

## Model Adaptive Learning: An Approach for Improving Object Recognition Efficiency

Diem Phuc Tran 

Duy Tan University, Vietnam

Corresponding author. Email: [phuctd@gmail.com](mailto:phuctd@gmail.com)

### ARTICLE INFO

Received: 11/03/2024  
Revised: 25/03/2024  
Accepted: 27/03/2024  
Published: 28/04/2024

### KEYWORDS

Deep Learning;  
Adaptive Learning;  
Object Detection;  
Auto Vehicle;  
Auto Robot.

### ABSTRACT

In recent years, research on artificial intelligence (AI) has experienced remarkable advancements. Many practical applications are the result of ongoing research, such as recognition technology, self-driving cars, translation, etc and most recently, the explosion of some AI-powered intelligent chatbots using large language models. Major corporations and research institutions worldwide are racing to develop AI models capable of the most accurate interaction based on user requests. However, despite achieving certain milestones, current AI models still fall short of the required intelligence to function similarly to the human brain. Based on research and experimentation, we propose an adaptive learning model that allows models to continuously learn during operation, select, and store previously acquired experiential knowledge to serve on-demand tasks. The proposed solution consists of four steps: (1) Initializing the initial recognition model; (2) Detection, recognition and collecting data from various instances of objects during operation based on object tracking; (3) Searching, selecting optimal models and hyperparameters on the discovered dataset; (4) Training and updating the model. The results of the proposed research could be a promising direction for the development of an adaptive learning model in advanced object recognition.

## Mô Hình Học Thích Ứng: Một Hướng Tiếp Cận Nhằm Nâng Cao Chất Lượng Nhận Dạng Đối Tượng

Trần Diễm Phúc 

Trường Đại học Duy Tân, Việt Nam

Tác giả liên hệ. Email: [phuctd@gmail.com](mailto:phuctd@gmail.com)

### THÔNG TIN BÀI BÁO

Ngày nhận bài: 11/03/2024  
Ngày hoàn thiện: 25/03/2024  
Ngày chấp nhận đăng: 27/03/2024  
Ngày đăng: 28/04/2024

### TỪ KHÓA

Học sâu;  
Học thích ứng;  
Nhận dạng đối tượng;  
Xe tự lái;  
Robot.

### TÓM TẮT

Trong thời gian gần đây, lĩnh vực nghiên cứu về Trí tuệ Nhân tạo (AI) đã đạt được những tiến bộ đáng kể. Có nhiều ứng dụng thực tế đã xuất hiện nhờ vào những thành tựu mới, như: công nghệ nhận dạng, xe tự lái, dịch thuật,... và đặc biệt là sự bùng nổ của các Chatbot sử dụng mô hình ngôn ngữ lớn. Các tập đoàn và viện nghiên cứu hàng đầu trên thế giới đang chạy đua phát triển các mô hình AI có khả năng tương tác chính xác theo yêu cầu. Tuy nhiên, mặc dù đã đạt được một số thành tựu nhưng đến nay, các mô hình AI vẫn chưa đạt được sự thông minh tương đương với bộ não con người. Dựa trên các nghiên cứu và thực nghiệm, chúng tôi đề xuất một phương pháp có khả năng thích ứng, giúp mô hình liên tục học trong quá trình vận hành. Mô hình này sẽ lựa chọn và lưu trữ tri thức từ những trải nghiệm thu thập trước đó, nhằm phục vụ cho các nhiệm vụ theo yêu cầu. Giải pháp đề xuất gồm 4 bước: (1) Khởi tạo mô hình nhận dạng ban đầu; (2) Xác định và thu nhận dữ liệu từ các trường hợp khác nhau của đối tượng trong quá trình học; (3) Tìm kiếm, lựa chọn các mô hình, các siêu tham số (hyperparameters) tối ưu trên tập dữ liệu vừa tìm được; (4) Huấn luyện và cập nhật lại mô hình. Kết quả của nghiên cứu đề xuất có thể là một hướng phát triển mới của một mô hình học thích ứng trong nhận dạng đối tượng nâng cao.

Doi: <https://doi.org/10.54644/jte.2024.1540>

Copyright © JTE. This is an open access article distributed under the terms and conditions of the [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/) which permits unrestricted use, distribution, and reproduction in any medium for non-commercial purpose, provided the original work is properly cited.

## 1. Giới thiệu

Công nghệ ngày càng phát triển và cùng với đó mỗi ngày chúng ta chứng kiến sự hình thành, phát triển của các phát kiến mới trong nhiều lĩnh vực, trong đó có trí tuệ nhân tạo. Những phát kiến này không chỉ dừng lại trong lĩnh vực nghiên cứu, mà còn được ứng dụng sâu rộng vào đời sống xã hội thông qua các sản phẩm mới. Trong lĩnh vực robotic và tự động hóa, nhiều nghiên cứu đã được tiến hành với trọng tâm là về robot, xe tự lái và các hệ thống hỗ trợ thông minh. Một trong những hướng nghiên cứu đầy hứa hẹn của các nhà khoa học là xây dựng mô hình thần kinh nhân tạo có khả năng thông minh, mô phỏng hoạt động của bộ não con người. Vì vậy, mạng thần kinh nhân chập (Convolutional Neural Network – CNN) đã và đang nhận được sự quan tâm lớn với khả năng học linh hoạt, tự động hóa và khả năng tự thích ứng theo môi trường cụ thể. Các mô hình mạng học sâu được đề xuất có thể kể đến như AlexNet [15], GoogleNet [26], ResNet [11], R-CNN [10], Fast R-CNN [9], Faster R-CNN [24],... là các mô hình được đề xuất với cấu trúc mạng lớn, lượng dữ liệu phục vụ huấn luyện khổng lồ và có khả năng nhận dạng được hầu hết tất cả các đối tượng hiện có với độ chính xác cao. Trong thời gian gần đây, các viện nghiên cứu lớn đã công bố nhiều mô hình AI có khả năng tổng hợp cao, đưa ra các công cụ hỗ trợ người dùng trên nhiều lĩnh vực như Chatbot [1], kỹ thuật/ hội họa [20], xây dựng video [25],... Các mô hình AI có thể tự khởi tạo nội dung theo yêu cầu đầu vào của người dùng với chất lượng tương đương các sản phẩm “hand make” chuyên nghiệp, như: Sona (của OpenAI), Imagine (của Meta),...

Tuy nhiên, mặc dù đạt được một bước tiến dài trong nghiên cứu trí tuệ nhân tạo, các mô hình AI vẫn chưa vượt qua được vách ngăn lớn và vẫn đề khó khăn nhất đặt ra cho các nhà nghiên cứu: khả năng tự học, khả năng tự thông minh. Đa số các mô hình AI hiện nay được đề xuất phần lớn phụ thuộc vào dữ liệu đầu vào, nếu dữ liệu đầu vào lớn thì khả năng học, phân tích, tổng hợp và đưa ra kết quả sẽ càng chính xác, sát với yêu cầu thực tế. Nhiều mô hình AI đang được huấn luyện từ các nguồn dữ liệu khổng lồ, đặc biệt là với sự hỗ trợ về phần cứng với các GPU được cải tiến về hiệu suất. Vì vậy, hầu hết các mô hình AI hiện tại đều ít nhiều phải có sự can thiệp của người dùng trong quá trình chuẩn bị, gán nhãn dữ liệu hoặc cung cấp nguồn khai thác dữ liệu.

Giải pháp đề xuất sẽ là một mô hình học thích ứng (Adaptive learning) mà ở đó, mô hình AI sẽ chỉ được hỗ trợ dẫn dắt ban đầu. Tương tự như một đứa trẻ lên 3 tuổi, sẽ được hướng dẫn và dạy dỗ một số nội dung cơ bản, đến một thời điểm bản thân đứa trẻ phải tự học hỏi và cập nhật kiến thức xung quanh để tồn tại và phát triển. Mô hình đề xuất có thể được gọi là mô hình học thích ứng nhận dạng đối tượng tổng quát (Advance Adaptive Object Recognition – AAOG). Mô hình AAOG phải giải quyết 03 bài toán cơ bản: (1) thích ứng về mô hình; (2) thích ứng về tham số và (3) thích ứng về mặt dữ liệu. Mỗi một mô hình nhận dạng (CNN) có các cấu trúc khác nhau, phù hợp và đưa đến độ chính xác với một nhóm đối tượng nhất định. Có thể kể đến các mô hình nhận dạng của các nhóm đối tượng như: nhận dạng phương tiện, nhận dạng người, nhận dạng chữ viết, xử lý ngôn ngữ tự nhiên,... Vì vậy, việc lựa chọn mô hình nhận dạng phù hợp với từng nhóm đối tượng sẽ là pha quan trọng để giải quyết bài toán đầu vào. Để giải bài toán về thích ứng dữ liệu, sau quá trình khởi tạo ban đầu, mô hình AAOG phải sử dụng trí thông minh có sẵn để nhận diện, học hỏi và tích lũy tri thức. Trong quá trình đó, mô hình AAOG phải tối ưu hóa về mặt dữ liệu, chọn lọc các dữ liệu và tri thức mới, loại bỏ các dữ liệu không cần thiết và tiến hành cập nhật cho mô hình hiện tại. Quá trình hoạt động không ngừng theo thời gian sẽ đưa đến nguồn tri thức kinh nghiệm khổng lồ cho AAOG. Ngoài ra, một mô hình AAOG sẽ được cài đặt các tham số huấn luyện ban đầu, tuy nhiên các tham số này chỉ phù hợp với các tập dữ liệu huấn luyện nhất định. Vì vậy, đối với các tập dữ liệu mới, phải đưa ra các tham số huấn luyện phù hợp nhằm tăng độ chính xác của mô hình AAOG trong quá trình huấn luyện. Các tập dữ liệu và tham số sẽ được gắn vào một mô hình nhận dạng được chỉ định, mô hình này sẽ được đánh giá và lựa chọn phù hợp với đối tượng cần nhận dạng. Giải quyết 03 vấn đề trên sẽ tạo nên một mô hình AAOG hoàn chỉnh, có khả năng thích ứng và tích lũy thông minh theo thời gian.

## 2. Các nghiên cứu gần đây

Các nghiên cứu trong thời gian gần đây về các hệ thống nhận dạng đưa đến các giải pháp với độ chính xác ngày càng cao, tốc độ huấn luyện và xử lý nhanh, ứng dụng trong nhiều các lĩnh vực: xe tự lái, robotic, xử lý ngôn ngữ tự nhiên,... Các giải pháp cơ bản có thể kể đến là sử dụng kết hợp các sensor và xử lý hình ảnh sử dụng các mô hình CNN. Các mô hình CNN được biết đến khá phổ biến như: mô hình AlexNet, mô hình GoogleNet, mô hình Microsoft ResNet, mô hình R-CNN, mô hình EfficientNet [21], mô hình Yolo (v1-9) [5], [28], ... đã đưa đến các nền tảng nghiên cứu căn bản. Dựa trên các mô hình này, hiện nay đã có nhiều hướng phát triển mới nhằm cải thiện độ chính xác và nâng cao tốc độ xử lý của hệ thống [23], [27], [31], [32]

Một số đề xuất theo hướng học thích ứng đã được các nhà nghiên cứu đề cập, trong đó nổi lên các hướng: Thích ứng về mô hình, Thích ứng về tham số và Thích ứng về dữ liệu.

**Thích ứng về mô hình:** Trong lựa chọn mô hình, các giải pháp tập trung vào việc lựa chọn tự động các kiểu mô hình nhận dạng mà không sử dụng một mô hình mặc định cụ thể (ví dụ như lựa chọn giữa CNN và SVM) [3], [22]. Việc lựa chọn các mô hình huấn luyện sẽ giúp giải quyết các trường hợp khác nhau của dữ liệu, đưa đến độ chính xác cao hơn. Ngoài ra, các giải pháp còn cho phép đánh giá: loại dữ liệu, kiểu dữ liệu,... để tự động chọn mô hình phù hợp [2], [17]. Tuy nhiên, các nghiên cứu và đánh giá về giải pháp thích ứng về mô hình đa số vẫn chỉ được đánh giá thực nghiệm trên các giải thuật truyền thống, chưa có đánh giá thực nghiệm sâu về việc sử dụng các mô hình CNN. Nguyên nhân có thể xuất phát từ sự hạn chế về hiệu suất của thiết bị phần cứng.

**Thích ứng về tham số:** là giải pháp tập trung vào các thuật toán và tham số của mô hình CNN, tạo ra các thay đổi thích ứng ở các lớp mô hình trong huấn luyện. Cụ thể, một số đề xuất xoay quanh việc xây dựng khung và nhúng chúng vào các vị trí lớp khác nhau để thay đổi trọng số trong huấn luyện và nhận dạng của mô hình CNN [12]. Các nghiên cứu khác tập trung vào việc thay đổi các tính năng, bao gồm căn chỉnh tính năng sâu cục bộ để tối giản mô hình [3] hoặc thay đổi tính năng giữa các lớp chập và tùy chỉnh các lớp [13], [18], [22]. Nhìn chung, các nghiên cứu nhằm tối ưu hóa cấu trúc của mô hình CNN và những thay đổi về cách bố trí của lớp cũng như tùy chỉnh các lớp đã đưa lại nhiều kết quả tích cực. Các mô hình mới sau khi huấn luyện có khả năng thông minh hơn với khả năng giữ lại các tính năng (features) quan trọng trong quá trình huấn luyện. Gần đây, một số đề xuất tập trung vào việc lựa chọn tự động các tham huấn luyện [16], [7], trong đó có giải pháp sử dụng một tập dữ liệu nhỏ ban đầu để đánh giá tính hiệu quả của các tham số, sau đó lựa chọn các tham số phù hợp và tiến hành huấn luyện đánh giá trên toàn bộ dữ liệu [7]. Hoặc lựa chọn ngẫu nhiên các tham số rồi tiến hành xác thực chéo trên một số lượng lần nhất định để lựa chọn tham số phù hợp với mô hình nhận dạng [16]. Ngoài ra, có giải pháp sử dụng kết hợp giải thuật tiến hóa để tự động tối ưu hóa cấu trúc của CNN bằng các siêu tham số [4]. Trong việc lựa chọn các siêu tham số cấu hình của một mô hình CNN, nổi bật là các phương pháp sử dụng các phương pháp lựa chọn ngẫu nhiên (Random Search) [8], phương pháp tìm kiếm dạng lưới hoặc phương pháp tìm kiếm sử dụng giải thuật Bayer [14], [33], [6]. Tuy nhiên, mặc dù được đánh giá là các đề xuất tối ưu trong việc thích ứng để đưa ra các cấu hình tốt nhất của mô hình nhận dạng, nhưng các đề xuất cũng đang dừng lại ở phần thực nghiệm với việc tìm kiếm số ít các tham số (hoặc các tham số cơ bản), nguyên nhân là nếu lựa chọn quá nhiều tham số sẽ ảnh hưởng lớn đến tốc độ xử lý của hệ thống.

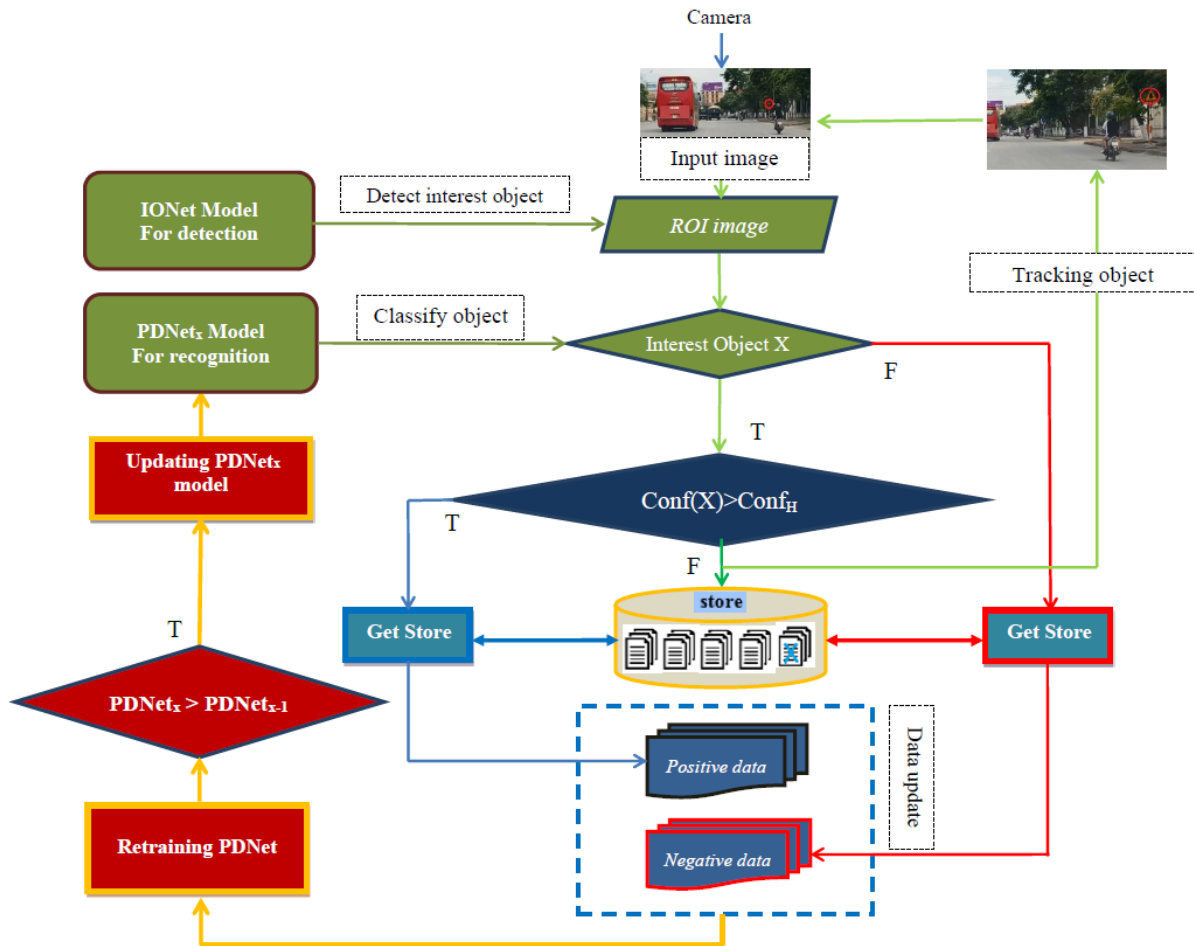
**Thích ứng về dữ liệu:** là nghiên cứu đề xuất mô hình có khả năng tự cải tiến về mặt dữ liệu, giúp mô hình CNN sẽ tự động cập nhật dữ liệu mà không cần sự can thiệp của con người. Thích ứng về dữ liệu được đánh giá là hướng nghiên cứu tiềm năng với việc mô hình CNN sẽ liên tục được học hỏi, cập nhật tương tự mô hình hoạt động của trí thông minh con người. Trong thích ứng về dữ liệu, có nhiều nghiên cứu theo hướng online tracking [19], [13] bằng cách điều chỉnh dữ liệu trong quá trình theo vết các đối tượng hoặc dựa vào việc trích xuất các đặc trưng (features) của đối tượng... Việc thích ứng về dữ liệu của một mô hình có phần tương quan với hoạt động của bộ não con người. Tuy nhiên, việc xử lý của bộ não con người là đa nhiệm, nghĩa là tại một thời điểm có thể nhận dạng và cập nhật nhiều loại đối tượng khác nhau. Ví dụ khi đi chuyển, con người sẽ xử lý và cập nhật: phương tiện giao thông, biển báo giao thông, cây cối, tòa nhà, người đi bộ, động vật, thời tiết,... Trong khi đó, đa số các nghiên cứu hiện nay chỉ mới dừng lại ở việc xử lý đơn nhiệm, nghĩa là mô hình chỉ nhận dạng một loại đối tượng

tại một thời điểm. Để giải quyết bài toán này cần một hệ thống đồ sộ với nhiều mô hình CNN được kết nối và hoạt động song song.

### 3. Các mô hình đã được đề xuất

#### 3.1. Mô hình thích ứng dữ liệu

Trong một đề xuất trước đây [29], nhóm tác giả đã đề xuất một mô hình có khả năng thích ứng và tự động cập nhật dữ liệu (Hình 1).



Hình 1. Mô hình học thích ứng về dữ liệu

##### 3.1.1. Định nghĩa các thành phần

Trước khi đi sâu vào từng khối chức năng của mô hình, các tác giả đề xuất một số khái niệm được sử dụng để làm rõ hơn các mô tả của luồng dữ liệu hệ thống:

(1) **Adaptive learning**: là khả năng tự học, tự thích ứng của một mô hình deep learning. Quá trình adaptive giúp mô hình tăng khả năng nhận dạng đối tượng so với mô hình ban đầu mà không cần sự can thiệp bổ sung dữ liệu và huấn luyện của chuyên gia.

(2) **Interest Object (IO)**: là đối tượng quan tâm cần nhận dạng, ví dụ như: con người, phương tiện giao thông, các đồ vật,...

(3) **Confident**: là giá trị đánh giá mức độ độ tin cậy khi một đối tượng được xác định là IO. Độ tin cậy của một đối tượng O được ký hiệu là  $Conf(O)$ .  $Confident_H$  là ngưỡng độ tin cậy cao.

(4) **Confident Tracking**: là quá trình theo vết khi một đối tượng được xác định là IO.

(5) **Lost Object (LO)**: là các đối tượng ban đầu được nhận dạng là đối tượng quan tâm IO nhưng có

độ tin cậy thấp, nó được theo vết qua n frame ảnh mà vẫn được nhận dạng là IO với độ tin cậy thấp và sau đó không còn xuất hiện trong khung hình tiếp theo nữa.

$$LO = \{O_1, O_2, \dots, O_n \mid Conf(O_i) < Confident_H \text{ với } i=1, \dots, n\}$$

(6) **Negative Object (NO)**: Là các đối tượng ban đầu được nhận dạng là đối tượng quan tâm IO (interest object) nhưng có độ tin cậy thấp (bé hơn  $Confident_H$ ), nó được theo vết (tracking) qua n frames và cuối cùng được nhận dạng không phải là IO.

$$NO = \{O_1, O_2, \dots, O_n \mid O_i \in IO \text{ and } Conf(O_i) < Confident_H \text{ với } i=1, \dots, n-1; O_n \notin IO\}$$

(7) **Positive Object (PO)**: Là các đối tượng ban đầu được nhận dạng là đối tượng quan tâm IO (interest object) nhưng có độ tin cậy thấp (bé hơn  $Confident_H$ ), nó được truy vết qua n frame ảnh và cuối cùng độ tin cậy vượt ngưỡng  $Confident_H$ .

$$PO = \{O_1, O_2, \dots, O_n \mid O_i \in IO \text{ and } Conf(O_n) \geq Confident_H\}$$

(8) **Store**: Là tập dữ liệu tạm thời chứa các tập dữ liệu theo từng đối tượng trong quá trình tracking. Một tập dữ liệu  $S_k$  được khởi tạo trong **Store** khi một đối tượng được xác định là  $IO_k$ . Một tập dữ liệu  $S_k$  sẽ bị xóa khi được xác định là **LO**, di chuyển vào **ND** nếu **IO** là **NO** hoặc di chuyển vào **PD** nếu **IO** là **PO**.

(9) **Negative data (ND)**: Là nơi chứa các tập dữ liệu **NO**.

(10) **Positive Data (PD)**: Là nơi chứa các tập dữ liệu **PO**.

(11) **Retraining**: Là quá trình huấn luyện lại mô hình khi sử dụng tập dữ liệu vừa thu thập được **PD** set và **ND** set khi số lượng các tập **PD** và **ND** đạt đến một giá trị nhất định.

(12) **Updating model**: là quá trình cập nhật lại mô hình của hệ thống khi mô hình mới có độ chính xác cao hơn mô hình trước đó sau khi đánh giá, chọn lọc.

### 3.1.2. Tổng quan của mô hình

Mô hình đề xuất tập trung vào giải pháp học thích ứng sử dụng mô hình CNN. Trong khuôn khổ của nội dung nghiên cứu trước đó, nhóm tác giả chỉ tập trung vào hai nhóm đối tượng là phương tiện tham gia giao thông và biển báo giao thông. Mô hình đề xuất sử dụng hai mô hình CNN: mô hình IONet phục vụ việc phát hiện các đối tượng và mô hình PDNet để xác định độ tin cậy và nhận dạng đối tượng.

Hệ thống được mô phỏng là hệ thống hoạt động liên tục theo thời gian. Trong quá trình hoạt động của mình, hệ thống sẽ liên tục phát hiện, nhận dạng và tổng hợp dữ liệu, huấn luyện để tự cập nhật mô hình nhận dạng (PDNet). Giả định tại một thời điểm trong quá trình hoạt động, hệ thống phát hiện đối tượng (phương tiện hoặc biển báo giao thông). Tuy nhiên, tại thời điểm đó, mô hình PDNet chỉ nhận dạng với độ tin cậy thấp với nhiều nguyên nhân khác nhau như: khoảng cách xa, bị che khuất hoặc trong điều kiện thời tiết xấu thiếu ánh sáng, bị mưa, nhiễu chuyển động,... Quá trình học thích ứng sẽ được kích hoạt với việc lưu trữ liên tục các hình ảnh có độ tin cậy thấp (tại **Store**) và tiếp tục theo vết đối tượng. Kết thúc quá trình theo vết đối tượng sẽ có ba trường hợp xảy ra: (i) Mất đối tượng (**Lost object**); (ii) Đối tượng nhận dạng sai (**Negative Object**); (iii) Đối tượng nhận dạng đúng (**Positive Object**). Trường hợp Mất đối tượng hệ thống sẽ tự động xóa các dữ liệu ảnh đã được lưu trữ trước đó trong quá trình theo vết (tracking). Trường hợp nhận dạng sai thì tập dữ liệu ảnh được lưu trữ sẽ được gán nhãn là **Negative Data**. Trường hợp nhận dạng đúng thì tập dữ liệu ảnh được lưu trữ sẽ được gán nhãn là **Positive Data**. Khi số lượng dữ liệu tại hai tập dữ liệu này đủ lớn, hệ thống sẽ bắt đầu quá trình huấn luyện lại. Tập dữ liệu thu nhận này sẽ được trộn với 30% tập dữ liệu trước đó, tạo thành tập dữ liệu hoàn chỉnh để huấn luyện lại mô hình PDNet. Mô hình PDNet mới sau khi huấn luyện sẽ được đánh giá với các mô hình trước đó. Nếu được đánh giá tốt hơn sẽ được sử dụng để thay thế mô hình cũ. Quá trình này được lặp lại liên tục trong quá trình hoạt động của hệ thống.

## 3.2. Mô hình thích ứng dữ liệu và tham số

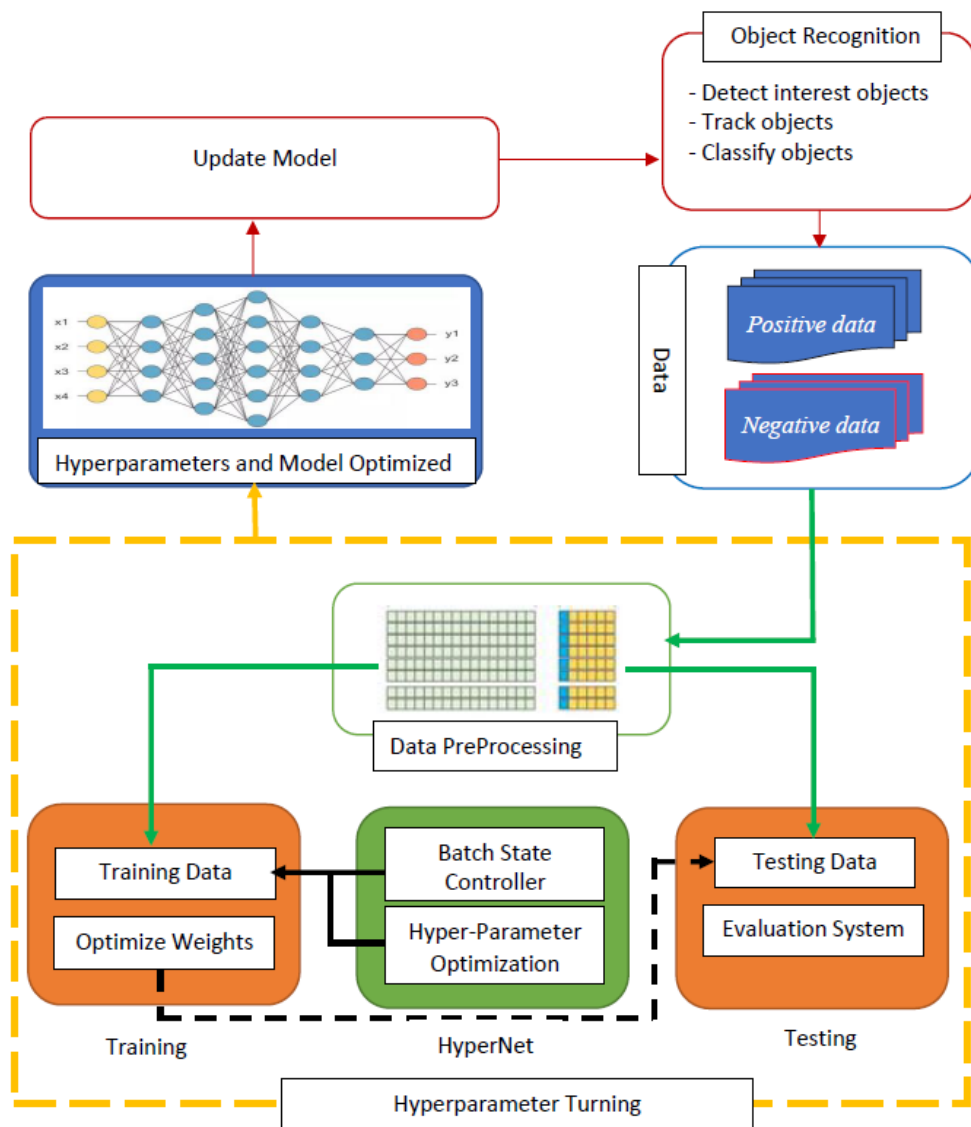
### 3.2.1. Giải quyết vấn đề

Mô hình hệ thống đề xuất trước đó (Hình 1) [29] có khả năng tự cải thiện khả năng nhận dạng và trở nên "thông minh" hơn. Tuy nhiên, ta nhận thấy trong mỗi lần hệ thống thu thập và cập nhật lại bộ dữ liệu mới (Retrain dataset), thì tập dữ liệu này có thay đổi so với tập dữ liệu trước đó. Trong khi đó, cấu

trúc của mô hình CNN là không đổi trong quá trình huấn luyện và nhận dạng. Chính vì vậy, các tác giả đã đề xuất giải pháp thay đổi tự động các siêu tham số của mô hình CNN, thích ứng và phù hợp với từng tập dữ liệu huấn luyện. Giúp mô hình PDNet cải nâng cao sự chính xác trong nhận dạng đối tượng.

### 3.2.2. Mô hình

Mô hình đề xuất [30] tiếp tục phát triển từ mô hình đã được đề xuất tại mục 3.1, tác giả sẽ không đề cập và trình bày lại. Nội dung thay đổi và bổ sung chủ yếu tập trung vào thay đổi khối chức năng **Retraining PDNet**. Trong khối chức năng này, bổ sung chức năng **HyperNet** cho phép thực hiện lựa chọn các siêu tham số phù hợp của mô hình huấn luyện một cách tự động bằng thuật toán bằng giải thuật **Bayers**. Mô hình đề xuất tổng thể của giải pháp được trình bày tại Hình 2.



**Hình 2.** Mô hình học thích ứng về tham số, dữ liệu

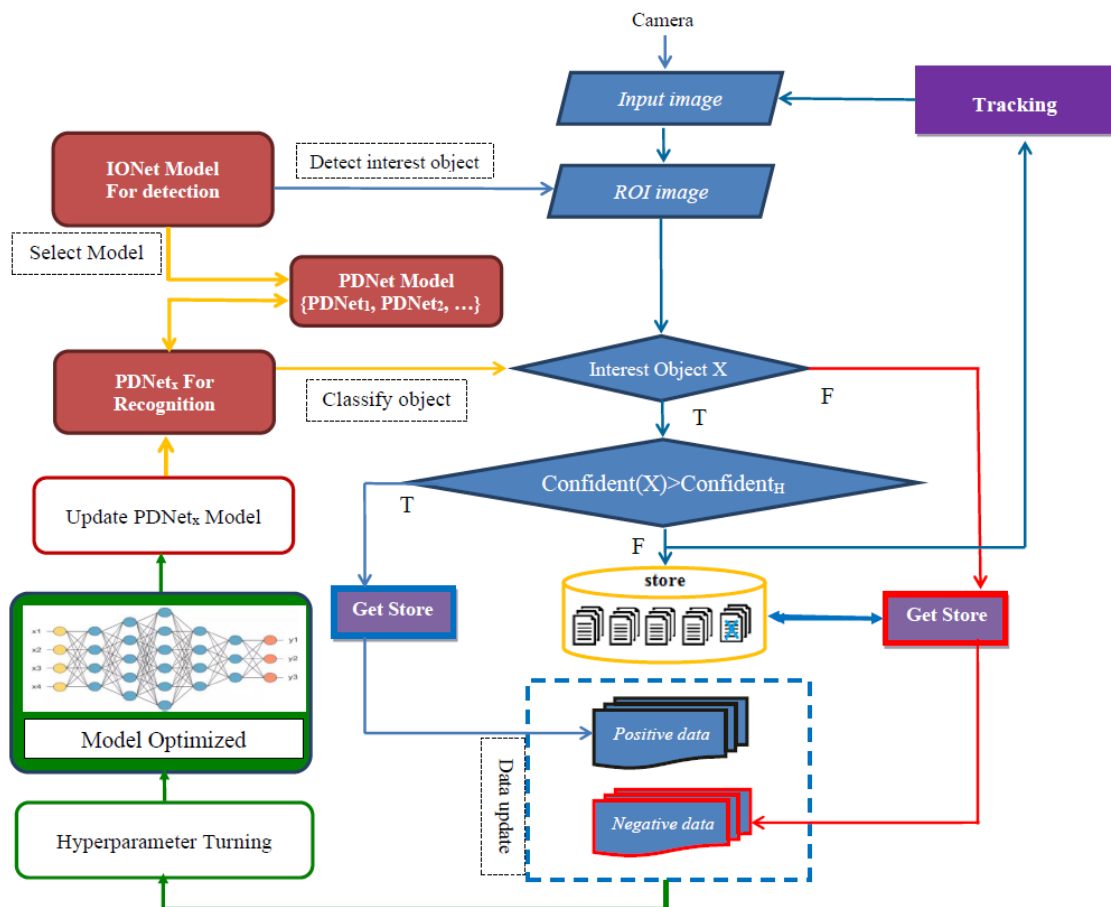
Trong suốt quá trình huấn luyện lại mô hình CNN trước đó, các thông số cấu trúc và huấn luyện ít thay đổi. Tuy nhiên, dữ liệu thu thập từ hệ thống thường xuyên biến đổi và được cập nhật. Do đó, theo lý thuyết có thể thấy việc điều chỉnh cấu trúc và tham số huấn luyện của mô hình CNN để phù hợp với dữ liệu mới là quan trọng. Tuy nhiên, vì tính đặc biệt của giải pháp được đề xuất, mô hình CNN sau khi tái huấn luyện (Retrained model) vẫn kế thừa sự "thông minh" từ mô hình trước đó, không thực hiện

việc điều chỉnh kiến trúc. Thay vào đó, giải pháp chú trọng vào sử dụng giải thuật Bayers để lựa chọn các siêu tham số phù hợp trong huấn luyện, từ đó tìm ra mô hình tối ưu nhất.

Có thể nói, các mô hình đề xuất trước đó (mục 3.1 và mục 3.2) đã cơ bản hình thành cơ chế học thích ứng. Tuy nhiên, mô hình PDNet không thay đổi trong suốt quá trình huấn luyện và hoạt động, vì vậy chỉ có khả năng nhận dạng trong một nhóm đối tượng nhất định. Về mặt lý thuyết, môi trường xung quanh có vô vàn đối tượng, mỗi đối tượng có các đặc điểm khác nhau, chỉ phù hợp với một mô hình và cấu trúc CNN nhất định. Chính vì vậy, chúng tôi đề xuất giải pháp thích ứng về mô hình nhằm hoàn thiện giải pháp học thích ứng trong nhận dạng đối tượng.

#### 4. Giải pháp đề xuất

Trong các công bố gần đây [29], [30], nhóm tác giả đã đề xuất mô hình học thích ứng dựa vào cải tiến dữ liệu và tham số. Tuy nhiên, các dữ liệu và tham số này được huấn luyện cho một mô hình cố định và chỉ phù hợp cho các nhóm đối tượng cụ thể nhất định. Trong khi thực tế có rất nhiều đối tượng xuất hiện trong quá trình hoạt động của hệ thống, đòi hỏi phải có các mô hình khác nhau, phù hợp mới có thể cải tiến và nâng cao độ chính xác. Có thể hiểu một cách khái quát cách thức hoạt động sẽ tương tự bộ não con người. Bộ não con người có nhiều khu vực, mỗi khu vực đảm nhận xử lý một số vấn đề nhất định như: ngôn ngữ, vận động, suy luận, thị giác,... Ví dụ như “Thùy trán” điều khiển các hoạt động vận động, giải quyết tình huống, khả năng phán đoán và kiểm soát cảm xúc, “Thùy đỉnh” thu nhận và xử lý thông tin từ bộ phận cảm giác, “Thùy thái dương” có chức năng ghi nhớ trực quan, hiểu về cảm xúc và ngôn ngữ, “Thùy chẩm” giúp con người có khả năng đọc và hiểu từ ngữ hay những thông tin liên quan đến tầm nhìn... Trên cơ sở các kết quả thực nghiệm trước đây kết hợp với thực tiễn phát triển, chúng tôi đề xuất mô hình tổng thể cho mô hình học thích ứng, được gọi là mô hình học thích ứng nhận dạng đối tượng tổng quát (Advance Adaptive Object Recognition – AAOG).



**Hình 3.** Mô hình đề xuất học thích ứng về mô hình, tham số và dữ liệu

Mô hình đề xuất là sự kết hợp và kế thừa của mô hình tại Mục 3.1 và mô hình Mục 3.2. Các kết quả thực nghiệm trước đây đã khẳng định các mô hình hoàn toàn khả thi và ứng dụng trong thực tế. Trên cơ sở các mô hình cũ, chúng tôi đề xuất cải tiến, bổ sung cho phép hệ thống AAOG có khả năng lựa chọn mô hình phù hợp với từng đối tượng được phát hiện trong quá trình hoạt động.

Có thể mô tả tổng quát như sau: Trong quá trình hoạt động liên tục của hệ thống, mô hình IONet ghi nhận một đối tượng X và tiến hành theo vết (tracking) đối tượng (hệ thống có thể nhận dạng nhiều đối tượng  $X_1, X_2, \dots$  cùng thời điểm). Đồng thời, hệ thống cũng sẽ lựa chọn một mô hình PDNet<sub>x</sub> phù hợp. Sự khác biệt nằm ở chỗ mô hình PDNet<sub>x</sub> không phải là cố định mà sẽ được lựa chọn phù hợp từ tập hợp các mô hình có sẵn, mỗi mô hình PDNet<sub>x</sub> sẽ phù hợp với một nhóm đối tượng X nhất định. Hệ thống sẽ tiếp tục trích xuất vùng quan tâm (Region of Interest - ROI) và sử dụng mô hình PDNet<sub>x</sub> để nhận dạng đối tượng. Nếu độ tin cậy khi nhận dạng đối tượng nhỏ hơn giá trị ngưỡng **Confiden<sub>H</sub>** thì hệ thống sẽ tiếp tục theo vết (tracking). Các dữ liệu thu nhận được trong quá trình theo vết sẽ được lưu trữ vào **Store**, gán nhãn tập dữ liệu **Store<sub>x</sub>**. Khi tin độ cậy vượt giá trị ngưỡng **Confiden<sub>H</sub>** thì hệ thống sẽ chuyển toàn bộ dữ liệu thu nhận của đối tượng đó sang tập **Positive data**. Ngược lại, nếu sau quá trình theo vết, đối tượng được nhận dạng không phải là X thì hệ thống sẽ chuyển toàn bộ dữ liệu của tập **Store<sub>x</sub>** sang tập **Negative data**. Khi số lượng dữ liệu thu thập được của tập **Positive data** và **Negative data** đạt được giá trị nhất định, sẽ được sử dụng để huấn luyện cho mô hình PDNet<sub>x</sub>. Quá trình huấn luyện lại mô hình PDNet<sub>x</sub> sẽ được tối ưu hóa thông qua quá trình tìm kiếm các siêu tham số của giải thuật Bayers. Toàn bộ quá trình tìm kiếm siêu tham số đã được mô tả chi tiết tại Hình 2, phần Hyperparameter Turning. Trên cơ sở các siêu tham số tối ưu vừa tìm được, mô hình PDNet sẽ được huấn luyện và cập nhật. Mô hình PDNet<sub>x</sub> sau khi được cập nhật sẽ thông minh hơn mô hình trước đó. Quá trình đó sẽ lặp lại liên tục trong suốt thời gian hoạt động của mô hình AAOG, lần lượt các mô hình PDNet sẽ được cải thiện khả năng nhận dạng.

Trong thời gian tới, khi điều kiện về hiệu suất phần cứng cho phép, chúng tôi đề xuất thực nghiệm mô hình AAOG như một AutoRobot với các đối tượng dữ liệu về giao thông như: phương tiện giao thông, biển báo giao thông, đường giao thông, người đi bộ, cây cối, tòa nhà, ... Với thời gian hoạt động đủ dài (từ 30 đến 90 ngày), một hệ thống AAOG có khả năng cải thiện sự thông minh theo thời gian.

## 5. Đánh giá giải pháp và những thách thức đặt ra

Trong 2 mô hình đề xuất trước đó [29], [30], các tác giả đã chứng minh bằng thực nghiệm với các nền tảng phần cứng phổ thông sẵn có như Bảng 1. Tại thời điểm đó, quá trình tìm kiếm siêu tham số theo mô hình đề xuất tại Hình 2 hệ thống phải mất từ 120 đến 170 giờ (trung bình từ 05-07 ngày) hoạt động liên tục để giải thuật Bayer chọn lọc được một bộ tham số tối ưu nhất trong một vòng lặp.

**Bảng 1. Cấu hình của thiết bị trong các đề xuất trước đây**

Thông số	Cấu hình chi tiết
CPU	Core I3 3.6Ghz
GPU	Geforce 1060 6Gb
RAM	16Gb
HDD	SSD 160Gb

Chính vì vậy, với đề xuất hiện tại, một số khó khăn và thách thức ngăn cản quá trình thực nghiệm và đưa ra kết quả. Có thể kể đến các khó khăn gặp phải đó là:

(1) Hiệu năng xử lý của thiết bị: Với việc lựa chọn mô hình PDNet<sub>x</sub> phù hợp với từng đối tượng nhận dạng, mô hình phải có tốc độ xử lý nhanh (CPU, GPU, RAM). Đặc biệt là hệ thống phải đảm bảo xử lý đa nhiệm, quá trình theo vết (tracking) và huấn luyện (retraining) được thực hiện cùng thời điểm cho nhiều đối tượng. Nghĩa là tại một thời điểm, có nhiều mô hình PDNet cùng thực hiện nhận dạng hoặc huấn luyện. Ngoài ra, với cấu hình thiết bị được cải tiến, có thể tăng số lượng siêu tham số (hyperparameter) để giải thuật Bayesian có thể tìm kiếm các siêu tham số tốt hơn. Có thể tăng từ 6 tham số như các đề xuất tại Mục 3.2 ('InitialLearnRate', 'L2Regularization', 'Momentum', 'MiniBatchSize',

'GradientThreshold' và 'GradientThresholdMethod') tăng lên 10 đến 20 tham số trong tổng số khoảng 28 tham số cơ bản của một mô hình CNN.

(2) Khả năng lưu trữ: Trong các mô hình đề xuất trước đây, các tác giả chỉ đặt giới hạn huấn luyện khi số lượng ảnh của đối tượng đạt ngưỡng khiêm tốn, 200 ảnh. Nghĩa là khi tập dữ liệu **Positive data** và **Negative data** đạt 200 ảnh, sẽ được trộn với 30% tập dữ liệu cũ để đạt được bộ dữ liệu hoàn chỉnh, phục vụ huấn luyện lại mô hình (retrain). Với việc nhiều mô hình PDNet<sub>x</sub> cùng thực hiện nhận dạng và huấn luyện lại, chỉ trong thời gian ngắn, bộ nhớ lưu trữ của hệ thống sẽ bị tràn, không còn không gian lưu trữ dữ liệu.

(3) Khả năng truyền dữ liệu trong môi trường mạng: Một giải pháp có thể giải bài toán về hiệu năng của thiết bị và khả năng lưu trữ đó là áp dụng mô hình đám mây (cloud). Tức là hệ thống AAOG chỉ thực hiện việc nhận dạng và phân loại đối tượng. Việc xử lý và tính toán được thực hiện tại máy chủ lớn, khác không gian địa lý. Tuy nhiên, việc này đòi hỏi khả năng đáp ứng về truyền tải dữ liệu với tốc độ cao, không giới hạn. Điều này chỉ có thể làm được với công nghệ truyền dữ liệu 5G, 6G.

(4) Kích thước của hệ thống AAOG: Một hệ thống AAOG có khả năng thích ứng và gia tăng sự thông minh liên tục theo thời gian đòi hỏi hệ thống đó phải có khả năng vận động, di chuyển trong quá trình hoạt động (time life). Để làm được điều đó, hệ thống phải tinh gọn, tối thiểu về kích thước nhưng vẫn phải đảm bảo hiệu năng (năng lượng, tốc độ xử lý, khả năng lưu trữ,...). Đây là một thách thức không hề nhỏ khi các nền tảng công nghệ thực tế vẫn chưa đáp ứng được do các máy chủ lớn (MainFrame) hoặc siêu máy tính (SupperComputer) có tốc độ xử lý lớn vẫn là các cỗ máy đồ sộ.

Mặc dù còn gặp không ít thách thức, nhưng với tốc độ phát triển như hiện nay, tin chắc rằng, trong thời gian ngắn sẽ khắc phục hoàn toàn các chướng ngại đang gặp phải và một hệ thống sử dụng AAOG hoạt động trong thực tế là hoàn toàn khả thi.

#### 4. Kết luận

Nội dung nghiên cứu và đề xuất mô hình hoạt động của một hệ thống học thích ứng nhận dạng đối tượng tổng quát - AAOG. Hệ thống có khả năng ứng dụng trong các auto robot hoặc phục vụ nghiên cứu xây dựng các mô hình AI tổng quát. Mặc dù chưa được thực nghiệm bằng các số liệu cụ thể, nhưng qua các thực nghiệm của mô hình đề xuất trước đó có thể khẳng định mô hình đề xuất tổng quát có khả năng ứng dụng trong thực tế.

Với sự phát triển mạnh mẽ của công nghệ, các nền tảng chip xử lý (CPU, GPU) và tốc độ mạng không dây sẽ không ngừng được nâng cao, là cơ sở quan trọng cho việc mở rộng các mô hình AI trong đó có ứng dụng các nền tảng của AAOG.

#### Xung đột lợi ích

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

#### TÀI LIỆU THAM KHẢO

- [1] J. Achiam *et al.*, "Gpt-4 technical report," *arXiv preprint arXiv:2303.08774*, 2023.
- [2] H. Bertrand *et al.*, "Hyperparameter optimization of deep neural networks: Combining hyperband with Bayesian model selection," in *Conférence sur l'Apprentissage Automatique*, 2017, Art no. 10497518.
- [3] K. Blix and T. Eltoft, "Machine learning automatic model selection algorithm for oceanic chlorophyll-a content retrieval," *Remote Sensing*, vol. 10, p. 775, 2018.
- [4] E. Bochinski *et al.*, "Hyper-parameter optimization for convolutional neural network committees based on evolutionary algorithms," in *2017 IEEE international conference on image processing (ICIP)*, 2017, pp. 3924-3928.
- [5] C. Y. Wang, I. H. Yeh, and H. Y. M. Liao, "YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information," *arXiv:2402.13616v2*, 2024.
- [6] G. Dikov and J. Bayer, "Bayesian learning of neural network architectures," in *The 22nd International Conference on Artificial Intelligence and Statistics*, 2019, pp. 730-738.
- [7] T. Domhan *et al.*, "Speeding up automatic hyperparameter optimization of deep neural networks by extrapolation of learning curves," in *Twenty-fourth international joint conference on artificial intelligence*, 2015, pp. 3460-3468.
- [8] A. C. Florea and R. Andonie, "Weighted random search for hyperparameter optimization," *arXiv preprint arXiv:2004.01628*, 2020.
- [9] R. Girshick, "Fast r-cnn," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1440-1448.
- [10] R. Girshick *et al.*, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 580-587.
- [11] K. He *et al.*, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770-778.
- [12] V. D. Hoang *et al.*, "Robust human detection using multiple scale of cell based histogram of oriented gradients and adaboost learning," in *International Conference on Computational Collective Intelligence*, 2012, pp. 61-71.

- [13] C. Huang *et al.*, "Learning policies for adaptive tracking with deep feature cascades," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 105-114.
- [14] L. Kotthoff *et al.*, "Auto-WEKA 2.0: Automatic model selection and hyperparameter optimization in WEKA," *Journal of Machine Learning Research*, vol. 18, pp. 1-5, 2017.
- [15] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, pp. 1097-1105, 2012.
- [16] N. Q. K. Le *et al.*, "Identification of clathrin proteins by incorporating hyperparameter optimization in deep learning and PSSM profiles," *Computer methods and programs in biomedicine*, vol. 177, pp. 81-88, 2019.
- [17] L. Li and A. Talwalkar, "Random search and reproducibility for neural architecture search," in *Uncertainty in artificial intelligence*, 2020, pp. 367-377.
- [18] M. Long *et al.*, "Transferable representation learning with deep adaptation networks," *IEEE transactions on pattern analysis and machine intelligence*, vol. 41, pp. 3071-3085, 2018.
- [19] H. Nam and B. Han, "Learning multi-domain convolutional neural networks for visual tracking," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 4293-4302.
- [20] A. Oksanen *et al.*, "Artificial intelligence in fine arts: A systematic review of empirical research," *Computers in Human Behavior: Artificial Humans*, p. 100004, 2023.
- [21] S. B. Punuri *et al.*, "Efficient net-XGBoost: an implementation for facial emotion recognition using transfer learning," *Mathematics*, vol. 11, p. 776, 2023.
- [22] S. Raschka, "Model evaluation, model selection, and algorithm selection in machine learning," *arXiv preprint arXiv:1811.12808*, 2018.
- [23] D. Reis *et al.*, "Real-time flying object detection with YOLOv8," *arXiv preprint arXiv:2305.0997*, 2023.
- [24] S. Ren *et al.*, "Faster r-cnn: Towards real-time object detection with region proposal networks," in *Advances in neural information processing systems*, 2015, pp. 91-99.
- [25] Z. Song *et al.*, "MovieLLM: Enhancing Long Video Understanding with AI-Generated Movies," *arXiv preprint arXiv:2403.01422*, 2024.
- [26] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1-9.
- [27] F. M. Talaat and H. ZainEldin, "An improved fire detection approach based on YOLO-v8 for smart cities," *Neural Computing and Applications*, vol. 35, pp. 20939-20954, 2023.
- [28] J. Terven and D. C. Esparza, "A comprehensive review of YOLO: From YOLOv1 to YOLOv8 and beyond," *arXiv preprint arXiv:2304.00501*, 2023.
- [29] D. P. Tran and V. D. Hoang, "Adaptive learning based on tracking and ReIdentifying objects using convolutional neural network," *Neural Processing Letters*, vol. 50, pp. 263-282, 2019.
- [30] D. P. Tran *et al.*, "Hyperparameter optimization for improving recognition efficiency of an adaptive learning system," *IEEE Access*, vol. 8, pp. 160569-160580, 2020.
- [31] S. Wang *et al.*, "Single-Stage Pose Estimation and Joint Angle Extraction Method for Moving Human Body," *Electronics*, vol. 12, p. 4644, 2023.
- [32] X. Wang *et al.*, "BL-YOLOv8: An improved road defect detection model based on YOLOv8," *Sensors*, vol. 23, p. 8361, 2023.
- [33] X. Zeng and G. Luo, "Progressive sampling-based Bayesian optimization for efficient and automatic machine learning model selection," *Health information science and systems*, vol. 5, pp. 1-21, 2017.



**Trần Diễm Phúc** is a Ph.D in Computer Science from Duy Tan University, Da Nang City, Vietnam. He is working in Information and Communications department of Quang Binh province. His main research interests include artificial intelligence, image processing, computer vision, auto vehicle and robotic.

Email: [phuctd@gmail.com](mailto:phuctd@gmail.com). ORCID:  <https://orcid.org/0000-0001-8335-6720>