


Article

Research on AI-Assisted Fire Risk Target Detection for Special Operating Conditions in Under-Construction Nuclear Power Plants

Zhendong Li ¹, Guangwei Liu ¹, Kai Yu ^{2,*} and Shijie Du ² 

¹ Zhangzhou Project Team, China Nuclear Power Engineering Co., Ltd., Zhangzhou 363300, China; lizda@cnpe.cc (Z.L.); liugw@cnpe.cc (G.L.)

² School of Safety and Environmental Engineering, Shandong University of Science and Technology, Qingdao 266590, China; 13139547391@163.com

* Correspondence: ykxfkd@sdust.edu.cn; Tel.: +86-13969849358

Abstract

In night-time construction scenarios of under-construction nuclear power plants, some yellow lights and open flames exhibit highly similar visual characteristics, resulting in frequent false alarms of fire sources. Such false alarm information tends to drown out real fire alarm signals, which not only severely disrupts construction operations but also endangers fire safety. To address this problem, this paper proposes an intelligent fire risk identification method based on an enhanced YOLOv8n (named YOLO-Fire). Specifically, shallow convolutional layers embedded with a coordinate attention mechanism are integrated into the Backbone of YOLOv8n; the Neck is optimised to improve the efficiency of multi-scale feature fusion; and the Head is enhanced to strengthen the localization and classification branches. Additionally, a composite loss function combining classification loss, regression loss, and similarity loss is designed, coupled with night-scene-specific data augmentation techniques and a two-stage progressive training strategy. Experimental results show that YOLO-Fire reduces the false alarm rate by 14.3%, increases the mean average precision (mAP@0.5) for open flames by 11.3% to 75.2%, and maintains an inference speed of over 85 frames per second (FPS). This study achieves an optimal balance between false alarm control, small object detection accuracy, and real-time processing efficiency, effectively resolving the misclassification issue between open flames and lights in night-time construction scenarios, and providing precise and efficient intelligent technical support for fire risk prevention and control during the construction phase of nuclear power plants.

Keywords: YOLO-Fire; deep learning; fire; nuclear power plant



Academic Editors: Haowei Yao, Xueming Shu, Xuecai Xie and Jun Hu

Received: 2 February 2026

Revised: 24 February 2026

Accepted: 2 March 2026

Published: 3 March 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

1.1. Research Background and Significance

Nuclear energy, as a clean and efficient form of power generation, relies on its safe and stable operation for the protection of the ecological environment and public welfare. Fire, being one of the most significant safety hazards throughout the entire operational lifecycle of a nuclear power plant, remains a core issue in nuclear safety prevention and control. The International Atomic Energy Agency (IAEA) explicitly states in its Nuclear Safety Standards series that thermal radiation, smoke propagation, and structural damage caused by fires may directly compromise nuclear safety barriers. In extreme circumstances, this

could lead to radioactive material leakage, causing long-term impacts on the surrounding environment [1].

Compared to operational nuclear power plants, fire prevention and control at nuclear power plants under construction face more severe and unique challenges. During construction, the nuclear island, conventional island, and ancillary facilities remain in an assembly state, presenting significant temporary and complex characteristics on-site. On one hand, construction zones involve extensive hot work such as welding and cutting. High-temperature welding spatter readily ignites nearby flammable materials like glass fibre insulation and paint. Issues such as improperly installed fire extinguishers and damaged emergency exit signage in certain areas further amplify the risk of fire propagation. Concurrently, the temporary firefighting system remains incomplete, with critical evacuation support facilities—such as positive pressure ventilation in stairwells within the nuclear island zone—not yet operational. Should a fire occur, high-temperature smoke could accumulate within the complex structure, hindering personnel evacuation and restricting firefighting and rescue routes, thereby significantly increasing the difficulty of incident response [2].

Night-time construction is a common practice for nuclear power plants under construction to meet deadlines and avoid high temperatures. However, low-light conditions present unique technical challenges for fire risk identification. Night operations rely on artificial light sources such as tower crane illumination and handheld lamps, which distribute light unevenly and create intense shadows and glare, significantly obscuring key fire indicators. A more pronounced issue arises from the yellow warning lights atop construction vehicles frequently traversing the site. Under low-light conditions, these lights exhibit highly similar visual characteristics to open flames—both manifesting as high-intensity point light sources. Furthermore, the dynamic flickering behaviour of both lights during vehicle movement creates visual confusion, rendering traditional detection systems ineffective at distinguishing between them [3]. This confusion of targets directly leads to severe false alarms in existing fire detection technologies during night-time operations: misidentifying the yellow roof lights as actual flames not only consumes limited fire emergency resources and disrupts normal construction operations, but may also cause personnel to become complacent towards alarm signals, potentially delaying response times when genuine fires occur [4].

Traditional fire detection methods are increasingly inadequate for meeting the specific night-time operational requirements of nuclear power plants under construction. Contact-based sensors relying on smoke and temperature exhibit slow response times and are prone to false alarms due to environmental factors such as construction dust and high-temperature operations. Conventional computer vision detection models, while offering real-time processing advantages, have not undergone customised optimisation for nuclear power plant construction environments. They exhibit inadequate feature extraction capabilities for small targets under low-light conditions and lack mechanisms to distinguish between objects with similar spectral signatures or morphological characteristics, resulting in persistently high false alarm rates [5]. Therefore, in response to the unique operational conditions of night-time construction at nuclear power plants under construction, the development of intelligent fire risk identification technology capable of precisely distinguishing between open flames and yellow roof lights not only fills the gap in existing detection methods for complex scenarios but also provides reliable technical support for safety management during the nuclear engineering construction phase. This holds significant engineering practicality and nuclear safety value for ensuring construction safety at nuclear power plants and reducing losses from fire incidents [6].

1.2. Research Question Formulation

In night-time construction scenarios at nuclear power plants under construction, false alarms triggered by vehicle roof lights and open flames are not coincidental occurrences. Rather, they result from the combined effects of target visual characteristics, environmental conditions, and existing technological limitations. Analysing target features reveals that the yellow warning lights atop engineering vehicles exhibit a high degree of visual similarity to open flames under low-light night-time conditions. Both manifest as high-intensity point light sources, with the luminous intensity of yellow warning lights exceeding 5000 cd, placing them within the same order of magnitude as the radiant intensity of open flames. Furthermore, during vehicle movement, the dynamic trajectory of warning lights and the flickering nature of flames create visual confusion, making it difficult for the human eye to rapidly distinguish between them under low-light conditions.

From an operational context perspective, numerous interfering factors exist within the night-time construction zones of nuclear power plants under construction. The metallic reflections from construction equipment, the haphazard distribution of temporary cables, and the intermittent flashes of welding sparks intertwine with the visual characteristics of yellow lights and open flames, further complicating target identification. Concurrently, night-time construction lighting systems predominantly employ localised intense light sources, resulting in severe uneven illumination and glare within images. This significantly obscures the edge contours and textural details of both warning lights and open flames, rendering traditional appearance-based recognition algorithms highly prone to failure.

From a technical shortcomings' perspective, existing fire detection models exhibit significant deficiencies. While mainstream YOLO series models demonstrate excellent performance in general object detection, they have not undergone customised enhancements for nuclear power plant construction scenarios [7,8]. On the one hand, the model exhibits insufficient capability for feature extraction from small targets. The yellow roof lights and open flames in the images predominantly constitute minute objects with pixel dimensions below 32×32 . The shallow feature extraction modules of existing models are unable to adequately capture their detailed information. On the other hand, the model lacks a mechanism for distinguishing between similar targets. The spectral differences between yellow lights and open flames—such as variations in colour and brightness—are not effectively utilised. Consequently, the classifier struggles to learn feature representations with sufficient discriminative power [9].

Existing object detection technologies demonstrate severe inadequacies in distinguishing between open flames and roof-mounted yellow lights when confronted with extremely small targets in low-light conditions, visually indistinguishable objects of different categories, and complex construction site backgrounds [10]. Concurrently, existing technologies lack customised optimisation solutions tailored to night-time construction scenarios at nuclear power plants under construction, rendering them incapable of accommodating the unique demands of such environments. The core challenge addressed by this research lies in overcoming the dual constraints of visual similarity and environmental complexity to develop an intelligent detection method capable of accurately identifying open flames while effectively eliminating interference from yellow roof lights.

1.3. Research Objectives and Innovative Contributions

This research aims to develop a precise and efficient fire risk identification technology tailored to the unique operational conditions of night-time construction at nuclear power plants under construction. The primary objective is to construct a detection model based on an enhanced YOLOv8n architecture, enabling accurate classification between open flames and roof-mounted yellow lights. This addresses the technical challenge of false alarms

caused by confusion between the two. Secondly, it seeks to enhance the model's capability to recognise small-sized targets in low-light environments and complex construction settings, ensuring high detection accuracy even when targets occupy a small proportion of the image or exhibit weakened visual features. Finally, it aims to balance model performance with real-time capability, optimising detection effectiveness while controlling computational complexity. This addresses the engineering requirement for real-time monitoring at nuclear power plant construction sites, providing reliable technical support for fire prevention and control during the construction phase.

2. Materials and Methods

Advancements in fire detection technology have centred on enhancing accuracy, reducing false alarm rates, and adapting to complex scenarios. This evolution has progressed from traditional contact-based detection to intelligent computer vision detection. Breakthroughs in small object detection and similar object discrimination techniques have laid the foundation for the application of YOLO series models in specialised industrial settings.

Conventional fire detection technologies rely on contact sensors, triggering alarms based on physical parameters such as smoke concentration and temperature changes. While these systems feature simple structures and lower costs, they exhibit significant limitations in nuclear power plant construction environments. Response times are affected by environmental medium conduction, making it difficult to rapidly capture early-stage fire signals. They are also susceptible to interference from environmental factors like construction dust and high-temperature operations, resulting in persistently high false alarm rates. Furthermore, their detection range is limited, failing to achieve comprehensive coverage in open construction spaces. As industrial settings demand greater real-time accuracy in fire detection, traditional contact-based technologies struggle to meet the complex prevention requirements of construction environments. This has driven the evolution of fire detection technology towards non-contact and intelligent solutions [11–13].

Computer vision-based fire detection technology has emerged as a research hotspot owing to its non-contact, large-scale monitoring capabilities. Early approaches relied on manually designed features, distinguishing open flames from backgrounds through colour, texture, and motion characteristics; however, these methods lacked robustness and were susceptible to variations in lighting and background interference. Following the advent of deep learning, detection methods based on convolutional neural networks gradually supplanted manual feature approaches. By enabling networks to autonomously learn deep visual features of fires, these techniques significantly enhanced detection performance. Object detection models such as Faster R-CNN, SSD, and YOLO achieved the localisation and classification of flames and smoke. In recent years, Transformer-based models and multimodal fusion techniques have further expanded the boundaries of application [14–17]. Multimodal fusion enhances detection accuracy by integrating multi-source data such as visible light and infrared, leveraging the distinct spectral characteristics of fires across different wavelengths. However, existing computer vision methods still face challenges under specific operational conditions. Factors including low illumination, small target dimensions, and the coexistence of objects with similar visual features can lead to inadequate feature extraction, resulting in missed detections and false alarms. This makes it difficult to meet the precise detection requirements for night-time construction scenarios at nuclear power plants under construction.

Small object detection and the differentiation of similar objects represent critical technical challenges within the field of computer vision. Research in this area provides vital support for fire detection in specialised scenarios. The core challenge in small object detection lies in the low pixel proportion of the target and the scarcity of detailed features.

Existing solutions focus on feature enhancement, multi-scale fusion, and data optimisation [18]. Feature enhancement techniques strengthen detail representation by deepening shallow networks or introducing attention mechanisms. Mechanisms such as coordinate attention and CBAM can improve the model's focus on small target regions [19]. Multi-scale fusion, exemplified by structures such as FPN, PAN, and BIFPN, achieves deep integration of high-level semantic and low-level detailed features by constructing feature pyramids. The lightweight SPP-Net is also widely employed to capture features of small objects at different scales. Similarity-based target discrimination focuses on extracting intrinsic feature differences. Multimodal fusion techniques leverage spectral characteristics across different light spectra for classification. Metric learning enhances discriminative power by optimising feature embedding spaces. Methods such as contrastive learning and triplet loss have been applied in relevant tasks [20–23]. These techniques offer potential solutions to the visual confusion between open flames and yellow roof lights. However, under the specific night-time operating conditions of nuclear power plants under construction, the key unresolved challenge remains how to efficiently adapt lightweight models by integrating both types of technology.

The YOLO series models, with their end-to-end detection architecture, high real-time performance, and balanced accuracy, are widely applied in industrial settings. From YOLOv1 to YOLOv8, the models have continuously evolved towards higher accuracy, faster speeds, and lighter weight [24–26]. As the lightweight variant within this family, YOLOv8n boasts a parameter count of merely 3.2 million and computational requirements of 8.7 gigaflops, conferring significant advantages for deployment on edge devices. The backbone employs a C2f module, enhancing feature extraction efficiency through parallel convolutions and residual connections. The neck utilises PAN-FPN for bidirectional feature fusion, adapting to multi-scale object detection. The head adopts a unified classification and regression design, combined with an anchor-free detection mechanism to simplify the training process. The loss function employs a combination of CIoU and Focal Loss to mitigate sample imbalance issues [27,28].

YOLOv8n has been successfully deployed in applications such as construction site safety monitoring, underground personnel detection, and forest fire prevention, demonstrating strong adaptability in scenarios demanding high real-time performance. Certain studies have enhanced its performance in specialised contexts by modifying its architecture. For instance, in construction site scenarios, refining the PANet architecture and incorporating the SPPF module increased small object detection recall by 22% [29]. In complex underground environments, the introduction of a multi-scale spatial attention enhancement mechanism and an adaptive spatial feature fusion module enhances the model's detection capabilities for uneven lighting and obscured targets. However, this model exhibits significant limitations in night-time construction scenarios at nuclear power plants under construction. Its shallow feature extraction capability fails to adequately capture the fine details of small open flames and roof-mounted yellow lights. The efficiency of multi-scale fusion remains limited, and the model lacks a mechanism to distinguish between targets with similar visual features. Consequently, its adaptability to low-light conditions, glare, and complex backgrounds is insufficient [30–32]. Existing research on YOLOv8n improvements has predominantly focused on enhancing performance in general scenarios, lacking customised optimisation for the specific operational conditions of nuclear power plants. It has not sufficiently integrated small object detection and similar object discrimination techniques, thereby failing to effectively resolve false alarms caused by open flames and roof-mounted yellow lights [33].

3. Results

3.1. Improved YOLOv8n Model Architecture

To address the detection requirements for small open flames and rooftop yellow lights during night-time construction at nuclear power plants under construction, the YOLOv8n network architecture has been customised with three core objectives: enhancing small target feature extraction, optimising the ability to distinguish similar targets, and improving adaptability in complex environments. By focusing on the three core modules—Backbone, Neck, and Head—and employing a layered design featuring shallow feature enhancement, multi-scale fusion optimisation, and detection branch adaptation, the approach specifically resolves the original model’s limitations in scenarios involving low illumination, small targets, and the coexistence of targets with similar visual characteristics. Neck, and Head modules, we implemented a layered design featuring shallow feature enhancement, multi-scale fusion optimisation, and detection branch adaptation. This specifically addresses the original model’s performance shortcomings in low-light conditions, small targets, and scenarios with coexisting targets possessing similar visual characteristics. Concurrently, we strictly controlled parameter and computational complexity growth to ensure the model meets real-time deployment requirements for edge devices at construction sites. The model structure is illustrated in Figure 1:

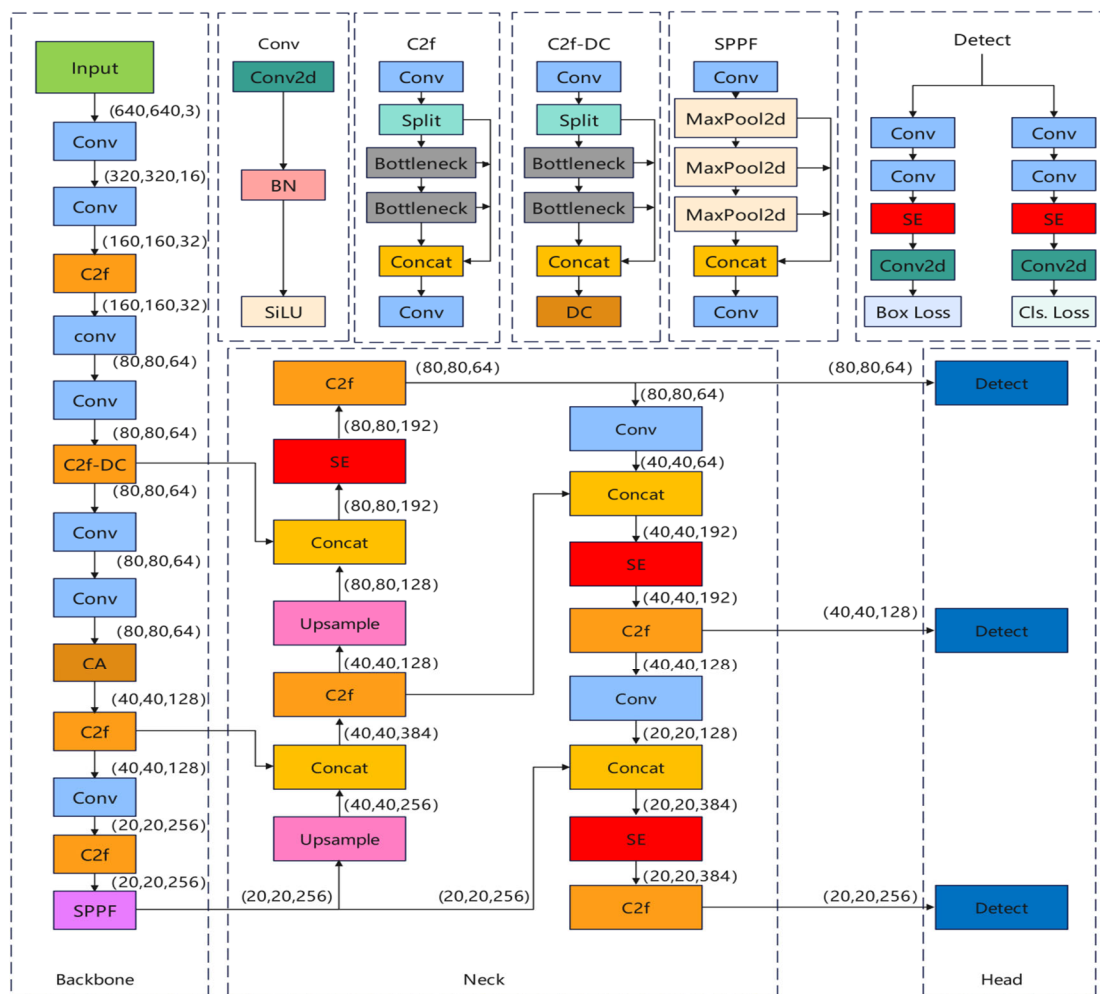


Figure 1. YOLO-Fire.

3.1.1. Backbone Enhancement: Strengthening Minor Objective Details and Scenario Adaptability

The backbone of YOLOv8n originally employed the C2f module for feature extraction. However, its shallow network structure proved inadequate for capturing fine details of small objects, struggled to address the issue of feature degradation in night-time scenarios involving open flames and yellow roof lights, and demonstrated limited adaptability to complex environments such as low illumination and glare. To address these shortcomings, four key enhancements were implemented [34,35].

(1) New shallow convolutional layers and detail enhancement. A 3×3 convolutional layer has been added after each of the two shallow C2f modules in the backbone architecture. The first new layer follows the C2f module with an output scale of 160×160 and 32 channels, maintaining the 32-channel count. Through a stride-1 convolution operation, it preserves feature map resolution while enhancing the capture of initial edges and luminance gradients in small objects. The second added layer follows the C2f module with an output scale of 80×80 and 64 channels, synchronously setting the channel count to 64 to further deepen shallow feature expression. Both new convolutional layers incorporate a Batch Normalisation (BN) layer with a Sigmoid-like Unilateral (SiLU) activation function. The BN layer stabilises feature distributions through standardisation, preventing information loss from sudden dimensional shifts. The SiLU activation maintains gradient flow in low-response regions, mitigating gradient vanishing to ensure critical details—such as small flames in low-light conditions and roof-mounted amber lights—remain discernible.

(2) Coordinate Attention Mechanism Embedding and Parameter Optimisation in Figure 2. A lightweight coordinate attention module is inserted between the C2f module (output scale 80×80 , 64 channels) and subsequent added convolutional layers. This module performs global average pooling along both width and height spatial dimensions, precisely capturing target positional information. This overcomes the limitation of traditional channel attention, which focuses solely on channel weights while neglecting spatial location. Following pooling, two independent 1×1 convolutional layers fuse channel features with positional features. The first layer reduces the number of channels to one-thirty-second of the original dimension, while the second layer restores them to 64 channels, generating a spatial attention weight map. Through this design, the model precisely focuses on regions containing small targets, enhancing local feature responses to open flames and roof-mounted yellow lights while suppressing irrelevant interference from construction site backgrounds—such as metallic reflections and tangled cables—thereby improving regional recognition of similar targets [36].

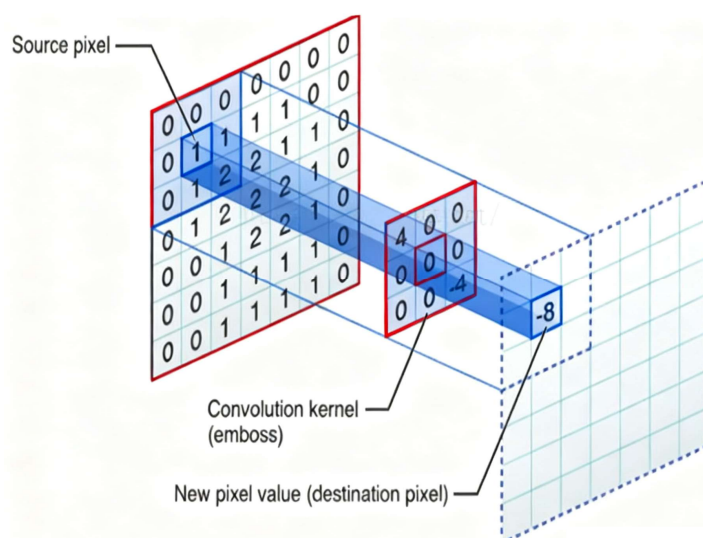


Figure 2. Ordinary Convolution.

(3) Deformable Convolution Replacement and Morphological Adaptation. A standard 3×3 convolution within the C2f module (output size 80×80 , 64 channels) is replaced with a deformable convolution as illustrated in Figure 3. This deformable convolution dynamically adjusts the receptive field shape by incorporating an offset learning mechanism within the convolution kernel. This adapts to the irregular flickering patterns of open flames and the dynamic movement characteristics of roof-mounted yellow lights during vehicle motion. To balance training stability and feature extraction capability, an incremental training strategy was adopted. Initially, the offset learning parameters were frozen. Once the backbone network converged, these parameters were gradually unfrozen. Concurrently, L1 regularisation was applied to the offsets to prevent overly discrete sampling point distributions. This adaptive adjustment enables the model to capture target features without relying on fixed receptive fields. It significantly enhances feature extraction accuracy for irregularly shaped small targets and reduces misclassification errors caused by variations in target morphology.

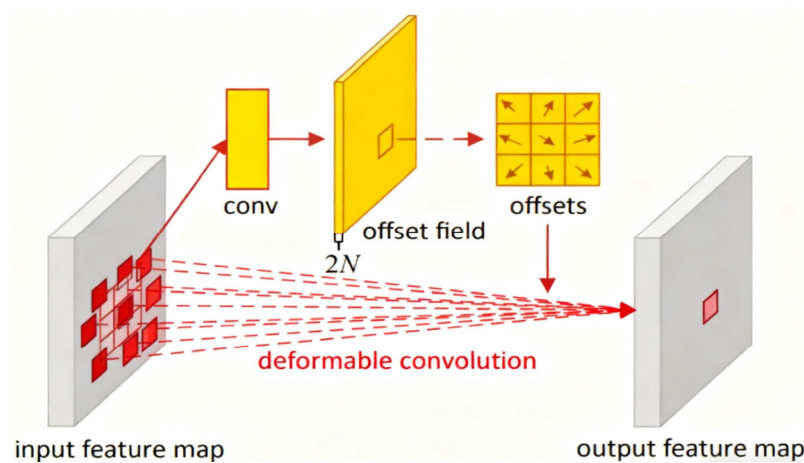


Figure 3. Variable-size Convolution.

(4) Shallow activation function optimisation: Replace the activation function in the first two C2f modules of the backbone from SiLU to LeakyReLU. Leaky ReLU preserves minute gradients on the negative half-axis, mitigating gradient vanishing issues caused by weak feature responses in low-light conditions. It delivers superior activation performance, particularly for faint brightness features such as open flames and vehicle roof lights within shadowed areas. Furthermore, Leaky ReLU's computational complexity remains comparable to SiLU, incurring no additional inference overhead and thus preserving the model's lightweight characteristics.

3.1.2. Neck Optimisation: Enhancing Multi-Scale Fusion Efficiency and Feature Discrimination

The original PAN-FPN architecture of YOLOv8n exhibited insufficient integration of shallow-level details with high-level semantic information during multi-scale feature fusion. This resulted in the dilution of small target features and a lack of reinforcement for distinguishing between similar target features. Consequently, optimisations were implemented across three dimensions: supplementing feature inputs, enhancing fusion efficiency, and strengthening feature discrimination [37].

Lightweight SPP-Net Insertion with Small Object Feature Enhancement: A lightweight SPP-Net module is inserted before the 80×80 scale feature map from the Neck layer enters the detection head. This module employs parallel pooling operations using 1×1 , 3×3 , and 5×5 pooling kernels to capture features of objects smaller than 10×10 pixels, between

10×10 and 20×20 pixels, and between 20×20 and 32×32 pixels respectively. This precisely adapts to the scale variation from minute open flames to medium-sized vehicle roof lights. Following pooling, a 1×1 convolutional layer reduces the output channel count to 32. This output is then concatenated with the original feature map. This approach preserves the richness of multi-scale features while mitigating computational overhead caused by channel expansion. The lightweight design ensures real-time performance remains unaffected. Furthermore, multi-scale pooling enhances robustness against object deformation, improving feature stability under low-light conditions [38].

Cross-layer feature fusion enhancement and detail supplementation: The output from the C2f module within the backbone network, featuring a scale of 160×160 and 32 channels, is introduced as supplementary input to the Neck layer. This feature map undergoes a 1×1 convolutional layer to reduce the channel count to 64, then undergoes concatenation and fusion with the high-level semantic features propagated top-down from the PAN-FPN. Following concatenation, a Batch Normalisation (BN) layer performs feature normalisation to mitigate fusion conflicts arising from distribution discrepancies between different feature levels. This design enables earlier, shallower-level detail features to participate in multi-scale fusion. It supplements texture and brightness features lost by small objects in low-light conditions, reinforcing the original feature expressions of open flames and vehicle roof lights. This provides additional detail support for subsequent discrimination.

Channel Attention Weighting and Feature Selection: Following the Concat operation on feature maps of scales 80×80 , 40×40 , and 20×20 at the three feature fusion nodes within the Neck, a lightweight SE channel attention module is inserted at each point. Each module acquires channel feature statistics via global average pooling, then learns channel weights through two fully connected layers. The first layer reduces channel count to 1/16th of the original dimension, while the second layer restores the full channel count. This dynamically elevates weights for fire-detection-relevant feature channels while suppressing background interference. For the 80×80 scale small-object feature map, the weight allocation precision of the attention module is further enhanced. By adjusting the learning rate of the fully connected layers, the feature channels distinguishing open flames from vehicle roof lights are significantly reinforced, laying the groundwork for distinguishing similar objects.

3.1.3. Head Adaptation: Enhancing Positioning Accuracy and Target Discrimination Capabilities

The original YOLOv8n detection head exhibited shortcomings in small object localisation accuracy and similar object classification precision. Two key adaptations were implemented to address scene requirements, alongside optimising the loss function to enhance training efficacy:

(1) **Regression branch enhancement and localisation accuracy improvement:** Following the shared convolutional layer for classification and regression within each detection branch, a dedicated coordinate attention module and 3×3 convolutional layer were added exclusively for the regression branch. The coordinate attention module leverages its spatial position capture advantage to refine small object location features, aiding bounding box coordinate prediction. The subsequent 3×3 convolution reduces channel dimensions to half the original size, followed by a 1×1 convolution to restore target dimensions. This deepens positional feature expression while controlling computational load. This refinement enables the model to predict bounding boxes with greater precision when handling small-area, positionally volatile objects such as open flames and rooftop yellow lights, reducing classification confusion caused by positioning errors. Concurrently, the regression loss employs a CIoU loss function with a dynamic weighting factor that adjusts loss weights according to target size. The regression loss weight for small objects is

increased to 1.5 times that of large objects, thereby enhancing the positioning accuracy of small targets [39].

(2) Supplementary Branch for Minuscule Object Detection and Loss Optimisation: Should the dataset contain minuscule open flames or roof-mounted yellow lights with pixel dimensions below 10×10 , a new detection branch with a 160×160 scale is added to the existing three detection branches. This branch adjusts the feature map from the C2f module within the backbone network (output scale 160×160 , 32 channels) via a 1×1 convolution to 64 channels. It is then directly connected to an independent detection head dedicated to detecting extremely small objects. Loss calculation for the new branch employs weighted processing: regression and classification losses for minute objects are each multiplied by a 1.5 weighting factor. A weighted IoU loss is introduced, assigning higher weight to object edge regions to enhance the model's learning of minute edge features. This design ensures the model focuses sufficiently on minute object features during training, preventing their loss from being masked by larger background objects, thereby improving detection recall for minute open flames and rooftop yellow lights.

(3) Classification Branch Enhancement and Distinction of Similar Targets: Following the shared convolutional layers in the classification branch, a lightweight channel attention module and contrastive learning loss term are introduced. The channel attention module further strengthens feature channels relevant to target categories while suppressing similar interfering features. The contrastive learning loss constructs positive-negative sample pairs to narrow the distance between features of similar targets and widen the distance between features of open flames and roof-mounted yellow lights, thereby improving classification accuracy for similar targets. The classification loss employs Focal Loss, which adjusts focus parameters to reduce the loss weighting of easily classified samples. This directs the model's attention towards learning features from challenging samples like open flames and roof-mounted yellow lights, thereby further lowering the false alarm rate [40].

3.2. Multimodal Feature Fusion and Attention Mechanism

To further explore the fundamental differences between open flames and rooftop yellow lights, and to resolve classification confusion arising from visual similarity, a visual-infrared multimodal feature fusion architecture with scene-adaptive attention mechanisms has been designed. This addresses the perception requirements for night-time scenarios at nuclear power plants under construction, enabling precise differentiation between similar targets.

3.2.1. Multimodal Data Input and Feature Extraction

Utilising dual-modality visible light and infrared data as model inputs, each data type corresponds to core features of the target across different dimensions: visible light imagery preserves visual details such as colour and texture, whilst infrared imagery reflects variations in thermal radiation intensity. Open flames, as high-temperature heat sources, exhibit intense thermal radiation responses in infrared imagery, whereas the yellow roof lights function as cold light sources with significantly lower thermal radiation intensity than open flames.

Following synchronous acquisition and registration, the dual-modality data are fed into independent feature extraction branches. The visible light branch employs a modified YOLOv8n backbone to generate multi-scale visual features; The infrared branch employs a lightweight convolutional network comprising three 3×3 convolutional layers and two C2f modules. Input infrared images undergo downsampling and feature enhancement through the convolutional layers, producing infrared feature maps (160×160 , 80×80 ,

$40 \times 40, 20 \times 20$) aligned in scale with the visible light branch. This ensures dimensional consistency between the dual-modal features.

3.2.2. Cross-Modal Feature Fusion Module

At the Neck stage, a bimodal feature fusion module is constructed employing a three-step process: feature alignment, weighted fusion, and feature refinement.

(1) Feature Alignment: A 1×1 convolutional layer adjusts the number of channels in the infrared feature map to match that of the corresponding scale visible feature map, eliminating channel dimension discrepancies. Interpolation techniques correct spatial offsets incurred during bimodal data acquisition, ensuring precise target position alignment.

(2) Weighted Fusion: Modal attention weight factors are introduced. Global average pooling extracts statistical information on the response intensity of bimodal features. A fully connected layer learns modal weights, enabling dynamic weighted summation of visible and infrared features. In low-light scenarios, infrared feature weights are automatically elevated to accentuate thermal radiation differences. Under relatively favourable illumination conditions, the contribution of visible feature details is enhanced.

(3) Feature Refinement: The fused features undergo processing through a 3×3 convolutional layer and a Batch Normalisation (BN) layer. This suppresses noise generated during modal fusion while enhancing the expression of effective features. The final output comprises a multi-scale fused feature map, providing more comprehensive target information for subsequent detection tasks.

3.2.3. Scene-Adaptive Attention Mechanism

Prior to the fusion features entering the detection head, a scene-adaptive attention module as depicted in Figure 4 is inserted to perform feature selection tailored to the complex environment of night-time construction at nuclear power plants under construction. This scene-adaptive attention module primarily comprises spatial attention, channel attention, and scene-aware adjustment.

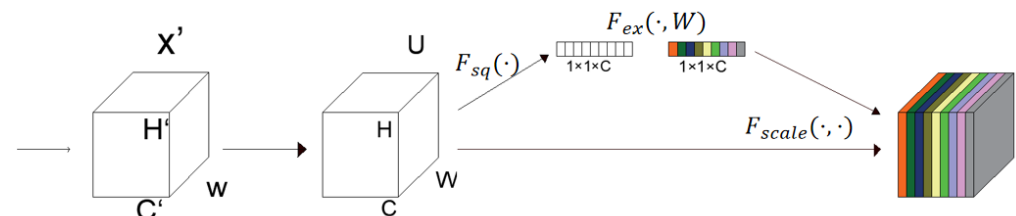


Figure 4. Scene-adaptive Attention Module.

Spatial Attention Branch: Generates a spatial weight map via two-dimensional Gaussian filtering, focusing on small, high-brightness target areas within the construction zone while suppressing feature responses from large, meaningless background regions such as tower crane structures and temporary buildings.

Channel Attention Branch: Based on statistical information from the dataset, pre-set feature channel priorities associated with open flames and vehicle roof lights. Through channel attention weight adjustments, it enhances discriminative channels such as thermal radiation features and dynamic shape characteristics while reducing interference from colour similarity channels.

Scene-Aware Modulation: The module incorporates an integrated scene brightness detection unit. It dynamically adjusts attention intensity based on the input image's average brightness—enhancing attention weights in low-illumination environments and moderately reducing them in high-illumination conditions. This ensures stable performance across varying lighting scenarios.

3.3. Loss Functions and Training Strategies

To accommodate small object detection, similar object discrimination, and complex night-time scenarios, the loss function has undergone multi-dimensional optimisation. Tailored training strategies have been designed to enhance the model's detection accuracy and robustness.

3.3.1. Multi-Component Loss Function Design

Construct a composite loss function L_{total} comprising classification loss, regression loss, and similarity objective comparison loss, with each component synergistically optimising model performance.

$$L_{total} = L_{reg} + L_{cls} + L_{contra} \quad (1)$$

Among these, L_{cls} denotes the modified Focal Loss classification loss, L_{reg} represents the weighted CIoU regression loss, and L_{contra} signifies the similarity-based target contrast loss. The contrast loss weight λ (set to 0.3) balances the optimisation contributions across loss components.

The classification loss function L_{reg} employs an enhanced Focal Loss, adjusting the focus parameter γ to 2.5 to further reduce the loss weighting for easily classifiable background samples. A category balancing factor is introduced, dynamically adjusting category weights based on the sample ratio of open flames versus rooftop yellow lights. This mitigates potential sample imbalance within the dataset, enabling the model to focus more intently on learning to classify these two similar object categories.

$$L_{cls} = -\frac{1}{N_{pos}} \sum_{i=1}^N \alpha_{c_i} \cdot (1 - p_{i,c_i})^\gamma \cdot \log(p_{i,c_i}) \quad (2)$$

N denotes the total number of annotated samples in the image, N_{pos} denotes the number of valid target samples (open flame + roof-mounted amber light); c_i denotes the true class of the i -th sample ($c_i = 0$ indicates roof-mounted amber light, $c_i = 1$ indicates open flame); p_{i,c_i} represents the model's predicted probability that the i -th sample belongs to class c_i ; α_{c_i} is the class balancing factor, dynamically adjusted according to sample proportions (if N_1 denotes open flame samples and N_0 denotes roof yellow light samples, then $\alpha_1 = N_0 / (N_0 + N_1)$, $\alpha_0 = N_1 / (N_0 + N_1)$); γ denotes the focus parameter, amplifying loss weighting for challenging samples.

The regression loss function L_{reg} employs CIoU Loss combined with a small object weighting factor. For small objects with pixel areas less than 32×32 , the regression loss weight is increased by a factor of 1.5 to enhance the localisation accuracy of small objects. A target edge distance penalty term is introduced into CIoU Loss to improve the fitting of bounding boxes to the edges of small objects, thereby reducing classification confusion caused by localisation errors.

$$L_{reg} = \sum_{i=1}^{N_{pos}} \omega_i \cdot (1 - CIoU_i + \beta \cdot d_{edge,i}) \quad (3)$$

$CIoU_i$ denotes the CIoU value between the predicted bounding box and the ground truth bounding box for the i -th sample, calculated according to the standard CIoU formula; ω_i represents the small object weighting factor: if the area S_i of the ground truth bounding box for the i -th sample is $< 32 \times 32$, then $\omega_i = 1.5$; otherwise $\omega_i = 1.0$; β denotes the edge distance penalty coefficient, set to 0.2 to enhance the fitting of the bounding box to the edges of small objects; $d_{edge,i}$ represents the mean edge distance between the predicted and ground-truth bounding boxes, calculated by averaging the corresponding distances along all four sides of the bounding box, measured in pixels.

Similar Object Contrastive Loss Function A new contrastive loss term is introduced to optimise the feature embedding space by constructing positive and negative sample pairs. Positive sample pairs comprise features from similar objects, while negative pairs combine features from open flames and vehicle roof lights. Contrastive Loss reduces distances between similar features and increases distances between dissimilar features, enhancing the discriminative power between object classes. The contrastive loss weight is set to 0.3 to prevent undue influence on the primary loss function's optimisation direction.

$$L_{contra} = \frac{1}{2M} \sum_{j=1}^M \left[(1 - y_j) \cdot \max(0, d_j - m_{neg})^2 + y_j \cdot \max(0, m_{pos} - d_j)^2 \right] \quad (4)$$

M denotes the total number of sample pairs constructed within a batch (with M/2 positive sample pairs and M/2 negative sample pairs randomly selected per batch); y_j denotes the label of the sample pair ($y_j = 1$ indicates a positive sample pair composed of features from the same target; $y_j = 0$ indicates a negative sample pair composed of features from open flame and roof-mounted yellow lights); d_j represents the Euclidean distance between the features of the j-th sample pair, $d_j = \|f_a - f_b\|$, where f_a and f_b denote the feature vectors of the two objects within the sample pair); m_{pos} represents the distance threshold for same-class sample pairs (set to 0.5), while m_{neg} denotes the distance threshold for different-class sample pairs (set to 2.0), thereby controlling the distribution range within the feature embedding space.

3.3.2. Scene-Adaptive Training Strategy

Tailored data augmentation schemes were designed to address the environmental characteristics of night-time construction scenarios. Glare interference was simulated by randomly adding light spots; uneven illumination was modelled using adaptive histogram equalisation to adjust local brightness levels; construction dust was replicated through Gaussian noise and fog effects; and target motion was simulated by randomly deforming open-flame samples and applying random translation and rotation to samples featuring yellow roof lights. Consistency in target labels is maintained throughout data augmentation to ensure enhanced samples retain real-world scene characteristics.

The first phase constitutes pre-training, employing a joint training approach using publicly available fire detection datasets and generic nuclear power plant construction datasets. This enables the model to learn universal features distinguishing fires from ordinary objects while initialising network parameters. The second phase involves fine-tuning using a proprietary dataset, focusing on optimising the model's ability to differentiate similar objects and detect small targets. The learning rate progressively decreases across both stages: an initial rate of 0.01 in stage one, adjusted to 0.001 in stage two. A cosine annealing learning rate scheduling strategy is employed to enhance training stability.

Domain adaptation loss is introduced to reduce feature distribution discrepancies between the public dataset (source domain) and the custom dataset (target domain), mitigating overfitting caused by sparse annotated data in specialised night-time scenarios. Domain adaptation training is achieved through gradient reversal layers, enabling the model to adapt to the domain-specific features of nuclear power plant night-time construction while learning the target classification task.

3.4. Dataset Construction

The existing public datasets lack samples of open flames and rooftop yellow lights during night-time construction scenarios at nuclear power plants under construction, rendering them inadequate for model training requirements. Consequently, a bespoke dataset has been constructed, with the specific workflow and specifications outlined below:

3.4.1. Data Collection

Data was collected at the construction site of the nuclear power plant under development in Zhangzhou. The dataset comprises two core target categories: open flames and yellow lights on vehicle roofs. Open flame samples were generated by simulating hot work operations during construction, encompassing diverse forms such as welding sparks and flames from small combustible materials. Roof-mounted amber light samples were collected from construction vehicles traversing the site, including warning lights atop various vehicle types such as cranes, concrete mixers, and transport lorries, covering static, moving, and turning states.

A total of 12,000 raw image and video data points were collected. Video data yielded 8000 images after frame extraction, while 4000 images were captured directly, resulting in 12,000 valid images. The ratio of the two target categories was controlled at 1:1.2, comprising 5450 open flame samples and 6550 roof-mounted yellow light samples, ensuring balanced learning for both types. The dataset incorporates diverse background interferences, such as metallic reflections, temporary lighting, tangled cables, and construction personnel, thereby enhancing the model's robustness against disturbances.

3.4.2. Data Annotation and Preprocessing

Using the Labelling annotation tool, annotations were performed according to the PASCAL VOC format. Bounding boxes were annotated for open flames and yellow roof lights in each image, ensuring precise enclosure of the target area with annotation errors not exceeding 2 pixels. Simultaneously, target category labels were annotated: open flames were labelled as 'fire', and yellow roof lights as 'yellow_light'.

Perform preprocessing operations on the collected images. First, crop the images to remove insignificant peripheral areas, uniformly resizing them to 640×640 pixels to meet model input requirements. Subsequently, normalise pixel values: visible light images are normalised to the [0,1] range, while infrared images are normalised to their maximum value. Finally, low-quality samples exhibiting severe blurring or occlusion were discarded, resulting in 10,000 high-quality annotated images retained for model training and validation.

3.4.3. Dataset Partitioning

The dataset was partitioned into training, validation, and test sets in an 8:1:1 ratio. The training set comprised 8000 images for model parameter learning. The validation set contained 1000 images for hyperparameter tuning and model performance monitoring during training. The test set comprises 1000 images for evaluating the model's final detection performance. During partitioning, consistency in scene distribution and target scale distribution across the three datasets was ensured to prevent evaluation distortion caused by data distribution bias.

4. Experiments and Results Analysis

4.1. Experimental Setup

The training iterations were set to 300, the batch size to 32, and the learning rate (lr_0) to 0.01. Stochastic gradient descent (SGD) was employed as the optimiser with momentum set to 0.937. The hardware and software configurations are detailed in Table 1:

Table 1. Experimental setup.

| Experimental Environment Configuration | Version Parameters |
|--|--|
| Operating System | Windows |
| Deep learning framework | Python torch 1.10.0 |
| CPU | Intel Core i7-13620H (2.4 GHz/L3 24 M) |
| Python | 3.10 |
| GPU | NVIDIA GeForce RTX 4060 |

4.2. Evaluation Criteria

The False Alarm Rate (FAR) serves as the core evaluation metric in this study, directly reflecting the model’s misclassification of yellow roof lights. It is calculated as the ratio of ‘the number of yellow roof light samples misclassified as open flames’ to ‘the total number of yellow roof light samples’. This metric focuses on the core research issue; a lower value indicates stronger model capability in distinguishing similar objects, thereby better meeting the practical requirements of nuclear power plant construction sites.

Precision: Measures the proportion of samples correctly classified as actual fires among those predicted as fires by the model. Calculated as the ratio of ‘true positive samples’ to ‘true positive samples + false positive samples’, reflecting the reliability of the model’s classification results.

Recall: Measures the proportion of all actual open flames successfully detected by the model. Calculated as the ratio of ‘true positive samples’ to ‘true positive samples + false negative samples’, it reflects the model’s ability to control missed detections of fire targets.

Mean Average Precision (mAP): The average precision value calculated based on the Precision-Recall curve, using an mAP@0.5 (IOU) threshold of 0.5 as the evaluation criterion. This metric reflects the model’s overall performance in both target localisation and classification.

4.3. Experimental Results

The P-R curves of the three models during the testing process are shown in Figure 5.

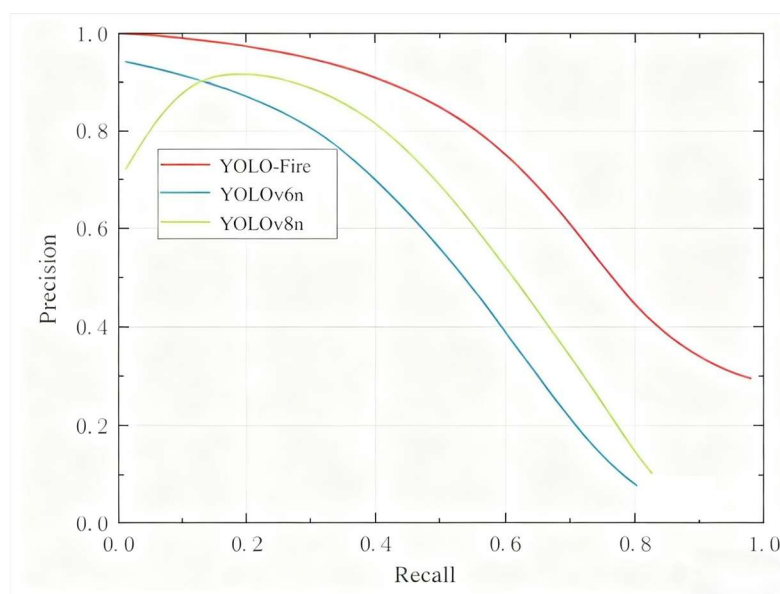


Figure 5. P-R.

The numerical comparison of YOLO-Fire and the reference model across various metrics is shown in Table 2:

Table 2. Experimental data.

| Model | R/% | mAP _{0.5} /% | FAR/% |
|-----------|------|-----------------------|-------|
| YOLOv6n | 50.6 | 50.1 | 42.3 |
| YOLOv8n | 63.1 | 63.9 | 34.9 |
| YOLO-Fire | 72.5 | 75.2 | 20.6 |

To more clearly demonstrate the improved detection performance of the YOLO-Fire algorithm compared to the original YOLOv8n model, images were extracted from the dataset for detection and visualisation. Detection was performed using YOLOv6n, YOLOv8n, and YOLO-Fire respectively, with the results shown in Figure 6:

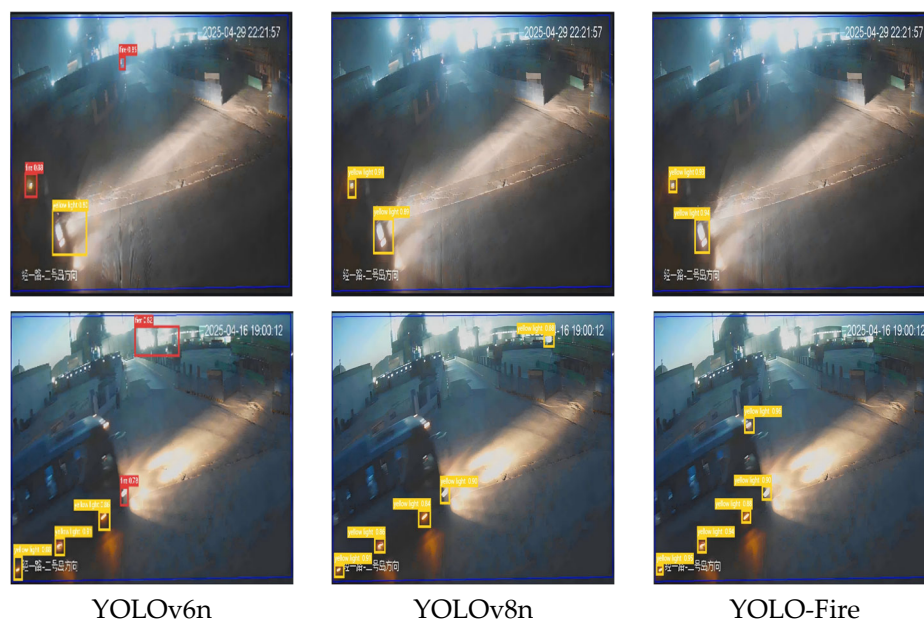


Figure 6. Visual comparison.

To clarify the contribution of each structural improvement and loss function optimisation to the final performance, ablation experiments were conducted under the same training and evaluation settings as the main experiment. The progressive configuration of the experiment is shown in Table 3, where the performance metrics (Recall, mAP@0.5, FAR) are consistent with the main experiment to ensure fair comparison.

Table 3. Ablation experiment results.

| Experimental Configuration | R/% | Map 0.5/% | FAR/% |
|----------------------------|------|-----------|-------|
| Baseline | 63.1 | 63.9 | 34.9 |
| +Backbone | 67.8 | 68.5 | 29.7 |
| +Neck | 70.3 | 72.1 | 25.3 |
| +Head | 71.9 | 74.3 | 22.1 |
| +Scene-Adaptive Attention | 72.5 | 75.2 | 20.6 |

5. Discussion

This paper addresses the issue of false detections caused by visual similarity, small target scale, and complex environments when distinguishing between open flames and rooftop yellow lights during night-time construction at nuclear power plants under construction. It focuses on scenario adaptation and precise differentiation, as detailed below:

(1) Customised enhancements to the YOLOv8n architecture address the original model's shortcomings in shallow feature extraction and weak discrimination of similar targets. A shallow convolutional layer is added to the backbone with embedded coordinate attention mechanisms, enhancing edge and texture detail capture for small targets under low-light conditions; Deformable convolutions adapt to irregular open-flame shapes and dynamic yellow light variations. Neck optimisation enhances multi-scale fusion efficiency to prevent feature dilution, while Head enhancements improve regression branch localisation accuracy. Strict parameter growth control throughout ensures precise alignment with edge device real-time deployment requirements.

(2) Optimised multi-dimensional loss functions and training strategies: Designed a composite loss function combining classification loss, regression loss, and similar object comparison loss. An improved Focal Loss focuses on challenging samples, weighted CIoU Loss strengthens small target localisation, while comparison loss widens feature distances between dissimilar objects. This synergistically optimises classification accuracy and localisation precision. Complemented by dedicated night-time scene data augmentation and two-stage progressive training, this approach effectively enhances model robustness in complex operational conditions while mitigating overfitting caused by scarce annotated data in specialised scenarios.

(3) Supported by a dedicated dataset for nuclear power plants under construction, we established a bimodal dataset covering critical areas such as the nuclear island construction zone and conventional island assembly zone. This dataset encompasses diverse weather conditions (e.g., moonless nights, light rain), target states (static/dynamic), and construction backgrounds (metallic reflections, cable interference), ensuring sample authenticity and representativeness. It fills the gap in annotated data for this specialised scenario, providing reliable support for model training and validation.

Three key innovations form a comprehensive technical chain encompassing model architecture, training optimisation, and data support. This approach specifically addresses core false alarm issues related to open flames and rooftop yellow lights while balancing engineering requirements for lightweight and real-time performance, achieving an equilibrium between technological innovation and field practicality.

6. Conclusions

This paper addresses the issue of false fire alarms caused by visual similarity, small target dimensions, and complex environments when distinguishing between open flames and rooftop yellow lights during night-time construction at nuclear power plants under construction. It investigates intelligent recognition methods and proposes YOLO-Fire, based on an enhanced YOLOv8n architecture. The primary research conclusions are as follows:

The research centres on enhancing small object feature extraction, identifying intrinsic differences between similar targets, and improving robustness in complex environments, achieving three key objectives: Firstly, a customised structural enhancement of YOLOv8n was implemented. This involved strengthening the shallow layers of the Backbone, optimising Neck fusion, and adapting Head branches. These modifications enhance the capture of small object details and positioning accuracy while controlling the number of parameters. Second, a composite loss function incorporating classification, regression, and comparison metrics was designed alongside a dedicated night-scene training strategy, enhancing the model's adaptability to challenging samples and complex environments. Third, a bespoke dual-modal night-time dataset for nuclear power plants was constructed, filling a data gap in specialised scenarios and providing robust training support.

Experimental results demonstrate that the improved model reduces false alarm rates by 14.3% compared to the original YOLOv8n on the proprietary dataset, specifically for open-flame detections (mAP@0.5 11.3%), while maintaining inference speeds exceeding 85 FPS to meet real-time deployment requirements for edge devices. Compared to benchmark models such as YOLOv6n and YOLOv8n, the improved model achieves optimal performance in balancing false alarm control, small object detection accuracy, and real-time capabilities.

The proposed technical solution effectively resolves false alarm challenges in fire detection during night-time construction at nuclear power plants under construction. It provides precise and efficient intelligent technical support for fire risk prevention during the construction phase, holding significant engineering application value. The research findings are not only applicable to nuclear power plant scenarios but also offer a reference technical approach for detecting similar small targets and visually confusing objects in industrial settings. Future research may further explore multi-sensor fusion techniques and model lightweight optimisation to enhance adaptability in extreme environments and flexibility for edge deployment.

Author Contributions: Conceptualization, Z.L.; Methodology, Z.L.; Software, Z.L. and G.L.; Formal analysis, G.L.; Investigation, K.Y.; Resources, K.Y.; Data curation, S.D.; Writing—original draft, K.Y. and S.D.; Visualisation, G.L.; Project administration, Z.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: Authors Zhendong Li and Guangwei Liu were employed by the company Zhangzhou Project Team, China Nuclear Power Engineering Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Xiao, X.; Liang, J.; Tong, J.; Wang, H. Emergency Decision Support Techniques for Nuclear Power Plants: Current State, Challenges, and Future Trends. *Energies* **2024**, *17*, 2439. [[CrossRef](#)]
2. Qi, B.; Liang, J.; Tong, J. Fault Diagnosis Techniques for Nuclear Power Plants: A Review from the Artificial Intelligence Perspective. *Energies* **2023**, *16*, 1850. [[CrossRef](#)]
3. Bushnaq, O.M.; Chaaban, A.; Al-Naffouri, T.Y. The role of UAV-IoT networks in future wildfire detection. *IEEE Internet Things J.* **2021**, *8*, 16984–16999. [[CrossRef](#)]
4. Muksimova, S.; Umirzakova, S.; Mardieva, S.; Abdullaev, M.; Cho, Y.I. Revolutionizing Wildfire Detection Through UAV-Driven Fire Monitoring with a Transformer-Based Approach. *Fire* **2024**, *7*, 443. [[CrossRef](#)]
5. Dilshad, N.; Khan, T.; Song, J. Efficient Deep Learning Framework for Fire Detection in Complex Surveillance Environment. *Comput. Syst. Sci. Eng.* **2023**, *46*, 749–764. [[CrossRef](#)]
6. Yar, H.; Khan, Z.A.; Hussain, T.; Baik, S.W. A modified vision transformer architecture with scratch learning capabilities foreffective fire detection. *Expert Syst. Appl.* **2024**, *252*, 123935. [[CrossRef](#)]
7. Csápaiová, N. The impact of vehicle fires on road safety. *Transp. Res. Procedia* **2021**, *55*, 1704–1711. [[CrossRef](#)]
8. Gaur, A.; Singh, A.; Kumar, A.; Kumar, A.; Kapoor, K. Video flame and smoke based fire detection algorithms: A literature review. *Fire Technol.* **2020**, *56*, 1943–1980. [[CrossRef](#)]
9. Celik, T.; Ozkaramanli, H.; Demirel, H. Fire Pixel Classification Using Fuzzy Logic and Statistical Color Model. In Proceedings of the 2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP'07, Honolulu, HI, USA, 15–20 April 2007.
10. Shamta, I.; Demir, B.E. Development of a deep learning-based surveillance system for forest fire detection and monitoring using UAV. *PLoS ONE* **2024**, *19*, e0299058. [[CrossRef](#)]
11. Xue, Z.; Lin, H.; Wang, F. A Small Target Forest Fire Detection Model Based on YOLOv5 Improvement. *Forests* **2022**, *13*, 1332. [[CrossRef](#)]

12. Khan, T.; Aslan, H.I. Performance evaluation of enhanced ConvNeXtTiny-based fire detection system in real-world scenarios. In Proceedings of the International Conference on Learning Representations (ICLR), Kigali, Rwanda, 2 March 2023.
13. Ergasheva, A.; Akhmedov, F.; Abdusalomov, A.; Kim, W. Advancing Maritime Safety: Early Detection of Ship Fires through Computer Vision, Deep Learning Approaches, and Histogram Equalization Techniques. *Fire* **2024**, *7*, 84. [[CrossRef](#)]
14. Li, S.; Yan, Q.; Liu, P. An efficient fire detection method based on multiscale feature extraction, implicit deep supervision and channel attention mechanism. *IEEE Trans. Image Process.* **2020**, *29*, 8467–8475. [[CrossRef](#)]
15. Abdusalomov, A.; Umirzakova, S.; Safarov, F.; Mirzakhilov, S.; Egamberdiev, N.; Cho, Y.I. A Multi-Scale Approach to Early Fire Detection in Smart Homes. *Electronics* **2024**, *13*, 4354. [[CrossRef](#)]
16. Yar, H.; Ullah, W.; Khan, Z.A.; Baik, S.W. An effective attention-based CNN model for fire detection in adverse weather conditions. *ISPRS J. Photogramm. Remote Sens.* **2023**, *206*, 335–346. [[CrossRef](#)]
17. Shorten, C.; Khoshgoftaar, T.M. A survey on Image Data Augmentation for Deep Learning. *J. Big Data* **2019**, *6*, 60. [[CrossRef](#)]
18. Hanson, S.; Andula, N.; Goulden, M.I.; Randerson, J.T. Human-ignited fires' results in more extreme wildfire behavior and ecosystem impacts. *Nat. Commun.* **2022**, *13*, 2717. [[CrossRef](#)] [[PubMed](#)]
19. Hou, Q.; Zhou, D.; Feng, J. IEEE/CVF Conference on Efficient Mobile Network Design. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021; pp. 13708–13717.
20. Hsu, W.Y.; Lin, T.Y. Rota-and-Scale-Aware YOLO for Pedestrian Detection. *IEEE Trans. Image Process.* **2021**, *30*, 934–947. [[CrossRef](#)]
21. Shafiq, M.; Gu, Z. Deep Residual Learning for Image Recognition: A Survey. *Appl. Sci.* **2022**, *12*, 8972. [[CrossRef](#)]
22. Han, Z.; Yue, Z.; Liu, L. 3L-YOLO: A Lightweight Low-Light Object Detection Algorithm. *Appl. Sci.* **2025**, *15*, 90. [[CrossRef](#)]
23. Yunusov, N.; Islam, B.M.S.; Abdusalomov, A.; Kim, W. Robust Forest Fire Detection Method for Surveillance Systems Based on You Only Look Once Version 8 and Transfer Learning Approaches. *Processes* **2024**, *12*, 1039. [[CrossRef](#)]
24. Wang, Z.; Wang, Z.; Zhang, H.; Guo, X. A Novel Fire Detection Approach Based on Cnn-Svm Using Tensorflow. In *Proceedings of the Intelligent Computing Methodologies: 13th International Conference, ICIC 2017, Liverpool, UK, 7–10 August 2017*; Springer: Berlin/Heidelberg, Germany, 2017; Part III 13.
25. Ayumi, V.; Noprison, H.; Ani, N. Forest Fire Detection Using Transfer Learning Model with Contrast Enhancement and Data Augmentation. *J. Nas. Pendidik. Tek. Inform. JANAPATI* **2024**, *13*. [[CrossRef](#)]
26. Seydi, S.T.; Saeidi, V.; Kalantar, B.; Ueda, N.; Halin, A.A. Fire-Net: A Deep Learning Framework for Active Forest Fire Detection. *J. Sens.* **2022**, *2022*, 8044390. [[CrossRef](#)]
27. Chetoui, M.; Akhloufi, M.A. Fire and Smoke Detection Using Fine-Tuned YOLOv8 and YOLOv7 Deep Models. *Fire* **2024**, *7*, 135. [[CrossRef](#)]
28. Li, J.W.; Chen, X.Y.; Chen, Z.Q.; Li, H.R.; Li, Y. Three-dimensional dynamic simulation system for forest surface fire spread and prediction. *Int. J. Wildland Fire* **2018**, *27*, 612–622.
29. Ma, Z.; Zhang, X.; Zheng, X.T.; Sun, M. Shufflenet v2: Practical guidelines for efficient CNN architecture design. In *Computer Vision—ECCV 2018; Lecture Notes in Computer Science*; Springer: Cham, Switzerland, 2018; pp. 116–131.
30. Wang, Z.; Chen, J.; Hoi, S.C.H. Deep Learning for Image Super-Resolution: A Survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *43*, 3365–3387. [[CrossRef](#)]
31. Gonçalves, L.A.O.; Ghali, R.; Akhloufi, M.A. YOLO-Based Models for Smoke and Wildfire Detection in Ground and Aerial Images. *Fire* **2024**, *7*, 140. [[CrossRef](#)]
32. Wahyono; Harjoko, A.; Dharmawan, A.; Adhinata, F.D.; Kosala, G.; Jo, K.H. Real-time forest fire detection framework based on artificial intelligence using color probability model and motion feature analysis. *Fire* **2022**, *5*, 23. [[CrossRef](#)]
33. Nguyen, D.T.; Nguyen, T.N.; Kim, H.; Lee, H.J. A High-Throughput and Power-Efficient FPGA Implementation of YOLO CNN for Object Detection. *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.* **2019**, *27*, 1861–1873. [[CrossRef](#)]
34. Hussain, M. YOLO-v1 to YOLO-v8, the Rise of YOLO and Its Complementary Nature toward Digital Manufacturing and Industrial Defect Detection. *Machines* **2023**, *11*, 677. [[CrossRef](#)]
35. Luan, T.; Zhou, S.; Zhang, G.; Song, Z.; Wu, J.; Pan, W. Enhanced Lightweight YOLOX for Small Object Wildfire Detection in UAV Imagery. *Sensors* **2024**, *24*, 2710. [[CrossRef](#)] [[PubMed](#)]
36. Romero, A.; Gatta, C.; Camps-Valls, G. Unsupervised Deep Feature Extraction for Remote Sensing Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 1349–1362. [[CrossRef](#)]
37. Sahoo, A.; Anari, M.S.; Varshney, A.; Agarwal, M.N.; Kanwal, N. FireNet-V2: Improved Lightweight Fire Detection Model for Real-Time IoT Applications. *Proc. Fire Sci.* **2023**, *218*, 2233–2242.
38. Wu, X.Q.; Yan, T.; Chen, Z.B.; Li, Z.; Chen, Y.; Yang, R.; Li, K.H. CacheTrack-YOLO: Real-Time Detection and Tracking for Fire Sources and Surrounding Tissues in Ultrasound Videos. *IEEE J. Biomed. Health Inform.* **2021**, *25*, 3823–3832. [[CrossRef](#)] [[PubMed](#)]

39. Alom, M.Z.; Taha, T.M.; Yakopcic, C.; Westberg, S.; Sidike, P.; Nasrin, M.S.; van Esesn, B.C.; Awwal, A.A.S.; Asari, V.K. The history began from alexnet: A comprehensive survey on deep learning approaches. *arXiv* **2018**, arXiv:1803.01164. [[CrossRef](#)]
40. Chen, T.-H.; Wu, P.-H.; Chiou, Y.-C. An Early Fire-Detection Method Based on Image Processing. In Proceedings of the 2004 International Conference on Image Processing, 2004, ICIP'04, Singapore, 24–27 October 2004.

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