Enhanced Road Surface Damage Detection Using Ground Penetrating Radar with an Optimized RT-DETR Model

Shuangxiu Zhang [1,*]
College of Artificial Intelligence
Southwest University
Chongqing 400715, China
[1]e-mail: zshuangxiu@163.com

Shiji Chen[2]
College of Artificial Intelligence
Southwest University
Chongqing 400715, China
[2]e-mail: csj841596779@email.swu.edu.cn

***Abstract*—Ground Penetrating Radar (GPR) is extensively used for non-destructive road surface damage detection but faces challenges due to the complex road environment affecting B-scan image quality. This paper introduces an enhanced road surface damage detection method leveraging an optimized RT-DETR backbone network, specifically ShuffleNet V2, to adapt to GPR image characteristics. The proposed method boosts feature extraction and real-time performance with a lightweight design, achieving 93.2% accuracy, 90.8% recall, and an mAP@0.5:0.95 of 0.626. It offers a 72.4% faster inference speed and 44.15% reduced computational cost compared to ResNet18, along with an 8M parameter reduction. This approach presents a superior solution for intelligent road damage detection, crucial for road maintenance and safety assessment.**

Keywords-Ground Penetrating Radar; Road Surface Damage Detection; Lightweight Network; RT-DETR; Object Detection

I. INTRODUCTION

His Road damage detection is a critical aspect of ensuring road safety, but traditional manual inspection and destructive testing methods suffer from low efficiency, limited coverage, and high costs, making them inadequate for modern rapid road inspection needs. Ground-penetrating radar (GPR) technology, with its non-contact, non-destructive, and high-sensitivity characteristics, has gradually become a key tool for road damage detection[1].By transmitting high-frequency electromagnetic waves and receiving reflected signals, GPR can quickly acquire internal structural information of the road, providing scientific support for road maintenance. However, the complexity of road materials and interference from environmental noise pose challenges for the analysis of GPR data and target detection.

GPR data is primarily presented in three forms: A-scan[2](one-dimensional depth signals), B-scan[3](two-dimensional cross-sectional images), and C-scan[4](three-dimensional target distributions). Among these, B-scan images, due to their intuitiveness and moderate data volume, have become key data for road damage detection. However, traditional methods based on manual feature extraction face limitations in handling noise interference and adapting to complex targets.

In recent years, deep learning technologies, particularly convolutional neural networks[5] (CNNs), have demonstrated significant advantages in feature extraction and multi-scale target recognition, greatly improving detection accuracy and generalization. To address the balance between real-time performance and accuracy, this paper proposes a lightweight GPR B-scan image detection method based on an improved RT-DETR model. Through lightweight network design and specific feature optimization strategies, the proposed method achieves high precision and speed at low computational costs, providing an efficient solution for intelligent road damage detection.

II. METHOD

A. Introduction to the RT-DETR Model

RT-DETR is a real-time end-to-end object detector based on Transformer, which effectively handles multi-scale features by decoupling intra-scale interaction and cross-scale fusion[6]. This significantly reduces the computational cost of DETR and outperforms advanced YOLO detectors of similar size in both speed and accuracy. RT-DETR omits the post-processing step, ensuring no inference delay while maintaining stable performance.

In the RT-DETR object detection model, the official implementation provides dozens of different models with backbones such as ResNet and HGNet. In this paper, RT-DETR-R18, which balances network depth and detection accuracy, is chosen as the base model. The specific network structure is shown in the Figure 1.

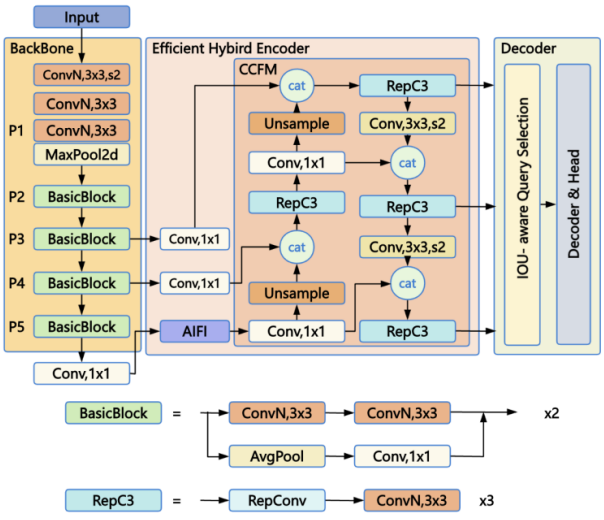

Figure 1. Network Structure of the RT-DETR-ResNet18 Model

979-8-3315-0976-7/25/$31.00 ©2025 IEEE

The model is composed of a backbone network, a hybrid encoder, and a decoder with auxiliary prediction heads. The backbone network uses the ResNet18 structure to extract features from the input image, taking the output features of the last three stages as inputs to the encoder, with output strides of 8, 16, and 32, respectively. The hybrid encoder consists of the AIFI module and the CCFM module: the AIFI module encodes the highest-level P5 features, while the CCFM module converts multi-scale features into a series of image features through feature fusion in both bottom-up and top-down paths. The decoder first utilizes an IoU-aware query module to select a fixed number of image features from the encoder's output sequence as initial object queries, which are then iteratively refined to generate prediction boxes and confidence scores.

B. *Lightweight Network Structure*

The backbone network of RT-DETR adopts the ResNet series, which consists of multiple residual blocks containing convolution, pooling, and activation functions. This structure affects the speed of feature extraction, making it less suitable for real-time object detection tasks. ShuffleNet V2[8], as a lightweight and efficient convolutional neural network designed for image classification and object detection, performs channel shuffle (Figure 2a) and grouped convolution (Figure 2b) operations within the ShuffleNet V2 unit. These operations significantly reduce the model's parameters and computational cost while maintaining accuracy, enabling it to adapt to resource-constrained hardware environments.

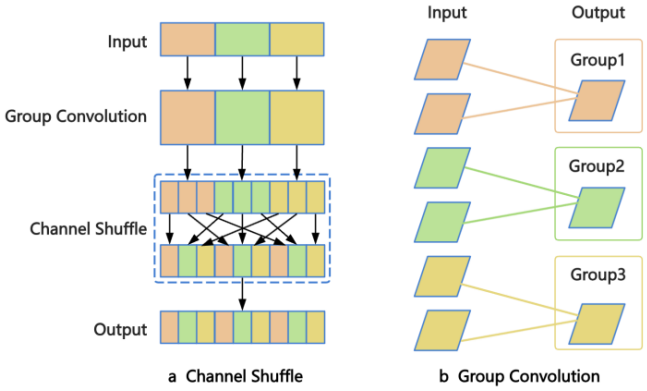

Figure 2. Core Structure of ShuffleNet V2

In Figure 3, Unit1 and Unit2 correspond to the cases of stride = 1 and stride = 2 in ShuffleNet V2, respectively. At the beginning of each unit, the input with c feature channels is split into two branches, containing c − c' and c' channels, respectively. Following G3, one branch remains unchanged. The other branch consists of three convolutions with the same number of input and output channels to meet the requirements of G1. After the convolution, the two branches are concatenated together. As a result, the number of channels remains unchanged (G1). Then, a "channel shuffle" operation is applied to enable information exchange between the two branches.

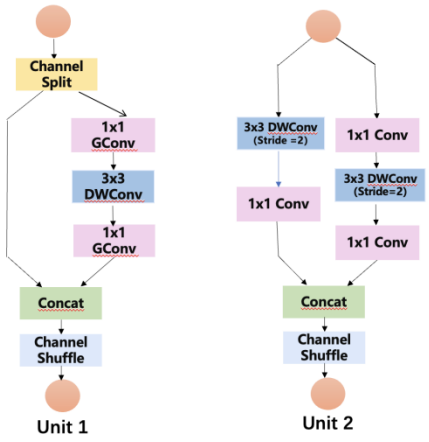

Figure 3. Structure of the Basic ShuffleNet V2 Unit and the Basic Spatial Downsampling Unit

C. *Improved Model Network Structure*

To address the issue of model deployment difficulties caused by the massive computational load of the Transformer architecture, this paper proposes improvements based on RT-DETR. Specifically, we adopt ShuffleNet V2 as the backbone network replacement, aiming to reduce the model's computational complexity by introducing a lightweight convolutional neural network structure, while maintaining detection accuracy. Compared to traditional ResNet series networks, ShuffleNet V2 significantly reduces the number of parameters and computational cost through optimized depthwise separable convolutions and channel shuffling operations, making it especially suitable for resource-constrained hardware environments and real-time object detection tasks. The overall network structure of the improved model is illustrated in Figure 4.

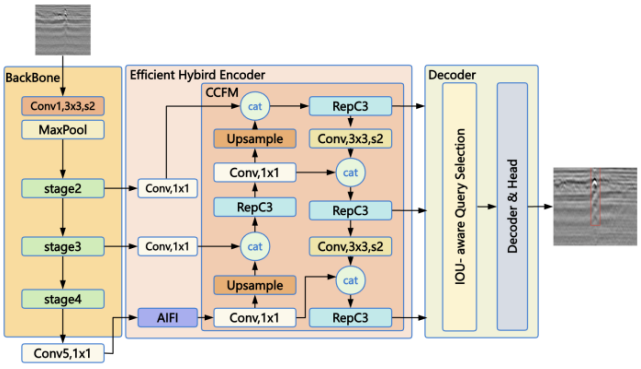

Figure 4. Network Structure of the Improved Model

III. EXPERIMENT

A. *dataset*

The experiments in this paper are based on a self-built dataset, supplemented with images collected from the internet and captured manually[9]. After screening, a total of 1,600 images were obtained, containing two types of defects: cracks and loose areas, named "crack" and "loose," respectively. A single image may contain multiple types of defects. The images were shuffled,

and following the construction method of the PASCAL VOC dataset, the images were strictly annotated using LabelImg according to labeling standards. The resulting category and location information were saved in XML files. Finally, the dataset was divided into training, validation, and test sets in a ratio of 7:2:1. Examples of defect images are shown in Figure 5.

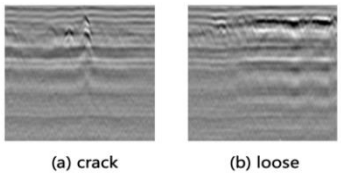

Figure 5. Example of GPR Defect Images

B. *Experimental Environment*

The experiments were conducted using the PyTorch deep learning framework, version 1.11.0. The operating system of the machine used for training is Ubuntu 20.04, with an RTX 4090D (24GB) graphics card and an 18 vCPU AMD EPYC 9754 128-Core Processor. CUDA 11.3 was used to accelerate the training process.

The experimental parameters are set as follows: the input image size is 640×640, with an initial learning rate of 1×10^{-4}, which undergoes linear warm-up for the first 2000 steps. The learning rate momentum is set to 0.9. The model was trained for 100 epochs on the training set with a batch size of 4. The AdamW optimizer was used, which is a variant of the Adam optimizer with weight decay, and the weight decay coefficient is set to 0.0001 to help prevent overfitting.

C. *Evaluation Metrics*

To analyze the model's performance in detecting road surface potholes, common metrics such as Precision (P), Recall (R), and Mean Average Precision (mAP) are selected as the main indicators for evaluating the algorithm's performance. Model parameters such as the number of Parameters (Param), Detection Speed (FPS), and Computational Complexity (GFLOPs) are used to assess the model's inference performance.

$$P = \frac{TP}{TP + FP}$$

$$R = \frac{TP}{TP+FN}$$

$$\text{mAP}=\frac{1}{\text{N}}\sum_{\text{i=1}}^{\text{N}}\text{AP}$$

$$\text{AP}=\int_{0}^{1}\text{P(R)dR}$$

D. *Experimental Results*

To compare the performance of different lightweight networks, we evaluated the effectiveness of replacing ResNet18 with MobileNet V2[7] and ShuffleNet V2 as the backbone networks.

Figure 6 shows the performance comparison curves of the RT-DETR model using ResNet18, MobileNet V2, and ShuffleNet V2 as backbones. The metrics include Precision, Recall, mAP@0.5, and mAP@0.5:0.95. From the Precision and Recall curves, it can be observed that RT-DETR-ShuffleNet V2 exhibits faster convergence and superior final performance compared to the other two backbone networks, demonstrating higher stability and accuracy.

For mAP@0.5 and the stricter mAP@0.5:0.95 metrics, ShuffleNet V2 also achieves faster improvement and higher final scores than the other models, with ResNet18 and MobileNet V2 showing similar but slightly lower performance compared to ShuffleNet V2.

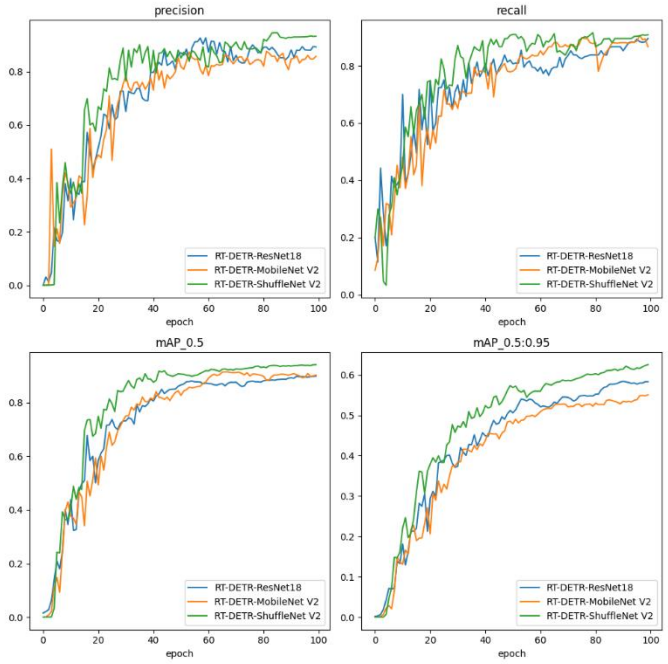

Figure 6. Performance Comparison Curves of Different Backbone Networks

TABLE I. and TABLE II. present the comparison results of precision, computational cost, params, inference speed, and model size when different networks are used as the backbone of the RT-DETR model.

TABLE I. PERFORMANCE COMPARISON OF DIFFERENT BACKBONE NETWORKS

Model	***BackBone***	***P***	***R***	***mAP@0.5***	***mAP@0.5:0.95***
RT-DETR-ResNet18	ResNet18	0.88	0.852	0.893	0.584
RT-DETR-MobileNet V2	MobileNet V2	0.857	0.876	0.903	0.55
RT-DETR-ShuffleNet V2	ShuffleNet V2	0.932	0.908	0.942	0.626

From TABLE I. it can be observed that replacing the backbone network leads to varying degrees of improvement in metrics such as precision, recall, and mAP, with the improvement range spanning 4% to 14%. When the backbone network is replaced with ShuffleNet V2, precision increases to

93.2%, recall rises to 90.8%, and mAP@0.5:0.95 improves by approximately 7.2%, significantly outperforming ResNet18 and MobileNet V2. The ShuffleNet V2 backbone also demonstrates excellent overall performance in more challenging object detection tasks.

TABLE II. COMPARISON OF COMPUTATIONAL COMPLEXITY AND MODEL RESOURCE CONSUMPTION FOR DIFFERENT BACKBONE NETWORKS

Model	***BackBone***	***GFLOPs***	***Inference speed (ms)***	***Model size (MB)***	***Params(M)*** /10^6
RT-DETR-ResNet18	ResNet18	56.9	7.6	40.5	19.9
RT-DETR-MobileNet V2	MobileNet V2	39.5	9.6	23.4	11.3
RT-DETR-ShuffleNet V2	ShuffleNet V2	31.8	13.1	24.7	11.9

From TABLE II. , it can be seen that when using the ResNet18 series as the backbone network, the model typically consumes a large amount of computational resources, has a higher number of parameters, and exhibits relatively slower inference speed. However, by replacing ResNet18 with the lightweight ShuffleNet V2 as the backbone network, the inference speed improves to 13.1ms, computational cost is reduced by 44.15%, and the number of parameters decreases by 8M, with a 72.4% increase in inference speed. This significantly reduces both computational cost and the number of parameters while maintaining the model's high performance.

Although MobileNet V2 has a minor advantage in model size,at 23.4MB compared to ShuffleNet V2's 24.7MB, ShuffleNet V2 demonstrates superior performance in terms of computational complexity and inference speed. Consequently, ShuffleNet V2 offers a more favorable balance between speed and accuracy, fulfilling the requirements for a lightweight design, and thus emerges as a more efficient option.

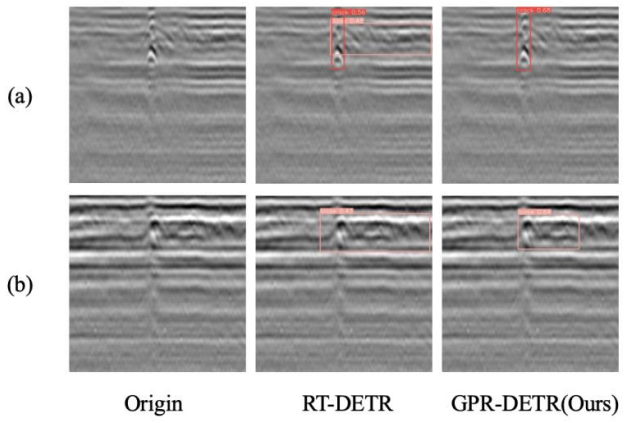

Figure 7. Comparison of Detection Performance Between RT-DETR and GPR-DETR Models on Actual GPR Defect Images.

Following a comparative analysis of various backbone networks, we selected ShuffleNet V2 as the optimal backbone due to its superior performance, and named the enhanced model GPR-DETR. Figure 7 illustrates a comparison of the detection performance of GPR-DETR with that of the baseline model on actual GPR defect images.

In actual road detection scenarios, our GPR-DETR model significantly outperforms the RT-DETR model in identifying loose pavement damage. As shown in Figure 5(a), the RT-DETR model experienced false positives, mistakenly identifying background areas as damage. In Figure 5(b), the GPR-DETR model achieved a confidence level of 0.64 for detecting loose damage, compared to the RT-DETR model's confidence level of only 0.43.

IV. CONCLUSIONS

To meet the requirements for accuracy and speed in detecting internal road defects using GPR images, a lightweight improved algorithm based on RT-DETR is proposed, effectively simplifying the RT-DETR model. By replacing the original ResNet series with a lightweight network as the backbone, experiments revealed that ShuffleNet V2 performed the best among mainstream lightweight networks. Experimental results show that, compared to the original model, GPR-DETR achieves improvements of 5.2, 5.6, and 4.9 percentage points in precision, recall, and mAP, respectively. While meeting accuracy requirements, the computational cost and parameter size of the RT-DETR model are significantly reduced, greatly enhancing inference speed.

REFERENCES

[1] Liu, Z., Wang, S., Gu, X., Wang, D., Dong, Q., & Cui, B. (2024). Intelligent assessment of pavement structural conditions: a novel femvit classification network for gpr images. IEEE Transactions on Intelligent Transportation Systems.

[2] Kim, J., Nam, B. H., & Youn, H. (2024, May). Sensitivity analysis on GPR A-scan data for cavity detection based on area estimation approach. In IOP Conference Series: Earth and Environmental Science (Vol. 1333, No. 1, p. 012008). IOP Publishing.

[3] Wang, B., Chen, P., & Zhang, G. (2023). Simulation of GPR B-scan data based on dense generative adversarial network. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 16, 3938-3944.

[4] Zhou, Y. , Zhang, J. , Hu, Q. , Zhao, P. , Ai, M. , & Huang, Y. , et al. (2024). Deep learning based method for 3d reconstruction of underground pipes in 3d gpr c-scan data. Tunnelling and underground space technology(Aug.), 150.

[5] Liu, H. , Wang, S. , Jing, G. , Yu, Z. , Yang, J. , & Zhang, Y. , et al. (2023). Combined cnn and rnn neural networks for gpr detection of railway subgrade diseases. Sensors (14248220), 23(12).

[6] Zhao, Y., Lv, W., Xu, S., Wei, J., Wang, G., Dang, Q., ... & Chen, J. (2024). Detrs beat yolos on real-time object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 16965-16974).

[7] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018). Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4510-4520).

[8] Ma, N., Zhang, X., Zheng, H. T., & Sun, J. (2018). Shufflenet v2: Practical guidelines for efficient cnn architecture design. In Proceedings of the European conference on computer vision (ECCV) (pp. 116-131).

[9] Chen, A., Zhou, X., Fan, Y., & Chen, H. (2023). Underground Diagnosis Based on GPR and Learning in the Model Space. IEEE Transactions on Pattern Analysis and Machine Intelligence.