JTE HCMUTE
ISSN: 1859-1272

TẠP CHÍ KHOA HỌC GIÁO DỤC KỸ THUẬT
Trường Đại học Sư phạm Kỹ thuật Thành phố Hồ Chí Minh
Website: https://jte.edu.vn
Email: jte@hcmute.edu.vn

Multi-Object Image Classification Using Deep Learning Method

Thi Dinh Nguyen*, Tran Bao Long Truong, Vuong Quoc Trung Ngo, Van Gia Bao Tran, Duong Tuan Nguyen, Phuong Hac Nguyen

Ho Chi Minh City University of Industry and Trade, Vietnam

*Corresponding author. *Email: dinhnt@huit.edu.vn*

ARTICLE INFO

Received: 11/03/2024
Revised: 24/03/2024
Accepted: 06/05/2024
Published: 28/10/2024

KEYWORDS

Image Classification;
Multi-object Image;
Deep Learning;
Object Recognition;
YOLOv8.

ABSTRACT

Image classification is an interesting topic for many scientists to improve the effectiveness of object recognition and image classification in computer vision. There are many techniques for image classification, in which deep learning methods have had many results in the problem of recognizing and classifying objects on an image. This paper performs a method for multi-object image classification using the YOLOv8 deep learning network. Firstly, each multi-object image is segmented into single-object images. Secondly, the identified image area, and then extracted feature vectors. Finally, the image is classified using the YOLOv8 deep learning network. An experiment conducted on the Flickr image set has shown better results than other methods and an average image classification result of 0.8872. Experimental results show that a proposed method using the YOLOv8 deep learning network for multi-object image sets is effective and can be applied to image data sets in many fields such as agriculture, traffic, and others.

Phân Lớp Ảnh Đa Đối Tượng Bằng Phương Pháp Học Sâu

Nguyễn Thị Định*, Trương Trần Bảo Long, Ngô Vương Quốc Trung, Trần Văn Gia Bảo, Nguyễn Dương Tuấn, Nguyễn Phương Hạc

Trường Đại học Công Thương Thành phố Hồ Chí Minh, Việt Nam

*Tác giả liên hệ. *Email: dinhnt@huit.edu.vn*

THÔNG TIN BÀI BÁO

Ngày nhận bài: 11/03/2024
Ngày hoàn thiện: 24/03/2024
Ngày chấp nhận đăng: 06/05/2024
Ngày đăng: 28/10/2024

TỪ KHÓA

Phân lớp ảnh;
Ảnh đa đối tượng;
Học sâu;
Nhận diện đối tượng;
YOLOv8.

TÓM TẮT

Phân lớp hình ảnh là chủ đề được nhiều nhà khoa học quan tâm để nâng cao hiệu quả nhận diện đối tượng và phân lớp hình ảnh trong lĩnh vực thị giác máy tính. Có nhiều kỹ thuật để phân lớp hình ảnh, trong đó phương pháp học sâu đã có nhiều kết quả trong bài toán nhận dạng và phân loại đối tượng qua hình ảnh. Trong bài báo này, một phương pháp đề xuất nhằm thực hiện phân lớp ảnh đa đối tượng sử dụng mạng học sâu YOLOv8. Đầu tiên mỗi ảnh đa đối tượng được phân đoạn thành các ảnh đơn đối tượng. Thứ hai, nhận diện và trích xuất véc-tơ đặc trưng. Cuối cùng hình ảnh được phân lớp bằng mạng học sâu YOLOv8. Thực nghiệm tiến hành trên bộ ảnh đa đối tượng Flickr đã cho kết quả tốt hơn một số phương pháp khác với kết quả phân lớp ảnh trung bình là 0.8872. Kết quả thực nghiệm cho thấy phương pháp đề xuất sử dụng mạng học sâu YOLOv8 cho bộ ảnh đa đối tượng là hiệu quả, có thể áp dụng được cho các tập dữ liệu hình ảnh thuộc các lĩnh vực khác nhau như nông nghiệp, giao thông và nhiều lĩnh vực khác.

Doi: https://doi.org/10.54644/jte.2024.1538

1. Giới thiệu

Xác định và phân lớp đối tượng bằng hình ảnh được ứng dụng nhiều trong các lĩnh vực như y tế, nông nghiệp, giao thông, v.v. cụ thể là bài toán phân loại bệnh qua ảnh y tế [1], phân loại bệnh trên cây trồng qua hình ảnh chụp được từ lá cây, phân loại mật độ giao thông bằng hình ảnh, v.v. Từ kết quả phân lớp hình ảnh, người dùng có thể ứng dụng vào các mục đích khác nhau trong cuộc sống như điều trị bệnh nhân, phát hiện bệnh cây trồng để điều trị hay chọn hướng đi phù hợp trong giao thông. Vì vậy, bài toán nhận diện và phân lớp hình ảnh bằng học sâu giúp nâng cao độ chính xác được thực hiện theo nhiều phương pháp khác nhau đã thu được những kết quả khả quan như k-NN [2], SVM [3], DNN [4]. Trong bài báo này, một phương pháp phân lớp ảnh đa đối tượng bằng phương pháp học sâu được trình bày và thực nghiệm với mạng học sâu YOLO.

Phương pháp học sâu được ứng dụng khá hiệu quả trong bài toán phân lớp ảnh như mạng học sâu DNN [1], CNN [5], YOLO [6]. Trong đó mạng học sâu YOLO được ứng dụng khá nhiều và hiệu quả trong bài toán phân lớp ảnh đa đối tượng. Hiện nay, mạng học sâu YOLO có nhiều phiên bản nhưng YOLOv8 đã có nhiều cải tiến về độ chính xác và thời gian nhận diện đối tượng trên ảnh so với các phiên bản trước đây. So sánh với các phiên bản trước mạng học sâu YOLOv8 thì phiên bản này có nhiều điểm cải tiến hơn về độ chính xác nhận diện đối tượng cũng như thời gian thực thi hệ thống, đặc biệt là áp dụng cho các bộ dữ liệu có kích thước khá lớn [7]. Vì vậy, bài báo đã sử dụng mạng học sâu YOLOv8 để nhận diện, phân lớp ảnh cho bộ ảnh đa đối tượng Flickr [8].

Đóng góp của bài báo gồm: (1) đề xuất phương pháp phân lớp ảnh đa đối tượng bằng mạng học sâu YOLOv8 nhận diện và phân loại ảnh đa đối tượng; (2) thực nghiệm phân lớp trên bộ ảnh đa đối tượng Flickr với kết quả phân lớp cao hơn một số phương pháp khác và các phiên bản trước của YOLO.

Phần còn lại của bài báo gồm: phần 2 trình bày các công trình nghiên cứu liên quan; phần 3 trình bày cơ sở lý thuyết về phân lớp hình ảnh và kiến trúc mạng học sâu YOLOv8; phần 4 trình bày kết quả thực nghiệm phân lớp hình ảnh trên bộ ảnh Flickr; kết luận và hướng phát triển được trình bày ở phần 5.

2. Các công trình nghiên cứu liên quan

Phân lớp hình ảnh có thể được thực hiện bằng nhiều phương pháp khác nhau như sử dụng thuật toán làng giềng gần nhất k-NN [2], thuật toán SVM [3], thuật toán DNN [4], thuật toán mạng học sâu YOLO [6] đã thu được nhiều kết quả khác nhau. Phân lớp dữ liệu bằng cấu trúc KD-Tree cũng đã mang lại hiệu quả trong một số bài toán như hình ảnh [18], dữ liệu văn bản [14]. Bên cạnh đó, mỗi phương pháp có những ưu, nhược điểm riêng và thích hợp với từng mẫu dữ liệu riêng. Vì vậy, một số công trình được khảo sát để đánh giá ưu, nhược điểm của từng phương pháp như sau:

Lia Farokhah và cộng sự (2020) [9] đã thực hiện phân lớp ảnh bằng thuật toán láng giềng gần nhất k-NN và thực nghiệm trên bộ ảnh Flower-17, tác giả đã chọn 4 thư mục ảnh đã thu được kết quả độ chính xác lần lượt là 64%, 73%, 76% tương ứng với (k = 1, 3, 5). Công trình này tác giả cũng đã chỉ ra những hạn chế cần cải tiến và giải pháp để nâng cao hiệu suất phân lớp hình ảnh. Faeze Sadati và cộng sự (2021) [10] đã thực hiện phân lớp hình ảnh trên bộ ảnh Flower-102 và Flower-17 với kết quả rất cao, đạt đến 96.47% và 97.64% tương ứng. Trong công trình này nhóm tác giả đã sử dụng kết hợp các kỹ thuật CNN và SVM để nâng cao hiệu quả phân lớp ảnh đầu vào. Sự kết hợp này được đánh giá cao trong các kỹ thuật phân lớp hình ảnh. Tuy nhiên, các phương pháp này chỉ hiệu quả cho tập ảnh đơn đối tượng, với các tập ảnh đa đối tượng còn nhiều hạn chế mà giải pháp cần được nghiên cứu và thực hiện với các bộ ảnh đa đối tượng như MS-COCO, Flickr, Visual Genome.

Ngoài ra, mạng học sâu YOLO mới đây đã được ứng dụng khá hiệu quả trong bài toán nhận diện đối tượng và phân lớp hình ảnh trên các tập ảnh đa đối tượng. Martin Štancel và cộng sự (2019) [6] đã trình bày khá chi tiết về kiến trúc mạng học sâu YOLO để áp dụng cho nhận diện từng đối tượng trên hình ảnh. Năm 2023, tác giả Thomas Stark và cộng sự [11] thực nghiệm và so sánh kết quả nhận diện đối tượng hoa bằng thuật toán YOLOv5 và YOLOv7. Kết quả thực nghiệm được đánh giá và so sánh trên cùng bộ dữ liệu để minh chứng hiệu quả của hai thuật toán này với hiệu suất trong vùng từ 93% đến

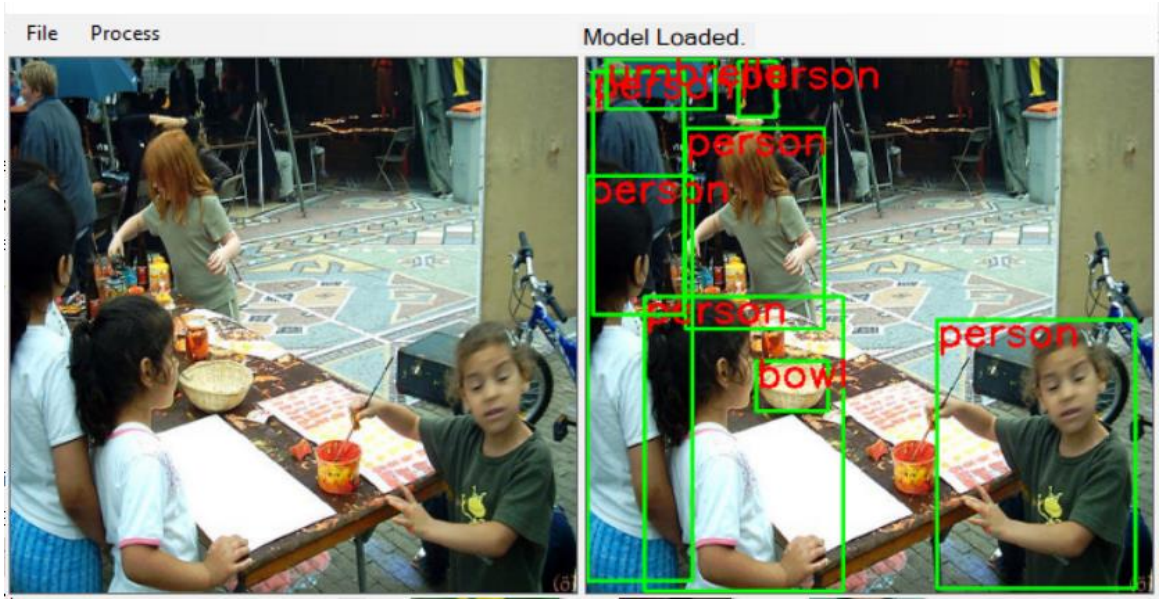
97%. Điều này cho thấy việc ứng dụng mạng học sâu YOLO vào các bài toán nhận diện đối tượng và phân loại đối tượng bằng hình ảnh là khá tốt.

Hiện nay, mạng học sâu YOLOv8 được ứng dụng cho bài toán nhận diện đối tượng và phân lớp hình ảnh được công bố bởi Naif Al Mudawi và cộng sự (2023) [12]. Công trình này nhóm tác giả đã kết hợp các nhóm đặc trưng cục bộ hình ảnh với mạng học sâu YOLOv8 để nhận diện và phân loại đối tượng trên hình ảnh để phân loại từng phương tiện giao thông trên mỗi ảnh thu được. Hiệu suất phân lớp phương tiện giao thông trong công trình này là khá cao trên 94.6% cho bộ ảnh VAID và trên 95.6% cho bộ ảnh VEDAI. Điều này chứng tỏ, với các bộ dữ liệu phức tạp thì mạng YOLOv8 đã đáp ứng được yêu cầu nhận diện và phân lớp khá tốt.

Từ những công trình nghiên cứu liên quan trên, một số thuật toán được sử dụng cho phân lớp hình ảnh như k-NN, DNN và CNN được ứng dụng khá hiệu quả và khả thi trên nhiều loại tập dữ liệu. Nếu kết hợp giữa các thuật toán này thì hiệu suất cao hơn, tuy nhiên thời gian huấn luyện mô hình chưa được phân tích và đề cập cũng như chi phí tài nguyên cho quá trình huấn luyện cũng chưa được làm rõ. Mạng học sâu YOLO đã tích hợp được nhiều tính ưu việt cho bài toán nhận diện từng đối tượng riêng biệt trên mỗi hình ảnh, từ đó phân lớp các đối tượng này cũng đã mang lại hiệu quả cao. Tuy nhiên, các phiên bản YOLO được khảo sát, trình bày đến YOLOv8 công trình này chúng tôi tiến thành thực nghiệm trên bộ ảnh Flickr với mạng học sâu YOLOv8 đã mang lại hiệu quả tương đối tốt so với một số phương pháp trước đây.

3. Mạng học sâu YOLOv8 cho nhận diện và phân lớp hình ảnh

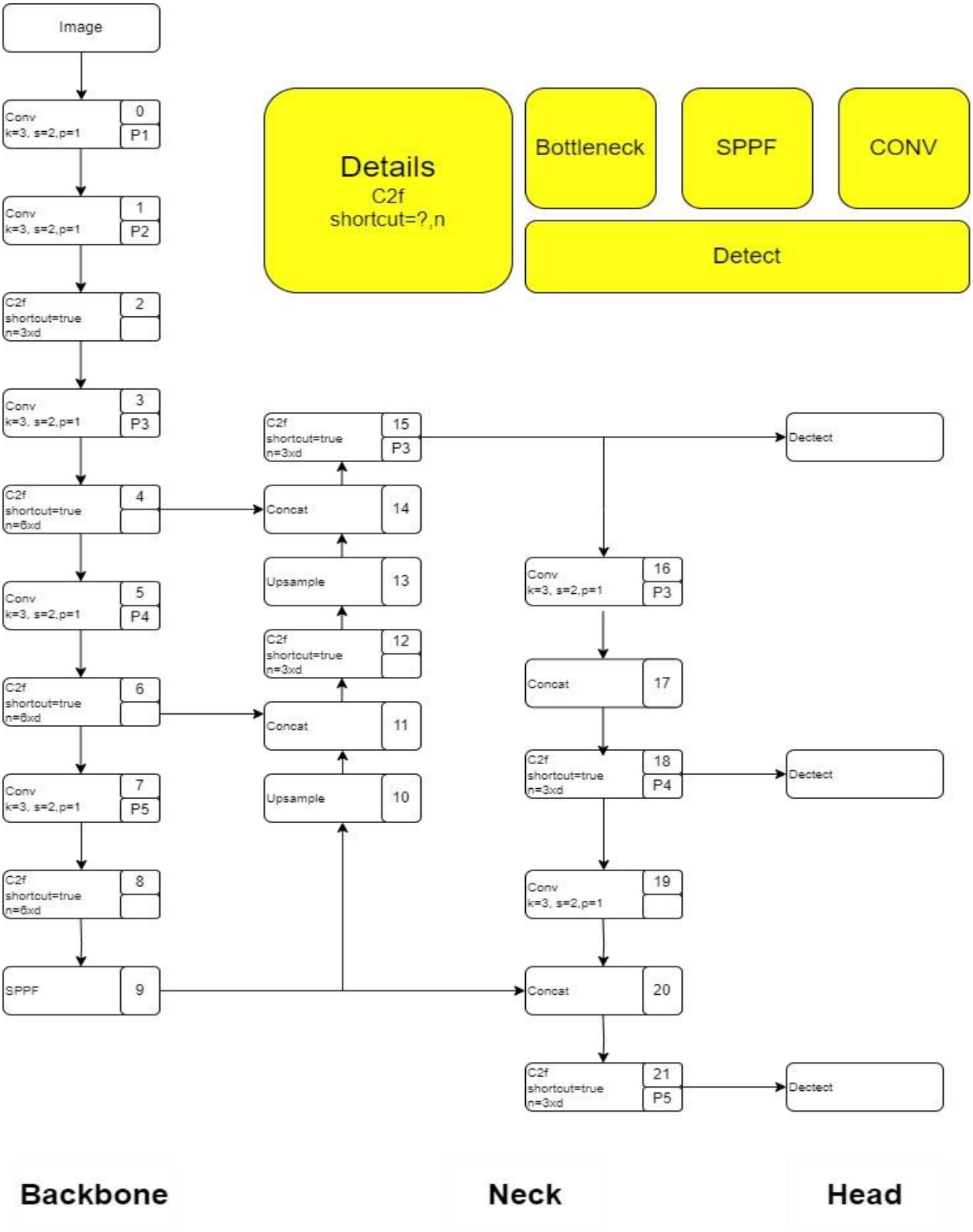
Mạng học sâu YOLO để nhận diện đối tượng, phân lớp đối tượng được đánh giá khá tốt trên nhiều bộ ảnh thực nghiệm và cả dữ liệu video. Qua các phiên bản của mạng học sâu YOLO đã có nhiều cải tiến về độ chính xác, thời gian thực thi và mạng YOLO được đánh giá tốt hơn cả mạng R-CNN trong một số bộ ảnh có đối tượng nhỏ, phức tạp.



Hình 1. *Nhận diện đối tượng trên ảnh bằng mạng học sâu YOLOv8 trên ảnh 438106.jpg*

Mạng học sâu YOLOv8 là phiên bản mới nhất của mạng YOLO được công bố gần đây và đã mang lại nhiều điểm tích cực so với các phiên bản trước như phát hiện đối tượng không dùng anchor, giới thiệu lớp tích chập C3 và tăng cường mosaic. Mô hình YOLOv8 được huấn luyện trước trên các bộ dữ liệu khá lớn như MS-COCO và ImageNet với tập dữ liệu phong phú và đa dạng các loại chủ đề ảnh, vì vậy ứng dụng khá hiệu quả cho các bộ ảnh tương tự. Mạng học sâu YOLOv8 cũng có tốc độ huấn luyện nhanh hơn, với độ chính xác cao hơn và kích thước mô hình nhỏ so với các phiên bản trước, đặc biệt là

vượt trội so với YOLOv5. Mô hình YOLOv8 có thể huấn luyện trên các máy có cấu hình GPU đơn, đây chính là ưu điểm mà YOLOv8 đã cải tiến cho người dùng dễ sử dụng tài nguyên lớn [7]. Một minh họa nhận diện đối tượng trên ảnh bằng mạng YOLOv8 được minh họa như Hình 1. Trong đó, mỗi đối tượng được nhận diện bởi một boundingbox và xác định phân lớp cho từng đối tượng trên ảnh.



Hình 2. *Kiến trúc mạng học sâu YOLOv8*

Để thực hiện nhận diện và phân lớp đối tượng bằng hình ảnh sử dụng mạng học sâu YOLOv8, đầu tiên với mỗi ảnh đầu vào gồm nhiều đối tượng sẽ được mạng YOLO nhận diện bằng các bounding box, từ đó trích xuất đặc trưng và qua mô hình huấn luyện để nhận diện và phân lớp cho đối tượng này. Hình 1 là một kết quả nhận diện và phân lớp cho ảnh 438106.jpg (Flickr).

Mạng học sâu YOLOv8 có nhiều ưu điểm giúp nâng cao hiệu suất nhận diện và phân lớp đối tượng, đó là: (1) hiệu suất YOLOv8 tốt, với kiến trúc mạng phức tạp, nó có khả năng xác định đối tượng và vị trí của đối tượng với độ chính xác cao. (2) đặc trưng đa dạng và đa nhiệm trên YOLOv8 đã hỗ trợ việc nhận diện và phân loại đối tượng trên các hình ảnh đa đối tượng khá tốt. Vì vậy, YOLOv8 có khả năng xử lý đa dạng các loại đối tượng và có thể được áp dụng trong nhiều ứng dụng khác nhau. (3) tích hợp thông tin từ nhiều tầng mạng, YOLOv8 sử dụng nhiều lớp tích chập và lớp kết nối đầy đủ để trích xuất đặc trưng từ các tầng khác nhau của mạng. Vì vậy, quá trình trích xuất thông tin chi tiết từ ảnh và nâng cao hiệu suất nhận diện của mô hình phân lớp đối tượng hình ảnh. Ngoài ra, phiên bản YOLOv8 có khả năng nhận diện được các đối tượng nhỏ, các đối tượng không rõ do môi trường ánh sáng khác nhau cũng có thể nhận diện được với mạng YOLOv8. Vì vậy, hiệu suất phát hiện, nhận diện và phân loại đối tượng trên ảnh đạt được hiệu suất khá tốt.

Theo mô hình YOLOv8 ở Hình 2 thì mạng học sâu YOLOv8 gồm các thành phần: (1) CSPDarknet53 Backbone đã được sửa đổi; (2) các mô-đun C2f thay thế CSPLayer trong mạng YOLOv5; (3) lớp tổng hợp (SPPF) tăng tốc tính toán bằng cách gộp các tính năng vào một bản đồ có kích thước cố định; (4) mỗi mạng tích chập (Conv) có lô chuẩn hóa và kích hoạt SiLU; (5) phần Head được tách rời để xử lý tính khách quan, phân loại và hồi quy. Đây là những thành phần cơ bản của mạng học sâu YOLOv8 và khác với phiên bản YOLOv5.

Ngoài ra, trong quá trình sử dụng mạng học sâu YOLOv8 một số hạn chế cần được cải tiến so với phiên bản YOLOv5 như: (1) kích thước mô hình lớn vì YOLOv8 sử dụng kiến trúc mạng phức tạp hơn, kích thước của mô hình cũng lớn hơn so với các phiên bản trước. Điều này có thể gây khó khăn trong việc triển khai mô hình này trên các thiết bị có tài nguyên hạn chế. (2) yêu cầu tài nguyên tính toán cao, việc huấn luyện và triển khai YOLOv8 yêu cầu nhiều tài nguyên tính toán lớn, điều này gây ra vấn đề về thời gian và phức tạp cho các ứng dụng có tài nguyên hạn chế. (3) YOLOv8 có kiến trúc mạng phức tạp, điều này có thể làm tăng độ khó trong việc tinh chỉnh tham số và tùy chỉnh mô hình cho các nhiệm vụ cụ thể.

Mạng học sâu YOLOv8 còn cung cấp mô hình phân đoạn ngữ nghĩa gọi là mô hình YOLOv8-Seg. Backbone trích xuất tính năng CSPDarknet53, mô-đun C2f thay vì kiến trúc cổ YOLO truyền thống. Mô-đun C2f được theo sau bởi hai phân đoạn Head nhằm học cách dự đoán mặt nạ phân đoạn ngữ nghĩa cho hình ảnh đầu vào. Mô hình này có các đầu phát hiện tương tự YOLOv8, bao gồm năm mô-đun phát hiện và một lớp dự đoán. Mô hình YOLOv8-Seg đã đạt được kết quả tiên tiến trên nhiều tiêu chuẩn phát hiện đối tượng và phân đoạn ngữ nghĩa [7]. Đây chính là những đặc điểm nổi bậc của mô hình mạng học sâu YOLOv8 mà các phiên bản trước đây chưa thực hiện được [13]. Ngoài ra mạng YOLOv8 còn thực hiện gán nhãn, huấn luyện và triển khai trên các bộ dữ liệu lớn. Vì vậy, các ưu điểm của YOLOv8 là nổi trội hơn so với các phiên bản trước đây, tuy nhiên cũng cần tài nguyên khá lớn để huấn luyện mô hình này.

4. Thực nghiệm và đánh giá

Môi trường thực nghiệm phân lớp ảnh đa đối tượng **ICYL** (*Image Classification by using YOLO*) được thực hiện trên nền tảng dotNET Framework 4.5, ngôn ngữ lập trình C#. Cấu hình máy tính: Intel(R) Core™ i5-5200U, CPU 2.7GHz, RAM 16GB và hệ điều hành Windows 10 Professional. Dữ liệu thực nghiệm là bộ ảnh đa đối tượng Flickr gồm 31,783 hình ảnh được mô tả như trong Bảng 1.

Bảng 1. *Mô tả dữ liệu thực nghiệm bộ ảnh Flickr*

Bộ ảnh	Tổng số ảnh	Số ảnh Training	Số ảnh Testing	Số ảnh đánh giá	Dung lượng
Flickr	31,783	29,000	1,783	1,000	9

JTE
HCMUTE
ISSN: 1859-1272

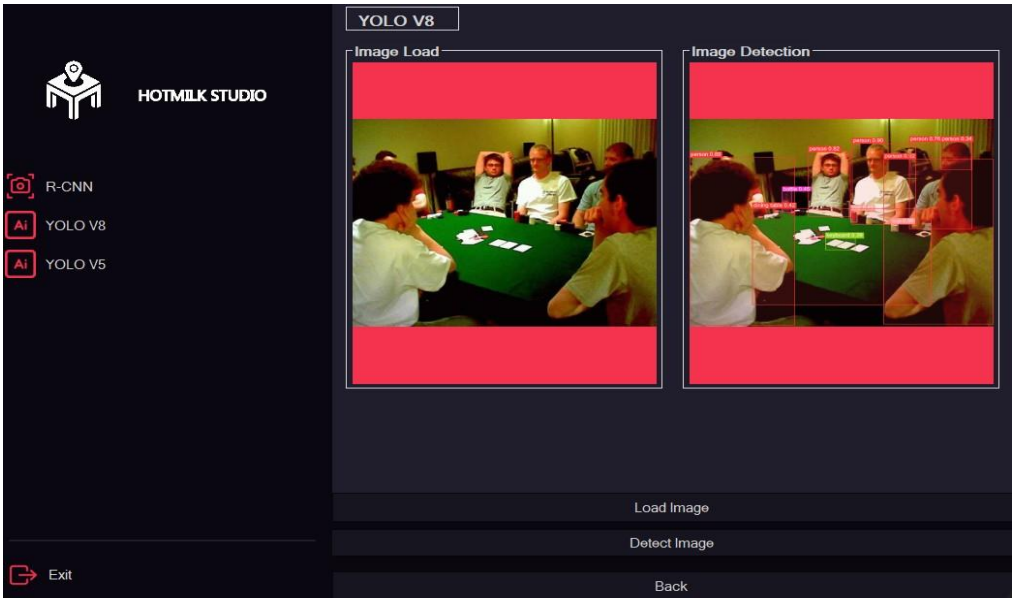
TẠP CHÍ KHOA HỌC GIÁO DỤC KỸ THUẬT
Trường Đại học Sư phạm Kỹ thuật Thành phố Hồ Chí Minh
Website: https://jte.edu.vn
Email: jte@hcmute.edu.vn

Thực nghiệm hệ phân lớp ảnh **ICYL** thu được kết quả độ chính xác phân lớp hình ảnh trên bộ ảnh Flickr với các thông số độ chính xác và thời gian thực thi được trình bày trong Bảng 2.

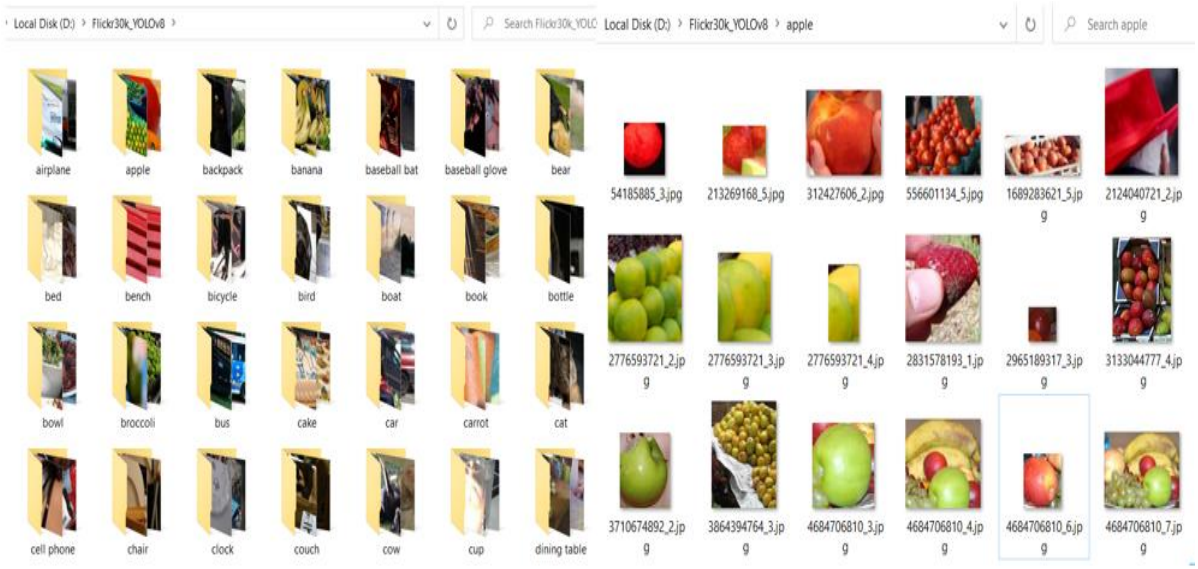
Bảng 2. *Hiệu suất phân lớp ảnh bằng YOLOv8 trên bộ ảnh Flickr*

Bộ ảnh	Số lượng ảnh kiểm thử	Số phân lớp	Hiệu suất phân lớp	Thời gian trung bình (ms)
Flickr	1,000	78	0.8872	32.28

Bảng 2 mô tả kết quả thực nghiệm về hiệu suất phân lớp ảnh đa đối tượng trên bộ ảnh Flickr. Trong đó số lượng kiểm thử là 1,000 ảnh đùn để do độ chính xác phân lớp theo từng loại đối tượng trên ảnh, với 1,000 ảnh này được thực hiện kiểm tra cho hệ **ICYL** thì kết quả thu được sau khi thực hiện nhận diện đối tượng và phân lớp thì có 78 phân lớp và hiệu suất phân lớp trung bình chung đạt 0,8872, thời gian thực hiện phân lớp trung bình của mỗi ảnh trên hệ ICYL là 32,28 (ms).



Hình 3. *Nhận diện và phân lớp ảnh bằng mạng học sâu YOLOv8*



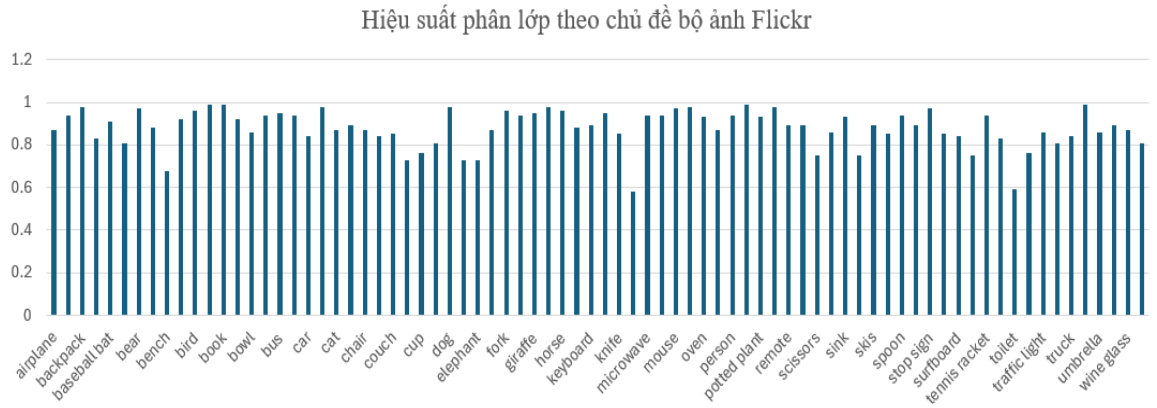
Hình 4. *Kết quả phân lớp trên bộ ảnh Flickr sử dụng mạng học sâu YOLOv8*

Hiệu suất phân lớp trên bộ ảnh Flickr đã được tác giả thực hiện kế thừa từ phương pháp tính hiệu suất phân lớp từ công trình [18]. Độ chính xác phân lớp là tỷ số giữa kết quả phân lớp thực thi trên hệ ICYL và kết quả phân lớp thủ công từ bộ ảnh Flickr.

Hình 3, mô tả giao diện thực nghiệm phân lớp ảnh cho mỗi ảnh đầu vào (Image Load), sau khi nhận diện đối tượng bằng boundingbox, trích xuất đặc trưng và phân lớp cho mỗi đối tượng thì kết quả được thể hiện (Image Detection). Đối với mạng YOLOv5 cùng thực nghiệm trên bộ ảnh Flickr thì kết quả nhận diện, phân lớp đối tượng trên ảnh có độ chính xác thấp hơn và thời gian thực thi trung bình lớn hơn so với mô hình YOLOv8 [15].

Kết quả phân lớp ảnh bằng mạng YOLOv8 thực nghiệm toàn bộ trên tập ảnh Flickr như Hình 4, kết quả thu được là 78 phân lớp ảnh gồm: airplane, apple, backpack, ..., wine glass, zebra. Mỗi thư mục chứa những hình ảnh cùng một phân lớp ảnh là các ảnh đối tượng được phân đoạn từ ảnh gốc.

Hiệu suất phân lớp cho bộ ảnh Flickr ứng với mỗi chủ đề trên được tính toán với kết quả đạt được minh họa bằng biểu đồ như Hình 4.



Hình 5. *Hiệu suất phân lớp theo chủ đề trên bộ ảnh Flickr sử dụng mạng học sâu YOLOv8*

Để minh chứng tính hiệu quả của mạng học sâu YOLOv8 với các phương pháp khác trong nhận diện và phân lớp hình ảnh. Một số công trình được chọn làm so sánh kết quả phân lớp ảnh trên cùng bộ dữ liệu ảnh Flickr được trình bày trong Bảng 3. Kết quả so sánh này minh chứng cho tính hiệu quả của mô hình phân lớp ảnh đa đối tượng sử dụng mạng học sâu YOLOv8 với một số kết quả khác đã công bố trong những năm gần đây.

Bảng 3. *So sánh hiệu suất phân lớp trên bộ ảnh Flickr với các phương pháp khác*

Phương pháp	Hiệu suất phân lớp ảnh
Deep Learning (Shadi Alijani, etc., 2022) [16]	0.8755
YOLOv5 – XceptionV3 (M. SAROJA, etc., 2023) [17]	0.8760
ICYL	**0.8872**

Kết quả phân lớp ảnh đa đối tượng bằng mạng học sâu YOLOv8 cao hơn một số phương pháp khác là bởi các lý do: (1) mạng học sâu YOLOv8 bổ sung thêm các đặc trưng cho nhận diện các đối tượng khó nhận diện do không đủ ánh sáng, kích thước nhỏ mà các phiên bản trước của YOLO chưa thực hiện được; (2) mạng học sâu YOLOv8 quá trình huấn luyện trên tập dữ liệu lớn nên góp phần hiệu suất phân lớp trên tập ảnh kiểm thử tốt hơn các phương pháp khác.

5. Kết luận và hướng phát triển

Trong bài báo này, một phương pháp để nhận diện và phân lớp hình ảnh đa đối tượng sử dụng mạng học sâu YOLOv8 được đề xuất và thực nghiệm trên bộ ảnh Flickr với kết quả phân lớp ảnh trung bình

là 0.8872. Mạng học sâu YOLOv8 đạt được hiệu suất nhận diện và phân lớp hình ảnh tốt hơn một số phương pháp khác thuật toán như k-NN, DNN, SVM. Đồng thời thời gian thực hiện trung bình thấp hơn các phiên bản trước của YOLO. Tuy nhiên, trong kết quả này cần cải tiến một số tham số và kết hợp thêm các kỹ thuật khác như k-NN hoặc SVM từ kết quả đầu ra của YOLOv8 thì kết quả phân lớp ảnh đa đối tượng sẽ cao hơn. Ngoài ra, bài toán xác định image captioning cho bộ ảnh Flickr cũng được nhóm nghiên cứu thực hiện trong tương lai dựa trên kết quả nhận diện và phân loại ảnh đối tượng bằng mạng học sâu YOLO và phân tích ngữ nghĩa hình ảnh cũng là một chủ đề cần được quan tâm và thử nghiệm trong tương lai.

Lời cám ơn

Chúng tôi xin trân trọng cám ơn Khoa Công nghệ thông tin, Trường Đại học Công Thương TP. HCM đã hỗ trợ về chuyên môn và tạo điều kiện về cơ sở vất chất giúp chúng tôi hoàn thành bài nghiên cứu này.

Xung đột lợi ích

Các tác giả tuyên bố không có xung đột lợi ích trong bài báo này.

Tuyên bố dữ liệu sẵn có

Dữ liệu hỗ trợ cho các khám phá của nghiên cứu này khi độc giả yêu cầu một cách hợp lý sẽ được tác giả liên hệ cung cấp.

TÀI LIỆU THAM KHẢO

[1] Y. Jiang *et al.*, "Breast cancer histopathological image classification using convolutional neural networks with small SE-ResNet module," *PloS One*, vol. 14, no. 3, p. e0214587, 2019.

[2] J. Guo and X. Wang, "Image classification based on SURF and KNN," in *2019 IEEE/ACIS 18th International Conference on Computer and Information Science (ICIS)*, 2019.

[3] M. A. Chandra and S. Bedi, "Survey on SVM and their application in image classification," *International Journal of Information Technology*, vol. 13, no. 5, pp. 1-11, 2021.

[4] S. Li *et al.*, "Deep learning for hyperspectral image classification: An overview," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 9, pp. 6690-6709, 2019.

[5] P. K. Mallick *et al.*, "Brain MRI image classification for cancer detection using deep wavelet autoencoder-based deep neural network," *IEEE Access*, vol. 7, pp. 46278-46287, 2019.

[6] M. Štancel and M. Hulič, "An introduction to image classification and object detection using YOLO detector," in *CEUR Workshop Proceedings*, 2019.

[7] J. Terven, D. M. C. Esparza, and J. A. R. González, "A comprehensive review of YOLO architectures in computer vision: From YOLOv1 to YOLOv8 and YOLO-NAS," *Machine Learning and Knowledge Extraction*, vol. 5, no. 4, pp. 1680-1716, 2023.

[8] Flickr. Dataset Flickr 2017. Available from: https://www.kaggle.com/datasets/hsankesara/flickr-image-dataset.

[9] L. Farokhah, "Implementasi K-Nearest Neighbor untuk Klasifikasi Bunga Dengan Ekstraksi Fitur Warna RGB," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*, vol. 7, no. 6, pp. 1129-1135, 2020.

[10] F. Sadati and B. Rezaie, "An improved image classification based on feature extraction from convolutional neural network: application to flower classification," in *12th International Conference on Information and Knowledge Technology (IKT)*, 2021.

[11] T. Stark *et al.*, "YOLO object detection models can locate and classify broad groups of flower-visiting arthropods in images," *Scientific Reports*, vol. 13, no. 1, p. 16364, 2023.

[12] N. Al Mudawi *et al.*, "Vehicle detection and classification via YOLOv8 and deep belief network over aerial image sequences," *Sustainability*, vol. 15, no. 19, p. 14597, 2023.

[13] B. Gašparović *et al.*, "Evaluating YOLOv5, YOLOv6, YOLOv7, and YOLOv8 in underwater environment: Is there real improvement?," in *8th International Conference on Smart and Sustainable Technologies (SpliTech)*, 2023.

[14] J. Zhang and H. Shi, "Kd-tree based efficient ensemble classification algorithm for imbalanced learning," in *2019 International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI)*, 2019.

[15] B. Selcuk and T. Serif, "A comparison of YOLOv5 and YOLOv8 in the context of mobile UI detection," in *International Conference on Mobile Web and Intelligent Information Systems*, 2023.

[16] S. Alijani, J. Tanha, and L. Mohammadkhanli, "An ensemble of deep learning algorithms for popularity prediction of Flickr images," *Multimedia Tools and Applications*, vol. 81, no. 3, pp. 3253-3274, 2022.

[17] M. Saroja and A. B. Mary, "Image Captioning Using Improved YOLO V5 Model and Xception V3 Model," 2023.

[18] N. T. Dinh and T. T. Van, "Image retrieval using YOLO deep learning network and KD-Tree Random Forest structure," in *Proceedings of the National Conference on Basic Research and IT Applications (FAIR22)*, 2022, ISBN: 978-604-357-119-6, doi: 10.15625/vap.2022.0244.

Nguyen Thi Dinh was born in 1983, graduated in Pedagogy Informatics Ho Chi Minh City University of Education in 2006, and received a Master's degree in industry Data transmission and computer network at Ho Chi Minh City Institute of Post and Telecommunications Technology Ho Chi Minh City in 2011. In 2023, she received a PhD degree in Computer Science from the University of Science, Hue, Vietnam.
Field research: image processing, image retrieval, and mechanics database.
Email: dinhnt@huit.edu.vn. ORCID: https://orcid.org/0000-0003-3428-3101

Truong Tran Bao Long was born in 2002 and is a fourth-year student majoring in Data Analysis at Ho Chi Minh City University of Industries and Trade.
Field research: image processing, image retrieval, and mechanics database.
Email: 2001200165@hufi.edu.vn. ORCID: https://orcid.org/0009-0001-3669-8565

Ngo Vuong Quoc Trung was born in 2002, and is currently a fourth-year student majoring in Data Analysis at Ho Chi Minh City University of Industries and Trade.
Field research: image processing, image retrieval, and mechanics database.
Email: 2001207135@hufi.edu.vn. ORCID: https://orcid.org/0009-0006-0438-3258

Tran Van Gia Bao was born in 2002, and is currently a fourth-year student majoring in Data Analysis at Ho Chi Minh City University of Industries and Trade.
Field research: image processing, image retrieval, and mechanics database.
Email: 2001207081@hufi.edu.vn. ORCID: https://orcid.org/0009-0009-8547-7281

Nguyen Duong Tuan was born in 2002, and is currently a fourth-year student majoring in Data Analysis at Ho Chi Minh City University of Industries and Trade.
Field research: image processing, image retrieval, and mechanics database.
Email: 2001207238@hufi.edu.vn. ORCID: https://orcid.org/0009-0006-0269-0924

Nguyen Phuong Hac was born in 1979, graduated in Ho Chi Minh City University of Science in 2002, and received a Master's degree in Hanoi University of Science and Technology in 2010.
Field research: image processing, image retrieval, and mechanics database.
Email: hacnp@huit.edu.vn. ORCID: https://orcid.org/0009-0007-1639-0620