

ĐỘ CHÍNH XÁC NHẬN DẠNG TRONG MÔ HÌNH FASTER R-CNN KHI CÓ NHIỀU

THE DETECTION ACCURACY OF THE FASTER R-CNN MODEL THROUGH INTERFERENCE ENVIRONMENTS

Nguyễn Văn Nam¹, Ngô Đình Thanh²

¹Trường Đại học Sư phạm Kỹ thuật - Đại học Đà Nẵng; nvnam@ute.udn.vn

²Trường Đại học Bách khoa - Đại học Đà Nẵng; ndthanh@dut.udn.vn

Tóm tắt - Thuật toán Faster 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 khác. Tuy nhiên, các kết quả đó hầu hết được thực hiện với ảnh đưa vào nhận dạng không bị nhiễu. Nghiên cứu này để đánh giá về độ chính xác nhận dạng của mô hình ở trạng thái bình thường và nhiễu. Để thực hiện việc này, tác giả đã huấn luyện cho mô hình nhận dạng 10 loài hoa và sau đó cho nhận dạng với 4 trạng thái khác nhau: Ảnh có ánh sáng tốt; ảnh bị che khuất 1/3; 1/2; ảnh thiếu ánh sáng. Phân tích phương sai ANOVA (Analysis of variance) và trắc nghiệm sự khác biệt nhỏ nhất LSD (Least Significant Difference) bằng phần mềm SPSS 20.0 để đánh giá độ chính xác nhận dạng trong 4 trạng thái. Kết quả cho thấy, độ chính xác của thuật toán phụ thuộc rất nhiều vào mức độ nhiễu; độ chính xác nhận dạng đạt 99,28%, 78,46%, 40,36% và 62,38% tương ứng với 4 trạng thái ở trên.

Từ khóa - Trí tuệ nhân tạo; thị giác máy tính; deep learning; nhận dạng đối tượng; xử lý ảnh.

1. Đặt vấn đề

Những năm gần đây, việc ứng dụng các mô hình Deep Learning vào trong thực tế được nhiều nhà khoa học quan tâm tham gia nghiên cứu, nổi trội trong đó là mô hình mạng nơ ron tích chập (Convolutional Neural Networks, CNN) [1], [2], [3]. Một ứng cử viên sáng giá để giải quyết các vấn đề như: xử lý dữ liệu đầu vào lớn, xử lý online, nâng cao độ chính xác và xử lý nhiễu đầu vào. Để thực hiện được những yêu cầu trên mô hình CNN cũng trải qua các cải tiến cụ thể: Mạng nơ ron tích chập khu vực (Regional convolutional neural networks, R-CNN) [4], [5]; Mạng nơ ron tích chập khu vực nhanh (Fast region-based convolutional neural networks, Fast R-CNN) [6]; Mạng nơ ron tích chập khu vực nhanh hơn (Faster region-based convolutional neural networks, Faster R-CNN) [7], [8].

Trong đó, mô hình Faster R-CNN không dùng thuật toán tìm kiếm chọn lọc để lấy ra các khu vực, mà nó thêm một mạng CNN mới gọi là mạng đề xuất khu vực (Region Proposal Networks, RPN) để tìm các khu vực [7]. Đầu tiên cả bức ảnh được cho qua mô hình huấn luyện trước để lấy bản đồ đặc trưng. Sau đó bản đồ đặc trưng được dùng cho RPN để lấy được các khu vực, sau khi lấy được vị trí các khu vực thì thực hiện tương tự Fast R-CNN [6].

Một kết quả thử nghiệm về thời gian của R-CNN [9] được thể hiện tại Hình 1. Hình 1 cho thấy, mô hình Faster R-CNN nhanh hơn hẳn các dòng R-CNN trước đó, vì vậy có thể dùng cho nhận dạng đối tượng ở thời gian thực.

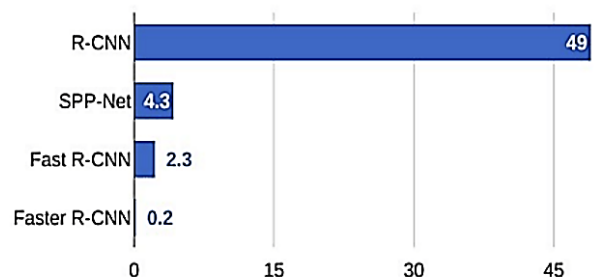
Độ chính xác nhận dạng là một yếu tố quan trọng của mô hình khi ứng dụng vào trong thực tế, khi đầu vào bị nhiễu (nhiều: hình ảnh trong môi trường trời tối, trời mưa hoặc ảnh bị che khuất một phần...) nó ảnh hưởng đến quá

Abstract - The Faster R-CNN algorithm is currently among the state-of-the-art in term of its speed and detection accuracy. However, most research on the accuracy of this algorithm is about noiseless images. This study, hence, conducts an accuracy assessment of the algorithm with both noisy and noiseless images. To this end, the algorithm is trained to classify ten flower species. Experiments are then implemented on images in four cases; images with non being in shadow, with ratio of 1/3, 1/2 of partially being in shadow and with totally being in shadow. Performance of the algorithm, via SPSS 2.0 software, is then analysed based on analysis of variance (ANOVA) and least significant difference (LSD). Experimental results show that the algorithm accuracy depends heavily on noise level. Detection accuracy achieves 99,28%, 78,46%, 40,36%, and 62,38% in cases of non being in shadow, 1/3, 1/2 of partially being in shadow, and totally being in shadow, respectively.

Key words - Artificial intelligence; computer vision; deep learning; identify the object; image processing

trình nhận dạng như thế nào? Trong nghiên cứu này sẽ cho thấy được sự ảnh hưởng của nhiễu ở đầu vào lên độ chính xác nhận dạng.

R-CNN Test-Time Speed



Hình 1. Thử thời gian R-CNN

Để thực hiện nghiên cứu trên, tác giả tiến hành huấn luyện cho mô hình Faster R-CNN nhận dạng mười loài hoa khác nhau, vì loài hoa rất phổ biến và dễ tiếp cận nên sẽ thuận lợi trong quá trình thu thập dữ liệu để phục vụ cho việc nghiên cứu. Hơn nữa, tác giả muốn hướng đến ứng dụng mô hình này vào phục vụ trong ngành lâm nghiệp, muốn huấn luyện mô hình này định danh các loài cây rừng từ hoa, lá, thân... Bởi vì muốn định danh được một số loài cây rừng thì hiện nay phải nhờ đến các chuyên gia mới định danh được. Ngoài ra, ứng dụng các mô hình nhận dạng định danh cho các loài cây rừng hay nhận dạng các loài động vật vẫn chưa được áp dụng phục vụ trong ngành lâm nghiệp. Việc điều tra, kiểm kê tài nguyên rừng hiện tại đang rất cần sự hỗ trợ của công nghệ hiện đại để giải quyết những khó khăn trong công tác quản lý tài nguyên rừng, công việc mà lâu nay hầu như do con người thực hiện. Để việc ứng dụng

vào thực tế mang lại được hiệu quả tốt thì cũng cần có các đánh giá trong quá trình thực nghiệm.

2. Nội dung thực hiện

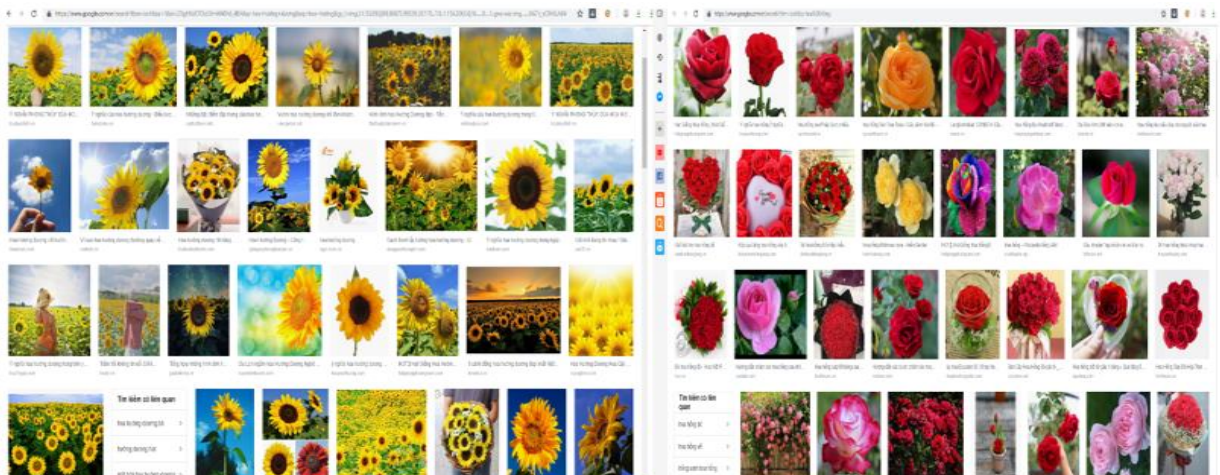
Để thực nghiệm mô hình đạt kết quả tốt thì quá trình thu thập tập dữ liệu để huấn luyện cho mô hình học là rất quan trọng. Ở nghiên cứu này, tác giả chọn thực nghiệm nhận dạng cho mười loài hoa nên việc thu thập dữ liệu chủ yếu được lấy từ mạng internet thông qua trang tìm kiếm Google. Đây cũng là một thuận lợi trong quá trình nghiên cứu.

2.1. Thu thập dữ liệu và gán nhãn cho các loài hoa

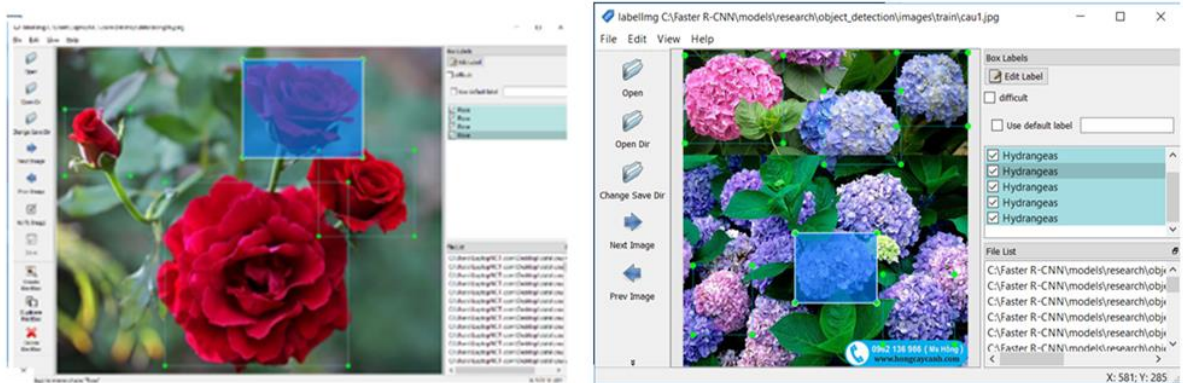
Tổng số hình ảnh thu thập để huấn luyện mô hình là 506 (ảnh) [10]. Các loài hoa được gán số thứ tự và được chia thành hai tập: tập dạy mô hình học (train) chiếm 80% trong tổng số hình ảnh; còn lại tập kiểm tra mô hình (test) chiếm 20%. Tập hình ảnh trong train và test được chọn một cách ngẫu nhiên.

Dữ liệu sau thu thập (train và test) được đưa vào phần mềm Labellmg để gán nhãn. Trong mỗi hình ảnh ta lấy các khu vực RPN rồi gán nhãn như Hình 3.

Số lượng hình ảnh từng loài hoa thu thập và gán nhãn



Hình 2. Thu thập dữ liệu hoa hồng và hoa hướng dương



Hình 3. Gán nhãn cho hoa hồng và hoa cẩm tú cầu

2.2. Môi trường thực nghiệm mô hình

Tác giả thực nghiệm trên máy tính PC main H310; Bộ xử lý (CPU): Core I7 8700es 6 nhân 12 luồng; Bộ nhớ (RAM): 16GB bus 1600GHz; Ổ cứng (SSD): 240GB; Card đồ họa (VGA): GTX 1060.

được thể hiện chi tiết tại Bảng 1.

Bảng 1. Bảng thu thập dữ liệu các loài hoa và gán nhãn

Số TT	Số Tên loài hoa	Số lượng tổng	Số lượng train	Số lượng test	Tên gán nhãn
1	Cầu ¹	55	44	11	Hydrangeas
2	Cúc ²	50	40	10	Chrysanthemum
3	Tiền ³	41	33	8	Gerbera
4	Dương ⁴	45	36	9	Sun Flower
5	Hồng ⁵	57	46	11	Rose
6	Ly ⁶	55	44	11	Lily
7	Mai ⁷	51	41	10	Apricot Blossom
8	Sen ⁸	55	44	11	Lotus
9	Sứ ⁹	56	45	11	Porcelain Flower
10	Tiên ¹⁰	41	33	8	Narcissus
Tổng		506	406	100	10 (loài hoa)

Chú thích: 1) Hoa cẩm tú cầu; 2) Hoa cúc; 3) Hoa đồng tiền; 4) Hoa hướng dương; 5) Hoa hồng; 6) Hoa ly; 7) Hoa mai; 8) Hoa sen; 9) Hoa sứ; 10) Hoa thủy tiên

2.3. Huấn luyện cho mô hình

Mô hình Faster R-CNN huấn luyện nhận dạng cho mười loài hoa có kiến trúc như sau: Lớp tích chập có kích thước 3x3; lớp lấy mẫu (RoI pooling) có kích thước 7x7; hàm kích hoạt ReLU (thông qua thư viện Keras) [11]; kích thước kernel 7x7x512.

3.2. Kết quả thực nghiệm nhận dạng khi ảnh bị che khuất 1/3

Tương tự, Bảng 3 ta có, số mẫu kiểm chứng đưa vào là 130 mẫu, tổng số mẫu mà mô hình nhận dạng được là 102 mẫu, 21 mẫu mô hình không nhận dạng được và 7 mẫu nhận dạng nhầm. Vậy ở trường hợp này độ chính xác tổng

thể của mô hình là 78,46%, 4 loài có độ chính xác 100%, 1 loài đạt 70%, những loài còn lại có độ chính xác dưới 70%. Riêng loài hoa ly có độ chính xác thấp nhất 27,27% với 11 mẫu đưa vào thì nhận dạng được 3 còn 8 mẫu không nhận dạng được.

Bảng 3. Bảng ma trận đánh giá độ chính xác kết quả nhận dạng khi ảnh bị che khuất 1/3

Kết quả nhận dạng	Đây là các loài được nhận dạng qua mô hình											
		Cầu	Cúc	Tiên	Dương	Hồng	Ly	Mai	Sen	Sứ	Tiên	Tổng
	Nhận dạng đúng	13	6	19	13	19	3	8	7	8	6	102
	Nhận dạng nhầm		1					1	2	3		7
	Không nhận dạng		2				8	3	1	3	4	21
	Tổng	13	9	19	13	19	11	12	10	14	10	130
Độ chính xác		100%	66,67%	100%	100%	100%	27,27%	66,67%	70%	51,14%	60%	78,46%

3.3. Kết quả thực nghiệm nhận dạng khi ảnh bị che khuất 1/2

Ở Bảng 4 ta có, tổng số mẫu đưa vào kiểm chứng là 109 mẫu, tổng số mẫu mà mô hình nhận dạng được 44 mẫu. Độ chính xác tổng thể của mô hình là 40,36%, không có loài nào đạt độ chính xác 100%, loài hoa hồng có độ chính xác

cao nhất 90%, loài hoa ly có độ chính xác thấp nhất 0%, 11 mẫu đưa vào thì 11 mẫu không nhận dạng được. Tổng số mẫu không nhận dạng được trong trường hợp này là 55 mẫu và có 10 mẫu nhận dạng nhầm. Khi ảnh bị che khuất 1/2 thì có hơn một nửa số ảnh mô hình không thể nhận dạng được.

Bảng 4. Bảng ma trận đánh giá độ chính xác kết quả nhận dạng khi ảnh bị che khuất 1/2

Kết quả nhận dạng	Đây là các loài được nhận dạng qua mô hình											
		Cầu	Cúc	Tiên	Dương	Hồng	Ly	Mai	Sen	Sứ	Tiên	Tổng
	Nhận dạng đúng	4	5	7	5	9	0	1	6	6	1	44
	Nhận dạng nhầm		1	5	1			1	2			10
	Không nhận dạng	7	3		6	1	11	10	2	6	9	55
	Tổng	11	9	12	12	10	11	12	10	12	10	109
Độ chính xác		36,36%	55,56%	58,33%	41,67%	90%	0%	8,33%	60%	50%	10%	40,36%

3.4. Kết quả thực nghiệm nhận dạng ảnh trong môi trường thiếu ánh sáng

Bảng 5 ta có, tổng số mẫu kiểm chứng đưa vào nhận dạng là 109 mẫu, tổng số mẫu mà mô hình nhận dạng được là 68 mẫu. Độ chính xác tổng thể của mô hình là 62,38%.

Trong đó, có 4 loài có độ chính xác 100%, loài hoa sứ có độ chính xác thấp nhất 10% với 10 mẫu đưa vào thì nhận dạng 1 mẫu, 4 mẫu nhận dạng sai (hoa ly: 3 mẫu; hoa mai: 1 mẫu) và 5 mẫu không nhận dạng được. Trong môi trường thiếu ánh sáng thì có đến 34 mẫu mô hình không nhận dạng được và 7 mẫu nhận dạng nhầm.

Bảng 5. Bảng ma trận đánh giá độ chính xác kết quả nhận dạng trong môi trường thiếu ánh sáng

Kết quả nhận dạng	Đây là các loài được nhận dạng qua mô hình											
		Cầu	Cúc	Tiên	Dương	Hồng	Ly	Mai	Sen	Sứ	Tiên	Tổng
	Nhận dạng đúng	10	9	7	10	8	2	10	4	1	7	68
	Nhận dạng nhầm			2					1	4		7
	Không nhận dạng			4		7	6		7	5	5	34
	Tổng	10	9	13	10	15	8	10	12	10	12	109
Độ chính xác		100%	100%	53,85%	100%	53,33%	25%	100%	33,33%	10%	58,33%	62,38%

3.5. Độ chính xác nhận dạng trong 4 trạng thái môi trường

Kết quả phân tích phương sai ANOVA và trắc nghiệm sự khác biệt nhỏ nhất LSD về kết quả độ chính xác nhận dạng trong 4 trạng thái môi trường được trình bày ở Bảng 6.

Kết quả ở Bảng 6 cho thấy, ĐCX của kết quả nhận dạng trong 4 trạng thái khác nhau khác biệt rất có ý nghĩa về mặt thống kê. Kết quả trắc nghiệm LSD cho thấy, ĐCX trong trạng thái có ánh sáng tốt khác biệt có ý nghĩa về mặt thống kê đối với các trường hợp còn lại.

Bảng 6. Kết quả phân tích phương sai và trắc nghiệm LSD về kết quả nhận dạng

	Trạng thái môi trường				Mức ý nghĩa (P)	LSD _{0,05}
	Ánh sáng tốt	Che khuất 1/3	Che khuất 1/2	Thiếu ánh sáng		
ĐCX (%)	99,28±3,5 ^a	78,46±24,7 ^b	40,36±28,1 ^c	62,38±34,7 ^{bc}	0,000	23,17

Ghi chú: Các kí tự khác nhau trên cùng một dòng thể hiện các số liệu khác biệt có ý nghĩa thống kê

4. Kết luận

Với 506 hình ảnh để huấn luyện cho mô hình Faster R-CNN học nhận dạng, kết quả sau thực nghiệm cho thấy ĐCX nhận dạng tổng thể của mô hình giảm dần khi độ nhiễu đầu vào nhận dạng tăng lên như khi ảnh bị che khuất 1/3 thì độ chính xác nhận dạng 78,46% còn khi ảnh bị che khuất 1/2 thì độ chính xác nhận dạng 40,36%, số mẫu mà mô hình không nhận dạng được cũng tăng tương ứng là 21 mẫu và 55 mẫu. Đối với ảnh trong môi trường thiếu ánh sáng thì độ chính xác 62,38% với 34 mẫu không nhận dạng được. Như vậy, với ảnh đưa vào nhận dạng không bị nhiễu thì mô hình cho kết quả nhận dạng chính xác cao nhất với 99,28% chỉ có 1 mẫu không nhận dạng được. Sự khác biệt về ĐCX nhận dạng trong 4 trạng thái môi trường nghiên cứu rất có ý nghĩa về mặt thống kê. Với kết quả này chúng ta có thể ứng dụng mô hình để nhận dạng các hệ động thực vật rừng phục vụ trong giảng dạy và nghiên cứu, đặc biệt là công tác kiểm kê, thống kê tài nguyên rừng. Hiện nay, trong công tác điều tra và kiểm kê tài nguyên rừng, việc nhận diện các loài thực vật, động vật, côn trùng, ... chủ yếu do con người thực hiện. Đây là một công việc thật sự vất vả khi công việc điều tra thường được thực hiện trên một diện tích lớn và số lượng điều tra nhiều. Hơn nữa, các chuyên gia phụ trách các lĩnh vực chuyên môn về việc định danh các loài thực vật, động vật, côn trùng, ... không nhiều, các đơn vị quản lý tài nguyên rừng đang thiếu các chuyên gia giỏi. Do đó, rất cần sự hỗ trợ từ công nghệ hiện đại như trí tuệ nhân tạo trong việc nhận dạng, định danh các loài động thực vật, việc làm mà lâu nay đều do con người đảm nhận. Việc làm này trong tương lai nếu do các drone thực hiện sẽ mang lại nhiều thuận lợi và hiệu quả hơn.

Những hạn chế đối với nghiên cứu này: Thứ nhất, môi trường thực nghiệm có cấu hình chưa đủ mạnh nên chưa thể thực hiện được khi dữ liệu đầu vào có dung lượng lớn. Thứ hai, chưa đánh giá được ảnh hưởng về khoảng cách chụp ảnh đến độ chính xác nhận dạng của mô hình. Thứ ba, chưa có các nghiên cứu tương tự để tác giả đưa ra đánh giá về tính hiệu quả trong nghiên cứu này.

Lời cảm ơn: Nghiên cứu này được tài trợ bởi Trường Đại học Sư phạm Kỹ thuật – Đại học Đà Nẵng cho đề tài mã số T2019-06-130, chủ nhiệm đề tài: Nguyễn Văn Nam, năm 2019.

TÀI LIỆU THAM KHẢO

- [1] J. Bouvrie, "Notes on Convolutional Neural Networks", *Massachusetts Inst. Technol. Cambridge, MA* 02139, 2006.
- [2] T. Liu, S. Fang, Y. Zhao, P. Wang, and J. Zhang, "Implementation of Training Convolutional Neural Networks", *arXiv:1506.01195v2*, 2015.
- [3] R. L. Galvez, A. A. Bandala, E. P. Dadios, R. R. P. Vicerra, and J. M. Z. Maningo, "Object Detection Using Convolutional Neural Networks", *IEEE Reg. 10 Annu. Int. Conf. Proceedings/TENCON*, vol. 2018-October, no. October, pp. 2023–2027, 2019.
- [4] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation", *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 580–587, 2014.
- [5] J. Liu, D. Wang, Z. Wei, L. Lu, L. Kim, and R. Summers, "Colitis detection on computed tomography using regional convolutional neural networks", *2016 IEEE 13th Int. Symp. Biomed. Imaging*, pp. 863–866, 2016.
- [6] R. Girshick, "Fast R-CNN", *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2015 Inter, pp. 1440–1448, 2015.
- [7] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks", *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, 2017.
- [8] R. Gavrilescu, C. Fo, C. Zet, and D. Cotovanu, "Faster R-CNN : an Approach to Real-Time Object Detection", *2018 Int. Conf. Expo. Electr. Power Eng.*, pp. 165–168, 2018.
- [9] F. R.-C. R-CNN, Fast R-CNN, "R-CNN Test-Time Speed", [Online]. Available: <https://towardsdatascience.com/r-cnn-fast-r-cnn-faster-r-cnn-yolo-object-detection-algorithms-36d53571365e..>
- [10] G. Drive, "Hình ảnh huấn luyện mô hình.rar", [Online]. Available: https://drive.google.com/open?id=1I5Fs9A9KmlcsjVrH2IH_0odII_ZLABIBd.
- [11] W. Pages, "ReLU (Rectified Linear Unit)", [Online]. Available: <https://missinglink.ai/guides/neural-network-concepts/7-types-neural-network-activation-functions-right/>.
- [12] G. Drive, "Hình ảnh đưa vào nhận dạng.rar", [Online]. Available: <https://drive.google.com/open?id=1xIeALJ31LMQTpj2tHWBojYp250i4A9du>.
- [13] G. Drive, "Hình ảnh sau nhận dạng.rar", [Online]. Available: https://drive.google.com/open?id=1DbAnEiIPL4uWQgSM_iYREIU2KyPN6YMF.

(BBT nhận bài: 07/9/2019, hoàn tất thủ tục phân biên: 25/11/2019)