

DOI:10.22144/ctujos.2025.236

MỘT HƯỚNG TIẾP CẬN PHÂN LOẠI CÔN TRÙNG SỬ DỤNG MÔ HÌNH HỌC SÂU YOLO

Nguyễn Hữu Hòa, Phạm Thế Phi và Thái Minh Tuấn*

Trường Công nghệ thông tin và Truyền thông, Đại học Cần Thơ, Việt Nam

*Tác giả liên hệ (Corresponding author): minhluan@ctu.edu.vn

Thông tin chung (Article Information)

Nhận bài (Received): 15/06/2025

Sửa bài (Revised): 16/11/2025

Duyệt đăng (Accepted): 24/11/2025

Title: An approach to insect classification using YOLO-based deep learning models

Author: Nguyen Huu Hoa, Pham The Phi and Thai Minh Tuan*

Affiliation(s): College of Information and Communication Technology, Can Tho University, Viet Nam

TÓM TẮT

Phân loại côn trùng chính xác là yếu tố quan trọng trong quản lý sâu bệnh và bảo vệ cây trồng. Nghiên cứu này đề xuất một phương pháp phân loại côn trùng sử dụng mô hình học sâu YOLOv11 với các cải tiến về kiến trúc và hiệu suất. Hai tập dữ liệu được sử dụng là IP102 và PEST204, trong đó PEST204 là phiên bản mở rộng từ IP102 với số lượng loài và chất lượng hình ảnh được cải thiện. Các biến thể của YOLOv11 được áp dụng kết hợp với học chuyển tiếp và tăng cường dữ liệu nhằm nâng cao độ chính xác. Mô hình được đánh giá qua các chỉ số như độ chính xác, thời gian xử lý và tài nguyên tính toán. Kết quả cho thấy YOLO11x-clas đạt độ chính xác cao nhất, trong khi YOLO11n-clas và YOLO11s-clas thể hiện tiềm năng triển khai trên thiết bị di động nhờ hiệu suất tốt và tốc độ nhanh. Ngoài ra, kỹ thuật Eigen-CAM được sử dụng để trực quan hóa vùng ảnh quan trọng, giúp tăng khả năng diễn giải mô hình. Kết quả cho thấy hiệu quả của YOLOv11 trong phân loại côn trùng, đóng góp vào phát triển hệ thống nông nghiệp thông minh.

Từ khóa: Học sâu, phân loại côn trùng, YOLO

ABSTRACT

Accurate insect classification is a key factor in pest management and crop protection. This study proposes an insect classification method using the YOLOv11 deep learning model, which features architectural and performance improvements. Two datasets are used: IP102 and PEST204, with PEST204 being an extended version of IP102 that includes more species and enhanced image quality. Variants of YOLOv11 are applied in combination with transfer learning and data augmentation to improve accuracy. The models are evaluated using metrics such as accuracy, processing time, and computational resource usage. Results showed that YOLO11x-clas achieves the highest accuracy, while YOLO11n-clas and YOLO11s-clas demonstrate potential for deployment on mobile devices due to their good performance and fast inference speed. Additionally, the Eigen-CAM technique is used to visualize important image regions, enhancing model interpretability. The findings confirm the effectiveness of YOLOv11 in insect classification, contributing to the development of intelligent agricultural systems.

Keywords: Deep learning, insect classification, YOLO

1. GIỚI THIỆU

Côn trùng đóng vai trò quan trọng trong hệ sinh thái nông nghiệp, vừa là tác nhân gây hại cho cây trồng, vừa là các loài thiên địch có ích. Các loài côn trùng gây hại có thể làm giảm sản lượng và chất lượng nông sản, gây tổn thất nghiêm trọng cho sản xuất nông nghiệp (Estruch et al., 1997). Ngược lại, những loài thiên địch hỗ trợ kiểm soát sâu bệnh một cách tự nhiên, góp phần cân bằng hệ sinh thái và hạn chế việc sử dụng thuốc bảo vệ thực vật. Tuy nhiên, dưới tác động của biến đổi khí hậu, ô nhiễm môi trường và việc lạm dụng hóa chất nông nghiệp, nhiều loài côn trùng có ích đang bị suy giảm nhanh chóng, trong khi các loài gây hại ngày càng phát triển mạnh (Oberti et al., 2016).

Phân loại và nhận dạng chính xác côn trùng là yếu tố then chốt trong việc xây dựng các biện pháp phòng trừ sâu bệnh hiệu quả và bền vững. Tuy nhiên, việc nhận diện côn trùng theo phương pháp truyền thống đòi hỏi chuyên môn sâu và tốn nhiều thời gian do sự đa dạng về loài, sự khác biệt hình thái tinh vi và các yếu tố ngoại cảnh như điều kiện ánh sáng, tư thế hay nền ảnh phức tạp (Al-Hiary et al., 2011). Do đó, nhu cầu phát triển các hệ thống tự động nhận dạng côn trùng ngày càng trở nên cấp thiết, nhất là trong bối cảnh phát triển nông nghiệp thông minh và bền vững.

Trong bối cảnh hệ thống nông nghiệp thông minh, phân loại côn trùng là một khâu trung gian quan trọng trong chuỗi giám sát – cảnh báo – ra quyết định. Kết quả phân loại theo loài cho phép hệ thống tự động ghi nhận thời điểm và mật độ xuất hiện của từng loài gây hại trên từng khu vực, từ đó cung cấp đầu vào cho các mô-đun dự báo bùng phát dịch hại và mô-đun khuyến nghị biện pháp xử lý phù hợp với từng giai đoạn sinh trưởng của cây trồng. Vì vậy, bài toán trong nghiên cứu này không chỉ dừng lại ở phân loại hình ảnh thuần túy, mà gắn trực tiếp với việc hỗ trợ ra quyết định trong quản lý sâu bệnh.

Những tiến bộ gần đây của công nghệ học sâu (deep learning) đã mở ra nhiều hướng tiếp cận mới cho bài toán phân loại hình ảnh nói chung và nhận dạng côn trùng nói riêng (Kirkeby et al., 2021; Rustia et al., 2021; Doan et al., 2022). Các mô hình học sâu, đặc biệt là các kiến trúc CNN hiện đại, đã chứng minh khả năng trích xuất đặc trưng mạnh mẽ, cho phép đạt độ chính xác vượt trội so với các phương pháp truyền thống. Trong số đó, dòng mô hình YOLO (You Only Look Once) nổi bật nhờ khả năng cân bằng tốt giữa tốc độ xử lý và độ chính xác, và đã được ứng dụng rộng rãi trong nhiều bài toán

thị giác máy tính (Wang et al., 2012; Krause et al., 2013).

Bài báo này đề xuất một hướng tiếp cận phân loại côn trùng dựa trên mô hình học sâu YOLO, cụ thể là sử dụng phiên bản mới nhất YOLOv11 với nhiều cải tiến về kiến trúc và hiệu suất. Các biến thể của YOLOv11 được khai thác để đánh giá khả năng phân loại trên hai bộ dữ liệu IP102 và PEST204, trong đó PEST204 là phiên bản mở rộng và làm sạch từ IP102 nhằm khắc phục các hạn chế về chất lượng dữ liệu. Bên cạnh đó, hướng tiếp cận còn kết hợp kỹ thuật học chuyển tiếp (transfer learning) và tăng cường dữ liệu (data augmentation) nhằm nâng cao hiệu quả huấn luyện, đồng thời áp dụng Eigen-CAM để phân tích và trực quan hóa các vùng ảnh trọng yếu, giúp tăng khả năng diễn giải mô hình.

Đóng góp chính của bài báo có thể tóm tắt như sau:

- Đề xuất và đánh giá một cấu hình áp dụng các biến thể YOLOv11-cls kết hợp học chuyển tiếp từ ImageNet và tăng cường dữ liệu cho bài toán phân loại côn trùng trên hai bộ dữ liệu IP102 và PEST204. Nghiên cứu cung cấp một phân tích định lượng có hệ thống về độ chính xác, thời gian suy luận và chi phí tính toán cho từng biến thể mô hình.
- Xây dựng bộ dữ liệu PEST204 mở rộng từ IP102 với 204 lớp côn trùng gây hại nông nghiệp. Các ảnh được kế thừa từ IP102 và thu thập bổ sung từ các nguồn hình ảnh nông nghiệp công khai, sau đó được lọc bỏ các ảnh nhiễu và rà soát lại nhãn để đảm bảo tính nhất quán. Bộ dữ liệu này góp phần cải thiện độ cân bằng giữa các lớp và chất lượng hình ảnh so với IP102, từ đó phản ánh tốt hơn các tình huống nhận dạng côn trùng trong thực tế.
- Đánh giá và phân tích chi tiết hiệu suất của các biến thể YOLOv11 trên hai tập dữ liệu IP102 và PEST204 thông qua nhiều chỉ số như độ chính xác Top-1, Top-5, thời gian suy luận và mức sử dụng tài nguyên.
- Ứng dụng phương pháp Eigen-CAM để trực quan hóa vùng trọng yếu trong quá trình phân loại, qua đó nâng cao khả năng giải thích kết quả mô hình.

2. TỔNG QUAN NGHIÊN CỨU LIÊN QUAN

Trong những năm gần đây, các phương pháp học sâu đã trở thành xu hướng chủ đạo trong bài toán phân loại côn trùng nhờ khả năng trích xuất đặc trưng tự động và hiệu quả từ hình ảnh. Các mô hình

mạng nơ-ron tích chập (CNN) đã được áp dụng rộng rãi và đạt được nhiều thành tựu đáng kể trong phân loại hình ảnh sâu bệnh và côn trùng nông nghiệp (Chithambarathanu et al., 2023).

Một số nghiên cứu đã chứng minh hiệu quả của việc ứng dụng CNN trong nhận dạng côn trùng. Hansen et al. (2020) đã áp dụng CNN để phân loại các loài bọ cánh cứng tại Anh, đồng thời cho thấy việc thu thập hình ảnh hành vi và môi trường sống của côn trùng giúp cải thiện đáng kể độ chính xác của mô hình. Xia et al. (2018) đã sử dụng CNN để trích xuất đặc trưng đa cấp nhằm thực hiện bài toán phân loại nhiều lớp côn trùng, đạt kết quả vượt trội so với các phương pháp học máy truyền thống. Ngoài ra, Li et al. (2020) đã tinh chỉnh mô hình GoogLeNet để phân loại 10 loài côn trùng trong điều kiện thực địa phức tạp, cho thấy khả năng ứng dụng thực tiễn cao của học sâu trong môi trường tự nhiên.

Nghiên cứu của Almryad and Kutucu (2020) đã áp dụng các mô hình deep learning như VGG16, VGG19 và ResNet50 để nhận dạng 10 loài bướm, đạt độ chính xác 79,5% ngay cả khi đối mặt với những thách thức như vị trí, góc chụp, khoảng cách, và nền ảnh phức tạp. Bên cạnh đó, Doan (2023) đề xuất kết hợp mô hình EfficientNet với phương pháp phân loại SVM, đạt độ chính xác cao trên các tập dữ liệu Xie24, D0 và IP102.

Các phương pháp khác cũng được phát triển nhằm nâng cao khả năng nhận diện côn trùng trong các điều kiện môi trường thay đổi như ánh sáng, chuyển động và phức tạp nền (Wang et al., 2017; Li et al., 2022). Tuy nhiên, các nghiên cứu này thường chỉ đạt hiệu quả cao trên các tập dữ liệu hẹp hoặc trong điều kiện kiểm soát, trong khi hiệu suất thực tế trên các bộ dữ liệu lớn và đa dạng như IP102 vẫn còn nhiều hạn chế.

Nhằm khắc phục những bất cập này, một số công trình nghiên cứu đã đề xuất các chiến lược tăng cường như kết hợp phương pháp chú ý (saliency methods) với CNN, hoặc xây dựng các mô hình tổng hợp (ensemble models) để cải thiện độ chính xác phân loại (Nanni et al., 2022).

Nhìn chung, mặc dù học sâu đã mở ra nhiều cơ hội cho bài toán phân loại côn trùng, các nghiên cứu hiện tại vẫn còn đối mặt với các thách thức lớn như: số lượng loài phong phú, hình thái tương đồng, dữ liệu huấn luyện không đồng đều và nhiễu trong môi trường tự nhiên. Để nâng cao hiệu quả nhận dạng, cần những hướng tiếp cận mới khai thác các mô hình

học sâu tối ưu hơn, đồng thời cải thiện chất lượng và độ đa dạng của bộ dữ liệu.

Ngoài họ mô hình YOLO, nhiều kiến trúc nhẹ khác như MobileNet, ShuffleNet và EfficientNet cũng được sử dụng rộng rãi cho các bài toán phân loại trên thiết bị di động nhờ số lượng tham số nhỏ và chi phí tính toán thấp. Tuy nhiên, trong phạm vi bài báo này, các biến thể YOLOv11-cla cho bài toán phân loại côn trùng được tập trung đánh giá một cách có hệ thống, do đó chưa tiến hành so sánh trực tiếp với các kiến trúc nhẹ khác. Việc mở rộng so sánh thực nghiệm với các mô hình như MobileNet, EfficientNet hay ShuffleNet được xem là hướng nghiên cứu tiếp theo.

Xuất phát từ những nhận định đó, nghiên cứu này đề xuất sử dụng các biến thể của mô hình YOLOv11, kết hợp với học chuyển tiếp và tăng cường dữ liệu trên bộ tập dữ liệu được nâng cấp PEST204, nhằm khắc phục các hạn chế của các phương pháp trước đây và nâng cao hiệu quả phân loại côn trùng trong điều kiện thực tế.

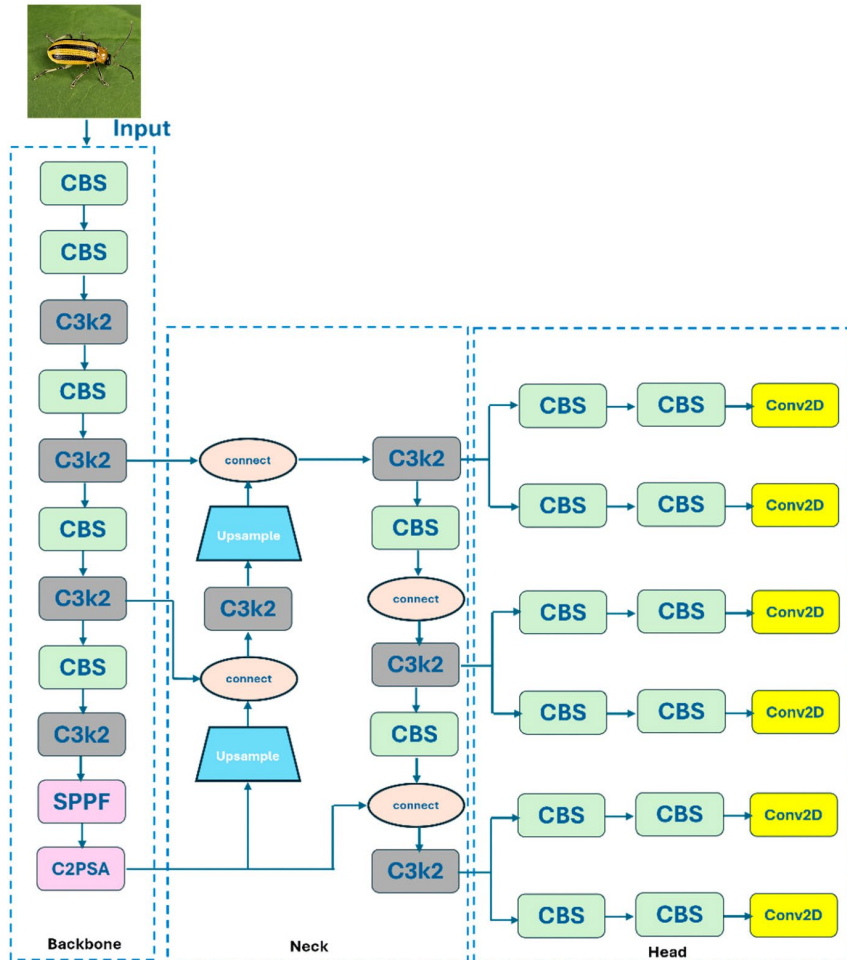
3. PHƯƠNG PHÁP

3.1. Mô hình YOLOv11

YOLO (You Only Look Once) là một dòng mô hình học sâu nổi bật trong lĩnh vực nhận dạng đối tượng nhờ khả năng cân bằng tốt giữa tốc độ và độ chính xác. Trong các phiên bản mới nhất, YOLOv11 tiếp tục kế thừa và phát triển các cải tiến kiến trúc nhằm tối ưu hóa hiệu suất cho nhiều nhiệm vụ thị giác máy tính, bao gồm phân loại hình ảnh, phát hiện đối tượng, phân đoạn và nhận dạng tư thế.

Khác với các phiên bản trước, YOLOv11 được thiết kế với ba thành phần chính: Backbone, Neck, và Head, mỗi thành phần được tối ưu hóa để tăng khả năng trích xuất đặc trưng, truyền tải thông tin và dự đoán đầu ra chính xác. Kiến trúc tổng thể của YOLOv11 dành cho bài toán phân loại côn trùng được minh họa trong Hình 1.

Trong kiến trúc Backbone, YOLOv11 sử dụng các khối CBS (Convolution-BatchNorm-SiLU) để trích xuất đặc trưng cơ bản từ hình ảnh đầu vào, kết hợp với các khối C3k2 nhằm khai thác các mối quan hệ ngữ cảnh cục bộ và toàn cục hiệu quả hơn. Cấu trúc C3k2 là một biến thể cải tiến từ khối C3 truyền thống với hai tầng convolution 3x3 thay vì ba, giúp giảm số lượng tham số mà vẫn duy trì hiệu quả trích xuất đặc trưng. Công thức tổng quát cho đầu ra của một khối C3k2 có thể biểu diễn như sau:



Hình 1. Kiến trúc tổng thể của mô hình YOLOv11 được sử dụng cho bài toán phân loại côn trùng

$$C3k2(X) = \text{Conv}3 \times 3(X) + \text{Conv}3 \times 3(\text{Split}(X))$$

Trong đó, X là đầu vào và phép Split chia X thành hai nhánh xử lý song song, sau đó cộng gộp để tái hợp thông tin.

Bên cạnh đó, YOLOv11 bổ sung mô-đun SPPF (Spatial Pyramid Pooling Fast) nhằm mở rộng trường nhìn nhận diện (receptive field) mà không làm tăng độ phức tạp tính toán đáng kể. Đồng thời, mô-đun C2PSA (Cross-Stage Partial with Self-Attention) được tích hợp để tăng cường khả năng nắm bắt các mối quan hệ không gian dài hạn thông qua cơ chế tự chú ý (self-attention), đặc biệt hữu ích trong việc phân loại các loài côn trùng có hình thái tinh vi.

Trong thành phần Neck, YOLOv11 áp dụng chiến lược kết nối skip (skip connections) kết hợp với các khối upsample và C3k2 nhằm tái kết hợp thông tin từ các tầng đặc trưng sâu và nông, giúp duy

trì độ chi tiết trong quá trình truyền đặc trưng. Phần Head của mô hình được tối ưu hóa cho nhiệm vụ phân loại với các tầng convolution 2D cuối cùng để sinh ra các xác suất phân lớp. Tổng thể, kiến trúc của YOLOv11 thể hiện một sự kết hợp hài hòa giữa việc khai thác các đặc trưng cục bộ và toàn cục, tăng cường khả năng học biểu diễn phức tạp mà vẫn đảm bảo tốc độ suy luận nhanh và khả năng tối ưu tài nguyên.

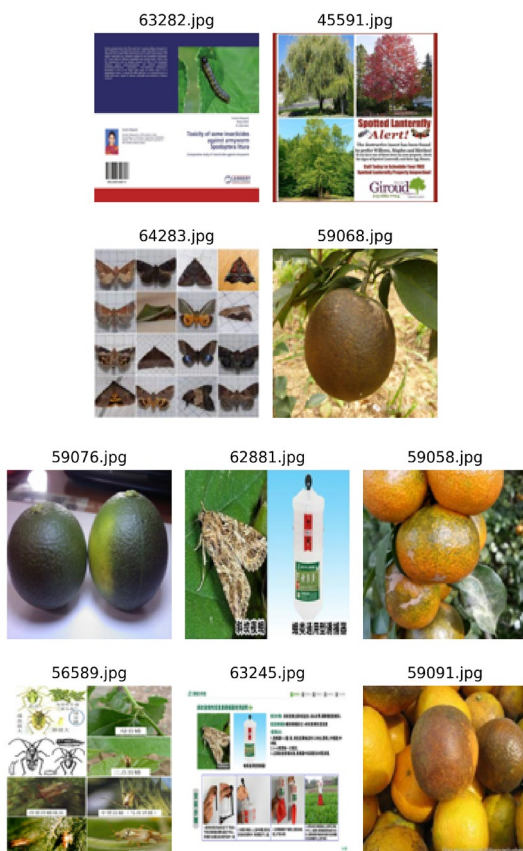
Hình 1 minh họa trực quan kiến trúc tổng thể của mô hình YOLOv11 áp dụng trong bài toán phân loại côn trùng. Hình ảnh đầu vào được xử lý qua chuỗi các khối CBS và C3k2 tại Backbone, sau đó thông tin đặc trưng được tái tổ hợp tại Neck thông qua các phép kết nối và upsampling. Cuối cùng, phần Head thực hiện việc phân loại bằng cách ánh xạ các đặc trưng tổng hợp thành các xác suất lớp đầu ra. Cấu trúc này cho phép mô hình khai thác tốt các đặc

trung đa cấp, từ đó nâng cao hiệu quả nhận dạng các loài côn trùng trong môi trường tự nhiên phức tạp.

3.2. Tập dữ liệu

3.2.1. Tập dữ liệu IP102

IP102 là một trong những bộ dữ liệu lớn nhất hiện nay phục vụ cho bài toán phân loại côn trùng, được công bố bởi Wu et al. (2019). Bộ dữ liệu này bao gồm hơn 75.000 hình ảnh thuộc 102 loài côn trùng gây hại khác nhau, được thu thập từ nhiều nguồn đa dạng như internet, báo chí, tạp chí khoa học và tài liệu nghiên cứu. IP102 phản ánh sự đa dạng sinh học trong môi trường tự nhiên và nông nghiệp, đồng thời thể hiện đặc trưng phân phối mất cân bằng giữa các lớp – một thách thức phổ biến trong việc huấn luyện các mô hình học sâu.



Hình 2. Một số hình ảnh trong bộ dữ liệu IP102 không chứa đối tượng côn trùng hoặc chứa nhiều vật thể khác gây nhiễu

Tuy nhiên, IP102 vẫn tồn tại một số hạn chế ảnh hưởng đến chất lượng dữ liệu huấn luyện. Cụ thể, trong bộ dữ liệu xuất hiện nhiều hình ảnh không chứa đối tượng côn trùng, hoặc chứa đồng thời nhiều vật thể không liên quan như thực vật, trái cây hoặc

sản phẩm thương mại, gây nhiễu trong quá trình học (Hình 2). Bên cạnh đó, một số loài chỉ có số lượng mẫu ảnh rất hạn chế, làm tăng nguy cơ overfitting và ảnh hưởng đến khả năng tổng quát hóa của mô hình học sâu.

Để khắc phục các vấn đề trên và cung cấp một nền tảng dữ liệu chất lượng cao hơn cho bài toán phân loại côn trùng, nghiên cứu này đã xây dựng và giới thiệu một bộ dữ liệu nâng cấp mang tên PEST204 với số lượng loài phong phú hơn và chất lượng hình ảnh được kiểm soát chặt chẽ.

3.2.2. Tập dữ liệu PEST204

Nhằm khắc phục các hạn chế tồn tại trong bộ dữ liệu IP102 và đồng thời cung cấp một nguồn dữ liệu chất lượng cao hơn cho bài toán phân loại côn trùng, nghiên cứu này đã phát triển bộ dữ liệu PEST204. Đây là một đóng góp quan trọng, giúp cải thiện đáng kể chất lượng và độ đa dạng của dữ liệu huấn luyện cho các mô hình học sâu.

Bộ dữ liệu PEST204 được xây dựng dựa trên IP102 với các cải tiến đáng kể như sau:

- (i) Số lượng loài côn trùng được mở rộng từ 102 lên 204, phản ánh sự phong phú hơn về mặt chủng loại sinh vật;
- (ii) Quá trình lọc dữ liệu được thực hiện kỹ lưỡng để loại bỏ các hình ảnh không chứa đối tượng côn trùng hoặc chứa nhiều vật thể hỗn tạp, nhằm giảm nhiễu và nâng cao tính chính xác cho nhiệm vụ phân loại;
- (iii) Các lớp dữ liệu được cân bằng lại, hạn chế tình trạng mất cân đối số lượng mẫu giữa các loài, từ đó cải thiện khả năng học và tổng quát hóa của mô hình.

(iv) Bộ dữ liệu PEST204 được xây dựng theo ba bước chính. Thứ nhất, các ảnh từ IP102 được sử dụng và loại bỏ các ảnh không chứa côn trùng hoặc có chất lượng kém (mờ, bị che khuất phần lớn đối tượng). Thứ hai, ảnh cho các lớp côn trùng ít mẫu được thu thập, bổ sung từ các nguồn hình ảnh nông nghiệp công khai và tài liệu trực tuyến, sau đó áp dụng cùng tiêu chí lọc nhiễu. Thứ ba, toàn bộ ảnh được rà soát và gán nhãn dựa trên mô tả lớp trong IP102 và các tài liệu nhận dạng côn trùng nông nghiệp, nhằm đảm bảo tính nhất quán giữa các lớp. Nhờ vậy, PEST204 có số lớp phong phú hơn và phân bố mẫu cân bằng hơn so với IP102.



Hình 3. Một số hình ảnh minh họa từ bộ dữ liệu PEST204, thể hiện sự đa dạng về chủng loại và hình thái côn trùng

Một số hình ảnh mẫu trong bộ dữ liệu PEST204 được trình bày trong Hình 3, cho thấy sự đa dạng về hình thái, kích thước và môi trường sống của các loài côn trùng. Điều này giúp mô hình học sâu có thể học được các đặc trưng phong phú và thích ứng tốt hơn với điều kiện thực tế, nơi các biến động môi trường và sự đa dạng sinh học diễn ra mạnh mẽ.

Nhờ những cải tiến nêu trên, PEST204 trở thành một nguồn dữ liệu phù hợp hơn để huấn luyện và đánh giá các mô hình học sâu hiện đại, đồng thời đóng vai trò hỗ trợ phát triển các hệ thống nhận dạng côn trùng thông minh trong lĩnh vực nông nghiệp và bảo vệ thực vật.

3.3. Chiến lược tăng cường dữ liệu

Tăng cường dữ liệu (Data Augmentation) là một kỹ thuật thiết yếu trong học sâu, nhằm cải thiện khả năng tổng quát hóa của mô hình bằng cách nhân rộng bộ dữ liệu huấn luyện thông qua các phép biến đổi hình ảnh hợp lý (Mikołajczyk et al., 2018; Shorten et al., 2019). Đặc biệt trong các bài toán phân loại côn trùng, dữ liệu hình ảnh thu thập từ môi trường tự nhiên thường không phong phú về số lượng và đa dạng về đặc trưng, dẫn đến nguy cơ overfitting khi huấn luyện các mô hình mạng nơ-ron sâu.

Chiến lược tăng cường dữ liệu trong nghiên cứu này được thiết kế nhằm phản ánh các biến thiên thực tế như thay đổi góc nhìn, thay đổi tỷ lệ kích thước, di chuyển vị trí côn trùng trong khung hình và đối xứng hình ảnh. Các kỹ thuật tăng cường cụ thể được áp dụng bao gồm:

- Xoay (Rotation): Xoay ngẫu nhiên hình ảnh trong khoảng $\pm 25^\circ$ để mô phỏng các góc chụp khác nhau.
- Thay đổi tỷ lệ (Scaling): Điều chỉnh tỷ lệ hình ảnh trong phạm vi $\pm 50\%$ nhằm làm phong phú các kích thước đối tượng.
- Dịch chuyển (Translation): Dịch chuyển ngẫu nhiên hình ảnh theo phương ngang và phương dọc, với độ lệch tối đa 10% so với kích thước gốc.
- Lật ngang (Horizontal Flipping): Lật đối xứng hình ảnh theo phương ngang với xác suất 50%.

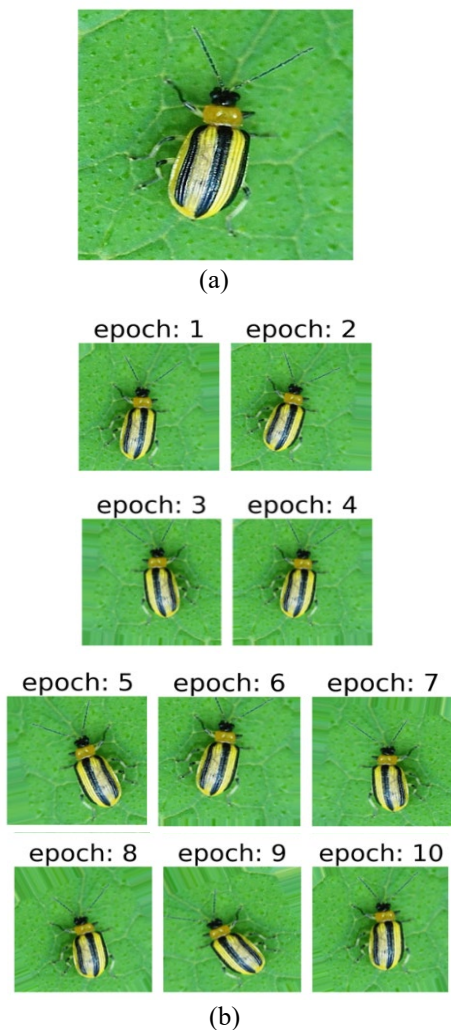
Giả sử x là một mẫu dữ liệu gốc và T là một hàm biến đổi đại diện cho các phép tăng cường dữ liệu, thì mẫu mới x' được biểu diễn dưới dạng: $x' = T(x)$; trong đó T thuộc tập hợp các phép biến đổi như xoay, thay đổi tỷ lệ, dịch chuyển và lật ngang.

Việc áp dụng đồng thời nhiều phép biến đổi trong mỗi epoch huấn luyện giúp mô hình tiếp xúc với các biến thể đa dạng của cùng một đối tượng, qua đó tăng cường khả năng học các đặc trưng bất biến (invariant features) và nâng cao độ bền vững của mô hình trước các biến động trong dữ liệu đầu vào.

Hình 4 minh họa một ví dụ về cách một ảnh côn trùng ban đầu được biến đổi thành nhiều phiên bản khác nhau thông qua các kỹ thuật tăng cường đã áp dụng. Nhờ chiến lược tăng cường dữ liệu hợp lý, mô hình học sâu có thể khai thác tối đa thông tin từ bộ dữ liệu ban đầu, giảm thiểu hiện tượng overfitting và cải thiện hiệu suất phân loại trên cả hai bộ dữ liệu IP102 và PEST204.

3.4. Ứng dụng học chuyển tiếp

Học chuyển tiếp (Transfer Learning) là một phương pháp phổ biến trong học sâu, cho phép khai thác các đặc trưng đã học từ một bài toán nguồn với dữ liệu lớn để giải quyết hiệu quả một bài toán đích có dữ liệu hạn chế (Li et al., 2022). Thay vì huấn luyện mô hình từ đầu, học chuyển tiếp tận dụng các trọng số đã được tối ưu trước trên các tập dữ liệu chuẩn như ImageNet, nơi mô hình đã học được các đặc trưng cơ bản như biên cạnh, kết cấu, hình dạng và màu sắc (Deng et al., 2009).



Hình 4. Minh họa về ảnh tăng cường thông qua kết hợp các phép xoay, thay đổi tỷ lệ, dịch chuyển và lật ngang trong 10 epoch

Ghi chú: (a) Ảnh gốc, (b) Các ảnh đã được tăng cường

Áp dụng học chuyển tiếp mang lại nhiều lợi ích quan trọng trong bài toán phân loại côn trùng:

- **Tiết kiệm thời gian huấn luyện:** Giảm số lượng epoch cần thiết để mô hình hội tụ.
- **Cải thiện độ chính xác tổng thể:** Các đặc trưng cơ bản đã học từ ImageNet hỗ trợ mô hình nhanh chóng thích ứng với đặc trưng hình ảnh của côn trùng.
- **Giảm yêu cầu dữ liệu huấn luyện lớn:** Đặc biệt hữu ích khi bộ dữ liệu côn trùng như IP102 hoặc PEST204 có sự phân bố không đồng đều giữa các lớp.

- **Hạn chế overfitting:** Các trọng số pre-trained đóng vai trò như một hình thức regularization ban đầu.

Trong nghiên cứu này, toàn bộ các mô hình YOLOv11 được khởi tạo bằng trọng số đã được huấn luyện trước trên tập dữ liệu ImageNet. Sau đó, các mô hình được tinh chỉnh (fine-tuning) trên hai tập dữ liệu IP102 và PEST204. Quá trình fine-tuning được thực hiện với tốc độ học (learning rate) phù hợp nhằm điều chỉnh các trọng số một cách mềm mại, vừa tận dụng các đặc trưng tổng quát đã học, vừa thích nghi với đặc điểm cụ thể của bài toán phân loại côn trùng.

Chiến lược fine-tuning trong nghiên cứu này chỉ cập nhật trọng số ở các tầng cao (các lớp fully connected hoặc convolution cuối cùng), trong khi các tầng thấp (early layers), vốn đã học được các đặc trưng cơ bản hữu ích, được giữ nguyên hoặc cập nhật rất chậm. Điều này giúp duy trì khả năng tổng quát của mô hình, đồng thời tập trung vào việc phân biệt các đặc trưng tinh tế giữa các loài côn trùng.

Nhờ ứng dụng hiệu quả kỹ thuật học chuyển tiếp, các mô hình YOLOv11 trong nghiên cứu đã đạt được kết quả huấn luyện ổn định, giảm thiểu hiện tượng overfitting, đồng thời cải thiện rõ rệt độ chính xác phân loại so với việc huấn luyện từ đầu.

4. KẾT QUẢ THỰC NGHIỆM

4.1. Thiết lập thực nghiệm

Các thực nghiệm được tiến hành nhằm đánh giá hiệu suất phân loại của các biến thể YOLOv11 trên hai tập dữ liệu IP102 và PEST204. Toàn bộ quá trình huấn luyện và đánh giá được thực hiện trong môi trường phần cứng và phần mềm như sau:

- Bộ xử lý (CPU): Intel Core i7-13700K, xung nhịp 2,4 GHz
- Bộ nhớ (RAM): 32 GB
- Bộ xử lý đồ họa (GPU): NVIDIA RTX 4070 Ti Super, 16 GB VRAM
- Hệ điều hành: Ubuntu 22.04
- Framework học sâu: PyTorch 2.6

Bộ dữ liệu được chia thành ba tập: tập huấn luyện (training set) chiếm 60%, tập xác thực (validation set) 30%, và tập kiểm tra (testing set) 10%. Việc phân chia được thực hiện ngẫu nhiên, đồng thời bảo đảm phân phối mẫu giữa các lớp được cân bằng nhằm duy trì tính đa dạng và khả năng đại diện của dữ liệu trong mỗi tập con.

Trong quá trình huấn luyện, các tham số chính được cấu hình như sau:

- Thuật toán tối ưu hóa: Stochastic Gradient Descent (SGD)
- Tốc độ học ban đầu (initial learning rate): 0,01
- Hệ số momentum: 0,9
- Hệ số suy giảm trọng số (weight decay): 0,0005
- Kích thước batch: 16 (là giá trị tối đa phù hợp với giới hạn bộ nhớ GPU)
- Số lượng epoch: 40

Việc lựa chọn các tham số trên dựa trên các nguyên tắc thực nghiệm tổng quát và kinh nghiệm phổ biến trong huấn luyện các mô hình học sâu hiện đại. Thuật toán SGD được lựa chọn nhờ khả năng tối ưu hóa ổn định và hiệu quả cho các bài toán phân loại quy mô lớn. Tốc độ học ban đầu 0,01, kết hợp với hệ số momentum 0,9, giúp tăng tốc độ hội tụ trong giai đoạn đầu huấn luyện mà vẫn duy trì tính ổn định. Hệ số weight decay 0,0005 đóng vai trò như một cơ chế regularization, góp phần hạn chế hiện tượng overfitting. Kích thước batch 16 được chọn là mức tối ưu trong giới hạn tài nguyên GPU, cân bằng giữa tốc độ huấn luyện và độ ổn định gradient. Số lượng epoch 40 được xác định dựa trên các thực nghiệm tiền kiểm, đảm bảo mô hình có đủ thời gian học hiệu quả mà không gây ra hiện tượng học quá mức (overfitting).

Các mô hình YOLOv11 được khởi tạo với trọng số đã huấn luyện trước trên tập dữ liệu ImageNet, sau đó được tinh chỉnh (fine-tuning) trên hai bộ dữ liệu IP102 và PEST204. Trong quá trình huấn luyện, chiến lược tăng cường dữ liệu được áp dụng đồng bộ để làm phong phú dữ liệu huấn luyện, tăng cường khả năng tổng quát hóa và cải thiện hiệu suất mô hình trên dữ liệu.

4.2. Độ đo đánh giá

Để đánh giá hiệu suất phân loại của các mô hình, nghiên cứu này sử dụng các độ đo tiêu chuẩn phổ biến trong học sâu, bao gồm:

- **Độ chính xác Top-1 (Top-1 Accuracy):** Tỷ lệ phần trăm số lượng mẫu mà dự đoán chính xác nhất (top-1) trùng khớp với nhãn thực tế.
- **Độ chính xác Top-5 (Top-5 Accuracy):** Tỷ lệ phần trăm số lượng mẫu mà nhãn thực tế nằm trong 05 dự đoán có xác suất cao nhất của mô hình.
- **Thời gian suy luận (Inference Time):** Thời

gian trung bình tính bằng mili-giây (ms) để mô hình đưa ra một dự đoán cho một mẫu đầu vào.

- **Số lượng tham số (Parameters):** Tổng số tham số học được của mô hình, biểu thị mức độ phức tạp của mạng nơ-ron.
- **FLOPs (Floating Point Operations):** Tổng số phép toán “floating point” cần thiết để xử lý một mẫu đầu vào, đại diện cho chi phí tính toán của mô hình.

Độ chính xác Top-1 và Top-5 được tính theo công thức sau:

$$\text{Top-1 Accuracy} = \frac{\text{Số lượng dự đoán chính xác}}{\text{Tổng số mẫu}} \times 100\%$$

$$\text{Top-5 Accuracy} = \frac{\text{Số lượng có nhãn đúng nằm trong 5 dự đoán hàng đầu}}{\text{Tổng số mẫu}} \times 100\%$$

Trong đó:

- Một dự đoán được coi là chính xác đối với Top-1 nếu nhãn được dự đoán cao nhất của mô hình khớp với nhãn thực tế.
- Đối với Top-5, dự đoán được coi là đúng nếu nhãn thực tế xuất hiện trong 05 nhãn dự đoán có xác suất cao nhất.

Ngoài ra, thời gian suy luận (Inference Time) và FLOPs được sử dụng để đánh giá hiệu quả tính toán và khả năng triển khai thực tế của mô hình, đặc biệt trong các ứng dụng yêu cầu tốc độ xử lý cao hoặc triển khai trên thiết bị có tài nguyên hạn chế.

Số lượng tham số học được cũng là một chỉ số quan trọng, phản ánh mức độ phức tạp và yêu cầu lưu trữ của mô hình. Các mô hình có số lượng tham số lớn thường có khả năng học mạnh hơn nhưng đồng thời tiêu tốn nhiều tài nguyên hơn trong quá trình huấn luyện và suy luận.

Việc kết hợp các độ đo về hiệu suất (Top-1, Top-5) và chi phí tính toán (Inference Time, Parameters, FLOPs) cho phép đánh giá toàn diện khả năng cân bằng giữa độ chính xác và hiệu quả của các mô hình.

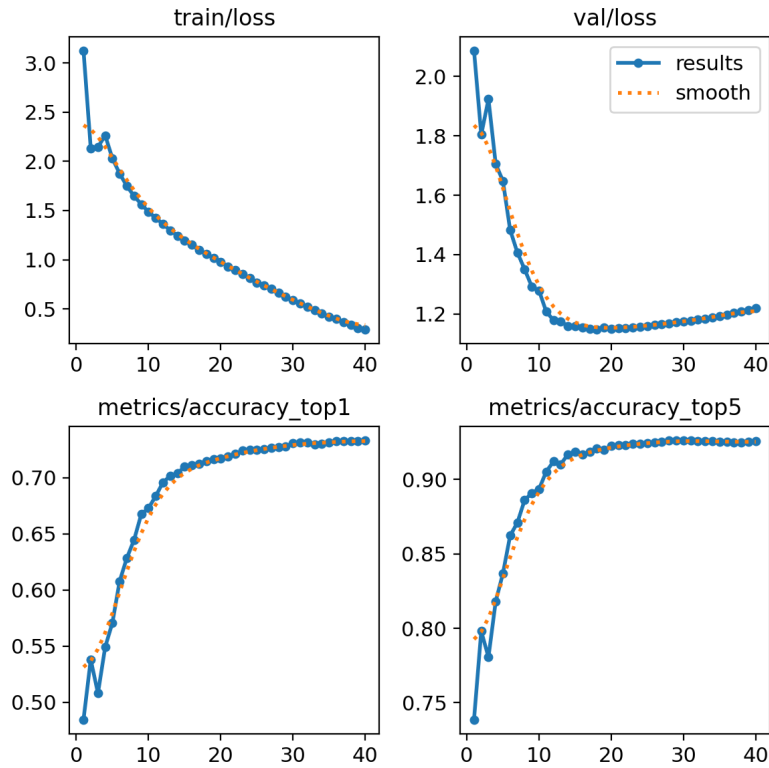
4.3. Kết quả trên tập dữ liệu IP10

Kết quả phân loại chéo của các mô hình YOLOv11 trên tập dữ liệu IP102 được tổng hợp trong Bảng 1. Các độ đo bao gồm độ chính xác Top-1, độ chính xác Top-5, thời gian tiền xử lý (preprocessing time), thời gian suy luận (inference time), mức sử dụng GPU và số lượng phép floating point (GFLOPs).

Bảng 1. Kết quả phân loại trên tập dữ liệu IP102 của các mô hình YOLO

Mô hình	Top-1 Accuracy	Top-5 Accuracy	Preprocess Time (ms)	Inference Time (ms)	GPU Usage	GFLOPs
YOLO11n-cls	71,5%	91,4%	0,8	0,3	1,7 GB	3,3
YOLOv8s-cls*	50,0%	75,7%	0,7	0,6	2,43 GB	12,6
YOLOv8s-cls	72,4%	91,6%	0,8	0,6	2,43 GB	12,6
YOLO11s-cls*	53,6%	76,8%	0,8	0,6	2,7 GB	12,1
YOLO11s-cls	73,3%	92,6%	0,7	0,6	2,7 GB	12,1
YOLO11m-cls	73,9%	92,9%	0,7	1,6	5,3 GB	39,4
YOLO11l-cls	74,7%	93,3%	0,5	2,0	6,7 GB	49,4
YOLO11x-cls	74,9%	93,4%	0,4	3,7	10,1 GB	110,4

(*): Các mô hình được huấn luyện không sử dụng học chuyển tiếp và tăng cường dữ liệu.



Hình 5. Biểu đồ hàm mất mát (loss) và độ chính xác (accuracy) của mô hình YOLO11s-cls trên tập dữ liệu IP102

Có thể nhận thấy rằng các biến thể YOLO11 đều đạt hiệu suất phân loại cao trên tập IP102. Trong đó, **YOLO11x-cls** đạt độ chính xác Top-1 cao nhất (74,9%) và Top-5 (93,4%), đồng thời có thời gian tiền xử lý ngắn nhất (0,4 ms). Tuy nhiên, mô hình này yêu cầu tài nguyên tính toán lớn nhất với mức sử dụng GPU lên đến 10,1 GB và số lượng phép toán dấu chấm động (GFLOPs) là 110,4.

Ở chiều ngược lại, YOLO11n-cls thể hiện ưu thế về hiệu quả tính toán với thời gian suy luận chỉ 0,3 ms và mức tiêu thụ GPU thấp nhất (1,7 GB), trong khi vẫn duy trì độ chính xác Top-1 ở mức 71,5%. Điều này cho thấy YOLO11n-cls có tiềm năng ứng dụng trong các hệ thống yêu cầu tốc độ cao hoặc triển khai trên các thiết bị có tài nguyên hạn chế.

Bảng 2. Kết quả phân loại trên tập dữ liệu PEST204 của các mô hình YOLO

Mô hình	Top-1 Accuracy	Top-5 Accuracy	Preprocess Time (ms)	Inference Time (ms)	GPU Usage	GFLOPs
YOLO11n-cls	77,1%	92,8%	0,5	0,4	1,46 GB	6,8
YOLO11s-cls*	63,1%	83,8%	0,7	0,6	5,13 GB	12,2
YOLO11s-cls	78,9%	93,8%	0,8	0,6	3,2 GB	12,2
YOLO11m-cls	79,8%	94,0%	0,7	1,6	5,7 GB	39,5
YOLO11l-cls	80,2%	94,2%	0,7	2,0	7,2 GB	49,5
YOLO11x-cls	81,1%	94,6%	0,4	3,8	10,8 GB	110,5

(*): Các mô hình được huấn luyện không sử dụng học chuyển tiếp và tăng cường dữ liệu.

So sánh với các mô hình YOLOv8s-cls, có thể thấy các biến thể YOLO11 đều vượt trội hơn về độ chính xác trong cùng điều kiện huấn luyện. Cụ thể, YOLO11s-cls đạt độ chính xác Top-1 là 73,3%, cao hơn mức 72,4% của YOLOv8s-cls. Đặc biệt, khi không sử dụng kỹ thuật học chuyển tiếp và tăng cường dữ liệu, độ chính xác Top-1 của YOLO11s-cls chỉ đạt 53,6%, thấp hơn đáng kể so với khi áp dụng đầy đủ các kỹ thuật hỗ trợ, cho thấy vai trò quan trọng của học chuyển tiếp và tăng cường dữ liệu trong quá trình huấn luyện.

Ngoài ra, thời gian suy luận của các mô hình duy trì ở mức thấp, dao động từ 0,3 ms đến 3,7 ms, phù hợp với yêu cầu của các ứng dụng thời gian thực.

Bên cạnh các kết quả định lượng, quá trình huấn luyện và đánh giá của mô hình YOLO11s-cls trên tập IP102 được minh họa thông qua Hình 5. Các biểu đồ thể hiện diễn tiến của hàm mất mát (loss) và độ chính xác (accuracy) theo số lượng epoch. Cụ thể, loss trên cả tập huấn luyện và tập kiểm định đều giảm ổn định theo thời gian, cho thấy mô hình học hiệu quả mà không xảy ra hiện tượng overfitting đáng kể. Độ chính xác Top-1 và Top-5 tăng dần và đạt trạng thái gần bão hòa sau khoảng 30 epoch. Điều này cho thấy mô hình đã học được các đặc trưng quan trọng từ dữ liệu và đạt được độ chính xác cao ổn định.

4.4. Kết quả trên tập dữ liệu PEST204

Kết quả thực nghiệm trên tập dữ liệu PEST204 được tổng hợp trong Bảng 2 với các độ đo tương tự như ở tập IP102, bao gồm độ chính xác Top-1, Top-5, thời gian tiền xử lý, thời gian suy luận, mức sử dụng GPU và GFLOPs.

So với tập IP102, hiệu suất phân loại trên tập PEST204 có sự cải thiện rõ rệt. Điều này cho thấy bộ dữ liệu PEST204 với sự sáng lọc và mở rộng hợp lý đã hỗ trợ mô hình học sâu khai thác thông tin đặc trưng tốt hơn.

Mô hình YOLO11x-cls tiếp tục đạt hiệu suất cao nhất với độ chính xác Top-1 là 81,1% và Top-5 là 94,6%, vượt trội so với các biến thể khác. Tuy nhiên, tương tự như trên IP102, mô hình này yêu cầu tài nguyên tính toán cao với mức sử dụng GPU hơn 10,8 GB và 110,5 GFLOPs.

Mô hình YOLO11n-cls, với mức sử dụng GPU thấp nhất (1,46 GB) và thời gian suy luận nhanh (0,4 ms), đạt độ chính xác Top-1 là 77,1%. Điều này cho thấy YOLO11n-cls vẫn là lựa chọn lý tưởng cho các hệ thống hạn chế tài nguyên nhưng yêu cầu độ chính xác tương đối cao.

Đáng chú ý, sự cải thiện hiệu suất của YOLO11s-cls sau khi áp dụng học chuyển tiếp và tăng cường dữ liệu là rất rõ rệt: từ 63,1% Top-1 (khi huấn luyện từ đầu) tăng lên 78,9% (khi áp dụng kỹ thuật hỗ trợ). Điều này càng khẳng định tầm quan trọng của các chiến lược tăng cường dữ liệu và học chuyển tiếp trong bài toán phân loại côn trùng trên tập dữ liệu thực tế.

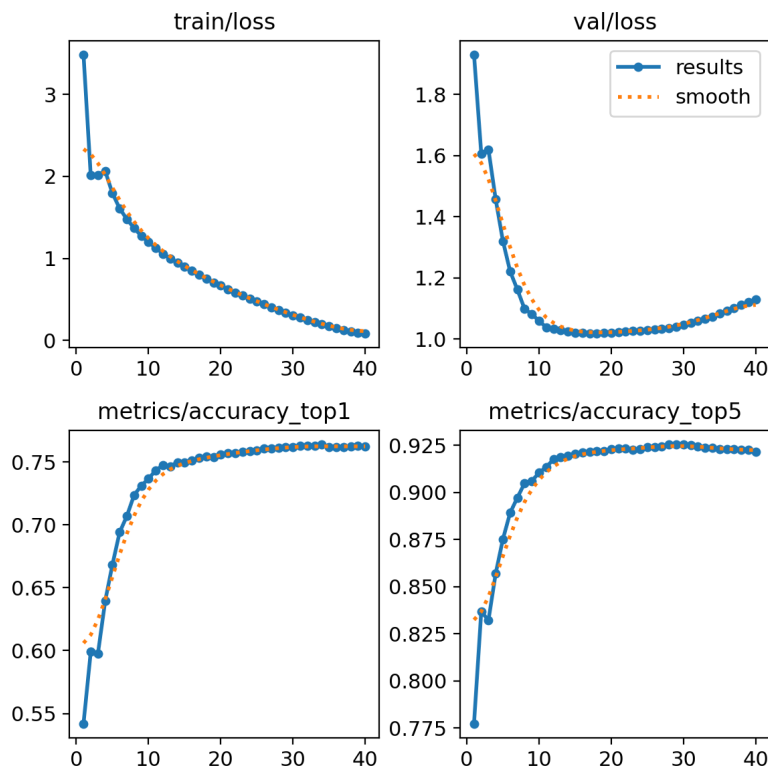
Nhìn chung, tất cả các mô hình đều đạt thời gian suy luận dưới 4 ms, cho thấy tiềm năng ứng dụng trong các hệ thống nhận dạng thời gian thực hoặc thiết bị di động.

Bên cạnh các kết quả định lượng, quá trình huấn luyện và kiểm định của mô hình YOLO11s-cls trên tập dữ liệu PEST204 được minh họa trong Hình 6. Các đồ thị mô tả sự thay đổi của hàm mất mát (loss) và độ chính xác (accuracy) theo từng epoch.

Quan sát Hình 6 cho thấy loss trên cả tập huấn luyện và kiểm định đều giảm mạnh trong những epoch đầu tiên và đạt mức ổn định sau khoảng 20 epoch. Độ chính xác Top-1 và Top-5 đều tăng liên tục trong suốt quá trình huấn luyện, tiến gần đến trạng thái bão hòa từ epoch thứ 30 trở đi. Các đường cong loss và accuracy giữa tập huấn luyện và kiểm định khá sát nhau, chứng tỏ mô hình học ổn định và không xảy ra hiện tượng overfitting.

Những kết quả này cũng cố thêm đánh giá rằng tập dữ liệu PEST204, với chất lượng và sự đa dạng

đã được cải thiện, giúp mô hình học sâu đạt hiệu suất cao và ổn định hơn so với tập IP102.



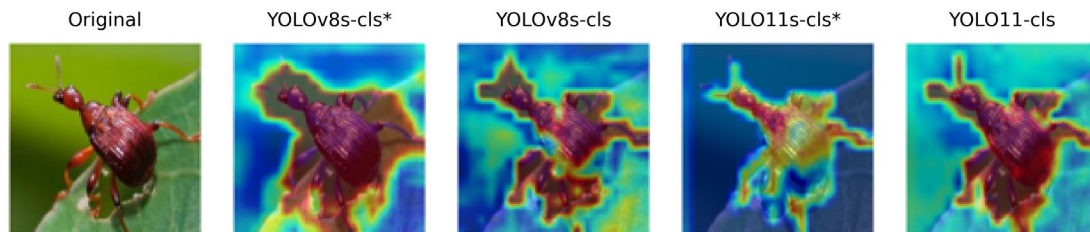
Hình 6. Biểu đồ hàm mất mát (loss) và độ chính xác (accuracy) của mô hình YOLO11s-cls trên tập dữ liệu PEST204

4.5. Phân tích trực quan bằng Eigen-CAM

Để hiểu rõ hơn về cơ chế hoạt động và trọng tâm chú ý của các mô hình phân loại côn trùng, nghiên cứu này áp dụng kỹ thuật trực quan hóa Eigen-CAM. Eigen-CAM là một phương pháp không yêu cầu sửa đổi kiến trúc mạng, cho phép trích xuất bản đồ nhiệt (heatmap) từ các lớp đặc trưng cuối cùng

nhằm xác định các khu vực quan trọng ảnh hưởng đến quyết định dự đoán của mô hình.

Không giống như các kỹ thuật trực quan hóa trước đây, Eigen-CAM khai thác các giá trị riêng (eigenvalues) để tạo ra bản đồ có khả năng tổng quát tốt hơn và phản ánh trực tiếp các đặc trưng nổi bật trong ảnh đầu vào.



Hình 7. Ví dụ trực quan hóa bằng Eigen-CAM: mô hình YOLO11s-cls tập trung vào các đặc trưng hình thái chính của côn trùng để đưa ra dự đoán chính xác

Trong nghiên cứu này, Eigen-CAM được sử dụng để phân tích các mô hình YOLO11-cls sau khi huấn luyện trên tập dữ liệu IP102. Hình 7 minh họa

một số kết quả phân tích trực quan trên ảnh mẫu. Các bản đồ nhiệt cho thấy vùng chú ý của các mô hình tập trung rõ nét vào đối tượng côn trùng chính,

trong khi các khu vực nền ít liên quan được giảm mức độ chú ý.

Đặc biệt, các mô hình YOLO11 cho thấy khả năng tập trung tốt hơn vào các chi tiết hình thái đặc trưng của côn trùng (như cánh, chân, màu sắc cơ thể), giúp phân biệt chính xác giữa các loài có hình dạng tương tự. Điều này phản ánh hiệu quả học sâu của các mô hình không chỉ ở cấp độ phân loại toàn ảnh mà còn ở cấp độ chi tiết cục bộ.

Kết quả phân tích bằng Eigen-CAM góp phần xác nhận rằng các mô hình YOLO11-cla không chỉ đạt hiệu suất phân loại cao về mặt định lượng mà còn có khả năng học được các đặc trưng hình ảnh phù hợp, hợp lý với bản chất thị giác của bài toán phân loại côn trùng.

5. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

Trong nghiên cứu này, một hướng tiếp cận phân loại côn trùng sử dụng các biến thể của mô hình học sâu YOLO11 được đề xuất và đánh giá. Bằng việc khai thác các kỹ thuật học chuyển tiếp và tăng cường dữ liệu hợp lý, các mô hình YOLO11 đã đạt được hiệu suất phân loại cao trên cả hai tập dữ liệu IP102 và PEST204.

Đồng thời, nghiên cứu đã đóng góp một tập dữ liệu mới có tên PEST204, được xây dựng dựa trên việc sàng lọc và mở rộng tập IP102 nhằm nâng cao chất lượng dữ liệu huấn luyện, giảm thiểu nhiễu và cải thiện độ đa dạng hình ảnh. Phân tích trực quan bằng Eigen-CAM cho thấy các mô hình không chỉ học dựa trên các đặc trưng ngẫu nhiên mà còn tập trung vào các chi tiết hình thái đặc trưng của côn trùng, từ đó tăng cường độ tin cậy trong việc phân loại.

TÀI LIỆU THAM KHẢO (REFERENCES)

- Ahmed, S., Hasan, M. B., Ahmed, T., Sony, M. R. K., & Kabir, M. H. (2022). Less is more: Lighter and faster deep neural architecture for tomato leaf disease classification. *IEEE Access*, 10, 68868-68884. <https://doi.org/10.1109/ACCESS.2022.3187203>
- Al-Hiary, H., Bani-Ahmad, S., Reyat, M., Braik, M., & Alrahmaneh, Z. (2011). Fast and accurate detection and classification of plant diseases. *International journal of computer applications*, 17(1), 31-38. <https://doi.org/10.5120/2183-2754>
- Almryad, A. S., & Kutucu, H. (2020). Automatic identification for field butterflies by convolutional neural networks. *Engineering*

Kết quả thực nghiệm cho thấy các biến thể YOLO11-cla đã đạt độ chính xác Top-1 lên tới 81,1% trên tập dữ liệu PEST204, cao hơn so với các mô hình thể hệ trước. Mô hình YOLO11n-cla, dù có quy mô nhỏ và yêu cầu tài nguyên thấp, vẫn duy trì được độ chính xác phân loại tốt, phù hợp với các ứng dụng thực tế yêu cầu xử lý thời gian thực trên thiết bị di động hoặc các hệ thống nhúng. Bên cạnh đó, tập dữ liệu PEST204 cũng chứng minh được hiệu quả trong việc cải thiện khả năng tổng quát hóa và hiệu suất của mô hình học sâu.

Trong tương lai, nghiên cứu sẽ tập trung vào việc phát triển các kiến trúc mạng nơ-ron nhẹ hơn nhưng hiệu quả hơn, nhằm tối ưu hóa khả năng triển khai trên các nền tảng thiết bị di động và IoT. Ngoài ra, kỹ thuật học liên miền (domain adaptation) sẽ được nghiên cứu để tăng khả năng tổng quát hóa của mô hình đối với dữ liệu đến từ các môi trường thực tế đa dạng. Việc mở rộng tập dữ liệu PEST204 với các hình ảnh côn trùng trong điều kiện ánh sáng yếu, nền phức tạp hoặc các góc chụp khó khăn cũng sẽ được xem xét nhằm nâng cao độ bền vững và tính ứng dụng thực tiễn. Một điểm hạn chế khác là nghiên cứu hiện tại mới tập trung so sánh các biến thể YOLOv11-cla với YOLOv8s-cla, chưa bao quát các dòng mô hình phân loại ảnh hiện đại khác như MobileNet, EfficientNet, ShuffleNet hay các kiến trúc transformer như ViT. Việc mở rộng so sánh thực nghiệm với các mô hình này giúp đánh giá sâu hơn điểm mạnh và hạn chế của YOLOv11-cla trong bối cảnh triển khai trên thiết bị biên và hệ thống nông nghiệp thông minh, và sẽ được dành cho các nghiên cứu tiếp theo. Cuối cùng, tích hợp các mô-đun phát hiện bất thường hoặc nhận dạng các loài côn trùng mới là một hướng đi tiềm năng để xây dựng hệ thống nhận diện côn trùng thông minh và linh hoạt hơn trong tương lai.

Science and Technology, an International Journal, 23(1), 189-195.

<https://doi.org/10.1016/j.jestech.2020.01.006>

- Chithambarathanu, M., & Jeyakumar, M. K. (2023). Survey on crop pest detection using deep learning and machine learning approaches. *Multimedia Tools and Applications*, 82(27), 42277-42310. <https://doi.org/10.1007/s11042-023-15221-3>

- Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009). Imagenet: A largescale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition* (pp. 248-255). <https://doi.org/10.1109/CVPR.2009.5206848>

- Doan, T. N. (2022). An efficient system for real-time mobile smart device-based insect detection. *International Journal of Advanced Computer Science and Applications*, 13(6).
<https://doi.org/10.14569/IJACSA.2022.0130605>
- Doan, T. N. (2023). Large-scale insect pest image classification. *Journal of Advances in Information Technology*, 14(2), 328-341.
<https://doi.org/10.12720/jait.14.2.328-341>
- Estruch, J. J., Carozzi, N. B., Desai, N., Duck, N. B., Warren, G. W., & Koziel, M.G. (1997). Transgenic plants: an emerging approach to pest control. *Nature biotechnology*, 15(2), 137-141.
<https://doi.org/10.1038/nbt0297-137>
- Hansen, O.L., Svenning, J.C., Olsen, K., Dupont, S., Garner, B.H., Iosifidis, A., Price, B.W. and Høye, T.T.. (2020). Species-level image classification with convolutional neural network enables insect identification from habitus images. *Ecology and Evolution*, 10(2), 737-747.
<https://doi.org/10.1002/ece3.5921>
- Kandel, I., & Castelli, M. (2020). The effect of batch size on the generalizability of the convolutional neural networks on a histopathology dataset. *ICT express*, 6(4), 312-315.
<https://doi.org/10.1016/j.ict.2020.04.010>
- Kirkeby, C., Rydhmer, K., Cook, S.M., Strand, A., Torrance, M.T., Swain, J.L., Prangma, J., Johnen, A., Jensen, M., Brydegaard, M. and Græsbøll, K. (2021). Advances in automatic identification of flying insects using optical sensors and machine learning. *Scientific reports*, 11(1), 1555.
<https://doi.org/10.1038/s41598-021-81005-0>
- Krause, J., Stark, M., Deng, J., & Fei-Fei, L. (2013). 3D object representations for finegrained categorization. In *Proceedings of the IEEE International Conference on Computer Vision Workshops* (pp. 554-561).
<https://doi.org/10.1109/ICCVW.2013.77>
- Li, C., Zhen, T., & Li, Z. (2022). Image classification of pests with residual neural network based on transfer learning. *Applied Sciences*, 12(9), 4356.
<https://doi.org/10.3390/app12094356>
- Li, Y., Wang, H., Dang, L. M., Sadeghi-Niaraki, A., & Moon, H. (2020). Crop pest recognition in natural scenes using convolutional neural networks. *Computers and Electronics in Agriculture*, 169, 105174.
<https://doi.org/10.1016/j.compag.2019.105174>
- Mikołajczyk, A., & Grochowski, M. (2018). Data augmentation for improving deep learning in image classification problem. In *2018 international interdisciplinary PhD workshop* (pp. 117-122). IEEE.
<https://doi.org/10.1109/IIPHDW.2018.8388338>
- Nanni, L., Manfè, A., Maguolo, G., Lumini, A., & Brahnam, S. (2022). High performing ensemble of convolutional neural networks for insect pest image detection. *Ecological Informatics*, 67, 101515.
<https://doi.org/10.1016/j.ecoinf.2021.101515>
- Oberti, R., Marchi, M., Tirelli, P., Calcante, A., Iriti, M., Tona, E., & Ulbrich, H. (2016). Selective spraying of grapevines for disease control using a modular agricultural robot. *Biosystems engineering*, 146, 203-215.
<https://doi.org/10.1016/j.biosystemseng.2015.12.004>
- Rustia, D. J. A., Chao, J. J., Chiu, L. Y., Wu, Y. F., Chung, J. Y., Hsu, J. C., & Lin, T. T. (2021). Automatic greenhouse insect pest detection and recognition based on a cascaded deep learning classification method. *Journal of Applied Entomology*, 145(3), 206-222.
<https://doi.org/10.1111/jen.12834>
- Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of big data*, 6(1), 1-48.
<https://doi.org/10.1186/s40537-019-0197-0>
- Wang, J., Lin, C., Ji, L., & Liang, A. (2012). A new automatic identification system of insect images at the order level. *Knowledge-Based Systems*, 33, 102-110.
<https://doi.org/10.1016/j.knosys.2012.03.014>
- Wang, R., Zhang, J., Dong, W., Yu, J., Xie, C., Li, R., & Chen, H. (2017). A crop pests image classification algorithm based on deep convolutional neural network. *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, 15(3), 1239-1246.
<https://doi.org/10.12928/telkomnika.v15i3.5382>
- Wu, X., Zhan, C., Lai, Y. K., Cheng, M. M., & Yang, J. (2019). IP102: A large-scale benchmark dataset for insect pest recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 8787-8796).
<https://doi.org/10.1109/CVPR.2019.00899>
- Xia, D., Chen, P., Wang, B., Zhang, J., & Xie, C. (2018). Insect detection and classification based on an improved convolutional neural network. *Sensors*, 18(12), 4169.
<https://doi.org/10.3390/s18124169>