MINISTRY OF EDUCATION AND TRAINING
DUY TAN UNIVERSITY

ADAPTIVE LEARNING SOLUTION BASED ON DEEP LEARNING FOR TRAFFIC OBJECT RECOGNITION

Major: Computer Science
Code: 9480101

SUMMARY OF DOCTORAL THESIS OF COMPUTER SCIENCE

DA NANG - 2022

The research has been completed at Duy Tan University

Reviewer 1: ....

Reviewer 2: ....

Reviewer 3: ....

Thesis defense will be conducted at the meeting of the Panel of Thesis assessment in ....

....

at..

Thesis can be research at:

- National Library of Vietnam
- Document Information Center of Duy Tan University

INTRODUCTION

Artificial intelligence (AI) is intelligence demonstrated by an artificial system. Artificial intelligence is everywhere today such as office applications, automatic answering systems, intelligent traffic management, smart home management, etc. Since the Computer hardware systems became increasingly capable, artificial intelligence has made great progress, applied more widely in all fields of life and society.

Artificial intelligence focuses on developing algorithms and applications that support human in decision making or self-decision making in the process of data identifying and acquiring. Object detection, Object action recognition and Human action recognition are one of the research targeted directions such as security surveillance systems, security, manual remote control systems, blind assist systems, sports data analysis systems, automated robots, self-driving cars [Hariyono, 2017], [Dollar, 2012], Stewart, 2016,], [Van-Dung Hoang, 2012] and so on. There have been many studies proposing many different solutions to artificial intelligence development such as heuristic algorithm, evolution algorithm, Support Vector Machine algorithm, Hidden Markov Model algorithm, expert method, neural network method [Yu, 2017],[Dalal, 2005], [Mittal, 2012], etc. Traditional solutions, yet all require human intervention and huge amounts of data to analyze and store but low accuracy and limited identification cases.

To overcome those shortcomings, machine learning with focusing on Deep Learning Method (Deep Learning) is now being applied in artificial intelligence in terms of object detection and action recognition.

Deep Learning has been a hotly debated AI topic. As a small category of machine learning, Deep Learning focuses on solving issues related to artificial neural networks in order to upgrade technologies such as voice recognition, image recognition and natural language processing. In just a few years, Deep

Learning has promoted progress in a variety of fields which are used to be very difficult to artificial intelligence researchers such as Object Perception, Machine Translation, voice recognition, etc.

However, despite of the fact that issues related to AI were solved, Deep Learning has still remained limitations that need to be settled.

- Firstly, to create a system capable of identifying a variety of objects, a huge amount of input data is required by Deep Learning to enable computers to learn. This process takes time with assistance of an extremely large processor which can be only processed by a large server system.

- Secondly, Deep Learning is still unable to recognize complex things like common social contacts. It, also, has trouble with detecting similar things because of having no technology good enough helping artificial intelligence to draw those recognition logically. Besides, integration of abstract knowledge into machine learning systems seem to be the challenging issues, such as information about what object is, what it is used for, how people use it, so on. In other words, machine learning has not acquired the usual knowledge like human yet.

The question is "How can a machine learning system learn the knowledge, select and update appropriate knowledge and then build a binding, stringed data set like human by itself?". Research on **Adaptive Learning** [Chandra, 2016], [Chatzilari, 2016], [Wang, 2017], [Zhang, 2017], [Liu, 2017], [Zhang, 2016] can be a solution to improve Deep Learning' limitations, exploring issues that Deep Leaning has not been able to do.

A comprehensive Adaptive Learning model will make an auto robot system being capable of self-learning and self-intelligence that emulate the way the human brain work. Under the device's operation, the intelligence of the system will increase over time. Accordingly, appropriate data will be automatically selected by the system with its retraining of the model and replacing of the old model

The proposed Adaptive Learning model could be promisingly applied in many different Auto Robot systems. Yet, in this research of a doctoral thesis, studying and experiment will be conducted on self-driving vehicles to simulate an operation process of an auto robot. Recognition objects of self-driving vehicles include objects in traffic such as other vehicles (motorcycles, cars, trucks, passenger cars, etc.), pedestrians, traffic signs, roadbed, roadside, etc.

CHAPTER 1. OVERVIEW OF ARTIFICIAL INTELLIGENCE

1.1 Artificial intelligence

There have been many different definitions of artificial intelligence, or AI in the world, specifically:

• By popular, artificial intelligence is is intelligence demonstrated by any artificial system. The term is often used to refer to computers with unspecified purpose and the science of theories and applications of artificial intelligence.

• According to Bellman (1978), artificial intelligence is the automation of activities that we associate with human thinking, activities such as decision-making, problem solving, learning, etc.

• Rich and Knight (1991): “Artificial intelligence is the study of how to make computers do things at which, at the moment, people are better".

Historically, each definition has its own right, but for simplicity we can get the idea of artificial intelligence as a computer science. It was built on a solid theoretical foundation and can be applied to automation of the intelligent behavior by computers. It makes computers acquire the human intelligence such as thinking, decision-making, problem solving, learning and self-adapting..

1.2 Domestic and international research

1.2.1 Domestic research

In Vietnam, from the 1990s to the early years of the 20th century, there were participation of the well-known researchers

Assoc. Prof. Ngo Quoc Tao, Assoc. Dr. Do Nang Toan, Assoc. Dr. Luong Chi Mai, etc in the field of AI research, especially image processing and recognition. Their research works include handwriting recognition [Dang Ngoc Duc, 2003], [P. A. Phuong, 2008], Vietnamese handwriting [Phạm Anh Phương, 2009], [Ngô Quốc Tạo, 2004], speech recognition, human face detection [Lam Thanh Hien , 2012], [Lê Thanh Hà, 2006 #51], [Đỗ Năng Toàn, 2011], simulation of the human body [Pham Ngoc Hung, 2017], etc. Most of the researches and publications exploit classic algorithms such as SVM, RandomForest, hidden Markov models, artificial neutral networks, and so on. These researches are considered as significant foundations for students and graduate students' reference. Along with the publication of researches, many publications on image processing and object recognition were also published.

After the first decade of the 20th century, AI growth, along with computer hardware, enables the fields of machine learning and object recognition to make advance. In Vietnam, however, in the first time, researches on Artificial neural networks and Convolution Neural Networks were still very primitive with no domestic research on this specific field. The researches and publications mostly come from oversea Vietnamese PhD students. From 2015 up to now, there have been many articles published on the international journal ISI, Scopus. These articles came from research groups such as Hanoi University of Technology [Quang, 2018], [Pham Ngoc Hung, 2017], Ton Duc Thang University, National University of Ho Chi Minh City, Duy Tan University - Da Nang, etc. Besides to the research groups of the institutes and labs, many independent research works have also been published, including researches assisting the fields of health, transport, agriculture and national defense such as autonomous cars, robots, and human action recognition, classifications, [Van-Dung Hoang, 2018], [Tri-Cong Pham, 2018], [Van-Dung Hoang, 2018], etc.

1.2.2 International research

The AI history and machine learning has gone through many phrases. The intelligence of the machine has been simulated and demonstrated by Alan Turing Since 1950. By 1955, John McCarthy, an American computer scientist and cognitive scientist, first coined the term “Artificial Intelligence”, meaning the science subject and intelligent computer engineering. One year later, he hosted the Dartmouth Conference, the first conference on this topic with the participation of experts from various universities and companies such as Carnegie Mellon University, Massachusetts Institute of Technology and IBM. Since then, the term "artificial intelligence" has been widely used.

Through many different stages, AI in general and the field of machine learning in particular are still growing, continuously fulfill their task of exploring many important algorithms such as Support vector machine, Random Forest, Neural network, K-mean, Decision tree, Booting, Hog, and so on. These algorithms are the fundamental for the growth of algorithms and applications in recognition, object classification, data processing, and so on. Along with the growth of computer hardware, the 1998s forward, Deep Learning and Convolution neural network which is one of the components of Machine learning, has made great progress with many application in life [Jiao, 2019], [Jiang, 2019], [Chowdhary, 2020], [Zhao, 2019], [Wu, 2020]. Yann LeCun is one of the pioneers in this particular field. LeNet, one of the most famous CNN networks, was developed by Yann LeCun in the 1998s. The structure of LeNet consists of 2 layers (Convolution + maxpooling) and 2 layers fully connected layer and the output (softmax layer) with the recognition accuracy up to 99%.

By 2012, AlexNet model [Krizhevsky, 2012] was introduced by Alex Krizhevsky and his colleagues. The AlexNet with a large margin (15.3% VS 26.2% error rates) is a CNN network that won the ImageNet LSVRC-2012 contest in 2012. AlexNet is a CNN training network with a very large number of parameters (60 million) compared to LeNet.

This was followed by new models proposed in turn, decreasing the error percentage, increasing the model's complexity with a deep architecture. The proposed models include VggNet 2014, GoogleNet 2014, MicrosoftResNet 2015, Densenet 2016, etc. In parallel with the improvement of network architecture, experimental training and recognition to almost all objects in reality were conducted by the models with high accuracy. For example, AlexNet is capable of identifying and classifying 1,000 different objects.

In addition, many works from research institutes and universities in the world have been published that proposed solutions to each AI specific problem on robotic, auto vehicles, etc. Each field, then, continue to be broken down by different levels for solving. For instant, the problem of self-driving cars can be classified into the following cases [Paul, 2016]:

- The lane - recognition problem for self-driving cars
- The on – road object recognition problem for self-driving cars
- The traffic sign recognition problem for self-driving cars
- The distance measurement problem for self-driving cars
- The pedestrian movement prediction for self-driving cars
- The obstacle recognition problem for self-driving cars

Up to now, it can be said that Deep learning network applied in AI has made a fairly long step on the "intelligent" path but been unable to be "self-intelligent" which is as a a big barrier . Solutions that enable systems to be capable to learn and be self-intelligent as human can. It is also the exploration direction included in the goal of this research thesis, contributing to the path of conquering new heights of artificial intelligence.

CHAPTER 2. RECOGNIZING OBJECTS BY DEEP LEARNING

The Chapter 2 focuses on studying and evaluating the performance of the CNN revolution by experiments conducted in the object recognition problems performed by AV in the traffic context. In which, the problem of pedestrian movement prediction and the problem of vehicle classification are emphasized

2.1 Problem: Pedestrian action prediction

2.1.1 Problem

Of all objects relating to the movement of autonomous vehicles, pedestrian recognition is considered the most difficult due to its complications in recognition and movement area and orbit. Therefore, it is a priority to precisely predict pedestrian's movement and walking speed to ensure the safety of pedestrians and vehicles. There are mainly three types of pedestrians: crossing, walking and waiting pedestrians. The three types involve in all possible interactions between pedestrians and autonomous vehicles. When pedestrians move or stand on the side roads, features are presented in their gestures, locations and scenes (roadway, side roads, road edges, etc..). Thus, it is possible to extract features from images of pedestrians and uses these features to train to predict and recognize pedestrian movement.

2.1.2 Solution

The proposed approach includes two phases which are: i) training a classifier model, which is used to predict pedestrian movement, with features extracted from CNN models (Figure 2.1); ii) with the frame image from real-time video of AV on the road, the order of process are: detecting pedestrians, extracting region of interest (ROI), extracting features of ROI and predict pedestrian movement in this ROI (Figure. 2.2). To extract features, CNN model of AlexNet is proposed [Krizhevsky, 2012]. To detect pedestrian, ACF algorithm is proposed [Dollár, 2009], [Dollar, 2012], [Dollár, 2014] to train and predict pedestrian movement, SVM model is proposed.

Figure 2.1 The process of extracted features with CNN model from image dataset

Figure 2.2 The process of pedestrian action prediction

The resolution of used camera is 2 Megapixels or more with the minimum resolution of collected images of 72 dpi.

2.1.3 Experimental evaluation

2.1.3.1 Extracting features and training classifier model

The experiment is carried out with about 3,000 images being extracted by CNN model. These images taken from the real street videos on the Internet were processed (selected and cut into the suitable frames) (http://youtube.com). There features are used for training of SVM classifier model. Table 2.1 shows the image and label datasets of extracted and trained features.

Table 2.1 Image and label datasets of extracted and trained features.

lass	Number	Label
Pedestrian crossing	1,000	Pedestrian_crossing
Pedestrian waiting	1,000	Pedestrian_waiting
Pedestrian walking	1,000	Pedestrian_walking

90% of images from each set is used for the training data and the rest 10% is used for the data validation.

2.1.3.2 Pedestrian detection and action prediction

With the input images, after using pedestrian detection ACF algorithm, the output is executed as in Figure 2.3. In case of the input images with many pedestrians in a frame, we extract ROI into a single image for action prediction by SVM classifier as shown in Figure 2.3. Each image in Figure 2.3 will be extracted features; finally, the system will rely on the SVM classification model to conduct action prediction of pedestrian and issue appropriate alerts for AV accordingly.

Figure 2.3 Pedestrians detected and ROI extracted.

The maximum results of rate-recognition after training and comparing with dataset in Table 2.1 are as follow:

Table 2.2 Maximum confusion matrix for pedestrian action prediction.

	Pedestrian crossing	Pedestrian waiting	Pedestrian walking
Pedestrian crossing	0.9796	0.0204	0
Pedestrian waiting	0.0612	0.9286	0.0102
Pedestrian walking	0.0102	0.0408	0.9490

The result of experiment in real-time video on the road gives minimum accuracy rate of 82%, maximum of 97% and the speed for processing reaching 0.6 second per pedestrian detected. They are promising results for potential self-driving.

2.2 Problem: Vehicle recognition

2.2.1 Problem

It is useful to detect and recognize vehicles in traffic control and separation. On the line of technology development, the need to travel and number of vehicles have increased. There are various problems in managing and separating vehicles which raises the need to apply automatic control systems with high level of precision. There are a number of solutions for monitoring systems and decision making in Intelligent Transportation Systems (ITS) such as sensors that help with reading data from devices attached to vehicles and use of the internet to network vehicles. However, some solutions could not apply in the reality due to limits in device production, internet bands and high expenses in establishment. Thus, it is essential to introduce automatic recognition and classification system for vehicles.

2.2.2 Solution

The proposed solution commences with the acquisition of images from the surveillance camera in ITS. Collected images are used to recognize objects of interest and determine the type of transportation. There are many methods for detecting vehicles, yet in this article, we focus on recognition models instead of detecting vehicles. By default, we use the semantic segmentation model based on Segnet's CNN architecture [Badrinarayanan, 2017], [He, 2015]. Vehicles detected will then be extracted to determine regions of interest (ROI). Area of interest is a sample of the vehicle. Depending on the proposed method, it is possible to use the CNN model as well as combine with data augmentation to improve accuracy. Recognition results are used in the ITS system to alert vehicles when they are not allowed to enter the limit line and to handle violations.

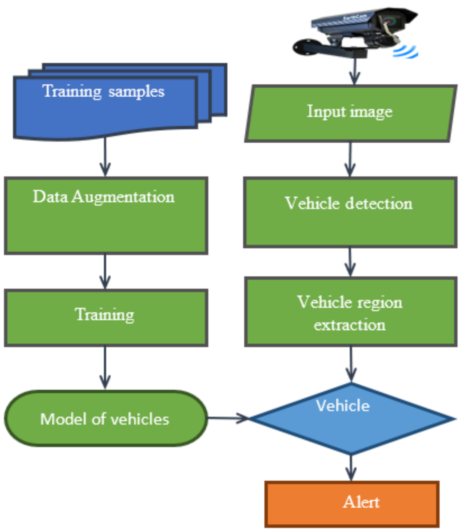

Figure 2.4 The overview architecture

2.2.3 Experimental evaluation

2.2.3.1 Experimental data

We have conducted experiments on a real database of vehicles including motors, cars, coaches, trucks taken from actual traffic situations in Nha Trang city, Khanh Hoa province, Vietnam. Camera systems typically receive signals in front of or behind the vehicles in traffic. This dataset is collected from different practical contexts on different traffic routes. The training dataset is divided into 4 different vehicle classes, including motors, cars, coaches, trucks simulated, with 8,558 vehicle images. Dataset is partitioned into 60% for training and the remaining 40% for evaluation as shown in Table 2.3.

Table 2.3 Training data

Categories	Number of samples			Sample size
	Overall	Train	Evaluation	
Motor	2673	1604	1069	128x128
Car	2808	1685	1123	128x128
Coach	1640	984	656	128x128

Truck	1437	862	575	128x128

Table 2.4 Training data after augmentation and balance data

Categories	Number of samples
Motor	16040
Car	16850
Coach	17712
Truck	17240

2.2.3.2 Training CNN

Result obtained after CNN model training is shown as follows:

(i) Filter parameters: The first convolution layer uses 64 filters, whose filter's weight is shown in Figure 2.5:

(ii) Convolution result: The sample images fed into the network through a convolution filter and the obtained data show components distinct from the original RGB image with various feature result, creating a variety of vehicle features. The output value of the convolution set contains a negative value, which should be normalized by linear adjustment. The output of some layers is shown below, with the input pattern of the motor sample (Figure 2.6).

Figure 2.5 The weight values of the filter of the first convolution layer. This layer consists of 64 filters size 7x7, each of which is connected to three RGB image input channels.

(a) (b) (c)

Figure 2.6 Some results of linear convolution and linear correction for the input images being motors. (a) The output of 64 convolutions at the first convolution layer; (b) The linear correction value after the first convolution layer; (c) The output of 64 samples at the second Convolution layer

2.2.3.3 Categorical vehicle recognition

Based on the experiment, three different methods have been evaluated on the same set of sample data as shown in Table 2.3. Methods include: (i) Traditional methods of HOG and SVM; (ii) CNN network; (iii) CNN network in combination with data augmentation.

The accuracy of the HOG and SVM method on the sample data set was 89.31%. Details of the sample size for each type and recognition result are shown in Table 2.5.

Table 2.5 Confusion matrix of vehicle recognition using HOG and SVM

	Motor		Car		Coach		Truck	
	1069		1123		656		575	
	#Num	Per(%)	#Num	Per(%)	#Num	Per(%)	#Num	Per(%)
Motor	**1029**	**97.26**	16	1.53	15	1.87	9	1.75
Car	25	2.36	**989**	**94.37**	77	9.59	32	6.23
Coach	1	0.09	23	2.19	**599**	**74.60**	33	6.42
Truck	3	0.28	20	1.91	112	13.95	**440**	**85.60**

The evaluated accuracy of the CNN method based on original data was achieved 90.10% on average, as shown in Table 2.6.

Table 2.6 Confusion matrix of vehicle recognition using CNN

	Motor		Car		Coach		Truck	
	1069		1123		656		575	
	#Num	Per(%)	#Num	Per(%)	#Num	Per(%)	#Num	Per(%)
Motor	**1026**	**95.98**	38	3.38	1	0.15	5	0.87
Car	32	2.99	**953**	**84.86**	17	2.59	24	4.17
Coach	6	0.56	104	9.26	**617**	**94.05**	58	10.09
Truck	5	0.47	28	2.49	21	3.20	**488**	**84.87**

The evaluated accuracy of the CNN method based on data augmentation was achieved 95.59% on average, as shown in Table 2.7.

Table 2.7 Confusion matrix of vehicle recognition using CNN and data augmentation

	Motor		Car		Coach		Truck	
	1069		1123		656		575	
	#Num	Per(%)	#Num	Per(%)	#Num	Per(%)	#Num	Per(%)
Motor	**1060**	**99.16**	11	0.98	0	0	1	0.17

Car	5	0.47	**1057**	**94.12**	8	1.22	13	2.26
Coach	0	0	41	3.65	**645**	**98.32**	51	8.87
Truck	4	0.37	14	1.25	3	0.46	**510**	**88.70**

In this study, we also evaluated the proposed CNN model to another traditional approach based on HOG feature descriptor and SVM classifier. Results of the comparison are shown in Figure 2.7.

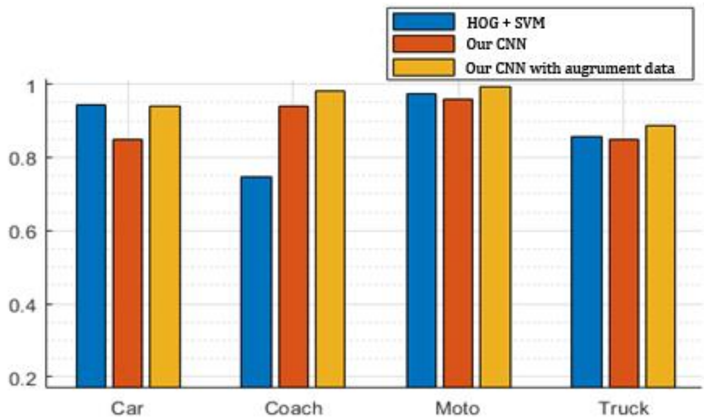

Figure 2.7 Comparison of HOG+SVM, CNN model and CNN with augmenting data

CHAPTER 3. DEVELOPMENT OF ADAPTIVE LEARNING TECHNIQUE IN OBJECT RECOGNITION

3.1. Overview of solutions

In this chapter, a solution will be suggested based on adaptive learning by CNN models. In this suggested method, the recognition model will automatically update by directly collecting data in the normal operation of an ADAS, training, comparing the accuracy and updating the model. The updating mission will focus on datasets that are different from those in previous training. The solution aims to update the old model so that it would be more adaptive and accurate. In the adaptive learning method, recognition systems can learn and add information by themselves without the help of experts in data labeling. Especially, thank to the increasingly developed online storage technology, development of infrastructure and data transmission solutions on new platforms available (5G, Cloud data, etc.), the problems of the proposed model are expected to be handled by storage and updating of online data. Suggested solutions include five main stages:

(1) Object detection with low reliability

(2) Object tracking in n images in following processes to identify if they are objects of interest.

(3) In case recognized objects with high reliability: label **Positive** for datasets recognized with low reliability in previous processes. In case recognized objects are not of interest, label **Negative** for all objected tracked in previous images.

(4) Establishing a training dataset based on the collective combination of training dataset and new dataset.

(5) Retraining and re-updating model if the new version has higher accuracy than the old one.

Trials were conducted to compare suggested model PDNet with modern models such as AlexNet and Vgg. Results showed that the suggested model have higher accuracy than a model that is self-taught over time. Further, the suggested adaptive learning model can be applied with conventional recognition models such as AlexNet and Vgg to improve their accuracy.

3.2 General Structure of the System

The general idea of adaptive learning for recognition model using CNN technology is illustrated in Figure 3.1. The recognition system can be applied to different types of objects. However, for convenience in analyzing the proposed method and functional blocks, we only apply to problems of vehicle and traffic sign classification to illustrate the idea.

There are two CNN models used in this method, the IONet model for the vehicle and traffic sign detection and the PDNet model for confidence determination and recognition.

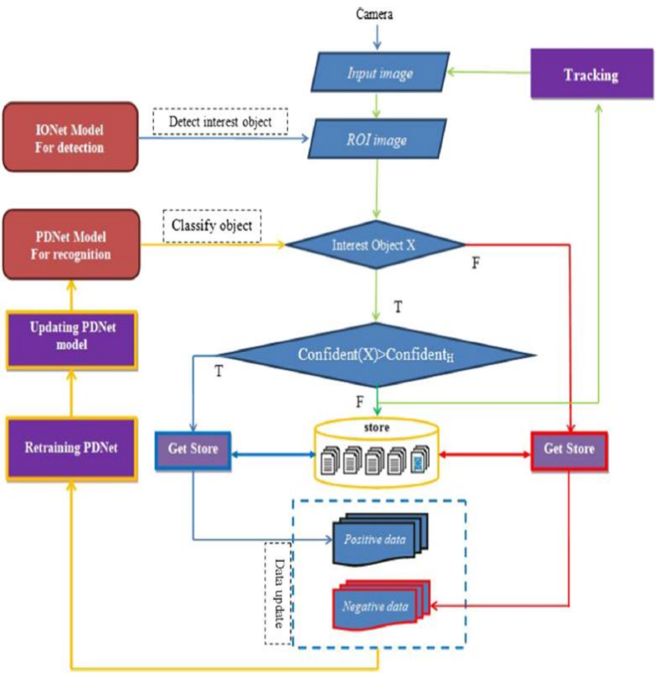

Figure 3.1 General flowchart of the system

Problem description Assume that we have trained two original CNN models, IONet and PDNet with the initial dataset. During the on the road journey, ADAS uses models to recognize vehicle, traffic sign and make appropriate decisions. However, during processing and recognizing, there are some cases in which the system recognizes vehicle and traffic sign with the low confidence score. This situation occurs when the system encounters data that is not similar to the trained dataset or the information is incomplete. The data is not homologous to original and the noise is often caused by long distance, vehicle and traffic sign obscured by other objects, warped or blurred signs, vehicles moving in conditions of light lack, rain, snow, motion noise, etc. This is the time to launch adaptive learning. The system will store images with low confidence score (*IO*) and continue to track (confident tracking) objects. The tracking process aims to identify cases: (i) Lost object; (ii) Negative Object; (iii) Positive Object. When the amount of data in the Positive Data and Negative Data sets is large enough, the retraining model CNN task is processed. The new trained model is selected and compared with the previously retrained models, the best of which is used to update the recognition model of the system. The Adaptive learning process is ongoing throughout the ADASs working process. Once updated, the new CNN model is able to recognize objects more accurately.

3.3 Experimental result

The model which was tested for relatively low configuration on a device system, corresponding to a configuration that can be equipped for self-driving vehicles in practice, is shown in Table 3.1.

Table 3.1 Device configuration

Device	Configuration
CPU	I3 3.6 GHz
GPU	Geforce 1060 6 Gb
RAM	16 Gb
HDD	SSD 160 Gb

There are changes in model accuracy when comparing vehicle and traffic sign recognition result of the initial model (*PDNet-Vehicle*$_0$, *PDNet-TrafficSign*$_0$) and retrained models (*PDNet-Vehicle*$_1$, *PDNet-TrafficSign*$_1$ and *PDNet-Vehicle*$_2$, *PDNet-TrafficSign*$_2$) on the Retrain dataset (70% data is used for train last model and 30% data received via Confident tracking object), shown in Figure 3.2

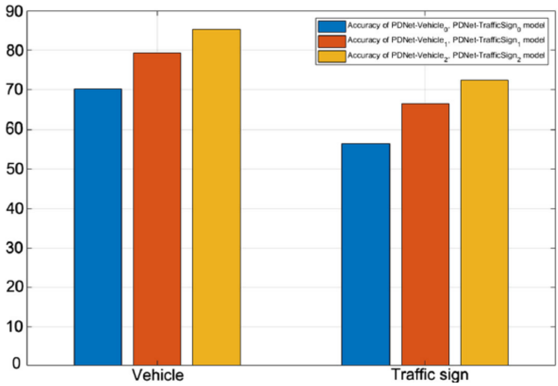

Figure 3.2 Comparing the accuracy of recognition results of retrained Vehicle and Traffic sign model

This section presents some experimental results of our proposed methods and some states of the art method of deep learning, such as AlexNet and Vgg. Initial results show that the PDNet model brings lower accuracy than that of AlexNet and Vgg models. However, after Adaptive learning process, the PDNet model brings higher accuracy than the original AlexNet and Vgg model (Figure 3.3). The processing speed of AlexNet and Vgg models is slower than that of PDNet model, since the PDNet model has a smaller input image size (64 × 64), while the AlexNet and Vgg models have large image sizes (227 × 227 and 224 × 224 respectively).

Our proposed adaptive learning method is also applied to AlexNet, Vgg models, whose results show that $Adap\text{-}AlexNet_1$, $Adap\text{-}AlexNet_2$ and $Adap\text{-}Vgg_1$, $Adap\text{-}Vgg_2$ models (after being retrained) bring higher accuracy than original $AlexNet_0$ and Vgg_0 (Figure 3.4). The results illustrated in the graphs Figure 3.3, Figure 3.4 show that no matter which model is used for training, the Adaptive learning process will improve that original model to bring asymptotically maximum accuracy over time.

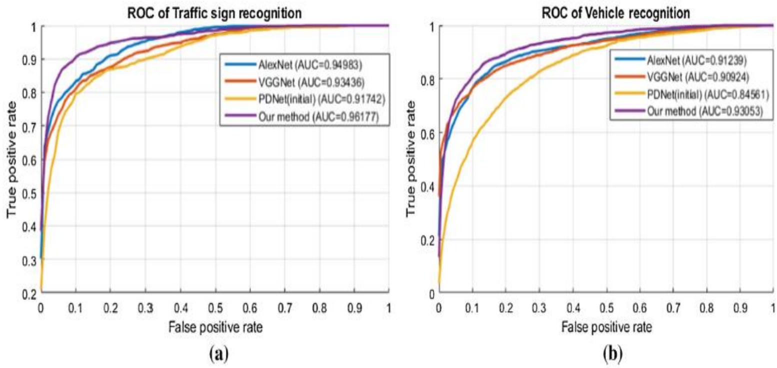

Figure 3.3 Comparison results of our proposed approach and other methods

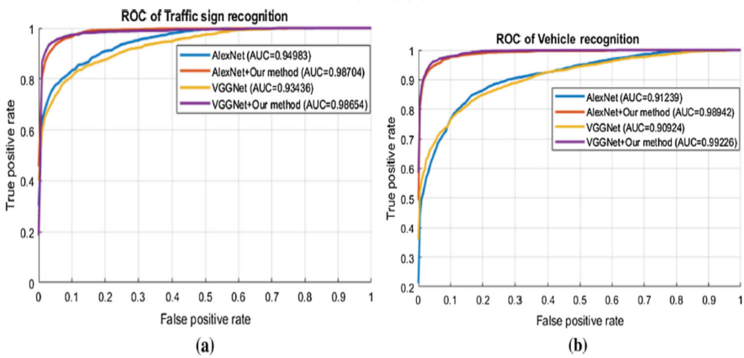

Figure 3.4 Comparison results by applying our adaptive learning to other methods

CHAPTER 4. OPTIMIZING HYPERPARAMETERS IN ADAPTIVE LEARNING

4.1. Problem of optimizing hyperparameters

Currently, studies on artificial intelligence and automotive systems focus on building solutions to optimize adaptive learning models and their parameters. Two main focused areas are model selection (e.g., CNN, ANN, LSTM and Segment) and model hyperparameters selection. However, in this thesis, parameter optimization of specific CNN models will be focused instead of training model selection.

In Chapter 2 and 3, current CNN models are trained on two types of parameters:

Optimization hyperparameters

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

Model hyperparameters

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

Model hyperparameters decide the changes of CNN models and change little during training processes of CNN models. Thus, to solve the problem of adaptive algorithms, solutions are sought to optimize the hyperparameters.

4.2. Solution overview

The training model and the proposed solution were inherited from the model proposed in Chapter 3. This proposed method makes a new contribution by changing the Retrained PDNet function block as illustrated in Figure. 3.1. The HyperNet function is added to enable the hyperparameter search for the training model, which improves recognition efficiency. The appropriate hyperparameter is automatically solved by the Bayesian approach. The overall proposed method is presented in Figure. 4.1

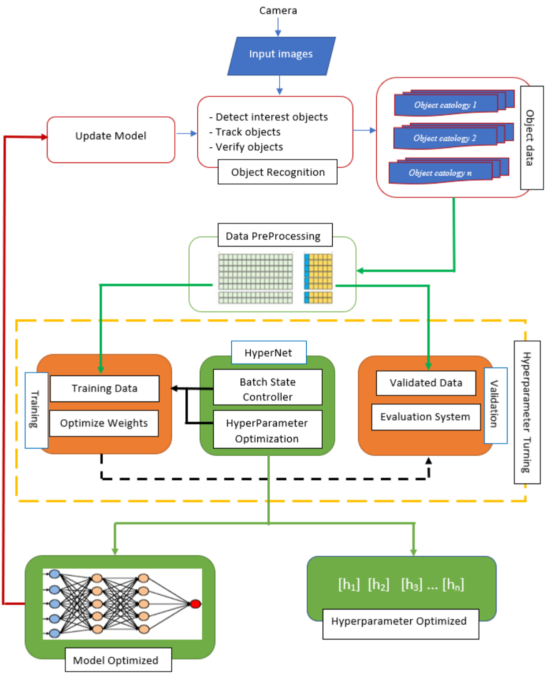

Figure 4.1. Overall proposed model

Furthermore, the data collected during ADAS movement is constantly changing and refreshed. There is no change in the structural parameters of the CNN model and training parameters in the retraining process of the previous CNN model. Therefore in theory, it is necessary to change the architecture of the CNN model and training parameters to ensure that they match with each new dataset. However, our adaptive solution for retraining the recognition model inherits its ‘intelligence’ from the previous model because of the nature of the proposed solution. Therefore, searching and changing the CNN model’s architecture is not suggested. The solution will focus on finding important hyperparameters of the training process. Then, the most equivalent and optimal model is expected to be found.

4.3 Experimental result

There are changes in model accuracy when comparing vehicle and traffic sign recognition result of the initial model (PDNet-Vehicle$_0$ and PDNet-TrafficSign$_0$) and optimal models (PDNet-Vehicle$_1$ and PDNet-TrafficSign$_1$ and PDNet-Vehicle$_2$ and PDNet-TrafficSign$_2$) shown in Figure 4.2.

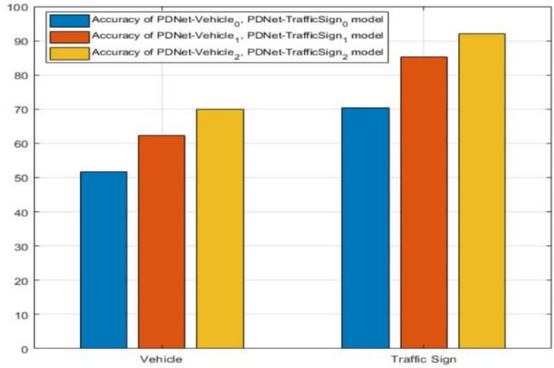

Figure. 4.2. Comparing the accuracy of recognition results of Vehicle and Traffic sign model

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

Models	Our method (%)	Previous method (%)
PDNet-Vehicle$_{0\ (\text{initial model})}$	51.77	51.77
PDNet-Vehicle$_1$	62.30	60.58
PDNet-Vehicle$_2$	69.98	68.41
PDNet-TrafficSign$_{0\ (\text{initial model})}$	70.46	70.46
PDNet-TrafficSign$_1$	85.19	84.93
PDNet-TrafficSign$_2$	92.90	90.36
AlexNet-Vehicle$_{0\ (\text{initial model})}$	66.14	66.14
AlexNet-Vehicle$_1$	88.24	86.61
AlexNet-Vehicle$_2$	90.75	90.40
AlexNet-TrafficSign$_{0\ (\text{initial model})}$	67.05	67.05
AlexNet-TrafficSign$_1$	88.78	87.73
AlexNet-TrafficSign$_2$	93.51	92.55
Vgg-Vehicle$_{0(\text{initial model})}$	71.46	71.46
Vgg-Vehicle$_1$	93.11	92.42
Vgg-Vehicle$_2$	94.78	94.14
Vgg-TrafficSign$_{0(\text{initial model})}$	70.46	70.46
Vgg-TrafficSign$_1$	95.27	94.74
Vgg-TrafficSign$_2$	95.53	94.74

Particularly, the application of Bayesian algorithm to search hyperparameters and model made the accuracy on PDNet and AlexNet, Vgg models higher than those of the similar models stated in the chapter 3, when being evaluated on the same dataset. The comparison results are shown in Table 4.1.

CONCLUSION AND DEVELOPMENT DIRECTION

1. Conclusion

The research results, which are presented in each chapter of the thesis, have been proved and confirmed through research works published in domestic and international conferences and journals. The research contents have been basically completed according to the stated objectives. In particular, outstanding contributions are:

(1) Having study and generalizing the indispensable fundamental role of traditional machine learning algorithms, the recent domestic and international researches on artificial intelligence, machine learning, Deep Learning object recognition techniques and Adaptive Learning techniques as well.

(2) The basic techniques of Deep Learning are demonstrated in the Chapter 2 (Pedestrian recognition, vehicle recognition,...). Through the simulation experiments of ADAS equipment in traffic, it has shown that the CNN models' ability to recognize is great when being trained. The research results in this chapter are considered as a foundation for an overall model development of an ADAS system which is capable of self-learning and become more intelligent.

(3) The main contribution of the thesis is to propose a comprehensive model for Adaptive Learning solution. The operation of the ADAS model demonstrated that an auto robot system is capable of self-learning and recognizing by simulation of the human brain. The proposed solution, along with adaptation and automatic updating of actual data, enables the system to change and adapt to the training hyperparameter set matched with the input data. It is this combination that has generated a quite

complete model for the Adaptive Learning solution of auto robot systems in the future.

(4) Through the experiments on the research contents, the author has collected and develop a dataset of many different objects such as a data set of actual pedestrians, a data set of pedestrian posture, a data set of traffic signs, and a dataset of vehicles as well. Because data for the experimental process are not available (including published famous datasets), the data sets of images stated in the thesis were in real ones which were collected directly from real movement of car on road or from internet videos.

(5) However, although there have been encouraged results, some following issues still remain to be solved to improve and prove the effectiveness of the Adaptive Learning model.

- A few numbers of experimental objects that have not covered many other cases. Limited images in the data set leaded to low accuracy of recognition model.

- Some parameter values for training are proposed default that have not been proved to bring the highest efficiency (For example: value of N image number at the start of retraining process of model, % of image data of the previous dataset is reused for next model training, etc.).

- The hyperparameter value range is only estimated through experiment does not value range need to be searched.

2. Development direction

The proposed model shows the Adaptive Learning solution of ADAS devices. However, it can be seen that further research and development in following different directions may be of potential:

- Extend objects for recognition to diversify the capabilities of the ADAS system or develop into a complete auto robot system capable to Adaptive Learning on all objects.

- Evaluate and search appropriate values replacing fixed values during training of Adaptive Learning model. Extend the search parameter range to increase the ability to select the appropriate parameters for retraining the model corresponding to the new data set. At the same time, the study will find a solution in which the complexity in the hyperparameter searching process of the proposed model is reduced with minimized time and improved processing efficiency.

- In the proposed model, the continuous adaptive learning process will enable the training dataset to rapidly increase in number. Thus, the point is to develop a lean solution with a selective training dataset in order to eliminate easy samples while prioritizing hard samples. This is expected to make the model possible to reduce training time and improve the accuracy and quality of the adaptive learning process.

- Develop a complete and large data set with a variety of different types of objects for the initial training of the Adaptive Learning model

LIST OF PUBLISHED SCIENTIFIC WORKS RELATED TO THE THESIS

1. Major publication papers

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. Supplementary publication papers

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.