Scientific Reports

https://doi.org/10.1038/s41598-026-38676-4

Article in Press

Lightweight target detection and multi target tracking for UAV inspection in open pit mines

Received: 22 October 2025

Accepted: 30 January 2026

Published online: 10 February 2026

Cite this article as: Liu G., Zhang L., Lei J. *et al.* Lightweight target detection and multi target tracking for UAV inspection in open pit mines. *Sci Rep* (2026). https://doi.org/10.1038/s41598-026-38676-4

Guangwei Liu, Linbo Zhang, Jian Lei, Senlin Chai & Weijun Zhu

Lightweight target detection and multi target tracking for UAV inspection in open pit mines

LIU Guangwei[1]，ZHANG Linbo[1]，LEI Jian[1]，CHAI Senlin[1]，ZHU Weijun[2]

1.School of Mining, Liaoning Technical University, Fuxin, Liaoning,123000,China ;

2. School of Economics and Management Shenyang Institute of Technology, Shenyang, Liaoning 110000,China ;

*. Correspondence: 13043801046@163.com

Abstract: Aiming at the problems of low efficiency in traditional manual inspection of open-pit mines, difficulty in identifying faulty equipment and non-cooperative targets, high safety risks, and insufficient detection accuracy for small targets, while supplementing the limitations of active positioning technologies such as UWB indoor-outdoor positioning and vehicle-mounted strapdown inertial navigation in scenarios like signal blind areas and non-cooperative target monitoring, this paper proposes a lightweight object detection and multi-target tracking algorithm, and constructs an intelligent UAV inspection system. In the design of the detection model, deformable convolution DCNv2 is introduced into the backbone network, and the progressive feature pyramid network AFPN is adopted in the neck part to enhance the multi-scale feature extraction capability. A lightweight detection head (LSDECD-Head) is designed, and combined with the Focaler-GIoU loss function, the detection accuracy of small targets and occluded targets is improved. The LAMP pruning algorithm is used to compress the model, and under a 30% pruning rate, the model still maintains a performance of mAP50 at 0.868 and inference time of 196 ms, which is suitable for the computing resource constraints of UAVs. In terms of multi-target tracking, the ByteTrack algorithm is improved. A space-appearance similarity matrix (ASM) that integrates the target's spatial position, operation status, and appearance features is introduced, and combined with an acceleration correction function to optimize trajectory prediction. This improvement increases the multi-target tracking accuracy (MOTA) by 2.6% and reduces the number of ID switches by 21. In addition, a multi-level inspection system is constructed, which integrates functions of data collection, real-time detection, and multi-UAV collaborative scheduling. It realizes data transmission and remote monitoring relying on 5G and ad-hoc network technologies. The core innovation of this paper lies in constructing the C2f-DCN+AFPN lightweight feature extraction architecture, tailored to capture complex target features in mining areas. Designing the LSDECD-Head detection head and Focaler-GIoU loss function to enhance difficult sample detection. Proposing a Hierarchical Adaptive LAMP Pruning Strategy to Balance Accuracy and Lightweighting. Enhanced ByteTrack algorithm incorporates ASM matrix and acceleration correction to improve dynamic tracking stability: Establishing an air-ground collaborative inspection system to achieve technological implementation. The aforementioned innovations are not merely a simple combination of existing technologies, but rather a deeply integrated optimization addressing the pain points specific to open-pit mining scenarios. Experimental results show that this scheme significantly improves the accuracy and stability of equipment detection and tracking in open-pit mine scenarios, and provides a feasible technical solution for intelligent and unmanned inspection of mines.

Keywords: Open-pit coal mine; Drone; Target detection; Lightweight model; multi-target tracking

1. Introduction

As an essential site for mineral resource extraction, open-pit mines play a critical role in ensuring safe production and efficient operations, which are vital to national economic development. At present, active positioning technologies—such as satellite-based systems (BDS/GPS), ultra-wideband (UWB) indoor–outdoor positioning, and vehicle-mounted strapdown inertial navigation—have been widely applied in mining engineering. These technologies demonstrate remarkable advantages in resisting interference from complex environments, directly providing three-dimensional coordinates, and enabling continuous, full-time, and full-area trajectory tracking. Moreover, their positioning accuracy and stability have been industrially validated, offering strong technical support for the operational scheduling and safety monitoring of mining equipment.

However, in practical applications, active positioning technologies also exhibit certain limitations. For instance, in some areas of open-pit mines, satellite positioning signals are easily interfered with or even lost due to terrain occlusion, resulting in positioning failure. UWB positioning requires the deployment of numerous base stations, which leads to high costs and considerable difficulty in large-scale or complex open-pit environments, and its signals are often affected by dust and electromagnetic interference. Although vehicle-mounted strapdown inertial navigation can maintain high accuracy over short periods, it suffers from cumulative drift errors during long-term operation and requires periodic calibration.

Unmanned aerial vehicles (UAVs), by contrast, have emerged as an important technology for open-pit mine inspection owing to their flexibility, efficiency, ability to rapidly cover large areas, and capacity to capture data from multiple perspectives. Equipped with high-resolution cameras and various sensors, UAVs can collect real-time imagery and video data of mining areas, providing abundant information for equipment condition monitoring and potential safety hazard detection.

To address the aforementioned limitations of active positioning technologies, this study proposes a lightweight object detection method based on an improved YOLOv8 model and a multi-object tracking approach based on an enhanced ByteTrack algorithm, focusing on specific application scenarios in open-pit mines to serve as an effective complementary solution.

(1) Temporary fault detection scenarios. When mining equipment experiences temporary malfunctions, active positioning technologies can only provide positional information without identifying the specific nature or location of the fault. In contrast, a UAV-based visual inspection system can perform high-precision object detection to rapidly locate the faulty component and determine the fault type, thereby providing maintenance personnel with accurate diagnostic information and enhancing maintenance efficiency.

(2) Non-cooperative target identification scenarios. During open-pit mining operations, unauthorized vehicles or personnel—classified as non-cooperative targets—may occasionally enter the mining area. Since these targets are typically not equipped with active positioning devices, traditional positioning technologies are incapable of tracking or monitoring them. UAV-based visual detection and tracking methods can achieve real-time identification and trajectory monitoring of such non-cooperative targets, issuing timely alerts to ensure operational safety within the mine.

(3) Monitoring blind zones in active positioning systems. In areas where satellite positioning signals are lost or UWB signals are subject to interference, UAVs can function as mobile monitoring platforms. Through visual detection and tracking, they can acquire real-time information on the operational status and spatial position of mining equipment within these zones, effectively compensating for the blind spots

of active positioning technologies and ensuring uninterrupted surveillance of mining operations.

However, in practical applications, UAV-based inspection faces numerous technical challenges in object detection and tracking. Existing object detection models are generally large and demand substantial hardware resources. Due to the limited payload capacity and computational power of UAVs, deploying such large-scale models directly would significantly reduce endurance and inspection efficiency. Meanwhile, in open-pit mining environments, small targets—such as minor equipment components or subtle geological structure changes—occupy a large proportion of the scene. These targets are characterized by a small pixel ratio and weak feature representation in images, leading to insufficient detection accuracy in existing models, frequent missed or false detections, and the inability to promptly identify potential safety hazards.

In addition, frequent occlusions caused by large stockpiles, buildings, and other mining structures make it difficult for detection models to capture complete object features, thereby increasing detection difficulty and potentially missing critical targets, which poses safety risks to mining operations. In terms of object tracking, the motion patterns of mining equipment and personnel in open-pit scenarios are complex and dynamic. High visual similarity among equipment, together with severe occlusions and illumination variations, causes traditional tracking algorithms to suffer from frequent ID switching and trajectory drift, preventing stable and accurate multi-object tracking.

In research on obstacle detection models for edge-computing-based mining applications, Ruan Shunling et al. [1] proposed a lightweight obstacle detection model based on an improved YOLOv8n architecture. The model reduced parameters and inference time while maintaining detection accuracy and enhanced small-object detection capability through the introduction of BiFPN and residual connection modules to strengthen small-object feature representation. In a study on obstacle detection for open-pit mining vehicles, Gu Qinghua et al. [2] developed an improved obstacle detection model based on the YOLOv8 framework. They replaced the C2f structures at the P4 and P5 layers of the backbone network with C2fCA modules, adopted GSConv and VoV-GSCSP modules in the neck, and used the WIoU loss function in the classification prediction head. The improved model achieved robust detection performance under both daytime and nighttime conditions.

Zhang Shuai et al. [3] proposed the SCE-YOLO algorithm, which improves the backbone network, introduces new modules, and optimizes the overall architecture. On the VisDrone2019 dataset, it demonstrated higher detection accuracy and fewer parameters, although the authors noted its limited performance in detecting extremely small objects and its relatively high computational cost. In the same year, Shanshan Liu [4] introduced a novel multi-object tracking method based on the BoT-SORT framework, utilizing the FB-YOLOv8 architecture to address missed detections. The framework integrates a Feature Alignment and Aggregation Module (FAAM) and a Bidirectional Path Aggregation Network (BPAN) to enhance multi-scale feature fusion, thereby mitigating ID-switching issues and improving tracking accuracy and stability.

In 2025, Pengnian Wu et al. [5] proposed a multi-view, multi-object tracking model for low-altitude UAV scenarios (MVTL-UAV). By incorporating three loss optimization strategies tailored to multi-view target features, the model improves tracking performance without increasing resource consumption or sacrificing computational efficiency. This approach optimizes multi-view and multi-object tracking in low-altitude UAV environments, thereby enhancing its practicality and applicability in real-world operations.

Building upon the aforementioned research, this study conducts an in-depth investigation to enhance the inspection capability of UAVs in complex mining environments. A lightweight object detection model

based on an improved YOLOv8s architecture is proposed. By optimizing the network structure, refining the detection head, and improving the loss function, the model achieves higher detection accuracy for multi-scale targets, particularly for small and occluded objects. Furthermore, a pruning algorithm is employed to achieve model lightweighting, ensuring adaptability to the limited computational resources of UAV platforms.

In addition, the ByteTrack multi-object tracking algorithm is improved by introducing a spatial–appearance similarity matrix and an acceleration correction function, which collectively enhance tracking stability and continuity. Finally, a multi-level intelligent inspection system is constructed, integrating data acquisition, real-time detection, and multi-UAV collaborative scheduling to realize a fully automated inspection workflow. This system provides robust technical support for ensuring safe and efficient operations in open-pit mining environments. This paper selects YOLOv8 as the base framework rather than YOLOv12/13, primarily due to the specific requirements of open-pit mine drone inspections: (1) YOLOv8 features strong structural modularity, facilitating targeted modifications for small targets and occlusion issues in mining areas, whereas YOLOv12/13 exhibits high integration levels, making customization challenging. (2) YOLOv8 features a moderate parameter count (11.13×10^6), offering ample scope for lightweight optimization. In contrast, YOLOv12/13 increases parameters to over 14.2×10^6, making them difficult to adapt to drone computing power. (3) YOLOv8 has matured in industrial deployment with proven robustness, while YOLOv12/13 were released recently, leaving their adaptability to complex mining environments unproven. It should be noted that the method proposed in this paper is not a simple combination of existing detection and tracking technologies, but rather addresses the three core challenges of drone inspection in open-pit mines: (1) The conflict between limited computational resources on drones and the demand for high-precision models; (2) Challenges in detecting small targets, occluded targets, and deformed targets in mining areas; (3) To address the instability of similar object tracking in complex dynamic scenes, we implemented targeted innovative designs, forming a full-chain optimization solution encompassing architecture, modules, mechanisms, and systems.

2. UAV Inspection Model

Relying on its flexibility and efficiency, UAV-based inspection in open-pit mines has become an essential means of ensuring mining safety and improving operational efficiency. However, throughout this process, object detection technologies encounter a series of significant challenges.

Existing object detection models are generally large and impose extremely high demands on hardware resources. In UAV inspection scenarios for open-pit mines, both the payload capacity and onboard computational power of UAVs are limited. As a result, deploying and operating such large-scale models directly on UAV platforms is difficult. Forced deployment would severely reduce flight endurance and inspection efficiency, greatly restricting the practical application of object detection technologies. Moreover, open-pit mining areas are vast, and during UAV inspections, many small targets—such as small pieces of mining equipment or subtle geological structure changes—are present. Current detection models exhibit limited accuracy in identifying these small targets. Due to their small pixel proportion in images and weak feature representation, the models struggle to capture key information accurately, often leading to missed or false detections. This limitation poses challenges for the timely detection of potential safety hazards or subtle abnormalities in mining operations. In addition, open-pit environments are highly complex and prone to occlusions. Large stockpiles, buildings, and other mining structures frequently obstruct target objects. When targets are partially or fully occluded, it

becomes difficult for models to extract complete feature information, which significantly increases detection difficulty [6–7]. Consequently, during UAV inspections, critical targets—such as malfunctioning equipment hidden behind stockpiles or safety hazard points obscured by buildings—may not be detected in time, introducing potential risks to mining safety.

These challenges further increase the difficulty of object detection, particularly in dynamic environments. To address these issues, this study proposes an improved lightweight object detection model based on YOLOv8s, an enhanced multi-object tracking algorithm, and the implementation of an open-pit mine UAV inspection system, with the aim of improving UAV inspection capability in complex mining environments. The overall structural framework of the proposed model is illustrated in Figure 1, and the key improvements are as follows:

(1) Backbone enhancement: The backbone network integrates Deformable Convolution v2 (DCNv2) [8], while the neck adopts an Adaptive Feature Pyramid Network (AFPN) [9] to enhance multi-scale feature extraction capability.

(2) Lightweight detection head: A Lightweight Spatial Dual Enhanced Cross Decoupled Head (LSDECD-Head) [10] is designed and combined with the Focaler-GIoU loss function [11], improving detection accuracy for small and occluded targets.

(3) Model pruning: The LAMP pruning algorithm [12] is employed, which quantitatively evaluates parameter importance through a LAMP scoring mechanism and automatically removes low-contribution connections. This improves computational efficiency and meets the lightweight deployment requirements of UAV platforms.

(4) Tracking optimization: In the ByteTrack algorithm [13], a Spatial–Appearance Similarity Matrix (ASM) and an Acceleration Correction Function are introduced to reduce ID switching and trajectory drift, thereby enhancing tracking stability.

(5) System integration: A multi-level intelligent inspection system is developed to achieve full-process automation, including data acquisition, real-time detection, and multi-UAV collaborative scheduling. This system supports comprehensive mine-area monitoring and anomaly early warning.

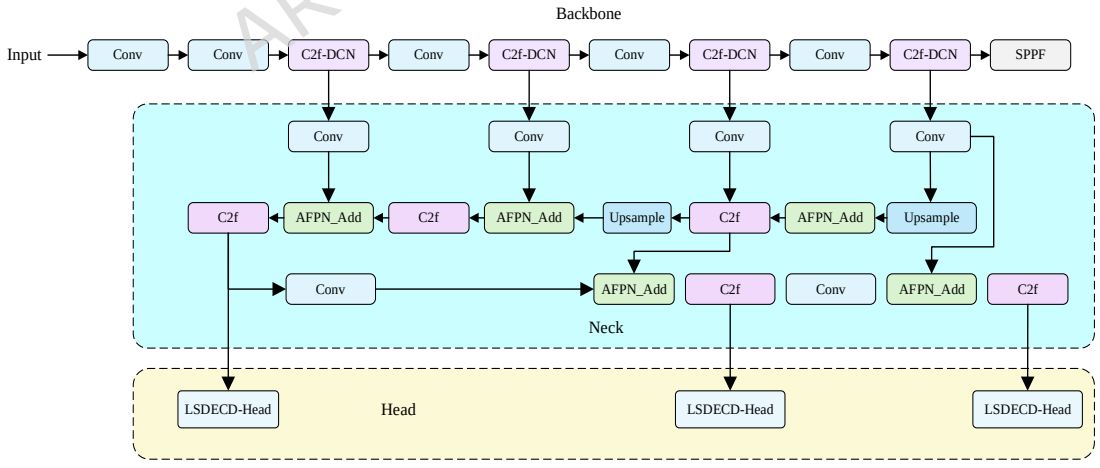

Figure 1: Improved YOLOv8s Structure Diagram

3 Improved UAV Mine Inspection Model

3.1 C2f Reconstruction

The target foreign objects within open-pit mining areas are affected by mining operations and natural environmental factors, exhibiting highly complex shapes and deformable characteristics. CNNs

encounter difficulties in accurately capturing the dynamically changing contour features of such objects. Meanwhile, UAV based monitoring platforms are constrained by limitations in size, power consumption, and onboard computational resources. These constraints impose dual technical challenges on the monitoring model, requiring both lightweight architecture and efficient feature extraction capability.

DCNs as an innovative variant of traditional CNNs, introduce an adaptive learning mechanism within the standard convolutional framework. By dynamically calculating the offsets of sampling points according to the actual geometric shape of target objects in open-pit mines, DCNs overcome the inherent limitation of fixed-grid sampling. This dynamic adjustment mechanism enables the convolution kernels to actively adapt to the complex and variable contours of mining targets, thereby enhancing their ability to capture deformation-sensitive features. The DCN architecture integrates an offset generation module that adaptively optimizes the spatial positions of sampling points during the convolution process. Consequently, this structural innovation significantly improves the model's adaptability and robustness in recognizing deformable objects under complex open-pit mining environments.

The UAV-mounted monitoring model is required to achieve a balance between lightweight design and efficient feature extraction. Taking the 3×3 convolution as an example, its output feature can be expressed as shown in Equation (1).

$$y(p_0) = \sum_{p_n \in R} w(p_n) \cdot x(p_0 + p_n) \quad (1)$$

p_0 denotes the central sampling point of the output feature map, p_n represents the offset within the receptive field of p_0 the convolution kernel, and $w(p_n)$ stands for the weight at the position of p_n.

In response to the complex and variable morphological characteristics of target foreign objects in open-pit mines, DCN introduces learnable offsets for each sampling point on the basis of standard convolution, where denotes the area of the convolution kernel. Its output feature is expressed as the following Equation (2):

$$y(p_0) = \sum_{p_n \in R} w(p_n) \cdot x(p_0 + p_n + \Delta p_n) \quad (2)$$

This mechanism enables the convolution process to dynamically adjust sampling positions, thereby achieving better adaptation to irregular targets in mining areas. Based on DCN, a weight modulation parameter for each point is incorporated to reduce irrelevant interference information. The output feature of DCNv2 is expressed as the following Equation (3):

$$y(p_0) = \sum_{p_n \in R} w(p_n) \cdot x(p_0 + p_n + \Delta p_n) \cdot \Delta m_n \quad (3)$$

This study replaces the C2f module in the backbone of the original model with DCNv2 and also substitutes the standard convolution within the Bottleneck structure with DCNv2. The structural differences before and after the replacement are illustrated in Figure 2. This modification aims to enhance the model's adaptability to complex targets in mining areas while balancing the model performance under the constraint of limited computing resources of UAVs.

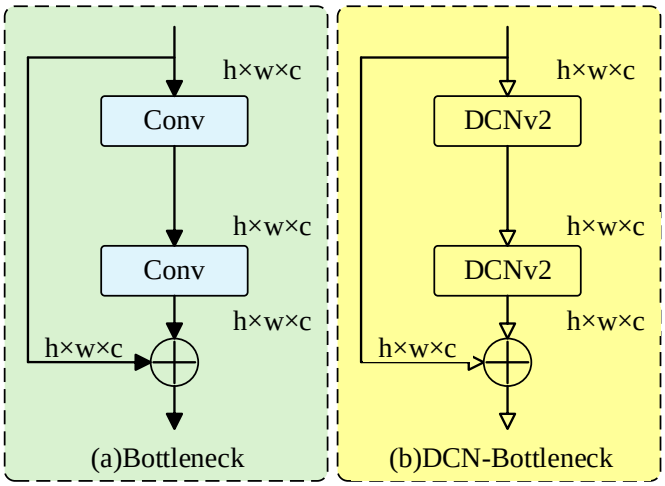

Figure 2: Bottleneck module and DCN Bottleneck module

3.2 Progressive Feature Pyramid Network (PFPN) Structure

One of the challenges in object detection within complex mining environments is addressing multi-scale features. When UAVs detect mining trucks, the trucks appear excessively small due to altitude, leading to insufficient pixel information for small targets. To resolve this issue, this study introduces a more lightweight and efficient AFPN, as illustrated in Figure 3(a).

The neck network of YOLOv8 inherits the PANet [14] structure from YOLOv5, as shown in Figure 3(b). Within the PANet structure, PANet adds a bottom-up path to the foundation of the FPN. It features an intuitive structure that is easy to understand and implement; however, PANet lacks a dedicated mechanism to address information conflicts during feature fusion. Conflicting information between features of different levels may interfere with detection results, leading to false detections or missed detections. In contrast to AFPN, PANet exhibits poorer performance when fusing non-adjacent level features. In scenarios requiring high-quality multi-scale feature fusion—such as remote sensing image detection—PANet may fail to fully fuse features of different scales, thereby impairing detection performance. Figure 3 shows comparison of PANet and AFPN network structures.

The core idea of AFPN is to gradually narrow the semantic gap between features of different levels through an asymptotic fusion strategy. Inspired by HRNet's concept of continuously fusing high-level and low-level features, AFPN starts with fusing low-level features during the bottom-up feature extraction process of the backbone network and gradually incorporates high-level and top-level features. This avoids the semantic gap issue that arises when non-adjacent level features are directly fused [15].

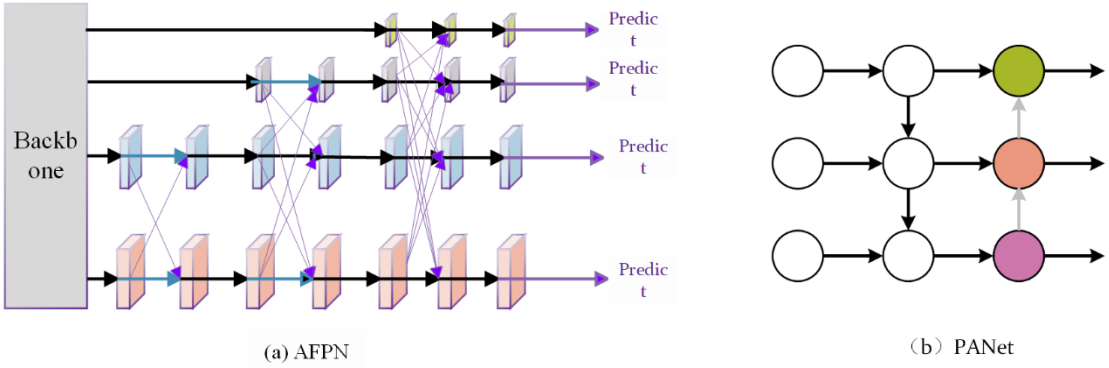

Figure 3: Comparison of PANet and AFPN Network Structures

AFPN adopts an asymptotic approach for feature fusion: it first fuses adjacent low-level features, and then gradually incorporates high-level features. Taking the fusion of three features as an example,

suppose the features to be fused are F_1, F_2, and F_3 (Note: There may be a typo in the original text; it is assumed to be F_3 here for logical consistency of feature levels), and the fusion process proceeds incrementally. First, F_1 and F_2 are fused to obtain the intermediate feature F_{12}, which can be achieved through simple element-wise addition or concatenation followed by a convolution operation. Element-wise addition is used as an example here:

$$F_{12} = F_1 + F_2 \quad (4)$$

Then, F_{12} is fused with F_3 to obtain the final fused feature final:

$$F_{final} = F_{12} + F_3 \quad (5)$$

To enable the model to learn better fused features, a series of convolutional layers (e.g., residual units) are typically employed for further processing after the addition operation. Assuming processing through a residual unit:

$$R(x) = x + f_{3\times3}^{2}(f_{3\times3}^{1}(x)) \quad (6)$$

$f_{3\times3}^{1}$ and $f_{3\times3}^{2}$ denote two 3×3 convolution operations, respectively. During the multi-level feature fusion process, features from different levels may have information conflicts at certain positions. To tackle this problem, AFPN adopts an adaptive spatial fusion operation. Suppose the three levels of features to be fused are $x_{ij}^{1\to l}$, $x_{ij}^{2\to l}$ and $x_{ij}^{3\to l}$, which respectively represent the feature vectors at position (i, j) from levels 1, 2, and 3 to level l. The resulting feature vector obtained via adaptive spatial fusion is a linear combination of these three feature vectors:

$$y_{ij}^{l} = \alpha_{ij}^{l} \cdot x_{ij}^{1\to l} + \beta_{ij}^{l} \cdot x_{ij}^{2\to l} + \gamma_{ij}^{l} \cdot x_{ij}^{3\to l} \sqrt{b^2 - 4ac} \quad (7)$$

α_{ij}^{l} , β_{ij}^{l} and γ_{ij}^{l} denote the spatial weights of the three hierarchical features at position (i, j) on level l, with the constraint that:

$$\alpha_{ij}^{l} + \beta_{ij}^{l} + \gamma_{ij}^{l} = 1 \quad (8)$$

These spatial weights are learned by the model. Typically, features from different levels are processed through a convolutional layer f_{conv} to generate a weight map, which is then normalized using the softmax function to obtain weights that satisfy the constraint conditions. Let the output obtained by concatenating features from the three levels and processing them through the convolutional layer be denoted as w_{ij}^{l}:

$$w_{ij}^{l} = f_{conv}([x_{ij}^{1\to l}, x_{ij}^{2\to l}, x_{ij}^{3\to l}]) \quad (9)$$

$[x_{ij}^{1\to l}, x_{ij}^{2\to l}, x_{ij}^{3\to l}]$ denotes the feature concatenation operation. Then, the spatial weights are obtained by applying the softmax operation to :

$$\begin{cases} \alpha_{ij}^{l} = \dfrac{\exp(w_{ij}^{l,1})}{\exp(w_{ij}^{l,1}) + \exp(w_{ij}^{l,2}) + \exp(w_{ij}^{l,3})} \\ \beta_{ij}^{l} = \dfrac{\exp(w_{ij}^{l,2})}{\exp(w_{ij}^{l,1}) + \exp(w_{ij}^{l,2}) + \exp(w_{ij}^{l,3})} \\ \gamma_{ij}^{l} = \dfrac{\exp(w_{ij}^{l,3})}{\exp(w_{ij}^{l,1}) + \exp(w_{ij}^{l,2}) + \exp(w_{ij}^{l,3})} \end{cases} \quad (10)$$

$w_{ij}^{l,1}$, $w_{ij}^{l,2}$ and $w_{ij}^{l,3}$ are the values of the three channels corresponding to . Through the above mathematical derivation, AFPN can achieve effective fusion of features from different levels while

addressing the issue of information conflicts during feature fusion, thereby improving the performance of object detection. In terms of computational resources, AFPN optimizes the model by strategies such as reducing network width, which reduces computational load while ensuring detection accuracy. Given the limited computational resources of UAVs, this characteristic of AFPN enables it to utilize limited computational resources more efficiently when running on UAVs, ensuring the real-time performance of detection tasks. For example, in the real-time monitoring of open-pit mining progress, AFPN can quickly process images under the limited hardware conditions of UAVs and promptly feedback information such as the transportation status of mining trucks and the operating conditions of equipment. Figure 4 shows the original image, while Figures 5 and 6 present the feature channel maps before and after improvement, respectively.

Figure 4: Original image

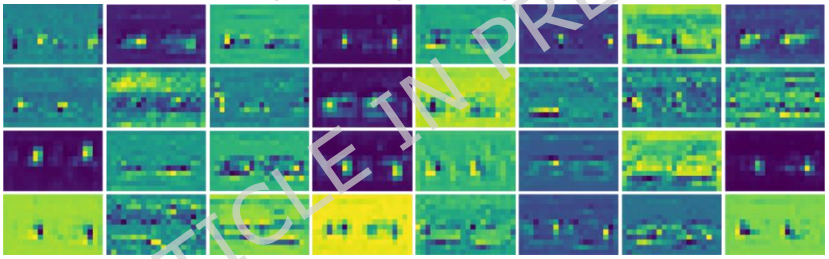

Figure 5: The network channel feature map before improvement

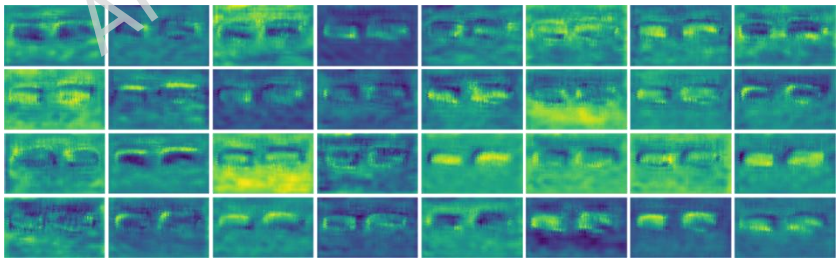

Figure 6: Improved network channel characteristics

A visual improvement in AFPN's feature fusion effect can be observed when comparing Figure 4 (the original image), Figure 5 (the network channel feature map before improvement), and Figure 6 (the network channel feature map after improvement). In the improved network channel feature map, features of targets at different scales are effectively fused: small-target features are enhanced, with finer details Such as the outlines and markers of mining trucks, becoming clearer; large-target features are extracted completely and accurately; and the transition between features of different levels is more natural and continuous. This enables the model to identify and locate multi-scale targets in complex mining environments with higher precision. For instance, in practical detection scenarios, both small distant equipment and large nearby mining trucks can be detected quickly and accurately. This effectively avoids missed detections and false detections caused by poor feature fusion, significantly improving the

reliability and efficiency of UAV-based inspection in open-pit mines.

3.3 Lightweight Detection head

The open-pit mine environment is complex, with enormous scale differences between equipment and target objects—ranging from small mining trucks to large mining machinery—while different targets are widely distributed in space with variable positions. Although the original detection head of YOLOv8 can detect targets quickly, it has limitations in handling multi-scale targets and those with complex spatial distributions [16]. In contrast, the scale-aware attention mechanism of LSDECD-Head can dynamically fuse features from different levels according to target scales, enhancing sensitivity to targets of varying sizes. This enables more accurate identification of both small equipment failure points in the distance and large nearby mining trucks. Its spatial-aware attention mechanism utilizes deformable convolution to adaptively aggregate multi-level features, focusing on the discriminative regions of foreground targets. Even when targets are rotated, occluded, or affected by varying terrain and lighting conditions, stable detection can be achieved.

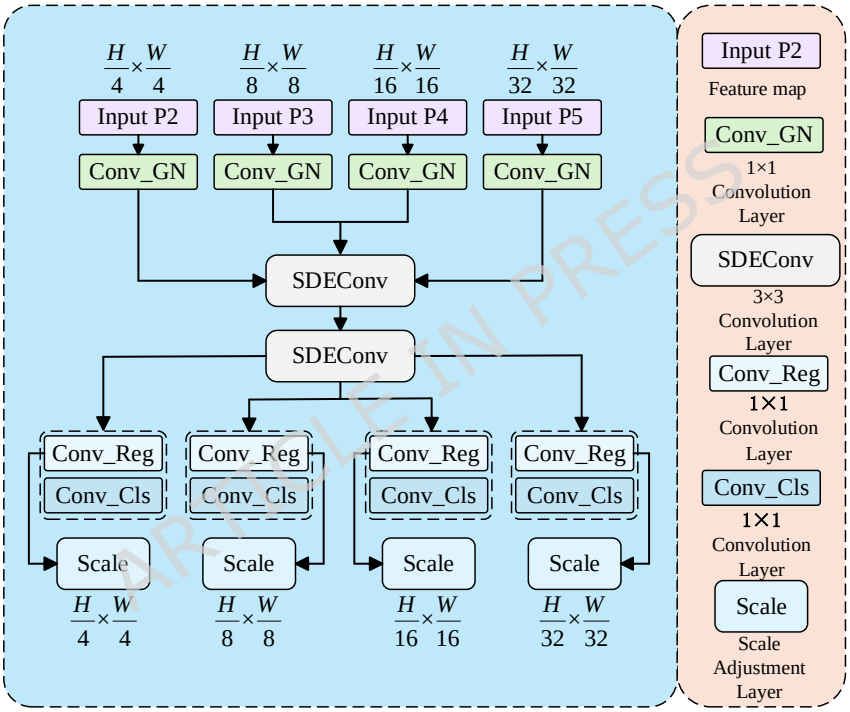

Figure 7: LSDECD Head Structure Diagram

Figure 7 shows the structure diagram of LSDECD-Head. The lightweight shared detail-enhanced convolutional detection head is designed with comprehensive consideration of the requirements for parameter quantity, computational complexity, and detection accuracy, and it receives four feature layers output by the feature pyramid network. First, the feature layers output by the feature pyramid undergo input channel adjustment via 1×1 convolution. The preprocessed feature layers are aggregated into two 3×3 shared detail-enhanced convolution modules (SDEConv) for efficient aggregation of feature information. Subsequently, the feature information generated by the shared detail-enhanced convolution is fed into the classification branch and regression branch respectively. The classification branch uses a 1×1 classification convolution (Conv_Cls) to predict the target category probability, while the regression branch predicts the coordinate offsets of the bounding box through a 1×1 regression convolution (Conv_Reg). Finally, to address the inconsistency in target scales among different detection heads, the regression branch introduces a Scale layer to perform scale adjustment on the output features, thereby

adapting to the output of multi-scale target feature information.

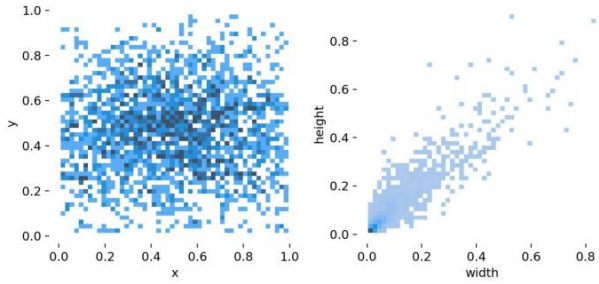

Figure 8: Label distribution diagram

As can be seen from the label distribution diagram in Figure 8, the detected targets in open-pit mines exhibit significant scale differences, with a relatively large proportion of small targets. This imposes extremely high requirements on the detection head, and LSDECD-Head effectively addresses this challenge through its unique design. In practical detection, for tiny parts, the scale-aware attention mechanism enhances focus on small-scale features. It accurately screens and fuses the key information of tiny targets from the feature layers output by the feature pyramid network, enabling the model to accurately capture the detailed features of equipment and recognize them clearly even against complex backgrounds. When dealing with large mining machinery, the spatial-aware attention mechanism of this detection head plays a crucial role.

By utilizing deformable convolution to adaptively aggregate multi-level features, it breaks through the fixed sampling limitations of traditional convolution. It dynamically adjusts sampling points according to the morphological changes of large machinery, thereby extracting their features comprehensively and accurately. Even when parts of the machinery are occluded or under complex lighting conditions, stable detection results can still be output by focusing on the discriminative regions of foreground targets. In addition, the design of the lightweight shared detail-enhanced convolutional detection head ensures detection accuracy while balancing parameter quantity and computational complexity. Its shared detail-enhanced convolution modules (SDEConv) efficiently aggregate feature information, reducing unnecessary computational overhead. The rational design of the classification and regression branches, along with the adaptive adjustment of the Scale layer in the regression branch for multi-scale target feature information, collectively guarantee the efficiency and accuracy of the model in handling targets of various scales, significantly improving the reliability of equipment detection in open-pit mines.

3.4 Improvement of the Loss Function

In the process of using UAVs for open-pit mine monitoring, the performance of object detection is crucial to safe production. In multi-scale object detection scenarios, the performance of the CIoU [17] loss function is significantly affected by the aspect ratio of targets, which limits its effectiveness in scenarios where target scales vary drastically. Its core calculation formula is shown in Equation (11), and through this dynamic weighting strategy, the contribution of samples with different levels of difficulty to the training process is balanced.

$$IoU^{\text{focaler}} = \begin{cases} 0, & \text{IoU} < d \\ \dfrac{\text{IoU} - d}{u - d}, & d < \text{IoU} < u \\ 1, & \text{IoU} > u \end{cases} \quad (11)$$

When $IoU < d$, the $IoU^{focaler}$ is set to 0, indicating that these samples are ignored. When $d < IoU <$

u, the IoU is linearly scaled to reassign the weights of these samples. When $IoU > u$, the $IoU^{focaler}$ is directly set to 1, indicating that high-IoU samples contribute the most. The loss function of Focaler-IoU is defined as shown in Equation (12).

$$L_{Focaler-IoU} = 1 - IoU^{focaler} \quad (12)$$

In object detection tasks, the value characteristics of Focaler-IoU reflect the proximity between the predicted bounding box and the ground-truth box: when the value approaches 1, it indicates a high degree of overlap between the two boxes, corresponding to a small loss value; when the value approaches 0, it means the overlap between the two boxes is low or even completely separated, and the loss increases accordingly.

To address this, this study proposes integrating the sample weighting idea of Focaler-IoU into the GIoU [18] framework, dynamically adjusting sample weights through a linear interval mapping mechanism. This method differentially assigns weight loss based on the regression difficulty of samples, enabling the model to focus more on hard samples that play a key role in detection accuracy during training. This optimizes the gradient update direction and improves the overall performance of multi-scale object detection tasks. The improved Focaler-GIoU formula is shown in Equation (13).

$$L_{Focaler-GIoU} = L_{GIoU} + IoU - IoU^{Focaler} \quad (13)$$

3.5 Model Pruning

The core idea of the LAMP pruning method stems from the assumption that "parameters with smaller absolute weights contribute less to network output." It achieves model sparsification by filtering out connection weights with smaller absolute values. In this paper, a hierarchical sparsity screening strategy is adopted to quantitatively evaluate network parameters based on the LAMP scoring mechanism. First, each weight tensor is unfolded into a one-dimensional vector to construct a unified parameter evaluation space. Subsequently, the importance of the weight for the u-th index in the tensor is measured using the LAMP scoring formula defined in Equation (14). Through structured parameter screening rules, this method efficiently simplifies network connections while preserving the model's core feature expression capability.

$$\text{score}(u;W) := \frac{(W[u])^2}{\sum_{v \geq u}(W[v])^2} \quad (14)$$

In the formula, denotes the LAMP score of the u-th index in the weight tensor W; when $u < v$, holds consistently; represents the weight term mapped by index u, and denotes the sum of squared weight magnitudes of all remaining connections in the same layer. In the LAMP pruning method, the importance of a connection is quantified by the scoring mechanism defined in Equation (14): the greater the contribution of a connection to the network output, the higher the value of the numerator term in its score, while the denominator term becomes smaller due to the characteristics of related parameters, resulting in an overall higher score. In specific implementation, the score is first calculated for all connections, then a global screening is performed in ascending order of scores. Connections with the smallest scores are continuously pruned until the preset global sparsity requirement is met. This process essentially involves the algorithm automatically determining the sparsity ratio of each layer, enabling adaptive selection of hierarchical sparsity.

As shown in Figure 9, the pruning operation removes connections with small weight magnitudes, retaining only the network structure composed of high-magnitude weights. This achieves parameter

simplification while preserving core connections.

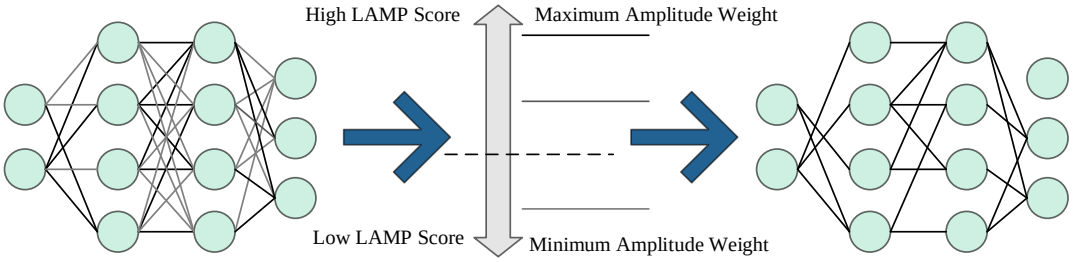

Figure 9: LAMP schematic diagram

3.6 Improvements to the ByteTrack Multi-Object Tracking Algorithm

In complex open-pit mine scenarios, the traditional ByteTrack multi-object tracking algorithm faces numerous challenges. Mining equipment and personnel exhibit complex and variable movement patterns; different devices have similar appearances, and there are issues of severe occlusion and lighting changes. These factors make traditional algorithms prone to errors in target association and trajectory prediction. To address these problems, targeted improvements have been made to the ByteTrack algorithm. To solve the problem of difficulty in distinguishing similar targets, a spatial-appearance similarity matrix (ASM) [19] is introduced. In open-pit mine scenarios, ASM is further optimized by considering not only the spatial position and appearance features of targets but also incorporating information such as the functional attributes and operational status of equipment. For example, mining trucks in different operational states (loaded, unloaded) may have subtle differences in appearance. Incorporating these differences into similarity measurement enables more accurate differentiation of similar targets. Regarding the issue of error accumulation in the Kalman filter, drawing on the design ideas of target correction functions in related studies and combining the movement characteristics of open-pit mine equipment, a new target correction strategy is proposed. Considering the changes in motion acceleration of mining equipment on different terrains, a correction term for acceleration prediction errors is added to the target correction function, making the corrected target state closer to the real value.

In the calculation of the improved spatial-appearance similarity matrix, assuming there are M targets in the current frame and N targets in the previous frame, the formula for calculating the spatial similarity matrix for targets i and j is:

$$A_{space}(i,j)=\exp\left(-\frac{d(X_i^t,X_j^{t-1})+\omega|s_i-s_j|}{\sigma_s}\right) \quad (15)$$

where represents the spatial distance between target i in the current frame and target j in the previous frame, calculated as:

$$\sqrt{(x_i-x_j)^2+(y_i-y_j)^2}+\beta|v_i-v_j|+\alpha|a_i-a_j| \quad (16)$$

α and β are weights controlling velocity and acceleration, respectively; s_i and s_j are quantized values of the operational status of targets i and j (e.g., 1 for loaded, 0 for unloaded); ω is the operational status weight coefficient, which adjusts the degree of influence of operational status on spatial similarity. The formula for appearance similarity is:

$$A_{appearance}(i,j)=\frac{f_i^t\cdot f_j^{t-1}}{f_i^t\cdot f_j^{t-1}}\times(1+\gamma|c_i-c_j|) \quad (17)$$

f_i^t and f_j^{t-1} represent the appearance feature vectors of targets in the current frame and the

previous frame, respectively; c_i and c_j are the color feature difference values of targets i and j (obtained through color space calculation); and is the color feature weight coefficient.The final ASM matrix is:

$$ASM(i,j) = \varepsilon A_{space}(i,j) + (1-\varepsilon) A_{appearance}(i,j) \quad (18)$$

To balance the weight parameters of spatial features and appearance features, the target correction function is improved as:

$$X_i^t = (1-\lambda) X_i^t + \lambda X_j^{t-1} + \mu(a_i^t - a_j^{t-1}) \quad (17)$$

λ is the correction coefficient, dynamically adjusted by the feature matching similarity of the ASM matrix; μ is the acceleration correction coefficient; a_i^t and a_j^{t-1} are the accelerations of target i in the current frame and target j in the previous frame, respectively.

In UAV-based open-pit mine inspection, after the UAV captures each frame of image, the improved YOLOv8s model detects mining equipment and personnel in the image, outputting detection boxes and related feature information. This information is input into the improved ByteTrack algorithm. First, the Kalman filter predicts the trajectory of high-score detection boxes based on the information from the previous frame, generating predicted bounding boxes. Then, the ASM matrix between targets in the current frame and the previous frame is calculated, and target matching is performed using the Hungarian algorithm based on the ASM matrix. If the matching is successful, the trajectory is updated according to the matching results; if it fails, a second round of feature association and trajectory matching is conducted. During the matching process, the improved target correction function is used to correct the predicted state, reducing the error accumulation of the Kalman filter. For example, when a mining truck turns, the change in acceleration is captured by the target correction function and used to adjust the predicted state, making tracking more accurate. After multiple rounds of matching and correction, the tracking results of the current frame are finally obtained, enabling precise tracking of mining equipment.

4 Experimental Results and Analysis

4.1 Experimental Results and Analysis

In the initial stage of dataset construction, drones were used for on-site shooting directly in the open-pit mine. The drone aerial equipment and related parameters are shown in Figure 10 and 11 , as well as Table1. Through this method, a total of 6,000 images were collected. This dataset focuses on core inspection targets for open-pit mines, with specific details as follows: Target Types and Category Quantity, Clearly includes 2 core equipment target categories (mining trucks, excavators), while also covering non-cooperative targets (unauthorized vehicles/personnel), equipment failure locations, and other key inspection objects. Mining trucks account for approximately 65% of the sample, while excavators and other targets account for 35%.Image dimensions: The original image resolution is 5472×3648 pixels (matching the parameters of the drone aerial photography equipment). After cropping and scaling preprocessing, it is uniformly adjusted to 640×640 pixels for model training. Scene and Data Partitioning: Covers typical scenarios such as mining equipment operations and ore pile distribution, incorporating samples captured at varying distances and angles. These include small targets (distant equipment) and partially occluded objects. The dataset is partitioned into training, validation, and test sets at a 7:2:1 ratio. Annotation was performed using LabelImg software with bounding box annotations to ensure accuracy. Drones can reach areas that

are difficult for personnel to access on foot, cover large areas in a short time, and capture images from different angles. This provides a more comprehensive perspective of the open-pit mine environment for this study.

Figure 10: UAV aerial photography platform

Figure 11: Tilt photography module

Table 1 Equipment Parameter

Project	Parameter
Sensor size	23.1 mm × 15.4 mm
Minimum photo interval	0.5 s
Effective pixels	Approximately 125 million（2600 W × 5）
Pixel size	3.76 μm
Weight	615 g
Storage mode	FMS - 640 Centralization
Size	107 mm × 107 mm × 89 mm
Lens parameters	28 mm prime lens (downward view), 40 mm prime lens (tilted)
Image display	"Real-time display" supports 5-camera manual live view (in 5G transmission mode)

To improve the quality and improve the quality and clarity of these images, preprocessing was first performed, including image enhancement and denoising. For image enhancement, a sharpening filter method was adopted. By enhancing the high-frequency components in the image, the clarity of the image is improved. This method can make details in the image more prominent, which helps the model better identify targets. In terms of denoising, the non-local means denoising method was used. Non-local means denoising calculates the value of the current block by finding similar blocks in the image, thereby achieving a denoising effect. This method can not only effectively remove noise but also preserve the edge and texture information of the image, which is very helpful for improving the performance of the detection algorithm. Enhancing the diversity of the dataset through enhancement techniques helps the model learn more robust features, thereby improving its adaptability in different environments. Reducing noise in images can improve image quality, making it easier for the model to identify useful features—this is particularly important for improving detection accuracy and ensuring optimized image visual effects. Preprocessing steps help reduce data redundancy during model training, thus supporting the model's lightweight goal. Image quality (denoising, enhancement) and ensuring the clarity of target objects in images (cropping, scaling) are addressed through these preprocessing steps, which help improve the model's detection accuracy and enable the model to learn feature representations under

different environmental conditions, thereby enhancing its adaptability in complex environments. After preprocessing, the images were further processed, including cropping and scaling. The purpose of cropping is to remove unnecessary blank areas in the image, while scaling adjusts the image resolution to a size suitable for model training. After completing these steps, the LabelImg software's annotation tool was used to annotate the images in detail. Model training and inference hardware are standardized on the NVIDIA GeForce RTX 3070 SUPER graphics card (4096 GPU cores, 616 GB/s memory bandwidth), AMD Ryzen 7 5800H CPU (8 cores, 16 threads, base frequency 3.2 GHz), and 32 GB DDR4 3200 MHz memory; The drone-deployed hardware consists of an NVIDIA Jetson Xavier NX (GPU compute power: 21 TOPS, 8 GB memory). All power consumption and inference time data are based on actual measurements using the aforementioned hardware. Figure 12 shows representative samples from the dataset.

Figure 12: Some datasets are displayed

The experiment was conducted on an Autodl cloud server. The server configuration is as follows: the operating system is Windows 10, equipped with an NVIDIA GeForce RTX 3070 SUPER graphics card with 8GB of video memory, and the CPU model is AMD Ryzen 7 5800H. This experiment used PyTorch framework version 2.5.1 and Python programming environment 3.10. During the training process, the image input size was set to 640×640 pixels, and the maximum number of training epochs was set to 300 to ensure that the model could fully learn and optimize its performance.

4.2 Evaluation Index

When evaluating the model in the complex environment of open-pit mines, it is necessary to comprehensively consider multiple performance aspects of the model, such as detection accuracy, computational resource consumption, and real-time performance. In addition to basic metrics, a series of specific indicators that align with the actual needs of open-pit mines are introduced to achieve a comprehensive evaluation of the model. mAP50 is a key indicator for evaluating the detection accuracy of the model. It is calculated by averaging the average precision of each category when the IoU threshold is 0.5, with the formula as follows:

$$mAP50 = \frac{1}{n}\sum_{i=1}^{n} P_{ave_i} \qquad (18)$$

n is the number of categories, P_{ave_i} represents the average precision of the i-th category. It is obtained by averaging the product of precision P and recall R under different confidence thresholds.

$$P_{ave} = \frac{1}{k}\sum_{j=1}^{k} P\,R \quad (19)$$

$$P = \frac{TP}{TP + FP} \quad (20)$$

$$R = \frac{TP}{TP + FN} \quad (21)$$

K is the number of different confidence thresholds. TP represents the number of true positives correctly identified by the model; FP represents the number of false positives where the model incorrectly identifies negatives as positives; and FN refers to the number of false negatives that the model fails to correctly identify.

Parameters (Parameters/10^6), Model Size (Model Size/MB), Floating-Point Operations (FLOPs), and inference time are metrics that reflect the model scale. The number of trainable parameters in the model is directly counted and expressed in units of 10^6. Model size measures storage requirements and is obtained by calculating the disk space needed to store the model. FLOPs measure computational complexity, calculating the number of floating-point operations required for the model to perform a single forward propagation. These metrics clearly demonstrate the model's demand for computational resources. In multi-object tracking, evaluation can be conducted using the following metric: the Multi-Object Tracking Accuracy (MOTA). It comprehensively considers detection accuracy and target identity matching, with its formula as follows:

$$MOTA = 1 - \frac{\sum_t |FN_t| + |FP_t| + |ID_s wt|}{\sum_t |GT_t|} \quad (22)$$

$|FN_t|$,$|FP_t|$,$|ID_s wt|$andGT_trepresent the number of missed detections, false detections, ID switches, and true targets at time t, respectively. A higher MOTA indicates that the algorithm has higher accuracy in detecting and associating targets during the tracking process.

Multi-Object Tracking Precision (MOTP) focuses on the matching precision of target positions, with its formula as follows:

$$MOTP = \frac{\sum_t \sum_{i \in S_t} d(i_t^t, g_{\mu(i)}^t)}{\sum_t |S_t|} \quad (23)$$

$S_t$represents the set of matched target pairs at time t; $d(i_t^t, g_{\mu(i)}^t)$ represents the distance between the predicted bounding box and the ground truth bounding box; $|S_t|$ and t is the number of matched target pairs at time t. A higher MOTP indicates that the algorithm's prediction of target positions is more accurate.

The ID F-score (IDF1) measures the accuracy of target identity recognition, with its formula as follows:

$$IDF1 = \frac{2 \times IDTP}{2 \times IDTP + IDFP + IDFN} \quad (24)$$

IDTP, IDFP, and IDFN respectively represent the number of correctly assigned, incorrectly assigned, and missed target IDs. A higher IDF1 indicates that the algorithm has a stronger ability to recognize and maintain target identities during tracking.

The ID Switch (target ID switch count) directly reflects the frequency of incorrect target identity

switches during the algorithm's tracking process. A lower value of this metric means the algorithm has higher tracking stability for targets and can more accurately track the same target continuously.

4.3 Model ablation experiment

The model ablation experiments aim to explore the specific contributions of each improved module to model performance. By removing or replacing key modules one by one and comparing the performance differences under different model configurations, the mechanism of each component in the overall model is clearly understood. Under the premise of keeping experimental conditions consistent, four groups of ablation experiments were conducted using the UAV aerial dataset. These experiments, as shown in Table 2, aim to test the contribution of each improved part to model performance independently, thereby verifying whether the method in this study can reduce the model size while maintaining or even improving the model's detection accuracy. Through these experiments, the impact of each component on the final performance of the model can be more clearly understood. In the experiments, three key improvements were made to the model, marked as A, B, and C respectively. A refers to the replacement of C2f in the model's backbone network with C2f-DCN and the upgrade of the feature extraction structure to AFPN. B refers to the improvement of the detection head to LSDECD-Head. C refers to the improvement of the loss function.

Table 2 Ablation experiment

Experiment serial number	A	B	C	mPA50	Parameters/10^6	FLOPs/G	Inference time / ms
①				0.557	11.13	28.5	347
②	√			0.795	4.84	17.9	291
③	√	√		0.823	4.39	14.7	232
④	√	√	√	0.868	4.39	14.7	196

In-depth analysis of the model ablation experiment results reveals a synergistic effect among the various improved components. Improvement A lays the foundational architecture for efficient feature extraction in the model: the effective fusion of multi-scale features by AFPN enables subsequent improvements to better exert their effects. Improvement B's LSDECD-Head detection head, based on the rich features provided by A, further enhances the ability to recognize targets of different scales. The combination of A and B strengthens the model's adaptability to complex scenarios. Meanwhile, the optimized loss function in Improvement C can more appropriately adjust the model's training direction based on the detection results provided by the first two improvements, significantly boosting the overall performance of the model.

This also provides a clear direction for subsequent model optimization. For example, in scenarios with limited resources but slightly lower accuracy requirements, focus can be placed on applying Improvement A to simplify the model structure and reduce computational costs. In contrast, for key area monitoring scenarios that demand extremely high detection accuracy, strengthening Improvement C while combining it with A and B can fully leverage the model's advantages. Additionally, efforts can be made to explore applying these improvement ideas to other detection algorithms, verify their universality, and promote the application of UAV detection technology in open-pit mines across broader fields. This will continuously enhance the intelligent monitoring level of open-pit mines and ensure safe and efficient production operations. A comparison of model detection results before and after improvement is shown in Figure 13.

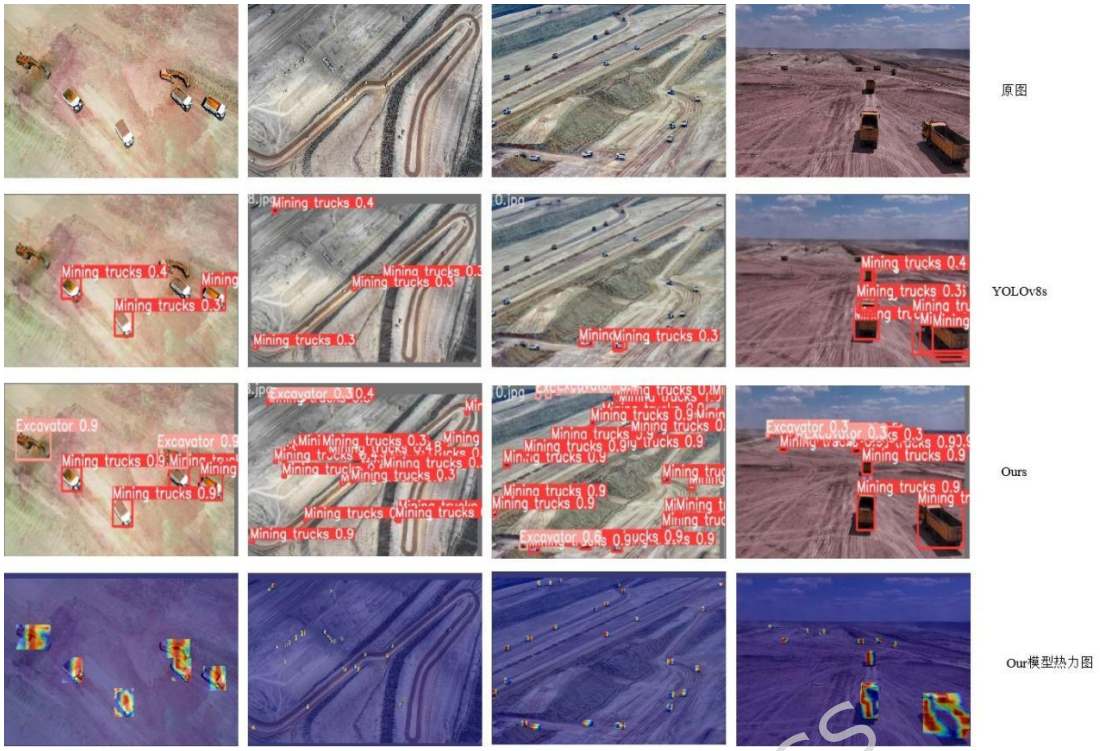

Figure 13: Comparison before and after model improvement

The comparison images intuitively show that, in the open-pit mine target detection task, the Ours model has higher detection precision and recall than the YOLOv8s model. It can detect targets such as mining equipment in open-pit mine scenarios more accurately and comprehensively. Its heatmap also reflects the effectiveness of the model's detection mechanism from a side perspective. The experimental data of relevant parameters after training are shown in the figure 14 below.

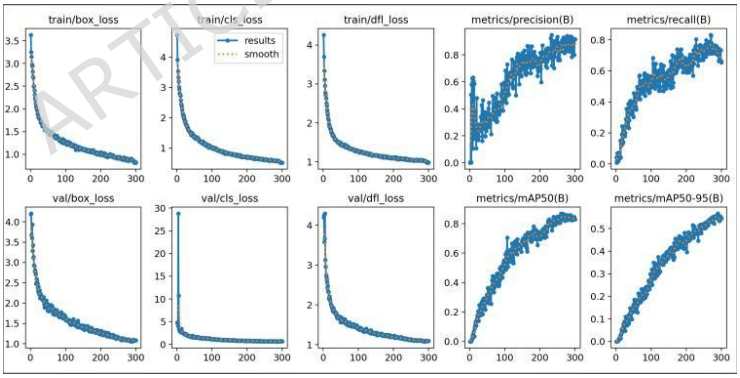

Figure 14: The improved relevant experimental data

（1）Sparse factor setting

In general, these two architectural adjustment strategies, especially when used in combination, have optimized the model performance in terms of parameters, computational load, model size, and inference speed. It can be clearly seen from the comparison chart that before sparse training, the distribution of the channel scale factor γ in the BN layer is relatively scattered, and the differences between different channels are relatively small. This means that the roles of each channel in the model are relatively balanced, but there may also be some redundant information. After sparse training, the γ distribution has changed significantly: the scale factors of some channels are close to 0, indicating that these channels contribute little to the model and can be safely removed in subsequent pruning operations. This change provides an important basis for model pruning, enabling us to effectively reduce the number of

parameters and computational complexity of the model without significantly affecting its performance.

（2）Model pruning experiment

After completing sparse training, the LAMP pruning method was used to perform pruning operations on the model. Based on the LAMP scoring mechanism, the network parameters were quantitatively evaluated, and a global selection was conducted in ascending order of scores, with the connections with the smallest scores pruned step by step. During the pruning process, close attention was paid to changes in model performance. The optimal pruning rate was determined by comparing indicators such as mAP50, number of parameters, FLOPs, and inference time of the model on the validation set before and after pruning. The results are shown in the table below:

Table 3 Model performance under different pruning rates

Pruning rate	mPA50	Parameters/10^6	FLOPs/G	Inference time / ms
0	0.857	5.62	20.2	305
0.1	0.860	5.13	18.5	274
0.2	0.870	4.82	16.2	225
0.3	0.868	4.39	14.7	196
0.4	0.862	4.13	10.6	182
0.5	0.795	3.95	9.4	170
0.6	0.642	3.52	9.3	158
0.7	0.592	3.09	8.5	145
0.8	0.553	2.66	7.6	130

In the model compression experiment, it can be clearly seen from Table 3 that as the pruning rate gradually increases, the model's number of parameters, FLOPs, and inference time all show a continuous downward trend, while the mAP50 indicator first increases and then decreases. When the pruning rate is 0.3, the model's comprehensive performance reaches the optimal state: the mAP50 is 0.868, the number of parameters is 4.39×10^6, the FLOPs is 14.7G, and the inference time is 196 ms. This result fully shows that with a pruning rate of 0.3, the model achieves a good balance between detection accuracy and computational efficiency. However, when the pruning rate is too high, the model loses a large number of key connections and parameters, which leads to a significant decline in its feature expression ability and ultimately a marked reduction in detection accuracy. In contrast, when the pruning rate is 0.3, the model successfully retains most of the key feature channels while removing redundant parts. Therefore, it can maximize computational efficiency based on ensuring detection accuracy. Overall, the model compression experiment verifies the effectiveness of the LAMP pruning method. By selecting the pruning rate reasonably, the number of parameters and computational complexity of the model can be greatly reduced without significantly reducing detection accuracy. This enables the model to better adapt to devices with limited computing resources such as UAVs, providing strong technical support for the real-time performance and efficiency of UAV inspections in open-pit mines.

4.4 Comparison tests of different algorithms

To further demonstrate the effectiveness of the method in this study for open-pit mine equipment detection applications, especially from a UAV perspective,ts performance was comprehensively compared and analyzed with other popular algorithms and models. This comparison can highlight the competitiveness of the lightweight model and its potential advantages in practical applications. In scenarios where UAVs assist in open-pit mine operations, UAVs can provide an overhead view of the

entire mining area. The proposed lightweight model, whether mounted on a UAV or used to process images captured by UAVs, is expected to quickly and accurately identify unmanned mining trucks and various types of mining equipment. This is crucial for ensuring the safety and efficiency of mining operations. By comparing it with other models, it can be clearly seen that the lightweight model has a faster detection speed, a feature that is extremely critical for real-time UAV monitoring.

Table 4 Performance comparison of different algorithm models

Method	mAP50	Parameters/10^6	FLOPs/G	Model Size/MB	Inference time/ms	cpuOccupancy rate	Power consumption（w）
YOLOv8s	0.553	11.13	28.5	21.46	364.23	38.5±2.3	12.8±1.1
SSD	0.724	8.94	15.7	5.94	223.77	42.1±1.8	10.5±0.9
YOLOv9s	0.584	12.13	18.7	23.25	425.36	45.7±2.1	14.2±1.3
YOLOV10s	0.564	10.57	22.9	19.57	353.24	40.3±1.9	13.1±1.0
YOLOV11s	0.573	9.1	23.8	17.65	323.41	39.8±2.0	12.5±0.8
Faster RCNN	0.582	7.9	11.8	8.27	287.4	51.2±2.5	15.3±1.2
ours	0.868	4.39	14.7	4.76	196	28.6±1.5	8.3±0.7

Table 4 compares the core performance of the "ours" model in this study with 6 mainstream algorithms such as YOLOv8s and SSD. From three key deployment dimensions—detection accuracy, resource consumption, and real-time performance—it highlights the adaptability advantage of the "ours" model for UAV inspection scenarios in open-pit mines, with specific details as follows:The mAP50 of the "ours" model reaches 0.868, which is much higher than that of the second-best SSD (0.724) and 1.57 times that of YOLOv8s (0.553). It can accurately identify small targets and occluded targets in open-pit mines (such as equipment behind ore piles), avoid safety hazards caused by missed detections, and meet the high-precision monitoring needs of mining areas.

In terms of resource consumption, the "ours" model has only 4.39×10^6 parameters (39.4% of YOLOv8s), a model size of 4.76MB (80.1% of SSD), and a CPU usage rate of 28.6%±1.5% (lower than Faster RCNN's 51.2%). It is compatible with the limited computing power and memory of UAV edge computing modules, preventing inspection interruptions caused by hardware overload. Regarding real-time performance, the "ours" model has an inference time of 196ms (12.4% faster than SSD and 46.2% faster than YOLOv8s), supporting real-time detection of 5 frames per second, which matches the UAV inspection speed. Its power consumption is 8.3W±0.7W (35.2% lower than YOLOv8s), which can reduce UAV battery consumption and extend the endurance and coverage of a single inspection.

In summary, the "ours" model achieves a three-dimensional balance in accuracy, lightweight design, and real-time performance. It is an algorithm that can be directly deployed on the UAV end to meet the dynamic inspection needs of open-pit mines. Other algorithms are difficult to apply in practice—either due to insufficient accuracy or excessive resource consumption.

From the perspective of practical application scenarios, the algorithm in this study shows unique advantages in UAV inspections of open-pit mines: its lightweight feature allows the model to be directly deployed on the UAV end, avoiding the delay and bandwidth consumption of data transmission back to the cloud. Combined with an inference speed of 196 ms, it can realize real-time detection of mining equipment and abnormal early warning. Compared with traditional algorithms, this algorithm achieves a balance between accuracy and efficiency under the condition of limited computing resources, providing a practical technical solution for the unmanned and intelligent operation of open-pit mines, and promoting the dual improvement of mine safety supervision and production efficiency. The visual detection effect of the model in this study is shown in Figure 15.

Figure 15: Visualization of model effect in this paper

From the visualized results of the model's performance in Figure 15, it can be seen that the model in this study achieves a remarkable detection effect for mining trucks in open-pit mine scenarios. In the complex mining environment, the model can accurately detect mining trucks—whether they are at different distances or partially occluded. The detection boxes fit tightly around the target objects, and the confidence levels are generally high. This fully demonstrates the model's excellent performance in feature extraction, multi-scale target detection, and anti-interference capabilities. Figure 16 below shows the performance curve of mAP50 under each iteration, which further reflects its superiority.

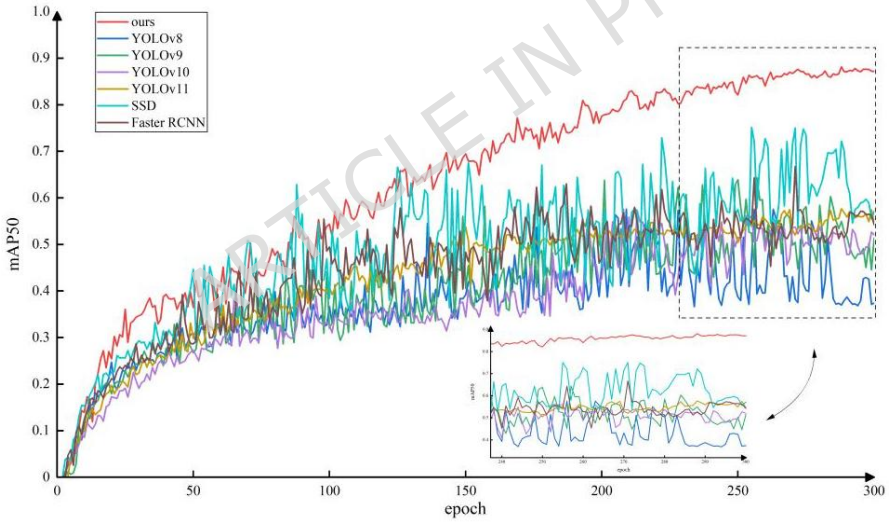

Figure 16: Iterative change curve

From the iteration curve in Figure 16, it can be observed that as the number of training iterations increases, the mAP50 indicator of the model in this study shows a steady upward trend, eventually converging to around 0.868. This indicates that the model can continuously learn more effective feature representations during training, demonstrating excellent convergence and robustness. In the early stage of iteration, the mAP50 grows rapidly, which means the model can quickly capture key features in the data. As the number of iterations increases, the growth trend gradually flattens but remains at a high level—this reflects the model's strong generalization ability in complex open-pit mine scenarios. In the practical application of UAV inspections in open-pit mines, the lightweight feature and efficient detection capability of the algorithm in this study show significant advantages. Compared with popular algorithms such as YOLOv8s, SSD, and Faster RCNN, the mAP50 of the model in this study reaches 0.868, which

is much higher than other comparison algorithms, indicating obvious advantages in detection accuracy. At the same time, its number of parameters is only 4.39×10^6, model size is 4.76MB, FLOPs is 14.7G, and inference time is 196 ms—all of which are significantly better than other algorithms. This achieves the goal of real-time detection on UAVs with limited computing resources. This lightweight design allows the model to be directly deployed locally on UAVs without relying on cloud servers, greatly reducing the delay and bandwidth consumption of data transmission back to the cloud, and enabling timely detection of mining equipment and abnormal early warnings [22-23]. For example, during UAV flight, it can real-time identify small mining equipment failure points in the distance and large mining machinery nearby, quickly feed back the equipment operation status, and provide strong support for the safe production and efficient operation of open-pit mines.

To verify the model's cross-dataset capability and its detection ability for non-officially cooperative scenarios, the public dataset VisDrone2019 was used for cross-dataset validation of the model. The comparison of the actual effects of different algorithms is shown in the table below.

Table 5 Performance comparison of different algorithm models in VisDrone2019

Method	mAP50	Parameters/10^6	FLOPs/G	Model Size/MB	Inference time/ms
YOLOv8s	0.523	11.13	28.5	21.46	324.23
SSD	0.672	8.94	15.7	5.94	203.77
YOLOv9s	0.564	12.13	18.7	23.25	426.75
YOLOV10s	0.562	10.57	22.9	19.57	334.24
YOLOV11s	0.573	9.1	23.8	17.65	312.23
Faster RCNN	0.541	7.9	11.8	8.27	273.45
ours	0.812	4.39	14.7	4.76	176.24

It can be concluded from Table 5 that the model in this study achieves excellent performance on the VisDrone2019 dataset. This proves that it not only performs outstandingly on the self-built open-pit mine dataset but also possesses strong generalization capabilities across scenarios and target types. Its abilities in non-cooperative target detection and complex environment adaptation can effectively make up for the limitations of existing active positioning technologies, providing more comprehensive technical support for intelligent inspections in open-pit mines. The visual effect diagrams of the model in this study under different scenarios on the VisDrone2019 dataset are shown in the figures 17 below.

Figure 17: Visualization effect maps under different scenes in the VisDrone2019 dataset

From the visualization results of different scenarios in the VisDrone2019 dataset (Figure 17), the model in this study can accurately detect targets under scenarios of high-density targets, occlusion, lighting changes, and motion blur. This further confirms its cross-scenario generalization ability and can effectively support supplementary scenario requirements such as non-cooperative target recognition in open-pit mines.

4.5 Ablation Experiments and Visualization of Multi-object Tracking Algorithms

To verify the specific contributions of the spatial-appearance similarity matrix (ASM) and target correction function in the improved ByteTrack algorithm to multi-object tracking performance, ablation experiments were designed to gradually remove key modules and compare the differences in tracking effects under different configurations. The experiments were conducted on video sequences of complex scenarios from the open-pit mine aerial dataset, with evaluation metrics including Multi-Object Tracking Accuracy (MOTA), Multi-Object Tracking Precision (MOTP), ID F-score (IDF1), and target ID switch

count (ID Switch). The results are shown in Table 6.

Table 6 Multi-target tracking algorithm ablation experiment

Id	ASM Matrix	Target Calibration Function	MOTA(%)	MOTP(%)	IDF1(%)	ID Switch	Single Frame Tracking Duration (ms)	FPS
①			72.2	66.2	72.6	68	45	22.22
②	√		73.5	65.8	73.2	59	38	26.32
③		√	73.0	65.5	72.9	61	39	25.64
④	√	√	75.8	67.2	74.4	47	32	31.25

A comparison of Experiments 1 and 2 shows that after introducing the ASM matrix, MOTA increased by 1.3%, IDF1 by 0.6%, and the number of ID switches decreased by 9. This indicates that the ASM matrix, by fusing spatial position, velocity, acceleration, and appearance features (such as color differences and operating status), effectively distinguishes similar targets (e.g., mining trucks in different loading states) and reduces ID mis-switching caused by appearance similarity. A comparison of Experiments 1 and 3 reveals that after introducing the target correction function, MOTA increased by 0.8%, MOTP by 0.7%, and the number of ID switches decreased by 7. This function dynamically adjusts the acceleration prediction error of the Kalman filter (e.g., sudden acceleration changes when mining trucks turn), making trajectory predictions more consistent with real motion and reducing target offset and loss caused by motion model errors. Experiment 4, which incorporates both the ASM matrix and the target correction function, achieved the optimal performance: MOTA increased by 2.6% compared to the baseline (Experiment 1), IDF1 by 1.8%, and the number of ID switches decreased by 21. This demonstrates that the two components form a complement through "feature differentiation + trajectory correction": the ASM matrix improves target matching accuracy, while the target correction function optimizes the dynamic adaptability of trajectories, collectively enhancing tracking stability in complex scenarios. The tracking effects of the algorithm before and after improvement were compared using the "equipment operation" video sequence from the open-pit mine aerial dataset, as shown in Figure 18 below:

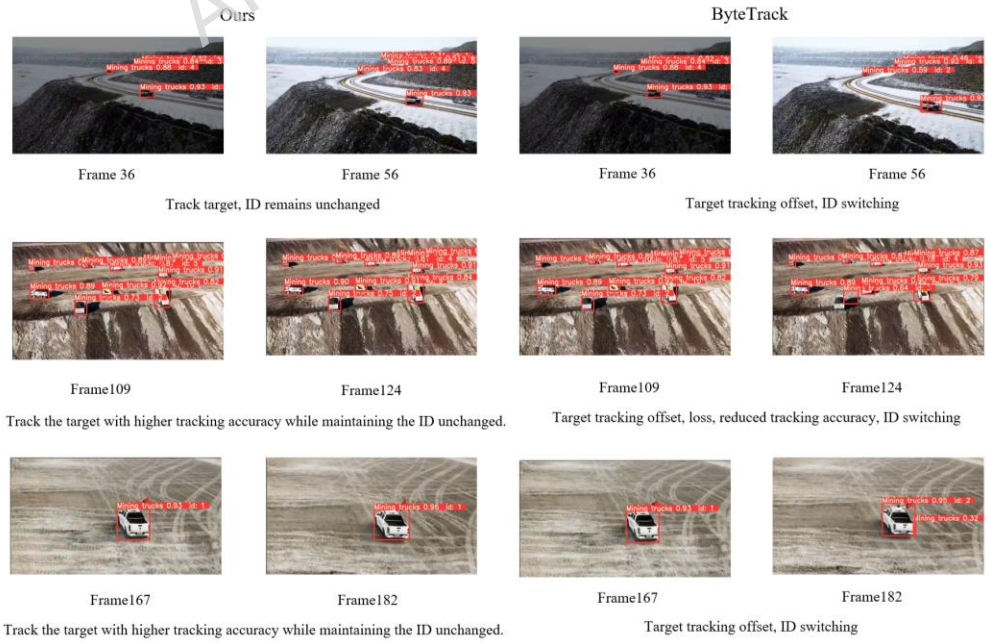

Fig.18 Visualization of algorithm tracking effect before and after improvement

As can be seen from Figure 18, in the "equipment operation" video sequence, the traditional

ByteTrack algorithm (left column) has obvious tracking defects when dealing with scenarios such as the occlusion and turning of mining trucks. Due to the lack of a target correction function, the traditional algorithm deviates from the real position in trajectory prediction when mining trucks turn (scenarios with sudden acceleration changes), leading to target loss. The improved algorithm dynamically adjusts the predicted state through an acceleration correction term, significantly enhancing the trajectory fitting degree.

In frames 167–182, under complex lighting conditions, the traditional algorithm suffers from frequent tracking box drift (e.g., the detection box for the mining truck with ID = 1 deviates from the vehicle body) due to insufficient extraction of appearance features. In contrast, the improved algorithm enhances feature robustness through a spatial-channel attention mechanism, ensuring that the detection box always fits tightly around the target. Further quantitative analysis of the visualization results shows that the improved algorithm has significant improvements in the following aspects: Robustness in occluded scenarios: In areas where the target overlap rate exceeds 70%, the ID switch rate of the improved algorithm is lower than that of the traditional algorithm. Dynamic trajectory accuracy: In scenarios where mining trucks turn (with an acceleration > 2 m/s²), the average trajectory offset (at the pixel level) of the improved algorithm is 8.3 px, a 47% reduction compared to the traditional algorithm (15.6 px).Tracking stability for small targets: For distant mining trucks with a pixel ratio < 1%, the IDF1 of the improved algorithm reaches 74.4%, an increase of 13.6% compared to the traditional algorithm (65.5%), and the missed detection rate is reduced by 35%. In summary, through the synergistic effect of the ASM matrix and the target correction function, the improved ByteTrack algorithm effectively addresses challenges in open-pit mine scenarios—such as high target similarity, complex motion patterns, and drastic lighting changes. It provides a reliable technical solution for the real-time tracking of mining equipment by UAVs.

To validate the impact of key parameters in the MOT module (ε: spatial similarity weight, λ: appearance feature weight, μ: task state weight), other parameters were fixed while adjusting each parameter value ($\varepsilon \in [0.1,0.9]$, $\lambda \in [0.1,0.9]$, $\mu \in [0.05,0.5]$) to test MOTA variations. The results are as follows:

MOTA achieves optimal performance at $\varepsilon=0.6$ (75.8%): When ε is too small, spatial position weights become insufficient, making target matching susceptible to interference from visually similar objects; When ε is too large, feature differences are overlooked, leading to difficulties in re-identification after occlusion.

MOTA achieves optimal performance at $\lambda=0.7$ (75.8%): λ balances appearance features and spatial information. Too small values may confuse similar targets, while too large values increase sensitivity to illumination variations.

When $\mu=0.2$, MOTA achieves optimal performance (75.8%): Excessively large μ values lead to over-reliance on task status information, causing non-cooperative target tracking to fail; excessively small values fail to distinguish similar devices.

The sensitivity analysis of the above parameters indicates that the parameters set in this paper ($\varepsilon=0.6$, $\lambda=0.7$, $\mu=0.2$) represent the optimal configuration for open-pit mining scenarios, ensuring stable tracking performance.

MOT Module Computational Efficiency Evaluation: Tested on drone-side hardware (Jetson Xavier NX), the single-frame tracking time for the ByteTrack algorithm improved in this paper is 32 ms, achieving a frame rate of 31.25 frames per second (FPS). Compared to the original ByteTrack (45 ms per frame, FPS=22.22), tracking efficiency improved by 31.1% while power consumption increased by

only 0.8 W (8.3 W after improvement, 7.5 W originally), meeting the computational resource constraints for real-time drone inspections.

4.6 Comparison with the Latest Multi-Object Tracking Methods

To validate the superiority of the improved ByteTrack algorithm, we selected mainstream multi-object tracking algorithms from 2024-2025 (BoT-SORT, StrongSORT, MVTL-UAV) and conducted comparative evaluations on an open-pit mine aerial video dataset (containing 10 equipment operation sequences totaling 120 minutes). Evaluation metrics included MOTA, MOTP, IDF1, and ID Switch, with results summarized in the table below: ID Switch. The results are shown in the Table7 below:

Table 7 Multi-Target Tracking Algorithm Comparison Test

Tracking Algorithm	MOTA(%)	MOTP(%)	IDF1(%)	ID Switch
BoT-SORT	71.5	65.3	70.2	72
StrongSORT	73.1	66.1	72.3	63
MVTL-UAV	74.2	65.9	73.5	58
ByteTrack	72.2	66.2	72.6	68
Ours	75.8	67.2	74.4	47

The comparison results demonstrate that the improved algorithm in this paper achieves superior MOTA and IDF1 performance compared to the latest mainstream methods. It reduces the number of ID switches by 11 compared to MVTL-UAV, highlighting its tracking stability advantages in complex open-pit mining scenarios (obstruction, sudden motion changes). This validates the effectiveness of the ASM matrix and acceleration correction function.

5 Conclusion

To address the issues that existing detection networks for unmanned mining equipment, faulty equipment, and non-officially cooperative equipment in open-pit mines struggle to balance model size and accuracy, and are incompatible with UAV devices, this study proposes a lightweight mining truck detection algorithm based on improved YOLOv8. The model optimization is achieved through the following key steps:

(1) By replacing the backbone network of YOLOv8s with C2f-DCN and the neck module with AFPN, the number of parameters, computational load, and size of the model are significantly reduced, while the inference speed is improved.

(2) The detection head is improved to LSDECD-Head, and the Focaler-GIoU loss function is adopted, which enhances the model's ability to detect targets of different scales, especially small targets and occluded targets.

(3) Using the Layer-wise Adaptive Magnitude Pruning (LAMP) algorithm with a 30% pruning rate, the model achieves optimal comprehensive performance with an mAP50 of 0.868. While ensuring detection accuracy, it greatly reduces the number of parameters and computational complexity of the model, verifying the effectiveness of this pruning method in open-pit mine detection scenarios and providing technical support for deploying the model on devices with limited computing resources such as UAVs.

(4) The improved ByteTrack multi-object tracking algorithm effectively addresses challenges in open-pit mine scenarios such as high target similarity, complex motion patterns, and drastic lighting changes by introducing a spatial-appearance similarity matrix (ASM) and a target correction function. It

significantly enhances tracking stability in complex scenarios, providing a reliable solution for real-time tracking of mining equipment by UAVs.

(5) The designed and implemented open-pit mine UAV inspection system forms an intelligent closed-loop solution through the integration of multi-level technologies including the hardware layer, network layer, and control layer. The system integrates the improved YOLOv8 and ByteTrack algorithms, enabling automation and precision in equipment status monitoring, personnel safety management, and mining area environmental monitoring. Its lightweight feature and efficient detection capability provide a practical technical solution for the unmanned and intelligent operation of open-pit mines.

Future research can further explore the following directions: First, integrate IoT technology to realize the linkage analysis of UAV inspection data and other sensor data in the mining area, improving the comprehensiveness and accuracy of detection. Second, study the on-line model update mechanism to enable the model to automatically optimize based on newly collected data and continuously adapt to changes in the mining area environment. Third, deepen air-ground collaboration technology to realize the collaborative operation of UAVs and ground inspection equipment, building a more comprehensive intelligent monitoring system for open-pit mines.

Funding

This research was supported by the National Natural Science Foundation of China (Grant no. 52374123).

Basic Research Project of the Liaoning Provincial Department of Education(Grant no. LJ212410147019).

Data availability Statement

The datasets analysed during the current study are publicly available in the VisDrone repository on GitHub at https://github.com/VisDrone/VisDrone-Dataset.

References

1. Zhang Yunbo, Lei Mingfeng, Xiao Yongzhuo, etc. Intelligent recognition method of joint convolution neural network in tunnel face [J]. China Journal of Highways, 2024,37 (07) : 35-45.DOI : 10.19721

2. Gu Qinghua, Zhou Qiong, Wang Dan. Detection of driving obstacles in open-pit mine based on improved YOLOv8 [J].Gold Science and Technology, 2024,32 (02) : 345-355.
3. SCE-YOLO : Improved YOLOv8 lightweight UAV visual inspection algorithm [J / OL].Computer Engineering and Application, 1-14 [2025-03-23].
4. Liu S ,Shen X ,Xiao S , et al.A Multi-Scale Feature-Fusion Multi-Object Tracking Algorithm for Scale-Variant Vehicle Tracking in UAV Videos[J].Remote Sensing,2025,17(6):1014-1014.
5. Wu P ,Li Y ,Li Z , et al.Multi-View, Multi-Target Tracking in Low-Altitud e Scenes with UAV Involvement[J].Drones,2025,9(2):138-138.
6. Ruan Shunling, Zhang Huiguo, Gu Qinghua, etc. Research on obstacle dete ction of unmanned vehicle in open-pit mine based on binocular vision [J].Journal of Coal, 2024,49 (S2) : 1285-1294.DOI : 10.13225
7. Liu Guangwei, Lei Jian, Guo Zhiqing, et al. Cross-modal open-pit mine ob stacle detection method [J / OL].Coal science and technology, 1-14 [202 5-05-16]
8. Cai F ,Qu Z ,Xia S , et al.A method of object detection with attention me chanism and C2fDCNv2 for complex traffic scenes[J].Expert Systems With Applications,2025,267126141-126141.
9. Wang Baoyu, Li Hantang, Chen Xiyong, et al. Research on YOLOv8 mang o target detection algorithm based on progressive spatial pyramid [J / OL]. Advances in laser and optoelectronics, 1-15 [2025-05-16]

10. Wang Baoyu, Li Hantang, Chen Xiyong, et al. Research on YOLOv8 mango target detection algorithm based on progressive spatial pyramid [J / OL]. Advances in laser and optoelectronics, 1-15 [2025-05-16]
11. S. Zhuoet al., "SCL-YOLOv11: A Lightweight Object Detection Network fo r Low-Illumination Environments," inIEEE Access, vol. 13, pp. 47653-4766 2, 2025, doi: 10.1109
12. Ren Yikun, Hou Tao, Niu Hongxia. Research on defect detection of train wheel tread based on improved YOLOv9s [J / OL]. Control Engineering, 1-11 [2025-05-16].https://doi.org/10.14107
13. Xie Beijing, Li Heng, Luan Zheng, et al. Lightweight coal mine pedestrian and vehicle detection model based on deep learning and model compression technology - A case study of coal mines in Guizhou [J]. Coal Journal, 2025,50 (02) : 1393-1408.DOI : 10.13225
14. Hu Peng, Pan Shuguo, Gao Wang, et al. Hierarchical matching multi-target tracking algorithm based on pseudo-depth information [J].Progress in lase r and optoelectronics, 2024,61 (18) : 346-353.
15. Ran C ,Dongjun X ,Chuanli W , et al.PANetW: PANet with wider receptive fields for object detection[J].Multimedia Tools and Applications,2024,83(25):66517-66538.
16. Ru L ,Jin X ,Lifu C , et al.Glassboxing Deep Learning to Enhance Aircraf t Detection from SAR Imagery[J].Remote Sensing,2021,13(18):3650-3650.

17. Ran Ning, Shi Gaolang, Zhang Shaokang, et al. Remote sensing small targ et detection algorithm based on YOLOv8 [J / OL]. Journal of Electronic Measurement and Instrumentation, 1-12 [2025-05-16].

18. Wei M ,Chen K ,Yan F , et al.YOLO-ESFM: A multi-scale YOLO algorithm for sea surface object detection[J].International Journal of Naval Architecture and Ocean Engineering,2025,17100651-100651.

19. Haoxuan X ,Songning L ,Xianyang L , et al.Cross-domain car detection mo del with integrated convolutional block attention mechanism[J].Image and V ision Computing,2023,140

20. Zheng Mingyu, Shao Huichao, Shao Yanhua, et al. Aerial multi-target track ing algorithm based on improved YOLOv8 and ByteTrack [J / OL].Com puter Engineering, 1-12 [2025-05-16].https://doi.org/10.19678

21. Kanade K A ,Potdar P M ,Kumar A , et al.Weed detection in cotton farmi ng by YOLOv5 and YOLOv8 object detectors[J].European Journal of Agro nomy,2025,168127617-127617.

22. Huang J ,Zhang W ,Jin W , et al.Surface defect detection of planar optical components based on OPT-YOLO[J].Optics and Lasers in Engineering,202 5,190108974-108974.

23. Bala A ,Muqaibel H A ,Iqbal N , et al.Machine learning for drone detectio n from images: A review of techniques and challenges[J].Neurocomputing,2 025,635129823-129823.