EAI Endorsed Transactions on Energy Web Research Article **EAI.EU**

AI-Enabled Lightweight Video Stream Processing for Electric Shock Prevention Monitoring in Smart Grid Edge Scenarios

Dainan Zhang[1], Xiaohui Li[1], Zijian Chen[1], Canshu Qiu[1], Hui Xu[1], Long Chen[2,*]

[1]Chaozhou Power Supply Bureau of Guangdong Power Grid Co., Ltd, Chaozhou 521000 Guangdong, China
[2]Guangzhou Vcmy Technology Co., Ltd., Guangzhou, 510000, Guangdong, China

Abstract

INTRODUCTION: Real-time early warning of electric shock risks is urgently needed in edge scenarios of smart grids. Existing visual methods are still limited by scene adaptability, end-to-end lightweight design, and insufficient field verification.
Objective: The objective of the study is to propose a lightweight video stream processing algorithm for power operation sites, balancing accuracy, latency, and deployment efficiency on edge devices.
METHODS: A constrained multi-objective model prioritizing safety risks is constructed, integrating risk-weighted ROI extraction, adaptive inter-frame filtering, lightweight image enhancement, ROI-aware YOLOv8n, pruning, and pipelined inference.
RESULTS: Data analysis shows that the model has 2.86M parameters, 3.12 GFLOPs of computation, a mAP@0.5 of 93.27%, and a Jetson Nano speed of 31.4 fps. In substations, the average mAP@0.5 is 92.14%, the early warning latency is 126 ms, and the false negative rate in rain, fog, low light, and occlusion scenarios is approximately 1%.
CONCLUSION: This method reduces redundant computation while maintaining high accuracy and low latency, providing an edge-intelligent solution for electric shock protection in smart grids.

Keywords: Smart grid; Energy Internet; electric shock prevention monitoring; AI-enabled edge computing; lightweight video stream processing; YOLOv8n; visual safety warning

Received on 12 March 2026, accepted on 31 May 2026, published on 18 September 2026

doi: 10.4108/ew.15089

1. Introduction

With the continued development of smart grids and the Energy Internet, power operation and maintenance are becoming more distributed, data-intensive, and safety-critical. Electric-shock incidents remain a persistent threat in power operations. Evidence from China's electric power industry and Brazilian distribution networks shows that these accidents continue to cause serious casualties and follow recurring patterns that call for targeted prevention [1,2]. Because manual supervision cannot continuously cover all camera views or issue immediate warnings, power-industry research has begun to combine edge computing with visual recognition of unsafe actions and substation workers' protective wear [3,4].

At the target-recognition level, studies have examined unsafe-behavior detection [5] and real-time PPE verification with worker–equipment association and temporal alarm logic [6]. Work on personnel safety in power-operation settings

*Corresponding author. Email: hzm2023@sina.com

now covers both safety-wear recognition and irregular-behavior detection [7,8].

Small targets, complex backgrounds, and temporal continuity remain central difficulties. Power-operation studies have improved small-object and multi-scale feature representation [9,10]. For continuous video, temporal inference has been applied to lineman safety monitoring [11], while substation inspection research uses YOLOv8 as the person detector within a spatiotemporal action-detection framework to distinguish normal from dangerous actions [12]. Substation PPE recognition has further been explored through graph models and multi-category lightweight detectors [13,14].

Deployment evidence is equally important. Power-transmission-corridor monitoring has also been implemented through a lightweight edge-AI framework that integrates network cameras, Jetson Nano devices, and adaptive frame sampling for local anomaly detection under embedded-resource constraints [15]. Domain-specific models now distinguish the wearing status of insulating gloves rather than only locating a glove [16]. Standardized PPE inspection focuses on whether the required equipment is worn correctly [17], whereas edge-deployed PPE systems support local real-time analysis and alarm generation [18]. Safe-approach-distance monitoring directly links visual perception to energized-area proximity risk [19]. Complementarily, electrical-construction violation detection combines PPE recognition with knowledge reasoning and progressive distillation, linking observed protective-equipment states to rule-based safety violations while reducing model complexity [20].

Based on this review, the deployment problem can be stated more clearly. Existing video-stream processing methods for electric-shock prevention have three linked limitations. First, scenario adaptability remains insufficient because many lightweight detectors are designed for general visual tasks and are not tailored to energized-area risk, small or occluded personnel targets, protective-equipment states, or unstable field illumination. Second, optimization is rarely end-to-end: preprocessing, detection, postprocessing, and inference scheduling are often improved separately, leaving redundant computation in the full video-stream link. Third, engineering validation is still limited, as many methods are evaluated on public images or PC-based simulations rather than on continuous substation video running on resource-constrained edge hardware. This paper addresses these limitations through a safety-prioritized lightweight algorithm that integrates risk-weighted ROI extraction, illumination-adaptive redundancy filtering, ROI-aware YOLOv8n optimization, target-aware structured pruning, and edge-side pipelined inference. The aim is to reduce redundant computation while maintaining the warning accuracy and response time required for practical electric-shock prevention.

The main contributions are as follows. First, electric-shock prevention monitoring is formulated as a safety-prioritized constrained multi-objective optimization problem, and an end-to-end lightweight processing architecture is implemented for smart-grid edge scenarios. Second, a scenario-aware preprocessing module is developed to allocate computation according to energized-area risk and to filter redundant frames under illumination changes. Third, YOLOv8n is improved with ROI-aware coordinate attention and target-aware structured pruning, and the resulting detector is coupled with pipeline inference for Jetson Nano deployment and substation field testing.

The remainder of this paper is organized as follows. Section 2 introduces the technical foundation and formulates the lightweight processing problem for electric-shock prevention scenarios. Section 3 describes the proposed full-link lightweight video-stream processing algorithm. Section 4 presents the experimental design, comparative and ablation results, and field verification. Section 5 concludes the paper and outlines future research directions.

2. Relevant Technical Foundation and Scenario Problem Modeling

2.1 Core Requirements and Scenario Constraints of Power Electric Shock Prevention Visual Monitoring

Power operation sites require electric-shock prevention monitoring for three practical tasks: (1) detecting personnel intrusion into live areas; (2) detecting whether workers are wearing insulating gloves, insulating boots, safety helmets, and other protective equipment as required; and (3) identifying high-risk behaviors such as unlicensed operation and illegal live operation. The warning latency for these tasks should be below 200 ms.

2.2 Fundamentals of Lightweight Video Stream Processing

Three technical components are relevant to this problem. The first is lightweight video preprocessing, which reduces unnecessary computation through ROI extraction, inter-frame redundancy filtering, and lightweight image enhancement. The second is compact detection-model design, including lightweight backbones, attention, pruning, quantization, and knowledge distillation. Coordinate attention preserves positional information with limited computational overhead [21], structured compression and quantization support real-time edge detectors [22], and lightweight low-light enhancement can improve dark frames without imposing a heavy front-end cost. Existing work therefore provides useful building blocks, but these components are usually optimized separately and seldom reflect the safety priority and scene-specific constraints of electric-shock prevention. This study adapts them as a unified solution for power-operation scenarios.

2.3 Mathematical Modeling of Lightweight Processing Problems for Power Electric Shock Prevention Scenarios

The lightweight video-stream processing problem for electric-shock prevention is formulated here as a safety-prioritized constrained multi-objective optimization problem. The safety-risk function is defined first to reflect the operating principle that the cost of a missed alarm is far higher than that of a false alarm:

$$R_{\text{risk}} = \omega_{\text{miss}} \cdot R_{\text{miss}} + \omega_{\text{false}} \cdot R_{\text{false}} \tag{1}$$

where R_{risk} is the comprehensive safety risk value of the electric shock monitoring system; the smaller the value, the better the system's safety performance. ω_{miss} is the weight of the missed detection risk, and ω_{false} is the weight of the false alarm risk. Considering power safety management requirements, missed detections may lead to direct electric shock accidents; therefore, $\omega_{\text{miss}} = 0.9$, $\omega_{\text{false}} = 0.1$ are taken. R_{miss} is the system missed detection rate, referring to the proportion of samples where high-risk events are not detected out of the total number of high-risk samples. R_{false} is the system false alarm rate, referring to the proportion of samples where normal events are mistakenly identified as high-risk events out of the total number of normal samples.

Second, a full-link computational cost function for a video stream is defined, which covers the entire process from input to output:

$$C_{total} = C_{pre} + C_{inf} + C_{post} \tag{2}$$

where C_{total} represents the total computational cost of processing a single frame of video stream, measured in FLOPs, and is a core indicator of lightweight design; C_{pre} represents the computational cost of the preprocessing module, including ROI extraction, inter-frame differencing, and image enhancement; C_{inf} represents the computational cost of the detection model inference, which is the core part of the total computational cost; C_{post} represents the computational cost of the postprocessing module, including non-maximum suppression and early warning logic judgment.

Subsequently, a real-time constraint function is defined at the edge to ensure a real-time response to early warnings:

$$T_{\text{total}} = T_{\text{pre}} + T_{inf} + T_{\text{post}} \leq \frac{1}{FPS_{\text{min}}} \tag{3}$$

where T_{total} represents the end-to-end processing latency of a single frame video stream, in milliseconds (ms); T_{pre}, T_{inf}, T_{post} represent the single-frame latencies for preprocessing, inference, and postprocessing, respectively; FPS_{min} represents the minimum real-time requirement, which is 25 fps in this paper, therefore the total latency per frame must be $\leq$ 40 ms.

Based on the above definitions, a multi-objective optimization function that prioritizes security risks is constructed as follows:

$$\min_X F(X) = \lambda_1 \cdot \frac{R_{\text{risk}}(X)}{R_{\text{risk}}^{\max}} + \lambda_2 \cdot \frac{C_{\text{total}}(X)}{C_{\text{total}}^{\max}} + \lambda_3 \cdot \frac{T_{\text{total}}(X)}{T_{\text{total}}^{\max}} \tag{4}$$

where X represents the set of end-to-end optimization parameters for the algorithm, including the ROI risk weights of the preprocessing module, inter-frame thresholds, structural parameters of the detection network, pruning quantization parameters, etc.; $R_{\text{risk}}^{\max}$ is the maximum allowable value of safety risk, which is taken as 0.02 in this paper; $C_{\text{total}}^{\max}$ is the maximum allowable value of computation per frame, which is taken as 6 G FLOPs in this paper; T_{total}^{max} is the maximum allowable value of latency per frame, which is taken as 40 ms in this paper; $\lambda_1, \lambda_2, \lambda_3$ are weight coefficients, satisfying $\lambda_1 + \lambda_2 + \lambda_3 = 1$, and taking into account the safety priority principle, $\lambda_1 = 0.5, \lambda_2 = 0.25$, $\lambda_3 = 0.25$.

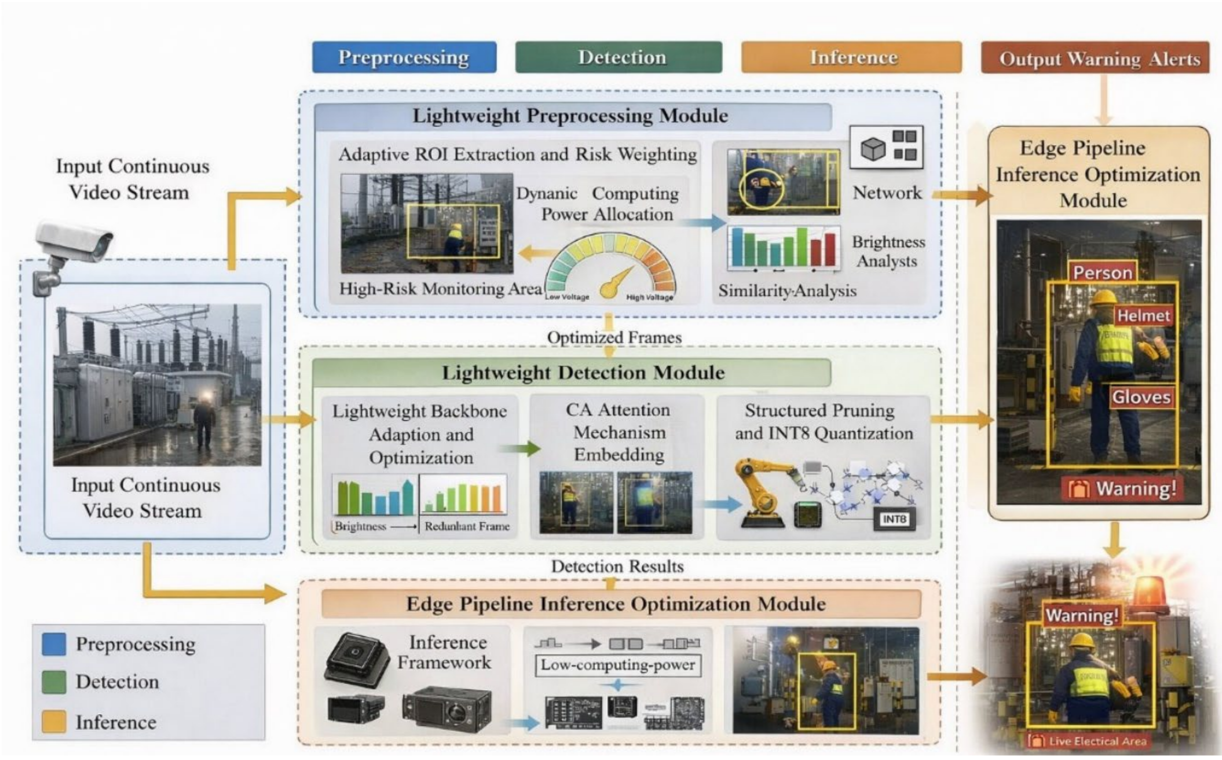

Figure 1. Algorithm Overall Architecture Diagram

Finally, the feasible region constraint of the optimization problem is defined to ensure that all solutions meet the hard requirements, as follows:

$$\begin{cases} R_{\text{risk}}(X) \le R_{\text{risk}}^{\max} \\ C_{\text{total}}(X) \le C_{\text{total}}^{\max} \\ T_{\text{total}}(X) \le T_{\text{total}}^{\max} \end{cases} \quad (5)$$

Accordingly, the algorithm seeks a Pareto-optimal solution under these constraints, balancing safety risk, computational cost, and processing latency for deployable edge monitoring.

3. Lightweight Video Stream Processing Algorithm Design for Electric Shock Prevention Monitoring

3.1 Overall Algorithm Architecture Design

Figure 1 illustrates the overall architecture of the proposed end-to-end lightweight video-stream processing algorithm. The framework contains three modules: power-scenario-adaptive lightweight preprocessing, electric-shock-prevention target detection, and edge-side pipelined inference optimization. Together, these modules cover the full path from video input to warning output and support full-link lightweight optimization.

The algorithm operates as follows. The input video stream first passes through lightweight preprocessing: adaptive risk-weighted ROI extraction identifies high-risk energized areas and allocates computation according to voltage level; illumination-aware inter-frame filtering determines whether full inference is necessary; and lightweight image enhancement improves low-light, rainy, and foggy frames. The processed image is then fed into a domain-specific detector for personnel, insulating gloves and boots, safety helmets, and live equipment; comparable electrical-PPE studies show that small and occluded protective items require dedicated feature treatment [14,16]. The detector combines ROI-aware coordinate attention [21] with target-aware structured pruning [22]. This fully local design is conceptually consistent with a power-corridor edge system that integrates camera input, adaptive frame sampling, and Jetson Nano inference [15].

3.2 Lightweight Preprocessing Module for Video Streams Adapted to Power Scenarios

Most monitoring cameras at power sites are fixed installations, and high-risk energized areas usually occupy stable positions with risk levels that vary by voltage. Because much of the background remains static, full-frame processing of every frame wastes edge computing resources. To address this, the proposed adaptive ROI extraction method assigns higher processing resolution to energized areas with higher risk and lower resolution to lower-risk regions. This concentrates computation where missed detections would be most harmful while reducing redundant background processing.

3.2.1 Adaptive ROI Extraction Algorithm for Energized Areas with Risk Weights

To account for differences in risk level among energized areas, an adaptive ROI extraction algorithm with risk weighting is proposed. The method assigns higher processing resolution to high-voltage, high-risk energized areas and lower resolution to low-risk regions, reducing computation on background areas while protecting detection accuracy where missed alarms are most critical. The dynamic resolution allocation formula is as follows:

$$S_i = S_{\text{base}} \cdot (1 + k \cdot w_i), w_i = \frac{U_i}{U_{\max}} \tag{6}$$

where S_i is the processing resolution of the i energized region, in pixels; S_{base} is the base processing resolution, which is 640×640 in this paper; k is the resolution adjustment coefficient, which is 0.5 in this paper to avoid excessively high resolution leading to a sharp increase in computational cost; w_i is the risk weight of the i energized region; the larger the value, the higher the risk of electric shock in that region; U_i is the rated voltage level of the i-th energized region, in kV; $U_{\max}$ is the highest voltage level in the monitored scene, in kV.

The implementation proceeds as follows: (1) pre-label each energized region in the monitoring image, record its voltage level, and calculate the corresponding risk weight and processing resolution; (2) mask the input frame to extract the ROI set for energized regions and personnel-passage regions; (3) scale each ROI according to Equation (6); and (4) merge the processed ROIs before sending them to the subsequent detection network.

3.2.2 Illumination-Adaptive Inter-Frame Redundancy Filtering Strategy

Because fixed monitoring images are often static, an illumination-adaptive inter-frame redundancy filtering strategy is designed. The strategy avoids full inference on frames without meaningful scene changes and reduces the sensitivity of fixed thresholds to illumination fluctuation. The adaptive threshold update formula is as follows:

$$T_r(t) = T_{r0} \cdot \left(1 + \alpha \cdot \frac{\sigma_I(t)}{\sigma_{I0}}\right), \alpha \in [0,1] \tag{7}$$

The core logic of the algorithm is as follows. The pixel change rate of the ROI is calculated through inter-frame difference. When scene illumination fluctuates (e.g., cloud cover or lighting changes), the grayscale standard deviation increases and the threshold is adjusted upward to avoid false triggers. When the scene is stable, the threshold is adjusted downward to retain sensitivity to small movements. When the ROI change rate is at least the current threshold, a significant scene change is identified and the frame is sent to the detection network; otherwise, the previous detection result is reused. The reuse strategy follows the multi-frame verification logic used in real-time PPE monitoring [6]. Edge PPE systems likewise show that video analysis and alarm generation can be performed locally [18]. In the present implementation, it reduces the number of full inferences by more than 90% in static scenes.

3.2.3 Lightweight Environmental Adaptive Image Enhancement

To handle low-light and rain/fog interference in outdoor power scenarios, a lightweight image-enhancement module based on adaptive gamma correction and guided filtering is designed. Lightweight curve-based enhancement has demonstrated that real-time correction can be achieved with very small models on constrained hardware. The adaptive gamma-correction formula used in this study is expressed as follows:

$$O(x,y) = 255 \cdot \left(\frac{I(x,y)}{255}\right)^{\gamma(x,y)}, \gamma(x,y) = \left(\frac{\mu_{ROI}}{128}\right)^{\beta} \tag{8}$$

where $O(x,y)$ is the pixel value of the enhanced image at position (x, y); $I(x,y)$ is the pixel value of the input image at position (x, y); $\gamma(x,y)$ is the adaptive gamma coefficient, dynamically adjusted according to the average brightness of the ROI; μ_{ROI} is the average grayscale value of the ROI, ranging from 0 to 255; β is the correction coefficient, which is set to 0.8 in this paper to balance the magnitude of brightness enhancement with the risk of noise amplification.

This algorithm only processes the ROI. When the average brightness of the ROI is below 128 (low light), $\gamma < 1$, increasing image brightness; when the brightness is above 128 (strong light/backlight), $\gamma > 1$, decreasing brightness, achieving adaptive image enhancement. Compared to traditional Retinex and dehazing algorithms, this algorithm reduces computational cost by more than 80%, while effectively improving image quality and avoiding interference from complex environments for subsequent detection.

3.3 Lightweight Detection Network Optimization Design for Electric Shock Prevention Targets

This study uses YOLOv8n as the baseline detector and performs multidimensional optimization for electric-shock-prevention targets. Power-operation studies show that small protective items require high-resolution feature retention [9], while substation PPE and insulating-glove detectors confirm the need to balance localization accuracy with model compactness [14,16]. The design therefore compresses the model while keeping the accuracy loss controllable.

3.3.1 Lightweight Backbone Network Adaptation Optimization

The YOLOv8n backbone network adopts a CSP Darknet structure. This study optimizes it in two ways. First, some standard convolutions are replaced with depth-wise separable convolutions to reduce parameters and computation while maintaining feature extraction. Second, redundant shallow channels are reduced while the deep-layer capacity for small-target features is retained. This choice follows evidence from power-operation and electrical-PPE detection that cuffs, gloves, belts, and other protective items are easily weakened by coarse or overly compressed features [9,14].

3.3.2 Improved Lightweight CA Attention Mechanism Embedded for ROI Perception

In electric-shock-prevention monitoring, core targets are concentrated in risk-defined ROIs and are often small or occluded. This study therefore embeds an ROI-aware variant of coordinate attention into the backbone. Coordinate attention separately encodes horizontal and vertical information so that positional cues are retained with low overhead [21]. The proposed variant further multiplies the learned attention by an energized-area mask. The attention weight is calculated as follows:

$$M_c(i,j) = \sigma\left(F_{1\times1}([X_h^c, X_w^c]) \cdot M_{ROI}(i,j)\right) \tag{9}$$

where $M_c(i,j)$ is the final attention weight of the c feature channel at position (i, j); σ is the sigmoid activation function, mapping the weights to between 0 and 1; F(1×1) is a 1×1 convolution operation used to fuse horizontal and vertical feature codes; X_h^c, X_w^c are the horizontal and vertical global feature codes of the c feature channel, respectively, inheriting the core structure of the CA attention mechanism; $M_{ROI}(i,j)$ is the spatial weight mask of the ROI, with a value of 1.2 within the ROI and 0.8 outside the ROI, strengthening the feature weights of energized regions within the ROI.

The improved CA attention mechanism is embedded in the C2f modules at the outputs of backbone layers 2, 3, and 4. This strengthens feature extraction for energized areas, personnel, and insulating equipment, especially when gloves or other items are small or partially occluded, which are documented difficulties in substation PPE detection [14,16]. The improved module uses about 30% of the parameters of the original CA implementation, so its effect on inference speed is limited.

3.3.3 Structured Pruning and INT8 Quantization for Electric Shock Prevention Target Perception

To further reduce model size and improve edge-side inference speed, a target-aware structured pruning method is introduced. Progressive multi-level distillation has also been applied to electrical-construction violation detection to reduce model complexity while preserving domain-specific detection performance [20]. Structured channel pruning is suitable for edge deployment because it reduces both computation and memory without requiring sparse-operation hardware [22]. Unlike generic pruning, the proposed strategy explicitly considers each channel’s contribution to personnel, insulating equipment, and live-equipment detection. The channel-importance score is defined as follows:

$$S_c = |\gamma_c| \cdot \frac{AP_c^{\text{target}}}{AP_c^{bg}} \tag{10}$$

where S_c is the importance score of the c convolutional channel; a larger value indicates a greater contribution of the channel to core target detection. γ_c is the scaling factor of the corresponding BN layer for this channel, reflecting the activation level of the channel. AP_c^{target} is the average accuracy of detecting core targets (personnel, insulating equipment, and live equipment) when this channel is activated. AP_c^{bg} is the average accuracy of the background category when this channel is activated.

The pruning implementation steps are: 1) Calculate the importance score S_c for each convolutional channel of the trained model; 2) Prune channels with S_c below the threshold, removing redundant channels; 3) Fine-tune the pruned model to restore model accuracy. This method removes 30% of redundant network channels and reduces model parameters by 25%, while limiting core-target accuracy loss to 0.28%, making it more effective than traditional BN-layer pruning.

After pruning, TensorRT INT8 quantization converts the trained 32-bit floating-point model into an 8-bit integer model. Quantization parameters are determined with a calibration dataset. This pruning-plus-quantization route follows established edge-detector deployment practice. With controllable accuracy loss, the model size is reduced to approximately one quarter of the original size, lowering computation and improving edge-device inference speed.

3.4 Pipeline Inference Optimization Strategy for Edge Adaptation

For ARM-based resource-constrained edge devices used in power-field applications, a multi-threaded pipelined inference strategy is designed to reduce serial-processing latency. A lightweight edge-AI framework for power-transmission corridors integrates network cameras, Jetson Nano devices, and adaptive frame sampling to provide local anomaly monitoring under embedded-resource constraints [15]. Unlike that adaptive frame-selection framework, the present implementation improves throughput through a four-stage producer–consumer pipeline executed entirely on the Jetson Nano. Edge PPE systems demonstrate local real-time analysis and alarm generation [18], while the progressively distilled electrical-construction violation detector in [20] was further evaluated on a Jetson Xavier NX, providing relevant embedded-deployment evidence for balancing domain-specific detection performance and computational efficiency. The optimal pipeline cycle is calculated as follows:

$$T_{\text{opt}} = \max\left(T_{\text{pre}}, T_{\text{inf}}, T_{\text{post}}\right), FPS_{\max} = \frac{1}{T_{\text{opt}}} \tag{11}$$

where T_{opt} represents the optimal cycle time for pipelined processing, in milliseconds; T_{pre}, T_{inf}, T_{post} represent the single-frame processing time for the preprocessing, inference, and postprocessing stages, respectively; $FPS_{\max}$ represents the theoretical maximum frame rate for pipelined parallel processing.

Video-stream processing is divided into four independent threads: video acquisition, preprocessing, model inference, and alert output. A producer-consumer pattern manages data exchange among the threads and forms a four-stage pipeline. When the stage times are balanced, the system frame rate is governed by the slowest stage rather than by the sum of all stage times; can improve throughput when the processing stages are appropriately balanced. The NEON instruction set, GPU acceleration, and inference-operator optimization are also used to improve hardware utilization on edge devices.

4. Experimental Design and Result Analysis

4.1 Experimental Environment and Dataset Description

4.1.1 Experimental Hardware and Software Environment

This study established both a PC-based simulation experimental environment and an edge-based deployment experimental environment. The detailed parameters are listed in Table 1. Model training was performed on a PC, whereas inference performance testing was conducted on both a PC and an edge device. The edge-based test fully simulated the actual deployment environment of a power field.

4.1.2 Dataset Construction

The self-built power electric-shock-prevention dataset was used as the primary experimental dataset. Images were collected from monitoring videos of multiple 220 kV substations, 10 kV distribution rooms, and outdoor transmission-line operation sites in China. The 12,000 valid images cover daytime, nighttime low light, backlight, light rain/fog, and occlusion. Five electric-shock-prevention targets were labeled: personnel, insulating gloves, insulating boots, safety helmets, and live equipment. The inclusion of insulating gloves and safety helmets is consistent with prior substation and electrical-construction PPE studies [14,16,20]. Personnel, insulating boots, and live equipment were additionally labeled to support intrusion detection, PPE-compliance assessment, and energized-area risk analysis specific to the present study. The dataset contains more than 38,600 labeled targets and was divided into training, validation, and test sets of 8,400, 1,800, and 1,800 images, respectively, following a 7:1.5:1.5 ratio. All annotations were completed according to the PASCAL VOC specification.

4.1.3 Evaluation Index System

Algorithm performance was evaluated from three dimensions: accuracy, lightweight performance, and real-time performance. The accuracy metrics included mAP@0.5, precision, recall, false-negative rate, and false-positive rate; lightweight metrics included parameter count, FLOPs, and model file size; real-time metrics included single-frame inference time, video-stream processing frame rate (FPS), and edge CPU utilization.

4.2 Model Training Details

The model was trained with transfer learning by initializing YOLOv8n weights pre-trained on the COCO dataset. The training settings were as follows: input image size, 640 × 640; batch size, 16; epochs, 100; optimizer, Adam; initial learning rate, 0.001; and cosine annealing for learning-rate scheduling. Mosaic augmentation, random flipping, random cropping, and brightness perturbation were used to improve generalization and robustness. After pruning and

Table 1. Experimental Environment Parameters

Environment type	Hardware configuration	Software environment
PC-based simulation environment	CPU: Intel i7-13700H, RAM: 32GB, GPU: NVIDIA RTX 4060 (8GB VRAM)	Operating system: Ubuntu 22.04; Deep learning frameworks: PyTorch 2.1.0, CUDA 12.1, cuDNN 8.9
Edge deployment environment	Core device: NVIDIA Jetson Nano (4-core ARM Cortex-A57 CPU, 128-core Maxwell GPU, 4 GB RAM)	Operating System: Ubuntu 20.04, JetPack 4.6.1, TensorRT 8.2, OpenCV 4.5.4

Table 2. Performance Comparison Results of Different Algorithms

Algorithm Model	Number of parameters (M)	FLOPs (G)	mAP@0.5 (%)	Model volume (MB)	Jetson Nano FPS (fps)
YOLOv5n	4.78	5.63	91.54	9.2	18.7
YOLOv7-tiny	6.21	6.12	92.03	12.1	17.2
YOLOv8n	3.94	4.47	93.59	7.6	22.6
MobileNetV2-SSD	4.12	3.82	88.76	8.1	20.3
EfficientDet-Lite0	3.88	4.21	90.25	7.3	19.5
The proposed algorithm	2.86	3.12	93.27	5.2	31.4

Table 3. Ablation Experiment Results

Experimental group	ROI extraction	Inter-frame filtering	Improve attention	Pruning Quantification	mAP@0.5 (%)	Jetson Nano FPS (fps)

Baseline Model	×	×	×	×	93.59	22.6
Experimental Group 1	√	×	×	×	93.57	27.3
Experimental Group 2	√	√	×	×	93.55	30.1
Experimental Group 3	√	√	√	×	93.82	29.4
Experimental Group 4 (complete proposed algorithm)	√	√	√	√	93.27	31.4

quantization, the model was fine-tuned for 10 additional epochs using the same settings to recover accuracy.

4.3 Comparative Experiment Design and Result Analysis

Five mainstream lightweight detection algorithms were selected as baselines to evaluate the comprehensive performance of the proposed algorithm. All algorithms were trained and tested under the same dataset and experimental environment. The comparison results are listed in Table 2.

Table 2 shows that the proposed algorithm has 2.86 M parameters, 27.4% fewer than YOLOv8n, and its FLOPs are reduced by 30.2%; the model size is 5.2 MB. For detection accuracy, its mAP@0.5 reaches 93.27%, only 0.32 percentage points lower than YOLOv8n. The AP for small targets such as insulating gloves and boots reaches 90.12%, which is 1.25 percentage points higher than YOLOv8n. On the Jetson Nano, the proposed algorithm runs at 31.4 fps, improving the frame rate by 38.6% over YOLOv8n and exceeding the 25 fps real-time requirement. CPU utilization remains at 58%, supporting stable operation on resource-constrained edge devices.

To test performance stability at different resolutions, the FPS of each algorithm was measured under multiple input resolutions (Figure 2). The horizontal axis represents input image resolution, and the vertical axis represents the average frame rate on the Jetson Nano. The proposed algorithm maintained the highest frame rate across resolutions, with a clearer advantage at higher resolutions.

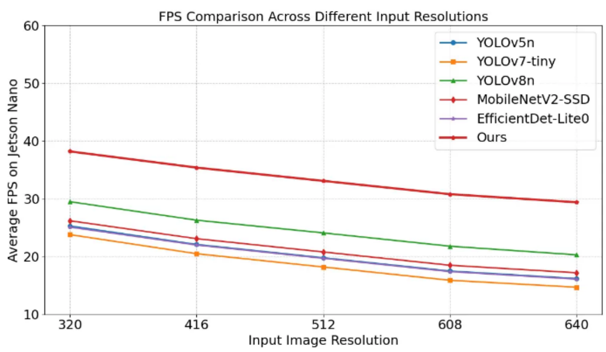

Figure 2. FPS variation curves of each algorithm under different input resolutions

To evaluate robustness in complex lighting, mAP variations of the proposed algorithm and YOLOv8n were measured under different illumination levels (Figure 3). Low-light enhancement research indicates that compact curve-estimation modules can improve downstream visual detection while preserving real-time execution. As illumination decreased, both algorithms lost accuracy, but the proposed algorithm showed the smaller decline. At 10 lux, it maintained an mAP@0.5 above 90%, indicating stronger low-light robustness than YOLOv8n.

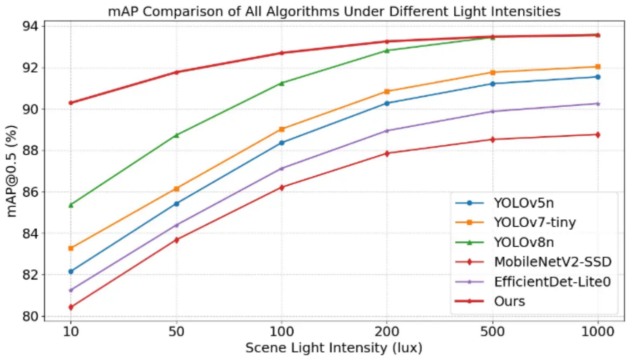

Figure 3. MAP variation curves of various algorithms under different light intensities

4.4 Ablation Experiment Design and Result Analysis

To verify the effectiveness of each innovative module proposed in this study, an ablation experiment was conducted based on the optimized lightweight YOLOv8n basic model. Each innovative module was added individually, and its impact on algorithm performance was tested. The results are presented in Table 3.

The ablation results in Table 3 show that ROI extraction increased the frame rate from 22.6 fps to 27.3 fps while reducing mAP by only 0.02 percentage points. Adding inter-frame redundancy filtering further increased the frame rate to 30.1 fps with the same small mAP loss, confirming that redundant static frames were effectively filtered. Embedding the improved CA attention mechanism increased mAP to 93.82%, while reducing the frame rate by only 0.7 fps. After pruning and quantization, the frame rate rose to 31.4 fps and mAP decreased by 0.55 percentage points, which remained within the acceptable accuracy range for the target application.

To analyze the effect of pruning rate, mAP and FPS were measured under different pruning ratios (Figure 4). The horizontal axis represents pruning rate, the left vertical axis represents mAP@0.5, and the right vertical axis represents FPS on the Jetson Nano. The best trade-off between accuracy and speed was obtained at a pruning rate of 30%.

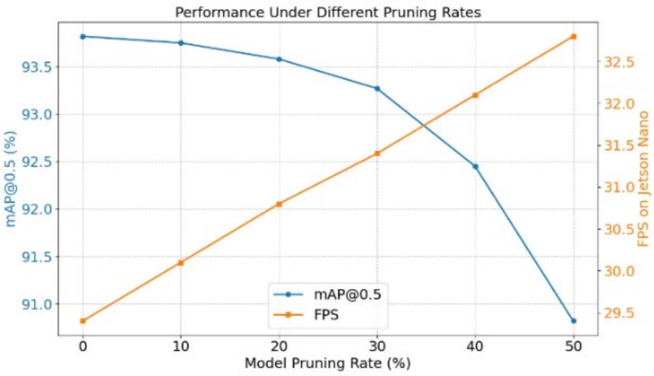

Figure 4. MAP and FPS variation curves of the algorithm under different pruning rates

To verify the performance of the inter-frame filtering module, the paper tested the single-frame processing latency under different frame change rates, and the results are shown in Figure 5. The horizontal axis of Figure 5 represents the frame change rate of the ROI, and the vertical axis represents the single-frame processing latency. When the change rate is less than 2%, the latency with the inter-frame filtering module is only one-third of that without the module, demonstrating a significant advantage.

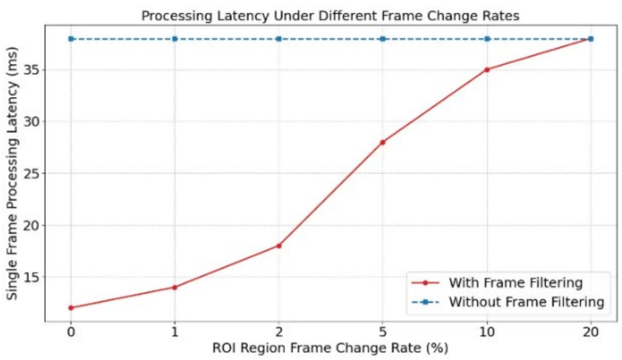

Figure 5. Average processing delay curves of a single frame under different frame change rates

4.5 Edge Deployment and Substation Field Pilot Test Verification

4.5.1 Field Test Environment and Scheme Design

To evaluate engineering feasibility in real power-operation scenarios, a 15-day field pilot test was conducted at a 220 kV smart substation in China. The test covered sunny weather, light rain, moderate rain, dense fog, and nighttime low-light conditions, as well as typical electric-shock prevention scenarios, including personnel intrusion, protective-equipment wearing detection, and partial occlusion by equipment during operation.

The on-site deployment plan was as follows. On the hardware side, an NVIDIA Jetson Nano edge-computing box was connected to three monitoring points in the existing substation video-surveillance system: the 110 kV gas-insulated switchgear (GIS) area, the energized area on the high-voltage side of the main transformer, and the outdoor switchyard. All monitoring points used fixed bullet cameras with a resolution of 1920 × 1080 and a frame rate of 25 fps. On the software side, the complete proposed algorithm and the baseline algorithms (YOLOv5n, YOLOv7-tiny, YOLOv8n, MobileNetV2-SSD, and EfficientDet-Lite0) were deployed on the edge box to establish a local processing flow of video acquisition, preprocessing, inference detection, and warning output. Warning signals were synchronized through the substation intranet to the handheld terminal of operation and maintenance personnel and to the central control room monitoring platform. During testing, raw video streams and algorithm outputs were collected simultaneously to compare detection performance in real field scenarios. The key metrics were mAP@0.5, false-negative rate, false-positive rate, average warning latency, and continuous-operation stability.

4.5.2 Basic Performance Testing Under Normal Conditions

Under normal daylight conditions, a 72-h continuous operation test and basic performance verification were conducted for the proposed algorithm and the baseline algorithms. The proposed algorithm maintained an average frame rate above 30 fps on real surveillance streams, with edge-box CPU utilization between 55% and 60% and memory utilization below 60%. No crashes, stuttering, or abnormal exits occurred during continuous operation. Among the baselines, YOLOv8n performed best, with an average frame rate of 22 fps and CPU utilization above 75%. The other baselines all remained below 20 fps, and YOLOv7-tiny and MobileNetV2-SSD showed frame-rate fluctuation and stuttering during prolonged operation.

For the electric-shock prevention task, the proposed algorithm achieved 99.2% accuracy for personnel intrusion detection and 98.7% accuracy for protective-equipment wearing detection, with an average warning latency of 126 ms. These results satisfy the real-time warning requirements for on-site violations in electric power safety work regulations. By comparison, YOLOv8n achieved 97.5% accuracy for personnel intrusion and 96.3% for protective-equipment detection, with an average warning latency of 158 ms. YOLOv5n and YOLOv7-tiny achieved 95.8% and 96.1% accuracy, respectively, with warning latencies above 180 ms. MobileNetV2-SSD and EfficientDet-Lite0 remained below 95% accuracy and had false-negative rates above 2%, which is insufficient for the target field-control requirements.

4.5.3 Specific Testing and Result Analysis for Complex Scenarios

Comparative tests were then conducted for common field interference scenarios, including rain, fog, target occlusion, and low light. The test conditions and performance results are as follows:

4.5.4 Rain and Fog Weather Scenario Test

Video clips under three typical rain/fog conditions (light rain, moderate rain, and dense fog) were selected from the substation site. For each condition, 200 valid 10 s clips were used to evaluate the proposed algorithm and all baseline algorithms. The results are listed in Table 4, and representative on-site detection results are shown in Figure 6.

Table 4. Algorithm Performance Comparison under Rain and Fog Weather Scenarios

Test Scenario	Algorithm Model	mAP@0.5 (%)	Missed report rate (%)	False alarm rate (%)
Light rain	YOLOv5n	87.23	3.15	3.89
	YOLOv7-tiny	86.75	3.52	4.12
	YOLOv8n	89.56	2.34	3.12
	MobileNetV2-SSD	83.47	4.28	5.03
	EfficientDet-Lite0	85.69	3.86	4.57
	The proposed algorithm	92.41	0.65	1.08
Moderate rain	YOLOv5n	84.15	4.32	4.98
	YOLOv7-tiny	83.67	4.76	5.23
	YOLOv8n	87.18	3.57	4.26
	MobileNetV2-SSD	80.32	5.69	6.15
	EfficientDet-Lite0	82.76	5.13	5.78
	The proposed algorithm	91.73	0.79	1.34
Dense fog	YOLOv5n	82.37	5.89	6.21
	YOLOv7-tiny	81.89	6.32	6.57
	YOLOv8n	86.24	4.82	5.19
	MobileNetV2-SSD	78.54	7.15	7.89
	EfficientDet-Lite0	80.92	6.78	7.23
	The proposed algorithm	91.35	0.87	1.52

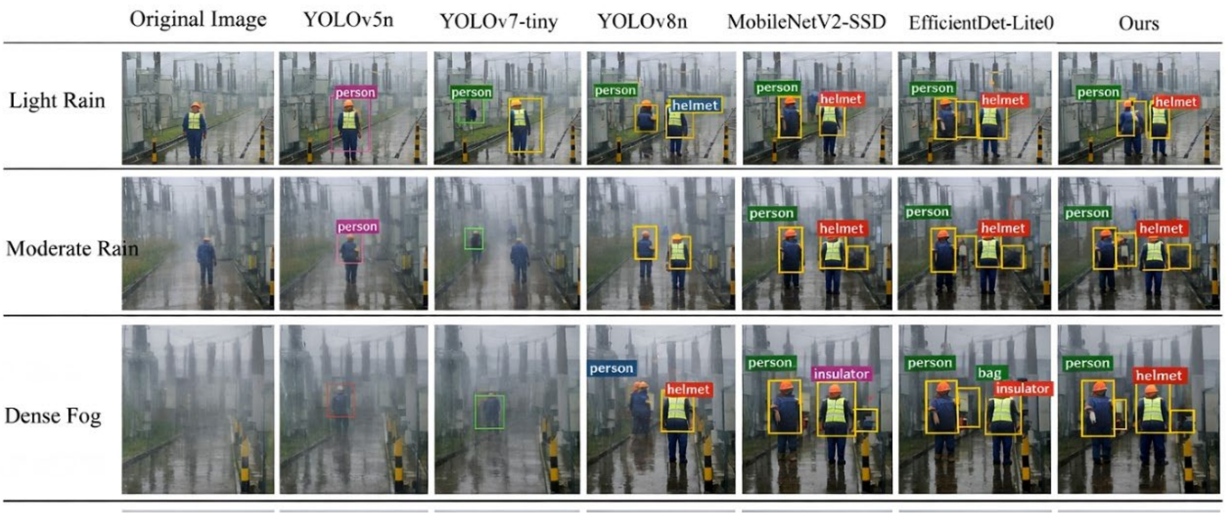

Figure 6. Visualization of detection results for various algorithms in rainy and foggy weather scenarios

Figure 6 shows the detection results of various algorithms in rainy and foggy weather scenarios. The horizontal row shows the original image and the detection results of different algorithms in the same scene, from left to right: original scene monitoring image, YOLOv5n, YOLOv7-tiny, YOLOv8n, MobileNetV2-SSD, EfficientDet-Lite0, and the algorithm proposed in this study; the vertical column shows the detection results of the same algorithm in different rainy and foggy scenarios, from top to bottom: light rain, moderate rain, and dense fog. In light rain scenarios, all algorithms could complete basic detection; however, MobileNetV2-SSD and EfficientDet-Lite0 exhibited slight false detections. In moderate rain scenarios, YOLOv5n and YOLOv7-tiny missed the detections of insulated equipment, and the confidence of YOLOv8n's detection box decreased. In dense fog scenarios, MobileNetV2-SSD and YOLOv7-tiny missed the detections of personnel, and YOLOv8n missed the detections of insulated safety helmets. Only the proposed algorithm could accurately identify workers and insulated equipment in energized areas and successfully trigger a red alert for personnel intrusion, without any missed or false detections.

The results show that rainy and foggy weather reduced the detection accuracy of all comparison models. In dense fog, MobileNetV2-SSD reached only 78.54% mAP and a false-negative rate above 7%, which does not satisfy power-safety monitoring requirements. YOLOv8n, the strongest baseline, also showed an mAP decrease of more than 7% and a false-negative rate close to 5%. By contrast, the proposed algorithm uses a lightweight image-enhancement module before detection, which suppresses rain/fog interference. Even in dense fog, its mAP remained above 91% and its false-negative rate stayed below 1%, indicating stronger scene robustness.

4.5.5 Target Occlusion Scenario Testing

To evaluate occlusion robustness, three common substation scenarios were simulated: mild occlusion (target occlusion area <30%), moderate occlusion (30%-60%), and severe occlusion (>60%). For each scenario, 150 samples were used to compare the proposed algorithm with the baseline algorithms. The results are presented in Table 5, and representative on-site detection results are shown in Figure 7.

Figure 7 compares detection results under different occlusion levels. Each row corresponds to mild, moderate, or severe occlusion, and each column compares the original image, YOLOv5n, YOLOv7-tiny, YOLOv8n, MobileNetV2-SSD, EfficientDet-Lite0, and the proposed algorithm. Under mild occlusion, all algorithms identified workers, but MobileNetV2-SSD and EfficientDet-Lite0 missed insulating gloves. Under moderate occlusion, YOLOv5n and YOLOv7-tiny misclassified protective-equipment-wearing status, while MobileNetV2-SSD and EfficientDet-Lite0 missed parts of the occluded worker. Under severe occlusion, YOLOv5n, YOLOv7-tiny, and MobileNetV2-SSD missed the main body of the worker, and YOLOv8n detected only partial personnel features while missing insulating equipment. The proposed algorithm identified partially occluded workers and insulating gloves more reliably and marked the nearby energized equipment. The results show that detection performance declined as occlusion severity increased. MobileNetV2-SSD achieved only 70.23% recall for personnel in heavily occluded scenes, creating a high false-negative risk. Even the strongest baseline, YOLOv8n, had a personnel recall below 80% under heavy occlusion. The proposed algorithm, with ROI-aware CA attention, retained stronger feature extraction for personnel and insulating equipment; its personnel recall stayed above 90% in heavily occluded scenes, reducing the false-negative risk in on-site safety monitoring.

Table 5. Algorithm Performance Comparison under Target Occlusion Scenarios

Test Scenario	Algorithm Model	mAP@0.5 (%)	Personnel testing recall rate (%)
Mild occlusion	YOLOv5n	90.23	93.15
	YOLOv7-tiny	89.76	92.48
	YOLOv8n	92.17	95.32
	MobileNetV2-SSD	86.54	88.76
	EfficientDet-Lite0	88.37	90.54
	The proposed algorithm	93.05	98.64
Moderate occlusion	YOLOv5n	86.15	87.32
	YOLOv7-tiny	85.67	86.54
	YOLOv8n	88.42	89.76
	MobileNetV2-SSD	82.18	83.45
	EfficientDet-Lite0	84.29	85.78
	The proposed algorithm	92.18	96.37
Severe occlusion	YOLOv5n	80.24	76.15
	YOLOv7-tiny	79.87	75.32
	YOLOv8n	82.35	78.41
	MobileNetV2-SSD	76.54	70.23
	EfficientDet-Lite0	78.69	73.56
	The proposed algorithm	89.62	91.25

Figure 7. Visualization of Detection Results of Various Algorithms in Target Occlusion Scenes

4.5.6 Nighttime Low-Light Scene Test

For nighttime substation inspection and emergency operation scenarios, three low-light conditions were tested: no supplementary lighting, weak supplementary lighting, and backlighting. For each condition, 200 valid video clips were used to compare the proposed algorithm with the baseline models. Figure 8 presents representative detection results. Each row corresponds to a different low-light condition, and the columns compare the original image, YOLOv5n, YOLOv7-tiny, YOLOv8n, MobileNetV2-SSD, EfficientDet-Lite0, and the proposed algorithm. In low-light scenes, YOLOv5n and YOLOv7-tiny produced shifted personnel boxes, while MobileNetV2-SSD and EfficientDet-Lite0 missed insulating equipment. Under backlighting, YOLOv8n missed part of the personnel target, and the other baselines produced false detections. Without supplementary lighting, MobileNetV2-SSD and YOLOv7-tiny failed to identify workers, and YOLOv8n missed violations such as missing insulating gloves. The proposed algorithm still identified worker violations and marked nearby energized equipment.

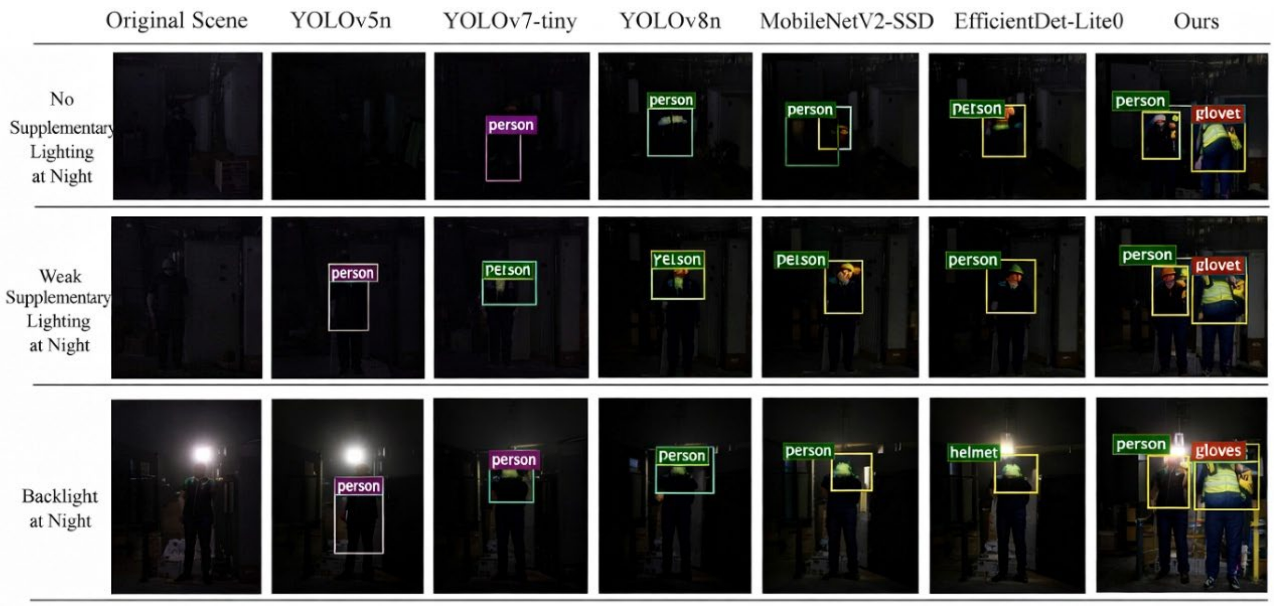

Figure 8. Detection results of various algorithms in low-light nighttime scenes

Under extremely low light without supplemental lighting, all baselines showed clear degradation: MobileNetV2-SSD reached only 80.12% mAP with a false-negative rate above 8%; YOLOv7-tiny reached 81.35% mAP with a 7.69% false-negative rate; YOLOv5n reached 83.27% mAP with a 6.45% false-negative rate; EfficientDet-Lite0 reached 84.19% mAP with a 5.87% false-negative rate; and YOLOv8n reached 85.37% mAP with a false-negative rate above 5%. By contrast, the proposed adaptive gamma-correction module improved target contrast in low-light images, enabling an mAP of 90.28% and a false-negative rate of 0.94%.

4.5.7 Summary of Overall Results of On-site Pilot Application

The substation pilot processed more than 360 h of monitoring video and covered typical operating conditions and field interference. The proposed algorithm was compared with YOLOv5n, YOLOv7-tiny, YOLOv8n, MobileNetV2-SSD, and EfficientDet-Lite0. In real substation scenarios, it achieved an average mAP@0.5 of 92.14% and an average warning latency of 126 ms. On the Jetson Nano edge device, the processing frame rate remained above 30 fps, satisfying the precision, latency, and stability requirements of electric-shock prevention monitoring. Compared with YOLOv8n, the frame rate improved by 38.6% under normal conditions; compared with all baselines, mAP improved by 5%-13% in complex scenarios, and the false-negative rate remained below 1%. The algorithm can be deployed on existing substation monitoring front-end devices without hardware upgrades, giving it practical engineering value for field application.

5. Conclusion

This study proposes a safety-prioritized lightweight video-stream processing algorithm for electric-shock prevention monitoring in smart-grid edge scenarios. The method integrates scenario-aware preprocessing, compact detection-network optimization, and edge inference acceleration to address the difficulty of balancing detection accuracy, warning latency, and deployment efficiency on resource-constrained edge devices. Experiments show that the proposed model reduces parameters and computational cost while maintaining detection accuracy. It reaches 31.4 fps on a Jetson Nano, and substation field tests achieve an average mAP@0.5 of 92.14% with an average warning latency of 126 ms. The method also maintains robust detection in rain, fog, low-light, and occlusion scenarios. The present work still has limitations: the algorithm is optimized mainly for fixed monitoring cameras, its performance in extreme weather such as heavy rain and dense fog requires further study, and the covered illegal-operation behaviors remain limited. Future work will improve adaptability to mobile operation scenarios, introduce multimodal data fusion for extreme environments, expand behavior-recognition categories, and test cross-hardware deployment.

Overall, the method offers a practical edge-intelligence path for electric-shock prevention in smart grids and Energy Internet applications. Its local perception, low-latency warning, and robustness under field interference can support safer and more scalable operation and maintenance. Future work will examine multi-camera collaborative perception, federated model updating, and integration with smart-grid digital twins.

References

[1] Chen S, Wang D, Zhang X, Shao B, Cao K, Li Z. A spatiotemporal analysis of personal casualty accidents in China's electric power industry. Heliyon. 2024;10(13):e33855. doi: 10.1016/j.heliyon.2024.e33855.
[2] de Souza DF, Martins WA, Martinho E, Tatizawa H. Accidents leading to electrical shocks in Brazilian electric power distribution: An analysis. IEEE Ind Appl Mag. 2024;30(3):61–67. doi: 10.1109/MIAS.2023.3328507.
[3] Yin Y, Lin J, Sun N, Zhu Q, Zhang S, Zhang Y, et al. Method for detection of unsafe actions in power field based on edge computing architecture. J Cloud Comput. 2021;10:17. doi: 10.1186/s13677-021-00234-w.
[4] Zhao B, Lan H, Niu Z, Zhu H, Qian T, Tang W. Detection and location of safety protective wear in power substation operation using wear-enhanced YOLOv3 algorithm. IEEE Access. 2021;9:125540–125549. doi: 10.1109/ACCESS.2021.3104731.
[5] Chen B, Wang X, Bao Q, Jia B, Li X, Wang Y. An unsafe behavior detection method based on improved YOLO framework. Electronics. 2022;11(12):1912. doi: 10.3390/electronics11121912.
[6] Lema DG, Usamentiaga R, García DF. Low-cost system for real-time verification of personal protective equipment in industrial facilities using edge computing devices. J Real-Time Image Process. 2023;20(6):111. doi: 10.1007/s11554-023-01368-7.
[7] Chen Z, Ma C, Ren J, Hao F, Wang Z. Research on the identification method of safety wearing of electric power workers based on deep learning. Front Energy Res. 2023;10:1091322. doi: 10.3389/fenrg.2022.1091322.
[8] Fang J, Li X. Object detection related to irregular behaviors of substation personnel based on improved YOLOv4. Appl Sci. 2022;12(9):4301. doi: 10.3390/app12094301.
[9] Lin Z, Chen W, Su L, Chen Y, Li T. HS-YOLO: Small object detection for power operation scenarios. Appl Sci. 2023;13(19):11114. doi: 10.3390/app131911114.
[10] Long X, Zheng Z, Liu R, Cui W, Chi Y, Zhang H, et al. Cascaded feature enhancement network model for real-time video monitoring of power system. Energy Rep. 2021;7:8485–8492. doi: 10.1016/j.egyr.2021.05.046.
[11] Ali Z, Park U. Real-time safety monitoring vision system for linemen in buckets using spatio-temporal inference. Int J Control Autom Syst. 2021;19(1):505–520. doi: 10.1007/s12555-019-0546-y.
[12] Liu C, Zhang W, Xu W, Lu B, Li W, Zhao X. Substation inspection safety risk identification based on synthetic data and spatiotemporal action detection. Sensors. 2025;25(9):2720. doi: 10.3390/s25092720.
[13] Zhao M, Barati M. Substation safety awareness intelligent model: Fast personal protective equipment detection using GNN approach. IEEE Trans Ind Appl. 2023;59(3):3142–3150. doi: 10.1109/TIA.2023.3234515.

[14] Zhang H, Mu C, Ma X, Guo X, Hu C. MEAG-YOLO: A novel approach for the accurate detection of personal protective equipment in substations. Appl Sci. 2024;14(11):4766. doi: 10.3390/app14114766.
[15] Nguyen DC, Nguyen TC. AI-powered edge-based safety monitoring system for power transmission corridors: A case study in Vietnam. Comput Netw. 2026;275:111881. doi: 10.1016/j.comnet.2025.111881.
[16] Tao C, Wang C, Li T. Detection research of insulating gloves wearing status based on improved YOLOv8s algorithm. J Eng Appl Sci. 2024;71(1):126. doi: 10.1186/s44147-024-00458-y.
[17] Li J, Zhao X, Zhou G, Zhang M. Standardized use inspection of workers' personal protective equipment based on deep learning. Saf Sci. 2022;150:105689. doi: 10.1016/j.ssci.2022.105689.
[18] Gallo G, Di Rienzo F, Garzelli F, Ducange P, Vallati C. A smart system for personal protective equipment detection in industrial environments based on deep learning at the edge. IEEE Access. 2022;10:110862–110878. doi: 10.1109/ACCESS.2022.3215148.
[19] Wang L, Wang B, Zhang J, Ma H, Luo P, Yin T. An intelligent detection method for approach distances of large construction equipment in substations. Electronics. 2023;12(16):3510. doi: 10.3390/electronics12163510.
[20] Ma B, Liang G, Rao Y, Guo W, Zheng W, Wang Q. Knowledge reasoning- and progressive distillation-integrated detection of electrical construction violations. Sensors. 2024;24(24):8216. doi: 10.3390/s24248216.
[21] Hou Q, Zhou D, Feng J. Coordinate attention for efficient mobile network design. In: Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2021 Jun 20–25; Nashville, TN, USA (virtual conference). p. 13708–13717. doi: 10.1109/CVPR46437.2021.01350.
[22] Choi K, Wi SM, Jung HG, Suhr JK. Simplification of deep neural network-based object detector for real-time edge computing. Sensors. 2023;23(7):3777. doi: 10.3390/s23073777.