanh khó phát hiện hơn và có thể gây nhầm lẫn. Mô hình MPCA (Multiplexed Coordinated Attention) [3] là một cải tiến của mô hình CA (Coordinate Attention), nắm bắt thông tin liên kênh, cũng như thông tin nhận thức vị trí và hướng, để đạt được việc định vị và nhận dạng mục tiêu chính xác hơn.

Hình 4. Thuật toán Non-max suppression [1]

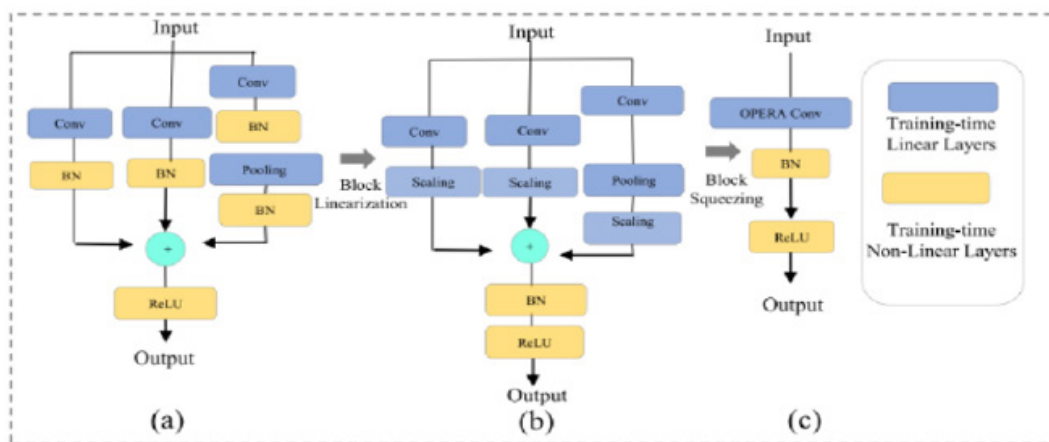


Như minh họa trong Hình 4, cơ chế chú ý MPCA cải tiến khác với CA bao gồm ba kênh, trong đó, giữ nguyên kênh X và Y của CA gốc, và hai kênh này đưa các bản đồ đặc trưng đi qua quá trình trung bình tích lũy theo hướng chiều cao và chiều rộng. Để tăng cường các đặc trưng chi tiết của chiều cao và chiều rộng, chúng tôi thực hiện một bước tích chập tiếp theo, sau đó, phân đoạn qua hàm sigmoid. Các kết quả phân đoạn sau đó được nhân với các kết quả tích chập ban đầu riêng lẻ, do đó, tăng cường trọng số. Nhánh kênh thứ ba được thêm vào, GAP (Global Average Pooling) nhấn mạnh việc thực hiện một lớp giảm mẫu trên toàn bộ thông tin đặc trưng, do đó, bảo tồn nhiều thông tin nền của hình ảnh hơn. Quá trình này của ba kênh ở

các giai đoạn thực hiện song song không chỉ nắm bắt thông tin liên kênh, mà còn thông tin vị trí liên quan đến hướng, giúp mô hình định vị và nhận dạng mục tiêu tốt hơn, đủ linh hoạt hơn.

OREPA (Online Convolutional Re-parameterization) là một cấu trúc mô hình cho việc tái tham số hóa tích chập trực tuyến, sử dụng một quy trình hai giai đoạn để giảm bớt khối lượng đào tạo lớn bằng cách nén các khối phức tạp của thời gian đào tạo thành một lần tích chập duy nhất. Điều này rất cần thiết cho việc đào tạo các mạng nơ-ron tích chập quy mô lớn với các cấu trúc phức tạp và cho phép chúng tôi tái tham số hóa mô hình theo cách tiết kiệm chi phí hơn. Cấu trúc của phương pháp OREPA [3] được hiển thị trong Hình 5.

Hình 5. Cấu trúc OREPA [3] với
(a) Prototype Block, (b) Linearized Block, (c) Training Block

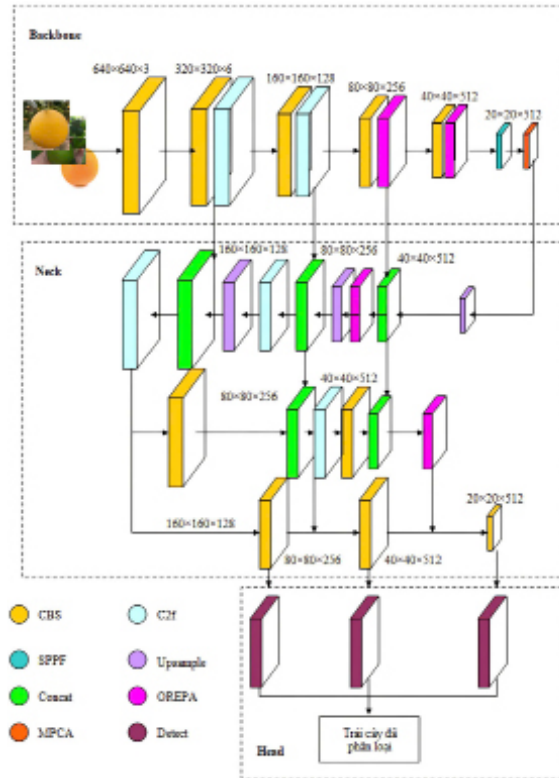


Như được minh họa trong các giai đoạn a đến b của Hình 5, giai đoạn đầu tiên của OREPA được gọi là tuyến tính hóa khối, bao gồm việc loại bỏ tất cả các lớp BN phi tuyến và thêm vào các lớp tỷ lệ tuyến tính. Mặc dù các lớp này chia sẻ các thuộc tính tương tự với lớp BN, chúng giúp tối ưu hóa sự đa dạng giữa các nhánh khác nhau. Từ b đến c cho thấy giai đoạn thứ hai của OREPA được gọi

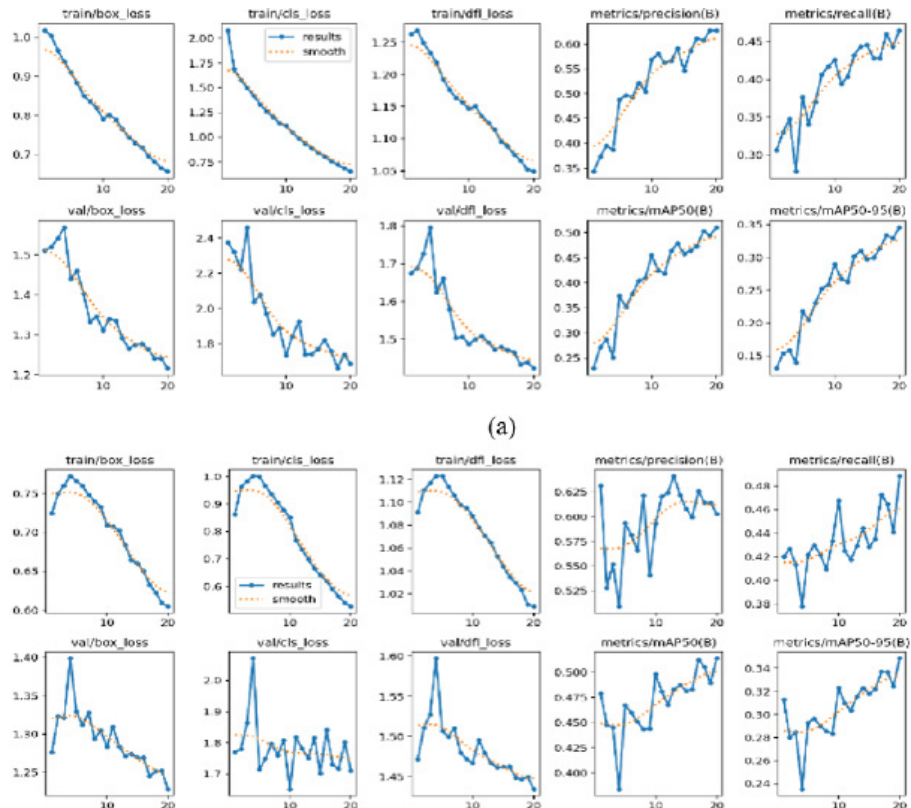
là nén khối, giảm sự phức tạp của các khối tuyến tính nhanh xuống thành một lớp tích chập duy nhất. Giai đoạn này giảm độ phức tạp của mạng trong khi vẫn giữ được thông tin quan trọng. Giai đoạn thứ hai bao gồm việc nén các tích chập nhiều lớp và các tích chập nhiều đường thành một tích chập duy nhất.

Sau cùng, thực hiện phân loại trái cây theo mô hình mô tả chi tiết tại Hình 6.

Hình 6. Mô hình phân loại đề xuất

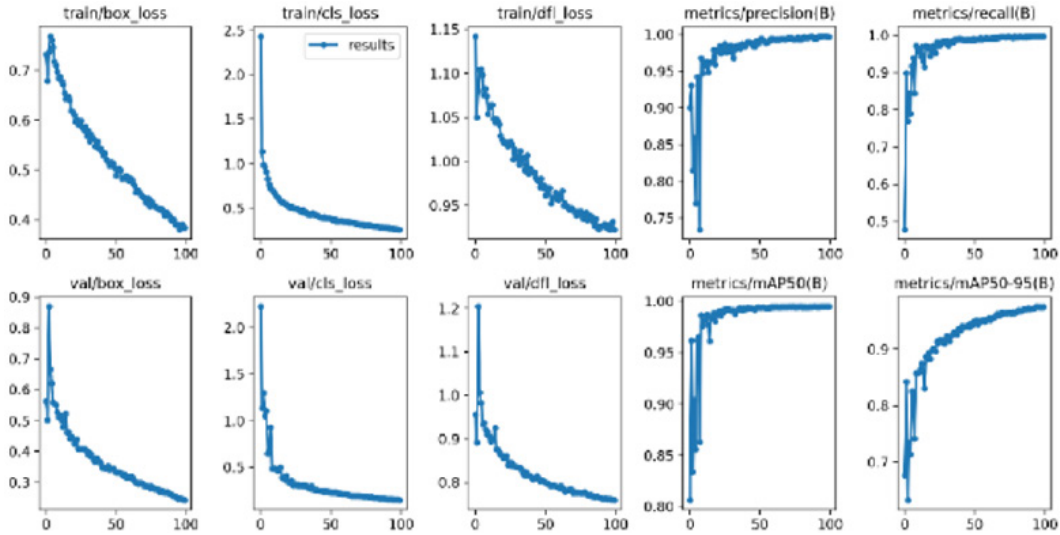


Hình 7. Kết quả theo dõi quá trình đào tạo với (a) kết quả theo dõi 20 epoch đầu tiên và (b) kết quả theo dõi 20 epoch tiếp theo



Sau khi thực hiện đào tạo và gỡ bỏ khóa các lớp trước đó, sử dụng tập dữ liệu huấn luyện để đào tạo toàn bộ mô hình đã được đề xuất, kết quả được thể hiện tại Hình 8.

Hình 8. Kết quả theo dõi quá trình đào tạo mô hình đề xuất



Dựa vào Hình 8 cho thấy, mất mát trong quá trình huấn luyện (Training Loss) và mất mát trong quá trình kiểm tra (Validation Loss) của tập dữ liệu train và valid đều đang hội tụ với sai số rất bé. Sai số giảm tương đồng trên cả 2 tập cho thấy, mô hình đã tránh được hiện tượng quá khớp (Overfitting).

Ngoài ra, để đánh giá khách quan bên cạnh theo dõi kết quả đào tạo qua các lần duyệt, thực hiện tính toán các chỉ số [4] cho toàn bộ quá trình đào tạo.

Độ chính xác của mô hình được tính theo công thức:

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Trong đó, TP là số lượng dự đoán đúng dương, TN là số lượng dự đoán đúng âm, FP số lượng dự đoán lỗi dương tính giả, FN là số lượng dự đoán lỗi âm tính giả.

Độ chính xác cho các dự đoán đúng dương chính xác được mô tả bởi:

$$precision = \frac{TP}{TP + FP}$$

Độ nhạy cho các phát hiện đúng chính xác của mô hình:

$$recall = \frac{TP}{TP + FN}$$

Hiệu suất mô hình được đánh giá qua chỉ số mAP như sau:

$$mAP = \frac{1}{N} \sum_{i=1}^N AP_i$$

Chỉ số mAP khi đạt trên 0.9 đã có thể nhận diện tốt trên thực tế, mô hình của chúng tôi để xuất đã đạt $mAP = 0.99$ (quan sát Hình 8) cho thấy mô hình đã đạt độ chính xác cao trong việc phân loại các đối tượng.

Hình 9. Ảnh huấn luyện đã được gán nhãn

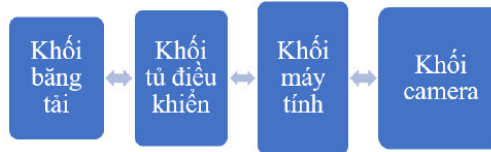


Bảng 2. Hiệu suất mô hình nhận diện

Loại trái cây	precision	recall	mAP50	MAP50-95
Cam	0.960 ± 0.034	0.971 ± 0.090	0.991 ± 0.040	0.985 ± 0.014
Chanh	0.930 ± 0.016	0.942 ± 0.017	0.941 ± 0.060	0.978 ± 0.006
Bưởi	0.910 ± 0.025	0.945 ± 0.011	0.935 ± 0.100	0.969 ± 0.012

4. Xây dựng mô hình dây chuyền

Mô hình gồm bộ điều khiển trung tâm là PLC, khối thu tín hiệu hình ảnh từ bên ngoài là Camera, sau đó, điều khiển hoạt động của các động cơ và hiển thị kết quả trên màn hình.

Hình 10. Sơ đồ khối hệ thống

Chức năng của từng khối:

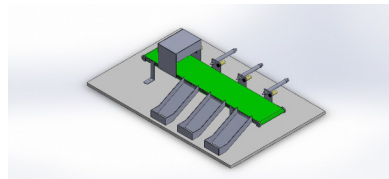
Khối tủ điều khiển (PLC): Có chức năng nhận tín hiệu hình ảnh từ khối camera đã được tiến hành xử lý và phân tích nhận dạng trên khối máy tính, đồng thời, nhận tín hiệu từ các cảm biến của khối băng tải sau đó ra lệnh cho các cơ cấu chấp hành thực hiện việc phân loại và đóng gói theo yêu cầu.

Khối camera: Có chức năng thu thập tín hiệu hình ảnh từ thực tế chuyển về tín hiệu điện và gửi

dữ liệu cho khối máy tính.

Khối máy tính: Có chức năng xử lý và phân tích các dữ liệu hình ảnh được thu thập từ khối camera, kết nối dữ liệu với bộ điều khiển và có chức năng hiển thị kết quả nhận được từ khối tủ điều khiển.

Khối băng tải: Là phần cơ khí của hệ thống nhận nhiệm vụ phân loại và đóng gói hoa quả. Nhận lệnh từ bộ xử lý khối băng tải thực hiện nhiệm vụ của mình.

Hình 11. Xây dựng bản vẽ 3D mô hình dây chuyền phân loại

5. Kết luận

Mô hình đạt độ chính xác cao trên các loại trái cây thuộc họ cam quýt (cam, chanh, bưởi), với chỉ số mAP đạt 0,99. Độ chính xác (precision) và độ nhạy (recall) cũng đạt mức vượt trội, đặc biệt với cam (precision = 0,960; recall = 0,971). Mô hình không chỉ nhận diện tốt các loại trái cây, mà

còn xác định chính xác vị trí của chúng trên dây chuyền. Tốc độ xử lý trung bình đạt 12–20 FPS trên thiết bị hỗ trợ GPU, phù hợp với yêu cầu phân loại theo thời gian thực trong các dây chuyền công nghiệp. Mô hình ứng dụng công nghệ xử lý ảnh và mạng nơ ron AI có thể mở rộng và phát triển để áp dụng vào thực tế cho các loại hoa quả khác.

Tài liệu tham khảo

- [1] Liu, Wei, et al., "Ssd: Single shot multibox detector", *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I* 14. Springer International Publishing, 2016.
- [2] Hussain, Muhammad, "YOLO-v1 to YOLO-v8, the rise of YOLO and its complementary nature toward digital manufacturing and industrial defect detection". *Machines* 11.7 (2023): 677.
- [3] Gai, Rongli, Yong Liu, and Guohui Xu, "TL-YOLOv8: A Blueberry Fruit Detection Algorithm Based on Improved YOLOv8 and Transfer Learning", *IEEE Access* (2024).
- [4] Liu, Qi, Yang Liu, and Da Lin, "Revolutionizing target detection in intelligent traffic systems: Yolov8-snakevision" 12.24 (2023): 4970.