applied sciences

MDPI

Study Protocol

Traffic Light Detection and Recognition Method Based on YOLOv5s and AlexNet

Chuanxi Niu and Kexin Li *

School of Civil Engineering and Transportation, Beihua University, Jilin 132000, China
* Correspondence: likenefu@beihua.edu.cn

Abstract: Traffic light detection and recognition technology are of great importance for the development of driverless systems and vehicle-assisted driving systems. Since the target detection algorithm has the problems of lower detection accuracy and fewer detection types, this paper adopts the idea of first detection and then classification and proposes a method based on YOLOv5s target detection and AlexNet image classification to detect and identify traffic lights. The method first detects the traffic light area using YOLOv5s, then extracts the area and performs image processing operations, and finally feeds the processed image to AlexNet for recognition judgment. With this method, the shortcomings of the single-target detection algorithm in terms of low recognition rate for small-target detection can be avoided. Since the homemade dataset contains more low-light images, the dataset is optimized using the ZeroDCE low-light enhancement algorithm, and the performance of the network model trained after optimization of the dataset can reach 99.46% AP (average precision), which is 0.07% higher than that before optimization, and the average accuracy on the traffic light recognition dataset can reach 87.75%. The experimental results show that the method has a high accuracy rate and can realize the recognition of many types of traffic lights, which can meet the requirements of traffic light detection on actual roads.

Keywords: traffic light; YOLOv5; image processing; AlexNet

Citation: Niu, C.; Li, K. Traffic Light Detection and Recognition Method Based on YOLOv5s and AlexNet. *Appl. Sci.* **2022**, *12*, 10808. https://doi.org/10.3390/app122110808

Academic Editor: Hui Yuan

Received: 18 September 2022
Accepted: 18 October 2022
Published: 25 October 2022

1. Introduction

Traffic light detection and recognition are important topics in the field of intelligent transportation. Detecting and recognizing the status of traffic lights through machine vision is one of the important application directions of image processing and pattern recognition research and is of great significance to the development of car-assisted driving systems and driverless systems. Driverless cars can use onboard cameras to acquire images of traffic lights and recognize their status to guide the vehicle to its next move. By acquiring traffic light status, drivers can be reminded by voice to avoid ignoring important traffic light status due to inattention, which can reduce traffic accidents to a certain extent. Due to the complexity and variability of the traffic light image background in the real traffic scene, the traffic light in the image occupies fewer pixels, and its feature structure is sparse, which increases the difficulty of algorithm recognition. Therefore, it is very important to study a more effective small-target detection algorithm for traffic signal detection [1].

At present, many scholars have conducted research on the detection and recognition of traffic lights. The research methods can be broadly classified into traditional methods based on image processing and deep learning methods based on convolutional neural networks.

In the field of traditional methods based on image processing, Xu et al. [2] generated colour, luminance, and edge feature maps and fused them to obtain salient maps based on the characteristics of traffic signals, then obtained candidate regions and performed target segmentation, and finally used the HOG feature extraction algorithm and SVM classifier for recognition; experiments showed that the detection and recognition rates of the algorithm

were above 97%. SHI et al. [3] used the pattern of detecting traffic light colour change to locate its area, and finally, the SVM classifier was used to identify it with an accuracy of 98%. Ji et al. [4] obtained candidate regions of traffic lights by the visual selective attention (VSA) model, then used HOG features of traffic lights and an SVM classifier to obtain the exact regions. The accuracy could reach more than 97%. Jang et al. [5] proposed a high-exposure technique based on the integration of low and normal-exposure techniques and used this as a tool for candidate area selection. Omachi et al. [6] first defined a mathematical model of traffic lights, then clustered each pixel of the input image into five types, and pixels with traffic light colours were used for Sobel edge detection; finally, the model and the detected edges were used for voting to detect traffic lights. The completed method can obtain 89% accuracy. Kim et al. [7] used colour adjustment, thresholding algorithms, and median filtering techniques to achieve traffic light detection. The method can obtain 90% and 81% accuracy in daytime and nighttime, respectively. Zhang et al. [8] used colour segmentation, speckle detection, and structural feature extraction to obtain candidate regions for the traffic light and scored the detected points and blocks in the image to finally achieve detection. The average detection accuracy of the method reached up to 96.07%.

However, traditional image processing methods have limitations in image detection and image recognition, and in real roads, the acquisition of traffic light images can be affected by light, weather, and other external environments. For example, the traffic light image of the city road at night due to the lens of the dazzling phenomenon, and the existence of the irregular shape of the halo, coupled with the jitter of the vehicle itself and other external environments on the traffic light interference and obscuration, etc., will make the obtained image produce more noise, defacement, etc., and may even cause the image to be incomplete, which greatly enhances the difficulty of traffic signal detection and recognition.

Deep learning has developed rapidly in recent years due to increased computing power and is increasingly being used in vision-based target recognition. Some researchers are beginning to favour the use of deep learning for traffic light detection and recognition. Li et al. [9] introduced contextual features into the field of traffic light recognition and combined the advantages of the convolutional neural network (CNN) algorithm to propose a recognition algorithm with gradual features. However, the model used in this algorithm is too simple, the number of parameters is small, the feature extraction ability is poor, and the detection results are not satisfactory. Kim et al. [10] applied the original SSD to the field of traffic light detection, first using a coarse-grained CNN to detect traffic light regions. The incorrect samples were then removed using a fine-grained CNN. The method is a complex process and only achieved a 68.0% accuracy and 29.5% recall for red circular signals in a homemade dataset, making it difficult to apply in practical scenarios. Muller et al. [11] improved the structure of SSD and proposed TL-SSD, which combines contextual and local information for the detection and recognition of traffic lights; it can detect small objects accurately, detecting 95% of the lights on their homemade dataset, but it can only reach ten frames per second, making it difficult to guarantee real-time performance. Lu et al. [12] borrowed the concept of a visual attention mechanism by first feeding a low-resolution image into the first network to determine the area where the traffic signal is located and then feeding a high-resolution image of the corresponding location into the second network and eventually determining the exact location of the traffic signal. However, the recognition accuracy on their homemade dataset is not very high. Niu et al. [13] proposed a framework for traffic light detection that combines traditional methods with deep learning. A heuristic candidate area selection module is first used to detect possible traffic light areas, and then a lightweight convolutional neural network classifier is used to classify the obtained results. Experimental results show that the framework is very fast in detection, with an accuracy of 96.6% of the training results on the LISA [14] (Laboratory for Intelligent and Safe Automobiles) dataset. However, it also has a 33.3% misdetection rate. Wang et al. [15] proposed an STL-YOLO traffic light detection model to address the problem of poor detection of small-scale targets by using two methods—"spanning feature fusion" and

"clustering scaling to obtain a new prior frame"—to improve the map by about 9%. Zhou et al. [16] proposed an improved YOLOv4-based algorithm for traffic light countdown digital detection and recognition. By improving the original YOLOv4 network model, the recognition accuracy was not only improved but also achieved the recognition of countdown digital signals. Pan et al. [17] implemented Faster-RCNN based traffic light detection and recognition and compared the effect of target detection and recognition under three different feature extraction networks: VGG16, ResNet50, and ResNet101, respectively. The experimental results show that the traffic light recognition method based on the Faster-RCNN framework (ResNet50) can achieve the optimal effect and its accuracy can reach 95.1%.

In summary, traditional nondeep learning methods achieve better results when using manually extracted traffic light colour and shape information for candidate area detection. However, the scene in which the traffic light is located is complex and changeable, and it is difficult to extract its features by a single algorithm, so the traditional algorithm is only effective for scenes with simple backgrounds and is not very practical. Although deep learning methods do not require the design of complex manual feature extraction methods, they are more generalizable to different scenarios and are more portable [18]. However, most of the current research in deep learning is on the detection and recognition of circular and arrow signals, while the detection of digital countdown lights is an important part of traffic lights, but few research scholars have conducted recognition studies for this type of signal. This paper, therefore, proposes a method based on YOLOv5s target detection and the AlexNet image recognition network to detect multiple types of traffic lights, both for circular and arrow signals and also for digital countdown lights.

The contributions of the proposed approach are as follows: (i) The use of two different algorithms in the target detection and image classification stages avoids the drawback of a single-target detection algorithm with a low recognition rate for small-target detection; (ii) traffic signals can be detected and identified in different installation directions; (iii) detection and identification of various types of traffic lights can be realized.

The remainder of the paper is structured as follows: Section 2 outlines the relevant work in the traffic light detection and recognition method proposed in this paper, focusing on the network structure of YOLOv5s and AlexNet. Section 3 outlines the specific experimental procedure of the method in this paper, including the acquisition of the dataset and the training of the network model. Section 4 presents an analysis of the results of traffic light detection and recognition. Finally, Section 5 concludes this study.

2. Materials and Methods

The original YOLOv5 network model already includes the detection of traffic lights, but YOLOv5 is a single-stage target detection algorithm, and this type of algorithm completes the detection and recognition of objects together, so there may be false detection and missed detection in complex scenes, as shown in Figure 1.

In Figure 1, the area positioned by the box is the result of target detection. Analysis of Figure 1 shows that in poor lighting conditions at night, relatively large targets such as pedestrians and vehicles can be accurately detected, while relatively small targets such as traffic lights are missed, and the exact location of the traffic light cannot be detected, making it impossible to classify the traffic light. Traffic light detection requires a high level of accuracy because it affects subsequent decisions on vehicle control [19]. Therefore, this paper proposes a method based on YOLOv5s target detection and AlexNet image recognition network, which processes target detection and image recognition separately. Through this method, the shortcomings of a single-target detection algorithm in terms of low recognition rate for small-target detection can be avoided, and the recognition of multiple types of traffic lights can be achieved at the same time.

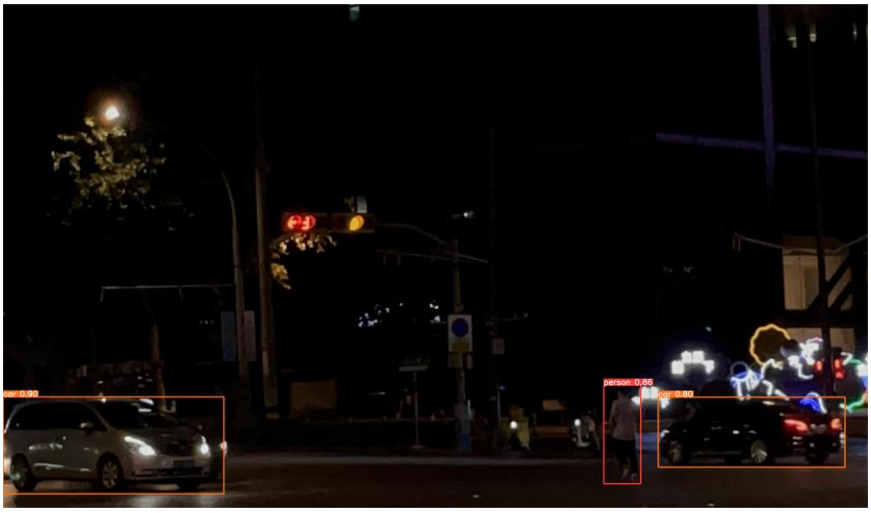


Figure 1. YOLOv5 misses on traffic light.

The flow chart of traffic light detection and recognition in this paper is shown in Figure 2. The solution first uses the trained YOLOv5s network to detect the location of the traffic light in the target scene, extracts the target area, and performs image processing to obtain a single traffic light image, and then feeds the obtained image into the trained AlexNet for image recognition and combines the output results to obtain the final information. This section is followed by an introduction to the network structure of YOLOv5s and AlexNet.

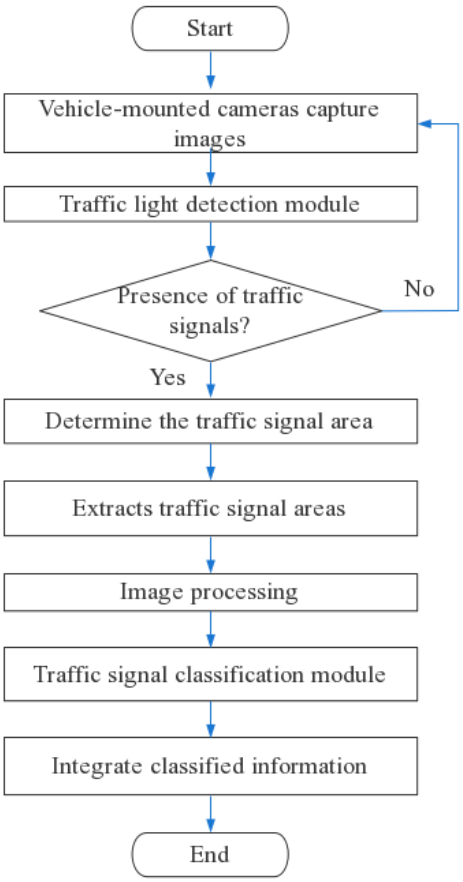


Figure 2. Traffic light detection and recognition flow chart.

2.1. YOLOv5s

Redmon et al. [20] proposed a deep learning framework in 2015 that can perform real-time target detection and run efficiently. YOLOv5 is divided into four versions: YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x. The performance comparison of its four versions is shown in Table 1.

Table 1. Performance comparison of different versions of YOLOv5.

Model	Size (Pixels)	mAP (0.5:0.95)	mAP (0.5)	Speed v100(ms)	Params (M)	FLOPs (G)
YOLOv5s	640 × 640	37.4	56.8	0.9	7.2	16.5
YOLOv5m	640 × 640	45.4	64.1	1.7	21.2	49.0
YOLOv5l	640 × 640	49.0	67.3	2.7	46.5	109.1
YOLOv5x	640 × 640	50.7	68.9	4.8	86.7	205.7

YOLOv5s has a high speed of detection and accuracy, and its model is tiny [21]. The smaller the network model, the lower the performance requirements for mobile and the easier it is to deploy [22]. Therefore, for the consideration of the weight file size, recognition accuracy, and detection speed of the network model, this paper selects YOLOv5s, which has the fastest detection speed and relatively high recognition accuracy for the study. The YOLOv5s network structure can be divided into four parts: Input, Backbone, Neck, and Prediction, and its network structure is shown in Figure 3.

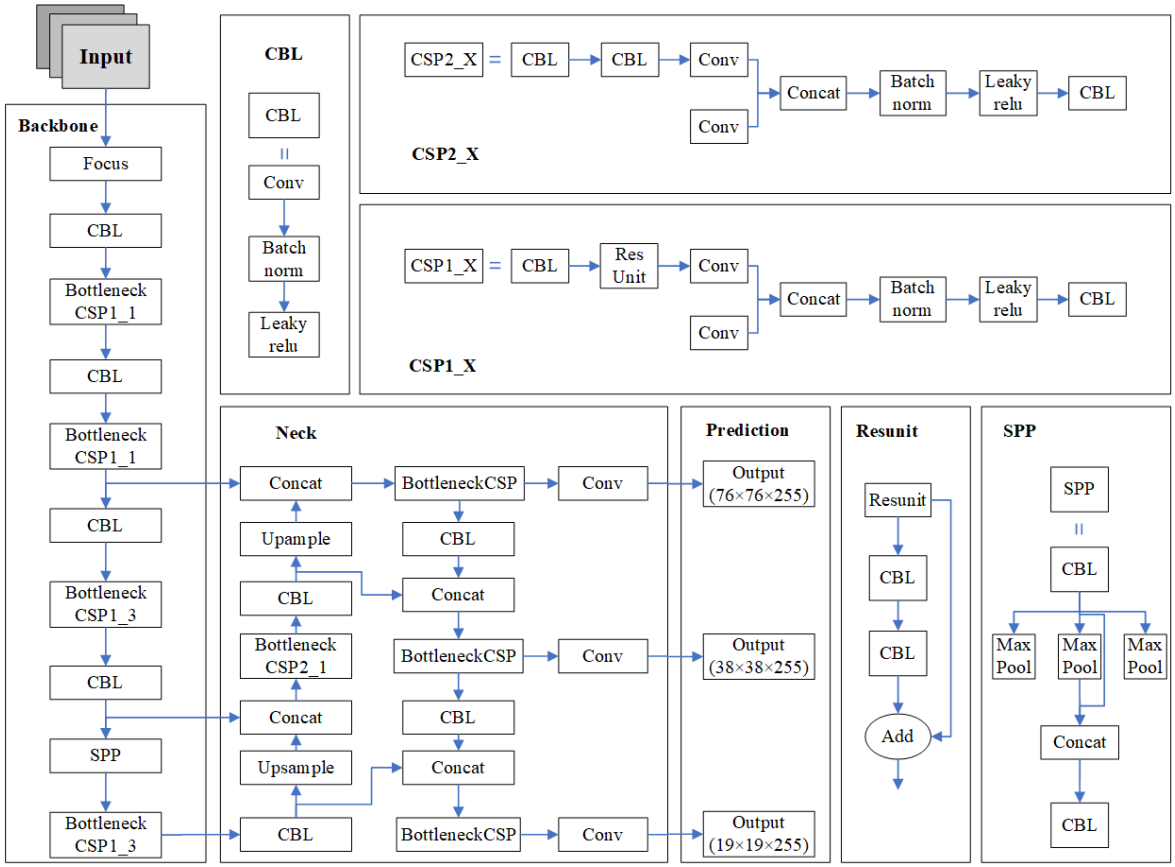


Figure 3. YOLOv5s network structure.

The Input side consists of a total of three parts: Mosaic data enhancement, adaptive anchor frame calculation, and adaptive image scaling. The backbone network forms a convolutional neural network to extract image features by aggregating different types

of image information [23]. It consists of three main components: Focus, the Cross Stage Partial Network (CSP) structure, and the SPP [24] module. The Neck network uses the ReLU activation function to improve network feature fusion by using the FPN [25] + PAN [26] network structure to aggregate features extracted from the backbone and detection networks [27]. On the prediction side, YOLOv5s uses the GIOU loss function to solve the problem of disjoint bounding boxes that the previous YOLO series could not optimize.

2.2. AlexNet

AlexNet was designed by Alex Krizhevsky et al. [28] and was the winner of the ImageNet large-scale visual recognition 2012 competition. AlexNet is a type of convolutional neural network which is now widely used in machine learning fields such as image recognition [29]. Its network model is shown in Figure 4.

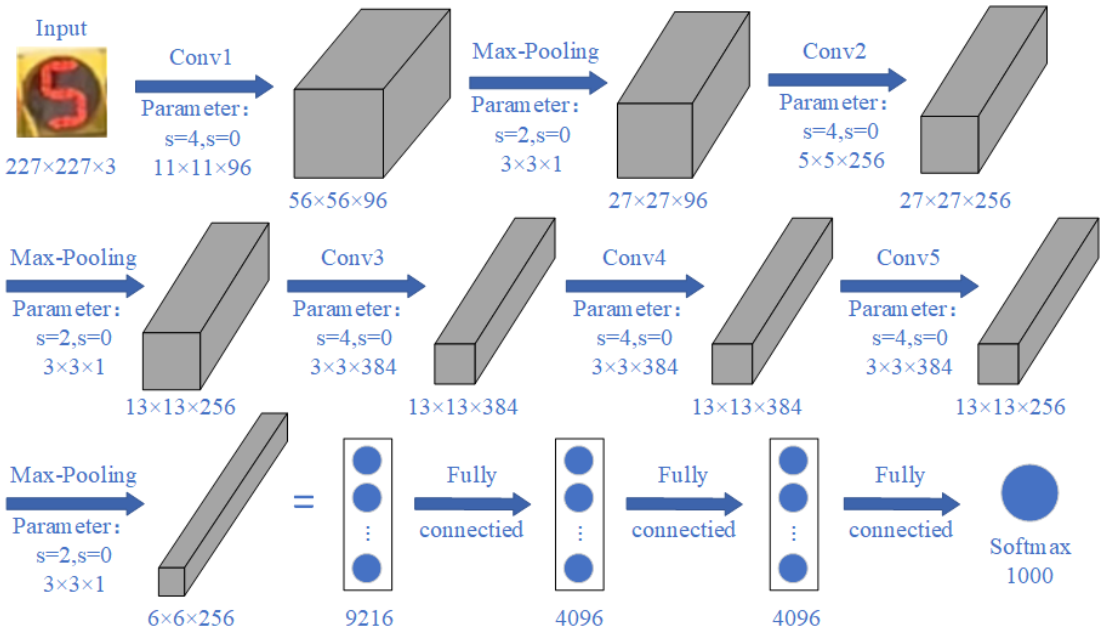


Figure 4. AlexNet network structure.

With its simple network structure and fast detection speed, AlexNet is suitable for real-time detection and recognition of traffic lights in traffic scenarios. The network contains eight learning layers, five of which are convolutional, three fully connected, and three maximum pooling layers, with the convolutional and pooling layers used in alternating connections. Instead of using the traditional Sigmoid activation function and Tanh activation function, AlexNet uses ReLU as the activation function for all convolutional and fully connected layers. However, since the value domain obtained after the ReLU function has no interval, AlexNet proposes local response normalization (LRN), which normalizes the data obtained from ReLU to suppress small neurons and improve the generalization ability of the model.

3. Experiment

The hardware platform used for the experimental part is Intel i5-12500H for the CPU, NVIDIA RTX 3050 for the GPU, window11 for the operating system, and Pytorch 1.11 for the training framework. The computing platform CUDA version is 11.3.1, and the deep neural network library CUDNN version is 8.0.5.

3.1. Traffic Light Dataset Construction

In this paper, sample traffic light images were collected from multiple intersections, and the sample images were captured at different times and directions at multiple inter-

sections using the rear camera of the mobile phone. The video is then frame-separated to obtain a sample image. Sample images of traffic lights were also collected on the web, and a total of 4200 colour images were finally used for traffic light detection, which includes complex traffic conditions such as night and foggy days. The network was trained on 3360 (80%) randomly selected datasets, and 840 (20%) were used as test sets to verify the performance of the network.

Selected images in the traffic light detection dataset, after image processing, were used as the dataset for image recognition, and a total of 1414 images were collected. There are 26 categories in total (red, green, and yellow round signals; red and green left turn signals; red and green 0–9 digit lights; and no signals).

3.2. Traffic Light YOLOv5s Detection Network Model Building

Traffic light detection mainly includes video frame extraction, image preprocessing, and image annotation, as well as YOLOv5s-based model training and target detection. The specific process is shown in Figure 5.

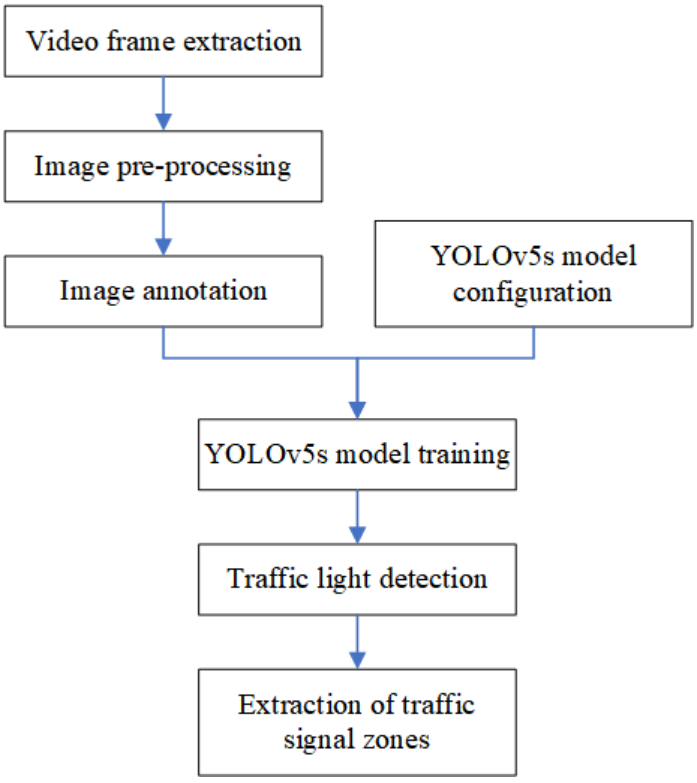


Figure 5. Traffic light detection flow chart.

As the target detection part only needs to determine the location of the traffic lights and does not require image recognition, only one class of image needs to be annotated, which greatly reduces the time spent on the annotation process and also reduces the number of samples required.

The Epoch of the YOLOv5s network was set to 100 for this experiment. When the YOLOv5s network model was finished training, a nighttime intersection image, as well as a foggy intersection image, were selected and fed into the trained YOLOv5s network model, and the output images are shown in Figures 6 and 7.

Figure 6. YOLOv5s detection night image output results.

Figure 7. YOLOv5s fog detection image output results.

The area checked by the box in both images is the detected traffic signal area. Analysis of the detection results of the two images shows that the network model is still very accurate in detecting traffic lights in relatively complex traffic scenes and has a high confidence level.

When a traffic light is detected, the network will also output the coordinates of the four points where the traffic light is located. Next, we will perform a series of image processing operations on the target based on the coordinates of these four points to meet the requirements that can be fed into AlexNet for final image recognition.

3.3. Traffic Light Image Processing

After determining the location of the traffic light, you first need to determine the direction of the traffic light installation. The installation of the traffic light is generally divided into vertical and horizontal installations, as shown in Figure 8.

Figure 8. Vertical and horizontal installation of traffic light.

Figure 8 shows a vertically mounted traffic light on the left and a horizontally mounted traffic light on the right.

Based on the detected traffic light area, the upper left coordinate point (X_1, Y_1) and the lower right coordinate point (X_2, Y_2) of the traffic light area is output, and then whether the traffic light is installed vertically or horizontally is determined according to Equation (1).

$$F = (X_2 - X_1) - (Y_2 - Y_1) \tag{1}$$

If $F > 0$, the width of the area is greater than the height, which means that the traffic signal is horizontally mounted, and conversely, if $F < 0$, the signal is vertically mounted.

When the traffic light is mounted horizontally, the image is cut by lines L_1 and L_2. The expressions for lines L_1 and L_2 are:

$$L_1 = (X_2 - X_1)/3 + X_1 \tag{2}$$

$$L_2 = 2(X_2 - X_1)/3 + X_1 \tag{3}$$

When the traffic light is installed vertically, the image is cut by lines L_3 and L_4. The expressions for lines L_3 and L_4 are:

$$L_3 = (Y_2 - Y_1)/3 + Y_1 \tag{4}$$

$$L_4 = 2(Y_2 - Y_1)/3 + Y_1 \tag{5}$$

Finally, a single traffic light image is obtained, as shown in Figure 9.

Figure 9. Results of image processing.

After processing the image, the original image is divided into three equal parts. Complex traffic signals are also split into individual simple traffic signals. This not only reduces the variety of samples but also reduces the time required to train the network. Finally, the obtained individual traffic light images are sequentially fed into the trained AlexNet for image recognition, and the recognition results are combined to obtain accurate traffic signals.

3.4. Traffic Light AlexNet Recognition Network Model Construction

A convolutional neural network (AlexNet)-based image recognition method is used for the problem of traffic light recognition. Before training the model, the collected individual traffic light images are first preprocessed, which is divided into two main parts: image resizing and data enhancement.

The default input for the AlexNet network is 227 × 227. As the size of the images in the dataset was not uniform, it was first necessary to standardize the resolution of the images and modify the resolution of all individual traffic light images to 227 × 227.

The number of samples in the dataset was increased by data augmentation. Common data enhancement methods include horizontal flipping, vertical flipping, random rotation, etc. However, considering the peculiarities of digital traffic light images, which may produce dissimilarity after geometric transformations such as flipping or rotation. Therefore, this experiment uses only pixel transformations to achieve data enhancement by adding salt-and-pepper noise and Gaussian noise. The processed image is shown in Figure 10.

Figure 10. Data enhancement results.

The original image is shown on the left in Figure 10, the image in the middle after adding salt-and-pepper noise, and the image on the right after adding Gaussian noise. After data enhancement, the final total number of samples was 4242. A total of 80% of the samples in each category were selected as the training set, and the remaining 20% constituted the test set in this paper.

After the image preprocessing was completed, the data were fed into the network for training. The learning rate set for this image recognition experiment was 0.001, and the epoch was 100.

4. Analysis of Experimental Results

To validate the performance of the model, evaluation metrics such as precision (P), recall (R), average precision (AP), and mean average precision (mAP) [30] are generally selected to evaluate the trained network model. Accuracy is the ratio of the correct positive class found to all positive classes found. On the other hand, recall rate describes how many real positive examples in the test set are selected by the dichotomy classifier, that is, how many real positive examples are recalled by the dichotomy classifier. The accuracy and recall rates are calculated as follows:

$$Precision = \frac{TP}{TP + FP} \times 100\% \tag{6}$$

$$Recall = \frac{TP}{TP + FN} \times 100\% \tag{7}$$

where TP indicates the number of targets correctly detected, FP represents the number of targets that were misdetected, and FN indicates the number of undetected samples. In general, the better the classifier is, the higher the average precision value is. The average precision value is the area under the precision–recall curve. The precision–recall curve in this paper is shown in Figure 11, where the epoch is the number of times all training datasets were trained.

Analysis of Figure 11 shows that the area under the precision–recall curve is as high as 0.993, which is a very desirable figure and proves that the average precision value of the training results of the model.

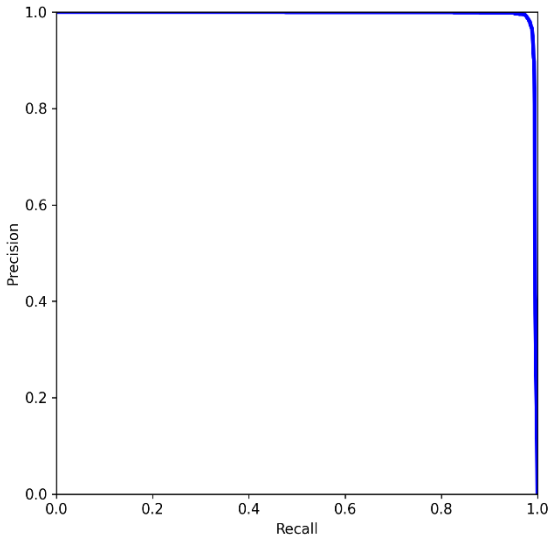


Figure 11. Precision–recall curve.

The average precision mean is the average of *AP*s of multiple categories. The mean average precision and average precision are calculated as follows:

$$AP = \int_0^1 P(R)dR \tag{8}$$

$$mAP = \frac{\sum PA}{NC} \tag{9}$$

The *mAP* is one of the most important metrics in the target detection algorithm, and its value must be in the interval [0, 1]. The larger, the better. The *mAP* curves for thresholds greater than 0.5 and for different IoU thresholds (from 0.5 to 0.95 in steps of 0.05) are shown in Figure 12.

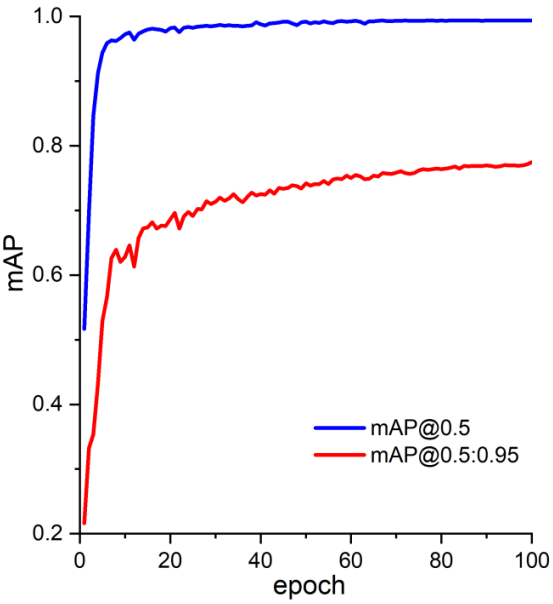


Figure 12. mAP curve.

Analysis of Figure 12 shows that the values of mAP@0.5 and mAP@0.5:0.95 gradually increase as the epoch gradually increases. The final mAP@0.5can reach 99.39%, while the mAP@0.5:0.95 can also reach 77.47%.

The loss value is another important measure of network performance and is used to gauge how close the network is to a perfect prediction when making predictions on the entire training image set. In YOLOv5s, GIoU is used as the loss function of the Bounding box, with smaller loss values representing more accurate detection of the box. The mean GIoU loss function curve is shown in Figure 13.

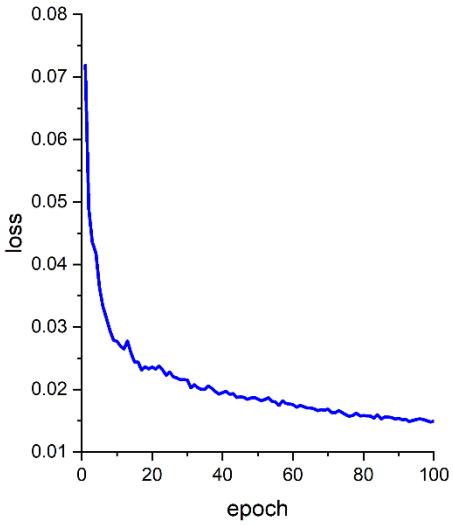


Figure 13. GIoU loss function mean value plot.

Analysis of Figure 13 shows that as the number of training sessions increases, the loss function gradually becomes smaller, and at about 25 training sessions the rate of decline decreases and begins to converge gradually, finally levelling off and reaching the ideal state.

In this paper, we also compare the performance exhibited by YOLOv5s, YOLOv4 [31], and Faster R-CNN [32] in traffic light detection by applied transfer learning. The test results are shown in Table 2.

Table 2. Performance comparison of Y6OLOv5s, YOLOv4, and Faster R-CNN.

Network	Loss	mAP/%
YOLOv5s	0.016	99.39%
YOLOv4	0.036	64.79%
Faster R-CNN	0.284	56.92%

By comparing the mAPs of the three methods, it is obvious that YOLOv5s have higher recognition accuracy compared with YOLOv4 and Faster R-CNN. To demonstrate the validity of the experimental results more accurately, we will use these three methods to detect the same image, as shown in Figure 14.

The detection results of YOLOv5s (a), YOLOv4 (b) and Faster R-CNN (c) are shown in Figure 14. From the comparison, it can be seen that YOLOv4 has the worst result and does not detect the traffic light. yolov5s can accurately locate the traffic light with high confidence. faster R-CNN can detect the traffic light, but the detected position deviates more and the result is not satisfactory.

In previous target detection experiments, the performance difference between YOLOv5 and YOLOv4 was not large, but there is a large difference in performance between the two in this experiment. We believe there are two reasons for this.

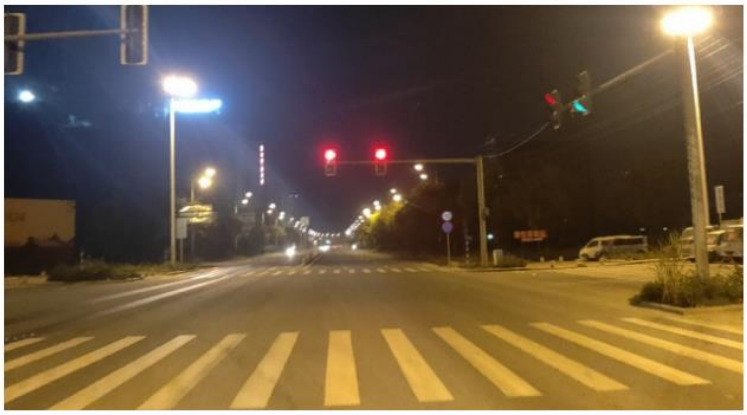
Figure 14. Assay results for YOLOv5s (**a**), YOLOv4 (**b**) and Faster R-CNN (**c**).

The first reason is that YOLOv5 uses methods for data enhancement such as Scaling, Colour Space Adjustment and Mosaic enhancement, while YOLOv4 uses many methods of information removal, such as Random Erase, Cutout, Hide and Seek. These methods are used to achieve data enhancement by masking and removing one or some areas of the image. The core requirement of information deletion methods is to avoid excessive deletion and retention of consecutive regions. However, when YOLOv4 performs information deletion on small-target images such as traffic lights, there is a high probability that the detection target will be completely deleted, such that the remaining information is not sufficient for target detection. Eventually, large detection errors may occur.

The second reason, YOLOv5 provides us with two optimization functions, Adam and SGD, which are generally a better choice when training smaller datasets, while SGD is more suitable for training large datasets. Since the dataset in this paper is small, the Adam optimization function is chosen for training YOLOv5. However, YOLOv4 uses the SGD optimization function by default. Therefore, the reason for the poor performance of YOLOv4 in this experiment may also be due to the use of the SGD optimization function.

The dataset in this paper contains a large number of low-light images. To further improve the performance of the network model, we chose to use ZeroDCE for the low-light enhancement algorithm, which has a superior runtime compared to similar models [33]. The images before and after processing are shown in Figure 15.

After using ZeroDCE for the low-illumination enhancement algorithm, we retrained the network model and compared the training results with the original model, and the comparison results are shown in Figure 16.



(a)

(b)

Figure 15. The original image (**a**) and the image processed by the low-light enhancement algorithm (**b**).

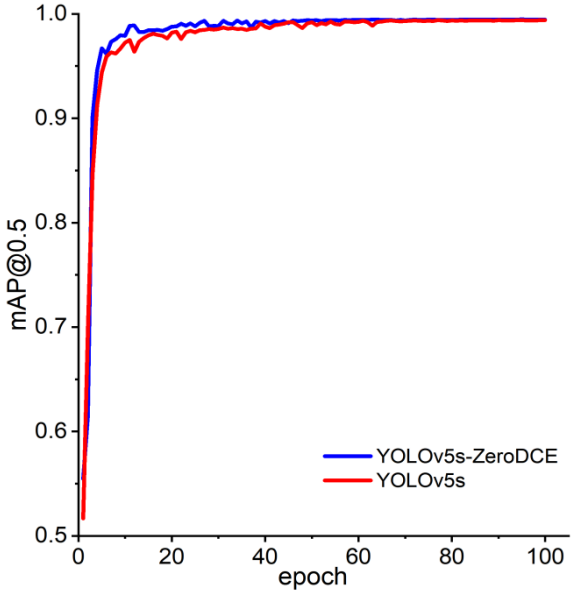


Figure 16. Comparison chart of mAP between yolov5s-ZeroDCE and YOLOv5s.

From the figure, we can see that the mAP of the model after performing the low-light enhancement algorithm is generally better than that of the original network model. The final mAP can reach 99.46%, which is an improvement of 0.07% compared with the original model. Therefore, when there are a large number of low-light images in the dataset, a reasonable low-light enhancement algorithm can improve the performance of the network model to a certain extent.

The accuracy and loss values are chosen as the evaluation metrics for the performance of the network model in the recognition task. The accuracy variation curve and loss value variation curve of the recognition model is shown in Figure 17.

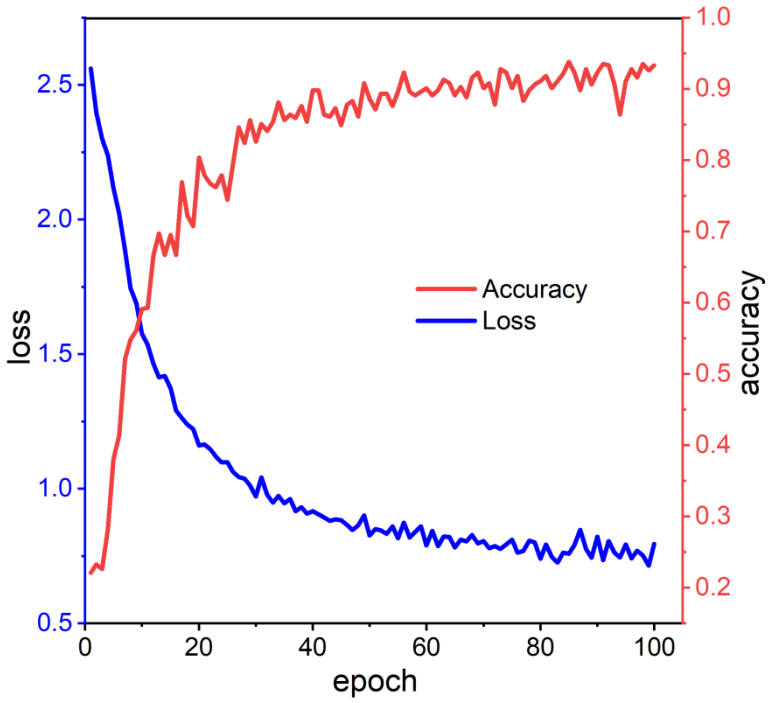


Figure 17. Loss and accuracy curves.

Analysis of Figure 17 shows that as the epoch increases, the loss value gradually decreases, while the accuracy rate gradually increases until 93.3% and stabilizes. As the loss value decreases, the accuracy can remain around the maximum value, meaning that the network is becoming better at distinguishing between categories even if the final prediction remains the same.

This paper also compares the traffic signal recognition performance of AlexNet, VGG16 [34], and ResNet18 [35] under the same parameters by applied transfer learning. The test results of the three types of classification networks are shown in Table 3.

Table 3. Performance comparison between AlexNet, VGG16, and ResNet18.

Network	Layers	Loss	Accuracy
AlexNet	8	0.668	0.878
VGG16	16	0.705	0.694
ResNet18	18	0.457	0.866

It can be seen from Table 3 that AlexNet has a higher recognition accuracy and a simpler network structure. Therefore, based on the careful consideration of recognition accuracy and detection speed, AlexNet was finally selected as the research object.

Once the traffic light recognition network was trained, we fed two preprocessed images into this network for validation, which is shown in Figures 18 and 19.

Figure 18. Validation result 1.

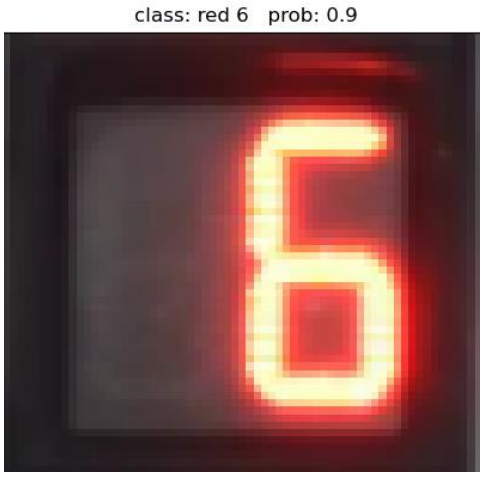


Figure 19. Validation result 2.

Figures 18 and 19 show pictures of the same set of traffic signals. The validation result obtained by the recognition network model in Figure 18 is "left red," with a confidence level of 91.5%. The validation result in Figure 19 is "red 6," with a 90% confidence level. By combining the verification results, the complete information about the group of traffic signals "red left, red 6" can be obtained. The specific meaning is 6 s countdown to the left turn red signal. The validation results show that the network model can identify both arrow signals and countdown signals with high accuracy.

5. Conclusions

In this paper, we propose a traffic light detection method that combines YOLOv5s target detection and AlexNet image recognition. In the traffic light detection stage, we use YOLOv5s to determine the location of traffic lights. Then, we determine the installation direction of the traffic signal by image processing method and output individual traffic signal images. Finally, for traffic signal recognition, we use the AlexNet classification network. Since the homemade dataset contains more low-light images, this paper also uses ZeroDCE low-light enhancement algorithm to optimize the dataset, and the performance of the trained network model is improved after the dataset optimization. By using different methods in the target detection and image recognition stages, the accuracy of the target detection algorithm for small target detection and recognition is effectively improved. We also compare YOLOv5s and AlexNet with the same type of networks, and the experimental results show that YOLOv5s and AlexNet not only have a simpler network structure but

also have a higher recognition accuracy, which is very suitable for real-time detection of traffic signals.

Although the proposed method is able to detect and recognize traffic signals, the current dataset available for training and testing the model is too small and homogeneous, and traffic light categories not available in the training model are still encountered in practice. In the next research, on the one hand, the number and variety of datasets need to be expanded to improve the breadth of traffic light recognition types. On the other hand, the overall structure of the network model needs to be improved to further optimize the detection and recognition speed of the model to better achieve the purpose of real-time traffic light detection.

Author Contributions: Conceptualization, C.N.; data curation, C.N.; formal analysis, C.N.; funding acquisition, C.N.; investigation, C.N.; methodology, C.N.; project administration, C.N.; resources, C.N.; software, C.N.; supervision, K.L.; validation, K.L.; visualization, C.N.; writing—original draft, C.N.; writing—review and editing, C.N. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Science and Technology Research Project of the Jilin Provincial Education Department, grant number JJKH20220057KJ.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Wang, Q.; Zhang, Q.; Liang, X.; Wang, Y.; Zhou, C.; Mikulovich, V.I. Traffic Lights Detection and Recognition Method Based on the Improved YOLOv4 Algorithm. *Sensors* **2022**, *22*, 200. [CrossRef] [PubMed]
2. Xu, M.; Zhang, C. Traffic Light Detection and Recognition Based on Salience Map. *Comput. Digit. Eng.* **2017**, *45*, 1397–1401.
3. Shi, X.; Zhao, N.; Xia, Y. Detection and classification of traffic lights for automated setup of road surveillance systems. *Multimed. Tools Appl.* **2016**, *75*, 12547–12562. [CrossRef]
4. Ji, Y.; Yang, M.; Lu, Z.; Wang, C. Integrating visual selective attention model with HOG features for traffic light detection and recognition. In Proceedings of the 2015 IEEE Intelligent Vehicles Symposium (IV), Seoul, Korea, 28 June–1 July 2015; pp. 280–285.
5. Jang, C.; Kim, C.; Kim, D.; Lee, M.; Sunwoo, M. Multiple exposure images based traffic light recognition. In Proceedings of the 2014 IEEE Intelligent Vehicles Symposium Proceedings, Dearborn, MI, USA, 8–11 June 2014; pp. 1313–1318.
6. Omachi, M.; Omachi, S. Detection of traffic light using structural information. In Proceedings of the IEEE 10th International Conference on Signal Processing Proceedings, Beijing, China, 24–28 October 2010; pp. 809–812.
7. Kim, Y.K.; Kim, K.W.; Yang, X. Real Time Traffic Light Recognition System for Color Vision Deficiencies. In Proceedings of the 2007 International Conference on Mechatronics and Automation, Harbin, China, 5–8 August 2007; pp. 76–81.
8. Zhang, Y.; Xue, J.; Zhang, G.; Zhang, Y.; Zheng, N. A multi-feature fusion based traffic light recognition algorithm for intelligent vehicles. In Proceedings of the 33rd Chinese Control Conference, Nanjing, China, 28–30 July 2014; pp. 4924–4929.
9. Li, Z.; Xu, G.; Guo, M. Traffic signal recognition method based on gradient content feature. *Transducer Microsyst. Technol.* **2018**, *37*, 38–40.
10. Kim, J.; Cho, H.; Hwangbo, M.; Choi, J.; Canny, J.; Kwon, Y.P. Deep Traffic Light Detection for Self-driving Cars from a Large-scale Dataset. In Proceedings of the 2018 21st International Conference on Intelligent Transportation Systems (ITSC), Maui, HI, USA, 4–7 November 2018; pp. 280–285.
11. Müller, J.; Dietmayer, K. Detecting Traffic Lights by Single Shot Detection. In Proceedings of the 2018 21st International Conference on Intelligent Transportation Systems (ITSC), Maui, HI, USA, 4–7 November 2018; pp. 266–273.
12. Lu, Y.; Lu, J.; Zhang, S.; Hall, P. Traffic signal detection and classification in street views using an attention model. *Comput. Vis. Media* **2018**, *4*, 253–266. [CrossRef]
13. Ouyang, Z.; Niu, J.; Liu, Y.; Guizani, M. Deep CNN-based Real-time Traffic Light Detector for Self-driving Vehicles. *IEEE Trans. Mob. Comput.* **2019**, *19*, 300–313. [CrossRef]
14. Philipsen, M.P.; Jensen, M.B.; Møgelmose, A.; Moeslund, T.B.; Trivedi, M.M. Traffic Light Detection: A Learning Algorithm and Evaluations on Challenging Dataset. In Proceedings of the 2015 IEEE 18th International Conference on Intelligent Transportation Systems, Las Palmas de Gran Canaria, Spain, 15–18 September 2015; pp. 2341–2345.
15. Wan, L.; Cui, S.; Su, B.; Song, Z. Detection and recognition of small scale traffic lights. *Transducer Microsyst. Technol.* **2022**, *41*, 149–152+160.
16. Zhou, K.; Zheng, Z.; Xiang, Y.; Zhao, M.; Tang, Y.; Song, J.; Shao, Y. Digital detection and recognition of traffic light countdowns based on improved YOLOv4. *Comput. Knowl. Technol.* **2022**, *18*, 7–9+21.

17. Pan, W.; Chen, Y.; Liu, B.; Shi, H. Traffic light detection and recognition based on Faster-RCNN. *Transducer Microsyst. Technol.* **2019**, *38*, 147–149+160.
18. Qian, H.; Wang, L.; Mou, H. Fast Detection and Identification of Traffic Lights Based on Deep Learning. *Comput. Sci.* **2019**, *46*, 272–278.
19. Yeh, T.-W.; Lin, H.-Y.; Chang, C.-C. Traffic Light and Arrow Signal Recognition Based on a Unified Network. *Appl. Sci.* **2021**, *11*, 8066. [CrossRef]
20. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.
21. Luo, Q.; Wang, J.; Gao, M.; He, Z.; Yang, Y.; Zhou, H. Multiple Mechanisms to Strengthen the Ability of YOLOv5s for Real-Time Identification of Vehicle Type. *Electronics* **2022**, *11*, 2586. [CrossRef]
22. Xing, J.; Pan, G. Research on improved YOLOv5s sign language recognition algorithm. *Comput. Eng. Appl.* **2022**, *58*, 194–203.
23. Wang, F.; Sun, Z.; Chen, Y.; Zheng, H.; Jiang, J. Xiaomila Green Pepper Target Detection Method under Complex Environment Based on Improved YOLOv5s. *Agronomy* **2022**, *12*, 1477. [CrossRef]
24. He, K.; Zhang, X.; Ren, S.; Sun, J. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *37*, 1904–1916. [CrossRef]
25. Lin, T.-Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature Pyramid Networks for Object Detection. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 936–944.
26. Liu, S.; Qi, L.; Qin, H.; Shi, J.; Jia, J. Path Aggregation Network for Instance Segmentation. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 8759–8768.
27. Deng, T.; Tan, S.; Pu, L. Research on traffic light recognition method based on improved YOLOv5s. *Comput. Eng.* **2022**, 1–13.
28. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. *Commun. ACM* **2017**, *60*, 84–90. [CrossRef]
29. Shi, C.; Tan, C.; Zuo, J.; Zhao, K. Expression Recognition Based on Improved AlexNet Convolutional Neural Network. *Telecommun. Eng.* **2020**, *60*, 1005–1012.
30. Wang, Y.; Ding, H.; Li, B.; Yang, Z.; Yang, J. Mask Wearing Detection Algorithm Based on Improved YOLOv3 in Complex Scenes. *Comput. Eng.* **2020**, *46*, 12–22.
31. Bochkovskiy, A.; Wang, C.Y.; Liao, H.Y.M. Yolov4: Optimal speed and accuracy ofobject detection. *arXiv* **2020**, arXiv:2004.10934.
32. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1137–1149. [CrossRef]
33. Al Sobbahi, R.; Joe, T. Comparing deep learning models for low-light natural scene image enhancement and their impact on object detection and classification: Overview, empirical evaluation, and challenges. *Signal Process. Image Commun.* **2022**, *109*, 116848. [CrossRef]
34. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
35. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.