

## Article

# Object Detection for Hazardous Material Vehicles Based on Improved YOLOv5 Algorithm

Pengcheng Zhu <sup>1</sup>, Bolun Chen <sup>1,2,\*</sup>, Bushi Liu <sup>1,†</sup>, Zifan Qi <sup>1</sup>, Shanshan Wang <sup>1,†</sup> and Ling Wang <sup>1</sup><sup>1</sup> Faculty of Computer and Software Engineering, Huaiyin Institute of Technology, Huaian 223003, China<sup>2</sup> Department of Physics, University of Fribourg, CH-1700 Fribourg, Switzerland

\* Correspondence: chenbolun@hyit.edu.cn

† These authors contributed equally to this work.

**Abstract:** Hazardous material vehicles are a non-negligible mobile source of danger in transport and pose a significant safety risk. At present, the current detection technology is well developed, but it also faces a series of challenges such as a significant amount of computational effort and unsatisfactory accuracy. To address these issues, this paper proposes a method based on YOLOv5 to improve the detection accuracy of hazardous material vehicles. The method introduces an attention module in the YOLOv5 backbone network as well as the neck network to achieve the purpose of extracting better features by assigning different weights to different parts of the feature map to suppress non-critical information. In order to enhance the fusion capability of the model under different sized feature maps, the SPPF (Spatial Pyramid Pooling-Fast) layer in the network is replaced by the SPPCSPC (Spatial Pyramid Pooling Cross Stage Partial Conv) layer. In addition, the bounding box loss function was replaced with the SIOU loss function in order to effectively speed up the bounding box regression and enhance the localization accuracy of the model. Experiments on the dataset show that the improved model has effectively improved the detection accuracy of hazardous chemical vehicles compared with the original model. Our model is of great significance for achieving traffic accident monitoring and effective emergency rescue.

**Keywords:** hazardous material vehicles; object detection; YOLOv5; attention mechanism

**Citation:** Zhu, P.; Chen, B.; Liu, B.; Qi, Z.; Wang, S.; Wang, L. Object Detection for Hazardous Material Vehicles Based on Improved YOLOv5 Algorithm. *Electronics* **2023**, *12*, 1257. <https://doi.org/10.3390/electronics12051257>

Academic Editor: Donghyeon Cho

Received: 8 February 2023

Revised: 28 February 2023

Accepted: 4 March 2023

Published: 6 March 2023



**Copyright:** © 2023 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

Chemical raw materials play a pivotal role in people's lives, industrial production and the development of science and technology. With the continuous improvement of China's industrialization, the demand for chemical raw materials from all walks of life is increasing day by day. However, some chemical materials are explosive, corrosive, flammable, toxic and other characteristics, and once leakage occurs, it may cause a certain degree of harm to the human body and the surrounding environment [1]. Therefore, it is vital to ensure the safety of hazardous chemicals. According to statistics from the China Federation of Logistics and Purchasing (CFLP), more than 1 billion metric tons of hazardous chemicals are transported by road in China every year, accounting for more than 60% of the total transport of hazardous chemicals, and this proportion is still rising [2]. According to data from the State Administration of Safety Supervision and the fire service, 77% of accidents occur during transportation [3]. The transport of hazardous chemicals has become one of the highest risks in the safety of hazardous chemicals. Therefore, in order to reduce casualties and property damage, it is imperative to supervise hazardous chemical vehicles in transit.

In recent years, deep learning and machine vision have developed rapidly, such as object detection and image segmentation [4]. This provides the technical basis for the identification of hazardous material vehicles. The use of vehicle detection technology to identify important roads in the detection of dangerous chemical vehicles, from road

management departments for the vehicle to real-time dynamic monitoring, can avoid or reduce the occurrence of traffic accidents, or in the event of dangerous chemical transport accidents, can provide timely and effective emergency rescue, as best as possible avoid secondary accidents, reduce casualties and limit property damage [5].

At present, vehicle recognition technology can be divided into two categories according to the required hardware and software basis: one is the recognition method using physical parameters and the other is the recognition method using image processing technology [6]. Vehicle recognition based on physical parameters has high requirements on hardware, and although the recognition accuracy of this method is high, its cost is great and the maintenance work on the hardware during its use is also difficult. Vehicle recognition based on image processing technology is the extraction of feature information (color, texture, size, shape, etc.) from vehicle images using certain methods where possible. Vehicle image recognition is a process of multiple operations on the vehicle image, by using a specific algorithm to convert the vehicle image into a feature vector representation, and the obtained feature information is differentiated by a classification algorithm [7].

In recent years, vehicle detection has been the focus of many researchers and there has been a proliferation of research on vehicle detection. Bochkovskiy et al. proposed a new YOLO architecture. Firstly, the CSPDarknet (Cross Stage Partial Darknet) is used as the backbone network and the SPP (Spatial Pyramid Pooling) is used for feature fusion for the first time. Then, the PAN (path aggregation network) structure is used as the neck of the model to do a further fusion of feature maps. Glenn Jocher proposes a YOLOv5 network model based on it, using adaptive anchor frames, automatic learning based on the training set as well as LeakyReLU and Sigmoid as activation functions. The SPPF (Spatial Pyramid Pooling Fast) layer is proposed to replace the SPP layer first. The algorithm can achieve fast and accurate detection [8]. Wang et al. proposed a new object detection algorithm. In this algorithm, the SPPCSPC (Spatial Pyramid Pooling Cross Stage Partial Conv) module is proposed for the first time and the original SPPF module is replaced by this module to achieve better feature fusion [9]. Woo et al. proposed a convolutional block attention mechanism, and the classical network model such as ResNet using this module can make the network more focused on the object and exhibits remarkable effects in the field of object detection [10]. Gevorgyan et al. proposed a new bounding box regression algorithm. This algorithm considers the direction of mismatch between the ground truth box and the prediction box and can effectively improve the accuracy of model inference [11].

Djenouri et al. proposed an improved regional convolutional neural network, which first uses a SIFT extractor to remove noise (set of outlier images) and then builds an improved regional convolutional neural network to detect vehicles at different scales, achieving improvements in detection accuracy and proposing a new hyperparametric optimization model based on evolutionary computation that can be used to optimize the deep learning framework of parameters [12]. Wang et al. proposed a new method for vehicle detection based on multi-sensor fusion. First, multiple sensors are combined to extract the target vehicle. Second, the potential area of the vehicle in the feature map is estimated according to the distribution of the target vehicle detected by the sensors, and predicting the region of interest (ROI) of the vehicle according to pixel regression. Finally, four new haar-like feature templates are developed to enhance the detection performance of vehicles [13]. This method can remarkably enhance vehicle detection performance in severe weather.

In the process of vehicle detection, current detection algorithms face challenges such as large calculations and unsatisfactory detection accuracy. Dong et al. proposed an improved lightweight YOLOv5 method for vehicle detection. The method introduces the C3Ghost and Ghost modules in the YOLOv5 neck network and the convolutional attention module in the backbone network to improve the detection accuracy of the algorithm, and then further considers CIoU loss as the bounding box regression loss function to speed up the bounding box regression and improve the localization accuracy of the algorithm [14]. The effectiveness and superiority of the method are demonstrated by example analysis and

comparison. Li et al. propose a hierarchical joint CNN model. The method uses a Faster R-CNN to extract multiple feature maps for the vehicle image; then, a CNN is used to train multiple feature maps, and finally, multiple classifiers are used to achieve the fine recognition of vehicles [15]. Mi et al. proposed a fusion algorithm of aggregated region classes and two-stage SVM classifiers for the detection of container trucks in ports. This method can display better truck detection performance than traditional methods [16].

In order to rapidly detect moving vehicles in road transportation, Chen et al. propose an SSD-based vehicle detection algorithm. The method replaces the backbone network of SSD (single-shot multibox detector) with the MobileNet-v2 network to achieve a lightweight model, and introduces an attention mechanism in the model to enhance the model's ability to extract vehicle features. Finally, a bottom-up feature fusion architecture is built based on a deconvolution module to enhance detection accuracy [17]. Kang et al. proposed a remote sensing satellite video motion vehicle fast detection method with an automatic region of interest constraint. Firstly, the region of interest of the moving vehicle is rapidly and automatically acquired; then, the fast detection of moving vehicles in the region of interest is achieved based on an improved Gaussian background subtraction under the region of interest constraint [18]. Zhang et al. proposed a detection algorithm based on sample adaptive segmentation. The method adopts different update strategies for different detection regions to achieve an adaptive update of background samples, and randomly penetrates the diffuse background points into the foreground region with a certain probability to update the background samples in its neighborhood, achieving the fast detection of vehicles in motion [19]. Alsanad et al. proposed a real-time truck monitoring algorithm based on YOLOv2. Aiming at the shortage of traditional methods that only pay attention to the position of the class target to predict its probability in the class, this method considers the whole image area for strong target detection, which improves the effectiveness of truck detection [20]. Zhang et al. proposed a non-maximum suppression method based on position priority to achieve the detection of mud trucks. This method designs a new bounding box matching algorithm to solve the problem of object loss when the IoU of two proposals is greater than a threshold and redefines the loss function to fit the improved method. Experiments prove the effectiveness of this method [21].

The main source of image data for vehicle detection is traffic surveillance images [22]. In recent years, vehicle detection algorithms have flourished, especially the YOLO family represented by the use of deep learning methods. However, these methods require manually labeling a large amount of data to train the network model [23]. Currently, most of the datasets are labeled with vehicles captured during the daytime, while vehicle images at night are scarce [24]. Li et al. proposed a domain adaptive (DA) method based on a Fast R-CNN. The method can increase the number of labeled nighttime vehicle images in the dataset by using the existing labeled daytime vehicle images to complement the unlabeled nighttime vehicle images, thus improving the detection capability of the model for nighttime vehicles [25]. Chen et al. propose an information fusion-based algorithm for detecting vehicles driving at night, using millimeter wave radar and vision sensors to detect vehicles ahead at night in order to provide comprehensive and reliable information for night-time driving safety [26]. Hua et al. proposed an improved YOLOv3 model algorithm based on dark channel defogging to address the problem of poor detection accuracy due to the serious influence of fog in the process of vehicle detection in foggy weather. First, the image is defogged by the dark channel algorithm to improve the clarity of the image, and then an attention mechanism is introduced to further the feature extraction of the feature map used for detection, which improves the algorithm's ability to mine feature information [27].

In this paper, an algorithm based on an improved YOLOv5 model is proposed for the detection of hazardous chemical vehicles. The main contributions include (1) integrating the attention module in the feature extraction network to enhance the quality of feature extraction for hazardous chemical vehicles by paying attention to spatial semantic information and channel semantic information; (2) replacing the spatial pyramid pooling layer

(SPPF) in the backbone network by the Spatial Pyramid Pooling Cross Stage Partial Conv (SPPCSPC), which can enhance the fusion ability of the model under different size feature maps; (3) replacing the CIoU loss in the loss function by SIoU loss, which can effectively accelerate the speed of bounding box regression and increase the localization accuracy of the algorithm.

## 2. Methodology

The YOLOv5 algorithm makes predictions based on the whole image, giving all the detection results at once [28]. YOLOv5 has four different sizes of network models, namely YOLOv5s, YOLOv5m, YOLOv5l and YOLOv5x. YOLOv5x has the largest model and thus has the highest detection accuracy. YOLOv5s has the smallest model, but it has the fastest detection speed. In this paper, YOLOv5s is chosen as the baseline model and improved to maximize detection accuracy while keeping the detection speed largely unchanged. This section details the architecture of the improved hazardous material vehicles detection model.

Specifically, the input images were first processed using Mosaic data enhancement, image scaling and adaptive initial anchor box calculation, and then the enhanced images were fed into the improved YOLOv5 network model for hazardous chemical vehicle detection. To further improve the semantic quality of the output of the feature from the feature extraction network, we replaced the original SPPF layer by the SPPCSPC in the backbone, and added attention modules to the backbone and neck of the network to focus on spatial features and channel features. The CIoU in the original bounding box loss is replaced with SIoU, effectively improving the inference accuracy.

### 2.1. Data Augmentation

To obtain a well-performing neural network model, a large amount of data is often required, but the task of acquiring new data is often time-consuming and labor-intensive [29]. The use of data augmentation techniques can make full use of computers to generate data and increase the amount of data, for example by using scaling, panning, rotating, color transformations, etc. It is beneficial for data augmentation to increase the number of training samples and the ability to increase the generalization power of the model by adding suitable noisy data [30].

In order to obtain excellent detection results, data augmentation is also used in YOLOv5. YOLOv5 uses Mosaic data augmentation to increase the amount of data for small targets in the dataset and enrich the number of samples by randomly flipping, scaling and cropping four images into a new image. The Mosaic data enhancement effect is shown in Figure 1.



**Figure 1.** Illustration of the architecture of the Mosaic data enhancement.

### 2.2. Feature Extraction Backbone Network

YOLOv5 uses the CSP-Darknet53 network as the backbone of the model, which is mainly composed of modules such as Focus, Convolution block and C3 module. Specifically,

the Focus module divides the feature data into four parts, each corresponding to two down samples, spliced in the channel dimension and then convolved to obtain a binary down-sampled feature map with no information loss, the structure of which is shown in Figure 2. The convolution module is the basic convolution unit of YOLOv5, which performs two-dimensional convolution, batch normalization and activation operations on the feature map, in turn. The C3 module consists of several modules called bottleneck residual connections, in which the feature map passes through two convolution layers and then performs an additive operation with the original feature map. This structure accomplishes the migration of residual features without increasing the channel depth. The structure of the C3 module is shown in Figure 3.

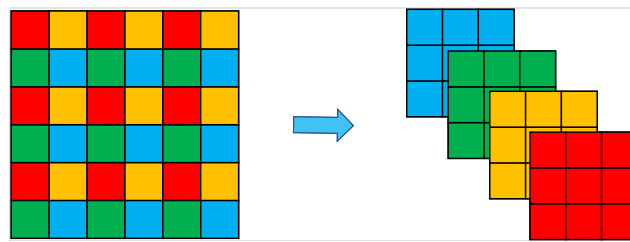


Figure 2. Illustration of the architecture of the Focus module.

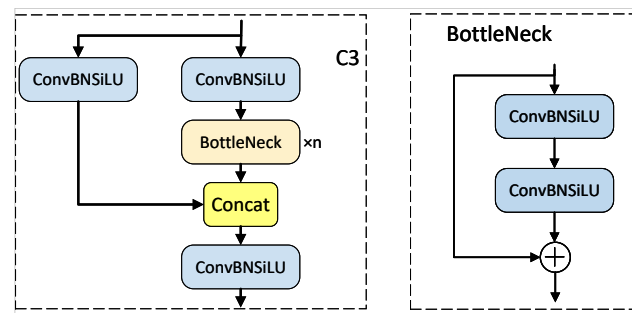


Figure 3. Illustration of the architecture of the C3 module.

The attention mechanism allows the model to better focus on vehicle information features and suppress non-vehicle information features, enabling the model to extract more accurate semantic information about the vehicle [31]. We add an attention module to the YOLOv5 backbone network to recalibrate the feature maps in order to enhance the feature representation capabilities. The attention module generates attentional feature map information in both channel and spatial dimensions, and then the feature map information is multiplied with the previous feature map for adaptive feature correction to produce the defined feature map. The architecture of the attention module is shown in Figure 4.

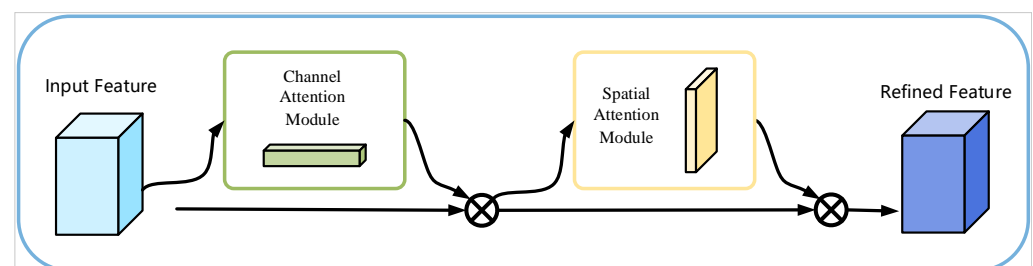


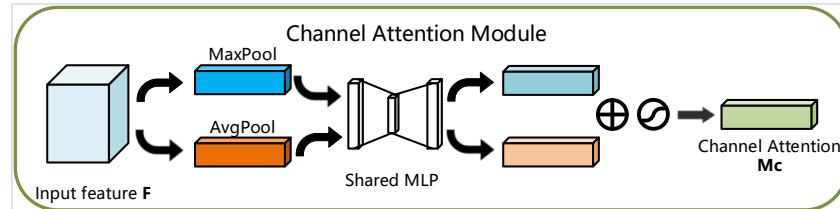
Figure 4. Illustration of the architecture of the attention module.

Channel attention mechanism focuses on the feature relationships between channels in the feature map to generate the channel attention, and its module structure is shown in Figure 5. Channel attention performs global maximum pooling and mean maximum pooling in the channel dimension of the feature map to aggregate the feature information in



the spatial dimension, respectively. Finally, a channel attention feature map of  $M_c \in R^{C \times 1 \times 1}$  is obtained. This feature map is used as the input of the spatial attention module. The specific formula is formulated as follows:

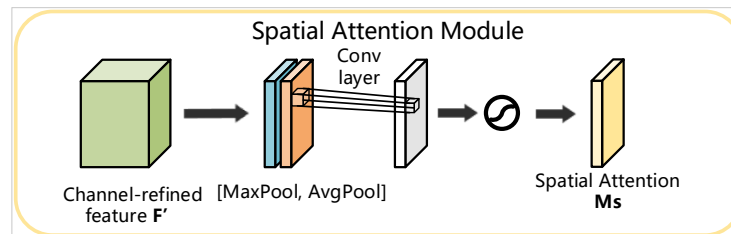
$$\begin{aligned} M_c(F) &= \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F))) \\ &= \sigma(W_1(W_0(F_{avg}^c)) + W_1(W_0(F_{max}^c))) \end{aligned} \quad (1)$$



**Figure 5.** Illustration of the architecture of the channel attention module.

After paying attention to the channel attention, the feature map is input into the spatial attention module, and its model structure is shown in Figure 6. First, the global max-pooling and global avg-pooling based on the channel dimension are performed, and then both tensors are spliced in the channel dimension. The spliced tensor depth is reduced to 1 in the channel dimension by a convolution operation. Spatial attention features are generated after sigmoid activation, and finally the original feature map is fused with the spatial attention features to obtain the feature map focusing on spatial information. The specific formula is formulated as follows:

$$\begin{aligned} M_s(F) &= \sigma(f^{7 \times 7}([AvgPool(F); MaxPool(F)])) \\ &= \sigma(f^{7 \times 7}(F_{avg}^s; F_{max}^s)) \end{aligned} \quad (2)$$



**Figure 6.** Illustration of the architecture of the spatial attention module.

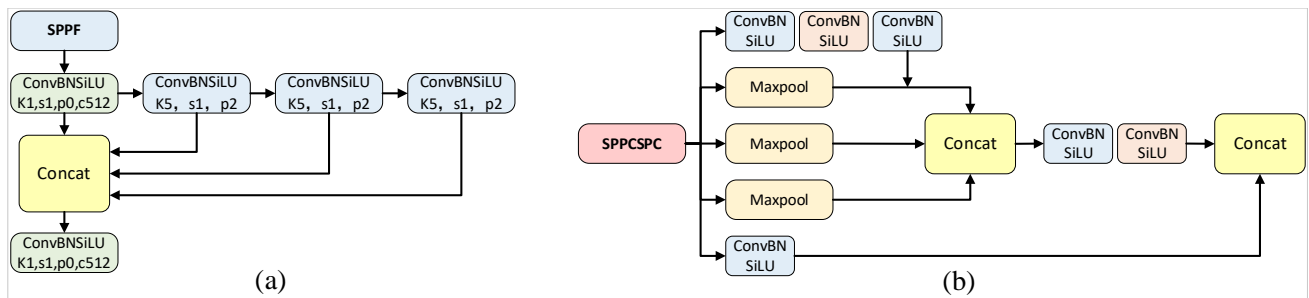
### 2.3. Neck

The neck of the YOLOv5 model consists mainly of Conv modules, C3 modules and SPP (Spatial Pyramid Pooling), and uses a PANet (Path Aggregation Network) as the neck for feature aggregation. Specifically, the SPP (Spatial Pyramid Pooling) performs max-pooling of different convolutional kernel sizes and integrates features. It can convert arbitrary-sized feature maps into fixed-sized features and effectively solve the problem of the repeated extraction of related features by convolutional neural networks. The structure of the SPPF (Feature Pyramid Pooling Fast) is shown in Figure 7a.

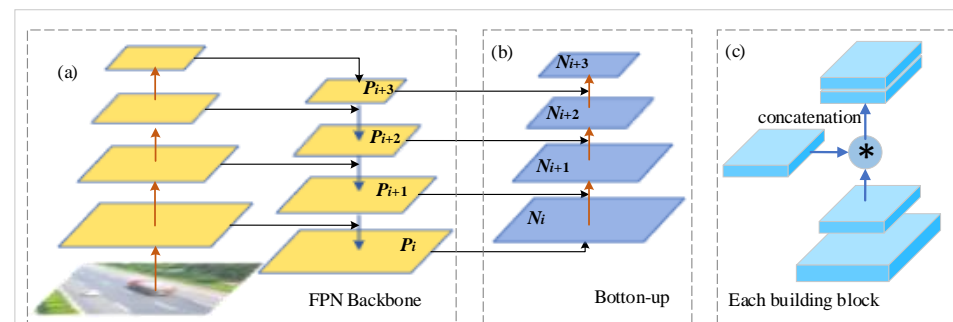
To further improve the capability of network feature fusion, we replaced the SPPF of YOLOv5 by the SPPCSPC module to enhance the detection capability of the model for different scales. The structure of SPPCSPC is shown in Figure 7b below.

In neural networks, the deeper the networks, the better the extraction of object feature information and the better the detection of the object by the model. However, the network model also makes the location information of the object blurred, and causes the loss of feature information for small objects as it continues to deepen. YOLOv5 adopts a PANet structure for the multi-scale fusion of features, which enables the bottom feature map to contain more semantic information of vehicles through top-down upsampling [32]. The

PAN structure achieves bottom-up subsampling through convolution. The PAN structure achieves bottom-up downsampling by convolution, so that the top-level feature map contains stronger information about the location of the vehicles. The specific process is shown in Figure 8. Through the feature aggregation, the feature maps of different sizes contain richer semantic information and location information, thus ensuring the accuracy of the detection of different sizes of hazardous materials vehicles.

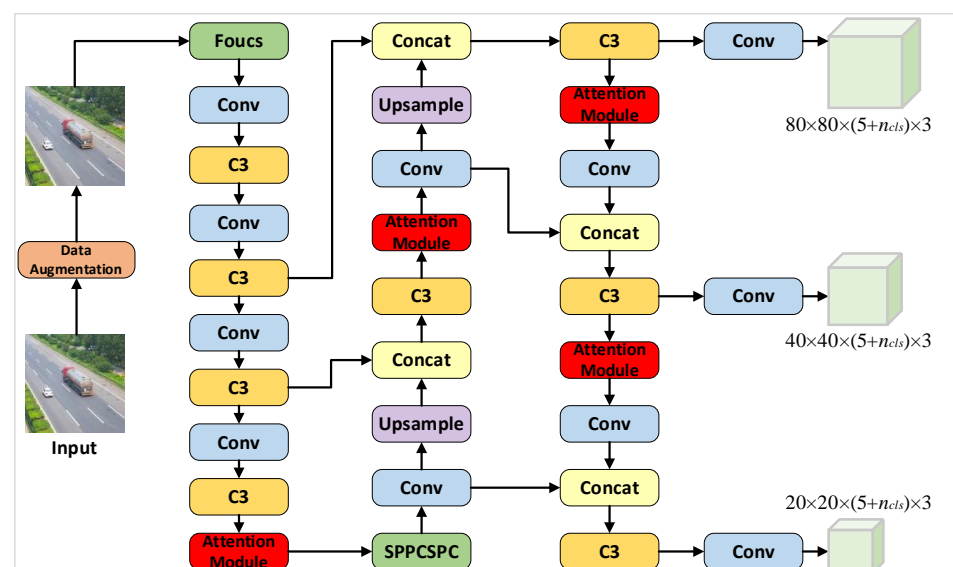


**Figure 7.** (a) Illustration of the architecture of the SPPF module. (b) Illustration of the architecture of the SPPCSPC module.



**Figure 8.** Illustration of Path aggregation network. (a) FPN Backbone. (b) Bottom-up path augmentation. (c) Each building block.

In order to enable the network to better learn the semantic information in the vehicle images, focusing on important information and suppressing useless information, we also introduced the attention mechanism into the neck structure of YOLOv5. The overall network structure we proposed is shown in Figure 9.



**Figure 9.** Illustration of the architecture of the designed vehicle detection model.

## 2.4. Loss Function

The loss function of YOLOv5 consists of three components: the confidence loss  $l_{obj}$ , the classes loss  $l_{cls}$  and the position loss of the bounding box  $l_{box}$ . The network divides the feature map into several cells, and each cell corresponds to a vector  $y = (t_x, t_y, t_w, t_h, p_o, c_1, c_2, c_3, c_4)$ , where  $t_x, t_y$  is used to calculate the offset between the prediction box and the corresponding anchor box, and  $t_w, t_h$  are used to calculate the width and height of the prediction box.  $p_o$  is the probability that the cell contains the object to be detected, and  $c_1, c_2, c_3, c_4$  are the prediction values of the four classes corresponding to the hazardous material vehicles dataset. The loss function is calculated as follows:

$$L_{v5}(t_p, t_{gt}) = \sum_{k=0}^K [\alpha_k^{balance} \alpha_{box} \sum_{i=0}^{S^2} \sum_{j=0}^B \Pi_{kij}^{obj} L_{CIoU} + \alpha_{obj} \sum_{i=0}^{S^2} \sum_{j=0}^B \Pi_{kij}^{obj} L_{obj} + \alpha_{cls} \sum_{i=0}^{S^2} \sum_{j=0}^B \Pi_{kij}^{obj} L_{cls}] \quad (3)$$

The confidence loss  $l_{obj}$  is formulated according to positive sample matching, including the object confidence score  $p_o$  in the prediction box and the intersection over union of the prediction box and the ground truth box. Both calculate the binary cross-entropy to obtain the final object confidence loss. The confidence loss  $l_{obj}$  is defined as follows:

$$l_{obj} = \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{obj} (\hat{C}_i \log(C_i) + (1 - \hat{C}_i) \log(1 - C_i)) - \lambda_{nobj} \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{nobj} (\hat{C}_i \log(C_i) + (1 - \hat{C}_i) \log(1 - C_i)) \quad (4)$$

Classes loss is similar to confidence loss in that classes loss is calculated from the classes score of the prediction box and the one-hot value of the ground truth box classes. The classes loss  $l_{cls}$  is defined as follows:

$$l_{cls} = \sum_{i=0}^{S^2} I_{ij}^{obj} \sum_{c \in classes} (\hat{P}_i(c) \log(P_i(c)) + (1 - \hat{P}_i(c)) \log(1 - P_i(c))) \quad (5)$$

The CIoU loss is used in the position loss of the prediction box. That takes into account three geometric factors of the bounding box regression function: overlap area, centroid distance and aspect ratio. The position loss  $l_{box}$  is defined as follows:

$$l_{box} = l_{CIoU} = 1 - CIoU = 1 - (IoU - \frac{d_o^2}{d_c^2} - \frac{v^2}{1 - IoU + v}), \quad (6)$$

$$v = \frac{4}{\pi^2} (\arctan \frac{w_{gt}}{h_{gt}} - \arctan \frac{b_w}{h_w})^2$$

where  $d_o$  is the centroid Euclidean distance between the bounding box and ground truth box,  $d_c$  is the diagonal distance between the bounding box,  $v$  is a parameter measuring the consistency of the aspect ratio,  $w_{gt}$  and  $h_{gt}$  are the width and height of the ground truth box, respectively, and  $w_p$  and  $h_p$  are the width and height of the prediction box, respectively.

Considering the possible directional mismatch between the prediction box and the ground truth box, we introduce a new bounding box position loss, replacing the original CIoU loss with the SIoU loss. This loss takes into account the vector angle between regressions and redefines the penalty metric, effectively reducing the total degrees of freedom of the loss. The SIoU loss consists of four costs: angle, distance, shape and IoU, which is calculated as follows:



The angle cost is defined as follows:

$$\begin{aligned}\Lambda &= 1 - 2 * \sin^2(\arcsin(x) - \frac{\pi}{4}), x = \frac{c_h}{\sigma} = \sin(\alpha), \\ \sigma &= \sqrt{(b_{c_x}^{gt} - b_{c_x})^2 + (b_{c_y}^{gt} - b_{c_y})^2}, \\ c_h &= \max(b_{c_y}^{gt}, b_{c_y}) - \min(b_{c_y}^{gt}, b_{c_y})\end{aligned}\quad (7)$$

Distance cost has been redefined based on the definition of angle cost:

$$\Delta = \sum_{t=x,y} (1 - e^{-\gamma \rho_t}), \rho_x = (b_{c_x}^{gt} - b_{c_x})^2, \rho_y = (\frac{b_{c_y}^{gt} - b_{c_y}}{c_h})^2, \gamma = 2 - \Lambda \quad (8)$$

Shape cost:

$$\Omega = \sum_{t=w,h} (1 - e^{-\omega_t})^\theta, \omega_w = \frac{|w - w^{gt}|}{\max(w, w^{gt})}, \omega_h = \frac{|h - h^{gt}|}{\max(h, h^{gt})} \quad (9)$$

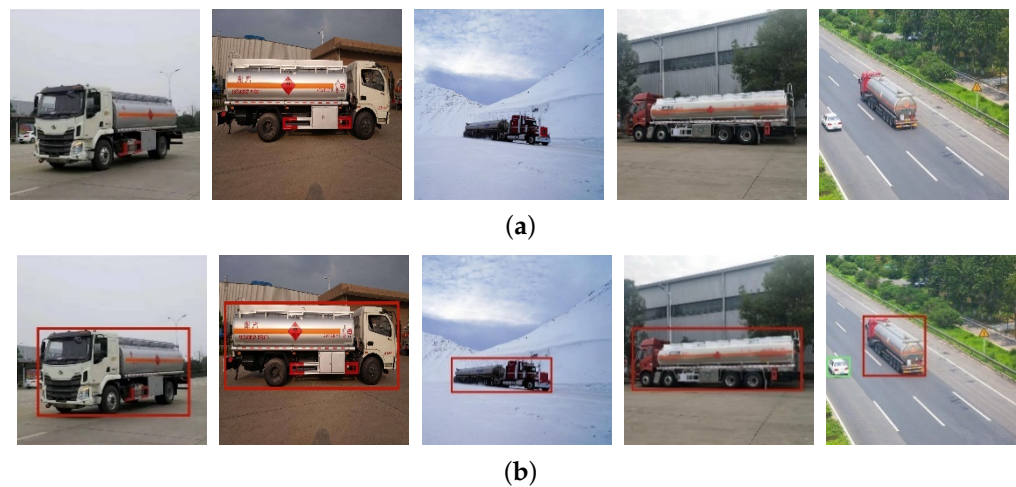
Finally, the regression loss function for the position loss of bounding box is written as follows:

$$l_{box} = 1 - IoU + \frac{\Delta + \Omega}{2} \quad (10)$$

### 3. Results and Discussions

#### 3.1. Dataset

In order to evaluate the accuracy of the algorithm, 4363 vehicle image samples were collected and the vehicles were classified into four categories, namely car, bus, truck and hazardous material vehicles, which included 2200 samples of hazardous material vehicles. The images were annotated using labelme, and then the label in json format was converted into text format in YOLO to generate  $id$ ,  $x$ ,  $y$ ,  $w$  and  $h$ . The training set, validation set and test set are divided according to the ratio of 6:2:2. Figure 10a shows a sample image of some of the hazardous chemical vehicles and Figure 10b illustrates the ground truths using bounding boxes.



**Figure 10.** The hazardous material vehicle object detection dataset. (a) Hazardous material vehicle images samples. (b) Ground-truth with bounding boxes.

#### 3.2. Network Configuration

The model was trained using the stochastic gradient descent (SGD) algorithm to update and optimize the weights of the network model. The specific parameters were set

as follows: the images were pre-processed and input into the network with a standard size of  $640 \times 640 \times 3$ , a batch size of 16, a learning rate of 0.01, a momentum parameter of 0.937, a weight decay factor of 0.0005 and a total of 150 training rounds for the hazardous material vehicles dataset.

### 3.3. Hazardous Material Vehicles Detection

We divided 20% of the dataset into test sets. To quantitatively assess the model's performance in detecting hazardous material vehicles, we used the following five evaluation metrics as measures of hazardous material vehicles detection: *Precision*, *Recall*, *F-score*, *mAP@0.5* and *mAP@0.5:0.95*. The *Precision* is defined as the ratio of the number of target vehicles correctly predicted to the number of vehicles predicted by the model as targets. The specific formula is formulated as follows:

$$Precision = \frac{TP}{TP + FP} \quad (11)$$

The *Recall* is defined as the ratio of the number of all target vehicles to the number of target vehicles correctly predicted. The specific formula is formulated as follows:

$$Recall = \frac{TP}{TP + FN} \quad (12)$$

where *TP* represents the number of target vehicles correctly predicted, *FP* represents the number of non-target vehicles wrongly predicted. *FN* represents the target vehicles that are not detected.

*F-score* provides an overall evaluation by comprehensively taking into consideration the precision and the recall metrics. The specific formula is formulated as follows:

$$F\text{-score} = 2 \times \frac{precision \times recall}{precision + recall} \quad (13)$$

The specific formulas of *mAP@0.5* and *mAP@0.5:0.95* are formulated as follows:

$$AP = \int_0^1 p(R) dR \quad (14)$$

$$mAP = \frac{\sum_{i=1}^N AP_i}{N} \quad (15)$$

where *mAP@0.5* denotes the mean AP of all vehicle categories when the value of the cross-merge ratio is 0.5, and *mAP@0.5:0.95* denotes the mean map for different cross-merge ratio cases.

In order to verify the effectiveness of the improved model, we compared it with some mainstream two-stage and one-stage models. The experimental results are shown in Table 1. It can be seen that the improved model has a significant advantage in terms of accuracy when compared with most of the network models, especially the precision metric, which is improved by 3.0 percent compared with the other best model.

**Table 1.** Performance of hazardous material vehicles detection between different algorithms. (The optional value of each index is bold).

Methods	Precision	Recall	F-Score	mAP@0.5	mAP@0.5:0.95
YOLOv5s	0.877	0.779	0.825	0.854	0.631
Faster R-CNN	0.603	<b>0.876</b>	0.711	0.846	0.579
SSD	0.757	0.788	0.772	0.815	0.613
YOLOv3	0.899	0.718	0.798	0.829	0.650
Ours	<b>0.929</b>	0.757	<b>0.834</b>	<b>0.867</b>	<b>0.661</b>

The proposed method obtained a very competitive hazardous material vehicle detection accuracy on the test set with a precision of 0.929, an mAP@0.5 of 0.867 and an mAP@0.5:0.95 of 0.661. Although there is a reduction in the recall metric, the overall performance is satisfactory. This illustrates that the inclusion of an attention mechanism allows for the network to better capture important features and weaken non-critical features. Secondly, the SPPCSPC has better feature fusion capability than the SPPF. The introduction of SIoU takes into account the vector angle between bounding box regression, achieving faster convergence and better performance in inference.

Specifically, due to the integration of an attention mechanism and a new spatial pyramidal pooling layer, the proposed model can achieve higher detection accuracy compared to the original model for relatively small vehicles. In addition, the method still performs well in the case of multiple vehicles in a single image. Overall, the method can effectively handle images of hazardous material vehicles at different scales and containing multiple vehicles.

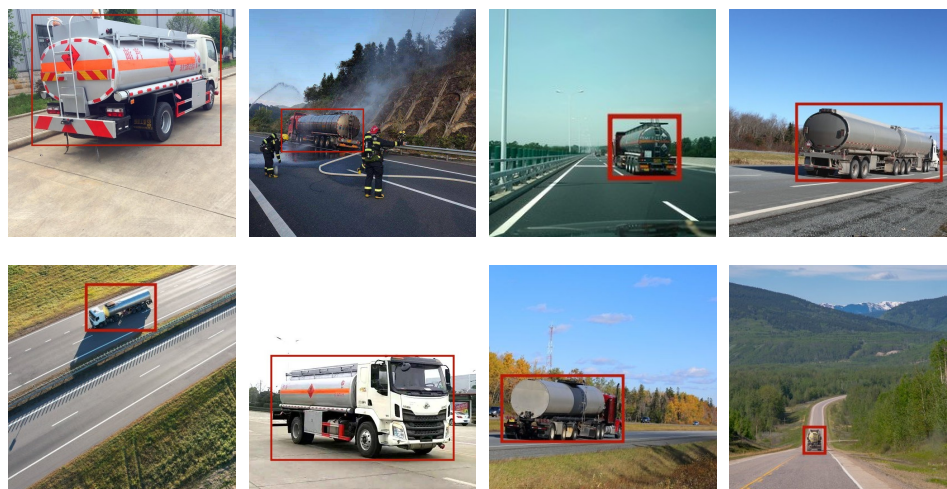
In order to more concretely represent the performance of the improved model, we list the precision, recall and F-score values of each class of the improved model and compare them with the initial model. The experimental results are shown in Table 2, where HM vehicle denotes the hazardous material vehicle. The experiments show that the improved model has evident advantages.

**Table 2.** Performance of per class detection between the original and the improved model.

Classes	Original Model			Improved Model		
	Precision	Recall	F-Score	Precision	Recall	F-Score
Bus	0.863	0.734	0.793	0.933	0.714	0.809
Truck	0.898	0.773	0.831	0.938	0.747	0.832
HM vheicle	0.881	0.822	0.850	0.943	0.814	0.874
Car	0.866	0.787	0.825	0.902	0.753	0.821

### 3.3.1. Single Hazardous Material Vehicle Detection Results

For visual inspection, Figures 11–13 show three subsets of hazardous material vehicle detection results taken from the test set and the content marked in the red boxes are the detected target vehicles. As shown in the images containing single hazardous material vehicles in Figure 11, vehicles of different sizes, especially the smaller ones, were correctly detected. This demonstrates the power of the SPPCSPC structure for the fusion of features of different sizes for the detection of differences in object size in images.



**Figure 11.** A subset of single hazardous material vehicle detection results.

### 3.3.2. Multi-Object Detection Results

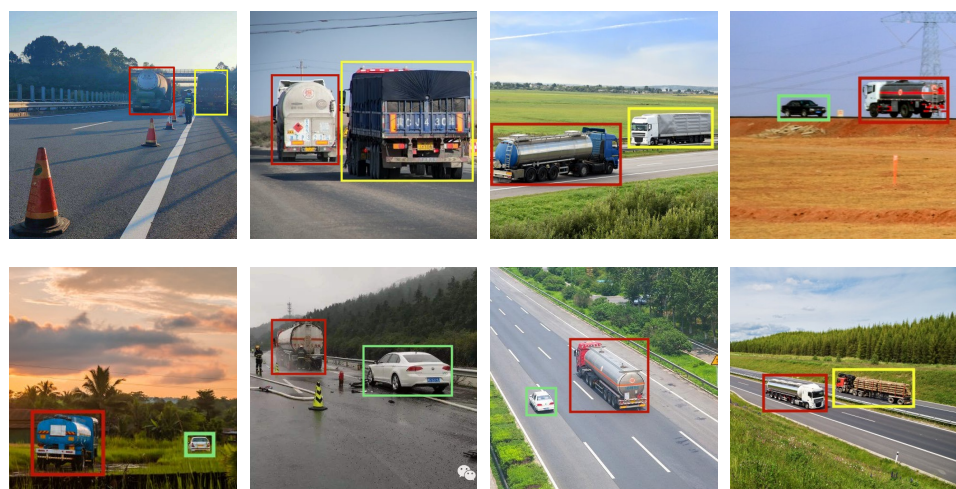
In addition, in the case of multiple hazardous material vehicles, the combination of clustering and feature pyramid pooling allows for the model to assign vehicles of different sizes under multiple objects to different feature layers, achieving the deep recognition for large sizes and the shallow recognition for small sizes, and finally completing the detection of multiple objects and correctly identifying the location of bounding boxes. The detection results of multiple hazardous chemical vehicles in the test set are shown in Figure 12.



**Figure 12.** A subset of multi-hazardous material vehicle detection results.

### 3.3.3. Multi-Category Vehicle Detection Results

Moreover, considering the complexity of road transport, where a road may contain several different categories of vehicles at the same moment, we designed the detection of different categories of vehicles to accommodate the diversity of vehicle changes on the road. Due to the inclusion of the attention mechanism, the model can focus on deeper semantic information about the vehicles and extract more accurate feature information of each category of vehicles, which makes the trained model have a better classification effect. As shown in Figure 13, the proposed method can effectively handle this situation in the process of hazardous chemical vehicle detection.



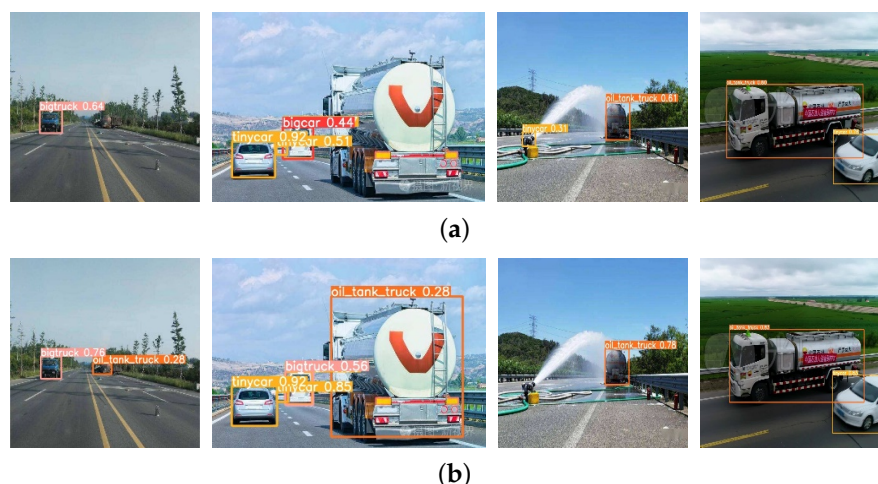
**Figure 13.** A subset of multiple categories vehicle detection results.



### 3.3.4. Comparison of Vehicle Detection Results

In order to compare the effect of the models between the original one and the one we proposed, we also used four sets of images of the scenes to qualitatively evaluate the detection effect of the model. The experimental image size is  $640 \times 640 \times 3$  for all images, with a confidence threshold of 0.25 and an NMS threshold of 0.45. The experimental results are shown in Figure 14.

In the first set of experimental images, the features of the hazardous material vehicles were blurred and small compared to the whole image, which made it difficult for the original model to detect them, but the model we proposed demonstrated remarkable results. In the second set of experimental images, the scene was more complex, with multiple vehicles, dense objects and serious occlusion. The original model missed and misidentified them, and the model we proposed accurately identified and located hazardous material vehicles, proving that the improved model is able to extract richer features. In the third set of experimental images, the original had serious false detections, while the model we proposed not only avoided false detections but also identified the hazardous material vehicles more accurately. In the fourth set of experimental images, although the original model can correctly detect the vehicle, it is not accurate enough in the bounding box regression. The method we proposed makes up for the deficiency of the original model by using the SIoU loss, which takes into account the angling aspect of the regressions and makes the detection more accurate.



**Figure 14.** Comparison of the improved model with the original. (a) Original YOLOv5 experimental result. (b) Improved YOLOv5 experimental results.

Overall, our method performs well in the detection of hazardous material vehicles under different conditions. The accuracy of detection and localization is generally higher than that of the original YOLOv5 model, both in complex scenarios and for the detection of small and multiple vehicles. This proves that the attention mechanism can indeed extract richer semantic information about hazardous material vehicles, and the SPPCSPC layer can improve the network feature fusion capability, while the introduction of SIoU can better improve the accuracy of vehicle localization.

### 3.4. Ablation Study

In order to further validate the detection performance of the algorithm proposed in this paper and to explore the effectiveness of each improved method, we evaluate the improved results step by step by adding new components to the original YOLOv5 model. The experimental results are shown in Table 3.

**Table 3.** Impact of individual components in the development of model.

Group	Attention	SPPCSPC	SIoU	Precision	Recall	F-Score	mAP@0.5	mAP@0.5:0.95
1				0.877	0.779	0.825	0.854	0.631
2	Y			0.882	0.779	0.827	0.864	0.647
3	Y	Y		0.884	0.781	0.829	0.867	0.659
4	Y	Y	Y	0.929	0.757	0.834	0.867	0.661

In this table, Attention represents the attention mechanism added to YOLOv5 in this paper, and SPPCSPC is the module that replaces SPPF (Spatial Pyramid Pooling Fast), while SIoU represents the modification of the intersection over union in the loss function. “Y” represents the introduction of the module. From Table 3, we can see that mAP@0.5 is improved by 1 percent and mAP@0.5:0.95 is improved by 1.6 percent after adding attention mechanism, and the overall performance is improved, although the recall value does not change. The overall performance is improved with the introduction of the SPPCSPC module, which indicates that the module has evident results when compared with SPPF when addressing features of different sizes. After the introduction of SIoU in the loss function, although the recall value decreases, the precision is improved substantially. This indicates that the SIoU loss has a significant effect on the regression of the bounding box. Compared with the original YOLOv5 model, the precision value is improved by 5.2 percent, the F-score improved by 0.9 percent, mAP@0.5 improved by 1.3 percent and mAP@0.5:0.95 improved by 3 percent.

#### 4. Conclusions

Hazardous chemical vehicles are highly dangerous during road transport, and the real-time and accurate detection of these vehicles during the driving process can effectively recognize traffic accidents and respond to them in time to avoid casualties and unnecessary property damage. In this paper, we propose an improved algorithm for detecting hazardous chemical vehicles based on YOLOv5. An attention mechanism is added to the network structure to suppress non-critical information by giving different weights to the feature layers for the purpose of selecting better features. The SPPCSPC layer is used for better feature fusion, and the SIoU loss is introduced to consider the vector angle between bounding box regressions, which effectively reduces the total degrees of freedom of loss. The experimental results show that the method not only exhibits better results for the detection of smaller hazardous chemical vehicles in images, but also achieves the correct recognition and accurate localization of hazardous chemical vehicles in complex scenes.

**Author Contributions:** Methodology, P.Z. and B.C.; validation, P.Z., B.C. and B.L.; data curation, Z.Q.; writing—original draft preparation, B.L. and Z.Q.; writing—review and editing, S.W. and L.W. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported in part by the Humanities and Social Sciences Project of the Ministry of Education of China under grant No. 22YJCZH014, National Natural Science Foundation of China under grant No. 61602202, Natural Science Foundation of Jiangsu Province under contract No. BK20160428 and Natural Science Foundation of Education Department of Jiangsu Province under contract No. 20KJA520008. Six talent peaks project in Jiangsu Province (Grant No. XYDXX-034) and China Scholarship Council also supported this work.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** Thanks are due to Cuiying Yu and Yue Zhao for their valuable discussion and the formatting of this manuscript.

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

**Sample Availability:** Samples of the codes are available from the authors.



## Abbreviations

The following abbreviations are used in this manuscript:

YOLO	You Only Look Once
SPP	Spatial Pyramid Pooling
SPPF	Spatial Pyramid Pooling Fast
SPPCSPC	Spatial Pyramid Pooling Cross-Stage Partial Conv
CFLP	China Federation of Logistics and Purchasing
CSPDarknet	Cross-Stage Partial Darknet
PAN	Path Aggregation Network
ROI	Region of Interest
CNNs	Convolutional Neural Networks
SSD	Single-Shot Multibox Detector
DA	Domain Adaptive
FPN	Feature Pyramid Network

## References

- Hou, J.; Gai, W.; Cheng, W.; Deng, Y. Hazardous chemical leakage accidents and emergency evacuation response from 2009 to 2018 in China: A review. *Saf. Sci.* **2021**, *135*, 105101. [\[CrossRef\]](#)
- Zhou, K.; Xiao, L.; Lin, Y.; Yuan, D.; Wang, J. A Statistical Analysis of Hazardous Chemical Fatalities (HCFs) in China between 2015 and 2021. *Sustainability* **2022**, *14*, 2435. [\[CrossRef\]](#)
- Du, L.; Feng, Y.; Tang, L.; Lu, W.; Kang, W. Time dynamics of emergency response network for hazardous chemical accidents: A case study in China. *J. Clean. Prod.* **2020**, *248*, 119239. [\[CrossRef\]](#)
- Deokjae, L.; Soomi, K.; Jeonghyeon, Y.; Gunil, S.; Byungtae, Y. A study on the improvement plan of transportation plan for safety management of hazardous chemical vehicles. *J. Korean Soc. Hazard Mitig.* **2017**, *17*, 151–157.
- Yao, X.; Zhang, Y.; Yao, Y.; Tian, J.; Yang, C.; Xu, Z.; Guan, Y. Traffic vehicle detection algorithm based on YOLOv3. In Proceedings of the 2021 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS), Xi'an, China, 27–28 March 2021; pp. 47–50.
- Ju, M.; Luo, H.; Wang, Z. An improved YOLO V3 for small vehicles detection in aerial images. In Proceedings of the 2020 3rd International Conference on Algorithms, Computing and Artificial Intelligence, Sanya, China, 24–26 December 2020; pp. 1–5.
- Chen, W.; Baojun, Z.; Linbo, T.; Boya, Z. Small vehicles detection based on UAV. *J. Eng.* **2019**, *2019*, 7894–7897. [\[CrossRef\]](#)
- Bochkovskiy, A.; Wang, C.Y.; Liao, H.Y.M. Yolov4: Optimal speed and accuracy of object detection. *arXiv* **2020**, arXiv:2004.10934.
- Wang, C.Y.; Bochkovskiy, A.; Liao, H.Y.M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. *arXiv* **2022**, arXiv:2207.02696.
- Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. Cbam: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8 September 2018; pp. 3–19.
- Gevorgyan, Z. SIoU loss: More powerful learning for bounding box regression. *arXiv* **2022**, arXiv:2205.12740.
- Djenouri, Y.; Belhadi, A.; Srivastava, G.; Djenouri, D.; Lin, J.C.W. Vehicle detection using improved region convolution neural network for accident prevention in smart roads. *Pattern Recognit. Lett.* **2022**, *158*, 42–47. [\[CrossRef\]](#)
- Wang, Z.; Zhan, J.; Li, Y.; Zhong, Z.; Cao, Z. A new scheme of vehicle detection for severe weather based on multi-sensor fusion. *Measurement* **2022**, *191*, 110737. [\[CrossRef\]](#)
- Dong, X.; Yan, S.; Duan, C. A lightweight vehicles detection network model based on YOLOv5. *Eng. Appl. Artif. Intell.* **2022**, *113*, 104914. [\[CrossRef\]](#)
- Trivedi, J.D.; Mandalapu, S.D.; Dave, D.H. Vision-based Real-time Vehicle Detection and Vehicle Speed Measurement using morphology and binary logical operation. *J. Ind. Inf. Integr.* **2022**, *27*, 100280. [\[CrossRef\]](#)
- Mi, C.; Wang, J.; Mi, W.; Huang, Y.; Zhang, Z.; Yang, Y.; Jiang, J.; Octavian, P. Research on regional clustering and two-stage SVM method for container truck recognition. *Discret. Contin. Dyn. Syst. Ser. S* **2019**, *12*, 1117–1133. [\[CrossRef\]](#)
- Chen, Z.; Guo, H.; Yang, J.; Jiao, H.; Feng, Z.; Chen, L.; Gao, T. Fast vehicle detection algorithm in traffic scene based on improved SSD. *Measurement* **2022**, *201*, 111655. [\[CrossRef\]](#)
- Kang, J.Z.; Wang, G.Z.; He, G.J.; Wang, H.H. Moving vehicle detection for remote sensing satellite video. *J. Remote Sens.* **2020**, *24*, 1099–1107.
- Li, J.; Xu, Z.; Fu, L.; Zhou, X.; Yu, H. Domain adaptation from daytime to nighttime: A situation-sensitive vehicle detection and traffic flow parameter estimation framework. *Transp. Res. Part C Emerg. Technol.* **2021**, *124*, 102946. [\[CrossRef\]](#)
- Alsanad, H.R.; Ucan, O.N.; Ilyas, M.; Khan, A.U.R.; Bayat, O. Real-time fuel truck detection algorithm based on deep convolutional neural network. *IEEE Access* **2020**, *8*, 118808–118817. [\[CrossRef\]](#)
- Zhang, Y.; Liu, D.; Su, C.; Liu, J. Location First Non-Maximum Suppression for Uncovered Muck Truck Detection. *IEICE Trans. Fundam. Electron. Commun. Comput. Sci.* **2022**. [\[CrossRef\]](#)

22. Zhang, J.; Guo, X.; Zhang, C.; Liu, P. A vehicle detection and shadow elimination method based on greyscale information, edge information, and prior knowledge. *Comput. Electr. Eng.* **2021**, *94*, 107366. [[CrossRef](#)]
23. Butt, M.A.; Riaz, F. CARL-D: A vision benchmark suite and large scale dataset for vehicle detection and scene segmentation. *Signal Process. Image Commun.* **2022**, *104*, 116667. [[CrossRef](#)]
24. Zhang, R.; Newsam, S.; Shao, Z.; Huang, X.; Wang, J.; Li, D. Multi-scale adversarial network for vehicle detection in UAV imagery. *ISPRS J. Photogramm. Remote. Sens.* **2021**, *180*, 283–295. [[CrossRef](#)]
25. Punyavathi, G.; Neeladri, M.; Singh, M.K. Vehicle tracking and detection techniques using IoT. *Mater. Today Proc.* **2022**, *51*, 909–913. [[CrossRef](#)]
26. van Ruitenbeek, R.; Bhulai, S. Convolutional Neural Networks for vehicle damage detection. *Mach. Learn. Appl.* **2022**, *9*, 100332. [[CrossRef](#)]
27. Qu, Z.; Gao, L.Y.; Wang, S.Y.; Yin, H.N.; Yi, T.M. An improved YOLOv5 method for large objects detection with multi-scale feature cross-layer fusion network. *Image Vis. Comput.* **2022**, *125*, 104518. [[CrossRef](#)]
28. Wu, W.; Liu, H.; Li, L.; Long, Y.; Wang, X.; Wang, Z.; Li, J.; Chang, Y. Application of local fully Convolutional Neural Network combined with YOLO v5 algorithm in small target detection of remote sensing image. *PLoS ONE* **2021**, *16*, e0259283. [[CrossRef](#)] [[PubMed](#)]
29. Niu, J.; Chen, Y.; Yu, X.; Li, Z.; Gao, H. Data augmentation on defect detection of sanitary ceramics. In Proceedings of the IECON 2020, the 46th Annual Conference of the IEEE Industrial Electronics Society, Singapore, 8–21 October 2020; pp. 5317–5322.
30. Kaur, P.; Khehra, B.S.; Mavi, E.B.S. Data augmentation for object detection: A review. In Proceedings of the 2021 IEEE International Midwest Symposium on Circuits and Systems (MWSCAS), Lansing, MI, USA, 9–11 August 2021; pp. 537–543.
31. Zhu, X.; Lyu, S.; Wang, X.; Zhao, Q. TPH-YOLOv5: Improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 2778–2788.
32. Kasper-Eulaers, M.; Hahn, N.; Berger, S.; Sebulonsen, T.; Myrland, Ø.; Kummervold, P.E. Detecting heavy goods vehicles in rest areas in winter conditions using YOLOv5. *Algorithms* **2021**, *14*, 114. [[CrossRef](#)]

**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.