RESEARCH ARTICLE

Underground Defects Detection Based on GPR by Fusing Simple Linear Iterative Clustering Phash (SLIC-Phash) and Convolutional Block Attention Module (CBAM)-YOLOv8

NIANNIAN WANG[ID]**, ZEXI ZHANG**[ID]**, HAOBANG HU**[ID]**, BIN LI, AND JIANWEI LEI**[ID]
Yellow River Laboratory, Zhengzhou University, Zhengzhou 450001, China
National Local Joint Engineering Laboratory of Major Infrastructure Testing and Rehabilitation Technology, Zhengzhou 450001, China
Collaborative Innovation Center of Water Conservancy and Transportation Infrastructure Safety, Zhengzhou 450001, China

Corresponding author: Zexi Zhang (zhangzexi22@163.com)

ABSTRACT Ground Penetrating Radar (GPR) is an effective non-destructive detection method, that is frequently utilized in the detection of urban underground defects because of its quick speed, convenient and flexible operation, and high resolution. However, there are some limitations to defect detection using GPR, such as less data, poor data quality, and complexity of data interpretation. In this study, an underground defect detection system based on GPR was established. First, a Simple Linear Iterative Clustering (SLIC)-PHash, a Data Augmentation (DA) optimization algorithm, was created to obtain high-quality datasets. Second, the Convolutional Block Attention Module (CBAM)-YOLOv8, a detection model, was produced for the recognition of defects. This model uses GhostConv and CBAM to create a lighter design that better focuses on target detection and increases efficiency. Finally, a one-click detection system was formed by fusing SLIC-Phsh and CBAM-YOLOv8, which were used for one-click GPR dataset optimization and defect detection. The developed system has the best detection mAP and F1 scores of 90.8% and 88.3%, respectively, compared to several well-known Deep Learning (DL)-based techniques. The results demonstrated that the system proposed in this paper can greatly improve detection efficiency and reduce detection time by achieving a good balance between detection speed and accuracy.

INDEX TERMS Ground penetrating radar (GPR), object detection, perceptual hashing (Phash) underground defects, YOLOv8.

I. INTRODUCTION

Ground Penetrating Radar(GPR) is a high-resolution and efficient geophysical non-destructive survey method that mainly uses the principle of reflection at the interface of two different dielectric constants during electromagnetic wave

The associate editor coordinating the review of this manuscript and approving it for publication was Francesco Benedetto[ID].

transmission [1]. They have been widely used in underground defects detection and other fields [2]. GPR forms radar images through several received echoes, and technicians infer the structure of underground media according to the characteristics of the reflected waveforms. However, manual interpretation of GPR images is a challenging task, and it is extremely time-consuming for professional technicians to interpret GPR image data manually [3], [4]. Meanwhile,

the Deep Learning (DL) method requires a large amount of data for training [5]; however, owing to factors such as the complexity of underground environments, obtaining a large amount of high-quality data directly is a challenge [6]. Increasing the amount and variability of GPR training data with Data Augmentation (DA) can enhance object detection performance on GPR image data. However, it brings new challenges to the capacity, throughput, maintainability, scalability, and energy management of existing storage systems, and the problems of high cost and low efficiency are prominent. In this case, it is of great significance to explore a fast and effective GPR image dataset optimization method and an automatic identification method for GPR data to improve efficiency and accuracy.

In recent years, GPR technology has made significant progress, including advances in data processing [7]. First in the early 21st century, owing to the ongoing growth and use of computer technology, Gamba et al. [8] employed Neural Networks (NN) to analyze GPR data. Shaw et al. [9] used a Multilayer Perceptron (MLP) network to automatically identify rebar GPR image data. Subsequently, with the continuous application of machine learning algorithms in GPR data processing, Pasolli et al. [10] used a genetic algorithm (GA) and a Support Vector Machine (SVM) to identify and categorize GPR images. Mass and Schmalzl [11] used the Viola-Jones learning algorithm from the ''OpenCV'' open-source library to apply GPR image recognition through transfer learning. deep convolutional neural Network (CNNs)-based [12] object detection techniques have become widespread for interpreting GPR data as Artificial Intelligence (AI) and computer vision have progressed [13]. Using 2-dimensional (2-D) GPR B-scans and deep CNNs, Bishop et al. [14] were capable of accurately classifying buried subsurface explosives. Pham et al. [15] applied the Faster-RCNN framework to the detection of buried objects in B-scan GPR images, which showed a significant improvement compared with classical computer vision methods. Zhang et al. [16] used a mixed deep CNN to accurately detect and locate water damage in asphalt pavement GPR images and achieved good performance and superiority. Zhang et al. [17] proposed a deep learning framework based on Generative Adversarial Networks (GANs) to generate new data, automatically learn features, and detect underground objects. Li et al. [18] utilized the RegNetY network for the intelligent recognition of defects in sewer pipeline networks, which can improve detection efficiency and have a stronger classification ability. The neural network framework based on YOLO has a fast detection speed, can detect small targets, and significantly improves detection accuracy and efficiency [19]. In 2020, Li et al. [20] used the YOLOv3 neural network built using the TensorFlow1.13.0 framework developed by Google to detect GPR images in real-time. Li et al. [21] first introduced the YOLO model as a DL model for GPR data-based hidden crack detection. The study shows that YOLO version 4 (YOLOv4) and 5 (YOLOv5), when compared to YOLO

version 3, make noticeable advancements even on smaller datasets, with YOLOv5 having the best mean Average Precision (mAP) values. Situ et al. [22] used the YOLOv5 model to detect the characteristics of sewer defects, which significantly improved detection accuracy and speed. Hu et al. [23] used the YOLOv5l model to achieve an improved detection performance for few-shot crack detection, serving as a valuable reference for small-dataset target detection. Subsequently, a new target detection model, YOLOv7, was introduced [24], which is a secondary development based on YOLOv5, with YOLOv7 demonstrating higher accuracy and speed, especially when processing large datasets. YOLOv7 incorporates an Efficient Layer Aggregation Network (E-ELAN) with expanded, shuffled, and merged cardinality structures, a network design innovation that enhances the model's learning capabilities while maintaining effective gradient flow [25]. In January 2023, Ultralytics released a significant update, YOLOv8. As a state-of-the-art (SOTA) model, YOLOv8 builds on the success of previous YOLO versions and introduces new features and improvements to further enhance performance and flexibility. This includes the introduction of the C2f module, which enhances the model's feature extraction ability, thereby improving accuracy [26].

DL has immense potential in GPR applications. However, the DL requires a substantial amount of training data. The scarcity of GPR data arises because of the limited presence of defect samples in standard engineering structures, as well as issues such as variability in detection equipment and complexity of subsurface environments [27]. Therefore, it is crucial to find methods to address the shortage of GPR defect data.

Common methods for GPR data augmentation include traditional techniques, methods based on Finite-Difference Time-Domain (FDTD) simulations, and deep learning model-based approaches. a) Traditional data augmentation methods mainly involve operations such as rotation, flipping, and cropping of input images, as well as changing the color space of the images. Liu et al. [13] employed horizontal flipping and image scaling by factors of 1.5 and 2 for data augmentation when detecting rebar in concrete using GPR. Zong et al. [28] used a combination of blurring, cropping, rotating, and mirror flipping to enhance the collected GPR data. Although these methods can significantly expand the dataset, considering the characteristics of GPR data, traditional augmentation may introduce changes that do not align with the true features of the data, thereby making it difficult for the model to learn effective features. For example, rotating GPR data might change the orientation of subsurface structures, leading to poor model performance on real data and a limited generalization ability. Similarly, blurring operations can obscure the details of the targets, thereby reducing the model's perceptual capabilities. b) FDTD-based methods expand GPR B-scan data by simulating target echo data under various scenarios [29]. However, the electromagnetic simulation required to generate GPR data is computationally

expensive and challenging to produce on a large scale. Additionally, the absence of background clutter, which typically arises from subsurface environmental noise in these simulations leads to discrepancies between the generated images and actual field data.c) With the development of deep learning, Generative Adversarial Networks (GANs) [30] have shown potential in creating images. This has inspired the use of GANs to generate GPR defect data and alleviate the problem of data scarcity in GPR DL applications [17], [31]. Yue et al. [32] used an improved GAN to create GPR images of rebar. Combining these images with real data improved the YOLOv4 detection accuracy by 10%. Zhao et al. [33] utilized the Wasserstein GAN to diversify radar image datasets and successfully generating a variety of GPR target classes. They strategically implemented the Wasserstein loss function, which contributed to a robust and steady training workflow. Xiong et al. [34] proposed a GPR-GAN approach that can adaptively adjust the network structure according to the data size, thereby showing strong generalizability. Although GAN-based methods are promising, their effectiveness across various devices, antenna frequencies, and for identifying a wide range of engineering defects still needs to be confirmed. Currently, traditional augmentation techniques can increase the training data and reduce overfitting when training deep learning models. However, they also face many challenges, such as a lack of realism compared with actual GPR images, insufficient image diversity, and instability during the training process.

Despite achievements in GPR data interpretation and enhancement, unresolved issues remain. For example, there is a mismatch between GPR image features and deep learning models, and traditional augmentation techniques often lack image diversity and can be unstable during training. As shown in Figure 1, this study introduces automatic underground defect detection based on GPR by fusing a simple linear iterative clustering phash (SLIC-Phash) and Convolutional Block Attention Module (CBAM)-YOLOv8. The main innovations of this study are summarized as follows:

- Addressing the current limitations of GPR DA, this study builds upon the simplest traditional data enhancement methods. We propose the SLIC-Phash data optimization technique, which filters the augmented data produced by traditional methods. This selection process retains high-quality, detail-rich GPR data and discards images with unrealistic target features and complex and ambiguous backgrounds, ensuring the model's generalizability and detection accuracy. Our approach is straightforward, does not require a high computing performance, and is readily integrable.
- To address the issue of subpar target detection in GPR, this study harnesses an improved YOLOv8 model to enhance the detection precision. This study employs a lightweight GhostConv network to refine the convolutional layers within the backbone and neck of the model and replaces the standard C2f with C2fCBAM, incorporating CBAM for added focus. This method reduces the

model complexity while maintaining the same receptive field, thereby boosting the real-time performance of the model. The integration of the attention mechanism further amplifies the feature representation, allowing the network to concentrate more on the targets of interest, which enhances the detection outcomes.

- A one-click processing system for the automated detection of underground defects in GPR images was introduced. The system has strong real-time capabilities and does not require high computational power. Utilizing this system, a variety of underground anomaly images (such as voids, pipelines, and loose soil) were examined, enabling the simultaneous augmentation and detection of multiple target categories. This resolves the issue present in the current methods, which can only generate data and detect single-category targets separately. The system was designed to deliver efficient and optimal target detection results.

The remainder of this study is organized as follows. In Section II, the primary theoretical approach to the proposed network is described. The dataset processing part of this study is presented in section III. The experimental setup for this study is described in Section IV. Section V describes the detailed training process of this study and compares some classical object-detection algorithms. Finally, a summary is presented in Section VI.

II. METHODS

In this study, underground defects detection based on GPR was developed by fusing simple linear iterative clustering phash (SLIC-Phash) and Convolutional Block Attention Module (CBAM)- YOLOv8. First, the images collected by GPR for municipal road disease detection were selected and sized uniformly. Underground defect images are generated by traditional DA methods, such as adding noise randomly, adjusting brightness and contrast randomly. To enhance the effectiveness of object detection and reduce the training time as much as possible, this study incorporated the expanded dataset into the SLIC-Phash deduplication model to obtain the optimal detection dataset. Second, the optimized data set is input into the CBAM-YOLOv8 object detection model, and finally a better detection result is obtained.

A. DATA OPTIMIZATION BASED ON SLIC-PHASH

The DA of GPR images using traditional methods is straightforward and can expand the dataset while introducing greater diversity. However, these methods may also induce changes that do not align with the true characteristics of real data, potentially reducing the generalizability of the model. Furthermore, blurring operations could obscure critical target details, thereby diminishing the model's perceptual acuity [35]. A common issue with GPR data is the significant deviation from the original data. To address this, the study introduces the SLIC-Phash model, which filters the augmented dataset to construct a high-quality GPR dataset. This
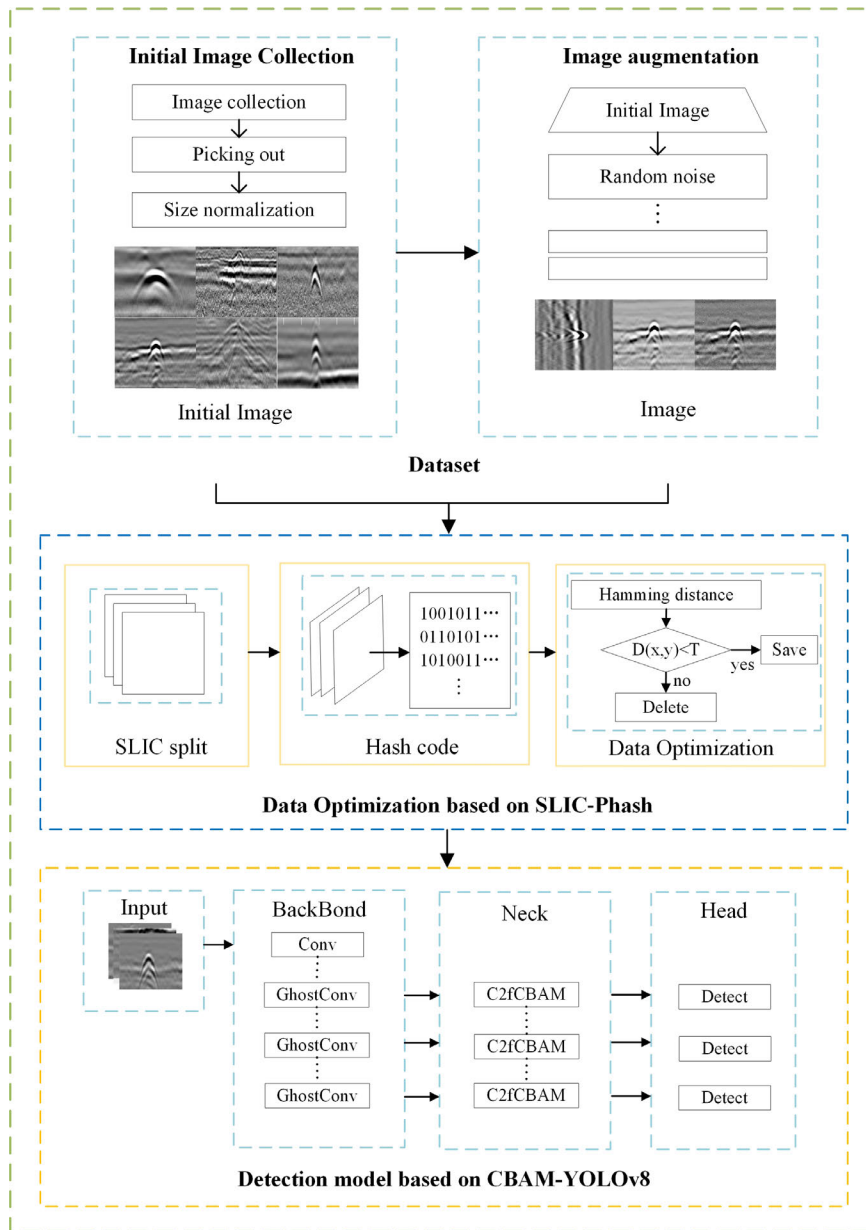
FIGURE 1. Framework of the proposed model.

model meticulously selects data that maintains fidelity to real-world features, ensuring that the enhanced dataset supports the development of more accurate detection models.

The Perceptual Hashing (Phash) algorithm encodes images into compact feature vectors, facilitating the easy comparison and deletion of hash codes. However, directly applying it to GPR data does not achieve the goal of dataset optimization. This study introduced superpixel segmentation prior to hash code generation. Superpixel segmentation divides an image into semantically continuous and compact regions. By leveraging superpixel segmentation with GPR image characteristics, local features and structural information can be partially extracted from the GPR images. Segmentation uses

hyperbolic features as constraints to ensure that adjacent pixels are grouped together into continuous regions.

After completing superpixel segmentation, the Phash algorithm was employed to encode each superpixel region. The system determines the similarity of images by setting a threshold and using the Hamming distance between hash codes. This approach preserves the useful image information and removes unnecessary image data. This ensures better generalization of the model during training and testing and reduces errors introduced by modifications that could alter the true characteristics of the target.

In addition, this method can eliminate low-quality GPR data with complex, ambiguous backgrounds that may

interfere with target recognition and localization. The detection accuracy and robustness of the model were improved by deleting these low-quality images.

1) SIMPLE LINEAR ITERATIVE CLUSTERING (SLIC)

Image superpixel segmentation has become an important tool in the visual field. It is a method that divides an image into multiple superpixel regions with similar color, textured and other features in each region, which is widely used in image processing and vision. These image regions retain the effective information of the region and will not affect the visual expression of the whole image. SLIC [36] employs k-means clustering to generate superpixels in a manner similar to [37]. By clustering pixels based on color similarity and proximity, SLIC creates superpixels. Although the method provided by SLIC is simple, it solves these problems and produces high-quality, compacted and almost uniform superpixel segmentation. Since the resulting superpixels are compact and orderly like cells, it is simple to express neighborhood features. Meanwhile, since most GPR radar images are generated as gray images, SLIC superpixel segmentation can not only segment color images, but also be compatible with gray images. SLIC provides additional benefits over other superpixel segmentation techniques in terms of running speed, compactness of created superpixels, and contour preservation. The procedure is as follows:

First, a uniform "$N \times N$" adjustment is made to the size of the radar image dataset using the linear interpolation method, which ensures that the image after scaling has good stability and robustness. It can also ensure that all images have the same length of perceptual hash code obtained by the proposed algorithm to facilitate the subsequent similarity comparison. A size-normalized GPR image was obtained. In this study, the normalized image was processed using Gaussian low-pass filtering to produce a secondary image, which reduces the influence of compression and other operations on the image. The calculation formula is as follows:

$$H(u, v) = e^{-D^2(u,v)/2D_0^2} \quad (1)$$

D_0 is the cutoff frequency and $D(u, v)$ is the distance from the center of the frequency rectangle.

The secondary image was segmented using SLIC superpixels. First, the origin points were initialized and dispersed uniformly over the picture in accordance with a predetermined number of superpixels. Pre-segment N pixels of the image into K superpixels of identical size. Each superpixel region has a size of N/K, whereas neighboring seed points have a step size of approximately $S = \sqrt{N/K}$. The seed points were reselected in a 3×3 area of the original seed points to prevent them from landing on the image edge and interfering with the clustering effect that followed. The gradient of all pixels in the vicinity were measured, and the point with the lowest gradient was chosen as the seed point. The search area is inside $2S \times 2S$, each pixel in the vicinity of each seed point is given a label, and this can speed up the convergence of the algorithm.
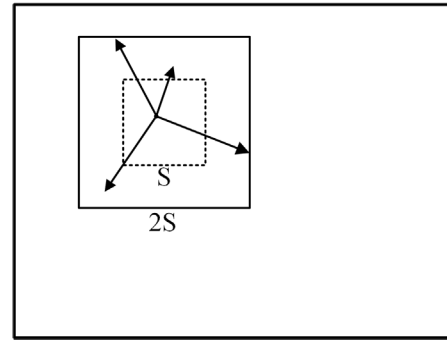

FIGURE 2. SLIC searches a limited region.

Final Distance Metric. It encompasses both spatial and color distances. The separation between each searched pixel and the seed point was determined. The point with the smallest distance between each element and its neighboring seed points is taken as the seed point of the pixel, that is, the cluster center, and iterative optimization is carried out until the cluster center of each pixel is no longer changed.

The pertinent equation is as follows:

$$d_c = \sqrt{(l_j - l_i)^2 + (a_j - a_i)^2 + (b_j - b_i)^2} \quad (2)$$

$$d_s = \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2} \quad (3)$$

$$D' = \sqrt{(\frac{d_c}{N_c})^2 + (\frac{d_s}{N_s})^2} \quad (4)$$

where D_c is the color distance, which is the pixel distance in the image laboratory color space. l, a and b are the components of images in the Lab color space. D_s denote the spatial distance. x and y are the coordinate components of the image coordinate system. D' is the final distance metric and the greatest color distance is N_c, which in this case is assumed to be 10.

As shown in Figure 3, Figure 3(a) and Figure 3(b) show the original image of the GPR underground defects and the image where the number of superpixels K is taken to be 218.

2) HASH CODE EXTRACTION

The GPR image set preprocessed by SLIC superpixel segmentation is transformed using DCT and then binarized. First, the image segmented by SLIC superpixels is transformed by the DCT. The DC component of each segmented sub-block was extracted. This process transforms an image from the pixel domain into the frequency domain. The DC component of each block is selected and the entire image is in the zigzag pattern order, as shown in figure 4. Finally, they were connected them to form a one-dimensional vector A_i with length M to represent the entire image.

Subsequently, the mean of the coefficient matrix Avg was calculated. The equation is expressed as follows:

$$Avg = \frac{\sum_{i=1}^{N} C_{eof}[i]}{N} \quad (5)$$
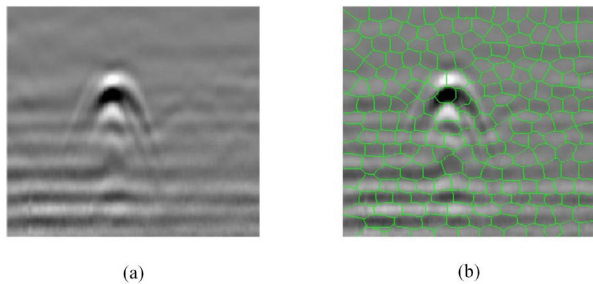
FIGURE 3. GPR images. (a) Original image; (b) Image obtained after SLIC superpixel segmentation.
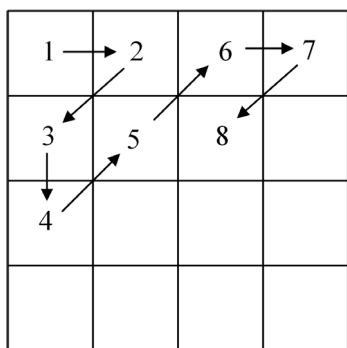

FIGURE 4. Zigzag mode.

The three coefficient matrices were then binarized. Where Avg is greater than or equal to 1 and less than 0. Finally, the DCT feature-aware hash H_i of the image is obtained as

$$H(i) = \begin{cases} 0, & C_{eof}[i] < Avg \\ 1, & C_{eof}[i] \geq Avg \end{cases} \quad (6)$$

Concatenate H_i is concatenated sequentially to generate a one-dimensional feature vector H of length M. The feature vector is the perceptual hash sequence value of the image.

3) DATA OPTIMIZATION

The Hamming distance is used as the foundation for assessing the similarity of pictures, because the final hash sequence produced by the method in this research is binary. The formula for Hamming distance D is as follows:

$$D = \sum_{i=1}^{L} (h_1(i) \oplus h_2(i)) \quad (7)$$

where $h_1(i)$ and $h_2(i)$ are the ith elements of the hash sequence H_1 and H_2 of the two images, respectively. $\oplus$ is XOR, L is the total length of the hash sequence.

The Hamming distance between the initial dataset and amplified dataset was compared, and the Hamming distance threshold was set as T:

If $D(x, y) \geq T$, then x and y are considered to be similar strings, and the images are removed from the augmented dataset.

If $D(x, y) < T$, then x and y are considered dissimilar strings and the image in the augmented dataset is saved.

B. CBAM-YOLOv8

Today, the YOLO family of algorithms has the most advanced real-time object detection systems. Compared to other detection methods, such as faster R-CNN with ResNet and SSD, it is more accurate and runs faster [38]. Manual interpretation of radar images is a challenging task. Expert technicians spending extensive time on the manual analysis of GPR image data is not only time-consuming but also difficult because of the complex backgrounds typically present in real GPR images. To address this issue, this study introduces CBAM-YOLOv8, based on the YOLOv8 framework, as illustrated in Figure 5. This study utilizes a lightweight GhostConv network to optimize the convolutional layers in the backbone and neck, reducing the number of convolutional operations and thus the computational complexity of the model, which accelerates the inference speed.

While reducing the complexity of the model, the approach maintains the same receptive field as conventional convolutional layers, which is beneficial for capturing spatial structural information from images. This reduction in storage requirements and improvement in detection efficiency also enhances the real-time capabilities of the system. As shown in Figure 6, the Neck portion of the network replaces the C2f module with a C2fCBAM module, thereby enhancing the feature representation. This enhancement allows the network to focus more on targets awaiting detection, thereby improving the overall effectiveness of detection.

1) YOLOv8 MODULE

YOLOv8, which represents the cutting edge in object detection models, accounts for the multi-scale nature of objects and utilizes three different scale detection layers to accommodate objects of varying sizes. YOLOv8 is a significant update of YOLOv5, which was open-sourced by Ultralytics on January 10, 2023. Compared with previous models in the YOLO series, YOLOv8 offers superior detection precision and speed.

The YOLOv8 algorithm is a rapid, single-stage object detection method comprising four parts: input, backbone, neck, and output segments. The backbone and neck sections likely drew inspiration from the YOLOv7 ELAN design, which replaced the C3 structure of YOLOv5 with a C2f structure that provides richer gradient flow. In addition, it adjusts the channel numbers for models of different scales. Compared with the C3 module of YOLOv5, the C2f module has fewer parameters and superior feature extraction capabilities. This approach retains the lightweight characteristics while capturing more abundant gradient flow information, significantly enhancing the model performance. The Head part of YOLOv8 has two major improvements compared to YOLOv5: 1) it has adopted the currently popular decoupled head structure (Decoupled-Head), which separates the
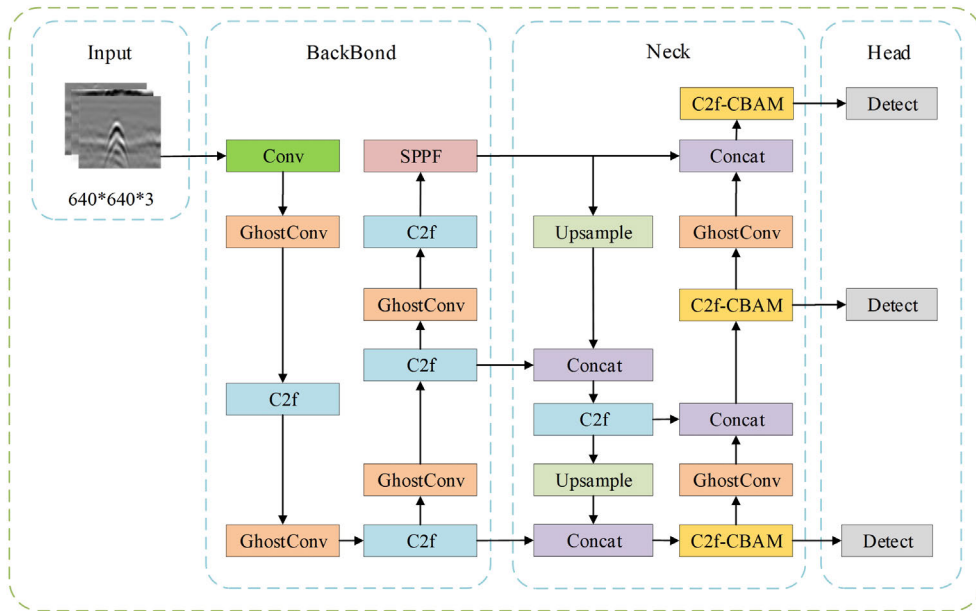
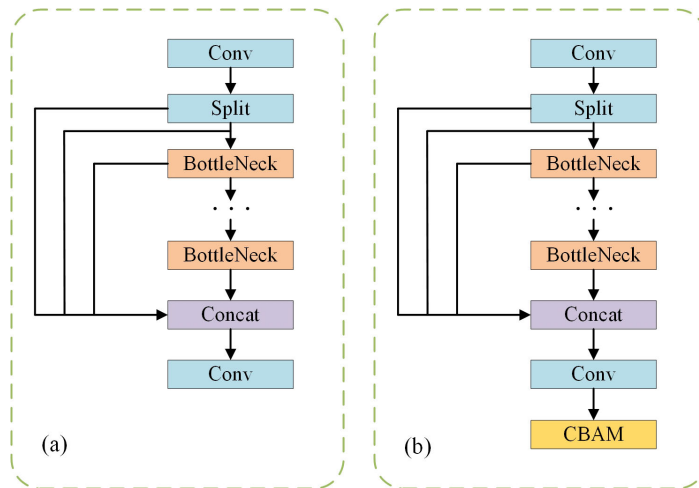**FIGURE 5.** Architecture of the CBAM-YOLOv8.

FIGURE 6. (a) C2f module; (b) C2fCBAM module.

classification and detection heads; 2) it has transitioned from an anchor-based to an anchor-free approach.

2) GhostConv

Timely and accurate interpretation of GPR defect images is crucial; however, owing to YOLOv8's large model parameters and high computational cost, it is challenging to deploy in resource-constrained environments. This study introduces a lightweight GhostConv network to optimize the YOLOv8 model. GhostConv is illustrated in the Figure 7. The Ghost-Conv model simplifies the process by initially using a few key filters to select features from the input image. Then, it applies simple and inexpensive operations to slightly tweak these features. Finally, everything is combined to create a complete set of features required for further processing. This streamlined approach saves computing resources without sacrificing too much detail. This approach reduces the learning cost of non-essential features by using a combination of a few convolutional kernels and cheaper linear operations to replace conventional convolution, thereby effectively reducing the demand for computational resources [35]. Therefore, by replacing the original Conv with GhostConv in the Backbone and Neck, this method substitutes the original network modules with a lighter and faster network, which, while reducing computational complexity, retains the model's sensitivity to GPR defect features as much as possible, thereby greatly improving the efficiency of model training.
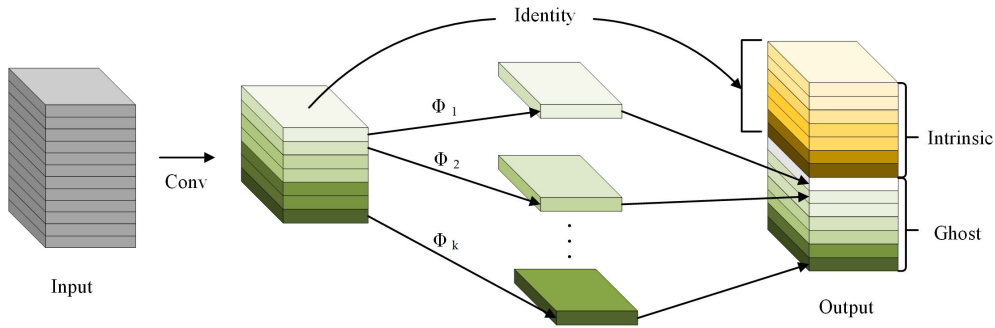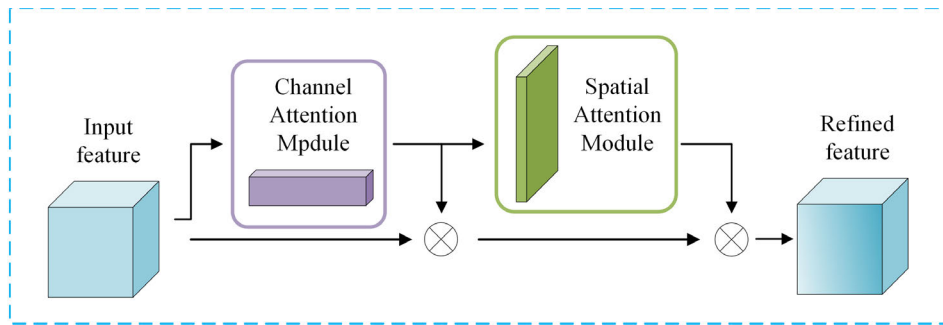
FIGURE 7. The GhostConv module.

FIGURE 8. CBAM structure diagram.

3) CBAM MODULE

The CBAM is a simple but effective attention module for feedforward CNN. It can be seamlessly integrated into any CNN architecture as a small generic module [39]. This method may highlight defects, suppress the characteristics of complex backgrounds, and emphasize the spatial positions of features in an image with complicated backgrounds [40]. Figure 8 depicts the structure of CBAM. The Spartial Attention Module (SAM) and Channel Attention Module (CAM) are two separate sub-modules. CAM and SAM Attention tasks were performed separately. This guarantees that it may be added as a plug-and-play module to the current network architecture while also conserving parameters and computing power.

As shown in Figure 8, the CBAM structure consists of four blocks. The output from the convolutional layer goes via the CAM and SAM modules. First, an intermediate feature figure $F \in R^{C \times H \times W}$. Then, a 1D map $M_C \in R^{C \times 1 \times 1}$ and a 2D map $M_S \in R^{1 \times H \times W}$ are created through CAM and SAM, respectively. The formula summarizing the full procedure is as follows:

$$F' = M_C(F) \otimes F \quad (8)$$
$$F'' = M_S(F') \otimes F' \quad (9)$$

where $\otimes$ denotes element-wise multiplication, F'' is the final output.

The CAM structure is shown in Figure 9(a). The channel attention map is produced by utilizing the inter-channel connections of the characteristics. Prior to the operation, the feature map is compressed by the CAM in the spatial dimension, resulting in a one-dimensional vector. [41]. First, the input feature map is subjected to width and height-based global max pooling and global average pooling [42]. Two spatial context descriptors were generated: F^c_{Avg} and F^c_{Max}. These two descriptors represent the average pooling feature and maximum pooling feature respectively, resulting in two $C \times 1 \times 1$ feature maps. A two-layer neural network was then fed to each layer. This neural network was shared between the two layers. Then, the features output by the MLP are added element-wise. After the sigmoid activation operation, the final channel attention feature is generated, namely M_C. Finally, the input features required by the spatial attention module are created by element-wise multiplication of the M_C by the input feature map F. Channel attention is computed as follows:

$$M_C(F) = \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F)))$$
$$= \sigma(W_1(W_0(F^c_{Avg})) + W_1(W_0(F^c_{Max}))) \quad (10)$$

where σ denotes the sigmoid function. $W_0 \in R^{C/r \times C}$, $W_1 \in R^{C \times C/r}$

The SAM is shown in Figure 9(b). Particularly for small targets, spatial attention is helpful in gathering spatial information. The link between various spatial positions is used by spatial attention to learn a 2D spatial weight map. The resulting spatial position was then multiplied to obtain more representative characteristics [43]. The input feature map
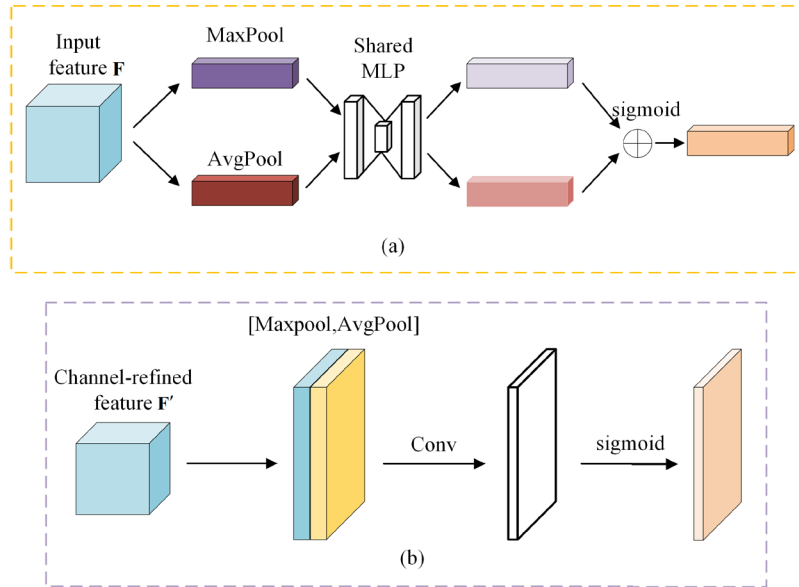
FIGURE 9. Schematic diagram of two attention submodules. (a) Channel Attention Module; (b) Spartial Attention Module.

F' of this module was generated using the previous CAM module. First, we consider channel-wise global maximum pooling and global average pooling. It obtains two $H \times W \times 1$ characteristics of the figure for the F_{Avg}^s and F_{Max}^s. The two feature maps are channel stitched. The dimensionality was decreased to one channel through a convolution of 7×7. Spatial attention feature M_S generated by sigmoid function. The final feature produced is calculated by multiplying the features of the module by its input features. The calculations for spatial attention are as follows:

$$M_S(F) = \sigma(f^{7 \times 7}([AvgPool(F)); MaxPool(F)]))$$
$$= \sigma(f^{7 \times 7}([F_{Avg}^c; F_{Max}^c])) \qquad (11)$$

where σ represents the sigmoid function and $f^{7 \times 7}$ represents the convolution operation with a 7×7 size.

III. DATASET

A. DATA COLLECTION

The algorithm model used in this study requires a large amount of image data. To ensure that the algorithm in this study can have high applicability, the GPR models of data collected in this study use the Greenview SIR-20, Leica DS2000 and Swedish Impulse CO-730. We collected GPR images in Zhengzhou municipal road detection using different types of GPR, and the original images were blurred and not of high quality owing to the complex environment. The high-definition GPR image dataset was created after rigorous screening. A total of 837 GPR images were selected for study in this dataset, which contained three types of diseases: pipeline, void, and loose. Complete manual data annotation was performed using the open-source software LabelImg. The original images were then randomly partitioned into

three sets: training set, validation set, and test set, at a ratio of 8:1:1 respectively.

B. REPEAT IMAGE GENERETION

To imitate the problem of useless data generated by DA, this study uses a series of measures to generate duplicate image data, as follows:

- Geometric transformation. Duplicate images were added by random flipping, rotation, and scaling of the image.
- Cropping. The original data set images were randomly cropped, and the size of the cropped image was about 0.8 to 0.9 times of the original image.
- Add noise. To create duplicate images, Gaussian and salt-and-pepper noises were randomly added to the images in the original dataset.
- Add filtering. Duplicate images were generated by randomly adding Gaussian low-pass, median and mean filtering to the images in the original dataset.
- Watermark. A watermark was randomly added to the GPR image dataset to generate duplicate images.
- Gamma correction. The radar image dataset was randomly gamma corrected to generate duplicate images.
- Brightness adjustment. The brightness of the images in the original dataset was randomly adjusted.
- Contrast adjustment. A contrast is randomly selected for adjustment.

After DA using the method, the final total number of images was 4695. In addition, the open-source software LabelImg was utilized for manual annotation of the augmented data, and a portion of the expanded image data is shown in Figure 10.
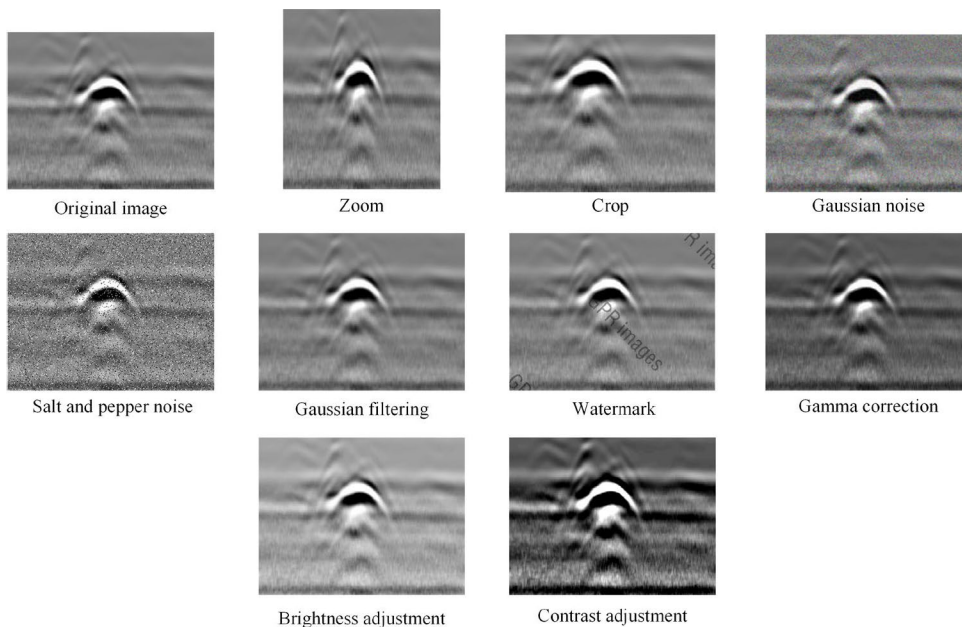
FIGURE 10. Partially duplicate images.

IV. EXPERIMENT

The experiments in this study were conducted on a PC with the following specifications: Intel® Xeon(R) Silver 4210 CPU @ 2.20 GHz, 20.0 GB of memory, and an NVIDIA GeForce RTX 2080Ti graphics card. The platform used for target detection runs on the Ubuntu 18.04 operating system and utilizes Python programming language version 3.8. To train the proposed model, we utilized pre-trained weights from coco128 and set the following parameters: a Hamming distance T of 24, an image size of 640*640, a batch size of 16, an initial learning rate of 0.01, and 300 epochs of training. Additionally, we configured the momentum as 0.937, the weight decay as 0.0005, and the scale for image resizing as 0.5. The accuracy of the model can be evaluated using various metrics, including precision, recall, F1-score@0.75, and mean Average Precision (mAP). These metrics can provide a comprehensive evaluation of the performance of a model in various situations.

A. MODEL INITIALIZATION

Data are crucial for the development of effective computer vision algorithms. In DL-based approaches, this typically entails collecting and annotating a large amount of images. YOLOv8 object detection relies on a large volume of data, and the detection effect of the YOLOv8 model can gradually increase with an increase in the dataset size. Transfer learning is a frequently used machine learning technique that involves applying the knowledge structure from a related domain to a target domain with the aim of enhancing learning outcomes [44]. coco128 was used as the dataset of the pre-trained model. The results were passed on to the YOLOv8 network for detecting underground defects, thereby enhancing the

training of the detection model. The parameters of the model were fine-tuned based on the pre-trained model using the real and simulated datasets. By guaranteeing the successful transfer of knowledge, a new network can be trained quickly and effectively. The proposed model has strong recognition capabilities for transfer learning even in complicated natural situations.

B. EVALUATION METRIC

To evaluate the results, the accuracy (ACC), precision (Pr.), recall (Re.) and F1-score can be utilized to assess the defect detection performance for each object. According to Equations (12)-(15), ACC is a basic metric that is the ratio of samples that are correctly predicted for all samples. The error detection rate was evaluated using Pr., while the miss rate be evaluated using Re. The F1 score is the harmonic mean of precision and recall and represents a combination of missed and false detections [45], [46]. In this case, TP refers to true-positive samples, wherein the model correctly predicts positive samples. False positives (FP) occur when the model incorrectly predicts negative samples as positive. True negative (TN) denotes the samples correctly predicted as negative by the model. Finally, false negatives (FN) refer to samples incorrectly predicted as negative by the model despite being positive in reality. Notably, TP considers the size of the observed bounding box in addition to the classification accuracy.

$$ACC = (TP + TN)/(TP + FP + TN + FN) \quad (12)$$

$$\text{Pr.} = TP/(TP + FP) \quad (13)$$

$$\text{Re.} = TP/(TP + FN) \quad (14)$$

$$F1score = (2 \times \text{Pr.} \times \text{Re.})/(\text{Pr.} + \text{Re.}) \quad (15)$$

TABLE 1. Model training result.

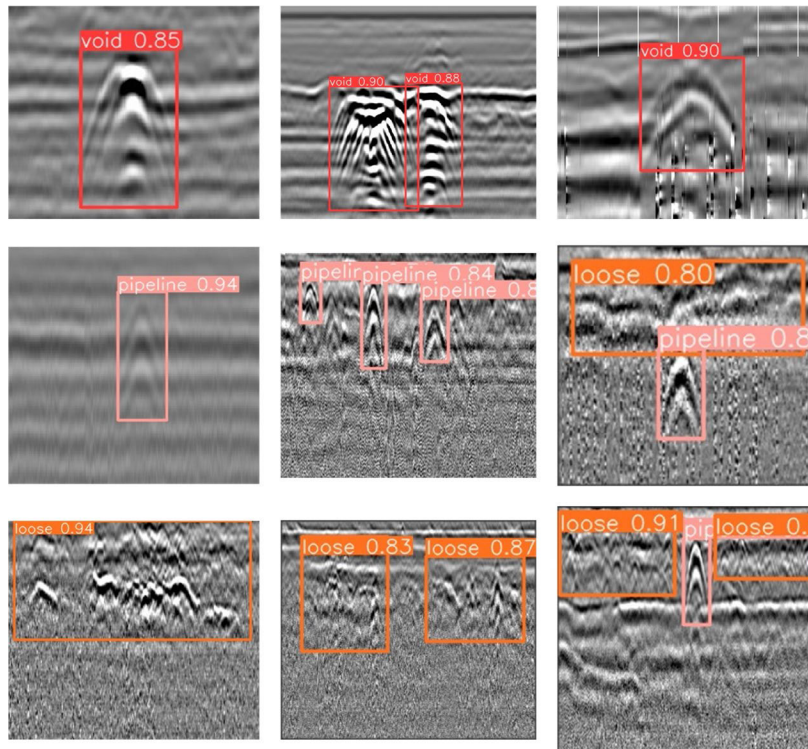| Types	Precision (%)	Recall (%)	mAP (%)	F1 score (%)	Training Duration (h)
pipeline	89.5%	88.8%	91.8%	89.1%	
void	93.6%	90.5%	94.8%	92.0%	0.953
loose	83.2%	84.2%	85.8%	83.7%	
all classes	88.8%	87.8%	90.8%	88.3%	

FIGURE 11. Output results of the CBAM-YOLOv8 model.

V. RESULTS

A. DETECTION EVALUATION

The average precision, recall, mAP (mean average precision), and F1 score values for the three types of defects were 88.8%, 87.8%, 90.8%, and 88.3% respectively, as indicated by the Table 1. Table 1 shows that pipeline and void identify the best, while loose identify less accurately. This is because the images of void and pipeline have obvious features, while the loose diseases are similar to the background and are not easy to distinguish, which is the case when manually judging. Through this model, the dataset is optimized. While increasing the amount of data, remove irrelevant data from the dataset to obtain an optimized dataset. Next, we leverage the improved YOLOv8 object detection algorithm to optimize the detection results to the maximum extent possible. Overall, the quantity of damage samples available is the most significant factor affecting the results. Additionally, the features of

the damage, such as size and shape, are also significant. With the exception of loose detection results, almost every metric achieves around 90%. This validates the model's excellent performance and fulfills the precision necessary for analyzing intricate defects.

The output results for the CBAM-YOLOv8 model presented in this study are shown in Figure 11. As can be observed from the figure, the optimized model demonstrates precise predictive capabilities for target diseases. Experimental evidence indicates that the model is not only versatile but also exhibits exceptional robustness and generalizability when processing various types of image scenes. Specifically, the model accurately identifies the location of the disease in images with single defect targets, even when the targets are situated against complex or ambiguous backgrounds. Moreover, the model effectively handled images containing multiple defects in the same category as well as complex
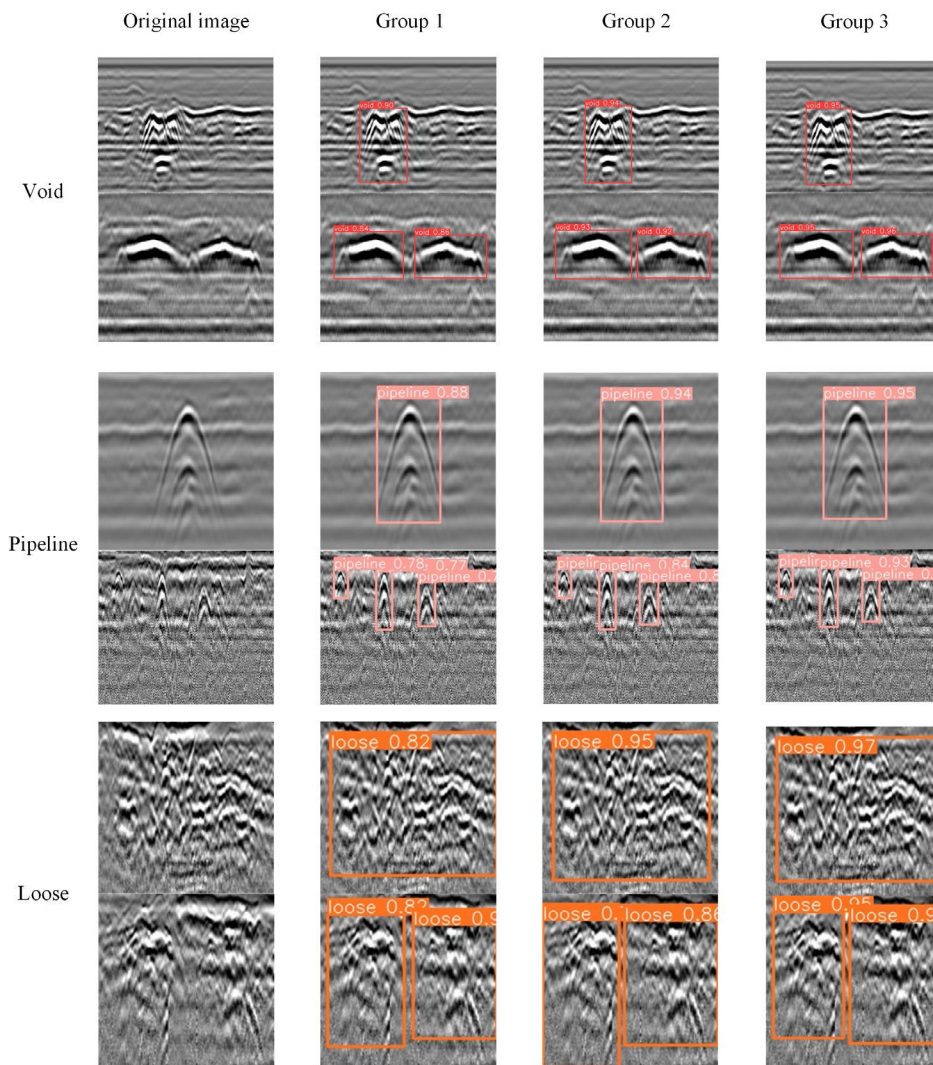
FIGURE 12. Three sets of experimental output results.

scenes with different types of defects. In such cases, the model correctly differentiates and labels each disease target, validating its potential and practical value in real-world applications. Consequently, this model provides a powerful method for advancements in the precision and application scope of disease-detection technology.

B. COMPARATIVE EXPERIMENT
In this study, the CBAM-YOLOv8 model was used for object detection experiments on the original and amplified datasets. As shown in Figure 12, The results of the three groups of experiments are compared. Group 1 is the object detection experiment on the original dataset, Group 2 is the object detection experiment on the unprocessed augmented dataset, and Group 3 is the object detection result on the optimized data set in Section V-A. It can be seen from the figure that all three groups of experiments can detect the features without misjudgment, where the recognition area in Group 1 is not

accurate, and the result of the target detection is much lower than that in Group 2 and 3. The object detection results for Group 3 were slightly higher than those for Group 2. The identified area was the most accurate among the three groups. The prediction box better matched the location and size of the feature. The feature region of each of the three image categories was enclosed by a prediction box, positioned with its upper edge coinciding with the vertex of the hyperbolic feature.

The results of the three sets of experiments are compared in Table 2. In this study, Precision, Recall, mAP and F1 score evaluation metrics were used to separately compare the three groups of experiments. As shown in Table 2, compared with the previous two groups, the total object detection results of the optimized dataset with the proposed deduplication algorithm under the same settings achieved 88.8% average precision, 87.8% recall, 90.8% mAP, and 88.3% F1 score. It is 5.4%, 7.0%, 6.3% and 6.3% higher than the Group 1

TABLE 2. Three groups of experimental training result.

Groups	Types	Precision (%)	Recall (%)	mAP (%)	F1 score (%)	Training Duration (h)
Group 1	pipeline	86.9	82.0	89.8	84.4	
	void	85.5	85.1	89.2	85.3	
	loose	77.7	75.2	74.5	76.4	0.569
	All classes	83.4	80.8	84.5	82.1	
Group 2	pipeline	88.6	86.1	90.9	87.3	
	void	87.4	86.7	89.6	87.0	
	loose	80.6	79.1	81.1	79.8	1.996
	All classes	85.5	84.0	87.2	84.7	
Group 3	pipeline	89.5	88.8	91.8	89.1	
	void	93.6	90.5	94.8	92.0	
	loose	83.2	84.2	85.8	83.7	0.953
	All classes	88.8	87.8	90.8	88.3	
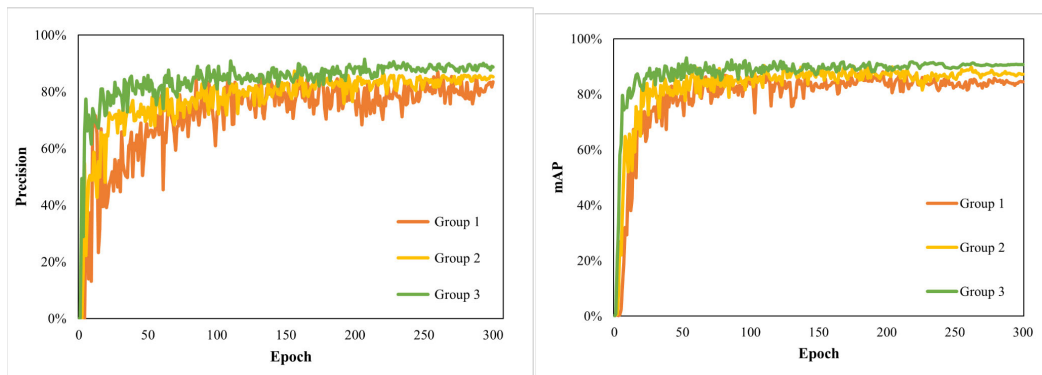
FIGURE 13. Evaluation comparison of different groups. (a) The result of Precision, (b) The result of mAP.

original dataset, and 3.3%, 3.8%, 3.6% and 3.6% higher than the Group 2 amplified data set. While improving the object detection effect, Group 3, using the proposed model, has a much shorter training time than Group 2's augmented dataset, saving 1.108h compared with Group 2. As demonstrated in Figure 13, The Precision and Recall curves of the three groups of experimental training processes are plotted. It can be clearly seen from the line chart that in these three groups of experiments, the curve of Group 3 rises the fastest, has the smallest oscillation and finally tends to be stable. The output result was the highest among the three groups. However, in the other two groups of curves, the rising rate is slow and the amplitude changes significantly, which persists until the end. In summary, according to the comparison results, the dataset optimized by the proposed deduplication algorithm has the best target detection effect, the most reasonable time consumption, and the highest efficiency.

C. COMPARISON ANALYSIS USING DIFFERENT MODELS
This study validates the improved performance of the model through thorough analysis and comparison with existing models. Different models were used for comparative analysis to demonstrate the advantages of the proposed CBAM-YOLOv8, such as YOLOv8, YOLOv5, MobileNet-SSD and Faster R-CNN. We used four evaluation metrics: precision, recall, mAP, and F1 score to evaluate the results. Under the same experimental conditions and according to the provided defects datasets, we conducted the same three groups of comparative experiments on object detection as described in Section V-B. The average defect detection results obtained from the different detection methods are presented in Table 4, with all comparative experiments conducted using the same training and testing datasets.

As shown in Table 3, the analysis of the results shows that the proposed model outperforms the other methods in improving the accuracy and efficiency of defect detection as much as feasible. Compared with the four groups of comparison tests of all methods, the optimization model part of this study has a notable increase in the detection results. Compared to the YOLOv8 algorithm, it significantly improves the detection accuracy while maintaining a minimal difference in the detection time. In contrast to the conventional YOLOv5

TABLE 3. Comparison of defects detection using different methods.

| Models	Groups	Precision (%)	Recall (%)	mAP (%)	F1 score (%)	Training Duration (h)
CBAM-YOLOv8	GROUP 1	83.4	80.8	84.5	82.1	0.569
	GROUP 2	85.5	84.0	87.2	84.7	1.996
	GROUP 3	88.8	87.8	90.8	88.3	0.953
YOLOv8	GROUP 1	82.6	75.5	80.6	78.9	0.527
	GROUP 2	84.1	81.3	85.5	82.7	1.767
	GROUP 3	87.9	84.3	88.6	86.1	0.864
YOLOv5	GROUP 1	75.5	72.7	74.8	74.1	0.489
	GROUP 2	83.4	81.5	84.9	82.4	2.094
	GROUP 3	86.6	83.3	86.3	84.9	1.019
MobieNet-SSD	GROUP 1	70.8	71.3	74.2	70.1	2.517
	GROUP 2	77.5	78.6	80.4	78.0	8.705
	GROUP 3	80.6	82.8	82.6	81.7	4.209
Faster RCNN	GROUP 1	65.8	66.7	67.0	66.2	3.133
	GROUP 2	73.1	71.8	72.8	72.4	10.688
	GROUP 3	76.9	78.0	78.1	77.4	5.162
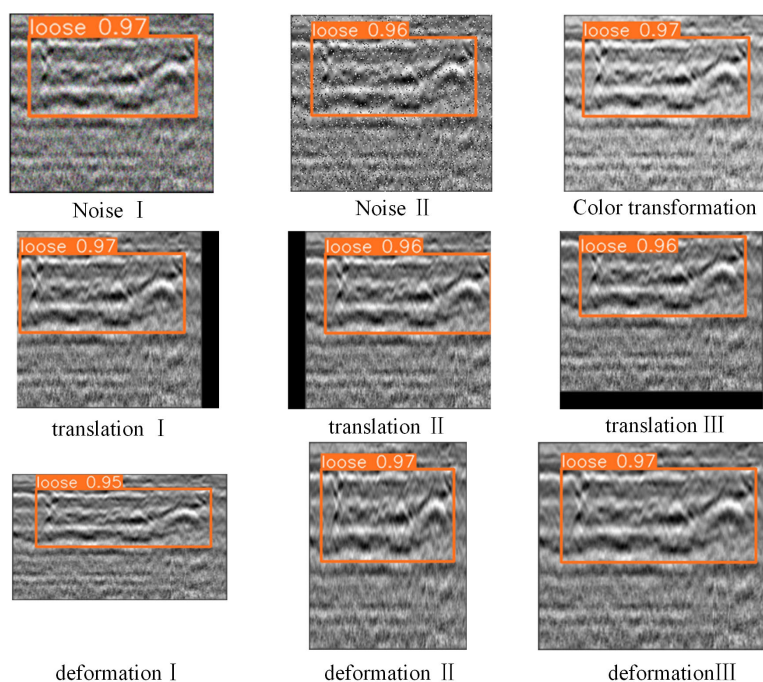
FIGURE 14. Analysis of model robustness.

algorithm, although the detection speed is marginally faster, the accuracy of detection is significantly lower than that of the proposed method. Some lightweight models, such as MobileNet-SSD, can reduce computation; however, the proposed method exhibits superior efficiency and accuracy. For traditional object detection methods, Faster RCNN requires more time but achieves a lower accuracy. The analysis of the results shows that, the detection results of the

CBAM-YOLOv8 model in this study are better. In addition, the detection rate is faster, which matches real-time needs. Specifically, the inference time (time to detect a single image) is substantially shorter than that of other approaches, exhibiting a high detection efficiency. In conclusion, our model differs from most current methods in that we significantly reduce the detection time while enhancing the detection results. By maximizing the balance between the detection

speed and accuracy, the proposed CBAM-YOLOv8 method outperforms existing methods.

D. ROBUSTNESS TESTING

To further highlight the robustness of the proposed method, we selected GPR loose images with the least obvious features and the lowest object detection effect as the data to verify the robustness of the method. We verified this in terms of noise, color transformation, translation, scaling, and deformation. As shown in Figure 14(a), we added Gaussian noise, salt-and-pepper noise, and brightness to the GPR image in this part, which has the capability to emulate the impact of various underground environments and noise. In figure 14(b), the image is shifted to the left, right and down, indicating that the location of the GPR image feature does not affect the recognition performance of the model. Figure 14(c) illustrates that the recognition performance of the model is not affected by the image scale deformation caused by horizontal and vertical stretching and scaling relative to the overall image size ratio. If the robustness test achieves good results for the image category with the least obvious features and the lowest detection effect, the robustness test results for the void and pipeline image data will also be good. In summary, it is evident that the proposed model exhibits remarkable stability and robustness. It can not only distinguish the features of GPR images, such as deformation, translation, and color transformation but also effectively adjust to specific noisy environments. Therefore, the object detection model developed in this study can help identify underground defect GPR images that are challenging for humans to distinguish because of their high levels of interference and blurriness.

VI. DISCUSSION

A. IMPORTANCE ANALYSIS OF THE MODEL FOR DETECING DEFECT

Currently, a vast range of data processing approaches has been developed for GPR object detection. However, the existing research still relies on manual work, and the challenges caused by the complex instabilities of underground scenes have not been fully addressed. DL-based GPR data analysis has not been able to keep pace with rapid advancements in other fields, mainly because of the inadequate availability of extensive radar databases. The precision of the evaluation results depends heavily on the quality of the data. Effectively obtaining high-quality datasets and detecting defect targets are of great significance for the development of GPR data analysis. This is shown in the proposed automated defect detection system. We can not only obtain high-quality amplified datasets, but also accurately detect underground defects. Moreover, the model can greatly reduce the amount of calculation and, save detection time, and the requirements for hardware and storage space are no longer harsh.

The integrated application of this method holds significant potential benefits, enabling efficient and accurate detection and localization of underground defects such as pipelines

---

Algorithm 1 SLIC-Phash

Input: Image Dataset G
Parameters: K = 218 – number of superpixels
S = 218 – hash size
T = 24 – the threshold of Hamming distance
Output: Deduplicated dataset G'
for img in G : // generate perceptual hash for each image
    superpixels = segment(img, K)
    hashes = []
    for Sp in superpixels
        hash = phash(sp, S)
        hashes.append(hash)
        avg_hash = mean(hashes)
        for hash in hashes :
            if hash >= avg_hash :
                hash = 1
            else:
                hash = 0
        end
    end
end
for img in G : // generate hash signature for each image
    Hashes = get_hashes(img)
    If scan_mode == 'zigzag':
        signature = zigzag(hashes)
        add signature to hash_table
end
for sig in hash_table : // delete images of low quality
    for other in hash_table :
        if distance(sig, other) <= T:
            remove other's img
            G' = get_unique_images(hash_table)
            Return G'
    end
end

---

and cables. It can assist urban planners in making informed decisions regarding the placement and layout of underground infrastructure, thereby enhancing the sustainability and effectiveness of urban development. Additionally, this method can be utilized for monitoring the health condition of underground structures, promptly identifying defects, and implementing appropriate maintenance measures to reduce safety risks and maintenance costs.

B. FUTURE WORK

In this study, an advanced image dataset optimization and detection system was proposed. However, despite its good performance on the current datasets, the model's generalization ability in new environments or under different geological conditions has not been fully tested. Second, further enhancement of real-time performance under extreme resource constraints might compromise accuracy. Finally, the current model primarily focuses on two-dimensional image features, which may not fully utilize the depth information available in GPR data. Considering these limitations, future research should explore cross-domain adaptation techniques to enable the model to adapt to different geological environments and detection conditions. Given the three-dimensional nature of GPR data, future models can incorporate deep

---

Algorithm 2 CBAM-YOLOv8

Input: Image img, number of classes nc
Parameters: nc = 3 – number of classes
img_size = 640 – input image size
batch_size = 16 – batch size for training
lr = 0.01 –initial learning rate
epochs = 300 –number of training epochs
Output: bboxes, class_ids, scores
for epoch in range(epochs) :
    for img_batch in DataLoader(batch_size) :
        Backbone // extract features from input image
            x_1 = conv(img, 64, kernel=3, stride=2) // first conv layer
            x_2 = Ghost_conv(x_1, 128, kernel=3, stride=2) // add Ghost
            x_3 = C2F(x_2) $*$ 3 // CSPDarknet53 to 2-Stage FPN module
            x_4 = Ghost_conv(x_3, 256, kernel=3, stride=2)
            x_5 = C2F(x_4) $*$ 6
            x_6 = Ghost_conv(x_5, 512, kernel=3, stride=2)
            x_7 = C2F(x_6) $*$ 6
            x_8 = Ghost_conv(x_7, 1024, kernel=3, stride=2)
            x_9 = C2F(x_8) $*$ 3
            x_{10} = SPPF(x_9) // spatial pyramid pooling
        Head // generate feature pyramids for detection
            p_3 = upsample(x_{10})
            p_3 = concat(p_3, x_6)
            p_3 = C2F(p_3) $*$ 3
            p_3 = CBAM(p_3) // channel attention
            p_4 = upsample(p_3)
            p_4 = concat(p_4, x_4)
            p_4 = C2F(p_4) $*$ 3
            p_4 = CBAM(p_4)
            p_5 = conv(p_4, 256, stride=2)
            p_5 = concat(p_5, p_3)
            p_5 = C2F(p_5) $*$ 3
            p_5 = CBAM(p_5)
            p_6 = conv(p_5, 512, stride=2)
            p_6 = concat(p_6, x_8)
            p_6 = C2F(p_6) $*$ 3
            p_6 = CBAM(p_6)
            bboxes, class_ids, scores = Detect([p_3,p_5,p_6], nc)
            Return bboxes, class_ids, scores
    end
end

---

learning with three-dimensional image processing techniques to make more comprehensive use of GPR data. Finally, research on GPR underground defect localization will be developed, and a complete underground defect assessment system will be established in accordance with existing standards.

VII. CONCLUSION

The key contribution of this study is to propose an underground defect detection based on GPR by Fusing Simple Linear Iterative Clustering Phash (SLIC-Phash) and Convolutional Block Attention Module (CBAM)-YOLOv8. It is used to optimize the underground defect image datasets and improve the detection efficiency. In short, the main novelties of this study are summarized as follows:

- This study introduces an optimized model for GPR defect images named SLIC-Phash. The model is based on traditional DA techniques that incorporate both SLIC and Phash algorithms. Through this integration, the model is capable of filtering out unrealistic target features and complex, ambiguous backgrounds in GPR image data with the aim of enhancing the quality of the dataset. This method has advantages such as simplicity, ease of integration, and minimal requirements for computer performance. Utilizing this optimized dataset not only significantly reduces the time required for target detection but also improves the accuracy of the detection results and reduces memory consumption.

- To achieve better target detection results, this study proposesd an object detection model called CBAM-YOLOv8. By incorporating the GhostConv network and CBAM attention mechanism, the YOLOv8 model was significantly improved in terms of detection accuracy, while the detection time remained nearly unchanged. The CBAM-YOLOv8 model presented in this paper achieves an average mAP of 90.8% and an F1 score of 88.3% in underground defect detection, which is superior to the target detection results of traditional models such as YOLOv8, YOLOv5, SSD, and Faster RCNN.

- A system that integrates dataset processing with object detection, termed the one-click GPR image detection system, was proposed. This system allows for the direct use of enhanced datasets to achieve optimal detection results. The system developed in this study significantly reduces the detection time and maximizes the balance between detection speed and accuracy. Empirical evidence demonstrates that the system is capable of simultaneously augmenting multiple target categories and detecting them precisely.

APPENDIX

See Algorithms 1 and 2.

REFERENCES

[1] E. Slob, M. Sato, and G. Olhoeft, "Surface and borehole ground-penetrating-radar developments," *Geophysics*, vol. 75, no. 5, pp. 75A103–75A120, Sep. 2010, doi: 10.1190/1.3480619.

[2] C. Caselle, S. Bonetto, C. Comina, and S. Stocco, "GPR surveys for the prevention of Karst risk in underground gypsum quarries," *Tunnelling Underground Space Technol.*, vol. 95, Jan. 2020, Art. no. 103137, doi: 10.1016/j.tust.2019.103137.

[3] J. M. W. Brownjohn, A. De Stefano, Y.-L. Xu, H. Wenzel, and A. E. Aktan, "Vibration-based monitoring of civil infrastructure: Challenges and successes," *J. Civil Struct. Health Monitor.*, vol. 1, nos. 3–4, pp. 79–95, Dec. 2011, doi: 10.1007/s13349-011-0009-5.

[4] U. Ozkaya, F. Melgani, M. B. Bejiga, L. Seyfi, and M. Donelli, "GPR B scan image analysis with deep learning methods," *Measurement*, vol. 165, Dec. 2020, Art. no. 107770, doi: 10.1016/j.measurement.2020.107770.

[5] D. Liu, L. Zhang, X. Jiang, C. Su, Y. Fan, and Y. Cao, "MEML: A deep data augmentation method by mean extrapolation in middle layers," *IEEE Access*, vol. 9, pp. 151621–151630, 2021, doi: 10.1109/ACCESS.2021.3125841.

[6] D. Yuan, M. Hong, and Z. An, "Visualizing GPR data using spatial-subband configuration," *IEEE Access*, vol. 6, pp. 54651–54659, 2018, doi: 10.1109/ACCESS.2018.2868797.

[7] J. Xie, F. Niu, W. Su, and Y. Huang, "Identifying coastal highway pavement anomalies using multiscale wavelet analysis in radar signal interpretation," *J. Civil Struct. Health Monitor.*, vol. 13, no. 1, pp. 49–65, Jan. 2023, doi: 10.1007/s13349-022-00595-z.

[8] P. Gamba and S. Lossani, "Neural detection of pipe signatures in ground penetrating radar images," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 2, pp. 790–797, Mar. 2000, doi: 10.1109/36.842008.

[9] M. R. Shaw, S. G. Millard, T. C. K. Molyneaux, M. J. Taylor, and J. H. Bungey, "Location of steel reinforcement in concrete using ground penetrating radar and neural networks," *NDT E Int.*, vol. 38, no. 3, pp. 203–212, Apr. 2005, doi: 10.1016/j.ndteint.2004.06.011.

[10] E. Pasolli, F. Melgani, and M. Donelli, "Automatic analysis of GPR images: A pattern-recognition approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 7, pp. 2206–2217, Jul. 2009, doi: 10.1109/TGRS.2009.2012701.

[11] C. Maas and J. Schmalzl, "Using pattern recognition to automatically localize reflection hyperbolas in data from ground penetrating radar," *Comput. Geosci.*, vol. 58, pp. 116–125, Aug. 2013, doi: 10.1016/j.cageo.2013.04.012.

[12] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017, doi: 10.1145/3065386.

[13] H. Liu, C. Lin, J. Cui, L. Fan, X. Xie, and B. F. Spencer, "Detection and localization of rebar in concrete by deep learning using ground penetrating radar," *Autom. Construct.*, vol. 118, Oct. 2020, Art. no. 103279, doi: 10.1016/j.autcon.2020.103279.

[14] L. E. Besaw and P. J. Stimac, "Deep convolutional neural networks for classifying GPR B-scans," *Proc. SPIE*, vol. 9454, May 2015, Art. no. 945413.

[15] M.-T. Pham and S. Lefèvre, "Buried object detection from B-scan ground penetrating radar data using faster-RCNN," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Valencia, Spain, Jul. 2018, pp. 6804–6807, doi: 10.1109/IGARSS.2018.8517683.

[16] J. Zhang, X. Yang, W. Li, S. Zhang, and Y. Jia, "Automatic detection of moisture damages in asphalt pavements from GPR data with deep CNN and IRS method," *Autom. Construction*, vol. 113, May 2020, Art. no. 103119, doi: 10.1016/j.autcon.2020.103119.

[17] X. Zhang, L. Han, M. Robinson, and A. Gallagher, "A GANs-based deep learning framework for automatic subsurface object recognition from ground penetrating radar data," *IEEE Access*, vol. 9, pp. 39009–39018, 2021, doi: 10.1109/ACCESS.2021.3064205.

[18] M. Li, M. Li, Q. Ren, H. Liu, and C. Liu, "Intelligent identification and classification of sewer pipeline network defects based on improved RegNetY network," *J. Civil Struct. Health Monitor.*, vol. 13, nos. 2–3, pp. 547–560, Mar. 2023, doi: 10.1007/s13349-022-00660-7.

[19] Y. Zheng, Y. Zhan, X. Huang, and G. Ji, "YOLOv5s FMG: An improved small target detection algorithm based on YOLOv5 in low visibility," *IEEE Access*, vol. 11, pp. 75782–75793, 2023, doi: 10.1109/ACCESS.2023.3297218.

[20] Y. Li, Z. Zhao, Y. Luo, and Z. Qiu, "Real-time pattern-recognition of GPR images with YOLO v3 implemented by tensorflow," *Sensors*, vol. 20, no. 22, p. 6476, Nov. 2020, doi: 10.3390/s20226476.

[21] S. Li, X. Gu, X. Xu, D. Xu, T. Zhang, Z. Liu, and Q. Dong, "Detection of concealed cracks from ground penetrating radar images based on deep learning algorithm," *Construct. Building Mater.*, vol. 273, Mar. 2021, Art. no. 121949, doi: 10.1016/j.conbuildmat.2020.121949.

[22] Z. Situ, S. Teng, X. Liao, G. Chen, and Q. Zhou, "Real-time sewer defect detection based on YOLO network, transfer learning, and channel pruning algorithm," *J. Civil Struct. Health Monitor.*, vol. 14, no. 1, pp. 41–57, Jan. 2024, doi: 10.1007/s13349-023-00681-w.

[23] N. Hu, J. Yang, X. Jin, and X. Fan, "Few-shot crack detection based on image processing and improved YOLOv5," *J. Civil Struct. Health Monitor.*, vol. 13, no. 1, pp. 165–180, Jan. 2023, doi: 10.1007/s13349-022-00632-x.

[24] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2022, pp. 7464–7475, doi: 10.1109/CVPR52729.2023.00721.

[25] V. Pham, D. Nguyen, and C. Donan, "Road damage detection and classification with YOLOv7," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Osaka, Japan, Dec. 2022, pp. 6416–6423.

[26] Y. Du, X. Liu, Y. Yi, and K. Wei, "Optimizing road safety: Advancements in lightweight YOLOv8 models and GhostC2f design for real-time distracted driving detection," *Sensors*, vol. 23, no. 21, p. 8844, Oct. 2023, doi: 10.3390/s23218844.

[27] Z. Tong, J. Gao, and D. Yuan, "Advances of deep learning applications in ground-penetrating radar: A survey," *Construct. Building Mater.*, vol. 258, Oct. 2020, Art. no. 120371, doi: 10.1016/j.conbuildmat.2020.120371.

[28] Z. Zong, C. Chen, X. Mi, W. Sun, Y. Song, J. Li, Z. Dong, R. Huang, and B. Yang, "A deep learning approach for urban underground objects detection from vehicle-borne ground penetrating radar data in real-time," *Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. XLII-2/W16, pp. 293–299, Sep. 2019, doi: 10.5194/isprs-archives-XLII-2-W16-293-2019.

[29] I. Giannakis, A. Giannopoulos, and C. Warren, "A machine learning scheme for estimating the diameter of reinforcing bars using ground penetrating radar," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 3, pp. 461–465, Mar. 2021, doi: 10.1109/LGRS.2020.2977505.

[30] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proc. 27th Int. Conf. Neural Inf. Process. Syst.*, vol. 2, Montreal, QC, Canada, 2014, pp. 2672–2680.

[31] G. Chen, X. Bai, G. Wang, L. Wang, X. Luo, M. Ji, P. Feng, and Y. Zhang, "Subsurface voids detection from limited ground penetrating radar data using generative adversarial network and YOLOV5," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. IGARSS*, Jul. 2021, pp. 8600–8603.

[32] Y. Yue, H. Liu, X. Meng, Y. Li, and Y. Du, "Generation of high-precision ground penetrating radar images using improved least square generative adversarial networks," *Remote Sens.*, vol. 13, no. 22, p. 4590, Nov. 2021, doi: 10.3390/rs13224590.

[33] D. Zhao, G. Guo, Z.-K. Ni, J. Pan, K. Yan, and G. Fang, "WAEGAN: A GANs-based data augmentation method for GPR data," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, pp. 1–5, 2023, doi: 10.1109/LGRS.2023.3323981.

[34] H. Xiong, J. Li, Z. Li, and Z. Zhang, "GPR-GAN: A ground-penetrating radar data generative adversarial network," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, pp. 1–14, 2024, doi: 10.1109/TGRS.2023.3337172.

[35] G. Yue, C. Liu, Y. Li, Y. Du, and S. Guo, "GPR data augmentation methods by incorporating domain knowledge," *Appl. Sci.*, vol. 12, no. 21, p. 10896, Oct. 2022.

[36] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274–2282, Nov. 2012, doi: 10.1109/TPAMI.2012.120.

[37] C. L. Zitnick and S. B. Kang, "Stereo for image-based rendering using image over-segmentation," *Int. J. Comput. Vis.*, vol. 75, no. 1, pp. 49–65, Jul. 2007, doi: 10.1007/s11263-006-0018-8.

[38] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6517–6525.

[39] S. Woo, J. Park, J. Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 3–19.

[40] C. Liu, H. Sui, J. Wang, Z. Ni, and L. Ge, "Real-time ground-level building damage detection based on lightweight and accurate YOLOv5 using terrestrial images," *Remote Sens.*, vol. 14, no. 12, p. 2763, Jun. 2022, doi: 10.3390/rs14122763.

[41] J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, "Squeeze-and-excitation networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 8, pp. 2011–2023, Aug. 2020, doi: 10.1109/TPAMI.2019.2913372.

[42] M. Lin, Q. Chen, and S. Yan, "Network in network," 2014, 10.48550/arXiv.1312.4400.

[43] H. Wang, Y. Fan, Z. Wang, L. Jiao, and B. Schiele, "Parameter-free spatial attention network for person re-identification," 2018, arXiv:1811.12150.

[44] F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong, and Q. He, "A comprehensive survey on transfer learning," *Proc. IEEE*, vol. 109, no. 1, pp. 43–76, Jul. 2020, doi: 10.1109/JPROC.2020.3004555.

[45] Y. Zhang and K. Yuen, "Crack detection using fusion features-based broad learning system and image processing," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 36, no. 12, pp. 1568–1584, Dec. 2021, doi: 10.1111/mice.12753.

[46] M. Martínez-Rojas, N. Marín, and M. A. Vila, "An approach for the automatic classification of work descriptions in construction projects," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 30, no. 12, pp. 919–934, Dec. 2015, doi: 10.1111/mice.12179.

NIANNIAN WANG was born in Henan, China, in 1989. She received the Ph.D. degree in structural engineering from the Dalian University of Technology, China, in 2019.

She is currently a Professor with the School of Water Resources and Civil Engineering, Zhengzhou University, China. She also serves as the Vice Chairperson of the Pipeline Inspection and Rehabilitation Committee of the China Municipal Engineering Association and a member of the Expert Committee of the Sino-U.S. Joint Trenchless Engineering Research Center. She has led more than ten research projects, including National Key Research and Development Program, National Natural Science Youth Fund, China Postdoctoral Special Fund, and China Postdoctoral Fund. Her main research interests include intelligent inspection of engineering structures, damage identification and assessment, and big data analysis.

Prof. Wang research results were awarded the Gold Prize of the First National Postdoctoral Innovation and Entrepreneurship Competition, the Second Prize of Scientific and Technological Progress in Hubei Province and Henan Province, and the ''Young Star'' of China International Trenchless Technology Symposium. She has published one monograph and 32 academic articles (25 SCI/EI articles). She received the 2019 ASCE Best Paper Award (the only award-winning paper from a university in mainland China).

HAOBANG HU was born in Hennan, China, in 1995. He received the B.S. degree in engineering from Zhengzhou University, Zhengzhou, China, in 2019, where he is currently pursuing the M.S. degree in engineering.

His research interests include the non-destructive detection based on ground penetrating radar and image recognition based on deep learning.

BIN LI was born in Henan, China, in 1993. He received the Ph.D. degree from the Dalian University of Technology.

His current research interests include safe operation and maintenance of underground water supply and drainage pipes. He is currently a member of the Chinese Society of Rock Mechanics and Engineering.

ZEXI ZHANG was born in Jiangxi, China, in 2000. He received the Graduate degree from the School of Civil Engineering, East China Jiaotong University, in 2017. He is currently pursuing the master's degree with the School of Water Resources and Civil Engineering, Zhengzhou University.

His research interest includes deep learning based on ground penetrating radar. He is mainly engaged in non-destructive detection of underground diseases.

JIANWEI LEI was born in Inner Mongolia, China, in 1993. He received the Ph.D. degree in safety and protection engineering from Zhengzhou University, Zhengzhou, China.

His research interests include non-destructive detection technique in engineering and numerical simulation methods (e.g., finite-difference time-domain method and symplectic methods).

• • •