

BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC DUY TÂN

**GIẢI PHÁP HỌC THÍCH ỨNG TRÊN NỀN
TẢNG MẠNG HỌC SÂU ỨNG DỤNG NHẬN DẠNG
ĐỐI TƯỢNG THAM GIA GIAO THÔNG**

Chuyên ngành : Khoa học máy tính

Mã số : 9480101

TÓM TẮT LUẬN ÁN TIẾN SĨ KHOA HỌC MÁY TÍNH

ĐÀ NẴNG - NĂM 2022

Công trình được hoàn thành tại Trường Đại học Duy Tân

Phản biện 1:

Phản biện 2:

Phản biện 3:

Luận án sẽ được bảo vệ trước Hội đồng chấm luận án họp tại ...

.....
vào lúc giờ ngày tháng năm

Có thể tìm hiểu luận án tại:

- Thư viện Quốc gia Việt Nam
- Trung tâm Thông tin tư liệu Trường Đại học Duy Tân

MỞ ĐẦU

Trí tuệ nhân tạo hay trí thông minh nhân tạo là trí tuệ được biểu diễn bởi một hệ thống nhân tạo. Ngày nay, trí tuệ nhân tạo xuất hiện ở khắp mọi nơi, trong ứng dụng văn phòng, trong các hệ thống trả lời tự động, trong quản lý giao thông thông minh, quản lý ngôi nhà thông minh... Cùng với sự phát triển của các hệ thống phần cứng máy tính, trí tuệ nhân tạo đã có những bước phát triển vượt bậc, ứng dụng ngày càng sâu rộng trong các lĩnh vực đời sống và xã hội.

Trí tuệ nhân tạo tập trung vào phát triển các giải thuật, ứng dụng hỗ trợ con người trong việc ra quyết định hoặc tự quyết định trong quá trình nhận biết và thu nạp dữ liệu. Một trong những hướng nghiên cứu đang được các nhà khoa học, các tập đoàn công nghệ hàng đầu nhắm đến là nhận dạng đối tượng, hành động của đối tượng và con người (Human Action Recognition), như: các hệ thống giám sát an ninh, hệ thống điều khiển từ xa bằng thao tác, hệ thống hỗ trợ người mù, hệ thống phân tích dữ liệu trong thể thao, robot tự động, xe ô tô tự lái [Hariyono, 2017], [Dollar, 2012], Stewart, 2016,], [Van-Dung Hoang, 2012]... Đã có nhiều nghiên cứu và đưa ra nhiều phương pháp khác nhau để phát triển trí tuệ nhân tạo (giải thuật heurictis, giải thuật tiến hóa, giải thuật Support Vector Machine, giải thuật Hidden Markov Model, phương pháp hệ chuyên gia, phương pháp mạng nơron,...), tuy nhiên các phương pháp truyền thống đều cần sự can thiệp của con người, đòi hỏi lượng dữ liệu phân tích, lưu trữ vô cùng lớn nhưng đưa đến độ chính xác không cao và chỉ giới hạn trong một số trường hợp nhận dạng nhất định.

Để khắc phục những hạn chế nói trên, trí tuệ nhân tạo trong nhận dạng đối tượng và hành động hiện nay sử dụng phương pháp học máy (Machine Learning), tập trung vào phương pháp học sâu (Deep Learning).

Deep learning đã và đang là một chủ đề AI được bàn luận sôi nổi. Là một phạm trù nhỏ của Machine Learning, Deep Learning tập trung giải quyết các vấn đề liên quan đến mạng thần kinh nhân tạo nhằm nâng cấp các công nghệ như nhận diện giọng nói, nhận dạng hình ảnh và xử lý ngôn ngữ tự nhiên. Chỉ trong vài năm, Deep Learning đã thúc đẩy tiến bộ trong đa dạng các lĩnh vực như nhận biết sự vật (Object Perception), dịch tự động (Machine Translation), nhận diện giọng nói,

... những vấn đề từng rất khó khăn với các nhà nghiên cứu trí tuệ nhân tạo.

Tuy nhiên, mặc dù đã giải quyết được các bài toán tương đối lớn của khoa học hiện đại về AI, Deep Learning vẫn còn tồn tại nhiều hạn chế cần phải khắc phục và nghiên cứu:

- Thứ nhất, để tạo ra một hệ thống có khả năng nhận dạng đa dạng các đối tượng, Deep Learning cần một lượng dữ liệu đầu vào khổng lồ để máy tính có thể học hỏi. Quy trình này mất nhiều thời gian và cần một bộ xử lý cực kỳ lớn mà chỉ có các hệ thống máy chủ cỡ lớn mới làm được.

- Thứ hai, Deep Learning vẫn chưa thể nhận biết được những thứ phức tạp, ví dụ như các mối liên hệ thông thường của xã hội. Chúng cũng sẽ gặp khó khăn khi cần nhận biết những thứ tương tự nhau. Lý do là vì hiện chưa có kỹ thuật nào đủ tốt để trí tuệ nhân tạo có thể rút ra những kết luận đó một cách logic. Bên cạnh đó, vẫn còn nhiều thách thức trong việc tích hợp kiến thức trừu tượng vào các hệ thống Machine Learning, ví dụ như thông tin về vật đó là gì, nó dùng để làm gì, người ta hay dùng nó ra làm sao... Nói cách khác, Machine Learning chưa có được những kiến thức thông thường như con người.

Câu hỏi đặt ra là làm sao để một hệ thống Machine Learning có thể học hỏi được các tri thức, chọn lọc và cập nhật được những kiến thức phù hợp và tự xây dựng một tập dữ liệu có tính ràng buộc, xâu chuỗi như con người. Hướng nghiên cứu **Adaptive Learning** [Chandra, 2016], [Chatzilari, 2016], [Wang, 2017], [Zhang, 2017], [Liu, 2017], [Zhang, 2016] có thể là một giải pháp nhằm cải tiến các hạn chế của Deep Learning, khai phá những vấn đề mà Deep Learning chưa làm được.

Một mô hình Adaptive Learning hoàn chỉnh sẽ cho phép hệ thống auto robot có khả năng tự học hỏi, tự thông minh mô phỏng theo hoạt động của bộ não con người. Sự thông minh của hệ thống sẽ được tăng lên theo thời gian theo quá trình hoạt động của thiết bị. Hệ thống sẽ tự động chọn lọc các dữ liệu được cho là phù hợp, tự huấn luyện lại mô hình và tự cập nhật thay thế mô hình cũ.

Trong khuôn khổ nội dung nghiên cứu của một luận án Tiến sĩ, nghiên cứu sinh sẽ tiến hành nghiên cứu và thực nghiệm trên các

phương tiện xe tự lái (ADAS) nhằm mô phỏng cho quá trình hoạt động của một autorobot. Đối tượng nhận dạng của phương tiện tự lái là các đối tượng tham gia giao thông như: phương tiện giao thông khác (xe máy, xe ô tô, xe tải, xe khách,...), người đi bộ, biển báo giao thông, lòng đường, lề đường, ...

CHƯƠNG 1. TỔNG QUAN VỀ TRÍ TUỆ NHÂN TẠO

1.1 Trí tuệ nhân tạo

Trên thế giới hiện có nhiều định nghĩa khác nhau về trí tuệ nhân tạo hay còn gọi là AI (Artificial Intelligence), cụ thể:

- Về tổng quan, trí tuệ nhân tạo là trí tuệ được biểu diễn bởi bất cứ một hệ thống nhân tạo nào. Thuật ngữ này thường dùng để nói đến các máy tính có mục đích không nhất định và ngành khoa học nghiên cứu về các lý thuyết và ứng dụng của trí tuệ nhân tạo.

- Bellman (1978) định nghĩa: trí tuệ nhân tạo là tự động hoá các hoạt động phù hợp với suy nghĩ con người, chẳng hạn các hoạt động ra quyết định, giải bài toán, ...

- Rich anh Knight (1991) thì cho rằng: Trí tuệ nhân tạo là khoa học nghiên cứu xem làm thế nào để máy tính có thể thực hiện những công việc mà hiện con người còn làm tốt hơn máy tính.

Mỗi khái niệm, định nghĩa đều có điểm đúng riêng, nhưng để đơn giản chúng ta có thể hiểu trí tuệ nhân tạo là một ngành khoa học máy tính. Nó xây dựng trên một nền tảng lý thuyết vững chắc và có thể ứng dụng trong việc tự động hóa các hành vi thông minh của máy tính. Giúp máy tính có được những trí tuệ của con người như: biết suy nghĩ và lập luận để giải quyết vấn đề, biết giao tiếp do hiểu ngôn ngữ, tiếng nói, biết học và tự thích nghi.

1.2 Nghiên cứu trong nước và quốc tế

1.2.1 Trong nước

Ở nước ta, từ thập kỷ 90 đến những năm đầu của thế kỷ 20, trong lĩnh vực nghiên cứu về trí tuệ nhân tạo đặc biệt là về lĩnh vực xử lý ảnh và nhận dạng, có sự tham gia của các nhà nghiên cứu nổi tiếng như PGS.TS Ngô Quốc Tạo, PGS. TS Đỗ Năng Toàn, PGS. TS Lương Chi

Mai... Có thể kể đến những công trình nghiên cứu nổi tiếng và có tính ứng dụng cao về nhận dạng chữ viết [Dang Ngoc Duc, 2003], [P. A. Phuong, 2008], chữ viết tay tiếng Việt [Phạm Anh Phuong, 2009], [Ngô Quốc Tạo, 2004], nhận dạng giọng nói, nhận dạng mặt người [Lam Thanh Hien , 2012], [Lê Thanh Hà, 2006], [Đỗ Năng Toàn, 2011], mô phỏng cơ thể người [Pham Ngoc Hung, 2017]...đá số các nghiên cứu và các công trình công bố đều sử dụng và khai phá các giải thuật truyền thống như SVM , RandomForest, mô hình Markov ẩn, mạng nơron nhân tạo,...Những nghiên cứu này là tiền đề và nền tảng quan trọng cho các sinh viên, học viên cao học và nghiên cứu sinh tham khảo, phát triển thành các công trình nghiên cứu lớn. Song song với việc công bố các công trình, các nhà khoa học cũng đã xuất bản nhiều ấn phẩm, sách chuyên khảo về lĩnh vực xử lý ảnh, nhận dạng đối tượng.

Sau thập kỷ đầu tiên của thế kỷ 20, cùng với sự phát triển của trí tuệ nhân tạo và phần cứng máy tính đã cho phép lĩnh vực học máy (machine learning) và nhận dạng đối tượng phát triển mạnh mẽ. Tuy nhiên ở nước ta, trong những năm đầu, nghiên cứu về mạng nơron nhân tạo, mạng học sâu nhân chập (Convolution Neural Network) vẫn còn rất sơ khai, hầu như chưa có các nghiên cứu trong nước về lĩnh vực cụ thể này. Các nghiên cứu và công bố hầu hết là của các nghiên cứu sinh người Việt Nam tại các nước. Từ những năm 2015 trở lại đây, đã bắt đầu xuất hiện các nhóm nghiên cứu như: Đại học Bách khoa Hà Nội [Quang, 2018], [Pham Ngoc Hung, 2017], Đại học Tôn Đức Thắng, Đại học Quốc gia TP. Hồ Chí Minh, Đại học Duy Tân – Đà Nẵng,...có nhiều bài công bố trên các tạp chí uy tín quốc tế ISI, Scopus.

Ngoài các nhóm nghiên cứu tại các viện và Lab của các trường đại học lớn, rất nhiều các công trình nghiên cứu độc lập cũng được công bố, trong đó nổi bật là các nghiên cứu về xe ô tô tự hành, robot, nhận dạng hành động con người, phân loại đối tượng, [Van-Dung Hoang, 2018], [Tri-Cong Pham, 2018], [Van-Dung Hoang, 2018],...hỗ trợ trong các lĩnh vực y tế, giao thông, nông nghiệp và quốc phòng.

1.2.2 Quốc tế

Lịch sử phát triển của trí tuệ nhân tạo và học máy đã có từ rất lâu. Từ năm 1950, Alan Turing đã mô phỏng và biểu diễn trí thông minh của máy. Đến năm 1955, John McCarthy, một nhà khoa học máy tính và

khoa học nhận thức của Mỹ đã lần đầu tiên đưa ra khái niệm *Trí tuệ nhân tạo*, mà theo đó có nghĩa là *bộ môn khoa học và kỹ thuật chế tạo máy thông minh*. Một năm sau đó, ông đứng ra tổ chức Hội nghị Dartmouth, hội nghị đầu tiên về chủ đề này. Các chuyên gia đến từ nhiều trường đại học và công ty khác nhau như Carnegie Mellon University, Massachusetts Institute of Technology và IBM đã tham gia hội nghị. Từ đó, khái niệm “Trí tuệ nhân tạo” được sử dụng rộng rãi.

Trải qua nhiều giai đoạn khác nhau, trí tuệ nhân tạo nói chung và lĩnh vực học máy nói riêng vẫn ngày càng một phát triển, nghiên cứu và khai phá nhiều giải thuật quan trọng như: Support vector machine, Random Forest, Neural network, K-mean, Decision tree, Booting, Hog...đây là các giải thuật nền tảng cho sự phát triển của các giải thuật và ứng dụng trong nhận dạng, phân loại đối tượng, xử lý số liệu,...Cùng với sự phát triển của phần cứng máy tính, trong những giai đoạn từ sau những năm 1998, DeepLearning, với mạng nơron nhân chập (Convolution neural network) – một trong những thành phần của Machine learning đã có những bước phát triển vượt bậc, tạo ra nhiều sản phẩm ứng dụng trong đời sống xã hội [Jiao, 2019], [Jiang, 2019], [Chowdhary, 2020], [Zhao, 2019], [Wu, 2020]. Một trong những người đi tiên phong trong lĩnh vực cụ thể này là Yann LeCun. Với LeNet là một trong những mạng CNN lâu đời nổi tiếng nhất được Yann LeCun phát triển vào những năm 1998s. Cấu trúc của LeNet gồm 2 lớp (layer) (Convolution + maxpooling) và 2 lớp kết nối đầy đủ (fully connected layer) và lớp xuất (output) là lớp softmax, độ chính xác trong quá trình nhận dạng lên đến 99%.

Đến năm 2012, Alex Krizhevsky và các cộng sự đã nghiên cứu và cho ra đời mô hình AlexNet [Krizhevsky, 2012]. AlexNet là một mạng CNN đã dành chiến thắng trong cuộc thi ImageNet LSVRC-2012 năm 2012 với tỷ lệ phần trăm lỗi đạt: 15.3% và 26.2%. AlexNet là một mạng CNN với một số lượng tham số rất lớn (60 triệu) so với LeNet.

Tiếp theo sau đó là các mô hình mới lần lượt được đề xuất, cải tiến tỷ lệ phần trăm lỗi, độ phức tạp của mô hình tăng lên, kiến trúc sâu hơn. Có thể kể đến các mô hình như: VggNet năm 2014, GoogleNet năm 2014, MicrosoftResNet năm 2015, Densenet năm 2016, ...Song song với cải tiến về kiến trúc mạng, các mô hình cũng đã cho thực nghiệm

huấn luyện và nhận dạng được hầu hết các đối tượng có trong thực tế với độ chính xác cao, ví dụ với AlexNet có thể nhận dạng và phân loại được 1.000 đối tượng khác nhau.

Ngoài ra, rất nhiều công trình từ các viện nghiên cứu, các trường đại học trên thế giới cũng được công bố, giải quyết từng bài toán cụ thể liên quan đến trí tuệ nhân tạo về robotic, auto vehicle, ... Mỗi lĩnh vực tiếp tục được chia nhỏ theo từng cấp độ khác nhau, chuyên biệt và giải quyết từng trường hợp cụ thể của bài toán thực tế. Ví dụ, đối với bài toán Xe ô tô tự lái có thể phân thành các trường hợp nghiên cứu [Paul, 2016]:

- Bài toán xe ô tô tự lái nhận dạng và phát hiện làn đường
- Bài toán xe ô tô tự lái nhận dạng và phát hiện các đối tượng tham gia giao thông
- Bài toán xe ô tô tự lái nhận dạng biển báo giao thông
- Bài toán xe ô tô tự lái tính toán khoảng cách với các đối tượng
- Bài toán xe ô tô tự lái phát hiện và dự báo hành động người đi bộ
- Bài toán xe ô tô tự lái phát hiện chướng ngại vật,...

Đến thời điểm hiện tại, có thể nói trí tuệ nhân tạo sử dụng mạng học sâu nhân chập đã tiến được một bước khá dài trên con đường “thông minh”, tuy nhiên vẫn đang còn một vách ngăn lớn chưa thể vượt qua, đó là khả năng “tự thông minh”. Các nhà khoa học vẫn đang tích cực nghiên cứu các giải pháp cho phép các hệ thống có khả năng tự học, tự cải thiện trí thông minh như khả năng của con người. Đó cũng là hướng khai phá mà mục tiêu của luận án nghiên cứu hướng đến, góp phần nhỏ bé trên con đường chinh phục những đỉnh cao mới của trí tuệ nhân tạo.

CHƯƠNG 2: NHẬN DẠNG ĐỐI TƯỢNG BẰNG HỌC SÂU

Nội dung của Chương 2 tập trung nghiên cứu và đánh giá bằng thực nghiệm hiệu năng của các mạng CNN trong bài toán nhận dạng đối tượng của thiết bị AV trong ngữ cảnh giao thông. Tập trung vào bài toán dự đoán hành động người đi bộ và bài toán phân loại phương tiện giao thông.

2.1 Bài toán dự đoán hành động người đi bộ

2.1.1 Bài toán

Trong tất cả các đối tượng xuất hiện trong quá trình di chuyển của xe tự lái, việc nhận dạng đối tượng người đi bộ được xem là khó khăn nhất vì độ phức tạp trong nhận dạng, phạm vi di chuyển và quỹ đạo di chuyển của người đi bộ. Chính vì vậy, khả năng dự đoán chính xác hành động của người đi bộ và tốc độ đưa ra các cảnh báo được đặt lên hàng đầu nhằm giải quyết bài toán với độ chính xác cao, đảm bảo sự an toàn cho người đi bộ và phương tiện. Qua nghiên cứu thực tế ta thấy rằng, có nhiều trường hợp người đi bộ khác nhau nhưng tựu chung lại có 3 trường hợp cụ thể là: người đi bộ sang đường (crossing pedestrian), người đi bộ đi dọc theo lề đường (walking pedestrian), người đi bộ đang đứng chờ sang đường (waiting pedestrian); 3 trường hợp này đại diện bao quát cho tất cả các trường hợp tương tác giữa người đi bộ và xe ô tô tự lái có thể xảy ra. Khi người đi bộ di chuyển (hoặc đứng yên) trên đường, thì các đặc trưng được thể hiện rõ giữa tư thế của người đi bộ, vị trí của người đi bộ và các ngữ cảnh của mỗi khung hình (đường, lề đường, mép đường,...); chính vì vậy, việc trích rút đặc trưng từ các hình ảnh có người đi bộ, sử dụng các đặc trưng để huấn luyện dữ liệu phục vụ cho việc dự đoán và nhận dạng hành động của người đi bộ là có thể thực hiện được.

2.1.2 Giải pháp

Giải pháp đề xuất gồm 2 bước: (i) Huấn luyện mô hình phân loại người đi bộ thông qua tư thế người đi bộ sử dụng các đặc trưng được trích xuất từ mô hình CNN (Hình 2.1); (ii) Từ các hình ảnh được trích xuất từ camera của AV khi di chuyển trên đường, thứ tự xử lý như sau: xác định người đi bộ, trích xuất vùng quan tâm (ROI), trích xuất đặc trưng của ROI và dự đoán hành động người đi bộ của ROI (Hình 2.2). Để trích xuất đặc trưng, sử dụng mô hình CNN của AlexNet [Krizhevsky, 2012]. Để xác định người đi bộ, sử dụng giải thuật ACF [Dollar, 2009], [Dollar, 2012], [Dollar, 2014] và để huấn luyện, dự đoán hành động người đi bộ sử dụng mô hình phân loại SVM.



Hình 2.1 Tiến trình trích xuất đặc trưng với mô hình CNN từ tập dữ liệu hình ảnh



Hình 2.2 Tiến trình dự đoán hành động người đi bộ

Thiết bị Camera được sử dụng trong quá trình thực nghiệm có độ phân giải từ 02 Megapixel trở lên, độ phân giải ảnh thu thập được tối thiểu từ 72 dpi.

2.1.3 Thực nghiệm

2.1.3.1 Trích xuất đặc trưng và huấn luyện mô hình phân loại

Quá trình thực nghiệm với 3,000 ảnh sử dụng được trích xuất đặc trưng bởi mô hình CNN. Các hình ảnh này được xử lý (chọn lọc và cắt thành các khung ảnh phù hợp) từ các video thực tế về đường phố được sưu tầm trên Internet. Các đặc trưng này được sử dụng để huấn luyện mô hình phân loại SVM. Bảng 2.2 hiển thị tập dữ liệu trích xuất đặc trưng và huấn luyện.

90% số lượng hình ảnh trong mỗi tập dữ liệu được sử dụng cho huấn luyện và 10% còn lại được sử dụng để đánh giá.

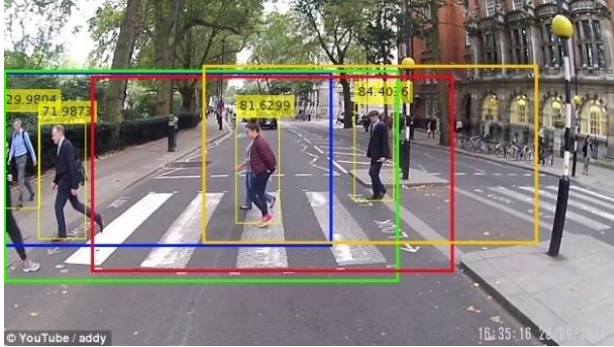
Bảng 2.2. Tập dữ liệu hình ảnh phục vụ trích xuất và huấn luyện đặc trưng

| Lớp | Số lượng | Nhãn |
|-----------------------------|----------|---------------------|
| Người đi bộ sang đường | 1,000 | Pedestrian_crossing |
| Người đi bộ đứng đợi | 1,000 | Pedestrian_waiting |
| Người đi bộ đi dọc lề đường | 1,000 | Pedestrian_walking |

2.1.3.2 Xác định người đi bộ và dự đoán hành động người đi bộ

Với hình ảnh thực tế thu được từ thiết bị camera của AV, sau khi xác định người đi bộ với giải thuật ACF, kết quả xử lý như hình 2.3. Trong trường hợp này, rất nhiều người đi bộ xuất hiện trong khung

hình, chúng tôi tiến hành trích xuất vùng quan tâm (ROI) thành các hình ảnh độc lập để tiến hành dự đoán hành động người đi bộ sử dụng mô hình phân loại SVM (Hình 2.3). Mỗi hình ảnh trong hình 2.3 sẽ được trích xuất đặc trưng và cuối cùng hệ thống sẽ sử dụng mô hình phân loại SVM để dự đoán hành động người đi bộ, đưa ra cảnh báo phù hợp cho AV



Hình 2.3. Xác định người đi bộ và trích xuất vùng quan tâm.

Độ chính xác cao nhất của quá trình phân loại và dự đoán hành động người đi bộ với tập dữ liệu tại Bảng 2.2 được thể hiện tại Bảng 2.3.

Kết quả thực nghiệm với camera thực tế khi AV di chuyển trên đường tối thiểu đạt 82%, cao nhất đạt 97% và tốc độ xử lý của hệ thống đạt 0,6 giây trên một người đi bộ được phát hiện.

Bảng 2.3. Ma trận đối sánh kết quả dự đoán hành động người đi bộ

| | Người đi bộ sang đường | Người đi bộ đứng đợi | Người đi bộ đi dọc lề đường |
|-----------------------------|------------------------|----------------------|-----------------------------|
| Người đi bộ sang đường | 0.9796 | 0.0204 | 0 |
| Người đi bộ đứng đợi | 0.0612 | 0.9286 | 0.0102 |
| Người đi bộ đi dọc lề đường | 0.0102 | 0.0408 | 0.9490 |

2.2 Bài toán nhận dạng phương tiện giao thông

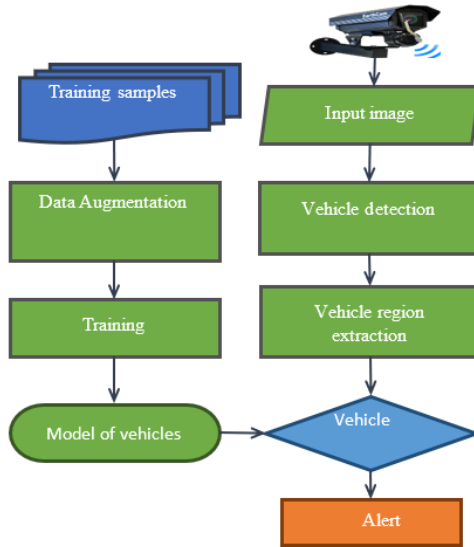
2.2.1 Bài toán

Phát hiện và nhận dạng loại xe trong các hệ thống giám sát giao thông, phân luồng giao thông có tính ứng dụng cao. Cùng với sự phát triển khoa học kỹ thuật thì nhu cầu đi lại và số lượng phương tiện ngày càng tăng. Với số lượng phương tiện giao thông lớn và không ngừng gia tăng như hiện nay, đã làm nảy sinh nhiều vấn đề trong việc kiểm soát và giám sát, phân luồng phương tiện phục vụ quản lý lưu thông được tốt hơn. Để giải quyết vấn đề này, nhu cầu đặt ra là áp dụng các hệ thống tự động để kiểm soát, giám sát và quản lý một cách hiệu quả với độ chính xác cao mà không cần có sự tham gia của con người. Có rất nhiều giải pháp được đưa ra đối với một hệ thống giám sát và hỗ trợ ra quyết định trong quản lý giao thông thông minh (Intelligent Transportation Systems - ITS) như: giải pháp sử dụng các sensor trên các phương tiện giao thông phục vụ đọc dữ liệu, giải pháp sử dụng kết nối đồng bộ các phương tiện giao thông trên nền tảng mạng internet (internet of things),... Tuy nhiên, nhiều giải pháp thực tế vẫn đang dừng lại ở các dạng mô hình lý thuyết do hạn chế trong việc chế tạo thiết bị, băng thông đường truyền và chi phí triển khai lớn. Chính vì vậy, việc đề xuất hệ thống tự động nhận dạng và phân loại phương tiện có độ chính xác cao thông qua phân tích, nhận dạng hình ảnh của mô hình học sâu là hết sức cần thiết.

2.2.2 Giải pháp

Giải pháp đề xuất bắt đầu với việc thu nhận hình ảnh từ camera giám sát trong ITS. Hình ảnh thu thập được sử dụng để nhận ra các đối tượng quan tâm và xác định phương tiện vận chuyển. Có rất nhiều phương pháp để phát hiện phương tiện, nhưng trong khuôn khổ nội dung nghiên cứu, chúng tôi tập trung vào các mô hình nhận dạng thay vì phát hiện phương tiện. Theo mặc định, chúng tôi sử dụng mô hình phân đoạn ngữ nghĩa dựa trên kiến trúc CNN của Segnet [Badrinarayanan, 2017], [He, 2015]. Các phương tiện được phát hiện sau đó sẽ được trích xuất để xác định các khu vực quan tâm (ROI). Khu vực quan tâm là một mẫu phương tiện, tùy thuộc vào phương pháp đề xuất, có thể sử dụng mô hình CNN cũng như kết hợp với tăng dữ liệu để nâng cao độ chính xác. Kết quả nhận dạng được sử dụng trong hệ thống ITS để cảnh báo các phương tiện khi di chuyển vào vạch giới hạn

và tránh bị xử lý vi phạm.



Hình 2.4 Kiến trúc tổng quan

2.2.3 Thực nghiệm

2.2.3.1 Dữ liệu thực nghiệm

Chúng tôi đã tiến hành thử nghiệm trên cơ sở dữ liệu thực tế về các loại xe bao gồm: xe mô tô, ô tô con, xe khách, xe tải được lấy từ các tình huống giao thông thực tế tại thành phố Nha Trang, tỉnh Khánh Hòa, Việt Nam. Hệ thống camera được sử dụng thu thập dữ liệu phía trước hoặc phía sau các phương tiện tham gia giao thông và được thu thập từ các bối cảnh thực tế khác nhau trên các tuyến đường giao thông khác nhau. Tập dữ liệu huấn luyện được chia thành 4 loại xe khác nhau, bao gồm: mô tô, ô tô con, xe khách, xe tải, với 8,558 hình ảnh phương tiện. Tập dữ liệu được chia thành 60% để huấn luyện và 40% còn lại để đánh giá như trong Bảng 2.4.

2.3.2.2 Huấn luyện mô hình CNN

Kết quả thu được sau khi đào tạo mô hình CNN được thể hiện như sau: (i) Tham số bộ lọc: Lớp tích chập đầu tiên sử dụng 64 bộ lọc, có trọng số của bộ lọc được thể hiện trong Hình 2.5:

Bảng 2.4. Dữ liệu huấn luyện

| Danh mục | Số lượng các mẫu | | | Kích thước |
|----------|------------------|------------|----------|------------|
| | Tổng | Huấn luyện | Đánh giá | |
| Xe máy | 2673 | 1604 | 1069 | 128x128 |
| Ô tô con | 2808 | 1685 | 1123 | 128x128 |
| Xe khách | 1640 | 984 | 656 | 128x128 |
| Xe tải | 1437 | 862 | 575 | 128x128 |

Bảng 2.5. Tập dữ liệu sau khi được tăng cường

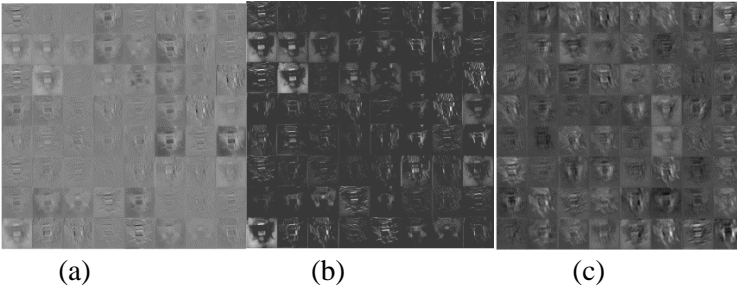
| Danh mục | Số lượng các mẫu |
|----------|------------------|
| Xe máy | 16040 |
| Ô tô con | 16850 |
| Xe khách | 17712 |
| Xe tải | 17240 |



Hình 2.5. Các giá trị trọng số của bộ lọc của lớp tích chập đầu tiên. Lớp này bao gồm 64 bộ lọc kích thước 7×7 , mỗi bộ lọc được kết nối với ba kênh đầu vào hình ảnh RGB.

(ii) Kết quả nhân chập: Hình ảnh mẫu được đưa vào mạng thông qua bộ lọc tích chập và dữ liệu thu được hiển thị các thành phần khác biệt với hình ảnh RGB ban đầu với các kết quả đặc trưng khác nhau, tạo ra

nhiều đặc trưng của phương tiện. Giá trị đầu ra của tập hợp chập có chứa giá trị tiêu cực, giá trị này sẽ được chuẩn hóa bằng điều chỉnh tuyến tính (Hình 2.6).



Hình 2.6. Một số kết quả của tích chập tuyến tính và hiệu chỉnh tuyến tính cho hình ảnh đầu vào là xe mô tô. (a) Đầu ra của 64 lần chập ở lớp chập đầu tiên, (b) Giá trị hiệu chỉnh tuyến tính sau lớp chập đầu tiên, (c) Đầu ra của 64 mẫu ở lớp chập thứ hai

2.3.2.3 Nhận dạng chủng loại phương tiện

Dựa trên kết quả thực nghiệm, ba phương pháp khác nhau đã được đánh giá trên cùng một tập dữ liệu mẫu như trong Bảng 2.4. Các phương pháp bao gồm: (i) Các phương pháp truyền thống của HOG và SVM; (ii) Mạng CNN; (iii) Mạng CNN kết hợp với tăng cường dữ liệu. Độ chính xác của phương pháp HOG và SVM trên tập dữ liệu mẫu là 89,31%. Chi tiết về kết quả nhận dạng được thể hiện trong Bảng 2.6.

Bảng 2.6. Bảng đối sánh kết quả nhận dạng phương tiện giao thông sử dụng HOG và SVM

| | Xe máy | | Xe con | | Xe khách | | Xe tải | |
|----------|-------------|--------------|------------|--------------|------------|--------------|------------|--------------|
| | 1069 | | 1123 | | 656 | | 575 | |
| | #SL | TL(%) | #SL | TL(%) | #SL | TL(%) | #SL | TL(%) |
| Xe máy | 1029 | 97.26 | 16 | 1.53 | 15 | 1.87 | 9 | 1.75 |
| Ô tô con | 25 | 2.36 | 989 | 94.37 | 77 | 9.59 | 32 | 6.23 |
| Xe khách | 1 | 0.09 | 23 | 2.19 | 599 | 74.60 | 33 | 6.42 |
| Xe tải | 3 | 0.28 | 20 | 1.91 | 112 | 13.95 | 440 | 85.60 |

Độ chính xác được đánh giá của phương pháp CNN dựa trên dữ liệu gốc đạt được trung bình 90,10%, như được thể hiện trong Bảng 2.7.

Bảng 2.7. Bảng đối sánh kết quả nhận dạng phương tiện giao thông sử dụng CNN

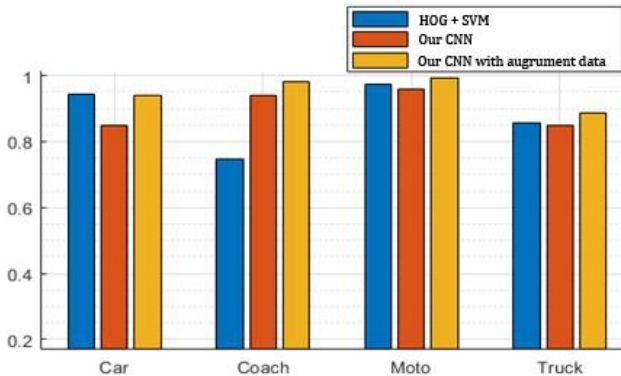
| | Xe máy | | Xe con | | Xe khách | | Xe tải | |
|----------|-------------|--------------|------------|--------------|------------|--------------|------------|--------------|
| | 1069 | | 1123 | | 656 | | 575 | |
| | #SL | TL(%) | #SL | TL(%) | #SL | TL(%) | #SL | TL(%) |
| Xe máy | 1026 | 95.98 | 38 | 3.38 | 1 | 0.15 | 5 | 0.87 |
| Ô tô con | 32 | 2.99 | 953 | 84.86 | 17 | 2.59 | 24 | 4.17 |
| Xe khách | 6 | 0.56 | 104 | 9.26 | 617 | 94.05 | 58 | 10.09 |
| Xe tải | 5 | 0.47 | 28 | 2.49 | 21 | 3.20 | 488 | 84.87 |

Đánh giá độ chính xác của phương pháp CNN dựa trên sự gia tăng dữ liệu đạt được trung bình là 95,59%, được thể hiện trong Bảng 2.8.

Bảng 2.8. Bảng đối sánh kết quả nhận dạng phương tiện giao thông sử dụng CNN và tăng cường dữ liệu

| | Xe máy | | Xe con | | Xe khách | | Xe tải | |
|----------|-------------|--------------|-------------|--------------|------------|--------------|------------|--------------|
| | 1069 | | 1123 | | 656 | | 575 | |
| | #SL | TL(%) | #SL | TL(%) | #SL | TL(%) | #SL | TL(%) |
| Xe máy | 1060 | 99.16 | 11 | 0.98 | 0 | 0 | 1 | 0.17 |
| Ô tô con | 5 | 0.47 | 1057 | 94.12 | 8 | 1.22 | 13 | 2.26 |
| Xe khách | 0 | 0 | 41 | 3.65 | 645 | 98.32 | 51 | 8.87 |
| Xe tải | 4 | 0.37 | 14 | 1.25 | 3 | 0.46 | 510 | 88.70 |

Trong nghiên cứu này, chúng tôi cũng đánh giá mô hình CNN được đề xuất theo một cách tiếp cận truyền thống khác dựa trên bộ mô tả tính năng HOG và bộ phân loại SVM. Kết quả so sánh được thể hiện trong Hình 2.7.



Hình 2.7. So sánh kết quả giữa các phương pháp HOG+SVM, mô hình CNN và CNN với tăng cường dữ liệu

CHƯƠNG 3. PHÁT TRIỂN KỸ THUẬT HỌC THÍCH ỨNG TRONG NHẬN DẠNG ĐỐI TƯỢNG

3.1. Bài toán và giải pháp tổng quan

Nội dung nghiên cứu của Chương là đề xuất một giải pháp dựa trên học thích ứng bằng mô hình CNN. Phương pháp đề xuất sẽ tự động cập nhật mô hình nhận dạng mới thay thế mô hình cũ bằng cách thu thập các tập dữ liệu trực tiếp trong quá trình hoạt động bình thường của một ADAS, tiến hành huấn luyện, so sánh độ chính xác và cập nhật mô hình mới. Nhiệm vụ cập nhật dữ liệu tập trung vào các mẫu dữ liệu ít giống với các mẫu được huấn luyện trước đó. Mục đích của giải pháp này là cập nhật mô hình bởi một mô hình mới thích ứng hơn, đạt được độ chính xác cao hơn. Trong phương pháp học tập thích ứng, hệ thống nhận dạng có khả năng tự học và bổ sung dữ liệu, mà không cần các chuyên gia gán nhãn dữ liệu. Đặc biệt, với công nghệ lưu trữ trực tuyến ngày càng phát triển, hạ tầng và giải pháp đường truyền dữ liệu trên các nền tảng mới (5G, Cloud data,...) thì việc lưu trữ, cập nhật dữ liệu trực tuyến sẽ giải quyết các khó khăn của mô hình đề xuất. Giải pháp đề xuất bao gồm 05 giai đoạn chính:

- (1) Phát hiện và nhận biết các đối tượng với độ tin cậy thấp;
- (2) Theo dõi các đối tượng trong n khung hình trong các tiến trình tiếp theo để xác định chúng có phải là đối tượng quan tâm hay không.
- (3) Trường hợp đối tượng được nhận dạng với độ chắc chắn cao: gán nhãn Positive (cùng loại đối tượng) cho các mẫu dữ liệu tương ứng với độ chính xác thấp được theo dõi (tracking) trong quá trình trước đó. Trong trường hợp các đối tượng được xác định không phải là đối tượng quan tâm, tiến hành dán nhãn là Negative cho tất cả các mẫu được theo dõi trong n khung hình trước đó;
- (4) Khởi tạo một tập dữ liệu huấn luyện dựa trên sự kết hợp có chọn lọc của dữ liệu được huấn luyện trước đó và dữ liệu mới.
- (5) Huấn luyện lại và cập nhật mô hình nếu so sánh có độ chính xác cao hơn.

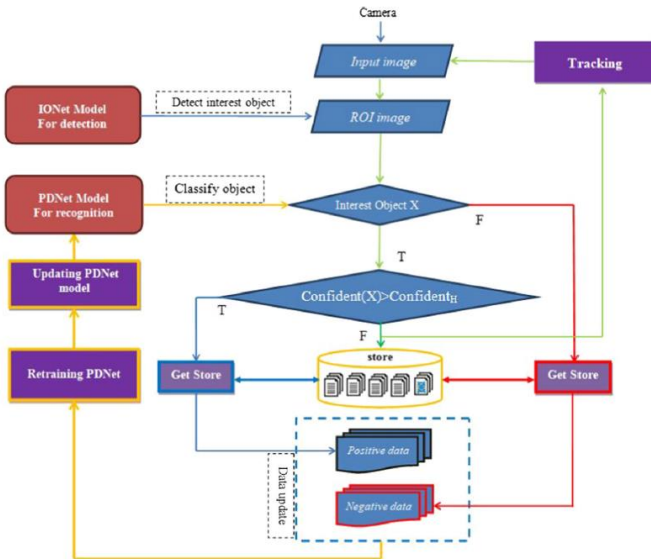
Tiến hành các thí nghiệm để so sánh kết quả của mô hình đề xuất mô hình PDNet và một số phương pháp hiện đại như AlexNet và Vgg. Các kết quả thử nghiệm chứng minh rằng phương pháp được đề xuất

cung cấp độ chính xác cao hơn khi mô hình tự học theo thời gian. Mặt khác, giải pháp học tập thích ứng được đề xuất có thể áp dụng cho các mô hình nhận dạng truyền thống như mô hình AlexNet và Vgg để cải thiện độ chính xác.

3.2 Kiến trúc tổng quan của hệ thống

Ý tưởng chung về học tập thích ứng cho mô hình nhận dạng sử dụng công nghệ CNN được minh họa trong Hình 3.1. Hệ thống nhận dạng có thể được áp dụng cho các loại đối tượng khác nhau.

Tuy nhiên, để thuận tiện trong việc phân tích phương pháp đề xuất, chúng tôi chỉ áp dụng cho các bài toán phân loại phương tiện và biển báo giao thông để minh họa cho giải pháp đề xuất. Có hai mô hình CNN được sử dụng trong phương pháp này, mô hình IONet để phát hiện phương tiện và biển báo giao thông và mô hình PDNet để xác định độ tin cậy và nhận dạng đối tượng.



Hình 3.1 Mô hình tổng quát của hệ thống

Mô tả vấn đề: Giả sử rằng chúng ta đã huấn luyện hai mô hình CNN ban đầu, IONet và PDNet. Trong suốt hành trình trên đường, ADAS sử dụng các mô hình để nhận biết phương tiện, biển báo giao thông và đưa ra quyết định phù hợp. Tuy nhiên, trong quá trình xử lý và nhận dạng, có một số trường hợp hệ thống nhận diện phương tiện và biển báo giao

thông với độ tin cậy thấp. Tình huống này xảy ra khi hệ thống gặp phải dữ liệu không giống với tập dữ liệu được đào tạo hoặc thông tin không đầy đủ. Dữ liệu không giống với dữ liệu gốc và bị nhiễu thường do khoảng cách xa, phương tiện và biển báo giao thông bị che khuất bởi các vật thể khác, biển báo bị cong vênh hoặc mờ, phương tiện di chuyển trong điều kiện thiếu ánh sáng, mưa, tuyết, độ rung do chuyển động,... Đây là thời điểm để bắt đầu học tập thích ứng. Hệ thống sẽ lưu trữ hình ảnh có điểm tin cậy thấp (IO) và tiếp tục theo dõi (tracking) đối tượng. Quá trình theo dõi nhằm xác định các trường hợp: (i) Mất đối tượng; (ii) Không phải đối tượng; (iii) Xác định được đối tượng. Khi lượng dữ liệu trong tập Positive Data và Negative Data đủ lớn, tác vụ huấn luyện lại mô hình CNN được xử lý. Mô hình được huấn luyện mới được chọn và so sánh với các mô hình đã được huấn luyện trước đó, mô hình tốt nhất được sử dụng để cập nhật mô hình nhận dạng của hệ thống. Quá trình học tập thích ứng diễn ra liên tục trong suốt quá trình làm việc của ADAS. Sau khi được cập nhật, mô hình CNN mới có thể nhận dạng các đối tượng chính xác hơn.

3.3 Kết quả thực nghiệm

Mô hình được thực nghiệm trên hệ thống thiết bị có cấu hình tương đối thấp, tương ứng với cấu hình có thể trang bị được cho các phương tiện xe tự lái trong thực tế, được thể hiện trong bảng 3.1.

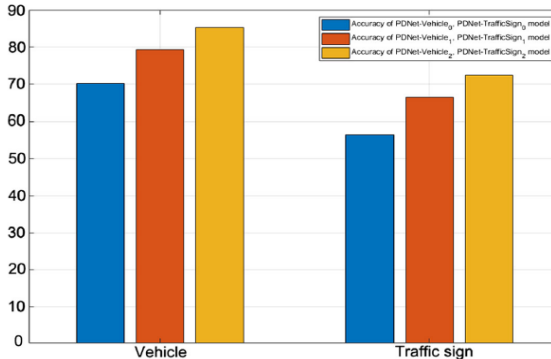
Bảng 3.1 Cấu hình thiết bị

| Thiết bị | Cấu hình |
|----------|-------------------|
| CPU | I3 3.6 GHz |
| GPU | Geforce 1060 6 Gb |
| RAM | 16 Gb |
| HDD | SSD 160 Gb |

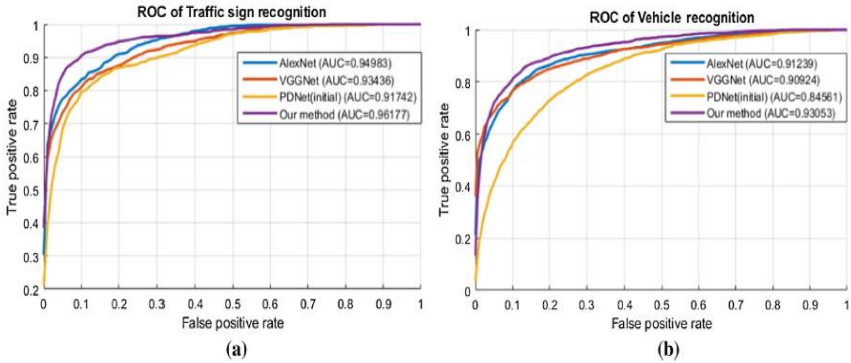
Kết quả thực nghiệm cho thấy có nhiều thay đổi về độ chính xác của mô hình khi so sánh kết quả nhận dạng phương tiện và biển báo giao thông của mô hình ban đầu ($PDNet-Vehicle_0$, $PDNet-TrafficSign_0$) và mô hình được đào tạo lại ($PDNet-Vehicle_1$, $PDNet-TrafficSign_1$ và $PDNet-Vehicle_2$, $PDNet-TrafficSign_2$) trong quá trình huấn luyện lại trên tập dữ liệu Retrain (70% dữ liệu được sử dụng lại từ mô hình huấn

luyện trước và 30% dữ liệu thu nhận được từ quá trình xác định độ tin cậy bằng theo vết đối tượng), được hiển thị trong Hình 3.2. Kết quả thực nghiệm cũng chứng minh tính hiệu quả khi so sánh kết quả thử nghiệm của các phương pháp được đề xuất của chúng tôi và một số mô hình lý tưởng của phương pháp học sâu, chẳng hạn như Alexnet và Vgg. Kết quả ban đầu cho thấy mô hình PDNet mang lại độ chính xác thấp hơn so với mô hình AlexNet và Vgg. Tuy nhiên, sau quá trình học tập thích ứng, mô hình PDNet mang lại độ chính xác cao hơn so với mô hình AlexNet và Vgg ban đầu (Hình 3.3). Tốc độ xử lý của mô hình AlexNet và Vgg chậm hơn so với mô hình PDNet, do mô hình PDNet có kích thước hình ảnh đầu vào nhỏ hơn (64×64), trong khi mô hình AlexNet và Vgg có kích thước hình ảnh lớn (227×227 và 224×224).

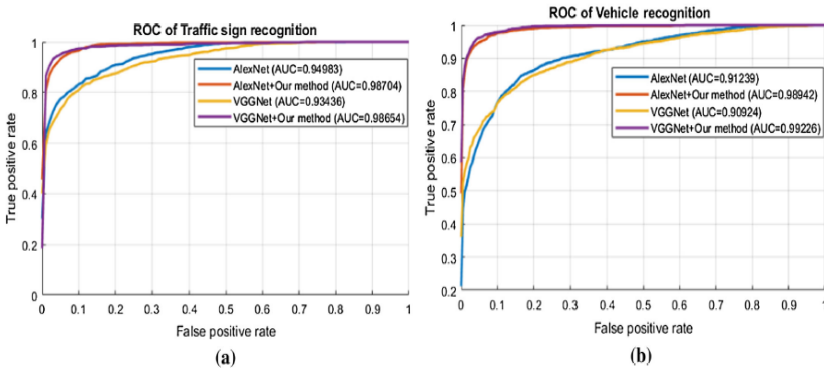
Phương pháp học tập thích ứng được đề xuất của chúng tôi cũng được áp dụng cho các mô hình AlexNet, Vgg, kết quả cho thấy các mô hình *Adap-AlexNet₁*, *Adap-AlexNet₂* và *Adap-Vgg₁*, *Adap-Vgg₂* (sau khi được đào tạo lại) mang lại độ chính xác cao hơn so với *AlexNet₀* và *Vgg₀* ban đầu (Hình 3.4). Kết quả được minh họa trong đồ thị Hình 3.3, Hình 3.4 cho thấy bất kể mô hình nào được sử dụng để đào tạo, quá trình học tập thích ứng sẽ cải thiện mô hình ban đầu đó để mang lại độ chính xác tiệm cận tối đa theo thời gian.



Hình 3.2 So sánh độ chính xác sau khi huấn luyện lại mô hình nhận dạng Phương tiện và Biển báo giao thông



Hình 3.3 So sánh kết quả của cách tiếp cận đề xuất của chúng tôi và các phương pháp khác



Hình 3.4 So sánh kết quả bằng cách áp dụng phương pháp học thích ứng với mô hình AlexNet và Vgg

CHƯƠNG 4. TỐI ƯU HÓA BỘ SIÊU THAM SỐ TRONG HỌC THÍCH ỨNG

4.1. Bài toán tối ưu siêu tham số

Trong giai đoạn hiện nay, nghiên cứu về trí tuệ nhân tạo nói chung và các hệ thống tự động nói riêng đang tập trung vào việc nghiên cứu, xây dựng các giải pháp tối ưu hóa các mô hình học máy và các tham số của các mô hình này. Trong đó, nổi lên hai xu hướng chính là lựa chọn mô hình - Model selection (CNN, ANN, LSTM, Segment,...) và lựa chọn

tham số của mô hình - Model hyperparametersselection. Tuy nhiên, trong khuôn khổ của luận văn nghiên cứu, chúng tôi không đi sâu vào các phương pháp lựa chọn mô hình huấn luyện mà tập trung vào các mô hình CNN cụ thể với việc tối ưu hóa các tham số của các mô hình này.

Trong các nội dung nghiên cứu của Chương 2 và Chương 3, ta nhận thấy tất cả các mô hình CNN hiện tại được huấn luyện với 02 loại tham số cơ bản:

Tham số huấn luyện (Optimization hyperparameters)

- *Learning rate:*
- *Mini-Batch Size*
- *Number of Epochs*

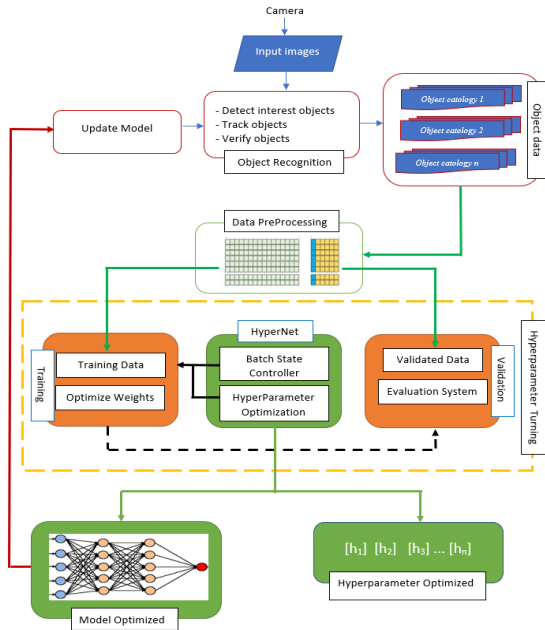
Tham số mô hình (Model hyperparameters)

- *Number of hidden units*
- *First hidden layer*
- *Number of layers*

Tham số mô hình quyết định sự thay đổi của mô hình CNN và trong quá trình huấn luyện của một mô hình CNN hầu như không có sự thay đổi. Vì vậy, để giải quyết bài toán thích ứng về giải thuật, tham số ta đi tìm các giải pháp tối ưu hóa tham số huấn luyện.

4.2. Tổng quan giải pháp

Mô hình huấn luyện và giải pháp đề xuất được kế thừa từ mô hình được đề xuất trong Chương 3. Phương pháp đề xuất này tạo ra một đóng góp mới bằng cách thay đổi khối chức năng Retrained PDNet được minh họa trong Hình 3.1. Chức năng HyperNet được thêm vào để cho phép tìm kiếm siêu tham số cho mô hình huấn luyện, giúp cải thiện hiệu quả độ chính xác trong nhận dạng. Siêu tham số thích hợp được tìm kiếm tự động bằng phương pháp Bayes. Phương pháp tổng thể được đề xuất được trình bày trong Hình 4.1

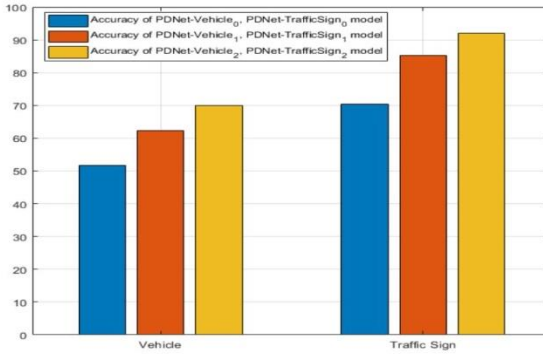


Hình 4.1. Mô hình đề xuất tổng thể

Ta thấy rằng, dữ liệu được thu thập trong quá trình chuyển động của ADAS liên tục thay đổi và làm mới. Tuy nhiên, các tham số về cấu trúc của mô hình CNN và tham số trong quá trình huấn luyện lại mô hình CNN không có sự thay đổi. Do đó về lý thuyết, cần phải thay đổi kiến trúc của mô hình CNN và các tham số huấn luyện để đảm bảo rằng chúng phù hợp với từng tập dữ liệu mới. Tuy nhiên, giải pháp học thích ứng được đề xuất tập trung vào huấn luyện lại mô hình nhận dạng, kế thừa sự 'thông minh' của nó từ mô hình trước đó. Do đó, việc tìm kiếm và thay đổi kiến trúc của mô hình CNN không được đề xuất. Giải pháp sẽ tập trung vào việc tìm kiếm các siêu tham số quan trọng của quá trình huấn luyện, phù hợp với từng tập dữ liệu mới được thu nhận.

4.3 Kết quả thực nghiệm

Có nhiều thay đổi về độ chính xác của mô hình khi so sánh kết quả nhận dạng phương tiện và biển báo giao thông của mô hình ban đầu ($PDNet-Vehicle_0$ và $PDNet-TrafficSign_0$) và mô hình tối ưu ($PDNet-Vehicle_1$, $PDNet-TrafficSign_1$ và $PDNet-Vehicle_2$, $PDNet-TrafficSign_2$) được hiển thị trong Hình 4.2.



Hình. 4.2. So sánh độ chính xác kết quả nhận dạng của mô hình PDNet-Vehicle và PDNet-TrafficSign

Table 4. 1 Results of proposed methods compared to those of the Chapter 3

| Models | Our method (%) | Previous method (%) |
|--|----------------|---------------------|
| PDNet-Vehicle ₀ (initial model) | 51.77 | 51.77 |
| PDNet-Vehicle ₁ | 62.30 | 60.58 |
| PDNet-Vehicle ₂ | 69.98 | 68.41 |
| PDNet-TrafficSign ₀ (initial model) | 70.46 | 70.46 |
| PDNet-TrafficSign ₁ | 85.19 | 84.93 |
| PDNet-TrafficSign ₂ | 92.90 | 90.36 |
| AlexNet-Vehicle ₀ (initial model) | 66.14 | 66.14 |
| AlexNet-Vehicle ₁ | 88.24 | 86.61 |
| AlexNet-Vehicle ₂ | 90.75 | 90.40 |
| AlexNet-TrafficSign ₀ (initial model) | 67.05 | 67.05 |
| AlexNet-TrafficSign ₁ | 88.78 | 87.73 |
| AlexNet-TrafficSign ₂ | 93.51 | 92.55 |
| Vgg-Vehicle ₀ (initial model) | 71.46 | 71.46 |
| Vgg-Vehicle ₁ | 93.11 | 92.42 |
| Vgg-Vehicle ₂ | 94.78 | 94.14 |
| Vgg-TrafficSign ₀ (initial model) | 70.46 | 70.46 |
| Vgg-TrafficSign ₁ | 95.27 | 94.74 |
| Vgg-TrafficSign ₂ | 95.53 | 94.74 |

Đặc biệt, việc áp dụng thuật toán Bayes để tìm kiếm siêu tham số và mô hình đã làm cho độ chính xác trên các mô hình PDNet và AlexNet, Vgg cao hơn so với các mô hình tương tự đã nêu trong chương 3 khi được đánh giá trên cùng một tập dữ liệu. Kết quả so sánh được thể hiện trong Bảng 4.1.

KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

1. Kết luận

Kết quả nghiên cứu được thể hiện trong từng chương của luận án, đã được chứng minh và khẳng định thông qua các công trình nghiên cứu được công bố trên các các hội thảo, tạp chí uy tín trong nước và quốc tế. Luận án đã hoàn thành cơ bản các nội dung nghiên cứu theo mục tiêu đề ra. Trong đó, nổi bật là các đóng góp:

(1) Tìm hiểu và tổng quát hóa vai trò nền tảng không thể thiếu của các giải thuật học máy truyền thống. Các nghiên cứu trong nước và quốc tế trong thời gian gần đây về trí tuệ nhân tạo, học máy, các kỹ thuật nhận dạng đối tượng bằng học sâu và kỹ thuật học thích ứng.

(2) Nội dung nghiên cứu của chương 2 đã chứng minh các kỹ thuật cơ bản của mạng học sâu. Thông qua các thực nghiệm mô phỏng của thiết bị ADAS khi tham gia giao thông, đã cho thấy khả năng nhận dạng hiệu quả của các mô hình CNN khi được huấn luyện. Kết quả nghiên cứu của chương là tiền đề để xây dựng một mô hình tổng thể của một hệ thống ADAS có khả năng tự học và thông minh hơn theo thời gian.

(3) Đóng góp nổi bật của luận án chính là đề xuất một mô hình tổng thể cho giải pháp học thích ứng. Thông qua hoạt động của mô hình ADAS, đã chứng minh khả năng tự học, tự nhận dạng của một hệ thống autorobot theo mô phỏng hoạt động của bộ não con người. Cùng với sự thích ứng và cập nhật dữ liệu thực tế một cách tự động, giải pháp đề xuất còn cho phép hệ thống thay đổi, thích ứng với bộ siêu tham số (hyperparameter) huấn luyện phù hợp với bộ dữ liệu thu nhận được. Chính nhờ sự kết hợp đó, đã tạo nên một mô hình khá hoàn chỉnh cho giải pháp học thích ứng của các hệ thống autorobot trong tương lai.

(4) Thông qua quá trình thực nghiệm các nội dung nghiên cứu, tác giả đã thu thập và xây dựng được bộ dữ liệu của nhiều đối tượng khác nhau như: tập dữ liệu thực tế người đi bộ, tập dữ liệu tư thế người đi bộ, tập dữ liệu một số biển báo giao thông, tập dữ liệu về phương tiện giao thông. Các tập dữ liệu này được tác giả thu thập trực tiếp trong quá

trình di chuyển thực tế của xe ô tô hoặc được thu thập từ các video thực tế từ internet.

Do các trường hợp dữ liệu phục vụ quá trình thực nghiệm không có sẵn (bao gồm các tập dữ liệu nổi tiếng đã được công bố) vì vậy các tập dữ liệu ảnh trong luận án được tác giả tự thu thập trực tiếp trong quá trình di chuyển thực tế của xe ô tô hoặc được thu thập từ các video thực tế từ internet

(5) Tuy nhiên, mặc dù đã đạt được nhiều kết quả đáng khích lệ nhưng vẫn còn một số vấn đề cần giải quyết nhằm hoàn thiện và chứng minh tính hiệu quả của mô hình học thích ứng:

- Số lượng các đối tượng thực nghiệm còn ít, chưa bao quát hết nhiều trường hợp đối tượng khác. Số lượng ảnh trong các tập dữ liệu chưa nhiều vì vậy độ chính xác trong mô hình nhận dạng còn thấp.

- Một số giá trị tham số phục vụ quá trình huấn luyện được đề xuất mặc định, chưa được chứng minh giá trị mang lại hiệu quả cao nhất (Ví dụ: giá trị số lượng ảnh N khi bắt đầu quá trình huấn luyện lại mô hình, tỷ lệ % dữ liệu ảnh của tập dữ liệu trước được sử dụng lại cho huấn luyện mô hình sau,...)

- Số lượng các tham số được đưa vào để tối ưu hóa còn ít (06 tham số), và miền giá trị bộ siêu tham số chỉ mang tính ước lượng thông qua thực nghiệm, chưa được chứng minh cụ thể khoảng giá trị cần tìm kiếm.

2. Hướng phát triển

Mô hình đề xuất đã thể hiện được giải pháp trong học thích ứng của các thiết bị ADAS, tuy nhiên ta có thể nhận thấy rằng khả năng tiềm tàng trong việc tiếp tục nghiên cứu và phát triển theo các hướng:

- Mở rộng các đối tượng nhận dạng nhằm đa dạng khả năng của hệ thống ADAS hoặc phát triển thành một hệ thống autorobot hoàn thiện có khả năng học thích ứng mọi đối tượng.

- Đánh giá và tìm kiếm các giá trị thích hợp thay thế các giá trị cố định trong quá trình huấn luyện mô hình học thích ứng. Mở rộng miền tham số tìm kiếm nhằm gia tăng khả năng lựa chọn các tham số phù hợp cho quá trình huấn luyện lại mô hình tương ứng với tập dữ liệu mới. Đồng thời, nghiên cứu phương án giảm độ phức tạp trong quá

trình tìm kiếm siêu tham số của mô hình đề xuất, giảm thời gian và nâng cao hiệu suất xử lý.

- Trong mô hình đề xuất, quá trình học thích ứng diễn ra liên tục sẽ làm cho tập dữ liệu huấn luyện ngày càng tăng nhanh về số lượng. Vì vậy, vấn đề đặt ra là xây dựng giải pháp tinh gọn, chọn lọc dữ liệu trong tập dữ liệu huấn luyện nhằm loại bỏ các mẫu dữ liệu dễ, ưu tiên học các mẫu dữ liệu khó, giúp cho mô hình giảm thời gian huấn luyện, nâng cao độ chính xác và chất lượng của quá trình học thích ứng.

- Xây dựng một bộ dữ liệu hoàn chỉnh, quy mô hơn với nhiều loại đối tượng phục vụ quá trình huấn luyện ban đầu của mô hình học thích ứng.

DANH SÁCH CÁC CÔNG TRÌNH KHOA HỌC ĐÃ CÔNG BỐ LIÊN QUAN ĐẾN LUẬN ÁN

1. Các công bố chính

- 1.1 "Pedestrian action prediction based on deep features extraction of human posture and traffic scene," in *Asian Conference on Intelligent Information and Database Systems*, 2018, pp. 563-572.
- 1.2 "Pedestrian activity prediction based on semantic segmentation and hybrid of machines," *Journal of Computer Science and Cybernetics*, vol. 34, pp. 113-125, 2018.
- 1.3 "Vehicle Categorical Recognition for Traffic Monitoring in Intelligent Transportation Systems," in *Asian Conference on Intelligent Information and Database Systems*, 2019, pp. 670-679.
- 1.4 "Adaptive Learning Based on Tracking and ReIdentifying Objects Using Convolutional Neural Network," *Neural Processing Letters*, vol. 50, pp. 263-282, 2019.
- 1.5 "Hyperparameter optimization for improving recognition efficiency of an Adaptive learning system", *IEEE Access*, vol. 08, pp.160569 - 160580, 2020.

2. Các công bố liên quan

- 2.1 "A solution based on combination of RFID tags and facial recognition for monitoring systems," in *2018 5th NAFOSTED Conference on Information and Computer Science (NICS)*, 2018, pp. 384-387.
- 2.2 "Personal Identification Based on Deep Learning Technique Using Facial Images for Intelligent Surveillance Systems," *International Journal of Machine Learning and Computing*, vol. 9, 2019.
- 2.3 "Meta-analysis of computational methods for breast cancer classification," *International Journal of Intelligent Information and Database Systems*, vol. 13, 2020.
- 2.4 "Deep Feature Extraction for Panoramic Image Stitching," in *Asian Conference on Intelligent Information and Database Systems*, 2020, pp. 141-151.