An object detection method for the work of an unmanned sweeper in a noisy environment on an improved YOLO algorithm

JunLiang Huo
South China University of Technology
BaiJun Shi (bjshi@scut.edu.cn)
South China University of Technology
YiHuai Zhang
Intelligent Transportation Thrust, The HongKong University of Science and Technology(Guangzhou)

Research Article

Keywords: Unmanned sweepers, YOLO-v5, Vibration response, Loss function, Attention mechanism

Posted Date: May 8th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2887763/v1

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Signal, Image and Video Processing on July 5th, 2023. See the published version at https://doi.org/10.1007/s11760-023-02654-4.

Abstract

Efficient and accurate object detection is crucial for the widespread use of low-cost unmanned sweepers. This paper focuses on the low-cost sweeper in practical working scenarios and proposes a traffic participant detection method based on an enhanced YOLO-v5 model. To train the model on noise knowledge, three types of noise were added to the data set based on the mathematical model's vibration response. The loss function was optimized to balance detection accuracy and real-time performance while focusing on traffic participant detection using YOLO-v5. CTDS and BFSA modules were proposed based on the attention mechanism to enhance the YOLO-v5 model. Comparative experiments demonstrated the effectiveness of the proposed method, with the enhanced YOLO-v5 model achieving a 4.5% higher mean average precision than the traditional YOLO-v5 network. Moreover, the proposed method can process images at a frame per second (FPS) of 89 while ensuring real-time performance, meeting the object detection requirements of actual sweeper.

1 Introduction

Resumption of work and production due to policy changes of COVID-19 epidemic, a lot of domestic waste and industrial waste are produced in the production activities in the industrial park. If this waste is not cleaned up in time, it will greatly affect the production activities of enterprises and hinder economic development. As a popular tool at the moment, unmanned sweepers are often applied to major industrial parks.

At present, unmanned sweepers have the advantages of high work efficiency, fast cleaning speed, all-weather work, and reduced labor costs. Therefore, unmanned sweepers have been applied to industrial parks. The market for unmanned sweepers is very broad, but the current unmanned sweepers have a high cost per unit, which increases the operating burden of industrial parks. As a result, some industrial parks cannot purchase unmanned sweepers and can only use higher prices. Hire sanitation workers to clean up. To further promote unmanned sweepers and popularize unmanned sweepers, it is necessary to compress and control the cost of unmanned sweepers. The most important thing for low-cost unmanned sweepers is to use low-cost suspension chassis devices without anti-shake functions camera. To ensure that low-cost unmanned sweepers can also complete work tasks, first of all, the traffic participant detection algorithm in the perception task needs to be optimized.

The motives of the enhanced YOLO-v5 model traffic participant detection method designed based on this paper are as follows: (1) Reduce costs. Since the COVID-19 epidemic has greatly affected the economy, enterprises need to reduce costs in all aspects and also maintain a clean environment in the park. It is also necessary to control costs, so it is very necessary to use low-cost unmanned sweepers. Low-cost unmanned sweepers can utilize affordable suspension chassis devices and cameras lacking anti-shake functionality, but this may introduce various types of noise to the images captured by the cameras, thereby negatively impacting the performance of the object detection task. Using a higher-performance processor for additional filtering will also increase the cost of the sweeper in disguise, which runs counter

to the manufacture of low-cost sweepers. (2) Reduce potential safety hazards. The working environment of unmanned sweepers is mostly industrial parks at night. Too long a cleaning time will lead to various accidents. Therefore, it is necessary to complete the cleaning quickly. If a general processor is used for filtering processing, it will not only fail to meet the real-time detection requirements but also prolong the working hours of the sweepers and increase the likelihood of accidents.

The main contributions of this work can be summarized as follows. Firstly, a novel traffic participant detection method based on YOLO-v5 is proposed for low-cost unmanned sweepers operating in environments with multi-noise interference. Secondly, this method performs data cleaning and noise processing on the BDD100K[1] image data set for the working environment of the sweeper, which avoids the need for additional computing power for filtering when the sweeper is working and uses higher computing power at a higher cost processor. At the same time, it enhances the diversity of the source data set.

The structure of this paper is outlined as follows: Section 2 provides an overview of related research. Section 3 presents a YOLO-based method for detecting traffic participants, as well as a noise-adding processing technique for the data set. In Section 4, we present various experiments on noisy datasets, as well as ablation, comparison, and actual detection experiments. Finally, Section 5 concludes the paper.

2 Related work

This section primarily summarizes the pertinent research content from two perspectives. The first aspect refers to object detection algorithms that rely on manual feature extraction, while the second aspect refers to object detection methods that use artificial intelligence and deep learning.

The first type of object detection involves manually designing features, filtering images to obtain feature maps, and then using classifiers for classification and regression. Traditional object detection algorithms include Harr[2] combined with Adaptive Boosting(Adaboost), Histograms of Oriented Gradients (HOG) combined with Support Vector Machine(SVM), Deformable Part Model(DPM). Matsumoto[3] proposed to use Self-Quotient Epsilon-Filter (SQEF) and HOG Combine to extract features, to realize that for images with illumination changes and damaged by noise, features can still be extracted, and more robust object detection is achieved from noise-damaged images, but if the lighting conditions change sharply, it will lead to HOG extraction performance deteriorates. Moubtahij et al.[4] describe a detection strategy that relies on the Adaboost algorithm and a Polynomial Image Decomposition (PID) technique. By applying PID decomposition, they effectively reduce noise and turbidity in underwater images, leading to improved visibility of objects in aquatic environments. However, since underwater noise is significantly different from conventional noise, the approach may be challenging to generalize to other types of scenes. Ali et al.[5] propose a rapid category-independent object detector that utilizes integrated modules to accelerate DPM steps. However, the real-time performance of this method is highly dependent on computing power, and it may not be feasible to achieve real-time performance when applied to a standard unmanned sweeper hardware system. To sum up, the traditional object detection method achieves specific problem-

specific analysis through the manual design of features. When the imaging quality is very good and the object quality is very high, the detection accuracy is sufficient, but the manual design of features is very difficult, and the designed features can generally only be based on prior knowledge, and it is not conducive to the current industrial development with an increasing degree of automation[6], and is less robust to different data.

The rapid development of GPUs and their high computing power has led to the emergence of deep learning in various fields. At present, the mainstream object detection methods include single step detection and two-step detection. two-step detection include: R-CNN[7], Fast-RCNN[8], Faster-RCNN[9], Mask-RCNN[10] these series of algorithms, Then, based on the two-stage classical algorithm, many new methods appeared in dealing with different noise situations. Ruxin et al.[11] propose a model. Their approach effectively removes noise from the image, but it requires high computing memory, which further exacerbates the drawbacks of the two-stage algorithm. Wang, Zhou[12] tackle the issue of poor detection caused by background environment noise by leveraging multi-layer feature fusion technology and extracting difficult cases. This approach effectively addresses the problem. However, further improvements are required for the categories in the data set, since this method necessitates extensive data-driven model learning. Lai et al.[13] propose a module in this paper, which is integrated into the object detection network based on the Mask R-CNN model. This approach enables the model to detect objects even in low lighting conditions. If real-time performance is not considered, the average accuracy (MAP) of the two-step algorithm can theoretically reach very high, but real-time performance cannot be ignored in practical applications.

The single-stage algorithm consists of Single Shot MultiBox Detector(SSD)[14], and You Only Look Once(YOLO). Since the YOLO-v1[15] coordinates of the anchor frame are directly predicted, the detection accuracy is very low, and there is a great prospect for improvement. YOLO-v2[16] based on YOLO-v1, the overall performance is still not satisfactory. YOLO-9000[17], YOLO-v3[18], YOLO-v4[19], YOLO-v5[20], YOLOX[21], YOLOP[22], YOLO-v7[23], YOLO-v5 achieves a good balance between accuracy and real-time performance, and it is widely used in industry, so there are a lot of available interfaces, which are convenient for maintenance and upgrades. Therefore, this paper chooses to improve based on the YOLO-v5 algorithm to realize object detection task for traffic participants of unmanned sweeper.

In a park setting characterized by bumpy roads and poor lighting conditions, achieving object detection with a satisfactory level of accuracy while meeting real-time detection requirements is challenging. Considering the advantages of the YOLO-v5 method in object detection and low memory usage. We propose an enhanced YOLO-v5 algorithm for object detection of traffic participants in park environments with bumpy roads and poor lighting conditions. First, establish the data set needed for this method, establish the four-degree-of-freedom model of the unmanned sweeper, add motion blur noise to the BDD100K data set according to the vibration response and the occurrence probability of bumpy road sections, and combine the working hours and working scenes of the unmanned sweeper, adding pepper noise to the BDD100K data set, combined with the distribution of random noise, adding Gaussian noise to the BDD100K data set. Then, the YOLO-v5s model was optimized and improved to perform well in

object detection for traffic participants in environments with rough roads and poor lighting conditions. Based on the characteristics of traffic participants, select the parameters of the model and generate an enhanced YOLO-v5 model to achieve the improvement effect and achieve the environmental perception task of unmanned sweeper. Finally, using a noisy test set, compare the accuracy and speed of YOLO-v3-tiny, YOLO-v5, YOLO-v7, and YOLO-P models with object detection. From the experimental results, the enhanced YOLO-v5 model achieves the highest MAP value, that is, the highest object detection accuracy, and provides a high-quality method reference for unmanned sweeper perception tasks in other working environments.

3 Method

3.1 Overall research approach

Aiming to meet the real-time and accuracy requirements for unmanned sweepers operating in environments with three different types of noise, an enhanced YOLO-v5 model is proposed for the unmanned sweeper object detection method, which is used for the vision of vehicle-mounted cameras object detection task. The Framework of object detection methods for unmanned sweeper in noisy environments is depicted in Fig. 1.

It mainly includes mathematical modeling of the sweeper and camera and solution of vibration response, quantitative expansion and modification of data set according to noise type, model training, and sweeper object detection based on enhanced YOLO-v5 model. First, use the Simulink simulator to mathematically model the sweeper and the camera with four degrees of freedom, and solve the noise corresponding to the vibration response. Then, according to the light and dark conditions of the actual working environment and common Gaussian noise, perform three noise ordered additions to the data set. The YOLO-v5 model was improved using the module proposed in this paper, and the expanded data set was utilized to train the enhanced YOLO-v5 model.

3.2 Traffic participant detection based on YOLO

Currently, the YOLO-v5 algorithm is widely used due to its good performance, as well as its versatility for industrial applications. The algorithm supports various industrial interfaces and can be easily maintained and updated.

3.2.1 YOLO-v5

The YOLO-v5 model consists of three modules: Backbone, Neck, and Head. Compared with YOLO-v4, feature fusion only performs fusion from small-scale feature maps to large-scale feature maps. The author also creatively introduced the PAN structure[24], It makes the fusion information richer and the feature fusion effect is better.

Backbone is composed of 5 ConvBlock_1 layers, 4 ConvBlock_4 layers, and 1 SPPF layer. The YOLO-v5 in its original form is depicted in Fig. 2.

3.2.2 Improvements proposed for YOLO-v5

Due to the real-time detection performance of YOLO-v5, it can meet the real-time requirements, it is considered a lightweight model within the YOLO-v5 series. Thus, we choose to enhance the performance of the network based on YOLO-v5s without compromising its real-time detection capabilities. The improvements are as follows:

(1)Propose CTDS module and embed it in feature extraction backbone network, combining Convolutional Block Attention Module(CBAM)[25] and Transformer attention mechanism[26], while improving the receptive field, the network emphasizes global information features and suppresses the excessive noise information to enhance the useful information, resulting in improved feature map quality. The large parameter problem caused by the mechanism, replaces convolution with Depthwise Separable Convolution[27].

(2)Propose BFSA module and embed it in feature fusion network. In the BiFusion Neck[28], we embed Stand-Alone Self-Attention[29], Change the up-sampled method to a transposed convolution with learnable parameters[30], Increasing the number of learnable parameters in the feature layer can improve the generalization of the network.

(3)Add a network decoupling module to embed it in the classification head to solve the coupling problem of classification and positioning problems[21].The structure of the enhanced YOLO-v5 model is depicted in Fig. 3.

(4)The Wise Intersection Over Union(WIOU) calculation formula is used to calculate the loss. Due to the blur problem of the noise image, the penalty of the model due to the influence of geometric factors is unreasonable. Using the WIOU calculation formula instead of the Complete Intersection over Union(CIOU) calculation formula can solve this problem in the model[31]. WIOU calculation formulas are shown in Eq. (1), Eq. (2).

$$L_{WIoUv1} = R_{WIoU} L_{IoU}$$

1

$$R_{WIoU} = \exp\left(\frac{(x - x_{gt})^2 + (y - y_{gt})^2}{W_g^2 + H_g^2}\right)$$

2

W_g, H_g represent the width and height of the minimum bounding box.

3.3 Image noise addition

Affected by low-cost suspension devices, cameras without anti-shake function, and working environment, the captured images mainly have three types of noise: motion blur noise, pepper noise, and Gaussian noise. If the influence of noise is eliminated in the image preprocessing stage, it will cause a high computing power burden. The computing power of unmanned sweeper is very limited, and sufficient computing power needs to be reserved for subsequent decision-making control units. Therefore, the network must learn about noise, it can also complete the detection task of traffic participants under the influence of noise. By increasing the training cost, the computing power burden of unmanned sweeper can be reduced. In this section, noise processing will be performed on the original data set.

Firstly, a four-degree-of-freedom mathematical model is established based on the sweeper and the camera. The mathematical modeling is illustrated in Fig. 4. The camera is installed on the vertical line of the center of gravity of the sweeper and is consolidated with the body, so the vibration response of the camera is consistent with the vibration response of the sweeper's center of gravity in the vertical direction. According to the analysis of vibration mechanics, the vibration response x_0 can be obtained from Eq. (3), Adding motion noise based on vibration response.

$$x_0(t) = \frac{z_1}{\left[\frac{z_2(k_5 - m_2w^2)}{z_2 + k_5 - m_2w^2} + \frac{z_3(-m_2w^2)}{z_3 - m_3w^2} - mw^2\right]} x_1$$

3

where z is the partial impedance of each degree of freedom.

We add two types of noise to the data set based on the working environment: salt-and-pepper noise and Gaussian noise. The Gaussian noise conforms to the Gaussian distribution, and the formula for the Gaussian distribution is shown below Eq. (4).

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

4

x represents the gray value of the feature point.

4 Experiment

4.1 Experimental configuration and training parameter settings

Table 1 shows the experimental environment set up for ensuring the efficient training of the enhanced YOLO-v5 model. The relevant training parameters settings : the initial learning rate is 0.001, the batch size is 8, and the Epoch is 100.

Table 1
Experiment environment

The operating system	Ubuntu18.04
CPU	12th Gen Intel Core i5-12400F
GPU	NVIDIA GeForce RTX 3050(8GB)
Memory	32GB
Deep learning platform	Pytorch 1.10.0

4.2 Performance evaluation index

To consider its practical application on the sweeper, and to evaluate the model's performance, we adopts commonly used performance evaluation metrics in object detection, including precision, recall, MAP, and FPS.

Precision, it is usually understood as the query accuracy, and can be calculated using Eq. (5).

$$\Pr ecision = \frac{TP}{TP + FP} \times 100\%$$

5

Recall, it is usually understood as the query completion rate, and can be calculated using Eq. (6).

$$\mathrm{Re}\, call = \frac{TP}{TP + FN} \times 100\%$$

6

AP is a graph drawn using precision as the ordinate and recalls as the abscissa. MAP provides an overall evaluation of the model's detection performance. The AP value for a specific object category i is denoted as AP_i, where i is the index of the category to be detected among k total categories. The formula to calculate AP_i is shown below Eq. (7)

$$mAP = \frac{\sum_{i=1}^{n} AP_i}{k}$$

7

We avoid using accuracy as an evaluation metric, as the proportion of samples in the BDD100K data set may vary across different object categories. TP means true positive number, FP means false positive number, FN means false negative number. FPS is a model speed performance parameter.

4.3 Image noise addition

By noise processing the pictures of the BDD100K data set, images similar to those collected in the actual work of the sweeper can be obtained. This section shows the processing effect of three types of noise. For motion noise, the horizontal displacement and vertical displacement are considered according to the vibration response. For and pepper noise, the gray value threshold is set to simulate the collected image in a dark environment. For Gaussian noise, the mean value is set to 0 and the variance is set to 50. Figure 5 illustrates the actual effect of processing images with noise, (a) represents the original image without any added noise, (b) represents the image after motion blur processing, (c) represents the image after pepper noise processing, (d) represents the image after Gaussian noise processing. By applying three types of noise to the original image, the resulting image can approximate a realistically collected image.

4.4 Data set construction

The BDD100K data set includes multiple labels such as traffic signs, road segmentation, and traffic participants. As traffic signs vary across regions, and road segmentation labels are not needed for this paper, the data set is cleaned and noise is added. Then, the training, validation, and test sets are re-divided to contain only objects as follow: cars, buses, pedestrians, trucks, bicycles, motorcycles, and cyclists. Table 2 displays the partitioned data set.

Table 2
Data set partition

Data set Type	image size	numbers
train	1280*720	55529
val	1280*720	7932
test	1280*720	15865

4.5 Verify the necessity of adding noise

To test the impact of noise on the model and evaluate whether the noise added in this article is effective, we conducted the following experimental verification. We trained the unimproved YOLO-v5s on the original noise-free data set and evaluated its performance on both the noisy test set and the noise-free test set. We randomly added the three types of noise discussed earlier to the noisy test set and evaluated the model's performance. As shown in Table 3, the results demonstrate a significant drop in the MAP and recall metrics, indicating that noise in the data set can lead to a decrease in the model's generalization ability. To assess the effectiveness of adding noise to the training set, we trained the unimproved YOLO-v5s on the noisy test set after adding all three types of noise to the training set. The results show that the MAP increased by 1.4% and recall increased by 1.2% compared to the original unimproved YOLO-v5s trained on the noise-free data set. In order to teach the model to learn from noisy data, we deliberately introduce noise information into the training set. Improving the generalization ability of the model is feasible. We utilize a noisy data set that includes three different types of noise, added in varying

proportions, to enhance the model's robustness when confronted with these types of noise simultaneously.

Table 3
Noise verification experiment results

Model	MAP50	Recall
A	77.2	86.8
B	52.3	46.3
C	53.7	47.5

A represents training an unimproved YOLO-v5s on a training set without noise, and test it on a test set without noise, B represents training an unimproved YOLO-v5s on a training set without noise, and test it on a test set with noise, C represents training an unimproved YOLO-v5s on a training set with noise, and test it on a test set with noise.

4.6 Ablation Experiment

In order to verify the effectiveness of the improvement module proposed in our paper for the YOLO-v5s network, we conducted ablation experiments on the noisy BDD100K data set to test the impact of individual improvement methods, as well as the fusion of various methods, on the performance of the model's main modification modules. Table 4 presents the comparison results obtained from the ablation experiments. The results indicate that the CTDS module, BFSA module, and decoupling head module proposed in our paper can effectively improve the performance of the model to varying degrees. The effectiveness of WIOU in improving the performance of object detection models has been demonstrated in the literature. To evaluate the impact of WIOU on the enhanced YOLO-v5 model, we conducted an ablation experiment, and the results are presented in Table 5. The utilization of WIOU not only enhances the performance metrics of the model but also improves its real-time detection capabilities.

Table 4
Ablation Experiment Results 1

Decoupling head module	CTDS module	BFSA module	Precision	Recall	MAP50	FPS
√			0.702	0.477	0.541	142
	√		0.721	0.453	0.540	151
		√	0.719	0.465	0.542	156
√	√		0.701	0.49	0.552	108
√		√	0.710	0.5	0.561	112
	√	√	0.710	0.495	0.557	116
√	√	√	0.730	0.514	0.582	89

Table 5
Ablation Experiment Results 2

Model	Precision	Recall	MAP50	FPS
enhanced YOLO-v5 (with WIOU)	0.730	0.514	0.582	89
enhanced YOLO-v5 (with CIOU)	0.722	0.508	0.574	86

4.7 Overall data set detection results

We select YOLO-v3, YOLO-v5, YOLO-v7, YOLO-P, and the enhanced YOLO-v5 model to do comparative experiments on the noisy BDD100K data set, among which YOLO-v3 and YOLO-v5 are often used as traffic participants. For object detection tasks, YOLO-v7 is a relatively new model with powerful performance. YOLO-P once performed best on the noise-free BDD100K data set. The comparison experiment is shown in Table 6. The enhanced YOLO-v5 model outperforms other detection algorithms in terms of MAP, achieving a MAP of 58.2% which is 4.5 percentage points higher than the unimproved YOLO-v5s model. Additionally, the model's FPS rate is also the highest among the tested models, satisfying the real-time monitoring requirements. Although YOLO-P has the highest MAP, because its Precision is too low, it is easy to cause object error detection, resulting in serious errors in the subsequent decision-making control process, so it cannot be used. which ensured the experimental theoretical basis of the model in practical application.

Table 6
Comparative test results

Model	Pictures	Labels	Precision	Recall	MAP50	FPS
YOLO-v3	15865	195233	0.551	0.355	0.356	52
YOLO-v5	15865	195233	0.709	0.475	0.537	217
YOLO-v7	15865	195233	0.6845	0.5042	0.5528	72
YOLO-P	15865	195233	0.045	0.875	0.712	75
Enhanced YOLO-v5	15865	195233	0.73	0.514	0.582	89

The experimental results demonstrate that the enhanced YOLO-v5 model can efficiently perform object detection of traffic participants under real-world conditions, including noise interference caused by vibration and insufficient lighting. The test results are presented in Fig. 6.

The enhanced YOLO-v5 algorithm for detecting traffic participants enables real-time detection using low-cost suspended devices, cameras without anti-shake functions, and in low-light conditions. Traditional object detection algorithms often perform poorly in complex environments with high levels of noise interference, whereas deep learning algorithms have demonstrated superior performance in such scenarios. When compared to the unimproved YOLO-v5 and other commonly used object detection models, the enhanced YOLO-v5 model shows a significant improvement in performance while maintaining real-time capabilities. The data set after adding noise is complete and can increase the robustness of the model, which can ensure that the sweeper can realize real-time traffic participant object detection under low-cost suspension devices, cameras without anti-shake function, and low light conditions with certain accuracy.

5 Conclusion

In practical applications, high-cost unmanned sweeper will increase the business burden of enterprises or companies, and it is very necessary to use low-cost unmanned sweeper. The most direct way to realize a low-cost unmanned sweeper is to use a low-cost suspension device and a camera without an anti-shake function, which will also bring three different types of noise impact, resulting in loopholes and defects in the object detection task of the sweeper. The algorithmic improvements and data set processing presented in this paper are of great significance for achieving low-cost unmanned sweeper.

The main contributions of our paper can be summarized as follows. First of all, In this paper, we propose three types of noise processing methods for the BDD100K data set, in which motion blur noise is added according to the mathematical model, pepper noise, Gaussian noise are configured according to the field environment parameters so that the data set obtains a large amount of noise information, which assisted in subsequent model training. Secondly, this paper proposes two modules: the CTDS module and the

BFSA module, which are used to embed the traditional YOLO-v5s model and use WIOU instead of CIOU to realize the enhanced YOLO-v5 model.

Subsequent research can focus on the following two points. First of all, it is essential to augment the original data set by incorporating real-world noise and expanding its scope to match the intended application scenario. Last but not least, the improved model in this paper cannot be applied to smaller-scale processors. How to further compress the model volume without affecting performance will be the next challenge.

Declarations

Data availability

Due to the nature of this study and in order to protect the privacy of study participants, participants of this study did not agree for their data to be shared publicly, so supporting data are not available.

Conflict of interest :The authors declare that there is no conflict of interest

Funding:The authors did not receive support from any organization for the submitted work.

Author Contributions: Material preparation, data collection, analysis and modification, experiment were performed by JunLiang Huo, the first draft of the manuscript was written by JunLiang Huo, All authors read and approved the final manuscript.

References

1. Yu, F., Chen, H., Wang, X., Xian, W., Chen, Y., Liu, F., Madhavan, V., Darrell, T.: BDD100K: A Diverse Driving Dataset for Heterogeneous Multitask Learning.In:2020 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp 2633-2642, (2020).https://doi.org/10.1109/CVPR42600.2020.00271
2. Papageorgiou, C.P., Oren, M., Poggio, T.: A General Framework for Object Detection.In:1998 6th International Conference on Computer Vision(ICCV), pp 555-562, (1998).https://doi.org/10.1109/ICCV.1998.710772
3. Matsumoto, M.: SVM-based Parameter Setting of Self-quotient -Filter and Its Application to Noise Robust Human Detection.In:2011 3rd International Conference on Agents and Artificial Intelligence(ICAART), pp 290-295,(2011).
4. Moubtahij, R.E., Merad, D., Damoisaux, J.L., Drap, P.: Mine Detection Based on Adaboost and Polynomial Image Decomposition.In:2017 19th International Conference on Image Analysis and Processing (ICIAP), pp 660-670,(2017).https://doi.org/10.1007/978-3-319-68560-1_59
5. Ali, A., Olaleye, O.G., Bayoumi, M.: Fast region-based DPM object detection for autonomous vehicles.In:2016 59th IEEE International Midwest Symposium on Circuits and Systems (MWSCAS),

pp 691-694,(2016).

6. Zou, Z., Shi, Z., Guo, Y., Ye, J.: Object Detection in 20 Years: A Survey. P IEEE, 257-276 (2019). https://doi.org/10.48550/arXiv.1905.05055
7. Girshick, R., Donahue, J., Darrell, T., Malik, J.: Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation.In:2014 27th IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp 580-587,(2014).https://doi.org/10.1109/CVPR.2014.81
8. Girshick, R.: Fast R-CNN.In:2015 IEEE International Conference on Computer Vision(ICCV), pp 1440-1448,(2015).https://doi.org/10.1109/ICCV.2015.169
9. Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE Transactions on Pattern Analysis & Machine Intelligence. 39(6), 1137-1149 (2017). https://doi.org/10.1109/TPAMI.2016.2577031
10. He, K., Gkioxari, G., Dollar, P., Girshick, R.: Mask R-CNN. IEEE T PATTERN ANAL. 42(2), 386-397 (2017). https://doi.org/10.1109/TPAMI.2018.2844175
11. Ruxin, Wang, Dacheng, Tao: Training Very Deep CNNs for General Non-Blind Deconvolution. IEEE T IMAGE PROCESS. 27(6), 2897-2910 (2018). https://doi.org/10.1109/TIP.2018.2815084
12. Wang, K., Zhou, W.: Pedestrian and cyclist detection based on deep neural network fast R-CNN. INT J ADV ROBOT SYST. 16(2), (2019). https://doi.org/10.1177/1729881419829651
13. Lai, K.C., Zhao, J., Liu, D.J., Huang, X.N., Wang, L.: Research on pedestrian detection using optimized mask R-CNN algorithm in low-light road environment. Journal of Physics: Conference Series. 1777(1), 12057 (2021). https://doi.org/10.1088/1742-6596/1777/1/012057
14. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C., Berg, A.C.: SSD: Single Shot MultiBox Detector.In:2016 14th European Conference Computer Vision(ECCV), pp 21-37, (2016).https://doi.org/10.1007/978-3-319-46448-0_2
15. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You Only Look Once: Unified, Real-Time Object Detection.In:2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp 779-788, (2016).https://doi.org/10.1109/CVPR.2016.91
16. Nakahara, H., Yonekawa, H., Fujii, T., Sato, S.: A Lightweight YOLOv2: A Binarized CNN with A Parallel Support Vector Regression for an FPGA.In:2018 ACM/SIGDA International Symposium on Field-Programmable Gate Arrays (FPGA), pp 31-40,(2018).https://doi.org/10.1145/3174243.3174266
17. Redmon, J., Farhadi, A.: YOLO9000: Better, Faster, Stronger.In:2017 30th IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp 6517-6525, (2017).https://doi.org/10.1109/CVPR.2017.690
18. Qu, H.Q., Yuan, T.Y., Sheng, Z.Y., Zhang, Y.: A Pedestrian detection method Based on YOLOv3 model and Image enhanced by Retinex.In:2018 11th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI). (2018).
19. Bochkovskiy, A., Wang, C.Y., Liao, H.: YOLOv4: Optimal Speed and Accuracy of Object Detection. (2020). https://doi.org/10.48550/arXiv.2004.10934

20. Mekhalfi, M.L., Nicolo, C., Bazi, Y., Rahhal, M., Alsharif, N.A., Maghayreh, E.A.: Contrasting YOLOv5, Transformer, and EfficientDet Detectors for Crop Circle Detection in Desert. IEEE GEOSCI REMOTE S. 19, (2022). https://doi.org/10.1109/LGRS.2021.3085139
21. Ge, Z., Liu, S., Wang, F., Li, Z., Sun, J.: YOLOX: Exceeding YOLO Series in 2021. (2021). https://doi.org/10.48550/arXiv.2107.08430
22. Wu, D., Liao, M.W., Zhang, W.T., Wang, X.G., Bai, X., Cheng, W.Q., Liu, W.Y.: YOLOP:You Only Look Once for Panoptic Driving Perception. 19(6), 13 (2022). https://doi.org/arXiv:2108.11250
23. Wang, C.Y., Bochkovskiy, A., Liao, H.: YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. (2022). https://doi.org/10.48550/arXiv.2207.02696
24. Liu, S., Qi, L., Qin, H., Shi, J., Jia, J.: Path Aggregation Network for Instance Segmentation.In:2018 31st IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp 8759-8768, (2018).https://doi.org/10.1109/CVPR.2018.00913
25. Woo, S.H., Park, J., Lee, J.Y., Kweon, I.S.: CBAM: Convolutional Block Attention Module.In:2018 15th European Conference on Computer Vision (ECCV), pp 3-19,(2018).https://doi.org/10.1007/978-3-030-01234-2_1
26. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Houlsby, N.: An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale, (2020). https://doi.org/10.48550/arXiv.2010.11929
27. Huang, G.J., Zhang, Y.L., Ou, J.Y.: Transfer remaining useful life estimation of bearing using depth-wise separable convolution recurrent network. MEASUREMENT. 176, (2021). https://doi.org/10.1016/j.measurement.2021.109090
28. Yung, N., Wong, W.K., Juwono, F.H., Sim, Z.A., IEEE: Safety Helmet Detection Using Deep Learning: Implementation and Comparative Study Using YOLOv5, YOLOv6, and YOLOv7.In:2022 International Conference on Green Energy, Computing and Sustainable Technology (GECOST), pp 164-170, (2022).https://doi.org/10.1109/GECOST55694.2022.10010490
29. Ramachandran, P., Parmar, N., Vaswani, A., Bello, I., Levskaya, A., Shlens, J.: Stand-Alone Self-Attention in Vision Models.In:2019 33rd Conference on Neural Information Processing Systems (NeurIPS). (2019).
30. Dumoulin, V., Visin, F.: A guide to convolution arithmetic for deep learning. (2016). https://doi.org/10.48550/arXiv.1603.07285
31. Cho, Y.J.: Weighted Intersection over Union (wIoU): A New Evaluation Metric for Image Segmentation. (2021). https://doi.org/10.48550/arXiv.2107.09858

Figures

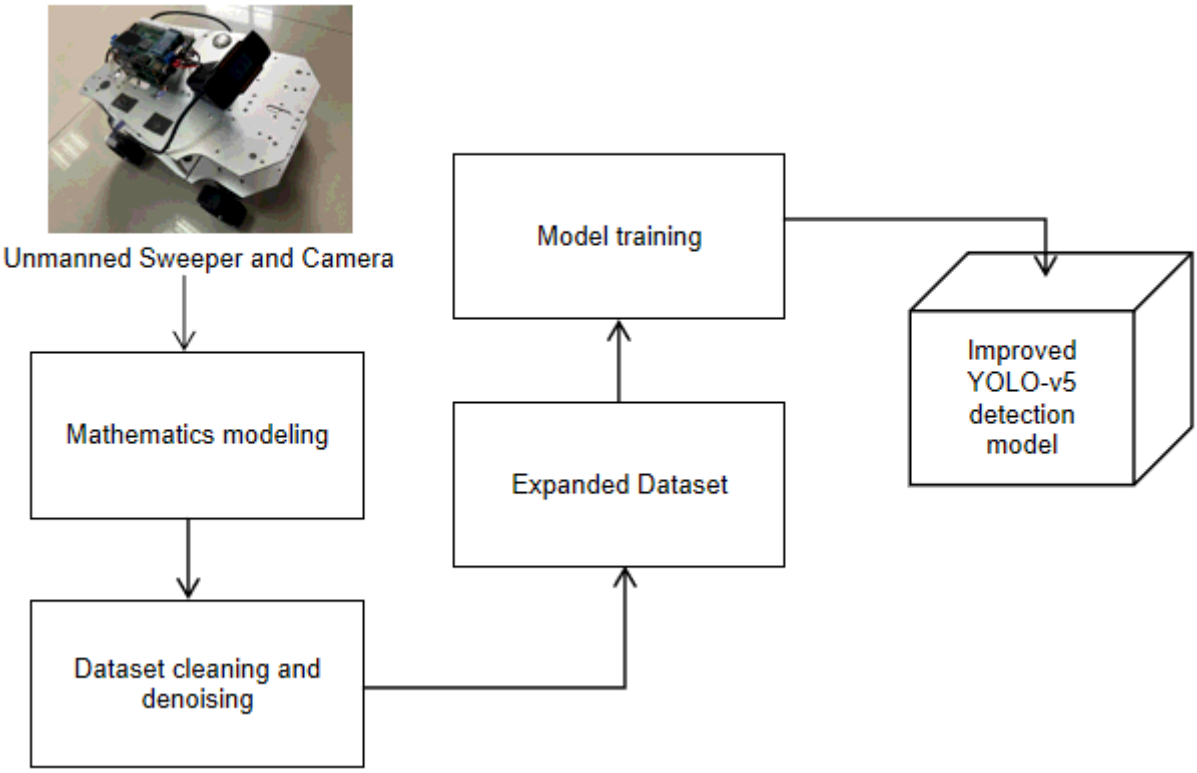

Figure 1

Framework of object detection methods for unmanned cleaning vehicles in noisy environments

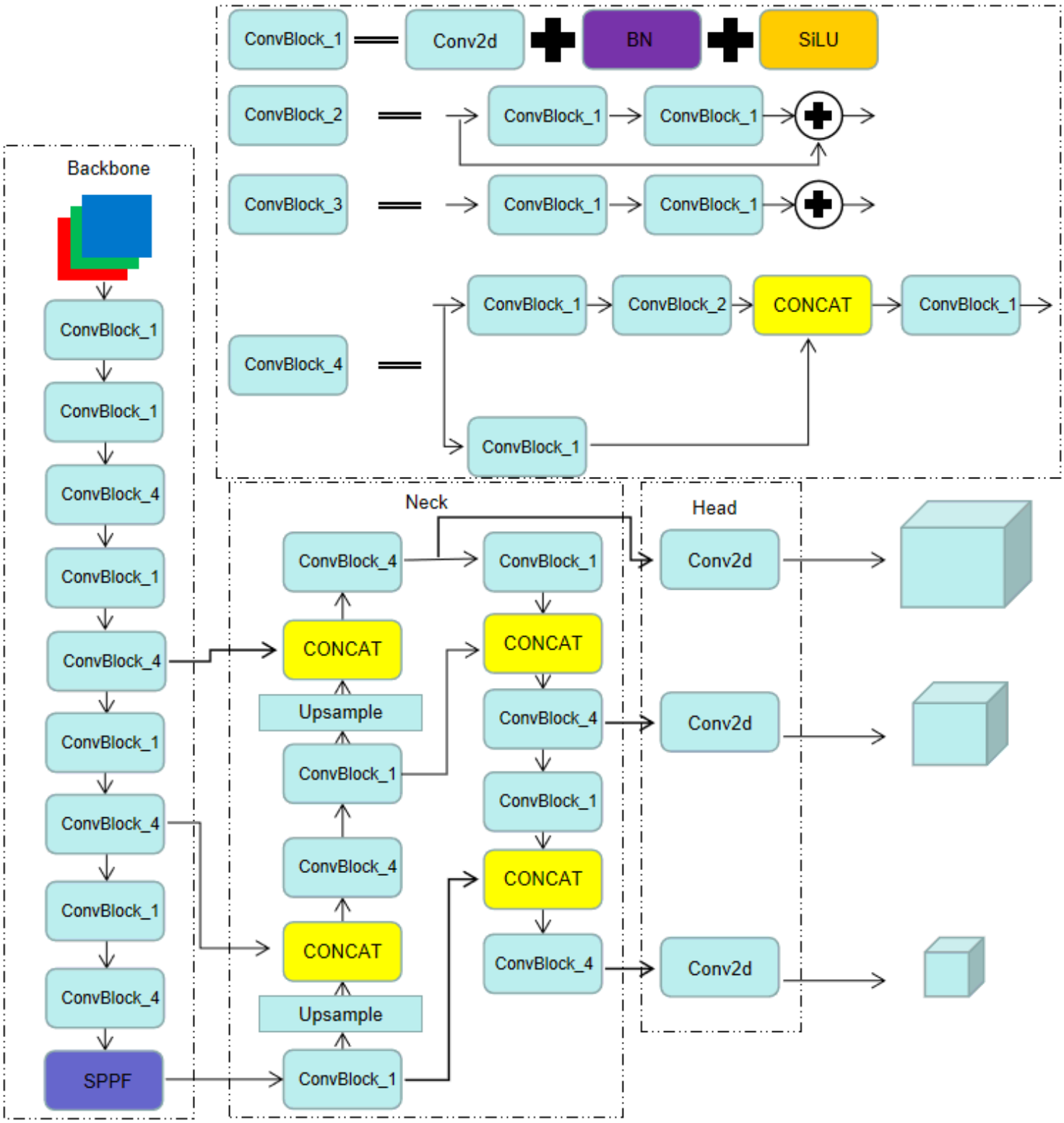

Figure 2

Original YOLO-v5 network structure is shown

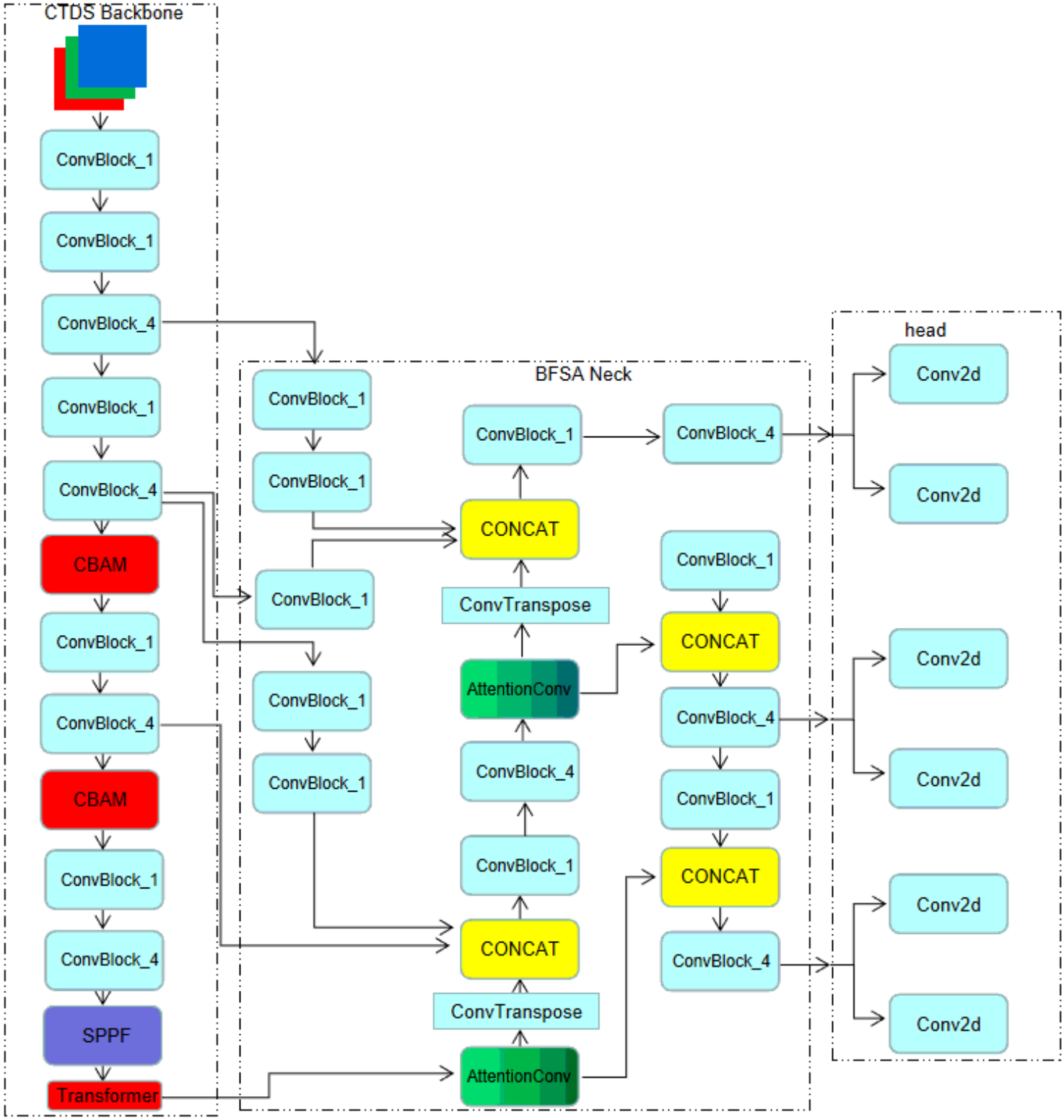

Figure 3

enhanced YOLO-v5 model structure

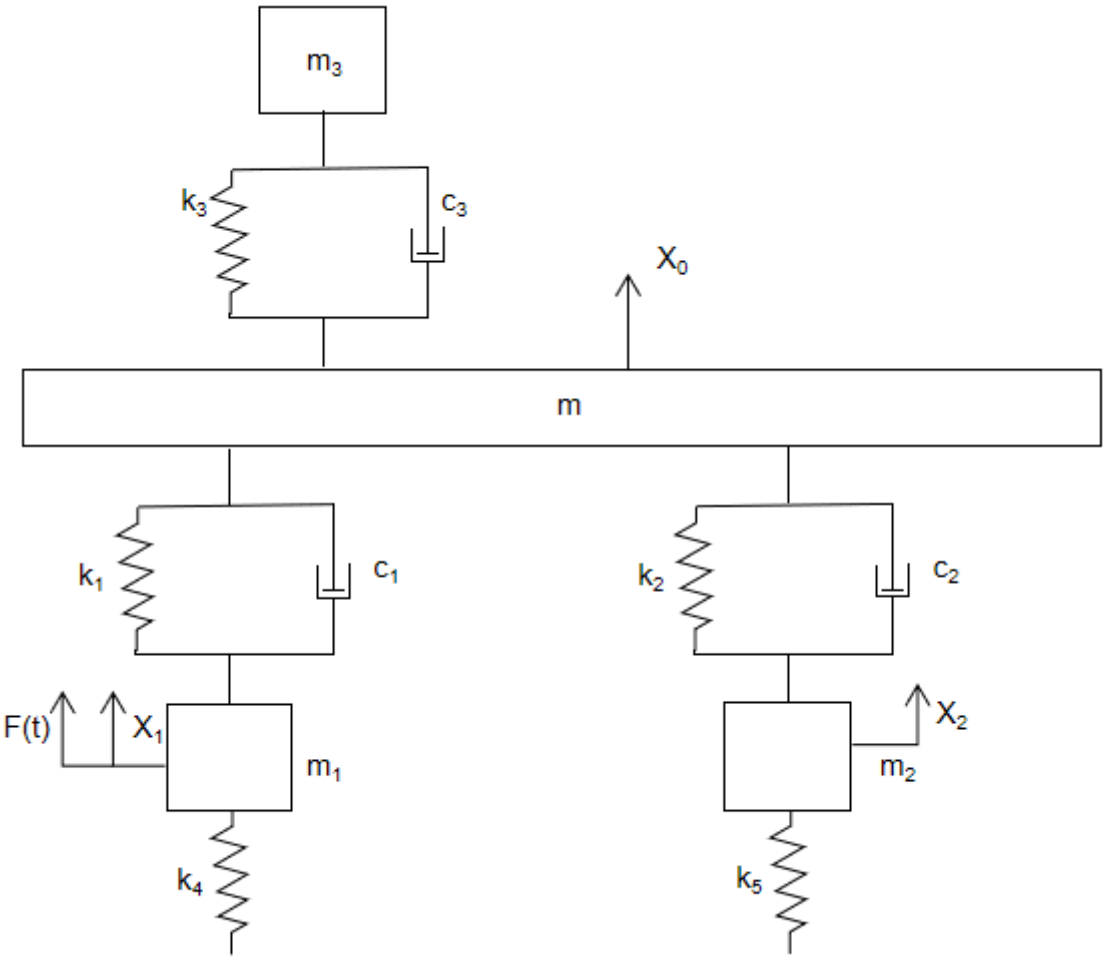

Figure 4

Mathematical modeling

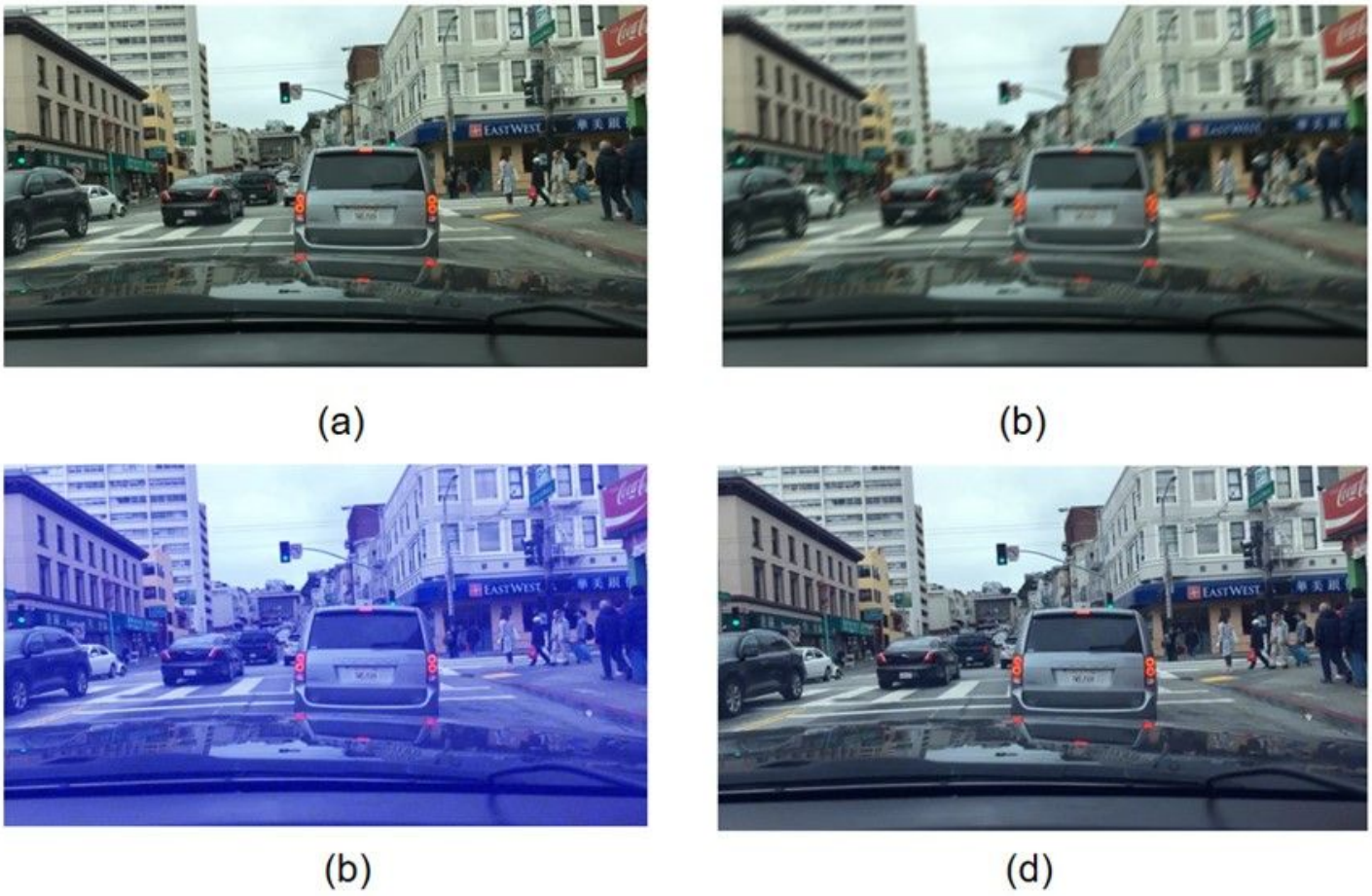

Figure 5

The actual effect of noise processing pictures

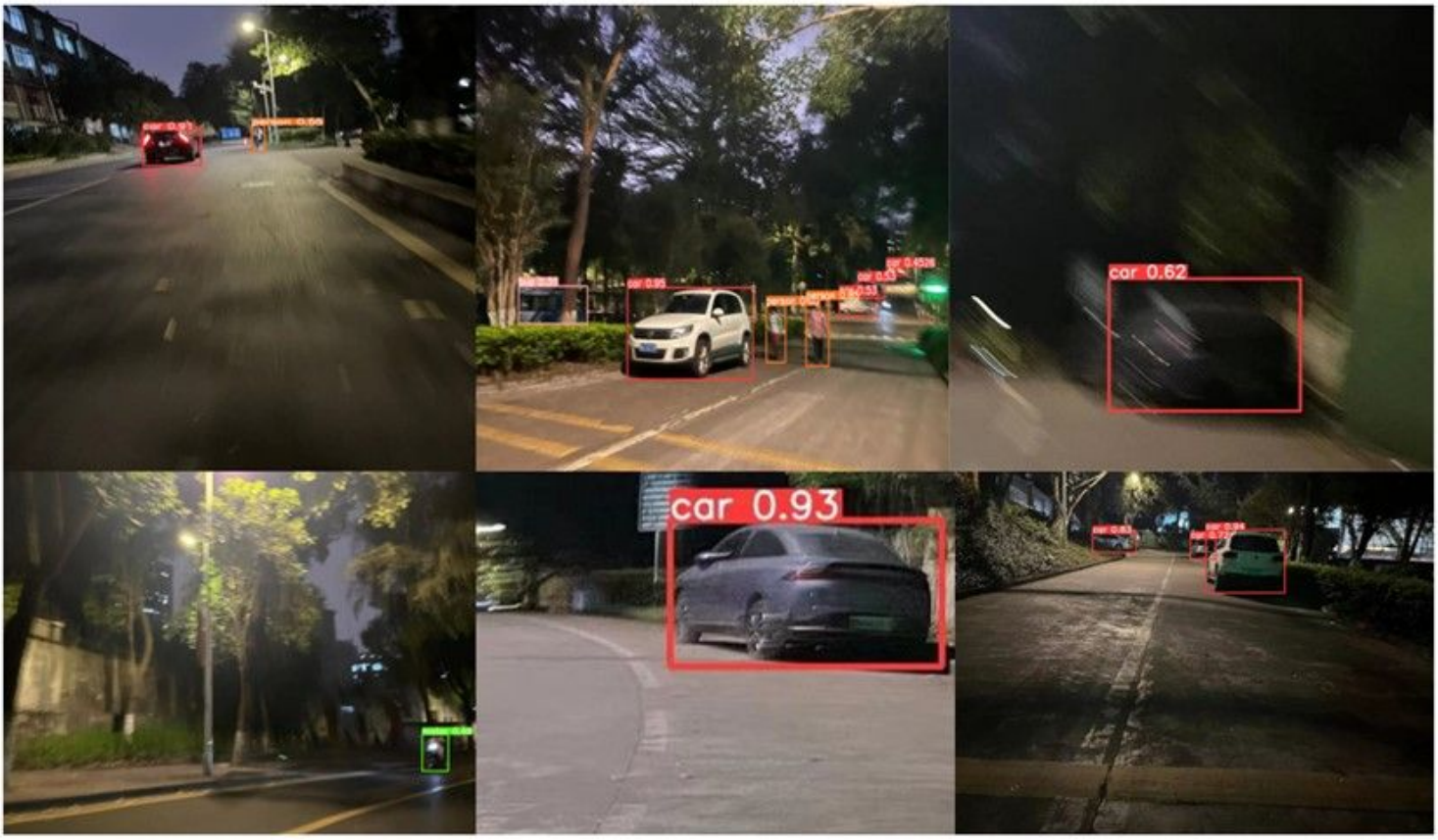

Figure 6

Actual work effect