applied sciences

MDPI

Article

YOLOv5-Sewer: Lightweight Sewer Defect Detection Model

Xingliang Zhao , Ning Xiao , Zhaoyang Cai * and Shan Xin

Beijing University of Civil Engineering and Architecture, Beijing 100044, China

* Correspondence: caizhaoyang@bucea.edu.cn

Abstract: In the field of defect detection in sewers, some researches focus on high accuracy. However, it is challenging for portable on-site devices to provide high performance. This paper proposes a lightweight sewer defect detection model, You Only Look Once (YOLO) v5-Sewer. Firstly, the backbone network of YOLOv5s is replaced with a stacked MobileNetV3 block. Secondly, the C3 module of the neck of YOLOv5s is improved with a C3-Faster module. Thirdly, to compensate for the accuracy loss due to the lightweight network, a channel attention (CA) and convolutional block attention module (CBAM) are added to the proposed method. Finally, the Efficient Intersection over Union (EIOU) is adopted as the localization loss function. Experimental validation on the dataset shows that YOLOv5-Sewer achieves a 1.5% reduction in mean Average Precision (mAP) while reducing floating-point operations by 68%, the number of parameters by 55%, and the model size by 54%, compared to the YOLOv5s model. The detection speed reaches 112 frames per second (FPS) with the GPU (RTX 3070Ti). This model successfully implements a lightweight design while maintaining the detection accuracy, enhancing its functionality on low-performance devices.

Keywords: sewer; defect detection; lightweighting; YOLOv5s

1. Introduction

At present, closed-circuit television (CCTV) inspection technology is widely employed in sewer inspection tasks. This technique involves workers analyzing videos or images captured by CCTV within the pipelines to identify internal defects in the sewer system. Nevertheless, the extended duration of these inspections can result in visual fatigue, thereby compromising the accuracy of defect identification. Consequently, the utilization of automatic detection methods has been recognized as an effective solution [1–5]. Numerous methodologies for the detection of defects in sewer systems have been proposed by researchers.

Ye et al. [6] introduced an image recognition algorithm for the diagnosis of sewer issues, employing feature extraction and machine learning techniques. This algorithm utilizes a support vector machine (SVM) for the categorization of sewer defects. On a similar note, Myrans et al. [7] proposed a machine learning algorithm that utilizes a random forest classifier ensemble to identify different fault types. Despite their effectiveness, both algorithms necessitate manual feature extraction, so their adaptability to the intricate environments encountered in practical engineering applications within sewer systems is limited.

Kumar et al. [8] utilized convolutional neural networks (CNNs) for the classification of sewer defects. Chen et al. [9] applied a classifier built with InceptionV3 to identify sewer defects. These methods based on single-label classifiers are not well suited for situations where sewer systems exhibit multiple defects simultaneously. The limitation in handling multiple defects underscores the need for further advancements in classification methodologies to address the complexity of real-world sewer environments.

In certain studies, researchers have employed semantic segmentation models to address the challenge of identifying multiple defects in sewer systems. Wang et al. [10] introduced a neural network named DilaSeg-CRF, which combines a CNN with dense

check for updates

Citation: Zhao, X.; Xiao, N.; Cai, Z.; Xin, S. YOLOv5-Sewer: Lightweight Sewer Defect Detection Model. *Appl. Sci.* **2024**, *14*, 1869. https://doi.org/10.3390/app14051869

Academic Editor: Lorenzo J. Tardón

Received: 26 January 2024
Revised: 20 February 2024
Accepted: 23 February 2024
Published: 24 February 2024

conditional random fields (CRF) to improve the accuracy in recognizing three types of sewer defects. Similarly, Pan et al. [11] proposed a semantic segmentation network called PipeUNet, designed for the identification of four typical defects. Despite their effectiveness in handling multiple defect types, it is worth noting that the detection speed of both methods is low. This limitation in detection speed may pose challenges, especially in applications where real-time monitoring and rapid analysis are crucial. Further research and development efforts are needed to optimize the efficiency of such semantic segmentation models for faster and more practical sewer defect identification.

Furthermore, some researchers have turned to object detection models for sewer defect detection tasks. Cheng et al. [12] put forward an automated method based on Regions with CNN Features (R-CNN), designed for high-precision and rapid sewer defect detection. Wang et al. [13] introduced a sewer defect detection model utilizing the Faster R-CNN algorithm, incorporating clustering analysis to effectively enhance the accuracy of defect detection in sewer systems. Despite the high accuracy achieved by these object detection models, there is a limitation in their consideration of on-site equipment performance. Addressing the practicality of these models with on-site tools and conditions is an essential aspect for further research and implementation in real-world sewer inspection scenarios.

In recent years, the field of object detection has seen notable advancements, with the You Only Look Once (YOLO) series of algorithms, as proposed by Redmon and Liu [14,15], demonstrating advantages such as high speeds, accuracy, and relatively low computational costs. Addressing the specific application to pipeline defect detection, Chen et al. [16] introduced an improved YOLOv5 algorithm, based on Cycle-GAN. This enhanced algorithm proves effective in identifying defective areas within pipelines. However, existing defect detection models pursue high accuracy and overlook the performance on low-performance devices.

Zhang at al. [17] introduced a lightweight method for the detection of sewer defects based on the improved YOLOv5, which can accurately identify four types of sewer defects. However, the model is still complex, with a large number of parameters, model size, and floating-point operations. YOLOv5, being a newer iteration in the YOLO series, outperforms many other object detection algorithms, including Faster R-CNN [18] and the Single-Shot Multi-Box Detector (SSD) [19], in terms of recognition performance. However, in practical sewer defect detection tasks, certain drawbacks persist. For on-site computing devices with limited computational power, the computational demands of YOLOv5 can be excessive, leading to suboptimal detection speeds. This limitation highlights the need for further optimization or alternative approaches to make such advanced algorithms more feasible for real-world sewer inspection scenarios.

This paper introduces a lightweight yet highly accurate model for the detection of sewer defects. While computer vision has been widely applied in inspecting buildings, dams, bridges, roads, industrial artifacts, and various pipelines (water, gas, petroleum), including sewers, this study stands out by addressing applications for field devices with limited computational capacities.

The approach is built upon an optimized version of You Only Look Once (YOLOv5) for the detection and classification model. Certain stages of the pipeline have been replaced with carefully crafted and tailored components. The resulting iteration is named YOLOv5-Sewer.

Experimental validation on the dataset, coupled with a comparison against alternative pipeline methods, demonstrates superior performance while maintaining acceptable accuracy levels. In conclusion, the proposed model successfully implements a more lightweight design for low-performance devices, achieving commendable detection accuracy and overall enhanced performance.

The rest of this paper is organized as follows. In Section 2, we introduce the method of model lightweighting and the method of compensating for accuracy losses. In Section 3, we introduce the dataset acquisition and improvements, the experimental environment, comparative experiments among different lightweight backbone networks, fusion experiments with various improvement modules, comparative experiments with other popular object detection models, detection speed comparisons on portable devices, and the visualization of the detection results. Finally, the conclusions drawn from this study are presented in Section 4.

2. YOLOv5-Sewer Network Design

2.1. Related Works

YOLOv5s is a lightweight model within the YOLOv5 framework. Many studies on sewer defect detection are based on YOLOv5 [2,17,20–22]. This paper introduces a novel and lightweight sewer defect detection model named YOLOv5-Sewer, based on the improved YOLOv5s, specifically designed for the recognition of eight types of defects commonly found in sewer systems. These defects encompass cracks, utility intrusion, obstacles, joint offset, debris, holes, buckling, and roots. The proposed model adopts MobileNetV3 block [23] stacking as the backbone network, aiming to reduce the model size and associated detection equipment costs. To further enhance the speed and accuracy in pipeline defect detection, the paper introduces the C3-Faster module, based on Point Convolution (PConv) and a FasterNet block [24,25], replacing the original C3 module in YOLOv5s.

Addressing the challenge of dark sewer backgrounds, the paper incorporates a convolutional block attention module (CBAM) [26] and channel attention (CA) [27] to mitigate their impact on the model. Moreover, the Efficient Intersection over Union (EIOU) localization loss function [28] is employed to enable the network to adapt effectively to multi-scale sewer defects. The experimental results validate the effectiveness of the proposed YOLOv5-Sewer model, demonstrating its capability to achieve lightweighting with high accuracy in sewer defect detection tasks.

2.2. YOLOv5-Sewer Network Architecture

The challenges posed by the complexity of sewer environments, homogenized images, and features such as medium- to long-distance dimness [29] present difficulties in sewer defect detection. These characteristics contribute to visual similarity, impacting the accuracy of defect detection and adding complexity to the network.

In response to these challenges, this paper proposes enhancements to YOLOv5s, a lightweight model known for its outstanding performance in object classification and detection. The objective is to tailor this model to the swift and effective detection of defects within sewer systems. The framework of the proposed model, YOLOv5-Sewer, is depicted in Figure 1. This framework aims to address the unique challenges posed by sewer environments, providing a robust and efficient solution for defect detection in such complex scenarios.

The enhancements to the YOLOv5s network in the proposed YOLOv5-Sewer model are as follows.

1. The original backbone of the YOLOv5s network is substituted with MobileNetV3 block stacking. This modification is aimed at reducing the network's complexity.
2. The C3 module is upgraded to the C3-Faster module, incorporating partial convolutions. This adjustment focuses the network on crucial regions of the feature map, enhancing its ability to capture relevant information.
3. To counteract the reduction in detection accuracy resulting from the lightweight design, a convolutional block attention module (CBAM) and channel attention (CA) modules are integrated. These attention mechanisms help the network to adapt to the challenging visual characteristics of sewer environments.
4. The Efficient Intersection over Union (EIOU) localization loss function is introduced, replacing the original Complete Intersection over Union (CIOU) loss function. This adaptation allows the model to effectively handle multi-scale sewer defects.

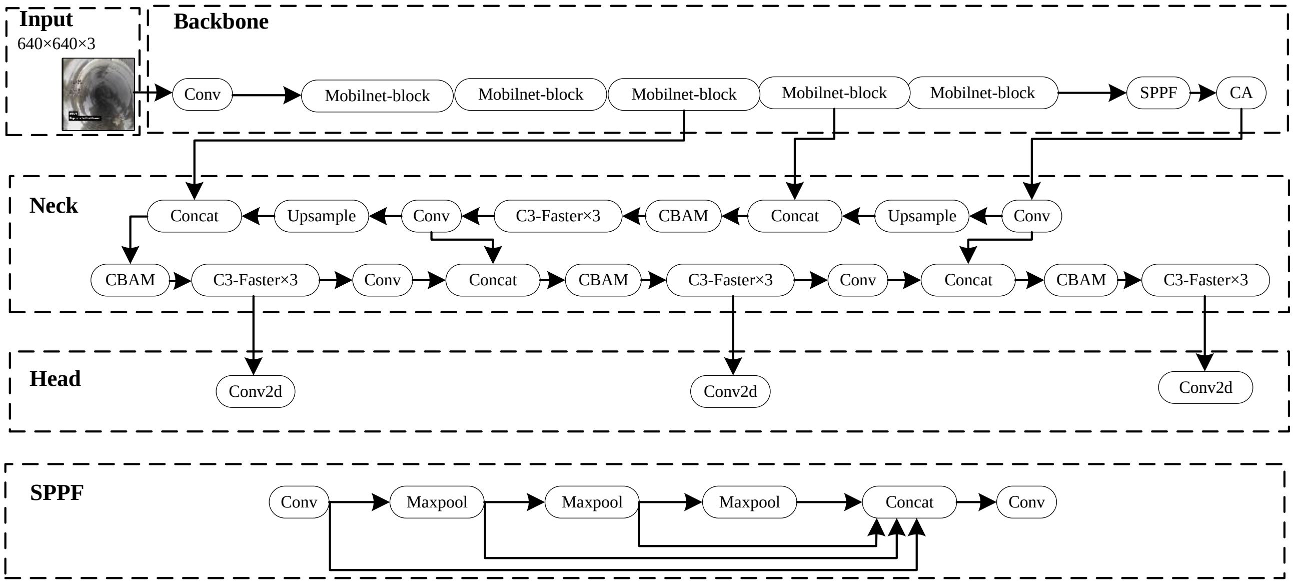

Figure 1. The framework of the YOLOv5-Sewer model.

The improved YOLOv5-Sewer model demonstrates robust capabilities in detecting pipeline defects. It successfully mitigates the impact of dim backgrounds in sewer environments, leading to enhanced accuracy in defect detection. Furthermore, the lightweight design reduces redundant computations, thereby lowering the training costs. This makes the model well suited for rapid defect detection in sewers on-site, particularly on low-configuration computing devices, owing to its optimized model size, computational efficiency, and detection speed.

2.2.1. Lightweight Network

In the traditional YOLOv5 series [30], the backbone utilizes the complex CSPDarknet53 network structure, which significantly increases the computational load of the model, resulting in slower detection speeds. This limitation poses challenges to the practical applicability of the model, particularly in on-site computing devices, where deploying large and intricate models for sewer defect detection can be problematic.

To address this issue, the proposed YOLOv5-Sewer model opts for a more lightweight backbone structure, utilizing MobileNetV3 block stacking. This choice aims to reduce the computing and memory requirements during training, albeit with a slight decrease in mean Average Precision (mAP). The MobileNet block, constituting the backbone, consists of three main components. Firstly, the input undergoes a 1×1 ordinary convolution to increase the dimensionality. Subsequently, depth-wise separable convolutions with different kernel sizes (3×3 or 5×5) are applied. Finally, a 1×1 ordinary convolution reduces the dimensionality for the output. This inverted residual structure helps to maintain the correlation between the input and output, and the low-dimensional features are extended to high-dimensional ones by depth-wise separable convolutions. The incorporation of a non-linear activation function enhances the interaction capabilities among channels, contributing to an efficient and lightweight backbone structure for sewer defect detection.

As illustrated in Figure 2, MobileNetV3 incorporates the Squeeze and Excitation Network (SENet) [31] after depth-wise separable convolutions. This integration allows for the adjustment of the weights in different channels, facilitating a focused emphasis on relevant channels. Additionally, to enhance the model's computation speed, some swish activation functions are replaced with hard swish activation functions.

The introduction of the Squeeze and Excitation Network enhances the adaptability of the MobileNetV3 backbone by enabling the model to dynamically adjust its focus on relevant channels, thereby improving its ability to capture important features. Moreover, the substitution of the swish activation functions with hard swish contributes to faster model computation. The combined effect of these enhancements results in a significantly advantageous and lightweight MobileNetV3 backbone for sewer defect detection. This lightweight design proves beneficial, particularly in environments with limited computational resources, making the proposed model well suited for deployment in scenarios where the computing capabilities are constrained.

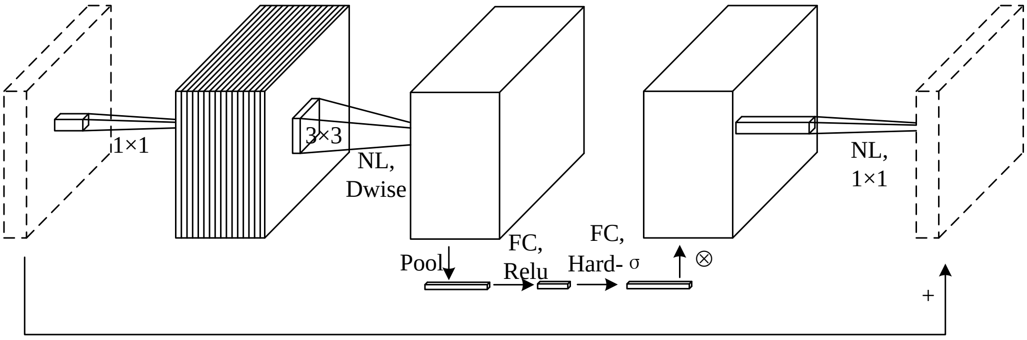

Figure 2. The structure of the MobileNetV3 block.

2.2.2. C3-Faster

Chen et al. [24] introduced the FasterNet block, a neural network module that incorporates the PConv convolution technique. This integration is designed to enhance the model's computation speed without compromising its accuracy. The structure of this module is illustrated in Figure 3. The utilization of PConv convolution within the FasterNet block contributes to efficient and rapid computation, making it a valuable component in the proposed model's architecture.

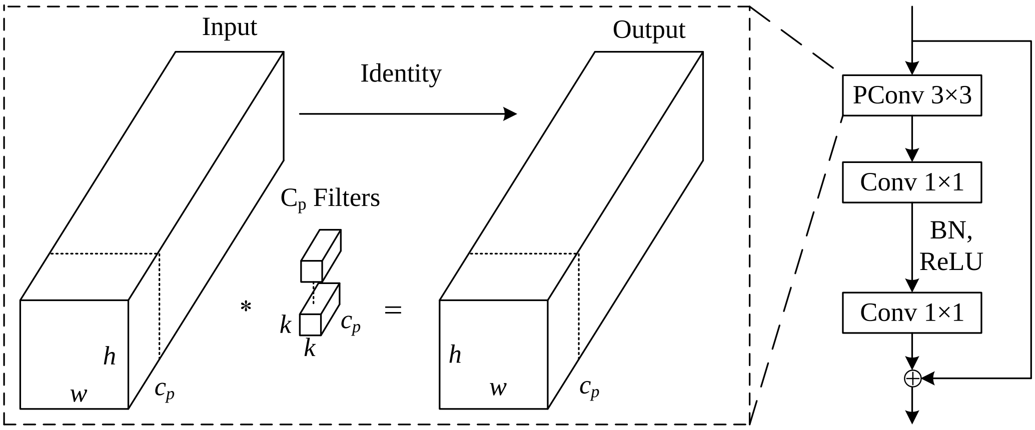

Figure 3. The structure of the FasterNet block.

The Point Convolution (PConv) operation is designed to perform convolution only on select input channels for spatial feature extraction while preserving the remaining channels. This approach aims to reduce memory access and minimize redundant computations. To fully and efficiently utilize all channel information, the proposed model introduces Pointwise Convolution (PWConv) after PConv. PWConv involves a 1 × 1 convolution operation along the channel dimension.

PWConv exhibits a T-shaped convolutional receptive field on the input feature map, focusing more on the central position compared to a regular convolution. This central position is distinguished by a higher Frobenius norm, indicating the presence of more crucial information in the feature map. By incorporating PWConv after PConv, the model strives to capture and emphasize significant spatial features, enhancing the overall effectiveness of the sewer defect detection process.

To enhance the network performance and reduce the number of parameters, the FasterNet block module is integrated into the original YOLOv5's C3 module, resulting in the C3-Faster module. The network structure is illustrated in Figure 4. The incorporation of the C3-Faster module into YOLOv5s significantly improves the operational speed of the network. Simultaneously, it successfully reduces the number of parameters, contributing to a more efficient and accurate model.

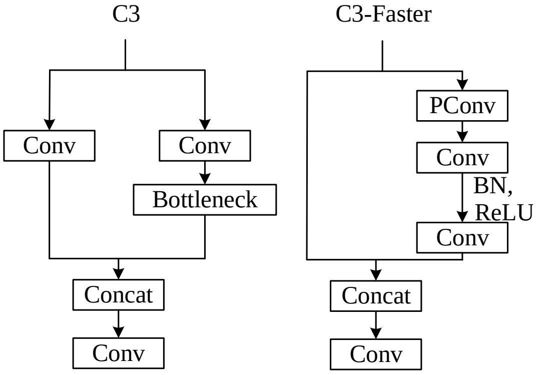

Figure 4. The structure of the C3 and C3-Faster blocks.

The combination of PConv and PWConv, along with the introduction of the FasterNet block module, plays a pivotal role in enhancing the performance of the sewer defect detection model. This integration not only effectively boosts the network speed but also yields satisfactory results in terms of model accuracy, making it a valuable improvement to the YOLOv5 architecture for sewer defect detection tasks.

2.2.3. Convolutional Block Attention Module

The convolutional block attention module (CBAM) is incorporated into the proposed model to address challenges posed by dim backgrounds and low-resolution images. CBAM enables the model to focus more on learning and extracting crucial features relevant to sewer defects. This becomes particularly crucial under low-resolution conditions, as CBAM is adept in capturing the features of sewer defects at lower resolutions. It takes into account differences between pixel categories, channel features, and contextual correlations. The structure of this module is illustrated in Figure 5.

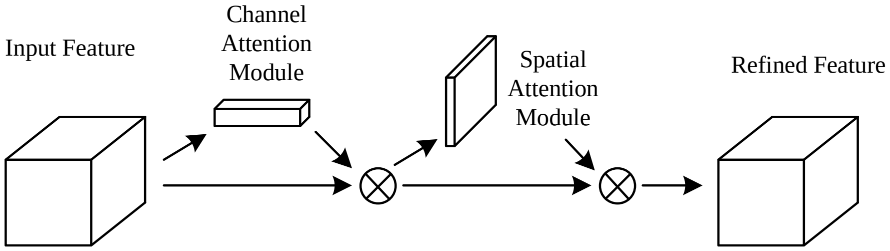

Figure 5. The structure of the convolutional block attention module.

The channel attention mechanism of CBAM can be expressed using the following Equation (1):

$$M_c(F) = \sigma(\mathrm{MLP}(avgpool(F)) + \mathrm{MLP}(maxpool(F))) \tag{1}$$

where F represents the input feature map; $avgpool$ and $maxpool$ denote average pooling and max pooling operations, respectively. MLP refers to a multi-layer perceptron, and σ represents the sigmoid function.

In the channel attention module, the input feature map undergoes both $avgpool$ and $maxpool$ operations based on width and height, resulting in two feature maps. These feature maps are then fed into a shared multi-layer perceptron (MLP), and the generated features are passed through the spatial attention module. Weighting is applied to the feature maps to emphasize crucial features. The spatial attention can be expressed using the following Equation (2):

$$M_S(F) = \sigma(f^{7\times7}([avgpool(F); maxpool(F)])) \tag{2}$$

The spatial attention module takes the output feature map from the channel attention module as input and performs channel-wise max-pooling and average-pooling operations to obtain two feature maps. These feature maps are then concatenated along the channel dimension and reduced to a single channel through a convolution operation. Subsequently, the spatial attention feature is generated using the sigmoid activation function. The final output feature map of the CBAM attention mechanism is obtained by the element-wise multiplication of the spatial attention module output feature map with the input feature map, as described by the following Equation (3):

$$Y = F \otimes M_C(F) \otimes M_S(F) \tag{3}$$

The integration of CBAM into the sewer defect detection model based on YOLOv5s serves to enhance the attention to crucial channels. This improvement results in the enhanced ability of the model to focus on defect features. The heightened attention provided by CBAM contributes to more accurate detection results in practical applications, especially in challenging scenarios with dim backgrounds and low-resolution images. The refined feature extraction facilitated by CBAM enhances the model's discriminative power, making it well suited for sewer defect detection in real-world settings.

2.2.4. Coordinate Attention

To overcome the limitations of CBAM, particularly in neglecting long-range dependencies, channel attention (CA) is integrated into the model. In contrast to CBAM, CA not only considers channel information but also incorporates position information related to the direction, providing a more comprehensive understanding of the relationships between defects at different locations. This holistic consideration enables the model to more accurately locate and identify sewer defects, taking into account not only the features within channels but also their spatial relationships.

Additionally, the lightweight design of CA allows it to be easily integrated into existing network structures without significantly increasing the computational complexity. The structure of CA is illustrated in Figure 6. This integration represents a thoughtful enhancement, addressing the limitations of CBAM and improving the model's capability to capture both channel-wise and spatial dependencies for more effective sewer defect detection.

The CA module is designed to address the limitations of CBAM in handling long-range dependencies. The module follows a series of operations to effectively capture spatial and positional information, contributing to the accurate localization of targets. The process is as follows.

1. Decomposition of Avg pool: To preserve spatial information and obtain positional information for long-range channel dependencies, CA decomposes the Avg pool by pooling separately in the X and Y directions. This generates feature maps of dimensions $C \times H \times 1$ and $C \times 1 \times W$, allowing the attention module to capture long-range dependencies along one spatial direction while preserving the positional information along the other spatial direction.
2. Feature map fusion: The obtained feature maps are fused through a concatenate operation.
3. Processing through convolution and activation: Through a 1×1 convolutional kernel and activation operation, the fused feature maps are processed.
4. Spatial split: A split operation is applied along the spatial dimension, dividing the feature maps into two parts, $C/r \times H \times 1$ and $C/r \times 1 \times W$.

5. Up-sampling and final attention vector: An up-sampling operation is performed through a 1 × 1 convolutional kernel, combined with the sigmoid activation function, to obtain the final attention vector.
6. Element-wise multiplication: The final attention vector is used to perform element-wise multiplication on the original feature maps, resulting in the final feature maps with attention weights in the X and Y directions.

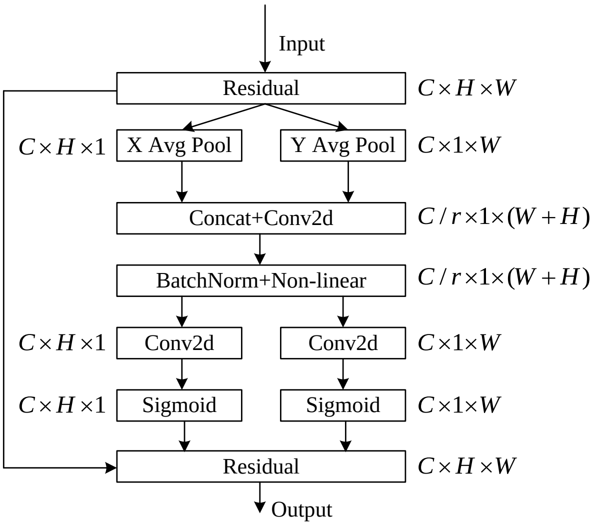

Figure 6. The structure of coordinate attention.

The CA attention module, with its ability to effectively handle long-range dependencies, proves to be a valuable enhancement over CBAM. Additionally, its lightweight design makes it a practical and efficient tool that can be seamlessly integrated into sewer defect detection models based on YOLOv5s.

2.2.5. Loss Function

The YOLOv5 network employs the Complete Intersection over Union (CIOU) as its localization loss function, taking into account the overlapping area, the distance between the center points, and the aspect ratio of the true and predicted bounding boxes, as shown in Equations (4)–(6).

$$L_{CIOU} = 1 - V_{IOU} + \frac{\rho^2(b, b^{gt})}{c^2} + \alpha v \tag{4}$$

$$v = \frac{4}{\pi^2}(\arctan\frac{w^{gt}}{h^{gt}} - \arctan\frac{w}{h})^2 \tag{5}$$

$$\alpha = \frac{v}{(1 - V_{IOU}) + v} \tag{6}$$

where V_{IOU} represents the Intersection over Union (IOU), ν represents the aspect ratio impact factor, α is the weight, and c represents the diagonal of the minimum rectangle containing both the true and predicted boxes. b and b^{gt} are the central points of the predicted bounding box and the ground-truth bounding box, respectively; ρ is the Euclidean distance between the center point of the predicted bounding box and the center point of the ground-truth bounding box; w^{gt} and h^{gt} are the width and height of the ground-truth boundary box, respectively; and w and h are the width and height of the predicted boundary box, respectively.

The EIOU loss function is built upon CIOU, considering the length and width influence factor of the predicted and true boxes, as shown in Equation (7).

$$L_{EIOU} = L_{IOU} + L_{dis} + L_{asp} = 1 - V_{IOU} + \frac{\rho^2(b, b^{gt})}{(w^c)^2 + (h^c)^2} + \frac{\rho^2(w, w^{gt})}{(w^c)^2} + \frac{\rho^2(h, h^{gt})}{(h^c)^2} \tag{7}$$

where L_{IOU} represents the loss of overlap between the prediction frame and the real frame, L_{dis} represents the loss of center distance between the prediction frame and the real frame, L_{asp} represents the loss of width and height of the predicted box and the real box, and w^c and h^c are the width and height of the smallest enclosing box that covers both boxes.

The EIOU loss function is composed of three parts, inheriting considerations from the loss function for the overlap area and center point distance between the true and predicted boxes. These three parts are designed to comprehensively address various aspects of bounding box localization. Below is a brief explanation of each part.

1. The first part is based on the CIOU loss, which considers the overlapping area. This component measures how well the predicted bounding box aligns with the truth bounding box.
2. The second part involves the calculation of the center point distance loss. This loss term evaluates the differences in the x and y coordinates of the center points of the predicted and true bounding boxes. Minimizing this distance contributes to accurate localization.
3. The third part deals with the aspect ratio loss. In this calculation, the lengths and widths of both boxes are separately computed. This component aims to minimize the differences in width and length between the predicted and true bounding boxes. This is particularly important in addressing variations in aspect ratio.

3. Experiment and Results Analysis

In this paper, the storm drain model dataset [32] and pipe-root dataset [33], both publicly available datasets, serve to evaluate the detection performance of the YOLOv5-Sewer model. These datasets are merged and optimized to form a comprehensive dataset capturing diverse scenarios relevant to sewer defect detection. The experiments integrate the YOLOv5s model with various enhancement methods to assess performance improvements. The evaluation process includes a thorough examination of the YOLOv5-Sewer model's detection accuracy and lightweight characteristics. Comparative analyses are conducted, contrasting the YOLOv5-Sewer model with other existing models. The assessment criteria consider not only the detection accuracy but also lightweight indicators, providing a comprehensive understanding of the model's performance.

The aim of these experiments is to objectively demonstrate the superior detection performance and lightweight characteristics of the YOLOv5-Sewer model compared to other models. The merged and optimized dataset allows for a robust evaluation of the model's capabilities in identifying sewer defects across various conditions and scenarios.

3.1. Sewer Defect Image Dataset

The storm drain model dataset, comprising 1059 images, covers various defects, including cracks, utility intrusion, obstacles, joint offset, debris, holes, and buckling, totaling 2246 annotated defects. The pipe-root dataset specifically focuses on root defects, featuring 415 images with 788 annotated defects. To maintain a 7:2:1 ratio for the training, validation, and testing sets within the storm drain model dataset, the images and annotation files from the pipe-root dataset are randomly split into corresponding sets based on the same ratio. This merging and random splitting process results in a combined dataset of 1474 images, covering eight defect categories. The training, validation, and testing sets include 949, 269, and 134 images, respectively. Partial images are shown in the following Figure 7.

The merged dataset exhibits significant variation in the annotated quantities across the different categories of pipeline defects. Notably, there are 788 annotations for roots, while holes have only 88 annotations. Due to this imbalance, the training of the YOLOv5s network on this merged dataset for 300 epochs yields mean Average Precision (mAP) of 77.9% for facility intrusion and 31.1% for holes. The substantial difference in annotation quantities can impact the training results. To mitigate this issue, data augmentation techniques, including image flipping and noise addition, are applied to images corresponding to categories with smaller annotation quantities. Simultaneously, data optimization involves removing some low-quality images related to categories with larger annotation quantities. These improvements result in a refined training set consisting of 1130 images, where the annotation quantities for various defect categories range between 400 and 500. This refined dataset aims to address imbalances and enhance the model's training performance.

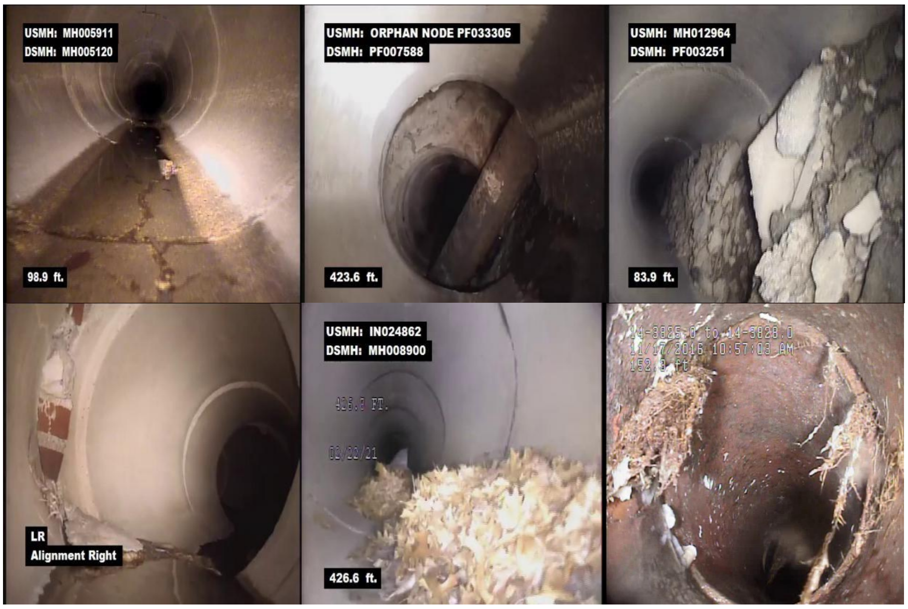

Figure 7. Partial images in the dataset.

3.2. Experimental Parameters and Experimental Configuration

The environment is configured on the Windows 11 operating system, and PyTorch is utilized as the primary framework. The specific details of the experimental environment configuration are presented in Table 1.

Table 1. Experimental environment configuration.

Name	Environment Configuration
Operating system	Windows 11
CPU	I7 12700
GPU	NVIDIA RTX 3070Ti
Memory	32G
GPU graphics memory	8G
Python version	3.8
Algorithm framework	Pytorch 1.10.0

3.3. Evaluation Indicators

The evaluation metrics include the mAP, FPS, model size, and parameters. The AP and mAP scores are calculated as shown in Equations (8) and (9). AP represents the area under the precision–recall (PR) curve, obtained by integrating the PR curve for each category. It provides the AP for each category, and mAP is the average AP across all categories.

$$V_{AP} = \int_0^1 p(r)dr \tag{8}$$

$$V_{mAP} = \frac{1}{m}\sum_{i=1}^{m} V_{AP_i} \tag{9}$$

3.4. Analysis of Experimental Results

3.4.1. Reducing the Number of Model Parameters Using MobileNetv3 Block

To address the constraints posed by the limited computing resources of on-site devices for sewer defect detection, modifications are made to enhance the model's lightweight characteristics. These adjustments include adopting MobileNetV3 block stacking as the backbone of the network. To assess the impact of this modification on the model's performance, horizontal comparative experiments are conducted with two other lightweight backbone networks: the Efficient Model (EMO) [34] and GhostNet [35].

The aim of these experiments is to compare the performance of the YOLOv5-Sewer model with MobileNetV3 block stacking against models utilizing EMO and GhostNet as backbones. The experimental results, presented in Table 2, offer insights into the relative impact of these lightweight backbone networks on the detection performance of the models. This comparative analysis is pivotal in identifying the most suitable lightweight backbone, allowing the model to achieve a lightweight design with minimal accuracy loss.

Table 2. Horizontal comparison experiments of backbone networks.

Test	Backbone			FLOPs/G	Param/M	Size/MB	mAP/%	FPS
	EMO	GhostNet	MobileNetV3					
1	×	×	×	15.8	7.03	14.5	85.5	149
2	√	×	×	29.6	4.32	9.1	82.7	93
3	×	√	×	6.0	3.26	7.0	78.6	107
4	×	×	√	5.9	3.53	7.5	80.3	136

√ denotes the use of this backbone; × denotes the absence of this backbone.

The implementation of lightweight backbones, including MobileNetV3, EMO, and GhostNet, in all three models successfully diminishes the parameter count. However, this reduction in the number of parameters has a trade-off, leading to a decrease in mean Average Precision (mAP). This is primarily attributed to the decreased number of convolutions in these lightweight backbones, leading to the weakened feature extraction capability of the backbone network.

Among the three models, the one enhanced based on GhostNet exhibits the most significant mAP reduction, reaching 6.9%. The model improved with EMO experiences the smallest decrease in mAP, which is 2.8%. However, this improvement is achieved at the cost of increased model computation and a reduction in the detection speed. The model enhanced with MobileNetV3 has a relatively low number of floating-point operations, with minor losses in accuracy and detection speed. It achieves the optimal overall performance.

To achieve a lightweight model design without excessively compromising the accuracy, further enhancements are applied to the YOLOv5s-MobileNetV3 model. These improvements involve the utilization of the C3-Faster module to reduce the parameters and computation. Attention modules are integrated, and the Efficient Intersection over Union (EIOU) is employed as the model's localization loss function to enhance the recognition accuracy. The iterative improvement process is geared towards optimizing the overall performance of the model, ensuring minimized accuracy loss while attaining a lightweight design. This makes it highly suitable for deployment on on-site operational devices dedicated to sewer defect detection.

3.4.2. Fusion Experiments

The YOLOv5s-MobileNetV3 model is sequentially integrated with C3-Faster, CA, CBAM, and the EIOU loss function for fusion experiments. They undergo training and validation on the dataset, and the results are shown in Table 3.

Table 3. Fusion experiment results.

Test	Module				FLOPs/G	Param/M	Size/MB	mAP/%	FPS
	C3-Faster	CA	CBAM	EIOU					
1	×	×	×	×	5.9	3.53	7.5	80.3	136
2	√	×	×	×	4.8	2.97	6.4	82.3	147
3	√	√	×	×	4.9	3.03	6.5	83.4	126
4	√	√	√	×	5.1	3.14	6.7	83.7	112
5	√	√	√	√	5.1	3.14	6.7	84.0	112

√ indicates the adoption of this module; × indicates the non-adoption of this module.

The experimental results illustrate that each improvement made to the YOLOv5s-MobileNetV3 model contributes positively to the final outcome. Specifically, the following outcomes are noted.

1. The integration of the C3-Faster module not only reduces the model computation, parameters, and size but also enhances the mean Average Precision (mAP) and detection speed.
2. The inclusion of attention mechanisms incurs minimal computational and parameter costs while improving the recognition accuracy.
3. The fusion of these modules results in positive optimizations across the overall performance metrics.

When all three modifications are combined, the best optimization effect is achieved, resulting in average accuracy of 83.7%. Additionally, employing the Efficient Intersection over Union (EIOU) as the model's localization loss function further improves the accuracy by 0.3%, reaching final accuracy of 84%. Despite a slight reduction in accuracy compared to the YOLOv5s model, the final model exhibits significant decreases in computation by 68%, parameters by 55%, and model size by 54%. Moreover, the detection speed reaches 112 frames per second (FPS).

These results emphasize the more lightweight design of the YOLOv5-Sewer model compared to other models and its acceptable accuracy loss. The substantial decrease in computational requirements and model size makes the optimized YOLOv5-Sewer model highly suitable for deployment on on-site operational devices for sewer defect detection, addressing real-world demands for rapid and efficient detection in sewer environments.

3.4.3. Contrast Experiments

In order to further validate the reliability of the YOLOv5-Sewer model for sewer defect detection, comparative experiments are conducted against popular object detection algorithms, including Faster-RCNN, YOLOv3-tiny [36], SSD, YOLOv7-tiny [37], and YOLOv5-GBC [17]. The experimental results, summarized in Table 4, highlight the performance of YOLOv5-Sewer in comparison to these algorithms.

Below are the key results of YOLOv5-Sewer compared to Faster-RCNN, YOLOv3-tiny, SSD, and YOLOv7-tiny on this dataset.

1. Faster-RCNN:
   - mAP improved by 7.4%;
   - Detection speed increased by 100 FPS;
   - Floating-point operations decreased by 98.6%;
   - Parameters decreased by 97.7%;
   - Model size decreased by 26.4%.
2. YOLOv3-tiny:
   - mAP improved by 6.9%;
   - Detection speed decreased by 51 FPS;
   - Floating-point operations decreased by 60.5%;
   - Parameters decreased by 63.9%;

- Model size decreased by 62.3%.

3. SSD:
    - mAP improved by 5.7%;
    - Detection speed increased by 78.3 FPS;
    - Floating-point operations decreased by 91%;
    - Parameters decreased by 88%;
    - Model size increased by 6%.
4. YOLOv7-tiny:
    - mAP decreased by 0.3%;
    - Detection speed increased by 60 FPS;
    - Floating-point operations decreased by 61.4%;
    - Parameters decreased by 47.9%;
    - Model size decreased by 45.5%.

Due to the different datasets and device models used, YOLOv5-Sewer and YOLOv5-GBC are compared in terms of lightweight indicators. The specific results are as follows:

- Floating-point operations decreased by 58.2%;
- Parameters decreased by 73.9%;
- Model size decreased by 85.4%.

Table 4. Comparison of experimental results.

Test	AP0.5/%								FLOPs/G	Param/M	Size/MB	mAP/%	FPS
	1	2	3	4	5	6	7	8					
a	70.1	86.1	70.3	80.7	84.9	62.3	79.2	79.1	370	137	9.1	76.6	12
b	70.6	88.4	70.1	79.3	83.2	65.4	81.5	78.3	12.9	8.7	17.8	77.1	163
c	71.6	90.4	73.2	86.4	86.4	60.1	87.2	70.7	63	26.3	6.3	78.3	33.7
d	72.0	92.1	83.0	90.9	94.0	72.1	92.7	77.3	13.2	6.03	12.3	84.3	52
e	77.8	96.1	80.4	92.8	88.9	69.8	90.2	76.4	5.1	3.14	6.7	84.0	112

1–8 represent cracks, utility intrusion, obstacles, joint offset, debris, holes, buckling, and roots, respectively; a–e represent Faster-RCNN, YOLOv3-tiny, SSD, YOLOv7-tiny, and YOLOv5-Sewer, respectively.

Taking into account the lightweight metrics and mean Average Precision (mAP), the YOLOv5-Sewer model achieves a lightweight design with acceptable accuracy trade-offs. It demonstrates improvements over Faster-RCNN and SSD in terms of mAP and detection speed, while also outperforming YOLOv3-tiny and Yolov7-tiny in terms of floating-point operations, parameters, and model size. Although YOLOv3-tiny has the highest detection speed, it occupies more resources and has lower accuracy. The comprehensive evaluation suggests that YOLOv5-Sewer is well suited for sewer defect detection, offering a favorable trade-off between lightweight characteristics and detection accuracy.

3.4.4. Detection Results

To enhance the realism of the on-site environment simulations, this study includes FPS comparative experiments using an Intel NUC Mini PC featuring an i5-1240P processor. The corresponding experimental results are presented in Table 5.

As indicated in Table 5, the YOLOv5-Sewer model demonstrates a notably superior detection speed compared to other models when deployed on the Intel NUC Mini PC with i5-1240P. Consequently, the lightweight model proposed in this article exhibits outstanding performance on portable devices.

The performance of the YOLOv5-Sewer model is validated on the Intel NUC Mini PC with i5-1240P and compared with other models, using the test set derived from the merged dataset. To evaluate the model's general applicability and robustness, images in the test set underwent adjustments in brightness and hue to simulate scenarios with inadequate lighting conditions. The training, validation, and testing sets included 1130, 269, and 268 images. The test results are depicted in Figure 8.

Table 5. Comparative experiments on the Intel NUC Mini PC with i5-1240P.

Model Name	FPS
Faster-RCNN	0.5
SSD	3.5
YOLOv5s	8
YOLOv7-tiny	9
YOLOv3-tiny	9
YOLOv5-Sewer	12

Model name represents the name of the defect detection model used, and FPS represents frames per second.

There are a total of 10 rows of images in Figure 8, with each two rows representing the detection results of a model. The first and second rows depict the detection results of the original image and the image under dim conditions, respectively. As shown in the image in column 7, there are underwater cracks. The dim condition image intentionally diminishes the brightness and tone of the original image. Based on the outcomes of the sewer defect detection, the YOLOv5-Sewer model can successfully identify eight types of sewer defects with a low miss rate, even in dimly lit conditions. Consequently, the YOLOv5-Sewer model exhibits strong performance in tasks related to sewer defect detection. Therefore, YOLOv5-Sewer successfully achieves model lightweighting with acceptable accuracy loss. The YOLOv5-Sewer model is more suitable for rapid defect detection on portable on-site devices.

In conclusion, the YOLOv5-Sewer model exhibits superior performance in the realm of sewer defect detection, boasting several advantages. These include a diverse range of categories, enabling the identification of eight types of sewer defects, such as cracks, utility intrusion, obstacles, joint offset, debris, holes, buckling, and roots. Additionally, the model showcases a fast detection speed, lightweight design, and acceptable accuracy on low-performance devices. Its universal applicability and robustness are evident as it successfully detects defects under dim conditions.

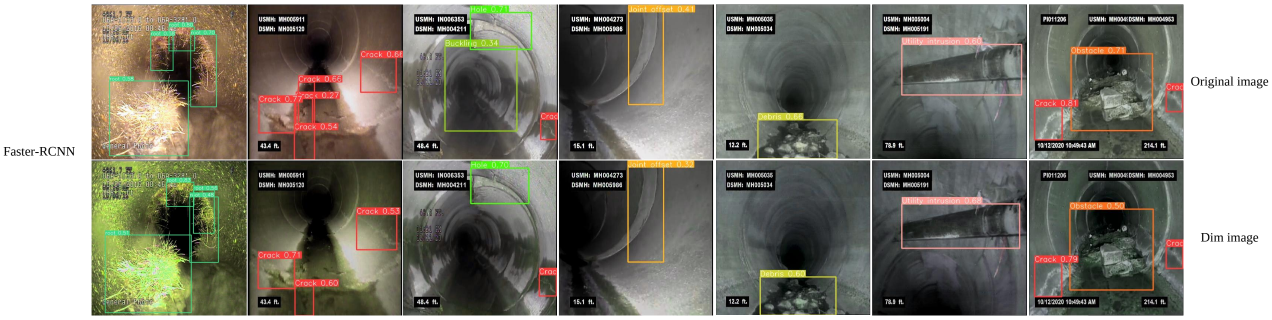

Figure 8. *Cont.*

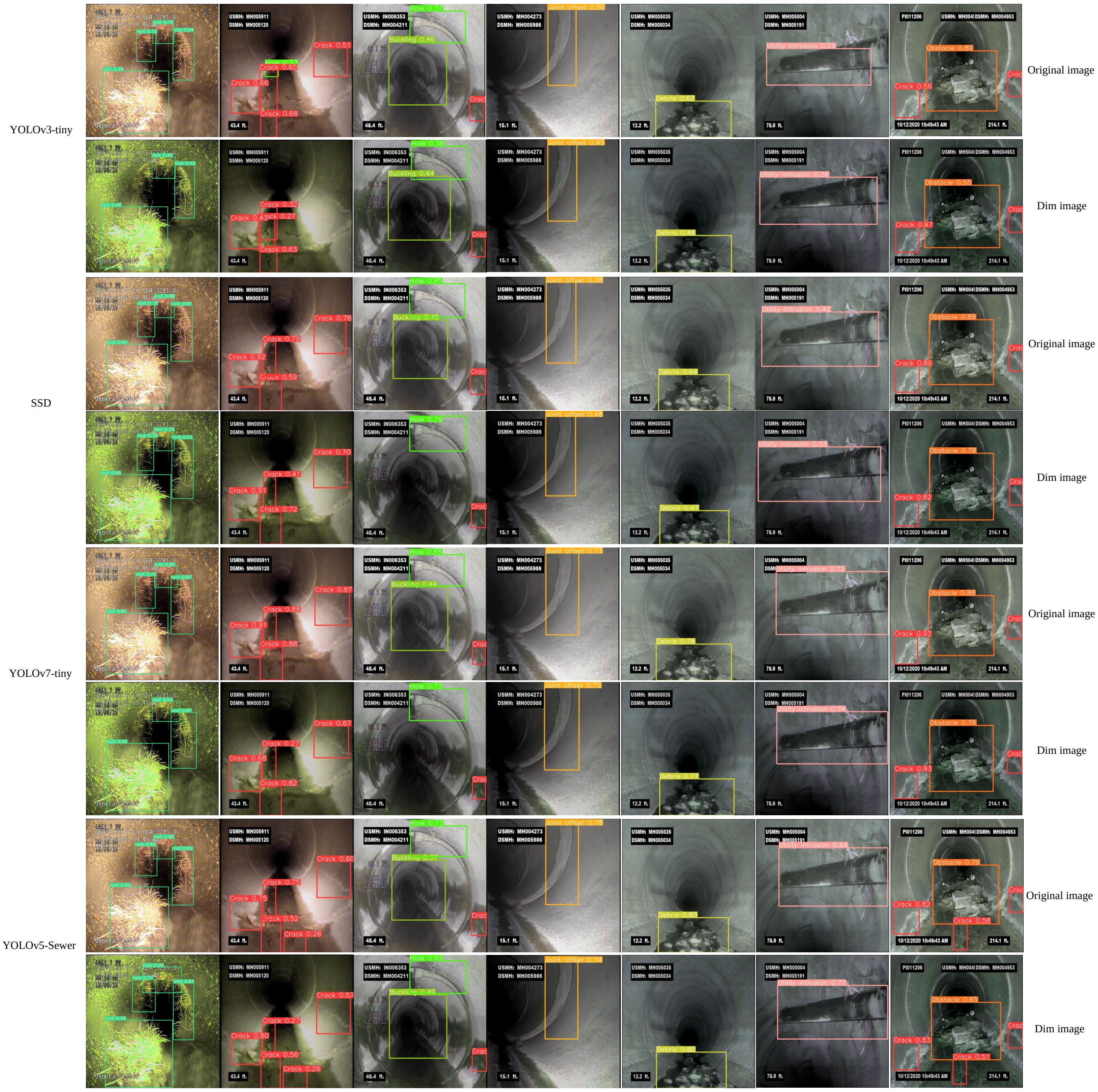

Figure 8. The detection results of sewer defects using different models, including original image and dim images.

4. Conclusions

In this paper, a lightweight and high-precision model named YOLOv5-Sewer is proposed for sewer defect detection in low-computational-capacity field devices. The model leverages lightweight MobileNetV3 block stacking as the backbone network, resulting in significant reductions in model size, parameters, and floating-point operations. C3-Faster, an improved C3 module, is introduced to further reduce redundant computations and focus on important regions, enhancing the computational efficiency. Additionally, attention modules, including CA and CBAM, are integrated to improve the feature extraction, and the EIOU loss function is employed for accurate localization, compensating for the decrease in detection accuracy due to the lightweight design. Finally, the detection performance of the YOLOv5-Sewer model is demonstrated to be close to that of the original YOLOv5s model, while outperforming other classical models in terms of model size, parameters, and the number of floating-point operations. The detection speed of this model on low-performance devices exceeds that of the original YOLOv5s.

However, this paper acknowledges that there is a certain deficiency in the training convergence speed of the YOLOv5-Sewer model. Further research is being conducted to improve the convergence speed during model training. This improvement aims to shorten the training time and improve the efficiency of the detection models in practical applications. In addition, the types of sewer defects and the number of images will be expanded in the future to ensure the universal applicability of the model. The continued reduction of the accuracy loss is also a focus of future research.

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

Funding: This research was funded by the Beijing Advanced Innovation Center for Intelligent Robots and Systems, grant number: 00921917001. This research was also funded by the BUCEA Post Graduate Innovation Project, grant number: 07081023002.

Institutional Review Board Statement: This study did not require ethical approval.

Informed Consent Statement: This study did not involve humans.

Data Availability Statement: The data presented in this study are available in the article.

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

References

1. Hawari, A.; Alamin, M.; Alkadour, F.; Elmasry, M.; Zayed, T. Automated defect detection tool for closed circuit television (cctv) inspected sewer pipelines. *Autom. Constr.* **2018**, *89*, 99–109. [CrossRef]
2. Zhang, J.; Liu, X.; Zhang, X.; Xi, Z.; Wang, S. Automatic Detection Method of Sewer Pipe Defects Using Deep Learning Techniques. *Appl. Sci.* **2023**, *13*, 4589. [CrossRef]
3. Yu, Y.; Safari, A.; Niu, X.; Drinkwater, B.; Horoshenkov, K.V. Acoustic and ultrasonic techniques for defect detection and condition monitoring in water and sewerage pipes: A review. *Appl. Acoust.* **2021**, *183*, 108282. [CrossRef]
4. Li, Y.; Wang, H.; Dang, L.M.; Song, H.K.; Moon, H. Vision-based defect inspection and condition assessment for sewer pipes: A comprehensive survey. *Sensors* **2022**, *22*, 2722. [CrossRef]
5. Czimmermann, T.; Ciuti, G.; Milazzo, M.; Chiurazzi, M.; Roccella, S.; Oddo, C.M.; Dario, P. Visual-based defect detection and classification approaches for industrial applications—A survey. *Sensors* **2020**, *20*, 1459. [CrossRef] [PubMed]
6. Ye, X.; Zuo, J.; Li, R.; Wang, Y.; Gan, L.; Yu, Z.; Hu, X. Diagnosis of sewer pipe defects on image recognition of multi-features and support vector machine in a southern Chinese city. *Front. Environ. Sci. Eng.* **2019**, *13*, 1–13. [CrossRef]
7. Myrans, J.; Everson, R.; Kapelan, Z. Automated detection of fault types in CCTV sewer surveys. *J. Hydroinform.* **2019**, *21*, 153–163. [CrossRef]
8. Kumar, S.S.; Abraham, D.M.; Jahanshahi, M.R.; Iseley, T.; Starr, J. Automated defect classification in sewer closed circuit television inspections using deep convolutional neural networks. *Autom. Constr.* **2018**, *91*, 273–283. [CrossRef]
9. Chen, K.; Hu, H.; Chen, C.; Chen, L.; He, C. An intelligent sewer defect detection method based on convolutional neural network. In Proceedings of the IEEE International Conference on Information and Automation, IEEE, Wuyishan, China, 11–13 August 2018; pp. 1301–1306.
10. Wang, M.; Cheng, J.C. A unified convolutional neural network integrated with conditional random field for pipe defect segmentation. *Comput.-Aided Civ. Infrastruct. Eng.* **2020**, *35*, 162–177. [CrossRef]
11. Pan, G.; Zheng, Y.; Guo, S.; Lv, Y. Automatic sewer pipe defect semantic segmentation based on improved U-Net. *Autom. Constr.* **2020**, *119*, 103383. [CrossRef]
12. Cheng, J.C.; Wang, M. Automated detection of sewer pipe defects in closed-circuit television images using deep learning techniques. *Autom. Constr.* **2018**, *95*, 155–171. [CrossRef]
13. Wang, M.; Cheng, J.C. Development and improvement of deep learning based automated defect detection for sewer pipe inspection using faster R-CNN. In Proceedings of the Advanced Computing Strategies for Engineering, Lausanne, Switzerland, 10–13 June 2018; Springer: Berlin/Heidelberg, Germany, 2018; pp. 171–192.
14. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.

15. Liu, R.; Ren, C.; Fu, M.; Chu, Z.; Guo, J. Platelet detection based on improved yolo_v3. *Cyborg Bionic Syst.* **2022**, *2022*, 9780569. [CrossRef]
16. Chen, K.; Li, H.; Li, C.; Zhao, X.; Wu, S.; Duan, Y.; Wang, J. An Automatic Defect Detection System for Petrochemical Pipeline Based on Cycle-GAN and YOLO v5. *Sensors* **2022**, *22*, 7907. [CrossRef] [PubMed]
17. Zhang, X.; Zhang, J.; Tian, L.; Liu, X.; Wang, S. A Lightweight Method for Detecting Sewer Defects Based on Improved YOLOv5. *Appl. Sci.* **2023**, *13*, 8986. [CrossRef]
18. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. *Adv. Neural Inf. Process. Syst.* **2015**, *28*, 1137–1149. [CrossRef]
19. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. Ssd: Single shot multibox detector. In Proceedings of the European Conference Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016; Springer: Berlin/Heidelberg, Germany, 2016; pp. 21–37.
20. Patil, R.R.; Calay, R.K.; Mustafa, M.Y.; Ansari, S.M. AI-Driven High-Precision Model for Blockage Detection in Urban Wastewater Systems. *Electronics* **2023**, *12*, 3606. [CrossRef]
21. Wang, T.; Li, Y.; Zhai, Y.; Wang, W.; Huang, R. A Sewer Pipeline Defect Detection Method Based on Improved YOLOv5. *Processes* **2023**, *11*, 2508. [CrossRef]
22. Huang, Q.; Zhou, Y.; Yang, T.; Yang, K.; Cao, L.; Xia, Y. A Lightweight Transfer Learning Model with Pruned and Distilled YOLOv5s to Identify Arc Magnet Surface Defects. *Appl. Sci.* **2023**, *13*, 2078. [CrossRef]
23. Howard, A.; Sandler, M.; Chu, G.; Chen, L.C.; Chen, B.; Tan, M.; Wang, W.; Zhu, Y.; Pang, R.; Vasudevan, V.; et al. Searching for mobilenetv3. In Proceedings of the International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 1314–1324.
24. Chen, J.; Kao, S.h.; He, H.; Zhuo, W.; Wen, S.; Lee, C.H.; Chan, S.H.G. Run, Don't Walk: Chasing Higher FLOPS for Faster Neural Networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023; pp. 12021–12031.
25. Yu, J.; Wang, C.; Xi, T.; Ju, H.; Qu, Y.; Kong, Y.; Chen, X. Development of an Algorithm for Detecting Real-Time Defects in Steel. *Electronics* **2023**, *12*, 4422. [CrossRef]
26. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. Cbam: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision, Munich, Germany, 8–14 September 2018; pp. 3–19.
27. Hou, Q.; Zhou, D.; Feng, J. Coordinate attention for efficient mobile network design. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 13713–13722.
28. Zhang, Y.F.; Ren, W.; Zhang, Z.; Jia, Z.; Wang, L.; Tan, T. Focal and efficient IOU loss for accurate bounding box regression. *Neurocomputing* **2022**, *506*, 146–157. [CrossRef]
29. Yougao, L.; Wei, H. Identification and feature extraction of drainage pipeline cracks based on SVD and edge recognition method. In Proceedings of the Electronic Information Technology and Computer Engineering, IEEE, Xiamen, China, 18–20 October 2019; pp. 1184–1188.
30. Wang, Z.; Jin, L.; Wang, S.; Xu, H. Apple stem/calyx real-time recognition using YOLO-v5 algorithm for fruit automatic loading system. *Postharvest Biol. Technol.* **2022**, *185*, 111808. [CrossRef]
31. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7132–7141.
32. new-workspace zyqyt. Storm Drain Model Dataset. 2021. Available online: https://universe.roboflow.com/new-workspace-zyqyt/storm-drain-model (accessed on 10 January 2024).
33. rootdataset. Pipe-Root Dataset. 2023. Available online: https://universe.roboflow.com/rootdataset/pipe_root (accessed on 10 January 2024).
34. Zhang, J.; Li, X.; Li, J.; Liu, L.; Xue, Z.; Zhang, B.; Jiang, Z.; Huang, T.; Wang, Y.; Wang, C. Rethinking mobile block for efficient neural models. *arXiv* **2023**, arXiv:2301.01146.
35. Han, K.; Wang, Y.; Tian, Q.; Guo, J.; Xu, C.; Xu, C. Ghostnet: More features from cheap operations. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 1580–1589.
36. Redmon, J.; Farhadi, A. Yolov3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767.
37. 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. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023; pp. 7464–7475.

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.