

Ứng dụng mô hình học sâu trong xác định các giai đoạn sinh trưởng của cây dưa lưới trồng trong nhà màng

Đặng Hoàng Anh Tuấn^{1*}, Nguyễn Minh Thắng²

¹Viện Ứng dụng Công nghệ

²Viện Khoa học và Công nghệ Việt Nam - Hàn Quốc

Ngày nhận bài 11/8/2021; ngày chuyển phản biện 16/8/2021; ngày nhận phản biện 14/9/2021; ngày chấp nhận đăng 16/9/2021

Tóm tắt:

Ngày càng có nhiều mô hình học sâu (deep learning - DL) được ứng dụng trong đời sống, xã hội như phân tích và dự đoán tài chính, giao thông thông minh, xe tự hành..., nhưng việc sử dụng hiệu quả công nghệ này để hỗ trợ trong canh tác nông nghiệp còn hạn chế. Bài báo trình bày kết quả triển khai một kiến trúc tiên tiến và gọn nhẹ có tên là YOLOv5 trong việc nhận dạng tự động các giai đoạn sinh trưởng của cây dưa lưới trong quá trình sinh trưởng, phát triển từ những hình ảnh thu thập được của hệ thống mạng camera lắp đặt trong nhà màng. Đề xuất nhận dạng hình ảnh này đạt độ chính xác trung bình 96% điểm F1 trong việc xác định 5 giai đoạn sinh trưởng quan trọng của cây dưa lưới bằng cách sử dụng một bộ dữ liệu huấn luyện và thử nghiệm hạn chế (2.818 ảnh cây dưa lưới). Kết quả sơ bộ cho thấy, YOLOv5 là một giải pháp DL gọn nhẹ và đầy hứa hẹn sau khi áp dụng kỹ thuật học chuyển giao (transfer learning). Hơn nữa, kiến trúc YOLOv5 còn thực thi tốt trên các thiết bị có cấu hình thấp, điều này có thể mở ra hướng tiếp cận mới trong việc nhận dạng các đối tượng khác nhau theo thời gian thực, được thực thi trực tiếp trên các thiết bị như smartphone, Jetson Nano, camera ip...

Từ khóa: dưa lưới, mô hình học sâu, nông nghiệp thông minh, thị giác máy tính, YOLOv5.

Chỉ số phân loại: 1.2

Giới thiệu

Để đảm bảo hiệu quả về năng suất cũng như chất lượng đáp ứng yêu cầu ngày càng cao và khắt khe hiện nay đối với các sản phẩm nông nghiệp, người nông dân cần phải có kiến thức và kinh nghiệm chuyên sâu về chăm sóc cây trồng, vật nuôi. Hiện nay, ngành nông nghiệp đang chuyển dịch từ quy mô hộ gia đình sang công ty và tập đoàn, với hình thức triển khai canh tác trên cánh đồng mẫu lớn; hệ thống các nhà màng khép kín với diện tích lên đến vài chục ha; trang trại chăn nuôi với quy mô lớn... Điều này đặt ra bài toán ứng dụng các công nghệ để quản lý, canh tác nông nghiệp hiệu quả, đáp ứng sự phát triển nêu trên.

Công nghệ trí tuệ nhân tạo (AI) là một thành phần quan trọng trong công nghệ nông nghiệp thông minh. Công nghệ AI được nghiên cứu và ứng dụng trong nông nghiệp nhằm phát hiện và cảnh báo sớm sâu bệnh trên cây trồng, nhận định tình trạng hạn, độ phì của đất qua việc phân tích ảnh vệ tinh, dự đoán sản lượng của mùa vụ... ngày càng được quan tâm và đầu tư. Trên cơ sở các kết quả/nhận định của AI, chúng ta có thể tham khảo để đưa ra các quyết định/chiến lược phát triển/phương án sản xuất phù hợp cho nông trại, hay cụ thể hơn là có thể xem xét và điều chỉnh chế độ canh tác cho cây trồng hay chăm sóc vật nuôi tại từng thời kỳ sinh trưởng. Vài năm gần đây, đã có một số công bố về việc ứng dụng AI trong trồng trọt như phát hiện các dấu hiệu sâu,

bệnh trên cây trồng như bệnh trên lá, nhận diện một số loại sâu, bọ gây hại... [1-4], nhưng với đối tượng là dưa lưới cho đến thời điểm hiện nay vẫn chưa thấy công bố nào.

Trong nghiên cứu này, chúng tôi trình bày kết quả ứng dụng công nghệ AI với việc áp dụng mô hình DL mới trong phát hiện và phân loại 5 giai đoạn sinh trưởng, phát triển chính của cây dưa lưới từ những hình ảnh thu thập được của hệ thống mạng camera lắp đặt trong nhà màng. Mô hình của chúng tôi được xây dựng bằng cách sử dụng kiến trúc DL tiên tiến, gọn nhẹ nhất hiện nay có tên là YOLOv5 (một kiến trúc DL tiên tiến trong lĩnh vực phát hiện đối tượng, cần ít tài nguyên để tính toán hơn so với các kiến trúc khác trong khi vẫn giữ được hiệu năng tốt [5, 6]) và được huấn luyện trên 2.818 hình ảnh cây dưa lưới do chúng tôi thu thập và gán nhãn [7].

Dữ liệu và phương pháp nghiên cứu

Chuẩn bị và xử lý tập dữ liệu

Nguồn dữ liệu chính trong nghiên cứu của chúng tôi được thu thập từ mạng camera thiết lập trong nhà màng trồng dưa lưới. Tổng số 2.818 hình ảnh cây dưa lưới với đủ các giai đoạn sinh trưởng, phát triển đã được thu thập, sau đó phân loại và gán nhãn thủ công với sự hỗ trợ của phần mềm VimageLabel (do nhóm nghiên cứu phát triển). Chú thích và hình ảnh ví dụ được hiển thị ở hình 1 (không sử dụng hình

*Tác giả liên hệ: Email: tuandha1980@gmail.com

Application of deep learning model in recognition of growth stages of *Cucumis melo* L. in greenhouse

Hoang Anh Tuan Dang^{1*}, Minh Thang Nguyen²

¹National Center for Technological Progress

²Vietnam - Korea Institute of Science and Technology

Received 11 August 2021; accepted 16 September 2021

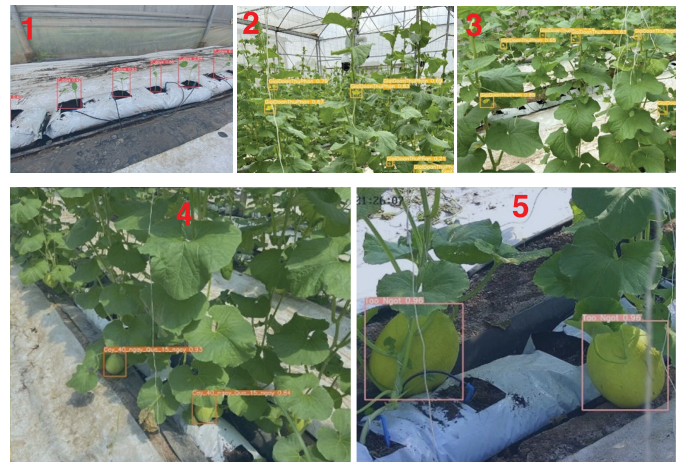
Abstract:

Despite the increasing application of deep learning (DL) models in various socioeconomics such as financial analysis and forecast, intelligent transport, self-driving, disease diagnosis, the effective use of this technology to support agricultural cultivation is still limited. This paper introduces the implementation of the lightest and state-of-the-art YOLOv5 architecture for automatic recognising of important growth stages of *Cucumis melo* L. from the camera images collected in the greenhouse. This image identification initiative achieved an average accuracy of 96% F1-score in the identification of the five growth stages of *Cucumis melo* L. using a limited set of training and testing data (total 2,818 images of *Cucumis melo* L.). These preliminary results lead to the conclusion that the YOLOv5 object detection and classification model is a truly lightweight and promising DL solution after the adoption of the transfer learning technique. Moreover, the YOLOv5 model can execute good performance on edge devices which may open up a new approach in different object detection and classification in real-time directly from a smartphone, Jetson Nano, IP camera...

Keywords: computer vision, *Cucumis melo* L., deep learning, smart agri, YOLOv5.

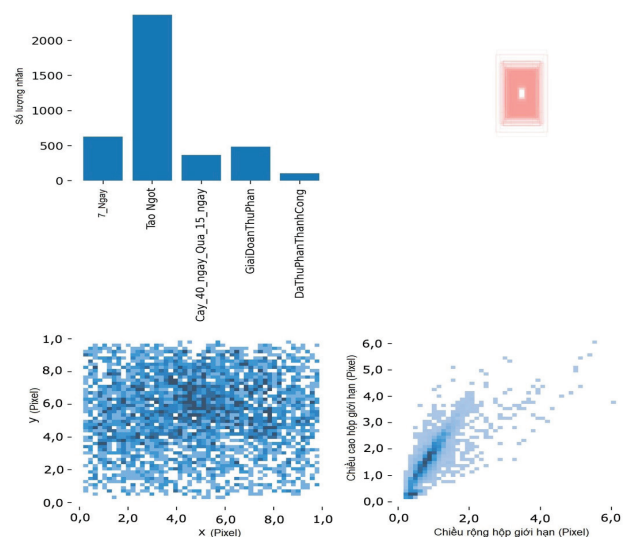
Classification number: 1.2

ảnh trong tập dữ liệu này). Thông qua việc nghiên cứu tài liệu và kinh nghiệm canh tác dưa lưới của Trang trại Linh An (Giang Biên, Long Biên, Hà Nội), chúng tôi xác định 5 giai đoạn sinh trưởng, phát triển quan trọng của cây dưa lưới trồng trong nhà màng cần được nhận biết gồm: 7 ngày; thụ phấn; đã thụ phấn; cây được 40 ngày và quả được 15 ngày; tạo ngọt. Với việc nhận biết chính xác các giai đoạn phát triển quan trọng này sẽ giúp người trồng có những chế độ canh tác phù hợp với nhu cầu sinh trưởng của cây.



Hình 1. Các giai đoạn đã nhận diện thành công. 1: giai đoạn 7 ngày; 2: giai đoạn thụ phấn; 3: giai đoạn đã thụ phấn; 4: giai đoạn cây được 40 ngày và quả được 15 ngày; 5: giai đoạn tạo ngọt.

Tương đương với 5 giai đoạn sinh trưởng, phát triển chính nêu trên, chúng tôi có 5 lớp nhận diện trong tập dữ liệu này. Hình ảnh trong tập dữ liệu của chúng tôi có nhiều kích cỡ khác nhau nhưng phần nhiều là độ phân giải 2.0 Megapixel. Sự phân bố các kích thước và số lượng dữ liệu được thể hiện ở hình 2.



Hình 2. Sự phân bố dữ liệu của các giai đoạn.

Kết quả hình 2 cho thấy, giai đoạn tạo ngọt (Tạo Ngọt) có số nhãn nhiều nhất (2.124), tiếp đến là 7 ngày (7_days) 512, thụ phấn (GiaiDoanThuPhan) 321, cây được 40 ngày và quả được 15 ngày (Cay_40_ngay_Qua_15_ngay) 164, đã thụ phấn (DaThuPhanThanhCong) 42. Ở đây có sự phân biệt dữ liệu như vậy là do việc thu thập dữ liệu của các giai đoạn có sự khó khăn khác nhau, như giai đoạn đã thụ phấn thành công quả rất nhỏ, bị che nhiều bởi lá cây, cộng với việc đặt camera ở xa so với kích thước nhãn của giai đoạn này.

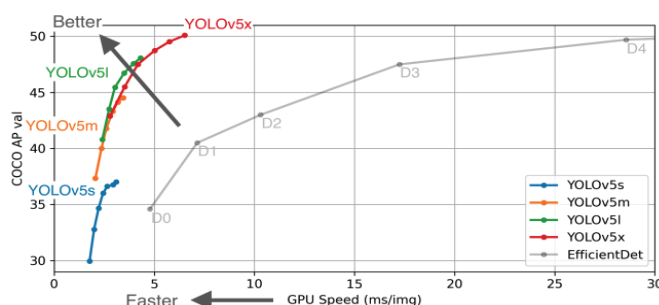
Tiếp theo kích thước của nhãn phân bố với chiều rộng ảnh từ 0,001x1.920 đến 0,6x1.920 pixel và chiều cao ảnh từ 0,001x1.080 đến 0,6x1.080 pixel. Các nhãn này tập trung chủ yếu ở kích thước rất nhỏ, chiều rộng từ 0,001x1.920 đến 0,2x1.920 pixel và chiều cao từ 0,001x1.080 đến 0,3x1.080 pixel.

Chúng tôi thực hiện xác thực chéo trong các thử nghiệm. Đây là quá trình tách tập dữ liệu thành 3 phần có kích thước bằng nhau và tiến hành 3 lần huấn luyện. Đối với mỗi lần huấn luyện, 20% dữ liệu được sử dụng để xác thực và 80% còn lại được sử dụng để huấn luyện mạng. Số vòng huấn luyện được lựa chọn tăng dần, tỷ lệ phần trăm đúng của các epoch khác nhau được phân bố ở bảng 1.

Bảng 1. Tỷ lệ % đúng của các epoch trong các lần huấn luyện.

Epoch	Các giai đoạn	7 ngày (%)	Tạo ngọt (%)	Cây được 40 ngày và quả được 15 ngày (%)	Thụ phấn (%)	Đã thụ phấn thành công (%)
100		93	90	85	45	35
200		98	95,5	92	88	86
300		99,8	97,6	98,3	96,3	88

Lý do chọn mô hình YOLOv5 cho nghiên cứu này gồm: hiện là mạng tiên tiến nhất trong lĩnh vực phát hiện đối tượng nhanh; kiến trúc nhẹ, điều này cho phép chúng tôi huấn luyện mô hình bằng cách sử dụng các tài nguyên tính toán nhỏ và chi phí thấp; kích thước nhỏ của mô hình có thể cho phép nó được ứng dụng với các thiết bị di động (nhận diện người, đồ vật...). Hình 3 cho thấy khả năng vượt trội của mô hình YOLOv5 so với các mô hình khác.



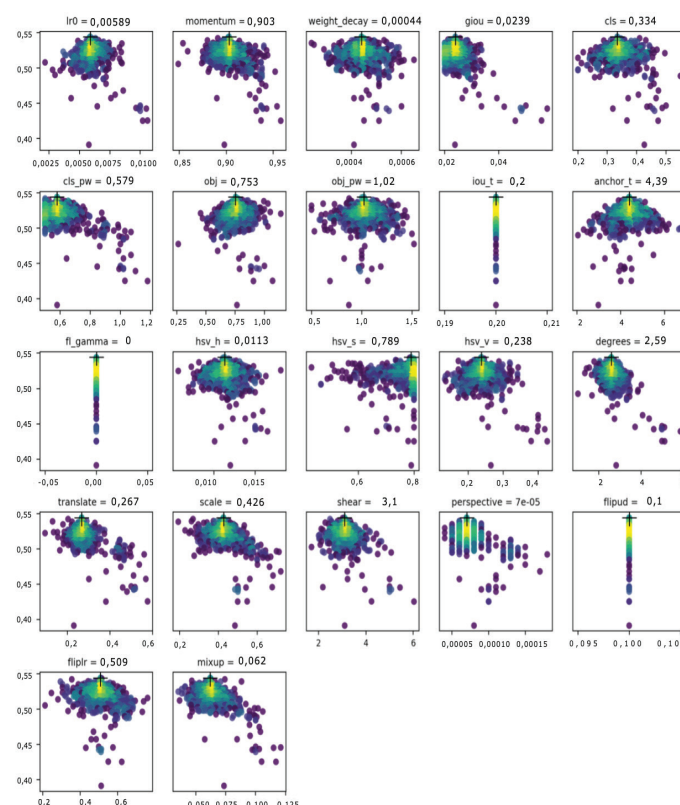
Hình 3. So sánh giữa các mô hình YOLOv5 và EfficientDet [8].

Các hình ảnh mà chúng tôi thu thập được đều có độ phân giải trung bình 1.920x1.080 pixel, cùng với độ chính xác của YOLOv5s6 là thấp hơn nhiều so với YOLOv5m6 và YOLOv5l6 khó giải quyết vấn đề nhận biết các đối tượng nhỏ (vấn đề nổi tiếng của YOLO) [9]. YOLOv5l6 và YOLOv5x6 lại có params rất lớn, gấp 2,15 và 3,94 lần YOLOv5m6 nên việc huấn luyện mô hình là rất mất thời gian và quan trọng nhất là không thực thi được trên thiết bị có cấu hình thấp như thiết bị Jetson AGX Xavier Developer Kit của chúng tôi. Vì vậy, chúng tôi đã chọn Pretrained Model YOLOv5m6 làm mô hình huấn luyện trong nghiên cứu này.

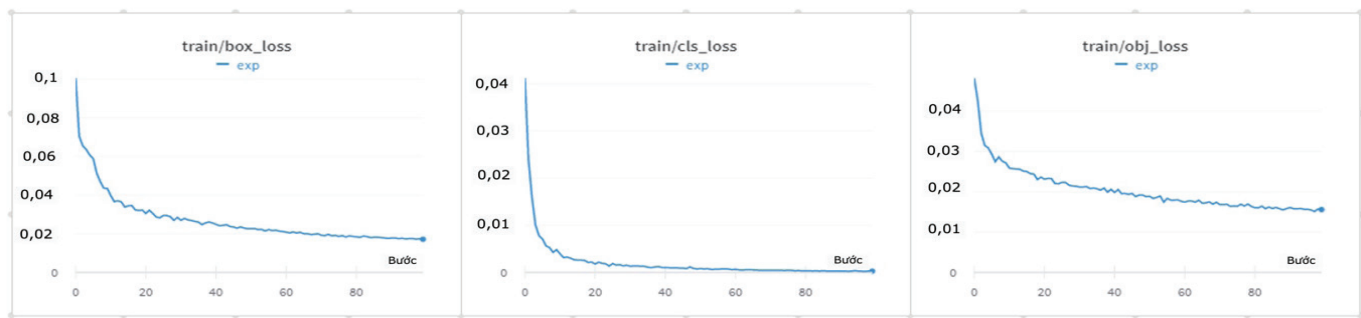
Huấn luyện mô hình

Việc đầu tiên của quá trình huấn luyện mô hình là điều chỉnh siêu tham số (hyper-parameter), điều này giúp xác định các tham số tối ưu hơn cho tập dữ liệu huấn luyện. Tiếp theo, sử dụng các hyper-parameter đã được điều chỉnh để huấn luyện mô hình và bắt đầu từ việc kiểm tra mô hình YOLOv5m6 đã được huấn luyện. Có khoảng 30 hyper-parameter được sử dụng trong YOLOv5m6 [10].

Sau quá trình huấn luyện, các hyper-parameter được mô hình hoá như ở hình 4. Mỗi hyper-parameter là một chấm nhỏ thể hiện sự phù hợp (trục y) tương ứng với các giá trị của nó (trục x); màu vàng cho biết mật độ của các hyper-parameter tập trung cao hơn; sự phân bố của các chấm nhỏ theo chiều dọc cho thấy một tham số bị vô hiệu và không làm ảnh hưởng đến các đối tượng đi qua nó.



Hình 4. Sự phân bố của các loại hyper-parameter.



Hình 5. Các thay đổi về hồi quy, phân loại và mất đối tượng của hộp giới hạn trong quá trình huấn luyện mô hình.

Sử dụng mô hình được huấn luyện trước là một kỹ thuật phổ biến trong thị giác máy được gọi là học chuyển giao [11]. Kỹ thuật này giúp đẩy nhanh quá trình huấn luyện và giữ cho sự tổng quát hóa ở mức cao. Trong các thí nghiệm, chúng tôi đã quan sát thấy rằng, số chu kỳ huấn luyện tối ưu là 300, sau đó có những thay đổi không đáng kể trong mô hình. Hình 5 cho thấy các hàm loss thay đổi như thế nào trong quá trình huấn luyện, kết quả được hiển thị trên một trong các phần tách để xác nhận chéo giữa các lần huấn luyện.

Hàm YOLOv5 loss là tổng của 3 hàm loss nhỏ hơn: loss hồi quy hộp giới hạn - hình phạt cho việc phát hiện hộp neo sai, lỗi bình phương trung bình được tính toán dựa trên vị trí hộp dự đoán (x, y, h, w); loss phân loại - Entropy chéo được tính toán để phân loại đối tượng; loss đối tượng - lỗi bình phương trung bình được tính cho điểm đối tượng - độ tin cậy (ước tính nếu hộp neo chứa đối tượng).

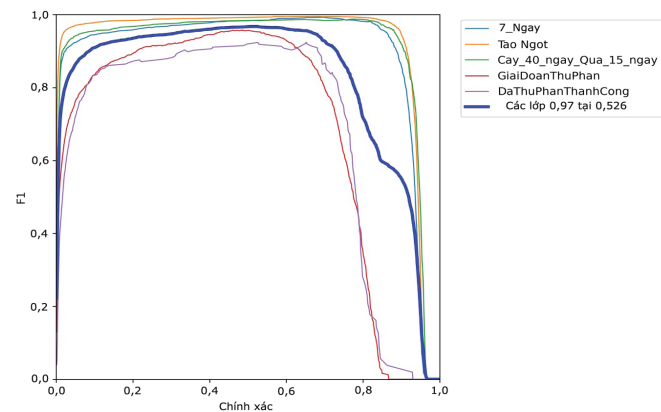
Phần mềm chính được sử dụng trong việc huấn luyện các mô hình là Python 3.8.8, với PyTorch 1.7.1, CUDA 11.1 và PyCharm Professional. Tất cả các lần lặp lại mô hình AI cũng như mô hình cuối cùng đều được huấn luyện bằng cách sử dụng card màn hình GPU GTX 1650. Chúng tôi đã thực hiện thử nghiệm trên Jetson AGX Xavier Developer Kit với phần mềm Linux được cài đặt sẵn. Thời gian huấn luyện cho một lần (100 epochs) trên dữ liệu huấn luyện mất 3 giờ (8,5 giờ cho toàn bộ quá trình xác nhận chéo).

Kết quả và thảo luận

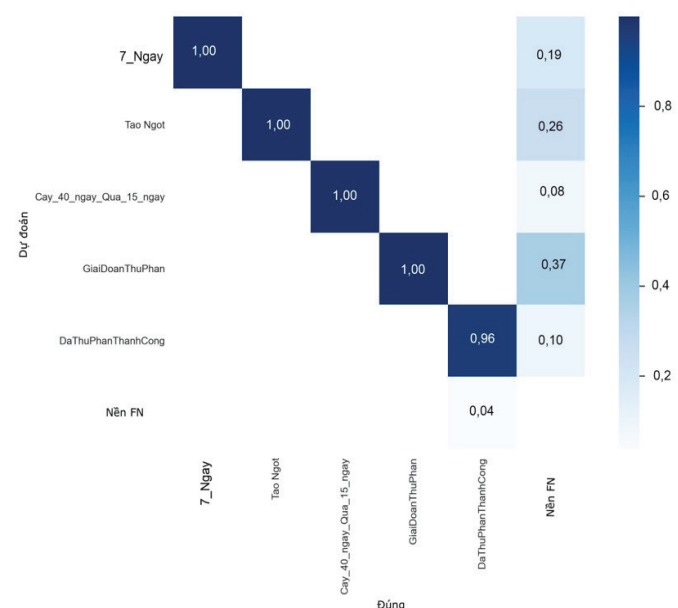
Sau các lần huấn luyện mô hình với một lượng dữ liệu hạn chế (2.818 ảnh), kết quả cho thấy, trung bình 96% điểm F1 trong việc xác định 5 giai đoạn sinh trưởng, phát triển quan trọng của cây dưa lưới. Mô hình YOLOv5 không chỉ có thể duy trì tỷ lệ phát hiện cao, mà còn đáp ứng được các yêu cầu phát hiện trong thời gian thực chính xác và nhanh chóng. Các kết quả được trình bày ở đây là kết quả trung bình từ 5 lần phân tách để thực hiện quá trình xác nhận chéo.

Từ dữ liệu nêu trên, chúng ta có thể thấy độ chính xác phát hiện đối với giai đoạn đã thụ phấn thành công là thấp nhất do có số lượng nhãn ít nhất và hình dạng rất nhỏ so với các đối tượng khác. Hơn nữa, lớp đã thụ phấn thành công có điểm F1 thấp nhất (0,88) vì ngoài sự phong phú của giai đoạn đã thụ phấn thành công thấp, thời điểm này còn dễ bị

nhầm sang giai đoạn thụ phấn. Cả 2 giai đoạn này đều có đối tượng nhận diện rất nhỏ và số nhãn ít. Do đó, giai đoạn thụ phấn và đã thụ phấn thành công thường bị phát hiện thiếu khi thực thi mô hình của chúng tôi (hình 6). Những khuyết điểm nêu trên sẽ được khắc phục ở các thí nghiệm trong tương lai.

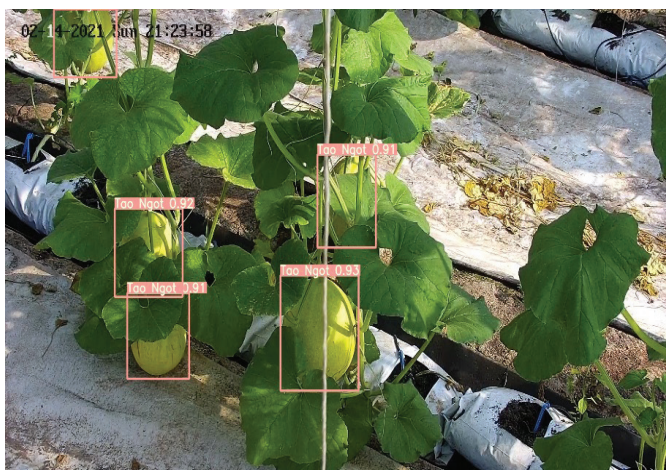


Hình 6. Độ chính xác của các giai đoạn ở tập thử nghiệm.

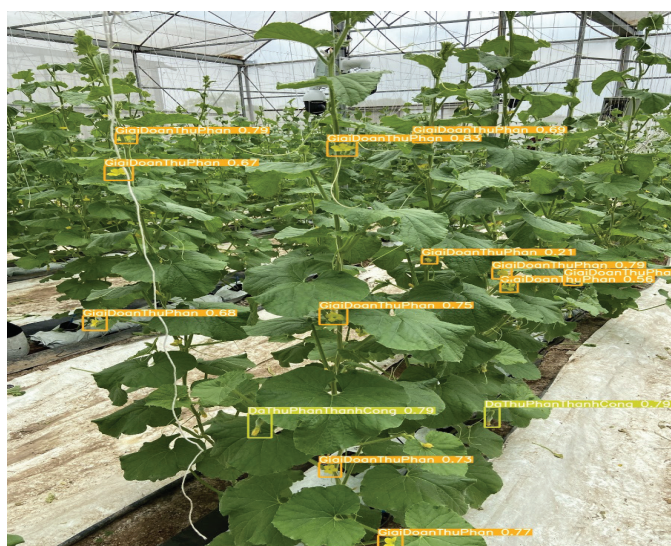


Hình 7. Ma trận của các giai đoạn ở tập thử nghiệm.

Các ví dụ nhận dạng thiếu: chúng tôi nhận thấy những nhược điểm của mô hình phân loại có 4 vấn đề phổ biến nhất là: đối tượng nhận dạng bị che khuất, nhận dạng quá xa, bị mờ và trùng với màu của lá.



Hình 8. Mô hình nhân dạng với giai đoạn tạo ngọt.



Hình 9. Mô hình nhận dạng với giai đoạn thụ phấn và giai đoạn đã thụ phấn thành công.

Kết luận

Kết quả sơ bộ đạt được khi sử dụng một lượng dữ liệu huấn luyện hạn chế, trung bình 96,0% điểm F1 trong việc xác định các giai đoạn sinh trưởng chính của cây dưa lưới trồng trong nhà màng. Những kết quả đầy hứa hẹn này cho thấy tiềm năng lớn trong việc sử dụng kiến trúc học sâu YOLOv5 để huấn luyện mô hình AI trong xác định các giai đoạn sinh trưởng, phát triển của cây dưa lưới nói riêng, các yêu cầu nhận dạng khác nói chung trong tương lai. Bộ dữ liệu hình ảnh càng lớn sẽ cho phép khả năng nhận diện của mô hình ngày càng tốt hơn, nhanh và chính xác hơn. Hơn nữa, kiến trúc YOLOv5 này đã được chúng tôi thử nghiệm thành công trên các thiết bị Jetson AGX Xavier Developer Kit và Jetson Nano, điều này có thể mở ra một hướng tiếp cận mới trong việc nhận dạng các đối tượng khác nhau theo thời gian thực được thực thi trực tiếp trên các thiết bị như smartphone, Jetson Nano, camera IP...

LỜI CẢM ƠN

Chúng tôi xin trân trọng cảm ơn Viện Khoa học và Công nghệ Việt Nam - Hàn Quốc, Viện Ứng dụng Công nghệ, Viện Nghiên cứu Rau quả Việt Nam, Trang trại Linh An đã giúp đỡ trong quá trình thực hiện nghiên cứu này.

TÀI LIỆU THAM KHẢO

- [1] <https://www.hindawi.com/journals/ddns/2020/2479172/>.
- [2] <https://www.hindawi.com/journals/complexity/2021/6663442/>.
- [3] <https://www.hindawi.com/journals/jece/2021/9981437/>.
- [4] <https://www.hindawi.com/journals/complexity/2021/6683255/>.
- [5] <https://arxiv.org/abs/1506.01497>.
- [6] <https://www.aaai.org/ocs/index.php/AAAI/AAAI17/paper/viewFile/14806/14311>.
- [7] <https://tinyurl.com/anhdualuoi>.
- [8] <https://github.com/ultralytics/yolov5/releases/tag/v5.0>.
- [9] <https://github.com/ultralytics/yolov5/issues/2009>.
- [10] <https://github.com/ultralytics/yolov5/issues/607>.
- [11] <https://github.com/ultralytics/yolov5/wiki/Train-Custom-Data>.