

## RESEARCH ARTICLE

# Cable Tunnel Waterlogging Detection for Low-Light and Interference Based on Three-Stage Recognition

JUN ZHU<sup>1,2</sup>, FENGLIAN LIU<sup>2</sup>, ZHIHANG XUE<sup>2</sup>, WENWEI LUO<sup>3</sup>, HAORAN PENG<sup>3</sup>, JUN HE<sup>1</sup>, ZHENGZHENG FU<sup>2</sup>, AND DONGHUI LUO<sup>2</sup>

<sup>1</sup>Sichuan Shuneng Electric Power Technology Company Limited, Chengdu 610095, China

<sup>2</sup>State Grid Sichuan Electric Power Research Institute, Chengdu 610095, China

<sup>3</sup>School of Information and Software Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China


Corresponding author: Haoran Peng (penghaoran23@163.com)

**ABSTRACT** Ensuring the stability of cable tunnels is crucial for power safety and urban reliability in light of the increasing demand for urban electricity. However, frequent waterlogging incidents within tunnels pose a significant threat to city safety by disrupting the power supply. Current methods have proven insufficient in timely and effective water detection within cable tunnels. While visual methods show promise, the complexity of cable tunnel environments degrades image quality under low-light conditions, intensifying the challenge of detection. To address these issues, this study proposes a novel hierarchical approach for waterlogging detection. The approach decomposes the problem into three subproblems: low-light image enhancement, image segmentation, and detection. Firstly, low-light image enhancement techniques are employed to improve image quality and enrich details for subsequent analysis. Next, image segmentation accurately delineates waterlogged road areas while mitigating interference from complex backgrounds. Finally, Faster R-CNN with ResNet as its backbone and integrated attention mechanisms enhances the model's capacity to identify essential features amidst complex backgrounds, significantly improving accuracy in detecting waterlogged areas. Experimental results demonstrate our method's superiority in accuracy and robustness for tunnel waterlogging detection tasks compared to traditional approaches. This proposed approach provides effective technical support for cable tunnel safety monitoring, contributing to ensuring urban power safety and reliability in the face of waterlogging challenges.

**INDEX TERMS** Waterlogging detection, low-light image enhancement, image segmentation, cable tunnels.

## I. INTRODUCTION

Power cables are a critical component of the power distribution system, typically laid in underground tunnels, providing a safe environment for the city's power systems. However, cables in underground tunnels face significant operational risks [1]. Some tunnels are shallow or have high groundwater levels, leading to poor waterproofing capabilities [2], [3]. Prolonged water accumulation can also erode tunnel linings, damage concrete structures, and even corrode cable insulation, affecting the normal operation of power systems. Waterlogging is where excessive amounts

The associate editor coordinating the review of this manuscript and approving it for publication was Sudhakar Radhakrishnan .

of water accumulate in an area. In urban areas and infrastructure such as tunnels, water accumulation can disrupt normal operations, cause structural damage, and pose safety hazards.

At present, there are mainly six methods for tunnel waterlogging detection: (1) using infrared thermal imaging technology to detect tunnel waterlogging [4], (2) using ultrasonic technology to detect tunnel waterlogging [5], (3) using lidar system [6], (4) using capacitive sensors [7], (5) using vision sensors with machine learning and artificial intelligence algorithms [8], and (6) combining multiple sensor networks [9], [10]. Since vision sensors have been widely used in cable tunnels, this paper will study the (5) direction mentioned above. In this paper, the object detection

method in computer vision is used to solve the problem of waterlogging identification in cable tunnels.

Traditional object detection methods [11], [12], [13], [14], [15], [16] in tunnel environments struggle with issues such as low light conditions, high background complexity, and the long-tail distribution of waterlogging data. Firstly, low light conditions and uneven lighting distribution within tunnels seriously deteriorate image quality [17], thereby affecting the accuracy of object detection algorithms. Secondly, the complexity of cable tunnel environments [18] with elements like cable conduits, maintenance equipment, and drainage ditches can obscure targets, increasing false positives or misses. For instance, the reflection from cable lines is frequently misidentified as waterlogging in water detection. Furthermore, the position and distance of surveillance cameras often result in waterlogged areas occupying only a minor portion of the image, which poses a significant challenge in detecting smaller targets. In addition, the variable conditions of tunnel environments, influenced by maintenance personnel activities, introduce further complexity to object detection efforts. Lastly, the training of object detection models faces significant challenges due to sample imbalance [19], characterized by a scarcity of waterlogged instances compared to non-waterlogged ones. This imbalance adversely impacts the model's ability to generalize and accurately identify waterlogged areas. Therefore, in such a specific environment as tunnels, traditional object detection methods need adjustments and optimizations to ensure the accuracy and reliability of detection.

To address the issues in waterlogging detection in cable tunnels, we propose a three-stage processing workflow aimed at enhancing the accuracy and robustness of detection. In the first stage, we enhance the low-light images from within the tunnel to recover rich image information, thereby improving the performance of subsequent tasks. Next, in the second stage, image segmentation is used to accurately delineate the road areas within the cable tunnel, effectively minimizing background interference from other fixed structures within the tunnel. Finally, in the third stage, waterlogging detection is carried out on the segmented road areas. In the waterlogging detection, we incorporate regularization and attention mechanisms into the Faster R-CNN, which effectively mitigates model bias and enhances the recognition of fine-grained features, significantly improving the model's generalizability and robustness across various scenarios.

Our contributions can be summarized as:

- We introduce a three-stage detection method focusing on image segmentation to precisely identify waterlogged areas in cable tunnels, improving accuracy and robustness.
- Our improved Faster R-CNN model integrates attention modules and label smoothing, significantly improving feature recognition and model stability.
- Experiments confirm our method outperforms traditional detection techniques in waterlogging detection, proving its effectiveness and practicality.

## II. RELATED WORK

### A. OBJECT DETECTION

With the advancement of deep learning, modern object detection methodologies increasingly rely on techniques such as proposals [11], [20], anchors [13], [14], and points [15], [16], [21] for accurate box regression and category classification, based on empirical data priors. The introduction of Faster R-CNN [11] marked a significant leap, merging candidate area proposals and object detection into a unified framework via the Region Proposal Network (RPN), thereby substantially improving detection precision and speed. The evolution of single-stage detectors, represented by the YOLO series [13] and SSD [22], facilitated real-time object detection by directly predicting object categories and locations, although they generally lag behind two-stage methods in accuracy.

The success of Transformers in natural language processing, attributed to their attention mechanism, has sparked interest in their application within the computer vision domain. DETR [23] pioneered the application of Transformers in object detection by discarding the conventional anchor and proposal mechanisms and proposing an end-to-end detection strategy. This approach highlighted Transformers' potential in complex scene object detection, inspiring further developments [24], [25], [26] such as Deformable DETR [26] and TSP-R-CNN [27], which have enhanced detection capabilities and efficiency beyond the foundational work of DETR. In this work, we study the problem in the task of waterlogging detection in cable tunnels and present our method.

### B. LOW-LIGHT IMAGE ENHANCEMENT

In recent years, deep learning-based methods for enhancing low-light images have seen notable advancements, primarily through supervised learning approaches to map low-light to normally-exposed images [28], [29], [30]. LLNet, for example, utilizes an autoencoder to enhance visibility in low-light conditions [31]. Studies have also employed multi-scale features to grasp global content and salient structures more effectively [32], [33], [34], while Retinex remains influential in guiding the decomposition of images into reflection and illumination maps [35], [36], [37]. Additionally, unfolding-based methods have focused on refining low-light enhancement priors [38], [39]. Beyond enhancements in sRGB space, the exploration of image space has led to a focus on realistic noise modeling to bridge the domain gap in synthetic training data, with the Poisson-Gaussian model playing a foundational role [40], [41], [42].

### C. IMAGE SEGMENTATION

Image segmentation is a core task in computer vision that aims to divide images into regions representing different instances or categories. This task traditionally splits into three subfields: semantic segmentation, instance segmentation, and panoptic segmentation. In semantic segmentation, early

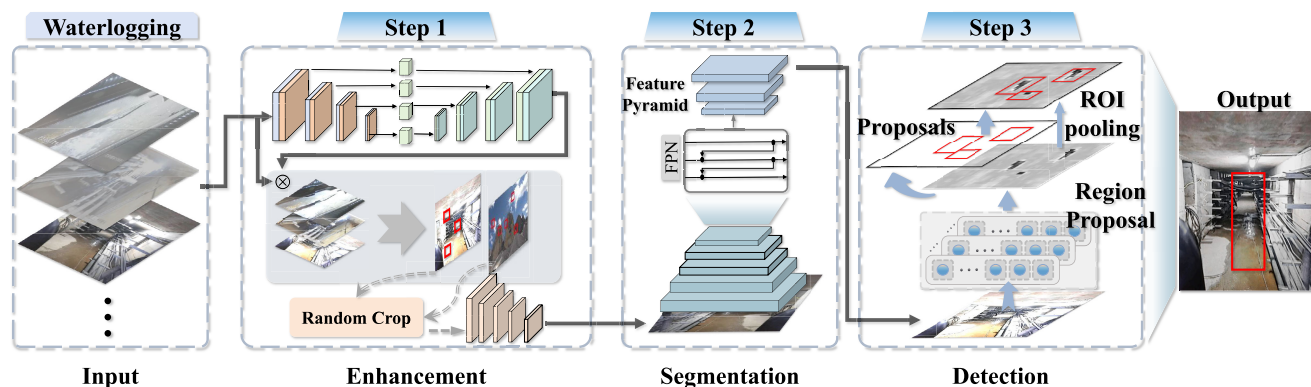


FIGURE 1. Overview of our framework.

methods used CNNs for pixel-level classification [43], [44], [45]. The success of transformers in language and visual tasks [23], [46] led to their adoption in semantic segmentation with great results [47], [48], [49]. For example, MaskFormer [50] approaches semantic segmentation as a mask classification problem. Instance segmentation has traditionally been seen as a mask classification problem too, where each instance gets a unique mask and category [51], [52], [53]. Panoptic segmentation [54] aims to unify semantic and instance segmentation. Early models like Panoptic-FPN [55] used separate branches for different tasks. Later research improved performance using transformer architectures [56], [57], [58]. Despite these advances, panoptic segmentation models have not fully reached their unification potential.

#### D. TUNNEL WATERLOGGING DETECTION

Traditional image processing techniques were among the first to be used for vision-based waterlogging detection. For example, an edge detection-based method uses the Canny edge detection algorithm to extract edge information and morphological operations to isolate waterlogged areas [59]. As the field progressed, machine learning methods began to be applied to tunnel waterlogging detection due to their ability to learn from data and make predictions. Reference [60] uses mask-RCNN to improve the accuracy of water identification. References [61], [62], [63], and [64] further improve the detection result. However, most of the previous studies on tunnel water detection focused on conventional tunnels such as subway tunnels. They did not study cable tunnels, let alone put forward solutions to the complex background of cable tunnels. This is the problem that this paper wants to solve.

### III. METHODS

Waterlogging detection in cable tunnel environments poses significant challenges due to low light conditions, reflective surfaces in the surroundings, and dynamic changes within

the scene. These factors complicate the task of distinguishing between waterlogged and dry areas, particularly when faced with varying water depths and the potential for the tunnel's complex infrastructure to visually merge with water reflections. To address these challenges, this study introduces a three-stage processing pipeline that begins with low-light image enhancement and is followed by waterlogged road segmentation before the detection stage. This strategy improves detection accuracy and robustness by dividing the complex problem into several sub-tasks and addressing the specific difficulties of each sub-task. Figure 1 shows the overview of our framework.

#### A. LOW-LIGHT IMAGE ENHANCEMENT

In environments like cable tunnels, image detection faces challenges like limited dynamic range and increased noise from high ISO settings, which impair algorithm performance. Additionally, obtaining paired images for training under variable lighting is often unfeasible. To address these issues, we introduce EnlightenGAN, an unsupervised method using generative adversarial networks for low-light enhancement without needing paired images. In this way, we can improve image quality and detail naturalness, which effectively helps us in the following segmentation and detection.

EnlightenGAN combines an attention-guided U-Net with a bi-discriminator mechanism for texture and structure enhancement, effectively bridging the gap between low and normal lighting conditions. To tackle spatial illumination variations, EnlightenGAN employs a novel global-local discriminator framework, which enhances adaptability by evaluating both overall and localized image patches. This ensures a balanced enhancement across images, avoiding local overexposure or underexposure.

In addition, for the global discriminator, the relativistic discriminator structure is utilized to estimate the probability that the real data is more realistic than the fake data and to guide the generator to synthesize a fake image that is more realistic than the real data. The function of the relativistic

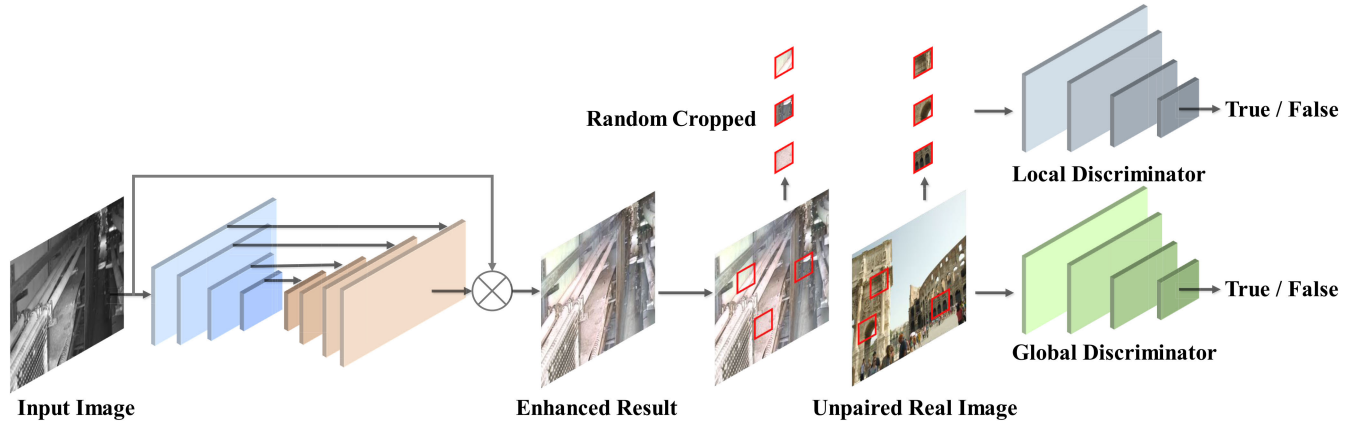


FIGURE 2. The overall architecture of EnlightenGAN.

discriminator is:

$$D_R(x_r, x_f) = \sigma(C(x_r) - E_{x_f \sim P_{\text{fake}}}[C(x_f)]), \quad (1)$$

$$D_R(x_f, x_r) = \sigma(C(x_f) - E_{x_r \sim P_{\text{real}}}[C(x_r)]), \quad (2)$$

where  $C$  is the discriminator network,  $x_r$  and  $x_f$  are sampled from the real and false distributions, respectively, and  $\sigma$  denotes the sigmoid function. It made a slight modification to the relativistic discriminator by replacing the sigmoid function with a least squares GAN (LSGAN) loss. Finally, the loss functions of the global discriminator  $D$  and generator  $G$  are:

$$\begin{aligned} \mathcal{L}_D^{\text{Global}} &= E_{x_r \sim P_{\text{real}}} \left[ (D_R(x_r, x_f) - 1)^2 \right] \\ &+ E_{x_f \sim P_{\text{fake}}} \left[ (D_R(x_f, x_r))^2 \right], \end{aligned} \quad (3)$$

$$\begin{aligned} \mathcal{L}_G^{\text{Global}} &= E_{x_f \sim P_{\text{fake}}} \left[ (D_R(x_f, x_r) - 1)^2 \right] \\ &+ E_{x_r \sim P_{\text{real}}} \left[ (D_R(x_r, x_f))^2 \right]. \end{aligned} \quad (4)$$

For the local discriminator, the original LSGAN loss is used as the adversarial loss by learning local patches randomly cropped from the output and the real image. This design enables the model to better simulate image features under normal illumination at both the global and local levels. The LSGAN as adversarial loss is as follows:

$$\begin{aligned} \mathcal{L}_D^{\text{Local}} &= E_{x_r \sim P_{\text{real-patches}}} \left[ (D(x_r) - 1)^2 \right] \\ &+ E_{x_f \sim P_{\text{fake-patches}}} \left[ (D(x_f) - 0)^2 \right], \end{aligned} \quad (5)$$

$$\mathcal{L}_G^{\text{Local}} = E_{x_r \sim P_{\text{fake-patches}}} \left[ (D(x_f) - 1)^2 \right]. \quad (6)$$

## B. WATERLOGGED ROAD SEGMENTATION

Following the initial low-light enhancement stage, we introduced a YOLOv8-based image segmentation module dedicated to processing enhanced images for accurately segmenting potential waterlogged areas on the road surface. Image segmentation for waterlogging effectively isolates complex backgrounds in the image, thereby reducing interference from non-target areas on the accuracy of

waterlogging detection. Thanks to the enhancement of EnlightenGAN under low-light conditions, the overall quality of the images has been significantly improved, including brightness, contrast, and clarity of details, which considerably simplifies the difficulty of image segmentation and enhances the precision of segmentation.

YOLOv8 introduces an anchor-free detection mechanism and an improved feature pyramid network, innovations that provide higher efficiency and accuracy when handling complex scenes like those found in cable tunnels. Additionally, YOLOv8 utilizes the CSPDarknet-53 as its backbone, optimizing gradient flow with the Cross Stage Partial (CSP) structure, and further improving model segmentation and detection performance with the Fast Spatial Pyramid Pooling (SPPF) structure and a decoupled head strategy. These technical features make YOLOv8 consistently demonstrate strong performance in segmentation tasks. Therefore, it is an ideal choice for us to segment images of road areas in cable tunnels. In this paper, the pre-training model of YOLOv8n-seg is used for training and reasoning.

## C. WATERLOGGING DETECTION

After segmenting the cable road to remove distractions like pipelines and lighting, we utilize our improved Faster R-CNN for waterlogging detection in cable tunnels. Faster R-CNN, being a two-stage detection method, offers superior accuracy over one-stage networks, effectively addressing multi-scale and small object detection challenges. Its incorporation of a Region Proposal Network (RPN) ensures precise detection performance, making it ideal for accurately identifying water accumulation. Enhanced with our proposed global residual attention module (GRAM) within its ResNet50 backbone, our improved Faster R-CNN adaptation improves feature extraction and recognition capabilities. The framework of our improved Faster R-CNN is shown in Figure 3.

The original Faster R-CNN combines deep feature extraction, region proposal generation, and consistent proposal sizing via the Region of Interest (ROI) Pooling, followed

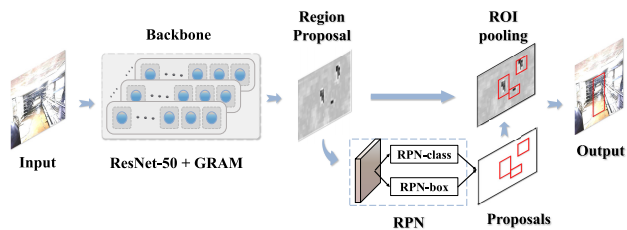


FIGURE 3. The overall architecture of our waterlogging detection.

by classification and bounding box regression in a unified framework. This process begins with deep feature extraction from the input image, followed by the generation of region proposals by the RPN. These proposals are then standardized through ROI Pooling, allowing for accurate object localization and classification in the final detection stages.

### 1) RESNET BACKBONE

In the traditional Faster R-CNN, VGG16 [65] is often used as the feature extraction network. However, with the increase in network depth, VGG16 faces issues such as increased training difficulty, vanishing gradients, and decreased gradient correlation. In contrast, the deep residual network ResNet50, with its strong feature representation capability, has become a more optimal choice. ResNet50 introduces residual connections, which not only reduce the model’s parameter count and computational burden but also effectively prevent information loss and gradient issues during the training of deep networks. These characteristics make ResNet50 highly suitable for object detection tasks that require high real-time performance and accuracy. In this paper, ResNet50 uses FrozenBatchNorm2d to save the training scale.

### 2) CHANNEL AND SPATIAL ATTENTION MECHANISM

To further enhance the performance of Faster R-CNN in complex scenarios such as waterlogging detection at the bottom of cable tunnels, we propose a global residual attention module (GRAM) based on the Convolutional Block Attention Module (CBAM) [66] and integrates GRAM on top of ResNet50. CBAM significantly improves the ability of network on important image features by introducing Channel Attention Module (CAM) and Spatial Attention Module (SAM). The channel attention mechanism assigns different importance weights to each channel of the feature map, while the spatial attention mechanism highlights the features at the target location in the image. This integration enhances the capacity of model for precise identification and localization of targets within images. Building upon CBAM, GRAM introduces redesigned channel attention and spatial attention submodules. It achieves improved attention effectiveness by reorganizing features to preserve the image’s dimensional information and leveraging convolutional layers for spatial information fusion. Notably, GRAM assigns

adequate weight to small object features within the image, enhancing its attention capabilities.

The channel attention submodule first applies batch normalization to emphasize prominent features, followed by a rearrangement operation to maintain the integrity of the three-dimensional information. Then, a two-layer multilayer perceptron is used to enhance the dependency between the channel and spatial dimensions, as shown in Figure 4. The process can be described by the following formula:

$$F_1 = W_\gamma (BN(F_1)), \tag{7}$$

$$F_2 = MLP(p(F_1)), \tag{8}$$

$$F_3 = \text{sigmoid}(\bar{p}(F_2)), \tag{9}$$

where  $W_\gamma$  represents the scaling factors for each channel,  $F_1$  and  $F_3$  represent the input and output feature maps, respectively,  $p$  and  $\bar{p}$  represent the three-dimensional rearrangement operations performed on the feature maps. Specifically, if the dimension of the input feature map is  $C \times W \times H$ , then the  $p$  operation changes the dimension of the feature map to  $W \times H \times C$ , while the  $\bar{p}$  operation restores it to the original  $C \times W \times H$  dimension.

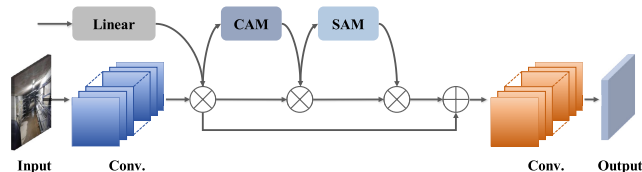


FIGURE 4. Convolutional block attention module diagram.

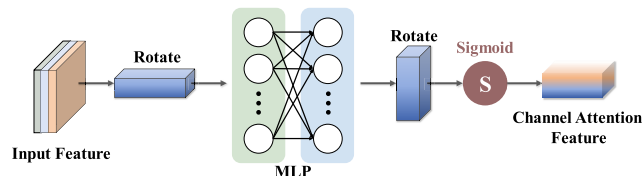


FIGURE 5. Channel attention module (CAM) diagram.

Within the CAM, as shown in Figure 5, the MLP is characterized by a two-layer architecture that employs the ReLU function for its activation mechanism. After processing by the first layer of the perceptron, the number of channels in the feature map is reduced to  $C/R$ , where  $R$  is the channel compression ratio. After processing by the second layer, the number of channels increases according to the channel expansion ratio, and in this module, the channel compression ratio and channel expansion ratio are kept consistent, so the final number of channels in the feature map remains unchanged. Moreover, BN represents batch normalization, a preprocessing step for the input feature map aimed at improving the stability and efficiency of model training. The specific formula is as follows:

$$BN(F) = \gamma \frac{F - \mu}{\sqrt{\sigma^2 + \epsilon}} + \beta, \tag{10}$$

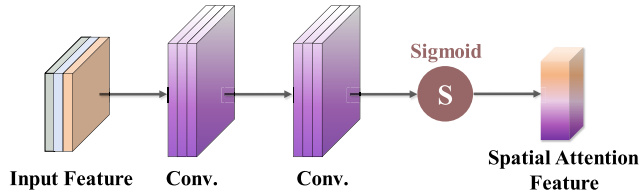


FIGURE 6. Spatial attention module(SAM).

where  $\mu$  and  $\sigma$  represent the mean and standard deviation of the batch data, respectively,  $\gamma$  and  $\beta$  are trainable scaling and shifting parameters.

Compared to the original CBAM, to preserve rich feature information, we eliminated the pooling operation in the spatial attention module and adopted pixel-level normalization, as depicted in Figure 6. Subsequently, spatial information is deeply integrated through two consecutive convolutional layers to enhance the representation capability of the feature map. The specific expression of this process is as follows:

$$F_1 = W_\lambda (BN_s (F_1)), \quad (11)$$

$$F_2 = \text{sigmoid} \left( \text{conv}^{5 \times 5} \left( \text{conv}^{5 \times 5} (F_1) \right) \right), \quad (12)$$

where  $F_1$  and  $F_2$  represent the input and processed output feature maps, respectively.  $BN_s$  represents the pixel-level normalization operation, and  $W_\lambda$  is the corresponding scaling factor, calculated similarly to the channel attention module. The difference is that the length of the scaling factor is determined by the total number of pixels in the feature map, not the number of channels. The feature map is first processed through the first convolutional layer, where the number of channels is reduced to  $C/r$ , with  $r$  being the same channel compression ratio as in the channel attention module. After processing through the second convolutional layer, the number of channels is restored to the original  $C$ , completing the integration and enhancement of spatial information in the feature map.

### 3) LABEL SMOOTHING

In order to refine the accuracy and enhance the stability of RPN head in the Faster R-CNN framework, this research introduces label smoothing as a crucial part of the class logits computation. It is integrated within the binary cross-entropy loss function, aiming to alleviate the model's overfitting and furnish regularization, thereby fostering improved generalization and bolstering the model's defenses against noisy data.

The implementation adjusts the target labels by a smoothing parameter  $\epsilon$ , effectively softening the binary targets. Instead of using hard 0 and 1 values, the targets are transformed towards a more moderate value that reflects the degree of uncertainty. Specifically, the adjusted target value  $T'$  is computed as:

$$T' = T \cdot (1 - \epsilon) + 0.5 \cdot \epsilon, \quad (13)$$

where  $T$  is original target value.  $T'$  effectively merging the target with a midpoint value, which is conventionally set at 0.5. Such a formulation of label smoothing diminishes the likelihood of the model's overreliance on absolute label values, thereby mitigating risks of overfitting. By incorporating this refinement into the class logits calculation of the RPN Head, the Faster R-CNN is endowed with a heightened capacity for discrimination, while simultaneously accommodating the labeling ambiguities frequently encountered in real-world scenarios.

Our method enhances image brightness in low-light conditions, adding semantic richness and employing a segmentation strategy for tunnel road surfaces to avoid extraneous interference. We enhance the Faster R-CNN by integrating ResNet and GRAM for deeper and more focused feature learning, along with label smoothing to refine the detection accuracy. This approach improves feature learning, stabilizes training against noisy data, and ensures precise detection in challenging cable tunnel environments, as illustrated in Figure 7.

## IV. EXPERIMENTS

We evaluated the performance of our proposed three-stage road segmentation flood detection method in a real cable tunnel. Comparative experimental results indicate that our method outperforms existing object detection methods in cable tunnel scenarios. Ablation experimental results demonstrate that each module in our method contributes to the improvement of flood detection tasks.

### A. EXPERIMENTAL SETUP

The experimental environment used in this study consisted of Ubuntu 20.04.6 LTS operating system with NVIDIA TITAN RTX GPU and 20GB video memory. The software environment for algorithm experiments was built on the PyTorch 1.13.0 framework using the Python programming language. The specific Python version employed was 3.8.18. For accelerated computations, CUDA 11 and the corresponding version of CUDNN were utilized.

#### 1) DATASET

To validate the feasibility and effectiveness of the proposed algorithm for cable tunnel waterlogging identification in our study, we created our dataset of cable tunnel waterlogging. The data was obtained from monitoring images of real-world cable tunnels in Chengdu, Sichuan Province, China. To ensure the applicability of the algorithm to different scenarios in cable tunnels, we selected images from different monitoring points in various cable tunnels, capturing different flood conditions and angles. In total, we collected hundreds of images of cable tunnels. We annotate the cable tunnel images of the dataset. The tunnel road annotation was performed for training and testing of the segmentation algorithm, and the waterlogging annotation was performed for training and testing of the detection algorithm. We divided the original set

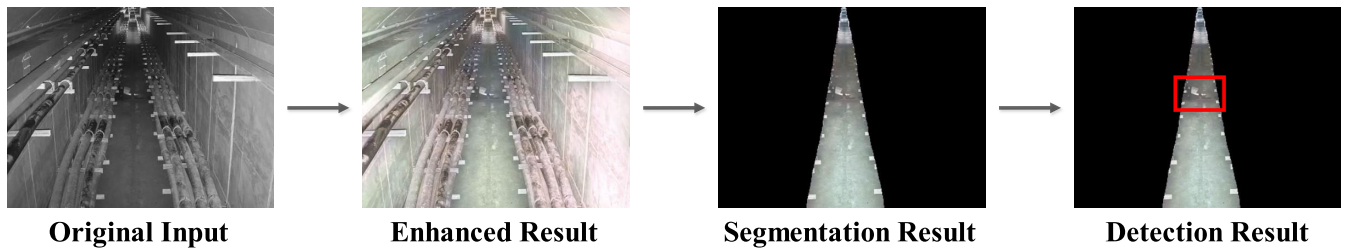


FIGURE 7. The pipeline of three-stage waterlogging detection process.

into a training set and a validation set using a split ratio of 0.85:0.15.

## 2) PERFORMANCE METRICS

In object detection and image segmentation tasks, IOU (Intersection over Union) and Average Precision (AP) are key metrics for evaluating performance. IOU measures the overlap between the predicted bounding box or segmentation result and the true bounding box or segmentation mask. Average Precision (AP) assesses the model's performance by calculating the average precision value at different recall levels, particularly at an IOU threshold of 0.50 and ranging from 0.50 to 0.95. Precision (P) also focuses on the model's accuracy in predicting the positive class, while Recall (R) represents the ratio of correctly detected objects to the total number of true objects. The F1 Score, as the harmonic mean of Precision and Recall, balances the two to evaluate a model's accuracy and coverage for imbalanced datasets comprehensively.

## B. COMPARATIVE EXPERIMENTS

To evaluate the detection performance of our proposed method for cable tunnel bottom, we compared it with the latest object detection methods, including YOLOv5, YOLOv8, and Faster R-CNN. The YOLOv5 model used in our comparison is the YOLOv5s-det model, with default parameters for YOLOv5s. Similarly, the YOLOv8 model used is the YOLOv8n-det model, with default parameters for YOLOv8n. The Faster R-CNN model is an improved Faster R-CNN model, utilizing ResNet50+FPN as the backbone.

For the YOLOv5s model, we trained it for 100 epochs, divided the entire dataset into 16 batches, set the initial learning rate (lr0) to 0.01, and the cyclic learning rate (lrf) to 0.01. The network settings for YOLOv8n are approximately the same as YOLOv5s.

As for the baseline Faster R-CNN, we trained it for 30 epochs with a learning rate of 0.02, batch size of 2, SGD momentum parameter of 0.9, and weight decay parameter of  $1e-4$ .

We conducted comparative experiments to evaluate the performance of our proposed method against the baseline methods, including YOLOv5, YOLOv8, and Faster R-CNN. The results of the comparative experiments are summarized in Table 1. In detail, we find that our method achieves the best results among all state-of-time methods on all metrics with a

TABLE 1. Model performance comparison.

Model	AP (50)↑	AP (50:95)↑	R↑	F1↑
Faster R-CNN	0.616	0.222	0.400	0.485
YOLOv5s	0.631	0.210	0.455	0.529
YOLOv8n	0.712	0.234	<b>0.489</b>	0.580
Ours	<b>0.871</b>	<b>0.296</b>	0.460	<b>0.602</b>

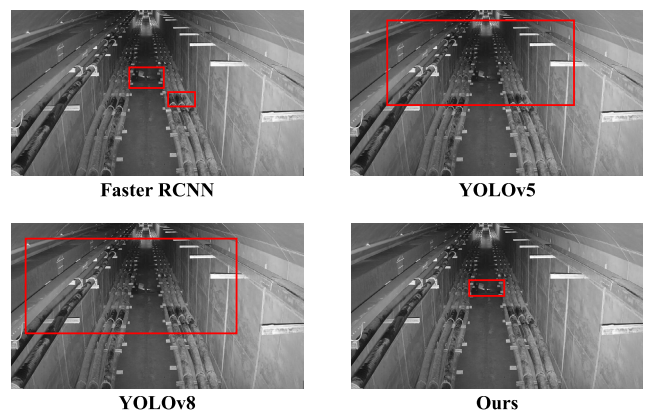
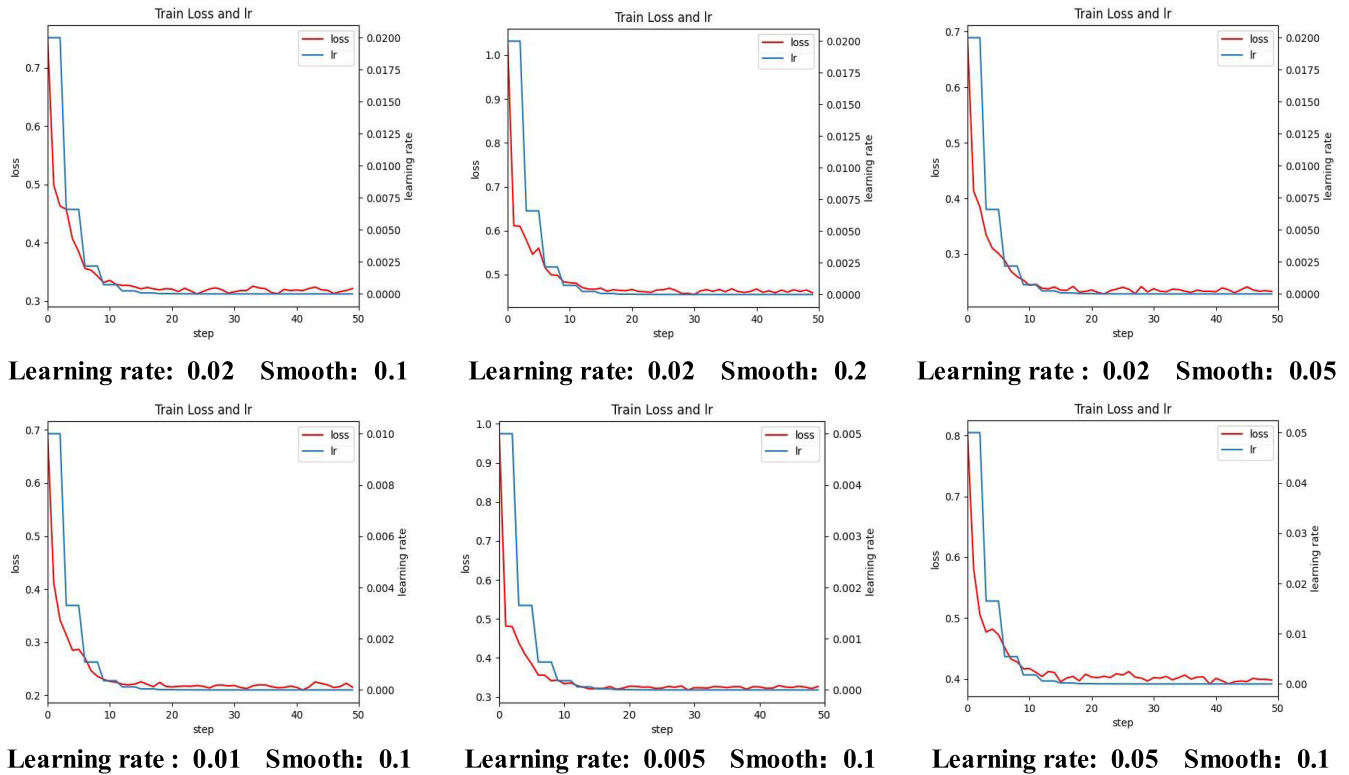


FIGURE 8. The detection result images for different models. It is noticeable that the detection results of the original Faster R-CNN method are influenced by various factors in the cable tunnel. On the other hand, the YOLO series methods tend to produce detection boxes that are larger than the actual water accumulation. In contrast, our method addresses these issues and offers improved detection performance.

large margin. In Figure 8, we present illustrative examples of detection results for various models.

Our approach uses image enhancement and an attentional backbone, which gives it a significant lead over other models in AP. Besides, the smoothness of the loss function enables a better improvement in the recall rate compared with the original Faster rcnn and YOLOv5. Although YOLOv8 slightly exceeds us in terms of recall rate, it can be seen from Figure 7 that YOLOv8's detection box is so large that it covers many non-water-standing areas. Such detection box makes its recall rate relatively high, and the detection box of our method happens to be the water-standing area. However, we are both ahead of other models in f1, proving the superiority of our model.

To investigate the impact of different hyperparameters on our algorithm, we conducted a series of combination experiments on the two most crucial hyperparameters: learning rate and smoothing parameter. The results of the experiments are illustrated in the Table 2 as shown.



**FIGURE 9.** Comparative analysis of training loss and learning rate across different hyperparameter settings. The varying combinations of learning rate and smoothing factor are investigated to discern their effect on the convergence behavior of the model during training. Each subplot corresponds to a specific pairing of learning rate and smooth factor, with the training loss and learning rate plotted over training iterations.

**TABLE 2.** Impact of hyperparameter tuning on model performance.

Model	AP (50)↑	AP (50:95)↑	R↑
Lr: 0.005, Smooth: 0.1	0.642	0.212	0.34
Lr: 0.05, Smooth: 0.1	0.640	0.251	0.40
Lr: 0.01, Smooth: 0.05	0.840	0.207	0.34
Lr: 0.02, Smooth: 0.05	0.817	0.275	0.42
Lr: 0.02, Smooth: 0.2	0.753	0.275	0.42
Lr: 0.02, Smooth: 0.1	<b>0.871</b>	<b>0.296</b>	<b>0.46</b>

From Figure 9, it is evident that the first combination of hyperparameters—learning rate at 0.02 and smoothing factor at 0.1—yields the best results. This setup shows the highest precision in detecting the objects at the basic intersection over a union (IoU) threshold of 0.5, as indicated by the highest Average Precision (AP) at 50% (AP50). Moreover, the superior performance is consistent across various IoU thresholds, from 0.5 to 0.95 (AP50:95), suggesting that this parameter tuning not only achieves the best balance in learning new features but also retains learned features effectively, leading to a more accurate and robust model.

When the learning rate is set to 0.02, the model exhibits the best performance in terms of AP(50), scoring 0.871. This suggests that a learning rate of 0.02 is a suitable choice, as it enables the model to achieve the highest precision in object detection at the basic Intersection over the Union (IoU) threshold of 0.5. Conversely, with a learning rate of 0.001,

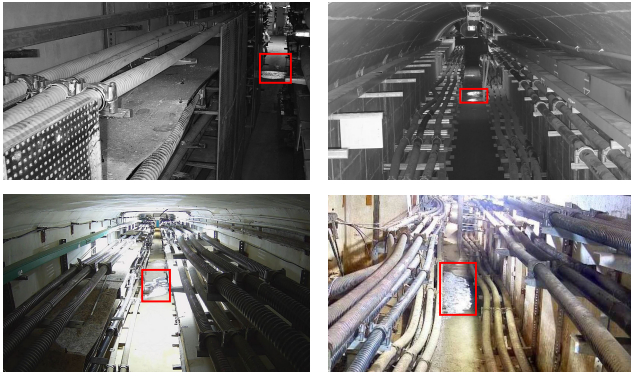
there is a noticeable decline in AP(50) to 0.642, which may indicate that the rate is too low for the model to learn adequately within the stipulated number of training epochs.

Models with a smoothing coefficient of 0.1 demonstrate the best results for AP(50), implying that a moderate smoothing coefficient helps balance the acquisition of new features with the retention of previously learned features. When the smoothing coefficient is increased to 0.2, there is a reduction in performance, likely due to excessive smoothing impeding the model’s ability to capture new features.

The AP(50:95) metric reflects the model’s average performance across a range of IoU thresholds from 0.5 to 0.95. Although some parameter settings occasionally achieve similar performance to the best settings (e.g., Lr: 0.02, Smooth: 0.05 with an AP(50:95) of 0.275), the initial parameter grouping with an AP(50) of 0.871 and AP(50:95) of 0.296 consistently outperforms the others. This indicates a higher accuracy in bounding box localization, suggesting that this parameter configuration yields a more precise detection model across various IoU thresholds.

Our method was employed to address diverse challenging conditions present in cable tunnels water accumulation scenarios, such as low illumination, lighting interferences, watermark disturbances, and complex background noise. The experimental results are presented in Figure 10. Extensive evaluations across a range of demanding environments

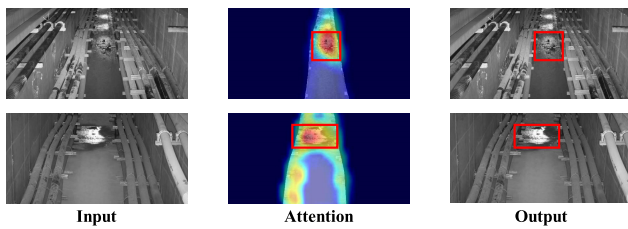




**FIGURE 10.** Performance evaluation of waterlogging detection in cable tunnels under challenging conditions. This figure demonstrates that our approach effectively overcomes these challenges.

indicate that our method not only robustly detects water hazards under suboptimal conditions but also maintains high accuracy and precision, illustrating its reliability and potential for practical applications in safety-critical infrastructure monitoring.

To further validate the effectiveness of the proposed method, this study employs EigenCAM technology for visual analysis of the model’s attention mechanism. As illustrated in Figure 11, the proposed model effectively focuses its attention on areas of water accumulation on the road surface, achieving precise detection.



**FIGURE 11.** Representative examples demonstrate our method’s focus on water accumulation through attention maps, significantly boosting detection accuracy and effectiveness.

**C. ABLATION STUDY**

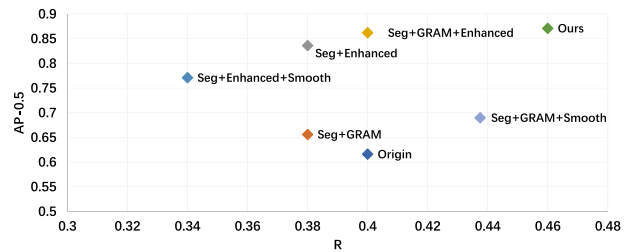
To verify the contributions of each component of our method, we design a series of ablation experiments, as presented in Table 3. Our results confirm that each module of

**TABLE 3.** Ablation study of the proposed module.

Model	AP (50)↑	AP (50:95)↑	R↑
Original (Faster R-CNN)	0.616	0.222	0.400
Seg+GRAM	0.656	0.284	0.380
Seg+Enhanced	0.836	0.259	0.340
Seg+GRAM+Enhanced	0.862	0.266	0.400
Seg+GRAM+Smooth	0.690	0.218	0.438
Seg+Enhanced+Smooth	0.771	0.247	0.340
Ours(Seg+GRAM+Enhance+Smooth)	<b>0.871</b>	<b>0.296</b>	<b>0.460</b>

our approach significantly improves waterlogging detection, highlighting the critical roles of low-light enhancement and data balance. The Enhance module notably boosts detection precision, while the GRAM and Smooth techniques improve recall. Low-light enhancement alone raises performance by approximately 20%, with GRAM increasing recall by 3-4% and further reducing false positives.

In Figure 12, we use a scatter plot to more clearly represent the contribution of each module in our method to the result. From this diagram, we can see that the enhanced module significantly improves precision, while GRAM and smooth contribute to the recall rate.



**FIGURE 12.** The scatter plot of the ablation study results where the horizontal axis is the recall and the vertical axis is the precision with an IOU of 50.

**V. DISCUSSION**

In this study, we introduce a novel waterlogging detection method tailored to the complex environment of cable tunnels. Our approach, which combines low-light image enhancement and a unique strategy for segmenting waterlogged surfaces, demonstrates significant improvements in accuracy and robustness against conventional methods. The use of GRAM further refines our model’s precision and recall, proving particularly effective in identifying subtle water accumulations.

Despite these advancements, challenges remain. Our method’s recall rate is below the desired threshold, which could hinder the practical application in cable tunnels where accurate detection of waterlogged areas is crucial. Additionally, the comprehensive three-stage approach, while effective, introduces time-consuming processes that could be optimized for quicker detection. Future work will aim to enhance the recall rate and streamline the detection process, addressing these limitations to better meet the operational demands within cable tunnels.

In recent studies of waterlogging detection, it is often the multi-sensor approach that combines advanced artificial intelligence and machine learning techniques that is most effective. However, this paper adopts a pure vision approach and tries to solve the problem only through visual sensors. Indeed, the combination of other multi-sensors can effectively solve the problem of waterlogging detection. The combination of multiple sensors can solve the recall problem in this paper. For example, ultrasonic sensors can find more areas of stagnant water than visual sensors can see.

However, the problems faced by multi-sensors are high cost and inconvenient maintenance. Vision cameras are already widely deployed in cable tunnels, while other sensor deployments still cost a lot to deploy. Therefore, it is necessary to study the pure vision method for waterlogging accumulation in cable tunnels.

## VI. CONCLUSION

In this study, we propose a novel three-stage approach for detecting waterlogging in cable tunnels. To overcome the challenge of feature blurring caused by low illumination, we introduce a data enhancement module that effectively improves the quality of images captured within the tunnel environment. By focusing on image segmentation specifically on the road surface of the cable tunnel, we successfully mitigate the impact of complex internal elements, such as cables, thereby reducing interference and improving the accuracy of waterlogging detection. Furthermore, we enhance the Faster R-CNN model by integrating the GRAM attention mechanism and incorporating label smoothing techniques. This enhancement significantly boosts the precision and recall of the detection system. Through extensive experiments conducted on a dedicated waterlogging dataset in cable tunnels, we demonstrate the superior performance of our proposed method compared to existing techniques.

Looking ahead, our future research will focus on further improving the recall rate while concurrently reducing the processing time of the system. This will enable more effective and efficient waterlogging detection in cable tunnels, contributing to enhanced safety and reliability in urban power systems.

## REFERENCES

- [1] F. Zhuang, C. Zupan, Z. Chao, and Z. Yanzheng, "A cable-tunnel inspecting robot for dangerous environment," *Int. J. Adv. Robotic Syst.*, vol. 5, no. 3, p. 32, Sep. 2008.
- [2] T. Xu, L. Xu, X. Li, and J. Yao, "Detection of water leakage in underground tunnels using corrected intensity data and 3D point cloud of terrestrial laser scanning," *IEEE Access*, vol. 6, pp. 32471–32480, 2018.
- [3] S. Qin, Z. Ma, J. Lin, Y. Xue, C. Jiang, X. Shang, and Z. Li, "New method for detecting risk of tunnel water-induced disasters using magnetic resonance sounding," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 6, pp. 843–847, Jun. 2018.
- [4] Z. Lu, F. Zhu, L. Shi, F. Wang, P. Zeng, J. Hu, X. Liu, Y. Xu, and Q. Chen, "Automatic seepage detection in cable tunnels using infrared thermography," *Meas. Sci. Technol.*, vol. 30, no. 11, Nov. 2019, Art. no. 115902.
- [5] Y. Yu, A. Safari, X. Niu, B. Drinkwater, and K. V. Horoshenkov, "Acoustic and ultrasonic techniques for defect detection and condition monitoring in water and sewerage pipes: A review," *Appl. Acoust.*, vol. 183, Dec. 2021, Art. no. 108282.
- [6] Q. Chen, Z. Kang, Z. Cao, X. Xie, B. Guan, Y. Pan, and J. Chang, "Combining cylindrical voxel and mask R-CNN for automatic detection of water leakages in shield tunnel point clouds," *Remote Sens.*, vol. 16, no. 5, p. 896, Mar. 2024.
- [7] A. Pai, C. H. Harshvardhan, B. S. Vaibhav, and T. R. Mahesh, "Implementation of the capacitive transducer for flood detection," in *Proc. 3rd Int. Conf. Innov. Sustain. Comput. Technol. (CISCT)*, Sep. 2023, pp. 1–5.
- [8] Z. Zhou, J. Zhang, and C. Gong, "Automatic detection method of tunnel lining multi-defects via an enhanced you only look once network," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 37, no. 6, pp. 762–780, May 2022.
- [9] L. Han, J. Chen, H. Li, G. Liu, B. Leng, A. Ahmed, and Z. Zhang, "Multispectral water leakage detection based on a one-stage anchor-free modality fusion network for metro tunnels," *Autom. Construct.*, vol. 140, Aug. 2022, Art. no. 104345.
- [10] P. Yu, H. Wu, C. Liu, and Z. Xu, "Water leakage diagnosis in metro tunnels by integration of laser point cloud and infrared thermal imaging," *Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. 3, pp. 2167–2171, Apr. 2018.
- [11] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 28, 2015, pp. 1–20.
- [12] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2999–3007.
- [13] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 779–788.
- [14] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6517–6525.
- [15] Z. Tian, X. Chu, X. Wang, X. Wei, and C. Shen, "Fully convolutional one-stage 3D object detection on LiDAR range images," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 35, 2022, pp. 34899–34911.
- [16] X. Zhou, D. Wang, and P. Krähenbuhl, "Objects as points," 2019, *arXiv:1904.07850*.
- [17] X. Cui, L. Ma, T. Ma, J. Yuan, X. Fan, and R. Liu, "Trash to treasure: Low-light object detection via decomposition-and-aggregation," in *Proc. AAAI Conference Artificial Intelligence*, 2024, vol. 38, no. 2, pp. 1417–1425.
- [18] T. Lin, X. Lin, F. Teng, and L. Wan, "Uncertainty analysis of UMRS parameters and its application for water detection in the tunnel," *J. Appl. Geophys.*, vol. 179, Aug. 2020, Art. no. 104116.
- [19] J. Zhao, M. Shi, G. Hu, X. Song, C. Zhang, D. Tao, and W. Wu, "A data-driven framework for tunnel geological-type prediction based on TBM operating data," *IEEE Access*, vol. 7, pp. 66703–66713, 2019.
- [20] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1440–1448.
- [21] X. Wang, K. Chen, Z. Huang, C. Yao, and W. Liu, "Point linking network for object detection," 2017, *arXiv:1706.03646*.
- [22] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single shot multibox detector," in *Proc. 14th Eur. Conf. Amsterdam, The Netherlands: Springer*, Oct. 2016, pp. 21–37.
- [23] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, "End-to-end object detection with transformers," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2020, pp. 213–229.
- [24] D. Jia, Y. Yuan, H. He, X. Wu, H. Yu, W. Lin, L. Sun, C. Zhang, and H. Hu, "DETRs with hybrid matching," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 19702–19712.
- [25] H. Zhang, F. Li, S. Liu, L. Zhang, H. Su, J. Zhu, L. M. Ni, and H.-Y. Shum, "DINO: DETR with improved DeNoising anchor boxes for end-to-end object detection," 2022, *arXiv:2203.03605*.
- [26] X. Zhu, W. Su, L. Lu, B. Li, X. Wang, and J. Dai, "Deformable DETR: Deformable transformers for end-to-end object detection," 2020, *arXiv:2010.04159*.
- [27] Z. Sun, S. Cao, Y. Yang, and K. Kitani, "Rethinking transformer-based set prediction for object detection," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 3591–3600.
- [28] L. Guo, R. Wan, G.-M. Su, A. C. Kot, and B. Wen, "Multi-scale feature guided low-light image enhancement," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2021, pp. 554–558.
- [29] Y. Jin, B. Lin, W. Yan, Y. Yuan, W. Ye, and R. T. Tan, "Enhancing visibility in nighttime haze images using guided APSF and gradient adaptive convolution," in *Proc. 31st ACM Int. Conf. Multimedia*, Oct. 2023, pp. 2446–2457.
- [30] Y. Jin, W. Yang, and R. T. Tan, "Unsupervised night image enhancement: When layer decomposition meets light-effects suppression," in *Proc. Eur. Conf. Comput. Vis.*, 2022, pp. 404–421.
- [31] K. G. Lore, A. Akintayo, and S. Sarkar, "LLNet: A deep autoencoder approach to natural low-light image enhancement," *Pattern Recognit.*, vol. 61, pp. 650–662, Jan. 2017.
- [32] L. Tao, C. Zhu, G. Xiang, L. Li, Y. Jia, and X. Xie, "LLCNN: A convolutional neural network for low-light image enhancement," *IEEE Vis. Commun. Image Process. (VCIP)*, vol. 1, no. 1, pp. 1–4, Jun. 2017.

- [33] F. Lv, F. Lu, J. Wu, and C. Lim, "MBLLEN: Low-light image/video enhancement using CNNs," *BMVC*, vol. 220, no. 1, p. 4, 2018.
- [34] W. Ren, S. Liu, L. Ma, X. Xu, X. Xu, X. Cao, and M. Yang, "Low-light image enhancement via a deep hybrid network," *IEEE Trans. Image Process.*, vol. 28, no. 9, pp. 4364–4375, Jul. 2019.
- [35] C. Wei, W. Wang, W. Yang, and J. Liu, "Deep retinex decomposition for low-light enhancement," 2018, *arXiv:1808.04560*.
- [36] Y. Wang, Y. Cao, Z.-J. Zha, J. Zhang, Z. Xiong, W. Zhang, and F. Wu, "Progressive retinex: Mutually reinforced illumination-noise perception network for low-light image enhancement," in *Proc. 27th ACM Int. Conf. Multimedia*, Oct. 2019, pp. 2015–2023.
- [37] L. Guo, R. Wan, W. Yang, A. Kot, and B. Wen, "Enhancing low-light images in real world via cross-image disentanglement," 2022, *arXiv:2201.03145*.
- [38] C. Zheng, D. Shi, and W. Shi, "Adaptive unfolding total variation network for low-light image enhancement," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 4419–4428.
- [39] W. Wu, J. Weng, P. Zhang, X. Wang, W. Yang, and J. Jiang, "URetinex-Net: Retinex-based deep unfolding network for low-light image enhancement," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 5891–5900.
- [40] A. Foi, M. Trimeche, V. Katkovnik, and K. Egiazarian, "Practical poissonian-Gaussian noise modeling and fitting for single-image raw-data," *IEEE Trans. Image Process.*, vol. 17, no. 10, pp. 1737–1754, Oct. 2008.
- [41] K. Wei, Y. Fu, J. Yang, and H. Huang, "A physics-based noise formation model for extreme low-light raw denoising," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 2755–2764.
- [42] H. Feng, L. Wang, Y. Wang, and H. Huang, "Learnability enhancement for low-light raw denoising: Where paired real data meets noise modeling," in *Proc. 30th ACM Int. Conf. Multimedia*, Oct. 2022, pp. 1436–1444.
- [43] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Apr. 2018.
- [44] B. Cheng, L.-C. Chen, Y. Wei, Y. Zhu, Z. Huang, J. Xiong, T. Huang, W.-M. Hwu, H. Shi, and U. Uiu, "SPGNet: Semantic prediction guidance for scene parsing," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 5217–5227.
- [45] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3431–3440.
- [46] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–14.
- [47] J. Jain, A. Singh, N. Orlov, Z. Huang, J. Li, S. Walton, and H. Shi, "SeMask: Semantically masked transformers for semantic segmentation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshops (ICCVW)*, Oct. 2023, pp. 752–761.
- [48] R. Strudel, R. Garcia, I. Laptev, and C. Schmid, "Segformer: Transformer for semantic segmentation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 7242–7252.
- [49] E. Xie, W. Wang, Z. Yu, A. Anandkumar, J. M. Alvarez, and P. Luo, "SegFormer: Simple and efficient design for semantic segmentation with transformers," in *Proc. NIPS*, Sep. 2021, vol. 13, no. 18, pp. 12077–12090.
- [50] B. Cheng, A. G. Schwing, and A. Kirillov, "Per-pixel classification is not all you need for semantic segmentation," in *Proc. NIPS*, Dec. 2021, pp. 17864–17875.
- [51] Z. Cai and N. Vasconcelos, "Cascade R-CNN: Delving into high quality object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 6154–6162.
- [52] K. Chen, J. Pang, J. Wang, Y. Xiong, X. Li, S. Sun, W. Feng, Z. Liu, J. Shi, W. Ouyang, C. C. Loy, and D. Lin, "Hybrid task cascade for instance segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 4969–4978.
- [53] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2980–2988.
- [54] A. Kirillov, K. He, R. Girshick, C. Rother, and P. Dollár, "Panoptic segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 9396–9405.
- [55] A. Kirillov, R. Girshick, K. He, and P. Dollár, "Panoptic feature pyramid networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 6392–6401.
- [56] B. Cheng, I. Misra, A. G. Schwing, A. Kirillov, and R. Girdhar, "Masked-attention mask transformer for universal image segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 1280–1289.
- [57] H. Wang, Y. Zhu, H. Adam, A. Yuille, and L.-C. Chen, "MaX-DeepLab: End-to-end panoptic segmentation with mask transformers," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 5459–5470.
- [58] H. Wang, Y. Zhu, B. Green, H. Adam, A. Yuille, and L.-C. Chen, "Axial-DeepLab: Stand-alone axial-attention for panoptic segmentation," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2020, pp. 108–126.
- [59] H. C. Peng, H. H. Zhu, and X. J. Li, "Detection of tunnel water leakage based on image processing," *Inf. Technol. Geo-Eng.*, vol. 1, pp. 254–262, Jun. 2010.
- [60] Y. Wu, M. Hu, G. Xu, X. Zhou, and Z. Li, "Detecting leakage water of shield tunnel segments based on mask R-CNN," in *Proc. IEEE Int. Conf. Archit., Construct., Environ. Hydraul. (ICACEH)*, Dec. 2019, pp. 25–28.
- [61] X. Gao, M. Jian, M. Hu, M. Tanniru, and S. Li, "Faster multi-defect detection system in shield tunnel using combination of FCN and faster RCNN," *Adv. Struct. Eng.*, vol. 22, no. 13, pp. 2907–2921, Oct. 2019.
- [62] L. Tan, X. Hu, T. Tang, and D. Yuan, "A lightweight metro tunnel water leakage identification algorithm via machine vision," *Eng. Failure Anal.*, vol. 150, Aug. 2023, Art. no. 107327.
- [63] W. Wang, X. Xu, and H. Yang, "Intelligent detection of tunnel leakage based on improved mask R-CNN," *Symmetry*, vol. 16, no. 6, p. 709, Jun. 2024.
- [64] J. X. Chen, X. Xu, G. Jeon, D. Camacho, and B.-G. He, "WLR-net: An improved YOLO-V7 with edge constraints and attention mechanism for water leakage recognition in the tunnel," *IEEE Trans. Emerg. Topics Comput. Intell.*, vol. 1, pp. 1–18, Jul. 2024.
- [65] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.
- [66] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis.*, Sep. 2018, pp. 3–19.



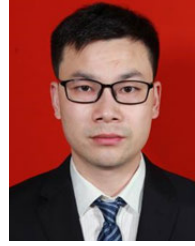
**JUN ZHU** received the Ph.D. degree in engineering from Southwest Jiaotong University, Chengdu, Sichuan, in 2015. He is currently a Senior Engineer with the State Grid Sichuan Electric Power Research Institute and a Technical Manager with Sichuan Shuneng Electric Power Technology Company Ltd. He has long been engaged in the research and application of electrical equipment performance (transmission line professional) technical supervision, power grid disaster prevention and mitigation, transmission line icing monitoring and early warning, dc transmission line metal electrochemical corrosion, and transmission cable operation and maintenance technology.



**FENGLIAN LIU** was born in Shanxi, China. She received the Diploma degree in electrical engineering and automation and the M.Sc. degree in high voltage and insulation engineering from Xi'an Jiaotong University, Xi'an, China, in 2009 and 2012, respectively. She started working as a Senior Engineer with the State Grid Sichuan Electric Power Research Institute, in 2012. Her current research interests include smart grids, transmission cable line, and intelligent operation inspection.



**ZHIHANG XUE** received the degree from the School of Energy, University of Electronic Science and Technology of China. In 2012, he joined the Power Science Research Institute, State Grid Sichuan Electric Power Research Institute. He has been engaged in disaster prevention and mitigation of power grid systems and live testing for a long time.



**JUN HE** received the bachelor's degree from the Southwest University of Science and Technology, in 2009. He is currently an Engineer and a Project Manager with Sichuan Shuneng Electric Power Technology Company Ltd. He has long been engaged in grid technology supervision, electric power safety tool testing, engineering quality inspection, information security testing and evaluation, and system planning and management.



**WENWEI LUO** received the bachelor's degree from Nanjing University of Aeronautics and Astronautics, in 2022. He is currently pursuing the degree with the University of Electronic Science and Technology of China. His research interests include computer vision and image enhancement.



**ZHENGZHENG FU** received the Ph.D. degree from Chongqing University, in 2021. She is currently an Engineer with the State Grid Sichuan Electric Power Research Institute. Her research interests include modeling and analyzing transient characteristics of power systems.



**HAORAN PENG** received the bachelor's degree from Dalian University of Technology, in 2023. He is currently pursuing the degree with the University of Electronic Science and Technology of China. His research interests include computer vision and visual language model.



**DONGHUI LUO** received the Ph.D. degree from Chongqing University, in 2019. He is currently a Senior Engineer with the Power Science Research Institute, State Grid Sichuan Electric Power Research Institute. His main research interests include lightning protection and grounding technology for power systems, and disaster prevention and mitigation technology for transmission equipment.

...