

Nhận Dạng Hoa Quả Thời Gian Thực Sử Dụng YOLOv8

Đoàn Duy Nhật Anh
Khoa Công nghệ Thông tin
Trường Đại học Đại Nam
Hà Nội, Việt Nam
anhddn.1571020003@gmail.com

ThS. Lê Trung Hiếu
Khoa Công nghệ Thông tin
Trường Đại học Đại Nam
Hà Nội, Việt Nam
hieult@dainam.edu.vn

Abstract—Trong bối cảnh nông nghiệp thông minh đang có những bước tiến vượt bậc, việc tự động hóa các quy trình sau thu hoạch là vô cùng quan trọng. Bài báo này trình bày việc phát triển và đánh giá một hệ thống nhận dạng đối tượng thời gian thực có khả năng phân loại 10 loại hoa quả phổ biến bằng mô hình học sâu YOLOv8 (You Only Look Once). Quá trình nghiên cứu bao gồm các giai đoạn chính: thu thập và chuẩn bị một tập dữ liệu hình ảnh đa dạng, gán nhãn tỉ mỉ cho từng đối tượng và huấn luyện mô hình YOLOv8. Kết quả thực nghiệm cho thấy mô hình đạt được độ chính xác trung bình (mAP@0.5) là 79.3%, trong đó một số lớp như 'Táo' đạt độ chính xác gần như tuyệt đối (99.5%). Với tốc độ xử lý nhanh, hệ thống chứng tỏ sự phù hợp cho các ứng dụng thời gian thực. Nghiên cứu này không chỉ khẳng định tiềm năng của YOLO trong việc tối ưu hóa chuỗi cung ứng nông sản mà còn mở đường cho các hệ thống kiểm tra chất lượng và phân loại tự động trong tương lai.

Index Terms—Nhận dạng hoa quả, YOLO, Học sâu, Thị giác máy tính, Nông nghiệp thông minh

I. GIỚI THIỆU

Nông nghiệp vẫn là một ngành kinh tế trọng điểm, và việc ứng dụng công nghệ hiện đại để nâng cao năng suất và chất lượng là một xu hướng tất yếu. Một thách thức lớn trong chuỗi cung ứng nông sản là giai đoạn sau thu hoạch, bao gồm phân loại, kiểm đếm và quản lý kho. Các công việc này thường tốn nhiều nhân công, chi phí cao và dễ xảy ra sai sót do con người.

Sự tiến bộ vượt bậc của Trí tuệ Nhân tạo (AI), đặc biệt trong lĩnh vực Thị giác Máy tính và Học sâu, mang đến những giải pháp đầy hứa hẹn cho vấn đề này. Các hệ thống nhận dạng đối tượng tự động có thể "nhìn" và "hiểu" hình ảnh, cho phép máy móc thực hiện các tác vụ phân loại với tốc độ và độ chính xác vượt trội so với con người.

Trong số các thuật toán nhận dạng đối tượng hiện đại, YOLO (You Only Look Once) nổi bật như một trong những kiến trúc hiệu quả nhất, cân bằng tốt giữa tốc độ và độ chính xác. Khác với các mô hình hai giai đoạn như R-CNN, YOLO xử lý toàn bộ hình ảnh trong một lần duy nhất, khiến nó trở thành lựa chọn lý tưởng cho các ứng dụng yêu cầu phản hồi thời gian thực, chẳng hạn như dây chuyền phân loại tự động hoặc robot nông nghiệp.

Xuất phát từ bối cảnh đó, nghiên cứu này nhằm xây dựng một hệ thống nhận dạng 10 loại hoa quả phổ biến bằng cách tận dụng sức mạnh của kiến trúc YOLOv8. Hệ thống được

thiết kế không chỉ để phân loại chính xác mà còn đảm bảo hiệu suất thời gian thực, tạo tiền đề cho các ứng dụng thực tiễn trong tự động hóa nông nghiệp.

II. CÁC CÔNG TRÌNH LIÊN QUAN

Lĩnh vực nhận dạng hoa quả tự động đã thu hút được sự quan tâm đáng kể của các nhà nghiên cứu trong những năm gần đây. Các phương pháp truyền thống thường dựa trên việc trích xuất các đặc trưng thủ công như màu sắc, kết cấu và hình dạng. Tuy nhiên, các phương pháp này hoạt động kém hiệu quả trong các điều kiện thực tế phức tạp với ánh sáng thay đổi, vật thể bị che khuất hoặc có nhiều vật thể.

Sự ra đời của Mạng Nơ-ron Tích chập (CNN) đã tạo nên một cuộc cách mạng. Các nghiên cứu ban đầu [1] đã sử dụng các kiến trúc như AlexNet và VGG để phân loại hình ảnh hoa quả, đạt độ chính xác cao đối với hình ảnh có một quả duy nhất nhưng thiếu khả năng định vị nhiều đối tượng.

Để giải quyết vấn đề định vị, các thuật toán nhận dạng đối tượng hai giai đoạn như R-CNN và các biến thể của nó (Fast R-CNN, Faster R-CNN) [2] đã được áp dụng. Các mô hình này đầu tiên đề xuất các vùng có khả năng chứa đối tượng và sau đó phân loại từng vùng. Mặc dù có độ chính xác cao, tốc độ xử lý chậm khiến chúng không phù hợp cho các ứng dụng thời gian thực.

Để khắc phục hạn chế này, các mô hình một giai đoạn như SSD (Single Shot MultiBox Detector) và YOLO [3] đã được giới thiệu. YOLO, đặc biệt là các phiên bản sau này như YOLOv5 và YOLOv8 [4], đã chứng tỏ hiệu quả vượt trội. Nó xử lý toàn bộ hình ảnh trong một lượt duy nhất, trực tiếp dự đoán các hộp giới hạn và xác suất lớp, do đó đạt được tốc độ cao trong khi vẫn duy trì độ chính xác cạnh tranh. Nghiên cứu của chúng tôi xây dựng dựa trên những công trình này, tập trung vào việc áp dụng phiên bản YOLOv8 mới nhất để phát triển một hệ thống vừa chính xác vừa nhanh chóng cho các ứng dụng nông nghiệp thực tiễn.

III. PHƯƠNG PHÁP ĐỀ XUẤT

A. Tổng quan hệ thống

Hệ thống được thiết kế theo một quy trình thị giác máy tính tiêu chuẩn, bao gồm bốn giai đoạn chính:

- 1) **Thu thập và Chuẩn bị dữ liệu:** Tìm kiếm hình ảnh của 10 loại hoa quả và gán nhãn bằng các hộp giới hạn.

- 2) **Huấn luyện mô hình:** Sử dụng bộ dữ liệu đã được gán nhãn để huấn luyện mô hình YOLOv8.
- 3) **Đánh giá mô hình:** Đo lường hiệu suất của mô hình đã huấn luyện bằng một tập kiểm thử riêng biệt và các chỉ số tiêu chuẩn.
- 4) **Ứng dụng thời gian thực:** Tích hợp mô hình vào một ứng dụng thực hiện nhận dạng qua webcam.

B. Thu thập và Chuẩn bị dữ liệu

Bộ dữ liệu được thu thập từ các nguồn trực tuyến như Kaggle, Roboflow và Google Images để đảm bảo tính đa dạng. Nó bao gồm 10 lớp hoa quả: Táo, Chuối, Nho, Kiwi, Xoài, Cam, Dứa, Mãng cầu và Dưa hấu. Công cụ LabelImg được sử dụng để gán nhãn. Để nâng cao chất lượng bộ dữ liệu và chống quá khớp, các kỹ thuật tăng cường dữ liệu như xoay, lật và điều chỉnh độ sáng ngẫu nhiên đã được áp dụng trong quá trình huấn luyện.

C. Kiến trúc và Huấn luyện mô hình YOLOv8

Chúng tôi sử dụng mô hình YOLOv8, bao gồm ba thành phần chính:

- **Backbone:** Một mạng dựa trên CSPDarknet53 để trích xuất đặc trưng.
- **Neck:** Một Mạng tổng hợp đường dẫn (PAN-FPN) để hợp nhất các đặc trưng từ các tỷ lệ khác nhau.
- **Head:** Một đầu ra tách rời (decoupled head) để dự đoán riêng biệt hộp giới hạn và nhãn lớp.

Bộ dữ liệu được chia thành 80% để huấn luyện và 20% để kiểm định. Mô hình được huấn luyện trong 100 epochs. Hàm mất mát là sự kết hợp của hàm mất mát phân loại và hàm mất mát hồi quy, được tối ưu hóa bằng thuật toán Adam.

IV. THỰC NGHIỆM VÀ KẾT QUẢ

A. Thiết lập thực nghiệm

Quá trình huấn luyện và kiểm thử được thực hiện trên máy tính cá nhân được trang bị GPU NVIDIA GeForce. Môi trường phần mềm dựa trên Python, PyTorch và framework YOLO của Ultralytics.

B. Phân tích kết quả

Quá trình huấn luyện được theo dõi qua các chỉ số khác nhau, như trong Hình 1. Các đường cong mất mát giảm dần trong khi các đường cong chỉ số (precision, recall, mAP) tăng lên, cho thấy mô hình đã học một cách hiệu quả.

Ma trận nhầm lẫn: Như được mô tả trong Hình 2, mô hình hoạt động cực kỳ tốt trên các lớp như 'Táo' (1.00) và 'Kiwi' (0.91). Tuy nhiên, nó có một số nhầm lẫn giữa 'Chuối' và 'Nho', có thể do sự tương đồng về hình ảnh trong một số trường hợp.

Đường cong Precision-Recall (PR): Đường cong PR (Hình 3) là một chỉ số hiệu suất quan trọng. Mô hình đạt được Độ chính xác Trung bình (mAP) là 0.793 ở ngưỡng IoU là 0.5. 'Táo' có AP cao nhất (0.995), trong khi 'Nho' là lớp thách thức nhất (0.677).

Đường cong F1-Confidence: Hình 4 cho thấy điểm F1 cao nhất là 0.75 đạt được ở ngưỡng tin cậy 0.411, đại diện cho sự cân bằng tối ưu giữa precision và recall để triển khai.



Fig. 1. Các chỉ số huấn luyện và kiểm định qua các epoch. Các đường cong mất mát giảm dần và các đường cong precision/recall/mAP tăng dần cho thấy mô hình học hiệu quả.

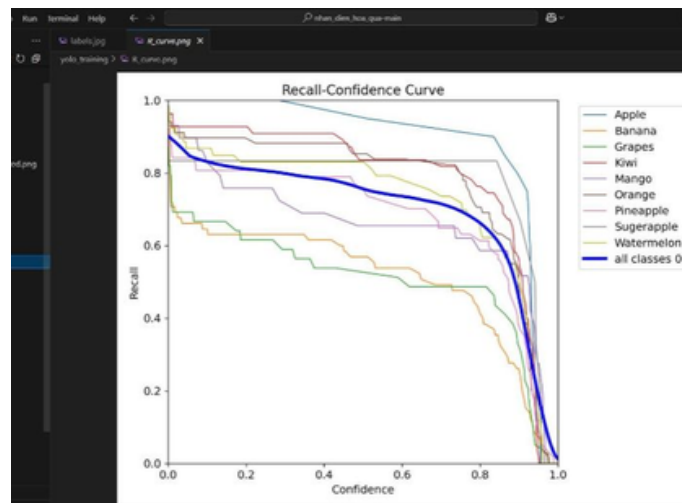


Fig. 2. Ma trận nhầm lẫn đã chuẩn hóa trên tập kiểm thử.

V. KẾT LUẬN

Nghiên cứu này đã chứng minh thành công việc xây dựng và xác thực một hệ thống nhận dạng hoa quả sử dụng YOLOv8, đạt được mAP@0.5 là 79.3

Tuy nhiên, mô hình có những hạn chế, bao gồm độ chính xác không nhất quán giữa các lớp. Công việc trong tương lai sẽ tập trung vào:

- 1) Mở rộng và cân bằng bộ dữ liệu.
- 2) Tinh chỉnh kiến trúc mô hình và các siêu tham số.
- 3) Tích hợp các tính năng nâng cao như phát hiện độ chín.

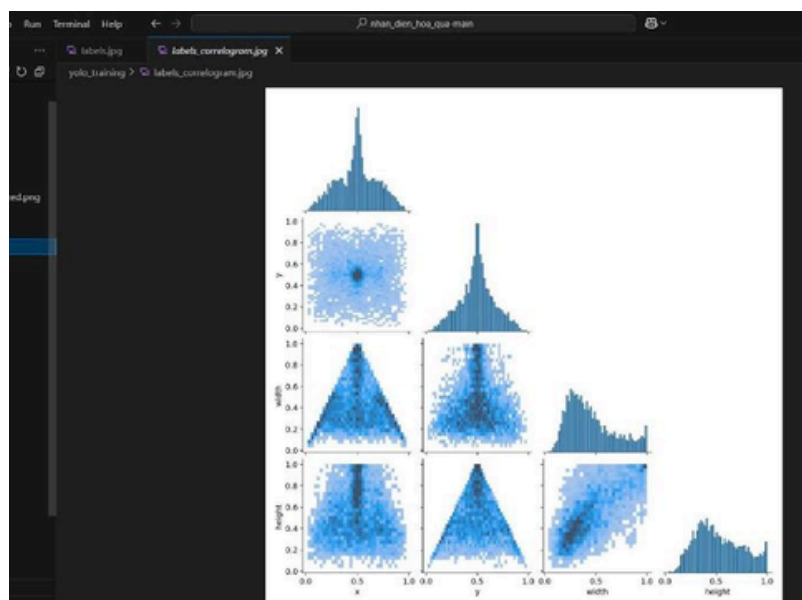


Fig. 3. Đường cong Precision-Recall cho tất cả các lớp.



Fig. 4. Điểm F1-Score theo ngưỡng tin cậy.

4) Triển khai mô hình trên các thiết bị nhúng cho các ứng dụng tại hiện trường.

Nghiên cứu này đóng vai trò là một bước nền tảng để phát triển các giải pháp nông nghiệp thông minh toàn diện hơn.

REFERENCES

- [1] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097-1105.
- [2] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Advances in neural information processing systems*, 2015, pp. 91-99.
- [3] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779-788.
- [4] G. Jocher, A. Chaurasia, and J. Qiu, "YOLO by Ultralytics," 2023. [Online]. Available: <https://github.com/ultralytics/yolov5>