applied sciences

MDPI

Article

Robust Miner Detection in Challenging Underground Environments: An Improved YOLOv11 Approach

Yadong Li [1,2], Hui Yan [2], Dan Li [3] and Hongdong Wang [1,3,*]

[1] School of Information and Control Engineering, China University of Mining and Technology, Xuzhou 221116, China; lb20060016@cumt.edu.cn
[2] School of Intelligent Engineering, Jiangsu Vocational College of Information Technology, Wuxi 214000, China
[3] School of Information Engineering, Xuzhou University of Technology, Xuzhou 221116, China; 11433@xzit.edu.cn
* Correspondence: zs10060188@cumt.edu.cn

Abstract: To address the issue of low detection accuracy caused by low illumination and occlusion in underground coal mines, this study proposes an innovative miner detection method. A large dataset encompassing complex environments, such as low-light conditions, partial strong light interference, and occlusion, was constructed. The Efficient Channel Attention (ECA) mechanism was integrated into the YOLOv11 model to enhance the model's ability to focus on key features, thereby significantly improving detection accuracy. Additionally, a new weighted Complete Intersection over Union (CIoU) and adaptive confidence loss function were proposed to enhance the model's robustness in low-light and occlusion scenarios. Experimental results demonstrate that the proposed method outperforms various improved algorithms and state-of-the-art detection models in both detection performance and robustness, providing important technical support and reference for coal miner safety assurance and intelligent mine management.

Keywords: underground coal mines; miner detection; efficient channel attention; adaptive confidence loss

check for updates

Citation: Li, Y.; Yan, H.; Li, D.; Wang, H. Robust Miner Detection in Challenging Underground Environments: An Improved YOLOv11 Approach. *Appl. Sci.* **2024**, *14*, 11700. https://doi.org/10.3390/app142411700

Academic Editor: Pedro Couto

Received: 4 November 2024
Revised: 13 December 2024
Accepted: 13 December 2024
Published: 15 December 2024

1. Introduction

In underground coal mining environments, complex geological conditions and limited lighting pose significant challenges to personnel safety. Research on underground personnel detection in coal mines is essential for preventing safety incidents and enabling efficient emergency responses. Accurate personnel detection not only enhances the level of intelligent safety management in coal mines but also significantly reduces the time costs and safety risks associated with manual inspections. Deep learning-based object detection technology can automatically and in real-time identify and locate workers in underground operations, greatly improving detection accuracy and efficiency. The advancement of this technology is expected to overcome the limitations of traditional monitoring methods, providing technical support for building a safer and more sustainable mining environment [1,2].

In recent years, the deep learning-based object detection technique has been quite prominent in many different scenarios. Meanwhile, it is difficult to apply the YOLO series of detection algorithms directly regarding complex environments, like those encountered in a coal mine, whose performances have gained broad adoption both in accuracy and real-time detection [3]. The underground coal mining environment is marked by poor illumination, significant interference from headlamps, occlusion, pervasive dust, and varied protective clothing used by miners; all these make detection of personnel harder. Additionally, the generalization capability of the detection model and the ability to detect small objects is poor. Therefore, more upgrades are required to strengthen the model and its reliability [4–6].

Recent aspects of coal mine safety development include the application of deep learning techniques to personnel detection and object recognition in an underground environment. Basically, many studies have focused on different aspects of personnel safety or monitoring with the help of deep learning, ranging from object detection models to performance assessment under constrained conditions. Wang et al. (2020) [7] and Zhang et al. (2020) [3] investigated the application of convolutional neural networks and deep learning algorithms for personnel detection in coal mines. These works highlighted how safety has been enhanced through the automatic detection of personnel, even when visibility is low, with higher accuracy than previous methods, which relied on historical data. Gao et al. (2021) [6] developed more state-of-the-art advances in the model for real-time detecting of personnel using YOLOv4 as a backbone, hence meeting the high-accuracy and real-time performance demands of the personnel identification scenarios. Surveillance systems for underground coal mines were improved by Cheng et al. [8] and Li et al. [9]. Cheng et al. developed a real-time surveillance system using object detection, while Li et al. used an improved YOLO model to identify personnel in poor light conditions. Therefore, this work tries to overcome some of these difficulties, such as those related to low illumination and/or trouble seeing in underground situations. Chen et al. (2022) [10] combined RFID technology with deep learning and proposed an optimized scheme for positioning and tracking personnel through underground mines, with safety being the concern. Xu et al. (2022) [11], on the other hand, focused on generalization capability in a complex UCM environment when proposing improvement technologies for deep learning small object detection in high-interference conditions. A real-time object detection and recognition system for coal mine safety monitoring was realized by Xu et al. [12], and, from improvements made in the study, demonstrated a much better way of preventing such incidents from happening. Zhang et al. [13] presented some optimization of deep learning models to increase the accuracy of object detection in occlusion and dust interference complex mining environments. Li et al. (2024) [14] designed a real-time personnel detection system for underground coal mine environments with deep learning-based systems. This paper greatly improves the efficiency and accuracy of underground mine safety personnel detection. Zhao et al. (2024) [15] also developed a system that integrated infrared and thermal imaging to monitor coal mine safety and detect personnel. For example, under bad visibility conditions in the coal mine environment, it can accurately measure the human body. In a nutshell, all these works constitute an important advance in the application of deep learning to coal mine safety. Therefore, the improvement will be in real-time detection, enhancement of the model under difficult conditions, and integration of technologies like RFID for boosting the safety of personnel and efficiency in monitoring.

YOLOv11, the latest version of the YOLO series, has been continually enhanced to improve detection accuracy and speed through progressive improvements to the model structure [16]. However, due to the complex environment in coal mines, relying solely on the original YOLOv11 might be insufficient for addressing the challenges of underground mining scenarios, such as illumination variations, strong light interference, dust occlusion, and diverse worker attire. In this paper, an improved version of YOLOv11 that integrates the Efficient Channel Attention (ECA) mechanism is proposed to enhance the model's feature extraction capability. The ECA mechanism is used to effectively capture inter-channel dependencies, thereby improving feature representation, particularly in highlighting target features and suppressing irrelevant information in complex backgrounds [17].

Furthermore, for the specific task of miner detection in underground coal mines, a weighted Complete Intersection over Union (CIoU) and an adaptive confidence loss function are introduced to improve localization accuracy and enhance model robustness. The adaptive weighting design helps the model focus on uncertain target regions, especially in low-light and occluded conditions.

The main contributions of this paper include the following:

1. A dataset for miner detection in underground coal mines has been innovatively constructed, encompassing complex environments such as low light, partial strong

light interference, and occlusion, which effectively supports the model's detection performance in real mining scenarios.

2. The ECA attention mechanism has been incorporated into YOLOv11, enabling the model to focus more effectively on key features, thereby improving detection accuracy in complex environments.
3. A weighted CIoU and adaptive confidence loss function have been proposed, where the adaptive weighting design enhances the model's attention to uncertain target areas, particularly under low-light and occlusion conditions.
4. A detailed analysis of the detection efficiency and accuracy of the improved model has been conducted, comparing it with various other attention mechanism enhancements, different backbone improvements of YOLOv11, and mainstream object detection models. Additionally, the performance of different algorithms under extreme conditions, such as reduced brightness and noise interference, has been tested. Experimental results show that the proposed method achieves significant performance improvement and strong robustness in miner detection tasks.

This study not only advances the technological development of intelligent safety detection in coal mines but also provides an important reference for the future intelligent management of mining operations, aiming to offer a higher level of safety assurance for coal mine workers.

2. Materials and Methods

2.1. Dataset Introduction

The standardized target detection dataset for the coal mine underground drilling site was gathered between July 2020 and August 2023 in a coal mine located in Binzhou City, Shaanxi Province [18]. An intrinsically safe law enforcement recorder was employed to capture footage of the drilling site. This equipment was capable of acquiring images with a maximum resolution of 4 million pixels, achieving a frame rate of 30 FPS, and the video format used was MPEG-4. The original dataset consists of 976 video clips, with a total duration of 161.8 h. The dataset used in this study was generated by randomly extracting video frames from the original footage at intervals of 30–50 frames using the OpenCV library. After data cleaning, annotation, and expert sampling and verification, the final dataset consists of 10,066 images of coal miners captured from various drilling sites and different environmental backgrounds. During the data cleaning process, we did not specifically address motion blur or low-quality images. This approach was intentional, as we aimed to provide a diverse and authentic dataset for researchers. These images cover three brightness levels (low, medium, and high) and are taken from front, left, and right angles. Additionally, the dataset includes complex scenes, such as target occlusions and blurring. All images have a resolution of 1280 × 720, with object scales ranging from 500 to 55,000 pixels. The majority of the targets have scales exceeding 3000 pixels, reflecting significant variation in the target sizes within the dataset, which meets the requirements for multi-scale object detection.

To enhance the accuracy of the training dataset, it is divided into a training set, validation set, and test set in an 8:1:1 ratio. It is important to ensure that the chosen training, validation, and test sets encompass all possible scenarios. This study employs the LabelImg tool to annotate the image dataset following the VOC format. To ensure high-quality annotations, the following four standards are strictly adhered to during image annotation. (1) Annotations should be made by drawing a box closely around the target's edge. (2) If the target is largely occluded or the overall size cannot be marked, annotations are omitted. (3) Even with minor occlusions, the target must still be annotated, with the obscured part manually added. (4) The annotation box should ideally not be positioned near the image's boundary. Consequently, for the issue of pedestrian detection in coal mines, the experimental results based on this data are more persuasive.

The sample images of the dataset are shown in Figure 1. In general, the images in the dataset are captured under low-light conditions, with significant challenges such as heavy occlusion and intense light interference caused by headlamps and other light sources. These factors contribute to the complexity and difficulty of accurately detecting miners in this specific environment. To further evaluate the algorithm's adaptability in challenging environments, we applied brightness reduction and blur to the images in the testing dataset, simulating real-world variations and assessing the algorithm's generalization capability.

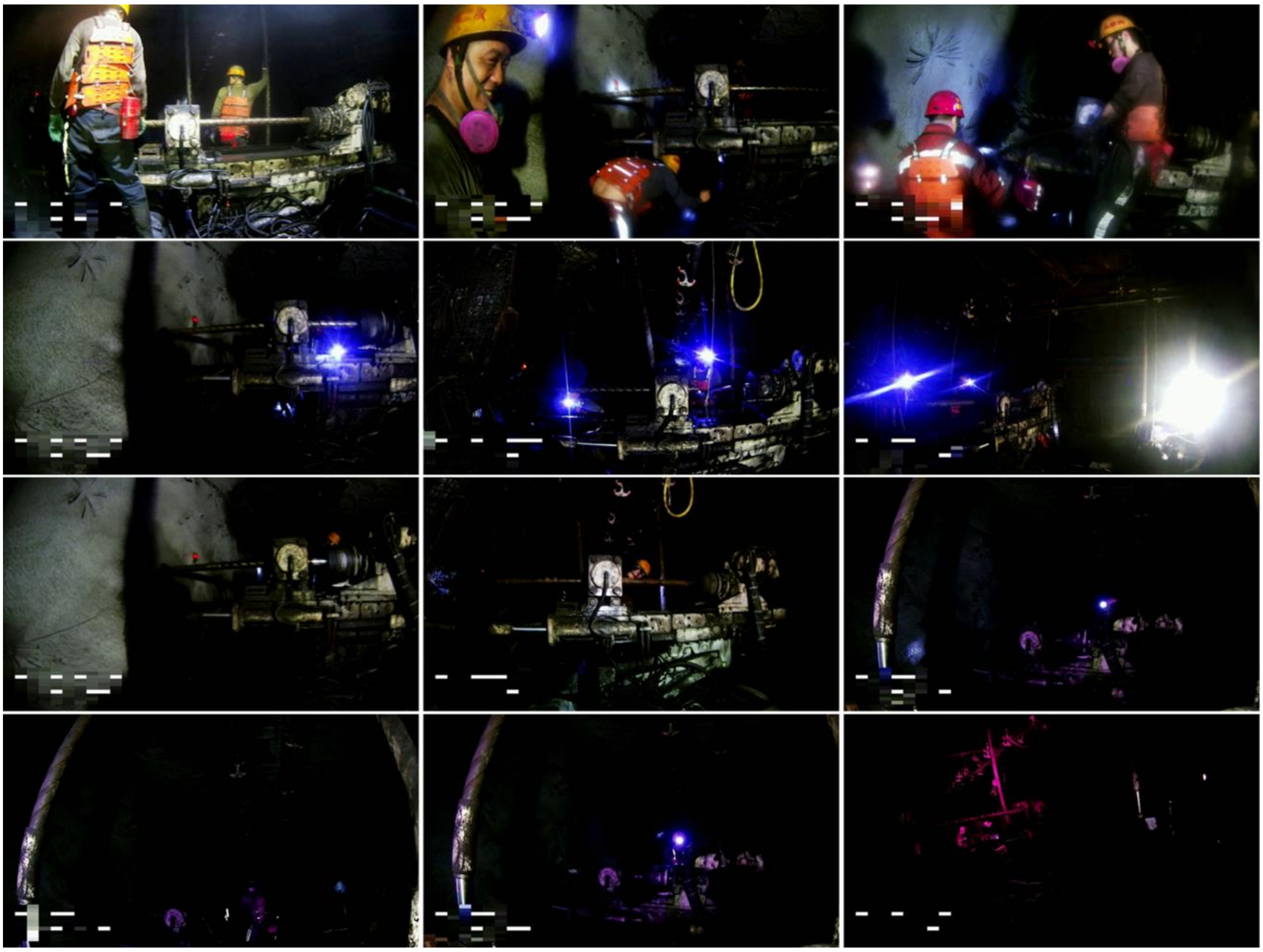


Figure 1. Representative images from the dataset.

2.2. Mine Worker Detection Network Structure

As YOLOv11 achieves structural advancements in the object detection domain, providing both real-time detection efficiency and precision [19], this paper selects it as the baseline network. Its architecture comprises three main modules: the backbone network, the neck network, and the head network, each incorporating innovative design elements to enhance performance. The backbone network adopts a transformer-based structure, effectively capturing long-range dependencies and global contextual information in images, which aids in detecting small targets and partially occluded objects. YOLOv11 introduces the C3k2 (Cross Stage Partial with a kernel size of 2) module, allowing feature partitioning through more efficient convolutional computations, thereby reducing computational load while improving feature representation capability. The neck network enhances detection performance further through multi-scale feature aggregation, utilizing the Spatial Pyramid Pooling Fast (SPPF) module and the Cross Stage Partial with Spatial Attention (C2PSA) module to integrate selective spatial attention mechanisms across different feature scales, enhancing accuracy in feature extraction and key information focus. The head network incorporates a dynamically adaptive detection head, which automatically adjusts computational resource allocation to improve detection efficiency in complex scenes. Moreover, the model introduces a Non-Maximum Suppression (NMS) replacement algorithm and dual label assignment mechanism, increasing accuracy in overlapping and dense target scenarios while reducing inference latency. Overall, the YOLOv11 model achieves higher frame rates and average precision, with efficient feature extraction and detection capabilities suitable for high-performance computing environments and adaptable for edge device deployment, broadening its potential in real-time applications [20].

This paper further proposes an improved network structure called the YOLOv11_ECA network, based on the YOLOv11 model, to further improve the detection precision in complex real-world scenes resulting from occlusion, low illumination, and partial interference caused by strong light. In this respect, it involves the mechanism of ECA above the SPPF layer to strengthen the learning ability in channel attention relations of great importance for maintaining the accuracy of occlusion condition detection. In addition, we propose a new loss function better suited to complex environmental conditions for single-class object detection in mining. This comprises weighted CIoU and an adaptive confidence loss component for better optimization of localization precision and robustness. This loss function mainly pays attention to the emphasis of certain predictions by dynamically adjusting its penalty scale according to the confidence level, thus allowing better results in unfavorable conditions. The use of a customized loss function is foreseen to increase model accuracy and robustness in real-world mining scenarios where object visibility could be compromised by varied lighting and physical obstructions. The improved network structure is shown in Figure 2. The detailed structures and algorithms of each module will be described in the following subsections.

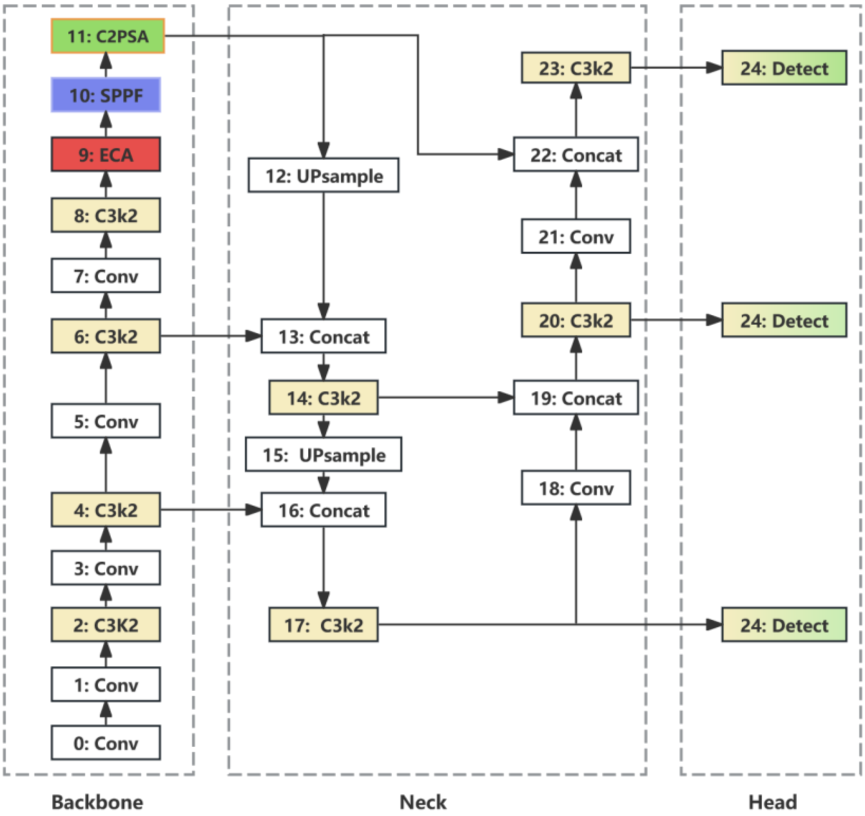


Figure 2. Improved network structure diagram.

2.2.1. ECA Attention Mechanism

The Efficient Channel Attention (ECA) mechanism dynamically adjusts the channel weights of input features, which empowers the network with stronger attention to key features [21]. Compared with the traditional squeeze-and-excitation mechanism, an ECA module avoids reducing the channel dimension and directly learns channel-wise attention through a one-dimensional convolution after global average pooling [22]. In a word, the key of the ECA mechanism is how to adaptively determine the size of the one-dimensional convolution kernel so that the range of local interactions between channels is equal to the dimension of the channel. Through this adaptive mechanism, the way to handle the feature automatically according to the number of feature channels can be learned on ECA; thus, effective feature learning is realized without tedious manual tuning of parameters.

The implementation steps of the ECA module are divided as follows, as illustrated in Figure 3: The ECA module first applies global average pooling (GAP) to the input feature map of dimensions (H × W × C), reducing it to a feature vector of size (1 × 1 × C), thereby capturing the global information for each channel. Subsequently, the adaptive convolution kernel size, computed using Equation (1), is employed to capture inter-channel dependencies. Finally, the Sigmoid function, as defined in Equation (2), is used to compute the activation values of the one-dimensional convolution output, thereby generating the channel-wise attention weights.

$$k = \psi(C) = \frac{\log_2 C + 1}{2} \tag{1}$$

$$\mathrm{Sigmoid}(x) = \frac{1}{e^{-x} + 1} \tag{2}$$

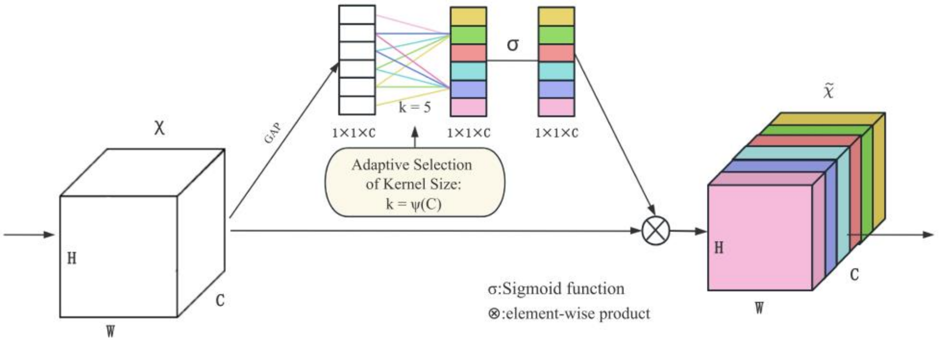


Figure 3. Schematic diagram of ECA module.

In YOLOv11, the ECA attention module was added above the SPPF layer of the back-bone network, enhancing the capture of significant information relevant to object detection. Because the ECA mechanism does not involve any dimensionality reduction, it can maintain more original information about features from the target objects, thus further strengthening the network for its feature representation capability [23]. In addition, one-dimensional convolution-based local interaction also allows the model to pay more attention to those important channel features that relate to object detection and adjust the receptive field adaptively, thus improving accuracy in detection and enhancing recognition capability, especially when faced with a complex background [24]. This lightweight design will greatly improve the network performance without adding too much computation cost or parameter count, hence supporting YOLOv11 in achieving high precision and efficiency.

The specific process for adding the ECA attention mechanism is as follows. First, create an ECA.py file in the *ultralytics/nn* folder and paste the core implementation code into it, then save the file. Next, import this module into the *__init__.py* file. Finally, complete the module import and registration in the *ultralytics/nn/tasks.py* file.

2.2.2. Weighted CIoU with Adaptive Confidence Loss Function (WCIoU-ACLoss)

In single-class object detection tasks of miners, especially under complex environments such as occlusion and low illumination, it is very important to design a loss function to improve the localization accuracy and enhance the robustness of the model. The proposed loss function consists of weighted CIoU [25] and adaptive confidence loss. As for the design rationale, it can be found below. In challenging environments with occlusion, the precision of localization is essential for object bounding boxes. Complete Intersection over Union integrates the distance between the centers of predicted and ground-truth boxes and their aspect ratio consistency into the standard IoU formulation. Additionally, it also incorporates a weighting factor into the complete gain of Intersection over Union to further improve robustness in complex conditions [26,27]. In the single-class detection, the loss weight will be dynamically adjusted with the variation of the predicted confidence score so that low-confidence regions will suffer higher penalties. The adaptive weighting design helps the models to pay more attention to the uncertain object regions, especially those under dark and occlusion circumstances. The formula of the weighted CIoU component of the loss function is given by Equation (3).

$$\mathcal{L}_{wCIoU} = \left(1 - \mathrm{IoU} + \frac{\rho^2\left(\boldsymbol{b_p}, \boldsymbol{b_g}\right)}{c^2} + \alpha v\right) \cdot w_{\text{occlusion}} \tag{3}$$

The parameters in the weighted CIoU formula are defined as follows. IoU represents the Intersection over Union between the predicted bounding box and the ground-truth box, indicating their overlap ratio. $\rho^2\left(\boldsymbol{b_p}, \boldsymbol{b_g}\right)$ is the Euclidean distance between the centers of the predicted ($\boldsymbol{b_p}$) and ground-truth ($\boldsymbol{b_g}$) bounding boxes, while c represents the diagonal length of the smallest enclosing box that contains both of these bounding boxes. Alpha and v are the adjustment factors used to ensure aspect ratio consistency between the predicted and ground-truth boxes. Additionally, an occlusion weight (w_occlusion) is applied to improve robustness under occlusion and low-light conditions. This weight, defined as w_occlusion = exp(−confidence), assigns higher importance to low-confidence situations, thereby enhancing the model's ability to handle challenging scenarios effectively.

The adaptive confidence loss function is provided in Equation (4). The confidence score represents the level of certainty that the model has regarding the predicted bounding box containing the target object. To enhance the model's sensitivity to uncertain predictions, an adaptive weighting factor ($w_{\text{confidence}}$) is incorporated into the loss function. This weighting factor is designed to increase for low-confidence regions and is formulated as $w_{\text{confidence}} = (1-\text{confidence})^{\gamma}$, where γ serves as a hyperparameter that controls the strength of the penalty for low-confidence areas. Typically, γ is set to 2, which effectively emphasizes regions with low confidence, thereby enhancing the model's robustness under challenging conditions and contributing to improved detection performance in adverse environments.

$$\mathcal{L}_{\boldsymbol{adptive\ confidence}} = -\log(\text{confidence}) \cdot w_{confidence} \tag{4}$$

The comprehensive loss function is given in Equation (5). λ is a balancing factor used to control the weight between localization loss and confidence loss. Under severe occlusion conditions, increasing the value of λ can help enhance localization accuracy. It is recommended that λ be initially set to 0.7 to prioritize localization loss, with further fine-tuning during training to optimize the model's overall performance.

$$\mathcal{L} = \lambda \cdot \mathcal{L}_{wCIoU} + (1 - \lambda) \cdot \mathcal{L}_{adptive\ confidence} \tag{5}$$

The steps for adding the loss function are as follows. The first step is to create and implement the weighted CIoU with adaptive confidence loss function code block, which will then be integrated into the model. Specifically, the existing bbox_iou function in the *ultralytics/utils/metrics.py* file will be fully replaced with the new implementation of the loss function. Subsequently, the forward function of the BboxLoss class in the *ultralytics/utils/loss.py* file will be modified to use the newly defined loss function for IoU calculation, replacing the previous IoU calculation method. Finally, the iou_calculation function in the *ultralytics/utils/tal.py* file will be updated to ensure consistency with the parameter settings used in loss.py. These changes will ensure that the new loss function is effectively integrated into the training pipeline and aligned with the existing framework.

3. Experiment and Results

3.1. Experimental Environment

The hardware configuration for this experiment includes an NVIDIA DGX-1 server equipped with Intel Xeon E5-2698 v4 CPU and NVIDIA Tesla V100-SXM2-32GB GPUs, sourced from NVIDIA (Santa Clara, California, USA), providing efficient computational power to support the training and inference of complex deep learning models. The software environment for experiments and testing is based on the Ubuntu 16.04 operating system, combined with CUDA 11.1 and cuDNN 8.0.5 GPU-accelerated libraries to fully utilize the parallel computing capabilities of the GPU. Additionally, PyTorch is used as the deep learning framework, facilitating rapid model development and optimization. This combination of hardware and software ensures experimental efficiency and reproducibility, providing a solid foundation for in-depth research.

3.2. Evaluation Metrics

In object detection, precision measures the proportion of correct detections, with the formula shown in Equation (6). Recall evaluates the model's ability to detect all relevant objects, with the formula shown in Equation (7), where higher values indicate fewer missed detections. Both metrics range from 0 to 1, with higher values indicating better model performance [28–30].

$$Precision = \frac{TP}{TP + FP} \tag{6}$$

$$Recall = \frac{TP}{TP + FN} \tag{7}$$

where TP (true positives) are correctly detected objects, FP (false positives) are incorrect detections of non-targets, and FN (false negatives) are actual objects missed by the model.

Mean Average Precision at 0.5 (mAP@0.5) is a key metric for evaluating the overall accuracy of a model, measuring accuracy across classes by combining localization and classification performance. The IoU (Intersection over Union) formula is shown in Equation (8). A detection is considered a true positive (TP) if its IoU with the ground truth exceeds 0.5; otherwise, it is labeled as a false positive (FP).

$$IoU = \frac{Area\ of\ Overlap}{Area\ of\ Union} \tag{8}$$

Mean Average Precision (mAP) is calculated by averaging the Average Precision (AP) values across all classes [31]. First, the AP for each class is computed by the area under the precision–recall curve. Then, mAP is obtained by averaging these AP values across classes. The formula is:

$$mAP = \frac{1}{N}\sum_{c\,=\,1}^{N} AP_c \tag{9}$$

In this formulation, N is the total number of classes, and *APc* is the Average Precision for class c. For a specific IoU threshold (e.g., 0.5), this is denoted as *mAP*@0.5, where *AP* is calculated for each class at IoU $\geq$ 0.5 [32].

Frames per second (FPS) measures a model's inference speed, indicating how many frames it can process per second. FPS is calculated by dividing the total number of frames processed by the total time taken. Higher FPS values signify faster processing, making the model suitable for real-time applications [33,34], with the formula:

$$FPS = \frac{Total\ Frames\ Processed}{Total\ Time} \tag{10}$$

3.3. Results and Analysis

To validate the effectiveness of the proposed algorithm, we conducted comprehensive experiments using a standardized coal mine underground drilling field miner detection dataset, aiming to evaluate the detection performance of the proposed method for miner detection in coal mine environments. The training process involved 300 epochs, with an initial learning rate of 1×10^{-2}. The Adam optimizer was employed, and the batch size was set to 16.

We conducted four sets of comparative experiments to verify the effectiveness and robustness of the improved algorithm. Specifically, we performed the following experiments: (1) a comparative analysis of the YOLOv11 algorithm improved with different attention mechanisms; (2) a performance comparison of the YOLOv11 algorithm using various backbone networks; (3) an analysis of the algorithm's performance under extreme conditions involving noise and reduced brightness; and (4) a comparison with existing state-of-the-art object detection algorithms. These experiments were designed to comprehensively evaluate the detection performance and robustness of the proposed method in different complex scenarios, thereby further validating the advantages of the improved algorithm.

3.3.1. Comparison with Different Attention Mechanisms

The proposed model demonstrates lower box localization losses during both training and validation phases compared to other YOLOv11 variants. As shown in Figure 4, all models exhibit a rapid decrease in loss at the beginning of training, followed by a stabilization phase, indicating convergence. The lower loss of the proposed model suggests superior generalization performance, particularly in challenging environments with occlusion and low illumination.

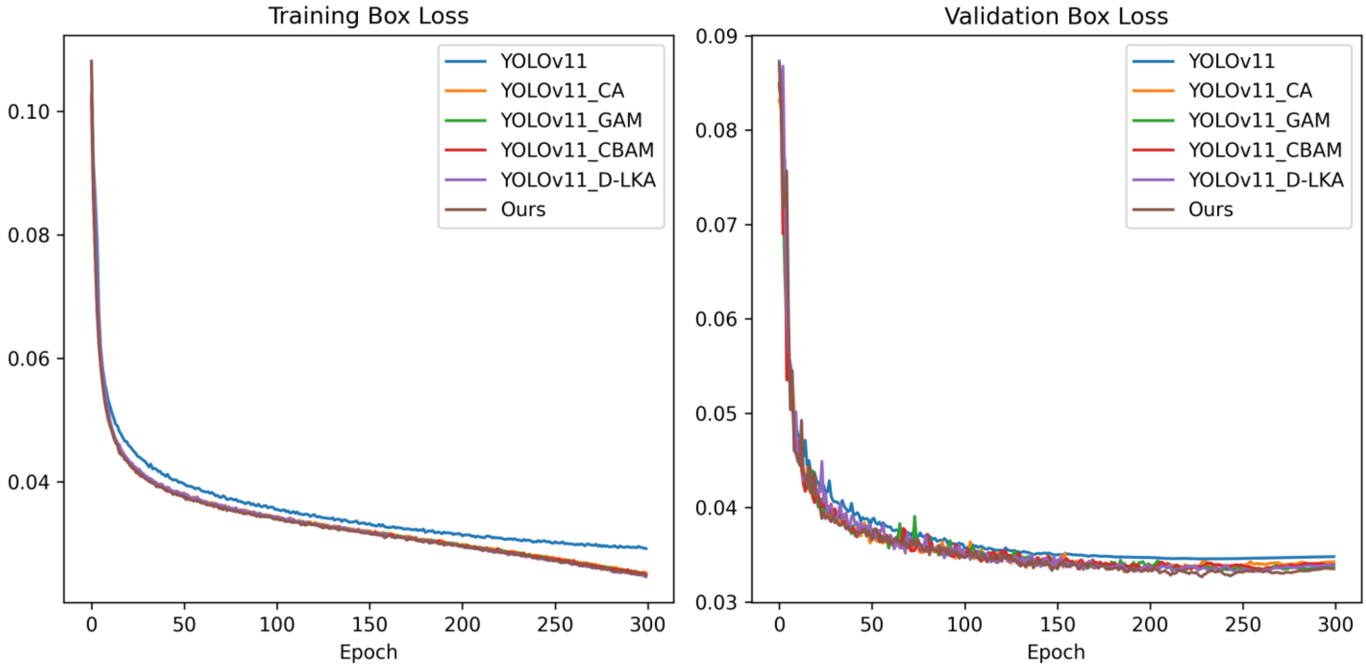


Figure 4. Training and validation loss curves for different attention mechanisms.

Figure 5 presents the evaluation results of different versions of YOLOv11 (e.g., YOLOv11 with CA, GAM, CBAM, and D-LKA attention mechanisms) and the proposed model on the validation set for object detection tasks. Except for YOLOv11, which shows significantly lower mAP within the IoU range of 0.5–0.95 compared to other algorithms, all models exhibit stable and similar performance metrics on the validation set after approximately 50 training epochs. This indicates that these improved algorithms enhance object detection capabilities in complex environments to some extent. However, further comparative analysis on the test set is required to comprehensively evaluate the generalization ability of each model.

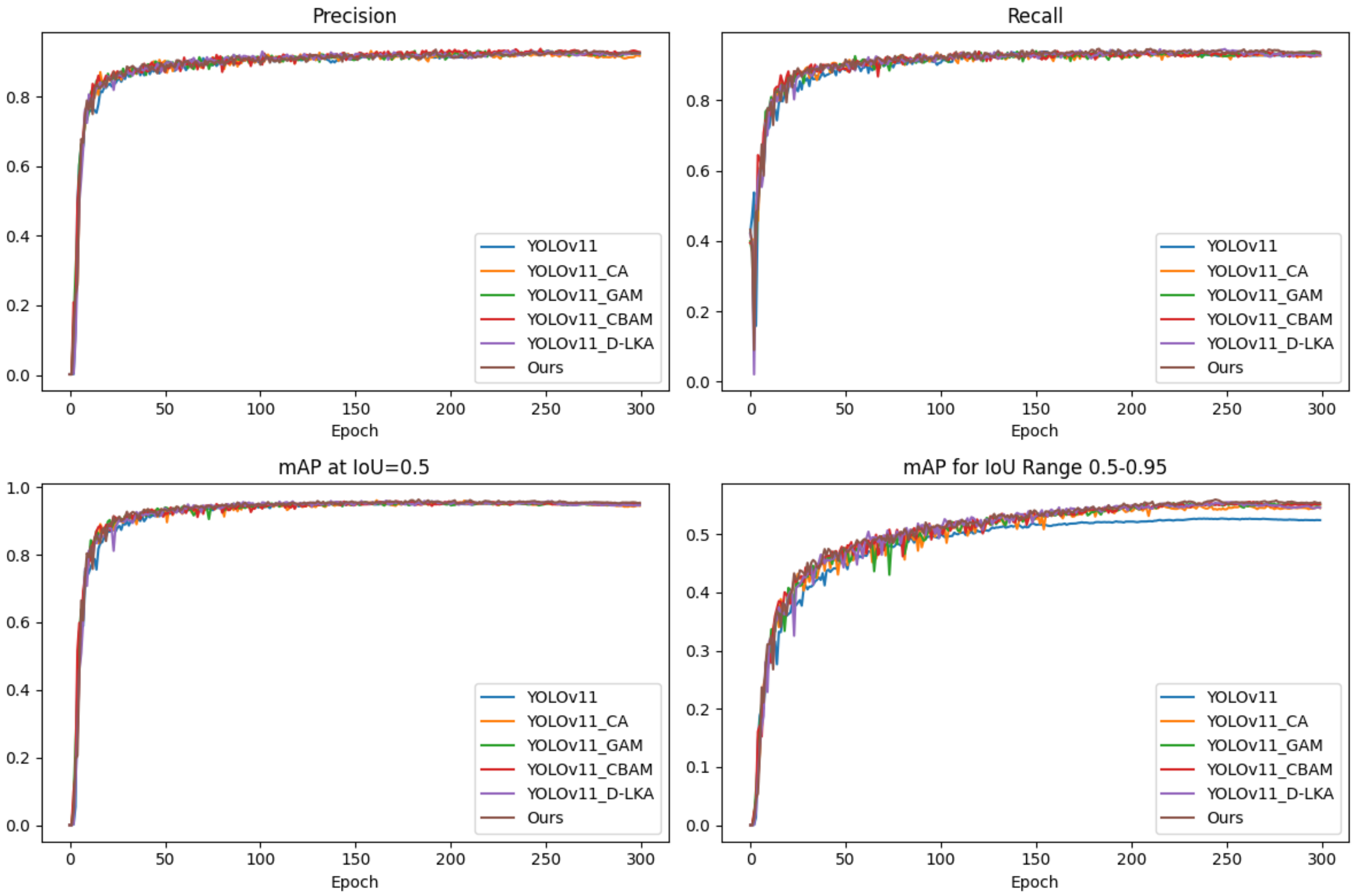


Figure 5. Validation set evaluation metrics for object detection under different attention mechanisms.

Table 1 illustrates the performance of various attention mechanisms on the test set across key metrics, including precision, recall, mAP@50, and FPS. The D-LKA model achieves the highest precision (0.934), closely followed by the proposed model at 0.932, indicating strong accuracy in positive detections. In terms of recall, the proposed model leads with a value of 0.934, reflecting superior sensitivity in identifying true positives. Furthermore, the proposed model outperforms all others in mAP@50, achieving a score of 0.958, which highlights its enhanced detection accuracy and robustness under complex conditions. The proposed model also demonstrates the highest FPS (59.6), significantly surpassing D-LKA's FPS of 23.6, where speed is reduced despite high precision. Overall, the proposed model achieves a balanced improvement in both accuracy and processing speed, making it the most effective among the compared models for complex object detection tasks.

Table 1. Performance comparison of different attention mechanisms on the test set.

Methods	Precision	Recall	mAP50	FPS
YOLOv11	0.914	0.925	0.938	51.6
YOLOv11_CA	0.920	0.925	0.939	51.1
YOLOv11_GAM	0.928	0.931	0.944	49.8
YOLOv11_CBAM	0.924	0.930	0.946	53.3
YOLOv11_D-LKA	0.934	0.927	0.946	23.6
Ours	0.932	0.934	0.958	59.6

3.3.2. Comparison with Different Backbone Architectures

The training and validation box loss curves for various object detection algorithms were analyzed, as depicted in Figure 6. The results indicate that the proposed method achieves the lowest box loss on both the training and validation datasets, demonstrating superior performance in terms of convergence speed and generalization ability. Specifically, the proposed model exhibits a lower final loss compared to other models, suggesting better fitting and reduced overfitting. YOLOv11_RepViT also performed well, but its validation loss was slightly higher than that of the proposed method. In contrast, other models, such as YOLOv11_PE-YOLO and YOLOv11_Retinexformer, exhibited relatively higher validation losses, which may imply limited generalization capabilities. Overall, the proposed method outperforms the other YOLOv11 variants, underscoring its effectiveness for object detection tasks.

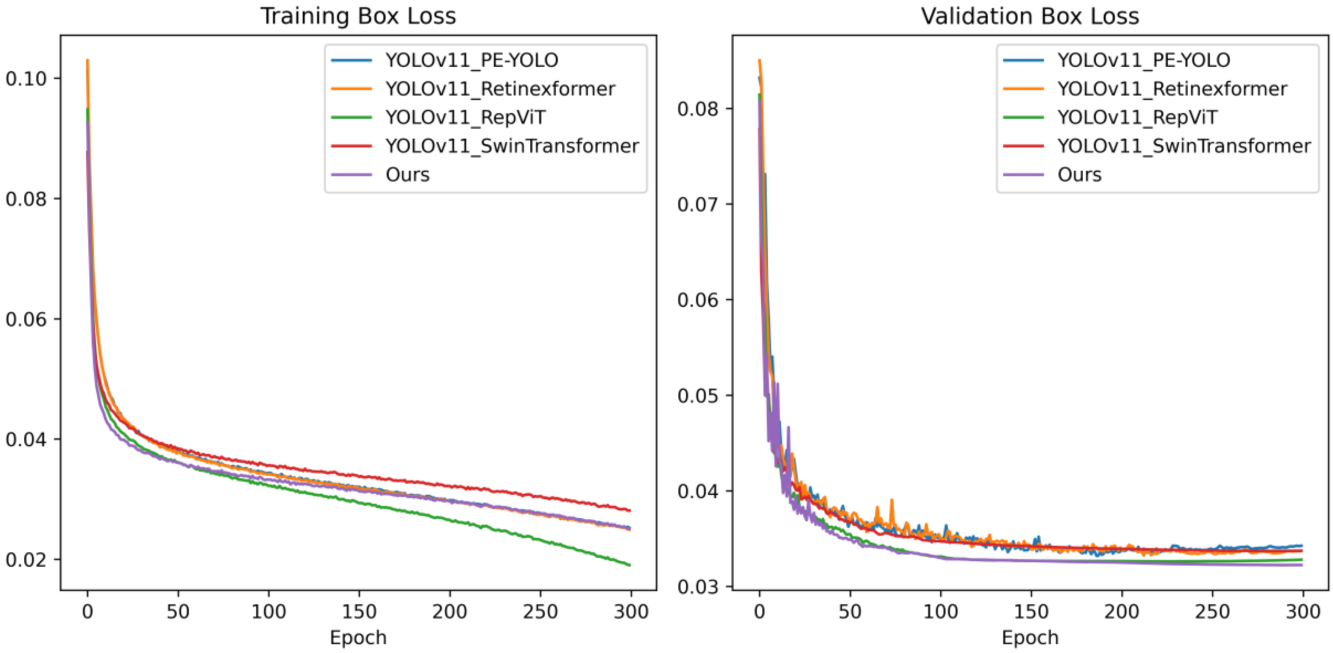


Figure 6. Training and validation loss curves for different backbone networks.

Figure 7 illustrates the convergence trends of various YOLOv11 models (including YOLOv11_PE-YOLO, YOLOv11_Retinexformer, YOLOv11_RepViT, YOLOv11_ SwinTransformer) and the proposed model ("Ours") in terms of precision, recall, mAP@IoU = 0.5, and mAP across the IoU range of 0.5–0.95. The results indicate that all models rapidly improve to high-performance levels in the early stages of training and gradually stabilize. The proposed model demonstrates superior performance across all evaluation metrics, particularly in precision and mAP@IoU 0.5–0.95, suggesting enhanced detection accuracy and robustness. This indicates that the proposed model has stronger generalization capability and localization precision across different IoU overlap conditions. In comparison, the other models show similar performance, with YOLOv11_Retinexformer exhibiting slightly lower precision and recall, possibly due to limitations in the suitability of its backbone. Overall, the proposed model demonstrates superior performance in complex detection tasks.

Table 2 presents a comparison of the test set performance for different backbone architectures, including precision, recall, mAP@0.5, and FPS. The proposed model achieves the highest precision (0.932) and recall (0.934), indicating superior accuracy in identifying true positives. Additionally, "Ours" exhibits the best mAP@0.5 score of 0.958, outperforming all other models, which implies better overall detection accuracy and robustness. In terms of inference speed, the proposed model significantly outperforms other backbones with an FPS of 59.6, which is nearly double that of the next fastest model (YOLOv11_SwinTransformer at 30.4 FPS). These results demonstrate the proposed model's advantage in terms of both detection performance and computational efficiency, making it well-suited for real-time and complex detection tasks.

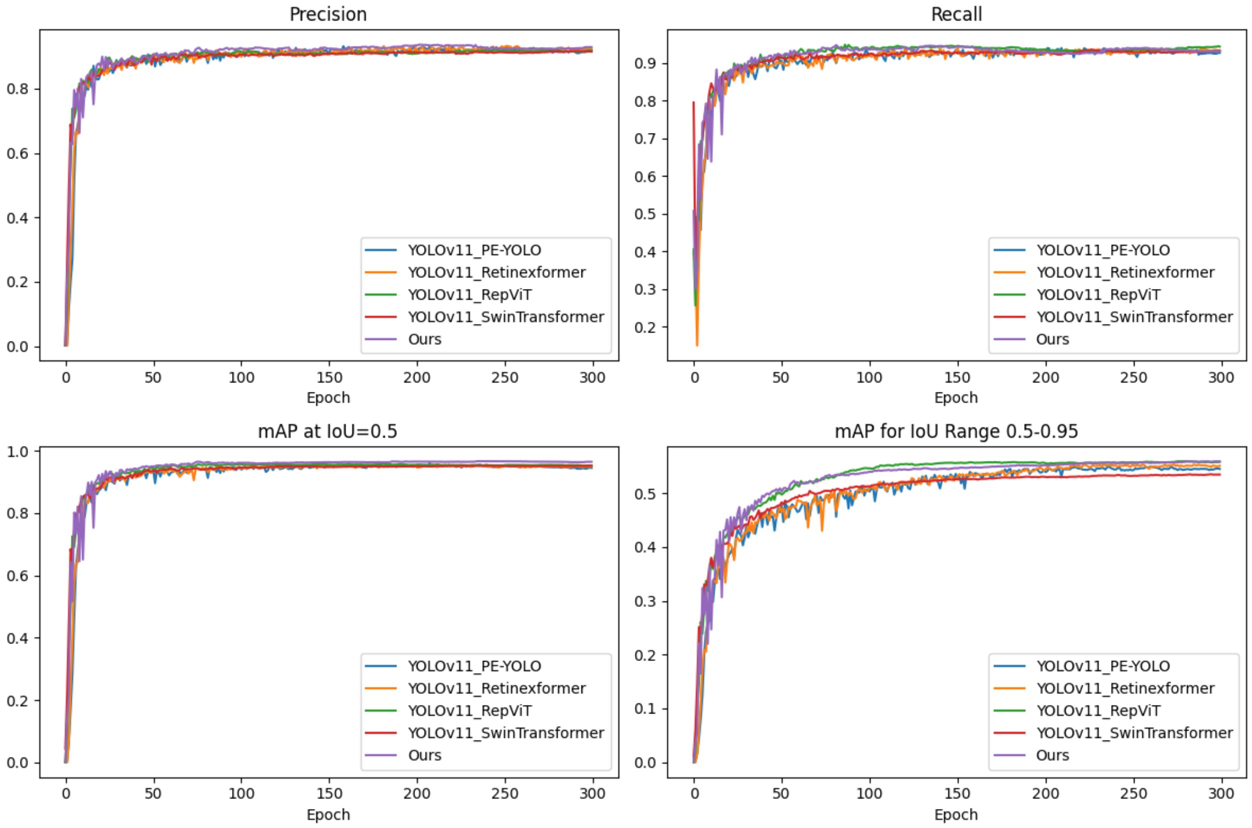


Figure 7. Validation set evaluation metrics for object detection with different backbone architectures.

Table 2. Comparison of test set performance for different backbone architectures.

Methods	Precision	Recall	mAP50	FPS
YOLOv11_PE-YOLO	0.920	0.932	0.946	10.1
YOLOv11_Retinexformer	0.921	0.928	0.946	16.8
YOLOv11_RepViT	0.926	0.933	0.947	24.7
YOLOv11_SwinTransformer	0.919	0.933	0.949	30.4
Ours	0.932	0.934	0.958	59.6

3.3.3. Comparison Under Challenging Visual Conditions

The test images were subjected to reduced brightness, set to half of their original value, and blurring to further validate the robustness and detection performance of the models in complex environments. The blurring operation uses Gaussian blur, and the size of the Gaussian blur kernel is 7 × 7. These enhanced testing conditions allow for an evaluation of the models' performance under challenges such as low illumination and image blurring, providing a more comprehensive understanding of their generalization ability and stability across different image qualities.

The comparison between Tables 1 and 3 reveals the impact of different attention mechanisms on object detection performance under normal and challenging conditions (brightness halved). Table 3 represents performance under halved brightness conditions; precision, recall, and mAP50 values for all methods decreased, reflecting the difficulty of detecting objects under low-light conditions. However, the proposed method still demonstrated outstanding performance, maintaining an mAP50 of 0.958 without any degradation compared to normal conditions, indicating strong robustness to changes in brightness. YOLOv11_CBAM achieved the highest recall (0.918) among other models, although its precision and mAP50 slightly declined compared to normal conditions. YOLOv11_D-LKA maintained a high precision (0.934) despite a drop in recall to 0.904, suggesting that its accuracy remained strong, but its sensitivity to detecting all instances decreased under lower brightness. Overall, the comparison results indicate that the proposed method is not only effective under normal conditions but also demonstrates high robustness in challenging lighting environments, maintaining higher detection accuracy compared to other attention mechanisms.

Table 3. Performance comparison of object detection under halved brightness with different attention mechanisms.

Methods	Precision	Recall	mAP50
YOLOv11	0.918	0.902	0.925
YOLOv11_CA	0.924	0.905	0.928
YOLOv11_GAM	0.926	0.910	0.932
YOLOv11_CBAM	0.925	0.918	0.937
YOLOv11_D-LKA	0.934	0.904	0.933
Ours	0.929	0.930	0.958

The comparison between Tables 2 and 4 highlights the impact of different backbone architectures on object detection performance under normal and challenging conditions (brightness halved). In Table 4, which presents the performance of different backbone architectures under halved brightness conditions, it can be observed that most methods experienced a decrease in both recall and precision, reflecting the challenges posed by low-light conditions. However, "our model" maintained the highest mAP50 (0.958), indicating its strong robustness to changes in brightness. Compared to Table 2, the precision of "our model" slightly decreased from 0.932 to 0.929, but it maintained a high recall (0.930), suggesting that it is less affected by low-light conditions compared to other models. In contrast, YOLOv11_PE-YOLO showed a slight increase in precision (0.931) but a small decrease in recall (0.929), while YOLOv11_Retinexformer, YOLOv11_RepViT, and YOLOv11_SwinTransformer all exhibited declines in both precision and recall, indicating their sensitivity to reduced brightness. Overall, the results from these two tables indicate that "our model" not only provides superior accuracy and efficiency under normal conditions but also maintains its robustness under challenging lighting conditions, outperforming other backbone architectures in low-light scenarios.

Table 4. Comparison of object detection metrics under halved brightness for different backbone architectures.

Methods	Precision	Recall	mAP_{50}
YOLOv11_PE-YOLO	0.931	0.929	0.947
YOLOv11_Retinexformer	0.923	0.911	0.932
YOLOv11_RepViT	0.930	0.914	0.934
YOLOv11_SwinTransformer	0.930	0.915	0.938
Ours	0.929	0.930	0.958

The comparison between Tables 1 and 5 highlights the impact of different attention mechanisms on object detection performance under both normal and challenging conditions, specifically blurred images. In Table 5, which presents the performance of various attention mechanisms under blurred conditions, it is evident that most methods experienced a decrease in recall, reflecting the difficulty posed by image blurring. The proposed model demonstrated the highest precision (0.936) and mAP50 (0.906), underscoring its robustness in the presence of blurred conditions. Compared to Table 1, the recall of the proposed model slightly decreased from 0.934 to 0.856, yet it still outperformed other models in this metric. In contrast, YOLOv11_D-LKA showed a slight improvement in precision (0.936) under blurred conditions but experienced a decrease in recall to 0.846. YOLOv11_GAM and YOLOv11_CBAM exhibited significant declines in recall, indicating their increased sensitivity to image blurring. Overall, the findings from Tables 1 and 5 indicate that the proposed model not only provides superior accuracy and efficiency under normal conditions but also maintains robustness in challenging conditions, such as blurred images, outperforming other attention mechanisms.

Table 5. Performance comparison of object detection under blurred conditions with different attention mechanisms.

Methods	Precision	Recall	mAP$_{50}$
YOLOv11	0.916	0.847	0.895
YOLOv11_CA	0.927	0.846	0.900
YOLOv11_GAM	0.931	0.816	0.885
YOLOv11_CBAM	0.921	0.829	0.885
YOLOv11_D-LKA	0.936	0.846	0.902
Ours	0.936	0.856	0.906

The comparison between Tables 2 and 6 highlights the impact of different backbone networks on object detection performance under both normal and challenging conditions (blurred images). In Table 6, which presents the performance of different backbone networks under blurred conditions, it can be observed that most models experienced a decline in recall and mAP50, reflecting the challenges posed by blurred images. However, the proposed model maintained the highest precision (0.936), recall (0.856), and mAP50 (0.906) under these conditions, demonstrating its robustness to blurred images. Compared to Table 2, the recall of the proposed model slightly decreased from 0.934 to 0.856, but it still outperformed all other models in this metric. YOLOv11_RepViT achieved the highest precision (0.944) under blurred conditions, but its recall (0.836) and mAP50 (0.896) were lower than those of the proposed model. YOLOv11_Retinexformer also performed well, with a recall of 0.841 and an mAP50 of 0.894, but did not surpass the proposed model. Overall, the comparison between these two tables indicates that the proposed model not only provides superior accuracy and efficiency under normal conditions but also maintains robustness under challenging conditions, such as blurred images, outperforming other backbone networks. This demonstrates the adaptability and effectiveness of the proposed model under different conditions.

Table 6. Performance comparison of object detection under blurred conditions with different backbone networks.

Methods	Precision	Recall	mAP$_{50}$
YOLOv11_PE-YOLO	0.911	0.704	0.816
YOLOv11_Retinexformer	0.912	0.841	0.894
YOLOv11_RepViT	0.944	0.836	0.896
YOLOv11_SwinTransformer	0.916	0.809	0.878
Ours	0.936	0.856	0.906

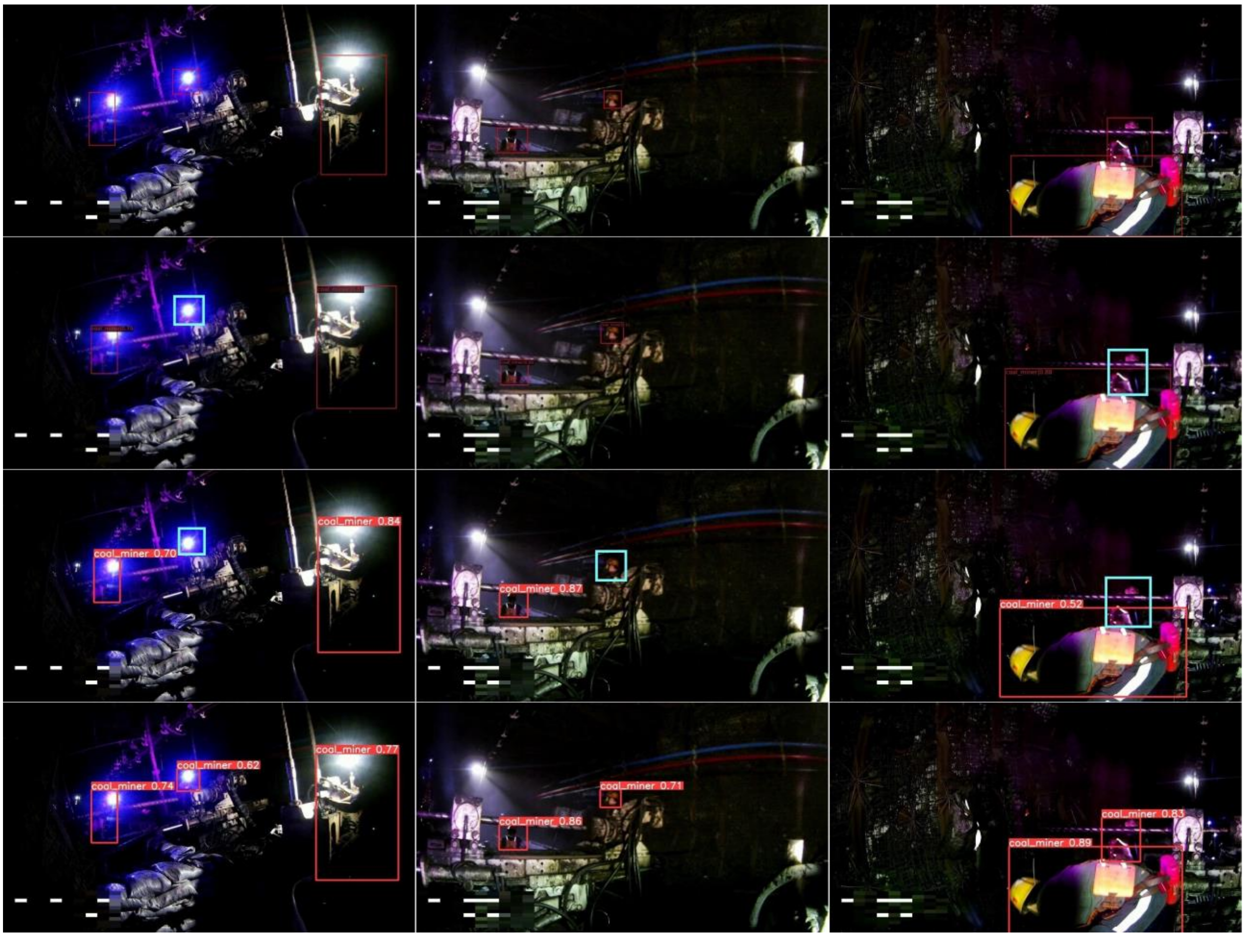
3.3.4. Comparison with State-of-Art Methods

The comparison in Table 7 highlights the performance differences between the proposed model and state-of-the-art object detection methods. The proposed model achieved the highest mAP50 (0.958), indicating superior detection accuracy compared to other methods. Although YOLOv8 had a slightly lower mAP50 (0.954), it achieved the highest frames per second (FPS) value of 62.5, demonstrating better real-time performance. Nevertheless, the FPS of the proposed model (59.6) was also very competitive, indicating a good balance between accuracy and efficiency. SSD and YOLOX performed well in terms of recall, with YOLOX achieving the highest recall (0.949). However, both models had lower precision and mAP50 values compared to the proposed model, suggesting that while these models excel in detecting more objects, their accuracy in correctly classifying detections is slightly inferior. Faster RCNN exhibited relatively low FPS (18.6), highlighting its limitations in processing speed, although its precision (0.886) and recall (0.936) were relatively competitive. YOLOv8 showed a balanced overall performance, with a precision of 0.927 and a recall of 0.929, reflecting a balance between detection accuracy and speed, but it still fell short of the proposed model in terms of mAP50. Overall, the proposed model outperformed state-of-the-art methods in terms of accuracy (mAP50) while maintaining competitive FPS, demonstrating an effective balance between high detection accuracy and efficiency. This balance makes the proposed model highly suitable for practical applications that require both high detection accuracy and fast processing speed.

Table 7. Comparison with state-of-the-art object detection methods.

Methods	Precision	Recall	mAP_{50}	FPS
SSD	0.916	0.912	0.920	56.5
YOLOX	0.897	0.949	0.947	29.8
Faster RCNN	0.886	0.936	0.920	18.6
YOLOv8	0.927	0.929	0.954	62.5
Ours	0.932	0.934	0.958	59.6

To further validate the superiority of "Ours", Figure 8 visualizes the detection results under challenging conditions such as low illumination, strong light interference, and severe occlusion. The blue bounding boxes are used to mark coal mine workers who were not detected by the algorithm, highlighting instances of missed detection. As shown in the figure, the proposed algorithm achieves the best detection results across various scenarios, with no false positives or missed detections. Figure 8a illustrates the detection performance under a complex situation involving simultaneous low illumination, strong light, and occlusion, where both YOLOX and YOLOv8 exhibit missed detections. Figure 8b,c present the detection results under low illumination and severe occlusion, where YOLOv8 misses detections in both Figure 8b,c, while YOLOX fails in Figure 8c. These results highlight the challenging nature of underground coal mine environments, which are often characterized by low illumination, headlamp interference, and occlusions from equipment and personnel, all of which can degrade the feature information of miners, resulting in missed detections by these algorithms.



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

Figure 8. Visual comparison of different object detection methods, comparison results of three different methods on YOLOX, YOLOv8 and ours stands for ground truth. The first row presents the ground truth image, while the subsequent rows (second, third, and fourth) illustrate the results obtained from the YOLOX, YOLOv8, and our proposed algorithm, respectively. (**a**) Images under strong light interference, (**b**) images with equipment occlusion, and (**c**) images with human occlusion.

3.3.5. Ablation Experiment

Table 8 shows the performance of different improved versions of the YOLOv11 model on the test set, focusing on the addition of the ECA attention mechanism and the weighted CIoU with adaptive confidence loss (WCIoU-ACLoss). The ablation study results indicate that both the ECA attention mechanism and WCIoU-ACLoss significantly enhance YOLOv11's performance. The ECA mechanism improves precision and mAP50, while the WCIoU-ACLoss function boosts recall and mAP50, although there is a slight decrease in precision. When both are combined, the model achieves the best performance in terms of precision, recall, and mAP50, demonstrating the effectiveness of these two improvements in enhancing detection accuracy and model robustness.

In summary, the proposed algorithm has demonstrated good effectiveness in object detection, evidenced by objective metrics and intuitive visual representations of the detection results. This has been made possible through the introduction of the ECA mechanism, in which model training can automatically balance channel importance and focus more on key features. This leads to significantly improved target feature identification under extreme conditions, such as low light and occlusion, due to enhanced modeling. Additionally, the introduction of a weighted Complete Intersection over Union (CIoU) with an adaptive confidence loss function equips the model to deal with uncertain target areas, thus making the approach more robust under challenging scenarios. These solve the model to flexibly address various problems in detection, so the overall performance of detection is tremendously improved, especially for complex and dynamic scenarios.

Table 8. Ablation Study.

Methods	Precision	Recall	$\mathbf{mAP_{50}}$
YOLOv11	0.914	0.925	0.938
YOLOv11_ECA	0.930	0.922	0.941
YOLOv11_WCIoU-ACLoss	0.915	0.934	0.954
YOLOv11_ECA_ WCIoU-ACLoss (Ours)	0.932	0.934	0.958

4. Conclusions

This study presents an innovative construction of a dataset for miner detection in underground coal mines, capturing complex environments including low-light conditions, partial strong light interference, and occlusion. The Efficient Channel Attention (ECA) mechanism was incorporated into YOLOv11, enhancing the model's ability to focus on salient features and thereby substantially improving detection accuracy under challenging conditions. Additionally, a weighted Complete Intersection over Union (CIoU) and an adaptive confidence loss function were introduced, which enhanced the model's focus on uncertain target regions, particularly in low-light and occluded environments, thereby improving its robustness. Comprehensive experimental comparisons with multiple improved algorithms and state-of-the-art detection models demonstrated that the proposed method achieved significant improvements in both detection performance and robustness in miner detection tasks, providing critical technical support and reference for ensuring coal miner safety and advancing intelligent mine management. In the future, the model will be extended to detect additional targets in underground coal mines, such as grippers, drill chucks, safety helmets, drill rods, etc., with the goal of enabling real-time identification of potential hazards and issuing alerts. This will significantly enhance the practical applicability and value of the model.

Author Contributions: Funding acquisition, D.L.; Methodology, Y.L.; Project administration, Y.L. and H.W.; Software, H.Y. and H.W.; Supervision, D.L.; Validation, Y.L. and H.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research is funded by the Natural Science Foundation of the Jiangsu Higher Education Institutions of China (Grant: 23KJA520013) and the Xuzhou Science and Technology Plan Project (Grant: KC22305).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data used in this article are publicly available. The source code supporting the findings of this study is openly available at GitHub via https://github.com/whdcumt/CoalMineWorker, under the MIT license, accessed on 8 December 2024.

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

Abbreviations

The following abbreviations are used in this manuscript:

YOLO	You Only Look Once
ECA	Efficient Channel Attention
CA	Channel Attention
GAM	Global Attention Mechanism
CBAM	Convolutional Block Attention Module

D-LKA	Dynamic Local Key Attention
PE-YOLO	Position-Enhanced YOLO
RepViT	Reparameterized Vision Transformer
CIoU	Complete Intersection over Union
IoU	Intersection over Union
SPPF	Spatial Pyramid Pooling Fast
FPS	Frames Per Second
C2PSA	Cross Stage Partial with Spatial Attention

References

1. Tian, S.; Wang, Y.; Hongxia, L.I.; Ma, T.; Mao, J.; Ma, L. Analysis of the causes and safety countermeasures of coal mine accidents: A case study of coal mine accidents in China from 2018 to 2022. *Process Saf. Environ. Prot.* **2024**, *187*, 864–875. [CrossRef]
2. Sun, Y.; Ma, Z.; Zhang, H. Research on target detection in underground coal mines based on improved YOLOv5. *J. Electron. Inform. Technol.* **2023**, *41*, 827–835.
3. Zhang, J.; Chen, L.; Tang, W. Deep learning algorithms for object detection in low-visibility environments: A case study in coal mines. *J. Min. Sci.* **2020**, *56*, 776–785.
4. Liu, Y.; Gao, P.; Chen, Z. Intelligent emergency response system in coal mines using deep learning and IoT technologies. *J. Loss Prev. Process Ind.* **2022**, *75*, 104683.
5. Yang, X.; Huang, S.; Feng, Y. AI-Powered Personnel Tracking for Emergency Response in Coal Mines. *Saf. Sci.* **2024**, *157*, 105948.
6. Gao, Y.; Liu, H.; Qian, F. Real-time deep learning-based personnel detection in coal mines with improved YOLOv4 model. *Eng. Geol.* **2021**, *288*, 106225.
7. Wang, J.; Li, X.; Xu, D. Enhancing safety in coal mines with automated personnel detection using convolutional neural networks. *IEEE Access* **2020**, *8*, 93472–93481.
8. Cheng, H.; Qian, F.; Liu, T. A real-time surveillance system for underground coal mines using deep learning object detection. *Eng. Geol.* **2021**, *276*, 105745.
9. Li, F.; Zhou, Y.; Sun, M. Enhanced YOLO-based detection for personnel identification in coal mines with adverse lighting conditions. *Saf. Sci.* **2022**, *146*, 105627.
10. Chen, Y.; Liu, X.; Li, Z. Personnel positioning and tracking in coal mines using RFID and deep learning. *Saf. Sci.* **2022**, *146*, 105572.
11. Xu, X.; Chen, Z.; Ma, J. Enhancing the generalization capability of object detection models in underground coal mining environments. *Int. J. Min. Sci. Technol.* **2022**, *32*, 455–464.
12. Xu, T.; Zhou, J.; Fang, H. Real-time object detection and recognition for safety monitoring in coal mines based on deep learning. *IEEE Trans. Ind. Inform.* **2023**, *19*, 1804–1812.
13. Zhang, L.; Liu, W.; Chen, Y. Addressing occlusion and dust interference in deep learning-based detection models for coal mines. *J. Min. Sci.* **2023**, *59*, 123–134.
14. Li, J.; Wang, H.; Zhang, Y.; Chen, F. A Real-Time Personnel Detection System in Underground Coal Mines Using Deep Learning. *Int. J. Min. Sci. Technol.* **2024**, *64*, 301–309.
15. Zhao, L.; Liu, Q.; Sun, Z.; Xu, K. Infrared and Thermal Imaging Combined System for Coal Mine Safety Monitoring and Personnel Detection. *IEEE Trans. Ind. Electron.* **2024**, *71*, 4110–4119.
16. Li, X.; Zhang, Y.; Huang, W. YOLOv11-based improved object detection model for autonomous driving in urban environments. *IEEE Trans. Veh. Technol.* **2023**, *72*, 4012–4025.
17. Wang, Q.; Wu, B.; Zhu, P.; Li, P.; Zuo, W.; Hu, Q. ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 14–19 June 2020; pp. 11534–11542.
18. Zhou, W.; Dong, L.; Ye, O.; She, X.; Duan, X.; Peng, Z.; Wang, S.; Zhao, N.; Guo, X. A dataset of drilling site object detection in underground coal mines. *China Sci. Data* **2024**, *9*, 1–10. [CrossRef]
19. Boesch, G. Advances in YOLOv11 for real-time object detection and tracking in industrial applications. *IEEE Trans. Pattern Anal. Mach. Intell.* **2024**, *46*, 220–233.
20. Li, X.; Chen, Y.; Sun, J. YOLOv11-based pose estimation for sports and fitness applications. *Pattern Recognit. Lett.* **2023**, *165*, 120–132.
21. Wang, Y.; Xu, L.; Zhang, H. An Improved ECA-Net for Real-Time Object Detection in Embedded Systems. *IEEE Trans. Circuits Syst. Video Technol.* **2021**, *31*, 3445–3457.
22. Wu, Q.; Lin, F. A Lightweight Network for Human Pose Estimation Using Attention Mechanisms. *Electronics* **2022**, *11*, 1187.
23. Zhou, X.; Wang, D.; Zhu, J. YOLOv11: Enhancements in Real-Time Object Detection with Improved Feature Representation. *J. Comput. Vis.* **2024**, *112*, 458–472.
24. Lin, T.Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature Pyramid Networks for Object Detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 936–944.
25. Zheng, Z.; Wang, P.; Liu, W. A Comprehensive Review on Loss Functions in Object Detection. *Electronics* **2021**, *10*, 2644.
26. Mao, Y.; Wang, T.; Guo, Y. Adaptive Confidence Loss for Object Detection with Uncertainty Estimation. *IEEE Trans. Neural Netw. Learn. Syst.* **2022**, *33*, 1564–1575.

27. Tian, Y.; Shen, C.; Chen, X. Focal Loss for Dense Object Detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *42*, 2986–2998.
28. Kong, Y.; Shang, X.; Jia, S. Drone-DETR: Efficient Small Object Detection for Remote Sensing Image Using Enhanced RT-DETR Model. *Sensors* **2024**, *24*, 5496. [CrossRef]
29. Bui, T.; Liu, J.; Cao, J. Elderly fall detection in complex environment based on improved YOLOv5s and LSTM. *Appl. Sci.* **2024**, *14*, 9028. [CrossRef]
30. Wang, C.; Wang, Y. A Real-Time Object Detection System Based on YOLOv5 for Smart Surveillance. *Sensors* **2022**, *22*, 1234.
31. Li, Y.; Wang, Z.; Sun, H. Object Detection in Large-Scale Remote Sensing Images with a Distributed Deep Learning Framework. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 4587–4599. [CrossRef]
32. Li, J.; Zhang, R.; Chen, Y. Real-Time Object Detection in Aerial Images Using YOLO with Attention Mechanism. *Pattern Recognit.* **2023**, *135*, 109–120.
33. Wang, X.; Liu, Y. A Review of Deep Learning Approaches for Object Detection in Remote Sensing Images. *Artif. Intell. Rev.* **2023**, *56*, 117–135.
34. Tang, L.; Li, T.; Xu, C. Stratigraphic division method based on the improved YOLOv8. *Appl. Sci.* **2024**, *14*, 9485. [CrossRef]

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