

NGHIÊN CỨU ỨNG DỤNG HỌC SÂU VÀO PHÂN LOẠI RÁC THẢI

● ĐẶNG VŨ KIM KÝ - NGUYỄN HUỲNH THÔNG - ÂU THÙY ÂN - PHẠM TẤN HƯNG

TÓM TẮT:

Bài báo đưa ra một thí nghiệm về một mô hình mạng nơ ron tích chập hiệu chỉnh (Convolution Neural Network, CNN) giúp phân loại rác thành 3 nhóm khác nhau là chai nhựa, lon, và thủy tinh bằng việc sử dụng lên tới 900 hình ảnh thu thập bằng tay, được chia ra thành 500 hình ảnh dùng cho việc phân tích và 400 hình ảnh dùng cho việc đánh giá kết quả. Mô hình cho kết quả tuyến tính giữa số lượng dữ liệu huấn luyện với độ chính xác trung bình (mean Average Precision, mAP) khi kiểm tra mô hình xuất ra. Từ đây, nghiên cứu thực hiện một thí nghiệm khác để đảm bảo hiệu quả của mô hình đã huấn luyện đạt được các yêu cầu đề ra, AI401 với 80% độ chính xác tổng quan.

Từ khóa: rác thải, mạng nơ ron tích chập hiệu chỉnh, mô hình học sâu, mã nguồn mở, Tensorflow Object Detection API.

1. Đặt vấn đề

Vào thời điểm hiện tại, sự nhận thức của người dân về phân loại rác đang là một vấn đề đáng quan tâm, bên cạnh vấn đề tốc độ tăng lượng rác thải cũng ngày một cao. Theo Bộ Tài nguyên và Môi trường, ở Thành phố Hồ Chí Minh, vào năm 2004, khoảng 5.500 tấn rác thải rắn các loại được thải ra mỗi ngày [1]. Vào năm 2014, lượng rác thải đã tăng lên thành 7.000 tấn một ngày với lượng rác thải trung bình là 1,02 kg/người/ngày [2].

Trong đời sống hiện đại, Trí tuệ Nhân tạo (Artificial Intelligent, AI) đã và đang đóng vai trò quan trọng trong nhiều lĩnh vực đời sống, được ứng dụng và tồn tại trong nhiều ngành như là lĩnh vực y tế, sức khỏe, nông nghiệp, nhận diện vật thể,... Bài báo này chủ yếu tập trung vào việc huấn luyện một mô hình phân loại rác tùy chỉnh và đánh giá độ chính xác của mô hình, như là một sự tham khảo khi

ta ứng dụng một mô hình học sâu đã huấn luyện trước đó (pre-trained Deep Learning model) trong tương lai.

2. Cơ sở dữ liệu

Tensorflow Object Detection API là một gói mã nguồn mở dùng để nhận dạng vật thể thay vì xây dựng từ đầu. Bên cạnh đó cũng có những tiêu chuẩn trong việc lựa chọn mô hình AI phù hợp tùy theo các mục đích khác nhau nhằm tối ưu độ chính xác đạt được, trong đó có mô hình phục vụ cho việc xử lý ảnh, mô hình cho nhận diện giọng nói, mô hình cho xử lý văn bản dữ liệu,... Bên cạnh đó, kiến trúc mạng học sâu MobileNets V2 được công bố bởi nhóm tác giả Sandler vào năm 2019, là mô hình thể hiện sự nhảy vọt trong hiệu năng so với phiên bản trước đó của mô hình này. Với việc tích hợp Linear Bottleneck và Inverted Residuals [3] trong liên kết giữa các lớp tích chập, Linear Bottleneck giúp tính

toán đầu vào và đầu ra, trong khi Inverted Residual giúp gói lại các đặc trưng trích xuất khi chuyển giao giữa các lớp.

Bộ dữ liệu được thực hiện với mục đích nhằm tối ứng dụng trước mắt là trong môi trường học đường, cụ thể là tại Trường Đại học Bách khoa cơ sở Lý Thường Kiệt, quận 10, bao gồm 3 nhóm chai nhựa, lon nước kim loại, chai thủy tinh, cùng với phương thức lấy ảnh nhằm đảm bảo sự đa dạng trong bộ dữ liệu (Hình 1). Phương thức lấy ảnh này bao gồm cả việc xoay hướng ngẫu nhiên, góc và cường độ chiếu sáng, độ phóng to thu nhỏ của vật thể để đạt số lượng tối đa hình ảnh cho từng vật thể. Mỗi nhóm có xấp xỉ 15 vật thể được chụp với các góc và hướng quay khác nhau. Tổng cộng là 500 tấm ảnh dùng cho việc huấn luyện mô hình như trong Bảng 1.

Dán nhãn bộ dữ liệu cần được dán nhãn trước khi tiến hành quá trình huấn luyện, một công cụ có tên là LabelImg [4] là một công cụ giúp thêm vào phụ chú đồ họa cho ảnh (Hình 2). Công cụ này được viết bằng ngôn ngữ lập trình Python, và sử dụng công cụ PyQt4 cho giao diện đồ họa của nó.

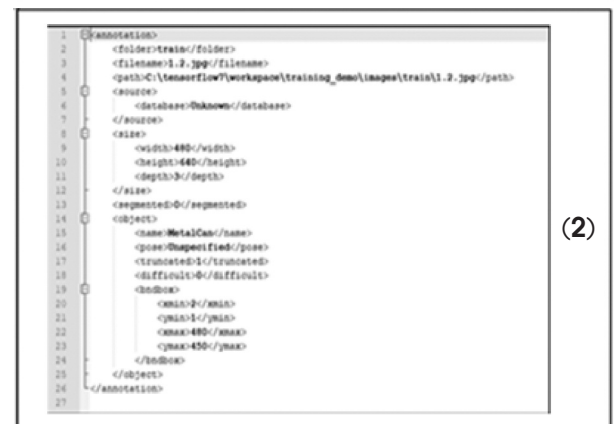
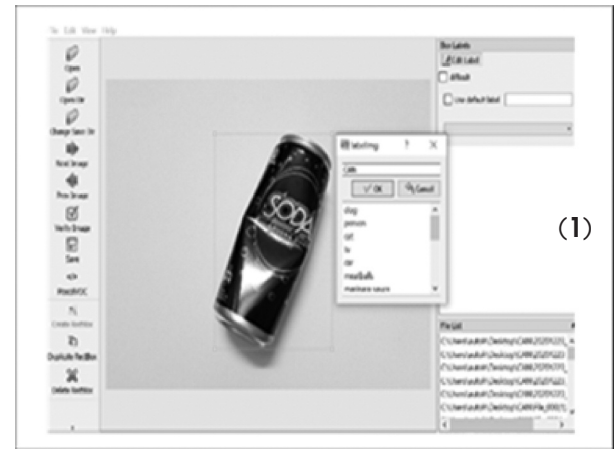
Hình 1: Một số ví dụ về bộ dữ liệu hình ảnh với các loại mẫu vật chọn mang ý nghĩa tượng trưng cho các nhóm chai nhựa, lon, chai thủy tinh.



Bảng 1. Số lượng các vật thể cũng như số lượng các tấm ảnh cho từng nhóm

STT	Nhóm	Loại	Số hình ảnh	Độ phân giải
1	Chai nhựa	14	166	640 x 640 pixels
2	Lon nước kim loại	15	166	
3	Chai thủy tinh	17	168	
	Tổng	48	500	

Hình 2: Vật thể (thuộc nhóm Lon) đang được dán nhãn bằng công cụ “Labelimg” (1); File XML chứa tên nhãn và tọa độ của Grounding box của vật thể bên trong tấm hình (2).



3. Thí nghiệm và kết quả

Nghiên cứu này với mục đích phân tích và đánh giá khả năng nhận diện của mô hình huấn luyện trích xuất được. Các mô hình đều được huấn luyện với các bộ dữ liệu khác nhau nhưng với cùng các

trọng số (hyperparameters) như trong Bảng 2 và Bảng 3. Nhóm nghiên cứu giữ nguyên các giá trị này do chúng đã hoàn toàn phù hợp và có thể được ứng dụng vào trong chính các mô hình mà nhóm đang muốn phân tích và đánh giá (Hình 3).

Nhóm nghiên cứu có thể tính toán ra 2 đường hồi quy tuyến tính thể hiện mối tương quan giữa mAP và số lượng ảnh dùng để huấn luyện mô hình theo thời gian (time) và số bước (steps).

Kết quả từ việc tính toán các giá trị tối ưu
Với các giá trị đo đạc được, nhóm nghiên cứu phác thảo ra biểu đồ như Hình 4 với trục hoành (x) thể hiện giá trị là số lượng tấm ảnh dùng để huấn luyện từng mô hình và trục tung (y) biểu thị cho các

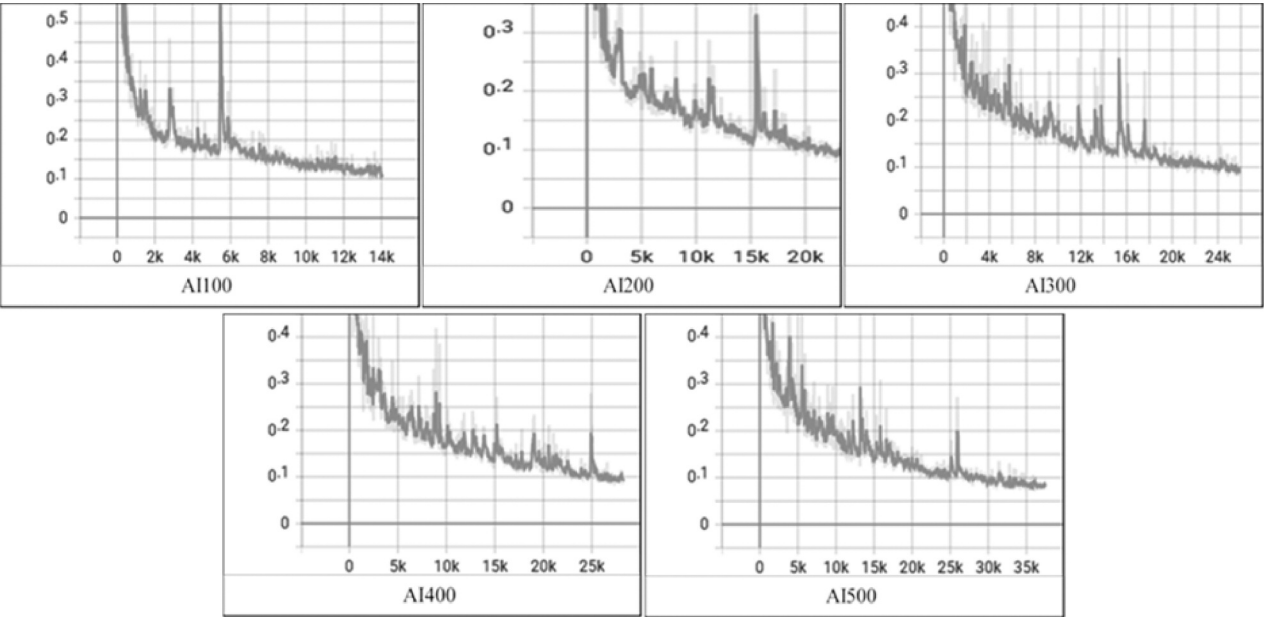
Bảng 2. Các giá trị trọng số đầu vào dùng cho việc huấn luyện

Thông số	Giá trị
Tỷ lệ học ban đầu (Initial Learning Rate)	0.08
Quán tính (Momentum)	0.9
Decay	0.997
Kích thước (Mini Batch Size)	6
Số chu kỳ (Number of Epochs)	833
Số bước tối đa (Maximum Steps)	50000
Trọng số (Weight)	4e - 05
Ngưỡng IoU (Matched IoU Threshold)	0.6

Bảng 3. Huấn luyện các mô hình với Điều kiện biên (Marginal Condition)

Mô hình học sâu (SSD MobileNets V2)	Tập dữ liệu huấn luyện	Điều kiện biên	Đầu ra (tên gọi)
Mô hình	100 hình	$\approx 0.1 \pm 3\%$	AI100
	200 hình		AI200
	300 hình		AI300
	400 hình		AI400
	500 hình		AI500

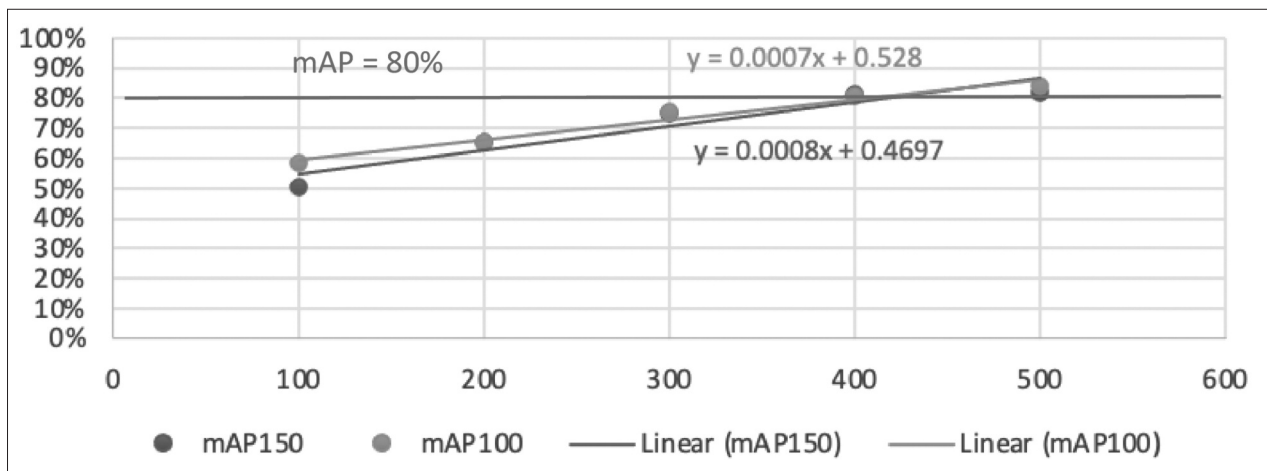
Hình 3: Hàm Mất mát chuẩn hóa trong quá trình huấn luyện cho các mô hình AI200, AI300, AI400, AI500.



Bảng 4. Các mô hình được đánh giá trên 2 bộ dữ liệu đánh giá với Thời gian (Time) và Số bước (Steps) cần cho quá trình huấn luyện

Tên gọi (x)	mAP với tập 150 hình (y1)	mAP với tập 100 hình (y2)	Thời gian (phút) (t)	Tổng số bước (z)
AI100	50.2%	58.3%	2 giờ	14000
AI200	65.1%	65.6%	3 giờ 22 phút	21000
AI300	75.2%	75.5%	3 giờ 51 phút	25880
AI400	81.4%	81.0%	4 giờ 36 phút	28300
AI500	81.6%	84.1%	5 giờ 31 phút	37400

Hình 4: Biểu đồ tương quan giữa mAP (%), (y)) và số lượng ảnh dùng cho việc huấn luyện (số ảnh, (x)).



giá trị như là mAP (%), Thời gian (phút), hoặc Số bước (bước), Thời gian (Time) (phút), Số bước (Steps) (bước). Với các giá trị trên, nhóm nghiên cứu phác thảo trên biểu đồ và xác định sự tuyến tính của tất cả các điểm phác thảo có được từ thí nghiệm và tính toán Hồi quy tuyến tính.

Việc tính toán để tìm ra điểm tối ưu với yêu cầu đặt ra là để đạt được giá trị chính xác mAP 80% cần thực hiện việc huấn luyện mô hình trên một bộ dữ liệu ảnh xác định, cũng như thời gian cần để huấn luyện, số bước cần thực hiện trong quá trình huấn luyện. Ta có:

- Cho bộ dữ liệu đánh giá 100 ảnh: $y = 0.0007x + 0.528$ and $R^2 = 0.9669 \Rightarrow x_1 \sim 389$ ảnh.

- Cho bộ dữ liệu đánh giá 150 ảnh: $y = 0.0008x + 0.4697$ and $R^2 = 0.8873 \Rightarrow x_2 \sim 413$ ảnh.

Và từ các giá trị x_1 và x_2 đạt được, ta có thể lấy

giá trị trung bình của 2 kết quả để nhận được do 2 đường hồi quy tuyến tính này gần như tương đồng nhau hoàn toàn, nên độ chênh lệch có thể được bỏ qua, với $x = 401$ ảnh. Tương tự, ta có thể phân tích giá trị tối ưu cho thời gian t : $t \sim 282$ phút và cho số bước cần cho việc huấn luyện k : $k \sim 30,780$ bước.

Nhóm nghiên cứu kết luận mô hình sẽ đạt giá trị tối ưu đề ra ở 80% mAP thì cần phải được huấn luyện 1 lần với 401 hình ảnh của 3 nhóm vật thể khác nhau với tổng số thời gian huấn luyện tính toán được là 282 phút (4 giờ và 48 phút), với tổng số bước cần là 30.780 bước.

Đánh giá khả năng nhận diện của mô hình nhận diện vật thể

Trong phần này, sau khi huấn luyện mô hình với giá trị tính toán tối ưu bên trên, nhóm nghiên cứu trích xuất mô hình nhận được và nhận được

các điểm khác biệt giữa giá trị ước tính và giá trị thực tế khi tiến hành thí nghiệm của các biến Thời gian và Số bước (Bảng 5). Ta có thể thấy sự khác biệt rõ ràng giữa giá trị giả thiết và thực tế của biến Thời gian, nhưng ngược lại là độ chênh lệch rất nhỏ giữa giả thiết và thực tế của biến Số bước cần cho việc huấn luyện (Bảng 6).

Điều này cho thấy, với cùng một số lượng hình ảnh dùng cho việc huấn luyện, thì quá trình giữa các lần huấn luyện là khác nhau, không có lần huấn luyện nào cho ra cùng một giá trị thời gian và số bước như nhau dù cùng một điều kiện biên và điều kiện đầu vào.

Sự khác biệt giữa các bộ dữ liệu này đối với mắt người là không nhiều, nhưng đối với cách máy tính và các thuật toán AI giải nghĩa một hình ảnh thì các hình ảnh là hoàn toàn khác nhau mặc dù được lấy hình từ cùng một mẫu vật thể. Vì vậy, thí nghiệm này cho thấy đánh giá độ chính xác của mô hình trích xuất trên một bộ dữ liệu mới hoàn toàn, từ đó, ta có thể dựa vào để đánh giá độ hiệu quả của mô hình tối ưu đề ra với kết quả đo đạt được theo Bảng 6 và Bảng 7.

Nhóm nghiên cứu so sánh giữa giá trị ước tính

và giá trị thực tế, nghiên cứu mong đợi kết quả đo được khi đánh giá độ chính xác đạt xung quanh 80% mAP. Từ đó, nhóm tác giả tách các giá trị đo đạt được và chia biểu đồ thành 2 vùng: vùng mong muốn (không màu) và vùng không mong muốn (màu hồng) (Hình 5).

4. Kết luận

Nghiên cứu đã đóng góp vào thư viện dữ liệu hình ảnh cho việc nhận dạng rác thải bao gồm 48 vật thể, với hơn 900 tấm ảnh được thu thập bên trong khu vực Trường Đại học Bách khoa Thành phố Hồ Chí Minh, Việt Nam. Tiếp đó, nhóm nghiên cứu đã huấn luyện thành công mô hình học sâu SSD MobileNets V2 với các điều kiện xác định và kiểm tra trên các tập dữ liệu nhằm chứng minh sự tuyến tính trong mối quan hệ giữa số lượng hình ảnh đưa vào huấn luyện và kết quả độ chính xác đưa ra, cùng với đó là sự độc lập giữa độ chính xác mà mô hình đưa ra so với bộ dữ liệu dùng để đánh giá độ hiệu quả. Cuối cùng, với mô hình AI401, tất cả các tập dữ liệu đánh giá đều đạt được ít nhất 80% mAP cho 3 nhóm - chai nhựa, lon kim loại, chai thủy tinh với 48 vật thể và đánh giá độ chính xác của mô hình đề ra ■

Bảng 5. Huấn luyện mô hình với Điều kiện biên (Marginal Condition) cùng với giá trị tính toán tối ưu về số lượng ảnh cần cho việc huấn luyện

Model Deep Learning (SSD MobileNets V2)		Tập dữ liệu huấn luyện	Điều kiện biên	Kết quả	
Mô hình	401 hình	$\approx 0.1 \pm 3\%$	AI401	Số bước = 34300	
				Thời gian = 445 phút	

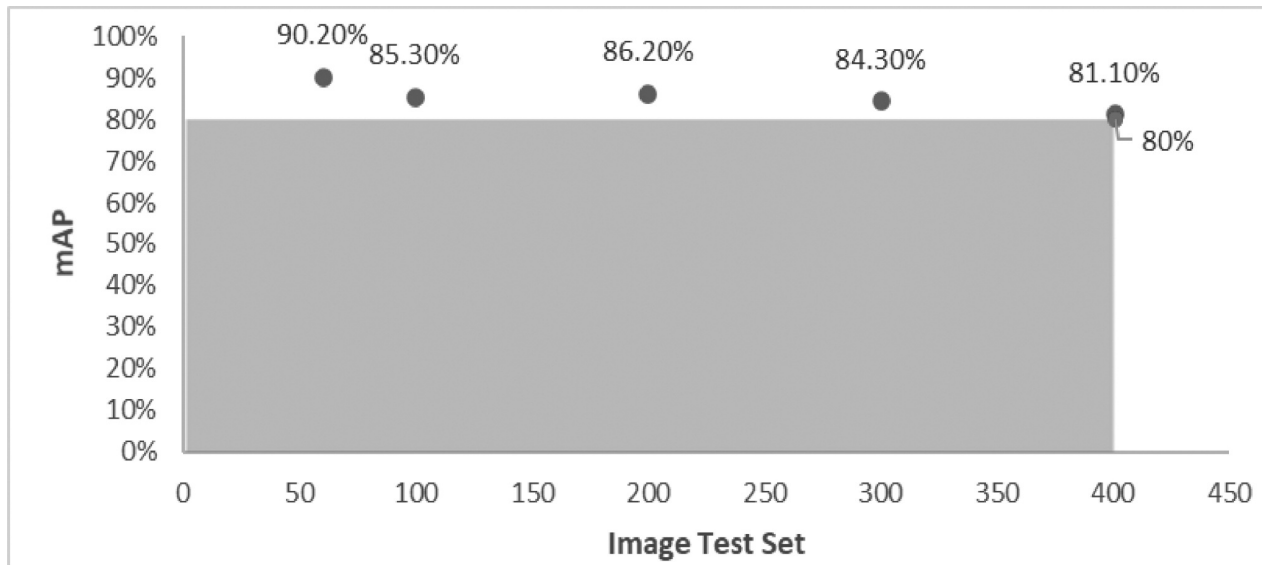
Bảng 6. Sự khác biệt giữa giá trị giả thiết tính toán và giá trị nhận được thực tế của các biến Thời gian và Số bước khi thực hiện tái thử nghiệm

	Giá trị dự đoán	Giá trị thật	Sai số (%)
Thời gian	282 phút	445 phút	57.8%
Số bước	30780 bước	34300 bước	11.4%

Bảng 7. mAP đo được (theo %) theo các bộ dữ liệu đánh giá khác nhau

Tập hình kiểm tra	60 hình	100 hình	200 hình	300 hình	401 hình
mAP	90.2%	85.3%	86.2%	84.3%	81.1%

Hình 5: Giá trị mAP đo được trên các bộ dữ liệu đánh giá với việc phân chia các vùng mong muốn (không màu) và vùng không mong muốn (màu hồng) cùng các điểm đo đặc được vẽ trên biểu đồ.



Lời cảm ơn:

Nghiên cứu này được tài trợ bởi Trường Đại học Bách khoa Thành phố Hồ Chí Minh (Việt Nam) với số hiệu To-ĐCDK-2020-16. Nhóm nghiên cứu cảm ơn sự hỗ trợ và cơ sở vật chất cung cấp bởi Trường Đại học Bách khoa Thành phố Hồ Chí Minh, Đại học Quốc gia Thành phố Hồ Chí Minh cho nghiên cứu này.

TÀI LIỆU THAM KHẢO:

1. VMENR (Vietnam Ministry of Environment and Natural Resources). (2014). *Canadian International Development Agency, Vietnam Environment Monitor: Solid Waste, Vietnam*.
2. DONRE (Department of Natural Resource and Management)-Ho Chi Minh City. (2014). *The report on overview of the SWM system of HCMC, Vietnam*.
3. Jiahui Yu, Yuning Jiang, Zhangyang Wang, Zhimin Cao, and Thomas Huang. Unitbox,. (2016). An advanced object detection network, *In Proceedings of the 24th ACM international conference on Multimedia*, ACM, pp. 516-520.
4. Shi, Lin; Zhong, Hao; Xie, Tao; Li, Mingshu. (2011). An Empirical Study on Evolution of API Documentation, *International Conference on Fundamental Approaches to Software Engineering*, Lecture Notes in Computer Science, ISBN 978-3-642-19810-6. Retrieved 22 July.

Ngày nhận bài: 3/1/2022

Ngày phản biện đánh giá và sửa chữa: 20/1/2022

Ngày chấp nhận đăng bài: 12/2/2022

Thông tin tác giả:

1. ĐẶNG VŨ KIM KÝ^{1,4}

2. NGUYỄN HUỲNH THÔNG^{2,4*}

3. ÂU THÙY AN^{2,4}

4. PHẠM TẤN HÙNG^{3,4}

¹Khoa Điện - Điện tử, Trường Đại học Bách khoa

²Khoa Kỹ thuật Địa chất và Dầu khí, Trường Đại học Bách khoa

³Khoa Môi trường và Tài nguyên, Trường Đại học Bách khoa

⁴Đại học Quốc gia Thành phố Hồ Chí Minh

A STUDY ON THE USE OF DEEP LEARNING FOR TRASH CLASSIFICATION

● **DANG VU KIM KY^{1,4}**

● **NGUYEN HUYNH THONG^{2,4}**

● **AU THUY AN^{2,4}**

● **PHAM TAN HUNG^{3,4}**

¹Faculty of Electrical and Electronic Engineering
Ho Chi Minh City University of Technology

²Faculty of Geotechnical Engineering and Petroleum
Ho Chi Minh City University of Technology

³Faculty of Environment and Natural Resources
Ho Chi Minh City University of Technology

⁴Vietnam National University - Ho Chi Minh City

ABSTRACT:

This paper presents an experiment on the use of a convolutional neural network (CNN) model to classify garbage into three different groups, namely plastic bottles, cans, and glass by using 900 manually collected images. In which, 500 images were used for analysis and 400 images were used for evaluation of results. This model gave a linear result which is between the amount of training data and the average accuracy (mean Average Precision, mAP). Based on these results, an other experiment was done to ensure that the trained model meets set requirements including AI401 with 80% overall accuracy.

Keywords: garbage, convolutional neural network, deep learning model, open source, Tensorflow Object Detection API.