MDPI

Article

Application of Target Detection Based on Deep Learning in Intelligent Mineral Identification

Luhao He [1,2,3], Yongzhang Zhou [1,2,3,*] and Can Zhang [1,2,3]

1   School of Earth Science and Engineering, Sun Yat-Sen University, Zhuhai 519000, China;
    helh8@mail2.sysu.edu.cn (L.H.); zhangc68@mail2.sysu.edu.cn (C.Z.)
2   Centre for Earth Environment and Resources, Sun Yat-Sen University, Zhuhai 519000, China
3   Guangdong Provincial Key Lab of Geological Process and Mineral Resources, Zhuhai 519000, China
*   Correspondence: zhouyz@mail.sysu.edu.cn

Abstract: In contemporary society, rich in mineral resources, efficiently and accurately identifying and classifying minerals has become a prominent issue. Recent advancements in artificial intelligence, particularly breakthroughs in deep learning, have offered new solutions for intelligent mineral recognition. This paper introduces a deep-learning-based object detection model for intelligent mineral identification, specifically employing the YOLOv8 algorithm. The model was developed with a focus on seven common minerals: biotite, quartz, chalcocite, silicon malachite, malachite, white mica, and pyrite. During the training phase, the model learned to accurately recognize and classify these minerals by analyzing and annotating a large dataset of mineral images. After 258 rounds of training, a stable model was obtained with high performance on key indicators such as Precision, Recall, mAP_{50}, and mAP_{50-95}, with values stable at 0.91766, 0.89827, 0.94300, and 0.91696, respectively. In the testing phase, using samples provided by the Geological and Mineral Museum at the School of Earth Sciences and Engineering, Sun Yat-sen University, the model successfully identified all test samples, with 83% of them having a confidence level exceeding 87%. Despite some potential misclassifications, the results of this study contribute valuable insights and practical experience to the development of intelligent mineral recognition technologies.

Keywords: mineral recognition; deep learning; machine learning; convolutional neural network; image recognition

1. Introduction

Mineral resources constitute a crucial lifeline for the development of human society and economic prosperity [1], and the advancement of intelligent mineral identification technology holds significant implications in facilitating the exploration, exploitation, and management of mineral resources [2]. Conventional methods of mineral identification typically depend on manual observation and experience, featuring low identification efficacy and vulnerability to subjective factors, thereby making it challenging to satisfy the demands of large-scale and high-efficiency mineral exploration. Nevertheless, along with the continuous advancement of deep-learning technology [3], deep-learning-based intelligent mineral identification technologies [4] has progressively emerged as a research hotspot and frontier.

Deep learning is a machine-learning method derived from artificial neural networks, which learns features from data through a multi-level neural network structure and automatically discovers data rules and patterns [5]. Zeng et al. (2020) [6] proposed a method combining mineral photo images and Moh hardness in a deep neural network to improve accuracy and expand the number of identified minerals. Experimental results showed that the top-one accuracy of the method reached 90.6% and top-five accuracy reached 99.6% for 36 common minerals. Wang et al. (2023) [7] studied the intelligent recognition of volcanic

rock slice images by using a deep residual contraction neural network model. The study investigated 11 basic types of volcanic rocks, and collected 12,000 high-definition images of rock slices using electron polarizing microscopy and a series of optimizations and improvements of network model types, with an accuracy of over 92% in the test set classification results; Zhang et al. (2019) [8] developed an intelligent recognition model for rock and mineral microscopy images by using an integrated machine-learning algorithm based on the Inception-v3 architecture. Different methods such as logistic regression, support vector machines, and multi-layer perceptrons were used to create a model that could distinguish minerals from the sample data of potassium feldspar, feldspar, plagioclase, and quartz. The accuracy rate is up to 90.9%; in addition, Zhang et al. (2023) [9] explored the use of electrochemical methods for mineral material identification in geology, emphasizing the need for efficient identification techniques in this field.

Target detection refers to the detection of the location and category of a specific target in an image or video, and the key is to achieve the target location positioning and category recognition simultaneously. By training on a large amount of data, the deep-learning model can automatically learn the features in the image, and can efficiently identify the target object, thus showing great potential in the target detection task [10]. In the field of intelligent mineral recognition, deep-learning-based target detection techniques can be developed by analyzing features in mineral images, realizing the automatic identification and classification of different minerals, and providing a whole new method for the exploration and management of mineral resources; deep-learning models such as YOLO (You Only Look Once) [11], RetinaNet [12], Faster R-CNN (Faster Region-based Convolutional Neural Network) [13], EfficientDet [14], and Mask R-CNN [15] have been widely used in object detection tasks, and have achieved an excellent performance. The research status of various methods is as follows: (1) YOLO: Traditional geological mineral identification methods have been supplemented by deep-learning models, such as YOLO-FIRI, which improves object detection in infrared images [16]. These models are also applied to ore image classification, which can identify mineral properties in real time [17]. YOLO-HR is an improved version of YOLOv5 that performs well in object detection in high-resolution images, including mineral detection [18]. Specifically, the YOLOv4 model has been customized for rock mineral identification in thin slices, improving the accuracy and speed of detection [19]. Object detection techniques such as YOLO-v7 have also been used in microfossil research to detect microfossil fish teeth and denticles, demonstrating the versatility of these models in various fields [20]. In addition, the application of deep learning in mineral detection extends to hydrothermal emission signatures, especially in seafloor massive sulfide deposits, which are valuable mineral resources [21]. On the other hand, the use of near-infrared (NIR) technology to rapidly determine the soil mineral nitrogen content has been explored, demonstrating multiple applications of image-sensing systems in mineral research [22]. Overall, integrating deep-learning models such as YOLO into mineral detection processes has demonstrated great potential in enhancing the accuracy, speed, and efficiency of mineral identification and classification in diverse geological and research settings [23,24]; (2) RetinaNet: RetinaNet is an enhanced convolutional neural network (CNN) trained using digital elevation data, which has been used in various applications in geology. In a study comparing Mars and Earth ice fights, RetinaNet was used to analyze the geological features on the two planets [25]. In addition, the GeoImageNet dataset is used to test the RetinaNet and other object detection models (such as Faster-RCNN), demonstrating its effectiveness in identifying natural features [26]. On the other hand, the researchers also explored the use of a RetinaNet model combined with Google Earth images to automatically identify ice buckets, demonstrating the ability of the model to detect large-scale geological structures [27]. RetinaNet application in geology extends to the detection and mapping of lunar rockfall, highlighting its potential in planetary research [28]. Furthermore, RetinaNet has been used to detect boulders in the lateral sweep sonar mosaic, demonstrating its versatile in various geological environments [29]. Overall, RetinaNet has become a valuable tool in geology, providing a powerful solution for object detection and geological feature identification in

different environments [30]; (3) Faster R-CNN: Ma et al. (2021) [31] proposed a classification network that fused VGG16 and MobileNet, and used the fusion network to optimize the Faster R-CNN target detection network. In addition, the Faster R-CNN model has been applied to the multi-objective intelligent recognition of lithographic thin slice images, and the backbone network is based on ResNet50 [32]. The effectiveness of Mask R-CNN in the segmentation of mineral particle instances in digital images has also been proved [33]. Liu et al. (2020) [34] used a variety of rock image data and a simplified VGG16 network to extract rock image features under the Faster R-CNN framework, and successfully trained a rock-type recognition system. The recognition accuracy of this system is over 96% for single-type rock, and over 80% for multi-type rock mixed image recognition; (4) Efficient-Det: Radulescu et al. (2024) [35] used a custom mineral dataset to compare quantized EfficientDet models with floating-point models, with the aim of optimizing mineral recognition for sustainable resource extraction. The research results emphasize the efficiency and effectiveness of the quantized EfficientDet model in mineral detection. Munteanu et al. (2022) [36] proposed the use of EfficientDet architecture for quantized deep learning in mineral recognition during mining processes. This method aims to improve the accuracy and speed of mineral detection in mining operations. The study demonstrates the potential of EfficientDet in improving the efficiency of mineral recognition tasks. Additionally, Jia et al. (2022) [37] explored underwater object detection based on an improved version of EfficientDet. The study focuses on the intelligent detection of marine life and emphasizes the importance of quickly and accurately detecting marine life for the marine economy. The study highlights the importance of EfficientDet in enhancing the detection capabilities for marine applications. Overall, EfficientDet has shown promising results in various object detection tasks, including mineral detection and marine life detection. Its scalability, efficiency, and accuracy make it a valuable tool for optimizing resource extraction processes and enhancing detection capabilities for marine applications [38,39]; (5) Mask R-CNN: Dong et al. (2021) [40] proposed a deep-sea nodule mineral image segmentation algorithm based on Mask R-CNN to enhance the segmentation performance. Similarly, deep-sea nodule mineral image segmentation algorithms have also been evaluated using various methods, such as U-Net, improved U-Net, Conditional Generative Adversarial Network (CGAN), and Mask R-CNN. Muhammad Ridwan Iyas et al. (2020) [41] proposed a deep-sea nodule mineral image segmentation algorithm based on Mask R-CNN and conducted a comparative analysis with different deep-learning methods such as U-Net and Generative Adversarial Network. Experimental results show that the Mask R-CNN-based method outperforms the methods based on U-Net, improved U-Net, and Conditional Generative Adversarial Network on the dataset. Koh et al. (2021) [42] demonstrated the application of transfer learning using Mask R-CNN for mineral segmentation, highlighting its potential for rapid and automatic segmentation. In addition, Caldas et al. (2024) [43] also discussed the use of Mask R-CNN for the phase characterization of pelleted feeds and identification of iron minerals in optical microscope images. Overall, YOLO is a model that performs object detection tasks in a single forward propagation, in an extremely fast but slightly less accurate manner; RetinaNet introduced Focal Loss to deal with category imbalance, which improved the detection accuracy of small targets, but the speed was slightly slower. Based on the regional proposal network (RPN), Faster R-CNN generates candidate regions and then performs accurate classification and regression, with a high accuracy but slow speed. EfficientDet uses EfficientNet as the backbone network, and BiFPN is used to improve the efficiency of multi-scale feature fusion, which demonstrates both efficient computation and high precision. Mask R-CNN adds an instance segmentation branch on the basis of Faster R-CNN, which can complete both target detection and segmentation tasks, but the computational complexity and speed are slow.

The YOLO series of algorithms [44] has been widely used in many fields, highlighting its strong versatility and robustness. The following are typical examples of research using YOLO algorithms in different fields, which further confirm their broad application prospects. Biology: The YOLO algorithm was employed by Abdullah et al. (2022) [45]

for the real-time detection of fish species in underwater videos, significantly improving the accuracy and efficiency of marine biodiversity research. Traffic Monitoring: Zuraim et al. (2021) [46] successfully applied the YOLO algorithm in traffic surveillance systems, achieving precise vehicle detection and tracking, which provided strong support for the analysis of traffic patterns and improvements in urban planning. Agriculture: Vilar-Andreu et al. (2024) [47] developed a pest detection system for crops based on the YOLO algorithm. This system allows for the timely identification and control of pests, effectively enhancing crop yield and quality. Medical Imaging: The YOLO algorithm was utilized by Prinzi et al. (2024) [48] to detect tumors in mammograms, greatly improving the accuracy of early tumor diagnosis and treatment outcomes. Industrial Applications: Reddy et al. (2024) [49] applied the YOLO algorithm in semiconductor manufacturing to achieve precise defect detection, effectively reducing production waste and enhancing production efficiency. As the eighth version in the YOLO series, the YOLOv8 algorithm [50] has undergone further optimizations and enhancements in its architecture. Specifically, YOLOv8 utilizes a deeper Darknet53 network as its backbone, aiming to improve the receptive field and feature representation capabilities of the network. Additionally, the algorithm incorporates feature fusion modules such as Spatial Pyramid Pooling (SPP) and Path Aggregation Network (PAN), which help the network effectively extract features at different scales, thereby enhancing its ability to detect small objects. To further improve the accuracy and robustness of object detection, YOLOv8 employs a strategy of cascading multi-scale feature maps. Moreover, by optimizing the size and aspect ratio settings of anchor boxes and integrating model ensemble and optimization strategies, the algorithm significantly boosts detection accuracy. The performance comparison of the YOLO series algorithms is shown in Figure 1, where the substantial improvements of YOLOv8 in metrics such as mean Average Precision (mAP) and computation time (ms/img) are clearly demonstrated. This study will be based on the YOLOv8 algorithm [51] and select the following seven minerals as the research objects in mineral intelligent identification. Because biotite, quartz, bornite, chrysocollae, malachite, muscovite, and pyrite have important representative and extensive application value in the field of geology and mineralogy, this is embodied in the following: (1) Biotite is a common silicate mineral in metamorphic and igneous rocks, and its existence can indicate the genesis and degree of metamorphism of rocks, and is also one of the indicators of prospecting; (2) Quartz is one of the most common minerals in the earth's crust, with a high hardness, transparency, and chemical stability, and has important applications in building materials, glass manufacturing, and other industrial fields; (3) Bornite is a typical copper sulfide, commonly found in copper deposits and mineralized veins. It has important mineral deposit value in geology, can be used as an important indicator mineral in copper exploration and development, and has important applications in mineral deposit exploration, ore dressing, and the metallurgical industry; (4) Chrysocolla is a copper-bearing silicate mineral, commonly found in the copper oxide mineralization zone. In geology, the presence of silico-malachite can indicate the formation environment and type of copper oxide deposit, and has a high value in decoration and handicraft production; (5) Malachite is also a carbonate mineral containing copper, which is commonly found in copper oxide mineralization zones. In geology and mineralogy, the presence of malachite can indicate the formation conditions and geological environment of copper oxide deposits, and also has a certain decorative and technological value; (6) Muscovite is a common silicate mineral, commonly found in metamorphic rocks and granites; (7) Pyrite is a typical sulfide iron ore, commonly found in igneous and sedimentary rocks.
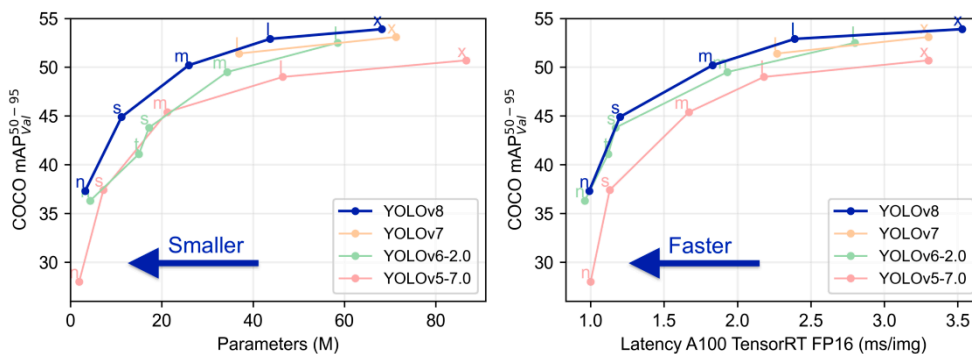
Figure 1. Performance comparison of different versions of the YOLO algorithm [44].

2. YOLOv8 Model

Since its introduction in 2023, YOLOv8 has attracted wide attention with its unique design concept and excellent performance [51]. As the latest generation of target detection frameworks, YOLOv8 not only inherits the consistent accuracy and real-time performance of the YOLO series, but also has achieved major breakthroughs in many aspects. Its overall structure is shown in Figure 1.

2.1. C3 Module and C2f Modules

YOLOv8 [51] introduces significant structural innovations, notably replacing the traditional C3 module from YOLOv5 with the more advanced C2f module. This upgrade, which enhances gradient flow, boosts both training efficiency and model performance. The result is a more lightweight, yet powerful model that excels in complex scenarios.

As shown in Figure 2a, the C3 module, inspired by CSPNet and residual structures, integrates a shunt mechanism with a BottleNeck residual module to enhance model performance. It comprises three convolution modules (Conv + BN + SiLU) and a flexible number of BottleNeck modules, with the final convolution layer doubling the channel count due to combined inputs from the main and secondary gradient flow branches.
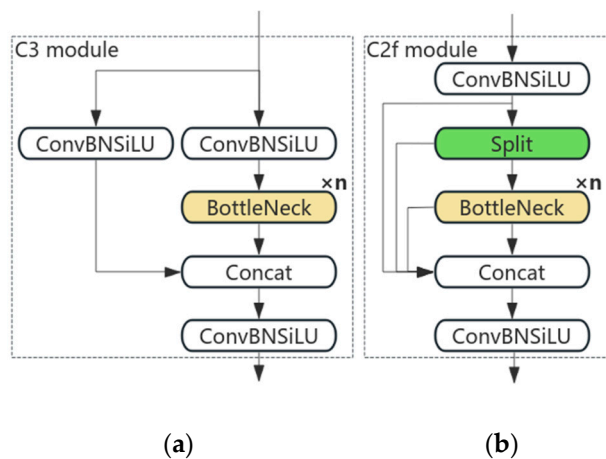

(**a**) (**b**)

Figure 2. C3 module and C2f module structure diagram [50]. (**a**) is mainly designed by combining the shunt idea and residual structure of CSPNet. (**b**) aims to maintain the lightweight of the model while providing richer gradient flow information, thereby improving the training and inference performance of the model.

In Figure 2b, the C2f module is depicted as a series of BottleNeck blocks, each with two convolution layers. The first layer processes the input feature map, which is then split, processed separately, and merged. This approach allows the model to capture richer context information, improving target identification accuracy. The second convolution layer refines the merged feature map before outputting it. The C2f module significantly

enhances model expression and accuracy, making it highly effective in practical target detection applications.

2.2. Backbone Network, Neck Network, and Head Network

As shown in Figure 3, the neck network in YOLOv8 serves as a crucial link between the backbone and head networks, playing a key role in feature fusion and processing. To enhance model performance, YOLOv8 incorporates several innovations in the design of its neck network. Drawing inspiration from YOLOv7's ELAN, it replaces the original C3 module with the more advanced C2f module. This change enriches gradient flow and improves feature utilization through additional cross-layer connections. YOLOv8 also optimizes the number of channels for different scales to better meet diverse detection requirements. Specifically, it replaces the original 6×6 convolution kernel in the neck with a 3×3 kernel, reducing computational load and improving feature extraction efficiency. Additionally, two convolutional layers are removed to simplify the network structure. The C2f module introduces more skip connections and split operations, further enhancing feature diversity and model expressiveness. These improvements collectively make YOLOv8 more accurate, efficient, and robust in target detection, particularly in complex scenarios.
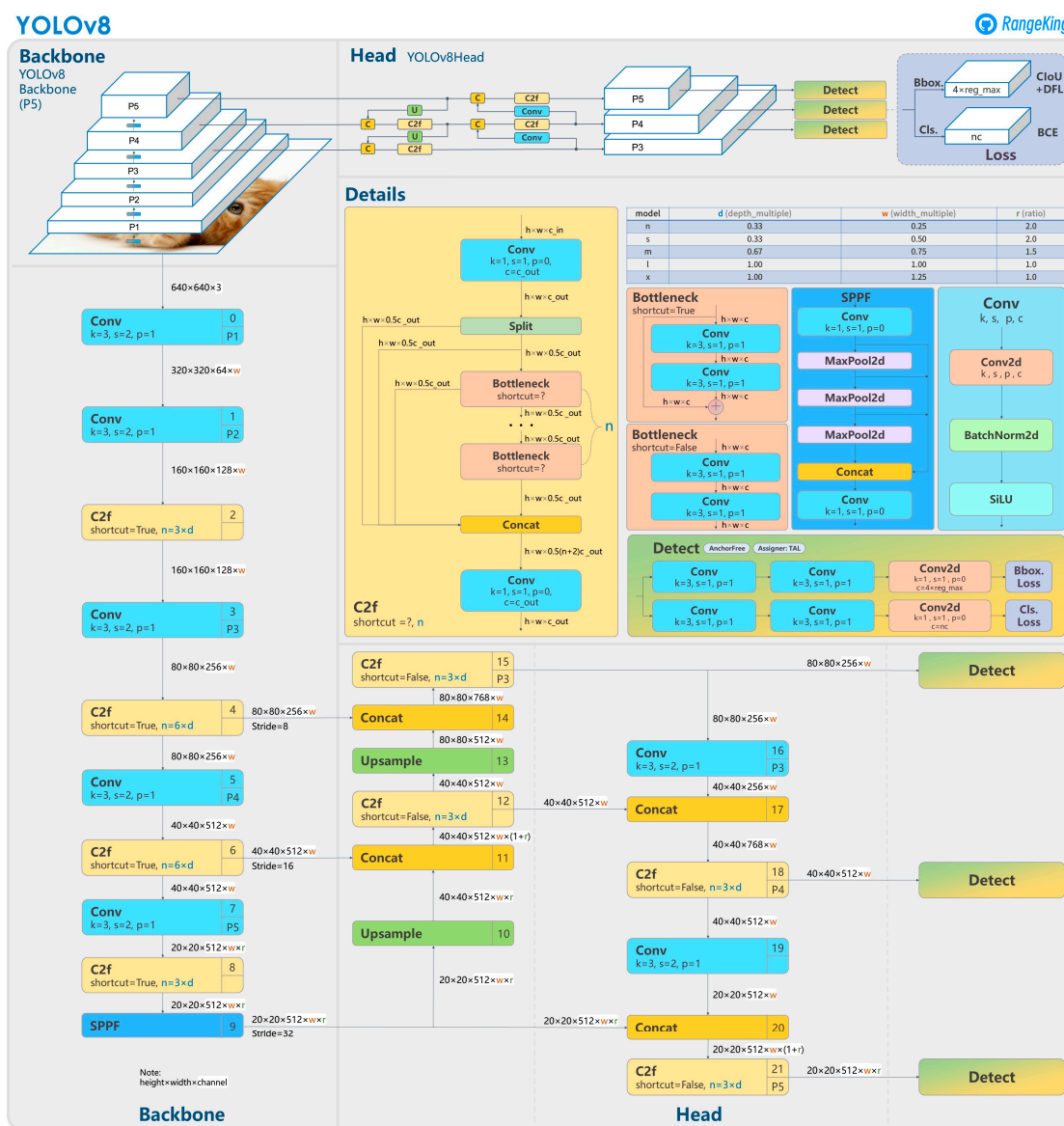

Figure 3. YOLOv8 overall structure diagram of the model [51].

2.3. Decoupled-Head

As shown in Figure 4, in terms of task division, YOLOv8 also demonstrates a well-thought-out design philosophy [51]. Traditional object detection frameworks tend to bundle classification and location tasks together and share the same set of parameters. In YOLOv8, however, the engineers boldly adopted a strategy of separating the classification and locating branches, combined with a "Decoupled-Head structure that does not share parameters". This design effectively avoids the inherent conflict between classification and regression tasks, thus improving the performance and accuracy of the model.
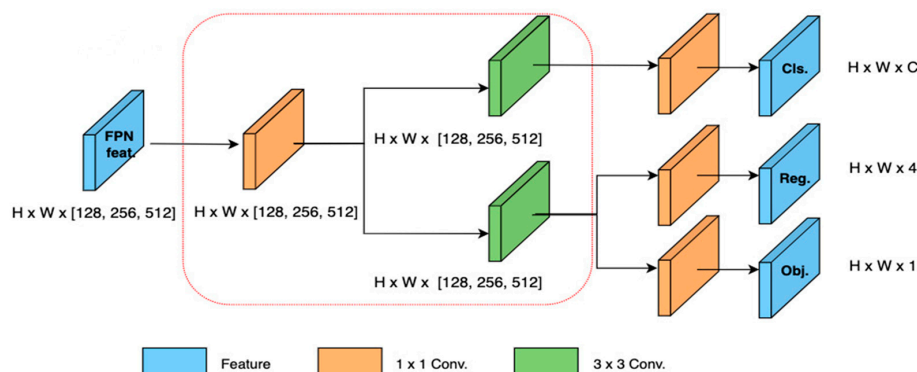

Figure 4. Flow chart of Decoupled-Head [46].

2.4. SPPF Module

As shown in Figure 5, the SPPF (Spatial Pyramid Pooling Fusion) module [52] enhances feature extraction by handling feature maps of varying sizes and improving target detection performance. It addresses the issue of information loss from fixed-size pooling by employing a pyramid pooling layer to capture multi-scale features. The SPPF module first applies multi-scale pooling to the input feature map, using a pyramidal structure to create grids of different scales and pooling each grid to generate diverse feature representations. These multi-scale features are then fused, typically through concatenation or summation, to form a comprehensive feature set. This fusion improves the model's ability to recognize objects at various scales. To reduce the computational complexity and parameter count, the module often includes a dimensionality reduction step, achieved through convolutional or fully connected layers. The output features, rich in spatial and semantic information, are then passed to subsequent network layers, enhancing target detection and recognition.
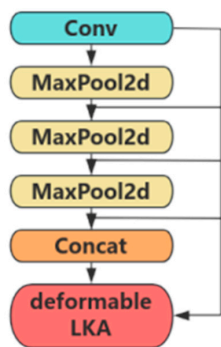

Figure 5. Flow chart of Decoupled-Head [47].

2.5. Label Assignment with a Loss Function

In the label allocation [53] and loss function [54] in the selection, YOLOv8 once again showed its unique insight. Traditional object detection frameworks often rely on anchor boxes for label assignment and loss calculation, but the settings of anchor boxes often need to be adjusted according to the specific dataset, and it is easy to introduce hyperparameters. To overcome this challenge, YOLOv8 abandons the frame-based idea

in favor of a more flexible frameless approach. This method not only simplifies the label allocation process, but also avoids the problems caused by the anchor box setting. To further ensure consistency between classification and regression tasks, YOLOv8 introduces the Task Alignment Learning (TAL) dynamic assignment strategy. This strategy can dynamically adjust the weights of classification and regression tasks according to the characteristics of the tasks, so that the model pays more attention to the important tasks in the training process. In addition, YOLOv8 [55] also combines DFL Loss with CIoU Loss as a regression loss function. DFL Loss focuses on the detection performance of objects at different scales, while CIoU Loss better measures the overlap between the predicted frame and the real frame. This combination enhances YOLOv8's performance in detection tasks.

3. Dataset Construction and Preprocessing

3.1. Acquisition of the Mineral Image Dataset

The deep feature learning ability of the model is closely related to its internal structure and the training data used. As the diversity of the training dataset increases, the more extensive the situation covers, and the more complete the features extracted by the model are. This not only enhances the generalization and extrapolation capabilities of the model, but also enables it to handle more complex classification tasks. This diverse trait is crucial for training, verifying, and identifying the effectiveness of models, especially in scientific research and field work. The sample sources of this dataset are relatively diverse. The main sources of this dataset are two public datasets (Minerals Identification Dataset [56] and Mineralogy Database [57]) and one non-public dataset on the Internet (on-site shooting at the Museum of Geology and Minerals, School of Earth Science and Engineering, Sun Yat-sen University). Seven minerals were selected, including biotite, quartz, bornite, chrysocolla, malachite, muscovite, and pyrite, with a total of 8540 photos. There are 6743 training sets, 1692 verification sets, and 1400 test sets. The classification of each dataset is shown in Table 1, and the sample dataset is shown in Table 2.

Table 1. Dataset classification table.

Mineral Species	Training Sets Num (pics)	Verification Sets Num (pics)	Test Sets Num (pics)
biotite	1073	290	200
quartz	1185	240	200
bornite	817	192	200
chrysocolla	740	182	200
malachite	998	258	200
muscovite	844	232	200
pyrite	1086	298	200

Table 2. Example of the dataset samples [56,57].

Table 2. *Cont.*

Training Sets	Verification Sets	Test Sets
quartz		
bornite		
chrysocolla		
malachite		

Table 2. *Cont.*

	Training Sets	Verification Sets	Test Sets
muscovite			
pyrite			

3.2. Data Preprocessing and Annotation

In this experiment, a series of data augmentation strategies was meticulously designed and implemented to enhance the generalization capability of the neural network, adhering to principles of scientific rigor. Compared to traditional methods of acquiring new data, data augmentation offers significant cost advantages. It expands the dataset through algorithmic processing, without the need for additional data collection and annotation efforts. This efficient and economical approach makes data augmentation a crucial technique in deep-learning experiments. Specifically, the augmentation techniques employed include flipping, translation, and rotation: (1) Flipping: Horizontal and vertical flips were applied to simulate variations in the orientation of minerals, allowing the model to learn from images in different directions. This is particularly beneficial for detecting minerals that may appear in varying positions within the image. (2) Translation: Small random translations were used to slightly shift the mineral images in different directions. This technique helps the model better handle scenarios where minerals are not perfectly centered in the image and ensures that the model does not become overly sensitive to the exact position of the minerals, thereby improving its ability to detect and classify minerals under less ideal conditions. (3) Rotation: Rotation augmentation was employed to simulate different perspectives of the minerals, accounting for their potential irregular shapes and orientations. This enhances the model's ability to detect and classify minerals from various angles, maintaining a high accuracy even when the target objects are rotated in different directions. These operations are intended to simulate the variations in object orientation, position, and angle that occur in real-world scenarios, thereby increasing the model's robustness to such changes.

In terms of data annotation, Labelme is used to manually annotate all data photos. The label format is as follows: <object-class> <x> <y> <width> <height>, where <x> and <y> are the center coordinates of the target box, and <width> <height> is the width and height of the target box, as shown in the following example: "2 0.502778 0.505556 0.994444 0.911111", "3 0.518500 0.521000 0.743000 0.676000", or "6 0.502953 0.505236 0.864173 0.981675 ".

4. Design and Training of Intelligent Mineral Recognition Mode

4.1. Model Training and Initial Parameters

1. Epochs: This parameter defines how many times the entire training dataset is passed through the model. Each epoch involves a full forward and backward pass to update the model's parameters. While the model was set to train for 500 epochs, the optimal performance was achieved at epoch 258.

2. Batch Size: This is the number of samples processed before updating the model parameters in each iteration. A larger batch size enhances GPU utilization and speeds up training, but may lead to high memory consumption. For our model, the batch size was set to 64 after several adjustments.

3. Image Size (Imgsz): This parameter specifies the dimensions of the input images used during training and inference. The image size for our model was set to 640 pixels after testing.

4. Pre-trained Weight File: We utilized YOLOv8x.pt, which includes weights optimized from training on a large-scale dataset. The "x" indicates a larger model variant with a deeper network and more parameters, enhancing performance.

5. Learning Rate: This hyperparameter controls the step size for parameter updates during training. We started with an initial learning rate of 0.0001 and employed a cosine learning rate scheduler to adjust the rate over time for improved convergence.

4.2. Model Evaluation

In object detection, multiple loss functions are usually used to optimize the object detection model, including box_loss (border loss), cls_loss (classification loss), and dfl_loss (detection feature learning loss).

Box_loss optimizes the prediction of the target bounding box in the object detection model, which measures the difference between the model's predicted bounding box and the real bounding box, usually by calculating the coordinate deviation between the predicted bounding box and the real bounding box. As shown in Figure 6, box_loss is usually very high at the beginning of training. The peak value of box_loss of both the training set and verification set is 0.71715 and 1.0679, respectively, in the fourth round. Since the initial parameters of the model are randomly initialized, the prediction of the target bounding box may be very inaccurate. As the training progresses, the model gradually learns better feature representation and target detection techniques, and box_loss gradually decreases. On the training set, the decline in box_loss is usually relatively stable, but there may be some fluctuations or oscillations before the model converges. The bounding loss fitting procedure on the validation set can provide an evaluation of the model's generalization ability. The model performs well on the verification set, and the box_loss of the verification set gradually decreases with the progress of training until it becomes stable, and, finally, the box_loss of the training set and the verification set is stable at 0.30024 and 0.24331, respectively. In the analysis of convergence and overfitting, the box_loss of both the training set and the validation set decrease steadily and become stable, indicating that the model may have converged and achieved a good performance on this task.

Cls_loss optimizes the prediction of the target category in the object detection model, which measures the difference between the predicted target category and the true category. As shown in Figure 7, the classification loss may be relatively high at the beginning of the training due to the random initialization of the model and the choice of the initial learning rate. As the training proceeds, the model may encounter some difficulties, such as unbalanced sample distribution, and confusion between target categories, leading to a rise in classification loss, and the observation of classification loss on the validation set may indicate that the generalization ability of the model is somewhat challenged. The peaks of cls_loss in both the training and validation sets appeared in the fourth epoch, with peaks of 1.3143 and 5.5793, respectively. The classification losses of both the training and validation sets eventually converge and stabilize, even though there is a rise in the early stages of training, which may indicate that the model finally achieves a good generalization

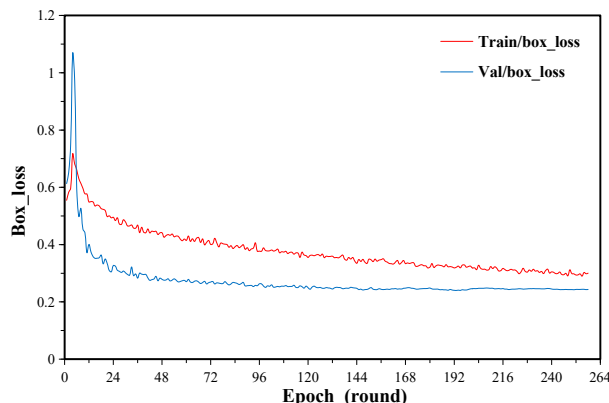performance. The cls_loss of the final training and validation sets were stable at 0.27971 and 0.30692, respectively.

Figure 6. Graph of the box_loss curves on the training dataset and the validation dataset.
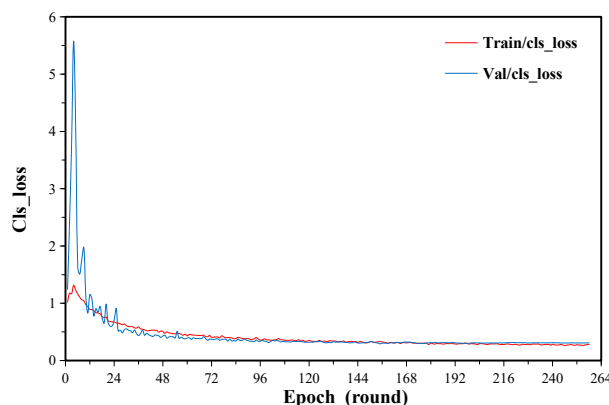

Figure 7. Graph of the cls_loss curves on the training dataset and the validation dataset.

Dfl_loss is used to enhance the model's ability to detect small targets. It usually uses specific techniques, such as introducing additional loss items or adjusting the network structure, to promote the model to better learn the feature representation of small targets and ensure that the features learned by the model have good discrimination and interpretability. As shown in Figure 8, similar to box_loss and cls_loss, dfl_loss peaks at the beginning of training, but, as the training gradually converges, the dfl_loss of the training set and validation set stabilize at 0.96787 and 0.90743, respectively.
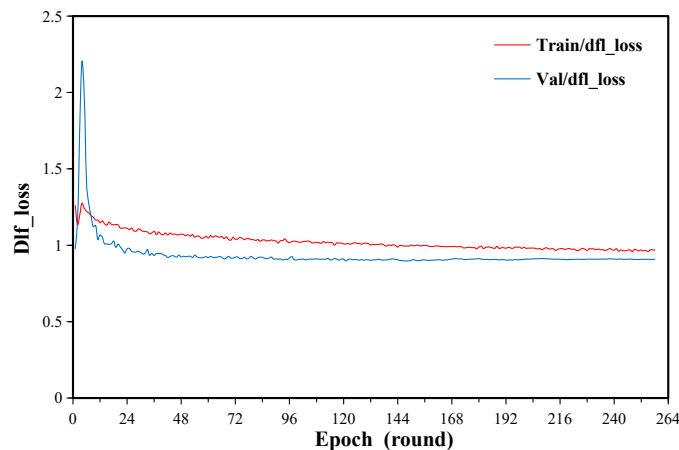

Figure 8. Graph of the dfl_loss curves on the training dataset and the validation dataset.

When evaluating the performance of a target detection model, an analysis of changes in precision, recall, mAP50 (average accuracy at 0.5 IoU threshold), and MAP50–95 (average accuracy at 0.5 to 0.95 IoU threshold) is critical. These indicators are calculated by the degree of match between the simulated test results and the real label, thus providing a quantitative standard for the performance evaluation of the model.

Precision: This refers to the proportion of the true positive sample in which the model predicts a positive sample. This index measures the prediction accuracy of the model for the target bounding box in the target detection task, which means that the proportion that the overlap between the prediction bounding box and the real target reaches or exceeds the set threshold, as shown in Equation (1); Recall: This refers to the ratio of the correct number of all true positive samples in the test result, as shown in Equation (2); mAP_{50}: This is the average accuracy under the 0.5 IoU threshold, which combines the precision and recall rates of different categories, as shown in Equation (3); mAP_{50-95}: This combines the average accuracy under different IoU thresholds to more comprehensively evaluate the detection performance of the model under different IoU thresholds, so it is a more rigorous evaluation index, as shown in Equation (4). As shown in Figure 9, after multiple rounds of optimization and testing, the final Precision value stabilized at 0.91766, 0.89827, 0.94300 for mAP_{50}, and 0.91696 for mAP_{50-95}. All of the above results show good stability and provide a reliable support for the subsequent applications.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{1}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{2}$$

$$mAP_{50} = \frac{1}{N}\sum_{i=1}^{N} AP_{i,0.5} \tag{3}$$

$$mAP_{50-95} = \frac{1}{N}\sum_{j=1}^{10} \left(\frac{1}{10} AP_{i,(0.5+0.005\cdot(j-1))}\right) \tag{4}$$

$$AP = \int_0^1 \text{Precision}(\text{Recall})d\text{Recall} \tag{5}$$

$$mAP = \frac{\sum_{j=1}^{c} AP_j}{c} \tag{6}$$
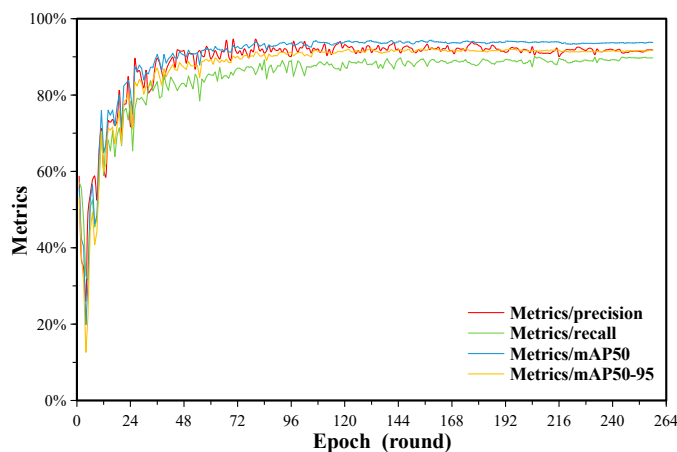

Figure 9. Each accuracy curve.

Here, TP (true positive) represents a positive sample that is correctly classified as positive; FP (false positive) represents a negative sample that is incorrectly classified as positive; FN (false negative) represents a positive sample that is incorrectly classified as negative; TN (true negative) represents a negative sample that is correctly classified

as negative; N represents the total number of categories; $AP_{i,0.5}$ represents the average precision of the ith category when the IoU threshold is 0.5; and $AP_{i,(0.5+0.05\cdot(j-1))}$ represents average precision of the ith category at the jth IoU threshold (from 0.5 to 0.95).

To further evaluate the performance of the model across the seven minerals, a detailed analysis using Precision–Recall (PR) curves was conducted, as illustrated in Figure 10. The PR curve provides a clear visualization of the trade-off between precision and recall at various confidence thresholds. The area under the PR curve (AUC-PR) is indicative of the model's overall performance across different thresholds; a larger area corresponds to a better performance. For each mineral, the mAP@0.5 metric reflects the model's detection capability within that category. Specifically, the mean Average Precision (mAP) is a crucial metric for assessing the overall performance of an object detection model. It represents the average precision across multiple classes, and the Intersection over Union (IoU) measures the overlap between the predicted and ground truth bounding boxes. An IoU $\geq$ 0.5 indicates that the predicted bounding box overlaps with the ground truth by at least 50%. The mAP@0.5, therefore, signifies the mean precision across all classes when the IoU threshold is set at 0.5. The following analysis interprets the Precision–Recall curves for each mineral based on their respective mAP@0.5 values and discusses the implications of these results: (1) Biotite (mAP@0.5 = 0.905): The PR curve for biotite shows a slight decrease in precision at higher recall levels. This indicates that, while the model is generally effective at detecting biotite, there are instances where it may generate false positives, which slightly reduces the overall precision. (2) Quartz (mAP@0.5 = 0.972): Quartz's PR curve approaches an ideal shape, maintaining a high precision even at high recall levels. This suggests that the model is highly reliable in detecting quartz, with very few false positives, making it an excellent performer in quartz detection tasks. (3) Chalcopyrite (mAP@0.5 = 0.896): The PR curve for chalcopyrite exhibits a more noticeable decline, particularly as recall approaches 1, indicating a sharper drop in precision. This suggests that the model's performance in detecting chalcopyrite is slightly inferior compared to other minerals, likely due to the more frequent false positives or difficulties in distinguishing chalcopyrite from other minerals. (4) Chrysocolla (mAP@0.5 = 0.971): The PR curve for chrysocolla resembles that of quartz, maintaining a high precision across most recall levels. This demonstrates the model's strong capability in accurately identifying chrysocolla, similar to its performance with quartz. (5) Malachite (mAP@0.5 = 0.981): The PR curve for malachite is nearly ideal, with consistently high precision across all recall levels. This indicates that the model is exceptionally proficient at detecting malachite, with almost no false positives, reflecting an outstanding detection capability. (6) Muscovite (mAP@0.5 = 0.890): Muscovite's PR curve shows a significant drop in precision as recall increases, particularly at higher recall levels. This suggests that the model's performance in detecting muscovite is relatively weaker, with a higher tendency for false positives. The decline in the PR curve indicates potential issues with the dataset or model that may require further optimization to improve precision and reduce false positives. (7) Pyrite (mAP@0.5 = 0.987): The PR curve for pyrite is nearly ideal, indicating that the model maintains a high precision and recall across almost all thresholds. This suggests that the model is extremely reliable in detecting pyrite, with virtually no false positives or missed detections. Overall, the average mAP@0.5 of 0.943 indicates that the model achieves a high level of performance across all seven minerals. Each mineral's mAP@0.5 value being close to or exceeding 0.9 demonstrates the model's robustness and adaptability in real-world applications. This high mAP@0.5 suggests that the model remains highly accurate even under challenging conditions, such as when minerals overlap or have blurred boundaries. For example, in the detection of muscovite embedded within a complex mineral vein, a high mAP@0.5 ensures precise boundary delineation, minimizing confusion with adjacent minerals. Moreover, the high mAP@0.5 values indicate that the model performs well across varying IoU thresholds. This is particularly important when minerals are partially occluded, overlapping, or have indistinct edges, as the model can consistently recognize and localize them accurately. For instance, in scenarios where chrysocolla and chalcopyrite

overlap, a high mAP@0.5 ensures that the model can accurately differentiate between the two minerals, correctly identifying their respective regions without confusion.
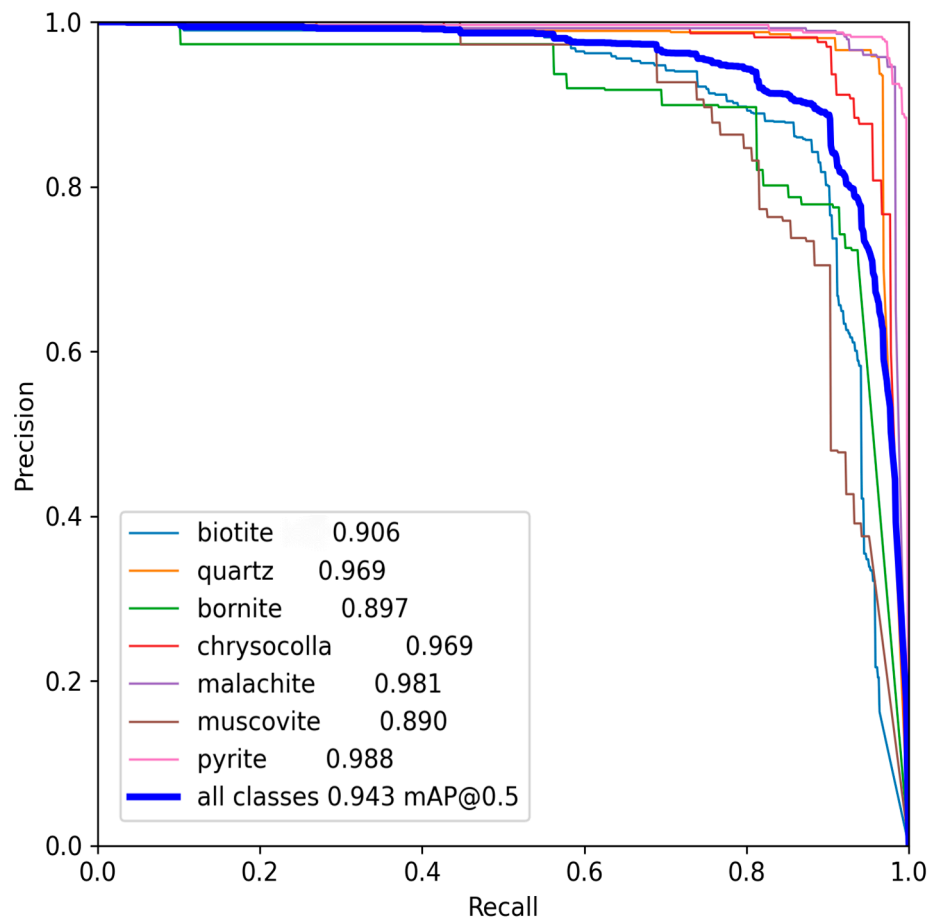

Figure 10. Precision–Recall curve.

5. Application Case Test

To validate the practical application of the trained model in mineral identification, high-resolution images of seven typical minerals were selected from a museum collection as test samples. These minerals included biotite, quartz, bornite, chrysocolla, malachite, muscovite, and pyrite, each possessing distinct identification features. For instance, biotite is characterized by its dark, platy structure; quartz by its transparency and morphological diversity; and bornite by its copper-green hue and metallic luster. As illustrated in Figure 11, the trained model successfully identified all test samples, with 83% of the samples achieving confidence levels above 87%. Furthermore, the model demonstrated rapid identification capabilities, facilitating the swift classification and recognition of large numbers of mineral specimens. The confidence levels for biotite, quartz, bornite, chrysocolla, malachite, muscovite, and pyrite were 82.30%, 84.75%, 80.08%, 83.12%, 85.25%, 79.82%, and 86.26%, respectively, yielding an average confidence level of 83.08%.

Upon further analysis of the experimental results, the model's performance across different mineral samples, as well as potential areas for improvement, can be examined more thoroughly: (1) Biotite (Confidence: 82.30%): Biotite's dark coloration and platy structure can be easily confused with other dark minerals or shadows, especially under uneven lighting conditions, potentially leading to reduced model confidence. Additionally, the reflective nature of biotite's cleavage planes may interfere with the model's detection, resulting in false positives or missed identifications. (2) Quartz (Confidence: 84.75%): Quartz's transparency and morphological diversity provide the model with rich feature information, enabling a relatively accurate identification. However, quartz's transparency can sometimes

cause confusion with the background or other transparent minerals, particularly under complex lighting conditions, which might explain why quartz's confidence level is not the highest. (3) Bornite (Confidence: 80.08%): The metallic luster and color of bornite can vary significantly under different lighting conditions, which may lead to inconsistent model performance across samples. Moreover, the similarity in color and luster between bornite and other metallic minerals, such as pyrite, could increase the difficulty of identification, resulting in lower confidence levels. (4) Chrysocolla (Confidence: 83.12%): Chrysocolla's vivid color and unique fibrous structure make it highly distinguishable from other minerals, aiding in model recognition. However, if chrysocolla is mixed with other green minerals, such as malachite, the model might experience confusion on specific samples, leading to a slight decrease in confidence. (5) Malachite (Confidence: 85.25%): The bright color and distinctive banded texture of malachite make it relatively easy for the model to recognize, accounting for its high confidence level. Nonetheless, the color similarity between malachite and chrysocolla, along with their frequent co-occurrence, might sometimes affect the model's precision. (6) Muscovite (Confidence: 79.82%): Muscovite's silvery-white luster and platy structure are prone to light reflection, which may cause fluctuations in recognition confidence under different lighting conditions. Additionally, muscovite's color and morphology closely resemble those of other mica group minerals (e.g., biotite), increasing the model's difficulty in differentiation and resulting in the lowest confidence level among the minerals tested. (7) Pyrite (Confidence: 86.26%): Pyrite's golden-yellow luster and unique cubic structure make it easily distinguishable from other minerals. These prominent features contribute to the model's high confidence, indicating that the model performs most reliably in identifying pyrite, with minimal confusion with other minerals.


Figure 11. Example of the test set samples.

"The variation in confidence levels across different minerals reflects the model's varied performance in recognizing different physical characteristics". Minerals with strong identification features, such as pyrite and malachite, tend to achieve higher confidence levels due to their distinctive color and morphology, which facilitate the easier differentiation by the model. Conversely, minerals like muscovite and bornite, whose features are more likely to be confused with other minerals or are unstable under certain lighting conditions, present greater challenges for the model, resulting in lower confidence levels. To address these discrepancies, further model optimization could focus on improving feature extraction accuracy and incorporating more diverse training samples to enhance overall model performance in practical applications.

6. Discussion

The deep-learning model proposed in this study demonstrates a high accuracy in mineral recognition tasks, with precision metrics exceeding 90%. This result highlights the model's significant performance advantages under theoretical and controlled laboratory conditions. However, the model has shown instances of misclassification when dealing with samples containing multiple minerals. For instance, as illustrated in Figure 12, a sample containing both malachite and copper ore was misclassified by the model as 36% malachite and 34% white mica. This misclassification may stem from the visual similarities between minerals, which complicate the model's ability to differentiate them accurately. Specifically, when multiple minerals are present in a single image, the model might incorrectly group these minerals into the same category, leading to erroneous results.
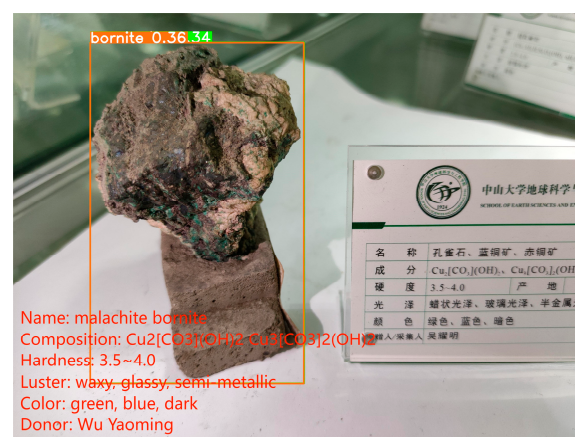
Figure 12. Miscalculation sample.

Further analysis reveals that the model's misclassification issues are predominantly associated with minerals exhibiting similar visual characteristics, underscoring the limitations of the current model in handling complex samples. To address these challenges, future research should consider several improvement strategies: First, increasing the diversity of training samples, particularly by incorporating more minerals with similar visual features, can aid the model in learning the subtle differences between these minerals. Second, optimizing the model architecture, such as by introducing deeper convolutional neural networks or attention mechanisms, can enhance the model's feature extraction capabilities and improve its handling of complex backgrounds and multi-mineral images. Finally, incorporating post-processing techniques, such as image segmentation or region enhancement methods, may help reduce misclassification and improve the model's accuracy.

Moreover, this study has primarily tested the model under controlled laboratory conditions, and its performance in practical application environments has not been thoroughly explored. Given the complexity of real-world scenarios, such as variations in lighting, background interference, and the presence of multiple minerals, future work should involve testing the model in real-world settings, such as active mining sites and geological survey environments. This will provide a better assessment of the model's stability and

effectiveness in practical applications and guide further model optimization. Field testing will offer a comprehensive understanding of the model's performance across different environmental conditions, ensuring its reliability and applicability in real-world scenarios.

In summary, although the model has demonstrated a commendable performance under current experimental conditions, further research is needed to address misclassification issues and enhance the model's applicability in real-world environments. These efforts will not only improve the model's robustness but also provide a solid foundation for the broader deployment of mineral recognition technology in practical applications.

7. Conclusions

This paper primarily introduces a deep-learning-based object detection model applied to intelligent mineral recognition. The model employs the YOLOv8 algorithm and is effective in identifying and classifying seven common minerals: biotite, quartz, chalcocite, silicon malachite, malachite, white mica, and pyrite. After 258 epochs of training, the performance metrics are notably stable, with Precision at 0.91766, Recall at 0.89827, mAP_{50} at 0.94300, and MAP_{50-95} at 0.91696. These results indicate that the model demonstrates a high accuracy and robustness in mineral recognition tasks. However, testing conducted using samples from the Geological and Mineral Museum at the School of Earth Sciences and Engineering, Sun Yat-sen University, revealed an average confidence level of 83.08%. It was observed that the model occasionally misclassifies samples containing multiple minerals. Such misclassifications are likely due to visual similarities among certain minerals in terms of shape, color, or luster, which challenge the model's ability to distinguish between them accurately. Despite these challenges, the results of this study provide valuable insights and practical experience for the advancement of intelligent mineral recognition technology.

References

1. Zhou, Y.Z.; Zuo, R.G.; Liu, G.; Yuan, F.; Mao, X.C.; Guo, Y.J.; Xiao, F.; Liao, J.; Liu, Y.P. The Great-leap-forward Development of Mathematical Geoscience During 2010–2019: Big Data and Artificial Intelligence Algorithm Are Changing Mathematical Geoscience. *Bull. Mineral. Petrol. Geochem.* **2021**, *40*, 556–573. [CrossRef]
2. Zhang, L.J.; Lu, W.h.; Zhang, J.D.; Peng, G.X.; Bu, J.C.; Tang, K.; Xie, J.C.; Xu, Z.B.; Yang, H.Y. Rock and mineral thin section identification based on deep learning. *Geosci. Front.* **2024**, *31*, 498–510. [CrossRef]
3. Zhou, Y.Z.; Zhang, L.J.; Zhang, A.D.; Wang, J. Big Data Mining & Machine Learning in Geoscience. Sun Yat-sen University Press: Zhuhai, China, 2018.
4. Xu, S.T.; Zhou, Y.Z. Artificial Intelligence Identification of Ore Minerals under Microscope Based on Deep Learning Algorithm. *Acta Petrol. Sin.* **2018**, *34*, 3244–3252. Available online: http://html.rhhz.net/ysxb/20181110.htm (accessed on 8 September 2018).
5. Alerigi, D.P.S.R.; Li, W. Method for providing rock characterization and classification for geo-exploration, involves applying deep learning models to newly received data that includes data, and predicting properties based on newly received geo-exploration data. *Saudi Arabian Oil Co* **2022**, D65162. Available online: https://webofscience.clarivate.cn/wos/alldb/full-record/DIIDW: 2022D65162 (accessed on 19 August 2022).
6. Zeng, X.; Ji, X.H.; Xiao, Y.C.; Wang, G.W. Mineral Identification Based on Deep Learning That Combines Image and Mohs Hardness. *Minerals* **2020**, *11*, 506. [CrossRef]
7. Wang, J.B.; Xue, L.F.; Gao, X. Identification Method of Volcanic Rock Slices Based on A Deep Residual Shrinkage Network. In Proceedings of the Fourth International Conference on Geoscience and Remote Sensing Mapping, Wuhan, China, 14–16 April 2023; Volume 12551, pp. 389–394. [CrossRef]

8. Zhang, Y.; Li, M.C.; Han, S.; Ren, Q.B.; Shi, J. Intelligent Identification for Rock-Mineral Microscopic Images Using Ensemble Machine Learning Algorithms. *Sensors* **2019**, *18*, 3914. [CrossRef]

9. Zhang, S.X.; Yang, Y.Y.; Sun, F.C.; Fang, B. Application of Image Sensing System in Mineral/Rock Identification: Sensing Mode and Information Process. *Adv. Intell. Syst.* **2023**, *5*, 2300206. [CrossRef]

10. Shi, C.J.; Zhang, W.M.; Chen, H.R.; Ge, L.L. Survey of Salient Object Detection Based on Deep Learning. *J. Front. Comput. Sci. Technol.* **2023**, *38*, 21–50. [CrossRef]

11. Joseph, R.; Santosh, D.; Ross, G.; Ali, F. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788. [CrossRef]

12. Tan, L.; Huangfu, T.R.; Wu, L.Y.; Chen, W.Y. Comparison of RetinaNet, SSD, and YOLO v3 for real-time pill identification. *BMC medical informatics and decision making.* **2021**, *21*, 1–11. [CrossRef]

13. Ren, S.Q.; He, K.M.; Ross, G.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1137–1149. [CrossRef]

14. Tan, M.; Pang, R.; Le, Q.V. EfficientDet: Scalable and Efficient Object Detection. In Proceedings of the 2020 IEEE Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 10778–10787. [CrossRef]

15. He, K.M.; Gkioxari, G.; Dollar, P.; Girshick, R. Mask R-CNN. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCVW), Venice, Italy, 22–29 October 2017; pp. 2980–2988. [CrossRef]

16. Zhang, J.Y.; Gao, Q.; Luo, H.L.; Long, T. Mineral Identification Based on Deep Learning Using Image Luminance Equalization. *Appl. Sci.* **2022**, *12*, 7055. [CrossRef]

17. Liu, Y.; Zhang, Z.L.; Liu, X.; Wang, L.; Xia, X.H. Ore image classification based on small deep learning model: Evaluation and optimization of model depth, model structure and data size. *Miner. Eng.* **2021**, *172*, 107020. [CrossRef]

18. Wan, D.H.; Lu, R.S.; Wang, S.L.; Shen, S.Y.; Xu, T.; Lang, X.L. YOLO-HR: Improved YOLOv5 for Object Detection in High-Resolution Optical Remote Sensing Images. *Remote Sens.* **2023**, *15*, 614. [CrossRef]

19. Pratama, B.G.; Qodri, M.F.; Sugarbo, O. Building YoloV4 models for identification of rock minerals in thin section. *IOP Conf. Ser. Earth Environ. Sci.* **2023**, *1151*, 012046. [CrossRef]

20. Mimura, K.; Nakamura, K.; Yasukawa, K.; Sibert, E.C.; Ohta, J.; Kitazawa, T.; Kato, Y. Applicability of Object Detection to Microfossil Research: Implications From Deep Learning Models to Detect Microfossil Fish Teeth and Denticles Using YOLO-v7. *Earth Space Sci.* **2024**, *11*, e2023EA003122. [CrossRef]

21. Mimura, K.; Nakamura, K.; Takao, K.; Yasukawa, K.; Kato, Y. Automated Detection of Hydrothermal Emission Signatures From Multibeam Echo Sounder Images Using Deep Learning. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2023**, *16*, 2703–2710. [CrossRef]

22. Ehsani, M.R.; Upadhyaya, S.K.; Slaughter, D.; Shafii, S.; Pelletier, M. A NIR Technique for Rapid Determination of Soil Mineral Nitrogen. *Precis. Agric.* **1999**, *1*, 219–236. [CrossRef]

23. Aswini, E.; Vijayakumaran, C. Auto Detector for Huanglongbing Citrus Greening Disease using YOLOV7. In Proceedings of the 2023 World Conference on Communication & Computing (WCONF), Raipur, India, 14–16 July 2023; pp. 1–6. [CrossRef]

24. Mei, H.; Wang, Q.Y.; Yu, L.; Zeng, Q. A deep learning-based algorithm for intelligent prediction of adverse geologic bodies in tunnels. *Meas. Sci. Technol.* **2024**, *35*, 096119. [CrossRef]

25. Williams, J.M.; Scuderi, L.A.; McClanahan, T.P.; Banks, M.E.; Baker, D.M.H. Comparative planetology–Comparing cirques on Mars and Earth using a CNN. *Geomorphology* **2023**, *440*, 108881. [CrossRef]

26. Li, W.W.; Wang, S.Z.; Arundel, S.T.; Hsu, C.Y. GeoImageNet: A multi-source natural feature benchmark dataset for GeoAI and supervised machine learning. *GeoInformatica* **2023**, *27*, 619–640. [CrossRef]

27. Shi, Z.X.; Mo, G.Q.; Cui, Y.R.; Yan, L.B.; Lu, Y.S.; Hou, L.N.; Lv, L.S.; Li, H.X. Automatic Identification of Cirques Based on RetinaNet Model and Pseudo-Color Image Fusion Method. *Adv. Space Res.* **2024**, *74*, 2930–2940. [CrossRef]

28. Bickel, V.T.; Lanaras, C.; Manconi, A.; Loew, S.; Mall, U. Lunar Rockfall Detection and Mapping usinng Deep Neural Networks. In Proceedings of the 50th Lunar and Planetary Science Conference, The Woodlands, TX, USA, 18–22 March 2019. Available online: https://hdl.handle.net/21.11116/0000-0003-4120-F (accessed on 18 March 2019).

29. Feldens, P.; Darr, A.; Feldens, A.; Tauber, F. Detection of boulders in side scan sonar mosaics by a neural network. *Geosciences* **2019**, *9*, 159. [CrossRef]

30. Bickel, V.T.; Loew, S.; Aaron, J.; Goedhart, N. A global perspective on lunar granular flows. *Geophys. Res. Lett.* **2022**, *49*, e2022GL098812. [CrossRef]

31. Ma, S.M.; Huang, W.H. Application of Deep Learning Algorithms in Determination of Trace Rare Earth Elements of Cerium Group in Rocks and Minerals. *Wirel. Commun. Mob. Comput.* **2021**, *2021*, 9945141. [CrossRef]

32. Wang, H.Y.; Cao, W.; Zhou, Y.Z.; Yu, P.P.; Yang, W. Multitarget intelligent recognition of petrographic thin section images based on faster RCNN. *Minerals* **2023**, *13*, 872. [CrossRef]

33. Arslan, E.A. Radio galaxy morphology classification with mask R-CNN. In Proceedings of the 2020 4th International Conference on Vision, Image and Signal Processing, Bangkok, Thailand, 9–11 December 2020; Volume 36, pp. 1–5. [CrossRef]

34. Liu, X.B.; Wang, H.Y.; Jing, H.D.; Shao, A.L.; Wang, L.C. Research on intelligent identification of rock types based on faster R-CNN method. *IEEE Access* **2020**, *8*, 21804–21812. [CrossRef]

35. Radulescu, M.; Dalal, S.; Lilhore, U.K.; Saimiya, S. Optimizing mineral identification for sustainable resource extraction through hybrid deep learning enabled FinTech model. *Resour. Policy* **2024**, *89*, 104692. [CrossRef]

36. Munteanu, D.; Moina, D.; Zamfir, C.G.; Petrea, S.M.; Cristea, D.S.; Munteanu, N. Sea mine detection framework using YOLO, SSD and EfficientDet deep learning models. *Sensors* **2022**, *22*, 9536. [CrossRef]

37. Jia, J.Q.; Fu, M.; Liu, X.F.; Zheng, B. Underwater object detection based on improved efficientdet. *Remote Sens.* **2022**, *14*, 4487. [CrossRef]

38. Edvardsen, I.P.; Teterina, A.; Johansen, T.; Myhre, J.N.; Godtliebsen, F.; Bolstad, N.L. Automatic detection of the mental foramen for estimating mandibular cortical width in dental panoramic radiographs. *UiT Nor. Arktiske Univ.* **2022**, *50*, 03000605221135147.

39. Kim, E.C.; Hong, S.J.; Kim, S.Y.; Lee, C.H.; Kim, S.; Kim, H.J.; Kim, G. CNN-based object detection and growth estimation of plum fruit (*Prunus mume*) using RGB and depth imaging techniques. *Sci. Rep.* **2022**, *12*, 20796. [CrossRef] [PubMed]

40. Dong, L.H.; Wang, H.L.; Song, W.; Xia, J.X.; Liu, T.M. Deep sea nodule mineral image segmentation algorithm based on Mask R-CNN. In Proceedings of the ACM Turing Award Celebration Conference-China, Hefei, China, 30 July–1 August 2021; pp. 278–284.

41. Iyas, M.R.; Setiawan, N.I.; Warmada, I.W. Mask R-CNN for rock-forming minerals identification on petrography, case study at Monterado. *E3S Web Conf.* **2020**, *200*, 06007. [CrossRef]

42. Koh, E.J.; Amini, E.; McLachlan, G.J.; Beaton, N. Utilising convolutional neural networks to perform fast automated modal mineralogy analysis for thin-section optical microscopy. *Miner. Eng.* **2021**, *173*, 107230. [CrossRef]

43. Caldas, T.D.P.; Augusto, K.S.; Iglesias, J.C.A.; Ferreira, B.A.P.; Santos, R.B.M.; Paciornik, S.; Domingues, A.L.A. A methodology for phase characterization in pellet feed using digital microscopy and deep learning. *Miner. Eng.* **2024**, *212*, 108730. [CrossRef]

44. Mao, Z.H.; Zhu, J.L.; Wu, X.; Li, J. Review of YOLO Based Target Detection for Autonomous Driving. *J. Comput. Eng. Appl.* **2022**, *58*, 68–77. [CrossRef]

45. Al Muksit, A.; Hasan, F.; Emon, M.F.H.B.; Haque, M.R.; Anwary, A.R.; Shatabda, S. YOLO-Fish: A robust fish detection model to detect fish in realistic underwater environment. *Ecol. Inform.* **2022**, *72*, 101847. [CrossRef]

46. Zuraimi, M.A.B.; Zaman, F.H.K. Vehicle detection and tracking using YOLO and DeepSORT. In Proceedings of the 2021 IEEE 11th IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE), Penang, Malaysia, 3–4 April 2021; pp. 23–29. [CrossRef]

47. Vilar-Andreu, M.; García, L.; Garcia-Sanchez, A.J.; Asorey-Cacheda, R.; Garcia-Haro, J. Enhancing Precision Agriculture Pest Control: A generalized Deep Learning Approach with YOLOv8-based Insect Detection. *IEEE Access* **2024**, *12*, 84420–84434. [CrossRef]

48. Prinzi, F.; Insalaco, M.; Orlando, A.; Gaglio, S.; Vitabile, S. A yolo-based model for breast cancer detection in mammograms. *Cogn. Comput.* **2024**, *16*, 107–120. [CrossRef]

49. Reddy, S.M.; Rakesh, K.; Aluvala, S.; Bindu, G.; Husseen, A. Fault Detection and Classification in Semiconductor Manufacturing for Sensor Screening Using Multi-Layer. In Proceedings of the 2024 International Conference on Distributed Computing and Optimization Techniques (ICDCOT), Deep Neural Network, Bengaluru, India, 15–16 March 2024; pp. 1–4. [CrossRef]

50. Reis, D.; Kupec, J.; Hong, J.; Daoudi, A. Real-time flying object detection with YOLOv8. *arXiv* **2023**, arXiv:2305.09972. [CrossRef]

51. Wu, T.Y.; Dong, Y.K. YOLO-SE: Improved YOLOv8 for Remote Sensing Object Detection and Recognition. *IEEE Trans. Appl. Sci.* **2023**, *13*, 12977. [CrossRef]

52. Zhang, M.H.; Wang, Z.H.; Song, W.; Zhao, D.F.; Zhao, H.J. Efficient Small-Object Detection in Underwater Images Using the Enhanced YOLOv8 Network. *IEEE Trans. Appl. Sci.* **2024**, *14*, 1095. [CrossRef]

53. Wang, X.L.; Girshick, R.; Gupta, A.; He, K.M. Non-local neural networks. In Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 7794–7803. [CrossRef]

54. Redmon, J.; Farhadi, A. YOLOv3: An Incremental Improvement. *arXiv* **2018**, arXiv:1804.02767. [CrossRef]

55. Li, J.F.; Yan, C.X. YOLOv8 with Multi Strategy Integrated Optimization and Application in Object Detection. *Autom. Mach. Learn.* **2024**, *5*, 23–31. [CrossRef]

56. Attallah, Y.; Minerals Identification & Classification. Kaggle. Retrieved 2023. Available online: https://www.kaggle.com/datasets/youcefattallah97/minerals-identification-classification (accessed on 8 February 2023).

57. Webmineral. Webmineral. Retrieved 2023. Available online: http://webmineral.com/ (accessed on 21 October 2023).