

Article

A Two-Stage Multi-Label Sewer Defect Identification Approach Using TResNet and CSRA Attention Module

Sunuan Wang, Chi Ma * and Zhongyuan Du

College of Mechanical and Electrical Engineering, Shaanxi University of Science & Technology, Xi'an 710021, China; wangsunuan@sust.edu.cn (S.W.)

* Correspondence: machi@sust.edu.cn

Abstract

To address the coexistence of multiple defect types and class imbalance in complex urban sewer scenarios, this study proposes a two-stage multi-label recognition method for sewer defect images. In the first stage, a lightweight binary convolutional neural network is used to distinguish between normal and defective images, thereby filtering out normal samples and reducing the computational burden. In the second stage, a TResNet-based multi-label classification model is developed, and a class-specific residual attention module is introduced to enhance defect-specific spatial feature representation. In addition, F1-Normal and F2-CIW are used to evaluate the recognition performance from both normal-pipeline recognition and defect-economic-impact perspectives. Experiments on the Sewer-ML dataset show that the proposed method achieves 90.88% F1-Normal and 53.93% F2-CIW, outperforming the baseline models in overall performance.

Keywords: sewer defect recognition; multi-label classification; TResNet; CSRA

1. Introduction

Urban sewer pipelines are an important component of urban infrastructure. China has a vast number of pipelines, many of which have been in service for a long time. Aging pipe networks are prone to leakage, collapse, and other accidents, thereby adversely affecting the environment, the economy, and public safety. Regular inspection and maintenance are therefore essential. At present, inspection mainly relies on CCTV technology, whereby pipeline robots acquire images that are subsequently interpreted by professional inspectors. However, manual inspection involves a heavy workload, low efficiency, and subjective bias. With the development of deep learning in image processing, automatic sewer defect recognition has gradually become feasible, and relevant studies have already achieved certain progress.

The internal environment of urban pipelines is highly complex, and different types of pipelines exhibit different structural characteristics, as shown in Figure 1. Compared with traditional digital image-processing methods based on manually extracted features, deep learning methods that automatically learn features perform better in image recognition tasks. Researchers both in China and abroad have applied such methods to sewer defect image recognition. Hassan et al. [1] used the classical AlexNet [2] for sewer image defect recognition. Similarly, Li et al. [3] applied an improved ResNet-18 [4] to end-to-end sewer defect recognition. Unlike the end-to-end methods of Hassan and Li et al., Chen Kefan et al. [5] divided sewer image defect recognition into two stages: in the first stage, a lightweight SqueezeNet [6] was used to determine whether an image contained a defect,



Academic Editor: Christos Bouras

Received: 5 June 2026

Revised: 30 June 2026

Accepted: 6 July 2026

Published: 9 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

and in the second stage, InceptionV3 [7] was used to identify the specific defect type. Xie et al. [8] likewise adopted a two-stage method for sewer image recognition.

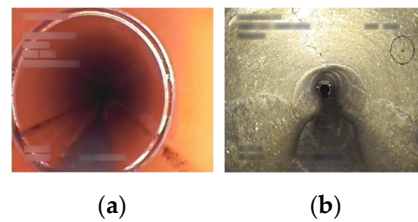


Figure 1. Sewer defect images: (a) deformation and displaced joint; (b) fracture, surface damage, and displaced joint.

Previous studies have mainly focused on single-defect recognition, whereas real-world images often contain multiple defects simultaneously, which constitutes a multi-label classification problem. Kumar et al. [9] trained an ensemble of small two-layer CNNs in a one-versus-one manner to classify four defect types. Zeng Baozhi et al. [10] proposed an improved label-graph-based GCN model for classifying seven defect types. Haurum et al. [11], after comparative analysis, selected TResNet [12] for multi-label classification and recognition.

In previous studies on multi-label sewer defect recognition, problems remained such as incomplete coverage of defect categories and insufficient recognition accuracy. To further address multi-label defect classification in complex urban sewer scenarios, this paper proposes a two-stage recognition method. In the first stage, a small CNN model is designed to determine whether defects exist in sewer images. In the second stage, a class-specific residual attention module (CSRA) [13] is introduced into the TResNet model to recognize the defect types present in sewer images.

2. Methodology

In real sewer image datasets, the number of normal samples is significantly larger than that of defective samples, resulting in a markedly imbalanced class distribution. To improve recognition performance, defect images are first preprocessed, and the recognition task is divided into two stages: normal-abnormal recognition and defect-type recognition. In the first stage, all defect categories are merged into a single abnormal class, which forms a binary classification problem together with the normal class, and a binary CNN model is used for discrimination. In the second stage, abnormal samples are subjected to multi-label defect classification using an improved TResNet network. The overall recognition procedure is shown in Figure 2.

2.1. Image Preprocessing

The original images have different resolutions, and all are uniformly resized to 224×224 pixels. To prevent overfitting and augment the data, horizontal flipping is applied with a probability of 50%; brightness, contrast, saturation, and hue are jittered within $\pm 10\%$; and normalization is performed. During inference, the images are likewise resized to 224×224 and normalized.

2.2. Normal-Abnormal Recognition Model

To address the severe class imbalance between normal and defective samples in sewer datasets, a lightweight convolutional neural network is designed for binary classification. This stage acts as a fast filter to remove defect-free images, reducing computational burden and improving the efficiency of subsequent multi-label recognition.

The structure of the binary classification network is shown in Figure 3. The network adopts a simple and stable structure with three convolutional layers and three fully connected layers. Max-pooling is applied after each convolution to downsample features and suppress overfitting. Zero padding is used to preserve edge information, where small defects often appear. Adaptive average pooling is employed before fully connected layers, and ReLU activation and dropout are adopted to further improve generalization.

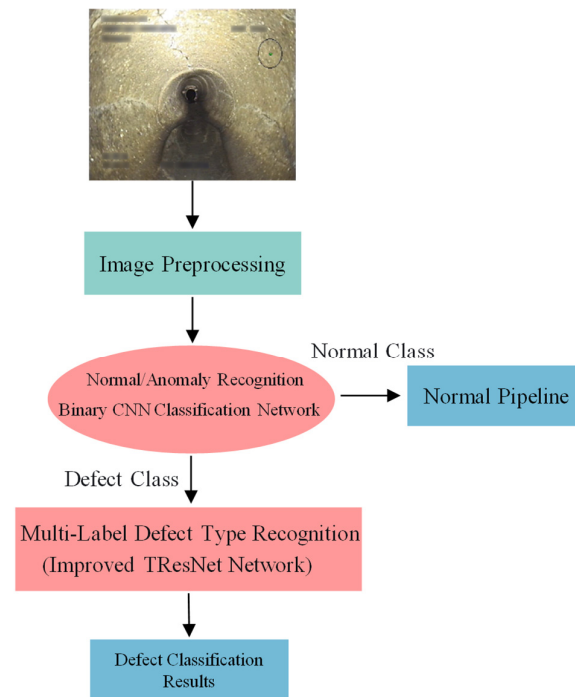


Figure 2. Sewer defect recognition method.

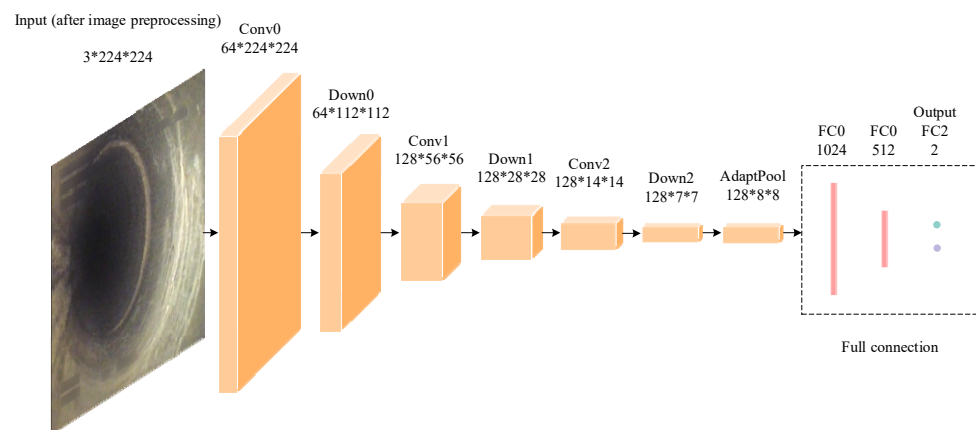


Figure 3. Binary classification network.

The model outputs two probabilities corresponding to the normal class and the abnormal (defective) class, respectively. It achieves efficient filtering with low computation, making it suitable for real-world sewer inspection systems.

2.3. Defect Type Recognition Model

TResNet exhibits fast GPU training and inference efficiency in practice [11] and has achieved excellent performance in multi-label image classification. To further improve sewer defect recognition accuracy under complex urban sewer scenarios, a CSRA module

specifically designed to improve multi-label classification performance is introduced at the end of the TResNet network for multi-label classification of sewer defect images.

2.3.1. TResNet Architecture

In recent years, many models have surpassed ResNet50 [4] in ImageNet accuracy, but their actual GPU throughput is lower, whereas ResNet50 offers a better balance between throughput and accuracy. TResNet is an improved architecture based on ResNet50 and achieves better performance under similar throughput. It includes three variants—M, L, and XL—and this study adopts TResNet-L. Its main improvements include the SpaceToDepth transformation layer, the Inplace-ABN layer, SE blocks, and deployment optimization.

ResNet50 initially uses a 7×7 convolution with stride 2 and a pooling layer for feature extraction and $4\times$ downsampling. TResNet replaces this module with a SpaceToDepth transformation layer [14] to reduce feature-map size, aiming to reduce the information loss caused by aggressive downsampling and thereby improve accuracy. The replacement relationship is shown in Figure 4.

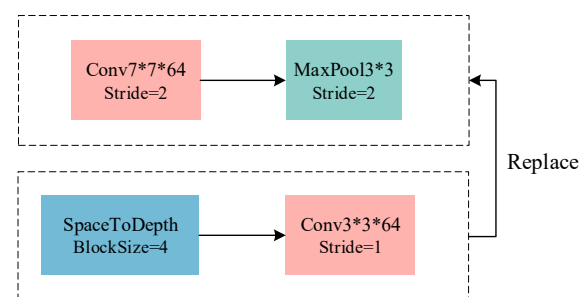


Figure 4. SpaceToDepth replacement relation.

TResNet uses an Anti-Alias Downsampling (AA) layer [15] to replace the original downsampling module, as shown in Figure 5.

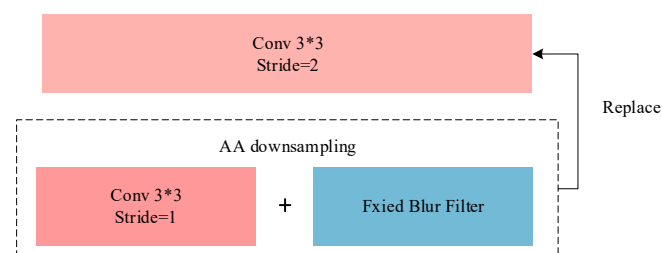


Figure 5. AA Downsampling replacement relation.

TResNet replaces all BatchNorm + ReLU layers with Inplace-ABN layers [16], substantially reducing memory requirements during training and thereby significantly increasing the feasible batch size. The activation function is changed from standard ReLU to Leaky-ReLU.

ResNet contains two basic building blocks: BasicBlock and Bottleneck. Bottleneck consists of two 1×1 convolutions and one 3×3 convolution, and generally yields higher accuracy. BasicBlock consists of two 3×3 convolutions, offers a larger receptive field, and requires fewer GPU resources. TResNet uses BasicBlock in the first two stages and Bottleneck in the last two stages to achieve a balance between speed and accuracy.

TResNet further adds Squeeze-and-Excitation (SE) blocks [17] to both building blocks to improve accuracy. The reduction ratio is set to 4 in BasicBlock and 8 in Bottleneck. The two block designs are shown in Figure 6.

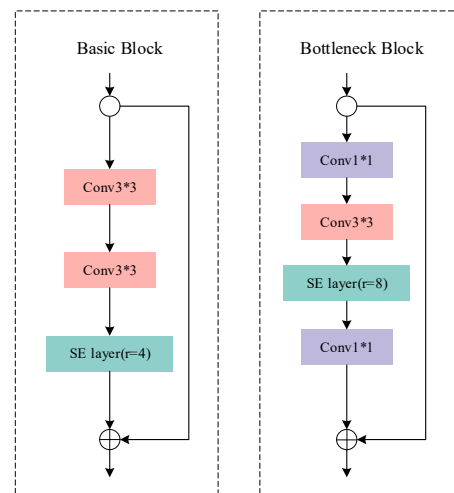


Figure 6. BasicBlock layer and Bottleneck layer design of TRResNet.

For a preprocessed sewer defect image I , sending it into the feature extractor Φ (TRResNet-L) yields a feature tensor $p \in \mathbb{R}^{d \times h \times w}$, where d , h , and w denote the channel dimension, height, and width of the tensor, respectively.

$$p = \phi(I; \theta) \quad (1)$$

In Equation (1), θ denotes the parameters of the TRResNet-L backbone. For an input sewer image with a resolution of $3 \times 224 \times 224$, the output feature tensor after TRResNet-L has a shape of $2432 \times 7 \times 7$.

2.3.2. Class-Specific Residual Attention Module (CSRA)

CSRA generates class-specific features for each category through a spatial attention scoring mechanism and combines them with class-agnostic average-pooled features. In this way, it can effectively capture the different spatial regions occupied by different classes of defects, thereby improving the performance of multi-label sewer defect classification.

After a sewer image passes through the TRResNet-L feature extractor, the output is a feature tensor of size $2432 \times 7 \times 7$, which can be decomposed into $x_1, x_2, \dots, x_{49}$ ($x_k \in \mathbb{R}^{2432}$). For i -th class ($i \in \mathbb{R}^{17}$), the class-specific attention score at j -th position of the feature tensor p ($j \in \mathbb{R}^{49}$) is defined as follows [13]:

$$s_j^i = \frac{\exp(Tx_j^T m_i)}{\sum_{k=1}^{49} \exp(Tx_k^T m_i)} \quad (2)$$

In Equation (2), T is a manually adjustable hyperparameter, and $m_i \in \mathbb{R}^{2432}$ denotes the classifier for the i -th class. The score s_j^i can be interpreted as the probability that the i -th class appears at position j .

After obtaining the probability that the i -th class appears at the j -th position, this probability is used to weight the original features, followed by a global summation over the corresponding dimension, yielding the class-specific feature vector [13] a^i :

$$a^i = \sum_{k=1}^{49} s_k^i x_k \quad (3)$$

In Equation (3), s_k^i denotes the attention score of the i -th class, and x_k denotes the vector decomposed from the feature map at each position.

The global feature of the feature map g can be obtained by averaging the features over all positions [18]:

$$g = \frac{1}{49} \sum_{k=1}^{49} x_k \quad (4)$$

Since the global feature vector g has been widely used and has shown good effectiveness, it is treated as the main feature vector, while a^i is treated as a residual component. After assigning a weight coefficient λ to the residual component a^i ($\lambda = 0.4$ in this paper) and adding it to g , the residual attention feature f^i for the i -th class is obtained [13]:

$$f^i = g + \lambda a^i \quad (5)$$

Finally, all such residual attention features f_i are multiplied by the classifier m_i of each class to obtain the final score vector [13]:

$$\hat{y} \triangleq (y^1, y^2, \dots, y^c) = (m^1 f^1, m^2 f^2, \dots, m^c f^c) \quad (6)$$

In Equation (6), C denotes the number of classes.

The above describes the single-head attention mechanism of CSRA. CSRA can be further extended to a multi-head attention mechanism, in which multiple different CSRA modules are combined additively, thereby yielding better classification results. The principle of this module is illustrated in Figure 7. The final score vector of the multi-head attention mechanism [13] y_s is given by:

$$\hat{y}_s = \sum_{h=1}^H y_{T_i} \quad (7)$$

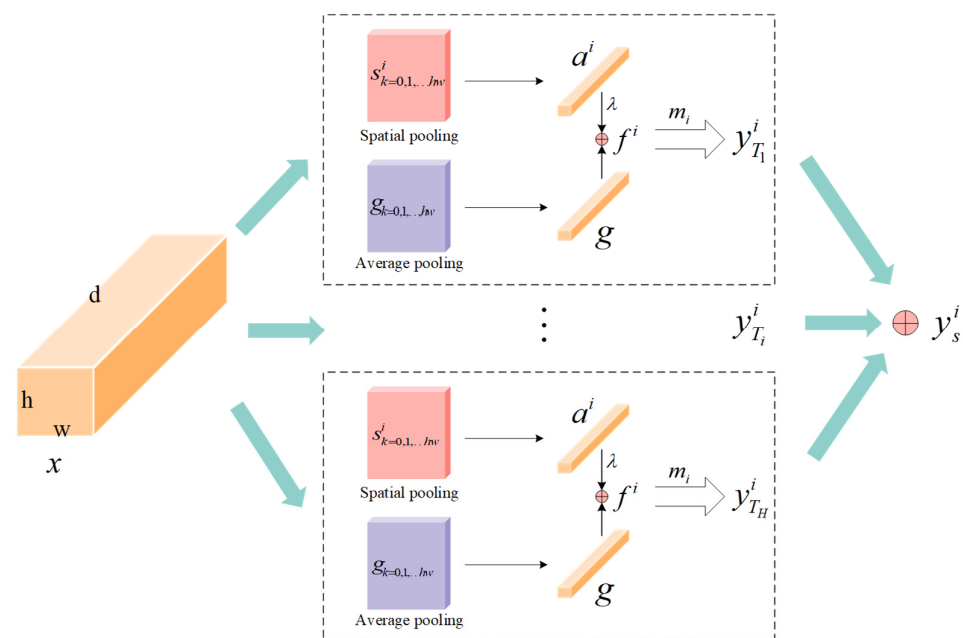


Figure 7. Multi-head CSRA module.

In Equation (7), T_i denotes the hyperparameter of the i -th attention head. After extending CSRA to a multi-head attention mechanism, the hyperparameters T_i can be determined according to the number of heads H . In this paper, a multi-head attention setting with $H = 4$ is adopted. According to Ref. [13], the hyperparameters of the attention

heads are chosen as $T_{1:4} = 1, 2, 4, \infty$. Among them, $T_4 = \infty$ has an effect equivalent to that of a global max-pooling layer.

The defect recognition model takes preprocessed images as input and uses TResNet-L for feature extraction. The backbone consists of five stages: the first stage uses SpaceToDepth for downsampling; the second and third stages repeat BasicBlock 4 and 5 times, respectively; the fourth stage repeats Bottleneck 18 times; and the fifth stage repeats Bottleneck 3 times. A four-head CSRA attention module is then used to capture spatial positional information and output the category prediction logits, as shown in Figure 8.

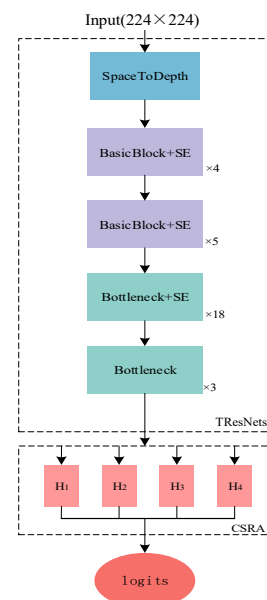


Figure 8. Multi-label classification network for defect type.

2.3.3. Loss Function

This paper adopts binary cross-entropy loss [19], which is commonly used in multi-label image classification, as shown in Equation (8):

$$L = \frac{1}{C} \sum_c -[w_c y_c \log(p_c) + (1 - y_c) \log(1 - p_c)] \quad (8)$$

In Equation (8), c denotes the number of defect categories; y_c denotes whether the current image contains defect class c ; p_c denotes the probability that the model predicts class c as positive; and w_c denotes the class weight for class c in the current image.

To address the imbalance in the number of different defect types in sewer images, different classes are weighted by w_c in the loss function, so that minority classes receive higher weights while majority classes receive lower weights, as shown in Equation (9):

$$w_c = \frac{N - N_c}{N_c} \quad (9)$$

In Equation (9), N denotes the number of images in the training set, and N_c denotes the number of images in the training set that contain the corresponding defect class.

3. Experimental Results and Analysis

To evaluate the performance of the improved model on the multi-label classification task for sewer defects, comparative experiments were conducted on the Sewer-ML dataset [11]. This section introduces the dataset, experimental environment, pa-

parameter settings, and evaluation metrics, and presents comparative analyses with representative methods.

3.1. Dataset Description

A total of 131,631 sewer images are used in this study, including 65,751 normal images and 65,880 defective images. The data split is shown in Table 1.

Table 1. Dataset distribution.

Type	Training Set	Test Set	Validation Set	Total
Normal images	51,909	6881	6961	65,751
Defect images	55,463	5325	5092	65,880
Total	107,372	12,206	12,053	131,631

The defects cover 17 categories [20], including cracks, surface damage, displaced joints, and others. In terms of sample distribution, surface damage, displaced joints, and construction-change-related defects account for relatively high proportions, specifically 24.68%, 41.19%, and 18.73%, respectively, whereas lateral reinstatement cuts and transition-profile-related defects account for relatively low proportions, specifically 0.44% and 0.08%, respectively. Considering that different defect types exert distinct economic impacts, each defect category should be assigned a corresponding economic impact score, which is determined by professional domain experts. Various pipeline defect datasets differ substantially in defect categories, annotation criteria, sample distribution, and application scenarios. Consequently, the category importance weight (CIW) exhibits strong dataset specificity and is not applicable for direct cross-dataset transfer. When adopting a new dataset, the CIW should be recalibrated comprehensively based on pipeline maintenance costs, safety risk levels, and local industrial inspection specifications. Based on economic impact scores, class weights (CIW) are assigned to each category, as shown in Table 2 [21].

Table 2. The defect type and its corresponding CIW.

Code	Defect Type	CIW
RB	Cracks, breaks, and collapses	1.0000
OB	Surface damage	0.5518
PF	Production error	0.2896
DE	Deformation	0.1622
FS	Displaced joint	0.6419
IS	Intruding sealing material	0.1847
RO	Roots	0.3559
IN	Infiltration	0.3131
AF	Settled deposits	0.0811
BE	Attached deposits	0.2275
FO	Obstacle	0.2477
GR	Branch pipe	0.0901
PH	Chiseled connection	0.4167
PB	Drilled connection	0.4167
OS	Lateral reinstatement cuts	0.9009
OP	Connection with transition profile	0.3829
OK	Connection with construction changes	0.3829

3.2. Experimental Settings

The main hardware configuration used in this experiment is as follows: CPU: Intel(R) Core(TM) i5-12400F; memory: 16 GB; GPU: NVIDIA GeForce RTX3060 with 12 GB video memory. The experiments were conducted under the Windows 10 operating system using

the PyTorch 1.13.1 and CUDA 11.7 deep learning framework. SGD with momentum was used for training. The maximum number of iterations was set to 100, and the batch size was set to 64. The learning rate was set to 0.1, the momentum to 0.9, and the weight decay to 0.0001. The learning rate was multiplied by 0.1 at the 30th, 60th, and 80th epochs.

3.3. Evaluation Metrics

The F_β score [22], which comprehensively considers both precision and recall, is commonly used to evaluate model performance. Its calculation is given in Equation (10):

$$F_\beta = (1 + \beta^2) \frac{\text{Prc} \cdot \text{Rcll}}{\beta^2 \text{Prc} + \text{Rcll}} \quad (10)$$

In Equation (10), Prc denotes model precision, Rcll denotes the recall of the model, and β is the weight assigned to recall.

In practical sewer image defect inspection, identifying images that contain defects is often more important than identifying the defect types within the images; that is, the recall of the model is more important than its precision. If both cannot be satisfied simultaneously, a higher weight should be assigned to recall. Therefore, $\beta = 2$ is adopted in this paper. Considering the economic impact of each defect type, the CIWs in Table 2 are incorporated into the F_2 evaluation metric, yielding the $F_{2\text{-CIW}}$ metric [11]:

$$F_{2\text{-CIW}} = \frac{\sum_{c=1}^C F_{2\text{-c}} \cdot \text{CIW}_c}{\sum_{c=1}^C \text{CIW}_c} \quad (11)$$

In Equation (11), CIW_c and $F_{2\text{-c}}$ denote the category importance weight and F_2 score of class c , respectively, and C denotes the number of defect categories. Because the normal class has no CIW, this paper simultaneously uses the F_1 -Score to evaluate normal pipelines, hereinafter referred to as $F_{1\text{-Normal}}$.

3.4. Comparative Experiments

ResNet-101 addresses the issues of gradient vanishing and model degradation prevalent in deep learning, and achieves outstanding performance in image classification tasks. It has become a baseline algorithm in the general field of image classification. To better verify the effectiveness of the proposed method, we conduct performance comparisons under identical experimental conditions between our algorithm, the representative original TResNet-L, and existing mainstream object detection algorithms including YOLOv8-L and RT-DETR v2. The detailed quantitative evaluation results of model performance are shown in Table 3.

Table 3. Comparison of experimental results.

Algorithm	$F_{1\text{-Normal}}$	$F_{2\text{-CIW}}$	Param/M	FPS
ResNet-101	90.78	50.46	90.31	50.03
TResNet-L	90.81	51.38	90.35	51.06
YOLOv8-L	89.21	50.77	43.7	42.5
RT-DETR v2	89.74	51.29	111.6	29.3
Ours	90.88	53.93	90.30	54.11

As can be seen from Table 3, the proposed method embeds a multi-head CSRA attention module into the TResNet model to focus on the information of different defect categories at diverse positions. It achieves better multi-label defect recognition capability

and superior model performance than YOLOv8-L, RT-DETR v2 and ResNet-101. The improvement in F_{2-CIW} indicates that the CSRA module is beneficial for economically important defect categories. By introducing class-specific spatial attention, the model can focus on different defect regions for different labels, which is particularly suitable for sewer images where multiple defects may coexist.

3.5. Performance of Individual Defect Classes

A comparison between model performance and economic impact is shown in Figure 9. The model performs well on FS, OB, and OK defects, which have both relatively high economic impact and high occurrence frequency. Functional defect classes such as FO, IS, and RO are recognized less effectively because of their more complex features, but their economic impact is relatively low. By contrast, RB and OS, which have the highest economic impact, still exhibit insufficient recognition accuracy. Improving the classification capability for these two defect types is therefore crucial for enhancing the overall score.

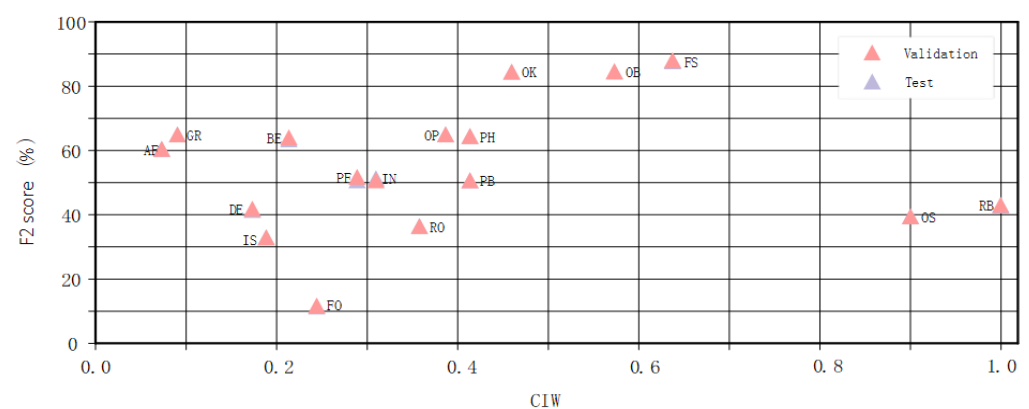


Figure 9. Class-wise performance of different defect categories.

To explore the effect of the number of attention heads H of the multi-head CSRA module on the multi-label recognition performance of underground sewer pipelines, we embed CSRA modules with different head numbers ($H = 1, 2, 4, 6$) into the proposed improved algorithm for experiments. The corresponding F_{2-CIW} test scores under different head numbers are illustrated in Figure 10. It can be seen that increasing the number of attention heads H obviously improves the multi-label defect recognition performance of pipeline images, and the model achieves the optimal performance when $H = 4$.

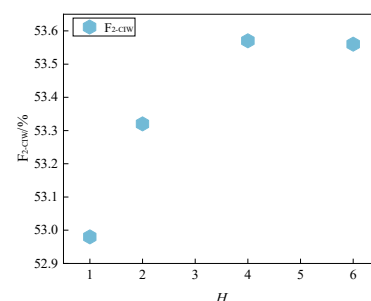


Figure 10. Effect of the number of CSRA heads on F_{2-CIW} performance.

3.6. Ablation Experiments

This experiment adopts a two-stage defect recognition method. In the first stage, all sewer defects are grouped into one category, and a binary classifier is used to judge the presence or absence of defects. In the second stage, fine-grained classification of specific defect types is carried out. To verify the effectiveness of the two-stage strategy

and explore the performance improvement brought by each module to the overall model, we design progressive ablation experiments. Each experiment adds a new module to the baseline model, and the model is evaluated via multiple metrics including $F_{1\text{-Normal}}$, $F_{2\text{-CIW}}$, Param/M and FPS. The relevant details are shown in Figure 11 and Table 4.

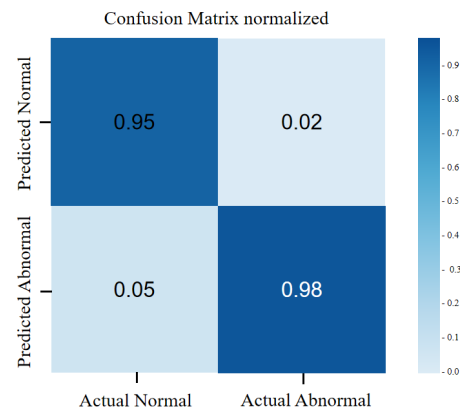


Figure 11. One-stage Classification Confusion Matrix.

Table 4. Ablation study results.

Method	$F_{1\text{-Normal}}/\%$	$F_{2\text{-CIW}}/\%$	Param/M	FPS
TResNet	78.67	50.76	89.2	53.1
TResNet + Stage 1 Filtering	88.75	51.36	89.8	52.7
TResNet + CSRA	85.74	53.12	90.1	52.3
Ours	90.88	53.93	90.3	54.1

Figure 11 shows that the recall rate of the model for abnormal samples is as high as 98%, and the missing detection rate is only 2%. In practical engineering applications, recall is more important than precision for the model. Overall, the model achieves superior classification performance on both positive and negative samples, and prioritizes the detection capability of abnormal samples, which conforms to the practical application requirements of this study. As shown in Table 4, the two-stage method achieves a significant improvement in $F_{1\text{-Normal}}$ scores, and the CSRA attention mechanism contributes positively to the promotion of $F_{2\text{-CIW}}$ scores. However, the improvement of $F_{2\text{-CIW}}$ is relatively marginal. Since it targets the multi-label classification of 17 defect categories, its performance is constrained by small-sample defects, complex defect features, and difficult-to-identify defects with high economic weights. The binary classification stage cannot break through the bottleneck of fine-grained defect classification, resulting in only limited performance gains.

4. Conclusions

Urban sewer defects seriously affect public safety and economic development. This paper proposes a two-stage multi-label classification algorithm. In the first stage, a binary classification network is designed to determine whether an image contains defects. In the second stage, TResNet is improved by introducing a CSRA attention module to enhance the classification performance of each defect category. Compared with the original TResNet-L, YOLOv8-L and other algorithms, the proposed model achieves superior multi-label recognition performance on the Sewer-ML sewer dataset.

Author Contributions: Conceptualization, S.W. and C.M.; methodology, S.W.; software, S.W.; validation, S.W., C.M. and Z.D.; formal analysis, S.W.; investigation, S.W.; resources, C.M.; data curation, S.W.; writing—original draft preparation, S.W.; writing—review and editing, C.M. and Z.D.; visual-

ization, S.W.; supervision, C.M.; project administration, C.M.; funding acquisition, C.M. All authors have read and agreed to the published version of the manuscript.

Funding: The authors thanks for supports of the Funded Project of Key Research and Development Plan of Xianyang City, Shaanxi Province L2025-ZDYF-KTH-002 and L2023-ZDYF-QYCX-018.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The Sewer-ML dataset analyzed in this research is a publicly available open dataset introduced in the published literature, which can be accessed from the official open-source channel stated in Ref. [11]. The trained model weights, experimental source code and intermediate processed data generated in this study are not publicly archived at present but are available from the corresponding author upon reasonable academic request.

Acknowledgments: The authors sincerely appreciate the open-source Sewer-ML dataset providers for publicly sharing the sewer defect image data, which provides reliable experimental resources for this research. We also thank the developers of the open-source PyTorch deep learning framework and the original authors of TResNet and CSRA modules for releasing relevant algorithm codes to support our model implementation. In addition, we are grateful to the anonymous reviewers and editorial staff for their valuable suggestions to improve the quality of this paper. During the preparation of this manuscript, the authors used AI writing auxiliary tools for partial language polishing and expression optimization. The authors have reviewed and edited all generated content carefully and take full responsibility for the content of this publication.

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

References

1. Hassan, S.I.; Dang, L.M.; Mehmood, I.; Im, S.; Choi, C.; Kang, J.; Park, Y.S.; Moon, H. Underground sewer pipe condition assessment based on convolutional neural networks. *Autom. Constr.* **2019**, *106*, 102849. [\[CrossRef\]](#)
2. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **2012**, *25*, 1097–1105.
3. Li, D.; Cong, A.; Guo, S. Sewer damage detection from imbalanced CCTV inspection data using deep convolutional neural networks with hierarchical classification. *Autom. Constr.* **2019**, *101*, 199–208. [\[CrossRef\]](#)
4. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*; IEEE: New York, NY, USA, 2016; pp. 770–778.
5. Chen, K. Research on Detecting Method of Sewer Defects Based on Vision. Master's Thesis, Harbin Institute of Technology, Shenzhen, China, 2019.
6. Iandola, F.N.; Han, S.; Moskewicz, M.W.; Ashraf, K.; Dally, W.J.; Keutzer, K. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5 MB model size. *arXiv* **2016**, arXiv:1602.07360.
7. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*; IEEE: New York, NY, USA, 2016; pp. 2818–2826.
8. Xie, Q.; Li, D.; Xu, J.; Yu, Z.; Wang, J. Automatic detection and classification of sewer defects via hierarchical deep learning. *IEEE Trans. Autom. Sci. Eng.* **2019**, *16*, 1836–1847. [\[CrossRef\]](#)
9. 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. [\[CrossRef\]](#)
10. Zeng, B.; Luo, J.; Xiong, Y.; Li, B. Graph Convolutional Network Detection Model for Pipeline Defects Based on Improved Label Graph. *Laser Optoelectron. Prog.* **2022**, *59*, 64–72. [\[CrossRef\]](#)
11. Haurum, J.B.; Moeslund, T.B. Sewer-ML: A multi-label sewer defect classification dataset and benchmark. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*; IEEE: New York, NY, USA, 2021; pp. 13456–13467.
12. Ridnik, T.; Lawen, H.; Noy, A.; Baruch, E.B.; Sharir, G.; Friedman, I. Tresnet: High performance gpu-dedicated architecture. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*; IEEE: New York, NY, USA, 2021; pp. 1400–1409.
13. Zhu, K.; Wu, J. Residual attention: A simple but effective method for multi-label recognition. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*; IEEE: New York, NY, USA, 2021; pp. 184–193.

14. Sandler, M.; Baccash, J.; Zhmoginov, A.; Howard, A. Non-discriminative data or weak model? On the relative importance of data and model resolution. In *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*; IEEE: New York, NY, USA, 2019; pp. 1036–1044.
15. Zhang, R. Making Convolutional Networks Shift-Invariant Again. *arXiv* **2019**, arXiv:1904.11486.
16. Buló, S.R.; Porzi, L.; Kotschieder, P. In-place activated batchnorm for memory-optimized training of DNNs. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*; IEEE: New York, NY, USA, 2018; pp. 5639–5647.
17. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*; IEEE: New York, NY, USA, 2018; pp. 7132–7141.
18. Lin, M.; Chen, Q.; Yan, S. Network in network. *arXiv* **2013**, arXiv:1312.4400.
19. Brownlee, J. *Probability for Machine Learning: Discover How to Harness Uncertainty with Python*; Machine Learning Mastery: Austin, TX, USA, 2019.
20. DANSK Vand-og Spildevandsforening. *Fotomanualen: TV-Inspektion af Afløbsledninger*; Dansk Vand og Spildevandsforening (DANVA): Skanderborg, Denmark, 2010; Volume 6, pp. 8–56.
21. DANSK Vand-og Spildevandsforening. *Fotomanualen: Beregning af Fysisk Indeks ved TV-inspektion*; Dansk Vand og Spildevandsforening (DANVA): Skanderborg, Denmark, 2005; Volume 3, pp. 9–16.
22. Van Rijsbergen, C.J. *Information Retrieval*; Butterworth-Heinemann: London, UK, 1979.

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.