

MÔ HÌNH FAST R-CNN CẢI TIẾN CHO GIẢI PHÁP NHẬN DẠNG, PHÁT HIỆN TRÁI DỨA THỜI KỲ CHÍN

THE IMPROVED FAST R-CNN MODEL FOR RECOGNITION AND DETECTION OF PINEAPPLE RIPENING

Trịnh Trung Hải^{1*}, Nguyễn Hà Huy Cường², Ninh Khánh Duy³

¹Trường Đại học Công nghệ Thông tin và Truyền thông Việt-Hàn - Đại học Đà Nẵng

²Trung tâm Phát triển Phần mềm - Đại học Đà Nẵng

³Trường Đại học Bách khoa - Đại học Đà Nẵng

*Tác giả liên hệ: tthai@vku.udn.vn

(Nhận bài: 22/1/2022; Chấp nhận đăng: 21/6/2022)

Tóm tắt - Trong những năm gần đây, khoa học công nghệ đã được ứng dụng mạnh mẽ và góp phần phát triển kỹ nguyên chuyển đổi số. Dịch bệnh COVID-19 đã kiểm chứng việc ứng dụng thành tựu công nghệ trong thực tế là rất hữu ích. Ngày càng có nhiều ứng dụng thông minh trong lĩnh vực nông nghiệp để thay thế nông nghiệp truyền thống. Đặc biệt là các ứng dụng trí tuệ nhân tạo vào tự động hóa các khâu sản xuất và bảo quản nông sản Việt Nam. Trong bài báo này, nhóm tác giả trình bày giải pháp nhận dạng phát hiện trái dứa vào thời kỳ thu hoạch sử dụng mô hình Fast R-CNN cải tiến, thực nghiệm được thực hiện tại các vùng, vườn trồng dứa ở Việt Nam với sản lượng thu hoạch khá lớn. Từ đó xây dựng hệ thống ứng dụng dự đoán trái dứa chín chạy trên nền tảng di động nhằm hỗ trợ người nông dân đạt hiệu quả hiệu quả kinh tế cao đối với cây dứa và các cây nông nghiệp chủ lực của Việt Nam.

Từ khóa - Nhận dạng trái cây; phát hiện trái dứa chín; Fast R-CNN cải tiến; dự đoán và phát hiện; mạng nơ-ron tích chập.

1. Đặt vấn đề

Trí tuệ nhân tạo (AI), Internet vạn vật (IoT), Điện toán đám mây (Cloud Computing) và dữ liệu lớn (BigData) là những trụ cột của cuộc cách mạng công nghiệp 4.0. Dựa trên những trụ cột này đã tạo ra các mô hình ứng dụng mới trong sản xuất, cũng từ đây đã thúc đẩy nhiều hoạt động, đã có tác động mạnh mẽ tới nền kinh tế số, chính trị số và đời sống xã hội số. Có thể nói rằng thời điểm này trí tuệ nhân tạo đang nhanh chóng trở thành một hình trong những lĩnh vực khoa học được mong đợi nhất, vì khả năng mang lại lợi ích cho nhiều ngành, nhiều lĩnh vực như công nghiệp, nông nghiệp, y khoa, giáo dục. Ứng dụng của thành tựu trí tuệ nhân tạo cho ngành nông nghiệp chính xác (Precision Agriculture) được sử dụng rộng rãi và mang nhiều kết quả to lớn như: Máy bay, xe kéo không người lái, robot hỗ trợ thu hoạch tự động, hệ thống đo độ ẩm của đất phục vụ cho tưới tiêu nông sản... Đặc biệt, nhiều nhà nghiên cứu đã chỉ ra rằng, việc sử dụng trí tuệ nhân tạo có thể làm tăng tốc độ và độ chính xác cho chẩn đoán, khuyến nghị người nông dân hạn chế sử dụng thuốc bảo vệ thực vật và đảm bảo được chất lượng an toàn thực phẩm.

Xử lý hình ảnh và thị giác máy tính là những nền tảng cơ sở đã được sử dụng để đáp ứng các nhu cầu trên. Nhận dạng

Abstract - Recently, high technology has been widely applied, contributed to the digital transformation society. The COVID-19 pandemic has proved that the application of technological achievements in the modern life is necessary. Nowadays, there are many intelligent agricultural applications that replaced the traditional agriculture. Especially, a number of artificial intelligence applications are used to automatize the agricultural production and storage in Vietnam. In this paper, we present a method of recognizing and detecting ripe pineapple by using the improved Fast R-CNN model at pineapple growing farms, areas in Vietnam which have high yield. On the basic of this, we create a ripe pineapple prediction system on mobile phone to assist farmers in achieving high economic efficiency for pineapple and other key agricultural crops in Vietnam.

Key words - The Fruit recognition; Detect ripe Pineapple; improved Fast R-CNN; recognition and detection; convolutional neural network.

trái cây thời kỳ thu hoạch từ dữ liệu ảnh số ở vùng trồng qua thiết bị camera là một trong trong số đó. Hiện nay, những nghiên cứu áp dụng này đang được triển khai đưa vào sử dụng ở một số nước tiên tiến. Tuy nhiên, do vấn đề về giá thành của sản phẩm, cũng như các thiết bị này không tích hợp được Tiếng Việt là một trong những hạn chế mà người nông dân vùng nông thôn không thể tiếp cận được. Việt Nam là nước nông nghiệp, hằng năm cung cấp một sản lượng lớn lúa, gạo phục vụ cho người dân thành phố và xuất khẩu ra thị trường thế giới. Ngoài mặt hàng lúa gạo trái cây cũng chiếm tỷ trọng lớn và đem lại thu nhập bình quân cho người nông dân khá cao so với các sản lượng nông sản khác. Việt Nam đang khoanh vùng trồng các loại cây chủ lực này nhằm hướng tới xuất khẩu ra thị trường thế giới. Trong đó, trái dứa (khám) đang được nhiều trồng nhiều ở Miền Trung và cả nước (Tiền Giang, Thanh Hóa, Quảng Nam.) và được đánh giá rất cao về giá trị dinh dưỡng, cũng như sản lượng thu hoạch. Vì thế, việc nghiên cứu áp dụng kỹ thuật học sâu để phát hiện và nhận dạng trái dứa trong thời kỳ chín có ý nghĩa khoa học, thực tiễn và có tính thời sự. Có thể nói nhận dạng trái cây, đặc biệt đối với trái dứa là một thách thức lớn vì dứa có những đặc điểm giống nhau về hình thái và màu sắc. Nhận dạng hình ảnh cần nhiều sức mạnh tính toán với các dữ liệu

¹ The University of Danang - Vietnam-Korea University of Information and Communication Technology (Trịnh Trung Hải)

² The University of Danang - Software Development Centre (Nguyễn Hà Huy Cường)

³ The University of Danang - University of Science and Technology (Ninh Khanh Duy)

thu thập lớn (Big Data). Phát hiện và phân loại đối tượng sử dụng các phép chuyển đổi không gian màu để tách các đối tượng ra khỏi môi trường xung quanh và cho phép lưu được những thông tin về cấu trúc của trái dứa. Sử dụng mạng nơ-ron tích chập nhằm trích xuất các đặc trưng phục vụ cho quá trình huấn luyện để tiến hành phân loại và dự đoán trái dứa.

Trái dứa là một loại cây ăn quả nhiệt đới hiện được trồng nhiều ở các nước Nam Mỹ. Một số nước nhiệt đới và một số nước cận nhiệt đới có sông ngòi tương đối nhiều như đảo Hawaii, Đài Loan, Việt Nam và các nước khác, chiếm 60% sản lượng dứa của thế giới. Dứa ở Việt Nam được phân thành nhiều nhóm dựa trên các đặc điểm hình thái, sự phát triển và chất lượng hương vị. Nền nông nghiệp của Việt Nam hiện đang xuất khẩu dứa loại 1 đi khắp thế giới. Đã góp phần tạo ra nền kinh tế phát triển thịnh vượng và mang lại thu nhập cho nhiều hộ gia đình.

2. Nghiên cứu liên quan

Các thuật toán cải tiến từ mạng nơ-ron tích chập đã chứng minh tính ưu việt cho phép đạt tới ngưỡng chính xác tuyệt đối, được sử dụng phổ biến và cho kết quả cao trong các bài toán phân loại hình ảnh. Mạng nơ-ron tích chập (CNNs) cho phép khả năng trích chọn đặc trưng của lớp tích chập, sử dụng nhiều lớp tích chập và lớp gộp để xử lý tăng độ sáng và chất lượng hình ảnh [1]. Thuật toán Fast R-CNN được cải tiến từ đó, với khả năng phát hiện đối tượng, thuật toán thực hiện 2 bước chính. Đầu tiên, sử dụng tìm kiếm có chọn lọc để tìm những vị trí phù hợp nhất của hộp giới hạn (Region of Interest hay - ROI), sau đó sử dụng các mô hình CNNs để trích xuất và nhận dạng các hộp giới hạn đó [2]. Với mô hình R-CNN kết hợp với phương pháp tìm kiếm chọn lọc để phát hiện các khu vực được đề xuất và là cơ sở cho sự ra đời mô hình Fast R-CNN [3]. Byoungjun Kim và cộng sự [4], đã nghiên cứu đề xuất sử dụng mạng nơ-ron sâu để cải thiện khả năng phát hiện chính xác trái Dâu Tây trong thử nghiệm môi trường bị che khuất, xen lẫn trong lá cây. Jose Luis Rojas-Aranda và cộng sự [5] sử dụng Deep Learning phân loại trái cây cho cửa hàng bán lẻ, với màu đơn sắc RGB và biểu đồ RGB từ phân cụm K-mean được sử dụng đầu mô hình. Joseph Redmon [6] đề xuất hướng tiếp cận mới để phát hiện đối tượng trong ảnh bằng cách sử dụng hộp giới hạn và xác suất lớp liên quan để tối ưu hóa hiệu suất phát hiện đối tượng trực tiếp trên ảnh.

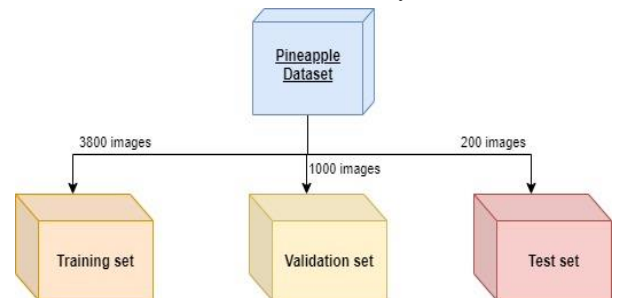
Trong lĩnh vực nông nghiệp các nhà nghiên cứu hiện nay đang rất quan tâm tới việc ứng dụng công nghệ cao trong vào trong các giai đoạn từ ương giống cho đến thu hoạch. Có thể thấy, đã có những sản phẩm công nghệ có thể hỗ trợ nông dân trong công việc chăm sóc định kỳ, tăng chất lượng đầu ra của trái cây. Với mong muốn có được nền nông nghiệp phát triển bền vững, vương tâm thế giới. Chính phủ Việt Nam luôn quan tâm và chú trọng đầu tư lĩnh vực nông nghiệp mũi nhọn này. Trên thế giới hiện nay, có các nhóm nghiên cứu của Horea và cộng sự [7], đã đề xuất một hệ thống huấn luyện theo kỹ thuật mạng nơ-ron sâu để xác định các loại quả từ hình ảnh. Nhóm Susovan Jana và cộng sự [8] đề xuất hệ thống tự động nhận dạng trái cây và rau quả sử dụng công nghệ thị giác máy tính và máy học. Nhóm nghiên cứu của Md Tohidul Islam và cộng sự

[9] đề xuất một hệ thống sử dụng mạng nơ-ron phức hợp để phân loại hình ảnh thực phẩm. Emmanuel Karlo Nyarko và cộng sự [10] đề xuất hệ thống nhận dạng quả đến kỳ để thu hoạch tự động bằng robot thông minh. Tao Yongting và Zhou Jun [11] đã đề xuất hệ thống nhận dạng để thu hoạch quả táo tự động bằng robot. Walter Maldonado Jr. và Jose Carlos Barbosa [12] đã đề xuất hệ thống chiết xuất các đặc điểm của quả xanh, vấn đề này được thực hiện với sự kết hợp của các kỹ thuật chuyển đổi mô hình màu, tạo ngưỡng, cân bằng biểu đồ, lọc không gian với các toán tử Laplace và Sobel và làm mờ Gaussian. Xiangqin Wei và cộng sự [13] đề xuất hệ thống hái trái cây trưởng thành bằng rô bốt. Sakib và cộng sự [14] đã đề xuất hệ thống nhận dạng trái cây bằng cách sử dụng Mạng neuron tích chập (CNN) và thị giác máy tính. Một số nghiên cứu liên quan sử dụng mô hình nhận dạng CNN và các mô hình cải tiến được đề xuất ứng dụng trong phát hiện và nhận dạng hình ảnh kỹ thuật số [15] – [20]. Qua đó cho thấy việc sử dụng trí tuệ nhân tạo, đặc biệt là xử lý hình ảnh trong lĩnh vực nông nghiệp, dịch vụ đang được chú trọng và sẽ có những bước đột phá thời gian đến.

3. Phương pháp

3.1. Tập dữ liệu đầu vào

Dữ liệu hình ảnh đầu vào được nhóm tác giả tự thu thập bằng các thiết bị quay chụp chuyên dụng. Số lượng hình ảnh đạt được là 5000 hình ảnh, được chia thành 3 tập dữ liệu nhỏ bao gồm: Tập huấn luyện, tập xác nhận và tập kiểm thử cùng với thẻ định danh thông tin. Trước khi sử dụng 1000 hình ảnh để đánh giá, trích 3800 hình ảnh đầu tiên để phục vụ cho huấn luyện. Tập dữ liệu kiểm thử được tạo bằng cách sử dụng 200 hình từ tập dữ liệu. Sơ đồ tập dữ liệu hình ảnh đầu vào được biểu thị theo Hình 1 dưới đây:



Hình 1. Tập dữ liệu hình ảnh đầu vào

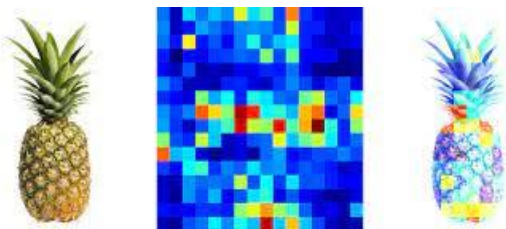
3.2. Tiền xử lý và làm giàu dữ liệu

Tiền xử lý là một bước quan trọng trong quá trình xử lý dữ liệu, giúp hình ảnh được đồng bộ về kích thước, chất lượng hình ảnh phục vụ cho việc huấn luyện. Cần phải xử lý hình ảnh để làm sạch dữ liệu hình ảnh đầu vào nhằm làm giảm thời gian huấn luyện và cải thiện hiệu suất. Sử dụng hình ảnh nhỏ hơn để huấn luyện có thể cắt giảm thời gian đáng kể, làm giàu dữ liệu nhằm tạo ra các mẫu dữ liệu mới từ dữ liệu hiện có. Từ đó tổng quát hóa các tình huống có thể xảy ra cho phép mô hình huấn luyện thể hệ mới được học hỏi từ các tình huống rộng hơn. Các hình ảnh lớn được cắt thành các hình ảnh có độ phân giải (416x416). Mô hình có thể lưu trữ nhiều dữ liệu hơn và xử lý nhanh hơn. Các bước xử lý các giai đoạn biến đổi khác nhau như xoay, tạo nhiễu, thay đổi độ sáng.

Bảng 1. *Tiền xử lý dữ liệu hình ảnh đầu vào*

Kích thước đầu vào	Số lượng	Tỷ lệ kích thước giảm
> 512x512 (medium)	504	416x416(~-1,23%)
>1920x1080 & < 512x512 (large)	848	416x416(~-2,59%)
>1920x1080 (jumbo)	3448	416x416(~-4,16%)
Tổng	4800	416x416

Quá trình làm giàu dữ liệu hình ảnh trái dứa cần cung cấp hình ảnh đã xử lý thô và xây dựng mô hình huấn luyện gắn nhãn và đánh dấu vị trí trái dứa có trong hình ảnh. Annotate với chức năng gắn nhãn đối với trái dứa sau khi được vẽ hình hộp chữ nhật bao quanh có thể nhìn thấy được, qua đó không chỉ thu được các hộp giới hạn mà còn cả các pixel tương ứng của mỗi trái dứa. Nhiệm vụ của gắn nhãn nhằm trích xuất các vùng đặc trưng dựa trên các đặc điểm hình thái như mắt, cuống, màu sắc, hình dạng, v.v. Kết quả cuối cùng sẽ là một tập dữ liệu định dạng gắn nhãn được lưu ở định dạng Microsoft COCO. Tại Hình 2 biểu diễn mật độ vị trí của trái dứa được gắn nhãn được biểu diễn bằng sơ đồ nhiệt.



Hình 2. *Biểu đồ nhiệt mật độ phân bố trái dứa trong hình*

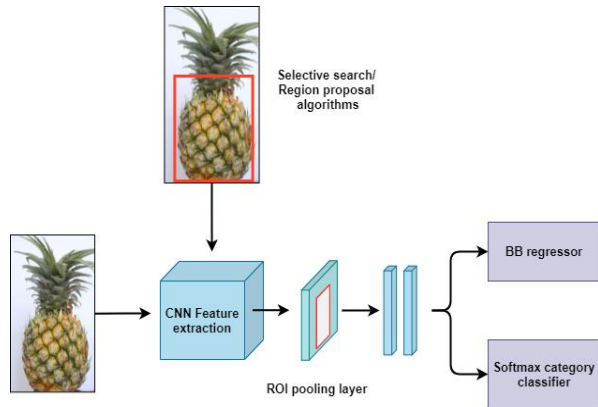
3.3. Mô hình huấn luyện Fast R-CNN

Mạng CNN được xây dựng để phân loại hình ảnh sử dụng kỹ thuật học sâu. Xây dựng mô hình CNN dễ dàng hơn, cùng với việc sử dụng thư viện Keras với ngôn ngữ Python. Mô hình sử dụng các lớp CNN như Conv2D & MaxPooling2D. Conv2D còn được gọi là 2D Convolution Layer. Lớp này tạo ra hàng chục đầu ra bằng cách tạo ra nhân chập trùng với lớp đầu vào. MaxPooling 2D là dùng cho hoạt động tổng hợp tối đa cho dữ liệu không gian. Dữ liệu không gian có thể được định nghĩa là biểu diễn thông tin về một đối tượng vật lý bằng các giá trị số. Việc chọn phần tử tối đa từ vùng của bản đồ điện tử được bao phủ bởi bộ lọc là toán hạng được thực hiện bởi lớp tổng hợp tối đa. Để giảm kích thước của vùng đặc trưng (Feature Map), các lớp gộp được sử dụng. Do đó, có thể kết luận rằng, việc sử dụng các lớp tổng hợp làm giảm số lượng các tham số cần tìm hiểu và khả năng tính toán được thực hiện trong mạng nơ-ron nhân tạo.

Thuật toán Fast R-CNN được đánh giá là mô hình nhận dạng khá tốt về độ chính xác phát hiện và tốc độ phát hiện. Đã có nhiều nghiên cứu đánh giá về độ chính xác mô hình này với các mô hình nhận dạng tương đồng khác như mô hình R-CNN [2]. Việc huấn luyện mô hình R-CNN có sẵn tốn khá nhiều tài nguyên tính toán, lưu trữ và thời gian vì mỗi hình ảnh yêu cầu phân loại các lớp với khoảng 2000 khu vực đề xuất, do đó, thời gian huấn luyện rất chậm và không thể áp dụng trong thời gian thực vì mỗi hình ảnh trong bộ thử nghiệm mất hơn 47 giây để xử lý và cho ra

kết quả dự đoán. Trong khi đó, Fast R-CNN thực hiện song song việc trích xuất đặc trưng với việc tách các vùng đề xuất 2000 khu vực vào CNN và thực hiện trích xuất từ bản đồ đặc trưng. Giảm được chi phí huấn luyện và thời gian đào tạo chỉ 0,32 giây trên cùng bộ dữ liệu như R-CNN, nên cho kết quả thực hiện nhanh hơn rất nhiều. Về thời gian phát hiện và nhận dạng bằng camera di động tương đồng với hình ảnh, vì mỗi giây camera sẽ gửi 24 khung hình/ 1 giây cho hệ thống. R-CNN xử lý 1 hình ảnh mất đến 47 giây là khá lâu nên không thể đáp ứng việc nhận dạng theo thời gian thực. Trong khi đó, Fast R-CNN chỉ mất khoảng 1s để trả về kết quả nên có khả năng xử lý camera di động và phát hiện đối tượng theo thời gian thực.

Fast R-CNN sẽ tạo ra một vùng đối tượng đặc trưng phức tạp, tải toàn bộ hình ảnh vào ConvNet (lớp chập + lớp tổng hợp tối đa). Sau khi kết hợp CNN, sẽ có một bản đồ đặc trưng nhỏ hơn đáng kể so với hình ảnh gốc, nhờ lớp ROI Pooling, lớp này sẽ đưa các vùng hình ảnh về kích thước hình vuông và định hình lại với cùng kích thước đầu ra. Tiếp tục qua các lớp được kết nối đầy đủ cho đến khi bạn đạt đến đầu ra vector tính năng ROI. Cuối cùng, để có được kết quả cuối cùng, mô hình sẽ dự đoán lớp và độ lệch của hộp giới hạn cho vùng phân mảnh đó.



Hình 3. *Kiến trúc hệ thống mô hình huấn luyện bằng thuật toán học sâu Fast R-CNN*

3.4. Hệ thống nhận dạng và phân loại trái dứa

Dựa trên mô hình huấn luyện đã hoàn thiện, nhóm tác giả đã phát triển một hệ thống có khả năng phân tích và đánh giá hình ảnh đầu vào. Bên cạnh đó, còn phát triển thêm giao diện người dùng thân thiện được xây dựng bằng Framework Django với ngôn ngữ lập trình Python. Người dùng chỉ cần chỉ định đường dẫn của hình ảnh đó trong máy tính, hệ thống sẽ đưa hình ảnh lên giao diện. Tiến hành nhận dạng, hệ thống sẽ kiểm tra hình ảnh có tồn tại không và tiến hành dự đoán với mô hình huấn luyện từ trước.



Hình 4. *Đầu vào và đầu ra khi nhận dạng vị trí của các trái dứa*

4. Đánh giá hiệu suất mô hình huấn luyện

Để đánh giá mô hình, nhóm tác giả so sánh kết quả từ mô hình Fast R-CNN trên tập xác nhận đã thu thập từ trước. Hệ số Intersection Over Union - IoU đạt được 0,5. Trường hợp này được coi là vị trí đạt được tiêu chuẩn. IoU được xác định bằng cách chia số pixel trong giao điểm cho toàn bộ pixel.

$$IoU = \frac{\text{Area of Overlap}}{\text{Area of Union}}$$

$$IoU = \frac{TP}{(TP+FP+FN)}$$

Độ chính xác trung bình (AP) là diện tích của biểu đồ bao gồm Precisions, Recall và mAP là diện tích trung bình của toàn bộ biểu đồ. Trong đó, đại lượng Precisions, Recall và F1 được sử dụng để đánh giá hiệu suất mô hình và được tính theo các công thức bên dưới:

$$\text{Precision} = \frac{TP}{TP+FP}; \text{Recall} = \frac{TP}{TP+FN}$$

$$F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Trong đó: F1 là đại lượng đánh giá độ chính xác trung bình điều hòa giữa precision và recall; Precision phản ánh mức độ chuẩn xác của mô hình dự đoán đúng (nhóm positive); Tỷ lệ precision càng cao thì kết quả dự đoán của mô hình càng tốt trong việc phát hiện trái dứa chín. Recall: đo lường tỷ lệ dự báo chính xác các trường hợp positive trên toàn bộ các dữ liệu mẫu dứa chín (positive). Để tính được recall thì phải biết trước nhãn của dữ liệu. Do đó, recall có thể được dùng để đánh giá trên tập train và validation vì đã biết trước nhãn. Trên tập test khi dữ liệu được coi như mới hoàn toàn và chưa biết nhãn thì sẽ sử dụng precision.

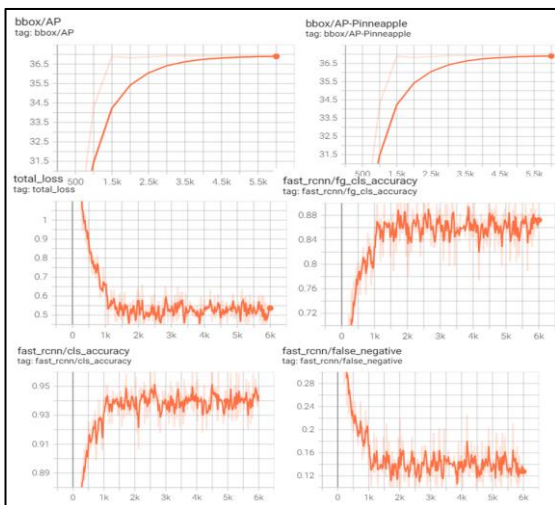
Các chỉ số TP, TN đại diện cho giá trị dự đoán chính xác, FP, FN đại diện cho giá trị dự đoán sai lệch. Cụ thể:

TP: Giá trị dự đoán chính xác; Khi mô hình dự đoán đúng (lớp Positive) cho cho trái dứa chín

FP: Giá trị dự đoán sai lệch; Khi mô hình dự đoán đúng (lớp Positive) cho trái dứa chín.

TN: Giá trị dự đoán đúng; Khi mô hình dự đoán sai (lớp Negative) cho cho trái dứa chưa chín.

FN: Giá trị dự đoán sai lệch; Khi mô hình dự đoán sai (lớp Negative) cho cho trái dứa chưa chín.



Hình 5. Sơ đồ hàm tính toán đánh giá hiệu suất mô hình huấn luyện

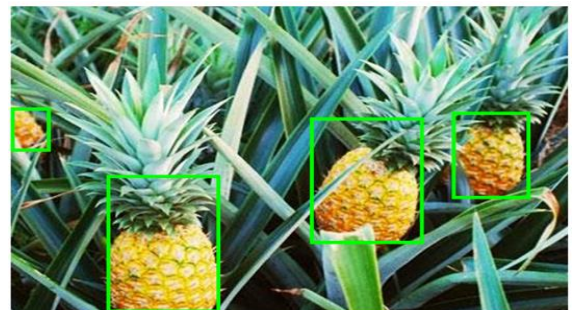
Hình 5 cho thấy, biểu đồ tính toán độ chính xác đạt tối đa từ 90% cho tới 95%. Với đường cong rõ rệt khi sử dụng α hệ số nhỏ để có thể tính toán và không bị vượt ngưỡng sau mỗi lần lặp. Sơ đồ biểu thị hàm tính toán mất mát và giá trị sai được giảm rõ rệt, gần đạt được ngưỡng tối ưu toàn cục. Việc sử dụng nhiều hình ảnh trái dứa với kích cỡ khác nhau được cho trước sẽ giúp kết quả nghiên cứu có thể đạt được độ chính xác cao. Kết quả nhận dạng trái dứa bằng hình ảnh được thử nghiệm tại Hình 6 (a) và đầu ra được hiển thị nhận dạng trái dứa với độ chính xác đạt 95,624% trong Hình 6 (b). Ngoài ra, Hình 7 và Hình 8 thể hiện kết quả huấn luyện từ mô phỏng thuật toán Fast R-CNN với khả năng phát hiện và nhận dạng hình ảnh trái dứa nằm xen lẫn trong vùng lá dứa được chụp trên cánh đồng dứa với kết quả dự đoán đạt được độ chính xác cao, đạt tỷ lệ 95,62) cho phép dễ có thể thực nghiệm tại nông trường thực tế.



Hình 6. (a) Dữ liệu đầu vào (b) đầu ra khi xác nhận độ chính xác nhận dạng từng đối tượng quả dứa



Hình 7. Bước huấn luyện thuật toán Fast R-CNN



Hình 8. Mô phỏng khả năng nhận dạng trái dứa bằng thuật toán Fast R-CNN

5. Kết luận

Bài báo tập trung cho việc ứng dụng kỹ thuật xử lý hình, thị giác máy tính và kỹ thuật học máy để nhận dạng hình ảnh trong lĩnh vực trái cây chủ lực của nông nghiệp Việt Nam với quy mô lớn. Kích thước, màu sắc, hình dạng, kết cấu và các điểm khuyết tật là những đặc điểm quan trọng

của trái dưa trong việc trích xuất đặc trưng. Với mục đích giúp cho người nông dân, giảm sức lao động bằng tay chân thay vào đó là kiểm tra trái dưa bằng hệ thống quan sát. Hệ thống được triển khai cung cấp kết quả xác thực, độ chính xác và tốc độ xử lý nhanh.

Trong nghiên cứu này, nhóm tác giả tập trung sử dụng mô hình học sâu (Deep Learning) để nhận dạng, phát hiện trái dưa bằng cách triển khai Fast R-CNN cải tiến. Kết quả thử nghiệm cho thấy, hệ thống hoạt động tốt để có thể nhận dạng và sẵn sàng cho các ứng dụng thực tế với chất lượng phát hiện (mAP) có tỷ lệ cao hơn R-CNN, đồng thời giảm tải bộ nhớ đệm khi truy xuất dữ liệu. Qua phân tích kết quả thực nghiệm đã chỉ ra được những điểm mạnh, điểm yếu trong thuật toán nghiên cứu trước đây và nghiên cứu cải tiến của nhóm tác giả. Trong tương lai, nhóm tác giả dự định sẽ triển khai cải tiến kết hợp giữa thuật toán Fast R-CNN và thuật toán đếm trái dưa dựa trên các hộp dự đoán với khả năng không lặp lại thông qua việc theo dõi đối tượng. Ngoài ra, làm cơ sở xây dựng hệ thống nhận dạng, phát hiện trái cây chín ở giai đoạn thu hoạch và khuyến nghị hỗ trợ người nông dân trong việc canh tác và tính toán sản lượng xuất khẩu.

Lời cảm ơn: Nhóm tác giả xin gửi lời cảm ơn sâu sắc tới Bộ Giáo dục và Đào tạo đã hỗ trợ dự án nghiên cứu này. Dự án nằm trong Chương trình Khoa học và Công nghệ của Bộ Giáo dục và Đào tạo. Mã số: B2021.DNA.09. “Nghiên cứu ứng dụng mô hình học sâu nhận biết thời kỳ chín của trái Quảng Nam - Đà Nẵng”.

TÀI LIỆU THAM KHẢO

- [1] Jiuxiang Gua,*, Zhenhua Wangb,*, Jason Kuenb, Lianyang Mab, Amir Shahroudyb..., “Recent Advances in Convolutional Neural Networks”, arXiv:1512.07108v6 [cs.CV] 19 Oct 2017, pp 01-38.
- [2] Girshick, R. “FAST R-CNN”, In *Proceedings of the IEEE international conference on computer vision*, 2015, 1440–1448.
- [3] Uijlings, J. R., Van De Sande, K. E., Gevers, T., and A. W. Smeulders, “Selective search for object recognition”. *Int. J. Comput. Vis.* 104, doi: 10.1007/s11263-013-0620-5, 154–171.
- [4] Byoungjun Kim, You-Kyoung Han, Jong-Han Park and Joonwhoan Lee1, “Improved Vision-Based Detection of Strawberry Diseases Using a Deep Neural Network”, 2020, <https://doi.org/10.3389/fpls.2020.559172>.
- [5] Jose Luis Rojas-Aranda, Jose Ignacio Nunez-VarelaJ. C. Cuevas-TelloGabriela Rangel-Ramirez. “Fruit Classification for Retail Stores Using Deep Learning”, 2021, doi:10.1007/978-3-030-49076-8_1.
- [6] Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi. “You Only Look Once: Unified, Real-Time Object Detection”, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, arXiv:1506.02640v5 [cs.CV] 9 May 2016.
- [7] Horea MURESAN an Mihai OLTEAN, “Fruit recognition from images using deep learning”, *Acta Univ. Sapientiae, Informatica* 10, June 2018, pp. 26-42.
- [8] Md Tohidul Islam, Sagidur Rahman, B.M. Nafiz Karim Siddique and Taskeed Jabid, “Image Recognition with Deep Learning”, *ICIIBMS*, 2018, pp. 106-110.
- [9] Emmanuel Karlo Nyarko, Ivan Vidovic, K. Radocaj and Robert Cupec, “A nearest neighbor approach for fruit recognition in RGB-D images based on detection of convex surfaces”, *Expert Systems With Applications*, Elsevier Ltd., July 2018, pp. 454-466, 24.
- [10] Tao Yongting and Zhou Jun, “Automatic apple recognition based on the fusion of color and 3D feature for robotic fruit picking”, *Computers and Electronics in Agriculture*, Elsevier B.V., July 2018, pp. 388-396.
- [11] Walter Maldonado Jr. and Jose Carlos, “Automatic green fruit counting in orange trees using digital images”, *Computers and Electronics in Agriculture*, Elsevier B.V., 27July 2016, pp. 572-581.
- [12] Nguyễn Thị Diệu Phương, “Ứng dụng mạng nơ ron trong nhận dạng mẫu và đánh giá chất lượng quả Xoài”, Luận văn thạc sỹ, Đại học Đà Nẵng, năm 2021.
- [13] Shadman Sakib, Zahidun Ashrafi and Md. Abu Bakr Siddique, “Implementation of Fruits Recognition Classifier using Convolutional Neural Network Algorithm for Observation of Accuracies for Various Hidden Layers”, *ArXiv e-Journal*, 01 April, 2019, pp. 1-4.
- [14] Muresan, H., and Oltean, M, “Fruit recognition from images using deep learning. *Acta Univ*”, *Sapientiae Inform*, 10, 2018, 26–42. doi: 10.2478/ausi.
- [15] Xiongwei Wu, Doyen Sahoo, Steven C.H. Hoi, “Recent Advances in Deep Learning for Object Detection”, *Computer Vision and Pattern Recognition*, Aug. 2019, arXiv:1908.03673v1 (cs.CV), pp 1-40.
- [16] K. Simonyan and A. Zisserman. “Very deep convolutional networks for large-scale image recognition”, *Computer Science, Computer Vision and Pattern Recognition*, In ICLR, 2015, pp 01-14.
- [17] Santagapita, P.R., Tylewicz, U. Panarese, V., Rocculi, P, Dalla, Rosa, M. “Nondestructive assessment of kiwifruit physic-chemical parameters to optimize the osmotic dehydration process”, A study on FT-NIR spectroscopy. *Journal of Biosyst. Eng.* 142, (2), 2016, 101-129.
- [18] Stajanko, Denis and Lakota, Miran and Hočevcar, Marko. *Estimation of number and diameter of apple fruits in an orchard during the growing season by thermal imaging*. *Computers and Electronics in Agriculture*. 42. 2004, 31-42.
- [19] Nguyen H.H.C., Luong A.T., Trinh T.H., Ho P.H., Meesad P., Nguyen T.T. “Intelligent Fruit Recognition System Using Deep Learning”. In: Meesad P., Sodsee D.S., Jitsakul W., Tangwannawit S. (eds) *Recent Advances in Information and Communication Technology* 2021. IC2IT 2021. Lecture Notes in Networks and Systems, vol 251. Springer, Cham. https://doi.org/10.1007/978-3-030-79757-7_2.
- [20] Nguyen, H.H.C., Nguyen, D.H., Nguyen, V.L., Nguyen, T.T., “Smart solution to detect images in limited visibility conditions based convolutional neural networks”, In: *Advances in Computational Collective Intelligence. ICCCI 2020. Communications in Computer and Information Science*, vol. 1287, pp. 641–650. Springer, Cham (2020). <https://doi.org/10.1007/978-3-030-63119-2-52>.