

ỨNG DỤNG MÔ HÌNH HỌC SÂU CHO VIỆC GIÁM SÁT ĐỘNG VẬT TRONG SỞ THÚ

Nguyễn Dũng¹, Lê Nguyễn Thủy Nhi¹,
Hoàng Văn Dũng², Lê Văn Tường Lâm³, Đinh Hoài Nam⁴

¹Khoa Công nghệ thông tin, Trường Đại học Khoa học, Đại học Huế

²Trường Đại học Sư phạm Kỹ thuật TP. HCM

³Đại học Huế

⁴Trường Trung cấp Đắk Lắk

*Email: nguyendung@hueuni.edu.vn

Ngày nhận bài: 5/12/2024; ngày hoàn thành phản biện: 8/12/2024; ngày duyệt đăng: 20/12/2024

TÓM TẮT

Việc giám sát động vật trong sở thú đóng vai trò quan trọng trong quản lý, bảo tồn và các loài động vật. Trong nghiên cứu này, chúng tôi áp dụng mô hình YOLOv8, một trong những giải pháp phát hiện đối tượng tiên tiến và hiệu quả nhất hiện nay. Mô hình được huấn luyện ban đầu trên bộ dữ liệu COCO và sau đó tiếp tục được tinh chỉnh (fine-tuning) trên bộ dữ liệu African-wildlife nhằm tối ưu hóa khả năng nhận diện các loài động vật hoang dã phổ biến trong sở thú như trâu, voi, tê giác và ngựa vằn. Sau quá trình tinh chỉnh, mô hình đạt được mAP (mean Average Precision) 94.76% tại ngưỡng IoU 0.5, thể hiện độ chính xác và độ tin cậy vượt trội trong điều kiện thực tế. Nhờ khả năng xử lý nhanh và chính xác, mô hình này mang đến giải pháp hiệu quả cho việc giám sát động vật trong sở thú theo thời gian thực.

Từ khóa: YOLO, giám sát, phát hiện đối tượng, fine-tuning.

I. GIỚI THIỆU

Việc giám sát động vật trong các sở thú luôn giữ vai trò đặc biệt quan trọng, nhất là đối với các loài động vật có nguy cơ tuyệt chủng hoặc cần được bảo tồn nghiêm ngặt. Các sở thú không chỉ đơn thuần là nơi lưu giữ và trưng bày động vật hoang dã, mà còn là trung tâm nghiên cứu sinh thái và bảo tồn động vật quý hiếm. Để đảm bảo việc theo dõi sức khỏe, số lượng và môi trường sống của các loài động vật, các nhà quản lý sở thú cần triển khai các hệ thống giám sát hiệu quả, tự động và đáng tin cậy, đặc biệt ở những khu vực rộng lớn hoặc với số lượng loài đa dạng. Trong đó, việc tích hợp hệ thống

camera giám sát với các mô hình trí tuệ nhân tạo hiện đại đang nổi lên như một phương pháp tiên tiến và hiệu quả, cho phép nhận diện động vật tự động trong thời gian thực, hỗ trợ đắc lực cho công tác quản lý và bảo tồn.

Các phương pháp truyền thống, như quan sát bằng mắt thường hoặc sử dụng cảm biến đơn giản, thường bộc lộ nhiều hạn chế, bao gồm sự phụ thuộc vào yếu tố con người, dễ dẫn đến sai sót và khó triển khai hiệu quả ở các khu vực rộng lớn hoặc khi số lượng động vật đông. Ngược lại, sự phát triển của các công nghệ hiện đại, đặc biệt là thị giác máy tính và học sâu, đã mang đến những giải pháp đột phá, giúp tự động hóa quá trình giám sát động vật một cách hiệu quả. Trong số đó, các mô hình phát hiện đối tượng tiên tiến như Faster-RCNN [1], RetinaNet [2], SSD [3], YOLO [4-10], đặc biệt là phiên bản gần đây YOLOv8 [11], đã chứng minh khả năng vượt trội cả về tốc độ lẫn độ chính xác. Các mô hình này không chỉ phát hiện và đếm động vật hiệu quả mà còn hoạt động tốt trong các điều kiện môi trường phức tạp, mở ra hướng đi mới cho việc quản lý và bảo tồn động vật trong thời đại công nghệ số.

Bài báo này trình bày việc ứng dụng mô hình YOLOv8 để giám sát động vật trong sở thú, với quy trình huấn luyện ban đầu trên bộ dữ liệu COCO (Common Objects in Context) [12] và tinh chỉnh [13-15] trên bộ dữ liệu African-wildlife [16]. Nội dung được tổ chức thành bốn phần chính: (1) Giới thiệu – nêu bối cảnh nghiên cứu, tầm quan trọng của việc giám sát động vật và mục tiêu bài báo; (2) Các nghiên cứu liên quan đến nội dung nghiên cứu; (3) Phương pháp áp dụng – trình bày cấu trúc mô hình YOLOv8, đặc điểm bộ dữ liệu, quy trình huấn luyện và các thông số kỹ thuật quan trọng; (4) Kết quả thực nghiệm – phân tích hiệu quả mô hình thông qua các chỉ số như mAP, tốc độ xử lý thời gian thực và so sánh trong các điều kiện môi trường khác nhau; (5) Kết luận – tóm tắt đóng góp chính và thảo luận về ứng dụng tiềm năng trong quản lý, bảo tồn động vật. Nghiên cứu không chỉ khẳng định hiệu quả của YOLOv8 mà còn mở ra hướng phát triển ứng dụng AI trong lĩnh vực này.

II. NGHIÊN CỨU LIÊN QUAN

Việc ứng dụng trí tuệ nhân tạo, đặc biệt là các mô hình thị giác máy tính hiện đại, trong giám sát động vật hoang dã và động vật trong sở thú đã thu hút sự quan tâm đáng kể trong những năm gần đây. Các nghiên cứu đầu tiên chủ yếu tập trung vào việc sử dụng cảm biến gắn trên động vật hoặc camera bẫy ảnh (camera trap) để thu thập dữ liệu phục vụ nhận diện hoặc phân loại loài [17-19]. Tuy nhiên, phương pháp này thường đòi hỏi quá trình xử lý hậu kỳ tốn thời gian và nhân lực.

Sự phát triển nhanh chóng của các mô hình học sâu, đặc biệt là các mô hình phát hiện đối tượng, đã mở ra khả năng giám sát động vật trong thời gian thực với độ chính xác cao. Các mô hình như Faster-RCNN [1], RetinaNet [2], SSD [3] đặc biệt là các cải tiến

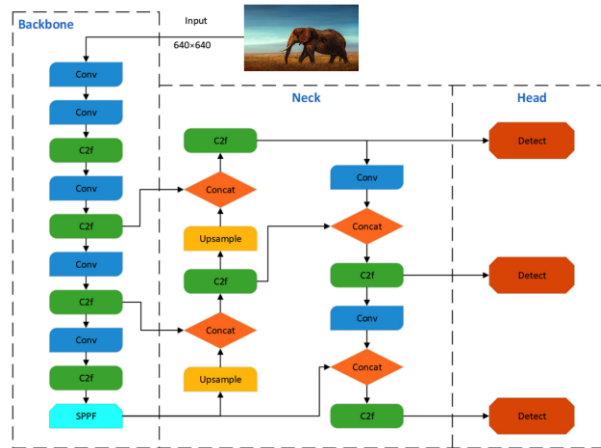
của mô hình YOLO [4-10] đã được sử dụng trong nhiều nghiên cứu để nhận diện động vật trong môi trường tự nhiên, khu bảo tồn hoặc sở thú. Chẳng hạn, Beery và cộng sự [12] đã đề xuất một phương pháp nhận diện động vật hoang dã trong các ảnh từ camera bằng cách sử dụng mô hình ResNet kết hợp với Faster-RCNN, cho thấy hiệu quả cao trong điều kiện ánh sáng và môi trường thay đổi. Trong nghiên cứu gần đây hơn, Norouzzadeh và cộng sự [20] đã áp dụng mô hình học sâu để tự động phân loại ảnh động vật trong tự nhiên, minh chứng cho tính khả thi và hiệu quả của các mô hình hiện đại trong giám sát động vật trên quy mô lớn. Ngoài ra, các phiên bản mới của YOLO như YOLOv5 và YOLOv8 đã được áp dụng thành công trong các hệ thống giám sát động vật với tốc độ xử lý nhanh, độ chính xác cao và khả năng triển khai trong điều kiện tài nguyên tính toán hạn chế [20].

Tuy nhiên, phần lớn các nghiên cứu hiện nay vẫn tập trung vào các mô hình có độ chính xác cao nhưng tốc độ xử lý còn hạn chế, hoặc ngược lại, ưu tiên tốc độ nhưng đánh đổi độ chính xác. Trong khi đó, bài toán giám sát động vật trong sở thú đòi hỏi cả hai yếu tố: phát hiện chính xác để đảm bảo theo dõi hiệu quả, và tốc độ xử lý nhanh để đáp ứng yêu cầu giám sát thời gian thực. Do đó, nghiên cứu này lựa chọn mô hình YOLOv8 – một kiến trúc hiện đại có khả năng cân bằng tốt giữa độ chính xác và tốc độ – nhằm ứng dụng vào việc phát hiện động vật trong môi trường sở thú phức tạp.

III. PHƯƠNG PHÁP ÁP DỤNG

3.1 Mô hình YOLOv8

YOLO (You Only Look Once) [4-11] là một trong những họ mô hình phát hiện đối tượng phổ biến nhất trong lĩnh vực thị giác máy tính. YOLOv8 mang lại nhiều cải tiến đáng kể về kiến trúc và hiệu suất so với các phiên bản trước đó. Khác với các mô hình phát hiện đối tượng khác, YOLO thực hiện việc phát hiện trong một lần xem duy nhất, giúp giảm thiểu thời gian xử lý nhưng vẫn duy trì được độ chính xác cao. Điều này đặc biệt quan trọng khi áp dụng trong các hệ thống giám sát động vật yêu cầu xử lý thời gian thực, nơi mà tốc độ và độ chính xác đều cần được đảm bảo. Hình 1 mô tả kiến trúc mô hình YOLOv8.



Hình 1. Kiến trúc mô hình YOLOv8

Một số đặc điểm nổi bật của YOLOv8 bao gồm:

- **Kiến trúc Anchor-free:** Trong các phiên bản trước của YOLO, việc xác định các hộp neo để giới hạn vùng phát hiện đối tượng là một khâu khá phức tạp. YOLOv8 loại bỏ yêu cầu này, giúp giảm độ phức tạp và tăng tính linh hoạt trong phát hiện đối tượng.
- **Backbone được cải tiến:** Mô hình sử dụng một backbone mới và hiệu quả hơn để rút trích đặc trưng từ hình ảnh, cho phép phát hiện tốt hơn các đối tượng ở các kích thước và khoảng cách khác nhau.
- **Hiệu suất cao trong thời gian thực:** Với khả năng xử lý nhanh chóng, YOLOv8 có thể phát hiện đối tượng trong các video hoặc hình ảnh được trích xuất từ camera giám sát một cách tức thời, đáp ứng tốt các yêu cầu của hệ thống giám sát động vật trong sở thú.

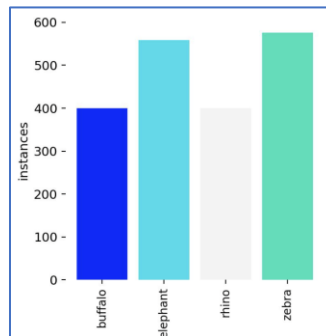
3.2 Bộ dữ liệu huấn luyện

Mô hình YOLOv8 được huấn luyện ban đầu trên bộ dữ liệu COCO, một trong những bộ dữ liệu phổ biến nhất trong thị giác máy tính, với hơn 80 lớp đối tượng khác nhau, bao gồm cả người, động vật và các vật thể thông thường. COCO không chỉ cung cấp một tập dữ liệu phong phú và đa dạng về hình ảnh mà còn giúp các mô hình học sâu được các đặc trưng của đối tượng trong nhiều tình huống khác nhau, bao gồm cả môi trường tự nhiên và nhân tạo. Tuy nhiên, để mô hình YOLOv8 có thể nhận diện và đếm các loài động vật trong sở thú một cách chính xác hơn, cần phải thực hiện thêm quá trình fine-tuning trên một bộ dữ liệu chuyên biệt hơn, trong trường hợp này là bộ dữ liệu African-wildlife. Bộ dữ liệu này tập trung vào các loài động vật hoang dã châu Phi, bao gồm nhiều loài như trâu nước, voi, tê giác, ngựa vằn. Hình 2 mô tả một số hình ảnh trong tập dữ liệu huấn luyện. Trong đó bộ dữ liệu được chia thành 2 tập: Tập dữ liệu huấn luyện chứa 1.052 hình ảnh, và tập dữ liệu xác thực bao gồm 225 hình ảnh.



Hình 2. Một số hình ảnh trong tập dữ liệu huấn luyện

Bằng cách fine-tuning mô hình trên bộ dữ liệu này, YOLOv8 có thể học thêm các đặc trưng cụ thể của các loài động vật thường gặp trong môi trường sở thú, từ đó nâng cao độ chính xác trong việc phát hiện và đếm động vật. Hình 3 mô tả số lượng đối tượng trên các lớp dữ liệu của bộ dữ liệu African-wildlife.



Hình 3. Phân phối đối tượng trên các lớp dữ liệu của African-wildlife

3.3 Quy trình huấn luyện

Quá trình huấn luyện mô hình YOLOv8 cho bài toán giám sát động vật trong sở thú bao gồm hai giai đoạn chính:

3.3.1 Giai đoạn 1 - Huấn luyện ban đầu trên bộ dữ liệu COCO

Trong giai đoạn này, mô hình được huấn luyện trên bộ dữ liệu COCO để học các đặc trưng cơ bản về hình dạng, kích thước và các mối liên hệ giữa các đối tượng. Bộ dữ liệu COCO cung cấp cho mô hình một nền tảng kiến thức chung về động vật và môi trường xung quanh chúng. Sau khi hoàn tất giai đoạn này, mô hình cho thấy tiềm năng đạt độ chính xác cao trong việc phát hiện các đối tượng động vật trong nhiều tình huống khác nhau.

3.3.2 Giai đoạn 2 - Fine-tuning trên bộ dữ liệu African-wildlife

Sau khi hoàn thành huấn luyện trên COCO, mô hình được fine-tuning trên bộ dữ liệu African-wildlife để cải thiện khả năng phát hiện các loài động vật đặc trưng của châu Phi, từ đó giúp mô hình thích ứng tốt hơn với các loài động vật trong sở thú. Trong quá trình fine-tuning, các lớp đầu của mô hình được giữ nguyên để bảo toàn kiến thức từ COCO, trong khi các lớp cuối được điều chỉnh để học thêm các đặc trưng từ dữ liệu động vật hoang dã. Quá trình huấn luyện và fine-tuning này giúp mô hình YOLOv8 không chỉ nhận diện chính xác các loài động vật mà còn có khả năng đếm chính xác số lượng động vật trong các điều kiện môi trường khác nhau, bao gồm cả sở thú.

IV. THỰC NGHIỆM VÀ KẾT QUẢ

4.1 Tham số huấn luyện

Trong quá trình huấn luyện mô hình YOLOv8 cho bài toán đếm động vật trong sở thú, chúng tôi đã sử dụng các tham số huấn luyện sau:

Bảng 1. Tham số huấn luyện của mô hình

Tham số	Giá trị
Learning Rate	0.00125
Batch Size	16
Epoch	50
Optimizer	AdamW
WeightDecay	0.0005
Momentum	0.937

4.2 Độ đo kết quả thực nghiệm

Nghiên cứu này sử dụng các độ đo sau để đánh giá hiệu suất của mô hình YOLOv8 trong giám sát động vật:

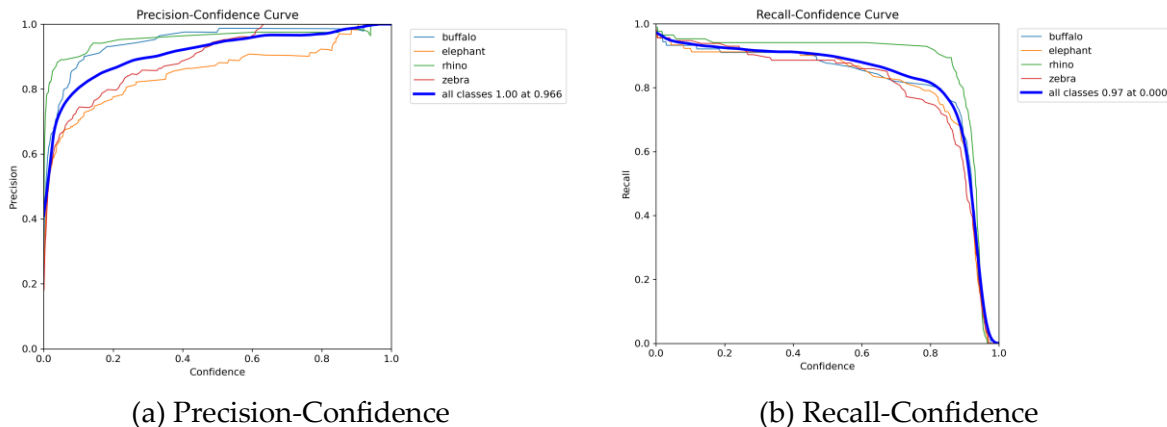
- **mAP (mean Average Precision):** Được tính trên nhiều ngưỡng IoU, phản ánh độ chính xác tổng thể của mô hình.
- **Precision:** Tỷ lệ giữa số đối tượng phát hiện đúng (TP) và tổng số đối tượng được phát hiện (TP + FP), cho biết mức độ chính xác của mô hình.
- **Recall:** Tỷ lệ giữa số đối tượng phát hiện đúng (TP) và tổng số đối tượng thực tế (TP + FN), đánh giá khả năng nhận diện đầy đủ các đối tượng.
- **F1-Score:** Chỉ số tổng hợp giữa Precision và Recall, thể hiện sự cân bằng giữa độ chính xác và khả năng phát hiện.

- **Thời gian dự đoán:** Đo lường tốc độ xử lý mỗi khung hình, đánh giá khả năng hoạt động thời gian thực.

4.3 Kết quả thực nghiệm

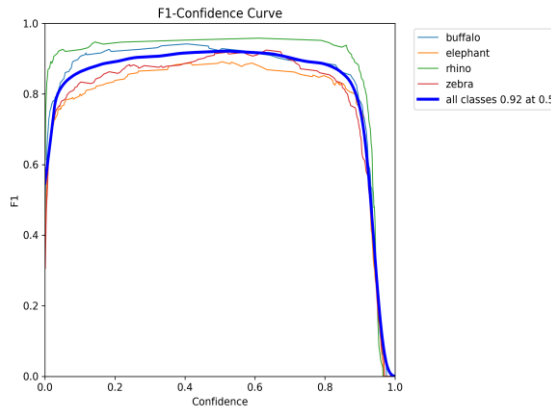
Hình 4 trình bày hai đồ thị Precision-Confidence và Recall-Confidence, phản ánh hiệu suất mô hình phân loại theo các lớp: buffalo, elephant, rhino và zebra. Khi ngưỡng độ tin cậy (confidence) tăng, độ chính xác (precision) cũng tăng – đặc biệt ở lớp buffalo và rhino – do giảm các dự đoán sai. Ngược lại, độ hồi tưởng (recall) giảm, cho thấy nhiều mẫu dương bị bỏ sót hơn. Đây là sự đánh đổi điển hình giữa precision và recall. Lớp rhino duy trì hiệu suất tốt nhất với cả hai chỉ số ở nhiều ngưỡng, trong khi elephant và zebra suy giảm mạnh khi confidence tăng. Đường cong tổng hợp cho thấy mô hình đạt precision tuyệt đối ở confidence 0.966 và recall khoảng 97% ở ngưỡng thấp nhất, phản ánh hiệu suất tốt nhưng vẫn có đánh đổi rõ rệt.

Hình 5 minh họa đồ thị F1-Confidence, cho thấy sự cân bằng giữa precision và recall. Lớp rhino tiếp tục nổi bật với chỉ số F1 ổn định và cao, trong khi elephant khó duy trì sự cân bằng. Đường cong tổng hợp đạt F1 tối đa 0.92 tại confidence 0.520 – được xem là ngưỡng tối ưu. Tuy nhiên, khi confidence tăng, F1 giảm rõ rệt, đặc biệt ở zebra và elephant, do precision tăng nhưng recall giảm, dẫn đến F1 giảm ở các ngưỡng cao.

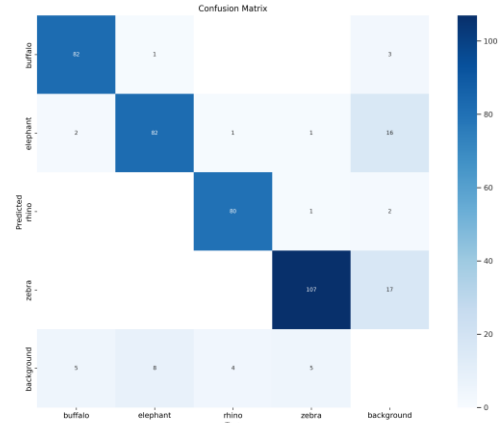


Hình 4. Đường cong Precision-Confidence (a) và Recall-Confidence (b)

Hình 6 mô tả ma trận nhầm lẫn khi dự đoán đối tượng. Hình này cho thấy mô hình hoạt động khá tốt với các loài như buffalo, rhino và zebra, với tỷ lệ dự đoán đúng trên 90%. Tuy nhiên, nhầm lẫn với nền là vấn đề lớn, đặc biệt với elephant và zebra, khi tỷ lệ bị nhầm với nền lần lượt là 42% và 45%. Điều này có thể do các đặc điểm hình dáng, màu sắc của chúng dễ bị trùng lặp với môi trường xung quanh. Điều này cho thấy mô hình gặp khó khăn trong việc phân biệt rõ giữa đối tượng và nền. Để cải thiện, mô hình có thể cần tăng cường dữ liệu hoặc áp dụng các kỹ thuật nhằm xử lý tốt hơn sự khác biệt giữa động vật và nền, giúp giảm tỷ lệ nhầm lẫn.

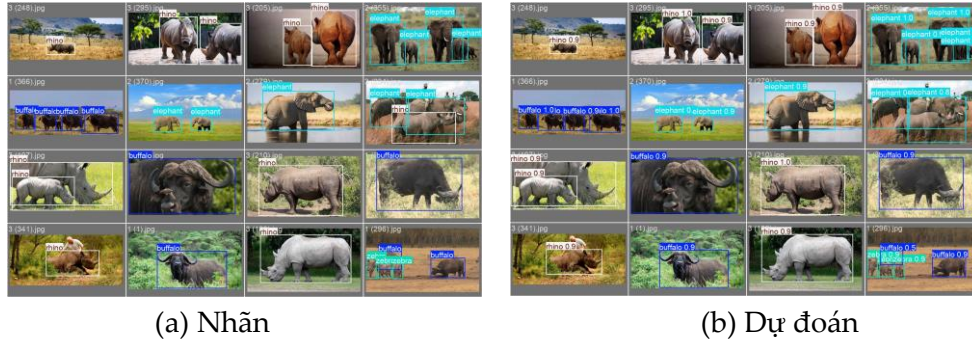


Hình 5. Đường cong F1-Score



Hình 6. Ma trận nhầm lẫn

Một số hình ảnh dự đoán cũng được mô tả như ở hình 7.



Hình 7. Một số kết quả phát hiện trên tập dữ liệu xác thực: (a) Nhận, (b) Dự đoán

Kết quả thực nghiệm của mô hình trên tập dữ liệu xác thực được thể hiện ở bảng 2 cho thấy hiệu suất cao với $mAP@0.5$ đạt 94.76% và $mAP@0.5-0.95$ là 81.16%, chứng tỏ mô hình có khả năng phát hiện đối tượng chính xác ở nhiều mức ngưỡng khác nhau.

Bảng 2. Kết quả thực nghiệm của mô hình trên tập dữ liệu xác thực

Tham số	Giá trị
$mAP@0.5$	94.76%
$mAP@0.5-0.95$	81.16%

Về mặt thông số kỹ thuật được mô tả ở bảng 3, mô hình có khoảng 43M tham số, 365 lớp mô hình, và đòi hỏi 165.4 Gflops để tính toán. Số lượng phép đạo hàm là khoảng 43M thời gian dự đoán trung bình mỗi lần cho 1 một ảnh là 184.5 ms, và mô hình dự đoán trên 4 lớp. Những thông số này thể hiện một mô hình có độ phức tạp cao nhưng vẫn đạt được hiệu suất tốt trên tập dữ liệu kiểm thử.

Bảng 3. Thông số kỹ thuật của mô hình thực nghiệm

Tham số	Giá trị
Số lượng tham số	43,632,924

Số lớp mô hình	365
Gflop	165.4
Số lượng phép đạo hàm	43,632,908
Thời gian dự đoán	184.5ms
Số lớp dự đoán	4

V. KẾT LUẬN

Nghiên cứu này đã trình bày quá trình huấn luyện và tinh chỉnh mô hình YOLOv8 cho bài toán phát hiện động vật trong sở thú. Thông qua huấn luyện ban đầu trên bộ dữ liệu COCO và tinh chỉnh trên bộ dữ liệu African-wildlife, mô hình đạt độ chính xác ấn tượng với mAP lên đến 94.76%, đồng thời vẫn đảm bảo khả năng xử lý thời gian thực với tốc độ trung bình 0.1845 giây/ảnh. Các đặc điểm nổi bật của YOLOv8, như kiến trúc anchor-free và cơ chế phát hiện nhanh, đã chứng minh sự phù hợp trong các điều kiện môi trường phức tạp, nhiều đối tượng như ở sở thú. Kết quả thực nghiệm cho thấy mô hình có tiềm năng lớn trong việc hỗ trợ các hệ thống giám sát tự động, góp phần nâng cao hiệu quả quản lý, theo dõi sức khỏe, hành vi và số lượng động vật – đặc biệt là các loài quý hiếm hoặc có nguy cơ tuyệt chủng. Trong tương lai, nghiên cứu có thể được mở rộng theo hướng áp dụng cho nhiều môi trường sống khác nhau (ví dụ: công viên hoang dã, khu bảo tồn thiên nhiên), đồng thời tích hợp thêm khả năng nhận diện hành vi hoặc bất thường nhằm phục vụ tốt hơn cho công tác bảo tồn và giám sát động vật một cách toàn diện và thông minh hơn.

LỜI CẢM ƠN

Nghiên cứu này được tài trợ bởi Đại học Huế qua đề tài mã số DHH2025-01-226.

TÀI LIỆU THAM KHẢO

- [1] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *Advances in neural information processing systems*, vol. 28, 2015.
- [2] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2980-2988.
- [3] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "Ssd: Single shot multibox detector," in *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14*, 2016: Springer, pp. 21-37.

- [4] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779-788.
- [5] J. Redmon and A. Farhadi, "YOLO9000: better, faster, stronger," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 7263-7271.
- [6] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," *arXiv preprint arXiv:1804.02767*, 2018.
- [7] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "Yolov4: Optimal speed and accuracy of object detection," *arXiv preprint arXiv:2004.10934*, 2020.
- [8] G. Jocher *et al.*, "ultralytics/yolov5: v3. 0," *Zenodo*, 2020.
- [9] C. Li *et al.*, "Yolov6 v3. 0: A full-scale reloading," *arXiv preprint arXiv:2301.05586*, 2023.
- [10] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 7464-7475.
- [11] G. J. a. A. C. a. J. Qiu, "Ultralytics YOLOv8," 2023. [Online]. Available: <https://github.com/ultralytics/ultralytics>.
- [12] T.-Y. Lin *et al.*, "Microsoft coco: Common objects in context," in *Computer Vision—ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13*, 2014: Springer, pp. 740-755.
- [13] M. Bansal, M. Kumar, M. Sachdeva, and A. Mittal, "Transfer learning for image classification using VGG19: Caltech-101 image data set," *Journal of ambient intelligence and humanized computing*, pp. 1-12, 2021.
- [14] T.-A. P. a. V.-D. Hoang, "Chest X-ray Image Classification using Transfer Learning and Hyperparameter Customization for Lung Disease Diagnosis," *Journal of Information and Telecommunication*, 2024, doi: 10.1080/24751839.2024.2317509.
- [15] H. O. Ikromovich and B. B. Mamatkulovich, "Facial recognition using transfer learning in the deep cnn," *Open Access Repository*, vol. 4, no. 3, pp. 502-507, 2023.
- [16] *African Wildlife Dataset*. [Online]. Available: <https://github.com/ultralytics/ultralytics/blob/main/ultralytics/cfg/datasets/african-wildlife.yaml>
- [17] A. Caravaggi *et al.*, "A review of camera trapping for conservation behaviour research," *Remote Sensing in Ecology and Conservation*, vol. 3, no. 3, pp. 109-122, 2017.
- [18] R. Steenweg *et al.*, "Scaling-up camera traps: Monitoring the planet's biodiversity with networks of remote sensors," *Frontiers in Ecology and the Environment*, vol. 15, no. 1, pp. 26-34, 2017.
- [19] M. A. Tabak *et al.*, "Machine learning to classify animal species in camera trap images: Applications in ecology," *Methods in Ecology and Evolution*, vol. 10, no. 4, pp. 585-590, 2019.
- [20] M. S. Norouzzadeh, A. Nguyen, M. Kosmala, A. Swanson, M. S. Palmer, C. Packer, and J. Clune, "Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning," *Proceedings of the National Academy of Sciences*, vol. 115, no. 25, pp. E5716-E5725, 2018.

APPLICATION OF DEEP LEARNING TECHNIQUES TO ZOO ANIMAL SURVEILLANCE

Nguyen Dung^{1*}, Le Nguyen Thuy Nhi¹,
Hoang Van Dung², Le Van Tuong Lan³, Dinh Hoai Nam⁴

¹University of Sciences, Hue University

²HCMC University of Technology and Education, Ho Chi Minh city

³Hue University, Hue city

⁴DakLak Polytechnic College

*Email: nguyendung@hueuni.edu.vn

ABSTRACT

Monitoring zoo animals is essential for effective species conservation, welfare management, and ecological research. In this study, we apply YOLOv8, one of the most advanced and efficient object detection models. The YOLOv8 model was initially trained on the COCO dataset and fine-tuned on the African Wildlife dataset to improve its recognition of common zoo species such as buffalo, elephants, rhinos, and zebras. After fine-tuning, the model achieved a mean Average Precision (mAP) of 94.76% at an IoU threshold of 0.5, demonstrating outstanding accuracy and reliability in real-world conditions. Thanks to its high-speed and precise processing capabilities, this model offers an effective real-time zoo animal monitoring solution.

Keywords: YOLO, Surveillance, Object detection, Fine-tuning.



Nguyễn Dũng sinh ngày 13/06/1988 tại Thừa Thiên Huế. Ông tốt nghiệp cử nhân Tin học tại trường Đại học Khoa học, Đại học Huế năm 2010. Năm 2013, ông tốt nghiệp thạc sĩ chuyên ngành Khoa học máy tính tại trường Đại học Khoa học, Đại học Huế. Hiện nay, ông là giảng viên Khoa Công nghệ thông tin, trường Đại học Khoa học, Đại học Huế.

Lĩnh vực nghiên cứu: Công nghệ phần mềm, trí tuệ nhân tạo, học máy, học sâu, cơ sở dữ liệu.



Lê Nguyễn Thủy Nhi sinh năm 1998 tại Huế. Bà tốt nghiệp Cử nhân ngành Công nghệ thông tin năm 2020 và Thạc sĩ ngành Khoa học máy tính năm 2023 tại Trường Đại học Khoa học, Đại học Huế. Bà hiện đang là giảng viên Khoa Công nghệ thông tin, Trường Đại học Khoa học, Đại học Huế.

Lĩnh vực nghiên cứu: Công nghệ phần mềm, học máy, học sâu, trí tuệ nhân tạo.



Hoàng Văn Dũng nhận bằng tiến sĩ ngành Kỹ thuật điện tử và Hệ thống thông tin tại Trường Đại học Ulsan, Hàn Quốc, năm 2015. Hiện nay, ông công tác tại Khoa Công nghệ thông tin, Trường ĐH Sư phạm Kỹ thuật TP. HCM.

Lĩnh vực nghiên cứu: Trí tuệ nhân tạo, Thị giác máy tính



Lê Văn Tường Lân sinh ngày 10/11/1974 tại Thừa Thiên Huế. Năm 1996, ông tốt nghiệp Đại học ngành Toán - Tin tại Trường Đại học Khoa học, Đại học Huế. Ông nhận bằng thạc sĩ Công nghệ thông tin tại Trường Đại học Bách Khoa Hà Nội năm 2002 và nhận học vị Tiến sĩ ngành Khoa học máy tính tại Trường Đại học Khoa học, Đại học Huế năm 2018. Hiện nay, ông công tác tại Ban Đào tạo và Công tác sinh viên, Đại học Huế.

Lĩnh vực nghiên cứu: Cơ sở dữ liệu, Công nghệ phần mềm, Khai phá dữ liệu.



Đinh Hoài Nam sinh năm 1977 tại Thừa Thiên Huế. Ông tốt nghiệp Cử nhân ngành Công nghệ thông tin năm 2003 Đại Học Đà Lạt, Thạc sĩ ngành Khoa học máy tính năm 2020 tại Trường Đại học Khoa học, Đại học Huế. Ông hiện đang là Trưởng khoa Công nghệ thông tin, Trường trung Cấp Đắc Lắc, TP Buôn Ma Thuột.

Lĩnh vực nghiên cứu: Công nghệ mạng máy tính, Công nghệ phần mềm.