

# *Ferrous Image Classification Based on YOLOv8*

Ang Li<sup>1</sup>, Xiaojun Qi<sup>1\*</sup>, Zhebin Yu<sup>1</sup>, Jialin Wei<sup>1</sup>, Wei Liu<sup>1</sup>, Wei Su<sup>1</sup>, Zongfa Li<sup>1</sup>

<sup>1</sup>China National Gold Group Geology Co., Ltd., Beijing, China

\*Corresponding Author. Email: 157231750@qq.com

**Abstract.** This paper, we propose a new mineral image classification method using YOLOv8 model enhanced by visual attention mechanism. The integration of attention blocks improves the model's ability to focus on relevant features, thereby reducing misclassification and improving accuracy, especially in complex and noisy environments. Experimental results using iron ore with different densities show that the attention-enhanced YOLOv8 outperforms traditional methods in ferrous iron type classification based on density, ash and microfraction. The proposed method significantly improves the efficiency of feature extraction and processing, which provides a promising solution for intelligent ore sorting in industrial applications.

**Keywords:** YOLOv8, Visual Attention Mechanism, Mineral Image Classification, Deep Learning, Intelligent Ore Sorting.

## 1. Introduction

The exploitation and application of mineral resources are entering a new phase due to rapid depletion from industrial growth, driving the need for advanced ore mining and processing technologies [1]. Intelligent ore sorting has become key, improving efficiency, safety, and sustainability in mining operations. Initially, this technology used high-tech sensors like X-ray Transmission (XRT) and X-ray Fluorescence (XRF) for particle identification and separation, offering high accuracy but facing limitations like high costs and radiation exposure [2].

Advances in digital image acquisition have shifted focus to machine vision-based sorting systems that process images to classify ore, reducing dependency on high-resolution images and manual feature extraction [3]. Both machine learning and deep learning methods are utilized, with deep learning models, especially CNNs, showing excellent performance in various classification tasks [4]. These systems automatically extract features and handle dynamic and static image classification more effectively than traditional methods [5].

However, multi-category mineral classification still faces challenges due to irrelevant information like dust and noise during the training phase, which can affect accuracy [6]. To overcome these issues, recent developments have incorporated visual attention mechanisms in CNN models, enhancing focus on essential parts of an image and improving overall classification accuracy [7].

This paper, we explore the use of visual attention in deep learning-based mineral image classification systems, in particular through different depths of YOLOv8 models tailored for ferrous ores at different density levels. The integration of visual attention modules aims to further improve

these models, improve the performance in complex classification scenarios, and provide a promising direction for improving intelligent ore sorting technology [8].

## 2. YOLOv8 improves the frame structure of the model

The YOLOv8 model encounters several challenges, including limited adaptability and insufficient feature extraction capabilities, which compromise its performance in detecting multi-scale, small, and distant objects. Furthermore, the information flow within the model is suboptimal, further hindering its effectiveness in these contexts.

This paper improves the YOLOv8-based autonomous target detection network through the integration of structural reparameterization, a bidirectional pyramid structure, and a redesigned detection pipeline. These enhancements are designed to enable high-efficiency and high-precision detection of multi-scale, small, and distant objects.

Structural reparameterization technology refines the network architecture, enhancing its capability for multi-scale and small target detection without compromising accuracy. This adjustment significantly improves the detection of targets across various sizes. Additionally, the bidirectional pyramid structure effectively processes multi-scale feature information, capturing both spatial and semantic details, which is crucial for detecting distant and small targets. The redesigned detection pipeline also optimizes information flow, boosting both the efficiency and accuracy of the model through mechanisms such as feature fusion and information transfer. Figure 1 presents the framework of the enhanced YOLOv8 model, with detailed explanations of these improvements provided below.

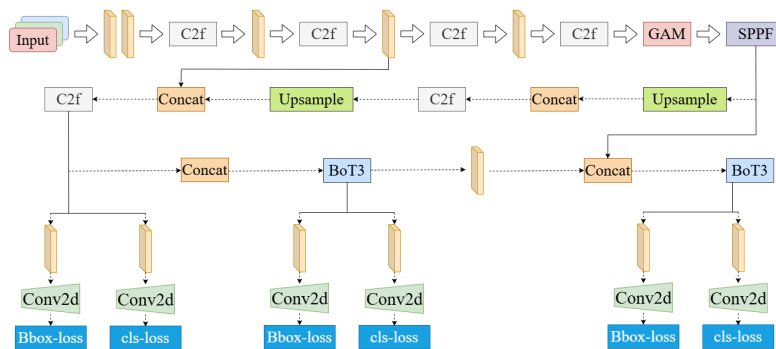


Figure 1: Improved YOLOv8 model structure

### 2.1. The introduction of backbone network introduces

Different Branch Blocks (DBB) models represent high-level architectural enhancements to the backbone network, specifically designed to improve feature extraction by merging multiple branches that target different scales, semantic information, and contextual aspects of an image. This modular approach facilitates the simultaneous extraction of different feature representations at different resolutions and semantic granularities. By using these branching structures to augment the original backbone, the network is able to effectively capture multi-scale and multi-semantic information, which significantly improves the ability to detect distant and small objects, an often challenging task in computer vision, especially in applications such as mineral detection.

In the context of smart mineral processing, the ability to detect and identify objects in the environment is essential to ensure mineral species and effective decision making. The detection must not only detect the ore itself, but also correctly identify critical environments as well as fine-

grained different ores and other potential tasks, some of which may be located at considerable distances or appear significantly reduced. Due to these factors, small and distant objects often exhibit low resolution, high levels of noise, or fine details, which make detection challenging. It is difficult for traditional methods to extract sufficient features from these objects, resulting in low detection accuracy. However, by incorporating the DBB module, the backbone network is able to adapt to these challenges by effectively handling multi-scale features and enhancing the visibility and recognition of distant and small objects in the scene.

In this paper, we propose a novel framework to integrate multiple DBB modules in the backbone network to enhance the detection capability of small size and distant targets. The incorporation of these DBB modules enables the network to exploit the complementary advantages of multi-scale and multi-semantic feature extraction to generate richer and more comprehensive feature maps. In order to further optimize the performance, the structure reparameterization technique is introduced into the C2f-DBB module to improve the detection speed and accuracy. This integration not only maintains the efficiency of the network but also contributes to a faster inference process, thus making the model more suitable for real-time applications. Figure 2 clearly illustrates the advantages of this approach, demonstrating the improvement in detection accuracy and processing efficiency using the proposed dbb enhanced network. With these innovations, a robust solution is provided to the key challenge of detecting tiny and distant targets in intelligent mineral processing systems.

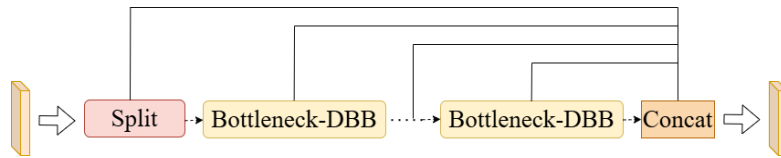


Figure 2: Structural reparameterization technology is introduced into the C2F-DBB module

## 2.2. The neck structure introduces a bidirectional pyramid structure network model

The YOLOv8 architecture utilizes a neck structure that refines feature maps from the backbone for object detection, performing key functions such as feature fusion, compression, enhancement, and adjustment to improve performance and efficiency. However, a limitation of the original YOLOv8 network's Path Aggregation Feature Pyramid Network (PAFPN) lies in its unidirectional operation. This unidirectional approach restricts the effective integration of multi-scale and semantic features, which can lead to incomplete feature extraction and reduced detection performance, particularly for objects at varying scales.

To address this, we propose the introduction of a bidirectional pyramid network into the neck structure, which maintains a similar transmission mode to the original YOLOv8, as illustrated in Figure 3. The bidirectional design allows for more flexible and effective feature integration by enabling information flow in both directions across the pyramid. This enhanced structure improves the network's ability to capture multi-scale and multi-semantic features, resulting in more comprehensive feature extraction and better performance in detecting objects across diverse scales and complexities. By overcoming the limitations of the unidirectional PAFPN, the bidirectional pyramid network significantly boosts the accuracy and robustness of autonomous target detection.

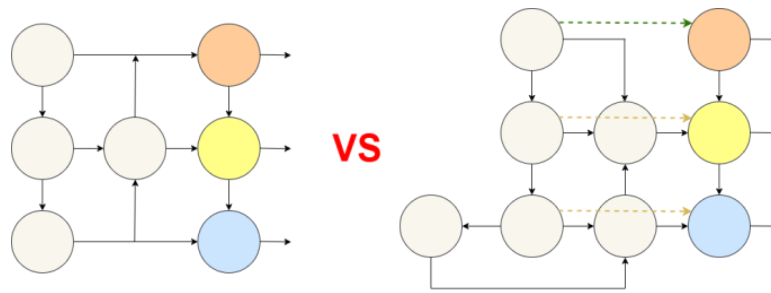


Figure 3: Comparison between the pyramid network of path aggregation features and the bidirectional pyramid structure network model

### 2.3. A new model of the inspection pipeline structure was introduced

The YOLOv8 neck structure plays a crucial role in transforming feature maps from the backbone network into optimized representations for object detection, performing operations such as feature fusion, compression, enhancement, and adjustment to improve both performance and efficiency. However, the original YOLOv8 network's Path Aggregation Feature Pyramid Network (PAFPN) is limited by its unidirectional design, which restricts the effective integration of multi-scale and semantic features. This limitation can hinder the network's ability to fully utilize relevant features, potentially affecting the accuracy of detection, especially in complex scenarios.

To overcome this challenge, a bidirectional pyramid network model is introduced, as illustrated in Figure 3. Unlike the unidirectional structure, the bidirectional design facilitates more flexible and effective feature integration by allowing information to flow in both directions across the feature pyramid. This enhanced feature aggregation enables the network to capture and combine multi-scale and multi-semantic information more comprehensively. As a result, the bidirectional pyramid network improves both the robustness and accuracy of object detection, addressing the limitations of the original PAFPN and enhancing overall network performance.

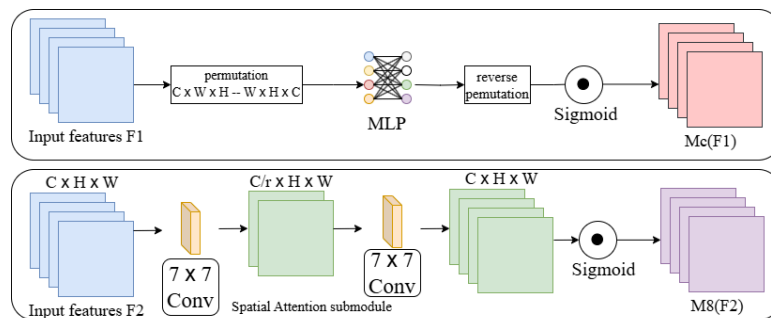


Figure 4: Schematic diagram of the structure of the new inspection pipeline structure model

## 3. Experimental results and discussion

In order to investigate the application potential of Convolutional neural Network (CNN) with visual attention mechanism in mineral image classification tasks, this experiment was conducted for three common coal particles in China: hematite, magnetite and limonite. Specifically, the ferrous particle size range selected in this study was from 13 to 25mm, and 20 kg of each category was manually screened for inclusion in the experiment.

In order to simulate the industrial separation process, the mineral particles were divided into four different density levels:  $<3.5 \text{ g/cm}^3$ ,  $3.5\text{g/cm}^3 - 4.5 \text{ g/cm}^3$ ,  $4.5\text{g/cm}^3 - 5.3 \text{ g/cm}^3$  and  $>5.3 \text{ g/cm}^3$  in this

study. These classifications are based on the density of the different iron subspecies. The ash and microfractions of the iron subspecies were analyzed before the experiment. The average ash content of each iron subspecies is shown in Table 1, and the results of the microcomponent analysis are shown in Table 2. These data can be used as a basis for understanding the characteristics of coal samples and help guide the mineral image classification process.

The purpose of this experiment is to explore how CNNS enhanced with visual attention mechanism can be used to improve the accuracy and efficiency of mineral image classification, especially to distinguish the type of ferrous iron based on density, ash and microfraction.

Table 1: Mass percentage and average ash of iron species in three density classes

Ferrous iron type	Ferrous iron property	Density level			
		<3.5g/cm <sup>3</sup>	3.5-4.5g/cm <sup>3</sup>	4.5-5.3g/cm <sup>3</sup>	>5.3g/cm <sup>3</sup>
Limonite	Ash Content	7.5%	22.8%	46.3%	85.7%
	Mass Percentage	33.3%	14.4%	20.3%	32.0%
Hematite	Ash Content	9.3%	24.3%	41.3%	87.4%
	Mass Percentage	27.6%	27.8%	11.0%	33.6%
Magnetite	Ash Content	7.1%	20.6%	40.3%	83.6%
	Mass Percentage	36.0%	23.1%	10.5%	30.5%

Table 2: Mineral Composition analysis of three density level Hematite, Magnetite and Limonite samples

Ferrous iron type	Ferrous iron property	Density level			
		<3.5g/cm <sup>3</sup>	3.5-4.5g/cm <sup>3</sup>	4.5-5.3g/cm <sup>3</sup>	>5.3g/cm <sup>3</sup>
Limonite	Limonite	67.3%	64.0%	52.6%	12.6%
	Hematite	16.4%	19.8%	22.8%	5.6%
	Magnetite	12.4%	10.5%	11.9%	4.1%
	Others	2.0%	5.7%	12.7%	77.7%
Hematite	Limonite	82.1%	74.3%	54.0%	11.0%
	Hematite	0.6%	3.7%	2.3%	0.4%
	Magnetite	13.1%	13.7%	21.8%	2.0%
	Others	4.2%	8.3%	21.9%	86.6%
Magnetite	Limonite	97.2%	88.0%	73.8%	0.0%
	Hematite	0.0%	0.0%	0.0%	15.6%
	Magnetite	0.9%	1.5%	2.6%	0.0%
	Others	1.9%	10.5%	23.6%	84.4%

The mineral composition of YOLOv8 in ferrous image classification is confused as Table 3.

The results show that the ferrous image classification based on YOLOv8 can obtain better classification results. In summary, the common YOLOv8 embedding confusion matrix of different attention blocks shows that embedding the attention block into the CNN model can effectively reduce the misjudgment rate, that is, improve the classification performance of the model for mineral images, but the misjudgment object of the CNN model will change accordingly.

Table 3: Ferrous classification result confusion matrix

	Limonite	Hematite	Magnetite	Others
Limonite	1319	43	17	19
Hematite	48	1304	29	26
Magnetite	19	37	1309	34
Others	14	16	45	1321

## 4. Conclusion

This paper proposes a novel approach to improve the accuracy and efficiency of mineral image classification using deep learning techniques, specifically focusing on integrating the visual attention mechanism in the YOLOv8 model. Adding attention blocks to the CNN effectively enhances the model's ability to focus on relevant features, thus reducing the misclassification rate and improving the overall performance, especially in complex and noisy environments.

Through extensive experiments on ferrous ores with varying density levels, we demonstrate that YOLOv8 outperforms traditional models in classification accuracy when visual attention mechanisms are incorporated. The results show that the enhanced model can effectively differentiate ferrous ores based on their density, ash content, and microfracture characteristics. Furthermore, the integration of advanced techniques, such as structural reparameterization and bidirectional pyramid networks, significantly improves feature extraction efficiency, making the model better suited for practical industrial applications.

Although the results are promising, several challenges remain, such as dealing with multi-class mineral classification with different grain sizes and densities. Future work will focus on further refining the attention mechanism, optimizing the network for higher accuracy, and exploring real-time deployment in industrial ore sorting systems.

Overall, this study has made significant progress in the field of intelligent ore sorting and mineral image classification, laying a foundation for future research to develop more robust, efficient, and accurate mining techniques.

## References

- [1] Sun Y, Ortiz J. An ai-based system utilizing iot-enabled ambient sensors and llms for complex activity tracking [J]. arXiv preprint arXiv: 2407.02606, 2024.
- [2] C. Robben, P. Condori, A. Pinto, R. Machaca, and A. Takala, "X-raytransmission based ore sorting at the san rafael tin mine," *Minerals Eng.*, vol. 145, Jan. 2020, Art. no. 105870, doi: 10.1016/j.mineng.2019.105870.
- [3] M. Massinaei, A. Jahedsaravani, E. Taheri, and J. Khalilpour, "Machine vision based monitoring and analysis of a coal column flotation circuit," *Powder Technol.*, vol. 343, pp. 330–341, Feb. 2019, doi: 10.1016/j.powtec.2018.11.056.
- [4] A. D. Gordon, L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and Regression Trees*, vol. 40, no. 3. Belmont, CA, USA: Wadsworth International Group, 1984.
- [5] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Trans. Inf. Theory*, vol. 13, no. 1, pp. 21–27, Jan. 1967, doi: 10.1109/TIT.1967.1053964.
- [6] Y. Liu, Z. Zhang, X. Liu, L. Wang, and X. Xia, "Ore image classification based on small deep learning model: Evaluation and optimization of model depth, model structure and data size," *Minerals Eng.*, 2021, Art. no. 107020, doi: 10.1016/j.mineng.2021.107020.
- [7] Z. Zhang, Y. Liu, Q. Hu, Z. Zhang, and Y. Liu, "Competitive votingbased multi-class prediction for ore selection," in *Proc. IEEE 16th Int. Conf. Autom. Sci. Eng. (CASE)*, Aug. 2020, pp. 514–519, doi: 10.1109/CASE48305.2020.9217017.

- [8] Wu B, Cai Z, Wu W, et al. AoI-aware resource management for smart health via deep reinforcement learning [J]. IEEE Access, 2023.