Enhanced Mineral Image Classification Using YOLOv8-CLS With Optimized Feature Extraction and Dataset Augmentation

Bofan Long*, Chunyang Zhou
School of Future Technology, China University of Geosciences (Wuhan), Wuhan, 430074, China
E-mail: Bofan_Long@outlook.com
*Corresponding author

Abstract: With the increasing global emphasis on ecological conservation and the sustainable use of resources, the effective utilization of mineral resources has become a pressing priority. Accurate mineral classification helps reduce resource waste, mitigate ecological impact, and improve processing efficiency. This paper proposes a deep learning model for mineral image classification based on the YOLOv8-CLS architecture, specifically targeting seven minerals: bornite, quartz, malachite, pyrite, muscovite, biotite, and chrysocolla, which are the focus of this study. The model was trained and tested on an open-source mineral image dataset, achieving Top-1 and Top-5 accuracy rates of 0.92053 and 0.99399, respectively, after 120 training epochs. During the testing phase, the model was evaluated on the test set, achieving a Top-1 accuracy of 0.90681 and a Top-5 accuracy of 0.99283, demonstrating high accuracy and stability. Although a decline in Top-1 accuracy to 0.80219 was observed when testing the model against a new batch of data and comparing it to the classic ResNet50 and ResNet101 models, the YOLOv8-CLS model still outperforms these models by 0.00684 and 0.04278, respectively, while also having lower performance overhead. Despite some remaining flaws, this study demonstrates that the YOLOv8-CLS model is more efficient than traditional models in intelligent mineral classification, contributing to resource efficiency and promoting the development of sustainable mining practices.

Povzetek: YOLOv8-CLS z optimizirano ekstrakcijo značilk in podatkovno augmentacijo omogoča bolj kvalitetno razvrščanje mineralnih slik kot ResNet50/101.

1 Introduction

Mineral resources play a vital role in economic and social development [1]. Intelligent mineral classification technology has actually made considerable developments in helping with mineral source research study, mining, and administration[2]. Existing mineral identification techniques predominantly rely on manual evaluations and expert judgments, resulting in reduced efficiency and potential biases, which are inadequate for massive expedition needs. However, the progression in deep learning technology has made intelligent mineral acknowledgment systems [4] a popular research study location and advanced subject. Specifically, we propose the use of the YOLOv8-CLS model to enhance the precision of mineral identification under real-world conditions, thereby optimizing resource utilization and supporting sustainable mining practices.

Deep learning, which uses the concepts of artificial neural networks, has actually changed the area of pattern acknowledgment through its multi-layered structure that automatically learns information patterns and relationships [5] Zeng et al. (2020) [6] suggested an ingenious approach by incorporating mineral image information and Mohs hardness into neural network architectures to improve classification accuracy. Their model showed excellent outcomes, achieving a Top-1

precision of 90.6% and a Top-5 precision of 99.6% when related to 36 generally encountered minerals. This method not only improves the precision however also expands the spectrum of identifiable mineral types. Building on these advancements, Wang et al. (2023) [7] checked out the application of a deep residual semantic network with a contractive architecture for classifying volcanic rock thin sections. Their research study, which included 12,000 high-resolution thin section images covering 11 significant sorts of volcanic rock, reported a classification precision exceeding 92% on the test collection, highlighting the effectiveness of this deep learning approach in geological image assessment. In parallel, Zhang et al. (2019) [8] developed a wise recognition model specifically made for the evaluation of rock and mineral images under a microscopic lense. This model shows considerable possibility in enhancing the accuracy and effectiveness of mineral classification in tiny monitoring. This model, utilizing the Inception-v3 architecture combined with conventional machine learning techniques such as logistic regression, help vector devices, and multilayer perceptrons, attained a precision of 90.9% for minerals consisting of quartz, feldspar, potassium feldspar, and plagioclase. Zhang et al. (2023) [9] furthermore increased this job by having a look at electrochemical methods for mineral recognition, highlighting the crucial task of exact identification

methods in geological research study. These looks into jointly highlight the expanding value of deep learning and expert system approaches in the specific classification of geological products, showing durable performance throughout different mineral and rock types.

Image classification jobs, which assign input photos to predefined groups, are generally utilized in medical diagnostics, self-governing driving, smart production, and other domains. In recent years, image classification models have actually progressed from standard machine learning techniques to innovative deep learning approaches utilizing convolutional neural networks (CNNs)[10]. Conventional image classification approaches depend on handcrafted feature extraction methods like SIFT and HOG, which face performance restrictions when handling facility, high-dimensional image information. However, with the widespread fostering of deep learning, specifically CNNs, automatic function learning has revolutionized image classification. Models such as ResNet, DenseNet, and EfficientNet represent modern improvements in image classification, resulting in significant renovations in precision and efficiency. ResNet makes use of recurring blocks to alleviate the disappearing slope concern in deep networks, which boosts both training depth and precision [11] DenseNet fully makes use of features from preceding layers through thick links, making it possible for much more reliable function reuse [12]. EfficientNet, created through neural architecture search (NAS), masters both efficiency and computational efficiency, causing its extensive application throughout numerous domains [13]. In the last few years, the Transformer-based Vision Transformer (ViT) model has additionally emerged in image classification tasks. By separating photos right into patches and applying the Transformer architecture initially made use of in natural language processing, ViT attains remarkable classification accuracy. Despite the high information and computational needs, ViT's ability to capture global image context improves classification performance. In the field of image classification, maximizing models for details application situations has actually become a key study focus. Networks like MobileNet provide high precision with less criteria and reduced computational prices, making them suitable for gadgets with restricted sources, such as smart devices and embedded systems [15]. As deep learning models remain to advance, image classification tasks are becoming significantly precise and efficient. Future study will concentrate on maximizing models for large-scale datasets and computational sources. In cases of source restraints, models will certainly be fine-tuned to supply efficient and exact services across different application circumstances.

The YOLO algorithm household [17] has obtained considerable traction throughout different markets because of its convenience and real-time item detection abilities. In the biological sciences, Abdullah et al. [18] made use of the YOLO algorithm in mix with underwater electronic cameras to identify fish varieties in real-time, significantly improving the accuracy and efficiency of aquatic biodiversity evaluations. This approach has set new criteria for eco-friendly monitoring, specifically in underwater environments. In web traffic management, Zuraim et al. [19] incorporated YOLO right into a website traffic surveillance system, making it possible for specific vehicle discovery and tracking. This improvement has actually offered beneficial insights right into traffic patterns, aiding in urban planning and congestion analysis, eventually contributing to enhanced framework development. In the farming industry, Vilar-Andreu et al. [20] carried out YOLO to create an insect discovery system that can determine and take care of plant bugs with high precision. The prompt control of insects has actually caused significant improvements in both plant top quality and return, underscoring the algorithm's influence on farming productivity. At the same time, in the clinical area, Prinzi et al. [21] applied YOLO to bust X-ray image evaluation, substantially boosting the detection of early-stage tumors. The enhanced accuracy in growth acknowledgment has actually resulted in far better treatment results and increased survival costs for people, revealing the formula's significance in medical diagnostics. In semiconductor production, Reddy et al. [22] utilized the YOLO formula to recognize concerns throughout manufacturing, which reduced product waste and increased functional efficiency. This application shows the formula's worth in high-precision commercial jobs. The present variation, YOLOv8 [23], presented structure renovations and much more enhanced its discovery capabilities, making it a much more reliable gadget throughout these different locations. These improvements have actually widened its applicability, reinforcing YOLO's standing as a leading formula in real-time detection across industries.

YOLOv8 builds upon the foundational YOLO style by consisting of several advanced approaches that enhance its detection abilities. At its core, YOLOv8 employs the Darknet53 network as its backbone, a key element that substantially boosts the responsive area and reinforces feature depiction. This design enables the model to capture more extensive information, boosting its performance in things detection jobs. This backbone supplies a robust framework for looking after intricate pictorial information. Along with the Darknet53 backbone, YOLOv8 integrates ingenious function blend modules, including Spatial Pyramid Pooling (SPP) and Path Aggregation Network (PAN). These modules are created to enhance the network's capacity to remove and process attributes throughout many ranges, as a result boosting its ability to handle different and elaborate visual details. This multi-scale function removal is particularly beneficial for finding little and comprehensive targets that might be challenging for earlier models. YOLOv8 a lot more fine-tunes its discovery capacities with the adoption of waterfall multiscale function maps, which increase the accuracy and efficiency of target detection in various real-world situations, such as untidy atmospheres. The algorithm also enhances discovery performance by changing the size and element ratio of support frameworks and integrating numerous model optimization strategies.

Figure 1 provides a relative analysis of the YOLO collection, illustrating YOLOv8's substantial improvements in mean Average Precision (mAP) and dealing with time (ms/img). This comparison highlights the substantial developments obtained with YOLOv8, showing both its enhanced precision and decreased computational time relative to previous variations. These enhancements highlight YOLOv8's remarkable performance and performance in product detection tasks. He et al. (2024) [16] developed an ingenious deep learning model using the YOLOv8, specially produced the target detection of 7 normal minerals. The model undertook significant training, extending 258 epochs, during which it showed significant performance improvements. By the end of the training treatment, the style got to stable and exceptional metrics: 0.91766 in accuracy, 0.89827 in recall, 0.94300 in mAP50, and 0.91696 in mAP50-95. These metrics mirror the model's strong capacity to precisely determine and categorize mineral samples. Moreover, in the screening stage, the variation efficiently acknowledged all samples, with a notable 83% of these examples revealing a self-confidence degree above 87%. This high degree of efficiency stresses the model's strength and reliability in specifically discovering and categorizing minerals in different conditions.

The YOLOv8-CLS algorithm used in this research study is a variation of the YOLOv8 algorithm collection particularly created for image classification tasks, which has unique benefits in this area. YOLOv8-CLS shares a number of function extraction layers with the YOLOv8 network and adapts them to the specific features of image classification tasks. Contrasted to the conventional ResNet network, the YOLOv8-CLS model has less specifications, reduced computational demands, and equivalent identification accuracy. The following seven minerals were selected for this study based on their geological significance and classification challenges: bornite, quartz, malachite, pyrite, muscovite, biotite, and chrysocolla.

1. **Bornite:** As a prominent copper sulfide mineral, bornite is widely distributed in copper deposits and plays a critical role in copper exploration and metallurgical processes. Its complex texture and variable appearance under different lighting conditions make it challenging to classify accurately.

2. **Quartz:** Quartz is one of the most abundant minerals in the Earth's crust, with significant industrial applications in glass manufacturing and construction. However, its transparency and diverse morphological forms often lead to misclassification, especially under complex lighting conditions.

3. **Malachite:** Known for its vibrant green color and ribbon-like texture, malachite is a key indicator of copper oxide mineralization. Its unique appearance facilitates identification, but its color similarity to other green minerals (e.g., chrysocolla) poses classification challenges.

4. **Pyrite:** Commonly referred to as "fool's gold," pyrite is a widespread iron sulfide mineral with a distinctive golden luster and cubic structure. Its visual similarity to other metallic minerals increases the difficulty of accurate classification.

5. **Muscovite:** A common silicate mineral found in metamorphic and granitic rocks, muscovite is characterized by its bright white luster and plate-like structure. However, its reflective properties and similarity to other mica minerals complicate its identification.

6. **Biotite:** Biotite is a dark, plate-like silicate mineral often found in igneous and metamorphic rocks. Its dark color and reflective surface make it prone to misclassification, especially under poor lighting conditions.

7. **Chrysocolla:** This copper silicate mineral is typically associated with copper oxide deposits and is valued for its bright blue-green color. Despite its distinctive appearance, chrysocolla can be confused with malachite due to their similar coloration and co-occurrence in nature.

The selection of these minerals reflects their geological importance in mineral exploration and processing, as well as the technical challenges associated with their accurate classification. By addressing these challenges, our study aims to contribute to the development of more robust and reliable mineral identification systems.
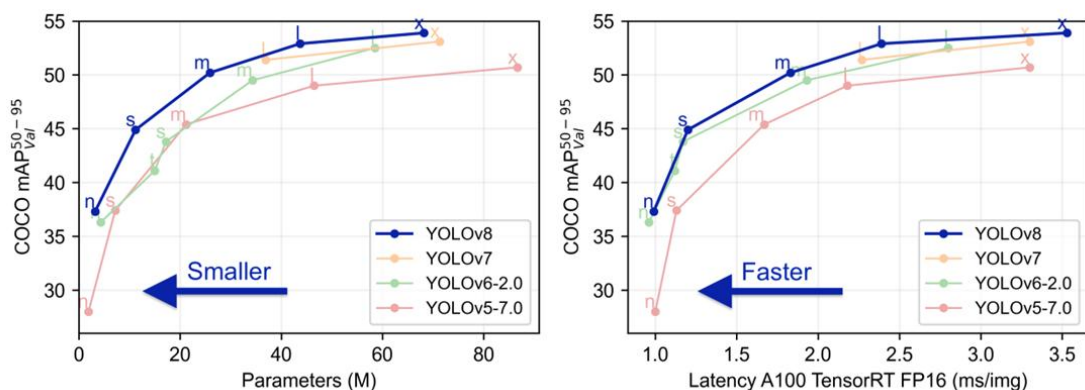

Figure 1: Performance comparison of different versions of the YOLO algorithm

2   YOLOv8 series model

YOLOv8, developed by Ultralytics, is the most widely used variant in the YOLO series, supplying notable enhancements in real-time discovery, classification, and division tasks. The YOLOv8 style incorporates a streamlined foundation network with a head network to improve both speed up and precision, keeping the performance of the YOLO series while boosting efficiency metrics, which has gathered considerable interest [24] YOLOv8 is versatile, sustaining a variety of jobs such as things discovery, image classification, circumstances division, and present estimate. This flexibility makes it incredibly versatile for a broad variety of computer system vision applications, from spotting objects in real-time to analyzing elaborate information in images and comprehending human positions. Its detailed capacities allow it to resolve varied difficulties in the field of computer system vision successfully. The total framework layout is highlighted in Figure 3.
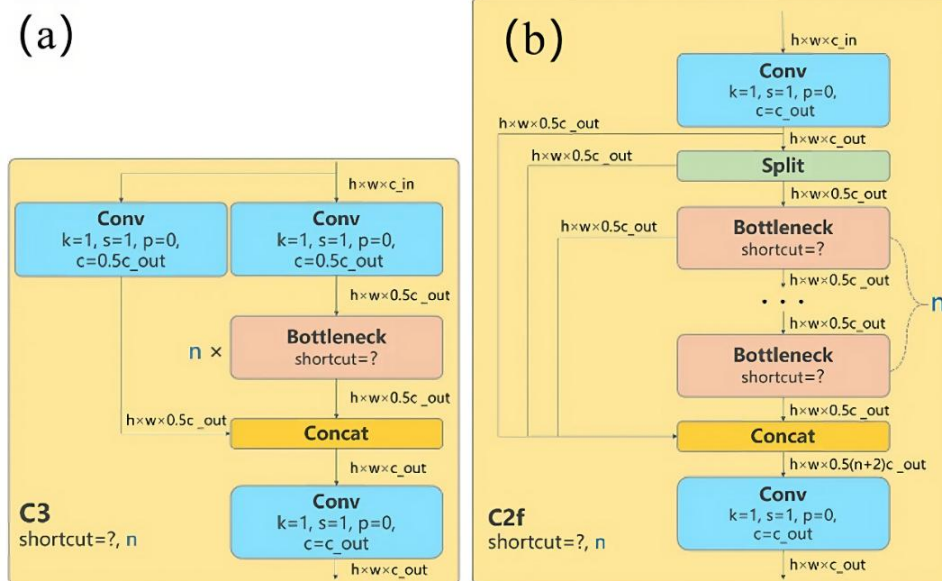
2.1 C3 module and c2f module

The C2f component, unlike YOLOv5's C3 component, includes fewer criteria and enhanced feature removal capabilities. In C2f, the input initially undergoes a convolutional layer with parameters k = 1, s = 1, p = 0, c = out, and is then processed with multiple Bottleneck layers complying with the recurring and backbone outputs. This upgrade enhances gradient circulation, therefore improving training rate and general model efficiency. This adjustment yields a model that is both lighter and a lot more robust, achieving superior efficiency in intricate circumstances. As shown in Figure 2(a), the C3 module substantially enhances the model's abilities by incorporating the shunt device with the BottleNeck recurring module. This assimilation is built on the style concepts of CSPNet and residual networks, leading to boosted feature extraction and enhanced general model efficiency. This module consists of three layers integrating convolution, batch normalization, and SiLU activation functions, plus a variable number of BottleNeck devices. The last convolutional layer boosts the network count by a variable of two, incorporating inputs from both the primary and second gradient flow branches.

Figure 2: The framework diagram of C3 component and C2f module

Caption:
(a) C3 Module: This part of the diagram illustrates the C3 module, which integrates the shunt concept of CSPNet with a recurrent architecture to enhance feature extraction capabilities. This configuration aids in managing gradient flow more effectively, contributing to improved model resilience and learning efficiency.
(b) C2f Module: This section shows the C2f module designed to keep the model architecture lightweight while enriching feature representation. It optimizes gradient flow across different layers, thus enhancing both the training and inference efficiency of the model. This module plays a crucial role in processing multi-scale features and maintaining high computational efficiency.[33]

As shown in Figure 2(b), the C2f module consists of a series of BottleNeck blocks, each containing two convolutional layers. The first layer is used to process the input function map, which is then split right into multiple parts, refined individually before being merged. Such a design enables the model to capture richer contextual information, thereby enhancing the accuracy of target recognition. Next, the second convolutional layer further refines the output of the merged feature map. The C2f component considerably boosts the model's expressive ability and precision, bring about outstanding efficiency in functional target detection applications. By enhancing feature depiction and refining discovery accuracy, the C2f module makes certain that the model masters

identifying and assessing complicated and varied targets in real-world scenarios.

2.2 Head network, neck network, and backbone network

As depicted in Figure 3, the YOLOv8 model comprises a head network responsible for producing last predictions and a neck network that works as an intermediary between the foundation and the head, concentrating on feature fusion and handling. YOLOv8 achieves boosted model performance through a number of style developments. Structure on the ELAN structure of YOLOv7, YOLOv8 introduces the innovative C2f module, which replaces the previous C3 component. This

improvement improves gradient flow and optimizes feature usage through added cross-layer connections. Moreover, YOLOv8 readjusts channel numbers to much better accommodate numerous discoveries demands and transitions from a 6x6 to a 3x3 convolutional kernel. This modification minimizes computational load while enhancing feature extraction efficiency. The network structure has likewise been structured by removing two convolutional layers. The C2f component additionally augments the model's abilities by incorporating added miss connections and splitting operations, thereby boosting function diversity and total expressiveness. Collectively, these innovations contribute to YOLOv8's superior accuracy, efficiency, and robustness in complex target detection tasks [16].

Figure 3: YOLOv8 overall model structure [16]
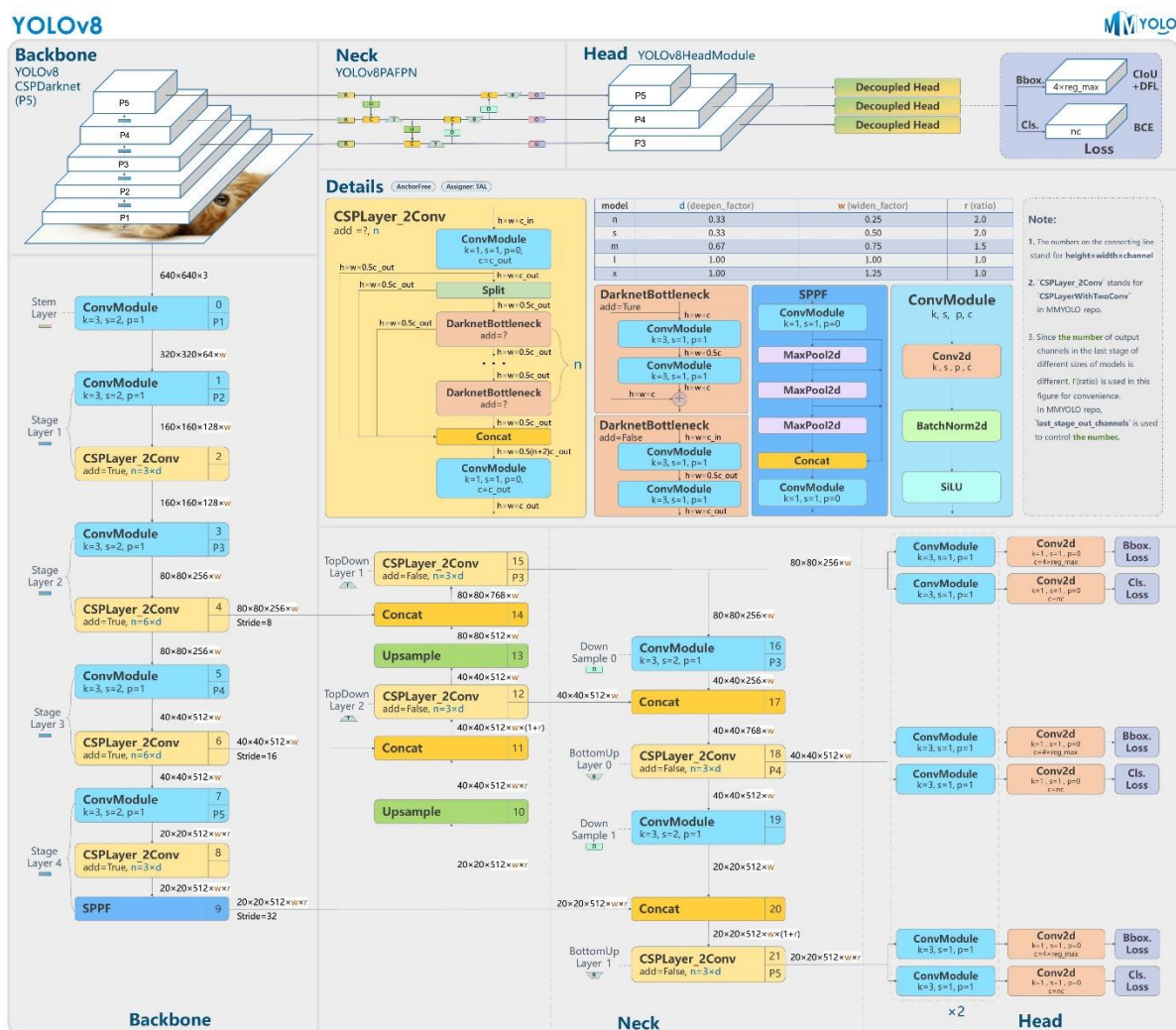
2.3 Uncoupling magnetic head

As illustrated in Figure 4, YOLOv8 introduces a well-considered design concept in task division, significantly advancing the architecture over traditional object detection frameworks. Traditionally, these frameworks combine classification and localization tasks, sharing the same parameter sets across both tasks. This commonality

often leads to mutual interference, where the optimization of one task can adversely affect the accuracy of the other. However, YOLOv8 engineers have innovated with the adoption of a "decoupled head structure" that distinctly separates the parameter sets for classification and localization tasks. This strategic separation allows each network component—the classification subnetwork and the regression subnetwork—to focus intensively on its specific function:

- **Classification Subnetwork:** This part is solely responsible for predicting the probability distribution of object classes. It specializes in discerning the varied categories of objects within an image, optimizing the network to handle the subtleties of class differentiation efficiently.
- **Regression Subnetwork:** Conversely, this subnetwork focuses exclusively on estimating the offsets for bounding box coordinates. It is finely tuned to improve the precision of object localization within the spatial domain of the image.

This decoupled approach mitigates the conflict inherent in performing these tasks concurrently within a single shared network. By reducing the inter-task interference, YOLOv8 enhances the model's overall precision and robustness—particularly under complex scenarios where traditional models might struggle with accuracy.The implementation of this decoupled head structure does not merely segregate the tasks; it refines the model's architecture to ensure that both classification and localization are handled more effectively without compromising on either. This design shift leads to a notable enhancement in model performance and accuracy, providing a more reliable system that excels in varied and dynamic environments. Thus, YOLOv8's decoupled head model significantly contributes to the model's efficacy, affirming its architectural evolution as a major leap forward in object detection technology.
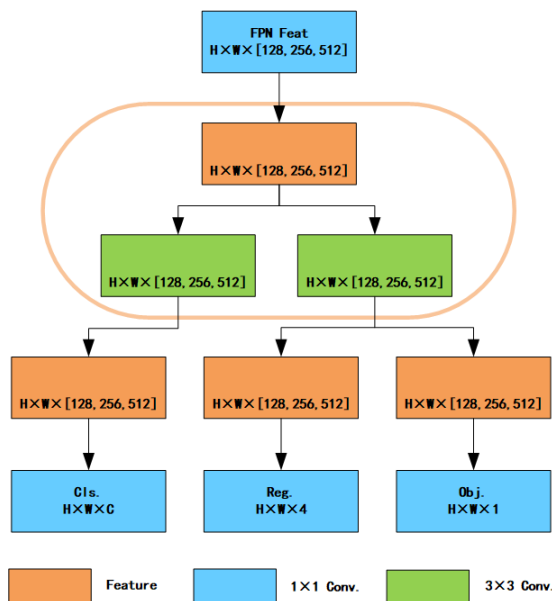

Figure 4: Flow chart of decoupled-head

2.4 SPPF module

Figure 5 highlights the capabilities of the Spatial Pyramid Pooling Fusion (SPPF) module in enhancing target detection. By refining feature maps across various dimensions, it significantly improves function extraction. The module utilizes a pyramid layer that captures and processes features at multiple scales, which helps to address the typical information loss seen with fixed-size pooling.At the start of its operation, the SPPF module applies multi-scale pooling techniques to the input feature map, segmenting it into grids of varying dimensions. Each grid is processed separately, enabling the creation of unique feature representations for each scale. This method enhances the module's ability to gather and integrate data from different spatial resolutions, leading to a richer and more robust feature extraction process. These multi-scale features are then combined using methods like concatenation or summation, merging them into a unified dataset that offers a detailed representation of the input. This integration is crucial for boosting the model's ability to detect and classify objects across diverse scales. To manage computational demands and reduce parameter count, the SPPF module incorporates a dimensionality reduction step, typically through a convolutional or fully connected layer. This reduction is essential for streamlining the processing phases. Finally, the enriched features, which contain extensive spatial and semantic details, are passed on to the next layers of the network. This transition significantly improves the model's effectiveness in identifying and classifying targets, thereby enhancing its performance and accuracy in complex detection scenarios.
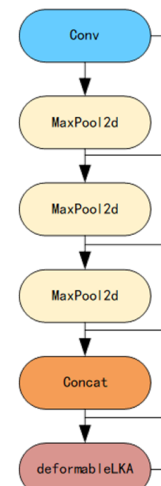

Figure 5: SPPF flow chart

2.5 Loss function-based label assignment

In its approach to loss function assignment, YOLOv8 introduces innovative strategies that distinguish it from conventional object detection systems, which typically rely on predefined anchor boxes for label assignment and loss computation. These traditional systems often require extensive hyperparameter tuning across different datasets, a challenge that YOLOv8 addresses with its adaptive strategies.

Task Alignment Learning (TAL): YOLOv8 incorporates Task Alignment Learning to synchronize the classification and regression tasks more effectively. Unlike conventional methods where these tasks may diverge, TAL aligns these tasks to enhance consistency and accuracy in the model's outputs. This alignment ensures that the learning objectives for both tasks are integrated, which helps in reducing discrepancies between classification scores and the predicted bounding

boxes' accuracy. The TAL method is not redundant; rather, it provides a novel way to ensure that adjustments in one task directly benefit the accuracy of the other, leading to more reliable and precise detection outcomes.

Distribution Focal Loss (DFL)[35]: Alongside TAL, YOLOv8 utilizes Distribution Focal Loss, which differs from standard Focal Loss by allowing the model to adapt more dynamically to the varying scales of objects within the dataset. DFL calculates loss based on the probability distribution of the class predictions, which helps in handling class imbalance more effectively — a common issue in object detection. This feature is particularly useful in mineral classification where object sizes can vary greatly.

CIoU Loss[34]: Furthermore, YOLOv8 combines DFL with Complete Intersection over Union (CIoU) loss. CIoU loss not only focuses on the overlap between the predicted and actual bounding boxes but also includes distance metrics between the box centers and aspect ratio terms. This comprehensive approach reduces the localization error and enhances the alignment of predicted boxes with ground truth, which is critical for achieving high precision in object detection tasks.

By integrating these advanced loss functions, YOLOv8 significantly refines its ability to detect and classify objects accurately across different scenarios and scales. The model's focus on dynamically adjusting its parameters based on the actual data it encounters makes it robust against the variations and complexities often found in diverse detection environments. This methodical enhancement of the loss function framework not only improves detection accuracy but also optimizes the overall training efficiency, making YOLOv8 a superior choice for modern object detection challenges.

2.6 Detailed overview of YOLOv8-CLS model and comparison with YOLOv8

2.6.1 Introduction to YOLOv8-CLS

YOLOv8-CLS is a specialized variant of the YOLOv8 model, tailored specifically for image classification tasks, unlike the broader object detection capabilities of the standard YOLOv8. This adaptation allows YOLOv8-CLS to focus on identifying and categorizing objects within an image without the additional computational overhead associated with bounding box predictions.

YOLOv8-CLS incorporates several key modifications that optimize it for classification:

Reduced complexity: The model simplifies the network architecture by omitting the Spatial Pyramid Pooling Fusion (SPPF) layers and the bounding box prediction modules found in YOLOv8.

Enhanced feature extractor: It employs a modified backbone that prioritizes features relevant for class categorization, enhancing the discriminative power of the network for classification tasks.

2.6.2 Comparative analysis with YOLOv8

While YOLOv8 is renowned for its efficiency and accuracy in object detection across various domains, YOLOv8-CLS refines the architecture to enhance image classification. Key differences include:

- **Model architecture:** YOLOv8-CLS uses a streamlined architecture with fewer convolutional layers, focusing solely on classifying the primary content of images. This results in faster processing times and reduced model size, suitable for deployment in environments with limited computational resources.

- **Optimization for classification:** The training process of YOLOv8-CLS has been optimized to maximize classification accuracy by eliminating loss functions unrelated to classification, retaining only the classification loss function to simplify the computational complexity of the loss function.

Figure 6 represent the architecture of YOLOv8-CLS, using different colors to distinguish the removed and modified components compared to the standard YOLOv8. Figure 7 provide a step-by-step visualization of the data flow within YOLOv8-CLS, emphasizing how input images are processed through the modified network to produce class predictions.
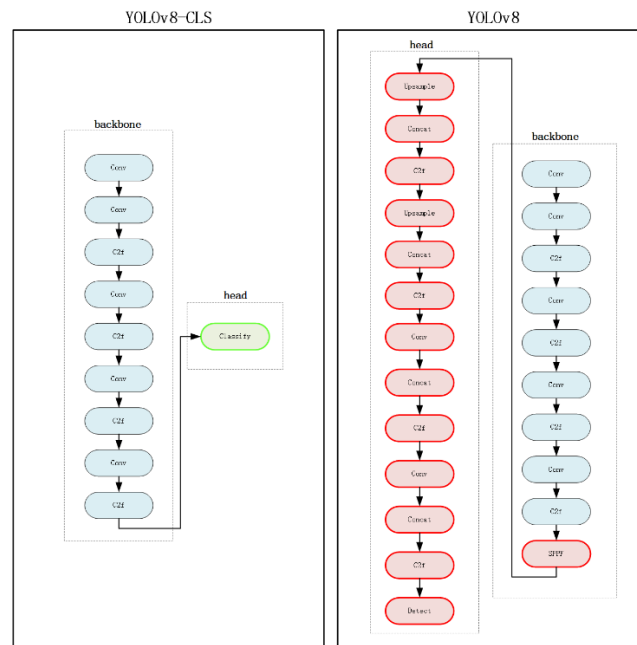
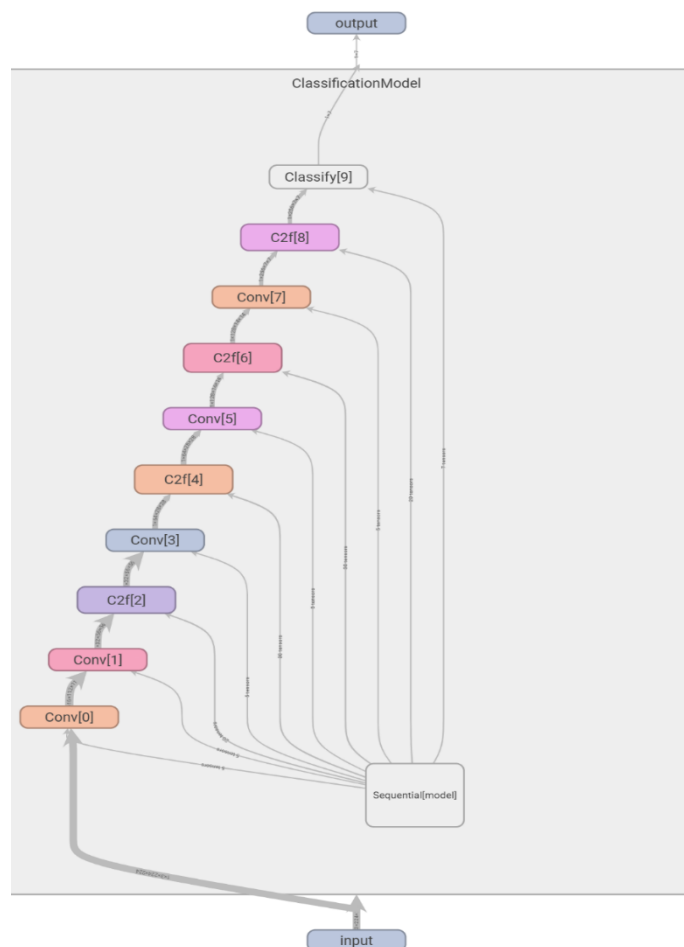Figure 6: Architectural differences between YOLOv8 and YOLOv8-CLS

Figure 7: Data processing flow in YOLOv8-CLS

3 Data preparation and augmentation

3.1 Collection of mineral image data

The model's capacity to discover deep attributes is highly affected by its interior style and the variety within the training information. As the training dataset broadens, the model removes more thorough attributes. This enhancement dramatically enhances the model's generalization and projection capabilities, enabling it to deal with even more elaborate and tough classification issues. By incorporating a varied series of features, the model becomes better furnished to handle intricate scenarios come across during training, recognition, and real-world identification tasks. Such a durable attribute collection is essential for ensuring the model's performance not just in scientific research yet additionally in sensible applications where precision and versatility are critical. This dataset was originated from an openly readily available mineral recognition dataset [27], making up seven minerals: bornite, quartz, malachite, pyrite, muscovite, biotite, and chrysocolla, causing an overall of 5,536 photographs after screening to get rid of unfavorable information. The dataset was divided into 3,872 images for the training set, 1,106 images for the validation collection, and 558 photos for the test set. The classification of the dataset for every mineral is presented in Table 1, while example pictures are shown in Table 2.

Table 1: Composition of the dataset [27]

Mineral species	Training Set (number of photos)	Validation Set (number of photos)	Test Set (number of photos)
biotite	721	206	103
quartz	812	232	116
bornite	290	83	42
chrysocolla	371	106	54
malachite	698	199	101
muscovite	235	67	35
pyrite	745	213	107
total	3243	1106	558

Table 2: Sample images for each of the 7 minerals in dataset [27]

Mineral species	training set	validation set	test set
biotite			
quartz			

bornite			
chrysocolla			
malachite			
muscovite			
pyrite			

3.2 Data augmentation

To enhance the accuracy of mineral image classification under varying conditions and improve model generalization, this study employs the built-in online data augmentation techniques provided by the Ultralytics library, integrated into the YOLOv8-CLS model. These techniques were specifically selected for their effectiveness in addressing the unique challenges posed by mineral image classification, where significant variations in texture, color, and form factor can impact classification accuracy.

Mosaic Stitching: This technique combines four different training images into a single composite image[37]. In the context of mineral images, which can exhibit considerable variability in scale and context within a single sample, mosaic stitching introduces diversity in background and scale. This method simulates a wide range of environmental conditions and teaches the model to recognize minerals across different contexts and juxtapositions, enhancing its ability to generalize to real-world scenarios. The mosaic parameter was set to 1.0, ensuring full application of this technique during training.

HSV Color Space Adjustment: Adjustments within the HSV (Hue, Saturation, Value) color space are crucial for mineral classification, as lighting conditions and camera settings can introduce substantial variations in color and brightness[37]. By randomizing these parameters during training, the model becomes more robust to such variations, ensuring it focuses on the intrinsic properties of the minerals rather than external

lighting conditions. The hue (h), saturation (s), and value (v) parameters were set to 0.015, 0.7, and 0.4, respectively, to simulate a variety of lighting scenarios while maintaining consistency in the model's learning.

Horizontal Flipping: This technique mirrors images along the vertical axis[37]. Given the isotropic nature of many minerals, where orientation does not affect their identity, horizontal flipping prevents orientation bias in the model. This helps improve the model's performance in environments where minerals may appear in arbitrary orientations. The flip horizontal (fliplr) parameter was set to 0.5, applying horizontal flipping with 50% probability during training. Vertical flipping was disabled by setting the flip vertical (flipud) parameter to 0.0, as vertical flips are generally not suitable for mineral images due to their natural orientation.

These data augmentation techniques address the specific challenges of mineral classification by increasing the diversity of training data representations. As a result, the model can perform robustly under a wide range of operational conditions. This approach aligns with recent advancements in adaptive data augmentation strategies, which dynamically adjust augmentation methods during training, thus enhancing the model's ability to generalize across various tasks. The chosen techniques and their parameters are consistent with the findings of recent research, such as Tang et al. [32], which highlights the effectiveness of real-time adaptive augmentation in improving the generalization capabilities of deep learning models.

4 Development and training of a smart mineral classification model

4.1 Model training and initial settings

In the training of the mineral classification model, our choice of epochs was guided by systematic empirical testing to determine the optimal point of convergence. Through multiple trials, we progressively adjusted the number of epochs and monitored the model's performance. We observed that by the 120th epoch, the model had already reached a plateau in learning, showing no significant gains in accuracy beyond this point. This experimental finding led us to set the maximum training limit at 120 epochs, as extending beyond this did not yield improvement and only increased computational costs unnecessarily.

The batch size of 128 was determined to be the most effective through these tests, balancing GPU utilization and memory consumption efficiently, thus allowing for smooth and effective training sessions. Similarly, the choice of an image dimension of 224 pixels was empirically optimized. Our tests showed that this dimension provided sufficient detail for the accurate classification of minerals while maintaining computational efficiency, which is critical for handling large datasets without compromising on speed or accuracy.

Regarding the learning rate, we settled on 0.00065 after experimenting with different rates to find a balance that allowed steady, effective learning without overshooting the model's capacity to converge. The application of a cosine scheduler was also an empirically motivated choice; it was selected for its effectiveness in fine-tuning the learning rate over time, which we noted helped in achieving better generalization by allowing more precise adjustments to the model weights in the later stages of training.

Additionally, our model uses **YOLOv8x-CLS** as the pre-trained weight, derived from training on a large dataset. The "X" option represents the largest model variant, with deeper layers and more network parameters, which typically leads to improved accuracy.

These parameters—each rigorously tested and chosen based on empirical evidence—collectively contributed to the robust performance of our model, ensuring that it not only learned efficiently but also generalized well to new, unseen data.

4.2 Model assessment

Unlike object detection, which utilizes multiple loss functions—including bounding box regression, classification loss, and objectness loss—to locate and categorize objects within an image, our mineral image classification model primarily relies on a single loss function—classification loss—to categorize each image into predefined mineral classes. This distinction emphasizes that our task focuses on classifying, not identifying individual mineral instances. This loss function is crucial for evaluating model performance, directing training, and enhancing the model. It measures how well the model's predictions match the true labels by quantifying the discrepancy between predicted and actual values. This measurement influences how the model updates its parameters, guiding the adjustment of weights and biases through optimization techniques like gradient descent to minimize errors and boost classification accuracy. As illustrated in Figure 8, the loss is initially high, reflecting poor model predictions at the start. Over the final 20 epochs, the loss values for both validation and test sets decrease and stabilize. For the training set, loss reduction is generally steady, though minor fluctuations are observed before convergence. The process of loss fitting on the validation set plays a crucial role in assessing the model's capacity for generalization. The model exhibited robust performance on this set, with the loss decreasing steadily and ultimately stabilizing. Final loss values were 0.26668 for the training set and 1.25980 for the validation set. The steady decline and stabilization of loss values in both convergence and overfitting analyses indicate that the model has likely converged and exhibits robust performance on the classification task.
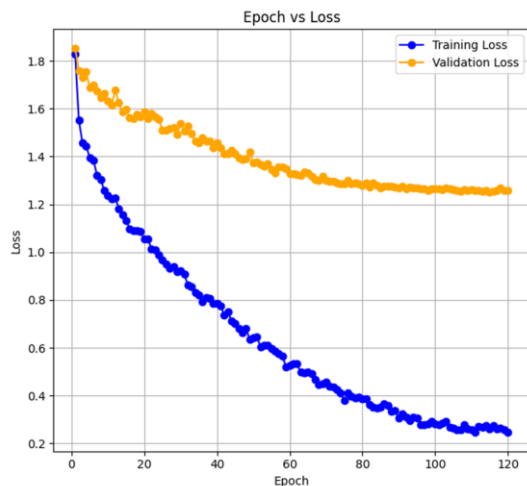
Figure 8: Loss curves for training and validation datasets

Top-1 Accuracy measures the rate at which the model's most confident prediction matches the actual label, while Top-5 Accuracy evaluates how often the true label appears among the model's top five predictions, as detailed in formulas (1) and (2). Figure 9 illustrates that, after extensive optimization and testing, the final Top-1 Accuracy and Top-5 Accuracy values reached stable levels of 0.92053 and 0.99399, respectively. These results demonstrate the model's exceptional precision and robustness in mineral image classification tasks.

$$\text{Top-1\_Accuracy} = \frac{\text{Number of correct predictions}}{\text{Total number of predictions}} \quad (1)$$

$$\text{Top-5 Accuracy} = \frac{\text{Number of times the true label is in the top 5 predicted labels}}{\text{Total number of predictions}} \quad (2)$$
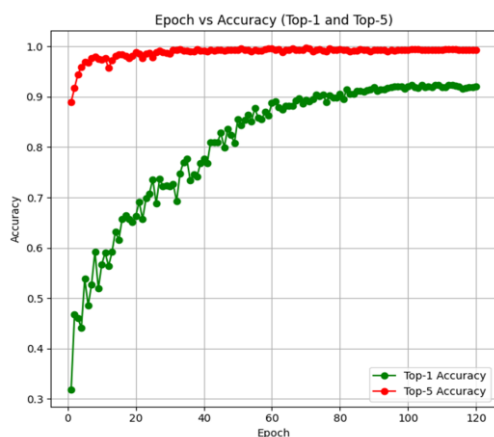

Figure 9: The curves for top-1 accuracy and top-5 accuracy

Confusion Matrix: ① A confusion matrix provides a summary of the prediction results for classification problems. The key aspect of a confusion matrix lies in its ability to quantify both correct and incorrect predictions using count values, further breaking them down by class. It visually illustrates the extent to which a classification model confuses different classes during prediction. Beyond merely identifying errors, the confusion matrix offers insight into the specific types of mistakes the model makes. This decomposition of results helps overcome the limitations of relying solely on classification accuracy as a performance metric. ② in the field of machine learning and statistical classification, a confusion matrix is a visualization tool primarily used in supervised learning, while in unsupervised learning, it is generally referred to as a matching matrix. Each column of the matrix represents the predicted instances of a given class, while each row corresponds to the actual instances of that class. The term "confusion matrix" originates from its function—making it easy to observe whether the model confuses different classes by misclassifying one as another.

Figure 10 displays the confusion matrix from this training experiment, providing several key insights: (1) There is no misidentification of backgrounds in the images. (2) The rate of false positives, where images are incorrectly classified, is maintained below 15%, demonstrating the model's overall precision. (3) The true positive rates for biotite, bornite, chrysocolla, malachite, muscovite, pyrite, and quartz are 0.96, 0.80, 0.90, 0.96, 0.66, 0.96, and 0.97, respectively. These rates reflect the model's accuracy in classifying each mineral, with true positive rates being crucial as they measure the model's ability to correctly identify each mineral type as present when it indeed is. This metric is particularly important in mineral classification to ensure that each type is accurately detected and not confused with others, which can directly impact mining and processing efficiency. (4) Bornite and muscovite have true positive rates below 0.9, indicating weaker recognition performance. This suggests specific issues with feature extraction for these minerals. The lower true positive rate for bornite and muscovite can be attributed to their complex physical properties, which may not be as distinct or may be similar to other minerals in the dataset, leading to challenges in effectively distinguishing them from others. This highlights the need for targeted improvements in the feature extraction process, such as enhancing the model's ability to handle variations in color, texture, and shape more distinctively for these minerals.
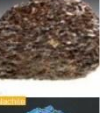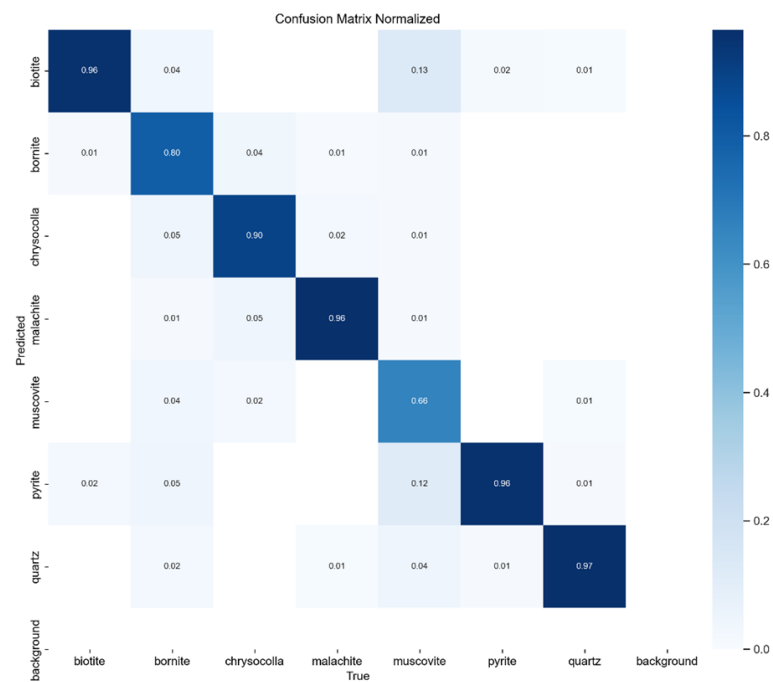
Figure 10: Confusion matrix of the validation set

5   Model testing

To assess the generalization capability of the trained model for mineral classification, we conducted testing using the test set from Table 1, which consists of 558 images. These images were not used in the training or validation stages, ensuring the evaluation was conducted on unseen data. The test set includes samples of 7 minerals: bornite, quartz, malachite, pyrite, muscovite, biotite, and chrysocolla. Each mineral has distinct features that challenge the model's identification capacities. For example, biotite is determined by its dark, plate-like framework, quartz is notable for its openness and variety of forms, and bornite is distinguished by its copper-green shade and metal sparkle. The performance of the model was examined by comparing truth labels with the predicted tags for each batch in the test set. The comprehensive results of this comparison are summarized in Table 3. Additionally, Figure 11 shows the complication matrix, which gives an extensive view of the model's precision and its efficiency across different mineral categories.

Table 3: Original and predicted labels for batches 0-2 in the testing.

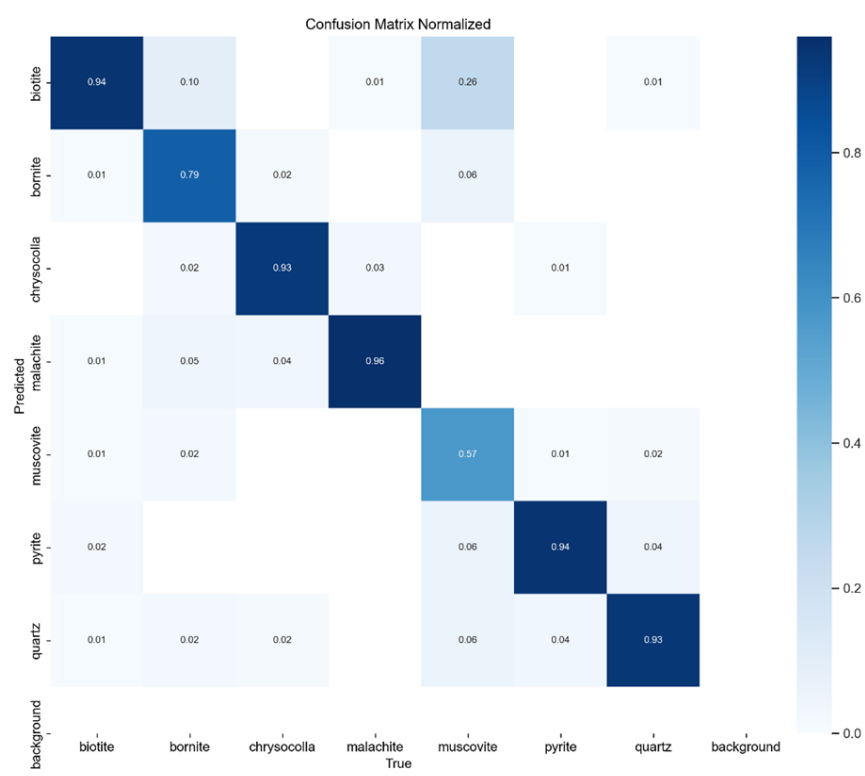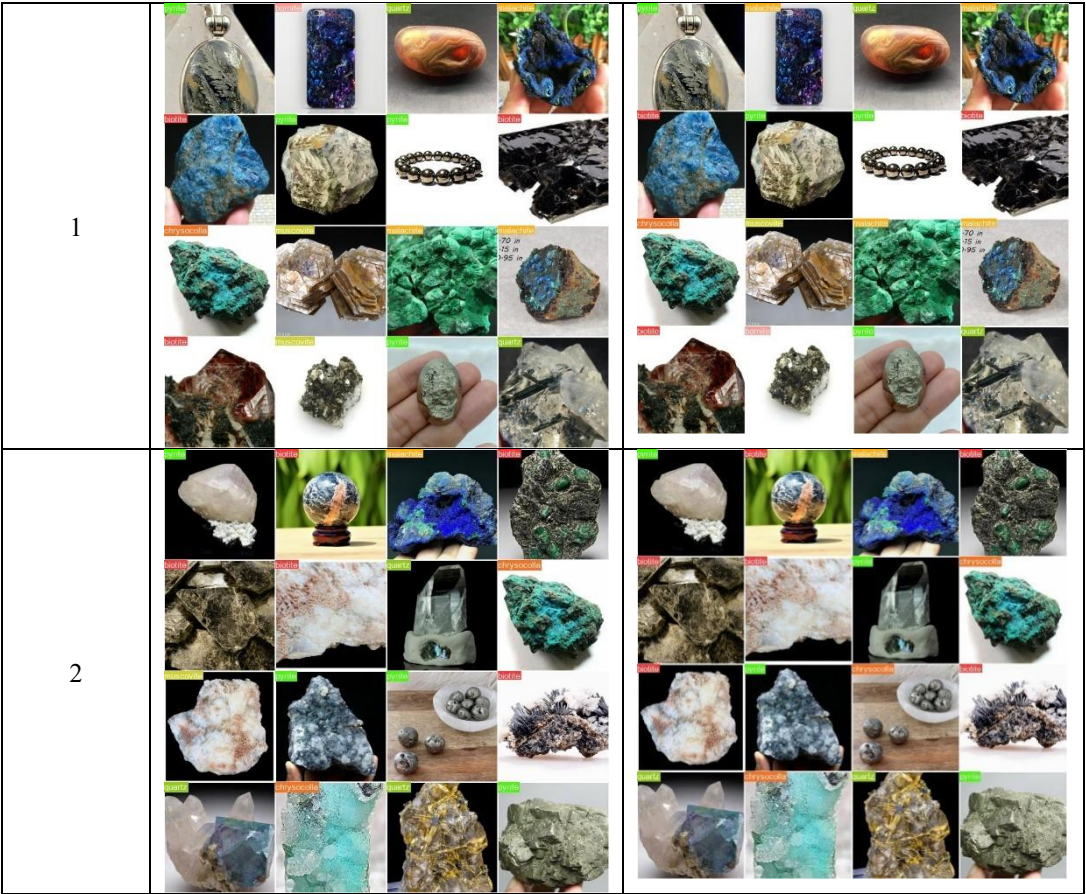| Batch	labels	predict
0		

Figure 11: Confusion matrix of the test set

An in-depth evaluation of the speculative outcomes yields useful insights right into the model's performance on numerous mineral samples and uncovers potential areas for enhancement: (1) Bornite (TP = 0.79): Bornite's complicated texture and the presence of impurities can obscure its features. The limited sample size in the

training set hampers effective feature learning, resulting in some false positives. (2) Quartz (TP = 0.93): The transparency and diverse morphology of quartz provide abundant feature information for accurate identification. However, its transparency can blend with the background or other transparent minerals, especially under complex lighting conditions, which can slightly affect the true positive rate. To address this issue, we attempted to simulate such conditions through data augmentation, such as adjusting HSV parameters, to enhance the model's ability to accurately recognize quartz under challenging lighting. The true positive rate of 0.93 for quartz in the test set aligns well with our expectations. (3) Malachite (TP = 0.96): Malachite's unique color and ribbon-like texture make it relatively easy to identify, resulting in a high true positive rate. However, its color similarity to silica malachite and their common co-occurrence can occasionally impact accuracy. (4) Pyrite (TP = 0.94): Pyrite's distinctive golden luster and cubic structure allow it to be easily distinguished from other minerals, contributing to a high true positive rate. (5) Muscovite (TP = 0.57): Muscovite's bright white luster and plate-like structure are prone to reflections, which can affect recognition accuracy under varying lighting conditions. Additionally, its similarity in color and morphology to other mica minerals, such as biotite, further complicates differentiation, leading to the lowest accuracy for this mineral in the model's test results. (6) Biotite (TP = 0.94): Biotite performed well in the test set but encountered certain challenges. Its dark, plate-like structure is often mistaken for other dark minerals or even shadows, particularly under weak or uniform lighting conditions, which may significantly reduce prediction accuracy. Moreover, the highly reflective surface of the mineral can cause misleading highlights, potentially leading to false positives during detection. These combined factors challenge the model's ability to accurately classify biotite, highlighting the need for improved strategies to address these issues. (7) Chrysocolla (TP = 0.93): Chrysocolla's distinctive bright color and fibrous structure facilitate its identification, resulting in high accuracy. However, its color similarity

to other green minerals, such as malachite, can lead to reduced recognition accuracy when mixed with them. Additionally, overlapping shadows and mineral co-occurrence may slightly reduce the overall accuracy of detection.

The variations in true positive rates for different minerals highlight the model's varying effectiveness in recognizing their distinct physical properties. Minerals with unique features, such as pyrite and malachite, benefit from their distinct colors and shapes, which result in higher classification accuracy. Pyrite, with its golden luster and cubic structure, and malachite, characterized by its vibrant green color and ribbon-like texture, are easily distinguishable, contributing to their high true positive rates. In contrast, minerals like muscovite and bornite present additional challenges. Muscovite's bright white luster and plate-like structure can reflect light in ways that complicate accurate recognition, and its similarity to other mica minerals further hinders its identification. Bornite's complex texture and the presence of impurities in some samples also add to the difficulty, resulting in lower recognition accuracy. These discrepancies highlight areas where the model's performance can be improved. To address these issues, future optimization efforts should focus on refining feature extraction techniques and incorporating a broader range of training examples. Enhancing the model's ability to handle varying lighting conditions and overlapping mineral characteristics will be crucial for improving its precision and efficiency in real-world applications.

6　Model comparison

To further investigate the performance gap between the YOLOv8-CLS model and classical models, we also introduced another open-source dataset with pre-defined training, test, and validation sets (as shown in Table 4) to explore the differences in generalization capabilities across various models. The following presents the specific experimental procedures and results.

Table 4: Composition of the dataset [36][27]

Mineral species	Training Set (number of photos)	Validation Set (number of photos)	Test Set (number of photos)
biotite	306	11	5
quartz	309	10	5
bornite	306	8	5
chrysocolla	639	21	9
malachite	318	10	5
muscovite	459	15	8
pyrite	585	17	10
total	2922	92	47

We compared the YOLOv8-CLS model with the classical ResNet50 and ResNet101 architectures. To control for variables and ensure the richness of the test

samples, we used the dataset from Table 1 for training ResNet50 and ResNet101 with the same training parameters and data augmentation settings. The models

were tested using the test set portion from Table 1 and the training set portion from Table 4. The test results are shown in Figures 12 and 13, and the performance comparison between the models is summarized in Table 5.

Table 5: Comparison of model performance [27]

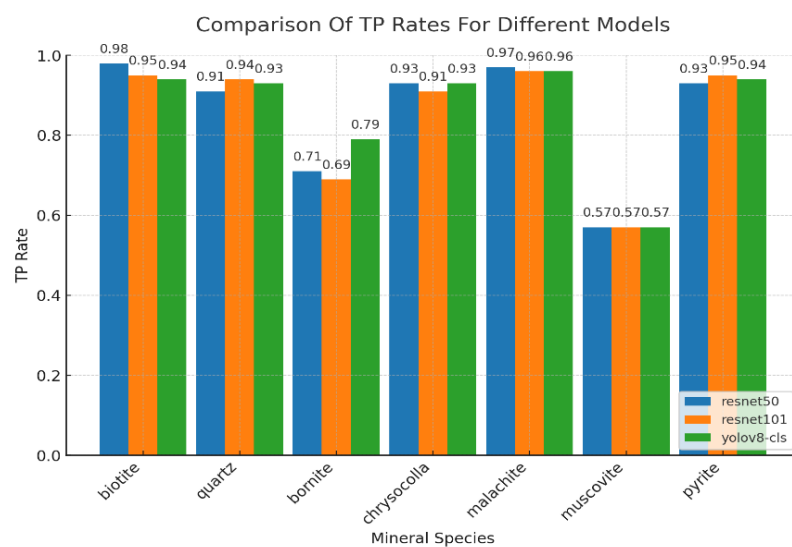| Model	GPU memory usage(training)	Top1 Accuracy in Table 1 test set	Top1 Accuracy in Table 4 train set	Parameters	GFLOPs
Resnet50	8.69GB	0.90502	0.79535	26113159	68.8 GFLOPs
Resnet101	12.52GB	0.90322	0.75941	45079175	129.4 GFLOPs
YOLOv8-CLS	1.68GB	0.90681	0.80219	1443847	3.3GFLOPs

Figure 12: Comparison of TP rates for different models on table1's test set
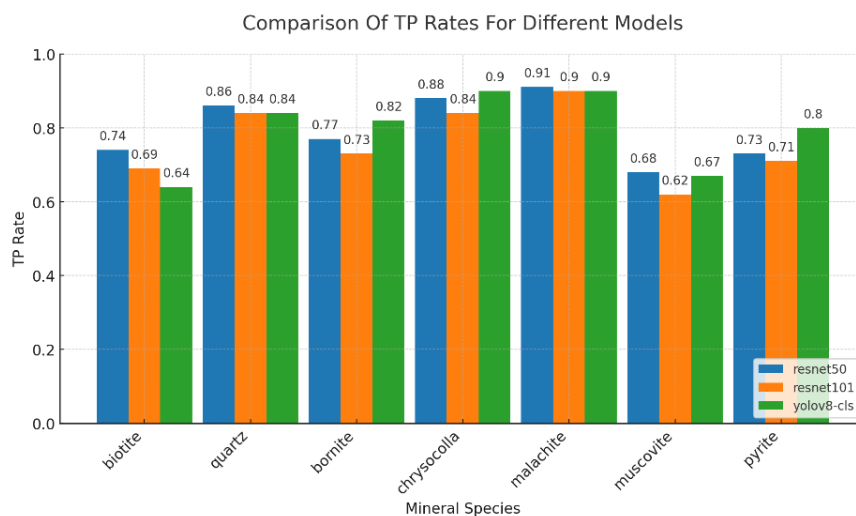

Figure 13: Comparison of TP rates for different models on table4's train set

In Table 5, we compare model performance using the memory usage during training, model parameter count, GFLOPs, and Top-1 Accuracy on both the test set from Table 1 and the training set from Table 4. Since Top-1 Accuracy is a stricter metric and more reflective of model performance differences, it was used as the

primary comparison criterion. From the results in Table 5, the YOLOv8-CLS-trained mineral classification model achieves the highest Top-1 Accuracy on both the Table 1 test set and the Table 4 training set compared to all other models, with values of 0.90681 (Top-1 Accuracy on Table 1 test set) and 0.80219 (Top-1 Accuracy on Table 4 training set), respectively. In comparison, ResNet50 and ResNet101 show a Top-1 Accuracy improvement over YOLOv8-CLS of 0.00179 and 0.00359 on the Table 1 test set, and 0.00684 and 0.04278 on the Table 4 training set, respectively.

In terms of memory usage during training, YOLOv8-CLS uses only 1.68GB of GPU memory, while ResNet50 and ResNet101 both exceed 6GB. Additionally, YOLOv8-CLS has the lowest model parameter count and GFLOPs. Specifically, YOLOv8-CLS has 20.85 and 39.21 times fewer GFLOPs than ResNet50 and ResNet101, respectively, which highlights the significant performance advantage of YOLOv8-CLS.

By analyzing Figure 12, we found that YOLOv8-CLS outperforms ResNet50 and ResNet101 in terms of true positive rates for bornite in the Table 1 test set by 0.08 and 0.1, respectively. However, for other minerals, the true positive rates were comparable. In Figure 13, YOLOv8-CLS shows a clear advantage in true positive rates for bornite and pyrite, but it lags behind ResNet50 and ResNet101 in the true positive rate for biotite. This indicates that YOLOv8-CLS has superior generalization ability compared to ResNet50 and ResNet101.

Furthermore, a comparison of Figure 13 with Figure 12 shows a significant decline in true positive rates for biotite and pyrite, while the true positive rates for bornite and muscovite improved. The true positive rates for other minerals showed slight declines. This change can be attributed to the shift in the test dataset. Upon visually inspecting the images in the Table 4 training set, we observed that the images in this set came from more complex environments that better reflect real-world application scenarios. This exposed a challenge for the model in accurately identifying biotite under complex lighting conditions, particularly low light, which aligns with our analysis in Section 5. The decrease in recognition accuracy for pyrite may be due to the intense reflections caused by its unique golden luster under strong lighting conditions with high contrast, which could also have contributed to the drop in performance.

Overall, YOLOv8-CLS demonstrates higher accuracy, lower performance overhead, and stronger generalization capability compared to ResNet50 and ResNet101 in the mineral classification task. However, it also exposes potential issues, as discussed in Section 5, particularly in handling specific lighting conditions and mineral feature overlap.

7　Discussion

The model developed in this experiment demonstrated an exceptionally high accuracy in mineral classification when tested on images from the Table 1 test set, achieving a Top-1 accuracy of approximately 90.68%. This result highlights the model's significant performance advantages in a controlled environment, demonstrating its effectiveness in both theoretical research and practical applications. The high precision not just verifies the model's robustness in controlled speculative setups yet likewise highlights its possibility for real-world usage, where accurate mineral classification is essential. Nevertheless, the model showed misclassification when managing mineral classification under particular illumination conditions. For instance, as shown in Figure 12, the model misclassified quartz as pyrite. This misclassification may arise from the openness buildings of quartz under particular illumination conditions engaging with the history, causing characteristics that carefully appear like those of pyrite. Although the confusion matrix in Figure 11 shows that the model achieves a high true positive rate of 0.93 for quartz, the misclassification examples in Figure 14 and the decline in true positive rate to 0.84 during model comparison indicate that there is still room for improvement in the model's recognition accuracy for quartz. This example also highlights performance issues under certain lighting conditions, which aligns with the conclusions from previous experiments. This points to a key area for improvement in our future work and emphasizes the limitations of relying solely on existing confusion matrices and other performance metrics to expose model weaknesses. Therefore, it is essential to test the model in real-world mineral classification scenarios, where practical testing in real-world environments can identify and address potential defects—this will be a primary focus of our future research. Similar challenges have been observed in Salima's research[30], where environmental factors such as lighting variations and image quality can significantly impact model performance.

Figure 14: Miscalculation sample (real mineral classification on left, misclassification on right)

Through the model comparison section, we further analyzed the performance of the YOLOv8-CLS model in mineral classification. This analysis demonstrates that the YOLOv8-CLS model, in comparison to ResNet50 and ResNet101, not only achieves higher accuracy and stronger generalization capabilities in mineral classification but also reduces hardware performance demands. However, it also reveals that the model's misclassification issues are not only related to minerals with similar visual features but are also closely tied to environmental lighting conditions. This highlights the model's limited ability to accurately distinguish mineral types in complex environments, emphasizing specific areas where its efficiency can be improved and providing insights for future improvements aimed at addressing these issues. SHE's study has suggested that integrating enhanced image processing techniques, such as Retinex-based image enhancement, can mitigate these challenges by improving feature extraction and contrast in low-light conditions [31]. To overcome these limitations, several improvements could be explored in future research study: (1) Expand the training dataset to consist of a larger range of minerals with similar appearance features, which will certainly help the model compare refined differences more effectively. (2) Enhance the model architecture by incorporating innovative techniques like deeper convolutional networks and interest devices. Deeper convolutional networks can boost function extraction by capturing a lot more intricate patterns and information, while interest mechanisms allow the model to focus on one of the most relevant parts of the image, even in the middle of intricate backgrounds. These improvements are expected to substantially improve the model's ability to process and properly identify complicated visual circumstances, leading to more trusted mineral identification. (3) Incorporate advanced post-processing techniques, such as sophisticated image segmentation and enhancement methods, to successfully lower misclassification prices and dramatically enhance total precision. Image division can assist isolate private minerals for more precise analysis, while improvement methods can clear up image information and improve feature visibility. SHE's research has demonstrated that

image enhancement methods, particularly those designed for low-contrast environments, can improve classification accuracy in vision-based tasks [31]. (4) Increase image sampling rates to capture more detailed mineral features, thereby providing the model with additional information for better feature extraction.

Furthermore, the testing conducted in this research study was mainly performed under regulated speculative conditions, which gave an organized setting for evaluating the model's performance. Nevertheless, the model's performance in real-world applications has actually not yet been completely examined. This constraint suggests that while the design demonstrates encouraging lead to managed settings, additional research study and recognition are needed to analyze its integrity and accuracy in different and practical scenarios. Real-world applications often involve unpredictable environmental factors, and Salima's study have highlighted the importance of adaptive deep learning approaches in handling such variations [30]. So Real-world testing will be crucial for validating the model's effectiveness and adaptability to varying environmental conditions and practical use cases. Given the challenges inherent in real-world settings-- such as varying lighting issues, history noise, and the exposure of many overlapping minerals-- it is important to carefully check and verify the variation under a lot more varied and dynamic conditions. This will help ensure that the model carries out dependably throughout a more comprehensive series of circumstances. Such useful examinations will certainly use a clearer understanding of the model's efficiency and its capability to get used to real-world complexities. This will certainly make certain the variation's durability and stability in diverse and much less regulated situations. For example, performing experiments in real mineral arranging manufacturing environments or in field treatments for geological exploration would provide a much more precise examination of the style's effectiveness and applicability. These real-world tests will certainly enable scientists to evaluate the style's effectiveness under differing conditions and recognize prospective problems related to complex histories, lighting changes, and mineral

problem. These field tests not just offer helpful suggestions for style optimization yet likewise expose its restraints and areas for improvement in reasonable applications, therefore making sure safe and reputable effectiveness in more variable and dynamic atmospheres.

Owing to constraints imposed by existing experimental infrastructure, our current investigation lacks extensive field validation in operational industrial settings to empirically evaluate the performance boundaries of our mineral identification framework under unconstrained environmental parameters. Through rigorous analysis of experimental datasets complemented by theoretical foundations established in SHE's seminal research on environmental interference mitigation in coal mining applications [31], we postulate that the primary challenges during practical deployment likely stem from four interdependent mechanisms: (1) suboptimal lighting conditions characterized by inadequate illumination intensity and reduced contrast ratios; (2) complex light-matter interactions arising from the anisotropic optical properties of translucent mineral specimens; (3) inter-class feature ambiguity in crystalline morphology descriptors; and (4) resolution-dependent attenuation of discriminative textural signatures. These compounded photometric and geometric variabilities could engender nonlinear error propagation through the feature extraction pipeline, potentially inducing cascading effects that degrade critical classification performance indicators such as Top-1 Accuracy and Top-5 Accuracy.

Finally, while the model demonstrates notable efficiency under the current experimental conditions, further research is essential to fully realize its potential. To ensure the model's broader application and performance, it is crucial to address possible misclassification issues and enhance its adaptability to real-world scenarios. By tacking these challenges, the model's robustness will be significantly improved. As suggested by SHE and Salima's prior research, integrating adaptive learning techniques and real-time enhancement algorithms may provide a viable path forward in enhancing classification accuracy in challenging environments[30] [31]. This, in turn , will provide a solid theoretical foundation and the necessary technical support for the widespread adoption of mineral recognition technology. As the model is further tested and refined in diverse practical settings, mineral recognition modern technology is expected to play an increasingly significant role in both industrial and research contexts. This methodological advancement demonstrates significant potential for enhancing precision in automated mineral categorization systems. The industrial implementation of our model could optimize operational efficiency in ore processing workflows through streamlined mineral classification processes, thereby mitigating resource depletion and minimizing material wastage. Furthermore, the framework exhibits extensible applicability in geological prospecting domains, where it may serve as a decision-support tool for field geologists by enabling real-time mineralogical analysis through integration with portable multispectral imaging devices. The architecture's inherent adaptability also suggests promising applications in automated petrographic analysis and remote sensing-based mineral mapping initiatives.

8   Conclusion

This paper presents an advanced intelligent mineral image classification model using deep learning techniques, specifically leveraging the YOLOv8-CLS algorithm. The model is designed to efficiently identify seven usual mineral kinds: bornite, quartz, malachite, pyrite, muscovite, biotite, and chrysocolla. After training for 120 epochs, the model's performance metrics showed substantial stabilization, with Top-1 precision and Top-5 accuracy reaching 0.92053 and 0.99399, particularly. These outcomes highlight the model's remarkable precision and robustness in mineral image classification tasks. The high-performance metrics highlight the variation's effectiveness and capacity for useful application in correctly figuring out a range of mineral types, showcasing its viability for use in different mineral acknowledgment situations.

Although the model performs excellently overall, some limitations were revealed when testing with an additional dataset during the model comparison section. The observed misclassifications are primarily attributed to the visual similarities among specific minerals such as shade, shape, and appeal. This similarity complicates the style's job of exactly distinguishing between these minerals, for that reason evaluating its total effectiveness in classification. Particularly when the mineral to be identified shares similar morphology and color with other minerals recognized by the model, the model may struggle to distinguish the subtle differences between them, leading to biased classification results. The decrease in the model's robustness under complex lighting conditions may be attributed to strong reflections and refractions occurring when light strikes the minerals. The conclusions drawn from this research study deal significant insights that are important for advancing smart mineral classification techniques. These findings provide a basic understanding that can help future r & d in this area. Via a thorough analysis of the source of model's errors and targeted optimization instructions, this research reveals the application possibility of deep learning models in mineral classification and information paths for enhancing variation efficiency in future research study. These results develop a solid basis for advancing clever mineral classification innovation and encouraging its functional release and more comprehensive use in real-world applications.

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.

[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.

[3] BATIN, Mikhail, et al. Artificial intelligence in life extension: from deep learning to superintelligence. Informatica, 2017, 41.4.

[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.

[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.

[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, https://doi.org/10.3390/min11050506.

[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, https://doi.org/10.1117/12.2668168.

[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, https://doi.org/10.3390/s19183914.

[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, https://doi.org/10.1002/aisy.202300206.

[10] L Chen, S Li, Q Bai, J Yang, S Jiang, Y Miao. Review of Image Classification Algorithms Based on Convolutional Neural Networks. Remote Sensing, 2021 - mdpi.com, https://doi.org/10.3390/rs13224712

[11] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 770-778).

[12] Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 2261-2269).

[13] GAAFAR, Alaa Sahl; DAHR, Jasim Mohammed; HAMOUD, Alaa Khalaf. Comparative analysis of performance of deep learning classification approach based on LSTM-RNN for textual and image datasets. Informatica, 2022, 46.5, https://doi.org/10.31449/inf.v46i5.3872.

[14] Dosovitskiy, A., Beyer, L., Kolesnikov, A., et al. (2021). An image is worth 16x16 words: Transformers for image recognition at scale. In Proceedings of the International Conference on Learning Representations (ICLR).

[15] Howard, A. G., Zhu, M., Chen, B., et al. (2017). MobileNets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861.

[16] He, L.; Zhou, Y.; Zhang, C. Application of Target Detection Based on Deep Learning in Intelligent Mineral Identification. Minerals 2024, 14, 873, https://doi.org/10.3390/min14090873.

[17] 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.

[18] 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, https://doi.org/10.1016/j.ecoinf.2022.101847.

[19] Zuraimi, M.A.B.; Zaman, F.H.K. Vehicle detection and tracking using YOLO and Deep SORT. In Proceedings of the 2021 IEEE 11th IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE), Penang, Malaysia, 3–4 April 2021; pp. 23-29, DOI: 10.1109/ISCAIE51753.2021.9431784.

[20] 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, DOI: 10.1109/ACCESS.2024.3413979.

[21] 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, https://doi.org/10.1007/s12559-023-10189-6.

[22] 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, DOI: 10.1109/ICDCOT61034.2024.10515343.

[23] Reis, D.; Kupec, J.; Hong, J.; Daoudi, A. Real-time flying object detection with YOLOv8. arXiv 2023, arXiv:2305.09972, https://doi.org/10.48550/arXiv.2305.09972.

[24] Wu, T.Y.; Dong, Y.K. YOLO-SE: Improved YOLOv8 for Remote Sensing Object Detection and Recognition. IEEE Trans. Appl. Sci. 2023, 13, 12977, https://doi.org/10.3390/app132412977.

[25] 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.

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

[27] Attallah, Y.; Minerals Identification & Classification. Kaggle. Retrieved 2023.

[28] 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. https://doi.org/10.3390/app14031095.

[29] Li, J.F.; Yan, C.X. YOLOv8 with Multi Strategy Integrated Optimization and Application in Object Detection. Autom. Mach. Learn. 2024, 5, 23-31, DOI: 10.23977/autml.2024.050104.

[30] Bourougaa-Tria, Salima, et al. Spubbin: Smart public bin based on deep learning waste classification an iot system for smart environment in algeria. Informatica, 2022, 46.8, https://doi.org/10.31449/inf.v46i8.4331.

[31] SHE, Dong. Retinex Based Visual Image Enhancement Algorithm for Coal Mine Exploration Robots. Informatica, 2024, 48.11, https://doi.org/10.31449/inf.v48i11.6003.

[32] Tang, Zhiqiang, et al. OnlineAugment: Online data augmentation with less domain knowledge. In: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part VII 16. Springer International Publishing, 2020. p. 313-329, https://doi.org/10.1007/978-3-030-58571-6_19.

[33] SOHAN, Mupparaju, et al. A review on yolov8 and its advancements. In: International Conference on Data Intelligence and Cognitive Informatics. Springer, Singapore, 2024. p. 529-545, https://doi.org/10.1007/978-981-99-7962-2_39.

[34] Z. Zheng, P. Wang, W. Liu, J. Li, R. Ye, and D. Ren, "Distance-iou loss: Faster and better learning for bounding box regression," in Proceedings of the AAAI conference on artificial intelligence, vol. 34, pp. 12993–13000, 2020, DOI: https://doi.org/10.1609/aaai.v34i07.6999.

[35] X. Li, W. Wang, L. Wu, S. Chen, X. Hu, J. Li, J. Tang, and J. Yang, "Generalized focal loss: Learning qualified and distributed bounding boxes for dense object detection," Advances in Neural Information Processing Systems, vol. 33, pp. 21002–21012, 2020, https://doi.org/10.48550/arXiv.2006.04388.

[36] MultiLabel classification - v1 2024-05-29 4:19pm https://universe.roboflow.com/finalproject-flqsu/multilabel-classification-vbxvr, https://doi.org/10.1007/978-3-319-41111-8_2.

[37] SHORTEN, Connor; KHOSHGOFTAAR, Taghi M. A survey on image data augmentation for deep learning. Journal of big data, 2019, 6.1: 1-48, https://doi.org/10.1186/s40537-019-0197-0.