minerals

MDPI

Article

The Segmentation of Tunnel Faces in Underground Mines Based on the Optimized YOLOv5

Chundi Ma [1], Kechao Li [1], Jilong Pan [1], Jiashuai Zheng [1], Qinli Zhang [1] and Chongchong Qi [1,2,*]

[1] School of Resources and Safety Engineering, Central South University, Changsha 410083, China
[2] School of Metallurgy and Environment, Central South University, Changsha 410083, China
* Correspondence: chongchong.qi@csu.edu.cn

Abstract: Tunnel faces in underground mines, as the front line of mining, play an important role in both mine safety and mining intelligence. However, the engineering quality of tunnel faces is still evaluated based on visual observations by technicians, which cannot guarantee safety and real-time performance. Therefore, there is an urgent need for a more effective method to detect the quality of tunnel face engineering. In this study, a high-performance and accurate tunnel face segmentation model was developed by applying the YOLOv5-seg computer vision model to an underground mine. By optimizing a classic Chinese underground mine image dataset through Sobel preprocessing and improving the network structure of the YOLOv5-seg model using the SimAM module, good predictive performance was achieved for tunnel face segmentation, with values of 0.97, 0.89, 0.80, and 0.78, respectively, achieved for the pixel accuracy, Dice coefficient, mask intersection over union (IOU), and box IOU on the test set. And the performance of this model outperforms all YOLOv5 models and U-net in the same task of tunnel face segmentation. Model interpretation and visualization further demonstrated the positive effect of the SimAM module on the model, and, finally, the segmentation results were used to evaluate the tunnel face engineering. Overall, this study's results provide a feasible, safe, and real-time method for accurately segmenting tunnel faces in underground mines and provide a reliable approach for data-driven applications of intelligent technology in mines in the future.

Keywords: underground mining; tunnel surface engineering; YOLOv5; SimAM

check for updates

Academic Editor: Amin Beiranvand Pour

Received: 23 January 2025
Revised: 23 February 2025
Accepted: 27 February 2025
Published: 28 February 2025

Citation: Ma, C.; Li, K.; Pan, J.; Zheng, J.; Zhang, Q.; Qi, C. The Segmentation of Tunnel Faces in Underground Mines Based on the Optimized YOLOv5. *Minerals* **2025**, *15*, 255. https://doi.org/10.3390/min15030255

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Mining, as a core industry for basic resources, is developing rapidly due to accelerating global industrialization [1]. However, due to the rapidly diminishing reserves of open-pit mineral resources and the enormous environmental damage caused by open-pit mining, underground mining will become the main mining method in the future [2]. The tunnel face of underground mines, at the forefront of the direct advancement of mining operations, often differs from its designed shape due to the complexity of the in situ geological structure, uncertainty in the blasting process, and the influence of many random factors such as the selection of heavy equipment and the differences in mechanical standards in the mining process [3–6]. These shape differences, which may be as small as a few centimeters of local concavity and convexity or as large as several meters or even a dozen meters of overall contour deformation, may affect the safety of mining operations. Therefore, accurate and rapid monitoring and analysis of the actual shape of the tunnel face is key to ensuring the orderly and stable progression of the mining process as a whole. At present, manual

monitoring remains the main approach used in the mining industry, requiring professionals to go deep into complex and dangerous underground environments to obtain information describing the tunnel face based on visual observations and manual recording [7]. This approach not only requires a considerable expenditure of manpower and material resources but it is also difficult to rapidly obtain accurate, real-time information using this method, thus representing a serious obstacle to the advancement of mine informatization and the application of intelligent technologies.

Computer vision technology, as a kind of non-contact perception technology, has the advantages of high efficiency, rapidity, and informatization capacity [8]. This approach primarily uses convolutional neural networks (CNNs) to combine and abstract features at different scales based on an image's pixel information, which then finally outputs the target information [9]. Therefore, computer vision technology is widely used in many fields to rapidly obtain information in real time without risking people's lives in high-risk settings such as underground mines [10]. Compared with the classic computer vision models such as Seg-Net, the YOLOv5 model stands out. YOLOv5 is known for its high calculation speed and is capable of real-time calculations. It is also highly accurate and capable of accurately segmenting targets in complex environments [11]. However, the application of computer vision models has been mainly focused on rock fractures on the tunnel face, and there is still a gap in the research on the segmentation of the tunnel face shape [12,13].

To fill this gap, the YOLOv5-seg model is used in this study for the segmentation of underground mine tunnel face images. YOLOv5-seg has a very high detection speed and can return results quickly and accurately [14]. This approach is also highly adaptable to varied lighting conditions, shooting angles, and complex backgrounds, thus making YOLOv5-seg a very suitable model for segmenting the tunnel faces of underground mines. In this study, YOLOv5-seg was trained and evaluated on image data from a Chinese classical underground mine. Considering that the low brightness and dusty environment of the tunnel face in underground mines can negatively affect the model accuracy, the images were further processed and the model network structure was optimized. Finally, the proposed model is interpreted and applied to evaluate the quality of a tunnel face project.

2. Materials and Methods

2.1. Dataset

2.1.1. Data Collection

In this study, the segmentation dataset collection was carried out at a classic Chinese underground mine. Considering the portability and stability required for future use by workers and the loading of the rock drilling jumbo, the GoPro HERO11 camera was used for dataset collection. After the scraper had cleared the ore, the workers entered the tunnel to record the image data. The following steps were then followed to obtain the image data: Considering the safety of the workers and the location of the camera loaded on the mine drilling jumbo, the shooting distance is 6–8 m from the tunnel face; make sure that the tunnel face is completely photographed. At the same time, make sure that the camera is facing the tunnel face directly to avoid the distortion affecting the accuracy of the image. Use a strong flashlight to assist in the illumination, making the outline of the tunnel face as clear as possible. A total of 176 pictures were taken of the working face, of which 119 pictures were retained after filtering to remove images that were blurred or did not show the complete tunnel face.

After data collection, to accurately analyze the morphological characteristics of the blasted tunnel face, the LabelMe software was used to perform pixel-level contour annotation, and the difference between the tunnel surface assessment of the raw image and the label image was found to be less than 5% (Figure 1). LabelMe is a graphical image labeling

tool that supports detailed contour depiction and labeling of objects in images. LabelMe not only generates labeling files in JSON or PNG formats that match the size of the image but also has an intuitive user interface that is easy to operate, which greatly improves the efficiency and accuracy of the labeling.

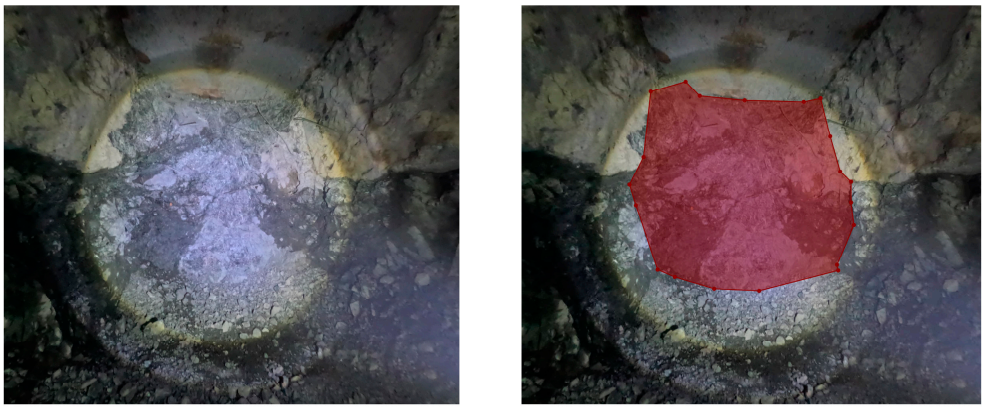


Figure 1. Original and labeled image: (**a**) Raw image of the tunnel face, (**b**) Labeled image of the tunnel face by LabelMe.

2.1.2. Image Preprocessing Methods

Image preprocessing can highlight image features, making the model more accurate and improving target segmentation efficiency [15]. In this study, histogram equalization, contrast-constrained adaptive histogram equalization (CLAHE), the Sobel filter, Laplacian operator, and Gaussian filtering (Figure 2) were applied and compared before model training, and the accuracies of the resulting models were compared in order to determine the optimal image preprocessing method.

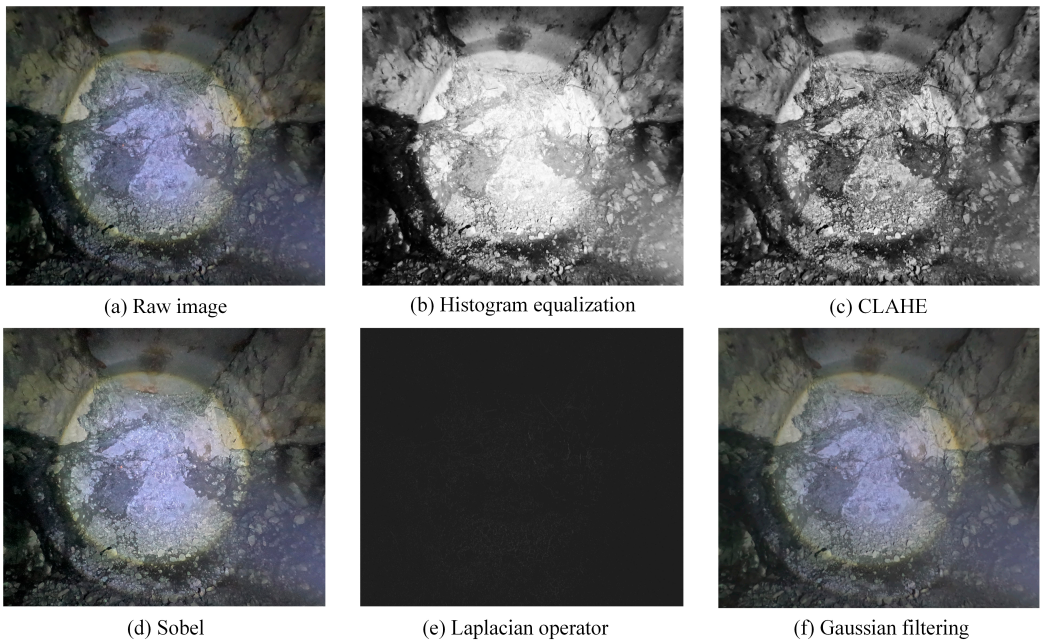


Figure 2. Diagram of different preprocessing methods: (**a**) Raw image of the tunnel face. (**b**) Diagram of raw tunnel face image processed by histogram equalization. (**c**) Diagram of raw tunnel face image processed by CLASH. (**d**) Diagram of raw tunnel face image processed by Sobel. (**e**) Diagram of raw tunnel face image processed by laplacian operator. (**f**) Diagram of raw tunnel face image processed by gaussian filtering.

All of the above preprocessing methods can improve the quality of the image, which in turn affects the model accuracy [16–20]. However, the Sobel filter, as a classic edge-detection method, can not only locate the edges but also provide gradient information by calculating the brightness gradients. When dealing with the image noise caused by the light reflection of suspended particles in the air, the Sobel filter, by virtue of its sensitive capture ability of local image changes, can effectively suppress such noise interference, more accurately identify and extract edges, and still perform outstandingly in edge-detection tasks in complex noise environments.

2.2. Segmentation Model

2.2.1. YOLOv5-Seg and U-Net

The YOLOv5-seg model adds a segmentation header, which is a lightweight convolutional network, to each predicted bounding box while maintaining the original multi-scale feature extraction capability of YOLOv5 [21]. This model is responsible for extracting the segmentation information from the feature map corresponding to the object contour. In turn, this allows YOLOv5-seg to perform pixel-level segmentation with high efficiency and accuracy.

The YOLOv5-seg network structure comprises three main parts (Figure 3), i.e., backbone, neck, and head. This structure enables YOLOv5-seg to have very high multi-scale training and inference ability while maintaining good performance on images of different sizes and in different environments [22]. YOLOv5-seg also has different depths of the network structure, thus making it more selective. YOLOv5-seg currently has four versions: YOLOv5s-seg (small), YOLOv5m-seg (medium), YOLOv5l-seg (large), and YOLOv5x-seg (extra-large). Their network structures become increasingly deep with increasing size to meet the different accuracy requirements of a range of scenarios across different fields [23].

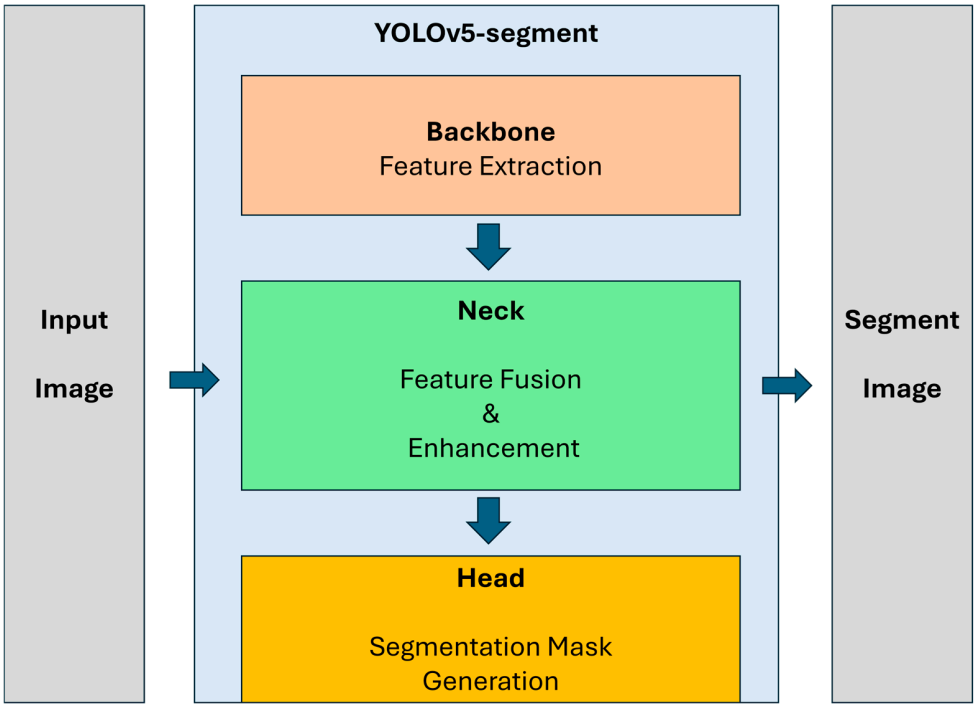


Figure 3. YOLOv5-seg structure diagram.

U-net has been widely used in numerous fields of research and is one of the better segmentation models [24]. This approach achieves efficient image segmentation via a symmetric network structure. This structure allows the network to perform accurate local pixel prediction while maintaining the image context information [25].

2.2.2. SimAM

SimAM (Simplified Axial Attention Module) is a self-attention mechanism designed to enhance the performance of CNNs in image processing tasks [26]. The core concept of the SimAM module is to adaptively retune the feature response of each channel using the self-attention mechanism, thus allowing the network to more intensively learn important feature information [27]. Compared with traditional attention mechanisms, SimAM is structurally simpler and more efficient (Figure 4).

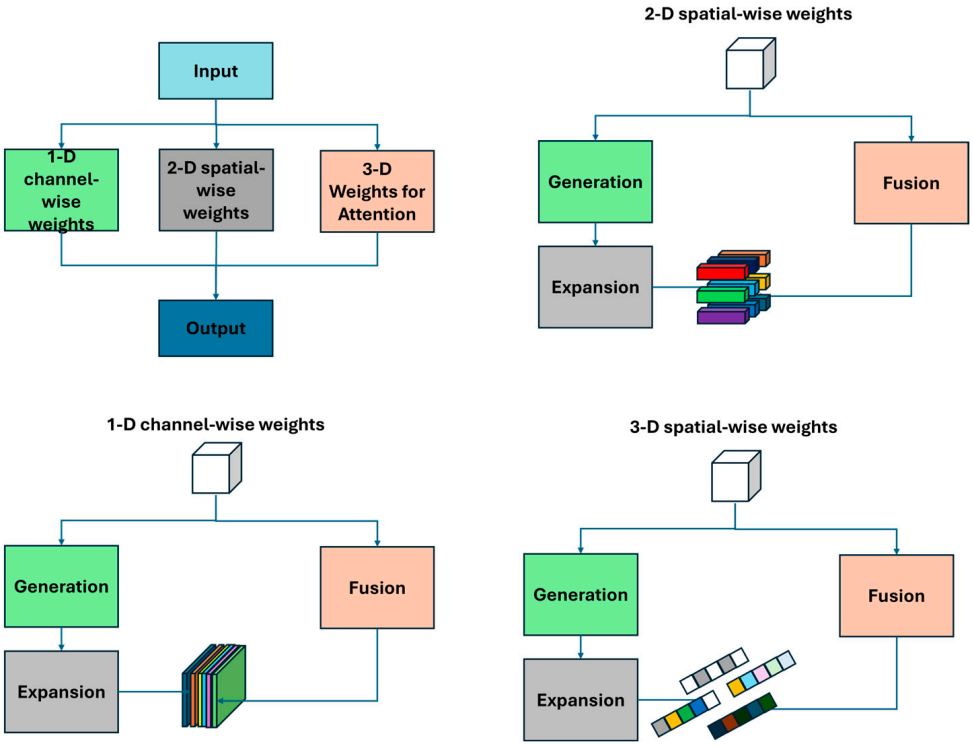


Figure 4. SimAM schematic diagram.

The SimAM module realizes the adaptive adjustment of features by calculating the self-attention weights for each spatial location in the feature map. For an input feature map $X \in R(C \times H \times W)$, SimAM computes the energy value E to capture the degree of feature activity at each spatial location, and the module then uses the computed energy value E to adjust the features of the original input X (Equation (2)), thereby enhancing important features and suppressing unimportant ones [28].

$$E_{i,j} = \sigma\left(\frac{1}{c}\sum_{k=1}^{C} X_{k,i,j}^2\right) \tag{1}$$

$$OutPut_{k,i,j} = \frac{1 - E_{i,j}}{E_{i,j}} \cdot X_{k,i,j} \tag{2}$$

$$\sigma(z) = \frac{1}{1 + e^{-z}} \tag{3}$$

where C is the number of channels, H and W are the spatial dimensions, and σ is the sigmoid function (Equation (3)) used to normalize E to the range of [0, 1]. SimAM reinforces important image features through a self-attention mechanism, allowing the network to focus more strongly on key information. This operation makes the resulting segmentation model more sensitive to the shapes and edges of objects in the image, thus improving the segmentation performance.

2.3. Model Evaluation and Interpretation

2.3.1. Model Evaluation

Performance assessment of segmentation models through evaluation metrics is a crucial step in the image segmentation task [29]. Evaluation metrics describe the accuracy of the model's ability to recognize and localize object contours in an image. Evaluation metrics also reflect the model's reliability and accuracy in practical applications [30]. The metrics commonly used for segmentation model evaluation include the Dice coefficient, pixel accuracy (PA), mask intersection over union (IOU), and box IOU. These four metrics have values closer to 1, indicating that the model performs better and the segmentation results are closer to the true values [31,32]. These metrics measure the model's segmentation performance from multiple perspectives and highlight areas for further model optimization and parameter adjustment.

The Dice coefficient is a statistical metric that measures the similarity of two samples and is widely used in image segmentation for model performance evaluation [33]. For the segmentation task, the Dice coefficient predicts the overlap between the segmentation results and the true segmentation labels (Equation (4)). The value of the Dice coefficient ranges from 0 to 1, where 1 indicates perfect segmentation results and 0 indicates no overlap. In practice, the Dice coefficient is an effective measure of how well the shape and size of the segmented region match the labeled region.

PA is one of the most intuitive evaluation metrics used in image segmentation. The PA metric measures the proportion of correctly categorized pixels to the total pixels in the segmented image (Equation (5)), thus reflecting the correctness of the model's classification at the pixel level. One limitation of the PA metric is in the case of unbalanced data, i.e., where the number of foreground and background pixels varies markedly; in this instance, the PA may be skewed in favor of the more numerous category [34]. Therefore, the PA method is often used in conjunction with other metrics to provide a more comprehensive performance assessment.

Mask IOU is an important metric for measuring the performance of segmentation models, especially in terms of evaluating the shape and edge alignment of an object [35]. Mask IOU calculates the ratio of intersection and concatenation between the predicted mask and the true mask [36] (Equation (6)). This metric is highly sensitive to the accuracy of the edge localization, thus making it useful in testing the model's ability to recognize the contours of an object. In practice, this metric is often used to evaluate segmentation performance for complex shapes and irregular objects.

Box IOU is also a widely used metric for evaluating object detection and segmentation models, especially in tasks where the accuracy of the predicted bounding box must be evaluated [37]. Box IOU measures the ratio of intersection and concurrency between the predicted bounding box and the true bounding box (Equation (7)) [38]. This metric is a useful indicator of the model's ability to localize objects, particularly in scenarios where the position and size of objects are important. A high box IOU value indicates that the model can accurately predict an object's location and extent, which is a key performance metric for many real-world applications.

$$Dice = \frac{2 \times |P \cap T|}{|P| + |T|} \tag{4}$$

$$PA = \frac{\sum_{i=1}^{C} TP_i}{\sum_{i=1}^{C} (TP_i + FP_i + TN_i)} \tag{5}$$

$$Mask\ Iou = \frac{|P \cap T|}{|P \cup T|} \tag{6}$$

$$Box\ IoU = \frac{\left|B_p \cap B_{gt}\right|}{\left|B_p \cup B_{gt}\right|} \tag{7}$$

where C is the total number of categories, TP_i is the number of true positive pixels (those correctly predicted to belong to category i), FP_i is the number of false positive pixels (those predicted to belong to category i that do not belong to category i), and FN_i is the number of false negative pixels (those that belong to category i but are not predicted to do so). B_p is the predicted bounding box and B_{gt} is the real bounding box.

Although there are structural differences between the models, the training parameter settings are uniform. In this experiment, the total number of model training rounds was set to 300 and to minimize the phenomenon of model overfitting, the number of waiting rounds of the early stopping mechanism was set to 50. In total, 12 images were processed in each batch during training. The resolution of the input images was set at 640 pixels, and a GPU device was used for model training to improve computational efficiency. To ensure the data can be efficiently delivered to the model during the data loading stage, the model was trained on the server shown in Table 1, which is the same as that used to run the model, and eight working threads were used to ensure efficient transport of data to the model. To optimize the parameters of the model and minimize the loss function, the stochastic gradient descent optimizer with a momentum parameter of 0.937 was used.

Table 1. Table of model metrics results for different preprocessing methods.

Preprocessing Method	Train Set				Validation Set			
	Dice	**PA**	**Mask IOU**	**Box IOU**	**Dice**	**PA**	**Mask IOU**	**Box IOU**
Origin	0.9529	0.9848	0.9106	0.9009	0.8388	0.9286	0.7972	0.8032
CLAHE	0.9641	0.9887	0.9308	0.9127	0.8746	0.9500	0.7971	0.8077
Histogram Equalization	0.9264	0.9762	0.8652	0.8691	0.8445	0.9304	0.7374	0.7681
Sobel	0.9676	0.9899	0.9374	0.9363	0.8922	0.9670	0.8068	0.8123
Laplacian Edge Enhancement	0.8976	0.9672	0.8165	0.8211	0.8627	0.9592	0.7603	0.7801
Gaussian Blurring	0.9082	0.9699	0.8336	0.7999	0.8485	0.9103	0.7975	0.8063

2.3.2. Network Interpretation Ability and Visualization

Neural network training is considered a black box process due to the unpredictable formation of neuron values and the lack of direct interpretability of neuron meanings. This is especially true for end-to-end networks, which are difficult to interpret and provide limited insight into the internal features learned in each convolutional layer [39]. In the studied image analysis process, the tunnel face image represents the input, and the segmentation and detection results are obtained directly as outputs; however, the intermediate learning process is hidden and cannot be analyzed. This poor interpretability greatly hinders further optimization of the network structure, evaluation of the robustness of each network layer, and the transferability and adaptability of the network to different applications. Network visualization can be applied, where internal features are converted into visually perceptible image patterns. In each visual neuron region, neurons can selectively respond to different features (color, shape, texture, etc.) to reconstruct the input image, which can be illustrated to visualize the integrated feature map learned by each layer of the network [40].

3. Results and Discussion

3.1. Model Performance

3.1.1. Selection of Preprocessing Method

In order to determine the optimal preprocessing method to increase the accuracy of the contour segmentation of the tunnel face after blasting, five state-of-the-art image

preprocessing methods (CLAHE, histogram equalization, Sobel filtering, Laplacian edge enhancement, and Gaussian blurring), were selected for comparison.

In this study, the original image of a Chinese underground mine was used as the dataset for training the YOLOv5l-seg model, and the benchmark performance data on the training and test sets were obtained, as shown in Table 1. On the training set, the model's Dice coefficient, PA, mask IOU, and box IOU values were 0.9529, 0.9848, 0.9106, and 0.9364, respectively, while the corresponding values of these metrics on the testing set were 0.8388, 0.9286, 0.7973, and 0.8032, respectively. The model training and evaluation metrics of the five preprocessing methods were then compared. The results show a significant degradation in the performance of the models trained by the histogram equalization, Laplacian edge enhancement, and Gaussian blurring methods on both the training and validation sets. This is due to the introduction of excessive noise or the loss of critical structural information during image processing, which affects the accuracy and robustness of the model of the hexagonal structure of the tunnel face.

In contrast, the models trained by the CLAHE and Sobel filtering preprocessing methods achieved better performance. Sobel preprocessing is performed because the Sobel filter can effectively enhance the edge features of the image and retain important structural information, thus significantly improving model performance [41]. For the training set with Sobel preprocessing, the Dice coefficient was 0.9676, the PA was 0.9899, the mask IOU was 0.9375, and the box IOU was 0.9364, representing improvements of 1.54%, 0.51%, 2.95%, and 3.94%, respectively, in these four metrics compared to the case with no preprocessing. Similarly, on the validation set, this model achieved a Dice coefficient of 0.8922, a PA of 0.9670, a mask IOU of 0.8069, and a box IOU of 0.8124, representing improvements of 6.37%, 4.13%, 1.21%, and 1.14%, respectively, relative to the model without preprocessing. The above results show that the use of the Sobel preprocessing method can significantly improve the model's accuracy and stability, thus making it suitable for training the tunnel face segmentation model.

3.1.2. Comparison of Segmentation Models

To find the optimal network structure model for the hexagonal tunnel face of the mine, this paper compares and analyzes models with different network depths, including YOLOv5-s, YOLOv5-m, YOLOv5-l, and YOLOv5-x, in addition to the U-Net model, which is widely used in the industry. The results are shown in Table 2. Among these models, the U-Net model achieves a Dice coefficient of 0.7514, a PA of 0.7128, a mask IOU of 0.7128, and a box IOU of 0.7268 on the training set, with equivalent test set values of 0.7288 0.7812,0.6518, and 0.6674, respectively, indicating poor model performance.

Table 2. Table of metrics for different models.

Model	Train Set				Validation Set			
	Dice	PA	Mask IOU	Box IOU	Dice	PA	Mask IOU	Box IOU
Yolo5s	0.9070	0.9696	0.8321	0.8211	0.8046	0.8717	0.8267	0.7996
Yolo5m	0.9152	0.9727	0.8455	0.8138	0.8049	0.8719	0.7971	0.7954
Yolo5l	0.9676	0.9899	0.9374	0.9363	0.8922	0.9670	0.8068	0.8123
Yolo5x	0.8772	0.9611	0.7846	0.7773	0.7960	0.8678	0.8125	0.8051
U-Net	0.7513	0.8611	0.7127	0.7267	0.7288	0.7812	0.6518	0.6674

Compared with the U-Net, the YOLOv5 models show significant advantages in all evaluation metrics. Among them, YOLOv5-l significantly outperforms the other models on both the training and test sets. Its Dice coefficient, PA, mask IOU, and box IOU values on

the training set improved by 28.79%, 14.95%, 31.52%, and 28.85%, respectively, and these metrics on the validation set improved by 22.41%, 23.78%, 23.78%, and 21.71%, respectively.

YOLOv5-l outperforms U-Net primarily because YOLOv5-l has a deeper network structure and stronger feature extraction capability, i.e., YOLOv5-l possesses more convolutional layers and complex feature maps that can more efficiently capture and understand the details and complex structures in the image, thus improving the model's detection accuracy and robustness [42]. In contrast, although U-Net has its advantages in semantic segmentation tasks, its shallow structure and limited number of feature maps limit its performance in image processing and segmentation tasks. This results in a lower performance than that of YOLOv5 models on both the training and test sets. In particular, in the key mask IOU and box IOU metrics, U-Net's performance is far inferior to that of YOLOv5-l, further validating the significant advantages of deep network structures in improving model accuracy and stability.

3.1.3. Network Structure Optimization

The advantage of the SimAM attention module is its ability to enhance the multi-attention mechanism through self-attention, thereby capturing the details and complex structures in the image more efficiently to improve model accuracy and robustness [43]. Therefore, in this study, the SimAM attention module is applied to improve the YOLOv5-l-seg model.

Since the backbone component forms the foundation of the YOLOv5 model framework, primarily performing the task of extracting features from images, placing the SimAM attention module in the backbone part can strengthen the important features in the early stages of feature extraction, which in turn has a global impact on the whole network's performance. Given that the feature map after the C3 module contains rich information about intermediate and high-level features, which helps in the attention mechanism's processing, and the SPPF module is mainly used for pooling processing, it is optimal to enhance the features before this point in the network to improve the model's multi-scale target detection performance. Therefore, the SimAM module is placed after the C3 module and before the SPPF module to improve the model's image processing accuracy and stability, with the specific structure shown in Figure 5.

The YOLOv5-l model including the SimAM attention module (referred to hereafter as YOLOv5-SimAM) shows improved model performance (Figure 6). On the training set, YOLOv5-SimAM achieves Dice coefficient, PA, mask IOU, and box IOU values of 0.9234, 0.9743, 0.8592, and 0.8478, respectively. On the validation set, the Dice coefficient was 0.8969, the PA value was 0.9691, the mask IOU value was 0.8142, and the box IOU value was 0.8498, while the equivalent values on the test set were 0.8886, 0.9718, 0.8016, and 0.7815, respectively. The Dice coefficient, PA, mask IOU, and box IOU values improved on the validation set by 0.52%, 0.21%, 0.90%, and 4.61%, respectively, and on the test set by 4.07%, 0.89%, 7.398%, and 4.03%, respectively. Relative to the YOLOv5-l model, the four evaluated metrics on both the validation and test sets improved significantly; however, they decreased slightly on the training set. This finding further demonstrates that the SimAM module successfully enhances the model's perception of the target area and feature extraction accuracy, thereby helping to improve the segmentation performance of the model and enhance its robustness and generalization.

The segmentation of the image based on the YOLOv5-SimAM model is illustrated in Figure 7—as shown, the model can effectively segment the hexagonal tunnel face.

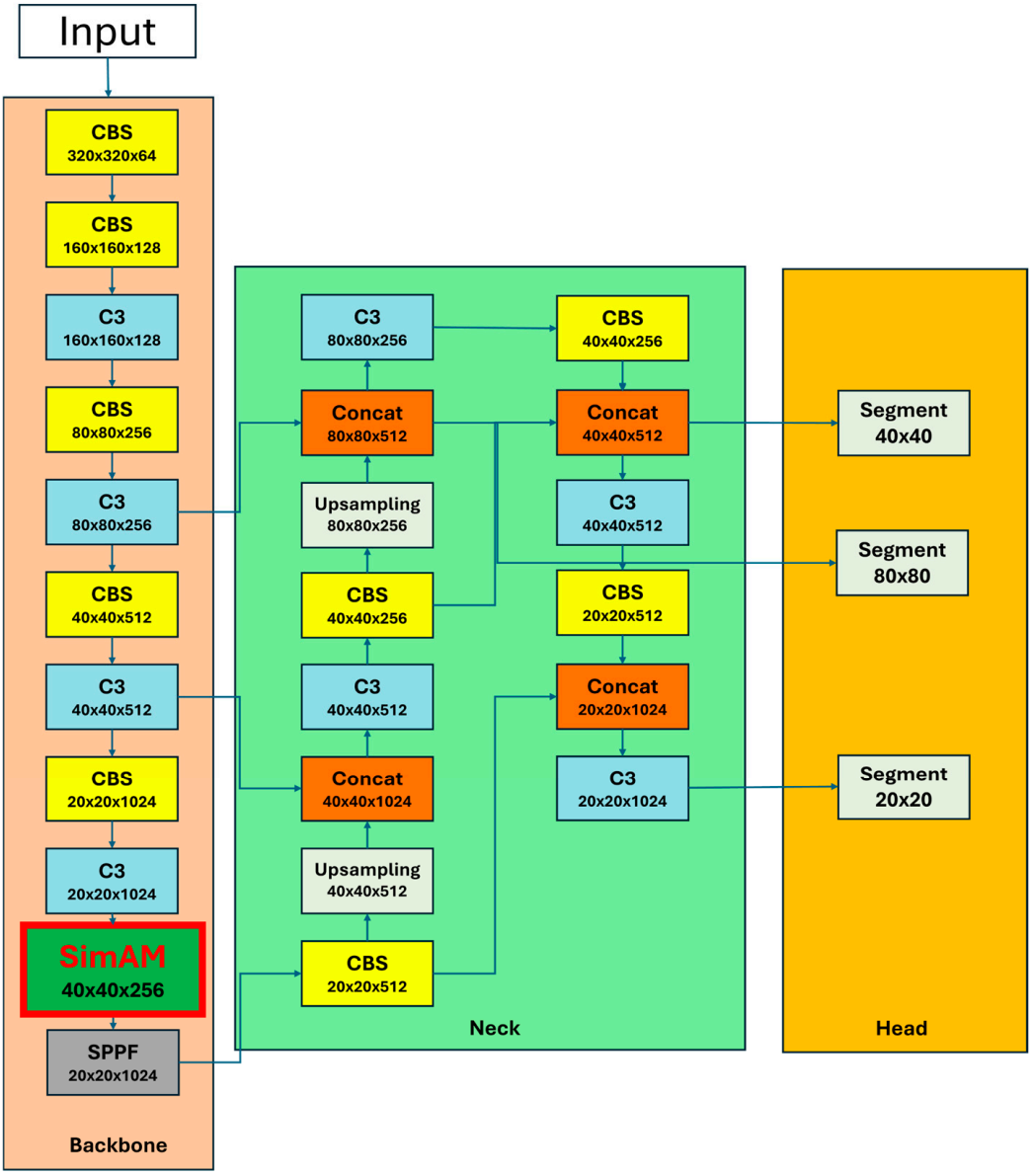


Figure 5. Optimized YOLOv5-l-seg model network structure diagram.

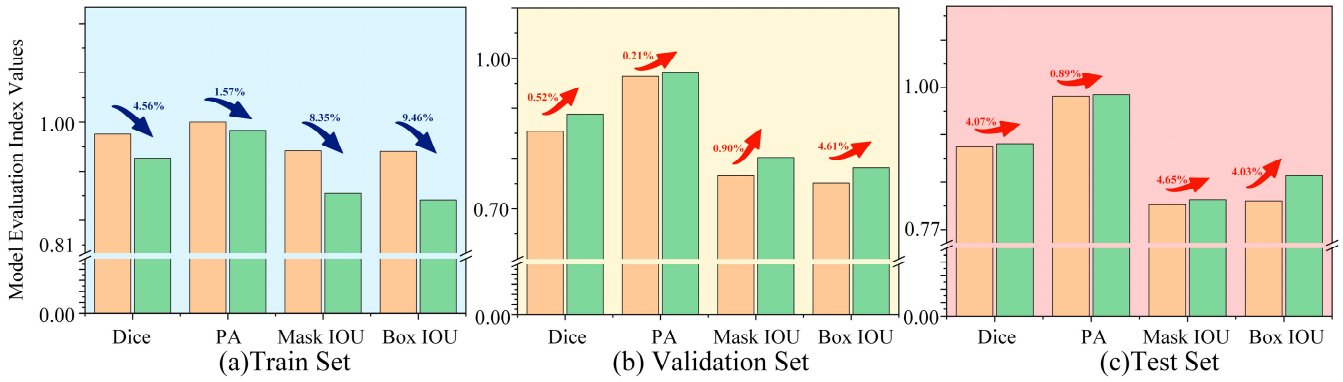


Figure 6. Comparison diagram of YOLOv5-l and YOLOv5-SimAM results: (**a**) Comparison diagram of YOLOv5-l and YOLOv5-SimAM results in train set. (**b**) Comparison diagram of YOLOv5-l and YOLOv5-SimAM results in validation set. (**c**) Comparison diagram of YOLOv5-l and YOLOv5-SimAM results in test set.



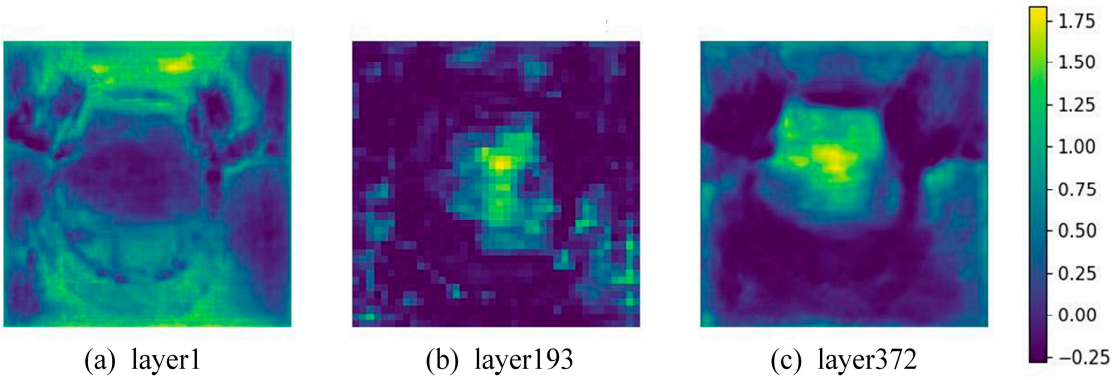
(a) Raw image (b) Sobel-processed image (c) Segmented image

Figure 7. YOLOv5-SimAM model segmentation diagrams: (**a**) Raw image of the tunnel face. (**b**) The tunnel face image processed by Sobel. (**c**) Segmentation result through YOLOv5-SimAm.

3.2. Model Interpretation and Application

3.2.1. Model Interpretation Visualization

To analyze the segmentation process of hexagonal tunnel face images in-depth, several of the main feature layers in the YOLOv5-SimAM network (layers 1, 193, and 372) were visualized to demonstrate the comprehensive feature maps learned in this study, as shown in Figure 8. Here, the main target features used for hexagonal mine tunnel face segmentation are shape and contour. In the shallow feature maps (Figure 8a), not only the shape of the tunnel face is shown but also cluttered background information is present. At deeper network levels, the extracted features are gradually blurred and abstracted (Figure 8b). Eventually, as shown in Figure 8c, the background information is heavily filtered out and the central tunnel surface region is highlighted in some feature maps. After upsampling and stacking, the hexagonal mine tunnel surface and the background information are accurately distinguished by several different blocks of pixels. The shape of the mine hexagonal tunnel face is clearly recognizable in the feature layer 372 map, while the background information has been effectively eliminated. It can also be observed that the model is less clear in segmenting the lower edge of the tunnel face compared to the upper edge. The main factor is that there is some remaining ore in the lower part of the tunnel face when the scraper clears the ore, which causes the lower edge features to be less obvious and affects the model segmentation accuracy.



(a) layer1 (b) layer193 (c) layer372

Figure 8. YOLOv5-SimAM model interpretation heat map: (**a**) Interpretation heat map of YOLOv5-SimAM at layer 1. (**b**) Interpretation heat map of YOLOv5-SimAM at layer 193. (**c**) Interpretation heat map of YOLOv5-SimAM at layer 372.

3.2.2. Tunnel Face Quality Evaluation

For underground mines, the more closely the construction of the tunnel face conforms to the engineering design program, the greater the stability of the rock mass surround-

ing the mine, thus helping to ensure the safety of underground workers and the mine's structure. Accordingly, it is essential to be able to evaluate tunnel face projects in real time. As the YOLOv5-l-SimAM model benefits from a rapid response time, high segmentation speed, and accurate segmentation, it is highly suitable for the real-time evaluation of mining projects.

The model segmentation is performed to obtain the predicted boxes and masks for the completed tunnel faces, with the adaptive masks constructed based on a standard top and bottom/waistline/height ratio of 4:5:6 for the tunnel faces in the current dataset. The center point of the box predicted by the model was used as the starting point for calculations to determine the dimensions of the adaptive mask. Using a width/height ratio threshold of 1.2, if the width/height of the box is >1.2, the mask size is computed based on the width of the box (Figure 9a). If the width/height of the box is ≤1.2, the mask size is calculated based on the height of the box (Figure 9b). This ensures that the mask can effectively cover the predicted tunnel face area. The mask obtained from model segmentation and the newly constructed hexagonal mask are then combined to calculate the Dice coefficient, PA, mask IOU, and box IOU, which evaluate different aspects of the quality of the tunnel face, thus providing a comprehensive assessment of the project overall.

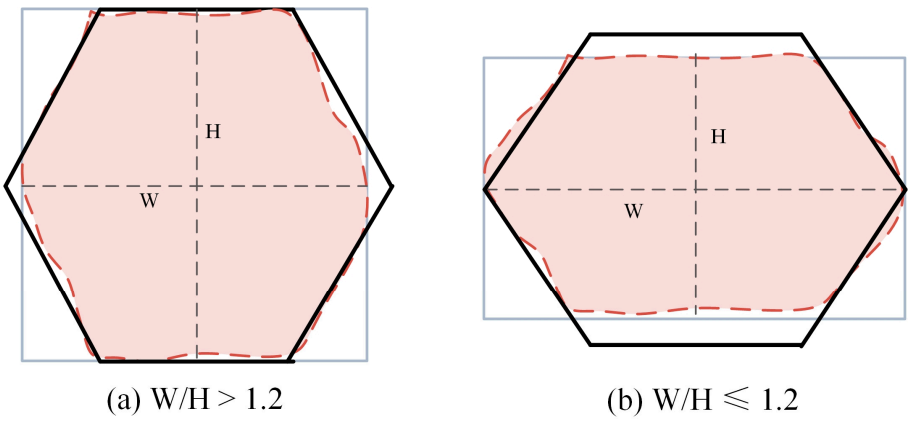
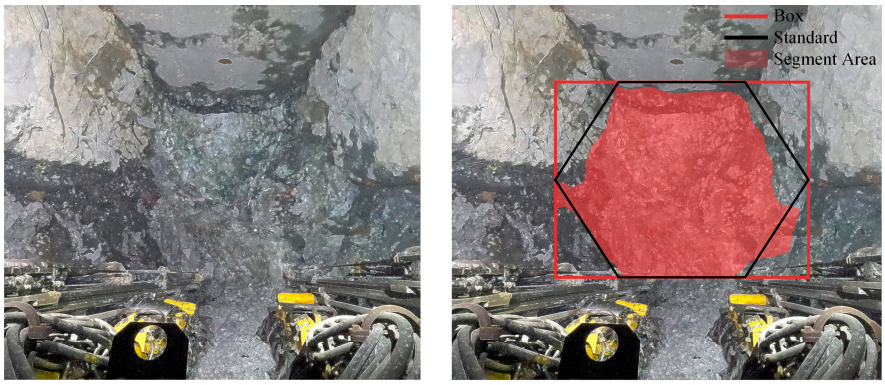


Figure 9. Mask generation baseline diagrams: (**a**) The diagram of mask generation baseline when W/H > 1.2. (**b**) The diagram of mask generation baseline when W/H ≤ 1.2.

Taking a tunnel face image after the ore extraction as an example (Figure 10), the split box width-to-height ratio of 1.3 exceeds the threshold, so the width is used as the basis for constructing the standard hexagon. The PA, Dice coefficient, mask IOU, and box IOU values were 0.91, 0.79, 0.71, and 0.72, respectively, indicating that the tunnel face project is satisfactory and meets the engineering design standards. However, although the overall standard is good, there is some divergence from the ideal state. This may be due to factors such as the rock hardness, leading to under-excavation and over-excavation, the development and distribution of joints and fissures, leading to the dissipation of energy during blasting operations, rock collapse after blasting, improper blasting program design, poor standardization due to human factors, etc. Overall, multiple factors can contribute to negative outcomes in terms of the quality of the tunnel face project.

To further improve the quality of the tunnel face project to be more in line with design standards, it is necessary to comprehensively consider factors including the characteristics of the host rock, the properties of the blasting program, and the standardization of the operation process. By evaluating these factors, targeted recommendations can be proposed to help achieve continuous process improvements. In this way, efficient and safe advancements in mining operations can be accomplished, ensuring that each tunnel face can meet the requirements of engineering standards in practice.



(a) Raw image (b)Evaluation result

Figure 10. Tunnel face engineering evaluation diagrams: (**a**) The raw image of tunnel face. (**b**) The result of evaluation tunnel face engineering through YOLOv5-SimAm.

3.2.3. Future Work and Limitations

The model proposed in this study is mainly applied to the field of mine informatization and intelligent blasting in the future. On the one hand, it can be used to evaluate the quality of tunnel face projects and accurately judge the quality of the project operation; on the other hand, combined with the blasting design scheme, it can realize the intelligent drilling and positioning of rock drilling carts, which can effectively improve the accuracy and safety of the blasting operation and provide strong support to the high efficiency and safety of mine mining. From the point of view of quality, speed, reliability and safety, the model shows certain advantages in practical application. By quickly processing a large amount of data, it can complete the assessment of the engineering quality of the tunnel face in a short period of time, saving time for the subsequent operations and ensuring the progress of the project. At the same time, its accurate positioning ability reduces blasting accidents caused by drilling deviation and improves safety.

The current model also has certain limitations. The model accuracy is high, but there is still some room for improvement. The quality of the dataset and the structure of the network can be further improved to reduce the gap with the real value. The size of the dataset also leads to the loss of model segmentation accuracy due to the rock structure changes caused by the complex stress state of the deep underground structure. Therefore, in future research, the dataset will be further collected to diversify the dataset, which will improve the generalizability, robustness, and accuracy of the model, so that the model can not only be applied in mine informatization and intelligent blasting but also contribute to the study of structural stability in underground spaces.

4. Conclusions

In this study, various analyses were performed based on a classic Chinese underground mine tunnel face dataset to build a model based on YOLOv5-seg capable of accurately segmenting the tunnel face. As part of the optimization process, the optimal preprocessing method was selected, and the model's network structure was adjusted. The resulting optimal YOLOv5-seg model allows for accurate and rapid segmentation of the tunnel face. This model was also interpreted visually and applied to the engineering assessment of tunnel faces. The main conclusions of this study are as follows:

(1) Sobel filtering was selected as the optimal preprocessing method for segmenting the tunnel face for the YOLOv5-seg model, with improved Dice coefficient, PA, box IOU, and mask IOU values of 0.89, 0.97, 0.81, and 0.81, respectively, achieved on the validation set.

(2) The SimAM module was used to further optimize the network structure of the YOLO-seg model, with increases of 4.07%, 0.89%, 7.40%, and 4.03% obtained for the Dice, PA, box IOU, and mask IOU metrics using the optimal model on the test set.

(3) The positive effect of SimAM on the model was proved by model interpretation visualization, and a tunnel face was evaluated based on image data, confirming the example tunnel face in row 10 of the classic Chinese underground mine met the required engineering standards.

Author Contributions: C.M.: Conceptualization, Methodology, Data curation, Software, Visualization, Writing—Original draft preparation, Writing—Reviewing and Editing; K.L.: Software; J.P.: Methodology; J.Z.: Methodology; Q.Z.: Conceptualization; C.Q.: Conceptualization, Data curation, Project administration, Writing—Original draft preparation, Writing—Reviewing and Editing, Funding. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Natural Science Foundation of Hunan Province, China (No. 2024JJ2074) and the Young Elite Scientists Sponsorship Program by CAST (No. 2023QNRC001). This work was also supported in part by the High Performance Computing Center of Central South University.

Data Availability Statement: The data that have been used are confidential.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Zhou, Y.; Liu, Y.; Niu, J. Role of mineral-based industrialization in promoting economic growth: Implications for achieving environmental sustainability and social equity. *Resour. Policy* **2024**, *88*, 104396. [CrossRef]
2. Worlanyo, A.S.; Jiangfeng, L. Evaluating the environmental and economic impact of mining for post-mined land restoration and land-use: A review. *J. Environ. Manag.* **2021**, *279*, 111623. [CrossRef] [PubMed]
3. Saadatmand Hashemi, A.; Katsabanis, P. Tunnel face preconditioning using destress blasting in deep underground excavations. *Tunn. Undergr. Space Technol.* **2021**, *117*, 104126. [CrossRef]
4. Odeyar, P.; Apel, D.B.; Hall, R.; Zon, B.; Skrzypkowski, K. A Review of Reliability and Fault Analysis Methods for Heavy Equipment and Their Components Used in Mining. *Energies* **2022**, *15*, 6263. [CrossRef]
5. Skrzypkowski, K.; Zagórski, K.; Zagórska, A.; Apel, D.B.; Wang, J.; Xu, H.; Guo, L. Choice of the Arch Yielding Support for the Preparatory Roadway Located near the Fault. *Energies* **2022**, *15*, 3774. [CrossRef]
6. Zhou, N.; Dong, C.; Zhang, J.; Meng, G.; Cheng, Q. Influences of mine water on the properties of construction and demolition waste-based cemented paste backfill. *Constr. Build. Mater.* **2021**, *313*, 125492. [CrossRef]
7. Seguel, F.; Palacios-Játiva, P.; Azurdia-Meza, C.A.; Krommenacker, N.; Charpentier, P.; Soto, I. Underground Mine Positioning: A Review. *IEEE Sens. J.* **2022**, *22*, 4755–4771. [CrossRef]
8. Cheng, Y.; Tian, Z.; Ning, D.; Feng, K.; Li, Z.; Chauhan, S.; Vashishtha, G. Computer vision-based non-contact structural vibration measurement: Methods, challenges and opportunities. *Measurement* **2025**, *243*, 116426. [CrossRef]
9. Zhang, Y.; Liu, Y.; Sun, P.; Yan, H.; Zhao, X.; Zhang, L. IFCNN: A general image fusion framework based on convolutional neural network. *Inf. Fusion* **2020**, *54*, 99–118. [CrossRef]
10. Tian, H.; Wang, T.; Liu, Y.; Qiao, X.; Li, Y. Computer vision technology in agricultural automation—A review. *Inf. Process. Agric.* **2020**, *7*, 1–19. [CrossRef]
11. Pedota, M. Big data and dynamic capabilities in the digital revolution: The hidden role of source variety. *Res. Policy* **2023**, *52*, 104812. [CrossRef]
12. Zhao, L.; Hao, S.; Song, Z. Context-aware semantic segmentation network for tunnel face feature identification. *Autom. Constr.* **2024**, *165*, 105560. [CrossRef]
13. Qiao, H.; Yang, X.; Liang, Z.; Liu, Y.; Ge, Z.; Zhou, J. A Method for Extracting Joints on Mountain Tunnel Faces Based on Mask R-CNN Image Segmentation Algorithm. *Appl. Sci.* **2024**, *14*, 6403. [CrossRef]
14. Luo, Y.; Huang, Y.; Wang, Q.; Yuan, K.; Zhao, Z.; Li, Y. An improved YOLOv5 model: Application to leaky eggs detection. *LWT* **2023**, *187*, 115313. [CrossRef]
15. Shahriar, M.T.; Li, H. A Study of Image Pre-processing for Faster Object Recognition. *arXiv* **2020**, arXiv:2011.06928.
16. Khan, M.F.; Khan, E.; Abbasi, Z.A. Image contrast enhancement using normalized histogram equalization. *Optik* **2015**, *126*, 4868–4875. [CrossRef]

17. Kuran, U.; Kuran, E.C. Parameter selection for CLAHE using multi-objective cuckoo search algorithm for image contrast enhancement. *Intell. Syst. Appl.* **2021**, *12*, 200051. [CrossRef]
18. Jana, S.; Parekh, R.; Sarkar, B. Chapter 3—A semi-supervised approach for automatic detection and segmentation of optic disc from retinal fundus image. In *Handbook of Computational Intelligence in Biomedical Engineering and Healthcare*; Nayak, J., Naik, B., Pelusi, D., Das, A.K., Eds.; Academic Press: Cambridge, MA, USA, 2021; pp. 65–91.
19. Mlsna, P.A.; Rodríguez, J.J. 4.13—Gradient and Laplacian Edge Detection. In *Handbook of Image and Video Processing*, 2nd ed.; Bovik, A.L., Ed.; Academic Press: Burlington, NJ, USA, 2005; pp. 535–553.
20. Deng, G.; Cahill, L.W. An adaptive Gaussian filter for noise reduction and edge detection. In Proceedings of the 1993 IEEE Conference Record Nuclear Science Symposium and Medical Imaging Conference, San Francisco, CA, USA, 31 October–6 November 1993; Volume 1613, pp. 1615–1619.
21. Khanam, R.; Hussain, M. What is YOLOv5: A deep look into the internal features of the popular object detector. *arXiv* **2024**, arXiv:2407.20892.
22. Qu, Z.; Gao, L.-y.; Wang, S.-y.; Yin, H.-n.; Yi, T.-m. An improved YOLOv5 method for large objects detection with multi-scale feature cross-layer fusion network. *Image Vis. Comput.* **2022**, *125*, 104518. [CrossRef]
23. Fan, Y.; Zhang, S.; Feng, K.; Qian, K.; Wang, Y.; Qin, S. Strawberry Maturity Recognition Algorithm Combining Dark Channel Enhancement and YOLOv5. *Sensors* **2022**, *22*, 419. [CrossRef]
24. Siddique, N.; Paheding, S.; Elkin, C.P.; Devabhaktuni, V. U-Net and Its Variants for Medical Image Segmentation: A Review of Theory and Applications. *IEEE Access* **2021**, *9*, 82031–82057. [CrossRef]
25. Jiang, Z.; Tahmasebi, P.; Mao, Z. Deep residual U-net convolution neural networks with autoregressive strategy for fluid flow predictions in large-scale geosystems. *Adv. Water Resour.* **2021**, *150*, 103878. [CrossRef]
26. Guo, M.-H.; Xu, T.-X.; Liu, J.-J.; Liu, Z.-N.; Jiang, P.-T.; Mu, T.-J.; Zhang, S.-H.; Martin, R.R.; Cheng, M.-M.; Hu, S.-M. Attention mechanisms in computer vision: A survey. *Comput. Vis. Media* **2022**, *8*, 331–368. [CrossRef]
27. Yang, L.; Zhang, R.-Y.; Li, L.; Xie, X. SimAM: A Simple, Parameter-Free Attention Module for Convolutional Neural Networks. In Proceedings of the International Conference on Machine Learning, Virtual, 18–24 July 2021.
28. Han, G.; Huang, S.; Zhao, F.; Tang, J. SIAM: A parameter-free, Spatial Intersection Attention Module. *Pattern Recognit.* **2024**, *153*, 110509. [CrossRef]
29. Wang, Z.; Wang, E.; Zhu, Y. Image segmentation evaluation: A survey of methods. *Artif. Intell. Rev.* **2020**, *53*, 5637–5674. [CrossRef]
30. Kovalkiy, E.; Kongar-Syuryun, C.; Sirenko, Y.; Mironov, N. Modeling of rheological deformation processes for room and pillar mining at the Verkhnekamsk potash salt deposit. *Sustain. Dev. Mt. Territ.* **2024**, *16*, 1017–1030. [CrossRef]
31. Chen, L.-C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In Proceedings of the Computer Vision—ECCV 2018, Munich, Germany, 8–14 September 2018; pp. 833–851.
32. Litjens, G.; Kooi, T.; Bejnordi, B.E.; Setio, A.A.A.; Ciompi, F.; Ghafoorian, M.; van der Laak, J.A.W.M.; van Ginneken, B.; Sánchez, C.I. A survey on deep learning in medical image analysis. *Med. Image Anal.* **2017**, *42*, 60–88. [CrossRef]
33. Liu, B.; Dolz, J.; Galdran, A.; Kobbi, R.; Ben Ayed, I. Do we really need dice? The hidden region-size biases of segmentation losses. *Med. Image Anal.* **2024**, *91*, 103015. [CrossRef]
34. Brar, K.K.; Goyal, B.; Dogra, A.; Mustafa, M.A.; Majumdar, R.; Alkhayyat, A.; Kukreja, V. Image segmentation review: Theoretical background and recent advances. *Inf. Fusion* **2025**, *114*, 102608. [CrossRef]
35. Cho, Y.-J. Weighted Intersection over Union (wIoU) for evaluating image segmentation. *Pattern Recognit. Lett.* **2024**, *185*, 101–107. [CrossRef]
36. Fernandes Junior, F.E.; Nonato, L.; Ranieri, C.; Ueyama, J. Memory-Based Pruning of Deep Neural Networks for IoT Devices Applied to Flood Detection. *Sensors* **2021**, *21*, 7506. [CrossRef] [PubMed]
37. Yang, L.; Zhang, K.; Liu, J.; Bi, C. Location IoU: A New Evaluation and Loss for Bounding Box Regression in Object Detection. In Proceedings of the 2024 International Joint Conference on Neural Networks (IJCNN), Yokohama, Japan, 30 June–5 July 2024; pp. 1–8.
38. Padilla, R.; Netto, S.L.; Silva, E.A.B.d. A Survey on Performance Metrics for Object-Detection Algorithms. In Proceedings of the 2020 International Conference on Systems, Signals and Image Processing (IWSSIP), Niteroi, Brazil, 1–3 July 2020; pp. 237–242.
39. Bai, Y.; Guo, Y.; Zhang, Q.; Cao, B.; Zhang, B. Multi-network fusion algorithm with transfer learning for green cucumber segmentation and recognition under complex natural environment. *Comput. Electron. Agric.* **2022**, *194*, 106789. [CrossRef]
40. Kumar, A.; Srivastava, S. Object Detection System Based on Convolution Neural Networks Using Single Shot Multi-Box Detector. *Procedia Comput. Sci.* **2020**, *171*, 2610–2617. [CrossRef]
41. Peng-o, T.; Chaikan, P. High performance and energy efficient sobel edge detection. *Microprocess. Microsyst.* **2021**, *87*, 104368. [CrossRef]

42. Wang, M.; Yang, W.; Wang, L.; Chen, D.; Wei, F.; KeZiErBieKe, H.; Liao, Y. FE-YOLOv5: Feature enhancement network based on YOLOv5 for small object detection. *J. Vis. Commun. Image Represent.* **2023**, *90*, 103752. [CrossRef]
43. Li, N.; Ye, T.; Zhou, Z.; Gao, C.; Zhang, P. Enhanced YOLOv8 with BiFPN-SimAM for Precise Defect Detection in Miniature Capacitors. *Appl. Sci.* **2024**, *14*, 429. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.