

# HỆ THỐNG NHẬN DẠNG QUẢ MĂNG CỤT CHÍNH DỰA TRÊN MÔ HÌNH FASTER R-CNN CẢI TIẾN

Trịnh Trung Hải<sup>1</sup>, Hồ Phan Hiếu<sup>2</sup>, Nguyễn Hà Huy Cường<sup>3</sup>, Ninh Khánh Duy<sup>4</sup>

<sup>1</sup>Trường Đại học Công nghệ Thông tin và Truyền thông Việt - Hàn, Đại học Đà Nẵng

<sup>2</sup>Văn phòng Đại học Đà Nẵng

<sup>3</sup>Trung tâm Phát triển Phần mềm, Đại học Đà Nẵng

<sup>4</sup>Trường Đại học Bách khoa, Đại học Đà Nẵng

thai@vku.udn.vn, hophanhieu@ac.udn.vn, nhhcuong@sdc.udn.vn, nkduy@dut.udn.vn

**TÓM TẮT:** Xu hướng công nghệ trí tuệ nhân tạo được triển khai ứng dụng mạnh mẽ trong mọi lĩnh vực đời sống kinh tế - xã hội, trong đó nông nghiệp (hay nền nông nghiệp thông minh) đang được Chính phủ Việt Nam quan tâm, chú trọng phát triển. Nhận thấy nền nông nghiệp truyền thống đang bị lạc hậu, tốn nhiều chi phí nguồn nhân lực. Chính vì thế, cần vận dụng những thành tựu khoa học để giải quyết các vấn đề tự động hóa các khâu trong quy trình trồng trọt và chế biến nông sản. Việc phân loại trái cây khi vào thời kỳ thu hoạch đang là phương pháp kỹ thuật mới và mang những yếu tố tích cực trong lĩnh vực công nghệ cao. Trong bài báo này, chúng tôi tập trung trình bày nghiên cứu giải pháp phân loại quả măng cụt chín thông qua các kỹ thuật máy học và áp dụng các mô hình học sâu, mạng nơron tích chập tiêu biểu là Faster R-CNN cải tiến dựa trên hình ảnh được chụp từ một số vùng trồng măng cụt. Sử dụng mạng đề xuất khu vực để trích xuất vùng hình ảnh chứa vật phẩm để phân loại và xác định vị trí của măng cụt và ROI (Vùng quan tâm). Thuật toán tổng hợp bằng cách thêm lớp Cân chỉnh ROI để tối ưu hóa dữ liệu tính năng trong quá trình huấn luyện. Qua kỹ thuật lập trình và mô phỏng thực nghiệm với phương pháp đề xuất của chúng tôi đã cải thiện cả tốc độ lẫn độ chính xác trong việc xử lý các tập dữ liệu lớn và phức tạp. Khi sử dụng bộ dữ liệu gồm 5.000 ảnh chụp quả măng cụt chín và nhiều quả măng cụt kết quả mô hình của chúng tôi vượt trội hơn so với phương pháp tiếp cận một giai đoạn về độ chính xác trong khi vẫn duy trì tốc độ thời gian thực với độ chính xác đạt tối đa từ 94% cho tới 98,7%.

**Từ khóa:** Nhận dạng quả măng cụt, xử lý hình ảnh, Faster R-CNN cải tiến, dự đoán và phân loại, CNNs.

## I. GIỚI THIỆU

Trí tuệ nhân tạo (AI), Internet vạn vật (IoT) hay dữ liệu lớn (Big Data) là những thành tựu nổi bật của CMCN 4.0 tạo ra sự chuyển biến đột phá về công nghệ khoa học phục vụ mô hình sản xuất mới, tác động mạnh mẽ tới kinh tế chính trị và đời sống xã hội. Trí tuệ nhân tạo đang nhanh chóng trở thành một trong những lĩnh vực khoa học rất tiềm năng được mong đợi nhất, sẽ đem lại lợi ích to lớn cho nhiều ngành công nghiệp. Ứng dụng trí tuệ nhân tạo vào các lĩnh vực, trong đó ngành Nông nghiệp được sử dụng rộng rãi và mang nhiều kết quả hữu ích như: Máy kéo không người lái, máy thu hoạch tự động, hệ thống đo độ ẩm của đất phục vụ cho tưới tiêu nông sản. Đặc biệt, nhiều nghiên cứu đã cho thấy việc sử dụng trí tuệ nhân tạo có thể làm tăng tốc độ và đo độ chính xác cho chẩn đoán nông nghiệp, đảm bảo chất lượng thực phẩm.

Xử lý hình ảnh và thị giác máy tính được sử dụng để giải quyết bài toán trong thực tế về phát hiện và nhận dạng đối tượng từ hình ảnh. Nhận dạng trái cây từ hình ảnh thuộc trong số đó, hiện nay đang được triển khai nghiên cứu và đưa vào thực tế. Nhận dạng trái cây tiêu biểu là măng cụt (Mangosteen) là một thách thức lớn vì quả măng cụt có những đặc điểm giống nhau về hình thái và màu sắc. Nhận dạng hình ảnh cần nhiều sức mạnh tính toán với nguồn dữ liệu thu thập lớn. 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 quả măng cụt chín. Sử dụng mạng nơron tích chập (CNNs) để trích xuất đặc trưng bề ngoài của quả măng cụt nhằm phục vụ cho quá trình huấn luyện để tiến hành phân loại và dự đoán quả măng cụt chín.

Măng cụt là một loại cây ăn quả nhiệt đới được trồng phổ biến ở các nước Đông Nam Á, ở Việt Nam măng cụt được trồng nhiều nhất ở Lái Thiêu, tỉnh Bình Dương. Quả măng cụt có giá trị dinh dưỡng và giá trị sử dụng rất cao; là một loại quả được xuất khẩu sang nhiều nước trên thế giới, mang về giá trị kinh tế lớn. Đặc biệt là thị trường Trung Quốc, hiện Việt Nam được phép xuất khẩu chính ngạch có 09 loại quả gồm chuối, xoài, măng cụt, dưa hấu, sầu riêng, nhãn, chôm chôm, thanh long, vải thiều, trong đó có măng cụt có tiềm năng rất lớn. Tuy vậy phần lớn việc trồng trọt, canh tác, thu hoạch, bảo quản, chế biến,... phục vụ xuất khẩu thực hiện thủ công, chưa được đầu tư nghiên cứu và hỗ trợ phương tiện kỹ thuật công nghệ cao để tăng năng suất, chất lượng quả chín và tăng giá trị xuất khẩu cho quả măng cụt chín, tạo nên nền phát triển thịnh vượng nhờ loại quả cây này và giúp mang lại thu nhập cho người nông dân.

Trong bài báo này, ngoài phần giới thiệu, chúng tôi trình bày một số nghiên cứu liên quan ở Mục II, phần giải pháp đề xuất trình bày ở Mục III, kết quả nghiên cứu được trình bày ở Mục IV, phần kết luận được trình bày ở Mục V và cuối cùng là tài liệu tham khảo.

## II. CÁC NGHIÊN CỨU LIÊN QUAN

Trong những năm gần đây, các thuật toán cải tiến từ mạng CNNs đang được ứng dụng ngày càng hiệu quả cho phép đạt tới ngưỡng chính xác tuyệt đối, cho kết quả cao trong các bài toán nhận dạng và phân loại hình ảnh, nhận

dạng văn bản,... Mạng 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 các đơn vị hình ảnh. Thuật toán Faster R-CNN được cải tiến từ đó, với khả năng phát hiện đối tượng, thuật toán thực hiện 2 bước chính: (1) Trích xuất bản đồ đặc trưng bằng cách cho tệp hình ảnh qua mô hình huấn luyện; (2) Bản đồ đặc trưng được dùng cho Mạng đề xuất khu vực (Region Proposal Networks - RPN) để lấy được các khu vực, sau khi lấy được vị trí các khu vực thì thực hiện tương tự Fast R-CNN [1, 2]. Một đề xuất khác You Only Look Once (YOLO) [3], một máy dò áp dụng mạng nơron duy nhất cho toàn bộ hình ảnh, sau đó chia hình ảnh thành các vùng và dự đoán các hộp giới hạn. Byoungjun Kim và cộng sự [4] sử dụng mô hình mạng thần kinh sâu (DNN) để cải thiện khả năng phát hiện bệnh trên trái dâu tây. 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 cách tiếp cận mới để phát hiện đối tượng sử dụng hộp giới hạn và xác suất lớp liên quan, tối ưu hóa end-to-end trực tiếp dựa trên hiệu suất phát hiện.

Trong lĩnh vực nông nghiệp công nghệ cao, nhóm nghiên cứu Horea và cộng sự [7] đã đề xuất một hệ thống huấn luyện theo kỹ thuật mạng thần kinh 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. Md Tohidul Islam và cộng sự [9] đề xuất một hệ thống sử dụng mạng nơron phức hợp để phân loại hình ảnh thực phẩm. Mạng nơron phức hợp sử dụng một mô hình máy tính để mô phỏng khả năng của bộ não thành cỗ máy nhận biết và phân biệt 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 rôbot thông minh. Tao Yongting và Zhou Jun [11] đã đề xuất hệ thống nhận dạng chính xác để thu hoạch quả táo tự động bằng rôbot. 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, Sobel và làm mờ Gaussian, phát triển một thuật toán đếm số quả cam xanh, có thể tạo ra công cụ hữu ích cho các nông dân trong quản lý trái cây trong vườn. Nhóm nghiên cứu Xiangqin Wei và cộng sự [13] đã mô tả phương pháp tự động trích xuất các đối tượng trái cây dưới môi trường phức tạp sử dụng rôbot hái trái cây. Nhóm nghiên cứu Sakib và cộng sự [14] đã đề xuất hệ thống nhận dạng trái cây bằng cách sử dụng Mạng CNNs và thị giác máy tính. Mô hình toán học được xây dựng và thực thi bằng python, kế thừa sử dụng thư viện TensorFlow, kết quả cho thấy, với số lớp ẩn phù hợp, CNN có thể đạt độ chính xác cao trong việc phân loại các loại trái cây [15, 16].

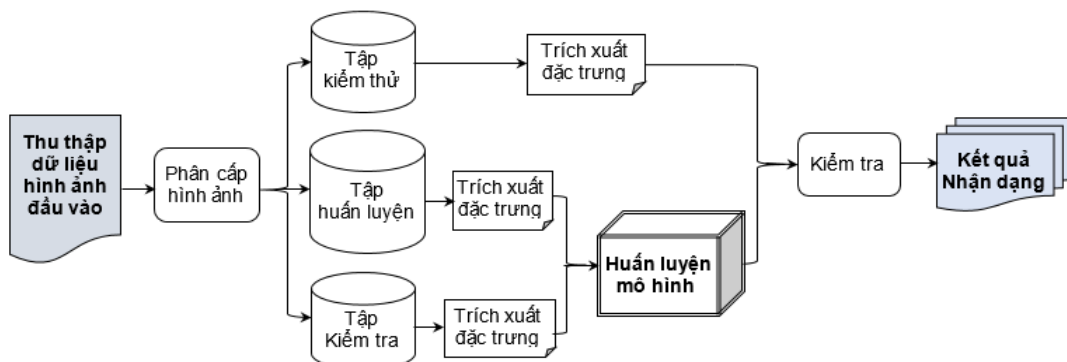
Đặc biệt, các nhà nghiên cứu hiện nay đang rất quan tâm tới việc ứng dụng công nghệ trí tuệ nhân tạo vào các mô hình nông nghiệp, có thể hỗ trợ nông dân trong công việc chăm sóc định kỳ, tăng chất lượng đầu ra của trái cây, từ đó có thể phát triển nền nông nghiệp vững mạnh.

Trong các công trình nghiên cứu của TS. Nguyễn Hà Huy Cường và cộng sự, đã nghiên cứu ứng dụng kỹ thuật YOLO-v5, trong việc nhận dạng và phát hiện đối tượng dựa vào hình ảnh, camera cho bốn lớp trái dứa ở các thời kỳ: Trái dứa non, trái dứa chưa chín (trái già), trái dứa chín, trái dứa hồng [17, 18]. Ngoài ra, nhóm chúng tôi cũng đã nghiên cứu thực nghiệm, đánh giá hiệu suất và so sánh ưu nhược điểm khi sử dụng mô hình Fast R-CNN, Faster R-CNN và YOLO trên tập số dữ liệu quả măng cụt đã xây dựng [19, 20].

### III. GIẢI PHÁP ĐỀ XUẤT

#### A. Thiết kế hệ thống

Hệ thống nhận dạng quả măng cụt sử dụng thị giác máy tính và kỹ thuật xử lý hình ảnh để phát hiện và phân loại quả măng cụt chín bằng cách sử dụng Mạng CNNs để phân tích, trích xuất đặc trưng các đặc điểm biểu hiện qua hình ảnh, đồng thời tăng cường dữ liệu hình ảnh và xử lý các thuộc tính của dữ liệu đầu vào. Chúng tôi sử dụng phương pháp tiếp cận mô hình Faster R-CNN cải tiến để khắc phục các hạn chế của R-CNN tăng tốc độ huấn luyện và cho ra kết quả có tỷ lệ chính xác cao. Đề xuất phương pháp cải tiến thêm lớp ROI Align để tối ưu dữ liệu trích xuất vùng đặc trưng trên bề mặt hình ảnh quả măng cụt và thu hẹp vùng quan tâm ROI từ toàn ảnh về 1 vùng ảnh nhỏ hơn ảnh gốc để xác định vị trí đối tượng chính xác hơn, giảm chi phí tính toán của mô hình [21, 22].



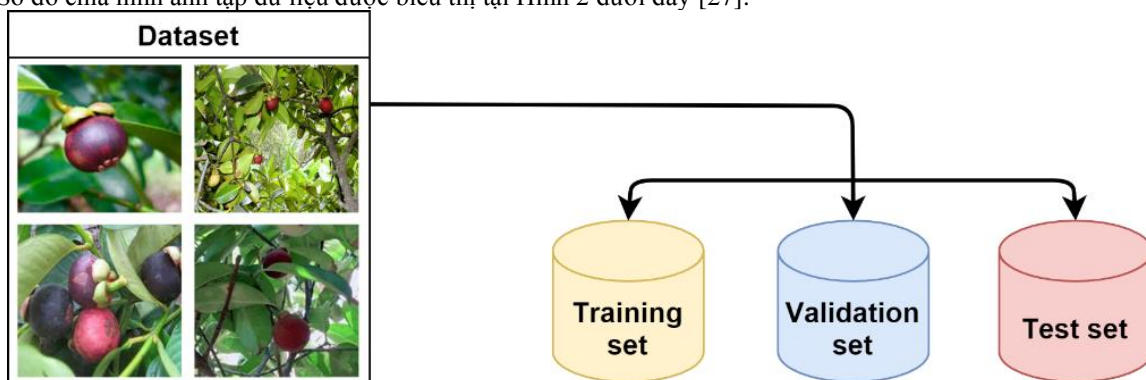
Hình 1. Quy trình thiết kế hệ thống nhận dạng quả măng cụt cơ bản

Hệ thống được xử lý các bước sau (Hình 1): Bước đầu tiên sẽ cung cấp một tập dữ liệu hình ảnh và tiến hành lưu trữ tập dữ liệu vào bộ nhớ hệ thống. Bước 2: Tập dữ liệu hình ảnh sẽ được tiền xử lý ảnh thô với các bước: loại bỏ tạp ảnh, các hình ảnh không nhìn rõ, đồng nhất kích thước ảnh và phân cấp dữ liệu ảnh (Tập huấn luyện, Tập kiểm tra và Tập kiểm thử). Bước 3: Hệ thống sẽ tiến hành trích xuất đặc trưng của hình ảnh, trên cơ sở màu sắc, hình dạng, kích thước và kết cấu bề mặt quả, để xác định các khuyết tật có trên quả măng cụt. Bước 4: Hệ thống nhận dạng sẽ kiểm tra, so khớp hình ảnh tồn tại trong tập dữ liệu và tiến hành dự đoán với mô hình huấn luyện hoàn chỉnh và mô hình Faster R-CNN cải tiến nhằm xác định độ chín của quả măng cụt. Hệ thống nhận dạng sẽ kiểm tra, so khớp hình ảnh tồn tại trong tập dữ liệu và tiến hành dự đoán nhằm xác định độ chín của đối tượng quả măng cụt trong ảnh. Bước 5: Trả về kết quả nhận dạng hình ảnh quả măng cụt chín hay chưa chín.

## B. Phương pháp xử lý

### 1. Tập dữ liệu quả măng cụt đầu vào

Dữ liệu hình ảnh đầu vào được chúng tôi tự thu thập bằng các thiết bị quay chụp chuyên dụng. Số lượng hình ảnh đạt được là 5.000 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 1.000 hình ảnh để kiểm tra đánh giá, chúng tôi trích 3.800 hình ảnh đầu tiên để phục vụ cho huấn luyện. Tập dữ liệu kiểm thử được tạo ra với việc sử dụng 200 hình từ tập dữ liệu. Sơ đồ chia hình ảnh tập dữ liệu được biểu thị tại Hình 2 dưới đây [27].

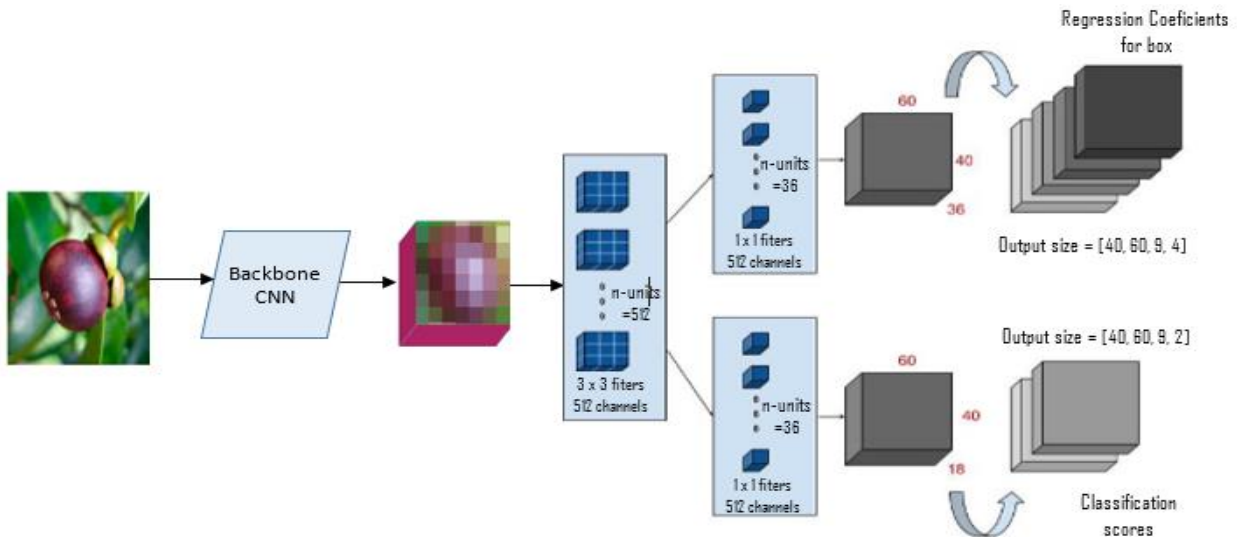


Hình 2. Cấu trúc phân bố của tập dữ liệu hình ảnh

### 2. Sử dụng mô hình huấn luyện Faster R-CNN

Hiệu suất cao do học sâu mang lại, một tập hợp con của học máy, nó đang trở nên rất phổ biến trên nhiều loại dữ liệu. CNN được xây dựng để phân loại hình ảnh sử dụng kỹ thuật học sâu. Xây dựng mô hình CNN dễ dàng hơn với việc sử dụng thư viện Keras trong 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 được sử dụng để tổng hợp tối đa dữ liệu, giảm kích thước của dữ liệu không gian bằng cách chọn giá trị lớn nhất trong một vùng cụ thể của dữ. Dữ liệu không gian là biểu diễn thông tin về một đối tượng vật lý bằng các giá trị số, trong đó mỗi giá trị tương ứng với một thuộc tính của đối tượng đó, ví dụ như tọa độ, độ cao, màu sắc, độ sáng, độ dốc... 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 bản đồ đối tượng, 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 nhân tạo [20, 21, 27].

Việc huấn luyện mô hình R-CNN có sẵn tồn tại rất nhiều tài nguyên và chậm vì mỗi hình ảnh yêu cầu phân loại các lớp với khoảng 2.000 đề xuất khu vực, do đó, thời gian huấn luyện rất lâu 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. Do đó, chúng tôi triển khai tính năng phát hiện đối tượng với R-CNN được nâng cấp lên Faster R-CNN để khắc phục các hạn chế của R-CNN và cải thiện tốc độ [24, 25, 26]. Sử dụng phương pháp kết hợp thuật toán Faster R-CNN được cải tiến để phân loại và xác định vị trí quả măng cụt chín, nhằm tăng tốc độ thời gian thực cho việc phát hiện đối tượng. Để truy xuất các đề xuất khu vực, Faster R-CNN tiếp tục sử dụng tìm kiếm có chọn lọc. Faster R-CNN không dùng thuật toán tìm kiếm chọn lọc để lấy ra các khu vực, mà nó thêm một mạng CNN mới đó là mạng RPN để tìm kiếm các khu vực, không tách đề xuất vùng 2.000 khỏi hình ảnh và thực hiện phân loại hình ảnh cho từng hình ảnh riêng biệt. Nó sẽ tạo ra một bản đồ 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 để đưa các vùng hình ảnh về kích thước hình vuông và định hình lại với cùng kích thước đầu ra và sử dụng lớp ROI Align để cải tiến việc lấy mẫu, ROI Align sử dụng trung bình có trọng số của các giá trị trong một ô vuông nhỏ hơn để tính toán giá trị của bản đồ đặc trưng cho vùng ROI. Tiếp tục qua các lớp kết nối đầy đủ cho đến khi đạt đến đầu ra vector đặc trưng ROI [22, 23, 27]. Để có được kết quả cuối cùng, mô hình sẽ dự đoán lớp và độ lệch của hộp giới hạn cho vùng phân ảnh đó.

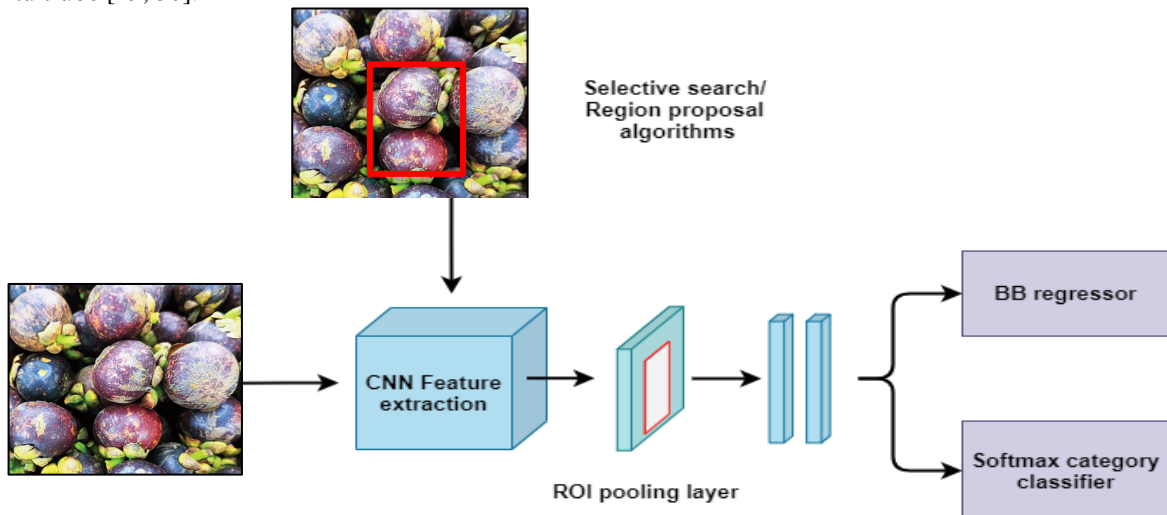


**Hình 3.** Trích xuất đặc trưng cho quả măng cụt sử dụng RPN

### 3. Xây dựng hệ thống nhận dạng quả măng cụt

Dựa trên mô hình huấn luyện đã hoàn thiện, chúng tôi đã phát triển một hệ thống có khả năng phân tích và đánh giá hình ảnh đầu vào bằng phương pháp, áp dụng kỹ thuật Faster R-CNN đã được đào tạo hoàn thiện (Re-Train); sử dụng mô hình dò tìm 2 giai đoạn Tow Stage Detector với thuật toán Mạng khu vực đề xuất (RPN) tinh chỉnh và khởi tạo bộ phân loại hình ảnh. Đầu tiên hình ảnh được đưa vào lớp ConvNet (lớp chập + lớp tổng hợp tối đa) để tạo ra một bản đồ đặc trưng phức hợp. Tiếp theo hình ảnh được đưa vào vùng đề xuất (RPN) để trích xuất ra những vùng có đối tượng, sau đó sử dụng lớp ROI Pooling để giới hạn các vùng hình ảnh định hình với cùng kích thước đầu ra, đồng thời chèn thêm lớp ROI Align để tối ưu dữ liệu trích xuất vùng đặc trưng trên bề mặt hình ảnh và cuối cùng chuyển qua các lớp Full Connected để phân loại và xác định vị trí đối tượng đưa ra kết quả các khung hộp đối tượng [27].

Bên cạnh đó, chúng tôi còn phát triển thêm giao diện người dùng trên các nền tảng ứng dụng Windows, IOS, Android... Được xây dựng bằng Framework Django với ngôn ngữ lập trình Python cho ra giao diện thân thiện và dễ dàng sử dụng. 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 [29, 30].



**Hình 4.** Kiến trúc hệ thống huấn luyện bằng thuật toán Faster R-CNN

## IV. KẾT QUẢ THỰC NGHIỆM VÀ ĐÁNH GIÁ HIỆU SUẤT

### A. Kết quả thực nghiệm

Chúng tôi đã tạo ra một hệ thống có khả năng phân tích và đánh giá ảnh dựa trên mô hình đã được huấn luyện. Các ảnh đầu vào bao gồm quả măng cụt còn chín và chưa chín được nhận dạng bằng Faster R-CNN, với các bộ phân loại được cải tiến để đảm bảo dữ liệu không mất quá nhiều trong quá trình huấn luyện. Thông qua các nghiên cứu thực tế, bộ phân loại có thể giúp các nông dân phân loại ảnh của quả măng cụt còn chín, tìm kiếm quả măng cụt ở những nơi kín đáo, nhỏ và phát hiện sâu bệnh gây hại cho chất lượng quả.

Chúng tôi đã huấn luyện trong 5.000 hình ảnh để kiểm tra độ chính xác trung bình (Mean Average Precision). Giá trị dự đoán được tính bằng hai chỉ số Precision và Recall, được đặc trưng bởi bốn tham số: True Positive (TP), True Negative (TN), False Positive (FP) và False Negative (FN). Ngoài ra còn sử dụng giá trị GTI (Global Template Image) thực hiện phép so sánh một bức ảnh mẫu với một ảnh khác và nhận dạng đối tượng trong ảnh đó. Hình 5 là biểu đồ 2D dựa trên hai chỉ số trên cho thấy độ chính xác cao nhất là 98,7%, sự biến thiên trung bình khoảng 95% và tỷ lệ False Negative giảm rõ rệt từ 16% xuống còn 11%. Điều này là cần thiết để xác định xem mô hình huấn luyện đạt độ chính xác cao hay không. Chúng tôi cũng đánh giá sự mất mát của việc xác định khung giới hạn và đầu ra phân loại để xác nhận rằng chúng thực sự hiệu quả hơn so với Fast-CNN. Khi xem xét tất cả các tình huống của thực nghiệm, vấn đề mất dữ liệu khá nhỏ và kỹ thuật đạt được kết quả kiểm tra rất tốt.

### B. Đánh giá hiệu suất

Để đánh giá mô hình, chúng tôi so sánh kết quả từ mô hình Faster R-CNN trên tập xác nhận đã thu thập từ trước. Hệ số IOU đạt được 0,5 hoặc lớn hơn GTI. 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. Precisions, Recall và mAP được sử dụng để đánh giá hiệu suất mô hình và được tính như dưới đây:

$$IoU = \frac{Area\_of\_Overlap}{Area\_of\_Union} \quad (1) \quad IoU = \frac{TP}{(TP + FP + FN)} \quad (2)$$

$$Precision = \frac{TP}{TP + FP'} \quad (3) \quad Recall = \frac{TP}{TP + FN} \quad (4)$$

$$F1 = \frac{2Precision \times Recall}{Precision + Recall} \quad (5)$$

trong đó, TP đại diện cho giá trị đạt tiêu chuẩn, FP đại diện cho giá trị không đạt tiêu chuẩn và FN đại diện cho giá trị sai. 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.

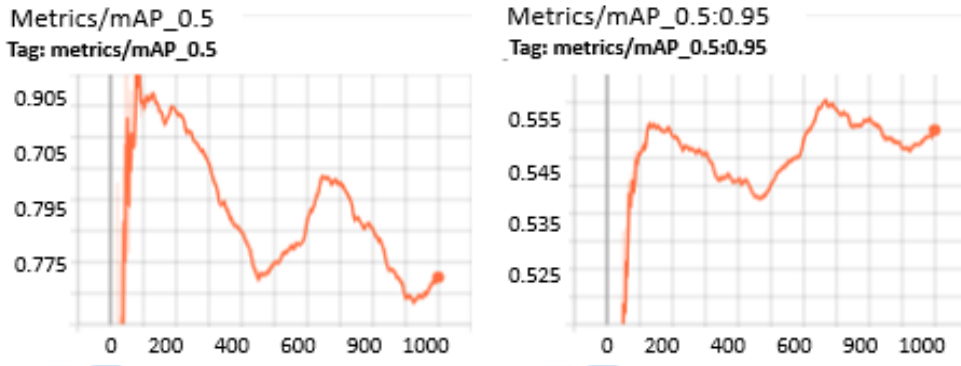
**Bảng 1.** So sánh độ chính xác và tốc độ xử lý trung bình của các mô hình trên cùng 1 lô hình ảnh

Mô hình thử nghiệm	Độ chính xác	Thời gian xử lý mỗi ảnh (s)
FPN	78,5%	0,136
R-CNN	73,6%	0,231
Mark R-CNN	89,6%	0,135
Fast R-CNN	87,4%	0,142
Faster R-CNN	94%	0,103
<b>Faster R-CNN Improved</b>	<b>98,7%</b>	<b>0,101</b>

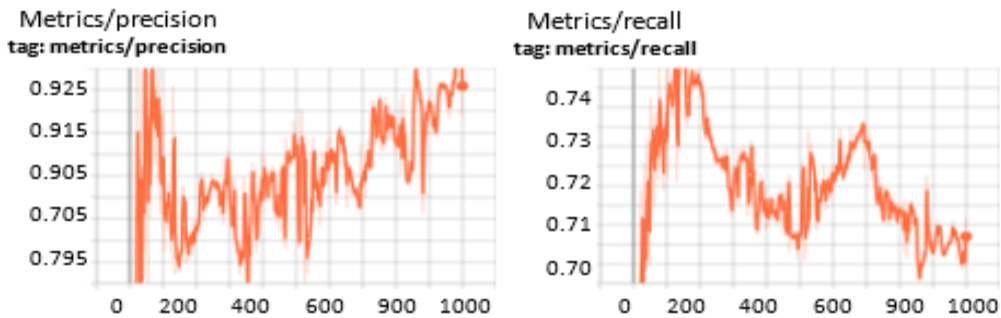
**Bảng 2.** Tỷ lệ chính xác của các lô hình ảnh khác nhau được thực hiện trên cùng mô hình Fater R-CNN

Lô dữ liệu	Độ đo đánh giá kết quả thực nghiệm		
	Precision (%)	Recall (%)	mAP (%)
Mangosteens_01	98,7	84,3	93,5
Mangosteens_02	92,2	88,5	78,2
Mangosteens_03	89,5	91,2	81,4
Mangosteens_04	87,2	93,0	86,2
Mangosteens_05	85,5	95,1	79,2

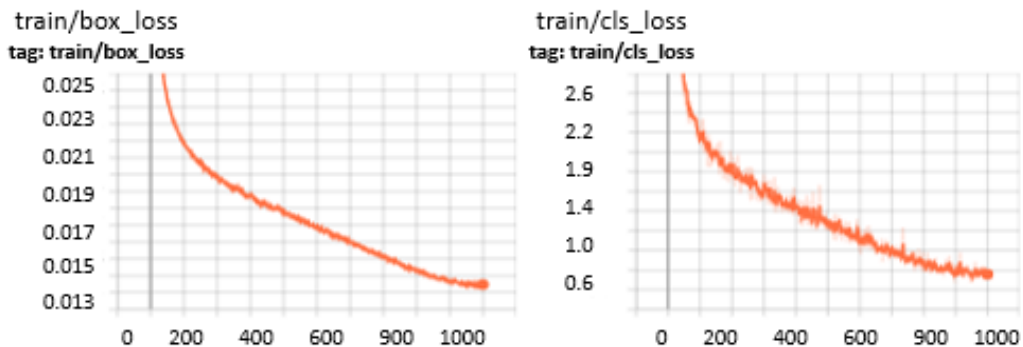
Số liệu tổng hợp ở Bảng 1 so sánh giá trị thời gian thực hiện xử lý trên từng ảnh đã thử nghiệm đối với các mô hình khác nhau nhằm đánh giá tốc độ xử lý của một số mô hình khác nhau thông qua bộ dữ liệu hình ảnh quả măng cụt được xử lý ảnh và đưa vào bộ dữ liệu. Trong đó, Fater R-CNN cho kết quả xử lý tương đối nhanh hơn so với các mô hình khác như FPN, R-CNN, Mark R-CNN, Fast R-CNN. Trong khi đó, số liệu ở Bảng 2 thể hiện giá trị độ đo đánh giá độ chính xác quá trình nhận dạng sử dụng mô hình Fater R-CNN ở các lô dữ liệu khác nhau. Mỗi lô chia trung bình 3.800 hình huấn luyện và tỷ lệ huấn luyện là 0,0001, độ chính xác đạt được trong kịch bản thử nghiệm đạt từ 94% đến 98,7% với tỷ lệ mất mát cross-entropy là 0,02985. Theo đó, chúng tôi đánh giá quá trình trích xuất đặc trưng có ảnh hưởng rất lớn đến độ chính xác của mỗi quá trình học và kiểm tra. Do đó, để đạt được độ chính xác tối đa, cần quan sát cẩn thận hơn để xác định mô hình được sử dụng trong quá trình trích xuất đặc trưng. Tuy nhiên, không phải tất cả các mô hình có độ chính xác tốt đều phù hợp để sử dụng trong trường hợp phát hiện đối tượng trong thời gian thực.



Hình 5. Biểu đồ thể hiện độ đo mAP để đánh giá hiệu suất nhận dạng của mô hình.

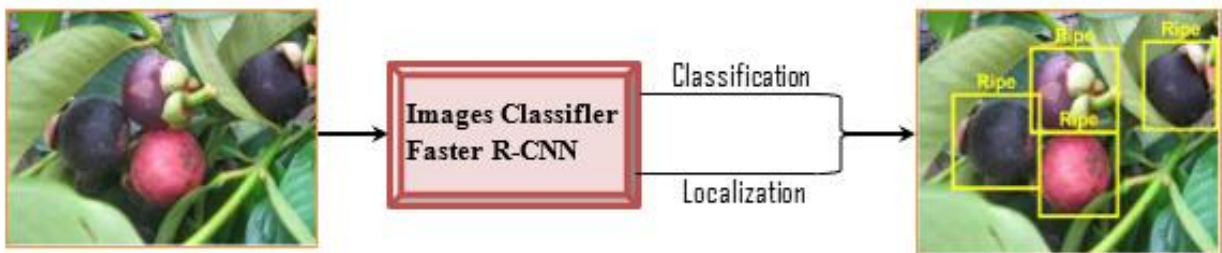


Hình 6. Biểu đồ thể hiện độ đo Precision để đánh giá hiệu suất nhận dạng của mô hình.



Hình 7. Biểu đồ thể hiện tỷ lệ sai lệch của hàm mất mát dựa trên mô hình huấn luyện

Hình 5, Hình 6, và Hình 7 cho thấy sơ đồ tính toán độ chính xác đạt từ tối đa từ 94% cho tới 98,7%. Với đường cong rõ rệt khi sử dụng  $\alpha$  hệ số nhỏ để có thể tính toán và không bị vượt ngưỡng sau mỗi lần lặp. Sơ đồ tính toán Loss và False Negative - biểu đồ giá trị sai được giảm rõ rệt; gần đạt được ngưỡng tối ưu toàn cầu. Việc sử dụng các hình ảnh quả măng cụt với nhiều kích cỡ được cho trước giúp kết quả nghiên cứu có thể đạt được độ chính xác cao. Kết quả nhận dạng quả măng cụt bằng hình ảnh được thể hiện thực nghiệm tại Hình 8 với hình ảnh đầu vào có nhiều quả măng cụt có màu sắc và trạng trạng thái khác nhau, sau xử lý cho ra kết quả nhận dạng là hình ảnh quả chín với nhãn Ripe và quả có đặc điểm biểu hiện chưa chín thì không được gán nhãn. Chúng tôi tạo ra mô hình mô phỏng dự đoán của thuật toán Faster R-CNN cho thấy khả năng phát hiện và phân loại của thuật toán được tối ưu cho mô hình nhận dạng đối tượng, dự đoán kết quả đạt được độ chính xác cho phép để có thể thực nghiệm tại nông trường thực tế.



Hình 8 (a)

Hình 8 (b)

Hình 8. Mô phỏng khả năng nhận dạng quả măng cụt bằng thuật toán Faster R-CNN

## V. KẾT LUẬN

Bài báo sử dụng các kỹ thuật xử lý hình ảnh, thị giác máy tính và đặc biệt ứng dụng mô hình học sâu cải tiến Faster R-CNN trong lĩnh vực nông nghiệp quy mô lớn. Kích thước, màu sắc, hình dạng, kết cấu và các điểm khuyết tật là những đặc điểm quan trọng của quả măng cụt trong việc trích xuất đặc trưng. Để phục vụ cho nhu cầu của người nông dân, giảm sức lao động bằng tay chân thay vào đó là kiểm tra quả măng cụt bằng hệ thống quan sát. Hệ thống được triển khai cung cấp kết quả xác thực, độ chính xác lên đến 98,7% và tốc độ xử lý nhận dạng hình ảnh rất nhanh đạt xấp xỉ 0,1s.

Trong nghiên cứu này, chúng tôi tập trung sử dụng mô hình học sâu để phát hiện và phân loại quả măng cụt bằng cách triển khai Faster R-CNN. 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. Việc nghiên cứu thuật toán học sâu đã đạt được, phát hiện được những điểm mạnh, điểm yếu trong thuật toán. Việc thu thập kinh nghiệm, kiến thức trong quá trình nghiên cứu và xây dựng mô hình nhận dạng quả măng cụt từ hình ảnh. Trong tương lai, chúng tôi dự định sẽ triển khai giữa thuật toán Faster R-CNN và thuật toán đếm quả măng cụt dựa trên các hộp dự đoán với khả năng không lặp lại thông qua việc theo dõi đối tượng; hỗ trợ người nông dân trong việc canh tác và tính toán sản lượng thu hoạch, đáp ứng nhu cầu, kế hoạch xuất khẩu.

## LỜI CẢM ƠN

Nghiên cứu này được tài trợ bởi Bộ Giáo dục và Đào tạo trong đề tài mã số B2023.DNA.19.

## TÀI LIỆU THAM KHẢO

- [1] R. Girshick, "Faster R-CNN," in *Proceedings of the IEEE international conference on computer vision (Santiago)*, pp. 1440-1448, 2015.
- [2] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards RealTime Object Detection with Region Proposal Networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137-1149, 2017.
- [3] J. Redmon, Divvala, S., Girshick, R., and Farhadi, A, "You only look once: unified, real-time object detection," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (Las Vegas, NV)*, pp. 779-788, 2016.
- [4] Byoungjun Kim, You-Kyoung Han, Jong-Han Park and Joonwhoan Lee1, "Improved Vision-Based Detection of Strawberry Diseases Using a Deep Neural Network," *Frontiers in Plant Science*, 11, 559172, 2021.
- [5] Jose Luis Rojas-Aranda, Jose Ignacio Nunez-VarelaJ. C. Cuevas-TelloGabriela Rangel-Ramirez, "Fruit Classification for Retail Stores Using Deep Learning," *Pattern Recognition: 12th Mexican Conference, MCP R 2020, Morelia, Mexico, in Proceedings 12*, Springer International Publishing, pp. 3-13, 2020.
- [6] Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 779-788, 2016.
- [7] Horea MURESAN an Mihai OLTEAN, "Fruit recognition from images using deep learning," *Acta Univ. Sapientiae, Informatica*, 10, pp. 26-42, 2018.
- [8] Susovan Jana, Saikat Basak and Ranjan Parekh, "Automatic Fruit Recognition from Natural Images using Color and Texture Features," *IEEE, Devices for Integrated Circuit (DevIC)*, pp. 620-624, 2017.
- [9] Md Tohidul Islam, Sagidur Rahman, B.M. Nafiz Karim Siddique and Taskeed Jabid, "Image Recognition with Deep Learning," *ICIIBMS*, pp. 106-110, 2018.
- [10] Emmanuel Karlo Nyarko, Ivan Vidovic, Kristijan 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*, pp. 454-466, 2018.
- [11] 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., pp. 388-396, 2017.
- [12] Walter Maldonado Jr. and Jose Carlos Barbosa, "Automatic green fruit counting in orange trees using digital images," *Computers and Electronics in Agriculture*, Elsevier B.V., pp. 572-581, 2016.
- [13] Xiangqin Wei, Kun Jia, Jinhui Land, Yuwei Li, Yiliang Zeng and Chunmei Wang, "Automatic method of fruit object extraction under complex agricultural background for vision system of fruit picking robot," *Optik*, pp. 5684-5689, 2014.
- [14] 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*, pp. 1-4, 2019.
- [15] Y. Tang, Chen, M., Wang, C., Luo, L., Li, J., Lian, G., et al., "Recognition and localization methods for vision-based fruit picking robots: a review," *Front. Plant Sci.* 11:510, 2020. DOI: 10.3389/fpls.2020.00510.
- [16] Muresan, H., and Oltean, M, "Fruit recognition from images using deep learning," *Acta Univ. Sapientiae Inform.*, 10, pp. 26-42, 2018. DOI: 10.2478/ausi-2018-0002.
- [17] H.H.C. Nguyen, 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*, Lecture Notes in Networks and Systems, Springer, Cham, vol 251, 2021.

- [18] H.H.C. Nguyen, 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*, Springer, Cham, vol. 1287, pp. 641-650, 2020.
- [19] N.H. H. Cuong, T. H. Trinh, D.-H. Nguyen, T. K. Bui, T. A. Kiet, P. H. Ho, et al., “An approach based on deep learning that recommends fertilizers and pesticides for agriculture recommendation,” *Int. J. Electr. Comput. Eng.*, vol. 12, no. 5, pp. 5580-5588, 2022.
- [20] N.H. H. Cuong, T. H. Trinh, Meesad, P., Nguyen, T.T., “Improved YOLO object detection algorithm to detect ripe pineapple phase,” *J. Intell. Fuzzy Syst.*, vol. 43, pp.1365-1381, 2022.
- [21] Thang B. Dinh, Van B. Dang, Duc A. Duong, Tuan T. Nguyen, Duy-Dinh Le, “Hand Gesture Classification Using Boosted Cascade of Classifiers,” *Proceedings of 4th IEEE International Conference RIVF 2006*, Ho Chi Minh City, pp 138-143, 2006.
- [22] P. Rajeshwari, P. Abhishek, P. Srikanth, T. Vinod, “Object Detection: An Overview,” *International Journal of Trend in Scientific Research and Development (IJTSRD)*, vol. 3, pp.1663-1665, 2019.
- [23] Xiongwei Wu, Doyen Sahoo, Steven C.H. Hoi, “Recent Advances in Deep Learning for Object Detection,” 2019.
- [24] Yuting Zhang, Kihyuk Sohn, Ruben Villegas, Gang Pan, Honglak Lee, “Improving Object Detection with Deep Convolutional Networks via Bayesian Optimization and Structured Prediction,” *arXiv:1504.03293 [cs.CV]*, 2016.
- [25] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” in *ICLR*, 2015.
- [26] Yongxi Lu, Tara Javidi, Svetlana Lazebnik, “Adaptive Object Detection Using Adjacency and Zoom Prediction,” *arXiv:1512.07711 [cs.CV]*, 2016.
- [27] Trung Hai Trinh, Xuan Thien Bui, Thu Huong Tran, Ha Huy Cuong, Nguyen, Khanh Duy Ninh, “Mangosteen Fruit Detection Using Improved Faster R-CNN,” *The First International Conference on Intelligence of Things (ICIT2022): Intelligence of Things Technologies and Applications, Part of the book series: LNDECT*, volume 148, Aug. 2022.
- [28] Keiron O’Shea, Ryan Nash, “An Introduction to Convolutional Neural Networks,” *arXiv:1511.08458 [cs.NE]*, 2015.
- [29] P.R. Santagapita, Tylewicz, U. Panarese, V., Rocculi, P, Dalla, Rosa, M, “Nondestructive assessment of kiwifruit physico-chemical parameters to optimize the osmotic dehydration process, A study on FT-NIR spectroscopy,” *Journal of Biosyst. Eng.*, 142, (2), pp. 101-129, 2016.
- [30] S. Kim, Y. Ji and K. Lee, “An Effective Sign Language Learning with Object Detection Based ROI Segmentation,” *Second IEEE International Conference on Robotic Computing (IRC)*, Laguna Hills, CA, pp. 330-333, 2018.

## ADVANCED RIPE MANGOSTEEN IDENTIFICATION SYSTEM BASED ON FASTER R-CNN MODEL

Trinh Trung Hai, Ho Phan Hieu, Nguyen Ha Huy Cuong, Ninh Khanh Duy

**ABSTRACT:** The trend of using technology in the Fourth Industrial Revolution is currently being implemented in all areas of social life, including agriculture, also known as smart agriculture, which the Vietnamese government is strongly focusing on. Traditional agriculture is seen as outdated and costly in terms of human resources. Therefore, scientific achievements need to be applied to solve the problems of automating the stages of the farming and processing of agricultural products. Fruit classification during harvest season is a new technical method that brings positive elements to the field of high technology. This article focuses on the research of a solution for classifying ripe Mangosteen fruit through machine learning techniques and the application of representative deep learning models, such as Faster R-CNN convolutional neural networks, based on images taken from some Mangosteen cultivation areas. The proposed network uses a region proposal network to extract image regions containing objects for classification and determining the position of Mangosteen and RoI (Region of Interest). The algorithm is synthesized by adding a RoI alignment layer to optimize feature data during training. Through programming techniques and experimental simulations, our proposed method has improved both speed and accuracy in processing large and complex datasets using the proposed methodology. When using a dataset of 5,000 Mangosteen images, our model outperforms the one-stage approach in terms of accuracy while still maintaining real-time speed with an accuracy rate ranging from 94% to 98.7%.