Traffic Sign Recognition From Digital Images Using Deep Learning

Jiawei Xing

A thesis submitted to the Auckland University of Technology
in partial fulfillment of the requirements for the degree of
Master of Computer and Information Sciences (MCIS)

2021

School of Engineering, Computer & Mathematical Sciences

Abstract

As a kind of road facilities, traffic signs are presented by various symbols with multiple backgrounds to guide, limit, and warn road users. It is one of the important parts in civilian transportations. Once a driver overlooks the signs, especially important ones, due to complexity of actual traffic scenes or influence of adverse weather conditions, which will lead to violation of traffic laws or regulations that may cause traffic accidents, result in casualties and property losses. Therefore, computer-based recognition of traffic signs is a vital part of real requirement of modern traffic systems, which takes not only predominant significance in our daily life, but also high values in academy research.

Due to the complexity of specific environment and increasingly severe haze in recent years, high requirements are put forward for traffic sign recognition (TSR). At the same time, the tolerance and adaptability requirements in practical application environment will also be improved. Based on the recognition of traffic signs, in this thesis, we propose an image defogging method by using guided image filtering. Then convolutional neural network (CNN) in deep learning is applied to recognize traffic signs under foggy weather condition, improve accuracy and speed of TSR. Our main contributions of this thesis are: (1) We propose a HSV color gamut defogging algorithm based on CLAHE algorithm. We offer a guided image filtering algorithm for image dehazing. Our result shows that the guided image filtering method is very effective in image defogging. (2) In this thesis, we proffer two deep learning algorithms for our experiments, one is Faster R-CNN, the other is YOLOv5, we compare the two methods to find that YOLOv5 is much suitable for real-time TSR. (3) In this thesis, we propound a new way to identify road signs. The improved YOLOv5 model is applied to satellite imaging for accurately detecting road signs on ground.

Keywords: TSR, Defogging, Faster R-CNN, YOLOv5

Table of Contents

List of Figures

List of Tables

Attestation of Authorship

I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person (except where explicitly defined in the acknowledgments), nor material which to a substantial extent has been submitted for the award of any other degree or diploma of a university or other institution of higher learning.

Signature: Date: 18 September 2021

Acknowledgment

First of all, I want to thank my parents. They are very happy that I can study at the Auckland University of Technology (AUT), and gave me financial support. This allows me to complete my research project at AUT successfully. At the same time, I also want to thank my friends and teachers, they gave me care and help in my ordinary study and life.

In this research project, I am very grateful to my supervisor Wei Qi Yan for his in-depth academic research in deep learning, meawhile, he is one of my amiable friends. With his help, I learned a lot about deep learning, the knowledge of visual object detection, and how to complete my master's degree under his supervision.

Jiawei Xing

Auckland, New Zealand

September 2021

Chapter 1
Introduction

In the first part, the motivation and background of our research project on TSR is introduced, in the second part, the current development of TSR is presented, our contribution, objective, and structure are detailed.

1.1 Background and Motivation

With continuous growth of the number of human beings and the constantly development of our communities, cars have become an indispensable means of transportation for people. The number of cars in various countries in this world is growing rapidly. From statistics in 2017, China's car owners have exceeded 200 million vehicles, an increase of 23.04 million vehicles compared to 2016, with a growth rate of 11.9%. By 2018, according to the latest statistics of China's automobile ownership, 31,720,000 motor vehicles have been successfully registered and registered, with 327 million vehicles, including 240 million vehicles and 369 million drivers. While people enjoy the convenience of cars, it brings a series of inevitable problems to road traffic.

Although there are traffic signs everywhere on the road to help us travel in real scenes as shown in Figure 1.1, the increasing number of cars makes our traffic more and more congested. What is more serious is the frequent occurrence of traffic accidents, which threatens our ordinary lives. At the same time, the increase in the number of cars also brings many social problems, especially the emission of automobile exhaust leads to the decline of air quality, the ecological environment is threatened. Extreme weather, especially haze weather and foggy environment lead to limited vision of drivers, it is easy to ignore important traffic signs as shown in Figure 1.2. This will induce drivers' consciousness of violating road traffic regulations and become the possible cause of traffic accidents.

Thus, better traffic condition on the urban traffic system is increasingly required, among them, the most important one is traffic accident (Litman, 2010). In order to reduce accidents, research work on driving assistance systems has become the focus of autonomous vehicles (Berkaya, Gunduz, Ozsen, et al., 2016).

Figure 1.1: Examples of traffic sign

Figure 1.2: A traffic sign in foggy weather

Due to the large population and different levels of motor vehicle drivers in various countries, speed limit and driving restriction cannot completely solve the problem. How to effectively alleviate the road traffic pressure without affecting the economic development, so as to improve the road traffic efficiency, reduce the occurrence of traffic accidents, and strengthen the road traffic safety, has become a problem all over the world, which has been paid close attention of various countries.

With the development of intelligent transportation system, experts and scholars have begun to investigate this issue, which has been risen globally in a short time (Wu, & Zhong, 2013). For those countries with large territory, road transport is still the main way of transportation, research work on a traffic management system, like Intelligent Traffic System (ITS), is bound to become a main direction of road traffic development. TSR (Lykele, 2014) as the main component, its theory is based on pattern recognition,

image processing, computer vision, and other computer technology. Its goal is to automatically acquire the current road information amid vehicle driving.

Throughout analysis of traffic signs in the obtained information, effective traffic sign information is accurately identified and transmitted to the driver. A driver can get the information of traffic signs at the first time, make the right response in time, ensure the driving safety and maintain the smooth road to the maximum extent. Therefore, with the rapid development, TRS has been paid much attention on academic communities (Timofte, 2009, Ai, 2014). However, the destruction of the ecological environment makes the frequent occurrence of bad weather such as haze in autumn and winter, which seriously restricts the development of intelligent transportation system.

The existing TSR algorithms greatly reduce the best outcomes in bad weather. The main reason is that the image quality obtained by using intelligent transportation imaging equipment in haze environment is poor, key image information is seriously lost (Zhu, & Wang, 2015). Therefore, it is of great value to study the recognition and detection of traffic signs in haze environment.

With the development of TSR, the performance of computers has also a qualitative leap, people apply computers to simulate and tackle a large amount of complex data. Intelligent driving assistant system simulates the mechanism of human visual system to complete object detection and object recognition (Shi, 2016) (Ma, Fu, Wang, & Yin, 2017) and other operations (Pan & Sun, 2013).

At present, a plenty of image processing methods have been offered to intelligent driving assistance to detect pedestrians, front and rear vehicles, road lines and lanes (Bertozzi. Broggi,1998, Handmann,2000) etc. However, identification and detection of road traffic signs are not very much. As the most intuitive information, road traffic signs in developed countries have their own mature standards. Road traffic signs include lane speed limit prompt, lane direction indication, lane condition warning, and other information, which is an indispensable part of intelligent assistant driving system. Efficient and accurate identification and detection of road traffic signs are conducive to alleviate traffic congestion and reduce the probability of traffic accidents.

In recent years, exploration and exploitation of deep learning have demonstrated

the great potential in deep learning. In order to prove the feasibility of deep learning, Hinton's research group participated in the ImageNet image recognition competition in 2012. They clutched the champion title through CNN-based AlexNet (Krizhevsky et al., 2017), by using the absolute advantage to crush the SVM method, it has an absolute advantage in classification performance and accuracy. After this achievement, a surge of deep learning was set off. In 2013, 2014, and 2015, through the ImageNet competition on image recognition, the improvement of deep network structure, training method and GPU hardware, deep learning has been continuously achieving remarkable progress in the fields besides image recognition, conquering the battlefield in various research areas. Deep learning plays a pivotal role in the accurate recognition and judgment of images, which is very suitable for recognition and detection of road traffic signs in this thesis.

In summary, the recognition and detection of traffic signs have a promising application in future, especially the research projects on the recognition of traffic speed limit signs in haze weather are in line with the needs of the current development of the automobile industry, which have high social and economic impact, and are conducive to promoting the transformation of scientific and technological knowledge. Therefore, the traffic sign detection in extreme weather based on deep learning is crucial, practical, and realistic.

1.2 Intelligent Recognition and Detection Systems for Traffic Signs

Owing to the development speed of automobile industry, a vast majority of automobile manufacturers such as Germany, the United States, and Japan have had relevant research institutions to start the research on automatic driving technology (Yurtsever, Lambert, Carballo, & Takeda, 2019). Among them, TRS system is an important part. The original intention of TRS system is to avoid incidents and assist or replace human drivers to complete related operations. The "Stanford Cart" project was firstly launched in 1961. It is recognized by the world as the first self-driving vehicle that utilized

vehicle mounted cameras and equipped with early artificial intelligence system to circumvent obstacles.

However, due to the hardware computing performance at that time and the imperfect related algorithms, the biggest problem of “Stanford Cart” is that it takes 20 minutes to move one meter forward. The earliest TSR in the world was developed by the Japanese team in 1987. Its main recognition object is speed limit sign. The classical algorithms such as threshold segmentation and template matching are profound to realize the recognition of speed limit sign. The whole recognition process spends 0.5 seconds (Fu, & Huang, 2010). After that, many teams from European and American countries devoted to the research of TSR. In 1994, Koblenz Landau University cooperated with Daimler Benz company of Germany developed a real-time road TSR system. The recognition speed reaches 3 frames / second for about 40,000 experimental images, the recognition accuracy is up to 98.0% (Priese, 1994).

In the 21st century, more and more research teams begin to devote to the research work in this field, the research methods tend to be more mature. In 2000, Osaka University in Japan has developed a driving assistance system that can recognize character signs and speed limit signs (Miura, 2000). The system took use of a wide-angle camera to find the possible object areas, and then utilized a long-distance camera to enlarge these possible areas, then applies threshold segmentation method to detect objects, and offered template matching to identify. There are 71 speed-limit traffic signs having been tested, the detection success rate is 97.2% and the recognition accuracy is 46.5%. In 2005, Gareth Loy laboratory in Sweden and Nick Barnes Institute of automation in Australia jointly participated in a set of TSR projects. The system adopts the symmetry and circular shape features of most traffic sign images to determine the centroid of the sign, and then realizes the recognition of the sign. The recognition accuracy is up to 95.0%. In 2010, the University of Massachusetts developed the TSR system, which was use of color threshold segmentation and principal component analysis algorithm to detect and recognize the object. The recognition accuracy of this system is as high as 99.2%. Moreover, the system achieves good results in low visibility weather conditions or slight object occlusion, but the time is long, which cannot

perform real-time TSR.

Compared with other countries, China started late in the field of intelligent transportation, the research work of TSR is much difficult. So far, there is no direct research outcomes which have been applied to practice. However, with the rapid development in recent years, many universities, research institutions, and enterprises in China have shifted their focus to TSR. Fleyeh (Hasan, 2008) have studied traffic sign detection and recognition technology earlier in China. Mathematical morphology method and morphological skeleton function were applied to extract features for warning signs, template matching algorithm was proffered to classify and recognize traffic signs. The system has good robustness, high complexity, and low recognition rate for traffic signs in natural scenes.

In 2009, Xiamen University China (Yuan, 2009) offered an improved model based on visual saliency regions to get saliency regions that may contain traffic signs, Histogram of Oriented Gradient (HOG) as visual features and Support Vector Machine (SVM) as the classifier were taken into account to detect these saliency regions. The detection rate reaches to 98.3%, the error recognition rate is 5.09%. The algorithm is able to deal with illumination, scaling, angle changing, and partial occlusion. The research work in TSR has always been a hot field in intelligent transportation, because of its difficulties, there are few satisfactory practical applications. The natural scene of traffic signs is complex, which is greatly affected by weather conditions. Especially in recent years, haze weather is increasing, the road visibility is getting lower and lower, and the signs have often fade, tilt, occlusion, adhesion, etc., which bring challenges in the identification of traffic signs. In addition, the algorithm requires high real-time performance, especially the speed-limit sign recognition has become a challenging job.

A complete TSR system comprises two parts, namely speed-limit sign detection and recognition. The main work of speed-limit sign detection is to find and segment the signs in the complex scene. The main work of speed-limit signs recognition is to use neural network and other classifiers to recognize the signs, so as to achieve the goal of TSR (Moutarde, 2007), (Nguwi, & Kouzani, 2008).

1.3 Contributions

This thesis takes use of deep learning methods for TSR and proposes an algorithm for the images taken in foggy weather. The contributions of this thesis are:

(1) Because different countries have different cultures, the traffic signs of each country are also different, so to improve the speed of recognition of traffic signs in New Zealand, we have created our own dataset of New Zealand traffic signs. In this thesis, we created two different datasets of New Zealand. One is driver's perspective dataset and another one is satellite image dataset. These two datasets help to improve our accuracy of TSR in New Zealand.

(2) We propose two defogging algorithms. The first is the HSV color gamut defogging algorithm based on the CLAHEPN algorithm. The second defogging method is guided image filtering, and then we compare the two methods. The final results show that our improved method has a better performance in the defogging.

(3) We improve Faster R-CNN model and YOLOv5 model which are popular in the field of detection. The experimental results show that the performance of our improved method for TSR after defogging is greatly improved compared with that before improvement. The specific methods are as follows: We take use of Leaky ReLU as the activation function of Faster R-CNN, which shows the nonlinear representation ability of neurons. At the same time, in the selection of feature extraction framework, we are use of GoogLeNet model to replace VGG model, which makes the improved model faster while ensuring the recognition accuracy. In the model of YOLOv5, we improve the loss function. In order to improve the detection ability of this model for small and far road signs, the constraint of adaptive balance between foreground and background is added to the loss function, so as to achieve better detection outcomes.

(4) We offer a new way to identify the road signs. As long as the improved YOLOv5 model is applied in the image captured from the satellite image to accurately detect the ground road signs from the perspective of the satellite image, the sign recognition results from the driver's perspective are improved, the mutual cooperation between the two is able to provide better results for TSR under the background of automatic driving.

1.4 Objective of This Thesis

At first, we need to collect massive data about traffic signs in New Zealand so that our model is able to get better results. Secondly, many external factors influence the accuracy of TSR. We find a defogging algorithm for defogging. This model has a higher accuracy rate for the foggy images. Since the probability of foggy days is very low, we won't have much time to shoot foggy images by ourselves, so we will find a slew of foggy images as our data sets. We will compare two different deep learning methods. At the same time, we will study the influence of fog on the recognition results, finally, we will find a method having fast recognition speed and high accuracy for TSR.

1.5 Structure of This Thesis

This thesis includes five chapters: In the first chapter, we describe the research background and significance of this topic, as well as the status of related topics in recent years. We analyze the main difficulties and research contents of this subject and expound the main technical route of this thesis. Finally, we make a brief summary of the organizational structure.

In the second chapter, we discuss basic theory and relevant methods related to this topic. We introduce the basic theory of fog removal method and deep learning method in traffic sign detection and recognition under fog weather condition, including the principle of haze formation, atmospheric scattering model, and the related components of convolution neural network. The relevant concepts of road sign recognition are introduced, which provide a theoretical basis for the subsequent chapters of this thesis.

In the third chapter, we mainly discuss the TSR experiments based on YOLOv5 and Faster R-CNN. Firstly, after analyzing and comparing the defogging algorithms, we propose an adaptive histogram equalization algorithm with limited contrast in S and V gamut in a guided image filtering algorithm. Secondly, GoogLeNet is used to replace the VGG network of Faster R-CNN, Leaky ReLU algorithm with better nonlinear expression ability is employed. Finally, we propose a YOLOv5 model based on the

adaptive balance between foreground and background loss. A plenty of experiments have proved that by balancing the weight of the network in the foreground and background, we make YOLOv5 much suitable for detecting small traffic signs in the distance. At the same time, we have better confidence in each object and accurately identify the type of traffic signs. We also recognize traffic signs from the perspective of the driver and satellite images. We found that using two perspectives for TSR can improve the efficiency of recognition. We see through our experimental results that the effective combination of the proposed defogging methods is able to effectively detect the road signs in foggy images better and faster than the original Faster R-CNN model.

In the fourth chapter, we mainly discuss the experimental results of Faster R-CNN and YOLOv5 on the recognition of traffic signs and compare the two methods.

In the fifth chapter, we analyze and discuss the contributions of this thesis. For example, we identify traffic signs on ground from satellite images. At the same time, we also analyze the shortcoming of our experiment and improvements of the algorithm.

In the sixth chapter, we depict the novelty and creativity of this thesis, we analyze the existing problems in this research field, and put forward new ideas for future research projects.

Chapter 2
Literature Review

This chapter is split into three parts. In the first part, the mechanism of haze on digital images and the atmospheric scattering model are investigated, the reasons for degradation of road sign recognition in haze weather are identified. In the second part, convolution neural network in deep learning, including convolutional layer, pooling layer, full connection layer, activation function, and corresponding optimizer is reviewed. In the third part, the process, purpose, and function of each part have been summarized.

2.1 External Factors Affecting TSR

With the development of artificial intelligence and deep learning, TSR is also being developed rapidly, many high-end models now include TSR driver assistance systems to help drivers on road safely. The current TSR has a high accuracy only on the images taken in sunny days, if traffic signs are unobstructed, especially in adverse weather (e.g., fogging, raining, snowing, etc.), challenging lighting conditions (e.g., night, direct sunlight, shadows, etc.). If traffic signs are obscured, false alarms or no recognition results output. We will then describe how to solve the problems, how the reasons affect TSR, and what the differences are.

2.1.1 Impact of Raining on TSR

TSR on the images taken in rainy days is a very difficult task because foggy is static and has not obvious motions, whereas raining or snowing has dynamic motions of visual objects (Garg, & Nayar, 2004). All these are very challenging for TSR, which already requires real-time object detection. The size, speed, and shape of raindrops are uncertain, the positions of the drops are rath random (Hnewa, & Radha, 2021). In the case of low lighting conditions, raindrops of shallow mass will fall through the air at a plodding speed. The drops in the air reduce visibility and obstruct the camera, which may cause the traffic signs before a camera has deformation that leads to be misidentified.

If these tiny water drops are appeared on car windows, they leave a trail of raindrops. If the rain is heavily, the raindrops will be fast enough to cover the windows like a curtain, which will make the traffic signs on the acquired images blurred. After the raining, water on the ground like a mirror to reflect traffic signs (Hasirlioglu, & Riener, 2018) which can make the TSR incorrect or unrecognizable. As a result of these problems, Chen et al. proposed a TSR system with a visibility enhancement module (Chen, Chang, Yu, & Chen, 2020), a convolutional encoder, mixed with a convolutional decoder and a dark channel (Nan, Gang, & Song, 2020). In the experiments, the accuracy rate was improved 50% by using the images from rainy days.

2.1.2 Impact of Illumination on Traffic Sign Recognition

Illumination is essential for TSR. During the daytime, if sun shines directly on a traffic sign or camera, it will be over exposed. At night, it is often too dark, the lights of the car will reflect on the signs amid driving at night. This will lead to incorrect recognition of traffic signs. In response to these problems, Greenhalgh et al. proposed a method for traffic detection that adjusts the intensity of the incoming light and allows for better TSR (Greenhalgh, & Mirmehdi, 2012). It effectively reduces the effect of shadows and illumination on traffic recognition that is combined with the geometric features of traffic signs to increase the recognition accuracy under various lighting conditions. Meanwhile, Kwangyong et al. proposed a byte-MCT and traffic sign-based AdaBoost classifier for TSR, which firstly applied SVM to verify the candidate regions and then CNN to extract visual features, this method achieved very good results with strong robustness under various lighting conditions.

2.1.3 Impact of Fog on Traffic Sign Recognition

While we shoot a photograph by using digital camera in the outdoor environment under haze weather, the ambient light will be seriously scattered, resulting in blurred image features, no feature extraction, and other related operations are employed (Li, Wang & Zheng, Zheng, 2014). Through the analysis of a large number of haze images and clear images of the same scene, we find that the haze images have specific characteristics, the analysis of these image characteristics automatically identify the image haze and classifier for pattern classification (Peng, Liu, & Dong, 2008). The characteristics are summarized as follows.

Firstly, the contrast of haze image is seriously reduced compared with the clear images, the attenuation degree of contrast shows exponential attenuation characteristics with the increase of scene structure information (Tatsumi, Yasuhiro, Akio, & Toshiharu, 2005).

Secondly, the salient features of the image are blurred, the contour and edge fade out (Provenzi, Fierro, Rizzi, Carli, Gadia, & Marini, 2007). The image colors appear

distortion and offset, the color saturation is reduced. The distribution of image pixel values is concentrated, the dynamic range shrinks and the pixel histogram presents uniform distribution (Rizzi, Gatta, & Marini, 2003). Finally, from the perspective of image frequency domain, there is a large amount of low-frequency information in the image, the high-frequency information is relatively reduced. In terms of visual effect and appearance, the grayscale level of the images changes evenly, the detailed information is reduced, the edge of the image contour is weakened (Dong, 2012).

The reasons of haze image quality degradation and blurring are summarized. In haze weather, there are suspended aerosol particles in the air. Aerosol particles will confuse the light transmission medium, cause serious light absorption and scattering, directly generate the energy attenuation carried by reflecting the light rays from the object surface in the scene (Deshmukh, Singh, & Lakha, 2016).

In the whole optical path of light rays to the imaging device, the ambient atmospheric light is also scattered to the imaging device so as to participate in the imaging of scene objects. The two factors work together to reduce the image contrast and the image resolution.

The influence of aerosol particles on light scattering varies with the size and shape of particles. At the same time, with the scene imaging in the environment, the particles themselves also participate in the scene imaging, which is regarded as image noise, resulting in the loss of image detail information (Guo, Cai, & Xie, 2011).

The light reflected by the object surface in the scene is scattered along the light path to the camera lens. At the same time, the light reflected by other object surfaces in the scene also participates in the imaging after multilevel scattering, resulting in image blurry (Hines, Rahman, & Jobson, 2005).

In the haze weather, radiation intensity of lights will be attenuated due to the cloud cover, results in the reduction of the light intensity in the environment, so that the brightness of the image collected in the haze weather is not enough (Xu, Guo, Liu, & Ye, 2012).

To sum up, the main reason of image blurry in haze environment is that there are a large number of suspended particles in the environment, which makes light scattering

seriously.

2.1.4 Atmospheric Scattering Model

Scattering refers to the phenomenon that while light passes through nonuniform medium, light ray deviates from the original propagation direction, the part that deviates from the given direction is the scattered light. The scattering particles in the atmosphere include aerosol, tiny ice crystals, tiny water droplets, etc. The scattering reflection is the main reason for the interference of the field of computer vision. In 1925, Koschmieder proposed that the low visibility of foggy image is generated by the absorption and scattering of reflections of particles in the atmosphere. In the actual image acquisition, the absorption of light by atmospheric particles is generally ignored, and only the scattering of light is considered.

In haze weather, the scattering of atmospheric particles will not only contribute to the loss of part of the reflected light from the object surface, but also add the atmospheric scattered light to the reflected light from the object. Under the combined results of the two reasons, the contrast and color characteristics of the images collected in foggy weather will be seriously attenuated (Hide, 1977). At the same time, based on scattering theory, the atmospheric scattering model has a far-reaching impact on the images. In 1999, Srinivasa et al (Nayar, & Narasimhan, 1999) explained the imaging process in foggy weather by establishing the corresponding mathematical model.

The lighting source received by the observation point from two parts: One is the object reflected light after the attenuation of the reflected atmospheric particles, the other is the atmospheric scattered light combined with the ambient light after the scattering of the particles. Throughout this model, we get the mathematical model as shown in Equation (2.1).

$$I(x,\lambda) = e^{-\beta(\lambda)d(x)}R(x,\lambda) + L_{\omega}\left(1 - e^{-\beta(\lambda)d(x)}\right) = D(x,\lambda) + A(x,\lambda) \quad (2.1)$$

where $I(x,\lambda)$ represents the foggy images obtained from the observation, $R(x,\lambda)$ indicates the original image to be recovered, λ is the wavelength of the light wave, x is the position of the pixel in the image, L_{ω} represents the atmospheric light value at

infinity, $t = e^{-\beta(\lambda)d(x)}$ is the transfer function.

The models are subdivided into atmospheric light scattering models and incident light attenuation models. The incident light attenuation model is a model of the phenomenon of incident light attenuation generated by scattering atmospheric particles. In haze weather, the attenuation of light is the main reason that leads to the decline of image quality. The model shows that the brightness of points in the scene decreases with the increase of depth of field.

Assuming that the incident beam has a unit cross-sectional area, where the beam travels a distance dx, its intensity change is written as Equation (2.2).

$$\frac{dE(x,\lambda)}{E(x,\lambda)} = -\beta(\lambda)dx \tag{2.2}$$

where $E(x,\lambda)$is the light intensity after attenuation, $\beta(\lambda)$is the scattering coefficient, assuming that the intensity of the undamped beam at $x = 0$ is $E_o(\lambda)$, by integrating Equation (2.2) from $x = 0$ to $x = d$, we get the following results.

$$E(d,\lambda) = E_0(\lambda)e^{-\beta(\lambda)d} \tag{2.3}$$

If the incident light is a point light source, then the intensity of the undamped light is I_0, the same integral from $x = 0$ to$x = d$ is performed for the differential equation, and the intensity of the point light source is divided by $x = d$ after attenuation is

$$E(d,\lambda) = \frac{I_0(\lambda)e^{-\beta(\lambda)d}}{d^2} = \frac{L_\omega\rho(x)}{d^2}e^{-\beta(\lambda)d} \tag{2.4}$$

The atmospheric light scattering model reflects the influence of ambient light scattering by atmospheric particles on image quality. After scattering by particles, a part of the ambient light will also enter the light received at one end of the observation point, which is the additional atmospheric light. The medium in the volume micro element is regarded as a light source, dwx^2 is the cross-sectional area, dx is the thickness. So dv is expressed as,

$$dv = dwx^2dx \tag{2.5}$$

Its intensity in the direction of observation point is,

$$dI(x,\lambda) = dvk\beta(\lambda) = dwx^2dxk\beta(\lambda) \tag{2.6}$$

where $\beta(\lambda)$ is still the scattering coefficient, k is the proportion constant of light source, which represents the relationship between scattering function and illuminance.

According to the attenuation equation of the light, the light intensity at the observation point is deduced as,

$$dL(x,\lambda) = \frac{dI(x,\lambda)e^{-\beta(\lambda)x}}{dwx^2} = k\beta(\lambda)e^{-\beta(\lambda)x}dx. \quad (2.7)$$

Then, by integrating Equation (2.7) in the range of $x = 0$ to $x = d$, the total atmospheric light intensity is obtained as follows,

$$L(d,\lambda) = k(1 - e^{-\beta(\lambda)d}). \quad (2.8)$$

Because the ambient light source generally comes from the sky which is considered to be near infinity relative to the observation point and the object point, let $d = \infty$,

$$L(d,\lambda) = L(\infty,\lambda) = L_\infty(\lambda) = k \quad (2.9)$$

The total light intensity of the observation point is

$$I(x) = \frac{L_\infty \rho(x)}{d^2} e^{-\beta d(x)} + L_\infty(1 - e^{-\beta d(x)}) \quad (2.10)$$

For the convenience of expression and calculation, let the reflected light of the object be $J(x) = L_\infty \rho(x)/d^2$, atmospheric transmittance is $t(x) = e^{-\beta d(x)}$, the object attenuation reflected light is expressed as $D(x) = J(x)t$, the atmospheric light is expressed as $A = L_\omega(1 - t) = A_\omega(1 - t)$. Therefore, the mathematical expression of the atmospheric scattering model under the final haze weather is shown as Equation (2.11).

$$I(x) = D + A = J(X)t(x) + A_\omega(1 - t(x)) \quad (2.11)$$

where $t(x)$ is the transmittance of the relevant medium, $I(x)$ represents the original image of the unprocessed fog, $J(x)$ is the image after defogging, $J(x)t(x)$ is the direct attenuation term, which takes account for the degree of attenuation of light reflected from an object in a medium, $A_w(1 - t(x))$ denotes the part of atmospheric scattered light attached.

2.2 Related Algorithms

2.2.1. Image Defogging Algorithm

Haze and other adverse weathers directly lead to the decline of image contrast, the

reduction of gray dynamic range, the reduction of definition, blurring, the coverage of detail information and other phenomena. Therefore, it is necessary to investigate how to restore the clear and fog free image. At present, the dehazing and defogging algorithms are qualitatively divided into two research directions: One is image restoration, which establishes a mathematical model based on the prior knowledge of image quality degradation and reverses the model to calculate the clear and fog free image (Huang, Huang, Gu, Liu, & Luo, 2017). The other direction is image enhancement, based on human visual needs as the standard, highlighting the image details, filtering noise information to restore the clear image (Feng, Li, Hua, 2017). The difference between them lies in that image restoration is to improve the comprehensibility of image from the perspective of image essence. Image enhancement is from the perspective of human visual sense, to improve the visual effect of the image, in order to meet the needs of human visual system.

In the field of visual object recognition, image enhancement has been the focus of digital image processing. The image enhancement algorithms include image enhancement based on gray transformation, image enhancement method based on histogram trimming (Li, 2011), image enhancement algorithm based on homomorphic filtering (Dong, 2018), image enhancement algorithm based on smoothing (Pal, 1981), image enhancement algorithm based on sharpening (Portniaguine, 2005.), image enhancement algorithm based on curvelet transform. These methods are classified into multiple categories based on frequency domain processing and algorithms based on spatial domain processing. From the perspective of image processing, it is grouped into local region-based image enhancement algorithms and global region-based image enhancement algorithms.

In the field of image enhancement, the algorithm is the most simple and easy to understand enhancement algorithm. The basic principle is to achieve image enhancement according to different gray transform functions. Histogram equalization also achieves the outcomes of image enhancement. It gets the overall image enhancement by reallocating the grayscale value of image pixels, however, it does not obviously show the details of the image.

The homomorphic filtering algorithm has brought fresh blood to image enhancement and made great progress (Xu, Gao, Liu, & Ye, 2012). Image smoothing and its corresponding low-pass filtering methods also enhance the image but cannot meet the needs of highlighting the details. Image sharpening algorithms and the corresponding high pass filter enhancement algorithms highlight the details of the image and highlight the local features. However, whilst enhancing the image effect, the original noises of the image are also highlighted, the new noises are introduced. In the field of image processing, the main function of wavelet transform is to highlight the image details, but it is difficult to avoid the extra noises in the processing, which makes the image enhancement effect is not ideal.

Another defogging algorithm based on mathematical model is from the perspective of image restoration. According to the prior knowledge of image quality degradation, the mathematical model is established, the clear and foggy-free image is restored. Among them, the design of haze image imaging model is based on analysis of prior knowledge of image quality degradation. Firstly, the interaction between the light reflected from the object surface in the scene and suspended particles in the propagation path is analyzed in detail. Then, the light energy decay model and the ambient light model are established. Under the interaction of the two models, the optical mathematical model of image imaging in haze weather is given, finally the scene restoration is realized (Liu, Yang, Wu, Zhuang, & Deng, 2016).

In the development of image defogging from the perspective of image restoration, a number of classic algorithms have been proposed. Firstly, the influence of atmosphere on scenes is analyzed and evaluated. The light scattering theory is the main reason for low image quality and blurred detail information in the whole process. The application of scattering theory and relevant mathematical models to the modeling of image degradation under haze weather conditions is able to restore clear images.

From macro point of view, the mathematical model based on the prior knowledge of image degradation plays a critical role in the haze image restoration. It explains the reason of image quality degradation from the perspective of image essence and has more advantages than other algorithms.

A new method was proposed to calculate the structural depth of field information of the current scene based on multiple images obtained under different weather conditions in the same scene to reconstruct clear images (Narasimhan, & Nayar, 2003). Schwartz proposed to calculate the ambient light, depth of field and other information based on multiple polarization images in the same scene (Shwartz, 2006). The algorithm has achieved good results in image defogging, but the disadvantage is that it needs to input additional image information, which limits its practical applications.

In order to overcome the weakness of the above algorithm, a defogging algorithm was proffered for grayscale and color image contrast enhancement from the perspective of enhancing image contrast, we assume that atmospheric light changes tend to be smooth with the increase of depth of field (Tan, 2008). In this algorithm, firstly, it is assumed that the surface shadow and the medium transmittance of the scene are not related to mathematical statistics, the reflectivity of the scene surface is estimated, the reflectance of the scene surface is estimated to derive the transmittance of the environmental light, resulting in image noise reduction (Fattal, 2008).

From the perspective of the differences between the outdoor clear image and the haze image, based on the analysis of a large number of clear images, the prior theory of dark primary color channel of image was recommended, the theory was applied to the image clarity and haze removal, the algorithm achieved satisfactory results (He, 2011). The light attenuation function is able to approach the maximum value, the local area of the image changes gradually, a fast and environment-related image defogging algorithm (Tarel, 2009) was proposed.

2.2.2 Traffic Sign Location Detection

The algorithm of traffic sign location is designed according to the characteristics of traffic signs. For example, in order to implement the recognition of speed limit signs, we need firstly segment the speed limit signs from the complex scene image to obtain the Region of Interest (ROI) from the traffic scene image that is to locate the position of the speed limit signs in the original scene image and extract them (Liang, 2013). Speed-limit sign has the characteristics of red circle, white background, black character,

symmetry, circle, and so on. Most of the detection methods primarily utilizes the color characteristics of the sign to obtain the area that conforms to the sign color, so as to narrow down the processing range. In these regions, the geometric features are used to find circular candidate regions, the region of interest is further reduced. Finally, the classifier was trained in advance which is used to determine whether these candidate regions contain speed limit signs, so as to realize the accurate positioning of speed-limit signs.

Color-based methods include direct color threshold segmentation algorithm proposed by Escalera (De, 1997) in 1997. In the Red, Green, Blue (RGB) color space, all the pixels of the image are segmented in a form by using the correlation algorithm, the corner detection method is used to judge the object area to determine whether the object area is a traffic sign. However, in RGB color space, light has great influence and interference, Arroyo (Maldonado-Bascon, 2007) converted RGB image to Hue, Saturation, Intensity (HIS) color space, histogram is applied to analyze saturation and hue. Finally, the color segmentation algorithm based on the corresponding color threshold is completed in HSI color space (Escalera, 2003). Besides colors, shape is another characteristic of traffic signs, the specifications of corresponding types of traffic signs used in a country should be unified throughout the country, even if the signs such as speed limit and no traffic have the same shape characteristics in different countries.

After obtained the candidate region by color, the geometric features of speed-limit sign, such as circle and symmetry, are used to further determine the candidate region. Indicator signs mainly provide certain information guidance, the function of indicator signs is gradually replaced by the navigation system. The edge chain coding and nonlinear least square estimation (Berkaya, Dunduz, Ozsen, Akinlar, & Gunal,2015) were employed to determine the circular traffic sign area. Hough transform based on gradient information (Psyllos, 2011) is proffered to detect the circular region in the image. Hough transform has a good effect in finding the circle, but it has a huge amount of calculation and high time complexity. At the same time, in the actual traffic scene, based on the observation angle of different traffic participants and whether the signs are placed regularly or not, the shape characteristics of traffic signs are not all regular

geometry compared with the perspective of the observer.

At this time, if we only judge by the shape of the sign, there will be a lot of interference factors in the actual scene. Therefore, most of the current shape feature-based detection algorithms are combined with color feature-based detection algorithms. For example, the traffic sign detection (Li, 2018) was completed according to the relevant color threshold in RGB space and the spatial features of shape contour.

2.2.3 Traffic Signs Recognition Algorithm

The last step of TSR is to classify the candidate regions obtained in the detection stage, judge their categories, and instruct the drivers. Due to the complexity of natural scenes where traffic signs are located, the variety of traffic signs, many traffic signs have low discrimination, which brings great challenges to the research of sign recognition algorithm (Stallkamp, 2012).

Most of early classification methods are based on template matching. This method needs to summarize the invariant and similar features of traffic signs in different scenarios, the corresponding feature extraction algorithm is executed to extract the feature representation of traffic sign image, and finally the matching algorithm is run for matching classification. The feature representation of this method needs to be set well, it is a tough problem to set the image feature representation precisely. Because of the complexity of scenes, the accuracy of matching algorithm is often affected by various factors in the actual scene.

With the constantly updating of satellite technology, traffic object detection based on satellite images has been investigated. In the early time, visual object recognition of satellite image was based on traditional methods. Subsequently, classification algorithms are mostly based on SVM, AdaBoost, and neural networks (Carrasco, 2012).

A speed-limit sign recognition algorithm was proposed based on SVM (Saadna, Behloul, & Mezzoudj, 2019), in order to increase the recognition rate, a self-learning particle swarm optimization algorithm was added to optimize the relevant parameters of SVM by adjusting the learning parameters to coordinate the global and local search ability of particles in the evolution process. Road extraction was based on remote

sensing image according to the geometric, radiation, and topological characteristics of roads (Huang & Zhang, 2009). The decision tree classifier is to segment the input image recursively. Its branches represent different segmentation paths, the leaves show the final classification results.

AdaBoost algorithm is an iterative algorithm, which cascades multiple weak classifiers to form a strong classifier. AdaBoost algorithm was fused with SVM classifier to implement accurate and real-time classification of traffic signs (He, & Xu, 2010). Vehicle detection was based on high-resolution satellite images (Eikil, & Aurdal, 2009). From the experimental results, it was said that though the image resolution is low, the detection result of the algorithm is ideal and close to the result of artificial classification. AdaBoost algorithm was based on Haar features to identify vehicles and combined with line detection technology so as to detect single vehicles (Leitloff, 2010). Compared with the method based on statistics only, the accuracy of this method was improved around 80%. Although the traditional methods have achieved satisfactory results in object recognition based on satellite remote sensing images, they need to specify visual features, the design is complex and lack of robustness to the diversity of visual objects.

At present, in the field of TSR, the most widely adopted classification algorithm is from artificial neural network. The neural network classifier is composed of two layers of networks (Wang, Yang & Xu, 2003). In the first layer, a BP network is employed to realize the rough classification of traffic signs; in the second layer, three BP networks are applied to subdivide the rough classification results and realize the recognition of traffic signs. In recent years, an algorithm combining Hu moment invariants and wavelet neural network was proposed to recognize objects. By extracting Hu moment invariants of images, wavelet neural network was applied to classify traffic signs (Qin, Huang, & Liu, 2005). Compared with BP neural network, wavelet neural networks have stronger learning ability and generalization ability.

With continuous development of artificial intelligence, the field of pattern recognition has made a breakthrough. Many experts have taken a keen interest in satellite image recognition. A fully symmetric convolutional neural network (Audebert,

2016) was profound to obtain details about all the low-level information, thus commence the semantic segmentation task of high-resolution remote sensing image. A multipath deconvolution method (Volpi and Tuia, 2017) was proposed to acquire low-level details and judge the edge of the object accurately. As a new field, there are still a plethora of problems in object recognition using this method. The ordinary images (Sherrah, 2016) are supplied to pretrain FCN, and test remote sensing images, which effectively improve the accuracy of object recognition in satellite images.

In addition, convolutional neural network has been paid more and more attention because it was automatically applied to extract all levels of features of objects without specified operations. A multiobject detection framework with rotation invariant convolution neural network (Gong, C, P. Zhou, & J. Han, 2016) is able to effectively detect various objects in remote sensing images which has been verified with a high performance. Multiscale convolution neural network to recognize speed-limit signs has achieved a remarkable recognition rate of 99.2% (Sermanet and Lecun, 2011). Therefore, deep learning has been employed to recognize traffic signs from satellite images, there is a sufficient room for developing it further.

2.3 Convolutional Neural Network

The inspiration of convolution neural network comes from the physiology of cat visual cortex. There are neuron cells which are extremely sensitive to external input in a certain area of cat visual cortex, this area is called receptive field (Hubel, 1962). From then on, visual cortex began to enter the field of view (FOV) and attracted attention. Convolutional neural network was firstly applied to handwritten digit recognition. The data model (Ripley, 1996) was trained with CNN and achieved good results in the experiment, which created a precedent for convolutional neural network to be widely employed in image, speech recognition, face recognition, and other fields.

Convolution neural network is improved and promoted on the basis of artificial neural network. Convolutional neural network consists of five parts: Full connection layer, convolution layer, input layer, output layer, and pooling layer. (Hubel, & Weisel,

1968).

2.3.1 Convolution Layer

The function of convolution layer is to extract the features of input data. The introduction of convolution layer reduces the number of weights that the traditional network needs to learn and relieve the burden of memory and calculation.

In order to enrich the extracted image features, each convolution layer contains multiple convolution kernels. The convolution parameter training process is consistent with the traditional neural network. Firstly, the output value of the network is obtained by forward propagation calculation, then the difference value is obtained by comparing with the actual value. The error of each node in each layer is calculated by backpropagation, the weight of the learning is optimized according to the obtained error.

Convolutional layers reduce the number of parameters mainly in two ways: One is to introduce the idea of local perception field (Hubel,1962), there is a correlation between TSR from the local to the global point of view. However, the correlation between pixels in the spatial adjacent domain is quite close, while the correlation between pixels far away is relatively weak. From this point of view, each neuron in each layer needs to perceive only the features of its own layer, only by perceiving the local image. Finally, the global information of the image is obtained by integrating the local information obtained in the next layer. After this operation, training parameters can be reduced. However, if the amount of training data is very large, there are still a plenty of parameters to learn after convolutions, there will be a great deal of problems in this operation. If each neuron only corresponds to a part of the image, the extracted content of this neuron cannot be applied to other neurons. In this case, the weight sharing method is taken. By keeping the parameters of each neuron consistent, the accuracy of the model will not be affected by the change of object position. The implicit principle is that local statistical characteristics of the image are consistent with other parts, we apply the local learning features to other parts.

The most important parameters of convolutional layers are the size of input image,

convolution step size, convolution kernel size, filling size, and so on. The specific operation process of convolution is shown in Figure 2.1. The 5×5 image in the figure representants the input feature map, which is denoted by letter B. The grayscale part (3×3) represents the convolution kernel, which is marked by the letter K. The length of each movement of convolution kernel is determined as the step length, which is expressed by letter s. The step size of convolution kernel moving horizontally is as same as that of convolution kernel moving vertically, where the step length is 1. Edge filling is not indicated. The default filling value is generally 0, the letter q is applied to indicate the size of edge filling.

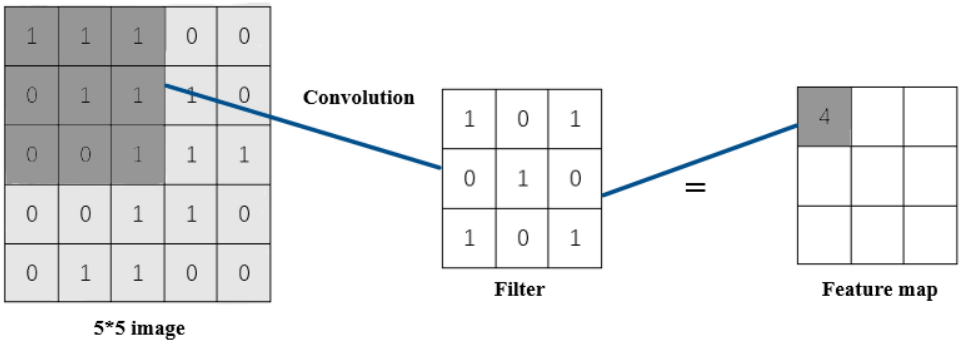


Figure 2.1: Convolution operation process

In general, the concept of convolution layer is proposed because convolutional layer has the characteristics of automatic image extraction and the number of parameters in the network are reduced, assists us to solve the difficult problem of feature extraction and huge training parameters.

2.3.2 Pooling Layer

The subsampling layer, also known as pooling layer, is usually located behind the convolutional layer. The main function of pooling layer is to reduce the useless feature information, retain the effective information, reduce the dimension of features, and achieve the purpose of compressing the number of data and parameters. On the one hand, reducing the size of the feature map is able to alleviate the computational complexity. Although the amount of computation is diminished after the previous convolution operation, it is still faced with the challenge of computational complexity

due to the complexity of the training data.

On the other hand, further reducing the number of parameters by pooling can prevent over fitting. Max pooling and average pooling are the two most popular pooling methods in pooling layer, their essence is a special convolution process. The specific operation is shown in Figure 2.2.

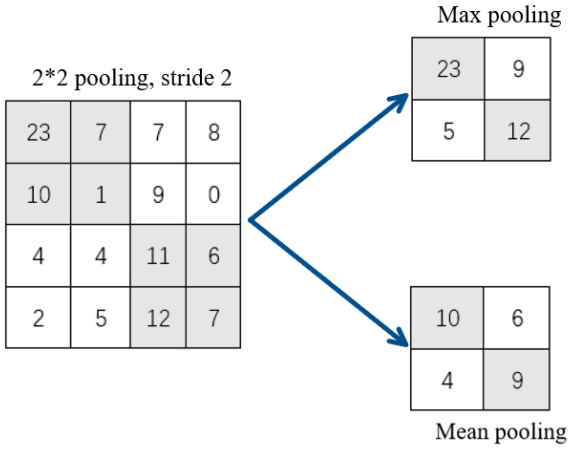


Figure 2.2: Pooling operation process

The upper side shows max pooling, only one of the weights in the convolution kernel is 1, the other weights are set to 0. The effect of maximum pooling is to cut the images down to a quarter of its original size, in which the maximum value of each 2×2 region is retained. The lower side represents average pooling, the weights set in the convolution kernel are all 0.25. The average pooling of the convolution kernel is to weaken the blur of original image to 1 / 4 of the original.

2.3.3 Full Connection Layer

The full join layer is in the last part of convolutional neural network. After a series of processes, the feature vectors of input data are extracted. The full connection layer allows all extracted features to be combined and turned into output values. That is to say, for all the neurons in CNN, it is necessary to connect them with all the neurons in the upper layer (Liu, Jia, Zhao, & Liu, 2019), finally transform the extracted features into an n × 1. The schematic diagram of the connection layer is shown in Figure 2.3.

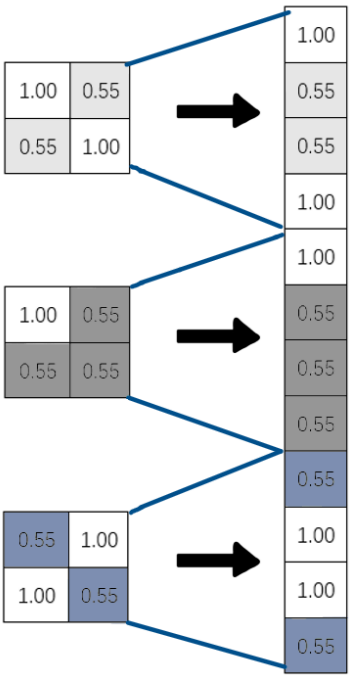


Figure 2.3 The principle of fully connected layer

2.3.4 Activation function

Activation function is a nonlinear factor in order to solve the problem of insufficient expression ability of linear model (Ajmi, Pérez, Ferchichi, & Zaafouri 2020). In the structure of neural network, an activation function needs to be added after each layer is superimposed. The activation functions of neural networks usually take use of nonlinear functions, such as sigmoid function, tanh function, and ReLU function. The expression of sigmoid function is shown in Equation (2.12).

$$h(a) = \frac{1}{1-e^{-a}} \tag{2.12}$$

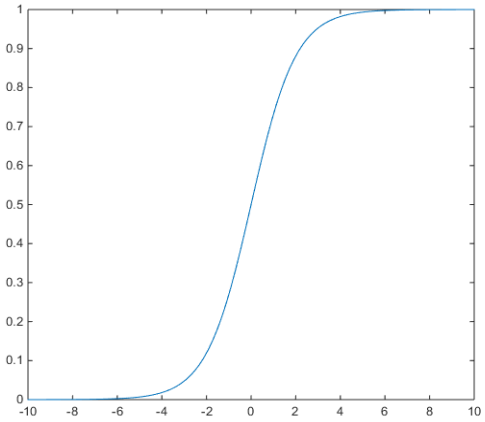


Figure 2.4: Sigmoid function

From the image of sigmoid function, we see that the function is employed to transform the input value into the range of 0 to 1.0, from the image, we see that the function is continuous, monotonic, and easily to be operated. Sigmoid function is broadly applied in the past, but it has two defects. Firstly, the output value of Sigmoid function is not symmetric with 0, which makes the calculation much complicated. Secondly, the function has soft saturation. If the input value is too large or too small, the weight gradient will be slowly changed to zero, that is, the gradient will be disappeared. Hyperbolic tangent function is evolved from sigmoid function, which is expressed in Equation (2.13).

$$h(a) = \frac{e^{a} - e^{-a}}{e^{a} + e^{-1}} \tag{2.13}$$

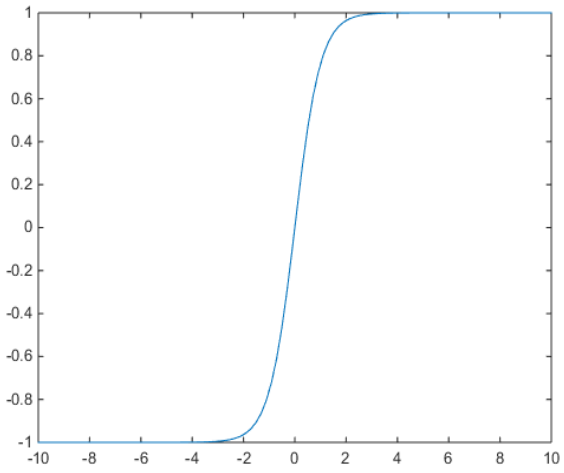


Figure 2.5: Tanh function

Tanh function is more popular used than sigmoid function in practice. Compared with sigmoid function, its output is 0-centered symmetry, the derivative of sigmoid function is much smaller than the derivative of tanh function, which shows that tanh function converges faster in the convolution layer, the obvious gradient change is during the training process. But the gradient vanishing problem due to activation fumctions like sigmoid function and tanh function still exists.

The linear corrected units (ReLU) are the most important activation function in recent years, its expression is shown as Equation (2.14).

$$f(x) = \begin{cases} 0, x \leq 0 \\ x, x > 0 \end{cases} = \max(0, x) \tag{2.14}$$

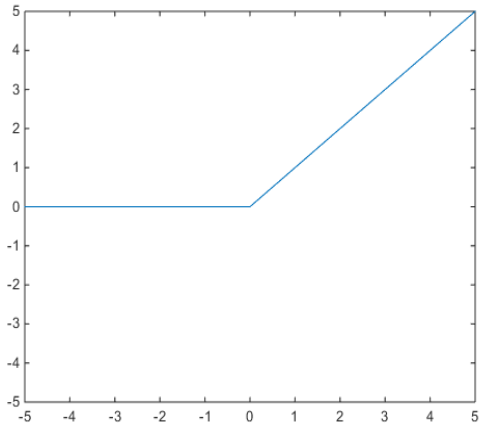


Figure 2.6: ReLU function

The image of ReLU function is shown in Figure 2.10. In Figure 2.10, we see that the calculation process of ReLU function is much simpler than that of tanh function and Sigmoid function. ReLU is particularly fast because it only needs to determine whether the result is greater than 0. In addition, compared with the first two activation functions, the convergence speed is obviously faster. Moreover, the problem of gradient vanishing is effectively alleviated while the amount of calculation is reduced.

Comparing the above three activation functions, we see that the output of sigmoid function is a positive real number, the output is the largest one around 0.0. The output of tanh function is a positive number or a negative number, while the input of ReLU function is only a number greater than 0.

2.3.5 Optimizer

In the process of backpropagation in deep learning, the optimizer constantly adjusts the parameters of loss function to make the input better fitting for the output and ensure the loss function approaches the global minimum. In essence, it is the problem of finding the optimal solution of functions, the core idea of current optimizer is to use the gradient descent method to optimize the parameters. The fastest converging direction of function value is the gradient direction. When deep learning model is set to resolve the minimum value of objective function, the direction moves along the descending direction of gradient, the results continuously approximate the optimal value. The two most important parameters of the optimizer are optimization direction and step size. The optimization direction is reflected as the gradient or momentum in the optimizer, the step size is reflected as the learning rate in the optimizer. The popular optimizers are Stochastic Gradient Descent (SGD), AdaDelta algorithm, and Adaptive Moment Estimation (Adam), Mini Batch Gradient Descent (MBGD), Adaptive Gradient Algorithm (AdaGrad), Batch Gradient Descent (BGD), etc.

The process of optimizing the parameters is split into four steps: First, the objective function, which we need to know how to calculate, corresponds to the gradient of the current parameters; Then the different momentum is calculated according to the parameter gradient; third, the descent gradient at the current time is calculated, finally the parameters are updated according to the descent gradient. In the last two steps, the algorithm is consistent, only there are differences in the first two steps. The gradient descent algorithm selects a sample randomly from the training set and does not involve momentum. The advantage of the algorithm is that it only learns one sample at a time, the training speed is fast.

It satisfies the local optimum or saddle point if the gradient is 0, which makes it impossible to update the parameters. A very primitive method in batch gradient descent experiments, which takes use of all the samples in each update. It provides a better representation of all the sample data and indicate more accurately where the extreme points are. However, this method will calculate the whole dataset each time while it is

being updated, which leads to the increase of computation and cannot add new samples. The improved small batch gradient descent method is based on BGD and SGD methods. A small number of samples from the dataset are selected to calculate the losses and update the network parameters.

On the one hand, it reduces the variance of parameter updating and makes the convergence stable. At the same time, it makes full use of the data for more effective gradient calculation. The disadvantage is that it cannot guarantee good convergence, which is essential to set a suitable learning rate. If the setting is too small, the convergence speed is too slow; if the setting is too large, it is easy to oscillate around the minimum. Adaptive learning algorithm adds the concept of the second-order momentum. SGD algorithm and its variant algorithm adjust each parameter for the same learning rate.

Deep neural network often contains a large number of parameters, not every parameter will be used. Therefore, through the adaptive learning rate algorithm, we set different learning rates according to the specific situation. Regarding frequently updated parameters, the learning rate is slower. With regard to the occasionally updated parameters, we hope to learn from each accidental sample, the learning rate should be larger. However, if iterated many times, the learning rate will be decreased, the second-order momentum will accumulate continuously, which will lead to the end of the training process ahead of time.

The adjustment of learning rate in AdaGrad algorithm is too radical, AdaDelta algorithm puts forward the idea of not only accumulating historical gradients but also focusing on the descending gradient of time window in the past. The Adam algorithm has a pile of benefits that controls the learning rate length and gradient direction in the algorithm, which prevents the problems such as gradient oscillations and saddle point stationarity.

2.4 Overview of Traffic Signs Recognition

2.4.1 Traffic Signs Recognition Process

In Figure 2.7, we show the workflow of TSR, which mainly includes three parts. The first part is for image preprocessing, which usually encapsulates image enhancement, image scaling, and other operations. The second part is the traffic sign detection, which includes two important steps: (1) The candidate regions are extracted; (2) The second classification of traffic signs is whether they are traffic signs or not; (3) The third part is the detailed classification of traffic signs, namely, the specific category of traffic signs.

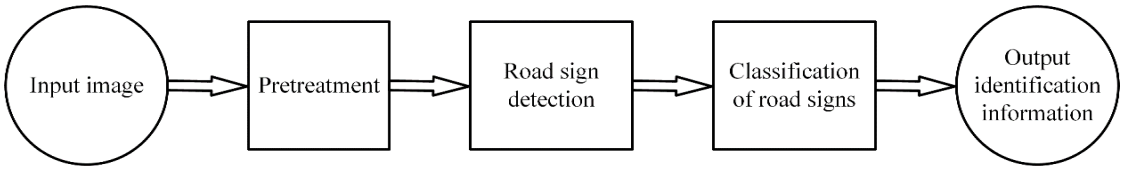


Figure 2.7: Workflow of TSR systems

2.4.2 Image Preprocessing

TSR is different from general object recognition. The irresistible influence of external factors such as haze, rain, and snow, all kinds of lighting conditions will affect the traffic sign image we get, which will make the image brightness too low or with noises, which will not be conducive to TSR. Image preprocessing effectively improves the quality of the image, reduces or even eliminates the interference factors that are not conducive to object recognition, and improves the efficiency and accuracy of traffic signs detection. Therefore, in order to improve the recognition accuracy of traffic signs, the image needs to be preprocessed prior to the sign recognition.

2.4.3 Traffic Sign Detection

The second part of TSR is traffic sign detection, which mainly includes two processes: (1) Extraction of candidate frames, (2) classification. The extraction of candidate box is to segment the part containing traffic signs in the original image according to the relevant features of traffic signs. The second classification is to extract the features of

the signs and determine whether the signs belong to the corresponding areas. The methods of traffic sign detection include shape detection, color segmentation, feature extraction combined with classifier and deep learning detection algorithm.

2.4.4 Classification of Traffic Signs

The third part is traffic sign classification, which is applied to determine the specific category of the segmented traffic signs area. Traffic signs classification is essentially multi-objects classification. The traditional classification methods encapsulate the combination of feature vector and classifier, template matching. With the continuous development of artificial intelligence, the application of deep learning in TSR has achieved better results (Yan, Zhao, Diao, & Wang, 2021). Compared with traditional methods, deep learning classification method is able to automatically learn features without human intervention. The more training data, the better classification performance. In particular, the appearance of convolution neural network enhances the ability of neural network feature extraction. It has become one of the most representative networks in deep learning, so it is widely applied to object classification.

Chapter 3
Methodology

In this chapter, we firstly introduce defogging algorithms, then we propose the defogging algorithm based on the improved HSV color gamut and the algorithm for guiding image filtering. We will compare these two algorithms. Then, we introduce how to adjust the parameters of Faster R-CNN and how to select the network. Afterward, we bring in the structure of YOLOv5 and how we can improve YOLOv5. Finally, we detail the evaluation methods.

3.1 Image Defogging Preprocessing Algorithm

According to the introduction and analysis of the previous chapter, we know that the TSR has a vast number of applications. There have been many breakthroughs in all TSR because of the growth of deep learning, but much of the research has centered on TSR in good visual conditions, the identification of signs in fog weather still needs further research. In this chapter, we will study the TSR in foggy weather from the perspective of fog image. The roadside identification in the whole foggy weather consists of four steps as shown in Figure 3.1.

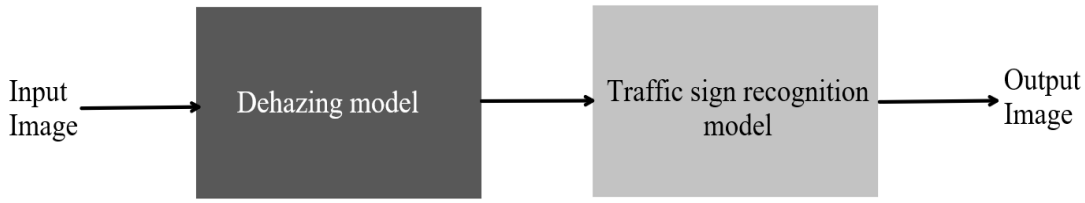


Figure 3.1: TSR process in foggy weather

As shown in Figure 3.1, there are two processes for TSR in foggy weather: (1) The fog in the image is removed by using the defogging model to get the clear image without fog. (2) After the previous step, we obtained the image after dehazing, we put it into the deep learning model for training, and then our model will recognize the traffic signs to be detected in the image. The following sections will be expanded for this process.

3.1.1 Dark Channel Prior Defogging Algorithm

The dark channel prior defogging algorithm (He, Sun, & Tang, 2011) was firstly proposed in 2009. The corresponding work has been awarded as the best paper award of IEEE CVPR in the same year.

In the dark channel, prior to defogging algorithm, the outdoor sunny images are analyzed and counted, in the non-sky part of these images, One or two of the three RGB color channels of each image have very low greyscale values or intensity. There are four reasons for this situation: The shadow of all kinds of glass, the projection of natural

objects. The brightly colored surface of visual objects, a dull surface of visual objects. As explained, the dark channel is shown as Equation (3.1).

$$J^{dark}(x) = \min_{y\in\Omega(x)}(\min_{c\in[r,g,b]} J^c(y)) \tag{3.1}$$

where a color channel in r,g,b is represented by C, Ω(x) represents the local area centered on *x* pixels (Chen, Ou, & Tian, 2019), J^{dark} is the image dark channel graph. At the same time, from the above prior knowledge, we can know that the gray value of the pixel in the dark channel image is very low, which is almost 0. So J^{dark} tends to 0.

In Equation (3.1), atmospheric light is assumed to be a known quantity. In fact, for any input image, 0.1% of the maximum grayscale intensity of pixels in the dark channel image correspond to the average grayscale value of pixels in the corresponding position of each channel of the original image, so as to obtain the atmospheric light value of each channel.

On the premise that atmospheric light is assumed to be known, the atmospheric scattering model is transformed into,

$$\frac{I^c(x)}{A^c} = t(x)\frac{J^c(x)}{A^c} + 1 - t(x) \tag{3.2}$$

where C means that each channel needs to be processed separately. At the same time, we start by considering the light transmission $t(x)$ as a constant and this constant is $\tilde{t}(x)$, then, the two sides of Equation (3.2) are filtered twice to export the final minima.

$$\min_{y\in\Omega(x)}(\min_{c}(\frac{I^c(x)}{A^c})) = \tilde{t}(x)\min_{y\in\Omega(x)}(\min_{c}(\frac{J^c(x)}{A^c})) + 1 - \tilde{t}(x) \tag{3.3}$$

where $\tilde{t}(x)$is a constant, the minimum value of $\tilde{t}(x)$ is calculated, J represents the original image to be solved, from the previous dark channel prior principle, we see that J^{dark} is close to 0. Combined with Equation (3.1), we have

$$\min_{y\in\Omega(x)}(\min_{c}\frac{J^c(y)}{A^c}) = 0. \tag{3.4}$$

By substituting Equation (3.4) into Equation (3.3), the estimated value of transmittance$\tilde{t}(x)$can be obtained. The calculation is conducted as follows,

$$\tilde{t}(x) = 1 - \min_{y \in \Omega(x)} (\min_{c} \frac{I^c(y)}{A^c}). \tag{3.5}$$

In actual cases, even in sunny days with good line of sight, there will still be tiny droplets and aerosol particles in the atmosphere. If all the fog is removed, it will have an impact on the realism of images. Therefore, a factor ω is introduced into Equation (3.5), whose value is between [0,1], Equation (3.5) thus becomes

$$\tilde{t}(x) = 1 - \omega \min_{y \in \Omega(x)} (\min_{c} \frac{I^c(y)}{A^c}). \tag{3.6}$$

The factor ω is generally 0.950 (Shi, Zhang, Zhou, & Cheng, 2021). At the same time, if $t(x)$ is very small, the value of J will be too large, resulting in a lot of noise in the whole image. Therefore, a threshold t_0 should be set, and if $t(x)$ is less than t_0, let $t(x) = t_0$. So, the final recovery is shown in Equation (3.7)

$$J(x) = \frac{I(x) - A}{\max(t(x), t_0)} + A \tag{3.7}$$

Figure 3.2 is a summary of the haze removal process of the dark channel prior algorithm.

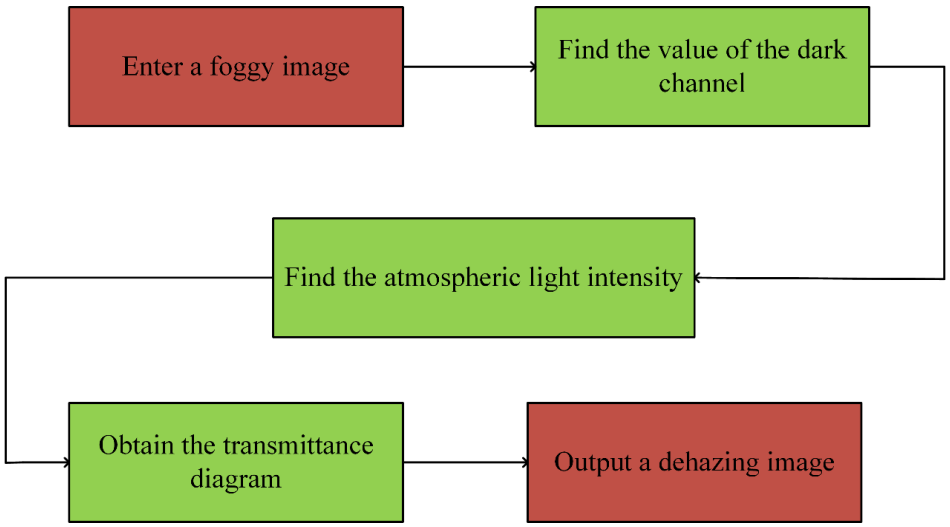


Figure 3.2: The workflow of the prior algorithm for fog removal

3.1.2 Histogram Equalization Defogging Algorithm

Histogram is a statistical report graph from the image meaning. It takes use of vertical line segments with different heights to represent the distribution of relevant statistical data (Setiawan, Mengko, Santoso, & Suksmono, 2013).

In image processing, histogram is a kind of spatial domain methods, the operation of image histogram is able to achieve the effect of image enhancement. Histogram equalization is a statical method. The mapping curve determined by the histogram is the cumulative distribution of the corresponding image.

Taken grayscale images into account, if the grayscale level of images is in the range [0, *L*-1], its histogram is a discrete function, which is expressed as Equation (3.8),

$$h(r_k) = n_k \qquad (3.8)$$

where k grayscale level is represented by using r_k, n_k represents the total number of pixels of this gray level whose gray level is r_k. Therefore, a normalized histogram is obtained by dividing the total number of pixels using the total number n of each level

$$P(r_k) = \frac{n_k}{n} \quad . \qquad (3.9)$$

The color image has its own histogram representation in RGB channel. MATLAB is applied to display the histogram of the image and its RGB three components in the real scene, as shown in Figure 3.3 and Figure 3.4.

Figure 3.3: Images taken in real scenes

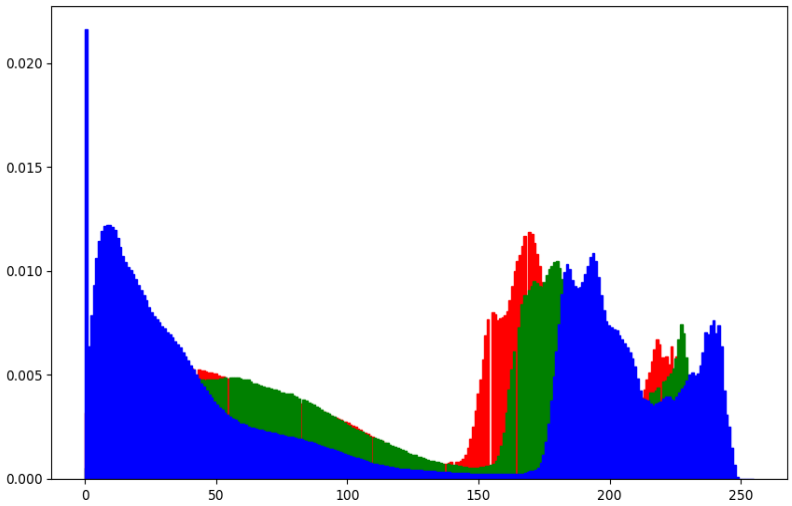


Figure 3.4: RGB component histogram

After analyzed the histograms of a huge number of images, it is found that if the image is too dark as a whole, the main components of the histogram will tend to the lower gray level part of the image. If the whole image is bright, the histogram will tend to the part with higher gray level. If the contrast of the image is low, the histogram range will be narrow, the distribution will be mainly in the middle of the grayscale intensity level. If the contrast is high, the histogram will have a wide range of grayscale levels.

According to the above analysis, we know that if we want an image with large dynamic range and large contrast, the pixels of the image should occupy more grayscale levels and be evenly distributed. Therefore, we need a transformation function to achieve this effect, which can be expressed as $T(r)$.

In the original image, each pixel r has a corresponding grayscale intensity s. Then $T(r)$ should be single valued and monotonically increasing in the range of [0,1]. Moreover, if r is in the range of [0,1], $T(r)$ is also in the interval [0,1]. The expression of $T(r)$ is shown as Equation (3.10),

$$s = T(r) = \int_0^r p_r(\omega) d\omega \tag{3.10}$$

where ω is the integral variable, the cumulative distribution regarding random variable

r is expressed as Equation (3.10), where $P_x(x)$ is the probability density function. Then $P_r(r), P_s(S)$ represent the probability density functions of r and s respectively. The results are shown as Equation (3.11).

$$P_s(s) = P_r(r) \left| \frac{dr}{ds} \right| \quad (3.11)$$

If $T(r)$ is given, $P_s(s)$ is obtained by using Equation (3.11). At the same time, according to the Leibniz criterion, we know that the derivative of the definite integral about the upper limit is the integral value of the upper limit.

$$\frac{ds}{dr} = \frac{dT(r)}{dr} = P_r(r) \quad (3.12)$$

Substituted Equation (3.12) to Equation (3.11), if the probability value is positive, we get that if s is in the range of [0,1], $P_s(s) = 1$, the value is 0 in other cases. The result of $P_s(s)$ is always uniform, regardless of the form of $P_r(r)$.

Histogram equalization takes use of the same histogram transformation method for whole image, which achieves good results if the image pixels are evenly distributed, but if the image contains obvious excessive or too dark regions, it may not achieve the ideal effect.

In order to solve this problem, adaptive histogram equalization is proposed. Adaptive histogram equalization is to calculate the transform function of each pixel to enhance the histogram of each pixel. In this method, the image will be divided into rectangular networks, the optimal contrast of each sub grid block needs to be calculated, generally segmented into 8 ×8 region to get the best result.

The adaptive histogram equalization method essentially calculates the histogram for each salient region of the image and then rearranges the brightness of the image, so it is ideal for increasing the local brightness in the image. However, this method also has the disadvantage that the enhancement must be read and will enhance very bright areas of the image, adding noise.

3.1.3 Defogging Algorithm Based on Improved HSV Gamut

Because adaptive histogram equalization will enlarge the noises in the homogeneous region of the image, an adaptive histogram equalization algorithm CLAHE was proposed (Yadav, Maheshwari, & Agarwal, 2014) with limited contrast. The difference between CLAHE and AHE (Kurak, & Charles, 1991) is that CLAHE limits the contrast amplitude. In the CLAHE algorithm, in order to overcome the defect of excessive noises caused by AHE, each small region needs to limit the contrast separately.

The specific improvement of CLAHE is to limit the contrast enhancement process of AHE. The contrast magnification around the specified pixel value in AHE depends on the slope of the transform function. The slope of the transform function is proportional to the slope of cumulative distribution function. So, limiting the slope of CDF (Cumulative Distribution Function) is equivalent to limiting the contrast.

In order to limit the slope of the cumulative distribution function, the histogram obtained from the sub block statistics should be trimmed by using the preset threshold before calculating the CDF. At the same time, the value of the clipping part is evenly distributed to the whole interval of the histogram to make the total area consistent with that before processing.

At the same time, in AHE and CLAHE, if only the mapping function in the block is applied to transform the pixels in the block, the final image will produce mosaic artefacts. In order to solve this problem, we need to use interpolation in the process of change. The process of interpolation is shown in Figure 3.5.

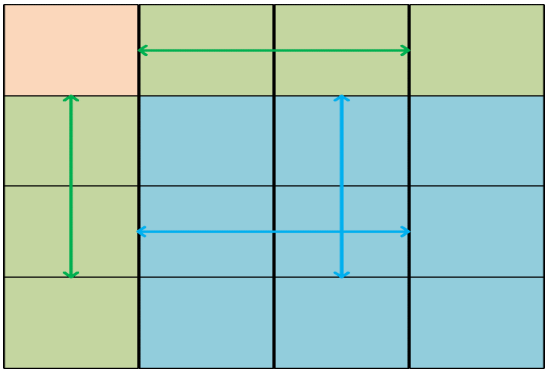


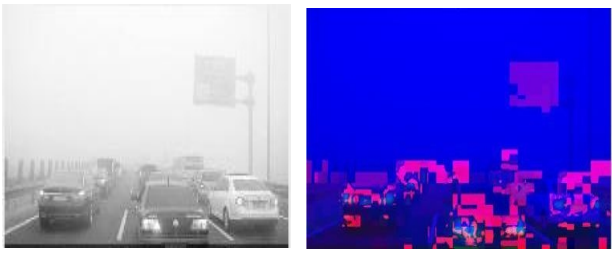
Figure 3.5: The diagram of interpolation operation in the process

The method to get the pixels in the blue area is: The mapping function of the four sub blocks around it is transformed to get the corresponding mapping values, and then the four mapping values are interpolated by using bilinear interpolation. The green region is the linear interpolation operation of the mapping value after the transformation of two words blocks. Finally, the red region can be detected by using the mapping value of the block directly.

In the process of CLAHE algorithm for color image processing, it is generally necessary to process RGB channels separately. Figure 3.6 shows the result of fog removal by using CLAHE algorithm for RGB color image of highway scene.

Figure 3.6: Removing fog from digital images with CLAHE algorithm in highway scene, (a) is the original image, (b) is the image after removing fog

In Figure 3.6, after RGB gamut is processed directly, the image may have color deviation problem, the use of CLAHE algorithm for defogging needs to calculate RGB channels at the same time, the operation is not very convenient. By considering the complexity of this operation, we remove the fog in HSV gamut. After the gamut conversion of the foggy image in Figure 3.6 (a), the three components and HSV values are shown in Figure 3.7.



(c) (d)

Figure 3.7: HSV diagram and H, S, V component diagram, where (a) H component diagram, (b) S component diagram, (c) V-component graph, (d) HSV graph

In HSV gamut, the meanings of the three components are: H stands for hue, S is saturation, and V refers to lightness. According to the theory of atmospheric scattering model (Ju, Ding, Ren, Yang, Zhang, & Guo, 2021), V component should be the main processing object. Most of the existing algorithms deal with three channels or only V channel. After analyzing the color changes of foggy and non-foggy images in natural environment, we find that there is little difference between foggy and non-foggy images in H-channel, but there is a big difference between V-channel and S-channel. So, we only deal with V and *s* channels and keep H channel unchanged. In this thesis, Python language is employed to implement the platform. The ideas are shown in Figure 3.8.

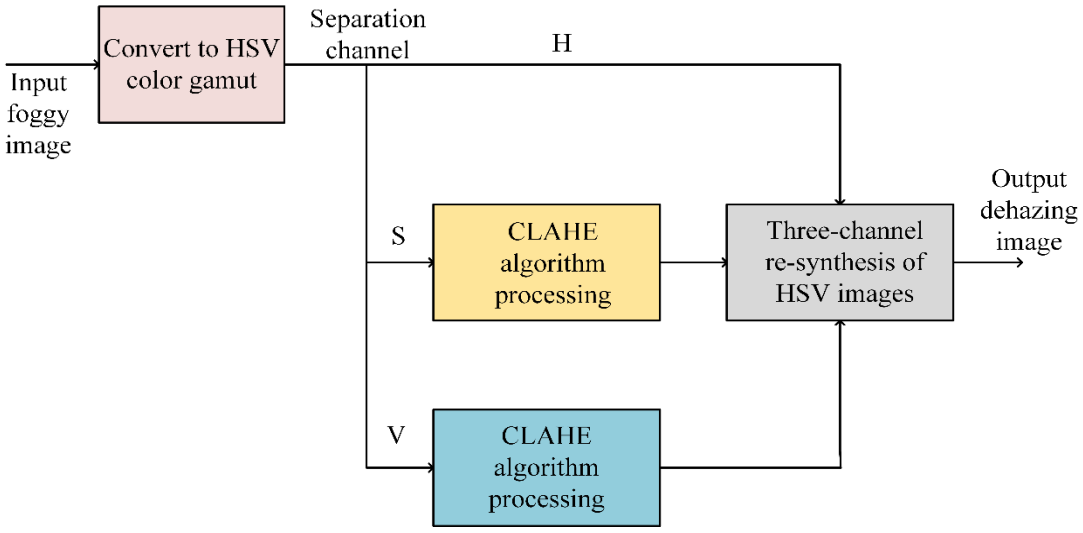


Figure 3.8: The framework of improving HSV gamut defogging algorithm

3.1.4 Guided Image Filtering

Image preprocessing is an essential step in the defogging process. In this step, we not only remove excess noises, but also enhance the image. The early image defogging algorithm is elementary and mainly divided into two types. One is based on histogram, a balanced defogging algorithm, another method is a defogging algorithm based on image restoration (Illingworth, & Kittler, 1988). This method compares the foggy image with the original image to create a new image. Although this method has an excellent defogging effect, the image after defogging is challenging to achieve the quality of the original image.

In this thesis, we are use of a new explicit image filter which is called guided filter. Our experiments show that guided image filtering has the function of defogging in TSR. Guided image filtering effectively solves the shortcomings of the two earlier methods. This method is introduced with the concept of a guided filter map. The guided filter map is an image downloaded from the Internet or the dehazing image itself, and we filter the image using the texture of the guided filter image. The final output image is as same as the target image, the quality of the image after defogging is the same as that of the original image. We are use of the characteristics of guided image filtering to achieve the effect of defogging and noise reduction of traffic signs.

In this experiment, we define the original image as p_i, the guiding image as l_i, and the output image as q_i. The relationship is linear as shown in Equation (3.13).

$$q_i = a_k I_k + b_k \quad i \in \omega_k \tag{3.13}$$

where a_k and b_k are specific factors, ω_k is a square window with a center point *k*, $i \in \omega_k$ guarantees that a_k is not too big. In order to ensure the guided image filtering has the best outcome, the difference between the original image and the output image needs to be minimized. Therefore, the cost function $E(a_k, b_k)$ is defined as Equation (3.14),

$$E(a_k, b_k) = \sum_{i,k\in\omega_k}(|(q_i - p_i)^2 - \varepsilon {a_k}^2|). \tag{3.14}$$

We see from equation (3.15) that a_k and b_k are obtained by using the least square method, where the minor $E(a_k, b_k)$, the better the output result.

$$a_k = \frac{\frac{1}{|\omega|}\sum_{i\in\omega_k} I_i q_i - u_k \bar{p_k}}{\sigma_k^2 + \varepsilon}, b_{k=}\bar{p_k} - a_k u_k \tag{3.15}$$

where u is the mean of I in W, σ is the variance of I in W, w is the number of pixels in the window. We input a_k and b_k into Equation (3.13),

$$q_i = \frac{1}{|\omega|}\sum_{i\in\omega_k}(a_k I_k + b_k) = \bar{a_i} I_i + \bar{b_i}. \tag{3.16}$$

3.2 Faster R-CNN Model for Traffic Sign Recognition

Visual object detection is an important application of deep learning, which is to mark the position of the object in the image, and put the object in the image. Visual object recognition generally needs two steps: (1) Find out where the object is, (2) Classify and identify what the object is. Due to the large size range of visual objects in high-resolution images, variable placement angle, and uncertain attitude, there are a great number of visual objects. Therefore, visual object detection is a complex problem. The most direct method is to construct a deep neural network, which takes the label and position of image as input, through convolutional neural networks and two full connection layers to identify the position of the object.

Figure 3.9 shows the workflow of object detection, in which the first full connectional layer outputs the classification score, exports the category of the detected object, the second full connection layer outputs the detected border position to resolve the position problem of the detected object in the image.

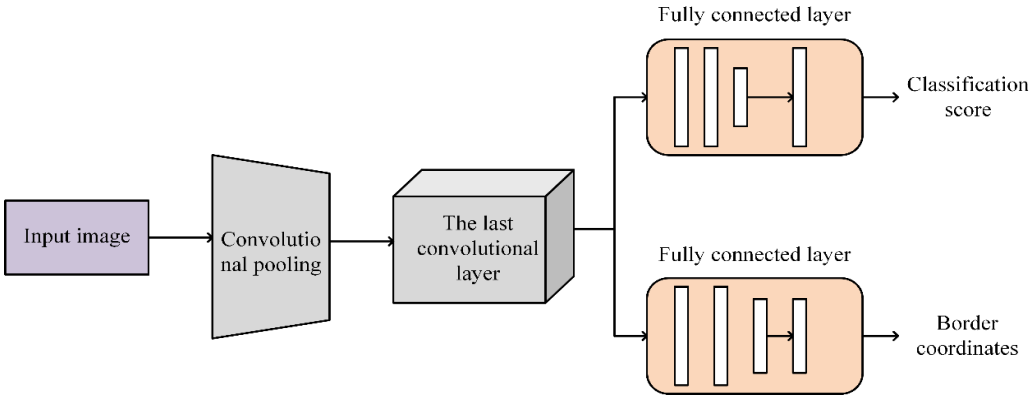


Figure 3.9: The workflow of visual object detection

In order to improve the efficiency of object detection, it promotes various deep learning models for this purpose. There are R-CNN (Rumelhart, McClelland, 1987) and its various improvements: Fast R-CNN (Girshick R, 2015), Faster R-CNN (Sun, He, & Girshick, Ren, 2015), as well as regression-based YOLO and SSD. In this chapter, we mainly introduce and analyze Faster R-CNN, a deep learning network, and make improvements on this basis.

3.2.1 Faster R-CNN Model

Faster R-CNN is an improvement of Fast R-CNN (Girshick R, 2015). In Faster R-CNN, instead of using selective search algorithm (Uijling, Sand, & Gevers, 2013) to extract candidate boxes, RPN (He, Zhang, Ren, & Sun, 2014) (region recommended network) was employed to segment candidate regions. RPN network provides the candidate areas with low quality and excellent quality for the network combined with the idea of pooling gold tower provided by using SPP-Net (Zeiler, & Fergus, 2013), not only improves the recall rate of the network, but also uplifts the recognition rate and accuracy of the network. Figure 3.10 shows the structure of Faster R-CNN net.

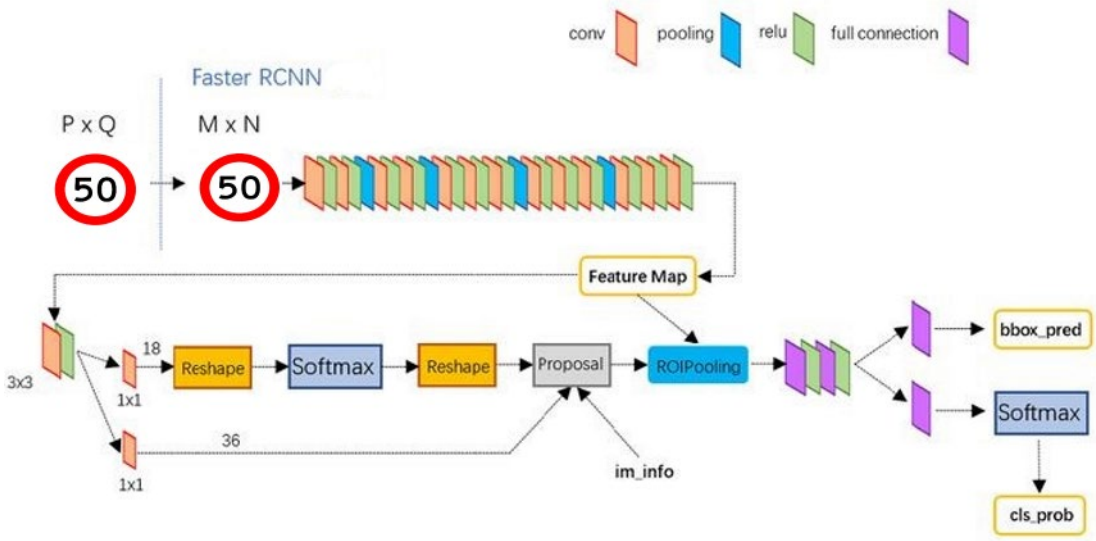


Figure 3.10: The structure of Faster R-CNN network

Figure 3.10 shows the network structure of Faster R-CNN. The specific identification process is as follows:

(1) After the image is input into CNN network, the convolutional layer is obtained. This convolutional layer has two purposes: 1) Input into RPN network as input; 2) input images to a specific fixed convolutional layer to get a more complex convolution layer dimension feature map.

(2) Through the RPN network, the scores of each regional recommendation and the corresponding regional recommendation are output, the scores of each regional recommendation are calculated, non-maximum suppression (NMS) (Hosang, Benenson 2017) is employed in the score, the threshold is 0.7, the final output is less than 300 candidate frames with good quality. The candidate frames are pooled to the same size with ROI pooling layer.

(3) The ROI in Step (2) and the high-dimensional feature map in Step (1) are input into the ROI pooling layer. Then, the corresponding features of each candidate region are obtained.

(4) Finally, the candidate features in Step (3) are input into the fully connected layer to obtain the score and border regression of each candidate region, the classification and border regression are trained jointly by using softmax loss and smooth L1 loss.

RPN is use of bounding box regressions and anchors (i.e., candidate box). It only needs to slide on the last convolution layer to get multiscale candidate regions. Figure 3.12 shows the network structure of RPN. The basic network is ZF (Simonyan, & Zisserman, 2015), the output dimension of conv5 layer is 256, corresponding to 256 characteristic graphs. Each sliding window predicts two candidate regions, the classification layer outputs two scores, namely, object probability. The layer outputs 4 coordinates, that is, the coordinates of the object boundary. Each anchor takes use of the current position as the origin and selects sliding windows of different sizes and scales. In Figure 3.12, three sizes and three length width ratios are employed, so that each corresponding sliding point has nine candidate regions, totally, nine candidate regions are generated. In this way, the candidate regions are translation invariant.

Faster R-CNN combines the original separation of boundary regression and classification, and uses the end-to-end method to detect the object, which not only

improves the accuracy of detection and recognition, but also beefs up the speed of recognition. The first Faster R-CNN model is associated with VGG (Szegedy, Liu, Jia, Semanet, Reed & Auguelov, 2014) model, and takes this model as an important feature extractor. However, the VGG network as the backbone of Faster R-CNN has larger scale and slower recognition speed. These shortcomings are not suitable for real-time TSR, which requires speed and accuracy. As a classifier, the performance of GoogLeNet (Maas, Hannun, & Ng, 2013) is not less than VGG model, its network scale is small. Therefore, we set the backbone model of Faster R-CNN to GoogLeNet.

3.2.2 Improved Faster R-CNN Network

Although the convergence speed of SGD is very fast, it is prone to neuron lifecycle during training. In the ReLU function, for the value is less than 0, the gradient of the neuron will always be 0, its weight cannot be updated, this is called “dead neuron”. In practice, if the learning rate is huge, more neurons in the network are likely to die, even if the learning rate is small, this situation is likely to happen. In order to solve the problem of neuron death, Leaky ReLU activation function is proposed.

$$Leaky\ ReLU = \begin{cases} x_i & if\ x_i \geq 0 \\ \alpha_i x_i & if\ x_i < 0 \end{cases} \tag{3.17}$$

In Equation (3.17), α_i is a small constant. x_i represents the input. If $x_i < 0$, the information of negative axis is not directly discarded like the ReLU function, but retained, thus solving the problem that the neurons of the ReLU function do not learn in the negative interval (Redmon, Divvala & Girshick, 2016). Figure 3.11 shows the Leaky ReLU function and the corresponding gradient function.

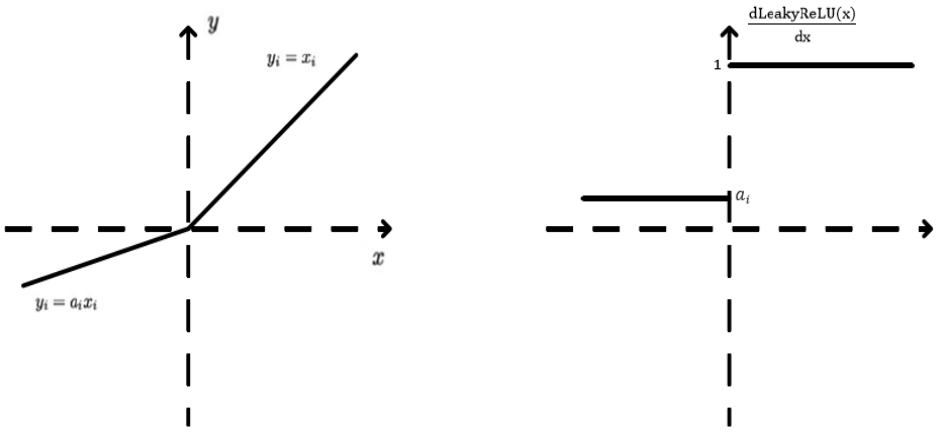


Figure 3.11: Leaky ReLU function and its gradient function

Leaky ReLU function is better than ReLU function. Therefore, in order to improve the performance of TSR methods, Leaky ReLU is employed to replace the original ReLU activation function in Faster R-CNN network. In this experiment, the parameter α_i is set to 0.02.

GoogLeNet model is a deep convolutional neural network model deepening on the depth and width of the network model based on the LeNet (Lecun, Bottou, & Haffner, 1998). We have improved the LeNet model so that the number of original CNN layers has increased to 22, the number of independent layers has also exceeded 100. The pixel sensing size of the GoogLeNet takes advantage of RGB color three channels. In order to enhance the features and reduce overfitting, we take use of ReLU after the convolution operation. Finally, we are use of softmax as the classifier. In order to reduce the thickness of the feature image, we improve the Inception structure as shown in Figure3.12.

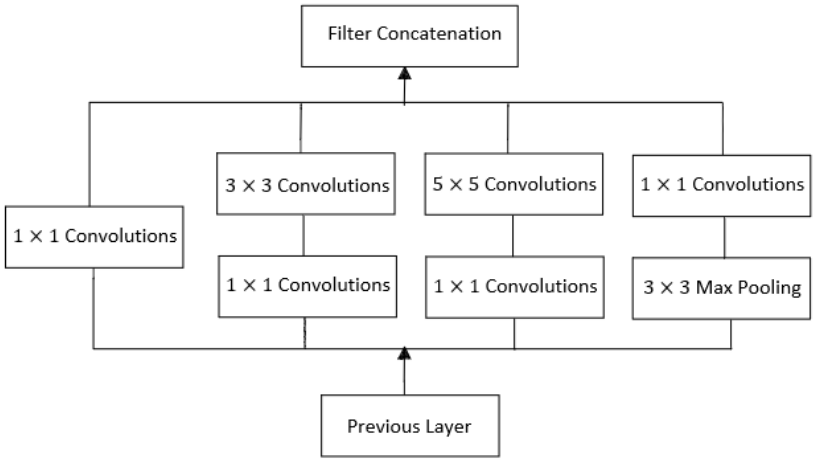


Figure 3.12: Inception model

The main principle is to replace the optimal local sparse structure of the image with a dense component. In this way, dimension reduction is realized effectively, so that the width and depth of the network are increased with the same computing resources, the parameters needed to be trained can be reduced and the over fitting problem can be reduced. Moreover, the architecture integrates with image features in various dimensions, which makes the features richer and easier for visual object recognition.

By using GoogLeNet, instead of VGG model as the feature extraction part of Faster R-CNN, the improved Faster R-CNN has smaller size and higher accuracy.

Table 3.1: The parameters of GoogLeNet

layer	**type**	**size**	**stride**
1	Conv	(7,7)	2
2	Max pooling	(3,3)	2
3	Conv	(3,3)	1
4	Max pooling	(3,3)	2
5	Inception(a)		
6	Inception(b)		
7	Max pooling	(3,3)	2
8	Inception(a)		

layer	type	size	stride
9	Inception(b)		
10	Inception(c)		
11	Inception(d)		
12	Inception(e)		
13	Max pooling	(3,3)	2
14	Inception(a)		
15	Inception(b)		

3.3 YOLOv5 Model for Traffic Signs Recognition

Under haze weather condition, the fuzzy problem of traffic sign images will lead to the decline of recognition accuracy, which has a threat to the safety requirements of autopilot. At the same time, the influence of the angle and size of the traffic sign will also lead to the decline of recognition accuracy. The rapidity of real-time detection also has high requirements for the recognition speed of the model. Considering the requirements, we improve the YOLOv5 model, which has more advantages in small object detection and takes both accuracy and speed into account, in order to better complete the traffic signs recognition and detection in haze weather. At the same time, we also improve YOLOv5 model for satellite images, another auxiliary perspective of traffic signs detection, in order to achieve better result.

3.3.1 Development of YOLO Family

You Only Look Once (YOLO) (Divvala, Redmon, Farhadi, & Girshick, 2016) algorithm is one of the excellent regression algorithms in deep learning object detection that is one of the representatives of single-stage algorithm. YOLO algorithm was firstly proposed in 2016. In the following two years, the team improved YOLO algorithm and released YOLOv2 (Farhadi & Redmon, 2017) and YOLOv3 (Farhadi & Redmon, 2018) successively. With the participation of more researchers, YOLOv4 (Chen, Lu, Liu, Li, & Qian, 2020) was developed in April 2020, its functionality is to improve significantly.

More than a month after the release of YOLOv4, YOLOv5 was released on GitHub platform, and officially was online YOLOv5.1.0 on June.

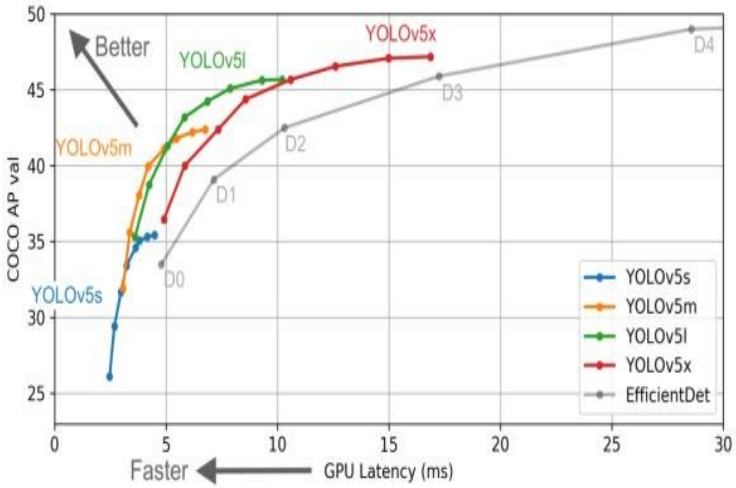


Figure 3.13: Comparison of YOLOv5 and EfficientDet on coco dataset

There are four net structures with various depth and width in YOLOv5. Figure 3.13 shows the comparison of map and reasoning speed between YOLOv5 and EfficientDet on COCO dataset. We see that YOLOv5 gives consideration to both speed and precision, its performance is excellent.

3.3.2 YOLOv5 Algorithm.

YOLOv5 was evolved from YOLOv4, the skeleton structure is very similar. YOLOv5 has three aspects: Backbone, neck, and prediction. Figure 3.14 shows the net structure of YOLOv5.

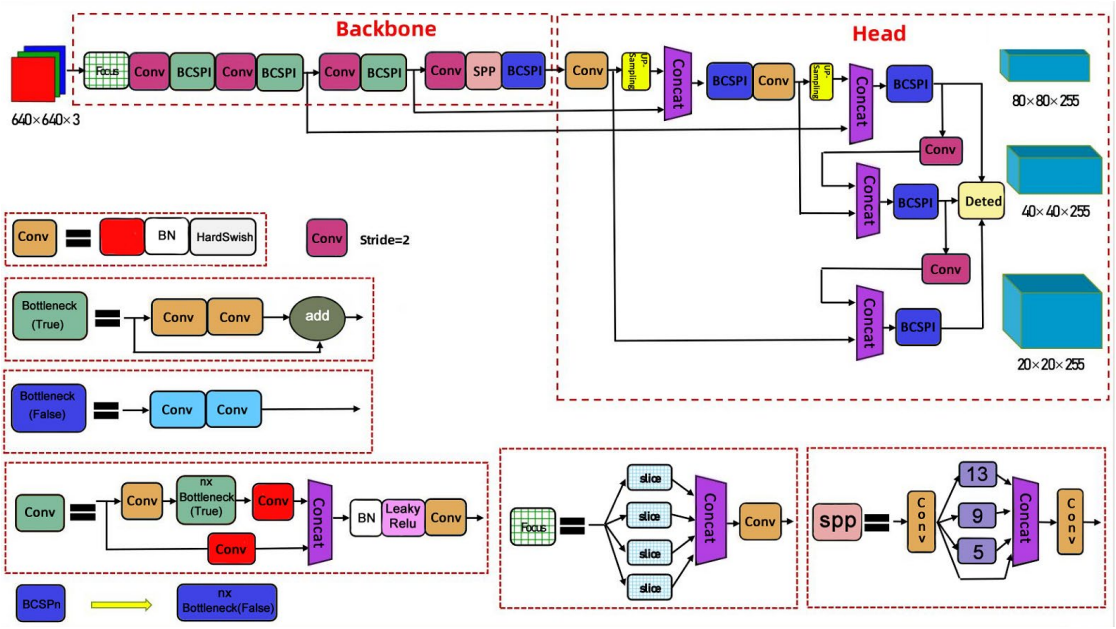


Figure 3.14: Network structure of YOLOv5

YOLOv5 is constructed by using PyTorch platform, which is divided into sub-modules, such as standard convolution conv module, focus module, spatial pyramid pool SPP module and two bottleneck modules. In order to construct a network of different sizes, only the parameters *depth_multiple* and *width_multiple* are implemented. We introduce the principle and function of each module.

The input side includes three parts: Mosaic data enhancement, image size processing, and adaptive anchor frame calculation. YOLOv5 and YOLOv4 both take use of mosaic method for data augmentation. This method is ideal for small object detection, which meets the needs of small object detection in this thesis. The size of the input image needs to be transformed into a fixed size in the YOLO algorithm, which was employed for training.

In this thesis, the standard size of images is 460 × 460 × 3. Before network training, we need to set the initial anchor frame. The initial anchor frame set of YOLOv5 is [116, 90,156,198,373,326], [30, 61, 62, 45, 59,119], [10, 13, 16, 30, 33, 23]. The network model is trained based on basic anchors to get the prediction frame, compared with the real frame. According to the difference, the model parameters are updated in reverse order and adjusted iteratively.

YOLOv3 and YOLOv4 have not a focus and their key step is sliced as shown in Figure 3.15. For example, the original image 416 × 416 × 3 is connected to the focus structure, it is converted into a 208 × 208 × 12 feature map by slicing operations, which is adjusted to a 208 × 208 × 32 feature map by using a 32 convolution kernel operation.

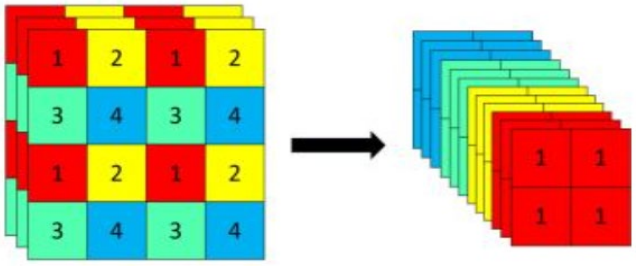


Figure 3.15: Slicing operation (Li, Zhao, Xu, Wang, Liu, & Qin, 2022)

The structure of FPN (Feature Pyramid Network) + PAN (Pixel Aggregation Network) is adopted in neck. The direction of the FPN is from top to bottom, which allows a perfect combination of high level features and low level features to obtain the ideal feature map before prediction. This ensures to pass high-level strong semantic features but only to enhance semantic information. There is no transmission of positioning information. So, PAN takes use of a bottom-up feature pyramid behind FPN to enhance semantic expression on multiple scales. The structure of FPN + PAN is shown in Figure 3.16

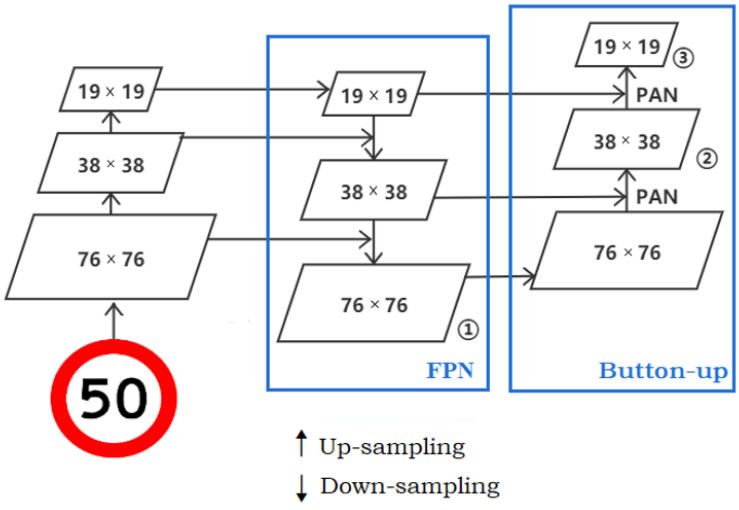


Figure 3.16: FPN + PAN structure

Prediction includes the non-maximum suppression (NMS) and loss function of bounding box regression. YOLOv5 is use of *IoU_Loss* as the loss function, the bounding box does not overlap often, so in this article, we use *CIoU_Loss* as the loss function. In the object detection and prediction stage, for the screening of numerous frames that appear, weighted NMS operation is used to obtain the optimal object frame.

3.3.3 Improved YOLOv5 Model

YOLO is constantly updated which has always been famous for its small size and fast recognition speed. Compared with other networks, it has the characteristics of small size and high stability. There is no process of obtaining a region proposal in YOLO network. Compared with CNN, YOLO is unified into a regression problem, while Faster R-CNN needs to divide the result into two parts: Classification problem and regression problem. Through comparisons, YOLOv5 outperforms than YOLOv4 in many aspects. In terms of backbone, YOLOv4 only takes use of CSP structure, and YOLOv5 is use of the *CSP1_X* structure on the backbone and the *CSP2_X* structure on the neck. At the same time, we have fine-tuned the parameters in YOLOv5. In this thesis, we set momentum to 0.950 and learning rate as 0.00120, Max epochs are set to 200, batch size is assigned as 16. During the experiment, there is a deviation between the sample and the experimental result, which affects the performance of the subsequent test. Hence, the automatic anchor frame detection in YOLOv5 is reset, we optimized the clustering algorithm to add random correction processing,

$$W_b = O_2^3 random[v1; v2] \times w_b \tag{3.18}$$

where O_2^3 means that two of every three cluster centers are randomly selected for correction, w_b is the width of the prior anchor point frame before correction, W_b is the width after correction. The numbers indicate the width and height of the anchor box respectively. It is observed that the minimum aspect ratio of the above clustering results is 0.527 and the maximum is 0.714. But for the data set in this project, 70% of the sample aspect ratio is between 0.720 and 1, 20% of the sample is between 0.600 and

0.700, 10% of the sample is between 0.600 and 0.700. The ratio is less than 0.600. From Equation (3.8), we see that the statistical results deviate from the clustering results.

In actual road conditions, traffic signs are minimal compared to buildings and cars, which may have problems such as occlusion and damage. This leads to a tiny proportion in the entire background. By using a general detector to detect traffic signs, there are a few of bounding boxes which do not include traffic signs. Because these bounding boxes have errors, the precision of TSR is not very high. Therefore, in this experiment, we have optimized the loss function. We mainly balance the foreground, the background loss is adjusted adaptively. The loss function includes classification loss and regression loss.

$$\begin{aligned} loss = & \sum_{i=0}^{S}\sum_{j=0}^{S}\sum_{k=0}^{B} E_{ijk}^{obj}\{\omega_{coord}[(x_{gt}-x_p)^2+(y_{gt}-y_p)^2+ \\ & (\sqrt{w_{gt}}-\sqrt{w_p})^2+(\sqrt{h_{gt}}-\sqrt{h_p})^2]\} \\ & +\sum_{i=0}^{S}\sum_{j=0}^{S}\sum_{k=0}^{B}\{\omega_{obj}E_{ijk}^{obj}[(C_{gt}-C_p)^2]+[\omega_{nobj}E_{ijk}^{nobj}C_p(C_{gt}-C_p)^2]\} \\ & +\sum_{i=0}^{S}\sum_{j=0}^{S}\sum_{k=0}^{B}(P_{gt}-P_p)^2 \end{aligned} \tag{3.19}$$

where S is the width and height of the feature map. There are three sizes of the feature map in this thesis: 52×52, 26×26, 13×13, B is the number of a priori boxes at each anchor point position; E_{ijk}^{obj} represents the anchor point whether the box is responsible for predicting the object, E_{ijk}^{noobj} means not responsible for predicting the object; x_{gt}, y_{gt}, w_{gt}, h_{gt} are ground trues, x_p, y_p, w_p, h_p are predicted values, indicating the coordinates of the object and Width and height (in pixels); C_{gt} and C_p represent true value confidence and prediction confidence, respectively; P_{gt} and P_p show classification true value probability and classification prediction probability, respectively; ω indicates the weight coefficient of each loss part for weights.

The parameters in this thesis are set as ω_{coord}=5.00, ω_{obj}=1.00, ω_{noobj}=0.5.00, the purpose of this setting is to reduce the loss of non-object areas and increase the loss of object areas; in order to further avoid the loss of background values to confidence. In

this thesis, C_p is also used as part of the weight to adjust the loss value of the background frame adaptively.

3.4 Traffic Signs Recognition of Satellite Images Based on the YOLOv5 Model

3.4.1 Traffic Signs Recognition in Satellite Images

With the progress of hardware, there are multiple ways to obtain traffic sign images. In terms of image acquisition methods, there are mainly two-folds: One is the road condition and traffic information taken by using optical camera on the ground, the other is the High Definition (HD) satellite imagery obtained by using satellite transmitting electromagnetic waves to the ground in space, the road sign image on the ground are also obtained from these images. Then, the deep learning algorithms are employed to extract features from the acquired images to get road object detection.

With the development of satellite technology, traffic object detection is more and more based on satellite remote sensing images. In the early research, a large number of researchers realized object recognition of satellite remote sensing images based on traditional methods. Huang et al (Huang, & Zhang, 2009) realizes road extraction from remote sensing images according to geometric, radiation and topological features of roads, and classifies them by using SVM (Saunders, & Stitson, 2002) method. The processing method of decision tree classifier is recursive segmentation of the input image. Its branches represent different segmentation paths and leaves represent the final classification results.

Therefore, the whole tree is the process of segmentation for vehicle detection based on high resolution satellite images (Eikvil, Aurdal, & Koren, 2009). Firstly, a rule-based method is applied to divide the image into normal region and shadow region. Then, the visual objects are classified by using a statistics-based method and the detection results are compared with the results of manual identification.

The experimental results show that though the image resolution is low and it is

difficult to classify objects manually, excellent algorithm recognition results and close to the results of artificial classification. A Haar-like feature-based AdaBoost algorithm (Leitloff, Hinz, & Stilla, 2010) was proposed to identify vehicles, combined with line detection technology to detect individual vehicles in the fleet.

Compared with the method based on statistics alone, the accuracy of this method has been improved to 80.0%. Although the traditional method has achieved good results in object recognition based on satellite remote sensing images, it needs to design features manually, the process is complex and lacks good robustness for the diversity of objects.

3.4.2 Satellite Image Object Recognition Network Based on YOLOV5

We choose YOLOv5 as our algorithm of road object recognition from satellite images for the following reasons: YOLOv5 incorporated Cross Stage Partial Network (CSPNet) (Wang, Liao, Wu, Chen, Wei, & Yeh, 2020) and created CSPDarkNet and added it to Darknet's backbone. CSPNet solves the gradient descent and information duplication in the backbone. And these changes are added to the feature map. The number of operations per second and model parameters are reduced. This not only reduces the size of the model, but also speeds up the calculation speed and accuracy. In the task of acquiring road object images for satellite radar sensors, object detection speed and accuracy are essential, the size of this compact model is also conducive to its reasoning efficiency on resource-poor edge equipment.

Secondly, the YOLOv5 applied Path Aggregation Network (PANet) (Wang, Liew, Zou, Zhou, & Feng, 2019) as its neck to boost information flow. PANet takes use of a new Feature Pyramid Network (FPN) structure. This kind of networks provide a bottom-up path so that low-level features can be effectively propagated. At the same time, the adaptive feature pool links various levels of features to form a feature network, which directly transfers different levels of feature information to the corresponding in the sub-network. PANet continuously improves the utilization of the positioning signal of the lower layer, which obviously enhances the location accuracy of the object.

Finally, the head of YOLOv5, namely the YOLO layer, because three different feature maps are created here, visual objects of different sizes are identified, predicted and processed. Traffic signs come in various types and sizes, and weather changes in real time affect radar sensors' ability to capture traffic images. Multiscale (Redmon, Farhadi, 2018) detection ensures that the model follows the scale changes in the process of vehicle travel and weather changes. The training objectives of our satellite-image-based TSR network take use of our improved YOLOv5 loss as shown in Equation (3.19).

3.5 Evaluation Methods

In this thesis, we make use of a variety of evaluation methods to evaluate our model. Firstly, the most commonly-used metrics in deep learning are precision and recall. There are four experimental results: TN (True Negative), FN (False Negative), FP (False Positive), TP (True Positive), TP represents the number of times that the prediction is greater than or equal to R and the actual value is greater than or equal to R, FN represents the number of times the actual value is greater than or equal to R, and the predicted value is less than R, FP represents the number of times the actual value is less than or equal to R, and the predicted value is greater than or equal to R, TN represents is the number of times the predicted value is less than R and the actual value is less than R, where R is the threshold standard for regression prediction. The precision we get is the prediction result, which indicates the probability that the positive sample is correct. The recall is applied to measure how many positive samples are accurately predicted. Therefore, the equation for recall and accuracy is as follows:

$$precision = \frac{TP}{TP + FP} \tag{3.20}$$

$$recall = \frac{TP}{TP + FN} \tag{3.21}$$

Average Precisions (AP) is calculated by Recall and Precision (Henderson, & Ferrari, 2017). Because the maximum value of Recall and Precision is one and the minimum value is zero, the range of AP is also between zero and one. Mean Average Precision is often harnessed with AP because their calculation methods are very similar.

AP is applied to evaluate the accuracy in each model, mAP is employed to evaluate the quality of the model in all categories, after the calculation of AP, mAP becomes very simple,

$$AP = \int_0^1 p(r)dr. \tag{3.22}$$

The mAP is related to AP, mAP is the average value of AP,

$$mAP = \frac{1}{c} \sum_{c \epsilon classes} \frac{TP}{TP+FP}. \tag{3.23}$$

Finally, we are use of Complete Intersection over Union (CIoU) to evaluate our model, CIoU is evolved from Generalized Intersection over Union (GIoU), they have broadly used indicators for evaluating the performance of detectors like the methods mentioned above (Rezatofighi, Tsoi, Gwak, Reid, & Savarese, 2019). CIoU solves the shortcoming that IoU cannot accurately reflect the overlap between the true value and the predicted value. Equations (3.24~3.25) are employed for the calculation method of CIoU and the calculation method of CIoU loss function.

$$CIoU = \frac{\rho^2(A,B)}{c^2} + \alpha v \tag{3.24}$$

and

$$L_{CIoU} = 1 - IoU + \frac{\rho^2(A,B)}{c^2} + \alpha v. \tag{3.25}$$

Chapter 4
Our Results

In this chapter, we introduce how we collect our own dataset. Then, we showcase the performance of Faster R-CNN and YOLOv5 in this experiment. Afterward we indicate the defogging effect of guided image filtering in this experiment. Finally, the experimental results of YOLOv5 and Faster R-CNN are compared.

4.1 Data Sources and Data Collection

Our dataset consists of 3,105 images and 5,536 instances in total. First of all, in the experiment of recognizing traffic signs based on the driver's perspective, we take into account of a dataset created by ourselves, including 12 traffic signs, which is shown in Table 4.1. We take use of the camera of GoPro with the resolution 2704×1520 pixels to record traffic sign videos in New Zealand at 60 frames per second. Then, we extract images with the interval of ten frames of the video. The resolution of the image is 1920×1080 which is stored in .JPG format.

Table 4.1: Types of traffic signs in the dataset (driver's perspective)

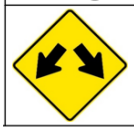
Bicycle Lane Ahead	Keep Left	Roundabout Give Way	Speed Limit 50	Give Way	No Parking Bus Stop
					
Road Diverges	**Roundabout Ahead**	**Children**	**Roundabout Chevron**	**Road Works**	**Pay Attention**
					

Figure 4.1: The examples of our dataset (driver's perspective)

In our experiment of identifying traffic signs based on satellite images, we take use of datasets that we've created by ourselves. There are 1,029 images captured from Google Earth which were annotated, each image is manually labelled with a traffic sign. This dataset mainly includes traffic signs like straight, right, left, give way, stop,

crosswalk, keep clear, etc., which is shown in Table 4.2. Among them, each sample of identifiers is not uniform. Among them, 60% signs are used for training, 20% are applied to verification, and 20% for testing.

Table 4.2: Types of traffic signs in the dataset (satellite image's perspective)

Go Straight	Turn Left	Turn Right	Go Straight/Left	Go Straight/Right	Slow Down
					
Give Way	**Crosswalk**	**Keep Clear**	**Stop**	**Bicycle Lane**	**Bus Lane**
					

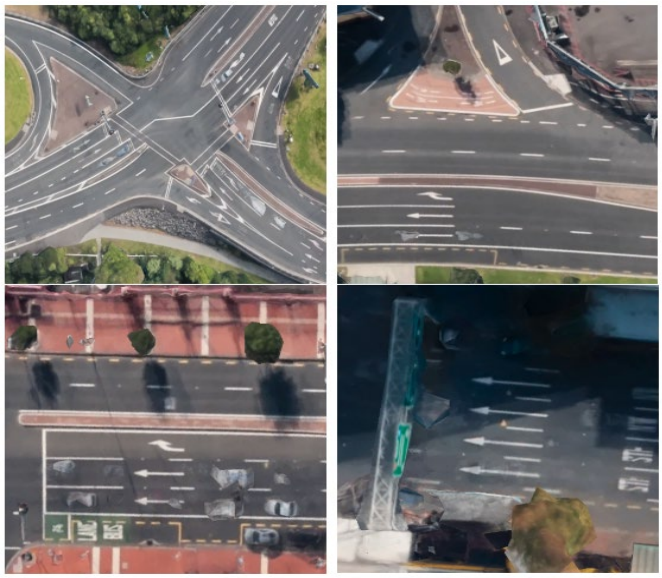


Figure 4.2: The examples of our dataset from satellite view

Due to the limited samples in our dataset, in order to make our experimental results more accurate, we also added German Traffic Sign Detection Benchmark (GTSDB) (Houben, Stallkamp, Salmen, Schlipsing, & Igel, 2013) to our data set. GTSDB is the dataset that suits for TSR very much since its release because it contains more than 1,200 traffic signs in four categories. GTSDB plus our dataset has more than 5,000 images totally, but there are no foggy ones. Thus, we have added three datasets: FROSI, FRIDA, and FRIDA2. The photographs are all synthesized by using computers, the visibility and area of the fog are under control. These three datasets all mimic the

perspective of real drivers, different types of fogs are randomly added, including cloudy hazes, cloudy heterogeneous hazes, heterogeneous fogs, and uniform fogs. The fogs in the FROSI dataset are grouped according to visibility from 50 meters to 400 meters, the traffic signs in the dataset are similar to real ones. There are 1,620 traffic signs placed at multiple heights. These three datasets assist us to improve the accuracy of the two deep learning methods on foggy days. In this thesis, in the experiment of traffic signs recognition from the driver's viewpoint, we combine the five datasets in total for our experiments.

Figure 4.3: The examples from FRIDA & FROSI dataset

4.2 Comparison and Analysis of Two Defogging Models

In this section, we show the defogging results of the improved HSV algorithm and the guided image filtering method in different scenarios. Figure 4.4 shows the output of each defogging algorithm.

From the comparisons, we see that the defogging algorithm based on the guided image filtering is much robust, the defogging result is much stable in different scenes. In the process of getting a better defogging result, it does not bring obvious color distortions or image darkening. On the contrary, it plays a central role in enhancing the colors. Therefore, in this thesis, we decide to take the guided image filtering as the final defogging algorithm.

Figure 4.4: The results of defogging methods in various scenes

Figure 4.5: The results of different defogging methods in the FROSI dataset

4.3 Experimental Analysis of Traffic Signs Recognition Based on Faster R-CNN

In this experiment, in order to compare the two methods of Faster R-CNN and YOLOv5, we take use of controlling variables, the same dataset and equipment were employed for training and evaluating the two methods. The GPU is NVIDIA RTX2070 associated with the CPU CORE i7-9th. In our experiments, we improved the parameters of Faster

R-CNN, set the momentum to 0.900, learning rate 0.0100, max epochs 200, batch size 24, weight decay 0.000300. At the same time, we are use of the fully connected layer and the ReLU activation function to extract the features of visual object.

In our experiments, we find that there are many types of backbones in Faster R-CNN, most of backbones have different characteristics for the model. Before our experiments, we decided to compare the backbone network of Faster R-CNN. We pick out the excellent performance which suits for our backbone network. Table 4.3 shows our experimental outcomes, which provides the recall and precision of three different nets. From Table 4.3, we see that the networks have various experimental outcomes. Among them, ResNet has the highest recall, GoogLeNet is the best among three networks in precision, mAP, and recognition speed. Thus, we finally choose GoogLeNet as the backbone network of Faster R-CNN.

Amid training models with our dataset, we find that there are visual objects on the road that resemble traffic signs, such as billboards, traffic lights, etc. These objects will affect the accuracy of our model. Therefore, in order to improve the accuracy of our model, we added a sample mining method to our training. Firstly, we classify the samples with a classifier, group those that are less than or equal to 0.7 as positive samples, and those that are greater than 0.7 as negative samples. The negative samples are stored as misclassified samples. Then, we find out the reasons for the misclassification and deal with them in a targeted way. Finally, the negative samples are trained, but this will cause the problem of data imbalance. Aftermaths, we add the hard positives to the positive samples so as to solve the problem of data imbalance. This not only increases the amount of data we have but also improves our recognition accuracy and gives us good experimental results.

Table 4.3: The traffic sign recognition results of different nets

Networks	Recall (%)	Precision (%)	mAP (%)	fps
VGGNet	88.2	89.1	90.2	16
GoogLeNet	88.7	93.2	95.3	17
ResNet	92.8	91.2	95.2	16

In Table 4.3, we compare the three backbone networks for the same dataset, where ResNet has the highest recall, but GoogLeNet has the highest precision and mAP as well as the fastest runtime.

In this experiment, we group the samples into sunny and foggy classes, and derive the differences between recall, precision and mAP with and without guided image filtering as shown in Table 4.4 and Table 4.5, respectively. Finally, the PR plots derived from PR curve are shown in Figure 4.6.

Table 4.4: Our experimental results of Faster R-CNN in various weather conditions

Weather	**Precision**	**Recall**	**mAP@0.5**
Sunny	0.971	0.975	0.974
Foggy	0.899	0.903	0.900

Table 4.5: Our experimental results of Faster R-CNN with guided image filtering

Weather	**Precision**	**Recall**	**mAP@0.5**
Sunny	0.964	0.958	0.962
Foggy	0.929	0.930	0.932

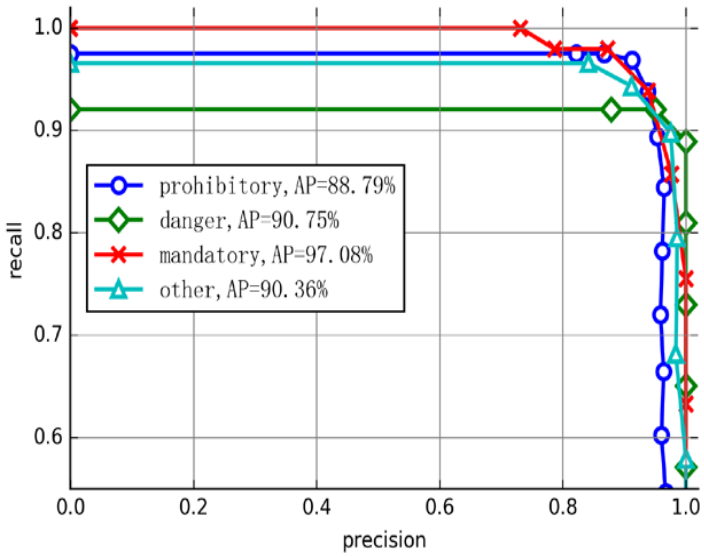


Figure 4.6: PR curve of our experimental results

In Table 4.5 and Table 4.6, we compare the accuracy and recall before and after using guided image filtering on sunny and foggy images. The Faster R-CNN has a higher precision and recall. In this thesis, we have chosen GoogLeNet as the backbone of the Faster R-CNN by comparison. We were use of a guided image filtering method to defog the foggy day images. By using guided image filtering, recognition accuracy on sunny images is much higher than that on foggy images. Throughout using guided image filtering to defog the image, though the accuracy on sunny days is 0.7% lower, the accuracy on foggy days is 3% higher, because the guided image filtering method not only removes the fog from the foggy images, but also adds a small amount of noises to the sunny images.

In addition, it is worth noting that the recognition speed of Faster R-CNN based on GoogLeNet is also faster, which is one frame per second higher than the original. To be more suitable for the requirements of fast and accurate detection, Figure 4.6 shows the four different classifications. The experimental results indicate that our method is able to identify traffic signs well in challenging scenarios.

Figure 4.7: TSR results regarding the improved Faster R-CNN model in foggy scenes

In Figure 4.7, we clearly see that foggy images have different scenes, the results of the improved Faster R-CNN for traffic sign detection fluctuate, the traffic signs to be detected are simple, the location and content in “exit” sign are detected. In complex scenes, the scene having multiple buildings or similar traffic signs, the recognition result of the improved Faster R-CNN is low, there are many cases of unrecognizable or incorrect recognition. In addition, in the case with foggy and far scenes, the accuracy of traffic sign detection is significantly lower than that in the near scene.

4.4 Experimental Analysis of Traffic Signs Recognition Based on YOLOv5

The dataset we took in this chapter is as same as the dataset in the previous one, the division method is also the same. Among them, 60% of the data is employed as the training set, 20% of the dataset is treated as the validation set, and 20% of the data is selected as the test set. The evaluation methods for the experiments are still recall and accuracy.

Table 4.5 shows the recognition experiment results of YOLOv5 if the guided image filtering is not adopted. From Table 4.5, we see that YOLOv5 has a high recognition precision and recall on image shots in sunny days, but the accuracy drops on an image from foggy days. At the same time, the recall decreases 6.5%, which is quite inadequate with the precision and recall on the images from sunny days. In order to improve the recognition accuracy, Table 4.6 shows the experimental results of YOLOv5 after using guided image filtering. Firstly, we see that the recognition precision has dropped 0.3%. The reason is that we have removed the foggy images and added noises to the traffic signs, but the recognition accuracy in foggy images is 3.9% higher than before, which effectively improves the recognition accuracy and recognition efficiency of our model on foggy images. The loss curve is shown in Figure 4.9.

Table 4.6: Our experimental results of YOLOv5 in various weather conditions

Weather	Precision	Recall	mAP@0.5
Sunny	0.955	0.957	0.960
Foggy	0.883	0.892	0.890

Table 4.7: Our experimental results of YOLOv5 used guided image filtering

Weather	Precision	Recall	mAP@0.5
Sunny	0.952	0.956	0.959
Foggy	0.922	0.927	0.929

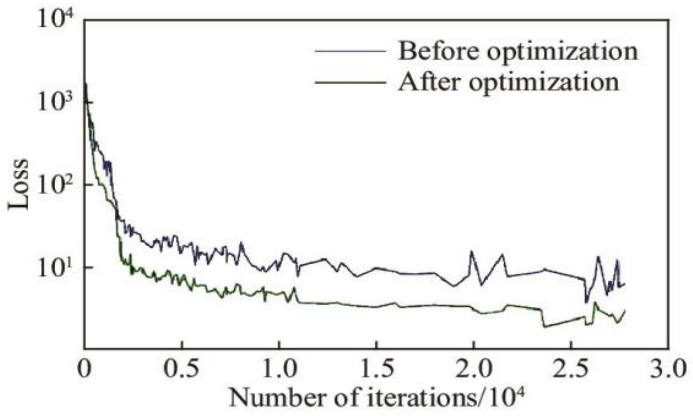


Figure 4.8: The loss curve

By improving the optimization ability of the loss function in the adaptive balance of foreground and background loss, YOLOv5 model, which is good at small object detection, is better at multiobject detection in complex scenes such as traffic sign images. At the same time, by observing the loss curves of the two models in Figure 4.8, we find that the loss function of improved YOLOv5 model converges better which is smoother on the whole. That is, the improved loss function makes the model better on the images including small objects after defogging and a variety of traffic signs, which accurately identifies them from the complex road conditions.

In order to further expand TSR, we have applied the improved YOLOv5 method to identify traffic signs from the driver's perspective while adding a satellite image view, which is employed to complement the driver's perspective recognition results to

improve the final recognition accuracy. The parameters of YOLOv5 have been adjusted. Figure 4.9 shows the recognition results of YOLOv5 with satellite images. Figure 4.10 is our detection result from sunny images. These four images were randomly selected from the video footages we shot. The four images in Figure 4.11 were acquired from a foggy video.

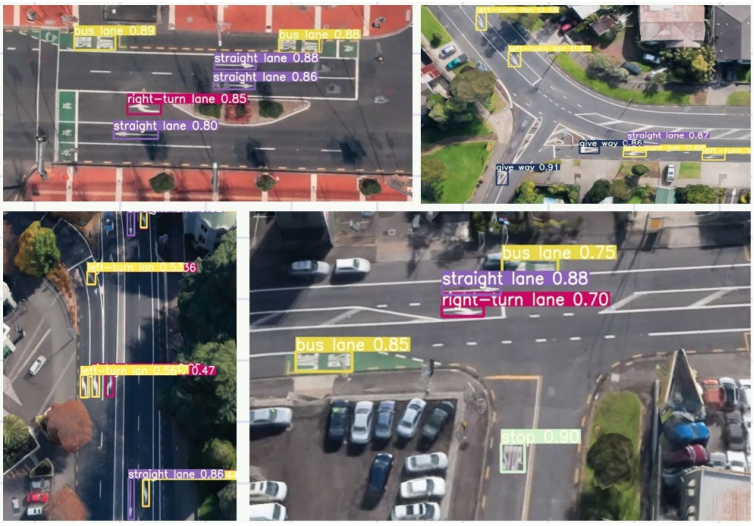


Figure 4.9: The TSR results from satellite images

Figure 4.10: The TSR result from driver's view

Figure 4.11: The TSR results from foggy images

4.5 Comparison Between Improved YOLOv5 and Improved Faster R-CNN Model

In this section, we compare the experimental results of improved YOLOv5 and improved Faster R-CNN. The datasets, experimental environment, and hyperparameters of these two methods are as same as those previous experiments.

The video we tested is composed of 2,590 frames after processing. YOLOv5 takes 0.009 seconds to process each frame, Faster R-CNN spends 21 seconds to process each frame. It costs much longer than YOLOv5. Under the same accuracy rate, YOLOv5 has a faster recognition speed. Because TSR is often employed for real-time detection and has high requirements for recognition speed, YOLOv5 is much suitable for TSR. Figures 4.12 and 4.13 show the recognition results of the two methods in the FRIDA dataset.

Figure 4.12: The TSR results based on FRIDA dataset with Faster R-CNN

Figure 4.13: The TSR result based on FRIDA dataset with YOLOv5

By observing the images shown in Figure 4.12 and Figure 4.13, we find that these two improved methods achieve accurate location detection for traffic signs in the background of complex foggy conditions. However, from one of the recognition results, we see that Faster R-CNN has two traffic signs that have not been recognized, but the method of YOLOv5 is perfect to be recognized. At the same time, thanks to the lighter model size and recognition speed. On the whole, the improved YOLOv5 performs better.

Chapter 5
Analysis and Discussions

In this chapter, we analyze and discuss the experimental results, mainly from the identification of precision and recall for comparisons, and also detail the defogging ability of guided image filtering.

5.1 Experimental Analysis

In this experiment, we find that Faster R-CNN obtains a higher accuracy-related score while receiving a lower loss score. For example, the value calculated by mAP is basically as same as that calculated by precision and recall, but the recognition speed is slow. YOLOv5 has a fast recognition speed while both precision and recall are high.

Since YOLOv5 has better performance, let's take YOLOv5 as an example. Figure 5.1 shows the changes of each indicator as to the number of iterations increases, where bounding box value decreases as the iteration increases at this point, mAP increases as the iteration increases at this point. It shows that the detection result of the network in this thesis is getting batter with the increase of iteration times. The precision and recall are also increased with the network parameter iteration. This indicates that the number of correct positive samples detected also increases with the number of iterations. In general, with the increase of the number of iterations, YOLOv5 method in this thesis is getting better and more stable for TSR.

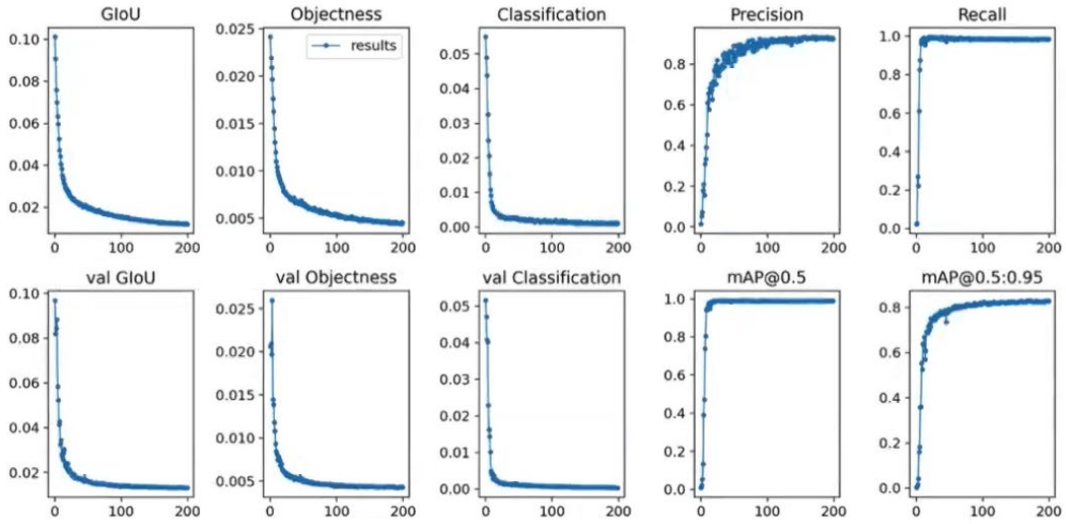


Figure 5.1: TSR results by using YOLOv5

Figure 5.2 is the PR curve of the experimental results, *y*-axis is the accuracy rate, and *x*-axis is the recall rate. We see that the PR curve is very close to the top right corner, indicating that our model is performing effectively. Therefore, TSR based on YOLOv5

has good performance which was well developed.

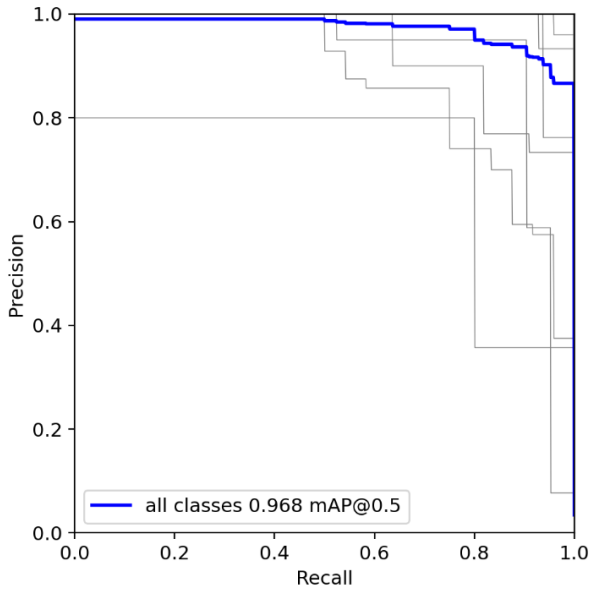


Figure 5.2: The PR curve of YOLOv5

(a) (b)

Figure 5.3: (a) Image guided filtering is not applied, (b) By applying image guided filtering

Figure 5.3 shows the experimental results by using YOLOv5 to recognize objects with and without guided image filtering. In Figure 5.3, we see that in the case of dense fog, our model only detects the traffic signs in the images without fog removal, there are some cases where the traffic signs in the distance cannot be detected. After defogging, our model is able to detect traffic signs with a long distance on the basis of the original, which improves the situation of missed detection and false detection. This

also shows the necessity of image defogging algorithm, which makes our YOLOv5 model robust and achieves a better result.

Chapter 6
Conclusion and Future Work

In this chapter, due to different results obtained by using two deep learning methods, we put forward our views on these two methods and found out a much suitable method for real TSR through our experiments. In this thesis, we find out the shortcomings of our proposed methods and envision our future work.

6.1 Conclusion

As an important part of intelligent driving, the detection and recognition of traffic signs visually have attracted more and more attention. At present, under low-light conditions, foggy weather, rainy weather, and other situations where digital cameras are blocked or blurred, the collected videos or images will become blurred, the recognition errors are much likely to be generated. Therefore, in these challenging environments, the TSR recognition accuracy is very low. For example, the accurate recognition of traffic signs in a haze environment is affected by the visibility of clear images, its recognition accuracy still needs to be improved.

We provide Faster R-CNN to complete the recognition and detection of traffic signs. In the experiment, we choose the most suitable network for TSR by backing off different backbone networks. Then we use cross-layer links and activation functions to construct feature maps more effectively and then perform feature extraction.

We offer the method of YOLOv5 to detect traffic signs. We improve the loss function, which improves the performance of YOLOv5. Throughout our experimental comparisons, we see that YOLOv5 is much important, though compared with Faster R-CNN, the accuracy rate drops slightly, but in the case of similar accuracy rates, a faster recognition speed is usually required in real TSR. YOLOv5 is a better choice for TSR. Faster R-CNN is much suitable for static object recognition.

After comparing various defogging methods, we found that the guided image filtering method has better defogging outcomes. Firstly, we conduct guided filtering on the image in the national policy of image preprocessing so that noises and fog in the image can be removed. Then, the processed image is trained by using a deep neural network. Guided image filtering is able to advance the precision of our model for object recognition in foggy weather.

6.2 Future Work

There are three aspects to our future work. Firstly, we need to continuously improve

our data set. Now, we only have a sunny dataset. In the next project, we will collect datasets in various lighting conditions such as foggy and raining days. Secondly, we compare more object recognition and detection methods in TSR. Finally, we will take advantage of more evaluation methods to evaluate our model which is able to intuitively discover the shortcomings of our model and make our model more robust and stronger.

References

Ai, C., Y. Tsai. (2014). Critical assessment of an enhanced traffic sign detection method using mobile LiDAR and INS technologies. *Journal of Transportation Engineering,* (5).

Ajmi, C., Pérez, J., Ferchichi, S., & Zaafouri, A. (2020). Deep learning technology for weld defects classification based on transfer learning and activation features. *Advances in Materials Science and Engineering, 2020(1): p. 1-16.*

Audebert, N., Saux, B.L., Lefèvre, S. (2016). Semantic segmentation of earth observation data using multimodal and multi-scale deep networks. *Asian Conference on Computer Vision.*

Berkaya, S. K., Gunduz, H., Ozsen, O., Akinlar, C., & Gunal, S. (2016). On circular traffic sign detection and recognition. *Expert Systems with Applications, 48,* 67-75.

Bertozzi, M., and A. Broggi., (1998). GOLD: A parallel real-time stereo vision system for generic obstacle and lane detection. *IEEE Transactions on Image Processing, 7*(1), 62-81.

Carrasco, J.P., la Escalera, A. d. l., de, and Armingol, J.M. (2012). Recognition stage for a speed supervisor based on road sign detection. *Sensors, 12*(9), 12153-12168.

Chambers, J., Yan, W., Garhwal, A., Kankanhalli, M. (2014) Currency security and forensics: A survey. *Multimedia Tools and Applications*, 74(11), 4013-4043.

Chen, W., Lu, S., Liu, B., Li, G., & Qian, T. (2020). Detecting citrus in orchard environment by using improved YOLOv4. *Scientific Programming, 2020*(1), 1-13.

Chen, X., Chang, C., Yu, C., & Chen, Y. (2020). A real-time vehicle detection system under various bad weather conditions based on a deep learning model without retraining. *Multidisciplinary Digital Publishing Institute (MDPI).*

Chen, Z., Ou, B., Tian, Q. (2019). An improved dark channel prior image defogging algorithm based on wavelength compensation. *Earth Science Informatics 12*(4):501-512.

Cui, W., Yan, W. (2016) A scheme for face recognition in complex environments. *International Journal of Digital Crime and Forensics* (IJDCF) 8 (1), 26-36.

De, L. E., Salichs, M.A., Arimingol, M. J., Moreno, L. E. (1997). Road traffic sign detection and classification. *IEEE Transaction on Industrial Electronics, 44*(6), 848-859.

Deshmukh, A., Singh, S., & Lakha, B. (2016). Design and development of image defogging system. *International Conference on Signal and Information Processing (IConSIP)*.

Dong, N. (2018). Research and application of image enhancement technology in traffic scene of fog and haze. *Journal of Interdisciplinary Mathematics, 21*(5), 1297-1301.

Eikvil, L., Aurdal, L., & Koren, H. (2009). Classification-based vehicle detection in high-resolution satellite images. *ISPRS Journal of Photogrammetry and Remote Sensing, 64*(1), 65-72.

Escalera, A., Armingol, J. M. A., Mata, M. (2003), Traffic sign recognition and analysis for intelligent vehicles. *Image and Vision Computing, 21*(3), 247-258.

Escalera, A., Moreno, L.E., Salichs, M.A., Armingol. (1997). Road traffic sign detection and classification. *IEEE Transactions on Industrial Electronics, 44*(6), 848-856.

Fattal, R.（2008）. Single image dehazing. *ACM Transactions on Graphics, 27*(3), pp. 1-9.

Feng, X., Li, J., & Hua, Z. (2020). Low-light image enhancement algorithm based on an atmospheric physical model. *Multimedia Tools Applications, 79, 32973-32997.*

Fu, M., Huang, Y. (2010). A survey of traffic sign recognition. *International Conference on Wavelet Analysis and Pattern Recognition.*

Garg, K., & Nayar, S.K. (2004). Detection and removal of rain from videos. IEEE

CVPR.

Girshick, R. (2015). Fast R-CNN. *IEEE International Conference on Computer Vision* pp. 1440-1448.

Gong, C., Zhou, P., Han, J. (2016). learning rotation-invariant convolutional neural networks for object detection in VHR optical remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing, 54*(12), 7405-7415.

Gowdra, N., Sinha, R., MacDonell, S., Yan, W. (2021) Maximum Categorical Cross Entropy (MCCE): A noise-robust alternative loss function to mitigate racial bias in Convolutional Neural Networks (CNNs) by reducing overfitting. *Pattern Recognition.*

Gowdra, N. (2021) *Entropy-Based Optimization Strategies for Convolutional Neural Networks.* PhD Thesis, Auckland University of Technology, New Zealand.

Greenhalgh, J., Mirmehdi, M. (2012). Traffic sign recognition using MSER and Random Forests. *Signal Processing Conference (EUSIPCO), 1935-1939*.

Gu, Q., Yang, J., Kong, L., Yan, W., Klette, R. (2017) Embedded and real-time vehicle detection system for challenging on-road scenes. *Optical Engineering*, 56 (6), 063102.

Gu, Q., Yang, J., Yan, W., Klette, R. (2017) Integrated multi-scale event verification in an augmented foreground motion space. Pacific-Rim Symposium on Image and

Video Technology, 488-500.

Gu, Q., Yang, J., Yan, W., Li, Y., Klette, R. (2017) Local Fast R-CNN flow for object-centric event recognition in complex traffic scenes. Pacific-Rim Symposium on Image and Video Technology, 439-452.

Guo, F., Cai, Z., & Xie, B. (2011). New Algorithm of Automatic Haze Removal for Single Image. *Central South University, China*.

Handmann, U., Kalinke, T., Tzomakas, C., Werner, M., Seelen, W.v. (2000). An image processing system for driver assistance. *Image and Vision Computing, 18*(5), 367-376.

Hasan, F. (2008). Traffic and Road Sign Recognition. *Dalarna University*.

Hasirlioglu, S., & Riener, A. (2018). Challenges in object detection under rainy weather conditions. *Intelligent Transport Systems, pp 53-65*.

He, K., Sun, J., Tang, X. (2011). Single image haze removal using dark channel prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 33*(12), 2341-2353.

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Transactions on Pattern Analysis & Machine Intelligence, 37*(9), 1904-16.

He, Y., & Xu, L. (2010). Automatic recognition algorithm of traffic signs in natural

scenes. *Microcomputer Information*, (04), P. 161-163.

Henderson, P., & Ferrari, V. (2017). End-to-end training of object class detectors for mean average precision. *Asian Conference on Computer Vision.*

Hide, R. (1977). Optics of the atmosphere: Scattering by molecules and particles. *Physics Bulletin*.

Hines, G., Rahman, Z., & Jobson, D. (2005). Single-scale retinex using digital signal processors. *IEEE VLSI,* 112~118.

Hnewa, M., & Radha, H. (2021). Object detection under rainy conditions for autonomous vehicles: A review of state-of-the-art and emerging techniques. *IEEE Signal Process. Mag. Signal Processing Magazine, IEEE. 38(1):53-67*.

Houben, S., Stallkamp, J., Salmen, J., Schlipsing, M., & Igel, M. Detection of traffic signs in real-world images: The German traffic sign detection benchmark. *International Joint Conference on Neural Networks (IJCNN), pp. 1-8.*

Hosang, J., Benenson, R., & Schiele, B. (2017). Learning non-maximum suppression. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Huang, D., Huang, W., Gu, P., Liu, P., & Luo, Y. (2017). Image super-resolution reconstruction based on regularization technique and guided filter. *Infrared Physics & Technology.*

Huang, X., & Zhang, L. (2009). Road centreline extraction from high-resolution

imagery based on multiscale structural features and support vector machines. *International Journal of Remote Sensing, 30*(8), 1977-1987.

Hubel, D. H., & Weisel, T.N. (1968). Receptive fields and functional architecture of monkey striate cortex. *The Journal of Physiology, 195(1):215-43.*

Illingworth, J., & Kittler, J. (1988). A survey of the Hough transform. *Computer Vision Graphics & Image Processing, 43*(2), 280-280.

Jiao, Y., Weir, J., Yan, W. (2011) Flame detection in surveillance. Journal of Multimedia 6 (1).

Ju, M., Ding, C., Ren, W., Yang, Y., Zhang, D., Guo, Y. (2021). Research of image enhancement algorithm based on histogram statistics. *Science Technology and Engineering.*

Koschmieder, H., & Rühle, H. Theory of horizontal visual range and danzig visibility measurements. *University of Texas, USA.*

Krizhevsky, A., Sutskever, I., Hinton, G.E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM, 60*(6), 84-90.

Kurak, J., & Charles, W. (1991). Adaptive histogram equalization: A parallel implementation. *IEEE Symposium on Computer-Based Medical Systems.*

Le, R., Nguyen, M., Nguyen, Q., Nguyen, H., Yan, W. (2020) Automatic data generation for deep learning model training of image classification used for augmented

reality on pre-school books. *International Conference on Multimedia Analysis and Pattern Recognition.*

Le, R., Nguyen, M., Yan, W. (2020) Machine learning with synthetic data – a new way to learn and classify the pictorial augmented reality markers in real-time. International Conference on Image and Vision Computing New Zealand.

Le, R., Nguyen, M., Yan, W. (2021) Training a convolutional neural network for transportation sign detection using synthetic dataset. *International Conference on Image and Vision Computing New Zealand.*

Le, R., Nguyen, M., Yan, W., Nguyen, H. (2021) Augmented reality and machine learning incorporation using YOLOv3 and ARKit. *Applied Sciences.*

Le, R. (2022) *Synthetic Data Annotation for Enhancing the Experiences of Augmented Reality Application Based on Machine Learning (PhD Thesis).* Auckland University of Technology, New Zealand.

Lecun, Y., Bottou, L., & Haffner P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE, 1998, 86*(11):2278-2324.

Leitloff, J., Hinz, S., Stilla, U. (2010). Vehicle detection in very high-resolution satellite images of city areas. *IEEE Transactions on Geoscience and Remote Sensing, 48*(7), 2795-2806.

Li, B., Wang, S., Zheng, J., & Zheng, L. (2014). Single image haze removal using

content-adaptive dark channel and post enhancement. *IET Computer Vision, 8*(2), 131-140.

Li, C., Yan, W. (2021) Braille recognition using deep learning. *International Conference on Control and Computer Vision.*

Li. C. (2022) *Special Character Recognition Using Deep Learning. Master's Thesis* Auckland University of Technology, New Zealand.

Li, F., Zhang, Y., Yan, W., Klette, R. (2016) Adaptive and compressive target tracking based on feature point matching. *International Conference on Pattern Recognition* (ICPR), 2734-2739.

Li, K., Li, S. (2011). Research of image enhancement algorithm based on histogram statistics. *Science Technology and Engineering.*

Li, P. (2018) *Rotation Correction for License Plate Recognition*. Master's Thesis, Auckland University of Technology, New Zealand.

Li, P., Nguyen, M., Yan, W. (2018) Rotation correction for license plate recognition. *International Conference on Control, Automation and Robotics*.

Li, Q., Zhao, F., Xu, Z., Wang, J., Liu, K., Qin, L. (2022) Insulator and damage detection and location based on YOLOv5. *International Conference on Power Energy Systems and Applications.*

Li, W. (2018) Traffic sign detection under complex light conditions. *Chinese Scientific*

Journal, 13 (2): 131-135.

Liang, C., Lu, J., Yan, W. (2022) Human action recognition from digital videos based on deep learning. *ACM ICCCV 2022.*

Liang, M., Yuan, M., Hu, X., Li, J., Liu, H. (2013). Traffic sign detection by ROI extraction and histogram features-based recognition. *International Joint Conference on Neural Networks (IJCNN),* 1-8.

Line, E., Lars, A., Hans, K. (2009). Classification-based vehicle detection in high-resolution satellite images. *ISPRS Journal of Photogrammetry & Remote Sensing, 64*(1), 65-72.

Litman, T., Burwell, D. (2010). Issues in sustainable transportation. *International Journal of Global Environmental Issues, 6*(4), 331-347.

Liu, H., Yang, J., Wu, Z., Zhang, Q., & Deng, Y. (2016). Fast single image dehazing based on interval estimation. *Journal of Electronics and Information Technology, 38*(2), P. 381-388.

Liu, M., Yan, W. (2022) Masked face recognition in real-time using MobileNetV2. *ACM ICCCV 2022.*

Liu, Q., Jia, R., Zhao, C., & Liu, X. (2019). Face super-resolution reconstruction based on self-attention residual network. *IEEE Access.*

Liu, X., Nguyen, M., Yan, W. (2019) Vehicle-related scene understanding using deep

learning. *Asian Conference on Pattern Recognition.*

Liu, X. (2019) *Vehicle-related Scene Understanding Using Deep Learning.* Master's Thesis, Auckland University of Technology, New Zealand.

Liu, X., Yan, W. (2020) Vehicle-related scene segmentation using CapsNets. *International Conference on Image and Vision Computing New Zealand.*

Liu, X., Yan, W. (2021) Traffic-light sign recognition using Capsule network. *Springer Multimedia Tools and Applications.*

Luo, Z., Nguyen, M., Yan, W. (2022) Kayak and sailboat detection based on the improved YOLO with Transformer. *ACM ICCCV 2022.*

Lykele, H., Ivo, M, C., With, de P.H.N. (2014). Exploiting street-level panoramic images for large-scale automated surveying of traffic signs. *Machine Vision & Applications. 25*(7), 1893-1911.

Ma, X., Fu, A., Wang, H., & Yin, B. (2018). Hyperspectral image classification based on deep deconvolution network with skip architecture. *IEEE Transactions on Geoscience and Remote Sensing,* pp. 4781-4791.

Ma, X. (2020) *Banknote Serial Number Recognition Using Deep Learning.* Master's Thesis, Auckland University of Technology, New Zealand.

Ma, X., Yan, W. (2021) Banknote serial number recognition using deep learning. *Springer Multimedia Tools and Applications*.

Maas, A. L., Hannun, A. Y., & Ng, A. Y. (2013). Rectifier nonlinearities improve neural network acoustic models. *Stanford University, USA*.

Maldonado-Bascon, S., Lafuente-Arroyo, P., Gil-Jimenez, H., Gomez-Moreno., Lopez-Ferreras, F. (2007). Road-sign detection and recognition based on support vector machines. *IEEE Transactions on Intelligent Transportation Systems, 8*(2), 264-278.

Mehtab, S., Yan, W. (2021) FlexiNet: Fast and accurate vehicle detection for autonomous vehicles-2D vehicle detection using deep neural network. *International Conference on Control and Computer Vision*.

Mehtab, S., Yan, W. (2022) Flexible neural network for fast and accurate road scene perception. *Multimedia Tools and Applications.*

Mehtab, S. Yan, W., Narayanan, A. (2022) 3D vehicle detection using cheap LiDAR and camera sensors. *International Conference on Image and Vision Computing New Zealand.*

Ming, Y., Li, Y., Zhang, Z., Yan, W. (2021) A survey of path planning algorithms for autonomous vehicles. *International Journal of Commercial Vehicles*.

Miura, J., Kanda, T., Shirai, Y. (2000). An active vision system for real-time traffic sign recognition. *IEEE Intelligent Transportation Systems,* 52-57.

Moutarde, F., Bargeton, A., Herbin, A., Chanussot, L. (2007). Robust on-vehicle real-

time visual detection of American and European speed limit signs, with a modular traffic signs recognition system. *Intelligent Vehicles Symposium,* 1122-1126.

Nan, N., Gang, R., & Song, R. (2020). Image defogging algorithm based on Fisher criterion function and dark channel prior. *International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*.

Narasimhan, S.G., Nayar, S.K. (2003). Contrast restoration of weather degraded images. *IEEE Transactions on Pattern Analysis & Machine Intelligence, 25*(6), 713-724.

Nayar, S. K., & Narasimhan, S.G. (1999). Vision in bad weather. *IEEE International Conference on Computer Vision.*

Nguwi, Y.Y., Kouzani, A.Z. (2008). Detection and classification of road signs in natural environments. *Neural Computing and Applications, 17*(3), 265-289.

Pal, S. A., King, R. A. (1981). Image enhancement using smoothing with fuzzy sets. *IEEE Transactions on Systems Man and Cybernetics, 11*, 494-501.

Pan, C., Sun, M., Yan, Z., Shao, J., Wu, D., Xu, X. (2013). Vehicle logo recognition based on deep learning architecture in video surveillance for intelligent traffic system. *International Conference on Smart and Sustainable City.*

Pan, C., Yan, W. (2018) A learning-based positive feedback in salient object detection. International Conference on Image and Vision Computing New Zealand.

Pan, C., Yan, W. (2020) Object detection based on saturation of visual perception. *Multimedia Tools and Applications*, 79 (27-28), 19925-19944.

Pan, C., Liu, J., Yan, W., Zhou, Y. (2021) Salient object detection based on visual perceptual saturation and two-stream hybrid networks. *IEEE Transactions on Image Processing.*

Peng, J., Liu, B., Dong, W., Wang, J., & Wang, Y. (2008). Method of image enhancement based on multi-scale retinex. *Laser and Infrared, 38(11), 1160~1163.*

Priese, L., Klieber, J., Lakmann, R., Rehrmann, V., Schian. R. (1994). New results on traffic sign recognition. *Proceedings of Intelligent Vehicles,* 249-254.

Provenzi, E., Fierro, M., Rizzi, A., Carli, L., Gadia, D., & Marini, D. (2007). Random spray retinex: A new retinex implementation to investigate the local properties of the model. *IEEE Transactions on Image Processing, 16*(1):162~171.

Psyllos, A., Anagnostopoulos, C.N., Kayafas, E. (2011). Vehicle model recognition from frontal view image measurements. *Computer Standards & Interfaces, 33*(2),142-151.

Qi, X., Huang, Y., & Liu, W. (2005). A sign recognition method based on moment invariants and wavelet neural network. *Journal of Changsha Jiao Tong University,* (02), 85-89.

Qin, Z., Yan, W. (2021) Traffic-sign recognition using deep learning. *International Symposium on Geometry and Vision.*

Radhakrishna, A., Yan, W., Kankanhalli, M. (2006) Modeling intent for home video repurposing. *IEEE MultiMedia* 13 (1), 46-55.

Redmon, J., & Farhadi, A. (2018). YOLOv3: An incremental improvement. IEEE *CVPR.*

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* pp. 779-788.

Redmon, J., Farhadi, A. (2017). YOLO9000: Better, faster, stronger. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7263-7271.

Redmon, J., Farhadi, A. (2018). YOLOv3: An incremental improvement. *IEEE CVPR.*

Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis & Machine Intelligence, 39*(6):1137-1149.

Rezatofighi, H., Tsoi, N., Gwak, J., Sadeghian, A., Reid, I., & Savarese, S. (2019). Generalized intersection over union: A metric and a loss for bounding box regression. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR),* pp. 658-666.

Ripley, B.D. (1996). Pattern Recognition and Neural Networks. *Cambridge University Press*.

Rizzi, A., Gatta, C., & Marini, D. (2003). A new algorithm for unsupervised global and local color correction. *Pattern Recognition Letters,24*(11):1663~1677.

Rumelhart, D. E., & McClelland, J.L. (1987). Learning Internal Representations by Error Propagation. *MIT Press, 318-362*.

Saadna, Y., Behloul, A., & Mezzoudj. (2019). Speed limit sign detection and recognition system using SVM and MNIST datasets. *Neural Computing and Applications, 31*(1).

Saunders, C. , Stitson, M. O., Weston, J. (2002). Support vector machine. *Computer Science, 1*(4):1-28.

Sermanet, P., Lecun. Y. (2011). Traffic sign recognition with multi-scale convolutional networks. *International Joint Conference on Neural Networks,* 2809-2813.

Setiawan, A. W., Mengko, T.R., Santoso, O.S., & Suksmono. (2013). Color retinal image enhancement using CLAHE. *International Conference on ICT for Smart Society.*

Shen, D., Xin, C., Nguyen, M., Yan, W. (2018) Flame detection using deep learning. *International Conference on Control, Automation and Robotics.*

Shen, Y., Yan, W. (2019) Blind spot monitoring using deep learning. *International*

Conference on Image and Vision Computing New Zealand.

Shi, S., Zhang, Y., Zhou, X., Cheng, J. (2021). Cloud removal for single visible image based on modified dark channel prior with multiple scale. *IEEE International Geoscience and Remote Sensing Symposium IGARSS.*

Shi, X., Fang, X., Zhang, D., Guo, Z. (2016). Image classification based on mixed deep learning model transfer learning. *Journal of System Simulation.*

Shwartz, S., Namer, E., Schechner, Y. Y. (2006). Blind Haze Separation. *IEEE*

Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. *IEEE CVPR.*

Stallkamp, J., Schlipsing, M., Salmen, J., Igel, C. (2012). Man vs. computer: Benchmarking machine learning algorithms for traffic sign recognition. *Neural Networks, 32,* 323-332.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., & Anguelov, D. (2015). Going deeper with convolutions. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR*), pp. 1-9

Tan, R.T. (2008). Visibility in bad weather from a single image. *IEEE Conference on Computer Vision and Pattern Recognition,* 1-8.

Tang, S., Zu, G., Nie, M. (2010). An improved image enhancement algorithm based on fuzzy sets. *International Forum on Information Technology and Applications,*

197-199.

Tarel, J.P., Hautière, N. (2009). Fast visibility restoration from a single color or gray level image. *IEEE International Conference on Computer Vision,* 2201-2208.

Tatsumi, W., Yasuhiro, K., Akio, K., & Toshiharu, K. (2005). An adaptive multi-scale retinex algorithm realizing high color quality and high-speed processing. *Journal of Imaging Science and Technology, 49*(5), 486~497.

Timofte, R., K. Zimmermann., L.V. Gool. (2009). Multi-view traffic sign detection, recognition, and 3D localization. *Machine Vision and Applications,* 633-647.

Uijlings, J. R. R., Sand, K. E. A. Van De., & Gevers, T. (2013). Selective search for object recognition. *International Journal of Computer Vision, 104*(2), 154-171.

Volpi, M., Tuia, D. (2017). Dense semantic labeling of sub-decimeter resolution images with convolutional neural networks. *IEEE Transactions on Geoscience & Remote Sensing, 55*(2), 881-893.

Wang, C., Liao, H., Wu, Y., Chen, P., Wei, J., & Yeh, I. (2020). CSPNet: A new backbone that can enhance learning capability of CNN. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops,* pp. 390-391.

Wang, K., Liew, J., Zou, Y., Zhou, D., & Feng, J. (2019). PANet: Few-shot image semantic segmentation with prototype alignment. *IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 9197-9206.

Wang, K., Yang, F., Xu, Z. (2003) Traffic sign recognition method based on neural network. *Journal of Liaoning University of Petroleum and Chemical Technology, 23 (001): 77-79.*

Wang, J., Yan, W., Kankanhalli, M., Jain, R., Reinders, M. (2003) Adaptive monitoring for video surveillance. International Conference on Information, Communications and Signal Processing.

Wang, J., Kankanhalli, M., Yan, W., Jain, R. (2003) Experiential sampling for video surveillance. ACM SIGMM International Workshop on Video surveillance, 77-86.

Wang, J., Yan, W. (2016) BP-neural network for plate number recognition. International *Journal of Digital Crime and Forensics* (IJDCF) 8 (3), 34-45.

Wang, J. (2016) Event-driven traffic ticketing system. Master's Thesis, Auckland University of Technology, New Zealand.

Wang, J., Ngueyn, M., Yan, W. (2017) A framework of event-driven traffic ticketing system. *International Journal of Digital Crime and Forensics* (IJDCF) 9 (1), 39-50.

Wang, J., Bacic, B., Yan, W. (2018) An effective method for plate number recognition. *Multimedia Tools and Applications,* 77 (2), 1679-1692.

Wu, J., L. Zhong. (2013). A new data aggregation model for intelligent transportation

system. *Advanced Materials Research, 671-674*(3), 2855-2859.

Xing, J., Yan, W. (2021) Traffic sign recognition using guided image filtering. *International Symposium on Geometry and Vision.*

Xing, J., Nguyen, M., Yan, W. (2022) The improved framework of traffic sign recognition by using guided image filtering. *Springer Nature Computer Science.*

Xu, H., Guo, J., Liu, Q., & Ye, L. (2012). Fast image dehazing using improved dark channel prior. *IEEE International Conference on Information Science and Technology.*

Yadav, G., Maheshwari, S., & Agarwal, A. (2014). Contrast limited adaptive histogram equalization-based enhancement for real time video system. *International Conference on Advances in Computing, Communications and Informatics (ICACCI).*

Yan, J., Zhao, L., Diao, W., & Wang, H. (2021). AF-EMS detector: Improve the multi-scale detection performance of the anchor-free detector. *Remote Sensing, 2021, 13*(2), 160.

Yan, W., Kankanhalli, M., Wang, J., Reinders, M. (2003) Experiential sampling for monitoring. *ACM SIGMM Workshop on Experiential Telepresence*, 70-72.

Yan, W., Wang, J., Kankanhalli, M. (2005) Automatic video logo detection and removal. *Multimedia Systems* 10 (5), 379-391.

Yan, W., Chambers, J. (2013) An empirical approach for digital currency forensics. *IEEE International Symposium on Circuits and Systems* (ISCAS), 2988-2991.

Yan, W., Chambers, J., Garhwal, A. (2014) An empirical approach for currency identification. *Multimedia Tools and Applications* 74 (7).

Yan, W. (2019) *Introduction to Intelligent Surveillance: Surveillance Data Capture, Transmission, and Analytics*. Springer.

Yan, W. (2021) *Computational Methods for Deep Learning: Theoretic, Practice and Applications*. Springer.

Yuan, X., Liu, L.F., Qu, Y.Y. (2009). Unifying visual saliency with HOG feature learning for traffic sign detection. *IEEE Intelligent Vehicles Symposium,* 24-29.

Yurtsever, E., Lambert, J., Carballo, A., & Takeda, K. (2020). A survey of autonomous driving: Common practices and emerging technologies. *IEEE Access*.

Zeiler, M. D., & Fergus, R. (2014). Visualizing and understanding convolutional networks. *ECCV*, 818-833.

Zhang, Q. (2018) *Currency Recognition Using Deep Learning*. Master's Thesis, Auckland University of Technology, New Zealand.

Zhang, Q., Yan, W. (2018) Currency detection and recognition based on deep learning. *IEEE International Conference on Advanced Video and Signal Based Surveillance*.

Zhang, Q., Yan, W., Kankanhalli, M. (2019) Overview of currency recognition using deep learning. *Journal of Banking and Financial Technology,* 3 (1), 59–69.

Zheng, K., Yan, Q., Nand, P. (2017) Video dynamics detection using deep neural networks. *IEEE Transactions on Emerging Topics in Computational Intelligence*.

Zhu, J., D. Wang. (2015). Fast smoothing technique with edge preservation for single image dehazing. *Computer Vision IET, 9*(6), 950-959.

Zhu, Y., Yan, W. (2022) Ski fall detection from digital images using deep learning. *ACM ICCCV, 2022.*

Zhu, Y., Yan, W. (2022) Image-based storytelling using deep learning. *ACM ICCCV, 2022.*

Zhu, Y., Yan, W. (2022) Traffic sign recognition based on deep learning. *Multimedia Tools and Applications.*