


Article

NTS-YOLO: A Nocturnal Traffic Sign Detection Method Based on Improved YOLOv5

Yong He ^{1,2}, Mengqi Guo ^{1,2,*} , Yongchuan Zhang ^{2,*}, Jun Xia ³, Xuelai Geng ², Tao Zou ² and Rui Ding ²

¹ Key Laboratory of Urban Land Resources Monitoring and Simulation, Ministry of Natural Resources, Shenzhen 518000, China; 990020050408@cqjtu.edu.cn

² Smart City College, Chongqing Jiaotong University, Chongqing 402247, China; 622220900019@mails.cqjtu.edu.cn (X.G.); 622220900029@mails.cqjtu.edu.cn (T.Z.); 632201110126@mails.cqjtu.edu.cn (R.D.)

³ Chongqing Academy of Surveying and Mapping, Chongqing 401121, China; geobook@163.com

* Correspondence: 622220900035@mails.cqjtu.edu.cn (M.G.); zhangyc@cqjtu.edu.cn (Y.Z.); Tel.: +86-189-6518-6521 (M.G.); +86-176-0231-7586 (Y.Z.)

Abstract: Accurate traffic sign recognition is one of the core technologies of intelligent driving systems, which face multiple challenges such as insufficient light and shadow interference at night. In this paper, we improve the YOLOv5 model for small, fuzzy, and partially occluded traffic sign targets at night and propose a high-precision nighttime traffic sign recognition method, “NTS-YOLO”. The method firstly preprocessed the traffic sign dataset by adopting an unsupervised nighttime image enhancement method to improve the image quality under low-light conditions; secondly, it introduced the Convolutional Block Attention Module (CBAM) attentional mechanism, which focuses on the shape of the traffic sign by weighting the channel and spatial features inside the model and color to improve the perception under complex background and uneven illumination conditions; and finally, the Optimal Transport Assignment (OTA) loss function was adopted to optimize the accuracy of predicting the bounding box and thus improve the performance of the model by comparing the difference between two probability distributions, i.e., minimizing the difference. In order to evaluate the effectiveness of the method, 154 samples of typical traffic signs containing small targets and fuzzy and partially occluded traffic signs with different lighting conditions at nighttime were collected, and the data samples were subjected to the CBAM, OTA, and a combination of the two methods, respectively, and comparative experiments were conducted with the traditional YOLOv5 algorithm. The experimental results showed that “NTS-YOLO” achieved a significant performance improvement in nighttime traffic sign recognition, with a mean average accuracy improvement of 0.95% for the target detection of traffic signs and 0.17% for instance segmentation.

Keywords: YOLO; deep learning; traffic road signs; detection; recognition



Academic Editor: Pedro Couto

Received: 30 December 2024

Revised: 16 January 2025

Accepted: 31 January 2025

Published: 4 February 2025

Citation: He, Y.; Guo, M.; Zhang, Y.; Xia, J.; Geng, X.; Zou, T.; Ding, R. NTS-YOLO: A Nocturnal Traffic Sign Detection Method Based on Improved YOLOv5. *Appl. Sci.* **2025**, *15*, 1578. <https://doi.org/10.3390/app15031578>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

With the rapid progress of autonomous driving technology and the increasing maturity of smart connected vehicles, safe and efficient night driving is increasingly dependent on the support of artificial intelligence. However, complex and unstable lighting conditions at night have a huge impact on the visibility and recognition rate of traffic signs. Under the combined effect of streetlights, headlights, and other light sources, it becomes more difficult to recognize traffic signs with small, fuzzy, and occluded targets, especially for automated driving systems that are highly dependent on visual sensors. Therefore, for

automatic driving systems, how to accurately recognize various traffic signs in nighttime environments is undoubtedly a key issue that needs to be continuously studied and solved in this paper.

Currently, the automatic detection and recognition of traffic signs is a key application in the field of computer vision and image processing, which involves the automated detection and interpretation of traffic signs in road images and video streams. This technology is capable of recognizing various types of traffic signs, such as speed limits, prohibition signs, and directional signs, thus providing important real-time information and warnings for driving. As a core component of self-driving car research, this technology area has been extensively studied, but in traditional research, the detection of traffic signs mainly relies on color segmentation [1–6] and shape detection [7–10] methods, and the traditional methods for the recognition of traffic signs are mainly template matching [11–13] and manual feature extraction [14–18]. However, these traditional methods have limitations in detection and recognition accuracy, sensitivity to changes in illumination and the viewing angle, and the ability to handle complex scenes. In recent years, the rapid development of deep learning technology has revolutionized traffic sign detection and recognition, and significant research results and progress have been achieved in the application of traffic sign detection and recognition by virtue of its efficient capability in image recognition and processing.

In traffic sign detection, Girshick et al. [19] proposed a candidate region-based target detection algorithm, RCNN, which utilizes a deep convolutional neural network to achieve target detection and semantic segmentation but suffers from the problem of repetitive feature extraction and storage and has low computational efficiency; subsequently, He et al. [20] proposed a spatial pyramid pooling network (SPP-Net) to improve the processing speed and quality of CNN feature extraction by computing the entire input convolutional feature mapping of the image and extracting features from different candidate regions on the feature map, while using a spatial pyramid pooling layer to eliminate the fixed size constraints of the network, thus improving the processing speed and the quality of CNN feature extraction. However, the training of SPP-Net is multi-stage rather than an end-to-end approach; the proposal of Fast RCNN [21] achieves end-to-end detection trained on shared convolutional features, and its VGG16 network is nine times faster than RCNN and three times faster compared to SPP-Net. Faster RCNN achieves a higher mAP by introducing the RPN [22] and FPN, which further improves the detection accuracy for small targets. In order to obtain a detection bounding box with a more accurate location, Jiang et al. proposed the IoU-Net network [23], which uses the network to train the IoU branch and extracts the localization confidence of each bounding box, which improves the accuracy of localization. But the IoU-Net network does not have a strong correlation with the commonly used loss, and it cannot accurately differentiate the alignment of the two objects. For this reason, Rezaatofighi et al. [24] proposed to use the GIoU loss function as the loss of the bounding box regression branch, which can improve the accuracy by 2% to 14%. Aiming at the detection of multi-scale targets and solving the real-time problem of detection, Wang J. [25] proposed an improved feature pyramid model using adaptive attention modules (AAMs) and feature enhancement modules (FEMs) to reduce the loss of information and enhance the characterization of the feature pyramid, which improved the detection performance for multi-scale targets in the YOLOv5 network under the premise of ensuring real-time detection. For the detection of small traffic sign targets, Yongliang Zhang [26] et al. designed multi-scale feature extraction, cascade feature fusion, and attention mechanism modules based on the YOLOv4 algorithm to address the complexity of traffic scene images and variations in traffic sign sizes, thereby enhancing the algorithm's localization and classification capabilities. Lanmei Wang [27] et al. proposed the CDF

and CDF-s traffic sign detection models, which achieve accurate small target detection without sacrificing the detection accuracy of medium and large objects. Jianming Zhang [28] et al. proposed a new neck-based multilayer interactive residual feature fusion network (MIRFFN), which effectively combines the spatial information of low-level feature maps with the semantic information of high-level feature maps and refines features by fusing different layers of the feature maps. However, for nighttime traffic sign detection in the face of small, fuzzy, and overlapping parts of the target, the existing model may still have inaccurate recognition and localization, and further research is required on more effective feature extraction and target localization methods.

In terms of automatic recognition, the studies of Fang [29], Ciresan [30], and Qian [31] demonstrated the efficient recognition and robustness of CNN-based methods on the GT-SRB dataset. Despite the high accuracy of these three networks, the activation functions they used were computationally inefficient and required a large number of multiplications on the hardware. In the study by Aghdam et al. [32], in order to reduce the amount of computation, the ReLU activation function and optimized network structure were used to divide the intermediate convolutional pooling layer into two groups, which reduced the number of parameters and improved the traffic sign recognition accuracy and real-time performance so that the recognition rate reached 99.51%. Xie et al. [33] used a cascaded convolutional neural network for traffic sign recognition. The proposed recognition algorithm divides the TSR into two phases: in the first phase, the network is trained according to the class labels of the signs, and the second phase trains the network on the shape and text of each class of signs. Although the recognition accuracy of this class of algorithms is high, the required running time is long and does not meet the real-time requirements of practical systems. Accordingly, Yi Shi [34] proposed a nighttime target recognition method based on infrared thermal imaging and the YOLOv3 target recognition framework, but its target recognition framework is not ideal for target recognition with overlapping, occlusion, and other interfering factors. To address the above problems, Qu S et al. [35] introduced the coordinate attention (CA) mechanism in the backbone network, used prediction headers to extract fine-grained features, and improved the accuracy of bbox regression by improving the localization loss function CIoU using Alpha-IoU. Wang Q et al. [36] used a dynamic label assignment strategy, Simple Optimal Transmission Assignment (SimOTA), in the label assignment process and, for the target size problem, proposed a feature fusion network, H-PFANet, to improve the recognition rate of overlapping and occlusion phenomena. Liu H et al. [37] introduced an optimization algorithm called ETSR-YOLO, which firstly improved the PANet of YOLOv5s by generating an additional high-resolution feature layer to enhance the multi-scale feature fusion and improve the recognition of small-sized objects; secondly, two improved C3 modules were introduced to suppress the background noise interference and enhance the feature extraction of the network; and finally, in the postprocessing stage, a Wise-IoU (WIoU) function was introduced to improve the learning ability and robustness of the algorithm. Yan Hai et al. [38] proposed a dedicated deep learning model that enhances the recognition of blurred traffic signs by using multi-scale convolutional stacking in the input layer. They analyzed the recognition performance of indication signs, prohibition signs, speed limit signs, and warning signs under different levels of occlusion based on the Chinese Traffic Sign Database. However, existing detection methods face issues such as slow result acquisition and low accuracy. To address these issues, G. Song [39] proposed an improved network based on lightweight convolutional neural networks, which improved the training speed of the network and completed traffic sign recognition faster and more accurately. Wenju Li et al. [40] proposed a traffic sign recognition algorithm that combines a CNN and an extreme learning machine (ELM). This method utilizes ResNet50 to extract image features, uses a region proposal network

(RPN) to generate proposals, and then classifies them using ELM, followed by regression prediction through a fully connected layer. Lin Shan et al. [41] proposed a traffic sign detection method based on a lightweight multi-scale feature fusion network. Shuen Zhao et al. [42] proposed a lightweight convolutional neural network recognition method for multi-interference scenarios. This approach enhances images through Gamma correction and histogram equalization, merges MobileNet-V2 and DeepLab-V3+ for segmentation, and finally uses Lw-CNN for the adaptive recognition of traffic signs and markings. However, the above research mainly focused on daytime scenes and research is lacking on nighttime scenes.

In summary, there is a relative lack of research on nighttime traffic sign detection and recognition, mainly focusing on daytime conditions. However, the complex and unstable lighting conditions at night result in features that are relied upon for daytime vehicle detection to be ineffective in nighttime environments, especially in long-distance, fuzzy, and partially occluded situations where the traffic sign recognition accuracy is not high. Therefore, this study improved traffic sign recognition methods, especially their performance in nighttime environments. By optimizing and adjusting advanced algorithms such as YOLO, combined with the in-depth analysis and processing of local nighttime traffic data, the system's recognition accuracy and robustness under low-light conditions were improved, thus ensuring safe driving in nighttime environments and providing more reliable technical support for self-driving cars.

2. Methods

In this paper, we propose a nighttime traffic sign recognition method, “NTS-YOLO”, which consists of three main parts, as shown in Figure 1. First, this paper adopted the unsupervised nighttime image enhancement technique proposed by Yeying Jin et al. [43]. It integrates the decomposition network and the light effect suppression network in a single unified framework, which enhances the brightness and contrast of the image and effectively improves the quality of the image under low-light conditions. Second, this paper introduced the Convolutional Block Attention Module (CBAM) attention mechanism on the basis of the YOLOv5 network structure, which is able to adaptively weight the channel and spatial features inside the model: the channel attention (CA) generates the global maximum and average features for each channel using global maximum pooling and average pooling operations, and then generates the global maximum and average features, and then learns the attention weights for each channel through a shared MLP network, thus emphasizing channels that contribute to the task and suppressing irrelevant channels; spatial attention (SA) generates a spatial feature map through maximum and average pooling, and then learns the attention weights for the spatial dimension through a convolutional layer, allowing the model to focus more on the pixel regions in the image that affect the classification decision. Finally, in this paper, the Optimal Transport Assignment (OTA) loss function was used to optimize the performance of the model in the target detection task, which first calculates the cost matrix between the two sets, which contains the costs between all the predicted and real frames, usually based on the distances between them in terms of the IoU values. Then, the optimal transmission algorithm was used to find the best matching solution that minimized the total cost. In this way, the accuracy of the predicted bounding boxes could be effectively optimized so that the model could predict the location of the target and the bounding boxes more accurately, thus improving the robustness and stability of the model in the target detection task.

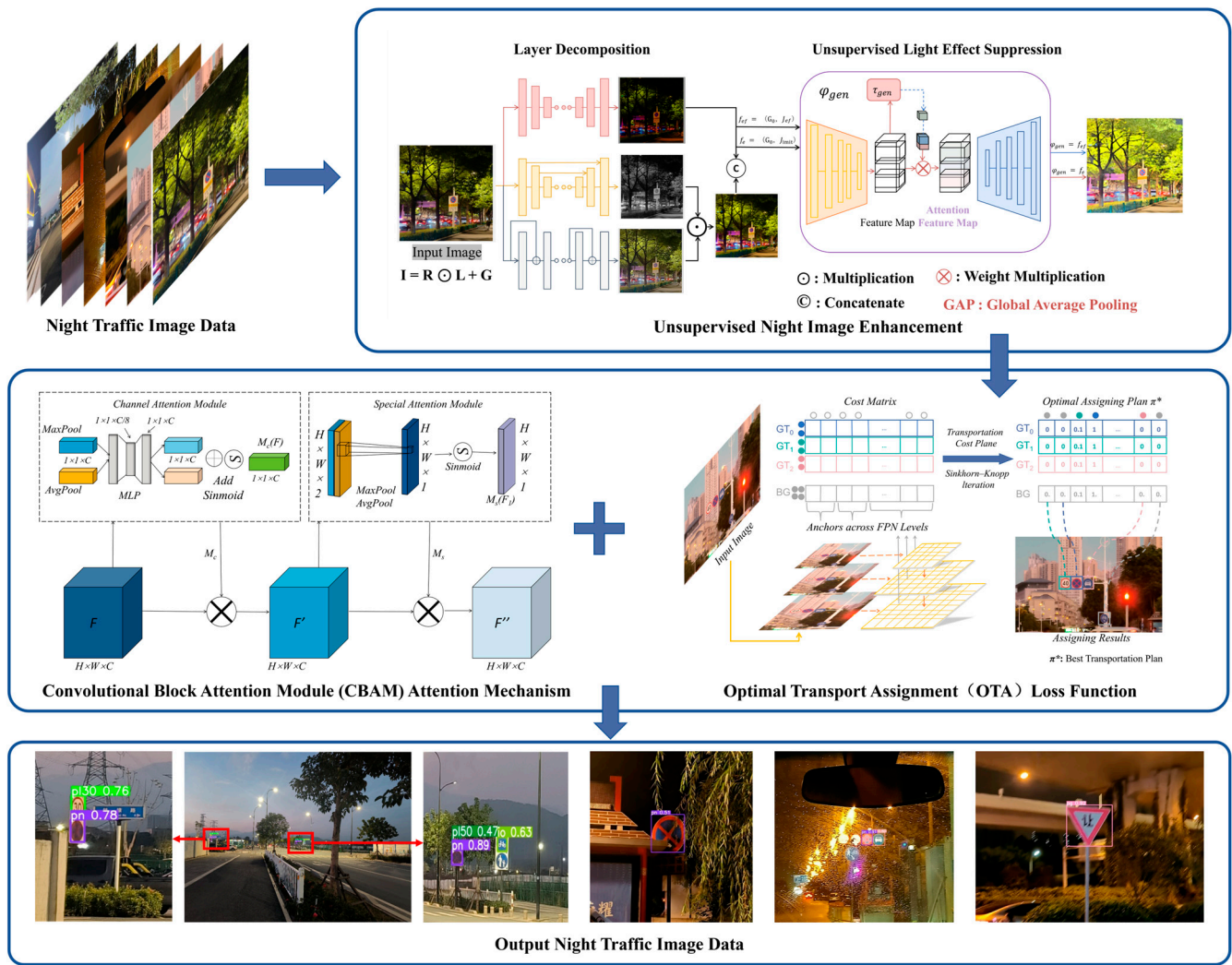


Figure 1. NTS-YOLO process diagram. The NTS-YOLO process diagram integrates unsupervised image preprocessing, the Convolutional Block Attention Module (CBAM), and Online Template Adaptation (OTA) to enhance the object detection performance beyond that of the standard YOLO model.

2.1. Image Preprocessing

Due to the influence of the light intensity strength, direction angle, and other factors at night, the visibility of traffic signs was significantly reduced, resulting in image feature details being difficult to identify, the signal-to-noise ratio being reduced, and the image as a whole being darker. Therefore, it was necessary to carry out the data enhancement preprocessing of the image to eliminate the factors that affected the image quality so that the captured image could be more efficiently extracted from the effective information. At present, image enhancement algorithms are commonly used to adjust the brightness, contrast, saturation, hue, etc., of an image to increase its clarity, reduce noise, etc. Common methods include histogram equalization, the Gamma transform, the Laplace transform, the Retinex algorithm, and image enhancement based on deep learning. Through these preprocessing steps, the recognition accuracy and reliability of traffic signs in the night environment can be effectively improved.

Existing nighttime visibility enhancement methods mainly focus on increasing the intensity of low-light regions; therefore, when these methods are applied to nighttime images containing light effects, they inevitably amplify the light effects and even further impair the visibility of the images. In this paper, we adopted the unsupervised nighttime image enhancement method proposed by Yeying Jin et al. [43], which integrates a decom-

position network and a light effect suppression network in a single unified framework with the goal of suppressing the light effect while increasing the intensity of the dark regions. Specifically, the image decomposition network divides the input image into a base layer and a detail layer. The base layer is designed to process the brightness and global information of the image, while the detail layer captures fine textures and local features. Building on this decomposition, the light effect suppression network leverages an attention mechanism to dynamically adjust high-light regions, effectively mitigating the adverse effects of glare and overexposure on detail extraction. This integrated approach not only enhances the visual quality of nighttime images but also significantly improves the visibility and discriminability of target regions. The structure diagram is shown in Figure 2.

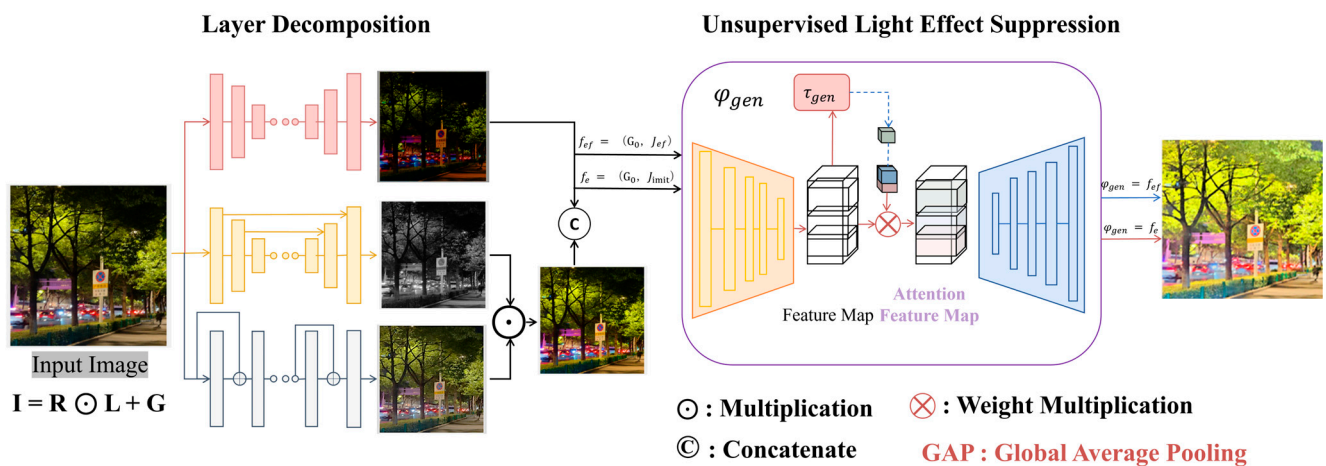


Figure 2. The unsupervised nighttime image enhancement method structure diagram. First, the input image is decomposed into three components: region information, R, light effects, L, and the general scene, G. Then, a generative model and attention mechanism are used to process the light effects and scene features separately. Finally, through global average pooling (GAP) and the weighted multiplication of feature maps, an enhanced image is generated that suppresses the light effects while preserving scene details.

To better adapt the enhancement method to a specific nighttime traffic sign dataset, this study conducted an optimization of its key hyperparameters to achieve a balance between brightness enhancement, contrast adjustment, and light effect suppression, as detailed in Table 1.

Table 1. Hyperparameter optimization and settings.

| Hyperparameter | Initial | Optimized | Optimization Objective |
|---------------------------------|------------|------------|---|
| Brightness Adjustment Range | [0.5, 1.5] | [0.8, 1.2] | Enhance brightness in dark areas while suppressing distortion in bright regions. |
| Contrast Adjustment Range | [0.7, 1.3] | [0.9, 1.1] | Strengthen edge features of target while avoiding excessively high background contrast. |
| Light Effect Suppression Weight | 0.5 | 0.6 | Balance the suppression of strong light effects with the enhancement of dark areas. |

The experimental results demonstrated that the optimized hyperparameters led to significant improvements in model performance. As shown in Figure 3, the adjustments to the brightness and contrast ranges enhanced the target regions while reducing the interference of glare and noise on the detection results. Furthermore, the optimization of the light effect suppression weight not only effectively suppressed interference from high-light regions but also improved the separability of features in dark areas.



Figure 3. Unsupervised preprocessing comparison chart. The collected nighttime traffic sign images were processed with data augmentation to improve the image quality.

2.2. Adding CBAM Attention Mechanism

Nighttime traffic sign recognition faces many challenges in practice, especially due to problems such as low light and ambient noise, which result in blurred images and difficulty in recognizing the edges and details of traffic signs. To remedy this problem, this study introduced the Convolutional Block Attention Module (CBAM) attention mechanism to enhance the model's ability to recognize and represent key features, which in turn improves the recognition accuracy and stability under low-light conditions. The CBAM focuses on important feature channels and key spatial regions in the image through the dimensions of channel attention (CA) and spatial attention (SA), as illustrated in Figure 4. The channel attention (CA) and spatial attention (SA) of the CBAM are shown in Equations (1) and (2).

$$CA(F) = \sigma(MLP(GAP(F)) + MLP(GMP(F))) \quad (1)$$

$$SA(F) = \sigma(f^{7 \times 7}([GAP(F); GMP(F)])) \quad (2)$$

where F in $CA(F)$ is the input feature map, σ is the sigmoid activation function, and MLP denotes the multilayer perceptron; GAP and GMP stand for global average pooling and global maximum pooling, respectively, which are used to compute the global statistical properties of each channel. The $f^{7 \times 7}$ in $SA(F)$ denotes the convolution operation with a 7×7 convolution kernel, and $[\cdot]$ denotes the stacking of feature maps.

Before the introduction of the CBAM, the output of the SPPF convolutional layer was the feature map F that was processed from the original input image after convolution, batch normalization, and the application of the ReLU activation function, as shown in Equation (3).

$$F = ReLU(BN(Conv(Input))) \quad (3)$$

After adding the CBAM, the output feature map F of the SPPF convolutional layer is further processed by the CBAM, and the network structure is shown in Figure 5. The CBAM first computes the attention weights of each channel through the CA part and then computes the spatial attention weights through the SA part. Finally, these two weights are multiplied with the original feature map F to obtain the final output feature map, as shown in Equation (4).

$$F' = F \times CA(F) \times SA(F) \quad (4)$$

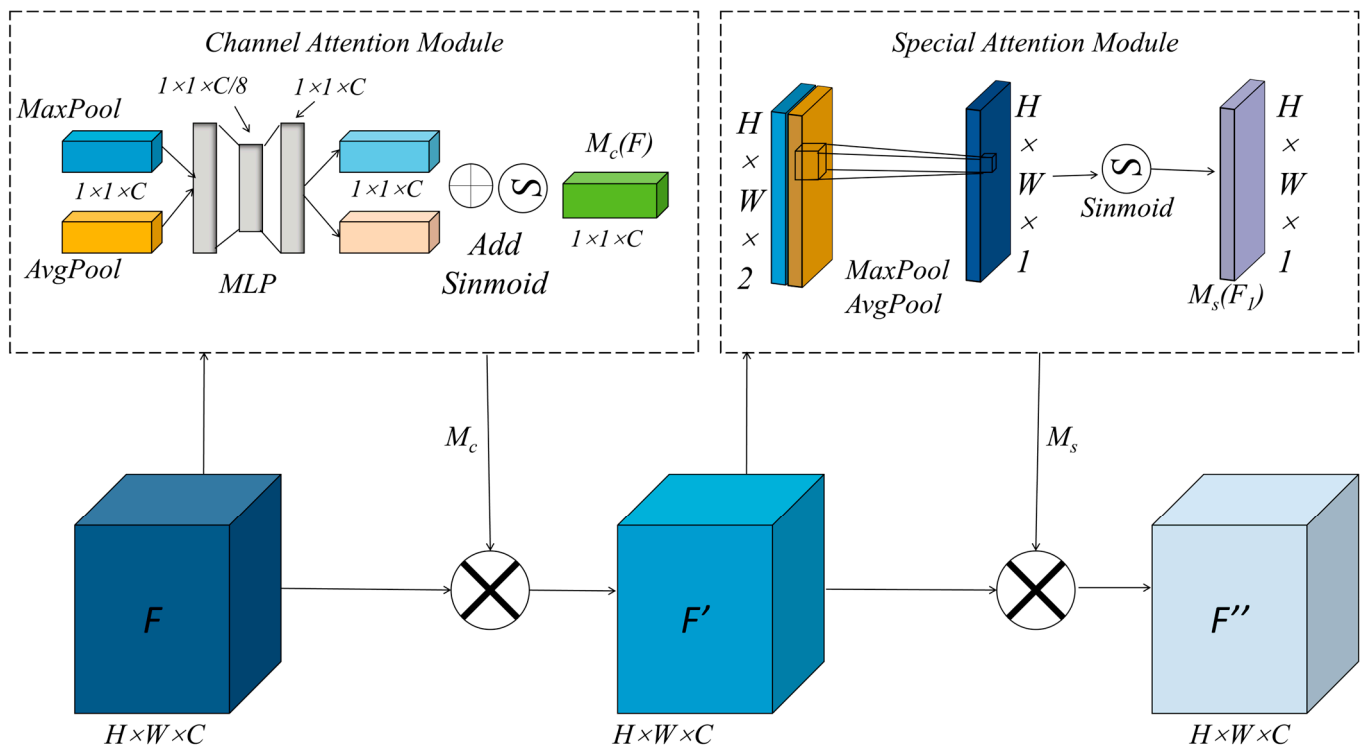


Figure 4. CBAM attention mechanism structure diagram. The CBAM processes the input feature maps through two main modules: the channel attention module and the spatial attention module. The channel attention module adaptively adjusts the weights of different channels to highlight important features, while the spatial attention module focuses on key regions in the spatial domain, thereby enhancing the model’s ability to recognize critical areas.

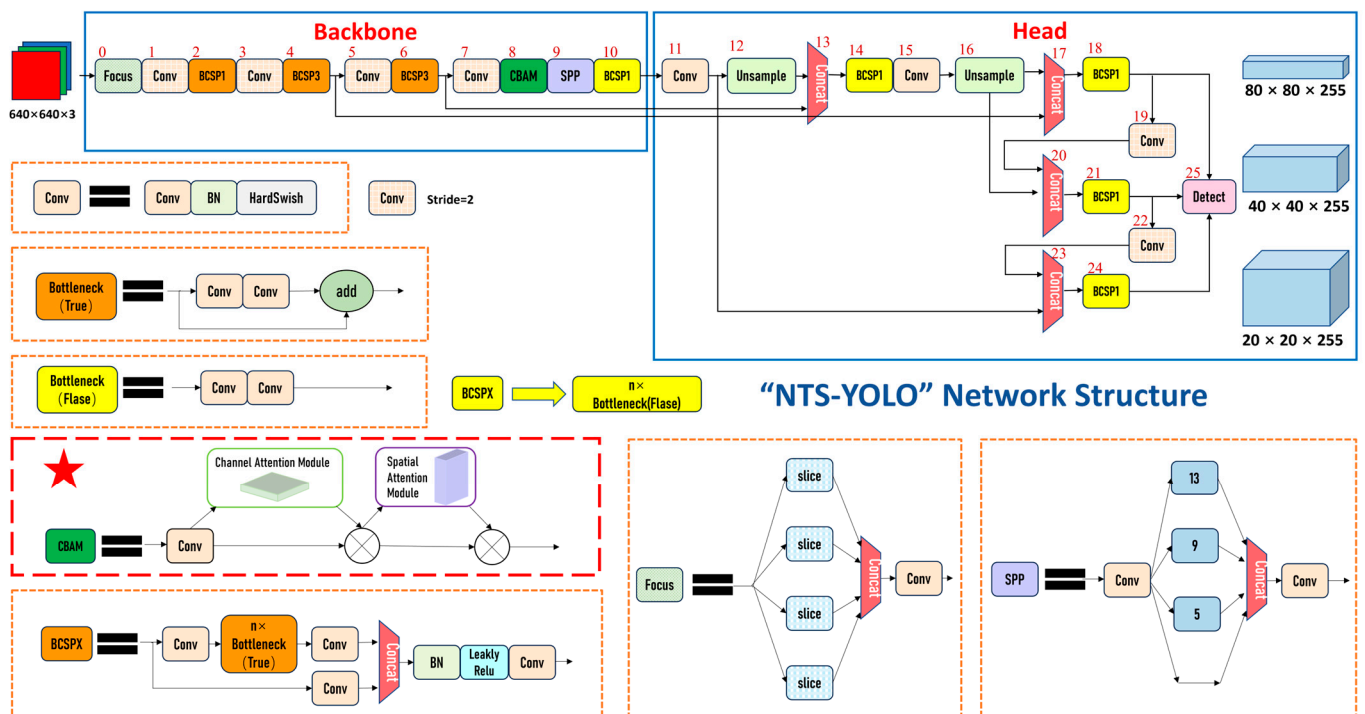


Figure 5. NTS-YOLO network structure. The NTS-YOLO network architecture integrates the CBAM attention mechanism, enabling the model to effectively focus on the region with the highest amount of information in the input image, significantly improving the real-time system’s object detection performance.

The introduction of the CBAM allows the model to focus more effectively on key channel features and to identify and emphasize important spatial regions in the image. In complex environments at night, identifying traffic signs, small targets, and possible occlusion situations, traditional feature extraction methods are often affected by uneven lighting and background noise. With the CBAM attention mechanism, the model is able to focus not only on important channel features but also on key spatial regions in the image when processing the feature map. This integration strategy improves the model's sensitivity to details and enhances its robustness under low-light conditions, which brings a significant performance improvement to the nighttime traffic sign recognition task and provides an effective solution to address the challenges of uneven lighting and background noise.

2.3. Optimization Loss Function

In standard loss functions (e.g., cross-entropy loss and mean square error loss), the optimization process of the model may be affected by incorrect label assignments, especially when the target sizes are small or overlapping. In order to improve the accuracy and efficiency of YOLOv5 in nighttime traffic sign recognition, this paper introduced an innovative loss function, the Optimal Transport Assignment (OTA) loss, as illustrated in Figure 6. Specifically, the OTA loss function is defined by the difference between the predicted bounding box and the real bounding box set as the optimal transport cost, which is computed by the following equation:

$$L_{OTA} = \sum_{i,j} T_{i,j} \cdot C_{i,j} \quad (5)$$

where $T_{i,j}$ is the amount of transmission between the predicted bounding box i and the real bounding box j , and $C_{i,j}$ is the corresponding cost, usually measured by the distance or difference between the two.

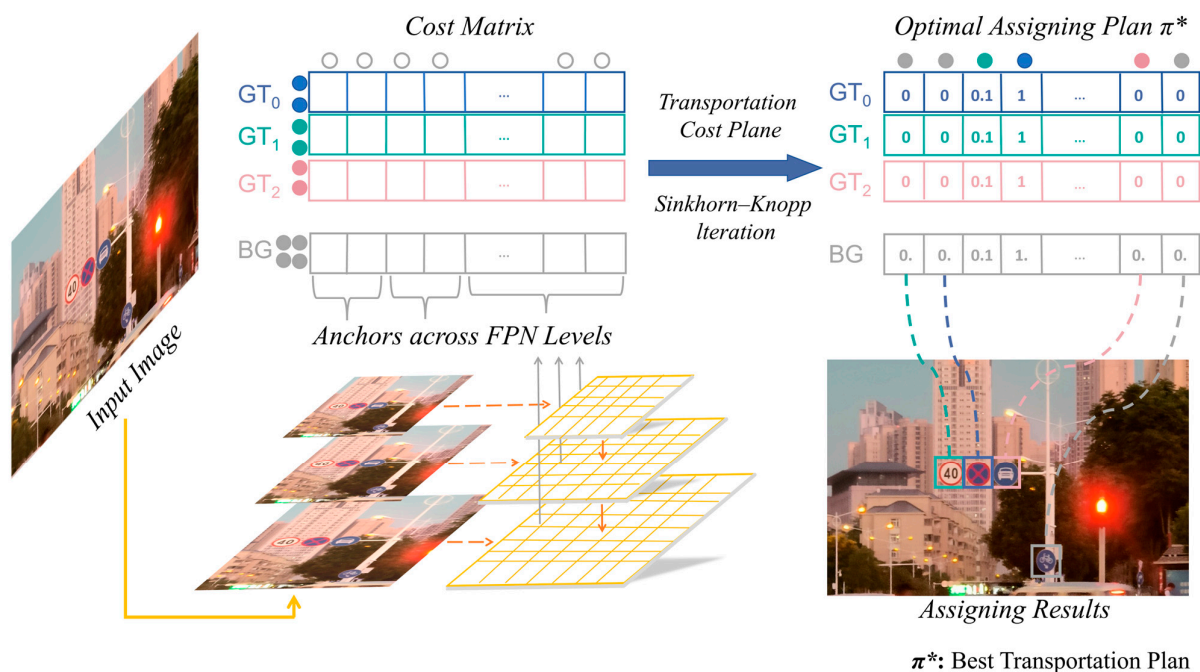


Figure 6. OTA loss function structure diagram. The figure shows the process of object detection across multiple ground truth layers (GT0, GT1, GT2) and the background (BG) representation. The input image is first divided into grids, with each grid corresponding to different ground truth values. These grids contain values that represent the presence of different objects in the scene, as indicated by the colored boxes for each ground truth and background category. The loss function operates by matching the ground truth layers with the background and object regions, optimizing the predictions across the different layers to suppress irrelevant areas and enhance the detection performance.

In this paper, we improved the accuracy of object detection by integrating the optimal transmission analysis (OTA) loss function into the YOLOv5 model. The OTA loss function was specifically designed to quantify the difference between the predicted bounding box and the real bounding box, and its computation is based on the amount of transmissions between the predicted bounding box and the real bounding box, T_{ij} , and the cost C_{ij} .

Furthermore, we introduced the OTA loss as a part of the training of the model, which not only took the alignment accuracy between predicted and real boxes, but also combined it with the traditional category loss, confidence loss, and bounding box loss in order to form a comprehensive loss function for the model. If the traditional YOLOv5 loss consists of the category loss (L_{class}), the confidence loss (L_{conf}), and the bounding box loss (L_{bbox}), then the improved total loss function (L_{total}) can be expressed as

$$L_{total} = \lambda_{class}L_{class} + \lambda_{conf}L_{conf} + \lambda_{bbox}L_{bbox} + \lambda_{OTA}L_{OTA} \quad (6)$$

Here, λ_{class} , λ_{conf} , λ_{bbox} , and λ_{OTA} are the weighting factors used to balance the different loss components.

2.3.1. Weighted Coefficient Optimization Method

To achieve a balance between the classification accuracy, bounding box confidence, target localization precision, and computational complexity, this study employed a grid search-based optimization method to determine the weighting coefficients. First, the weight of a specific loss component was adjusted individually while keeping the weights of the other components fixed, and its independent impact on the model performance was observed. Next, based on single-factor optimization, the optimal weight values for each loss component were combined, and their performance under the total loss function was verified. Specifically, the classification loss (λ_{class}) weight was varied within the range of [0.1, 0.6] with a step size of 0.1, the confidence loss weight (λ_{conf}) within [0.1, 0.5] with a step size of 0.1, the bounding box loss weight (λ_{bbox}) within [0.2, 0.4] with a step size of 0.1, and the OTA loss (λ_{OTA}) weight within [0.1, 0.3] with a step size of 0.1.

2.3.2. Model Improvement Basis and Optimal Configuration

The weight optimization experiments in this study were based on the YOLOv5 model enhanced with the CBAM attention mechanism. The specific experimental results are shown in Table 2.

Table 2. CBAM weighted coefficient configuration table.

| Weighted Coefficient Configuration | mAP-0.5 (%) | mAP-0.5:0.95 (%) | Speed (it/s) |
|---|-------------|------------------|--------------|
| $\alpha = 0.4, \beta = 0.3, \gamma = 0.3, \delta = 0.2$ | 93.27 | 73.60 | 2.57 |
| $\alpha = 0.3, \beta = 0.3, \gamma = 0.3, \delta = 0.2$ | 93.10 | 72.88 | 2.58 |
| $\alpha = 0.4, \beta = 0.2, \gamma = 0.3, \delta = 0.2$ | 93.12 | 72.84 | 2.58 |

The experimental results demonstrated that when the classification loss weight (λ_{class}) was set to 0.4, the model's ability to recognize small target classes was significantly enhanced, especially improving the classification accuracy in complex nighttime scenarios. Setting the confidence loss weight (λ_{conf}) to 0.3 effectively increased the confidence of detection boxes, optimizing both precision and recall. Similarly, setting the bounding box loss weight (λ_{bbox}) to 0.3 improved the localization accuracy of bounding boxes. When the OTA loss weight (λ_{OTA}) was set to 0.2, the model achieved optimal performance in small target detection and robustness in complex scenes, while avoiding the efficiency loss caused by excessive computational complexity.

As a result, the optimal weight configuration was $\alpha = 0.4$, $\beta = 0.2$, $\gamma = 0.3$, $\delta = 0.2$. With this final configuration, the mAP-0.5 metric on the validation set increased by 0.95%, and the mAP-0.5:0.95 metric improved by 0.17%, while the reduction in the processing speed was controlled within 8.4%. This ensured that the model met real-time requirements while maintaining high accuracy.

3. Results

3.1. Experimental Environment and Evaluation Indicators

In this study, the experimental environment consists of a high-performance computer configured with an Intel Core i7 processor, 32 GB of RAM, and an NVIDIA GeForce RTX 4060 graphics card. The GPU's memory capacity and computational power ensured the efficient execution of complex models with high-resolution inputs, providing excellent stability and efficiency for deep learning model training and inference tasks. Additionally, in order to efficiently conduct deep learning experiments, PyTorch 2.0.1 was selected as the main deep learning framework in this paper, and CUDA technology was utilized for accelerating the model training and inference, which ensured computational efficiency and a good data processing capability during the experimental process.

For evaluation metrics, this study focused on three aspects: the model accuracy, computational efficiency, and resource utilization efficiency. The model accuracy was evaluated by standard metrics such as the precision, recall, and mAP value to ensure that the model had good performance on different types of datasets. Meanwhile, the training and inference times of the model were recorded in detail to assess its computational efficiency. In addition, the CPU and GPU usage, as well as the memory occupancy, were also taken into account to comprehensively evaluate the resource utilization efficiency of the model in real-world application scenarios.

3.2. Construction and Processing of Datasets

In this research, 599 nighttime images from the CCTSDB2021 dataset were referenced [44], wherein 80% of the images (479) were utilized as the training set and 20% (120) served as the validation set. Given the relatively fewer types of nighttime traffic signs, to augment the data diversity, 9170 daytime road scene images from the TT100K dataset [45] were also referenced, divided into a training set (7208 images) and a validation set (1962 images) at an 8:2 ratio. Moreover, 154 self-collected nighttime road scene images were employed as an independent test set, which covered scenarios such as small targets, occluded signs, and blurred scenes. Among the 154 images, there were 60 prohibition signs, 50 warning signs, and 44 instruction signs. Small target signs and occluded signs account for 19.5% and 20.8% of the total targets, respectively. Small targets were primarily concentrated in long-distance scenarios or under low-light conditions, while occlusions were mainly caused by trees, vehicles, or light reflections. Some sample data are shown in Figure 7. By amalgamating nighttime and daytime data from different datasets, the aim was to enhance the model's traffic sign recognition accuracy and generalization capability under various lighting conditions. Such a dataset construction approach provided a more comprehensive and representative data foundation for this study.

This dataset included three categories of prohibited signs, warning signs, and directional signs, and the number of labeled classes reached 221, which basically covered a variety of common traffic signs in China, as shown in Figure 8. In order to increase the diversity and robustness of the dataset, this paper carried out data enhancement processing, including image rotation, scaling, panning, and brightness adjustment operations to simulate different observation angles and lighting conditions; the test set was specially designed to include traffic signs in special environments, such as small targets at night, occlusion,

blurring, etc., which was aimed at evaluating the performance of the model under complex conditions and verifying its recognition accuracy and robustness under insufficient lighting, with blurred or obscured targets, and so on. The purpose of the test set was to evaluate the model's performance under complex conditions and verify its recognition accuracy and robustness with low-light, blurred, or obscured targets, which helped to evaluate the model's performance more comprehensively.

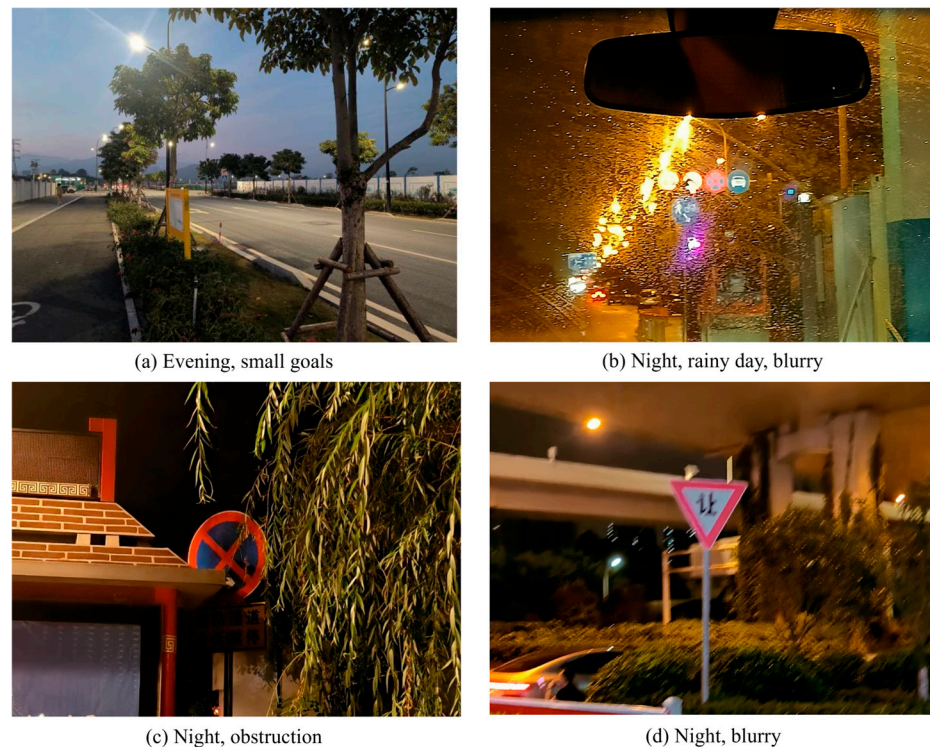


Figure 7. Dataset example. Partial display of night traffic sign data, including small targets, partial occlusion, blurriness, and rainy days.

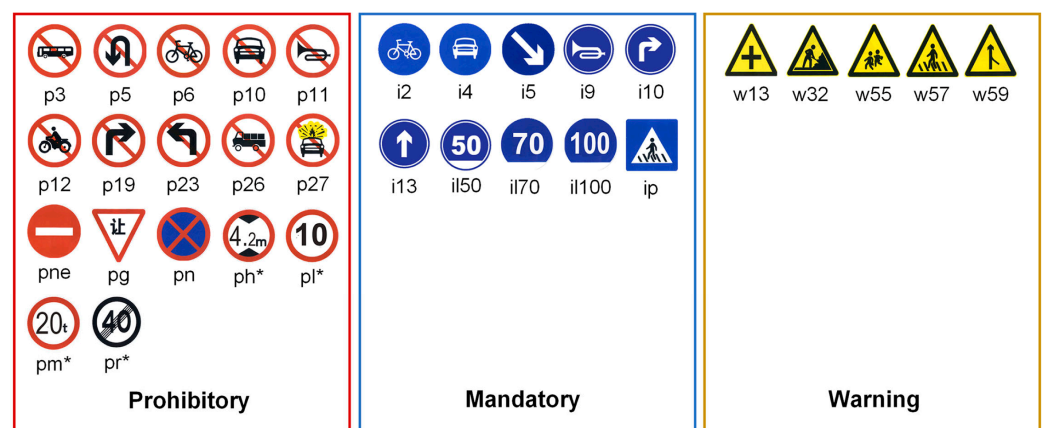


Figure 8. Dataset classification. There were three categories in total, namely warning, prohibition, and indication, and the traffic sign dataset was divided according to these three categories.

3.3. Model Training Configuration

To optimize the performance of the improved model in nighttime traffic sign detection tasks, this study conducted experimental validation on the configurations of the input resolution and batch size. The input resolution and batch size directly affect the model's detection accuracy, training stability, and hardware resource utilization. To determine the

optimal configuration, combinations of input resolutions (640×640 , 1024×1024 , and 1280×1280) and batch sizes (4, 8, and 16) were tested. The results are shown in Table 3.

Table 3. Model training configuration comparison table.

| Input Resolution | Batch Size | GPU Memory Usage (GB) | mAP-0.5 (%) | mAP-0.5:0.95 (%) | Convergence Stability | Speed (Epochs/s) |
|--------------------|------------|-----------------------|-------------|------------------|-----------------------|------------------|
| 640×640 | 8 | 1.9 | 90.34 | 67.12 | Stable | 6.7 |
| 1024×1024 | 8 | 4.2 | 91.88 | 70.04 | Stable | 8.3 |
| 1280×1280 | 8 | 4.7 | 92.84 | 72.44 | Relatively Stable | 9.3 |
| 1280×1280 | 8 | 6.2 | 93.27 | 73.6 | Stable | 11.5 |
| 1280×1280 | 8 | — | — | — | — | — |

The experimental results indicated that increasing the input resolution significantly improved the model's detection accuracy, particularly in small object detection and complex background scenarios. Specifically, at a resolution of 1280×1280 , the mAP-0.5 on the validation set reached 93.27%, representing an improvement of 1.39% and 2.93% compared to resolutions of 1024×1024 and 640×640 , respectively. This demonstrates that higher resolutions can capture target details more clearly, contributing to better detection performance for small objects.

Secondly, the batch size had a significant impact on memory usage, the training stability, and the convergence speed. At a resolution of 1280×1280 , a batch size of four resulted in a memory usage of approximately 4.7 GB and an mAP-0.5 of 92.84%. However, the insufficient sample size for gradient updates caused slight performance degradation due to instability. When the batch size was set to eight, the memory usage was approximately 6.2 GB, and the mAP-0.5 on the validation set reached 93.27%, with stable gradient updates and efficient training. However, when the batch size was increased to 16, the memory requirement exceeded the limits of the experimental device, making training impossible. Therefore, at a resolution of 1280×1280 , a batch size of eight was the optimal configuration in terms of performance, efficiency, and hardware resource utilization.

In summary, the configuration of an input resolution of 1280×1280 and a batch size of eight could effectively enhance the model performance while maintaining the training efficiency and stability within the limits of hardware resources.

3.4. Model Training and Optimization

In the preparatory stage of model training, we spent a lot of time preprocessing the training dataset in detail. Firstly, the dataset was partitioned into a validation set and a training set according to the ratio of 2:8. For the images in the training set, we performed pixel-level scaling and normalization so that the image size and pixel value distribution could meet the input requirements of the model. In order to increase the generalization ability and noise resistance of the model, we also performed a series of data enhancement operations on the training images, including random cropping, flipping, and color contrast and brightness adjustment.

The CBAM was embedded into the backbone network of YOLOv5, specifically inserted at the output position between the ninth layer of the backbone network and the Spatial Pyramid Pooling Fast (SPPF) module. The CBAM collaboratively optimizes the feature extraction process through channel attention and spatial attention. The channel attention module generates weights for each channel via global pooling operations, enhancing the focus on critical channels. The spatial attention module generates spatial attention weights through per-channel pooling, enabling the model to concentrate more effectively on target regions.

In the detection head, this study replaced the original fixed IoU threshold assignment strategy with the OTA loss function. OTA dynamically optimizes the positive and negative sample assignment process by first constructing a cost matrix of matches between predicted boxes and ground truth boxes, taking into account the IoU loss, classification loss, and the distance between the centers of bounding boxes. Then, the optimal transport algorithm is used to solve the matching problem, ensuring a more reasonable allocation of positive and negative samples, thereby improving the accuracy of bounding box predictions in object detection.

After integrating the OTA loss with other loss components, this study fine-tuned the weight proportions to ensure a balance among the loss components, further enhancing the model's performance.

To ensure the stability and efficiency of the training process, this paper adopted the strategy of periodic validation. By evaluating the model on an independent validation set, we were able to monitor the loss changes during the training and validation process in real time and adjust the training parameters, such as the learning rate and weight decay, accordingly. In addition, we calculated key evaluation metrics such as the precision rate, recall rate, and F1 value to comprehensively assess the performance of the model. Regular evaluations on independent test sets helped this paper verify the model's generalization ability on unknown data and ensure its reliability in real-world applications.

3.5. Ablation Study

This study evaluated the performance of the improved YOLOv5 model on the nighttime traffic sign recognition task. All experiments were conducted under uniform settings: the number of iterations was set to 120, the batch size was eight, and the input image resolution was 1280×1280 .

3.5.1. Performance Analysis of Ablation Study

In order to verify the effectiveness of the improved nighttime traffic sign recognition algorithm proposed in this paper, ablation experiments were conducted. Firstly, we compared the addition of the CBAM attention mechanism and the OTA loss function alone, respectively, and then applied them both to the YOLOv5 model and observed the performance in various cases. The results are shown in Table 4. In this paper, mAP-0.5 was chosen as the evaluation index, which could comprehensively assess the performance of the model in the target detection task.

Table 4. Comparison of experimental results of different methods.

| Model | CBAM | OTA | Accuracy (%) | Recall (%) | mAP-0.5 (%) | mAP-0.5:0.95 (%) | Speed (s) | GPU |
|---------------|------|-----|--------------|------------|-------------|------------------|-----------|-----|
| YOLOv5 | - | - | 85.98 | 88.71 | 92.39 | 72.40 | 2.38 | 5.2 |
| YOLOv5 + CBAM | ✓ | - | 89.34 | 87.31 | 92.44 | 72.59 | 2.58 | 5.5 |
| YOLOv5 + OTA | - | ✓ | 87.45 | 87.31 | 92.43 | 72.69 | 2.58 | 5.4 |
| NTS-YOLO | ✓ | ✓ | 90.59 | 88.81 | 93.27 | 73.60 | 2.57 | 5.7 |

In the experiments, we first used the original YOLOv5 model for the nighttime traffic sign recognition task and obtained a benchmark performance with an mAP-0.5 metric of 92.39%. Next, we added the CBAM attention mechanism and OTA loss function to the YOLOv5 model and conducted the experiments separately. The results showed that the addition of the CBAM attention mechanism slightly improved the model's mAP-0.5 score to 92.44%, an increase of 0.056%. This indicates that the CBAM mechanism helped the

model to better focus on the key regions of the image, such as the edges and shapes of the signs, which improved the performance of the model.

On the other hand, when the OTA loss function was introduced, the mAP-0.5 metric of the model was further improved to 92.43%, an increase of 0.052%. This indicates that the optimized loss function could effectively optimize the training process of the model and improve the accuracy of target detection. The OTA loss function achieved a better match by minimizing the cost between the predicted bounding box and the true bounding box, allowing the model to more accurately predict the target location and the bounding box, thus improving the robustness and stability of the model.

And when the CBAM and OTA were simultaneously applied to the YOLOv5 model, it could be observed that the mAP-0.5 score was improved to 93.27%, which was an increase of 0.88% compared to the original model.

However, it is worth noting that the introduction of the CBAM and OTA increased the computational complexity of the model, resulting in a reduction in the processing speed from 2.38 it/s to 2.58 it/s, approximately a 7.98% decrease. This speed reduction is mainly attributed to the attention computation in the CBAM and the dynamic matching optimization in the OTA loss. While these modules improve the model's adaptability to small targets and complex scenarios, they also add an extra computational burden and memory usage. Specifically, under the 1280×1280 resolution setting, the NTS-YOLO model occupied approximately 5.7 GB of GPU memory, which was 0.5 GB more than the original YOLOv5. This increase in memory usage was primarily due to the attention computation in CBAM and the matching optimization in the OTA loss. In terms of GPU utilization, the NTS-YOLO model maintained a utilization rate of approximately 72%, compared to 65% for YOLOv5, demonstrating higher resource usage. This indicates that the improved model makes more efficient use of hardware computing resources while achieving high-performance operation in the current hardware environment.

Nevertheless, this speed reduction remains within an acceptable range, especially considering the significant improvement in accuracy. Under the 1280×720 resolution setting, "NTS-YOLO" achieved a processing speed of approximately 11.5 frames per second, meeting the requirements of most real-time application scenarios, such as intelligent traffic monitoring and autonomous driving systems. In these scenarios, systems often need to accurately recognize nighttime traffic signs within a limited time to support decision-making and ensure safety. This demonstrates that "NTS-YOLO" strikes a balance between performance and efficiency, proving its practicality and effectiveness in nighttime traffic sign detection tasks.

3.5.2. Comparative Experiments of Different Models

To further verify the advantages of the NTS-YOLO model, it was compared with other mainstream object detection algorithms on the test set of the CCTSDB2021 nighttime dataset. The results are shown in Table 5.

As shown in the results in Table 5, NTS-YOLO outperformed other mainstream models in key metrics such as the precision (90.59%) and mAP-0.5:0.95 (73.60%) under complex nighttime scenarios. Compared to ETSR-YOLO, NTS-YOLO achieved a 0.55% improvement in the mAP-0.5 and a 1.39% improvement in the mAP-0.5:0.95. Although it was slightly slower than YOLOv7 and YOLOX in terms of the detection speed, its performance of 2.57 s per frame still meets the real-time requirements of practical applications, while achieving significant advantages in the detection accuracy.

Table 5. Performance comparison of mainstream object detection models.

| Model | Accuracy (%) | Recall (%) | mAP-0.5 (%) | mAP-0.5:0.95 (%) | Speed (it/s) | GPU |
|--------------|--------------|------------|-------------|------------------|--------------|-----|
| Faster R-CNN | 81.23 | 83.45 | 87.12 | 65.30 | 3.12 | 6.8 |
| YOLOv3 | 83.15 | 85.21 | 89.50 | 68.42 | 2.85 | 5.8 |
| YOLOv6 | 86.50 | 87.90 | 91.40 | 70.60 | 2.65 | 5.6 |
| YOLOv7 | 88.20 | 88.50 | 92.10 | 71.85 | 2.50 | 5.4 |
| YOLOX | 87.80 | 88.30 | 92.05 | 71.40 | 2.47 | 5.3 |
| ETSR-YOLO | 87.85 | 88.01 | 92.72 | 72.21 | 2.45 | 5.7 |
| YOLOv5 | 85.98 | 88.71 | 92.39 | 72.40 | 2.38 | 5.2 |
| NTS-YOLO | 90.59 | 88.81 | 93.27 | 73.60 | 2.57 | 5.7 |

3.5.3. Comparative Analysis Under Different Nighttime Scenarios

In order to comprehensively analyze the performance of the method proposed in this paper, four different nighttime scenarios were selected, as shown in Figure 9.

As represented in Figure 9a, the recognition accuracy of the YOLOv5 base model was low in the evening long-distance small target scenarios, and only speed limit and prohibition signs could be recognized, with a 58% recognition rate for a speed limit of 30 and an average recognition rate of 83% for prohibition signs. With the introduction of OTA and the CBAM alone, the recognition rate was improved, and an additional indicator sign was recognized; specifically, the recognition rate of a speed limit of 30 was improved to 74% and 31%, the average recognition rate of prohibited signs was improved to 86% and 75%, and the recognition rate of bike lanes was improved to 59% and 63%, respectively. With the simultaneous introduction of the OTA and CBAM mechanisms, the model was able to recognize all signs, with a recognition rate of 76% for a speed limit of 30, 89% for prohibited signs, and 63% for bike lanes. In addition, the model was also able to successfully recognize a traffic sign (speed limit of 50) where the small target was partially obscured by a tree, with a recognition rate of 47%.

As represented in Figure 9b, the original YOLOv5 model, as well as the introduction of the CBAM alone, failed to detect the signs in the nighttime occluded scenario. The possibility of wrong detection still exists with the introduction of the OTA loss function alone. After the introduction of the CBAM and OTA, “NTS-YOLO” could successfully recognize the prohibited signs with a recognition rate of 51%.

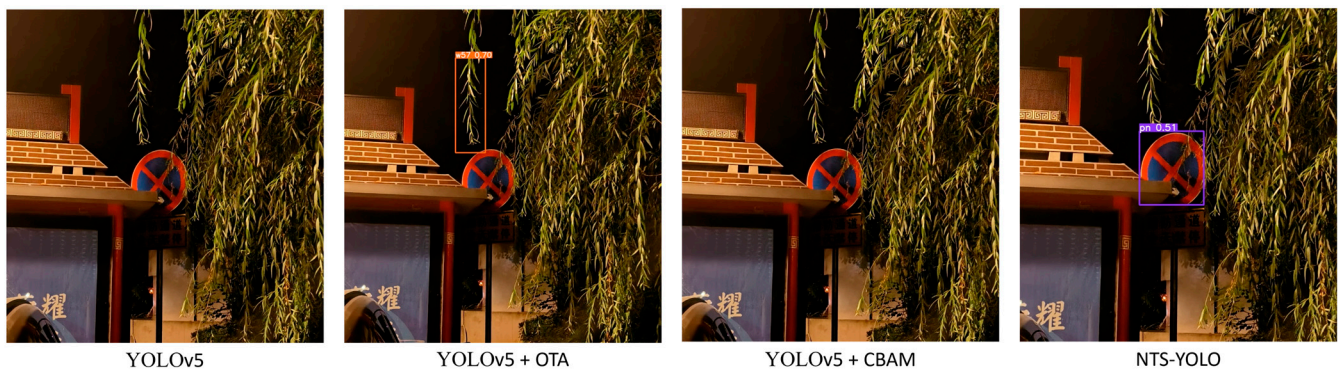
As shown in Figure 9c, the recognition rates of YOLOv5, YOLOv5 + OTA, and YOLOv5 + CBAM were 88%, 89%, and 89%, respectively, for the warning sign “let” in the nighttime blurred scene. After the introduction of OTA and the CBAM, the recognition rate of “NTS-YOLO” reached 92%, which was an improvement of four to five percentage points.

As represented in Figure 9d, in the nighttime rainy blur scenario, YOLOv5 and YOLOv5 + CBAM did not detect the sign, and YOLOv5 + OTA recognized the prohibited sign with a recognition rate of 51%. With the introduction of OTA and the CBAM, “NTS-YOLO” was able to successfully recognize two traffic signs with recognition rates of 91% and 60%, respectively.

A comprehensive analysis of the experimental results showed that the improvements introduced by adding the CBAM attention mechanism and the OTA loss function had significant advantages in recognizing traffic signs under complex conditions such as dusk, occlusion, and blur. The CBAM mechanism helps the model to more accurately capture key features, thereby improving the accuracy of object detection. Meanwhile, OTA optimizes the learning process by adjusting the loss function, further enhancing the model’s recognition performance. The combination of these two mechanisms provides the model with more precise object detection capabilities, effectively improving the overall recognition performance.



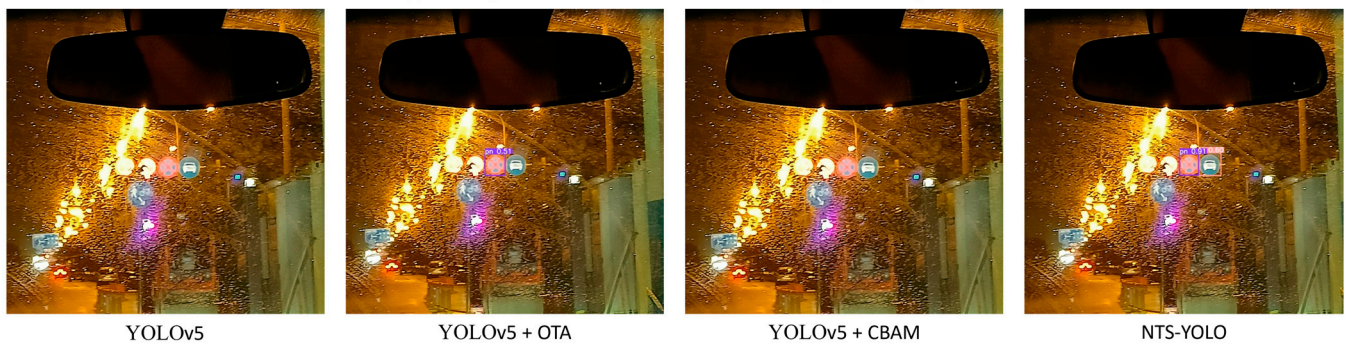
(a) Comparison results of distant small targets in the evening



(b) Comparison results of occluded targets at night



(c) Comparison results of blurred images at night



(d) Comparison results of blurred images on a rainy night

Figure 9. Comparison of results in different scenarios. There were four types of scenes in total, namely small targets in the evening, obstructed targets at night, blurred targets at night, and rainy days at night.

4. Conclusions

This study improved the nighttime traffic sign recognition network to enhance its performance in complex nighttime scenarios. Firstly, we introduced the CBAM attention mechanism and the OTA loss function based on the YOLOv5 model to address the challenges of recognizing traffic signs in nighttime environments caused by insufficient lighting, target blurriness, and occlusion. The experimental results show that through the unsupervised nighttime image enhancement method, this paper successfully improved the illumination balance and reduced the background noise level of the dataset, which significantly improved the recognition accuracy of traffic signs.

Secondly, after introducing the CBAM attention mechanism, the recognition accuracy of the model in the nighttime environment was improved by 0.056%. Meanwhile, the application of the OTA loss function resulted in a 0.052% improvement in the target detection accuracy. However, the introduction of the additional attention mechanism and the optimized loss function resulted in a slight decrease of 8.4% in the model processing speed. Nevertheless, this performance degradation is acceptable considering the significant improvement in the recognition accuracy. In practical applications, the relationship between accuracy and the processing speed needs to be weighed to achieve optimal performance.

However, it is worth noting that the dataset used in this study was primarily based on Chinese traffic signs, which to some extent limits the model's applicability in international scenarios. Chinese traffic signs exhibit regional characteristics in terms of their shape, color, and text symbols. For example, red circles are used for prohibition signs, and blue rectangles are used for instruction signs. In contrast, traffic signs in other countries may differ significantly in design. For instance, Europe uses white signs with red borders widely, while North America predominantly adopts yellow diamond-shaped warning signs. These design differences may lead to a decline in the recognition accuracy when the model is applied to cross-country scenarios, thereby limiting its applicability in international traffic sign detection tasks.

In the future, to address the limitations of this study and improve the model, the following directions will be pursued:

- (1) Further optimize the computational complexity of the CBAM and OTA modules. By adopting lightweight attention mechanisms or sparsification optimization strategies, the additional computational load can be reduced to improve the inference speed.
- (2) Expand the internationalization of the training dataset. By incorporating diverse datasets that encompass traffic sign characteristics from more countries (e.g., GT-SRB and LISA), the model's cross-regional applicability can be enhanced, thereby improving its generalization ability.

Author Contributions: Conceptualization, Y.H. and M.G.; methodology, Y.H. and M.G.; software, Y.H. and M.G.; validation, Y.H., M.G., Y.Z. and J.X.; formal analysis, Y.Z., M.G. and Y.H.; investigation, Y.Z., M.G. and Y.H.; resources, Y.Z., M.G. and Y.H.; data curation, Y.H., M.G. and X.G.; writing—original draft preparation, M.G., X.G., T.Z. and R.D.; writing—review and editing, Y.H., M.G., Y.Z. and J.X.; visualization, M.G., Y.Z. and J.X.; supervision, Y.H. and Y.Z.; project administration, Y.H. and M.G.; funding acquisition, Y.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Open Fund of Key Laboratory of Urban Land Resources Monitoring and Simulation, Ministry of Natural Resources (Grant No.KF-2022-07-012).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Sequence data supporting the results of this study have been deposited in figshare with primary access via <https://doi.org/10.6084/m9.figshare.25816276.v1>.

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

References

- Benallal, M.; Meunier, J. Real-time color segmentation of road signs. In Proceedings of the CCECE 2003—Canadian Conference on Electrical and Computer Engineering. Toward a Caring and Humane Technology (Cat. No.03CH37436), Montreal, QC, Canada, 4–7 May 2003; pp. 1823–1826.
- Janssen, R.; Ritter, W.; Stein, F.; Ott, S. Hybrid Approach For Traffic Sign Recognition. In Proceedings of the Intelligent Vehicles 93 Symposium, Tokyo, Japan, 14–16 July 1993; pp. 390–395.
- Ruta, A.; Li, Y.; Liu, X. Real-time traffic sign recognition from video by class-specific discriminative features. *Pattern Recognit.* **2010**, *43*, 416–430. [\[CrossRef\]](#)
- Ruta, A.; Porikli, F.; Watanabe, S.; Li, Y. In-vehicle camera traffic sign detection and recognition. *Mach. Vis. Appl.* **2009**, *22*, 359–375. [\[CrossRef\]](#)
- Greenhalgh, J.; Mirmehdi, M. Real-Time Detection and Recognition of Road Traffic Signs. *IEEE Trans. Intell. Transp. Syst.* **2012**, *13*, 1498–1506. [\[CrossRef\]](#)
- Ritter, W.; Stein, F.; Janssen, R. Traffic sign recognition using colour information. *Math. Comput. Model.* **1995**, *22*, 149–161. [\[CrossRef\]](#)
- Duda, R.O.; Hart, P.E. Use of the Hough transformation to detect lines and curves in pictures. *Commun. ACM* **1972**, *15*, 11–15. [\[CrossRef\]](#)
- Loy, G.; Zelinsky, A. Fast radial symmetry for detecting points of interest. *IEEE Trans. Pattern Anal. Mach. Intell.* **2003**, *25*, 959–973. [\[CrossRef\]](#)
- Belaroussi, R.; Tarel, J.-P. Angle vertex and bisector geometric model for triangular road sign detection. In Proceedings of the 2009 Workshop on Applications of Computer Vision (WACV), Snowbird, UT, USA, 7–8 December 2009; pp. 1–7.
- Li, H.; Sun, F.; Liu, L.; Wang, L. A novel traffic sign detection method via color segmentation and robust shape matching. *Neurocomputing* **2015**, *169*, 77–88. [\[CrossRef\]](#)
- Grigorescu, C.; Petkov, N. Distance sets for shape filters and shape recognition. *IEEE Trans. Image Process* **2003**, *12*, 1274–1286. [\[CrossRef\]](#)
- Betke, M.; Makris, N.C. Fast object recognition in noisy images using simulated annealing. In Proceedings of the IEEE International Conference on Computer Vision, Cambridge, MA, USA, 20–23 June 1995; pp. 523–530.
- Khan, J.F.; Bhuiyan, S.M.A.; Adhami, R.R. Image Segmentation and Shape Analysis for Road-Sign Detection. *IEEE Trans. Intell. Transp. Syst.* **2011**, *12*, 83–96. [\[CrossRef\]](#)
- Zaklouta, F.; Stanculescu, B. Real-time traffic sign recognition in three stages. *Robot. Auton. Syst.* **2014**, *62*, 16–24. [\[CrossRef\]](#)
- Aziz, S.; Mohamed, E.A.; Youssef, F. Traffic Sign Recognition Based On Multi-feature Fusion and ELM Classifier. *Procedia Comput. Sci.* **2018**, *127*, 146–153. [\[CrossRef\]](#)
- Xue, B.; Li, W.; Song, H.-Y.; Fang, A.-Q.; Peng, J.-T.; Wang, P.-J.; Guo, H.-Y. Review on Feature Extraction of Traffic Sign Recognition. *J. Graph.* **2019**, *40*, 1024–1031.
- Houben, S.; Stallkamp, J.; Salmen, J.; Schlipsing, M.; Igel, C. Detection of traffic signs in real-world images: The German traffic sign detection benchmark. In Proceedings of the 2013 International Joint Conference on Neural Networks (IJCNN), Dallas, TX, USA, 4–9 August 2013; pp. 1–8.
- Wang, G.; Ren, G.; Wu, Z.; Zhao, Y.; Jiang, L. A hierarchical method for traffic sign classification with support vector machines. In Proceedings of the 2013 International Joint Conference on Neural Networks (IJCNN), Dallas, TX, USA, 4–9 August 2013; pp. 1–6.
- Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 580–587.
- 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. [\[CrossRef\]](#)
- Girshick, R. Fast R-CNN. In Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015; pp. 1440–1448.
- 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. [\[CrossRef\]](#)
- Jiang, B.; Luo, R.; Mao, J.; Xiao, T.; Jiang, Y. Acquisition of Localization Confidence for Accurate Object Detection. In Proceedings of the Computer Vision—ECCV 2018: 15th European Conference, Munich, Germany, 8–14 September 2018; Part XIV; Springer: Munich, Germany, 2018; pp. 816–832.
- Rezatofighi, H.; Tsoi, N.; Gwak, J.; Sadeghian, A.; Reid, I.; Savarese, S. Generalized Intersection Over Union: A Metric and a Loss for Bounding Box Regression. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–19 June 2019; pp. 658–666.

25. Wang, J.; Chen, Y.; Dong, Z.; Gao, M. Improved YOLOv5 network for real-time multi-scale traffic sign detection. *Neural Comput. Appl.* **2023**, *35*, 7853–7865. [\[CrossRef\]](#)
26. Zhang, Y.; Lu, Y.; Zhu, W.; Wei, X.; Wei, Z. Traffic sign detection based on multi-scale feature extraction and cascade feature fusion. *J. Supercomput.* **2022**, *79*, 2137–2152.
27. Wang, L.; Wang, L.; Zhu, Y.; Chu, A.; Wang, G. CDFF: A fast and highly accurate method for recognizing traffic signs. *Neural Comput. Appl.* **2022**, *35*, 643–662.
28. Zhang, J.; Yi, Y.; Wang, Z.; Alqahtani, F.; Wang, J. Learning multi-layer interactive residual feature fusion network for real-time traffic sign detection with stage routing attention. *J. Real-Time Image Process.* **2024**, *21*, 176. [\[CrossRef\]](#)
29. Fang, C.Y.; Fuh, C.S.; Yen, P.S.; Cherng, S.; Chen, S.W. An automatic road sign recognition system based on a computational model of human recognition processing. *Comput. Vis. Image Underst.* **2004**, *96*, 237–268. [\[CrossRef\]](#)
30. Ciresan, D.; Meier, U.; Masci, J.; Schmidhuber, J. Multi-column deep neural network for traffic sign classification. *Neural Netw.* **2012**, *32*, 333–338. [\[CrossRef\]](#)
31. Qian, R.; Yue, Y.; Coenen, F.; Zhang, B. Traffic sign recognition with convolutional neural network based on max pooling positions. In Proceedings of the 2016 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), Changsha, China, 13–15 August 2016; pp. 578–582.
32. Aghdam, H.H.; Heravi, E.J.; Puig, D. Toward an optimal convolutional neural network for traffic sign recognition. In Proceedings of the International Conference on Machine Vision, Barcelona, Spain, 19–21 November 2015.
33. Xie, K.; Ge, S.; Ye, Q.; Luo, Z. Traffic Sign Recognition Based on Attribute-Refinement Cascaded Convolutional Neural Networks. In Proceedings of the Advances in Multimedia Information Processing-PCM 2016, Xi'an, China, 15–16 September 2016; pp. 201–210.
34. Yi, S.; Nie, Y.; Zhang, Y.; Zhao, Q.; Zhuang, Y. Nighttime Target Recognition Method Based on Infrared Thermal Imaging and YOLOv3. *Infrared Technol.* **2019**, *41*, 970–975.
35. Qu, S.; Yang, X.; Zhou, H.; Xie, Y. Improved YOLOv5-based for small traffic sign detection under complex weather. *Sci. Rep.* **2023**, *13*, 16219. [\[CrossRef\]](#)
36. Wang, Q.; Li, X.; Lu, M. An Improved Traffic Sign Detection and Recognition Deep Model Based on YOLOv5. *IEEE Access* **2023**, *11*, 54679–54691. [\[CrossRef\]](#)
37. Liu, H.; Zhou, K.; Zhang, Y.; Zhang, Y. ETSR-YOLO: An improved multi-scale traffic sign detection algorithm based on YOLOv5. *PLoS ONE* **2023**, *18*, e0295807. [\[CrossRef\]](#)
38. Yan, H.; Pan, S.; Zhang, S.; Wu, F.; Hao, M. Sustainable utilization of road assets concerning obscured traffic signs recognition. *Proc. Inst. Civ. Eng. Eng. Sustain.* **2024**, 1–11. [\[CrossRef\]](#)
39. Song, G. An Improved Traffic Sign Recognition Algorithm Based on Deep Learning. In Proceedings of the 2021 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS), Xi'an, China, 27–28 March 2021; pp. 1–4.
40. Li, W.; Na, X.; Su, P.; Zhang, Q. Traffic sign detection and recognition based on CNN-ELM. *J. Phys. Conf. Ser.* **2021**, *1848*, 12106.
41. Lin, S.; Zhang, Z.; Tao, J.; Zhang, F.; Fan, X.; Lu, Q. Traffic Sign Detection Based on Lightweight Multiscale Feature Fusion Network. *Sustainability* **2022**, *14*, 14019. [\[CrossRef\]](#)
42. Zhao, S.; Gong, Z.; Zhao, D. Traffic signs and markings recognition based on lightweight convolutional neural network. *Vis. Comput.* **2024**, *40*, 559–570. [\[CrossRef\]](#)
43. Jin, Y.; Yang, W.; Tan, R.T. Unsupervised night image enhancement: When layer decomposition meets light-effects suppression. In Proceedings of the European Conference on Computer Vision, Tel Aviv, Israel, 23–27 October 2016; pp. 404–421.
44. Zhang, J.; Zou, X.; Kuang, L.-D.; Wang, J.; Sherratt, R.S.; Yu, X. CCTSDB 2021: A more comprehensive traffic sign detection benchmark. *Hum.-Centric Comput. Inf. Sci.* **2022**, *12*, 106129.
45. Zhu, Z.; Liang, D.; Zhang, S.; Huang, X.; Li, B.; Hu, S. Traffic-sign detection and classification in the wild. In Proceedings of the IEEE conference on computer vision and pattern recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2110–2118.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.