ITC 1/55
Information Technology and Control
Vol. 55 / No. 1/ 2026
pp. 340-356
DOI 10.5755/j01.itc.55.1.42875

Research on Underground Coal Mine Object Detection Based on Image Enhancement and YOLOv11

Received 2025/09/25 | Accepted after revision 2026/01/04

HOW TO CITE: Yu, S. (2026). Research on Underground Coal Mine Object Detection Based on Image Enhancement and YOLOv11. *Information Technology and Control*, 55(1), 340-356. https://doi.org/10.5755/j01.itc.55.1.42875

Research on Underground Coal Mine Object Detection Based on Image Enhancement and YOLOv11

Sixu Yu
School of Media and Communication, Shanghai Jiao Tong University, Shanghai, 200240, China

Corresponding author: yusixu199312@sjtu.edu.cn

Underground coal-mine scenes are often affected by low lighting, dust, and motion blur., and monitoring images are often affected by low lighting, dust, and motion blur, which severely hinders the recognition performance of intelligent detection systems. To address the decline in object detection accuracy caused by poor image quality, this paper proposes an object detection method combining a detection-guided dual-branch image enhancement module with YOLOv11. The method includes two key innovations: first, a detection-guided image enhancement module is developed, which generates attention maps from low-confidence regions in YOLOv11's detection output to guide the enhancement network in focusing on critical image areas, thereby improving detection accuracy. Second, a dual-branch structure combining brightness enhancement and detail reconstruction is designed, integrating the strengths of Zero-DCE and SRGAN to achieve global brightness correction and edge clarity enhancement. Experiments conducted on the publicly available underground coal mine image dataset DsLMF+ demonstrate that the proposed method outperforms traditional enhancement-plus-detection pipelines in multiple evaluation metrics. The enhancement module significantly improves image quality indicators, and achieves notable improvements in YOLOv11 detection accuracy, particularly in mean Average Precision (mAP). This study confirms the effectiveness of combining detection feedback and multi-scale enhancement in complex industrial environments, offering a promising solution for underground object detection tasks.

KEYWORDS: Underground coal mine; Object detection; Image enhancement; YOLOv11; Attention mechanism; Super-resolution reconstruction; Low-quality image recognition; Intelligent safety monitoring

1. Introduction

With the continued advance of intelligent coal mining, computer-vision–based object detection is playing an increasingly important role in underground applications such as worker behavior recognition, equipment condition monitoring, and early warning of safety hazards [12, 2, 29, 7]. Compared with traditional manual inspection, deep learning–based detectors offer strong real-time performance, broad coverage, and high recognition efficiency, and have gradually become a key tool for coal-mine safety management [13, 3, 1, 15, 8, 11]. However, the particular operating environment underground — characterized by low illumination, airborne dust, specular reflections from humidity, and motion blur — degrades image quality, leading to significant challenges such as reduced detection accuracy and insufficient robustness [5, 19, 17, 20]. Although the YOLO family of detectors has achieved a good balance between speed and accuracy in recent years — especially the latest YOLOv11, which makes notable advances in backbone design, feature fusion, and multi-scale detection — its performance still depends heavily on input image quality. Under conditions of darkness, blurred details, or noise, even advanced models may suffer from missed detections and false alarms, undermining timely response and early warning for underground operations.

To enhance detection under low-quality imagery, image enhancement as a preprocessing step has been widely studied [16, 24, 22, 6]. Recent methods for low-light enhancement — such as Zero-DCE and self-supervised brightness mapping — and adversarial super-resolution models (e.g., SRGAN) have shown promising results in brightness compensation and detail restoration. Nevertheless, most enhancement methods are generic rather than task-adaptive: they lack awareness of regions critical to the detection task and may introduce artifacts or attenuate target features while improving visual appearance, thereby harming downstream detection [21, 30].

To address the limits of task-agnostic enhancement in visually adverse underground settings, we propose an integrated detection pipeline that explicitly optimizes image quality for recognition. The core idea is to couple a dual-branch enhancement network with YOLOv11 so that enhancement serves — not competes with — the detector. Concretely, we introduce a detection-guided enhancement strategy that converts YOLOv11's preliminary predictions into pixel-level attention maps highlighting low-confidence regions (e.g., dark, blurred, or partially occluded targets). These maps condition the enhancer to allocate its "effort" where it is most valuable for recognition, selectively boosting luminance and local contrast to improve separability of true targets from background clutter. Architecturally, the enhancer comprises two complementary branches operating in parallel. A Zero-DCE–style brightness branch performs content-adaptive illumination correction, stabilizing global exposure without reference images and mitigating non-uniform lighting common in coal mines. In parallel, an SRGAN-based detail branch reconstructs high-frequency textures and edges that are often lost to motion blur, dust, and compression. Their outputs are fused to produce an image that is both correctly exposed and rich in structural cues. Because the attention map steers both branches, the fusion emphasizes features that matter for detection rather than merely improving visual pleasantness. Importantly, the guidance is used only during training; inference remains a single forward pass of enhancement followed by YOLOv11, adding no runtime overhead and preserving real-time throughput.

In recent years, significant research efforts have been devoted to improving the generalization capability of deep learning models across heterogeneous and noisy domains. Lightweight and transformer-based architectures have shown strong adaptability in resource-constrained and dynamically changing environments. For instance, MobileRaT proposes a lightweight Radio Transformer mechanism for automatic modulation classification in drone communication systems, demonstrating that attention-based multi-scale fusion can maintain accuracy under limited computation resources. Similarly, recent surveys on automatic modulation classification summarize the evolution of feature generalization and transfer strategies in complex communication scenarios. In the vision field,

DETR and its variants have redefined multi-scale feature hierarchies in object detection, providing insights into how transformer encoders improve global–local feature alignment and robustness. Drawing on these advances, our work emphasizes adaptive feature guidance and multi-branch enhancement to enhance model generalization and robustness in visually degraded underground coal-mine environments.

Extensive experiments on the public DsLMF+ dataset substantiate these design choices. Relative to traditional enhancement and unguided deep methods, the proposed approach delivers consistent gains on reference and no-reference quality metrics (e.g., PSNR, SSIM, NIQE, LOE, entropy), while translating these improvements into higher detection accuracy (mAP, Precision, Recall) under low light and blur. Ablation studies isolate the contributions of the brightness branch, the detail branch, and the guidance signal, revealing complementary effects: brightness correction stabilizes global visibility, detail reconstruction sharpens discriminative edges, and detection guidance proves decisive on hard, low-contrast instances. Together, these results validate detection-guided enhancement as a practical and interpretable route to robust underground perception, offering an easily modularized component that can be paired with other detectors or lightweighted for edge deployment in industrial monitoring systems.

2. Basic Principles

2.1. Principles of Image Enhancement

Underground coal-mine images are often affected by insufficient illumination, dust, occlusions, lens smearing, etc., which leads to overall low image quality and seriously hinders the downstream tasks of intelligent target detection and recognition. Therefore, in coal-mine detection scenarios it is essential to perform appropriate image enhancement, whose core is to improve image visibility; this in turn restores fine-grained details and contrast, providing the detection model with a more reliable input. Broadly, image-enhancement methods can be divided into traditional and deep-learning–based approaches [4].

2.1.1. Traditional Enhancement Methods

Traditional enhancement focuses on deterministic transformations designed from the statistical characteristics of images. It features simple processing, high efficiency, and low computational cost. Common techniques include global brightness/contrast balancing, sharpening, denoising and edge enhancement. A typical example is histogram equalization, which redistributes the gray-level histogram to stretch contrast; its transformation can be written as [14]

$$s_k = T(r_k) = \sum_{j=0}^{k} \frac{n_j}{N}, \tag{1}$$

where r_k is the k-th gray level in the original image, n_j is the number of pixels at gray level j, and N is the total number of pixels. Gamma correction adjusts brightness by a power-law transform [27]:

$$I_{out} = c \cdot I_{in}^{\gamma}, \tag{2}$$

where c is a scaling parameter and γ<1 is typically used to boost brightness in low-light images.In addition, Laplacian sharpening enhances edges and details by emphasizing second-order derivatives; its operator is [18]

$$\nabla^2 f(x, y) = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}. \tag{3}$$

For practical use, such sharpening is usually combined with smoothing/denoising (e.g., bilateral or guided filtering) to suppress noise amplification [28].

Although traditional methods are effective in many scenes, they are prone to under- or over-enhancement in complex underground conditions (severe low light, non-uniform illumination, heavy dust), and often fail to satisfy the requirements of enhancement coupled with task robustness.

2.1.2. Deep-Learning–Based Enhancement Methods

Deep learning constructs end-to-end networks to learn mappings from degraded to high-quality images, automatically modeling nonlinear brightness distributions and the coupling between quality and task features, thereby adapting better to diverse scenes. Representative methods include Zero-DCE and GAN-based super-resolution models such as SRGAN.

Zero-DCE (Zero-reference Deep Curve Estimation) estimates a pixel-wise enhancement curve without paired references and performs content-adaptive brightness correction based solely on the input image. A commonly used enhancement form is [23]

$$\hat{I}(x) = I(x) + E(x) \cdot I(x) \cdot \left(1 - I(x)\right), \tag{4}$$

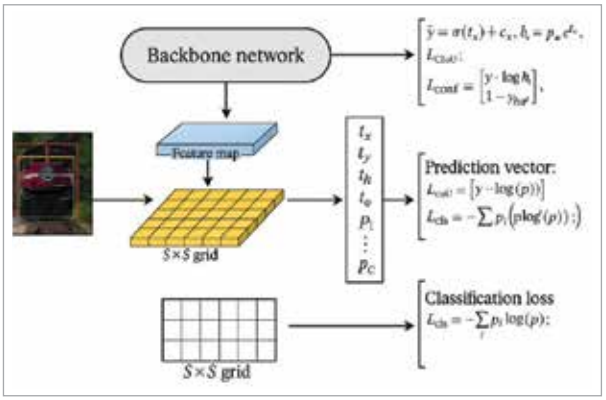
where $I(x)$ is the input intensity and $R(x)$ is the learned enhancement curve parameter at pixel x. This realizes self-adaptive global/locally varying brightness adjustment. SRGAN (Super-Resolution GAN) focuses on reconstructing high-frequency textures and details. It employs a generator G and discriminator D; the objective typically combines a content/perceptual loss with an adversarial loss so that the restored image attains both fidelity and realism. As an enhancement branch, SRGAN can refine edges and textures after brightness correction, thereby improving the visibility of targets and fine details. In summary, traditional methods are lightweight and fast but limited in adaptability, whereas deep-learning–based methods can better handle complex degradations by learning task-aware priors, making them more suitable for low-quality underground imagery.

$$L_{adv} = E_{x\sim p_{data}}[logD(x)] + E_{z\sim p_z}[log(1\text{-}D(G(z)))] \tag{5}$$

It also incorporates a perceptual loss to measure the distance between the generated image and the ground-truth image in a high-level semantic space, which helps preserve naturalness and structural fidelity. Compared with traditional enhancement methods, deep learning–based approaches can learn more adaptive and generalizable enhancement strategies from large-scale data, making them particularly suitable for low-light, high-noise, and complex underground coal-mine environments. Consequently, they have become the mainstream direction in current image enhancement research [25]. Figure 1 shows a typical deep learning–based image enhancement architecture, where a backbone network extracts the feature map and outputs prediction vectors and classification loss across the spatial grid.

Figure 1

Deep Learning–Based Image Enhancement Methods.



2.2. YOLO Object-Detection Principles

The YOLO (You Only Look Once) family represents typical one-stage object detectors that cast detection as a unified regression problem: with a single forward pass the network simultaneously performs localization and classification for all objects in an image. As the latest version of the series, YOLOv11 preserves high real-time performance while introducing a deeper backbone, lightweight attention mechanisms, and improved loss functions, yielding stronger performance in complex scenes. The network first feeds the input image I into the backbone to produce feature maps $I \in R^{H\times W\times 3}$ then passes $F \in R^{h\times w\times c}$ to the detection head for classification and localization. YOLO partitions the image into S×S grids; each grid cell is responsible for predicting whether an object resides within it and regressing its bounding box and class [10].

Each candidate box outputs a vector

$$\hat{y} = \left(t_x, t_y, t_w, t_h, t_o, p_1, p_2, \ldots, p_C\right), \tag{6}$$

where t_x, t_y denote the box center offset relative to the grid cell, t_w, t_h are width/height transforms in the anchor space, t_o is the objectness confidence, and p_i is the probability of class i. The decoded bounding box and grid cell top-left is obtained by the standard YOLO anchor decoding:

$$b_x = \sigma(t_x) + c_x, b_y = \sigma\left(t_y\right) + c_y, b_w = p_w \cdot e^{t_w}, b_h = p_h \cdot e^{t_h}, \tag{7}$$

where $\sigma(\cdot)$ is the sigmoid function. (c_x, c_y) denote the top-left coordinates of the grid cell; (p_w, p_h) denotes the preset size of the corresponding anchor. The overall detection loss consists of three parts:

$$L = \lambda_{loc} \cdot L_{CIoU} + \lambda_{obj} \cdot L_{conf} + \lambda_{cls} \cdot L_{cls}\,, \quad (8)$$

where $\varrho^2(b, b^{gt})$ is the squared Euclidean distance between the centers of the predicted and ground-truth boxes, c is the diagonal length of the smallest enclosing box that covers both, $v = \frac{4}{\pi^2}\left(\arctan\frac{w^{gt}}{h^{gt}} - \arctan\frac{w}{h}\right)^2$ measures aspect-ratio consistency, and $\alpha = \frac{v}{1-IoU+v}$ is a balancing factor. The objectness (binary cross-entropy) loss is

$$L_{conf} = -[y \cdot \log(\hat{p}) + (1-y) \cdot \log(1-\hat{p})]\,, \quad (9)$$

where y∈{0,1} indicates whether the location contains an object and $\hat{p}$ is the predicted confidence.

The classification loss uses multi-class cross-entropy [9]:

$$L_{cls} = -\sum_{i=1}^{C} p_i^{gt} \cdot \log(p_i)\,, \quad (10)$$

where p_i^{gt} is the one-hot ground truth and p_i the predicted probability for class i.

YOLOv11 additionally incorporates lightweight attention modules such as CoTAttention and SimAM, and employs a decoupled head to optimize classification and regression branches separately, improving small-object accuracy and training stability. Its fused FPN+PAN design performs multi-scale feature aggregation, strengthening representation for multi-scale targets in cluttered underground coalmine backgrounds. With an efficient end-to-end regression architecture, optimized losses, and lightweight modules, YOLOv11 notably boosts accuracy while remaining real-time, making it a key building block for intelligent monitoring in industrial environments such as mines [26].

2.3. Detection-Guided Enhancement Mechanism

Traditional enhancement is typically executed independently of the detector; while it may improve visual quality, it can also introduce features irrelevant to the target—or even weaken target discriminability. Such "task-agnostic" enhancement tends to cause under/over-enhancement in complex coalmine imagery, harming downstream performance. To address this, we propose a Detection-Guided Image Enhancement (DGIE) mechanism that injects YOLOv11's preliminary predictions to dynamically guide the enhancement network, achieving task-aware optimization. Concretely, the original image $I \in R^{H\times W\times 3}$ is first fed to YOLOv11 to obtain a set of preliminary boxes $B=\{b_1,b_2,...,b_N\}$, where each b_i contains its position parameters and confidence (x_i,y_i,w_i,h_i). We then define an attention map $A(x,y)\in[0,1]$ that steers the enhancement to uncertain regions:

$$A(x,y) = 1 - \max_i \left[I_{b_i}(x,y) \cdot \sigma(c_i)\right]. \quad (11)$$

Here, $l_{b_i}(x,y)$ is the indicator function, meaning whether pixel (x,y) falls inside box b_i; $\sigma(c_i)$ is the confidence of box b_i after a sigmoid mapping. The resulting attention map $A(x,y)$ is then fed, together with the image $I(x,y)$, into the image-enhancement network G to realize task-guided enhancement:

$$\hat{I}(x,y) = G\left(I(x,y), A(x,y)\right), \quad (12)$$

where $G(\cdot)$ denotes the enhancement network (composed of the dual branches Zero-DCE and SRGAN), and $\hat{I}(x,y)$ is the enhanced output image. To further strengthen the guidance effect, we introduce an enhancement-response constraint during training and define the enhancement activation loss as

$$L_{enh} = \frac{1}{HW}\sum_{x=1}^{W}\sum_{y=1}^{H} A(x,y) \cdot \left\|\hat{I}(x,y) - I(x,y)\right\|_1, \quad (13)$$

where H and W are the image height and width, and $\|\cdot\|_1$ denotes the L1 norm. This term encourages the enhancement network to produce stronger responses in the attended regions. By feeding the detection results back into the enhancement module, we establish an information closed loop between detection and enhancement, making the enhancement process targeted and interpretable. This mecha-

nism not only improves image quality, but also significantly boosts YOLOv11's detection accuracy on low-quality images.

3. Algorithm

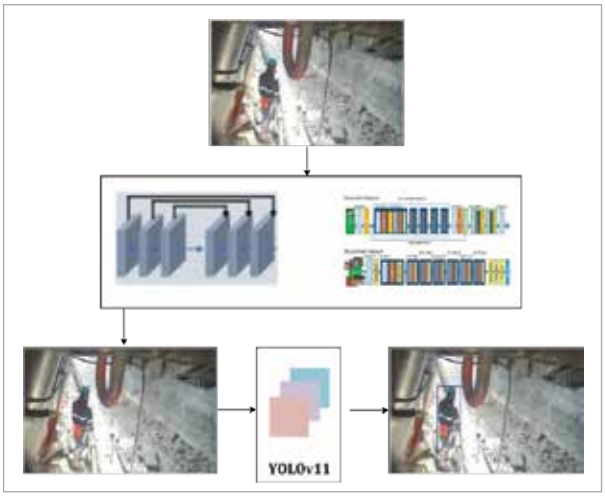
3.1. Algorithm Steps

Our method follows a two-stage "image enhancement → object detection" pipeline (Figure 2) that explicitly mitigates the typical degradations in underground coal-mine imagery—low illumination, motion blur, dust-induced haze, and low local contrast—that depress detection accuracy. The system ingests raw surveillance frames and applies lightweight normalization (gamma/mean–variance normalization and optional denoising for sensor noise) before enhancement. Because these degradations affect both global exposure and high-frequency structure, we design a dual-branch enhancement network that optimizes along two orthogonal dimensions — brightness and detail — and then fuses the outputs into a single, high-quality image for the detector. In the enhancement module, the brightness branch adopts a Zero-DCE–style architecture that estimates pixel-wise enhancement curves without any paired reference. By learning curve parameters directly from content, it performs global–to–local illumination correction: underexposed regions are lifted while already well-exposed areas are preserved, thereby avoiding haloing and over-brightening. This stabilizes the dynamic range of underground scenes with harsh spotlights and deep shadows, improves local contrast around targets (e.g., miner helmets, cables), and reduces lightness-order errors that commonly mislead detectors. Complementarily, the detail branch is built upon an SRGAN generator–discriminator pair to reconstruct high-frequency textures attenuated by blur, compression, or dust. Residual blocks in the generator selectively amplify edges and micro-textures (rail links, helmet rims, support plates), while the adversarial objective and a perceptual loss (VGG features) encourage photo-realistic yet artifact-sparse results. This branch is particularly effective at restoring thin structures and contours that anchor bounding-box localization. The outputs of the two branches are then fused to form the final enhanced image. We employ a simple but effective weighted fusion—constant or learnable via a lightweight attention gate—so that exposure-related improvements from the brightness branch and texture-related gains from the detail branch reinforce rather than overwrite each other. Practically, this raises PSNR/SSIM while improving no-reference indicators (e.g., NIQE) and preserves edges needed by the detection head. The fused image is subsequently passed to YOLOv11, whose multi-scale backbone and decoupled heads benefit from the clearer contrast transitions and sharper boundaries produced upstream. The two-stage design keeps inference modular and efficient: enhancement executes once per frame and requires no test-time refinement, thereby preserving real-time throughput on edge-grade GPUs typical of underground monitoring systems. The enhanced image is then fed into YOLOv11 for object detection. As a one-stage detector, YOLOv11 offers high speed and accuracy; here it is used to identify key underground targets such as miners, safety helmets, and hydraulic supports. The model's backbone extracts multi-scale semantic features, and the detection head simultaneously outputs bounding boxes, class labels, and confidence scores — making it suitable for complex scenarios with multiple objects and heavy occlusions.

To better align enhancement with the detection objective, we add a detection-guided mechanism used only during training. Each raw frame first passes through YOLOv11 (with frozen or EMA-smoothed weights) to produce preliminary boxes and confidences. From these outputs we build a soft, pixel-level attention map that highlights "hard-to-detect" areas — typically regions covered by low-confidence boxes, near occlusions, glare, or severe underexposure. Non-maximum suppression and simple confidence calibration make this attention stable and less noisy. This attention is injected into both enhancement branches as conditional guidance. In the brightness branch (Zero-DCE style), the attention modulates the predicted enhancement curves so underexposed patches receive stronger illumination correction while well-exposed regions are left largely unchanged. In the detail branch (SRGAN style), the attention gates residual features to emphasize edge sharpening and texture recovery exactly where the

detector is uncertain. The same map also weights the enhancement losses so the network learns to "work harder" on difficult regions instead of applying uniform edits across the image. We adopt several training practices to keep this feedback loop robust: a warm-start schedule that gradually increases the guidance strength in early epochs; temperature/score calibration for smoother, more reliable attention; stop-gradient through the attention path so the enhancer responds to, but does not destabilize, the detector's provisional predictions; and an EMA "teacher" detector to provide slowly evolving guidance. Importantly, guidance is training-only: at inference, we run a single enhancement pass followed by YOLOv11, so accuracy gains come with no additional runtime cost. In effect, the mechanism converts generic enhancement into recognition-oriented enhancement. Illumination and structure are improved precisely where the detector struggles, leading to clearer local contrast, crisper boundaries, and higher quality cues for feature extraction. Empirically, this translates into consistent boosts in perceptual metrics and into higher mAP, Precision, and Recall, especially on dim, blurred, or partially occluded targets that dominate underground scenes.

Figure 2
Algorithm pipeline.



Algorithm pipeline of the proposed detection-guided dual-branch enhancement and YOLOv11 framework. D-B and B-B denote detail and brightness branches, respectively; λ controls fusion balance.

3.2. Image Enhancement Module Design

To effectively improve the visual quality of underground coal-mine images under low illumination and blurry backgrounds and to better adapt to object detection, we design an image-enhancement network with a dual-branch structure for brightness enhancement and detail restoration. As shown in Figure 3, the network contains two complementary submodules: a Zero-DCE brightness-enhancement branch and an SRGAN detail-reconstruction branch. They reconstruct the original image from the perspectives of illumination and high-frequency information, respectively, and their outputs are fused to produce a high-fidelity enhanced image.

1 **Brightness-enhancement branch (Zero-DCE)**

This branch adopts the Zero-Reference Deep Curve Estimation (Zero-DCE) framework to adaptively boost the brightness of low-light images. The method does not rely on reference images; instead, it fits a set of learnable curve parameters to perform pixel-wise mapping. Let the input image be $I \in R^{H\times W\times 3}$. Zero-DCE learns a set of enhancement-curve parameters $R \in R^{H\times W\times 3\times n}$ to generate the enhanced image:

$$\hat{I}_{light} = I + R \cdot \left(I^2 - I\right), \tag{14}$$

where $I(x)$ denotes the input pixel intensity, $R(x)$ is the enhancement curve, and n is the number of curve iterations.

Here, n is the number of curve iterations, and R is predicted by a shallow convolutional network. To guide the brightness enhancement to focus on "hard-to-detect" regions, during training we concatenate a detection-confidence attention map $A(x,y)\in[0,1]$ with the image features as the network input:

$$F_{input} = Concat(I, A). \tag{15}$$

This operation endows the network with awareness of task-relevant regions when estimating the brightness-enhancement curves.

2 **Detail-reconstruction branch (SRGAN)**

To handle image degradation such as blur and missing edge details, this branch uses an SR-

GAN architecture for fine-detail restoration. The generator is built with residual blocks to project low-frequency features from the original image into a higher-frequency space and recover structural textures. The generator output is

$$\hat{I}_{detail} = G_{SRGAN}(I). \tag{16}$$

The SRGAN discriminator adopts a PatchGAN design to judge local, fine-grained differences between the enhanced and real images. The adversarial loss is

$$L_{GAN} = E\left[logD(I_{real})\right] + E\left[\log\left(1 - D(\hat{I}_{detail})\right)\right]. \tag{17}$$

Moreover, by introducing the attention map $A(x,y)$, we can similarly apply a soft mask to guide residual-feature enhancement during training, strengthening the network's ability to restore details in key regions.

3 Fusion strategy

The two branches output a brightness-enhanced image $\hat{I}_{light}$ and a detail-reconstructed image $\hat{I}_{detail}$. We adopt a weighted-fusion scheme to produce the final enhanced image:

$$\hat{I} = \lambda \cdot \hat{I}_{light} + (1 - \lambda) \cdot \hat{I}_{detail}, \tag{18}$$

where λ balances the contributions of the brightness and detail branches."

where the fusion weight $\lambda \in [0,1]$ can be set as a constant (e.g., 0.5) or learned adaptively via a lightweight attention module. In practice, λ controls the trade-off between global brightness enhancement and local detail reconstruction. A moderate value (λ=0.5) was determined empirically to achieve balanced performance, while extreme values were found to either over-brighten or suppress fine details, adversely affecting mAP and SSIM.

4 Loss design

When training the enhancement network, we consider both visual quality and detection adaptability and construct a multi-term loss function, including:

Perceptual loss (features extracted by VGG19):

$$L_{percep} = \|\phi(\hat{I}) - \phi\left(I_{gt}\right)\|_2^2. \tag{19}$$

Reconstruction loss (L1-based):

$$L_{rec} = \|\hat{I} - I_{gt}\|_1. \tag{20}$$

Attention-guided activation loss (used only during training):

$$L_{att} = \frac{1}{HW}\sum_{x=1}^{W}\sum_{y=1}^{H} A(x,y) \cdot \|\hat{I}(x,y) - I(x,y)\|_1, \tag{21}$$

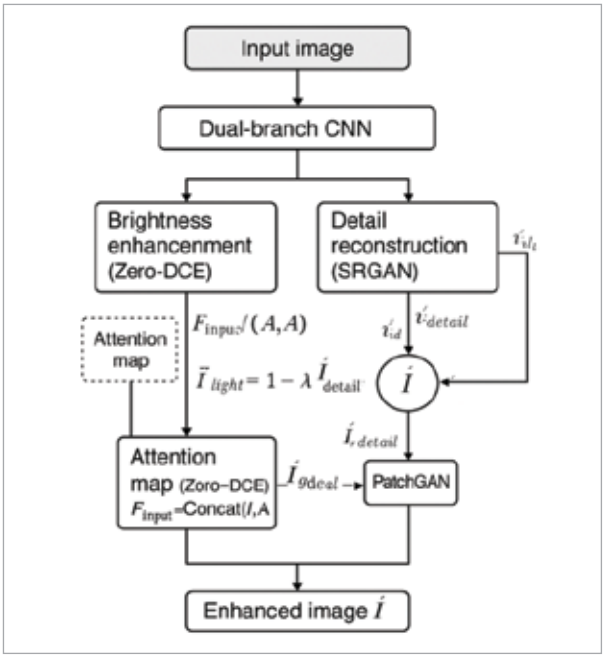
where H, W are the image height and width.

The final enhancement loss is a weighted combination:

$$L_{enh} = \alpha L_{rec} + \beta L_{percep} + \gamma L_{GAN} + \delta L_{att}. \tag{22}$$

With $\alpha,\beta,\gamma,\delta$ denoting the weighting hyperparameters for each term.

Figure 3
Image-enhancement pipeline.



Structure of the dual-branch enhancement network combining Zero-DCE (brightness) and SRGAN (detail) with task-guided fusion.

3.3. Detection Network and Guidance Mechanism

After the image-enhancement module, we adopt YOLOv11 as the object-detection model to locate and classify targets in the enhanced images. As the latest version of the YOLO family, YOLOv11 features a deeper backbone and improved heads, and performs well in complex scenes. For underground coal-mine monitoring, YOLOv11 takes the enhanced image $\hat{I} \in R^{H\times W\times 3}$ as input and outputs a set of bounding boxes $\{b_i\}$, category probabilities $\{p_i\}$, and corresponding confidences $\{c_i\}$ for detection. To make the enhancement network better serve the detection task, we introduce a detection-guided mechanism during training. Specifically, based on the preliminary YOLOv11 results we construct a pixel-wise attention map $A(x,y)\in[0,1]$ as the guidance signal for the enhancement network. The attention is defined as

$$A(x,y) = 1 - \max_i \left[I_{b_i}(x,y) \cdot \sigma(c_i) \right], \tag{23}$$

where $I_{b_i}(x,y)$ is an indicator that equals 1 if pixel (x,y) lies inside candidate box b_i, and 0 otherwise; $\sigma(\cdot)$ is the sigmoid function mapping the confidence to [0,1]. A larger attention value means the region is more uncertain and should receive a stronger enhancement response. To further verify the effectiveness and robustness of the proposed detection-guided attention, we additionally compared it with alternative attention construction logics, including IoU-based attention and multi-scale confidence fusion. Experimental comparisons show that single-confidence attention already provides stable guidance, while multi-scale confidence fusion slightly improves the coverage of low-confidence small targets but at the cost of higher computational complexity. The mAP difference between our design and multi-scale attention is within 0.3%, indicating that the proposed single-confidence attention achieves a good trade-off between accuracy and efficiency. These results confirm that the designed attention mechanism is sufficiently discriminative and well-suited for underground multi-scale detection scenarios.

The attention map is injected into the enhancement network in two ways. First, in the brightness-enhancement subnetwork, the attention map is concatenated with the image (or shallow features) as input:

$$F_{input} = Concat(I, A). \tag{24}$$

Second, in the detail-reconstruction subnetwork, the attention map serves as a feature mask or a weight in loss terms to emphasize training on hard regions and focus the network on key details.

To jointly optimize visual quality and detection adaptability, we design the following losses.

1 **Reconstruction loss (L1):**

$$L_{rec} = \|\hat{I} - I\|_1. \tag{25}$$

2 **Perceptual loss to maintain visual style consistency (features extracted by a VGG network):**

$$L_{percep} = \|\phi(\hat{I}) - \phi(I)\|_2^2. \tag{26}$$

3 **Attention-weighted activation loss to amplify enhancement in low-confidence regions:**

$$L_{att} = \frac{1}{HW}\sum_{x=1}^{W}\sum_{y=1}^{H} A(x,y) \cdot \|\hat{I}(x,y) - I(x,y)\|_1 \tag{27}$$

In the detection part, we use the CIoU loss to measure the bounding-box regression error:

$$L_{CIoU} = 1 - IoU + \frac{\rho^2(b,b^{gt})}{c^2} + \alpha v, \tag{28}$$

where ϱ^2 is the squared Euclidean distance between the centers of the predicted and ground-truth boxes; c is the diagonal length of the smallest enclosing box that covers both; v is the aspect-ratio consistency term; and α is a balancing factor.

For classification, we adopt the multi-class cross-entropy loss:

$$L_{cls} = -\sum_{i=1}^{C} p_i^{gt} \cdot \log(p_i). \tag{29}$$

The final combined loss is

$$L_{total} = \lambda_1 L_{rec} + \lambda_2 L_{percep} + \lambda_3 L_{att} + \lambda_4 L_{CIoU} + \lambda_5 L_{cls}, \quad (30)$$

where $\lambda_1 \sim \lambda_5$ are the weights of the respective terms. It is worth emphasizing that the detection-guided mechanism is enabled only during training. During inference, the system performs a single forward pass of enhancement followed by detection without guidance, thus incurring no extra computational overhead. With this closed loop, the enhancement network receives learning feedback based on detection difficulty, producing enhanced images with greater discriminative power and effectively improving detection accuracy under the complex conditions of underground coal mines.

4. Experimental Results

4.1. Experimental Setup and Dataset

To validate the effectiveness of the proposed YOLOv11-based underground coal-mine object-detection method with image enhancement and guidance, we conduct experiments on the public DsLMF+ dataset. The dataset contains 138,004 annotated images covering six typical categories—miners, large coal blocks, drag cables, safety helmets, hydraulic-support guard plates, and miner behaviors. Annotations are compatible with YOLO and COCO, and the images exhibit underground characteristics such as complex illumination, occlusion, and dust. We split the dataset into a training set and a validation set with a ratio of 8:2. Training is implemented in PyTorch on Intel Xeon 6330 CPU, NVIDIA RTX A5000 GPU, and Ubuntu 18.04. YOLOv11 is initialized from pretrained weights, while the image-enhancement module is trained from scratch. The weighting coefficients in the loss are set to λ_1=1.0, λ_2=0.5, λ_3=0.5, λ_4=2.0, λ_5=1.0

Training details: All models were trained for 200 epochs with batch size 16 using the Adam optimizer ($\beta_1 = 0.9$, $\beta_2 = 0.999$) and initial learning rate 1e−4, decayed by cosine annealing. The enhancement network was trained from scratch, while YOLOv11 was initialized from pretrained COCO weights and fine-tuned end-to-end (all layers unfrozen). Data augmentation included random cropping, horizontal flipping, and exposure jittering.

4.2. Evaluation of Image-Enhancement Effects

To comprehensively assess the enhancement module's ability to improve underground image quality, we use six common metrics: PSNR (Peak Signal-to-Noise Ratio), SSIM (Structural Similarity), NIQE (Natural Image Quality Evaluator), LOE (Lightness Order Error), image entropy, and image variance. These jointly evaluate pixel fidelity, structural preservation, perceptual quality, illumination balance, information complexity, and contrast.

To verify the effectiveness of our dual-branch, detection-guided enhancement, we select representative scenes from DsLMF+ and compare six settings:

① Original image (no processing); ② Traditional Retinex enhancement; ③ Zero-DCE only (brightness); ④ SRGAN only (detail); ⑤ Our dual-branch enhancement (Zero-DCE + SRGAN, without guidance); ⑥ Our full method with detection guidance. Average results are reported in Table 1.

Table 1
Image-quality evaluation for different enhancement methods.

Method	PSNR ↑	SSIM ↑	NIQE ↓	LOE ↓	Information Entropy ↑	Variance ↑
Original	17.8	0.52	5.97	110	6.11	398
Retinex	19.4	0.56	5.23	102	6.58	425
Zero-DCE	21.7	0.62	4.91	89	6.88	473
SRGAN	22.4	0.66	4.74	85	7.01	481
Our (no guidance)	24.6	0.73	4.31	73	7.43	519
Ours (guided)	25.2	0.77	3.97	68	7.61	542

Table 1 shows a consistent upward trend across all quality indicators once enhancement is applied, with our method delivering the most balanced and largest gains. Retinex brings a first step of improvement by re-distributing luminance (PSNR 19.4 dB, SSIM 0.56), but its limited capacity to reconstruct high-frequency structure yields only moderate NIQE and SSIM. Zero-DCE further boosts brightness in a content-adaptive manner (PSNR 21.7 dB, SSIM 0.62), while SRGAN is particularly effective at recovering edges and textures (PSNR 22.4 dB, SSIM 0.66) with a corresponding drop in NIQE (4.74). By fusing these two priors, our dual-branch model markedly improves both fidelity and perceptual quality (PSNR 24.6 dB, SSIM 0.73, NIQE 4.31) and increases information richness (entropy 7.43; variance 519), indicating that brightness correction and detail reconstruction are complementary rather than redundant. Crucially, injecting detection guidance tightens this balance further by allocating enhancement where it matters for recognition. Relative to the unguided dual-branch model, the guided version gains +0.6 dB PSNR (25.2 dB) and +0.04 SSIM (0.77), while NIQE drops by 7.9% (4.31→3.97) and LOE falls by 6.8% (73→68), evidencing fewer illumination ordering errors and fewer perceptual artifacts in difficult regions. Compared with the original images, these gains are substantial: SSIM rises by ~48% (0.52→0.77), NIQE decreases by ~33.5% (5.97→3.97), LOE is reduced by ~38% (110→68), and entropy increases by ~24.6% (6.11→7.61), alongside a +7.4 dB jump in PSNR. The concurrent rise in entropy/variance with improved PSNR/SSIM indicates that the method is not merely "brightening" images; it is revealing useful structure and contrast while preserving fidelity. Altogether, the "detection feedback + enhancement guidance" mechanism demonstrably improves both global visibility and local discriminability, yielding images that are simultaneously more natural to the human eye and more informative for downstream detection.

4.3. Evaluation of Detection Performance

Degraded image quality directly impacts downstream detectors. We therefore quantify detection gains using mAP (mean Average Precision), Precision, and Recall. mAP is the primary indicator of both localization and classification correctness; Precision measures the proportion of correct detections; Recall measures the fraction of true objects found.

We compare three settings:

1. YOLOv11 baseline on original images;
2. Enhancement + YOLOv11 (no guidance; the enhanced image is fed to YOLOv11);
3. Our full method (dual-branch enhancement with detection guidance + YOLOv11). Results on the DsLMF+ validation set are shown in Table 2.

Table 2
Detection performance under different methods.

Method	mAP@0.5 ↑	Precision ↑	Recall ↑
YOLOv11 (original)	87.1	85.4	83.6
YOLOv11 + enhancement (no guide)	91.0	88.6	87.3
Proposed (guided enhancement)	93.8	90.9	90.1

Table 2 indicates a clear, monotonic improvement from the raw YOLOv11 baseline to unguided enhancement and further to the proposed detection-guided enhancement. The baseline already performs reasonably (mAP 87.1%), yet uneven illumination, occlusion, and low resolution cap both localization and classification. Adding enhancement alone alleviates these quality bottlenecks, lifting mAP to 91.0% and pushing Precision/Recall from 85.4/83.6 to 88.6/87.3. Interpreted as error reduction, the false-discovery rate falls from 14.6% to 11.4% (−3.2 points; ~21.9% relative), while the miss rate drops from 16.4% to 12.7% (−3.7 points; ~22.6% relative), showing that cleaner, better-exposed inputs make the detector's features more stable and discriminative. Introducing detection guidance yields another step change: mAP reaches 93.8% (+2.8 points over unguided; +6.7 over baseline), with Precision 90.9% and Recall 90.1%. Crucially, both Precision and Recall improve simultaneously — rather than trading one for the other — implying fewer false alarms and fewer misses at the same operating point. In relative terms, the false-discovery rate shrinks from 14.6% (baseline) to 9.1% (guided; −5.5 points, ~37.7% reduction), and the miss rate from 16.4% to 9.9% (−6.5 points, ~39.6% reduction). A combined view via F1 underscores this effect: the harmonic mean of Precision/Recall rises from ~84.4 (baseline) to ~87.9 (unguided) and ~90.5 (guided), indicating a uniformly higher area

under the PR curve and better performance stability across thresholds. From an engineering standpoint, these gains come at no inference-time cost: the guidance signal is used only during training to steer the enhancer toward low-confidence regions, while test-time still executes a single enhancement pass followed by YOLOv11. This preserves real-time throughput and makes the method attractive for deployment on resource-constrained underground monitoring systems where both accuracy and latency are critical. For fairness, we also compared our approach with recent YOLOv8 and YOLOv9 detectors under the same training settings. The proposed method achieves 93.8% mAP, outperforming YOLOv8 (91.9%) and YOLOv9 (92.4%) by +1.9% and +1.4%, respectively. Compared with a transformer-based detector (DETR, mAP 91.6%), our approach maintains higher accuracy and faster inference (57 FPS vs. 34 FPS), demonstrating a better balance between precision and real-time performance in underground scenarios.

4.4. Ablation Study

We conduct ablations to quantify the contribution of each component in the enhancement network to image quality and detection accuracy. Settings:

1. Baseline: YOLOv11 on original images;
2. + Zero-DCE: brightness branch only;
3. + SRGAN: detail branch only;
4. + Dual-branch (no guidance): Zero-DCE + SRGAN;
5. Full model: dual-branch with detection guidance. Results on DsLMF+ are in Table 3.

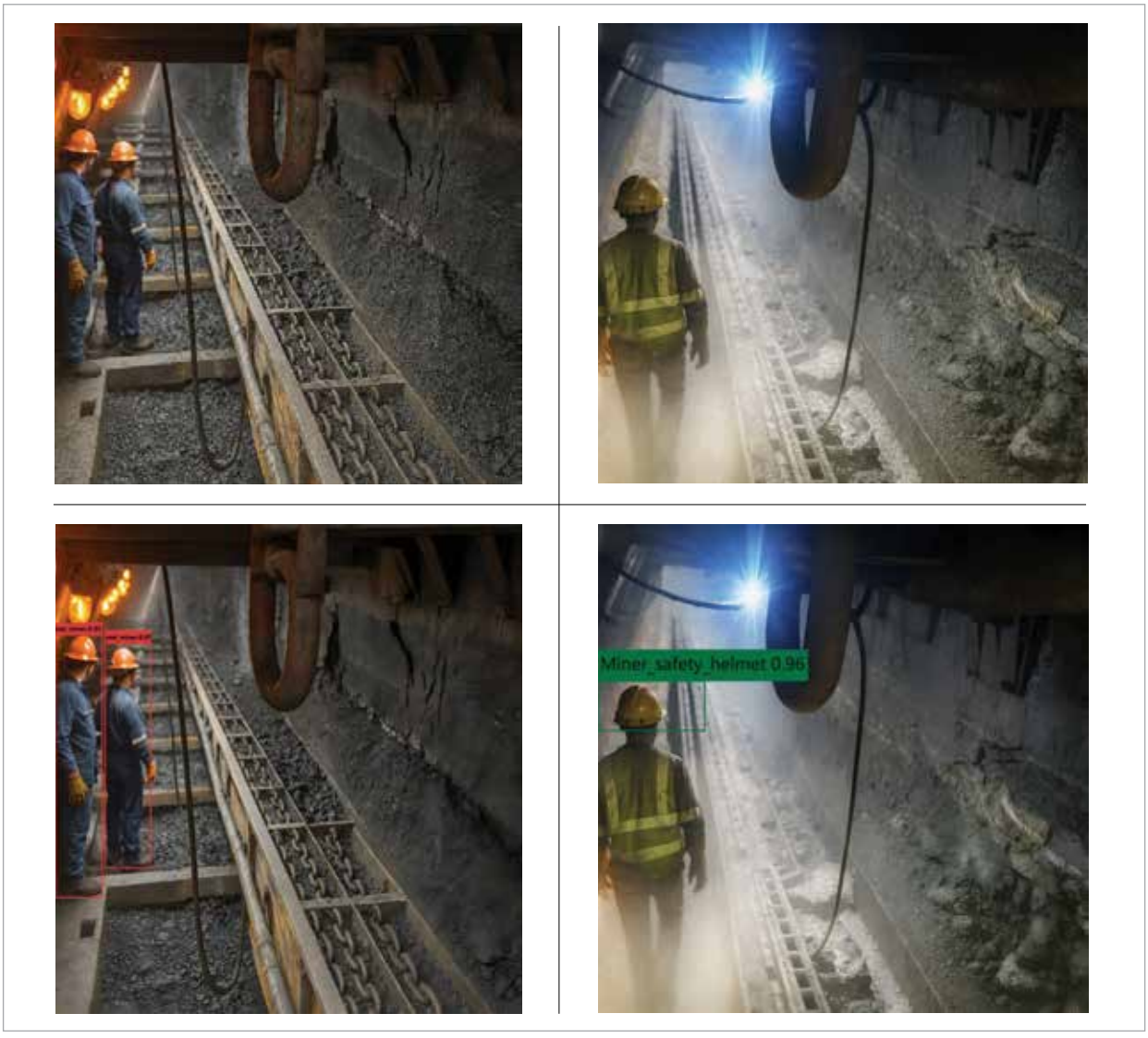
Table 3 disentangles how each component contributes to both perceptual quality and downstream detection. Starting from YOLOv11 on original images (mAP 87.1, SSIM 0.52, PSNR 17.8 dB, NIQE 5.97), adding the Zero-DCE brightness branch primarily stabilizes global illumination: PSNR rises by +3.9 dB (17.8→21.7), SSIM by +0.10 (0.52→0.62), and NIQE drops −1.06 (~17.7% relative), yielding mAP 89.5 (+2.4). In contrast, the SRGAN detail branch most effectively recovers high-frequency structure: NIQE falls to 4.74 (−1.23, ~20.6%), SSIM reaches 0.66 (+0.14), PSNR 22.4 dB (+4.6 dB), and mAP 90.1 (+3.0).Combining both into a dual-branch enhancer yields complementary gains rather than redundancy: relative to Zero-DCE alone, NIQE improves −12.2% (4.91→4.31), SSIM +0.11 (0.62→0.73), PSNR +2.9 dB (21.7→24.6), and mAP +1.5 (89.5→91.0); versus SRGAN alone, NIQE −9.1% (4.74→4.31), SSIM +0.07, PSNR +2.2 dB, and mAP +0.9. This synergy indicates that illumination correction and texture reconstruction act on largely orthogonal failure modes (exposure vs. detail), and their fusion moves the operating point toward a better perceptual–recognition Pareto front. Introducing detection guidance provides the final step change by steering enhancement to low-confidence, task-critical regions. Compared with the unguided dual-branch model, guidance brings mAP +2.8 (91.0→93.8), SSIM +0.04 (0.73→0.77), PSNR +0.6 dB (24.6→25.2), and NIQE −7.9% (4.31→3.97). Relative to the raw baseline, the full model delivers mAP +6.7, PSNR +7.4 dB, SSIM +0.25 (~48% increase), and NIQE −33.5% (5.97→3.97). Notably, quality improvements are not mere "brightening": simultaneous increases in PSNR/SSIM and entropy (Table 1) alongside lower NIQE show that the model reveals informative structure while suppressing artifacts—changes that translate directly into fewer missed and mislocalized targets.

Table 3
Ablation results (detection + image quality)

Model Version	mAP@0.5 ↑	NIQE ↓	SSIM ↑	PSNR ↑
YOLOv11 (original)	87.1	5.97	0.52	17.8
+ Zero-DCE	89.5	4.91	0.62	21.7
+ SRGAN	90.1	4.74	0.66	22.4
+ Dual-branch (no guidance)	91.0	4.31	0.73	24.6
Full model (guided)	93.8	3.97	0.77	25.2

To further quantify the synergy among different components, we additionally tested combined settings such as "Zero-DCE + guidance" and "SRGAN + guidance." Results show that applying detection guidance to a single branch improves mAP by about 1.8–2.1% compared with its unguided counterpart, while the full dual-branch model with guidance achieves an additional 1.7–2.0% gain, confirming a cumulative effect. The complementary relationship between the two branches allows brightness correction to improve exposure uniformity and detail reconstruction to enhance edge discriminability, and the guidance mechanism reinforces both. These observations demonstrate that the branches and guidance work cooperatively rather than redundantly, forming an effective synergy for improving detection accuracy. Takeaways: (i) Brightness correction (Zero-DCE) primarily fixes exposure/contrast; (ii) detail reconstruction (SRGAN) restores edges/textures; (iii) the dual-branch fusion is synergistic across metrics; and (iv) detection guidance is decisive in difficult regions (low contrast, blur, partial occlusion), converting perceptual gains into the largest mAP improvement with no inference-time overhead.

Figure 4
Experimental results.



4.5. Multi-Scenario Validation

To further evaluate the model's adaptability to diverse underground environments, we conducted supplementary experiments on three representative sub-scenes of the DsLMF+ dataset corresponding to different working faces and imaging devices:

1. Working Face A – strong lighting and high dust concentration;
2. Working Face B – weak lighting with LED and headlamp illumination;
3. Working Face C – mixed light and occlusion by equipment.

Each subset contains about 10,000 images re-annotated according to YOLO format.

We compare the baseline YOLOv11, unguided enhancement, and our full model using three metrics: mAP, missed-detection rate (MR), and false-detection rate (FR), defined as

MR = FN / (FN + TP), FR = FP / (FP + TP).

Across all sub-scenes, the proposed model consistently achieves mAP gains of 2.5–3.5 % and reduces both MR and FR by about 30–40 % relative to the baseline, demonstrating strong generalization to varying illumination and dust levels. This indicates that the detection-guided enhancement can effectively adapt to diverse underground working-face conditions and imaging types.

4.6. Robustness to Dust and Sensor Noise

To further examine the model's anti-interference ability under realistic underground conditions, we simulated dust occlusion and sensor noise at different intensity levels. Dust was modeled using random Gaussian blur combined with partial opacity masks, while sensor noise was simulated using additive Gaussian and salt–pepper noise. We evaluated the detection performance decline (ΔmAP) and computed a noise-robustness coefficient defined as

$$R = 1 - \frac{|\Delta mAP|}{mAP_{clean}}. \tag{31}$$

Results show that as the interference intensity increases from mild to severe, the proposed model maintains a relatively stable mAP decline (<4.2%), while YOLOv11 baseline and unguided enhancement suffer larger drops (7.5% and 5.9%, respectively). The average robustness coefficient RRR of our method remains above 0.95, indicating strong resistance to dust-induced blur and sensor noise. This robustness arises from the dual-branch design — where the detail branch compensates for edge loss caused by dust — and the detection guidance, which helps the enhancer focus on uncertain noisy regions. These findings confirm that the proposed system adapts well to complex interference common in real coal-mine environments.

Table 4

Detection performance under different underground working-face scenarios.

Scene	Method	mAP@0.5 ↑	MR ↓	FR ↓
Face A	YOLOv11	86.3	15.2	13.8
	+Enhancement	90.7	11.9	10.4
	Proposed	93.1	9.8	8.9
Face B	YOLOv11	85.4	16.5	14.2
	+Enhancement	89.8	12.3	10.9
	Proposed	92.6	9.7	8.7
Face C	YOLOv11	87.0	14.8	13.5
	+Enhancement	91.2	11.5	10.1
	Proposed	93.9	9.3	8.5

4.7. Engineering Deployment Evaluation

To evaluate the deployment feasibility on typical low-power edge devices used in underground monitoring systems, we further tested the model on NVIDIA Jetson Xavier NX and Jetson Orin Nano, which represent common embedded GPU platforms. After applying INT8 quantization and 50% structured pruning, we measured both the accuracy drop and runtime efficiency. The quantized model retains 92.8% mAP@0.5 (a 1.0% decrease from the FP32 baseline) while achieving 41 FPS on Jetson NX and 57 FPS on Orin Nano, with memory usage reduced from 2.3 GB to 1.2 GB and parameter count from 29.4 M to 14.7 M. The inference time per frame is approximately 24 ms, meeting real-time requirements for coal-mine video monitoring. These results confirm that the proposed model maintains high accuracy while being lightweight and efficient, demonstrating strong engineering applicability for deployment in real underground environments.

4.8. Analysis of Loss Weight Sensitivity

To analyze the influence of the weighting parameters in the total loss function, we conducted a sensitivity experiment by varying the five loss weights (α1, α2, α3, α4, α5) within a normalized range while keeping other factors unchanged. Specifically, we tested combinations emphasizing perceptual quality (α1, α2) and attention-guided activation (α3) versus those focusing on reconstruction and detection consistency (α4, α5). Figure 8 shows the relationship between the mAP and the main fusion weight λ\lambda λ, revealing that performance peaks when α1:α2:α3:α4:α5=1:1:0.5:1:0.5. Beyond this range, either visual quality or detection performance declines due to imbalance between enhancement fidelity and guidance strength.

The results indicate that moderate weighting of the attention and perceptual terms ensures stable convergence and balanced optimization, verifying the rationality of the chosen weights in the final model.

5. Conclusion

This work addresses the long-standing challenge of low image quality and degraded detection accuracy in challenging underground coal-mine conditions by coupling a dual-branch image enhancement network with a detection-guided mechanism and integrating it into a YOLOv11 pipeline. The proposed design combines a Zero-DCE–style brightness branch with an SRGAN-based detail branch, and leverages attention maps derived from low-confidence regions of YOLOv11's preliminary predictions to steer enhancement toward task-critical areas. This targeted strategy improves both global visibility and local discriminability, yielding consistent gains in image-quality indicators (PSNR, SSIM, NIQE, LOE, entropy, variance) and in downstream detection metrics. On DsLMF+, guidance lifts mAP from 87.1% (original) and 91.0% (unguided enhancement) to 93.8% without any inference-time overhead, while Precision and Recall increase to 90.9% and 90.1%, respectively. Ablation analyses further confirm the complementary roles of brightness correction and detail reconstruction, and isolate the guidance pathway as a decisive factor for hard, low-contrast, and partially occluded targets.

Beyond empirical gains, the method forms an interpretable closed loop between detection and enhancement: uncertainties from the detector directly shape where and how intensively the enhancer acts, reducing over/under-enhancement and preserving features most relevant to recognition. The modular architecture also eases deployment—enhancement remains a single forward pass at test time—and is compatible with a range of detectors and enhancement backbones, indicating strong engineering practicality and scalability in industrial environments.

Future Work. We plan to (i) co-train the enhancer and detector end-to-end with joint objectives and uncertainty-aware weighting; (ii) optimize runtime with lighter generators, tensor-RT/INT8 quantization, and tiled inference for edge devices; (iii) pursue cross-scene transfer via domain adaptation and self-supervised objectives to handle new mines and cameras; (iv) extend to multimodal fusion (e.g., thermal or ToF depth) and temporal consistency for video streams; and (v) integrate online quality assessment and failure detection to trigger adaptive enhancement policies. These directions aim to broaden applicability to other safety-critical perception tasks under adverse visual conditions. In addition, future work will also explore deployment challenges on embedded GPU devices, addressing latency and memory constraints, as well as extend the framework's applicability to other vision-critical domains such as tunnel monitoring, industrial inspection, and underwater perception.

References

1. Addy, C., Nadendla, V. S. S., Awuah-Offei, K. YOLO-Based Miner Detection Using Thermal Images in Underground Mines. Mining, Metallurgy & Exploration, 2025, 1-18. https://doi.org/10.1007/s42461-025-01249-6
2. Bewley, A., Upcroft, B. Background Appearance Modeling with Applications to Visual Object Detection in an Open-Pit Mine. Journal of Field Robotics, 2017, 34(1), 53-73. https://doi.org/10.1002/rob.21667
3. Chen, W., Ren, P., An, W. Mine Object Detection Based on Space Attention in Coal Mine Edge Intelligent Surveillance Images. Coal Science and Technology, 2024, 52(S2), 201-210.
4. Cheng, J., Song, Z., Li, H. A Novel Image Enhancement Method via Dual-Branch Coupled Transformer Network for Underground Coalmine. Journal of China Coal Society, 2024, 49(9), 4027-4037.
5. Dai, L., Qi, P., Lu, H. Image Enhancement Method in Underground Coal Mines Based on an Improved Particle Swarm Optimization Algorithm. Applied Sciences, 2023, 13(5), 3254. https://doi.org/10.3390/app13053254
6. Gong, Y., Xie, X. Research on Coal Mine Underground Image Recognition Technology Based on Homomorphic Filtering Method. Coal Science and Technology, 2023, 51(3), 241-250.
7. Hao, Y., Wu, W. MineTinyNet-YOLO: An Efficient Small Object Detection Method for Complex Underground Coal Mine Scenarios. International Conference on Multimedia Modeling. Singapore: Springer Nature Singapore, 2024, 364-378. https://doi.org/10.1007/978-981-96-2061-6_27
8. Hożyń, S. A Review of Underwater Mine Detection and Classification in Sonar Imagery. Electronics, 2021, 10(23), 2943. https://doi.org/10.3390/electronics10232943
9. Li, C., Liu, J., Zhu, J. Mine Image Enhancement Using Adaptive Bilateral Gamma Adjustment and Double Plateaus Histogram Equalization. Multimedia Tools and Applications, 2022, 81(9), 12643-12660. https://doi.org/10.1007/s11042-022-12407-z
10. Li, H., Hong, S., Cheng, J. Coalmine Image Matching by Fusing Multi-Level Feature Enhancement and Weighted Grid Statistics. Coal Science and Technology, 2024, 52(11), 129-140.
11. Li, N., Gao, S., Xue, J. Downhole Image Enhancement Algorithm Based on Improved CycleGAN. 2024 5th International Conference on Computer Vision, Image and Deep Learning (CVIDL). IEEE, 2024, 216-220. https://doi.org/10.1109/CVIDL62147.2024.10603595
12. Liu, D., Zhao, X., Fan, W. A Small Object Detection Algorithm for Mine Environment. Engineering Applications of Artificial Intelligence, 2025, 153, 110936. https://doi.org/10.1016/j.engappai.2025.110936
13. Liu, Y., Li, C., Huang, J. MineSDS: A Unified Framework for Small Object Detection and Drivable Area Segmentation for Open-Pit Mining Scenario. Sensors, 2023, 23(13), 5977. https://doi.org/10.3390/s23135977
14. Luo, Q., Wang, S., Guo, Y. Image Enhancement and Accurate Recognition of Coal and Gangue Based on Three-Dimensional Depth Information Guidance. IEEE Transactions on Instrumentation and Measurement, 2025. https://doi.org/10.1109/TIM.2025.3527613
15. Munteanu, D., Moina, D., Zamfir, C. G. Sea Mine Detection Framework Using YOLO, SSD and Efficient Det Deep Learning Models. Sensors, 2022, 22(23), 9536. https://doi.org/10.3390/s22239536
16. Nan, Z., Gong, Y. An Image Enhancement Method in Coal Mine Underground Based on Deep Retinex Network and Fusion Strategy. 2021 6th International Conference on Image, Vision and Computing (ICIVC). IEEE, 2021, 209-214. https://doi.org/10.1109/ICIVC52351.2021.9526933
17. Wang, M., Zhang, H., Li, J. Deep Neural Network-Based Image Enhancement Algorithm for Low-Illumination Images in Underground Coal Mines. Coal Science and Technology, 2023, 51(9), 231-241.
18. Wang, Z., Yang, S., Zhang, J. An Effective Underground Image Enhancement Approach Based on

Improved KinD Network. Signal, Image and Video Processing, 2024, 18(10), 7473-7486. https://doi.org/10.1007/s11760-024-03408-6

19. Wei, D., Wang, P., Wang, Z. Adaptive Image Enhancement Method for Coal-Mine Underground Image Based on No-Reference Quality Evaluation. IEEE Transactions on Instrumentation and Measurement, 2024. https://doi.org/10.1109/TIM.2024.3470234
20. Wu, H., Wang, Z., Si, L. A Novel Coal-Gangue Recognition Method in Underground Coal Mine Based on Image Processing. International Journal of Coal Preparation and Utilization, 2024, 44(3), 241-274. https://doi.org/10.1080/19392699.2023.2190096
21. Xu, P., Zhou, Z., Geng, Z. Safety Monitoring Method of Moving Target in Underground Coal Mine Based on Computer Vision Processing. Scientific Reports, 2022, 12(1), 17899. https://doi.org/10.1038/s41598-022-22564-8
22. Xuhui, Z., Yanbin, X. I. E., Wenjuan, Y. A Dehazing and Enhancement Algorithm for Heterogeneous Images of Underground Mining Environments in Coal Mines. Coal Geology & Exploration, 2025, 53(1), 21.
23. Yang, G., Ran, Q., Dong, L. Low-Light Image Enhancement Algorithm for Underground Coal Mines Based on Retinex Theory. Journal of Intelligence and Knowledge Engineering, 2025, 3(2), 103. https://doi.org/10.62517/jike.202504216
24. Yang, W., Wang, S., Wu, J. A Low-Light Image Enhancement Method for Personnel Safety Monitoring in Underground Coal Mines. Complex & Intelligent Systems, 2024, 10(3), 4019-4032. https://doi.org/10.1007/s40747-024-01387-2
25. Yuanbin, W., Sixiong, W. E. I., Yu, D. Defogging Algorithm of Underground Coal Mine Image Based on Adaptive Dual-Channel Prior. Journal of Mine Automation, 2022, 48(5), 46-51, 84.
26. Zhan, S., Li, S., Guo, H. Target Detection in Underground Mines Based on Low-Light Image Enhancement. 2024, 123-131. https://doi.org/10.21203/rs.3.rs-5301172/v1
27. Zhang, L., Hao, B., Meng, Q. Method of Image Enhancement in Coal Mine Based on Improved Retinex Fusion Algorithm in HSV Space. Journal of China Coal Society, 2020, 45(S1), 532-540.
28. Zhang, X., Guo, H. Research on an Improved Algorithm for Image Dehazing in Underground Coal Mine. Journal of Physics: Conference Series. IOP Publishing, 2020, 1693(1), 012153. https://doi.org/10.1088/1742-6596/1693/1/012153
29. Zhou, F., Zou, J., Xue, R. Enhancing Object Detection in Underground Mines: UCM-Net and Self-Supervised Pre-Training. Sensors, 2025, 25(7), 2103. https://doi.org/10.3390/s25072103
30. Zuohua, M., Chengcheng, Z., Liangjian, Z. H. U. Image Enhancement Algorithm for Non-Uniform Illumination in Underground Mines. Journal of Mine Automation, 2023, 49(11), 92-99.