CBAM-Enhanced YOLOv5 for Automated Detection of Urban Underground Drainage Pipe Defects

Ying He
College of Transportation and Municipal Engineering, Inner Mongolia Technical University of Construction Hohhot, Inner Mongolia 010010, China
E-mail: heyinghy@hotmail.com

Keywords: deep learning, defect and damage, drainage pipe, YOLOv5, loss function

Received: September 18, 2025

The timely detection and repair of defects and damages in underground drainage pipes are crucial for the normal operation of cities. Focusing on the defect detection and damage localization of urban underground drainage pipes, this paper introduced the Convolutional Block Attention Module (CBAM) to the You Only Look Once version 5 (YOLOv5) algorithm to enhance its ability of feature extraction. Then, several different loss functions were compared. Experimental analyses were carried out using the sewer-ML dataset. The results showed that among different versions of the model, the YOLOv5l model had better overall performance. Compared with the Squeeze-and-Excitation and coordinate attention modules, the CBAM had a better optimization effect on the YOLOv5 algorithm, bringing a 5.7% mean average precision improvement. The detection effect obtained when Softmax Intersection over Union (SIoU) was used as the loss function was better than efficient Intersection over Union (EIoU) and Focal EIoU. When CBAM and SIoU were used for optimization together, the improved YOLOv5 algorithm achieved a mean average precision of 93.37% and a frame rate of 85 frames per second, which had an advantage over the other algorithms. The method can be used in practice.

Povzetek: Izboljšani YOLOv5 z modulom CBAM natančno in hitro zazna poškodbe v kanalizacijskih ceveh (93,37% natančnost, 85 FPS).

1 Introduction

Underground drainage pipes play a very important role in the collection and transportation of rainwater and sewage. During long-term operation, defects and damages such as rupture and deposition are inevitable [1], which pose certain hidden dangers for urban development and may cause disasters such as waterlogging and collapse. Therefore, regular inspection and repair of underground drainage pipes are necessary to ensure the normal operation of the city [2]. Manual pipe inspection is less efficient and prone to errors. If the defects and damages of drainage pipes can be detected more intelligently and automatically, the efficiency and quality of inspection can be greatly improved. Therefore, it is necessary to study the detection of defects and damages in urban underground drainage pipes. Most of the current drainage pipe inspections are carried out by filming the interior of the pipe through closed-circuit television and combining methods such as image processing and machine learning [3]. You Only Look Once version 5 (YOLOv5) is currently a widely used deep learning algorithm in defect detection. In terms of its improvement, its combination with the Convolutional Block Attention Module (CBAM) mechanism is relatively common. Fu et al. [4] introduced the CBAM mechanism into the backbone part of YOLOv5 for the helmet monitoring of electric bicycle riders. On a self-built dataset, they found that compared with the original YOLOv5s model, the proposed model achieved an improvement of 1.89% in the overall mean average precision (mAP). Pang et al. [5] also found that combining the CBAM mechanism with YOLOv5 can significantly improve the efficiency and accuracy of the model in solar cell defect detection. Lv et al. [6] also used the CBAM-combined YOLOv5 structure in the disease detection of apple tree leaves and achieved an improvement in the detection effect. In order to further improve the detection accuracy and speed of defects and damages in urban underground drainage pipes, this paper designed a detection method based on a deep learning algorithm. The CBAM mechanism and Softmax Intersection over Union (SIoU) loss function were introduced into the YOLOv5 algorithm. It is assumed that this improvement can enhance the detection performance of the YOLOv5 algorithm and improve the detection accuracy and speed. The assumption was verified through experiments on the dataset, with the expectation of providing a new available method for the management and construction of urban drainage systems.

2 Related works

Table 1: Related works

	Method	Dataset	Detection performance
Xiao et al. [4]	The improved cumulative sum model	Laboratory indoor simulation	The method had a relatively fast detection speed for defects in urban drainage pipes, which can reduce the detection costs.
Wang [5]	An improved detection method based on semantic segmentation labeling	Self-built datasets	The method had a mean average precision of 72.8, a precision of 84.0%, and a recall rate of 63.7%.
He et al. [9]	AlexNet and ResNet50	Self-built datasets	The two methods achieved an accuracy of 92.00% and 96.50% respectively for the test set and an accuracy of 85.41% and 87.94% respectively for real cases.
Huang et al. [10]	An improved convolutional neural network	The pipeline defect dataset collected in real scenarios	The method achieved an accuracy of 90.2%.

3 Design of an algorithm for detecting defects and damages in underground drainage pipes

3.1 YOLOv5 algorithm

Among deep learning algorithms, the YOLO series algorithms have good applications in target detection, including face recognition and autonomous driving [11]. YOLOv5 is the mainstream model. Compared with versions like YOLOv4, YOLOv5 is simpler to use and can achieve multi-scale detection, which is more efficient. Based on these advantages, YOLOv5 is a preferred algorithm for many targets detection tasks [12]. After YOLOv5, YOLOv7 and YOLOv8 introduced some new complex modules, which have a higher computational complexity and also place higher demands on computing resources and memory. As a well-verified benchmark, YOLOv5 has already gained a wide consensus on its performance. Moreover, due to a mature and stable codebase and community ecosystem, YOLOv5 has a clear modular design that makes it easy to modify. Therefore, this paper designed a defect detection and damage localization method for underground drainage pipes based on the YOLOv5 algorithm. The YOLOv5 algorithm mainly consists of the following parts.

(1) Input: The image to be detected is divided into four feature maps, and they are concatenated in the channel dimension to reduce the number of parameters.

(2) Feature extraction: it includes three modules: CBS, C3, and Spatial Pyramid Pooling Fast (SPPF).

① CBS: it consists of a Convolution (Conv), a Batch Normalization (BN), and a Sigmoid Linear Unit (SiLU) activation function, and their respective functions are extracting image features, preventing overfitting, and learning more complex features.

② C3: Three CBS modules + one BottleNeck module for extracting more detailed features;

③ SPPF: it is used to combine local and global features.

(3) Feature fusion: Use the Feature Pyramid Network (FPN) + Path Aggregation Network (PAN) structure to fuse shallow graphic features with deep semantic features and;

(4) Head: it includes three convolutional modules corresponding to three feature layers.

3.2 An improved YOLOv5 algorithm

Considering the complexity of actual drainage pipe images, in order to further meet the needs of defect loss detection, this paper improved the YOLOv5 algorithm by introducing a CBAM after the C3 module and optimizing the loss function of the bounding box to enhance the detection effect. The details are as follows.

(1) CBAM

The attention mechanism enables models to focus more on the important information in the input, thereby achieving higher performance, and has good applications in many areas of research such as speech recognition [13] and image processing [14]. The environment of urban underground drainage pipes is very complex, which increases the difficulty of feature extraction. Therefore, the attention mechanism can be applied to focus more on the information useful for defect loss detection. The CBAM is a lightweight structure [15] that can be quickly embedded into many models, and it combines two modules for precise feature extraction as follows.

① Channel attention module (CAM): Features of different channels can be weighted to enhance the representation of important features, calculated as follows:

$$M_C(F)=\sigma\left(MLP\left(AP(F)\right)+MLP\left(MP(F)\right)\right)=\sigma\left(W_1\left(W_0\left(F_{avg}^c\right)\right)+W_1\left(W_0(F_{max}^c)\right)\right)$$

where M_C denotes the channel attention feature map, σ is the sigmoid function, MLP is the shared neural network, AP denotes the mean feature point, MP denotes the maximum feature point, F^c_{avg} is the mean-pooling feature, and F^c_{max} is the max-pooling feature.

② Spatial attention module (SAM): The spatial dimension of the feature map can be weighted to enhance attention to important features at different positions, complementing the feature information focused on by CAM. The final feature map is obtained after integration. The calculation formula is:

$$M_S(F)=\sigma\left(f^{7\times7}([(F);MP(F)])\right)=f^{7\times7}\left(\left[F^c_{avg};F^c_{max}\right]\right),$$

where M_S represents a spatial attention feature map and $f^{7\times7}$ represents convolution operation.

(2) Loss function

The complete Intersection over Union (CIoU) damage function [16] is used in YOLOv5, which has better stability compared to the traditional IoU but also has problems such as high computational complexity and poor performance in small target boxes. Therefore, the following improved versions of IoU are used in this paper.

① Efficient Intersection over Union (EIoU) [17]: Considering the position and shape of the of the target box and depth features, it can more accurately reflect the differences between the target box and the real box. The calculation formula is:

$$L_{EIoU}=1\text{-}IoU+\frac{\rho^2(b,b^{gt})}{c^2}+\left(\frac{v}{(1\text{-}IoU)+v}\right)v,$$
$$v=\frac{4}{\pi^2}\left(\arctan\frac{w^{gt}}{h^{gt}}-\arctan\frac{w}{h}\right)^2,$$

where $\rho\left(b,b^{gt}\right)$ is the Euclidean distance between the center points of the real box and the predicted box, c is the diagonal length of the minimum bounding rectangle, w^{gt} and h^{gt} are the length and width values of the real box, and w are the length and width values of the predicted box.

② Focal EIoU: It is based on EIoU and combined with the Focal Loss function [18]. It can pay more attention to difficult-to-classify samples:

$$L_{FocalEIoU}=IoU^{\gamma}L_{EIoU},$$

where γ is a hyperparameter controlling the curve curvature, usually 0.5-2.0.

③ SIoU: Based on IoU, combined with the Softmax loss function, it can compare multiple categories of detection boxes to obtain the optimal detection box:

$$L_{SIoU}=1\text{-}IoU+\frac{\Delta+\Omega}{2},$$

where Δ is distance cost, exploring the distances of different bounding boxes from different centers as much as possible:

$$\Delta=\sum_{t=x,y}\left(1\text{-}e^{-\gamma\rho t}\right)=2\text{-}e^{-\gamma\rho_x}\text{-}e^{-\gamma\rho_y},$$
$$\rho_x=\left(\frac{b^{gt}_{c_x}\text{-}b_{cx}}{c_w}\right)^2,$$
$$\rho_y=\left(\frac{b^{gt}_{c_y}\text{-}b_{cy}}{c_h}\right)^2,$$

where c_w and c_h are the width and height of the minimum bounding rectangle, $b^{gt}_{c_x}$ and $b^{gt}_{c_y}$ are the coordinates of the real box center, b_{cx} and b_{cy} are the coordinates of the predicted box center, γ: γ=2-Λ, Λ is the angle cost, $\Lambda=1\text{-}2sin^2\left(arcsin\left(\frac{c_h}{\sigma}\right)\text{-}\frac{\pi}{4}\right)$, minimizing the number of distance-related variables to the greatest extent, and Ω is the shape cost, which is used to penalize the difference in the aspect ratio between the real box and the predicted box.

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

where θ is the sensitivity controlling the shape cost, usually 4.

4 Results and analysis

4.1 Experimental settings

The experiment was conducted in a Windows 10 environment, and the specific configuration is presented in Table 2.

Table 2: Experimental configuration.1

Central processing unit (CPU)	Intel TM i5-11400F CPU
Graphics processing unit (GPU)	GeForce RTX1080TI GPU
Acceleration module	CUDA 11.1
Deep learning framework	PyTorch 1.7.0

For the underground drainage pipe defect damage detection algorithm, the parameters are set as Table 3.

Table 3: Parameter settings.

Epochs	300
Batch size	16
Image size	640 × 640
Optimizer	Stochastic gradient descent
Initial learning rate	0.001

The experiment used the Sewer-ML dataset [16], with images from actual drainage pipe inspection projects. Data distributions in the dataset are shown in Table 4.

Table 4: Data distributions in the Sewer-ML dataset.

	Training	**Validation**	**Test**	**Total**
Normal	552,820	68,681	69,221	690,722
Defective	487,309	61,365	60,805	609,479
Total	1,040,129	130,046	130,026	1,300,201

The Sewer-ML dataset included 18 different types of defect and damage, five of which were selected in this paper.

Sediment: Impurities and silt settle at the bottom of the drainage pipe to form sediment. As the volume expands, the area of water flow through the pipe decreases, weakening the pipe's ability to transport rainwater and sewage.

Crack: When a pipe breaks due to external forces such as compression, rainwater and sewage will seep out and pollute the water environment.

Cut: The cut at the connection of the branch pipe and the main pipe is uneven, with gaps appearing at the edge of the cut. This causes rainwater and sewage to seep out, polluting the groundwater. Soil also flows in, forming sediment.

Disconnection: The joints of the two ends of the pipe are not fully joined, causing the joints to shift and create a gap, which will allow rainwater and sewage to seep out and pollute the water environment.

Obstacle: They may be foreign matters carried in by rainwater or sewage, or other material that fall off and block in the pipe, resulting in a reduction in water flow area.

The five selected types exhibit a high degree of diversity in visual characteristics, covering various challenges ranging from slender small targets (cracks), complex texture targets (tree roots) to large-area irregular targets (sediments). This provides a testing benchmark for comprehensively evaluating the generalization and robustness of the model. Moreover, these five types of defects have sufficient and high-quality annotation data, ensuring the effectiveness of model training and the statistical reliability of the evaluation results. The images in the dataset were enhanced using operations such as translation, rotation, and cropping to obtain 3,000 images of each type. Image quality was improved through histogram equalization and sharpening. They were labeled using the LabelImg tool [20]. Moreover, ten-fold cross-validation method [21] was used to divide the dataset into a training set, a validation set, and a test set. For the detection effect of the algorithm, the IoU threshold value was set as 0.5, and samples below 0.5 were considered negative cases. The following evaluation indicators were used:

(1) precision: $Precision=\frac{TP}{TP+FP}$

(2) recall rate: $Recall=\frac{TP}{TP+FN}$

(3) average precision (AP): area under the precision-recall curve, $AP=\int_0^1 P(R)dR$,

(4) mAP: $mAP=\frac{\sum_{i=1}^{N} AP_i}{N}$,

(5) frames per second (FPS) [22]: the number of images detected per second, which is used to reflect the detection speed of a model.

In the above equations, TP is the number of positive samples detected as positive, FP is the number of negative samples detected as positive, and FN is the number of positive samples detected as negative.

4.2 Result analysis

The YOLOv5 algorithm was divided into different versions based on network width and depth. Experiments were conducted on different versions. The obtained mAP and FPS are shown in Table 5 and Figure 1.

Table 5: Comparison of different versions of YOLOv5.2

	mAP/%	FPS	Parameter quantity
YOLOv5n	78.12	93	1.90×10^6
YOLOv5s	81.77	90	7.20×10^6
YOLOv5m	83.94	82	2.12×10^7
YOLOv5l	85.51	81	4.65×10^7

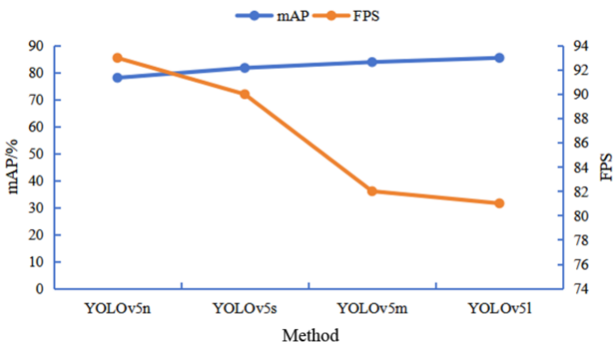

Figure 1: Comparison of different versions of YOLOv5.

As shown in Table 5 and Figure 1, with the expansion of model size, the mAP for the defect detection and damage localization also increased. The mAP of the YOLOv5n algorithm was 78.12%, and the mAP of the YOLOv5l algorithm was 85.51%, which was 7.39% higher than the YOLOv5n algorithm. However, the increase in scale also affected the detection speed; the larger the scale, the slower the detection speed. The comparison between YOLOv5m and YOLOv5l showed an increase of 1.57% in *mAP* and a decrease of one in FPS, indicating that the difference in detection speed between them was not significant. Therefore, in the subsequent experiments, the YOLOv5l version was used.

The effects of different attention mechanisms on detection results were compared (Table 6, Figure 2).

Table 6: Comparison of different attention mechanisms (Unit: %).3

	Baseline	Squeeze-and-Excitation (SE)	Coordinate Attention (CA)	CBAM
Sediment	85.33	87.64	88.46	90.88
Crack	71.21	73.36	74.41	81.21
Cut	94.56	95.16	96.77	98.97
Disconnection	95.19	96.93	96.14	97.43
Obstacle	81.26	83.56	85.57	87.56
mAP	85.51	87.33	88.27	91.21

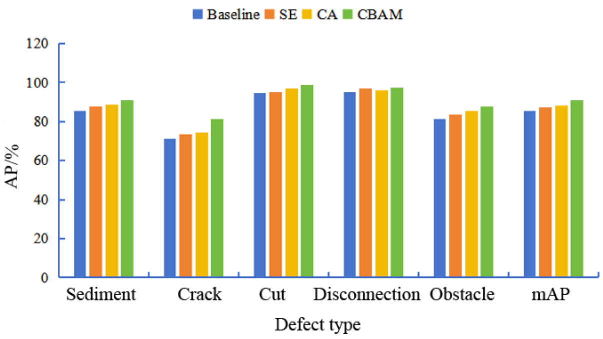

Figure 2: Comparison of different attention mechanisms.

From Table 6 and Figure 2, it can be seen that among different types of defect detection and damage localization, the detection of cuts and disconnections was more accurate. This may be because the characteristics of disconnections and cuts are more obvious and easier to identify. Cracks vary in size and direction and are easily confused with the background, making detection difficult. Similarly, sediments are located at the bottom of the pipe with blurred boundaries and are not easy to detect, and obstacles are also easily confused with sediment. The mAP was improved after the addition of different attention mechanisms compared to the baseline, and the CBAM had the best performance. SE only considered the importance of channel pixels and lacked attention to channel positions. CA performed limited on complex tasks. CBAM combined channel attention and spatial attention and achieved high accuracy in the detection of complex defects (cracks and obstacles), demonstrating its advantages.

The effects of different loss functions on the detection results were compared (Table 7, Figures 3 and 4).

Table 7: Comparison of different loss functions (unit: %).4

	Baseline	EIoU	Focal EIoU	SIoU
Sediment	85.33	86.16	86.89	87.03
Crack	71.21	72.64	72.23	73.07
Cut	94.56	96.17	96.68	97.03
Disconnection	95.19	96.55	96.91	97.28
Obstacle	81.26	82.33	82.94	83.19
mAP	85.51	86.77	87.13	87.52

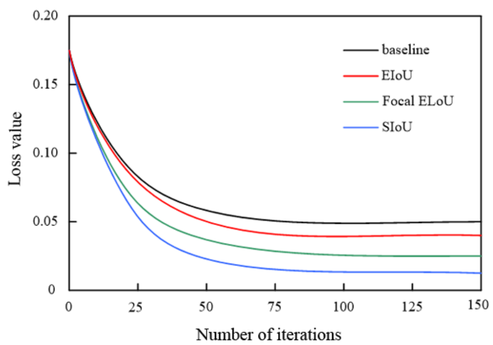

Figure 3: Loss value curve comparison chart.

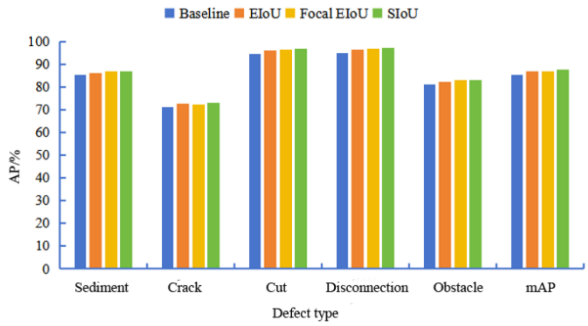

Figure 4: Comparison of different loss functions.

From Table 7 and Figures 3 and 4, it can be seen that after improving the loss function, the convergence of the network became faster, the detection performance was improved to a certain extent, but to a small extent. Among the three loss functions, SIoU exhibited the greatest improvement in detection performance and had the best convergence effect, and its mAP had an improvement of 2.01% compared to the baseline. Therefore, SIoU can be used instead of the original CIoU in YOLOv5 to achieve performance improvement.

The effect of the improvement on the detection results was determined by the ablation experiment (Table 8, Figure 5).

Table 8: Ablation experiments.5

	mAP/%	FPS
Baseline	85.51±2.16	81±1.21
YOLOv5+CBAM	91.21±3.33	79±0.77
YOLOv5+SIoU	87.52±2.77	83±1.45
YOLOv5+CBAM+SIoU	93.37±3.56*	85±1.17*

Note:* indicates $p < 0.05$ compared to the other method

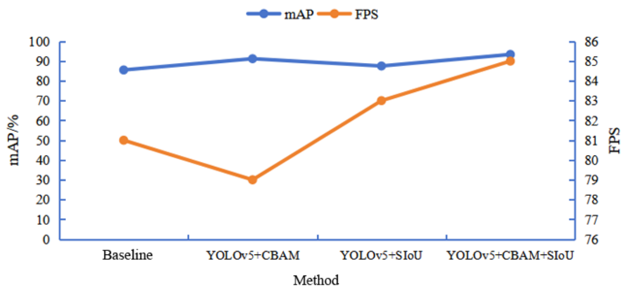

Figure 5: Ablation experiments.

According to Table 8 and Figure 5, the introduction of CBAM brought a 5.7% mAP improvement, and the FPS dropped from 81 to 79, indicating that the introduction of CBAM was beneficial to the improvement of detection accuracy, but it affected the detection speed to some extent. From this perspective, CBAM could effectively enhance the feature extraction ability, but it increased the computational load. The introduction of SIoU brought a 2.01% mAP improvement, and the FPS increased from 81 to 83, possibly because SIoU had a fast convergence speed. By introducing structural priors (angles and shapes), SIoU guided the model to converge to a state with greater “geometric regularity” during the training process, providing higher-quality predicted bounding box proposals, thus significantly reduced the computational latency and achieved an improvement in throughput. The combined use of CBAM and SIoU achieved the best results, achieving a mAP of 93.37% and a FPS of 85. Moreover, the statistical significance results showed that compared the results of YOLOv5+CBAM+SIoU with the other methods, the p value was less than 0.05, demonstrating the reliability of the improvement to the YOLOv5 algorithm.

The method proposed was compared with some other deep learning-based detection methods (Table 9 and Figure 6).

Table 9: Comparison with other detection methods.6

	mAP/%	FPS
Single-Shot Multibox Detection (SSD) [23]	59.87	54
YOLOv3 [24]	75.59	68
YOLOv4 [25]	81.21	72
Faster regional-based convolutional neural network (R-CNN) [26]	84.93	38
The improved YOLOv5	93.37	85

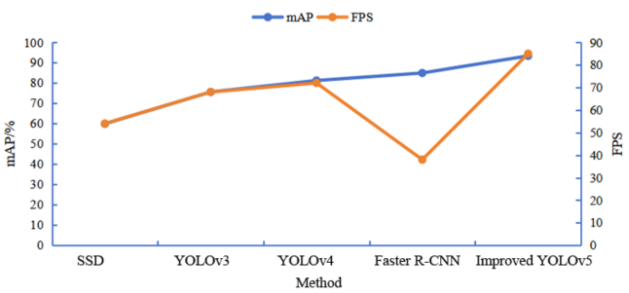

Figure 6: Comparison with other detection methods.

From Table 9 and Figure 6, it can be seen that the SSD performed poorly in drainage pipe defect detection and damage localization, with a mAP of 59.87% only and an FPS of 54. The feature fusion of SSD was relatively simple, while the proposed method realized the bidirectional fusion of deep and shallow features. When facing defects of different scales, the proposed method was more robust. Although the Faster R-CNN algorithm had high accuracy (mAP = 84.93), this came at the expense of speed. Its FPS was only 38. Faster R-CNN is a two-stage detector. Its serial process is computationally complex and the inference speed is slow, while YOLOv5 can achieve end-to-end fast inference. Compared with YOLOv3 and YOLOv4, the improved YOLOv4 algorithm proposed in this paper is advanced in terms of network architecture and data augmentation. The introduced CBAM and SIoU also bring significant performance improvements. Generally speaking, the proposed method was significantly superior to the other methods in terms of both precision and speed.

5 Discussion

In the detection of defects and damages in urban underground drainage pipes, this paper designed a YOLOv5 algorithm improved by combining the CBAM mechanism and SIoU, and verified its detection performance of using the sewer-ML dataset. The results showed that compared with other attention mechanisms or other loss functions, the selected CBAM and SIoU both had advantages. Comparisons with other deep learning detection methods showed that as a single-stage detector, SSD has insufficient ability to extract the diverse defect features of the pipes, resulting in many false detections and missed detections. As a two-stage detector, Faster R-CNN has many calculation steps and takes a long time, failing to meet the requirements of rapid response and efficient inspection. As early versions of the YOLO series, YOLOv3 and YOLOv4 are also inferior to YOLOv5 in terms of precision and speed. The combination of the CBAM mechanism with YOLOv5 can help the YOLOv5 network focus more accurately on the key features of the defects, providing the subsequent detection head with more informative and less noisy features. SIoU reconsiders the cost of bounding box regression, which helps the model generate prediction boxes with a higher degree of fit with the real boxes, bringing faster convergence and better convergence effects to the model. Under the synergistic effect of CBAM and SIoU, the precision and speed of the improved YOLOv5 algorithm for the detection of defects and damages in urban underground drainage pipes were improved.

Based on the research results, the designed algorithm can be applied in the detection of defects and damages in actual urban underground drainage pipes. For example, it can be applied in inspection robots to detect complex underground drainage pipes, process video streams in real time, and accurately locate and quantify defects, which makes it possible to formulate point-to-point repair plans and greatly saves the costs of excavation and repair. For the municipal system, the intelligent detection is conducive to the municipal department in formulating predictive maintenance plans. According to the detection results, priority maintenance can be carried out on high-risk pipe sections, thus avoiding accidents such as road collapse and urban waterlogging, and having a profound impact on enhancing urban safety and reducing operation and maintenance costs.

In future actions, experiments will be carried out on a more diverse dataset of underground drainage pipe defects and tested under more realistic conditions, such as occluded defects and noisy images. Meanwhile, further research will also be conducted on the deployment of the algorithm in the actual environment.

6 Conclusion

This paper designed a method based on the YOLOv5 algorithm for defect damage detection of urban underground drainage pipes. Through experiments on the dataset, it was found that the proposed method effectively balanced accuracy and speed. The detection accuracy for different defect loss types was above 80%, mAP reached 93.37%, and the FPS was 85. It can be applied in actual urban sewage management to achieve better detection of defects and damages in underground drainage pipes.

References

[1] Huang R, Tao Z, Lin Y, Wei J, Zhou W, He Y (2024). Current Situation of Drainage Pipe Network in China and its Detection Technology: a Brief Review. *Polish Journal of Environmental Studies*, 33(1), pp. 19–29. https://doi.org/10.15244/pjoes/166899.

[2] Li Y, Wang H, Dang LM, Song HK, Moon HJ (2023). Attention-guided multiscale neural network for defect detection in sewer pipelines. *Computer-Aided Civil & Infrastructure Engineering*, 38(15), pp. 2163. https://doi.org/10.1111/mice.12991.

[3] Yougao L, Wei H (2020). Identification and feature extraction of drainage pipeline cracks based on SVD and edge recognition method. *2019 3rd International Conference on Electronic Information Technology and Computer Engineering (EITCE)*, pp. 1184–1188. https://doi.org/10.1109/EITCE47263.2019.9095153.

[4] Fu SY, Wei D, Zhou LY (2024). Helmet Detection Algorithm of Electric Bicycle Riders Based on YOLOv5 with CBAM Attention Mechanism Integration. *International Workshop on Advanced Computational Intelligence and Intelligent Informatics*, pp. 43–56. https://doi.org/10.1007/978-981-99-7593-8_5.

[5] Pang X, Zeng J, Yiyang C, Liu C, Fan X, Li S, Gong M (2025). Defect Detection Network for TOPCon Solar Cells Based on Improved YOLOv5 and CBAM Mechanism. *International Conference on Big Data*, pp. 75–87. https://doi.org/10.1007/978-3-031-77088-3_6.

[6] Lv M, Su WH (2023). YOLOV5-CBAM-C3TR: an optimized model based on transformer module and attention mechanism for apple leaf disease detection. *Frontiers in Plant Science*, 14, 1323301. https://doi.org/10.3389/fpls.2023.1323301.

[7] Xiao Y, Meng J, Yan H, Wang J, Xin K, Tao T (2024). Novel drainage pipeline breakages detection based on MEMS inertial sensor: From mechanism to application. *Journal of Cleaner Production*, 473(Oct.1), pp. 1–14. https://doi.org/10.1016/j.jclepro.2024.143546.

[8] Wang J, Wang C, Liu Y, Zhang X, Huang J (2025). Detection and effectiveness of improved drainage pipe defects–based semantic segmentation labeling methods. *Journal of Pipeline Systems Engineering and Practice*, 16(4). https://doi.org/10.1061/JPSEA2.PSENG-1898.

[9] He Z, Dang N, Zhou X, Li J. (2025). Automatic defect classification method of drainage pipe based on image processing and deep learning. *International Conference on Hydraulic and Civil Engineering & Engineering Safety and Disaster Prevention Forum*, pp. 1007–1013. https://doi.org/10.1007/978-3-031-90717-3_101.

[10] Huang QH, Li BA, Lv XQ, Zhang ZJ, Liu KH. (2020). Research on Pipeline Video Defect Detection Based on Improved Convolution Neural Network. *Journal of Physics: Conference Series*, 1576(1), 1–6. https://doi.org/10.1088/1742-6596/1576/1/012028.

[11] Budzan S, Szwedka M (2024). Integration of YOLO detection algorithm with trajectory prediction of pedestrians for advanced driver assistance system. *Przegląd Elektrotechniczny*, 100(12), pp. 53–58. https://doi.org/10.15199/48.2024.12.12.

[12] Zhou K, Zheng G, Zhai H, Lv X, Zhang W (2024). An Improved Distraction Behavior Detection Algorithm Based on YOLOv5. *Computers, Materials & Continua*, 81(11), pp. 2571–2585. https://doi.org/10.32604/cmc.2024.056863.

[13] Ha MH, Nguyen DC, Chan L Q, Chen OTC (2024). Emotional Inference from Speech Signals Informed by Multiple Stream DNNs Based Non-Local Attention Mechanism. *EAI Endorsed Transactions on Industrial Networks & Intelligent Systems*, 11(4), pp. 1. https://doi.org/10.4108/eetinis.v11i4.4734.

[14] Zhang K, Liu Y, Hu K (2025). RAW image reconstruction based on multi-scale attention mechanism. *Journal of Beijing University of Aeronautics and Astronautics*, 51(1), pp. 257–264. https://doi.org/10.13700/j.bh.1001-5965.2022.0959.

[15] Ye R, Qian Y, Huang X (2024). RT-CBAM: Refined Transformer Combined with Convolutional Block Attention Module for Underwater Image Restoration. *Sensors*, 24(18), pp. 17. https://doi.org/10.3390/s24185893.

[16] Yang Y, Hu S, Ke Y, Zhou R (2023). Flame smoke detection algorithm based on YOLOv5 in petrochemical plant. *International Journal of Intelligent Computing and Cybernetics*, 16(3), pp. 502-519. https://doi.org/10.1108/IJICC-11-2022-0291.

[17] Wang H, Liu Y, Luo H, Luo Y, Zhang Y, Long F, Li L (2024). Camellia oleifera trunks detection and identification based on improved YOLOv7. *Concurrency & Computation: Practice & Experience*, 36(27), pp. 1–14. https://doi.org/10.1002/cpe.8265.

[18] Li C, Chen D, Chen J, Dai H (2021). A Cross-Layer Fusion Multi-target Detection andRecognition Method Based onImproved FPN Model inComplex Traffic Environment. *CAAI International Conference on Artificial Intelligence*, pp. 323–334. https://doi.org/10.1007/978-3-030-93046-2_28.

[19] Haurum J B, Moeslund T B (2021). Sewer-ML: A Multi-Label Sewer Defect Classification Dataset and Benchmark. *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA*, pp. 13451–13462. DOI:10.48550/arXiv.2103.10895.

[20] Deka B, Chakraborty D (2024). UAV Sensing-Based Litchi Segmentation Using Modified Mask-RCNN for Precision Agriculture. *IEEE Transactions on Agrifood Electronics*, 2(2), pp. 509–517. https://doi.org/10.1109/TAFE.2024.3420028.

[21] Tan JH, Bhandary SV, Sivaprasad S, Hagiwara Y, Bagchi A, Raghavendra U, Rao AK, Raju B, Shetty NS, Gertych A, Chua KC, Acharya UR (2018). Age-related Macular Degeneration detection using deepconvolutional neural network. *Future Generation Computer Systems*, 87(OCT.), pp. 127–135. https://doi.org/10.1016/j.future.2018.05.001.

[22] Ruiz-Beltran CA, Romero-Garces A, Gonzalez M, Sanchez Pedraza A, Rodriguez-Fernandez JA, Bandera A (2022). Real-time embedded eye detection system. *Expert Systems with Applications*, 194, pp. 1–10. https://doi.org/10.1016/j.eswa.2022.116505.

[23] Sarma K S R K, Sasikala C, Surendra K, Erukala S, Aruna SL (2024). A comparative study on faster R-CNN, YOLO and SSD object detection algorithms on HIDS system. *AIP Conference Proceedings*, 2971(1), pp. 13. https://doi.org/10.1063/5.0195857.

[24] Apoorva, Babu S, Rani A, Srivastva A, Reddy A (2024). On stream face mask and helmet detection system using YoloV3 algorithm. *AIP Conference Proceedings*, 2742(1), pp. 8. https://doi.org/10.1063/5.0191204.

[25] He J, Wang W, Yang N (2024). Research on Track Fastener Service Status Detection Based on Improved Yolov4 Model. *Journal of Transportation Technologies*, 14(2), pp. 212-223. https://doi.org/10.4236/jtts.2024.142013.

[26] Wang Q, Wang Z, Li J, Yang Y, Liu P (2020). Improvement of Faster-RCNN Detection Algorithms for Small Size Line Accessory Equipment. *Journal of Physics: Conference Series*, 1453, pp. 1–6. https://doi.org/10.1088/1742-6596/1453/1/012007.