ISSN: 2688-8645

Research on Improved YOLOv8 Based on EMA and DyHead for Mine Target Detection

Shixian Wei, Jiaqi Shu, Kejie Zhao, Siqi Zhu, Wenjie Su, Lu Zhang

College of Electrical and Information Engineering, Quzhou University, Quzhou Zhejiang 324000, China

Abstract

As an important part of intelligent mine construction, target detection technology is crucial to ensure mine safety production. However, in practical applications, due to the drastic changes in lighting, severe dust interference, small target size and complex dynamic background in the mine environment, the information obtained by traditional detection algorithms often has large errors, leading to missed detection and false detection. In order to solve this problem, this study proposes a YOLOv8-ED model enhanced by EMA attention mechanism and DyHead dynamic detection head. By embedding EMA attention after the C2f module of the backbone network, the model's ability to extract key features in low-light and high-occlusion scenes is enhanced. The original detection head is replaced with DyHead to realize unified modeling and dynamic feature fusion in three dimensions of scale, space and task. Experiments on the self-built mine violation dataset and PASCAL VOC2012 dataset show that the mAP@0.5 of YOLOv8-ED is improved by 3.2% and 2.0% respectively compared with the original YOLOv8. This research provides an effective technical solution for intelligent mine safety monitoring and has important significance for improving the overall performance of mine detection systems.

Keywords

Mine Target Detection; YOLOv8; EMA Attention Mechanism; DyHead.

1. Research Status and Significance

With the continuous development of intelligent mining technology, mine safety monitoring is gradually transforming from manual inspection to intelligent visual monitoring. As a core technology of intelligent monitoring, target detection can automatically identify unsafe behaviors and hidden dangers in the mine, which is of great significance for preventing mine accidents and protecting miners' lives. [4]

At present, most mine target detection systems use traditional deep learning algorithms such as YOLOv5 and YOLOv7 [5]. However, the mine environment has the characteristics of uneven lighting, dense smoke, large differences in target scales and complex dynamic backgrounds, which make these algorithms have problems such as weak feature focusing ability, high missed detection rate of small targets and poor anti-interference ability [3]. In addition, the traditional Kalman filter algorithm used in some positioning systems has poor performance in complex nonlinear environments, which further affects the accuracy of target detection.

The research of this project aims to meet the market demand for high-performance mine detection systems and improve the accuracy and robustness of target detection in complex mine environments. Aiming at the limitations of traditional YOLOv8 model in mine scenes, this study adopts the method of combining attention mechanism and dynamic detection head to enhance the model's ability to extract key features and detect multi-scale targets. This method can effectively solve the problem that the target position cannot be accurately locked in

ISSN: 2688-8645

complex nonlinear environments, and provide technical support for the construction of intelligent mines. [3]

2. YOLOv8-ED Enhanced Model Design

When the mobile robot captures dynamic targets, it needs to locate the target and plan the path to achieve an efficient capture process. Similarly, when the mine detection system identifies targets, it needs to extract effective features from complex backgrounds and accurately locate targets of different scales. The specific research contents of this study include the design of EMA attention mechanism, the improvement of DyHead dynamic detection head, and the construction of YOLOv8-ED overall model.

2.1. EMA Attention Mechanism

The EMA (Efficient Multi-Scale Attention) mechanism is a lightweight attention module based on cross-space learning, which avoids the feature loss caused by dimension reduction in traditional channel attention and enhances the global context perception ability through multi-scale branch structure. [1]

EMA adopts a three-parallel branch structure to extract features: two 1×1 convolution branches perform 1D global average pooling along the horizontal and vertical directions respectively to efficiently extract global spatial context information; a 3×3 convolution branch captures multi-scale local feature descriptions to enhance the adaptability to targets of different scales. Through cross-dimensional interaction to aggregate the features of each branch, the pixel-level pairwise relationship is captured, and the global context information of the target area is highlighted. [1]

In this study, EMA is embedded after the three C2f modules of the YOLOv8 backbone network to adaptively assign attention weights on high-level feature maps, effectively distinguish targets from complex backgrounds, and significantly enhance the model's feature expression ability in low-light and smoke occlusion scenes.

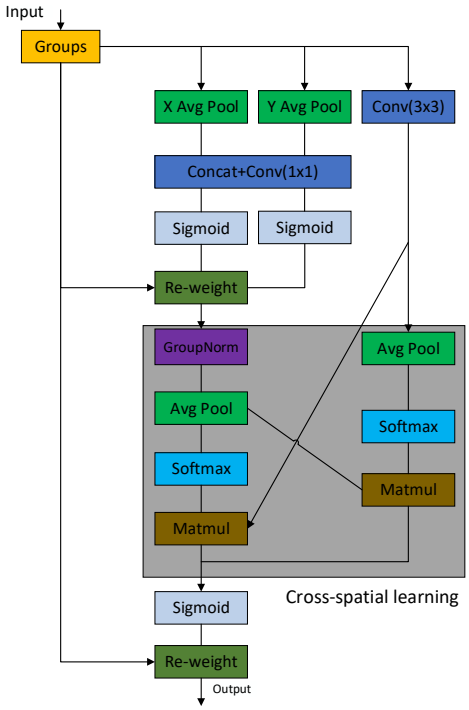

Figure 1. Schematic Diagram of EMA Attention Structure

ISSN: 2688-8645

2.2. DyHead Dynamic Detection Head

The original YOLOv8 detection head adopts a decoupled design, but has weak adaptability to multi-scale and occluded targets. Especially in mine scenes, small targets and targets occluded by smoke are prone to missed detection. DyHead optimizes the feature expression of the detection head through a three-dimensional perception enhancement mechanism [2]:

Scale perception enhancement: Dynamically adjust weights according to the semantic importance of different scale features to realize adaptive fusion of multi-scale features;

Spatial perception enhancement: Use deformable convolution to learn sparse sampling positions, focus on key target areas, and suppress background noise interference;

Task perception enhancement: Adaptively select activated channels through a switch control mechanism to optimize feature extraction for target classification and bounding box regression tasks respectively [2].

In this study, DyHead is used to replace the original YOLOv8 detection head, and the three-layer multi-scale detection mechanism is retained, which significantly improves the positioning accuracy and classification stability of small and occluded targets in mines.

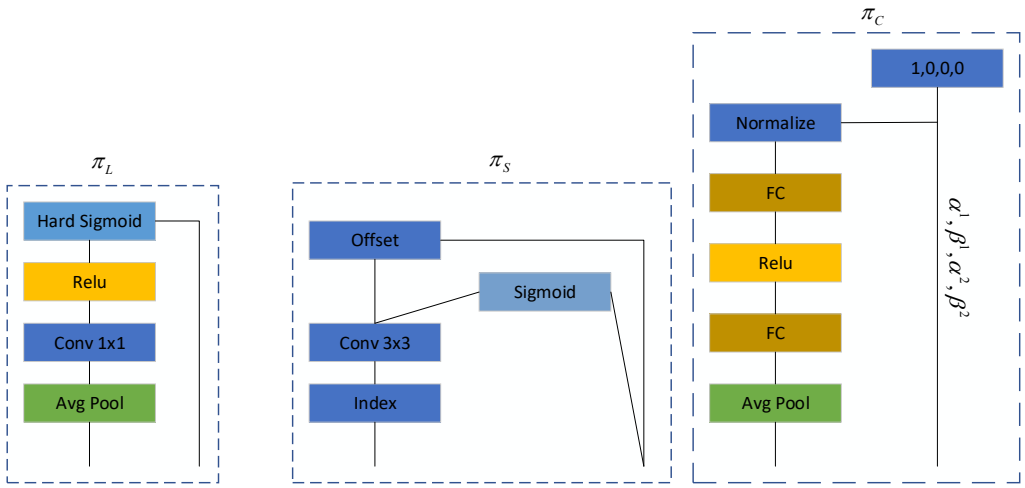

Figure 2. Schematic Diagram of DyHead Structure

2.3. Overall Structure of YOLOv8-ED Model

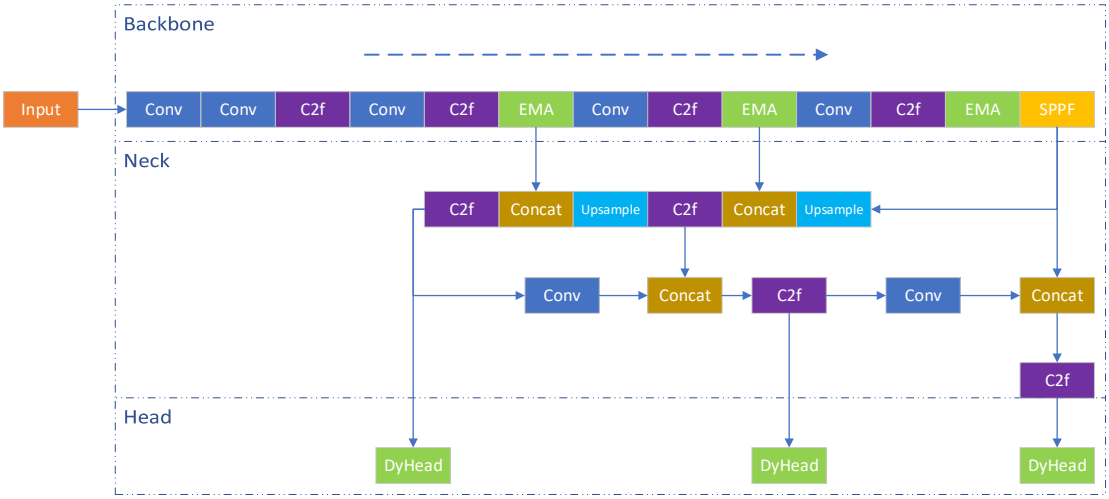

Figure 3. Schematic Diagram of YOLOv8-ED Model Structure

The YOLOv8-ED model takes YOLOv8n as the benchmark, and its overall structure is divided into four parts:

ISSN: 2688-8645

Backbone network: Adopt CSPDarknet structure to extract multi-scale features through C2f modules;

Attention enhancement module: Embed EMA attention after the three C2f modules respectively to strengthen key feature extraction [1];

Neck network: Adopt BiFPN bidirectional feature pyramid to realize efficient fusion of multi-scale features;

Detection head: Adopt DyHead dynamic detection head to output target classification and bounding box regression results [2].

EMA optimizes the feature extraction link, and DyHead optimizes the detection output link. The two work together to realize the full-link accuracy improvement from features to detection.

3. Mine Target Detection Experiment

Path planning can make the mobile robot most efficient in the current task environment. Similarly, reasonable model design can make the target detection system achieve the best performance in complex mine environments. For mine target detection, this study designs a series of experiments to verify the effectiveness of the proposed method.

3.1. Experimental Dataset

Self-built mine violation dataset: The mine scene images are collected using Hikvision DS-2DC 4423 IW-D camera, with a total of 3000 images, marking 5 types of violations: not wearing a helmet, not wearing a miner's lamp, illegal use of mobile phones, smoking, and illegal operation of power supplies. It is divided into training set (2100 images), validation set (300 images) and test set (600 images) according to the ratio of 7:1:2.

Public dataset: PASCAL VOC2012, which contains 20 categories of general targets, is used to verify the generalization performance of the model.

3.2. Experimental Environment and Parameter Setting

Hardware environment: Intel Xeon Gold 6348 CPU @ 2.60GHz, NVIDIA A40 graphics card (48GB video memory);

Software environment: Ubuntu 18.04 LTS operating system, PyTorch 2.1.2 deep learning framework;

Training parameters: batch size=8, epochs=500, initial learning rate=0.01, cosine annealing learning rate decay strategy, AdamW optimizer.

Evaluation indicators: mean average precision (mAP@0.5), number of model parameters (M), floating point operations (GFLOPs).

3.3. Ablation Experiment

In order to verify the effectiveness of each improved module, ablation experiments are carried out on the self-built mine dataset, and the results are shown in Table 1.

Table 1. Ablation experiment results on mine dataset

Model	Image Size	Parameters	GFLOPs	mAP@0.5/%
baseline	640*640	3.01M	8.2	0.852
+EMA	640*640	3.03M	8.4	0.863
+DyHead	640*640	3.49M	9.8	0.875
+EMA+DyHead	**640*640**	**3.51M**	**9.9**	**0.884**

It can be seen from Table 1 that:

ISSN: 2688-8645

Introducing EMA attention alone increases mAP@0.5 by 1.1%, only adding 0.02M parameters and 0.2GFLOPs of computation [1];

Replacing DyHead detection head alone increases mAP@0.5 by 2.3%, and the increase in parameters and computation is controllable [2];

The synergistic effect of the two increases mAP@0.5 by 3.2%, proving the complementarity and synergistic optimization effect between the modules.

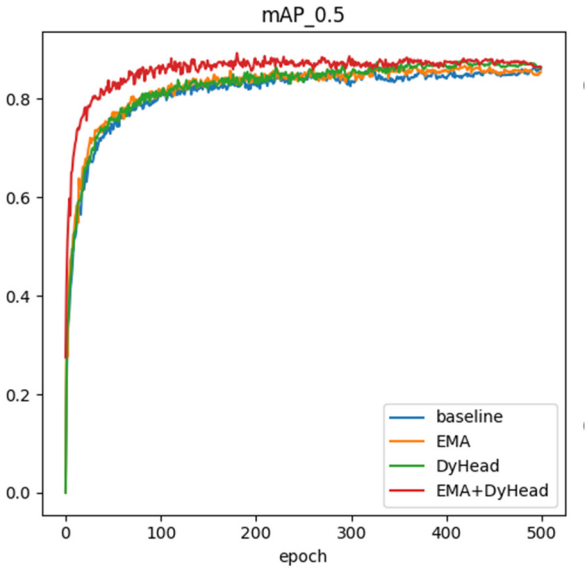

Figure 4. Average Accuracy of Different Models in Ablation Experiments

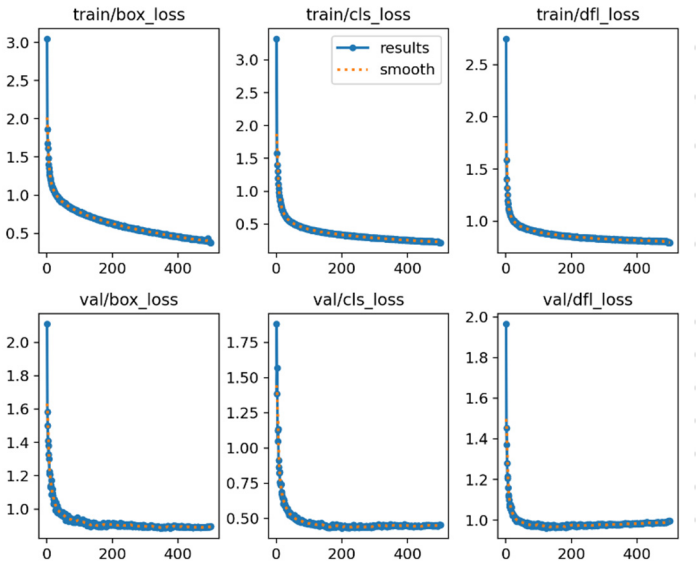

Figure 5. YOLOv8-ED Loss Function Curve

3.4. Attention Mechanism Comparison Experiment

On the VOC2012 dataset, compare EMA with mainstream attention mechanisms such as SE, CBAM, BAM, TRA and SiAM, and the results are shown in Table 2.

It can be seen from Table 2 that EMA attention increases mAP@0.5 by 1.4% with only a small increase in computation, which is better than all other comparison attention mechanisms, proving its superiority in feature enhancement [1].

ISSN: 2688-8645

Table 2. Comparison results of different attention mechanisms

Model	Image Size	Parameters	GFLOPs	mAP@0.5/%
baseline	640*640	3.01M	8.2	0.642
+EMA	**640*640**	**3.03M**	**8.4**	**0.656**
+SE	640*640	3.02M	8.2	0.644
+CBAM	640*640	3.10M	8.2	0.649
+TRA	640*640	3.01M	8.2	0.650
+BAM	640*640	3.04M	8.2	0.651
+SiAM	640*640	3.01M	8.2	0.643

3.5. Generalization Performance Experiment

Ablation experiments are carried out on the VOC2012 dataset to verify the generalization performance of YOLOv8-ED, and the results are shown in Table 3.

Table 3. Ablation experiment results on VOC2012 dataset

Model	Image Size	Parameters	GFLOPs	mAP@0.5/%
baseline	640*640	3.01M	8.2	0.642
+EMA	640*640	3.03M	8.4	0.656
+DyHead	640*640	3.49M	9.8	0.657
+EMA+DyHead	**640*640**	**3.51M**	**9.9**	**0.662**

The experimental results show that YOLOv8-ED improves mAP@0.5 by 2.0% compared with the baseline model on the general dataset, proving that the model is not only suitable for mine scenes, but also has stable performance in general target detection tasks [3] [5].

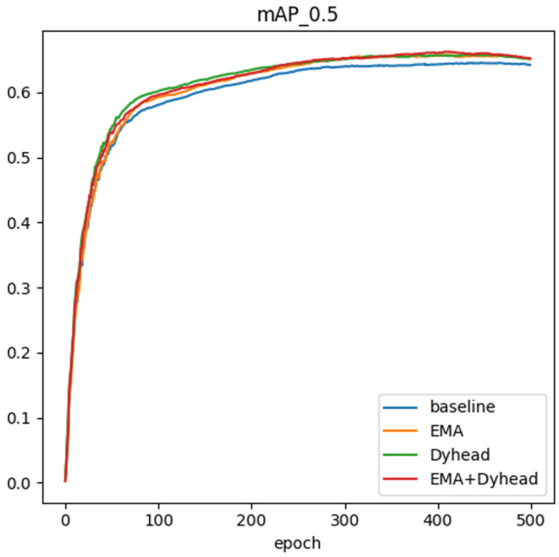

Figure 6. Average Accuracy of Different Models in Ablation Experiments

4. Conclusion and Prospect

In the mine target detection task, comprehensive consideration of feature extraction ability, multi-scale target detection performance and anti-interference ability are key factors to ensure the high efficiency of the system. The comparison and experimental verification of different methods can provide effective guidance and support for the application of target detection technology in complex mine environments. As an effective feature enhancement method,

ISSN: 2688-8645

attention mechanism is of great significance to improve the detection accuracy of the model. However, in practical applications, it is necessary to fully consider the characteristics and requirements of the system and choose the appropriate model improvement method.

Future research directions include further improving the performance of the model in more complex mine environments, and combining multi-sensor fusion technology to improve the perception ability of the system. In addition, the combination of lightweight technology and deep learning can further reduce the computational complexity of the model, so as to achieve more accurate and efficient mine target detection on embedded devices.

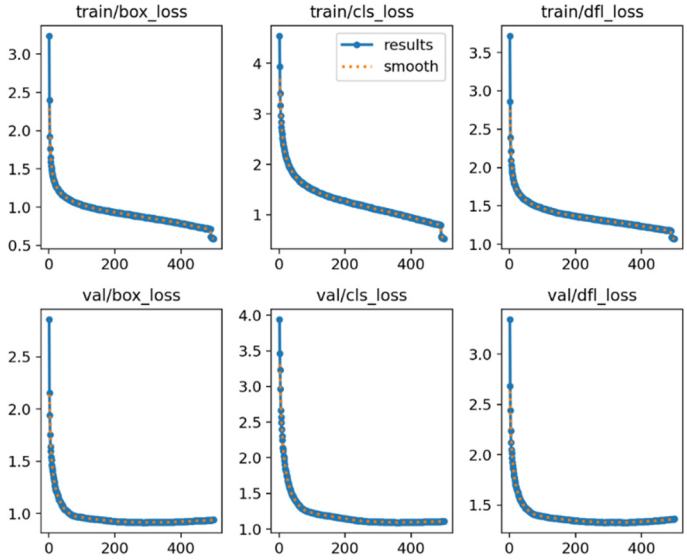

Figure 7. YOLOv8-ED Loss Function Curve

Acknowledgments

This work was partially supported by National innovation and entrepreneurship training program for College Students (No. 202511488031).

References

[1] Ouyang D, He S, Zhang G, et al. Efficient multi-scale attention module with cross-spatial learning[C]//2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2023: 1-5.

[2] Dai X, Chen Y, Xiao B, et al. Dynamic head: Unifying object detection heads with attentions [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 7373-7382.

[3] Zhang F, Zhang J R. Research review of intelligent mine target detection technology based on deep learning[J]. Coal Science and Technology, 2024, 52(6): 1-14.

[4] Yuan L. Research on China's coal mine safety development strategy[J]. China Coal, 2021, 47(06): 1-6.

[5] Shao X Q, Li X, Yang T, et al. Underground personnel detection and tracking algorithm based on improved YOLOv5s and DeepSORT[J]. Coal Science and Technology, 2023, 51(10): 291-301.