Hindawi
Advances in Civil Engineering
Volume 2024, Article ID 8810092, 13 pages
https://doi.org/10.1155/2024/8810092

Research Article

Research on Intelligent Detection and Segmentation of Rock Joints Based on Deep Learning

Lei Peng,[1] Haibo Wang ,[2] Chun Zhou,[1] Feng Hu,[1] Xiaoyang Tian,[1] and Zhu Hongtai[1]

[1]*Railway No. 7 Engineering Group Co., Ltd. (Guangzhou Engineering Company), Guangzhou 510760, China*
[2]*School of Aeronautics and Astronautics, Sun Yat-Sen University, Guangzhou, China*

Correspondence should be addressed to Haibo Wang; wanghb63@mail2.sysu.edu.cn

Received 24 May 2023; Revised 25 November 2023; Accepted 1 April 2024; Published 20 May 2024

Academic Editor: Camelia Delcea

The current methods for detecting joints on tunnel face rely primarily on manual sketches, which are associated with issues of low detection efficiency and subjectivity. To address these concerns, this paper presents an intelligent recognition and segmentation algorithm based on Mask R-CNN (mask region-based convolutional neural network) for detecting joint targets on tunnel face images and automatically segmenting them, thereby improving detection efficiency and objectivity of the results. Additionally, to tackle the challenge of low detection accuracy in existing image processing methods, particularly for complex tunnel joint surfaces in dark environments, the paper introduces a path aggregation network (PANet) to enhance the fusion capability of feature information in Mask R-CNN, thereby improving the accuracy of the intelligent detection method. The algorithm was trained on a dataset of 800 tunnel face images, and the research findings demonstrate that it can quickly detect the position of joints on tunnel face images and assign masks to the joint pixel regions to achieve joint segmentation. The mean average precision (mAP) of the detection boxes and segmentation in the 80 test set images were 58.0% and 49.2%, respectively, which outperforms the original Mask R-CNN algorithm and other intelligent recognition and segmentation algorithms.

1. Introduction

Amid the rapid evolution of infrastructure construction, the field of tunnel and underground engineering has witnessed substantial growth opportunities. Tunnels, as subterranean engineering structures, face intricate construction environments characterized by challenges such as fractured rock formations, elevated temperatures, and the presence of underground water. These factors collectively contribute to diminished construction efficiency and heightened safety risks. Therefore, the accurate delineation of the tunnel face holds paramount importance. The description of the tunnel face represents a crucial yet intricate undertaking. This process serves as the foundation for assessing the quality of the surrounding rock within the tunnel. Such an evaluation provides a robust basis for making informed decisions during tunnel construction. The provision of an accurate and efficient tunnel face description can significantly expedite the progress of tunnel construction endeavors, while simultaneously ensuring the safety of the entire operation. Tunnel face rock joints are geological structural features, including fissures, cracks, and faults, found on the rock surface exposed during tunnel excavation. These joint characteristics serve as indicators of the overall integrity of the rock mass at the tunnel face. Therefore, it is of great significance to detect joints on the tunnel faces accurately for tunnel construction.

The conventional approach to joint detection on tunnel faces heavily relies on manual observations and descriptions, which are notably inefficient. In the wake of rapid advancements in science and technology, computer-based image processing methods have emerged as valuable tools for identifying geological features in fractured rock masses. For instance, Fam and Hu [1–4] have employed digital image processing techniques, including edge detection, threshold segmentation, and the Hough transform, to discern surface cracks on rock masses. These methodologies have been packaged into software solutions for the benefit of engineers. However, a drawback of this method lies in its intricate and nonintuitive processing procedures. Digital image processing techniques have also found application in the realm of joint identification within tunnel faces [3]. Ye et al. [4–6] effectively processed tunnel face images to extract distinct joint profiles, effectively supplanting traditional geological sketches. In a similar vein, Li et al. [7, 8] directly extracted joint information from tunnel face images using structural plane processing software such as SIR6.0 and the OpenCV platform. Nevertheless, the challenging tunnel environment often leads to image quality degradation due to dust and other factors. To mitigate these challenges, Zhou et al. [9] harnessed infrared photography technology to obtain clear tunnel face images. Subsequent to image capture, digital image processing technology was employed to filter, equalize, and binarize these images, with joints represented as straight lines in this approach. Wang et al. [10] ventured into the combination of digital photogrammetry and image processing technology to investigate the geometric properties of rock mass structural planes, albeit acknowledging the need for efficiency and accuracy improvements. Further enhancing the precision of tunnel face analysis, Yang et al. [11] employed pixel-scale digital image technology, offering a more accurate and flexible alternative to traditional sketching methods. In a different vein, Yang et al. [12] established survey stations and markers at the tunnel site and harnessed the ShapeMetrix3D imaging system to capture face images. Zhuang et al. [13] introduced a method that combines digital image processing and machine vision to obtain comprehensive information on the rock mass in tunnel excavation faces. Their technique involves generating a 3D model, which, when analyzed in tandem with surface structure line sketches, enables the quantitative characterization of rock mass joint features. Leng et al. [14] delved into the utilization of image processing technology for boundary extraction in tunnel excavation faces, providing a foundation for categorizing surrounding rock based on extracted parameters like joint group numbers, average crack spacing, and occurrence. While digital image processing technology has been instrumental in tunnel excavation face joint extraction, its applicability is contingent upon stringent conditions, often necessitating controlled lighting and clear joint targets. Furthermore, the abundance of thresholding segmentation algorithms in digital image processing technology can complicate field applications, requiring frequent switching and adjustments to attain optimal segmentation results. Additionally, the output of this technology often requires complex postprocessing due to joint targets being treated as integral wholes. As a result, there is an urgent need for the development of more adaptable and robust technical solutions capable of thriving in complex lighting environments akin to those encountered in tunnels and achieving pixel-level segmentation of joint targets [15–18].

The rapid advancement of artificial intelligence has ushered in a wave of intelligent recognition algorithms [19–22], which have been seamlessly integrated into the conventional civil engineering sector. This integration has ushered in a new era of intelligent development for engineering tasks.

The realm of computer vision comprises four fundamental image processing tasks: image classification, semantic segmentation, object detection, and instance segmentation [23]. Image classification enables the categorization of images, yet it falls short in pinpointing specific details related to tunnel excavating face joints. Semantic segmentation excels at segmenting all joints within an image but lacks the ability to differentiate between individual joints since they are treated as a unified entity. Object detection can identify joints pertaining to distinct individuals but does not offer segmentation capabilities. In contrast, instance segmentation, building upon object detection, provides the means to achieve segmentation of different joint targets. In practical applications at tunnel sites, the detection results are instrumental in computing critical parameters such as the count of joint groups and the spacing between joints. Hence, for the intelligent detection task concerning tunnel excavating face joints, it is imperative to detect and segment each joint target individually. Consequently, this study advocates the utilization of an instance segmentation algorithm, specifically Mask R-CNN, as opposed to a semantic segmentation algorithm, to efficiently detect and segment tunnel excavating face joint targets.

The current digital image processing methods employed for the identification of tunnel excavating face joints are afflicted by the issue of intricate fine-tuning. While they succeed in eliminating the subjective element associated with manual depiction, they still necessitate manual intervention and adjustment, resulting in a recognition and detection process that lacks true intelligence and is bound by certain constraints [24]. In a bid to transcend the constraints inherent in conventional digital image methods, this study advocates the adoption of Mask R-CNN, an instance segmentation algorithm. By doing so, it accomplishes the intelligent detection and segmentation of tunnel excavating face joints while simultaneously sidestepping the subjectivity of manual depiction and the limitations of established digital image processing methods [25, 26].

2. Intelligent Tunnel Face Recognition Method Based on Mask R-CNN

Mask R-CNN, an extension of the Faster R-CNN framework, seamlessly incorporates instance segmentation into the object detection process [27]. Its framework is shown in Figure 1. It operates through three pivotal steps: first, it extracts features from the input image using a deep neural network; second, it employs a region proposal network (RPN) to suggest candidate object regions; finally, it further refines these regions through ROIAlign and generates precise binary masks for each detected object. This approach not only enables Mask R-CNN to identify objects in an image but also provides pixel-level segmentation masks, rendering it a highly valuable tool for tasks like joint recognition in tunnel excavating face images.

2.1. Main Network. The main network comprises two crucial components: Resnet101 and feature pyramid network (FPN). Resnet101 is structured with Conv blocks and Identity blocks, each serving distinct roles in the network. Conv blocks are responsible for performing convolution operations, enhancing the depth of the network to extract critical feature information from tunnel face images. This critical feature information encompasses the vital characteristics of

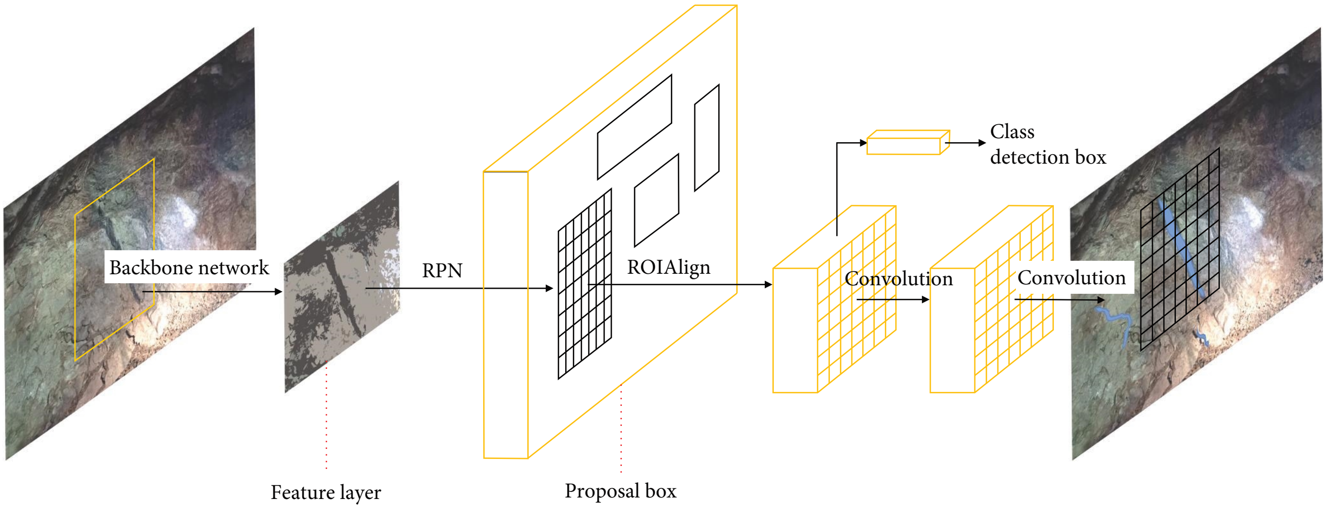


Figure 1: The network framework of Mask R-CNN algorithm.

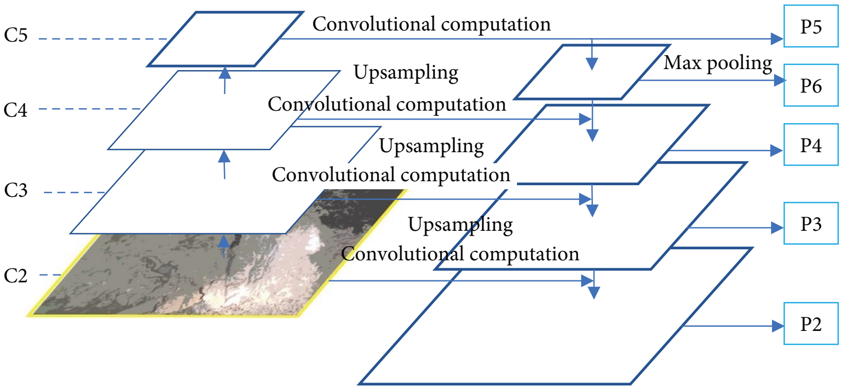

C5
C4
C3
C2
Convolutional computation
Upsampling
Convolutional computation
Upsampling
Convolutional computation
Upsampling
Convolutional computation
Max pooling
P5
P6
P4
P3
P2

Figure 2: FPN network framework.

joints in tunnel face images, including pixel values, dimensions, and shapes. Utilizing pixel data, multiple feature maps of varying sizes are computed and generated. Multiple feature maps of varying sizes refer to a set of distinct two-dimensional arrays, each capturing different spatial resolutions and patterns within an image. These feature map pixel points can be viewed as secondary pixels, offering a representation of the original image's pixel characteristics. These feature maps are denoted as C2–C5, with C1 being excluded from further processing due to its limited receptive field and semantic information, rendering it ineffective at capturing the object structure and features within images. This approach facilitates the detection of objects of various sizes. The study introduces a targeted improvement in the design of feature map sizes, tailoring them to the dimensions of palm images, in order to better align with the requirements of palm image joint recognition tasks.

The FPN excels at amalgamating feature information from various levels into a multiscale and multisemantics feature pyramid, as depicted in Figure 2. The "latlayer" operation involves sampling the high-resolution feature map to match the spatial scale of the low-resolution feature map and subsequently adding them together to yield the fused feature map. In the context of Mask R-CNN, FPN's output consists of characteristic maps denoted as P2, P3, P4, P5, and P6, which are effective layers utilized in the detection of objects of diverse scales.

2.2. Region Proposal Network. The region proposal network (RPN), a lightweight neural network, employs a sliding window approach to systematically scan the input image and pinpoint regions of interest (ROIs) that have the potential to contain target objects [28]. These ROIs are often referred to as "anchors" and are designed in various sizes. To extract joint profiles from images of the tunnel excavating face, this study leverages face label files to compute joint size information within the existing facial images. It then analyzes the dimensions and aspect ratios of joint label boxes, allowing for specific adjustments to be made in the design of anchor sizes and aspect ratios. These adjustments are geared towards achieving a closer alignment with the shape of the facial joint, thereby enhancing the accuracy of joint detection.

In the RPN training algorithm, the intersection-over-union (*IOU*) ratio is calculated between each anchor and the searching box (i.e., the ratio of the intersection area between the two boxes to the union area) to determine the anchor category. If the *IOU* is greater than the set threshold, the anchor is considered positive; if it is less than the threshold, the anchor is negative. The RPN training process involves two critical tasks: anchor classification training and searching box position regression training. The training error L_R is expressed as follows:

$$L_R = L_{Rc} + L_{Rr}, \tag{1}$$

where L_{Rc} represents the error function of anchors when conducting the classification training and L_{Rr} represents the error function of anchors box when conducting regression training. They are expressed by the following formulas:

$$L_{Rc} = \frac{1}{N_{Rc}} \sum_i l_{rc}(p_i, p_i^*), \tag{2}$$

$$L_{Rr} = \lambda \frac{1}{N_{Rr}} \sum_i p_i^* l_{rr}(s_i, s_i^*), \tag{3}$$

where i represents the number of anchors; p_i represents the probability that the anchor at number i is predicted to be a positive class; p_i^* represents the true label value of the anchor at number i; when the anchor is a positive sample, $p_i^* = 1$, and when the anchor is a negative sample, $p_i^* = 0$; s_i is a vector containing four elements, which are the center coordinate, width, and height of the anchor box at number i, and s_i^* represents a four-dimensional vector containing the corresponding elements of the label box. N_{Rc} and N_R represent the batch data volume of RPN stage classification and regression training, λ is the super parameter, L_{Rc} is the cross entropy loss function, and the expression of the regression loss function L_{Rr} is as follows:

$$l_{Rr}(s_i, s_i^*) = R(s_i - s_i^*), \tag{4}$$

where R expression is as follows:

$$R(x) = \begin{cases} 0.5x^2 & |x| < 1 \\ |x| - 0.5 & |x| \geq 1 \end{cases}. \tag{5}$$

The definitions of the four-dimensional vector $\mathbf{s}_i$ and $\mathbf{s}_i^*$ are as follows:

$$\begin{cases} \mathbf{s}_i = (s_x, s_y, s_w, s_h), \mathbf{s}_i^* = (s_x^*, s_y^*, s_w^*, s_h^*) \\ s_x = (x - x_a)/w_a, s_y = (y - y_a)/h_a \\ s_w = \ln(w/w_a), s_h = \ln(h/h_a) \\ s_x^* = (x^* - x_a)/w_a, s_y^* = (y^* - y_a)/h_a \\ s_w^* = \ln(w^*/\omega_a), s_h^* = \ln(h^*/h_a) \end{cases}, \tag{6}$$

where $x, y, w,$ and h represent the center coordinates, width, and height of the box, respectively; $x, x_a,$ and x^* correspond to the prediction box, anchor box, and real box, respectively. This rules are also appliable to the $y, w,$ and h.

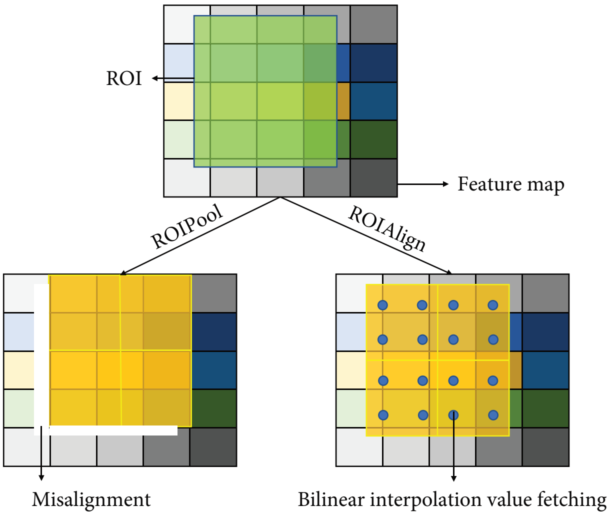


Figure 3: Comparison of ROIAlign and ROIPool.

After training, RPN can predict the type of anchor (positive or negative) and make preliminary adjustments to the position of the box. The improved regional suggestion network (RPN) attains the extraction of possible joint regions in the face feature image and generates a suggestion box as the input of the next stage.

2.3. ROIAlign Regulation. The region of interest is derived from the RPN in the form of a proposed bounding box. The anchor boxes initially generated by the RPN exhibit variations in size and are subsequently fine-tuned using a positional adjustment model. To ensure consistent sizing for classification purposes, the study employs ROIAlign. In contrast to the ROIPool module utilized in the Faster R-CNN algorithm [29], ROIAlign effectively addresses the issue of pixel displacement by incorporating bilinear interpolation values. This results in optimized box sizes and the preservation of a more comprehensive set of original information. For a detailed depiction of ROIAlign, refer to Figure 3.

The ROIAlign module improves ROI size adjustment, effectively preventing the loss of feature information and facilitating accurate recognition and segmentation of joint targets from images of tunnel excavating faces.

2.4. Result Prediction. The Mask R-CNN network further processes the image information provided by the adjusted ROI. The main tasks of the Mask R-CNN are briefly introduced as follows: (1) classification: for the dataset of tunnel excavating face, a specific label of ROI, namely "joint," is given, which differs from the two categories (positive and negative) in the RPN stage. (2) Fine-tuning of the prediction box position: based on the fine-tuning in the RPN stage, the position, length, and width of the prediction box are further adjusted to better fit the target. (3) Generate mask: the pixels belonging to the target object in the prediction box are identified and marked to form a mask. This helps to obtain fine recognition and segmentation of the joint targets from the images of the tunnel excavating face.

The error function L_{ROI} set in the training process at this stage is as follows:

$$L_{ROI} = L_{ROIc} + L_{ROIr} + L_{\text{mask}}, \tag{7}$$

where L_{mask} is the mask loss function is the average binary cross entropy loss function. The loss function L_{ROIc} of the classification L_{ROIc} is expressed as follows:

$$l_{ROIc} = -\ln(p_u), \tag{8}$$

where p_u represents the probability that it is categorized to u. In the above formula, L_{ROIr} is the loss function of the L_{ROIr} classification. It is expressed as follows:

$$l_{ROIr} = \lambda[\mu \geq 1] l_{Rr}, \tag{9}$$

where $[u \geq 1]$ means that when $u \geq 1$, the value is 1, otherwise it is 0.

To sum up, five error functions are set in the RPN stage and the result prediction stage, which are as follows:

$$L = L_{Rc} + L_{Rr} + (L_{ROIc} + L_{ROIr} + L_{\text{mask}}). \tag{10}$$

2.5. Intelligent Palm Recognition Method Based on Improved Mask R-CNN. Although the FPN can enhance feature extraction ability through feature fusion, the long calculation path from bottom to top, including a 101-layer network, is not conducive to feature information transmission, especially for diverse joint targets in complex tunnel environments. To address this issue, this study proposes the path aggregation network (PANet) [30].

The path aggregation network (PANet) enhances the feature hierarchy within a neural network by introducing an improved aggregation path. Unlike the original FPN, where features are transmitted from bottom to top through a multilayer network structure, potentially leading to lengthy information paths and information loss, PANet shortens this path. It achieves this by incorporating precise positional signals from lower levels, facilitating more effective integration and aggregation of feature information. As a result, PANet improves the network's capability to handle diverse targets, particularly in complex tunnel environments.

Figure 4 provides an overview of the path aggregation network structure. The red dashed line corresponds to the original FPN transmission path, where the bottom-up transmitted features traverse a multilayer network structure. The enhanced aggregation path method, depicted by the green dashed line, incorporates precise positional signals at lower levels to bolster the entirety of the feature hierarchy. This approach effectively shortens the information pathway between lower and upper level features, thereby mitigating information loss.

Building on the advantages of PANet, this study integrates it into FPN and proposes a PA-FPN network to improve the recognition and segmentation performance of the intelligent tunnel excavating face joint recognition model.

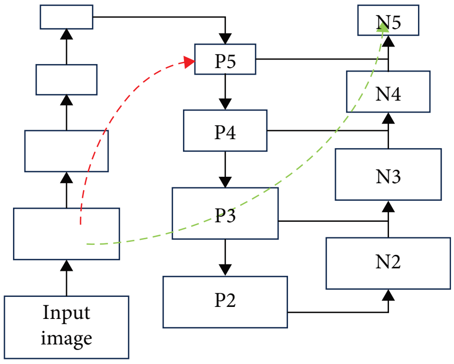


Figure 4: The framework of path aggregation network.

3. Image Collections of Tunnel Excavating Face

3.1. Image Collections. Regarding data collection, the acquisition of a more extensive dataset of original images proves highly beneficial when training deep learning models. These original images should authentically represent the diverse spectrum of tunnel face types commonly encountered in construction scenarios. This mandates that the data collection process spans tunnels characterized by varying rock strata, a wide range of lighting conditions, distinct geological compositions, and a multitude of construction conditions.

For this study, an extensive dataset of tunnel excavating face images, obtained from various tunnel construction sites in China, was meticulously compiled. After a rigorous selection process, a total of 400 original images were chosen, all standardized to a uniform size of 2,048 × 2,048 pixels. Figure 5 showcases some of the sample images from this dataset. Notably, these collected tunnel excavating face images exhibit a wide range of environmental conditions, including diverse angles, lighting scenarios, potential trolley interference, and variations in shadowing, among other factors. This diversity in environmental conditions substantially bolsters the robustness of the intelligent joint recognition algorithm proposed in this study.

In the realm of deep learning tasks, data augmentation plays a pivotal role in enhancing model performance. This technique involves various methods such as rotation, scaling, flipping, and brightness adjustments to artificially increase the diversity of the training dataset. By doing so, it enables the model to generalize more effectively and mitigates the risk of overfitting, as it learns from a broader spectrum of image variations. In the specific case of tunnel face images, where challenging working conditions and brief data capture opportunities are prevalent, data augmentation becomes even more crucial. It empowers the deep learning model to recognize and adapt to a wider array of real-world scenarios, ultimately bolstering its accuracy and resilience for tunnel construction applications.

Hence, in order to enhance the diversity of the acquired images from different tunnels, various image augmentation techniques are employed, including flipping, cropping, and brightness adjustments, as illustrated in Figure 6. Initially, we subjected the collected pool of 400 original images to a process of curation and elimination, resulting in the retention of

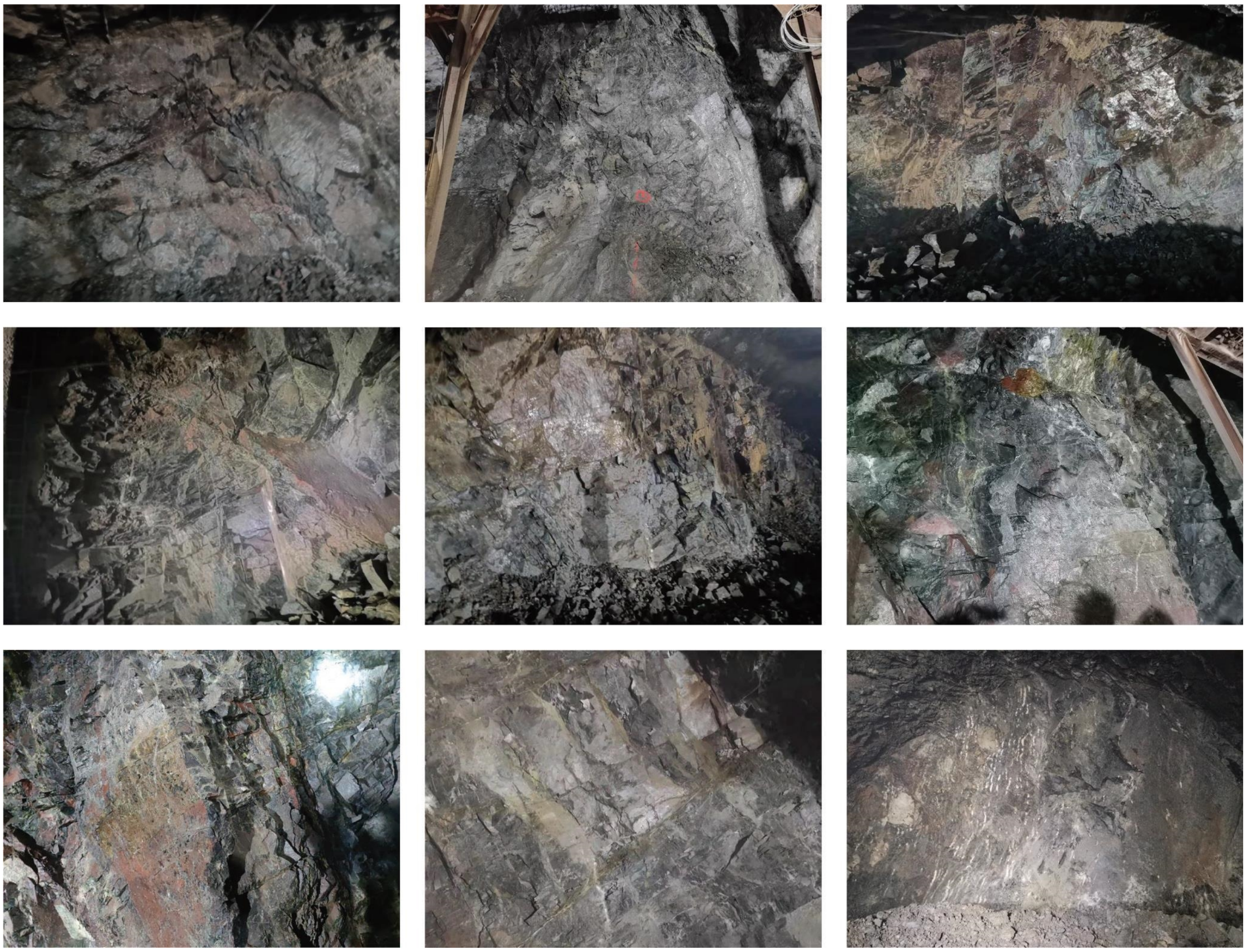


Figure 5: Typical image of tunnel excavating face.

200 high-quality tunnel face images. It is noteworthy that, at this stage, we refrained from immediately applying augmentation techniques such as flipping and lightening. Instead, we took a step-wise approach. The initial set of 200 images was proportionally split into training, validation, and test sets at a ratio of 7 : 2 : 1, yielding quantities of 140, 40, and 20 images, respectively. Subsequently, each of these subsets underwent augmentation processes, including flipping and lightening, expanding them to 560, 160, and 80 images, respectively. This strategic sequencing is aimed at fortifying the robustness of the trained model. By deferring the augmentation until after the initial partition, we mitigate the risk of the model encountering familiar images during the validation and testing phases, thus avoiding overfitting. Furthermore, the dataset is currently undergoing continuous updates and expansions, with a slight shortage noted in the quantity of high-quality images. Consequently, the morphological features of the joint targets in tunnel face images are not yet fully comprehensive.

3.2. Labels for Datasets. Accomplishing high-quality object annotation in tunnel face images is a formidable undertaking, particularly when the inherent features of these objects are not distinctly evident, thus augmenting the intricacy of the annotation process. Nevertheless, the caliber of annotated data holds paramount importance for the triumph of deep learning models. As the term "artificial intelligence" implies, this process essentially embodies the concept of being "half human, half intelligent," with exceptional human annotation serving as the cornerstone for crafting exceptional models. Consequently, it is vital to underscore the pivotal role of annotation quality. To ensure the attainment of superior annotation outcomes, it is imperative to employ stringent training and guidance, institute quality control and periodic assessments, involve multiple annotators, contemplate the integration of automation-assisted tools, and institute a continuous feedback and enhancement mechanism. These measures are instrumental in ensuring the precision, uniformity, and dependability of annotations, thereby heightening the efficacy of deep learning models.

Recognizing joints in tunnel excavating face images deviates from traditional image classification methods. Conventional joint segmentation techniques necessitate manual identification and marking of the target object in each image to produce labels. This meticulous process can consume approximately 10 min per image, ensuring the creation of

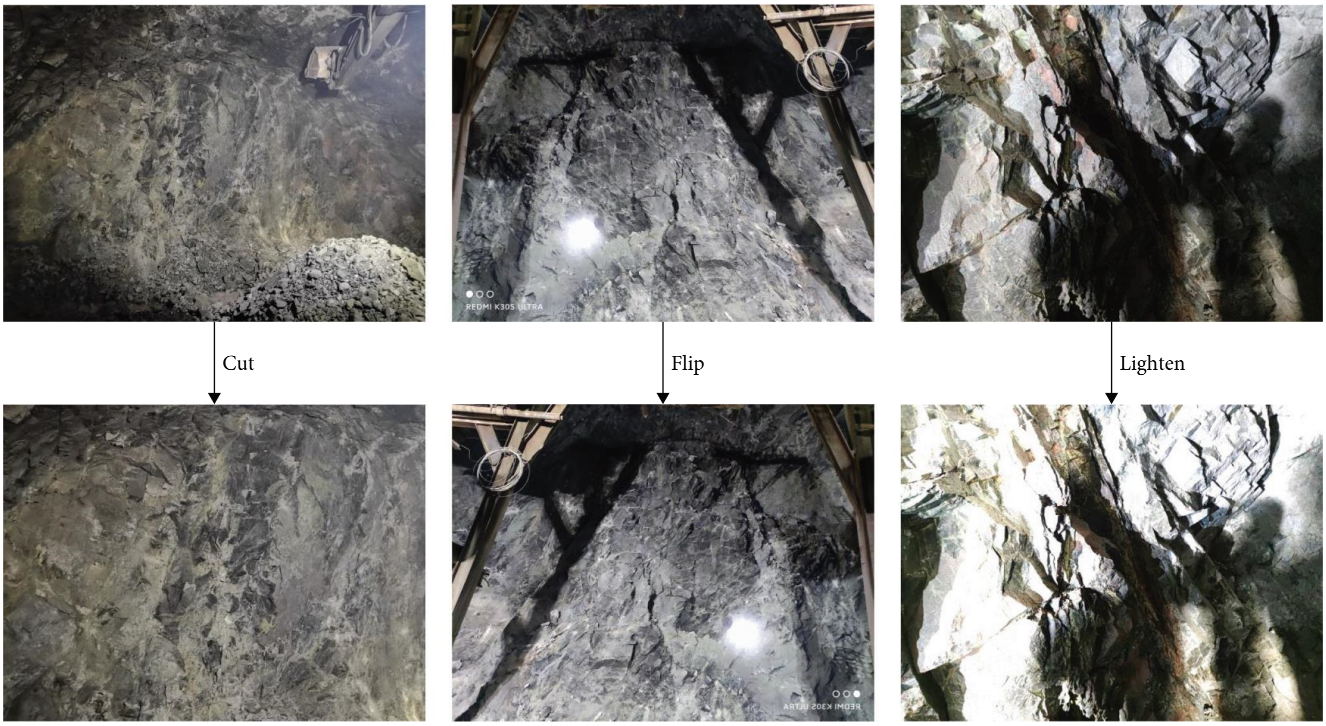
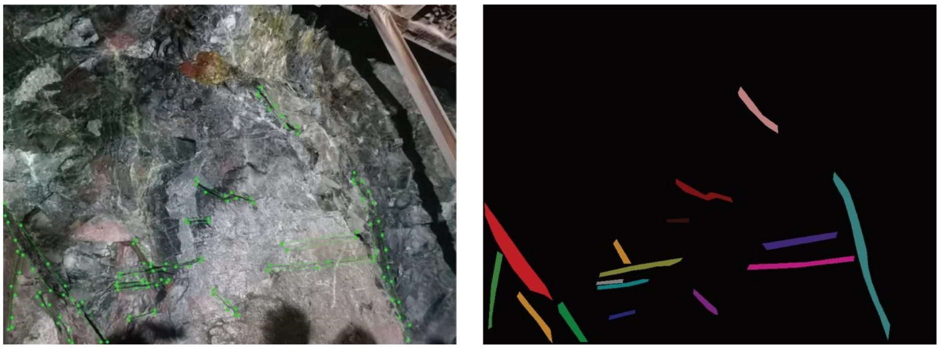


Figure 6: Data enhancement method.



(a)

(b)

Figure 7: Image marked and labeling method. (a) Picture annotation process. (b) Generate picture labels.

high-quality joint information. To streamline this task, our study leverages Labelme annotation software to annotate the joints present in tunnel excavating face images, as depicted in Figure 7(a). The annotated joints are depicted as polygonal contours, constructed from dots and lines. Once the annotation process for each image is complete, JSON files containing joint position and name details are generated. Subsequently, these files are transformed into the Coco dataset format utilizing built-in code, leading to the creation of trainable binary label images, as illustrated in Figure 7(b).

4. Instance Segmentation Experiment of the Tunnel Excavating Face Images

The process of intelligently recognizing and segmenting tunnel face joints can be broken down into two primary steps. First, the model for intelligent recognition and segmentation of tunnel excavating face joints is trained using sample data from the training and validation sets. Following this training phase, the performance of the model is assessed and evaluated using data from the test set. The implementation of the intelligent recognition and segmentation algorithm is carried out using the Python programming language. The system platform utilized is Windows, with Python version 1.8.0. The computer system boasts 32 GB of RAM and 24 GB of GPU memory, ensuring efficient processing of the algorithm.

4.1. Evaluation Indicators. There are multiple evaluation metrics that can be used to assess the performance of machine learning tasks, such as instance segmentation. Confusion matrix, precision, and other metrics can be employed for this purpose.

4.1.1. Confounding Matrix. When performing instance segmentation tasks using the deep learning approach, the confusion matrix is used to calculate the intersection over union (IOU) of the prediction box and all real boxes provided by the intelligent segmentation model for a single test image, similar to the calculation in the RPN stage mentioned in Section 1.2. The resulting matrix has the prediction boxes in the column direction and the real boxes in the row direction [26]. Typically, the IOU threshold is set to 50%. If the IOU is greater than 50%, the prediction box is considered to have successfully detected the target. If the category predicted by the prediction box matches the real box, it indicates a correct classification; otherwise, the classification is deemed incorrect. The confusion matrix provides a visual representation of the detection performance of the intelligent segmentation model.

4.1.2. Precision, Recall, and Average Precision. Precision (P), recall (R), and average precision (AP) are commonly used evaluation indicators in classification tasks. The expressions are as follows:

$$P = \frac{TP}{TP + FP}, \tag{11}$$

$$R = \frac{TP}{TP + FN}, \tag{12}$$

$$AR = \int f(R) dR, \tag{13}$$

where TP represents the number of prediction boxs when both the prediction box and the real box of the test set image are positive, that is, the effective number of detection; FP indicates the number of prediction boxs when the prediction box of the test set image is positive and the real box is negative, that is, the number of detection failures; FN refers to the number of predicted boxs in the case that the real box of the test set image is not detected, that is, the number of missed detections; and $f(R)$ represents the relationship function between R and P obtained from the test data.

To evaluate the performance of the established intelligent recognition segmentation model, a single test image is used to calculate the IOU and generate the confusion matrix. The prediction boxes in the column direction of the confusion matrix are sorted based on their classification confidence levels in descending order [26]. Each classification's confidence level is used as the classification threshold in turn to calculate the corresponding precision rate P and recall rate R. The calculation results are then used to draw the $P-R$ curve and calculate the AP value.

However, the AP value of a single test image does not provide a precise evaluation of the performance of the intelligent recognition and segmentation model. Therefore, the mean average precision (mAP) value of the entire test set of images is calculated as the comprehensive evaluation index.

Furthermore, in the context of the majority of semantic segmentation tasks, the typical evaluation metrics employed include $mIOU$ (mean intersection over union) and the $F1$ score. These metrics are expressed as follows:

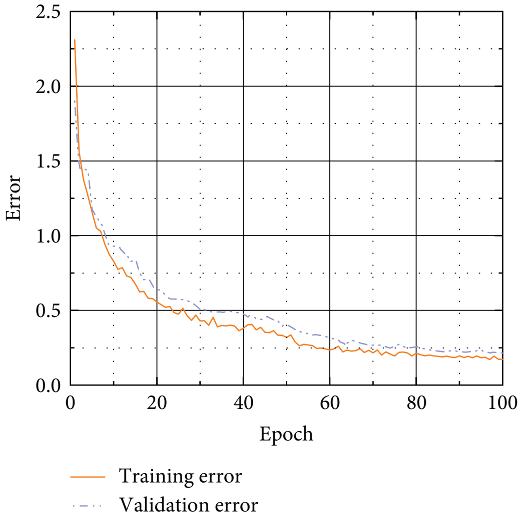


Figure 8: Training error and validation error curves.

$$mIOU = \frac{\sum (TP/(TP + FP + FN))}{N}, \tag{14}$$

$$F1 = \frac{\sum_{1}^{N} ((2P \times R)/(P + R))}{N}, \tag{15}$$

where N represents the total number of classes.

4.2. Hyperparameter Setting. In deep learning algorithms, we encounter two types of parameters: weight parameters and hyperparameters. Weight parameters are subject to continuous optimization throughout the training process, adapting to the data. On the other hand, hyperparameters are predetermined settings used to fine-tune the training process. Consequently, fine-tuning hyperparameters is essential to achieve an optimal intelligent recognition and segmentation model.

In this study, the learning rate was configured at 0.001, playing a critical role in governing the step size for weight updates during neural network training. Additionally, the picture size hyperparameter, which determines the training dimensions set within the deep learning network, significantly impacts model performance and training speed. To ensure network quality, the picture size was established at $1{,}024 \times 1{,}024$ pixels. The model underwent training for a total of 100 epochs, with each epoch signifying the complete cycle of training all the samples within the training dataset once.

5. Experimental Results and Analysis

5.1. Model Training Results. When training our intelligent recognition segmentation algorithm, we generate a training log file that records the training error and validation error data, illustrated in Figure 8. After 100 training iterations, a

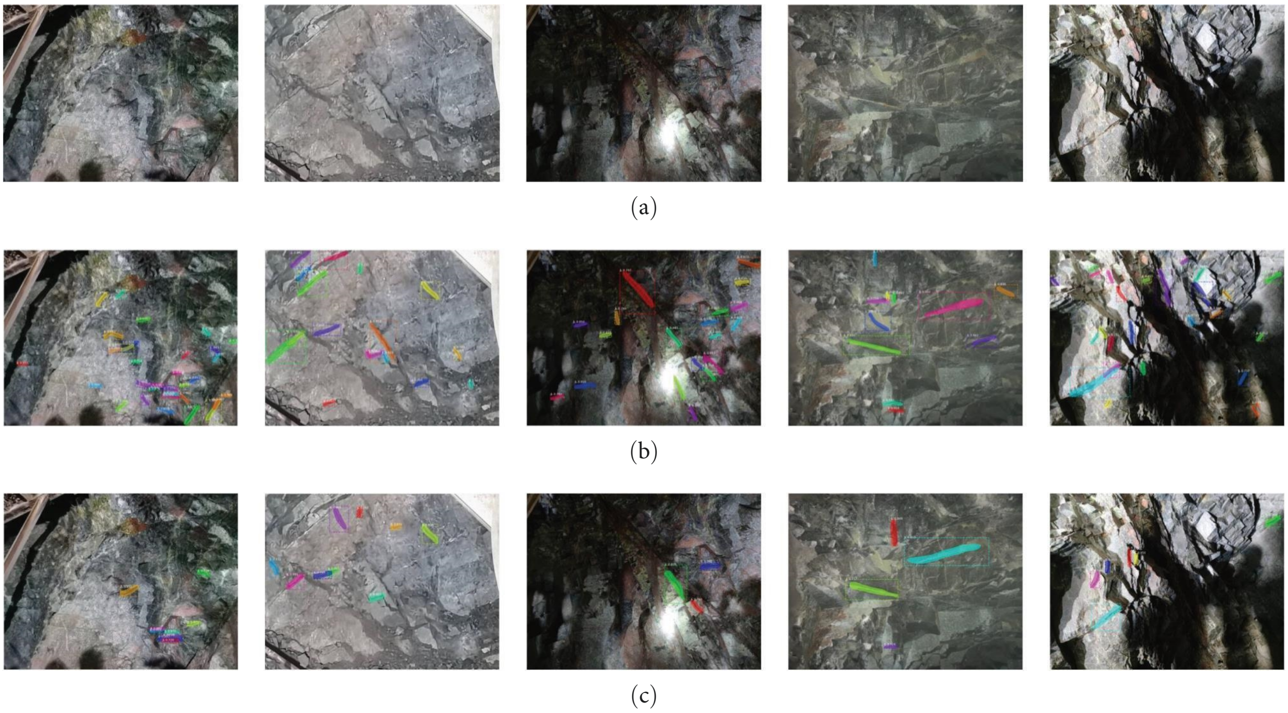


FIGURE 9: Intelligent joint recognition of tunnel face pictures in each scenario. (a) Original images, (b) our method, and (c) Mask R-CNN.

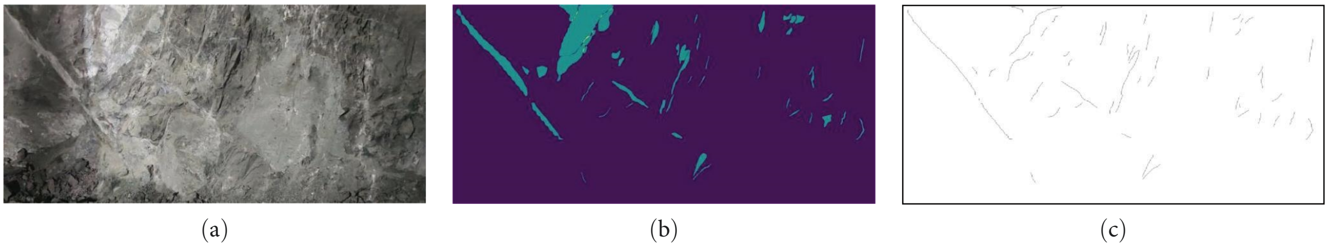


FIGURE 10: Skeletonization process of the detection result. (a) Original image, (b) detection result, and (c) skeletonization process.

noticeable trend emerges. The training error and validation error gradually approach each other and stabilize. This suggests that our training model has achieved a state of "convergence." In simpler terms, the model's performance has plateaued, signifying that it has acquired the most effective approach to perform the task.

Ultimately, our training error registers at 0.17, while the validation error stands at 0.21. These values are remarkably close, indicating that our model delivers consistent performance across various datasets. This is a noteworthy point, as it demonstrates the model's adaptability beyond the training data, showcasing "robustness." In other words, our model remains stable and reliable in its performance.

In conclusion, our model exhibits both stability and robustness throughout the training and validation processes. This implies that it can consistently deliver strong performance across diverse datasets, not merely confined to the training data. These results underscore the effectiveness and feasibility of our algorithm, providing a solid foundation for further applications and research.

5.2. Model Test Results and Evaluation

5.2.1. Model Test Results. The performance assessment of the intelligent recognition and segmentation model was meticulously conducted using a carefully selected test dataset. As illustrated in Figure 9(a)–9(c), this set of test images showcases a variety of real-world scenarios, reflecting the model's capability to operate under diverse environmental conditions. In Figure 9(a)–9(c), the top row corresponds to the detection samples, the middle row showcases the detection results of the method proposed in this paper, and the bottom row illustrates the detection results of the original Mask R-CNN. Looking at the comparison of the detection results, the algorithm proposed in this paper predicts more intricate details, highlighting the enhanced effect of PANet.

Given the marked difference between the actual width of joints and the predicted width, postidentification refinement is essential. This refinement process includes skeleton extraction, as depicted in Figure 10(a)–10(c), which outlines the workflow for skeletal processing of a typical image's identification result.

One of the notable strengths of the intelligent recognition and segmentation model, designed for the specific task of tunnel excavating face joints, is its consistent ability to detect the majority of joint targets with remarkable accuracy. What sets this model apart is its capacity to do so without the need for manual adjustments or interventions, a testament to its true intelligence. Regardless of whether it is a dimly lit environment or a well-illuminated one, the model showcases its versatility by delivering dependable results.

In light of the model's commendable performance, there are nuanced findings that warrant closer scrutiny. First, although the model is generally highly accurate, it occasionally misses visible joint targets, offering an opportunity for fine-tuning to enhance its sensitivity and adaptability to diverse target appearances.

Second, during the segmentation of joint targets, the masks generated by the model tend to be slightly wider than the actual joints in some instances. While this does not significantly impede the model's functionality, addressing this detail could contribute to further refining its performance.

Regarding the issues of missed detections and wider masks, these concerns can be attributed to a combination of factors that necessitate a thorough investigation. Complex tunnel joint structures, variations in lighting conditions, environmental interferences during construction, and inconsistencies in annotation quality are among the factors that can impact algorithm performance. Complex joint structures often present diversity and variations in images, making some joints challenging to differentiate or containing subtle feature changes. Additionally, fluctuating lighting conditions can lead to changes in image brightness, contrast, and shadows, posing challenges for algorithms to accurately capture precise mask boundaries. Environmental factors such as vibrations, dust, and occlusions during construction can further complicate algorithm performance. Moreover, the inconsistency in annotation quality, especially when multiple annotators are involved in dataset creation, can lead to inaccurate annotations, ultimately affecting algorithm performance. To address these challenges, a multifaceted approach is necessary, encompassing improvements in annotation quality, adjustments to algorithm parameters, the selection of appropriate deep learning models, and active diversification of the dataset to better cope with the complexities of tunnel environments.

To overcome the limitations of the algorithm, it is vital to both expand the dataset and elevate the quality of annotations. Expanding the dataset can be achieved through diverse data collection, data augmentation techniques, and the utilization of transfer learning. Meanwhile, the improvement of annotation quality can be realized by engaging domain experts, adopting iterative refinement processes, and leveraging crowdsourcing approaches. These measures promise more accurate and resilient algorithms, broadening their scope of applicability, mitigating the risk of overfitting, and ultimately enhancing the research's practicality and real-world significance.

Prediction box	1	2	3	4	5	6	7	8	9	10
1	89%	0	0	0	0	0	0	0	0	0
2	0	0	0	85%	0	0	0	0	0	0
3	0	0	91%	0.3%	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	2.2%	0
5	0	2.5%	0	0	88%	0	0	0	0	0
6	0	0	0	0	0	0	93%	0	0	3.5%
7	0	0	0	0	0	95%	0	0	0	0
8	0	0	0	0	0	0	3.3%	72%	0	0
9	0	0	0	0	0	0	0	0	87%	0

Ground truth

Figure 11: Detect the confounding matrix of the test set images.

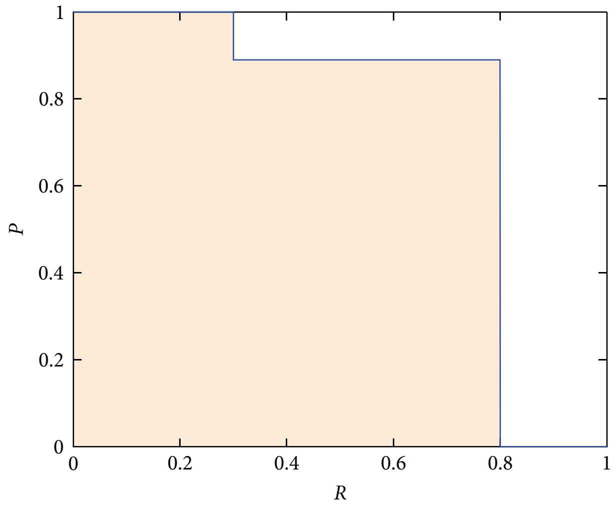


Figure 12: The line chart of *P*–*R*.

5.2.2. Model Evaluation. To thoroughly evaluate the performance of our intelligent recognition and segmentation model, we conducted a meticulous assessment using a carefully selected set of test images. The goal was to ensure objectivity and comprehensiveness in our evaluation process. This evaluation involved several critical steps.

First, each image in the test set underwent detection by our intelligent recognition and segmentation model. Subsequently, we calculated the intersection-over-union (*IOU*) value, representing the extent of overlap between the predicted bounding box and the ground-truth box for each image. This analysis resulted in the creation of a confusion matrix, as depicted in Figure 11. A scrutiny of the *IOU* values unveiled key insights.

For instance, we observed that the *IOU* between the fourth predicted bounding box and all the ground-truth boxes consistently fell below 50%, indicating instances of false detection. Similarly, the *IOU* values between the second and tenth ground-truth boxes and the predicted boxes were also below 50%, further emphasizing occurrences of false detections. It is important to note that, despite these nuances, the algorithm effectively detected the majority of joint

Table 1: The *mAP* (mean average precision) of different algorithms.

Algorithm	*mAP* of detection box (%)	*mAP* of segmentation (%)
The algorithm proposed in this paper	**58.0**	**49.2**
Mask R-CNN	47.3	38.1
Cascade R-CNN	49.5	38.8
Yolact	45.5	35.2
Mask Scoring R-CNN	50.2	41.1

Bold values represent the model evaluation metrics obtained using the algorithm proposed in this paper.

Table 2: The *mIOU* and *F*1 of different algorithms.

Algorithm	*mIOU*	*F*1
The algorithm proposed in this paper	**71.6**	**73.2**
DeepLabV3+	68.3	69.5
U-net	60.2	61.1

Bold values represent the model evaluation metrics obtained using the algorithm proposed in this paper.

targets, generally aligning with engineering requirements. However, these findings underscore the significance of paying close attention to the specific intricacies of joint detection in images.

Furthermore, the precision rate (*P*) and recall rate (*R*) were calculated using the formula outlined in Section 3.1.2, resulting in the creation of a *P*–*R* line chart, as presented in Figure 12. The calculated average precision (*AP*) value was determined to be 0.75. We extended this assessment by iteratively calculating the *AP* value for 80 images within the test set. The average *mAP* value obtained from this set of 80 images was 0.58.

These findings indicate that the intelligent recognition and segmentation model employed in this study effectively fulfills its objective of intelligent detection. Notably, when compared with other intelligent recognition and segmentation models, including the original Mask R-CNN model, our proposed method demonstrates superior detection accuracy advantages. For instance, the average detection value of the Mask R-CNN model within the COCO dataset is only 0.43, underscoring the strengths of our approach.

In conclusion, this performance evaluation has provided valuable insights into our model's strengths and areas for improvement. We are committed to enhancing the model by expanding our dataset to encompass a wider range of scenarios and optimizing annotation quality. These steps will contribute to improved detection accuracy and robustness. Our research continues to strive for an efficient and effective intelligent detection method, particularly in the realm of tunnel excavating face joints, to meet evolving application demands.

5.3. Comparative Test Results. To furnish more compelling evidence of the effectiveness of the proposed algorithm, a series of comparative experiments were conducted. These experiments included evaluations of the original Mask R-CNN algorithm, as well as other widely used instance segmentation algorithms, in addition to the algorithm presented in this study.

Table 1 presents the mean average precision (*mAP*) values of each algorithm. The results show that the proposed algorithm outperforms the traditional Mask R-CNN algorithm, achieving higher detection box and segmentation *mAP* values (58.0%, 49.2%), indicating that the introduced PANet improves the performance of the original algorithm and is more suitable for joint detection and segmentation tasks on tunnel excavating face.

Furthermore, this study compares its algorithm with three popular instance segmentation algorithms, namely Cascade R-CNN [31], Yolact [32], and Mask Scoring R-CNN [33], as shown in Table 1. The detection box and segmentation *mAP* values of Cascade R-CNN and Mask Scoring R-CNN are (49.5%, 38.8%) and (50.2%, 41.1%), respectively, which show an improvement compared to the traditional Mask R-CNN algorithm but are still inferior to the proposed algorithm. However, the Yolact algorithm has the lowest detection box and segmentation *mAP* values and the worst performance in tunnel excavating face joint detection, making it challenging to perform intelligent detection tasks in complex environments. It is worth mentioning that this algorithm has a fast detection rate and broad application prospects in relatively simple application scenarios and object detection tasks. In summary, the proposed algorithm's performance is superior to that of the traditional Mask R-CNN algorithm and various currently popular instance segmentation algorithms, demonstrating its effectiveness and superiority in tunnel excavating face joint detection and segmentation tasks.

In addition, commonly employed semantic segmentation algorithms, including DeepLabV3+ and U-net, were incorporated into the comparison. The performance of each network was assessed using two evaluation metrics, *mIOU* and *F*1, and the results are presented in Table 2. It is clear that the evaluation metrics for DeepLabV3+ and U-net are (68.3, 69.5) and (60.2, 61.1), respectively. In contrast to the evaluation metrics of the algorithm introduced in this paper (71.6, 73.2), they still fall slightly short. This reaffirms the algorithm's competitive edge among common semantic segmentation algorithms, as demonstrated in this study.

6. Conclusion

This study presents a deep learning-based intelligent recognition and segmentation algorithm for tunnel excavating face joints using the Mask R-CNN algorithm and Resnet101 as the main network for feature extraction. To improve the fusion ability of FPN for feature information, the path aggregation network is introduced. The proposed algorithm can intelligently, quickly, and accurately detect multiple types of tunnel excavating face joints in complex on-site environments. The main conclusions are summarized as follows:

(1) In this study, a database comprising 800 images of tunnel excavating faces has been created. The joint structures in the raw images are identified and labeled using the polygon mapping method. The labeled joint structures provide valuable information, such as joint profiles and pixel features, for training the learning network.

(2) Targeted improvements are made to the algorithm based on the specificity of the palm joint recognition and segmentation task, and the introduction of path aggregation networks has improved the fusion ability of FPN for feature information. The proposed algorithm can detect joint information in tunnel excavating face photos, locate the position of joints through detection boxes, and segment pixels belonging to joints through masks. The algorithm has strong anti-interference ability and can be applied to intelligent detection and segmentation of tunnel excavating face joints in complex tunnel environments.

(3) The performance of the proposed algorithm is evaluated on 80 sample images in the test set using indicators such as confusion matrix, accuracy, and recall. The calculated mean average precision (*mAP*) values of the detection frame and segmentation were 58.0% and 49.2%, respectively. Compared with the Mask R-CNN algorithm and several current popular instance segmentation algorithms, the performance was excellent, demonstrating the effectiveness and superiority of the proposed algorithm in the joint detection and segmentation task of tunnel excavating face.

Moreover, the proposed intelligent recognition segmentation model can be directly applied to joint detection tasks of railway and highway tunnel excavating faces in complex environments. Combined with hardware development, intelligent recognition devices such as drones and robots can be utilized to address the subjectivity and low efficiency issues of traditional sketching methods, providing technical and theoretical support for the intelligent development of tunnel construction.

Data Availability

Data will be made available on request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (No. 52278422 and No. 51978677).

References

[1] L. M. Fam and N. Li, “Preliminary study on rock mass fracture measurement method based on digital photography technology,” *Journal of Rock Mechanics and Engineering*, vol. 5, pp. 792–797, 2005.

[2] G. Hu and Q. K. Jin, “Research on computer image processing technology of rock mass natural fractures,” *Non-Ferrous Metals*, vol. 5, pp. 39-40, 2004.

[3] S.-S. Leu and S.-L. Chang, “Digital image processing based approach for tunnel excavation faces,” *Automation in Construction*, vol. 14, no. 6, pp. 750–765, 2005.

[4] Y. Ye and S. W. Meng, “Research on digital logging and recognition technology of geological information of tunnel face,” *Journal of Beijing Jiaotong University*, vol. 31, no. 1, pp. 59–62, 2007.

[5] B. Leng, Y. Zhang, and H. Yang, “Rapid identification method of rock mass fracture in tunnel face,” *Journal of Southwest Jiaotong University*, vol. 56, no. 2, pp. 246–252+322, 2021.

[6] J. Luo and D. G. Liu, “Research on image processing technology for parameters of development degree of surrounding rock structural plane,” *Computer Engineering and Science*, vol. 35, no. 4, pp. 75–80, 2013.

[7] S. C. Li, H. L. Liu, and L. P. Li, “Quantitative characterization method and engineering application of rock mass structure of mining face based on digital image,” *Journal of Rock Mechanics and Engineering*, vol. 36, no. 1, pp. 1–9, 2017.

[8] P. Y. Li, K. Zhao, and Z. D. Chen, “Research on geological structure information extraction of tunnel face based on image processing,” *Information Technology of Civil Construction Engineering*, vol. 9, no. 6, pp. 67–72, 2017.

[9] C. L. Zhou, H. H. Zhu, and X. J. Li, “Infrared photography and image processing of tunnel face in NATM construction,” *Journal of Rock Mechanics and Engineering*, vol. S1, pp. 3166–3172, 2008.

[10] F. Y. Wang, J. P. Chen, and G. D. Yang, “Solution models of geometrical information of rock mass discontinuity based on digital close range photogrammetry,” *Journal of Jilin University (Earth Science Edition)*, vol. 42, no. 6, pp. 1839–1846, 2012.

[11] M. Yang and W. X. Wang, “Rock discontinuity spacing measurement based on image technique,” *Journal of Computer Applications*, vol. 30, no. 158, pp. 146-147, 2010.

[12] T. H. Yang and S. K. Chen, “Rock mass structure digital recognition and hydro-mechanical parameters characterization of sandstone in Fangezhuang coal mine,” *Chinese Journal of Rock Mechanics*, vol. 28, no. 12, pp. 2482–2489, 2009.

[13] X. Zhuang, C. Baolin, and F. Jinyang, “Digital recognition method and application of tunnel face rock mass structure based on machine vision 3D reconstruction technology,” *Journal of Railway Science and Engineering*, vol. 16, no. 4, pp. 1001–1007, 2019.

[14] B. Leng, W. G. Qiu, and G. Wang, “Digital image processing in tunnel engineering application research in geological analysis,” *Railway Standard Design*, vol. 11, pp. 77–81, 2013.

[15] H. W. Huang and Q. T. Li, “Image recognition of shield tunnel leakage disease based on deep learning,” *Journal of Rock Mechanics and Engineering*, vol. 36, no. 12, pp. 2861–2871, 2017.

[16] L. F. Li, W. F. Ma, and L. Li, “Research on bridge crack detection algorithm based on deep learning,” *Journal of Automation*, vol. 45, no. 9, pp. 1727–1742, 2019.

[17] Y. Lin, Z.-Y. Yin, X. Wang, and L. Huang, “A systematic 3D simulation method for geomaterials with block inclusions from image recognition to fracturing modelling,” *Theoretical and Applied Fracture Mechanics*, vol. 117, Article ID 103194, 2022.

[18] M.-F. Lei, Y.-B. Zhang, E. Deng et al., “Intelligent recognition of joints and fissures in tunnel faces using an improved mask region-based convolutional neural network algorithm,” *Computer-Aided Civil and Infrastructure Engineering*, vol. 39, no. 8, pp. 1123–1142, 2023.

[19] Y. Lin, X. Wang, J. Ma, and L. Huang, “A finite-discrete element based appoach for modelling the hydraulic fracturing of rocks with irregular inclusions,” *Engineering Fracture Mechanism*, vol. 261, Article ID 108209, 2022.

[20] Y. Lin, X. Wang, J. Ma, and L. Huang, “A systematic framework for the 3D finite-discrete modelling of binary mixtures considering irregular block shapes and cohesive block-matrix interfaces,” *Powder Technology*, vol. 398, Article ID 117070, 2022.

[21] Y. Lin, C. Li, J. Ma, M. Lei, and L. Huang, “Effects of void morphology on fracturing characteristics of porous rock through a finite-discrete element method,” *Journal of Natural Gas Science and Engineering*, vol. 104, Article ID 104684, 2022.

[22] L. Huang, J. Ma, M. Lei, L. Liu, Y. Lin, and Z. Zhang, “Soil-water inrush induced shield tunnel lining damage and its stabilization: a case study,” *Tunnelling and Underground Space Technology*, vol. 97, Article ID 103290, 2020.

[23] K. He, G. Gkioxari, and P. Dollár, “Mask r-cnn,” in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2961–2969, IEEE, 2017.

[24] L. Huang, S. Huang, and Z. Lai, “On optimizing site investigation program using centroidal Voronoi tessellation and random field theory,” *Computers and Geotechnics*, vol. 118, Article ID 103331, 2020.

[25] L. Huang, S. Huang, and Z. Lai, “On the energy-based criteria for defining slope failure considering spatially variable soil properties,” *Engineering Geology*, vol. 264, Article ID 105323, 2020.

[26] Y. Lin, J. Ma, Z. Lai, L. Huang, and M. Lei, “A FDEM approach to study mechanical and fracturing responses of geo-materials with high inclusion contents using a novel reconstruction strategy,” *Engineering Fracture Mechanism*, vol. 282, Article ID 109171, 2023.

[27] L. Raphael and D. Y. Mery, “Mapping fire blight cankers and autumn blooming in pear trees using faster R-CNN,” *Precision Agriculture*, vol. 1, Article ID 25, 2024.

[28] M. F. Lei, Y. B. Zhang, and W. D. Wang, “Research on rock lithology mask R-CNN intelligent identification method and application,” *Journal of Railway Science and Engineering*, vol. 19, no. 11, pp. 3372–3382, 2022.

[29] S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: towards real-time object detection with region proposal networks,” in *Advances in Neural Information Processing Systems*, vol. 28, Curran Associates, Inc., 2015.

[30] S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia, “Path aggregation network for instance segmentation,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 8759–8768, IEEE, Salt Lake City, UT, USA, 2018.

[31] Z. Cai and N. Vasconcelos, “Cascade R-CNN: high quality object detection and instance segmentation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 5, pp. 1483–1498, 2021.

[32] Y. X. Lin, Z. S. Lai, J. J. Ma, and L. C. Huang, “A combined weighted Voronoi tessellation and random field approach for modeling heterogeneous rocks with correlated grain structure,” *Construction and Building Materials*, vol. 416, Article ID 135228, 2024.

[33] Z. Y. Yue, L. C. Huang, Y. X. Lin, and M. F. Lei, “Research on image deformation monitoring algorithm based on binocular vision,” *Measurement*, vol. 228, Article ID 114394, 2024.