

NHẬN DẠNG HẠT ĐIỀU TRONG DÂY CHUYỀN ĐÓNG GÓI VÀ KIỂM TRA CHẤT LƯỢNG SẢN PHẨM DỰA TRÊN MẠNG NƠ RON TÍCH CHẬP

CASHEW NUT IDENTIFICATION IN PACKAGING AND QUALITY INSPECTION LINES BASED ON CONVOLUTIONAL NEURAL NETWORK

Đinh Quốc Quân¹, Nguyễn Thu Hà¹,
Nguyễn Đình Bảo¹, Nguyễn Vũ Linh¹, Phạm Văn Hùng^{1,*}

TÓM TẮT

Bài báo này trình bày về một thuật toán nhận dạng và phân loại hạt điều trong dây chuyền đóng gói và kiểm tra chất lượng sản phẩm dựa trên mạng nơ-ron tích chập YOLOv3 (phiên bản 3). Bài báo đã xây dựng một tập dữ liệu để phân loại hạt điều tốt, hoặc bị vỡ hoặc bị sót vỏ (chưa tách hết vỏ lụa). Các kết quả thực nghiệm cho thấy thuật toán đề xuất có khả năng nhận dạng hạt điều với độ chính xác trung bình mAP là 97,34%.

Từ khóa: Hạt điều, mạng nơ-ron tích chập, nhận dạng hạt điều, thị giác máy tính.

ABSTRACT

This paper presents an algorithm to identify and classify cashew nuts in the packaging and quality inspection lines based on the YOLOv3 convolutional neural network (version 3). In addition, we describe here a dataset for classifying good, broken, and shelled cashews. The experimental results show that the proposed algorithm can identify cashew nuts with an mean average precision (mAP) over our dataset is of 97.34%.

Keywords: cashew nut, Convolutional Neural Network (CNN), cashew nut identification, computer vision.

¹Trường Đại học Công nghiệp Hà Nội

*Email: phamvanhung@hau.edu.vn

Ngày nhận bài: 20/3/2022

Ngày nhận bài sửa sau phản biện: 19/4/2022

Ngày chấp nhận đăng: 29/8/2022

1. GIỚI THIỆU

Hạt điều là một trong các loại hạt được tiêu dùng phổ biến trên thế giới, bởi nhân điều chứa nhiều vi chất và khoáng chất có lợi cho sức khỏe, mà các thực phẩm khác không có, như P, Cu, Mg và có hàm lượng chất béo không bão hòa thấp nhất trong các loại hạt, giúp duy trì hàm lượng cholesterol có lợi [1]. Trong công nghiệp chế biến hạt điều, vấn đề then chốt ảnh hưởng đến giá thành sản phẩm là năng suất và chất lượng hạt điều sau chế biến.

Trước đây, nhiều công nghệ điều khiển và tự động hóa đã được áp dụng trong công nghệ hấp, sấy, bóc tách hạt. Tuy nhiên việc phân loại hạt điều đạt chất lượng thường sử dụng các máy rung để loại bỏ các hạt không đảm bảo kích thước và việc kiểm tra lại chất lượng dựa trên kích thước, màu sắc do người công nhân thực hiện bằng tay.

Ngày nay, các camera kết hợp công nghệ xử lý ảnh, cũng như trí tuệ nhân tạo (AI - Artificial Intelligent) đã và đang được nghiên cứu để hỗ trợ con người trong việc phân loại và kiểm tra chất lượng hoa quả nói chung và hạt điều nói riêng. Ứng dụng mạng nơ-ron DCNN (Deep Convolutional Neural Networks) và mô hình phát hiện đối tượng [2] trong việc phân loại cà chua và phát hiện vùng bị bệnh được đề cập trong [3] hay ứng dụng hệ chuyên gia (ES - Expert System) phát hiện dưa chuột bị bệnh ở tài liệu [4]. Còn về nhận dạng hạt điều, tài liệu [5] nghiên cứu hệ thống phân loại hạt điều cho kết quả bước đầu là đã phân loại được hạt điều tuy nhiên độ chính xác chưa cao (90%). Ngoài ra một số công ty đã thương mại một số thiết bị phân loại màu năng suất hàng tấn/h như Tailo (Trung Quốc), WESORT (Thụy Điển), IGSP (Hàn Quốc), nanoSorter (Ấn Độ) với độ chính xác 95%. Các sản phẩm này có năng suất lớn nhưng hiệu suất chưa cao, thường được sử dụng để phân loại thô và cần thêm nhân lực kiểm tra sản phẩm trước khi đóng gói hàng hóa (Qua khảo sát tại nhà máy OLAM Việt Nam). Như vậy có thể thấy việc áp dụng ứng dụng công nghệ cao định hướng công nghiệp 4.0 như công nghệ xử lý ảnh, robot, trí tuệ nhân tạo vào công nghiệp chế biến thực phẩm đã và đang được nghiên cứu trên thế giới.

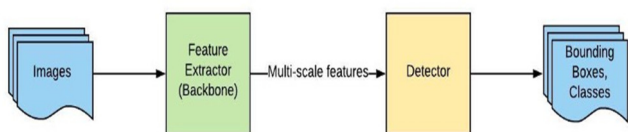
Ở nước ta đã có một số công trình nghiên cứu áp dụng xử lý ảnh và mạng nơ-ron trong phân biệt và nhận dạng các khuyết điểm các loại hoa quả như [6, 7, 8]. Tuy nhiên, các công trình nghiên cứu về nhận dạng, phân loại chất lượng hạt điều còn hạn chế, cần được bổ sung, đặc biệt là vấn đề nhận dạng ở chế độ động khi hạt điều chuyển động trên băng tải trong quá trình chế biến sản phẩm.

Trong bài báo này, chúng tôi đề xuất một thuật toán dựa trên mạng nơ-ron tích chập để phát hiện và nhận dạng những hạt điều không đảm bảo chất lượng, chẳng hạn như bị sứt vỏ hoặc bị cháy. Trước hết, chúng tôi xây dựng tập dữ liệu bao gồm các hình ảnh hạt điều thu được từ camera được gắn ở dây truyền đóng gói và kiểm tra sản phẩm và thực hiện gán nhãn tương ứng. Sau đó, chúng tôi đề xuất sử dụng mô hình mạng nơ-ron YOLOv3 (phiên bản 3) để thực hiện huấn luyện trên tập dữ liệu tự xây dựng. Cuối cùng, chúng tôi tiến hành đánh giá định lượng hiệu quả của mô hình đề xuất trên tập dữ liệu kiểm tra cũng như đánh giá định tính từ những video thu được bởi camera gắn trên dây truyền đóng gói hạt điều thực tế.

Phần còn lại của bài báo được tổ chức như sau: Phần 2 trình bày về kiến trúc mạng nơ-ron YOLOv3 được sử dụng để phát hiện và nhận dạng hạt điều. Phần 3 trình bày về tập dữ liệu cũng như các kết quả đánh giá, trong khi phần 4 đưa ra những kết luận, đánh giá và hướng phát triển tiếp theo của nghiên cứu.

2. MÔ HÌNH MẠNG NƠ-RON YOLOv3

YOLO (Only Look Once) [9] là một kiến trúc mạng nơ-ron tích chập (CNN - Convolutional Neural Networks) được sử dụng khá phổ biến trong phát hiện và phân loại đối tượng bởi ưu điểm là tốc độ nhanh hơn so với nhiều kiến trúc khác mà vẫn duy trì được độ chính xác tốt. Với YOLO, ta có thể nhận dạng được đối tượng và xác định được vị trí của nó trong bức ảnh. Mạng YOLO (hình 1) gồm 2 khâu chính là "Feature Extractor" và "Detector", hay thường gọi tương ứng là khâu "trích xuất đặc trưng" và khâu "Phát hiện". Khi đưa một bức ảnh nhất định vào khâu trích xuất đặc trưng, ta sẽ thu được tập hợp các đặc trưng (feature map) khác nhau (multi-scale features). Sau đó tiếp tục cho qua khâu "Detector" ta sẽ thu được các thông tin về vật thể trong ảnh như loại (class) và đường bao quanh vật thể (bounding box).



Hình 1. Sơ đồ tổng quát kiến trúc mạng YOLO

Bảng 1. So sánh về độ chính xác trên hàng tỷ các phép tính thường (Ops) và các phép tính số thực trên giây (BFLOP/s) và số khung hình trên giây (FPS) với các mạng khác nhau

Backbone	Top-1	Top-5	Ops	BFLOP/s	FPS
Darknet-19	74,1	91,8	7,29	1246	171
ResNet-101	77,1	93,7	19,7	1039	53
ResNet-152	77,6	93,8	29,4	1090	37
Darknet-53	77,2	93,8	18,7	1457	78

Hiện nay, có nhiều phiên bản khác nhau của YOLO, trong ba phiên bản đầu của Joseph Redmon là YOLOv1 đến v3 thì phiên bản v3 có độ chính xác trong nhận dạng đối

tượng tốt nhất. Vì vậy, trong nghiên cứu này, chúng tôi sử dụng YOLOv3 (phiên bản 3) cho việc phát hiện những hạt điều không đảm bảo chất lượng trên dây truyền đóng gói và kiểm tra sản phẩm.

Nếu như YOLOv2 sử dụng Darknet-19 thực hiện "Feature Extractor" các bức ảnh (hình 2) thì YOLOv3 thực hiện với Darknet-53, nhờ vậy tốc độ của Darknet-53 chậm hơn nhưng độ chính xác cao hơn khi so sánh với Darknet-19 và hiệu quả hơn khi so sánh với ResNet-101, ResNet-152 (bảng 1). Từ bảng 1, chúng ta có thể thấy rằng Darknet-53 nhanh hơn 1,5 lần so với ResNet101, nhanh hơn 2 lần so với ResNet-152 nhưng độ chính xác vẫn đảm bảo.

Type	Filters	Size	Output
Convolutional	32	3 × 3	256 × 256
Convolutional	64	3 × 3 / 2	128 × 128
1x	Convolutional	32	1 × 1
	Convolutional	64	3 × 3
	Residual		128 × 128
Convolutional	128	3 × 3 / 2	64 × 64
2x	Convolutional	64	1 × 1
	Convolutional	128	3 × 3
	Residual		64 × 64
Convolutional	256	3 × 3 / 2	32 × 32
8x	Convolutional	128	1 × 1
	Convolutional	256	3 × 3
	Residual		32 × 32
Convolutional	512	3 × 3 / 2	16 × 16
8x	Convolutional	256	1 × 1
	Convolutional	512	3 × 3
	Residual		16 × 16
Convolutional	1024	3 × 3 / 2	8 × 8
4x	Convolutional	512	1 × 1
	Convolutional	1024	3 × 3
	Residual		8 × 8
Avgpool		Global	
Connected		1000	
Softmax			

Hình 2. Kiến trúc Darknet-53 của YOLOv3

Darknet-53 gồm có 23 residual units, trong đó mỗi residual unit gồm có một 3 × 3 và một 1 × 1 convolutional layer. Tại cuối mỗi residual unit, một phép cộng theo từng phần tử (element-wise) được thực hiện giữa vec-tơ đầu vào và đầu ra. Sau mỗi convolutional layer là một batch normalization và một hàm kích hoạt Leaky Relu. Bên cạnh đó, sau mỗi residual units là một convolutional layer với bước nhảy là 2 để giảm kích thước feature map và do đó giảm số lượng tham số cho mô hình.

YOLOv3 sử dụng kiến trúc Feature Pyramid Network (FPN) để phát hiện vật thể ở nhiều tỉ lệ khác nhau. Cụ thể như với ảnh đầu vào có kích thước 416 × 416 thì đầu ra của YOLOv3 là ba "output map" với kích thước tương ứng là 13 × 13, 26 × 26 và 52 × 52. Thông thường để nhận dạng những đối tượng kích thước lớn thì dùng "output map" có kích thước nhỏ, và ngược lại nhận dạng đối tượng kích thước nhỏ thì dùng "output map" có kích thước lớn. Nhờ vậy, YOLOv3 có thể dự đoán chính xác hơn so với các phiên bản trước đó YOLOv1 và YOLOv2.

Theo như thuật toán YOLOv3 thì mỗi ô lưới trên mỗi output map có nhiệm vụ dự đoán ba hộp bao quanh, như vậy số lượng hộp được dự đoán trên một bức ảnh sẽ là:

$$(13 \times 13 + 26 \times 26 + 52 \times 52) \times 3 = 10647 \tag{1}$$

Các hộp bao quanh này được YOLOv3 dự đoán dựa theo các hộp neo (anchor box), đã được xác định trước và bao quanh vật thể một cách tương đối chính xác (cơ sở ước lượng). Mỗi một vật thể trong hình ảnh huấn luyện được phân bố về một hộp neo. Nếu có từ hai hộp neo trở lên cùng bao quanh một vật thể thì hộp neo có phần giao (Intersection Over Union) với hộp sự thật (truth bounding box) cao nhất sẽ được lựa chọn.

Như được biểu diễn trong hình 3: xét một ô lưới có tọa độ (c_x, c_y) thì hộp dự đoán (nét đứt đoạn) từ hộp neo (màu xanh) có tọa độ và kích thước được xác định như sau:

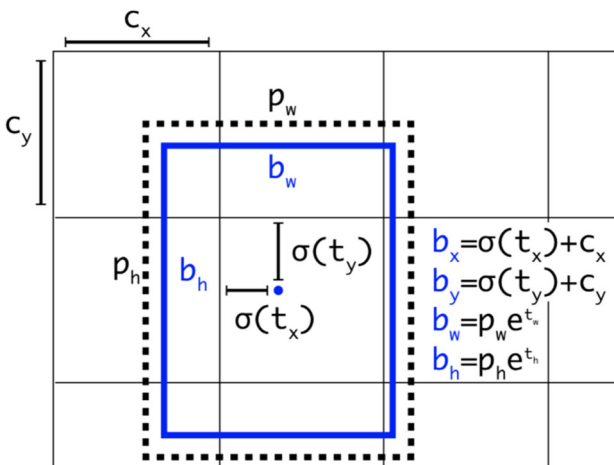
$$b_x = \sigma(t_x) + c_x$$

$$b_y = \sigma(t_y) + c_y$$

$$b_w = p_w e^{t_w}$$

$$b_h = p_h e^{t_h}$$

YOLOv3 dự đoán bốn tham số (t_x, t_y, t_w, t_h) , trong đó hai tham số đầu là độ lệch (offset) so với góc trên cùng bên trái và hai tham số sau là tỷ lệ so với hộp neo.



Hình 3. Công thức ước lượng bounding box (màu xanh) từ anchor box (đường nét đứt) và cell mà hộp đó thuộc về

Mặt khác, với mỗi hộp dự đoán, YOLOv3 cũng dự đoán xác suất mà hộp đó có chứa vật thể và xác suất lớp (classes) của vật thể đó. Nói cách khác, đầu ra của mô hình YOLOv3 là một véc-tơ gồm các thành phần sau đây:

$$y^T = [p_0, \langle t_x, t_y, t_w, t_h \rangle, \langle p_1, p_2, \dots, p_c \rangle]$$

Trong đó, p_0 là xác suất dự đoán vật thể xuất hiện trong hộp.

$\langle t_x, t_y, t_w, t_h \rangle$ là các tham số để xác định hộp dự đoán (hình 3).

$\langle p_1, p_2, \dots, p_c \rangle$ là véc-tơ phân phối xác suất dự đoán của các lớp.

Như phân tích ở trên, YOLOv3 có thể dự đoán ra rất nhiều hộp bao quanh khác nhau, nên những ô lưới có vị trí gần nhau thì khả năng các hộp dự đoán bị chồng chéo (overlap) là rất cao. Vì vậy, YOLOv3 sử dụng thuật toán non-max suppression (NMS) để giảm bớt các hộp dự đoán này.

Thuật toán NMS thực hiện theo hai bước như sau: Đầu tiên là loại bỏ các bounding box có xác suất chứa vật thể nhỏ hơn 0,5. Sau đó lựa chọn những hộp có xác suất chứa vật thể là cao nhất và loại bỏ tất cả các hộp có IOU với hộp này lớn hơn một giá trị ngưỡng nào đó.

Quá trình huấn luyện của YOLOv3 là quá trình tối ưu hàm mất mát multi-part loss function. Hàm mất mát này là tổng hàm mất mát của bounding box dự báo so với thực tế (localization loss) và hàm mất mát của phân phối xác suất (confidence loss). Khác với hai phiên bản trước đó, YOLOv3 sử dụng logistic classifier thay vì softmax cho việc classify đối tượng. Điều này làm cho YOLOv3 có hiệu quả tốt hơn nếu các label không "mutually exclusive", tức là có thể có đối tượng cùng thuộc 2 hay nhiều class khác nhau.

3. ĐÁNH GIÁ KẾT QUẢ

3.1. Tập dữ liệu

Để quá xây dựng tập dữ liệu và phân loại chúng tôi xét hạt điều được xét gồm 3 loại như sau (hình 4):

- Chất lượng loại 1 (Good): Hạt điều có màu sắc đồng nhất, có thể có trắng, vàng nhạt, ngà nhạt, hình dạng còn nguyên không bị vỡ và không bị cháy sém vàng
- Chất lượng loại 2 (Bad_1): Hạt điều có nhân đã bị nứt vỡ mảnh ở 1 số chỗ trên thân hạt điều.
- Chất lượng loại 3 (Bad_2): Hạt điều có nhân bị sém vàng, có thể có màu vàng, nâu nhạt, ngà nhạt, xám tro nhạt hoặc ngà đậm



Hình 4. Hình ảnh minh họa các loại hạt điều: Good (hình bên trái), Bad_1 (hình giữa), Bad_2 (hình bên phải)

Để xây dựng tập dữ liệu, chúng tôi tiến hành ghi lại ba video về hạt điều trên dây chuyền đóng gói sản phẩm. Tiếp theo, chúng tôi thực hiện trích xuất những hình ảnh từ những video này với tần suất 30 khung hình/1 ảnh và thu được 424 hình ảnh tất cả. Các hình ảnh này sau đó được gán nhãn thủ công về loại (tốt, vỡ, sứt vỏ lụa) và vị trí của hạt trong bức ảnh thông qua ứng dụng labellmg [10]. Cuối cùng, chúng tôi sử dụng phương pháp tăng dữ liệu bằng cách xoay các hình ảnh này góc 90° và 180° và thu được một bộ dữ liệu gồm có 1272 ảnh cùng với dữ liệu gán nhãn về vị trí và loại tương ứng.

3.2. Kết quả đánh giá

Với tập dữ liệu được mô tả ở phần trên, chúng tôi tiến hành chia ngẫu nhiên theo tỉ lệ 70% huấn luyện (training), 10% xác nhận (validation) và 20% để thử nghiệm (testing).

Việc thử nghiệm tiến hành trên Google Colab, trong đó ảnh vào của YOLOv3 là width=608, height=608; còn classes=3 (ba loại hạt là tốt, vỡ, sót vỏ lụa) và filters=(3 + 5)×3=24.

Bảng 2. Số lượng các hình ảnh đã được gán nhãn cho mỗi loại hạt điều

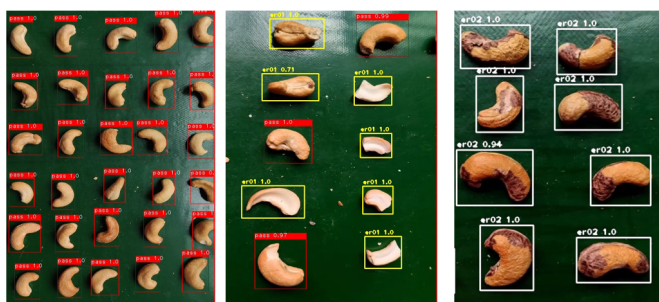
Hạt điều	Số lượng chú thích
Tốt (Good)	3819
Vỡ (Bad_1)	1412
Sót vỏ lụa (Bad_2)	1132
Tổng cộng	6360
Số lượng ảnh	1272

Thực hiện đào tạo sử dụng tập dữ liệu "training" và "validation" Quá trình huấn luyện được thực hiện dựa trên tập dữ liệu huấn luyện và tập dữ liệu xác nhận. Sau khi kết thúc quá trình này, kết quả cho ở file "yolov3-cashew.weights" được sử dụng để phân loại và nhận dạng hạt điều sau này.



Hình 5. Mô hình phân loại hạt điều

Sau khi huấn luyện xong, áp dụng thuật toán nhận dạng vào mô hình băng tải thực (hình 5) với tốc độ băng tải trong khoảng 0 – 0,6cm/s thu được một số kết quả thu được trên tập dữ liệu kiểm tra như được biểu diễn trong hình 6.



Hình 6. Kết quả phân loại hạt điều trong dây chuyền đóng gói và kiểm tra chất lượng sản phẩm

Trong đó:

- pass - hạt đạt
- er01 - hạt lỗi 1 (hạt bị vỡ)
- er02 - hạt lỗi 2 (hạt chưa bóc vỏ)

Những kết quả này cho thấy mô hình YOLOv3 sử dụng có thể phân loại khá chính xác hạt điều là tốt, bị vỡ hay sót

vỏ lụa, mặc dù những hạt điều này đang di chuyển trên băng tải đóng gói sản phẩm.

Tiếp theo, để đánh giá mức độ hiệu quả, các tham số AP50 và mAP (mean Average Precision) được sử dụng. Kết quả cụ thể như ở bảng 3.

Bảng 3. Kết quả nhận dạng hạt điều với tập dữ liệu kiểm tra.

	backbone	AP _{Good}	AP _{Bad 1}	AP _{Bad 2}	mAP
YOLOv3	Darknet53	97,98%	97,08%	96,96%	97,34%

Kết quả đánh giá cho tập dữ liệu kiểm tra cho thấy độ chính xác cho phát hiện các hạt tốt là 97,98%, hạt bị vỡ là 97,08% và hạt bị sót vỏ lụa là 96,96%. Điều đó có nghĩa là độ chính xác trung bình (mAP) của mô hình đề xuất là 97,34%, tỉ lệ nhận dạng như vậy là khá cao và hoàn toàn có thể áp dụng trong dây chuyền đóng gói và kiểm tra chất lượng hạt điều thực tế.

4. KẾT LUẬN

Bài báo này đã trình bày về một thuật toán để phát hiện và nhận dạng các hạt điều là tốt, bị vỡ hay bị sót vỏ trong dây chuyền đóng gói và kiểm tra chất lượng sản phẩm. Một tập dữ liệu ảnh hạt điều được gán nhãn về vị trí và loại đã được xây dựng. Mô hình mạng nơ-ron tích chập đề xuất là YOLOv3, và các kết quả đánh giá cho thấy mô hình có thể nhận dạng hạt điều với độ chính xác trung bình là khoảng 97,34%. Độ chính xác này là khá cao và cho thấy tính khả thi của thuật toán đề xuất trong việc ứng dụng công nghệ xử lý ảnh vào trong dây chuyền đóng gói và kiểm tra chất lượng hạt điều thực tế.

Trong bài báo này, chúng tôi mới chỉ ứng dụng mô hình YOLOv3 mà chưa có cải tiến nào. Vì vậy, trong nghiên cứu tiếp theo, chúng tôi sẽ cải thiện mô hình YOLOv3, thử nghiệm và so sánh với một số mô hình mạng nơ-ron khác, cũng như tích hợp thuật toán đề xuất trên những hệ thống nhúng có cấu hình thấp.

LỜI CẢM ƠN

Bài báo này được hỗ trợ và tài trợ bởi trường Đại học Công nghiệp Hà Nội theo đề tài mã số 15-2021-RD/HĐ-ĐHCN.

TÀI LIỆU THAM KHẢO

[1]. "Copyright © INC International Nut and Dried Fruit Council", 2015, https://www.nutfruit.org/files/tech/1516013457_Cashew_Tech_Kit_French_Final_170105.pdf

[2]. Y. Le Cun, B. Boser, J. S. Denker, et al., 1989. *Backpropagation applied to handwritten zip code recognition*. Neural Computation, vol. 1, no. 4, pp. 541–551.

[3]. Qimei Wang, Feng Qi, Minghe Sun, Jianhua Qu, Jie Xue, 2019. *Identification of Tomato Disease Types and Detection of Infected Areas Based on*

Deep Convolutional Neural Networks and Object Detection Techniques. Computational Intelligence and Neuroscience, Article ID 9142753, 15 pages <https://doi.org/10.1155/2019/9142753>

[4]. Ahmed Rafea, Salwa El-Gamal, Reda Abd El Whab, 2004. *Integrating Diagnostic Expert System with Image Processing via Loosely Coupled Technique*. Central Laboratory for Agricultural Expert System(CLAES).

[5]. Bordekar L., Velingkar H., Fernandes E., Bandekar H.H., Harmalkar A.G., Pinto B.J., 2018. *Cashew Nut Grade Identification and Quality Testing Using Machine Learning*. 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT), 661-664.

[6]. Truong Quoc Bao, Nguyen Van Vung, Truong Quoc Dinh, 2016. *Detect and recognize defects on the surface of mango peel*. 9th Fundamental and Applied IT Research (FAIR'9), DOI: 10.15625/vap.2016.00070.

[7]. Nguyen Minh Triet, Truong Quoc Bao, Truong Quoc Dinh, 2017. *Grapefruit leaf pests detection and recognition automatically using image technology*. Can Tho University Journal of Science: 88-95.

[8]. Nguyen Dac Thanh, 2017. *Nhan dang va phan loai hoa qua trong anh mau*. Master thesis, University of Engineering and Technology - VNU.

[9]. J. Redmon, A. Farhadi, 2018. *Yolov3: An incremental improvement*. arXiv preprint arXiv:1804.02767.

[10]. <https://pypi.org/project/labellmg/?msclkid=d4255650ad1b11ec9f1bb0cfaa07188e>

AUTHORS INFORMATION

**Dinh Quoc Quan, Nguyen Thu Ha, Nguyen Dinh Bao,
Nguyen Vu Linh, Pham Van Hung**

Hanoi University of Industry