Int. J. Simul. Multidisci. Des. Optim. **15**, 7 (2024)

https://doi.org/10.1051/smdo/2023023

IJSMDO
Available online at:
https://www.ijsmdo.org

Modelling and Optimization of Complex Systems with Advanced Computational Techniques

RESEARCH ARTICLE OPEN ACCESS

Investigation on road underground defect classification and localization based on ground penetrating radar and Swin transformer

Jinke An[1], Li Yang[2], Zhongyu Hao[2], Gongfa Chen[1,*], and Longjian Li[1]

[1] School of Civil and Transportation Engineering, Guangdong University of Technology, Guangzhou, 510006, China
[2] JSTI Group Guizhou Engineering Survey and Design Co., Ltd, 510800, China

Received: 19 October 2023 / Accepted: 8 December 2023

Abstract. In response to the low detection efficiency and accuracy of traditional manual methods for detecting road underground defects, this paper proposes an intelligent detection method based on ground penetrating radar (GPR). This method integrates the detection, classification, and localization of road underground defects. The approach uses Swin Transformer as a feature extraction network and utilizes the YOLOX object detection algorithm as a road underground defect detection model. It enables the detection of defect regions in three types of defect images: voids, non-compact areas, and underground pipelines. In addition, the collected radar signals are processed by Fourier transformation to obtain time-domain spectra and frequency-domain spectra, which are further analyzed to generate signal classification data set to achieve the defect classification. Finally, based on the relative positional relationship between the detected defect images and the GPS information collected by the GPR, the real positions of the defects on the map are automatically determined using the APIs provided by Amap (AutoNavi map). Experimental results show that this method achieves a precision and recall rate of 94.2% and 99.1%, respectively, for the detection of road underground defects, with an average precision of 94% and an average classification accuracy of 90%. The method significantly improves the accuracy and speed of road underground defect detection while meeting engineering requirements, making it highly valuable for practical road underground defect detection tasks.

Keywords: Road underground defect / group penetrating radar / Swin transformer object detection algorithm / Fourier transform / GPS intelligent localization

1 Introduction

In recent years, there have been a number of accidents involving road collapses throughout the country, including on highways, urban roads, and rural roads, among others. According to China's National Geological Hazard Bulletin, from January 2019 to January 2022, there were over 717 road collapse accidents nationwide, resulting in a total of 108 deaths and 174 injuries. The rate of road collapse accidents has been increasing at an annual rate of 80.93%. These incidents pose a significant safety hazard to cities, causing substantial economic losses, and posing a serious threat to public property safety, as shown in Figure 1. Common causes of road collapse include underground defects such as voids, non-compact areas, and damaged underground pipelines [1]. For instance, aging and ruptured underground drainage pipes can be affected by rainwater, wastewater, and groundwater, leading to erosion and soil erosion, and ultimately to ground collapse [2,3]. To prevent further similar accidents, it is necessary to regularly detect and monitor the underground conditions of roads, and to promptly prevent and repair potential underground defects. Common road detection methods include GPR, multi-channel surface wave methods, high-density resistivity methods, and seismic imaging methods [4]. GPR has developed rapidly in recent years due to its convenience, wide applicability, high detection efficiency, and intuitive results. It has become a primary means of detecting road underground defects. Xu et al. [5] successfully detected underground voids using GPR and analyzed the causes of void formation. In summary, the application of GPR in road detection is becoming increasingly frequent. Besides detecting road surface cracks, GPR can also detect abnormalities such as internal road reinforcement, underground voids, and pipelines.

* e-mail: gongfa.chen@gdut.edu.cn

(a) Road collapse in Jiangsu (b) Road collapse in Lanzhou

Fig. 1. Road collapse accidents.

Most isolated targets and underground pipelines often exhibit hyperbolic features in radar images. In the early stages, methods such as Hough transform, Viola-Jones algorithm, edge detection, and thresholding were commonly used to process radar data and extract hyperbolic features. These methods still relied on technicians to identify the extracted features, maintaining a detection accuracy of around 0.758 and remaining highly dependent on human subjectivity [6]. With the continuous development of computer vision technology, researchers have applied machine learning methods to the study of hyperbolic feature extraction and defect identification and classification [7]. Lei et al. [8] combined the framework of Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) neural networks to identify the position of reinforcing bars in concrete and calculate their diameter. Tong et al. [9] developed multi-level CNN and cascaded CNN models for the automatic classification and identification of roadbed defects (uneven settlement, potholes, cracks), achieving classification accuracies of over 90% for both CNN structures. Machine learning, in addition to its application in road engineering, can also be applied in various other fields. Teng et al. [10] generated a large number of random bridge structure models, used CNN to extract damage features of bridges, and applied CNN to damage detection of randomly generated models. The results show that using acceleration signals as CNN input gives excellent detection results, with an accuracy of up to 99.4%, overcoming the limitations associated with individual structures. J. Kers et al. [11] used artificial neural networks and a hybrid genetic algorithm to establish a relationship model between filler density and polymethylmethacrylate powder content. They designed a new composite from recycled glass fiber reinforced plastic. These cases illustrate the powerful generalization capabilities of machine learning and showcase outstanding practical applications in diverse fields.

In addition to the recognition and classification of radar defect images, researchers have also studied defect localization and intelligent detection processes [12]. Zhang et al. [13] developed the Incremental Random Sampling (IRS) method for radar image preprocessing and established a hybrid CNN model using ResNet-50 and YOLOX frameworks for road moisture detection. The results showed that this method not only achieved automation of image sample selection but also identified and located moisture-associated damage with an accuracy of 92%. Inspired by transfer learning, Lei [14] trained the Single Shot Multibox Detector (SSD) model using collected samples of reinforcing bars and underground pipelines. By testing the model's performance, the target detection accuracy reached 91%.

At present, GPR has been widely used in road inspections, but it still faces significant challenges due to its reliance on manual interpretation and analysis of radar data. It has been observed that the use of deep learning and object detection algorithms yields significant results in classifying and detecting targets with distinct hyperbolic features such as underground pipelines and reinforcing bars [8], or clearly discernible dielectric constants of underground water. However, there is limited research into the features of other road underground defects, such as voids and non-compact areas, and the use of deep learning and object detection algorithms fails to meet the requirements of classifying and detecting multiple types of defect targets. For road defect localization, traditional methods mostly rely on neural network approaches [15], which are trained using collected defect images. However, these methods suffer from long detection times, low localization accuracy, and lacking the capability to perform real-time localization of road underground defects using GPS. This method does not meet the practical engineering requirements for road defect prevention. Therefore, it is imperative to investigate the characteristic patterns of road underground defects reflected in GPR signals and images, and to develop an efficient, intelligent, and applicable defect analysis system. This system aims to achieve intelligent detection of road underground defects, reduce reliance on experiential judgment, improve defect detection accuracy and efficiency, and reduce time and economic costs.

This paper addresses the detection requirements for road underground defects and analyzes and summarizes the issues related to object detection and defect features. Focusing on three types of defects, namely voids, non-compact areas, and underground pipelines, the following researches were conducted:

– Investigate the characteristic patterns of different road underground defects in images and signals based on the radar data collected by GPR.
– Investigate defect localization methods based on object detection, with improvements being made to the Swin Transformer algorithm from the perspective of feature extraction networks and detection networks. The model

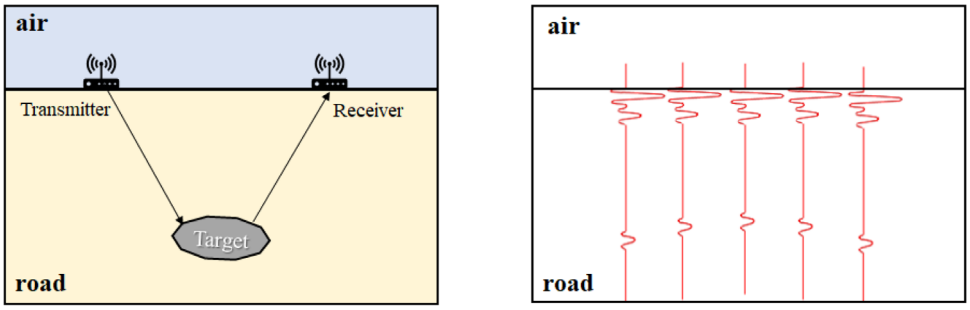


Fig. 2. The principle of ground penetrate radar.

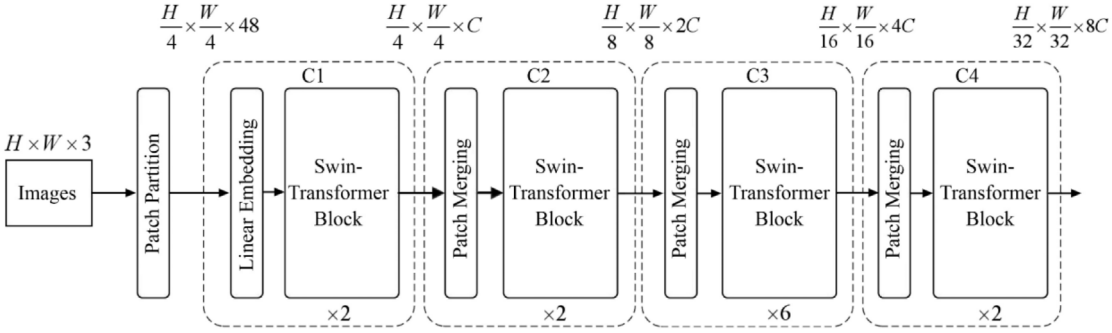


Fig. 3. The architecture of a Swin transformer.

will be evaluated using indicators such as recall, precision, and average precision, resulting in the optimal object detection algorithm suitable for detecting road underground defects in GPR images.
– Investigate the classification and recognition of radar signals using deep learning, obtaining the optimal classification and recognition network by evaluating the performance under different network parameters.
– Investigate the relationship between measurement points and defect points based on the GPS information obtained from GPR, thereby obtaining the actual positions of the defect points.

2 Methods

2.1 Working principles of ground penetating radar

When GPR operates, it emits electromagnetic waves downwards from the transmitting antenna. As the electromagnetic waves propagate underground, they encounter anomalies that cause them to reflect back. The reflected electromagnetic waves are received by its receiving antenna. The waveform diagram composed of the received echo signals from the GPR is shown in Figure 2a, with the characteristics in Figure 2b reflecting the detection of road underground defects [16].

2.2 Proposed deep learning approach

In this experiment, building on the existing YOLOX and Mask R-CNN models for road underground defect detection, the backbone networks of both models were replaced by the Swin Transformer and attention mechanisms were introduced to extract road underground defect features. This approach resulted in the optimal detection network model suitable for road underground defect detection.

2.2.1 Swin transformer

After the breakthrough achieved by the Transformer model in natural language processing (NLP) tasks [17], Transformer has gradually become the popular architecture model. Consequently, it has been applied to the field of computer vision. Prior to the Swin Transformer model [18] employed, Vision in Transformer (ViT) [19] was the first to apply Transformer to image classification. However, ViT did not consider the inherent characteristics of visual signals, making it less suitable for region-level or pixel-level tasks such as object detection and semantic segmentation. Therefore, the Swin Transformer algorithm was proposed for use in object detection and semantic segmentation tasks.

The Swin Transformer network architecture employs the Windows Multi-Head Self-Attention (W-MSA) and Windows Shifted Multi-Head Self-Attention (WS-MSA) mechanisms to achieve a hierarchical Transformer. This enables the extraction of multi-scale features from images similar to CNNs, thereby serving as a backbone network for visual tasks such as object detection and image segmentation.

Figure 3 illustrates the network architecture of the Swin Transformer. The model is constructed with 4 stages using the block configuration [2,2,6,2]. For an input image (H, W, 3), it is initially divided into patches of size 4×4, where each patch has a dimension of $4 \times 4 \times 3 = 48$.

Therefore, the input image dimension becomes (H/4, W/4, 48). The image is then passed through a Linear Embedding layer, its dimensions are transformed to (H/4, W/4, 96). The Swin Transformer Block with self-attention calculations is then applied to these patches, forming "Stage 1". As the network progresses, each Patch Merging layer divides the input feature map into 2×2 patches, reducing the resolution by a factor of 2, similar to pooling operations in CNNs. The transformed features are then passed through Swin Transformer Blocks, maintaining a resolution of (H/8, W/8), which is referred to as "Stage 2". This process is repeated twice to form "Stage 3" and "Stage 4", resulting in output resolutions of (H/16, W/16) and (H/32, W/32), respectively.

In the Swin Transformer model, the Swin Transformer Block consists of a window-based multi-head attention mechanism and a shift window-based multi-head attention mechanism, as shown in Figure 4. The left module in Figure 4 represents the window-based multi-head attention calculation, which includes the Layer Norm layer, W-MSA unit, Multi-Layer Perceptron (MLP) layer, and the skip connections. The right module represents the shift window-based multi-head self-attention calculation, which consists of the Layer Norm layer, SW-MSA unit, the MLP layer, and the skip connections. The Swin Transformer Block only extracts image features without changing the dimension of the feature maps.

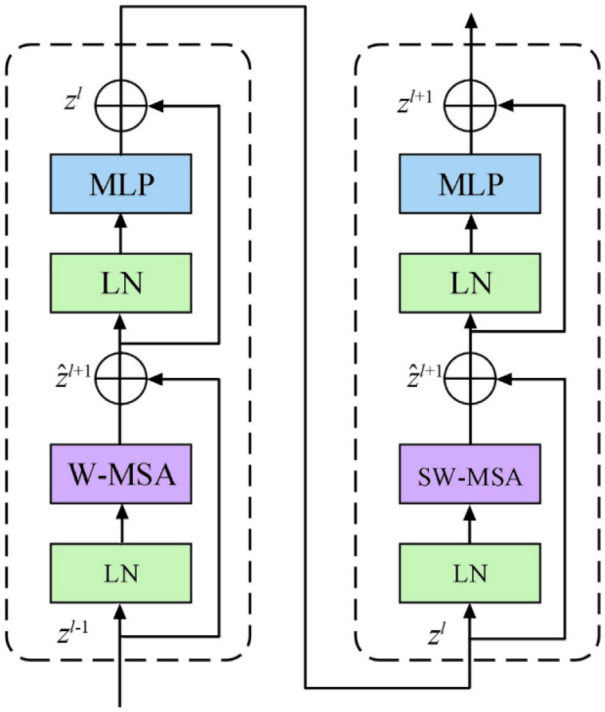


Fig. 4. Two successive Swin transformer blocks.

2.2.2 YOLOX

The YOLO series has evolved with the development of object detection technology. YOLOX [20], based on YOLOv5, introduces strategies such as decoupled head, anchor-free and Simplified Optimal Transport Assignment (SimOTA) to improve the detection accuracy of the network model. The structure of the YOLOX algorithm can be divided into three main parts: Backbone, Neck and Head, as shown in Figure 5.

The backbone network of YOLOX adopts CSPDarknet, which is based on ResNet and consists of modules such as Focus, Conv, CSPLayer, and SPPBottleneck. The Focus module slices the input image horizontally and vertically, taking every second pixel in the sliced images and concatenating them into four independent feature layers. This concatenation increases the channel depth fourfold while reducing the dimensions by half. Conv is the fundamental convolution unit in YOLOX, consisting of Conv2D functions, Batch Normalization (BN) and the SiLU activation function, which performs 2D convolution, normalization, and activation operations sequentially. CSPLayer utilizes the CSPnet structure, which splits residual blocks into two parts: one as a backbone for stacking residual blocks, and the other as residual edges. After minimal processing, these two parts are concatenated together. This approach exploits the rich gradient information and mitigates the gradient vanishing problem associated with increasing network depth. The SPPBottleneck module is responsible for performing convolution, pooling, and concatenation operations on the input. Within this module, basic convolution units (Conv) are used for convolution, and different-sized max-pooling kernels are used for feature extraction in the pooling layer.

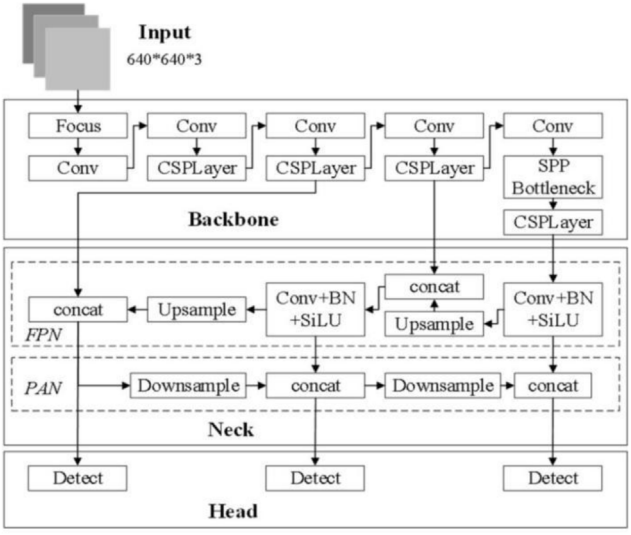


Fig. 5. Structure of YOLOX.

The Neck consists of the Feature Pyramid Networks (FPN) [21] and the Path Aggregation Network (PAN). The FPN is used to construct high-level semantic feature maps, conveying strong semantic features from top to bottom. PAN enhances the feature hierarchy through a bottom-up approach, utilizing precise low-level location signals to strengthen the entire feature hierarchy, thereby shortening the information path between lower-level and higher-level features. The combination of FPN and PAN facilitates the fusion of low-level and high-level features, as shown in Figure 6.

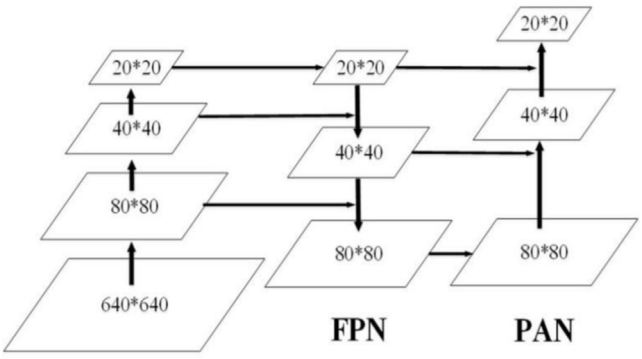


Fig. 6. Structures of FPN and PAN.

The Head is the detection structure of YOLOX, which uses three enhanced features obtained from the Neck to predict large, medium, and small targets respectively. Unlike previous YOLO models that used a Coupled Head structure, where classification and regression were integrated into a 1×1 convolution, YOLOX employs a Decoupled Head for classification and regression. It begins with a 1×1 convolution to reduce the dimensionality of the output. The classification and regression parts each use two 3×3 convolutions to improve prediction accuracy and accelerate network convergence.

2.2.3 Mask R-CNN

Mask R-CNN [22] is an object detection framework based on Faster R-CNN, which is capable of simultaneously detecting objects and generating semantic segmentation masks for these objects. As shown in Figure 7, the structure of Mask R-CNN can be divided into three parts: the Backbone Network, the Region Proposal Network (RPN), and the Fully Convolutional Network (FCN).

In the Mask R-CNN architecture, the input image is first passed through the backbone structure, ResNet+FPN, which generates a series of feature maps at different scales. These feature maps are then fed into the RPN, which generates a set of candidate proposal boxes. These candidate boxes are then aligned using Region of Interest (RoI) Align, allowing each candidate box to correspond to a specific region on the feature map. The aligned feature maps fed into fully connected layers to obtain class labels and bounding box coordinates for each candidate box.

At the same time, the output of RoI Align is also fed into the FCN to obtain object segmentation masks. The FCN network performs convolution and upsampling operations on the aligned feature maps, generating segmentation masks of the same size as the candidate boxes. Once the segmentation masks have been generated, operations on the masks provide specific shape and position information about the objects. This information can be used to refine the predictions for object classification and bounding boxes. Finally, each candidate box is assigned a class label, bounding box coordinates, and a segmentation mask.

2.3 Obtaining GPS information of defect points

2.3.1 Coordinates of defect points and measurement points

The processed images are then fed into the Swin Transformer network for detection to obtain the bounding box information $[x, y, h, w]$ of the defects, where x and y represent the coordinates of the top-left corner of the bounding box relative to the top-left corner of the image, and h and w represent the height and width of the bounding box, respectively. At this point, the X-coordinate of the defect center point D in the image is calculated as shown in Figure 8:

$$D_x = x + \frac{w}{2}. \tag{1}$$

Using MATLAB to read the pixel value information of the processed image, it can be observed that the recorded measurement points are represented as 5×5 matrices with pixel values of 255 (The RGB channels of the measurement points are $255 \times 255 \times 255$). Therefore, the center point coordinates of the 5×5-regions in the image, where all pixels have a value of 255, correspond to the X-axis coordinates of the measurement points. Using MATLAB functions such as 'bwboundaries' and 'regionprops', the center point coordinates of all 5×5 regions with pixel values of 225 can be obtained, as shown in Figure 9. This results in an array of X-axis coordinates for all measurement points, denoted as $[T_1, \ldots, T_n, \ldots, T_{2n}]$. Additionally, assuming the different value between the X-axis coordinate D_x of the defect center point and the array of X-axis coordinates for all measurement points $[T_1, \ldots, T_n, \ldots, T_{2n}]$ is represented as the array $[f_1, \ldots, f_n, \ldots, f_{2n}]$.

2.3.2 GPS coordinates of defect points

By subtracting the X-axis coordinate D_x of the defect center point from the array of X-axis coordinates for all measurement points $[T_1, \ldots, T_n, \ldots, T_{2n}]$, we obtain the array $[f_1, \ldots, f_n, \ldots, f_{2n}]$. The absolute values of $f_1, \ldots, f_n, \ldots, f_{2n}$ can be used to determine the measurement point T that is closest to the defect point D, and the positive and negative signs of f indicate the relative position of the defect point D with respect to the measurement point T.

Each measurement point has corresponding GPS information, and based on the concept of linear interpolation, the GPS information of the defect point can be determined and shown in Table 1. The solution formula is shown in equation (2).

$$GPS_D = \frac{D_x - T_{n+1}}{T_n - T_{n+1}} \times GPS_T + \frac{D_x - T_n}{T_{n+1} - T_n} \times GPS_{T+1}. \tag{2}$$

2.4 Analysis of road underground defect classification

2.4.1 Radar grayscale image features

By analyzing the processed grayscale image, the characteristics of different defects in the radar image can be identified. The shape and arrangement of the stripes in the

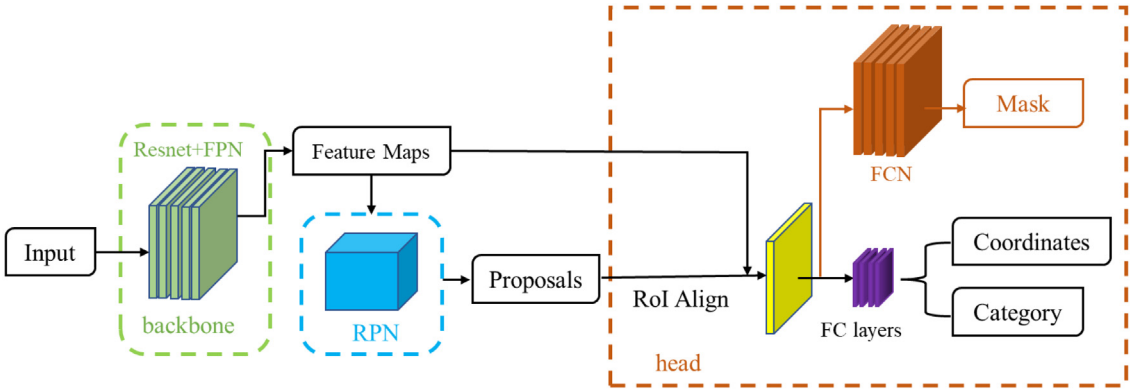


Fig. 7. Structure of Mask R-CNN.

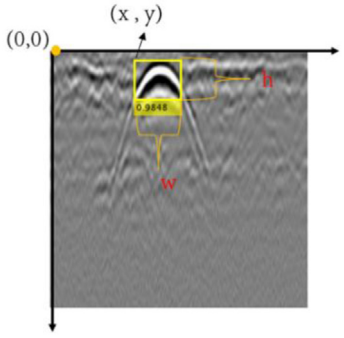


Fig. 8. Defect detection box.

radar image can effectively differentiate between different types of defects, providing a basis for further defect analysis.

- Healthy road: In the image of a healthy road, the lower part is uniformly gray without any prominent black or white blocks. There are no significant reflections of abnormal electromagnetic waves or weak signal amplitudes, as shown in Figure 10.
- Void defect: Void defects appear as alternating black and white blocks in the lower part of the radar image. The black and white blocks are relatively thin and have a relatively uniform shape, often forming a pattern of multiple hyperbolic curves or straight lines, as shown in Figure 11.
- Non-compact defect: Non-compact defects appear as alternating black and white blocks in the lower part of the radar image. The black and white blocks are relatively thin and the blocks are disordered and intermingled, forming multiple wave-like segments that are not neatly arranged. However, it is difficult to discern the compactness of the defect from the image alone, as shown in Figure 12.
- Underground pipelines: In Figure 13, underground pipelines appear in the radar image as alternating black and white blocks in the lower part. The black and white blocks are relatively thin. For PVC pipes with diameters of 160 mm and 50 mm, the black and white blocks have a relatively uniform shape, forming two parallel hyperbolic curves with some distance between them. In the case of a 15 mm diameter steel pipe, only one hyperbolic curve is observed due to the inability of the electromagnetic waves to penetrate the metal and to be completely reflected at the surface of the steel pipe.

2.4.2 Time-domain and frequency-domain characteristics of radar reflection waves

Although the use of the object detection network alone meets the requirements for speed and accuracy in detecting defect targets in the radar grayscale image, there is still a reliance on the accuracy of human interpretation for defect classification. If the human interpretation is not accurate, the corresponding object detection network will also have inaccuracies in determining the defect types. Therefore, more intuitive and stable features need to be obtained for each defect category.

Figure 14 shows the radar signal reflections of the normal underground and three types of road underground defects. The first waveform (highlighted by a green circle) represents the direct wave propagating through the surface, while the subsequent waveform represents the radar reflection wave (highlighted by a black circle). Since the direct wave is the same for all image types, we can ignore the direct wave and focus only on the reflection wave to more intuitively differentiate the signal characteristics of each image category. It can be observed that when there is no defect in the underground, the signal remains in a stable state. However, when abnormal defects are present in the underground, the signal reappears with enhanced amplitude at the locations of the abnormalities. This indicates that the amplitude of the radar reflection wave can distinguish whether there are abnormal defects in the underground.

From Figure 14, it is evident that the void signal exhibits distinct and dense amplitudes at the defect locations. The underground pipeline signal shows a regular waveform at the defect locations. Since the differences between the non-compact signal and the normal signal are not clearly manifested in the time domain, it is necessary to transform the time-domain signals into frequency-domain signals using the Fast Fourier Transform (FFT), as shown in Figure 15. It can be observed that the normal signal appears smoother in the frequency domain, while the non-compact defect signal appears slightly rougher in comparison.

By combining the time-domain and frequency-domain approaches, the collected signals from the GPR can be classified into the categories of voids, non-compact and underground pipelines, and corresponding datasets can be generated. Finally, these time-domain signal datasets are

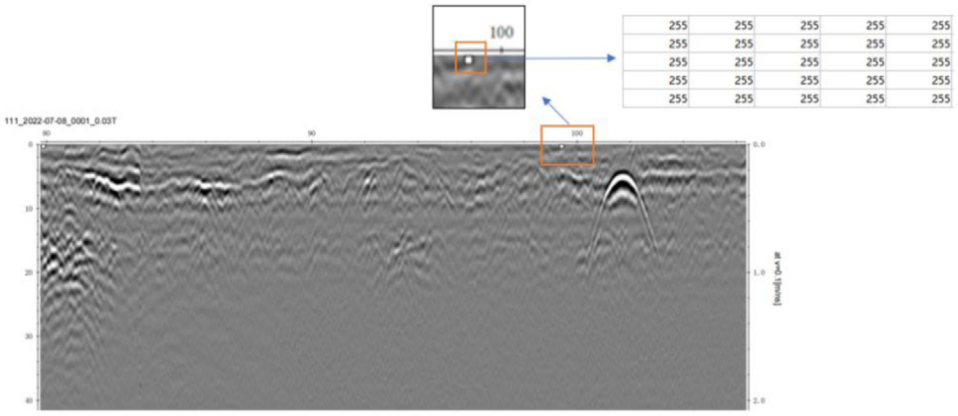


Fig. 9. Measurement point information.

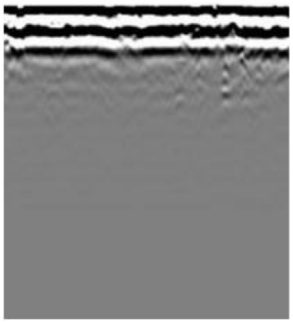


Fig. 10. Radar image of a healthy road.

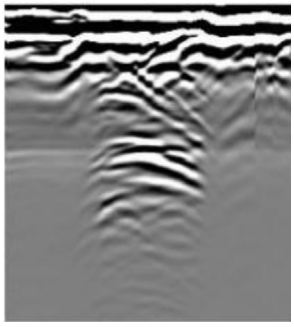


Fig. 11. Radar image of void defect.

fed into a transfer learning-based classification network model for training, resulting in a more accurate defect classification network and obtaining the classification accuracy for each type of defect.

The discrete FFT algorithm is used to transform the dynamic signals from the time-domain to the frequency-domain, decomposing the temporal waveform of the signal into a frequency spectrum to obtain more intuitive features. For a signal $x(t)$, its Fourier transform $X(\omega)$ can be calculated using equation (3):

$$X(\omega) = \int_{-\infty}^{+\infty} x(t) e^{-j\omega t} dt. \tag{3}$$

3 Road underground defects detection experiment

3.1 Data collection

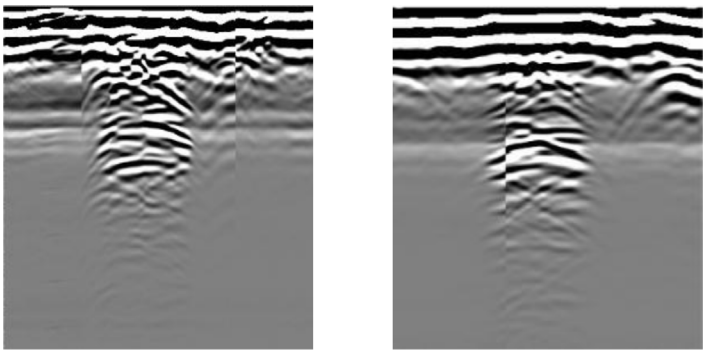
The data acquisition system consists of two parts: hardware and software. The hardware part utilizes the Impulse Radar CO1760 (ImpulseRadar, Gothenburg, Sweden), a real-time sampling impulse radar, which includes the radar host, power supply, ranging device and acquisition equipment. The data acquisition software part uses ViewPoint (Impulse Radar). The hardware components of the GPR device are shown in Figure 16 and their related parameters are shown in Table 2.

Table 1. Coordinate-to-parameter information.

Image coordinates	*GPS* coordinates
T_n	GPS_T
D_x	GPS_d
T_{n+1}	GPS_{T+1}

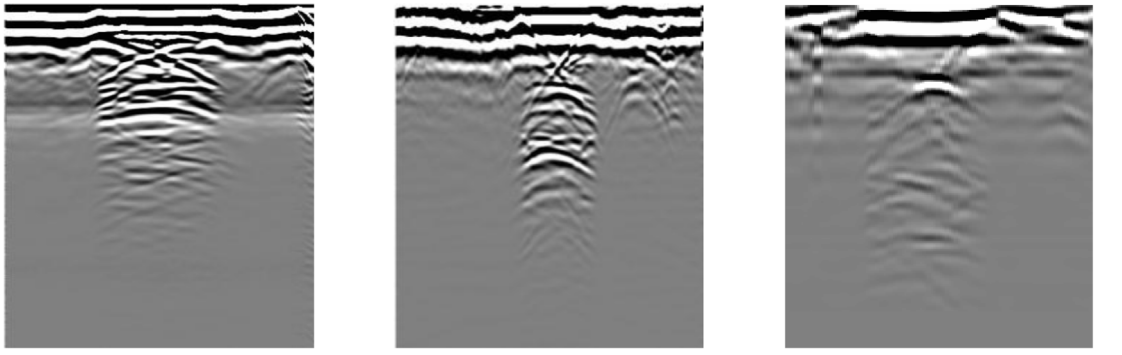
Before starting the detection process, it is necessary to lay out the measurement lines and place measurement point markers along the lines. During the detection process, the radar is pulled along the established measurement lines, ensuring that the ranging wheel rotates to trigger data acquisition. When a measurement point marker is passed over, it is recorded in the ViewPoint (Impulse Radar) data acquisition software to ensure that the recorded markers in ViewPoint match the markers placed in the field, as shown in Figure 17 and Figure 18.

The data collected by the GPR are mainly stored in the following file formats: mrk, cor, iprb, and iprh. The mrk and cor files are used to record the GPS information of the measurement points and the measurement lines, making it easier to determine the real position of anomalies on a map based on the relative positional relationship between the anomalies and the measurement points. The IR_BatchConv (Zond Software, Lausanne, Switzerland) software is used to



(a) Non-compact (b) Compaction 50%

Fig. 12. Radar image of porous defect.



(a) PVC pipe with diameter of 160mm (b) PVC pipe with diameter of 50mm (c) steel pipes with diameter of 15mm

Fig. 13. Radar image of underground pipelines defect.

convert the iprb and iprh format files collected by the radar device into rd format files, by adjusting the relevant parameters, the radar signal waveform can be transformed into a grayscale image [23], which is used for subsequent defect detection and analysis. After organizing and summarizing the data, some of the results are shown in Table 3.

3.2 Sample augmentation

The training performance of neural networks heavily relies on the quantity and quality of the data, as insufficient training data can easily lead to overfitting. Overfitting refers to a situation where the model can accurately predict the training set results but performs poorly on unknown samples in the validation and test sets, indicating a lack of generalization ability of the network.

Due to the limited amount of road underground defect image data and significant variation in the number of each type of defect, it is necessary and effective to employ data augmentation techniques to achieve the desired training performance and prevent overfitting. Common data enhancement techniques include geometric transformations and color space enhancements.

Geometric transformations involve translating, flipping, rotating, and scaling the images. These transformations do not alter the pixel values of the images, but rearrange the pixels on the image plane, ensuring reshaping of the image without losing pixel information.

Color space enhancement refers to the representation of grayscale or RGB images using other methods, such as HSV (Hue, Saturation, Value). For road underground defect images, controlling the hue, brightness, and saturation of the image within the range of 0 to 1, and contrast within the range of 1 to 3, can preserve image features to the maximum extent while expanding the sample set. Figure 19 demonstrates the effects of data enhancement on the same image.

3.3 Road underground defect detection based on Swin transformer

3.3.1 YOLOX model based on Swin transformer

Due to the hierarchical Transformer modules of the Swin Transformer model, which can extract features at different scales, and the windowing strategy of the Swin Transformer, which is advantageous for capturing rich contextual

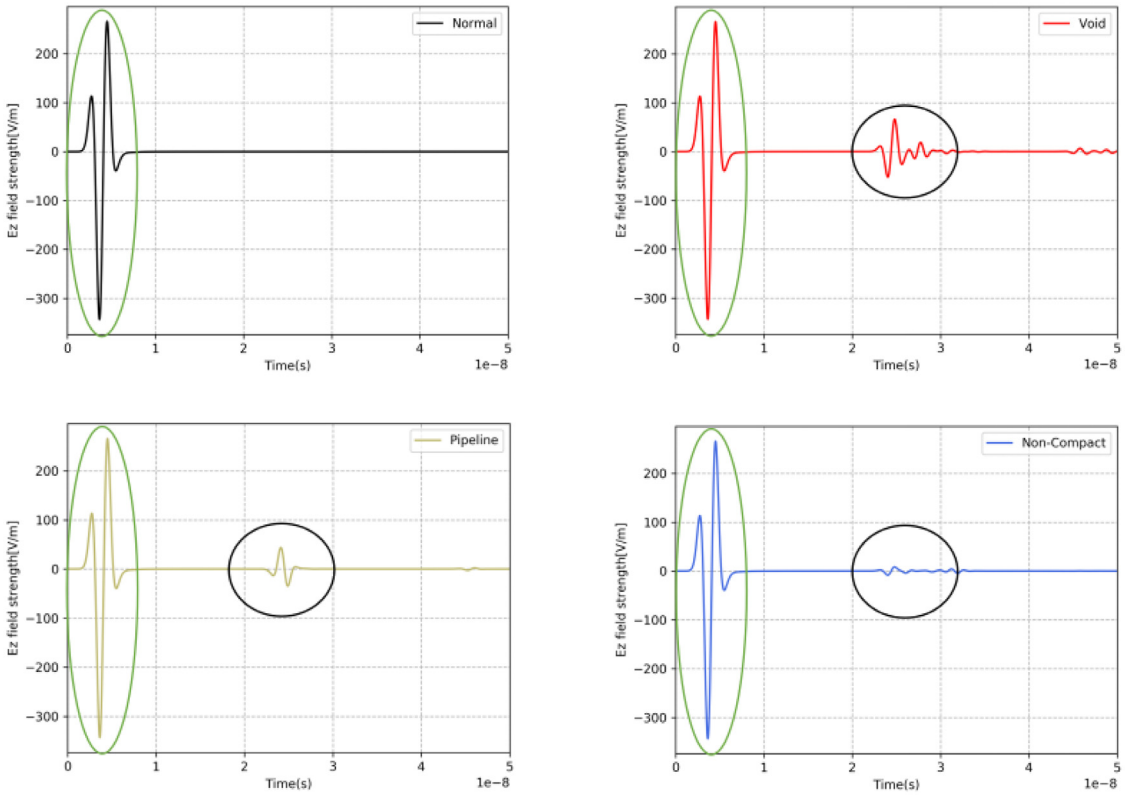


Fig. 14. Time-domain signals of a healthy road and three types of defects.

information in images, the backbone network of the original YOLOX model, CSPDarknet, was replaced by the Swin Transformer for the extraction of underground defect features, as shown in Figure 20.

The Swin Transformer is located in the YOLOX backbone network. The Patch partition module is responsible for receiving the defected image information from the road underground, and the C2, C3, and C4 modules output feature maps of sizes 80×80, 40×40, and 20×20, with feature channel numbers of 192, 384, and 768, respectively. As the network layers deepen, the semantic information of the features shifts from low-dimensional to high-dimensional. Each layer of the network causes some loss of feature information. Therefore, it is necessary to fuse features from different layers to complete the semantic information. As shown on the right side of Figure 20, feature maps of size 20×20 and 40×40 are upsampled and convolved to produce feature maps of the same scale as the previous level. Feature maps of the same scale are horizontally concatenated to achieve feature fusion. After convolution, P3 and P4 are obtained. P3 and P4, after convolution and down sampling, produce feature maps of the same scale as P4 and P5. After further convolution, feature maps N4 and N5 are obtained, implementing the bidirectional fusion concept of the PAFPN network.

3.3.2 Mask R-CNN model based on Swin transformer

In the Mask R-CNN model, a convolutional neural network is typically used to extract features from images. However, in this study, the backbone network was replaced by the Swin Transformer model to generate feature maps, which are then fed into the Region Proposal Network (RPN). The RPN generates candidate object bounding boxes to obtain the Region of Interest (RoI). The RoI is processed using the RoIAlign layer, which produces fixed size feature maps. Finally, classification and regression are performed as shown in Figure 21.

4 Intelligent detection results of road underground defects

4.1 Evaluation indicators

For the trained network models, commonly used evaluation indicators for object detection models include:

Precision (Pr): The ratio of correctly detected or segmented target samples or pixels to the total number of detected or segmented target samples or pixels. It represents the ratio of correctly identified images with defects to the total number of identified images.

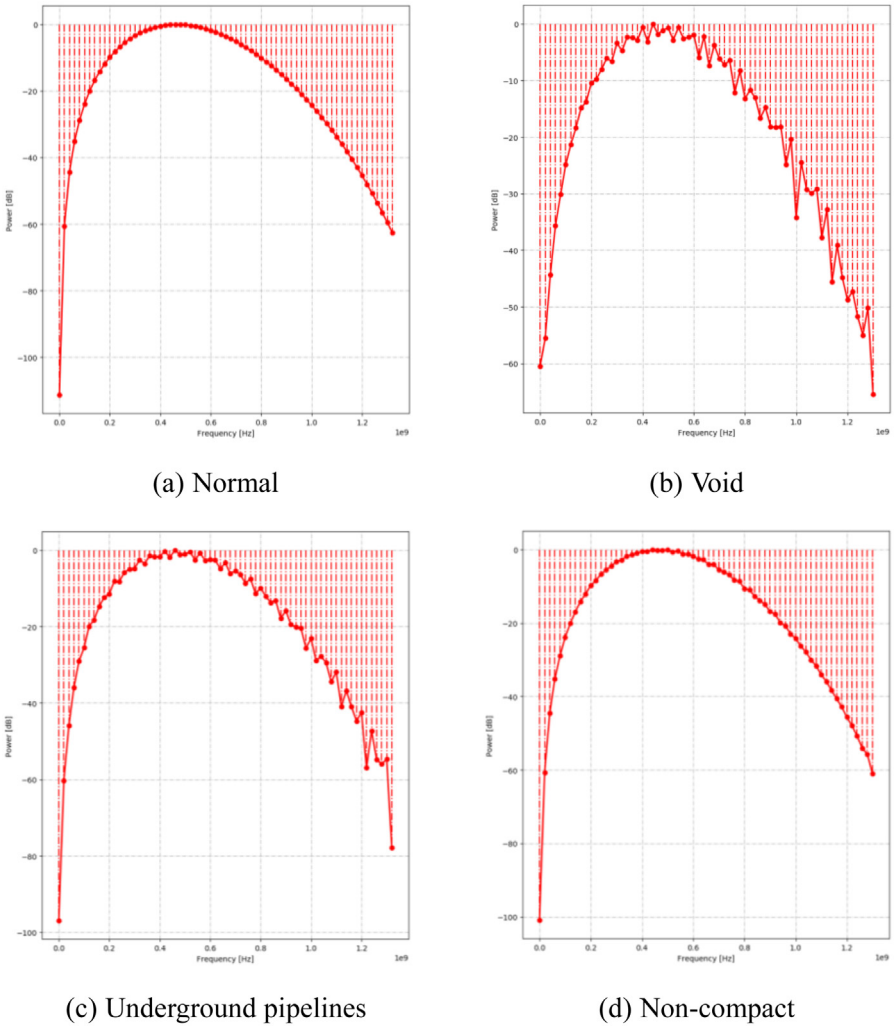


Fig. 15. Frequency-domain signals of a healthy road and three types of defects.

Fig. 16. Hardware equipment.

Table 2. Experimental parameters.

Parameters	Value
Antenna frequency	170,600 MHZ High-frequency
Sampling points	1024/Sc
Antenna movement speed	2 km/h
Scanning speed	64 times/s
Trigger mode	Ranging Wheel
Dielectric constant	8.0
Time windows	20 ns

Recall (Re): The ratio of correctly detected or segmented target samples or pixels to the actual number of target samples or pixels. It represents the ratio of correctly identified images with defects to the total number of images with defects.

Average Precision (AP): The integral of the P-R curve, which is a comprehensive evaluation indicator for recall and precision. The value of AP ranges from 0 and 1, with a higher value indicating higher accuracy of detection.

Table 3. Summary of engineering testing results.

Defect number	Measurement line	Defect mileage location/m	Defect length/m	Defect type	Defect depth/m
1	1	28-52	2.4	Void	0.5-1.4
2	1	71-84	1.3	Non-compact	0.5-1.3
3	1	190-200	1.0	Non-compact	0.5-1.8
4	1	248-254	0.6	Void	1.2-1.6
5	1	284-288	0.4	Void	0.8-1.8
6	2	0-20	2.0	Non-compact	0.5-1.1
7	2	20-50	3.0	Void	0.5-1.3
8	2	72-88	1.6	Underground pipelines	0.5-0.8
9	2	105-130	2.5	Non-compact	0.5-0.8
10	2	140-188	4.8	Underground pipelines	0.6-0.8
11	2	197-219	2.2	Non-compact	0.5-1.3
12	2	221-245	2.4	Void	0.55-1.4
13	2	249-275	2.6	Non-compact	0.5-1.9
14	3	36-64	2.8	Underground pipelines	0.5-1.3
15	3	73-87	1.4	Non-compact	0.5-1.6
16	3	111-116	0.5	Void	0.5-1.3
17	3	201-211	1.0	Underground pipelines	0.5-1.1
18	3	226-235	0.9	Void	0.6-1.1
19	3	254-266	1.2	Void	0.9-1.2
20	3	266-274	0.8	Non-compact	0.5-1.1
21	3	275-284	0.9	Void	0.7-1.4
22	4	0-17	1.7	Underground pipelines	0.5-1.3
23	4	42-110	6.8	Non-compact	0.3-1.4
24	5	0-7	0.7	Void	1.7-1.9
25	5	28-35	0.7	Void	0.8-1.1
28	6	12-28	1.6	Non-compact	0.4-1.3
29	6	36-47	1.1	Underground pipelines	0.7-1.4
30	6	67-71	0.4	Void	1.4-1.7
31	7	0-15	1.5	Void	1-1.8
32	7	4-26	2.2	Non-compact	0.3-1.1
33	7	30-34	0.4	Void	1.0-1.6
34	7	41-43	0.2	Void	1.2-1.6

The specific formulas used to calculate these indicators are as follows:

$$Pr = \frac{TP}{TP + FP}, \tag{4}$$

$$Re = \frac{TP}{TP + FN}. \tag{5}$$

Among them, TP represents the cases where an image with a defect is correctly identified as defective, i.e., the image has a defect and the model also recognizes it as a defect. FP represents the cases where an image without a defect is mistakenly identified as having a defect. FN represents the cases where an image with a defect is mistakenly identified as not having a defect.

$$AP = \int_0^1 P(R) = \sum_{k=1}^{N} \quad \mathsf{Pr} \quad (k)\Delta R(k), \tag{6}$$

where $k = 1, 2, \ldots, N$; N is the number of samples; $Pr(k)$ is the precision of the kth sample, $\Delta R(k)$ is the recall between the kth sample and the $(k\text{-}1)th$ sample.

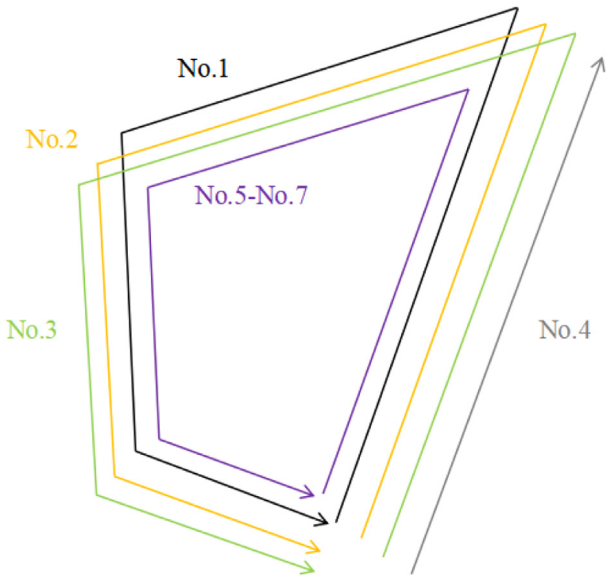


Fig. 17. Measurement lines layout.

Fig. 18. Field testing.

4.2 Road underground defects based on Swin transformer model

The images collected from the field were enhanced to augment the dataset, resulting in a total of 1433 images, including both normal and defect images. A subset of 20% of the dataset, consisting of 287 images, was selected as a test set for intelligent underground defect detection.

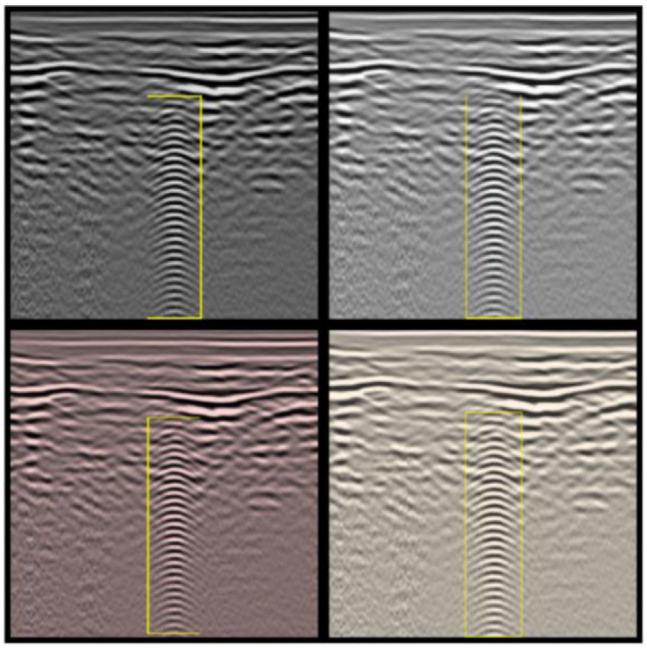


Fig. 19. Sample augmentation.

Based on the Swin Transformer feature extraction network for object detection, and keeping the dataset the same for comparison, Table 4 presents the precision and recall of two backbone networks: YOLOX and Mask R-CNN. The YOLOX model achieved a precision of 94.2% and a recall of 99.1%, showing an improvement of 5.2% and 7.5% respectively compared to the Mask R-CNN model. Furthermore, as shown in Figure 22, the average precision for Mask R-CNN was 79%, while YOLOX achieved an average precision of 94%, demonstrating its superiority in feature processing over Mask R-CNN. These results confirm the feasibility of using YOLOX as the backbone network in this study.

Figure 23 illustrates the partial detection results of the Swin-YOLOX model on the test dataset, showing the detected bounding boxes and confidence scores. It can be observed that the detection network accurately identifies the positions and extents of the defects. Moreover, it assigns higher confidence scores to the defect regions, thus validating the accuracy and feasibility of the model.

4.3 Road underground defect classification and recognition based on vibration signals

The time-domain signals collected by the radar were processed to extract prominent waveform segments, resulting in a total of 300 signals representing three types of defects. These 300 sets of radar signals were divided into three subsets: training set, validation set, and test set. The training set comprised 80% of the dataset, with 240 sets of radar time-domain signals used as the input to train a neural network classification model. The model achieved a classification accuracy of 90% for the three types of defects:

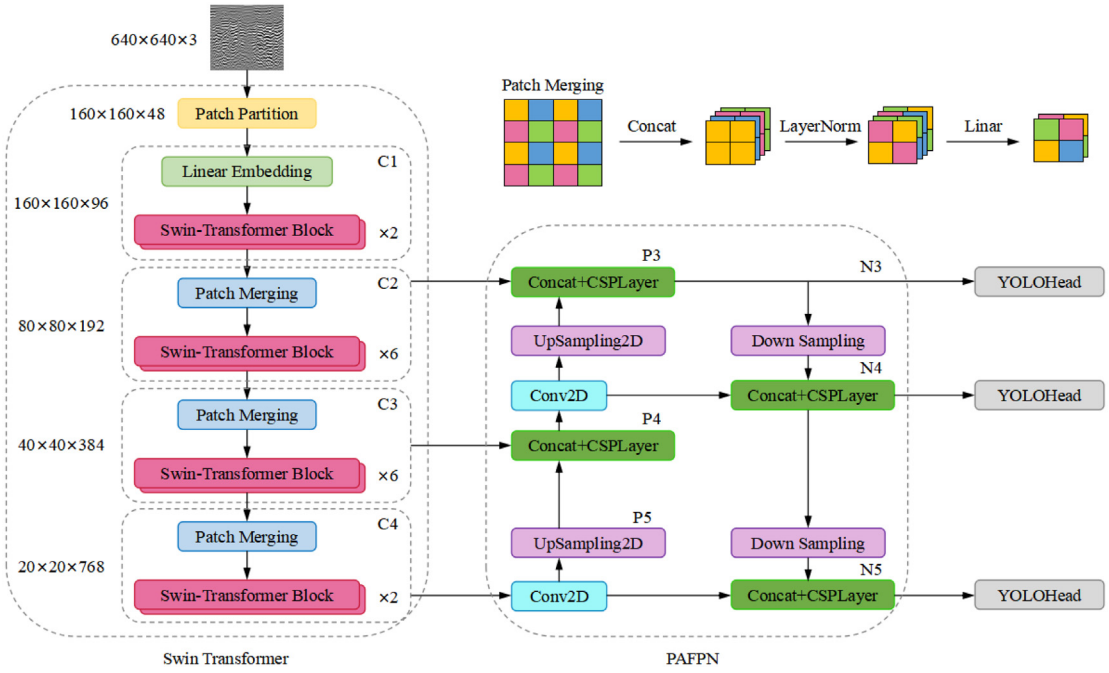


Fig. 20. Network structure of YOLOX based on Swin-T.

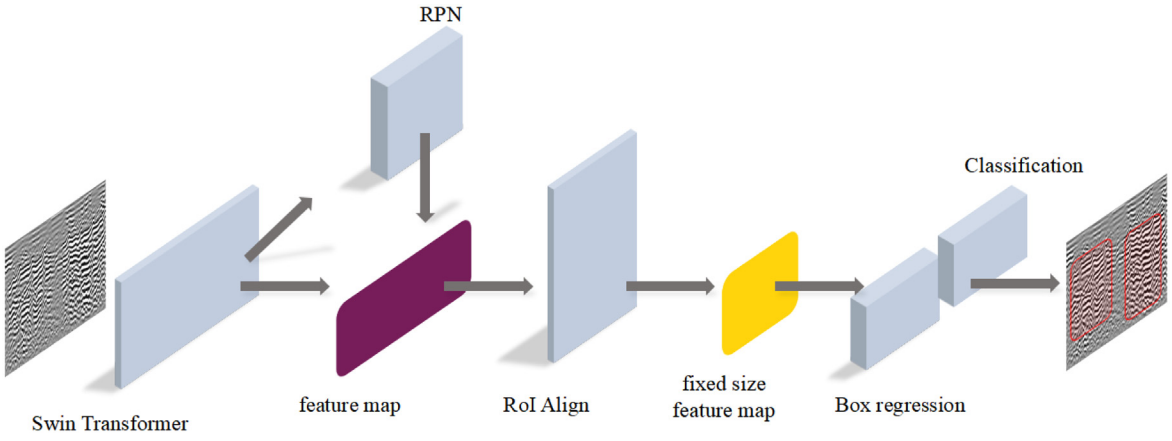


Fig. 21. Network structure of Mask R-CNN based on Swin-T.

voids, non-compact areas, and underground pipelines, as shown in Table 5. This demonstrates the feasibility of using this method for signal classification.

4.4 Intelligent localization of road underground defects based on GPS

The detected grayscale images containing defects are imported into the system together with the GPS information of the collected measurement lines and points, as shown in Figure 24. The system automatically locates the position of the defect center on the actual map. It also displays the previous and next measurement points, which are measurement point 5 and measurement point 6 respectively. The detailed address of the defect location is: 946 Township Road, Zhenlong Town, Huiyang District, Huizhou City, Guangdong Province.

Table 4. Training results of two network structures.

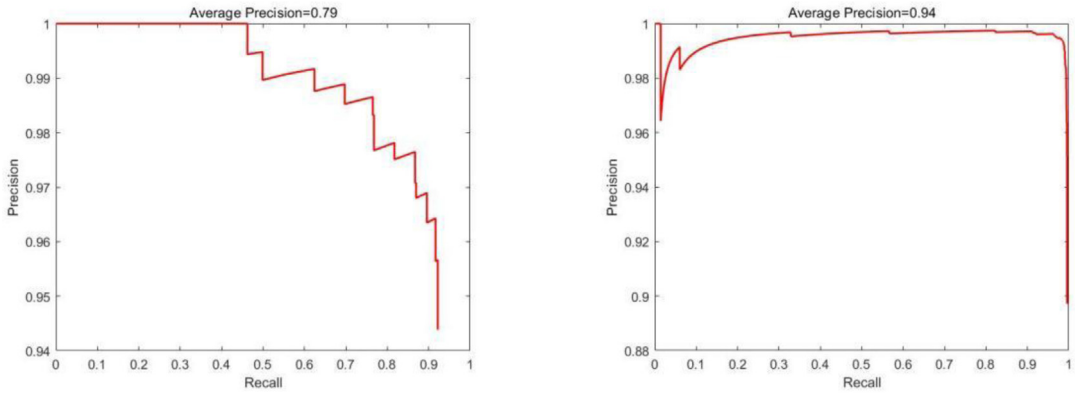
Network structures	Mask R-CNN	YOLOX
Pre (%)	89%	94.2%
Recall (%)	91.6%	99.1%

5 Discussion and conclusions

5.1 Discussion

5.1.1 Data diversity

This paper utilizes image enhancement techniques to augment the dataset of road underground defects, thereby addressing the issue of insufficient training samples. To further improve the generalization ability of the detection



(a). Swin-Mask R-CNN P-R curve (b). Swin-YOLOX P-R curve

Fig. 22. The P-R curves of Swin-Mask R-CNN and Swin-YOLOX models.

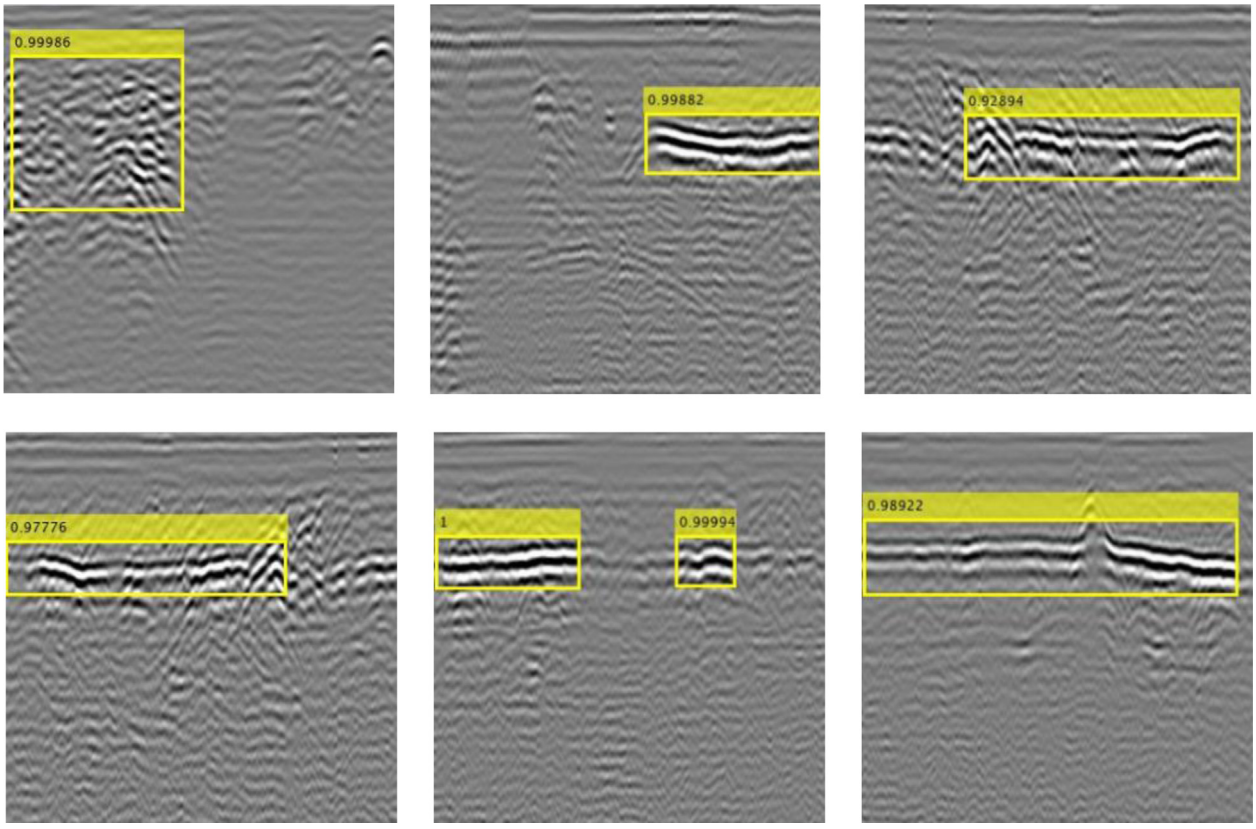


Fig. 23. Detection performance of the Swin-YOLOX structure for defect.

network and to build better deep learning models, the problem of data diversity becomes particularly important. With the development of image processing techniques, future considerations include the use of various techniques such as Generative Adversarial Networks (GANs) and Poisson blending to generate new images, thereby increasing the quantity and quality of the training set. This will enhance the diversity of image data and improve the detection results of the algorithm, enabling better application in engineering projects.

5.1.2 Defects feature indicators

This paper mainly focuses on the research of intelligent detection methods for radar images and radar time-domain signals. In future studies, data fusion techniques, such as integrating radar reflection wave data with image data, can be considered to obtain more defect feature indicators and enhance the effectiveness of intelligent detection.

Table 5. Classification results of time-domain signals.

	Voids	Underground pipelines	Non-compact areas	Total
Correct	19	16	19	54
Error	1	3	2	6
Total	20	20	20	60
Accuracy	95%	85%	90%	90%

Fig. 24. Localization of road defect based on GPS.

5.2 Conclusions

This study investigates the features and patterns of three typical road defects in radar images and signals. The shapes and arrangements of bands in radar images effectively discriminate defect types, while the amplitude and frequency of radar reflection waves can classify the defect. This indicates that grayscale images and radar reflection wave signals can be used as a basis for determining the presence of anomalies in radar data and can be applied in practical engineering projects.

The Swin Transformer object detection algorithm is employed in this study, which approaches the problem from both the feature extraction and detection network perspectives. Using transfer learning, the performance of the network is examined under different parameter settings for defect classification and detection. The results show that compared to the Mask R-CNN being used as the backbone network model, the YOLOX model achieves improvements of 5.2% in precision, 7.5% in recall, and 15% in average precision. Consequently, the Swin-YOLOX model is considered to be effective for road underground defect detection, offering higher detection accuracy. The final choice of the detection network model is Swin-YOLOX.

Furthermore, this study utilizes GPS positioning technique to determine the actual locations of detected defects on the road. By contrasting the GPS information of the measurement points and the defect areas in images, the relative positional relationship between the measurement points and the defect points is calculated to obtain precise location information for defects on the actual road. This information is crucial for maintenance and prevention efforts to detect actual road defects.

References

1. R. Knight, Ground penetrating radar for environmental applications, Annu. Rev. Earth. Planet. Sci. **29**, 229–255 (2001)
2. X.L. Lu, X.Q. Hu, C.P. Luo, Z.Y. Xu, X. Liao, L.H. Liu et al., A rapid imaging method of the seismic back-scattered wavefield for urban road near-surface anomalous structures, Near Surf. Geophys. **20**, 315–328 (2022)
3. T. Thongprapha, K. Fuenkajorn, J. Daemen, Study of surface subsidence above an underground opening using a trap door apparatus, Tunn. Undergr. Sp. Tech. **46**, 94–103 (2015)
4. C. Rodeick, Roadbed void detection by ground penetrating radar, Highw. Heavy Constr. **127**, 60–61 (1984)
5. H. Liu, Z.S. Shi, J.H. Li, C. Liu, X. Meng, Y.L. Du et al., Detection of road cavities in urban cities by 3d ground-penetrating radar, Geophys. **86**, WA25–WA33 (2021)
6. T. Noreen, U. Khan, Using pattern recognition with HOG to automatically detect reflection hyperbolas in ground penetrating radar data, *Proceedings of the IEEE International Conference on Electrical and Computing Technologies and Applications (ICECTA)*, 2017, pp. 1–6
7. A. Giannopoulos, Modelling ground penetrating radar by gprmax, Constr. Build. Mater. **19**, 755–762 (2005)
8. W. Lei, J. Luo, F. Hou, L. Xu, R.Q. Wang, X.Y. Jiang, Underground cylindrical objects detection and diameter identification in GPR B-scans via the CNN-LSTM framework, Electronics **9**, 1804 (2020)
9. Z. Tong, J. Gao, Z. Han, Z.J. Wang, Recognition of asphalt pavement crack length using deep convolutional neural networks, Road Mater. Pavement. Des. **19**, 1334–1349 (2018)
10. S. Teng, X. Chen, G. Chen et al., Structural damage detection based on convolutional neural networks and population of bridges [J], Measurement, **202**, 111747 (2022)
11. J. Kers, J. Majak, Modelling a new composite from a recycled GFRP [J], Mech. Compos. Mater. **44**, 623–632 (2008)
12. C.C. Zhang, Z.O. Zhou, Research on automatic target detection and orientation of ground penetrating radar in shallow subsurface application, J. Electron. Inf. Technol. **27**, 1065–1068 (2005)
13. J. Zhang, X. Yang, W.G. Li, S.B. Zhang, Y.Y. Jia, Automatic detection of moisture damages in asphalt pavements from GPR data with deep CNN and IRS method, Autom. Constr. **113**, 103119 (2020)
14. W.T. Lei, F.F. Hou, J.C. Xi, Q.Y. Tan, M.D. Xu, X.Y. Jiang et al., Automatic hyperbola detection and fitting in GPR B-scan image, Autom. Constr. **106**, 102839 (2019)
15. X.G. Pan, J.P. Shi, P. Luo, X.G. Wang, X.O. Tang, Spatial as deep: Spstial CNN for traffic scene understanding. Thirty-Second AAAI Conference on Artificial Intelligence, 2018, Vol. 32, pp. 1–6
16. H.B. Hu, H.Y. Fang, N.N. Wang, D. Ma, J.X. Dong, B. Li et al., Defects identification and location of underground space for ground penetrating radar based on deep learning, Tunn. Undergr. Sp. Tech. **140**, 105278 (2023)

17. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. Gomez et al., Attention is all you need. *NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems*, 2017, pp. 6000–6010
18. Z. Liu, Y.T. Lin, Y. Cao, H. Hu, Y.X. Wei, Z. Zhang et al., Swin transformer: Hierarchical vision transformer using shifted windows. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 10012–10022
19. A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner et al., An image is worth 16×16 words: Transformers for image recognition at scale, arXiv preprint arXiv **2010**, 11929 (2020)
20. Z. Ge, S.T. Liu, F. Wang, Z.M. Li, J. Sun, Yolox: Exceeding yolo series in 2021, arXiv preprint arXiv **2107**, 08430 (2021)
21. J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 3431–3440
22. K. He, G. Gkioxari, P. Dollár, R. Girshick, Mask R-CNN. Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 2961–2969
23. P. Asadi, M. Gindy, M. Alvarez, A machine learning based approach for automatic rebar detection and quantification of deterioration in concrete bridge deck ground penetrating radar b-scan images, KSCE J. Civ. Eng. **23**, 2618–2627 (2019)

Cite this article as: Jinke An, Li Yang, Zhongyu Hao, Gongfa Chen, Longjian Li, Investigation on road underground defect classification and localization based on ground penetrating radar and Swin transformer, Int. J. Simul. Multidisci. Des. Optim. **15**, 7 (2024)