scientific reports

Check for updates

OPEN

A high-precision segmentation method based on UNet for disc cutter holder of shield machine

Dandan Peng, Guoli Zhu✉ & Zhe Xie

Visual positioning plays a pivotal role in enabling robotic disc cutter replacement for the shield machine. However, underground operational challenges—including low illumination, high dust concentrations, and irregular sand deposition on the surface of the disc cutter and its holder—severely compromise recognition accuracy. To address this, we propose a multi-mechanism enhanced UNet model for robust segmentation of the disc cutter holder under heterogeneous surface conditions. Experimental comparisons with mainstream semantic segmentation models demonstrate that the Res-UNet achieves superior training efficiency and segmentation accuracy. Ablation studies further reveal optimal performance when utilizing a hybrid loss function (dice loss + cross-entropy loss) paired with the Adam optimizer. By integrating attention mechanisms, we develop the Res-UNet-CA architecture, which achieves state-of-the-art metrics on independent test sets: accuracy (99.45%), precision (98.9%), recall (99.11%), F1-score (99%), and mIoU (98.63%). The Res-UNet-CA model significantly outperforms other semantic segmentation models in prediction quality, offering an innovative solution for shield machine disc cutter holder detection.

Keywords Shield machine, Disc cutter holder, Semantic segmentation, Attention mechanism

A shield machine serves as a critical equipment in tunnel excavation, utilizing disc cutters on the cutterhead to fracture rock and soil during operation. Due to rapid wear rates, disc cutters require frequent replacement. However, current manual replacement processes exhibit inefficiency and safety risks, substantially hindering tunneling productivity and escalating operational costs. To address these limitations, automation-driven solutions such as disc cutter replacement robots have emerged. Visual localization constitutes a critical component of robotic cutter replacement, particularly as disc cutters inherently lack distinct positioning features. Consequently, surface characteristics of the cutter holder could be employed for localization. The core objective of cutter holder segmentation lies in deploying image detection algorithms to isolate the holder from complex backgrounds, achieving results comparable to manual segmentation precision.

Deep learning has gained extensive adoption across diverse domains and offers innovative approaches to unresolved challenges in shield construction. For instance, it has been applied to segment cracks in tunnel lining images for leakage detection[1] and to analyze muck characteristics for real-time geological monitoring during excavation[2]. Semantic segmentation is an important application of deep learning in image processing, which aims to assign a class label to each pixel in the image. The foundational framework for modern semantic segmentation models is the fully convolutional network (FCN) proposed by Long et al. (2015)[3]which replaces traditional CNN fully connected layers with convolutional layers to generate spatial outputs adaptable to arbitrary input sizes. The typical image semantic segmentation algorithms commonly used are UNet[4]DeeplabV3[5], DeeplabV3 +[6], LRASPP[7]PSPNet[8]DANet[9]SegFormer[10]etc. Among these, the UNet network, which is widely used in the field of medical lesion detection, is favoured by researchers because of its high detection accuracy and simple network structure. Many improvements and attempts have been made to apply it in different fields.

Shallow UNet architectures often suffer from insufficient feature learning and low segmentation accuracy, while excessive network depth may induce performance degradation. The unique residual structure of the Resnet network can effectively alleviate this problem[11]. Integrating ResNet with UNet (Res-UNet) addresses both depth limitations and performance decline under extreme depth conditions, thereby enhancing segmentation capability[12]. Xu et al. proposed a segmentation model for soil crack images that combines deep Res-UNet and the attention gate. The model can effectively identify soil cracks under uneven illumination conditions[13]. Feng et al. proposed a lightweight Res-UNet method, which can achieve accurate segmentation of reflection ferrograms and has good anti-interference performance[14]. Res-UNet has also been adapted for building change

The School of Mechanical Science & Engineering, Huazhong University of Science and Technology, Wuhan 430074, China. ✉email: glzhu@mail.hust.edu.cn

detection[15]coronal loop identification[16]and near-infrared image colorization[17]. Additionally, its applications extend to the segmentation of tar-rich coal macerals particles[18]the segmentation of paint craquelures in traditional polychrome paintings[19]the classification of satellite images in complex urban surfaces[20]and the classification of tree species[21].

The integration of attention mechanisms into UNet architectures has significantly enhanced performance across diverse visual tasks[22–24]. These mechanisms optimize feature extraction by emphasizing discriminative patterns while suppressing irrelevant information. Attention modules (such as SE, CBAM, and ECA) are commonly used. The SE module is a typical implementation of the channel attention mechanism. By learning adaptive channel weights, the model pays more attention to useful information[25]. For instance, Yu et al. achieved improved semantic segmentation by embedding SE blocks into UNet[26]. However, SE lacks spatial awareness and performs optimally only in channel-rich scenarios. The CBAM module focuses on images from both spatial and channel aspects, aiming to enhance the ability of convolutional neural networks to focus on images[27]. Li et al. proposed a flood-submerged area extraction method based on the UNet combined with the CBAM module. The introduction of the CBAM module improves the segmentation accuracy of the network[28]. The computational complexity of the CBAM module is higher than that of the SE module, which requires more computing resources. Wang et al. improved the SE block and proposed an efficient channel attention (ECA) module, which added a small number of parameters compared to the SE module but achieved significant performance gains[29]. Introducing the ECA module in the network can strengthen information interaction and fusion, effectively extract local and global features so that the model can focus more on areas that are difficult to separate, and obtain better segmentation results[30,31]. The ECA module has some limitations in dealing with global context dependencies and channel spatial relationships. The coordinate attention (CA) module was proposed by Hou et al. in 2021. It introduces spatial attention while introducing channel attention and embeds location information into the channel attention[32].

Despite significant advancements in semantic segmentation across domains, visual inspection of the disc cutter holder in tunnel engineering remains underexplored, with no standardized datasets established for this specific task. This study pioneers a dedicated deep segmentation framework for shield machine cutter holder. In this paper, a high-precision segmentation model (Res-UNet-CA) of the cutter holder of the shield machine based on the UNet framework is proposed. The Resnet50 is used as the backbone feature extraction network for down-sampling, and the CA module is added at the bottom of the network to effectively extract local and global information. In addition, multi-scale feature fusion is used to strengthen feature extraction during up-sampling. The primary contributions of this article are as follows.

1) Architectural Innovation: By integrating ResNet50's residual blocks into UNet's downsampling path, we adopt Res-UNet model to address cutter holder segmentation. This hybrid architecture prevents gradient degradation through residual learning while enhancing feature representation. UNet's skip connections further enable hierarchical feature fusion across encoder-decoder stages, improving segmentation precision.
2) Optimization Strategy: A hybrid loss function combining dice and cross-entropy losses enhances training stability and segmentation consistency, effectively mitigating overfitting risks.
3) Attention Enhancement: The incorporated Coordinate Attention (CA) module captures long-range spatial dependencies while preserving positional integrity through directional encoding, enabling precise localization of the cutter holder. Experimental validation demonstrates Res-UNet-CA's superiority over state-of-the-art models in both segmentation quantitative metrics and visual prediction quality on our cutter holder dataset.

Dataset construction

Image acquisition

The cross-section size of the disc cutter changing room in the shield head of the shield machine is 1 m × 1 m, and the distance between the camera and the cutter holder is not more than 1 m when the vision system performs image acquisition. The space in the disc cutter changing room is narrow, and the movement of the robot is limited. In order to simulate the movement of the robot in the disc cutter changing room, we built a four-axis motion platform in the laboratory, as shown in Fig. 1, to collect the image of the cutter holder. The degrees of freedom of the platform are the translation in the XYZ direction and the rotation in the Z-axis direction. The imaging acquisition system integrates three core hardware components: camera, lens, and illumination modules. An industrial-grade camera (Manta G-1236, Allied Vision GmbH) was deployed, featuring a Sony IMX304 CMOS sensor with 4112 × 3008 resolution. To accommodate the large field of view (FOV) requirements, a Navitar NMV-8M1.1 wide-angle lens with 8.5 mm focal length was optically coupled to the camera. Considering the specular reflection characteristics of metallic components, two parallel and symmetrically arranged LED light bars were installed bilaterally along the vertical direction of the camera axis to achieve enhanced uniformity in ambient lighting[33].

Given the constrained workspace for disc cutter change and the substantial physical dimensions of the cutter holder (661 mm × 467 mm), the target occupies at least a quarter of the camera's field of view during imaging. Compared to small-target segmentation tasks, large-target segmentation of the cutter holder requires less data complexity. Consequently, a dataset of 100 images suffices for this task, split into 80 training and 20 testing samples.

Image augmentation

In the shield construction site, although the high-pressure water gun is used to clean the surface of the cutter holder, there will still be rust, soil cover, and other conditions. The illumination in the shield machine is unstable, and it is difficult to install a large light source due to the limited space in the disc cutter changing room, leading

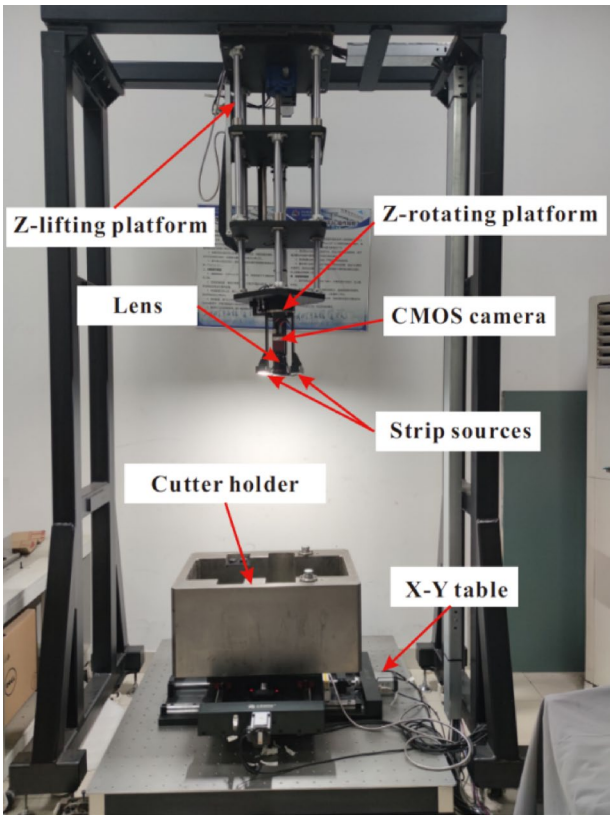

Fig. 1. The cutter holder image acquisition platform.

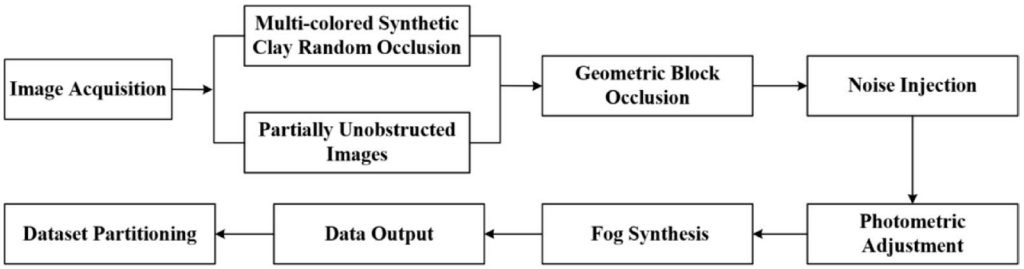

Fig. 2. The flowchart of image data processing.

to uneven brightness of the collected cutter holder images. The underground high-humidity environment will cause the camera lens to produce fog, coupled with the interference of air dust, resulting in blurred images of the collected cutter holder. All of the above will increase the difficulty of segmentation of the target cutter holder.

In order to simulate the on-site environment as realistically as possible, the data processing shown in Fig. 2 was carried out on the collected images. In the laboratory environment, we use the simulated soil to randomly block the surface of the cutter holder during image acquisition. After the acquisition, the collected image is further enhanced on the computer to approximate the image of the cutter holder collected in the real environment. The specific operation is as follows. Firstly, the surface of the cutter holder is randomly occluded using occlusion blocks such as rectangles or circles. Secondly, Gaussian white noise with a standard deviation of 0 to 50 distribution is randomly added to the images. Then, the brightness and contrast of the images are randomly adjusted. Finally, the central point synthetic fog method is used to randomly add different concentrations of fog to the cutter holder images. The partially occluded local images and the enhanced images of the obtained cutter holder data set are shown in Fig. 3.

Overall network structure

The UNet network implements the encoding and decoding process based on FCN and U-shaped structure. There is no fully connected layer in the whole network, which is composed of convolution and pooling layers. The encoding process, known as the down-sampling stage, performs feature extraction on the image. The up-

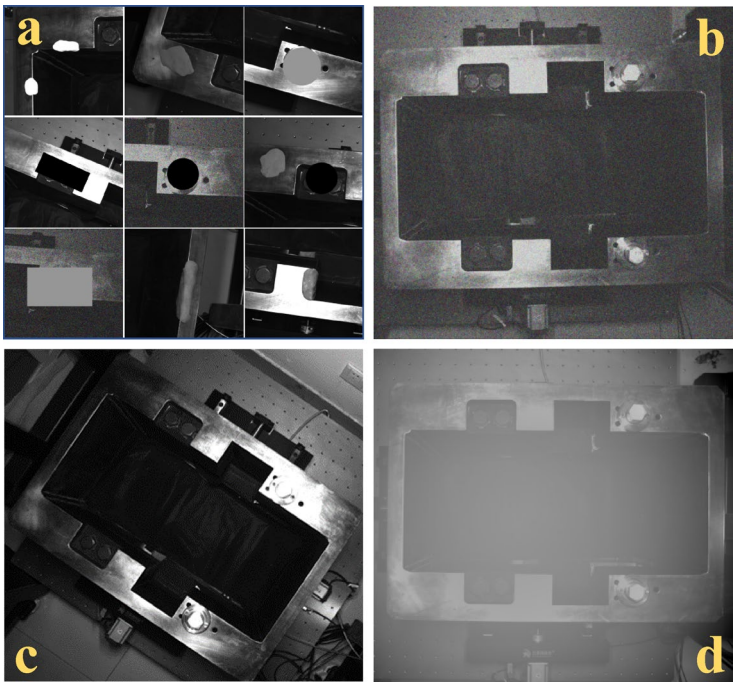

Fig. 3. The local screenshots of the samples of the dataset. Fig. a shows the local images of the cutter holder occluded by occluded blocks and simulated mud. Figs. b, c and d show the surface of the cutter holder with noise, uneven illumination distribution and fog occlusion, respectively.

sampling stage decoding process uses the 'skip connection' to transfer the features extracted from the down-sampling process to the up-sampling layer to achieve multi-scale information fusion, which can obtain more image detail features.

The multi-level downsampling structure of Unet can gradually extract the multi-scale features of cutter holder wear and soil coverage, and can resist texture blurring caused by low underground illumination. During decoding, jump connections are utilized to fuse shallow high-resolution features and deep semantic features, suppressing strong noise while ensuring the edge of the cutter holder. The encoding and decoding symmetric structure and skip connection mechanism of UNet essentially construct a parameter-efficient feature reuse system, which has strong adaptability to small samples and is suitable for tool changing scenarios with relatively simple environments. The u-shaped structure proposed in this paper is shown in Fig. 4, which mainly comprises three parts: the main feature extraction module, the attention module and the enhanced feature extraction module.

The main feature extraction module

The original UNet architecture predominantly employed VGG as its backbone. This study adopts ResNet, which enhances VGG's framework through residual connections while retaining its small-kernel convolutional layers. The residual blocks utilize skip connections to mitigate gradient vanishing in deep networks, enabling accuracy improvement through depth escalation. Crucially, ResNet achieves parameter efficiency and computational economy while enabling deeper architectures compared to VGG. Compared with Resnet34, Resnet101, and other networks, Resnet50 is more suitable for the segmentation task of cutter holder images, considering its comprehensive performance in segmentation speed, parameter number, and recognition ability.

The attention module

The human eye can quickly scan the global image to select the target area that needs to be focused on and focus on the area to obtain more target detail features while suppressing other useless information. The attention mechanism aims to determine which part of the input needs attention and allocate limited information processing resources to the critical part. The attention mechanism is generally divided into channel attention mechanism and spatial attention mechanism, and the combination of the two. In practice, the specific choice of which kind of attention needs to be considered comprehensively according to the specific application scenarios. Moreover, the attention module is a plug-and-play module that can be placed behind any feature layer.

Coordinate attention is an efficient mechanism that retains the important direction information generated while capturing channel information. It can also improve the sensitivity of the network to feature location recognition. The schematic diagram of the CA module is shown in Fig. 5, which mainly generates accurate position information coding for channel relationships and remote dependencies through coordinate information embedding and coordinate attention.

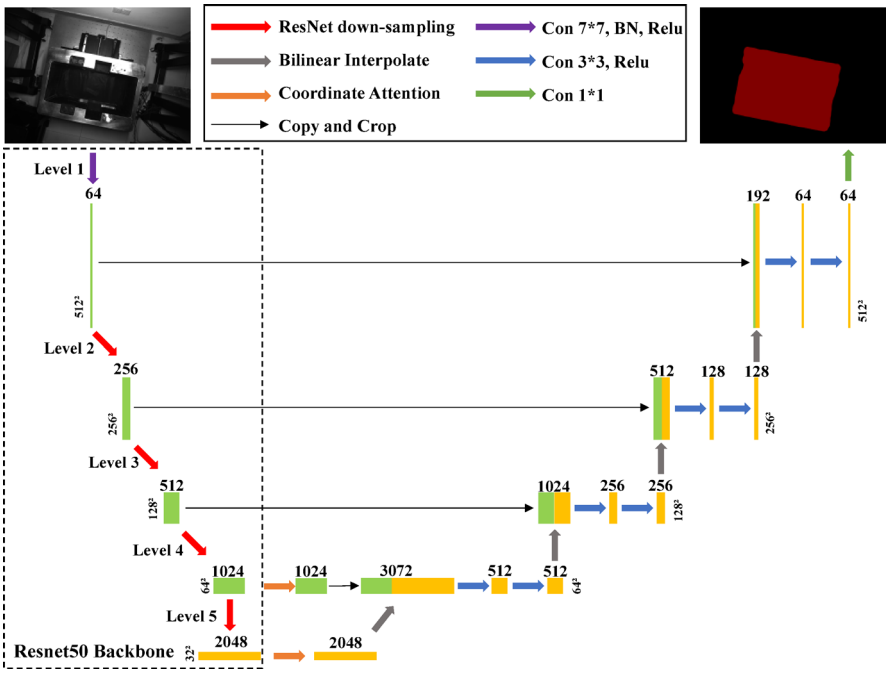

Fig. 4. Res-UNnet-CA network structure.

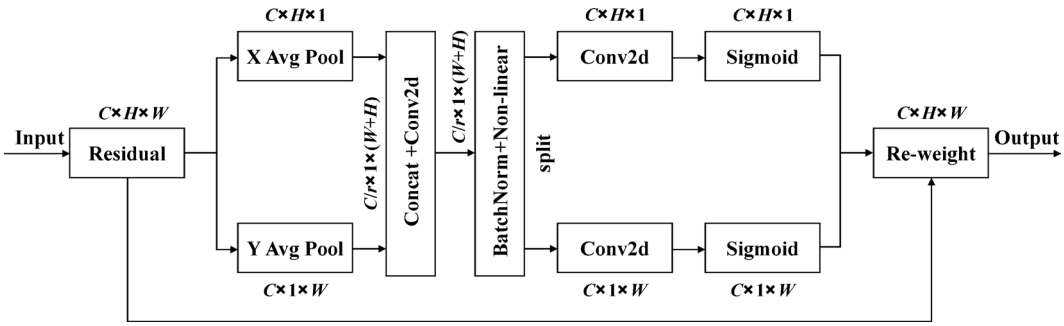

Fig. 5. The structure of the coordinate block. 'X Avg' and 'Y Avg' represent one-dimensional horizontal and one-dimensional vertical global pooling, respectively. $H \times W$ is the size of the feature map. r is the reduction ratio.

The enhanced feature extraction module

The right half of the network is the enhanced feature extraction module, the up-sampling part. It is mainly composed of four small modules. Each module follows the typical architecture of the convolutional network and contains a bilinear interpolation, which can amplify the feature layer and then fuse with the effective feature layer obtained by down-sampling. It also includes two 3×3 convolutions, each with a ReLU function. Through the down-sampling process, we obtained five effective feature layers and stacked them with the feature layers obtained by the enhanced feature extraction module to achieve feature fusion. Finally, a 1×1 convolution is used to map each 64-component feature vector to the foreground or background category.

Experiments and results

Experimental environment and evaluation indicators

The segmentation models used in this study are trained on a server with running 256G memory. It is implemented using the Python 3.7.0 programming language under the PyTorch 1.11.0 deep learning framework. The server has two NVIDIA RTX 3090 GPUs and two Intel Xeon Gold 6242 Processors @ 3.10 GHz. The operating system is Windows 10, and parallel computing is realized by CUDA 11.6.

In order to evaluate the effectiveness and accuracy of the cutter holder segmentation method proposed in this paper, the experiments evaluate the cutter holder segmentation results from accuracy, precision, recall, F1_score, IoU (intersection over union), mIoU (mean intersection over union) and other indicators. Each indicator is defined as follows.

Learning rate	0.01
Optimizer	SGD
Momentum	0.9
Loss function	Cross-entropy
Batch size	16
Number of epochs	100

Table 1. Hyperparameters used for the segmentation models.

Network	FCN	LR-ASPP	DeepLabV3	DeepLabV3+	PSP-DANet	SegFormer	Res-UNet
Training Time(h)	1.58	1.48	1.6	1.09	1.02	0.36	0.87

Table 2. Training time of the five segmentation networks.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

$$F1\_Score = \frac{2 \times \ Precision \times \ Recall}{Precision + Recall} \tag{4}$$

$$IoU = \sum_{i=0}^{k} \frac{p_{ii}}{\sum_{j=0}^{k} p_{ij} + \sum_{j=0}^{k} p_{ji} - p_{ii}} = \frac{TP}{TP + FN + FP} \tag{5}$$

$$MIoU = \frac{1}{k+1} \sum_{i=0}^{k} \frac{p_{ii}}{\sum_{j=0}^{k} p_{ij} + \sum_{j=0}^{k} p_{ji} - p_{ii}} \tag{6}$$

where *i* denotes the true value, *j* denotes the predicted value, and p_{ij} denotes the prediction of *i* to *j*.

Comparison of Res-UNet with other semantic segmentation models

Based on the cutter holder dataset, we compare the segmentation performance of the Res-UNet model with the main part of Resnet50 with several other commonly used models for semantic segmentation. These segmentation models are trained on the open-source framework PyTorch. The hyperparameters of all models are shown in Table 1. The model's results were saved and evaluated every ten epochs during training. Moreover, the latest model can also be loaded to resume training if there is a training termination.

Transfer learning serves as a widely adopted approach in deep learning to enhance model generalization without requiring extensive additional datasets. Standard segmentation models typically employ pre-trained networks and optimize all parameters post-weight initialization, accelerating convergence while boosting performance. The Res-UNet variant distinguishes itself by exclusively training later network layers during initial phases: the first 50 epochs freeze core feature extraction modules while fine-tuning subsequent layers, followed by full parameter optimization in the final 50 epochs. The average time of the segmentation models after multiple trainings on the cutter holder training set is shown in Table 2.

We use the above segmentation models to compare their accuracy, precision, recall, F1_score, IoU, and mIoU on the cutter holder test set. As shown in the experimental results of Fig. 6 and Tabel 6, the Res-UNet model performs best, and the values of its evaluation indicators are higher than other segmentation models. Compared with the second-performing DeepLabV3 + model, its accuracy is 99.07%, an increase of 4.08%. In addition to the similar precision value, the other indicators of Res-UNet have increased by more than 8% points. The integrated PSP-DANet framework employs PSPNet for swift coarse localization of the cutter holder, followed by DANet's sub-pixel precision refinement within ROI regions. This combined approach achieves optimal segmentation efficiency among compared methods, though with marginally lower accuracy than Res-UNet.

Comparison of different loss functions

Loss functions serve as critical optimization benchmarks in deep learning, where their minimization drives model convergence and prediction error reduction. The choice of loss functions significantly influences model performance, particularly when network architectures are fixed. Identifying the optimal loss function for guiding networks toward superior solutions demands systematic comparative analysis under defined structural parameters.

Pixel-level cross-entropy is the most commonly used loss function in image semantic segmentation tasks. Its specific expression is as follows.

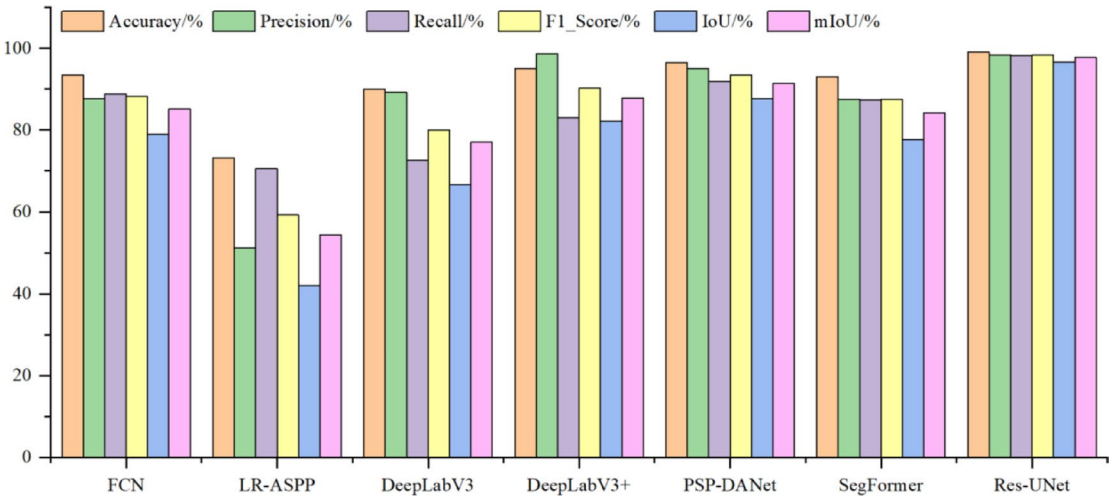

Fig. 6. Comparison of evaluation indicators of different semantic segmentation models.

$$L^{CE} = -\sum_{i=1}^{N} y_i \log p_{y_i'} \tag{7}$$

where N represents the number of pixels in the image, y_i and y_i' represent the label value and the predicted value, respectively, and $p_{y_i'}$ represents the probability of the predicted value.

The focal loss function is a new loss function proposed by He et al. for the imbalance of training samples and the difficulty of samples[34]. Its specific expression is as follows.

$$L^{Focal} = -\alpha\,(1 - p_{y'})^{\gamma} \log p_{y'} \tag{8}$$

where α is used to weight the loss of the sample of different categories, the effect of parameter γ is that when the probability of sample prediction is large, the loss of easy-to-classify samples will be significantly reduced.

The dice loss function can measure the similarity between the predicted and the real segmented images, thereby improving the segmentation effect. Generally, the use of dice loss alone cannot achieve good results, and we generally use it in combination with cross-entropy loss or focal loss in practice. The specific expression of the dice loss is as follows.

$$L^{DC} = 1 - \frac{2 \times |y \cap y'| + smooth}{|y| + |y'| + smooth} \tag{9}$$

where y and y' represent the label and predicted values, respectively.

Based on the cutter holder data set, we compare the segmentation results of the Res-UNet model with different loss functions through experiments to select the most suitable loss function. We compared the effects of CE loss, focal loss, dice loss and a total of five loss functions combining dice loss with the first two loss functions on network training. In the training process, the settings of other hyperparameters except the loss function are the same as those in Table 1.

Figure 7 shows the loss function smoothing, accuracy, and mIoU curves obtained in the model training stage when different loss functions are used. The effect of the loss function is compared by accuracy and mIoU values and their convergence smoothness. It can be seen from these three sets of curves that when the loss function is focal, the convergence speed of the loss function is the fastest, but the segmentation accuracy is the worst. The combination of dice loss and ce loss has the best effect, and its accuracy and mIoU curve convergence is the fastest, and the obtained values of the two are also the highest. The abrupt curve shifts around epoch 50 originate from full-parameter optimization activation in the latter training phase, reflecting expected behavioral transitions during model fine-tuning. The Res-UNet model with the combination of dice loss and ce loss as the loss function uses the cutter holder test set to obtain the results of accuracy, precision, recall, F1_score, IoU, and mIoU values through experiments, and the results are 99.23%, 98.56%, 98.84%, 98.7%, 97.27%, and 98.11%, respectively (as shown in Table 4). Compared with the evaluation index results of the Res-UNet model with ce loss as the loss function in Fig. 5, they were increased by 0.16%, 0.17%, 0.59%, 0.38%, 0.57%, and 0.4%, respectively.

Comparison of different optimizers

Optimizers in deep learning backpropagation direct loss function parameters toward global minima by updating along the gradient's steepest descent path, forming the foundation of gradient-based optimization. While stochastic gradient descent (SGD) accelerates training through rapid convergence toward local extrema, its performance suffers from oscillations along steep dimensions and sluggish progress in flat regions. The Adam algorithm addresses these limitations by integrating first-order momentum to dampen oscillations and second-

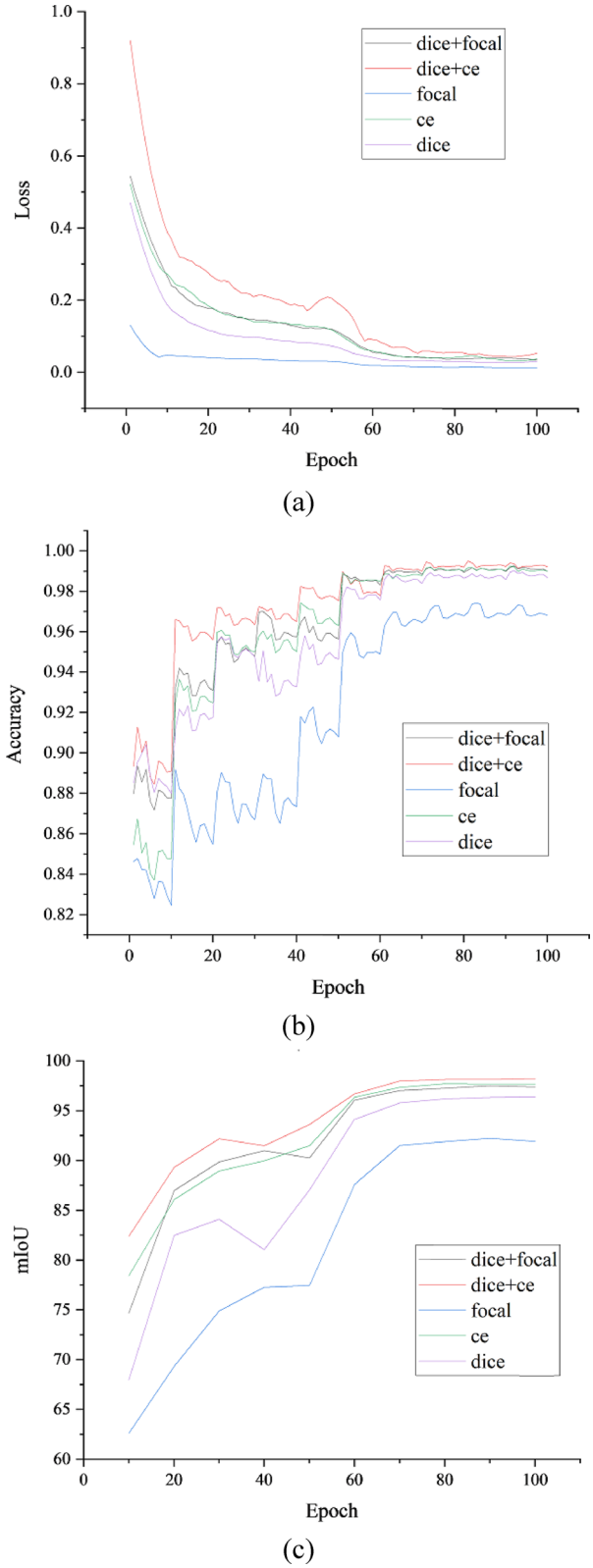

Fig. 7. Loss (**a**), accuracy (**b**) and mIoU (**c**) curves using different loss functions.

order momentum for adaptive learning rate adjustment, achieving superior computational efficiency and faster convergence compared to SGD through gradient-aware parameter updates.

Based on the cutter holder data set, we use the Res-UNet model to replace the SGD optimizer with the Adam optimizer for training and testing experiments, and the obtained comparison results of loss functions are shown

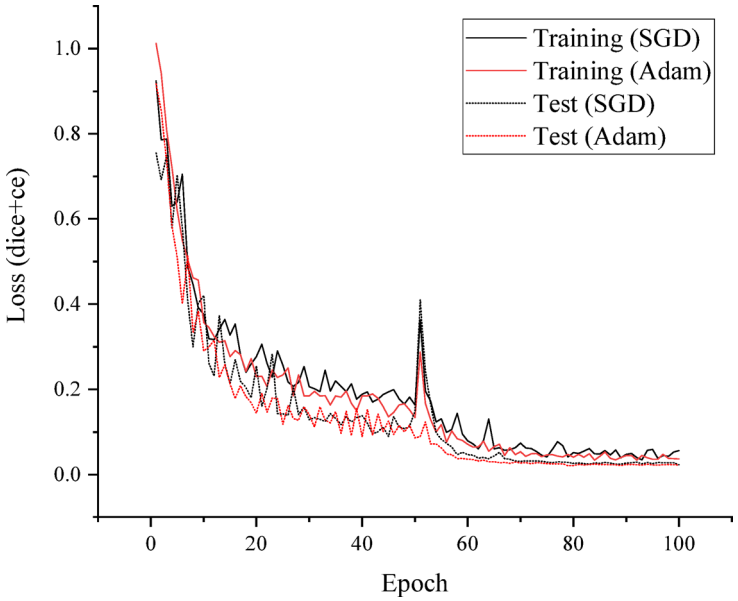

Fig. 8. Loss curves using different optimizers.

Model	Optimizer	Accuracy	Precision	Recall	F1_Score	IoU	mIoU
Res-UNet	SGD	0.9923	0.9856	0.9884	0.9870	0.9727	0.9811
	Adam	0.9927	0.9863	0.9874	0.9868	0.9740	0.9820

Table 3. Comparison of segmentation accuracy using different optimizers.

Models	Batch size	Training epochs	Trainable params	Params size (MB)	Optimizer	Initial learning rate
Res-UNet	16	100	43,859,010	167.31	Adam	0.0001
Res-UNet-CBAM	16	100	45,169,926	172.31	Adam	0.0001
Res-UNet-ECA	16	100	43,859,022	167.31	Adam	0.0001
Res-UNet-CA	16	100	44,842,434	171.06	Adam	0.0001

Table 4. The parameters assigned to each deep learning model.

in Fig. 8. When the Adam optimizer is used, the convergence of the loss function is faster and more stable, and the loss values are smaller for network training and testing.

As shown in Table 3, after replacing the optimizer of the Res-UNet model, each segmentation evaluation index of the network has also been improved. In the subsequent experiments, we use the Adam method to optimize the loss function to verify the performance of the improved model.

Comparison of different attention mechanisms

In order to study the effectiveness of the coordinate attention added in this paper, we add several different attention mechanisms to the Res-UNet network for comparison. After adding different attention mechanisms, the parameters of each model are shown in Table 4. After adding the ECA module, the number of parameters of the Res-UNet model increases the least, and the calculation amount of the network increases less. Compared with the Res-UNet-CBAM model, the Res-UNet-CA model has a relatively small increase in the number of parameters. To minimize the increase in parameters, we incorporated the attention mechanism only in the Level 4 and Level 5 skip connection layers of the downsampling path. The target cutter holder occupies a substantial portion of the image due to its considerable physical size. Incorporating an attention mechanism at the base of the UNet network effectively captures the target's global features. Introducing attention mechanisms across all connection layers significantly increases computational overhead while providing negligible improvement in segmentation accuracy.

Based on the cutter holder data set, the models with different attention mechanisms are compared through experiments, and the increments of each segmentation evaluation indicator of each model after adding attention are shown in Table 5. The results show that adding attention mechanisms to the Res-UNet model can improve network performance, and the Res-UNet-CA model has the best performance.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1_Score (%)	IoU (%)	mIoU (%)
Res-Unet-CBAM	0.11	0.07	0.12	0.09	0.20	0.13
Res-Unet-ECA	0.03	0.15	−0.03	0.05	0.12	0.08
Res-Unet-CA	**0.18**	**0.27**	**0.37**	**0.31**	**0.63**	**0.43**

Table 5. Incremental comparison of each segmentation accuracy index of each model after adding different attention. Bold face indicates the best performance.

Model	Accuracy	Precision	Recall	F1_Score	IoU	mIoU
FCN	0.9346	0.8771	0.8883	0.8827	0.7900	0.8516
LR-ASPP	0.7315	0.5113	0.7050	0.5927	0.4212	0.5437
DeepLabV3	0.8997	0.8922	0.7256	0.8003	0.6672	0.7708
DeepLabV3+	0.9499	0.9871	0.8303	0.9019	0.8213	0.8782
PSP-DANet	0.9642	0.9503	0.9188	0.9343	0.8767	0.9143
SegFormer	0.9302	0.8758	0.8733	0.8746	0.7771	0.8424
Res-UNet-CA	**0.9945**	**0.9890**	**0.9911**	**0.990**	**0.9803**	**0.9863**

Table 6. Comparison of segmentation evaluation indexes between the Res-UNet-CA and other segmentation methods. Bold face indicates the best performance.

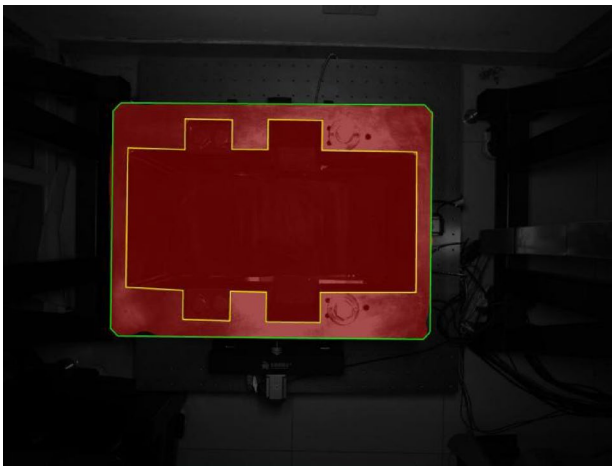

Fig. 9. The segmentation effect of Res-UNet-CA model.

The segmentation prediction effect of the Res-UNet-CA model

The proposed Res-UNet-CA method and other mainstream segmentation approaches were comparatively evaluated on the cutter holder test set, with their segmentation performance metrics summarized in Table 6. The results show that the segmentation performance of the Res-UNet-CA model is significantly better than that of other segmentation models.

The effect of using the Res-UNet-CA model to segment the cutter holder is shown in Fig. 9. The red area represents the segmented cutter holder, the green polygon represents the actual outer contour of the marked cutter holder, and the yellow frame line represents the inner contour feature of the cutter holder. Although the segmented outer contour of the cutter holder cannot coincide with the actual outer contour to a certain extent, the segmentation results can always ensure the integrity of the inner contour of the cutter holder so that the disc cutter changing robot can accurately locate the cutter holder according to the inner contour feature. Therefore, the Res-UNet-CA model's segmentation effect can meet our engineering needs.

A random subset of collected cutter holder images underwent predictive processing to visually compare segmentation performance across different models, as illustrated in Fig. 10. It can be seen from the figure that the segmentation effect of Res-Unet-CA is significantly better than that of several other models, especially the edge segmentation effect, making the obtained cutter holder boundary contour more precise and complete.

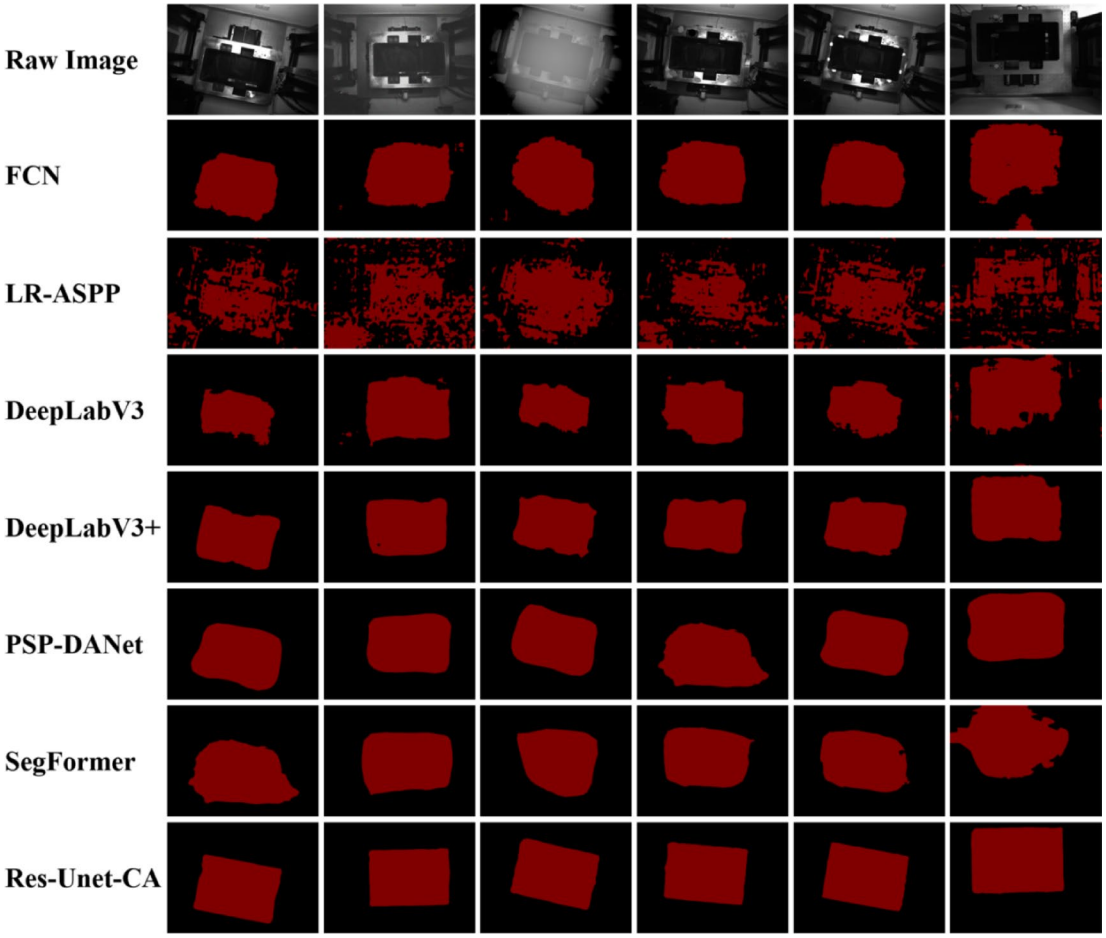

Fig. 10. Visual comparison of prediction results of different semantic segmentation models.

Conclusion

This study presents a segmentation framework for the disc cutter holder of the shield machine based on deep learning. Building upon the U-Net architecture, the proposed method incorporates ResNet50's residual units in place of standard convolutional blocks, facilitating multi-scale hierarchical feature learning while establishing direct propagation pathways from shallow to deep layers through skip connections with identity mapping. We compare the Res-UNet model with several commonly used semantic segmentation models on the cutter holder dataset. Experimental results demonstrate that the Res-UNet architecture outperforms benchmarked state-of-the-art models in segmentation metrics while exhibiting dual advantages in computational efficiency, achieving the most optimal balance between accuracy and training time expenditure among compared methods.

In order to improve the accuracy of cutter holder segmentation, several loss functions and two optimizers are compared through experiments. The results show that when the loss function selects the mixed loss function combined with dice loss and ce loss, and the optimizer selects Adam, the segmentation evaluation indicators of the network have the highest value and the best effect.

Attention modules are integrated at the network's lower layers to enhance segmentation performance by enabling autonomous focus on critical image features. Based on the cutter holder data set, the influence of different attention mechanisms on the network is compared, and it is found that the addition of coordinate attention has the most significant improvement in network performance. The coordinate attention mechanism enhances the network's positional awareness of the cutter holder by preserving directional cues during channel dependency encoding. From the prediction results of each model on the cutter holder images, the segmentation effect of Res-UNet-CA is better than that of other segmentation models. Moreover, the segmented cutter holder image contains the complete internal contour features on its surface, which meets the actual engineering requirements of our subsequent positioning work of the disc cutter changing.

The proposed method achieves high segmentation accuracy in controlled laboratory environments. However, its performance in real-world construction site scenarios requires further validation due to significant environmental complexities. Future work will focus on collecting on-site image data to enhance the model's generalizability under practical operating conditions.

Data availability

The data that support the findings of this study are available from the author, Dandan Peng, d201880277@hust.edu.cn upon reasonable request.

Received: 25 February 2025; Accepted: 4 July 2025
Published online: 05 July 2025

References

1. Zhao, S., Zhang, D., Xue, Y., Zhou, M. & Huang, H. A deep learning-based approach for refined crack evaluation from shield tunnel lining images. *Autom. Constr.* **132**, 103934 (2021).
2. Zhang, D., Fu, L., Huang, H., Wu, H. & Li, G. Deep learning-based automatic detection of muck types for Earth pressure balance shield tunneling in soft ground. *Computer-Aided Civ. Infrastruct. Eng.* **38**, 940–955 (2023).
3. Shelhamer, E., Long, J. & Darrell, T. Fully convolutional networks for semantic segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **39**, 640–651 (2017).
4. Ronneberger, O., Fischer, P. & Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. Preprint at (2015). https://doi.org/10.48550/arXiv.1505.04597
5. Chen, L. C., Papandreou, G., Schroff, F. & Adam, H. Rethinking Atrous Convolution for Semantic Image Segmentation. Preprint at (2017). https://doi.org/10.48550/arXiv.1706.05587
6. Chen, L. C., Zhu, Y., Papandreou, G., Schroff, F. & Adam, H. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. Preprint at (2018). https://doi.org/10.48550/arXiv.1802.02611
7. Howard, A. et al. Searching for MobileNetV3. in. *IEEE/CVF International Conference on Computer Vision (ICCV)* 1314–1324 (2019). (2019). https://doi.org/10.1109/ICCV.2019.00140
8. Zhao, H., Shi, J., Qi, X., Wang, X. & Jia, J. Pyramid Scene Parsing Network. in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 6230–6239 (2017). 6230–6239 (2017). (2017). https://doi.org/10.1109/CVPR.2017.660
9. Fu, J. et al. Dual Attention Network for Scene Segmentation. in. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* 3141–3149 (2019). (2019). https://doi.org/10.1109/CVPR.2019.00326
10. Xie, E. et al. SegFormer: simple and efficient design for semantic segmentation with Transformers. In *ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS 34 (NEURIPS 2021)* Vol. 34 (eds Ranzato, M., Beygelzimer, A., Dauphin, Y., Liang, P. S., Vaughan, J. W. et al.) (nips), 2021).
11. Yang, T., Song, J., Li, L. & Tang, Q. Improving brain tumor segmentation on MRI based on the deep U-net and residual units. *J. X-Ray Sci. Technol.* **28**, 95–110 (2020).
12. Zhao, J., Wu, Y., Zhang, Q. & Liao, J. Two-Stage channel Estimation for MmWave massive MIMO systems based on ResNet-UNet. *IEEE Syst. J.* **17**, 4291–4300 (2023).
13. Xu, J. J. et al. Automatic soil crack recognition under uneven illumination condition with the application of artificial intelligence. *Eng. Geol.* **296**, 106495 (2022).
14. Feng, S. et al. Wear debris segmentation of reflection ferrograms using lightweight residual U-Net. *IEEE Trans. Instrum. Meas.* **70**, 1–11 (2021).
15. Li, S. et al. MF-SRCDNet: Multi-feature fusion super-resolution Building change detection framework for multi-sensor high-resolution remote sensing imagery. *Int. J. Appl. Earth Obs. Geoinf.* **119**, 103303 (2023).
16. Wang, Y., Liang, B. & Feng, S. Coronal loop detection using multiscale convolutional neural networks. *ApJS* **270**, 4 (2023).
17. Liu, Y., Guo, Z., Guo, H. & Xiao, H. Learning to colorize near-infrared images with limited data. *Neural Comput. Applic.* **35**, 19865–19884 (2023).
18. Fan, J. et al. Macerals particle characteristics analysis of tar-rich coal in Northern Shaanxi based on image segmentation models via the U-Net variants and image feature extraction. *Fuel* **341**, 127757 (2023).
19. Yuan, Q., He, X., Han, X. & Guo, H. Automatic recognition of Craquelure and paint loss on polychrome paintings of the palace museum using improved U-Net. *Herit. Sci.* **11**, 65 (2023).
20. Fan, Y., Ding, X., Wu, J., Ge, J. & Li, Y. High spatial-resolution classification of urban surfaces using a deep learning method. *Build. Environ.* **200**, 107949 (2021).
21. Cao, K. & Zhang, X. An improved Res-UNet model for tree species classification using airborne High-Resolution images. *Remote Sens.* **12**, 1128 (2020).
22. Fan, Z. et al. ResAt-UNet: A U-Shaped network using ResNet and attention module for image segmentation of urban buildings. *IEEE J. Sel. Top. Appl. Earth Observations Remote Sens.* **16**, 2094–2111 (2023).
23. Huang, A., Jiang, L., Zhang, J. & Wang, Q. Attention-VGG16-UNet: a novel deep learning approach for automatic segmentation of the median nerve in ultrasound images. *Quant. Imaging Med. Surg.* **12**, 3138150–3133150 (2022).
24. Chen, X., Zhang, K., Wang, W., Hu, K. & Xu, Y. Intelligent identification of tunnel water leakage based on super-resolution reconstruction and triple attention. *Measurement* **225**, 114009 (2024).
25. Hu, J., Shen, L., Albanie, S., Sun, G. & Wu, E. Squeeze-and-Excitation networks. Preprint at (2019). https://doi.org/10.48550/arXiv.1709.01507
26. Yu, H., Men, Z., Bi, C. & Liu, H. Research on field soybean weed identification based on an improved UNet model combined with a channel attention mechanism. *Front Plant. Sci* **13**, (2022).
27. Woo, S., Park, J., Lee, J. Y. & Kweon, I. S. CBAM: Convolutional Block Attention Module. Preprint at (2018). https://doi.org/10.48550/arXiv.1807.06521
28. Li, W. et al. UNet combined with attention mechanism method for extracting flood submerged range. *IEEE J. Sel. Top. Appl. Earth Observations Remote Sens.* **15**, 6588–6597 (2022).
29. Wang, Q. et al. ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks. Preprint at (2020). https://doi.org/10.48550/arXiv.1910.03151
30. Li, W. et al. Cross-Scene Building identification based on Dual-Stream neural network and efficient channel attention mechanism. *IEEE J. Sel. Top. Appl. Earth Observations Remote Sens.* **17**, 6920–6932 (2024).
31. Li, J. et al. Eres-UNet++: liver CT image segmentation based on high-efficiency channel attention and Res-UNet++. *Comput. Biol. Med.* **158**, 106501 (2023).
32. Hou, Q., Zhou, D. & Feng, J. Coordinate Attention for Efficient Mobile Network Design. Preprint at (2021). https://doi.org/10.48550/arXiv.2103.02907
33. Singh, S. A., Kumar, A. S. & Desai, K. A. Vision-based system for automated image dataset labelling and dimension measurements on shop floor. *Measurement* **216**, 112980 (2023).
34. Lin, T. Y., Goyal, P., Girshick, R., He, K. & Dollár, P. Focal Loss for Dense Object Detection. Preprint at (2018). https://doi.org/10.48550/arXiv.1708.02002

Acknowledgements

The work presented in this paper was financially supported by the National Key Research and Development

Program of China (Grant No. 2022YFC3802300).

Author contributions

D.P. performed the experimentation and initial writing. G.Z. and D.P. looked after the whole research work and did the complete review of the manuscript. Z.X. prepared all visualization and also performed data preparations. All authors reviewed the manuscript.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to G.Z.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.