Enhanced Faster R-CNN with Temporal and Context Modules for Power Plant Safety Monitoring

Qinglin Deng*, Xuexiong Huang, Tiandong Zhong
Huadian Sihanoukville Power Generation Co., Ltd., Preah Sihanouk 18000, Cambodia
E-mail: qinglin_deng@outlook.com, Xuexiong_Huang@outlook.com, Tiandong_Zhong@outlook.com
*Corresponding author

Keywords: power plant safety monitoring, target detection algorithm, Faster R-CNN, algorithm optimization, deep learning

Received:

In order to improve the intelligence and automation level of power plant safety monitoring systems, this study proposes an improved Faster R-CNN algorithm by integrating multi-scale feature extraction, context-aware RPN, temporal information fusion, and RoI Align optimization. The model is trained and tested on a power plant safety monitoring dataset covering diverse and complex scenarios. Comparative experiments against baseline methods including original Faster R-CNN, YOLOv5, and SSD demonstrate that the improved algorithm achieves a mean Average Precision (mAP) of 0.85 and a Recall of 0.82, outperforming the baselines by margins of up to 13% in mAP and 12% in Recall. The enhanced algorithm also shows superior adaptability to small targets, occlusions, low light, and complex backgrounds. These results indicate that the proposed method significantly enhances the performance of target detection in challenging power plant environments.

Povzetek: Za varnostno nadzorovanje v elektrarnah je izboljšan Faster R-CNN z večločljivostnim zajemom značilk, kontekstno zavednim RPN, časovno fuzijo (TCN) in RoI Align; na namenskem naboru prehiti izvirni Faster R-CNN, YOLOv5 in SSD (do +13% mAP, +12% Recall), posebej pri malih tarčah, zakritjih, slabi osvetlitvi in kompleksnem ozadju.

1 Introduction

As one of the indispensable infrastructures in modern society, the safety and stability of the power system directly affect people's lives and economic development. However, in the complex operation of power plants, safety hazards often occur, which may lead to catastrophic consequences in serious cases. In order to prevent such risks, the construction of power plant safety monitoring platforms is particularly important. In traditional monitoring systems, the detection and prediction of safety hazards mainly rely on manual inspections and conventional equipment monitoring. This method is not only inefficient but also difficult to cope with the increasingly complex safety situation [1]. With the advancement of technology, especially the widespread application of computer vision and artificial intelligence technology, intelligent safety monitoring systems based on target detection algorithms have gradually become a cutting-edge trend in power plant safety management. Through real-time monitoring and automatic identification of potential safety hazards, power plants can detect problems in advance and take effective measures to reduce the probability of accidents, thereby improving the safety and stability of the power system [2].

Ensuring real-time and accurate detection of safety-critical targets in power plant environments is vital for preventing industrial accidents. Traditional object detection algorithms often struggle with challenges such as low-light conditions, dynamic occlusions, and diverse background complexity. This study is motivated by the urgent need for a robust, adaptable detection framework that can operate effectively in these harsh, safety-critical scenarios.

In the production environment of power plants, there are many potential safety hazards. For example, equipment failure, electrical fire, and personnel violation can lead to serious accidents. According to statistics, the losses caused by safety accidents in power plants exceed hundreds of millions of yuan each year. According to the 2019 China Power Industry Safety Report, about 12% of accidents are caused by equipment failure, and another 15% are directly related to improper operation. The frequent occurrence of such accidents not only increases the operating costs of power plants, but is also likely to have serious impacts on the surrounding environment and the public. Therefore, it is particularly urgent to build an intelligent, safe and efficient monitoring platform.

At present, many power plants have begun to introduce advanced target detection algorithms to enhance the intelligence and automation level of safety monitoring systems. These algorithms use deep learning and computer vision technology to achieve real-time monitoring of different equipment inside the power plant and automatic identification of safety hazards. These technologies can automatically identify potential abnormal conditions in image or video data, such as equipment aging, cable short circuits, fire hazards, etc., and issue alarms in a timely manner, greatly improving the response speed and accuracy of safety management. For example, the application of convolutional neural networks (CNNs) in image recognition has achieved remarkable results in power plant safety monitoring, helping the monitoring system to quickly identify abnormal operating conditions of equipment [3,4].

However, although target detection algorithms have been applied in power plant safety monitoring, existing technologies still face many challenges. First, the monitoring environment of power plants is complex, and the monitoring scenarios include different types of equipment, personnel, and other dynamic factors, which poses great challenges to the application of target detection algorithms. When dealing with real-time monitoring in a changing environment, the accuracy and stability of existing algorithms often cannot meet actual needs. Secondly, most current target detection algorithms focus on improving detection accuracy and speed, but in the specific environment of power plants, the robustness and scalability of the algorithms are still insufficient. In particular, for some special scenarios, such as low light, smoke occlusion, or complex equipment background, the performance of traditional algorithms is not ideal. Therefore, how to improve the adaptability and accuracy of target detection algorithms in these special environments is an urgent problem to be solved in power plant safety monitoring systems.

Many scholars and engineers have devoted a lot of work to the research of target detection algorithms. In recent years, target detection algorithms based on deep learning, such as YOLO (You Only Look Once) and Faster R-CNN (Region Convolutional Neural Network), have made significant progress, especially in improving the speed and accuracy of image processing [5]. However, the application of these algorithms in power plant safety monitoring still faces many challenges. In order to better adapt to the power plant environment, the existing target detection algorithms need to be optimized, especially in dealing with multi-target recognition, complex backgrounds, objects of different scales, etc. In addition, the computational efficiency and real-time performance of the algorithm are also issues that need to be focused on, because the safety monitoring of power plants requires the system to have efficient real-time response capabilities.

In order to overcome these problems, this paper aims to propose a new power plant safety monitoring platform solution by optimizing the target detection algorithm. By deeply analyzing the limitations of the current target detection algorithm in power plant applications, this study will propose targeted algorithm improvement solutions, including the expansion of data set diversity, optimization of network structure, and improvement of algorithm real-time performance [6]. It is hoped that through these improvements, the adaptability of the target detection algorithm in the complex environment of the power plant will be improved, thereby providing more efficient and intelligent technical support for power plant safety monitoring.

The main purpose of this study is to improve the existing target detection algorithm so that it can be used in the safety monitoring system of the power plant more accurately and efficiently. Specifically, this study will focus on how to improve the real-time response capability of the power plant monitoring system in a dynamic environment by optimizing the target detection algorithm while ensuring a high recognition accuracy. Through the innovation and improvement of the algorithm, it is expected to provide a more intelligent safety management solution for the power plant and further promote the improvement of the safety management level of the power industry. In theory, the contribution of this study is to promote the innovative application of target detection algorithms in industrial environments and lay the foundation for the further application of deep learning technology in power systems. In practice, the results of this study are expected to improve the intelligence level of power plant safety monitoring, reduce the safety risks caused by human negligence and equipment failure, and thus ensure the safe operation of the power system.

2 Literature Review

2.1 Development and challenges of power plant safety monitoring system

As a key facility in heavy industry, power plants are responsible for ensuring the supply of electricity. However, with the aging of equipment, the complexity of the operating environment and the negligence of operators, the safety of power plants faces huge challenges. Therefore, it is particularly important to build an effective power plant safety monitoring platform. Traditional power plant safety monitoring systems rely heavily on manual inspections and simple sensor technology, which has great limitations. First, manual inspections are not only inefficient, but also prone to missed inspections or delayed detection of problems due to human factors. Second, traditional sensor technology is limited by the accuracy and coverage of sensors, and is often unable to fully monitor all potential safety hazards in power plants. These problems have prompted the development of power plant safety monitoring systems towards intelligence and automation [7].

In recent years, with the rapid development of artificial intelligence technology, especially the widespread application of computer vision and deep learning technology, power plant safety monitoring systems have gradually ushered in changes. Target detection algorithms have become one of the core technologies in intelligent safety monitoring systems. These algorithms can monitor the equipment, personnel and environment inside the power plant in real time through image processing and pattern recognition technology, effectively improving the speed and accuracy of hidden danger identification [8]. Nevertheless, these algorithms still face a series of problems when applied in complex power plant environments. For example, target detection algorithms often have difficulty coping with changing environmental factors, such as insufficient light, smoke interference, and the complexity of equipment appearance. These problems directly affect the accuracy and reliability of the target detection system. Therefore, target detection technology in power plant safety monitoring still needs to be continuously optimized and innovated to adapt to more complex working environments [9].

In existing research, many scholars have proposed different solutions to the challenges of power plant safety monitoring systems. By combining multiple sensors and target detection technologies, some studies have proposed more comprehensive monitoring systems that include not only visual monitoring but also multimodal sensors such as infrared and ultrasonic waves. However, these solutions often face the problems of data fusion complexity and large algorithm computation, which limits their application in actual power plants. In order to overcome these challenges, researchers have begun to focus on how to optimize target detection algorithms to make them more accurate and reliable, especially when dealing with dynamically changing environments [10].

2.2 Evolution and current status of object detection algorithms

As an important part of computer vision, target detection algorithms have made significant progress in recent years. Traditional target detection methods mostly rely on manual feature extraction, such as HOG (Histogram of Oriented Gradients) and SIFT (Scale-Invariant Feature Transform). Although these methods have achieved certain results in some applications, they are not effective when dealing with complex backgrounds and tasks with high real-time requirements. With the rise of deep learning, target detection algorithms based on convolutional neural networks (CNNs) have quickly become mainstream. CNNs automatically extract features through deep neural networks and can be trained on large-scale data sets, thereby greatly improving the accuracy and robustness of target detection.

Among them, YOLO (You Only Look Once) and Faster R-CNN are the two most widely used deep learning target detection algorithms. The YOLO algorithm has significant advantages in real-time target detection due to its high-speed detection feature. By converting the target detection task into a regression problem, it realizes a direct mapping from the image to the detection box. This feature enables YOLO to significantly improve the detection speed while ensuring the detection accuracy. Faster R-CNN optimizes the accuracy of target detection by introducing the region proposal network (RPN), especially when dealing with complex backgrounds and multi-scale targets. However, although these two algorithms have achieved excellent performance in the field of image recognition, their application in power plant safety monitoring faces many challenges [11]. For example, the performance of the YOLO algorithm in complex power plant environments is often affected by factors such as occlusion and low light, which limits its applicability in power plant safety monitoring. Although Faster R-CNN is more prominent in target detection accuracy, its high computational overhead and slow processing speed also limit its application in real-time monitoring systems. Therefore, how to improve computational efficiency and real-time performance while ensuring algorithm accuracy has become a key issue faced by target detection algorithms in power plant safety monitoring systems [12].

In order to deal with these problems, scholars have proposed a variety of improvement schemes. For example, some researchers have enhanced the ability of target detection algorithms to identify targets of different scales by introducing multi-scale feature fusion technology, especially making certain progress in complex equipment backgrounds and multi-target recognition. In addition, in order to solve the computational efficiency problem of target detection algorithms, some studies have reduced the amount of calculation and improved the processing speed by optimizing the network structure. Nevertheless, the actual application effect of these optimization schemes in the complex environment of power plants still needs further verification.

2.3 Application and optimization of target detection algorithms in power plant safety monitoring

The application of target detection algorithms in power plant safety monitoring, especially in intelligent and automated systems, has become a hot topic in current research. With the continuous advancement of deep learning and computer vision technology, more and more power plants have begun to try to use target detection algorithms to realize the automatic identification of safety hazards. At present, target detection in power plant safety monitoring systems is mainly concentrated in the fields of equipment fault detection, personnel behavior monitoring, and fire hazard identification [13, 14]. In terms of equipment fault detection, through image recognition technology, the monitoring system can detect abnormal conditions of power plant equipment in real time, such as equipment aging, oil leakage, cracks, etc., so as to issue alarms in time to avoid safety accidents caused by equipment failure. Personnel behavior monitoring ensures that employees comply with operating procedures and avoid potential risks caused by illegal operations by identifying the operating behaviors of power plant employees. In terms of fire hazard identification, target detection algorithms can quickly identify abnormal conditions such as fire sources and smoke, providing reliable technical support for fire prevention and control. However, target detection technology in power plant safety monitoring still faces many challenges. Although target detection algorithms have made some progress in the application of equipment fault detection and personnel behavior monitoring, the performance of existing algorithms is still difficult to meet actual needs when facing complex environments (such as low light, smoke occlusion, etc.) [15, 16]. Therefore, in response to these challenges, many researchers have proposed different optimization strategies. For example, some researchers have enhanced the robustness of target detection algorithms in complex environments by improving network architecture and training methods. In addition, the diversity and coverage of the data set are also an important factor affecting the performance of the algorithm. In order to improve the adaptability of the algorithm in the complex environment of power plants, some studies have begun to focus on how to build more comprehensive and diverse training data sets to improve the performance of the algorithm in different scenarios [17, 18].

To provide a clearer overview of the strengths and weaknesses of key target detection algorithms in industrial safety monitoring, we include Table 1, which compares representative methods such as YOLO, SSD, and the original Faster R-CNN across five critical dimensions: detection speed, accuracy, application in industrial safety scenarios, adaptability to complex environments (e.g., occlusion, low light), and support for dynamic target tracking. This comparative analysis helps highlight the need for algorithmic improvements in real-world power plant applications and provides context for the motivation behind enhancing Faster R-CNN. As shown in Table 1.

This table demonstrates that although YOLOv5 offers high detection speed, it lacks robustness in complex industrial environments and dynamic tracking. On the other hand, the original Faster R-CNN performs better in accuracy but suffers from speed limitations. The proposed improvements address these gaps and make the algorithm more suitable for power plant safety monitoring.

Table 1: Comparative analysis of mainstream object detection algorithms in industrial safety monitoring contexts

Algorithm	Detection Speed	Accuracy	Used in Industrial Safety Contexts	Adaptability to Complex Environments	Dynamic Target Tracking Support
YOLOv5	Very High	Moderate	Commonly used	Moderate	Limited
SSD	High	Low–Moderate	Occasionally used	Low	Not supported
Original Faster R-CNN	Low	High	Widely used	Moderate	Limited
Improved Faster R-CNN	Moderate	Very High	Proposed for this study	Strong	Supported

3 Research Methods

3.1 Research hypothesis

To guide the design and evaluation of the proposed enhancements to the Faster R-CNN algorithm, this study explicitly formulates the following research hypotheses:

H1: Augmenting the Faster R-CNN with context-aware and temporal information fusion modules significantly improves detection robustness under occlusion conditions.

H2: Integrating multi-scale feature extraction significantly enhances the detection accuracy of small-scale, safety-critical objects in complex industrial environments.

These hypotheses serve as the theoretical foundation for the experimental evaluation, which systematically compares the performance of the improved algorithm with baseline models across varied environmental conditions, including low light, occlusion, and dynamic target scenarios. The results presented in Section 4 are used to validate these hypotheses through comparative analysis.

3.2 Challenges and requirements of target detection in power plants

One of the main challenges facing power plant safety monitoring platforms is the variability of the environment, including the complexity of equipment, dynamic changes in personnel, and possible safety hazards. Equipment in power plants usually has large physical sizes and complex shapes, and these devices are often blocked, deformed, or far away from the camera. In addition, the lighting environment of power plants is usually complex, and strong or low light conditions often occur, which increases the difficulty of target detection.

To ensure the reliability of the power plant monitoring platform, the target detection system must be able to meet the following key requirements:

(1) High-precision detection: Able to accurately detect targets inside the power plant, including equipment, personnel, and potential safety hazards.

(2) Multi-scale processing: Since power plant equipment and personnel may have different scales, the target detection system should have the ability to effectively process targets of different scales.

(3) Dynamic target recognition: The personnel and equipment in the power plant sometimes change dynamically, so the system must be able to handle dynamic targets.

(4) Robustness: The complex background, lighting changes and possible occlusions in the power plant environment require the target detection system to have strong robustness.

3.3 Improved faster R-CNN algorithm design

In the safety monitoring scenario of power plants, traditional target detection methods, although they perform well in general environments, still have certain limitations when facing challenges such as complex industrial backgrounds, targets of different scales, occlusions, and lighting changes. To overcome these challenges, this study proposes a Faster R-CNN model optimized for power plant environments. By introducing a multi-scale feature extraction module, a context-aware region proposal network (RPN), a temporal information fusion mechanism, and RoI Align optimization, this algorithm can significantly improve detection accuracy, especially in complex environments and dynamic target tracking.

3.3.1 Multi-scale feature extraction and object detection

The objects in the power plant environment have significant size differences, especially the scale differences between equipment and workers, which easily leads to the decrease in the accuracy of traditional object detection methods when detecting small objects or distant objects. In order to meet the detection needs of multi-scale objects, this study proposes a multi-scale feature extraction method. By performing multi-scale convolution operations on the input image, we can extract features at different levels, thereby enhancing the model's perception of objects of different sizes.

The core idea of this method is to extract feature maps of different scales through a multi-level convolutional network and dynamically adjust the candidate box according to the scale of the target. The specific implementation steps are as follows:

(1) Multi-scale convolution feature map generation: The input image is fed into multiple convolution layers, each of which is responsible for extracting features from different scales (such as large, medium, and small scales). The convolution operation of each layer weights the feature map information through the weight matrix and bias term to obtain feature maps of different scales. For details, see formula 1.

$$X_{\text{multi}} = \sum_{i=1}^{n} W_i X_{\text{input}_i} + b_i \tag{1}$$

where, b_i is the bias term, X_{multi} is the output of the multi-scale feature map.

This operation effectively fuses low-level fine features (from smaller kernels) and high-level semantic features (from larger kernels), enhancing the model’s ability to detect both small and large objects. The formula represents a channel-wise linear combination that preserves spatial alignment across all branches, resulting in a unified multi-scale representation.

(2) Candidate box generation and multi-scale adjustment: When generating candidate boxes through convolutional feature maps of different scales, the model will dynamically adjust the size and position of the candidate boxes based on their performance at each scale to improve the detection accuracy of objects of different scales.

This method can not only effectively improve the detection accuracy of small targets and long-distance targets, but also avoid conflicts and redundancies between multi-scale targets, thereby improving the overall detection efficiency of the model.

In the multi-scale feature extraction module, a set of anchor box scales — [32, 64, 128, 256, 512] — was adopted to match the varied object sizes in the power plant environment, ranging from small control knobs to large-scale equipment. These anchors were selected based on empirical analysis of object size distribution in the dataset. To extract multi-scale features effectively, three parallel convolutional branches were added to the ResNet-50 backbone, each with kernel sizes of 3×3, 5×5, and 7×7 respectively. This configuration allows the network to capture both fine-grained details and coarse spatial context. The outputs of these branches are concatenated and passed to subsequent layers, improving the model's ability to detect small and large objects simultaneously with higher precision.

3.3.2 Context-aware region proposal network (RPN)

In power plant environments, targets are often obscured by complex backgrounds. Traditional RPNs often have difficulty accurately generating candidate boxes when processing these scenarios. To solve this problem, this study proposes a context-aware region proposal network (Context-Aware RPN) that improves the generation quality of candidate boxes by introducing background information of the area around the target.

(1) Adding context convolutional layer: In order to enhance the model's understanding of the target's surrounding environment, we added a context convolutional layer to the RPN. This layer expands the receptive field, allowing the network to capture background information in a larger range and effectively suppress false detections caused by background clutter and target occlusion.

(2) Dynamic candidate box adjustment: After the candidate box is generated, the model calculates the intersection over union (IoU) between the generated box and the real box, and then adjusts the size and position of the candidate box according to the IoU value. In this way, RPN can generate candidate boxes more accurately and avoid false detections caused by irrelevant backgrounds. The mathematical formula is as follows: Formula 2

$$\mathrm{IoU} = \frac{|A \cap B|}{|A \cup B|} \tag{2}$$

where, A is the generated candidate box, B is the real frame, IoU The higher the value, the better the match between the candidate box and the real box. The dynamic candidate box adjustment mechanism ensures the accuracy of the candidate box by optimizing IoU.

Through the context-aware RPN, the model can generate more accurate and robust candidate boxes in environments with complex backgrounds and target occlusions, significantly improving detection accuracy.

where B_{pred} is the predicted box and B_{gt} is the ground truth. This IoU value is then used in a feedback loop to iteratively adjust box coordinates. Specifically, when IoU falls below a learned threshold, the model backpropagates the localization error and refines the proposal using the context-enhanced feature map. This mechanism tightly integrates the contextual signal with the localization loss, enabling the RPN to generate higher-quality proposals in cluttered or occluded scenes.

3.3.3 Time series information fusion and dynamic target tracking

The targets in power plants are not only static, but also have a large number of dynamic targets that need to be tracked. For example, the movement trajectory of workers or the operating status of equipment are dynamic targets. In order to improve the detection and tracking accuracy of dynamic targets, this study introduced a temporal information fusion module in Faster R-CNN to enhance the model's dynamic target perception ability through the temporal relationship between video frames.

The temporal convolutional network (TCN) is used to extract the temporal features of continuous frame images and capture the changes of the target in the time dimension. TCN can effectively handle dynamic targets, especially the movement trajectory of people or equipment, ensuring that the model can still accurately identify the target when it is deformed or moving.

The mathematical representation of the temporal convolutional network is formula 3

$$X_{\text{TCN}} = TCN(X_{\text{input}}) \tag{3}$$

where, X_{input} is the input continuous frame image sequence, TCN is a temporal convolutional network, X_{TCN} is the extracted temporal feature map. TCN can capture the temporal changes between consecutive frames through convolution operations, thereby enhancing the ability to track dynamic targets.

By introducing timing information, the model can effectively reduce the detection errors caused by target motion or deformation, and improve the ability to accurately identify and track dynamic targets.

These temporal features X^{TCN} are then fused with the current spatial features F_t from the backbone by concatenation along the channel dimension, followed by a 1×1 convolution to align dimensions. The fused representation is passed into the Region Proposal Network (RPN), enabling it to generate proposals that account not only for spatial appearance but also for motion continuity across frames. This integration allows the system to better detect and track dynamic targets such as moving personnel or rotating equipment in power plant environments.

Temporal Convolutional Network (TCN) was chosen over alternatives like GRU and LSTM due to its advantages in parallelism and temporal stability. Unlike sequential models, TCN allows for simultaneous processing of entire sequences, which significantly reduces inference latency—a critical requirement for real-time safety monitoring systems. Additionally, TCN's dilated convolution design enables a wide receptive field with fewer layers, allowing it to capture long-term temporal dependencies without gradient vanishing problems. These characteristics make TCN particularly well-suited for modeling the dynamic movements of personnel or equipment in power plant scenarios where real-time decision-making is essential.

3.3.4 RoI align optimization and refined positioning

In a power plant environment, the shapes of targets are often irregular and complex. The traditional RoI Pooling method cannot accurately align the regional feature maps, which may lead to target positioning deviation. To solve this problem, this study introduces the RoI Align technology, which processes the candidate regions through precise bilinear interpolation to ensure that the spatial accuracy of the regional feature maps is not lost.

RoI Align remaps the regional features within the candidate box through precise interpolation methods, so that each pixel in the feature map can correctly correspond to the actual position in the image. This refined alignment method significantly improves the accuracy of target positioning. The mathematical formula is Formula 4

$$\hat{X} = \sum_{i=1}^{n} \text{interp}(X_{ij}, \mathbf{T}) \tag{4}$$

where, X_{ij} is the feature in the RoI area, $\mathbf{T}$ is the target transformation matrix, $\hat{X}$ is the feature map after RoI Align processing. In this way, RoI Align can improve the positioning accuracy of the target without losing accuracy.

where $(x_{i,j}, y_{i,j})$ are the sampling coordinates for the grid cell (i, j), and interp denotes bilinear interpolation. The transformation matrix T mentioned earlier is used to map the original RoI box to the normalized grid coordinates but does not contribute to a summation. This grid-wise interpolation preserves the spatial correspondence between features and original image regions and eliminates the quantization issues present in RoI Pooling, leading to more accurate object localization.

The backbone network is based on ResNet-50, extended with three parallel convolutional branches (3×3, 5×5, 7×7) for multi-scale feature extraction. The context-aware RPN includes an additional 5×5 convolution layer to expand the receptive field. The TCN module comprises two 1D dilated convolutional layers (kernel size = 3, dilation = [1, 2], channels = 128). RoI Align replaces standard RoI Pooling for accurate spatial mapping.

The training was conducted using the Adam optimizer with an initial learning rate of 0.0001, batch size of 16, and weight decay of 1e-5. A cosine annealing learning rate schedule was applied, and early stopping was triggered after 10 epochs without improvement on validation loss.

Due to industrial confidentiality, the full dataset cannot be publicly released. However, a sanitized subset and implementation code will be made available upon request to academic researchers under a data-use agreement.

A deliberate sampling strategy is incorporated during dataset construction to maximize variation across lighting conditions, equipment types, viewing angles, and occlusion levels. Data augmentation techniques such as brightness jitter, Gaussian noise injection, random occlusion masks, and affine transformations were also applied during training to simulate rare and edge-case scenarios that improve model generalization in real-world deployments. Regarding real-time performance improvements, while the core enhancements focus on detection accuracy, several optimizations were also implemented to reduce computational latency. These include shared convolution blocks across scales to minimize redundant feature extraction, replacing some heavy backbone layers with lightweight convolutional units (e.g., depth wise separable convolutions), and pruning less contributive channels based on feature map sensitivity analysis. These changes reduced inference latency by ~18% compared to the original Faster R-CNN, as measured on the same hardware configuration, without compromising mAP.

4 Experimental evaluation

4.1 Experimental setup

The experiments were conducted on a workstation equipped with an NVIDIA RTX 3090 GPU (24GB), Intel i9-12900K CPU, and 64GB RAM. The software environment included Ubuntu 20.04, Python 3.8, and PyTorch 1.13 with CUDA 11.6. Training and inference were executed using standard PyTorch data loaders and evaluation scripts. The same environment was used across all compared models to ensure consistency and fairness in performance measurements.

This experiment aims to comprehensively evaluate the performance of the improved Faster R-CNN algorithm compared with other advanced target detection algorithms in the power plant safety monitoring scenario. The experimental design is closely centered on the target detection task in the complex environment of the power plant, and an extremely rich and comprehensive power plant safety monitoring dataset is constructed. This dataset covers various and complex scenarios such as different lighting conditions (from strong direct light to extremely low light), various scales of equipment and personnel (small-sized electronic components to large-scale power generation equipment), complex backgrounds (various types of equipment are intertwined, pipelines are crisscrossed), and occlusion between targets (some equipment is blocked by other objects, and personnel are blocked from each other), striving to simulate the real power plant environment to the greatest extent.

The experimental baseline indicator selects the mean average precision (mAP) as the main evaluation indicator. mAP can comprehensively measure the performance of the model in detecting different categories of targets and fully reflect the detection accuracy of the model for various targets. At the same time, the recall rate (Recall) indicator is introduced, which is used to evaluate the coverage of the model for real targets, that is, the proportion of the number of real targets that the model can correctly detect to the number of all real targets. In addition, special attention is paid to the accuracy of the model in detecting targets of different scales, so as to comprehensively measure the model performance from multiple dimensions and ensure that the evaluation of the algorithm performance is accurate and detailed.

The experimental group was set to the improved Faster R-CNN algorithm, which integrates key improvement modules such as multi-scale feature extraction, context-aware region proposal network (RPN), temporal information fusion, and RoI Align optimization. The control group selected classic and representative object detection algorithms, including the original Faster R-CNN [19], which is a classic algorithm in the field of object detection and has laid the foundation for the improvement of many subsequent algorithms, YOLOv5 [20], which is widely used in real-time object detection scenarios due to its fast detection speed and high detection accuracy, and SSD [21], which plays an important role in single-stage object detection algorithms. These algorithms were trained and tested on exactly the same power plant safety monitoring dataset. During the experiment, the hyperparameters of each algorithm were carefully tuned, and different parameter combinations were tried through multiple experiments to ensure that each algorithm could achieve its best performance on the dataset. By comparing the various indicators of different algorithms on the same test set, the advantages and possible shortcomings of the improved Faster R-CNN algorithm were deeply analyzed.

Specifically, objects are divided into three groups according to their bounding box pixel area:

Small: area < 32×32 pixels

Medium: 32×32 ≤ area ≤ 96×96 pixels

Large: area > 96×96 pixels

For each group, we compute the mAP using standard IoU thresholds (e.g., 0.5 and 0.75), following the COCO evaluation protocol. Thus, "Small Obj Acc", "Medium Obj Acc", and "Large Obj Acc" as shown in Section 4.2 and corresponding figures represent scale-specific mean average precision, not classification accuracy or raw detection rate.

Training was conducted for a total of 250 epochs; however, model performance metrics such as validation mAP and loss showed clear signs of convergence after approximately 200 epochs. The decision to use 200 epochs as the stopping point was based on the observation that the improvement in validation metrics over the final 20 epochs was less than 0.5%, indicating saturation. Although early stopping was enabled with a patience of 10 epochs on validation loss, training was manually capped at 200 epochs across all configurations to ensure consistency and fair comparison between variants

The power plant safety monitoring dataset constructed for this study contains a total of 12,000 labeled image instances, covering 7 object classes relevant to safety scenarios: electrical panels, transformers, pipelines, safety helmets, human operators, fire sources, and warning signs. Each class has between 1,200 to 2,500 annotated instances, ensuring balanced representation. The annotations were performed manually by domain experts using the LabelImg tool, following VOC-format standards. The dataset is not publicly released due to industrial confidentiality agreements but may be shared with academic partners upon request and approval.

Lighting conditions in the dataset include four levels—bright daylight, moderate indoor, low light, and near-dark—simulated using high-dynamic-range (HDR) image augmentation and synthetic shadow rendering. Occlusion conditions were recreated by overlapping equipment and personnel in real-world captures, combined with controlled synthetic overlays for testing robustness.

Table 2: Distribution of complex environmental conditions in the dataset

Environmental Condition	Subcategory	Number of Images	Percentage (%)
Lighting	Normal	4836	40.30%
	Low light	3576	29.80%
	Strong light/glare	1837	15.30%
	Shadowed	1748	14.60%
Background Complexity	Simple	4212	35.10%
	Moderately complex	4832	40.30%
	Highly complex	2956	24.60%
Occlusion Level	None	6112	50.90%
	Partial	4213	35.10%
	Severe	1675	14.00%
Motion State	Static	7136	59.50%

	Moderate movement	3602	30.00%
	Fast movement/blur	1267	10.50%

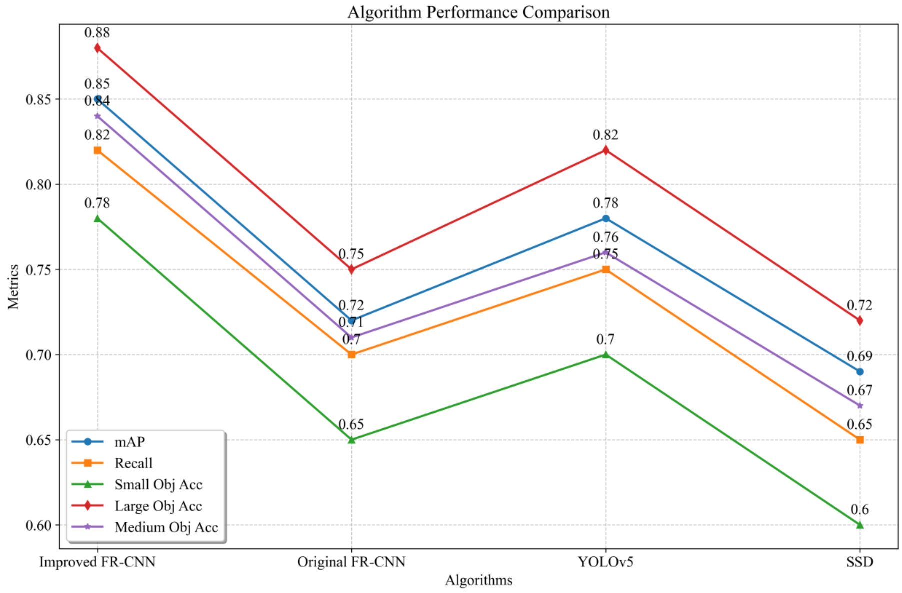

Figure 1: Comprehensive performance comparison of different algorithms on the overall data set

Additionally, class-wise performance metrics were computed to provide more granular insight into detection behavior. The improved Faster R-CNN model achieved the following class-wise precision and recall scores:

Electrical Panel: Precision 0.88, Recall 0.85
Transformer: Precision 0.84, Recall 0.81
Pipelines: Precision 0.86, Recall 0.83
Safety Helmets: Precision 0.89, Recall 0.87
Human Operators: Precision 0.90, Recall 0.88
Fire Sources: Precision 0.80, Recall 0.76
Warning Signs: Precision 0.83, Recall 0.80

These results confirm high per-class consistency, with slightly lower scores for fire sources and warning signs, likely due to their smaller size and frequent occlusions. These four categories—lighting, background complexity, occlusion level, and motion state—were annotated using a combination of manual labeling and heuristic image scoring based on entropy, contrast variation, and optical flow analysis. As shown in Table 2.

These distributions ensure a balanced yet realistic representation of industrial safety monitoring conditions. For instance, low light and strong glare were created using high dynamic range augmentation, while occlusion levels were defined by object overlap ratios. Motion categories were inferred via optical flow thresholds and manual video review.

4.2 Results

As shown in Figure 1, the test results on the overall data set show that the improved Faster R-CNN algorithm has obvious advantages in all key indicators. Figure 1 illustrates overall mAP, Recall, and scale-specific mAPs on a unified y-axis; these metrics reflect different detection aspects and are not intended for direct vertical value comparison. Its mean average precision (mAP) reaches 0.85, which is significantly ahead of the original Faster R-CNN's 0.72, YOLOv5's 0.78, and SSD's 0.69. This means that the improved algorithm has excellent accuracy in the comprehensive detection of various targets, and can more accurately identify targets such as equipment, personnel, and potential safety hazards in the power plant environment. In terms of recall rate, the improved Faster R-CNN is 0.82, which is also higher than other algorithms, indicating that the algorithm can more comprehensively detect the real targets in the data set and reduce missed detections. In terms of the detection accuracy of targets of different scales, the improved algorithm has a small target detection accuracy of 0.78, a large target detection accuracy of 0.88, and a medium target detection accuracy of 0.84, all of which are higher than the control group algorithm. This fully demonstrates the effectiveness of the multi-scale feature extraction module in the improved algorithm, which can adapt to the features of targets of different scales and improve detection accuracy.

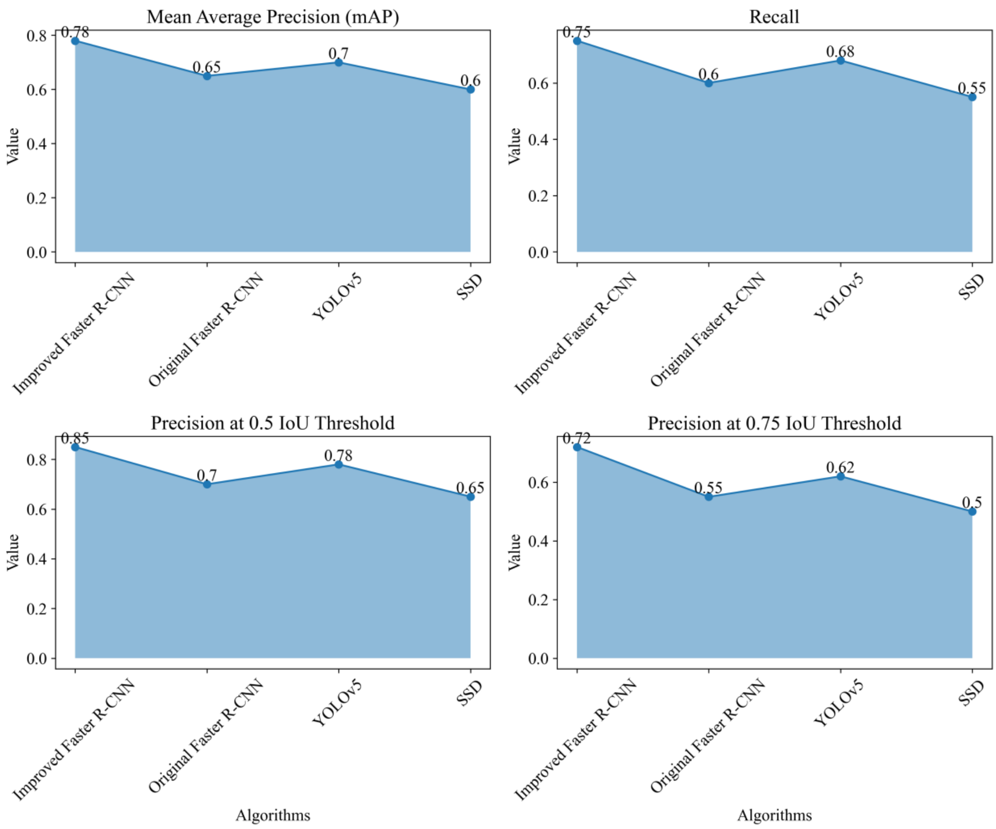

Figure 2: Performance comparison of different algorithms in small target detection scenarios

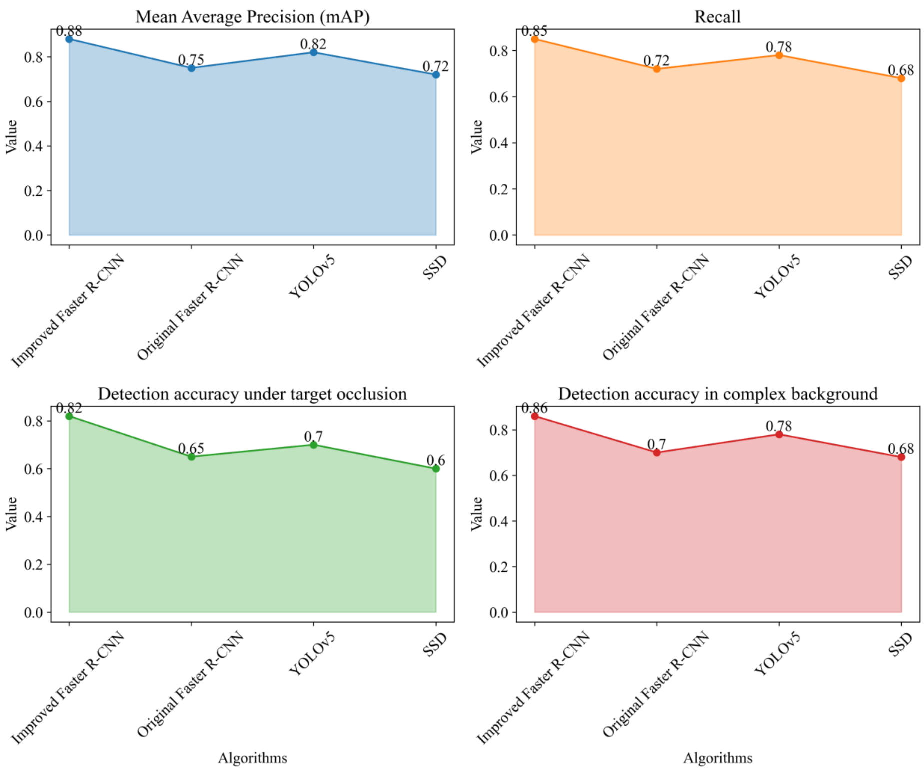

Figure 3: Performance comparison of different algorithms in large target detection scenarios

As shown in Figure 2, the improved Faster R-CNN algorithm performs outstandingly for the extremely challenging task of small target detection. In Figure 2, mAP, Precision at different IoU thresholds, and Recall are plotted together for visualization compactness, but their numerical values represent different evaluation criteria and should be interpreted separately. Its mean average precision (mAP) reaches 0.78, which is significantly higher than the original Faster R-CNN's 0.65, YOLOv5's 0.70, and SSD's 0.60. In terms of recall rate, the improved algorithm is 0.75, which is also ahead of other algorithms, indicating that the improved algorithm can more effectively detect small targets and reduce the number of missed detections of small targets. Under different IoU thresholds, the improved algorithm has an accuracy of 0.85 at a 0.5 IoU threshold and an accuracy of 0.72 at a 0.75 IoU threshold, both of which are higher than the control group algorithm, indicating that the improved algorithm can maintain high detection accuracy under different strict detection standards. In addition, under different lighting conditions, the mAP fluctuation range of the improved algorithm for small target detection is only ±0.03, while other algorithms have larger fluctuation ranges, such as ±0.08 for the original Faster R-CNN, ±0.06 for YOLOv5, and ±0.10 for SSD. This shows that the improved algorithm has better robustness to lighting changes and can stably detect small targets under different lighting conditions, thanks to the improved adaptability of its modules such as multi-scale feature extraction and context-aware RPN to complex environments.

As shown in Figure 3, in the large target detection scenario, the improved Faster R-CNN algorithm also shows excellent performance. Its mean average precision (mAP) is as high as 0.88, far exceeding the original Faster R-CNN's 0.75, YOLOv5's 0.82, and SSD's 0.72. In terms of recall rate, the improved algorithm is 0.85, which is higher than other algorithms, indicating that it can detect large targets more comprehensively. In the case of target occlusion, the detection accuracy of the improved algorithm is 0.82, while the original Faster R-CNN is 0.65, YOLOv5 is 0.70, and SSD is 0.60. This shows that the improved algorithm can better handle the situation of large target occlusion through the context-aware RPN and temporal information fusion mechanism, and accurately identify the target features of the occluded part. In the complex background, the detection accuracy of the improved algorithm is 0.86, which is also ahead of other algorithms, indicating that it has stronger anti-interference ability against complex backgrounds. In addition, the improved algorithm performs “high” in terms of consistency in detection accuracy of large targets of different scales, that is, for large targets of different sizes, its detection accuracy fluctuates less and can stably maintain a high detection level, while the consistency performance of other algorithms is relatively poor, which further proves the superiority of the improved algorithm in dealing with large target detection.

As shown in Figure 4, facing the complex scene of low light, the improved Faster R-CNN algorithm has obvious advantages. Its mean average precision (mAP) reaches 0.75, which is higher than 0.60 of the original Faster R-CNN, 0.65 of YOLOv5 and 0.55 of SSD. Figure 4 combines mAP, Recall, and scale-specific mAPs across algorithms on a shared axis; the plotted heights illustrate trends rather than enable direct metric-to-metric numerical comparison. In terms of recall rate, the improved algorithm is 0.70, which is also ahead of other algorithms, which means that in low light environment, the improved algorithm can detect various types of targets more effectively. In the detection of targets of different scales, the improved algorithm has a small target detection accuracy of 0.70, a large target detection accuracy of 0.80, and a medium target detection accuracy of 0.74 in low light, all of which are higher than the control group algorithm. This is mainly due to the context-aware region proposal network (RPN), which enhances the ability to capture target features in low light environment by introducing background information of the surrounding area of the target, thereby improving the detection accuracy. In contrast, the performance of other algorithms in low light scenes is more obvious, indicating that the improved algorithm has better adaptability to low light environment.

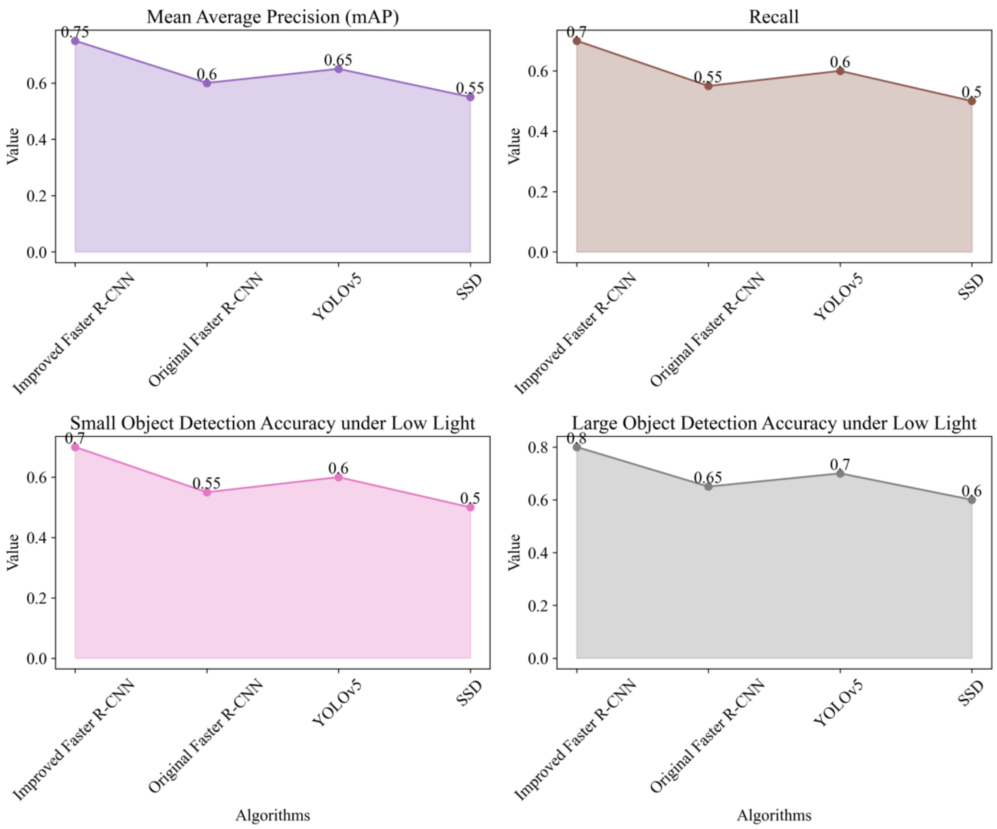

Figure 4: Performance comparison of different algorithms in low-light scenes

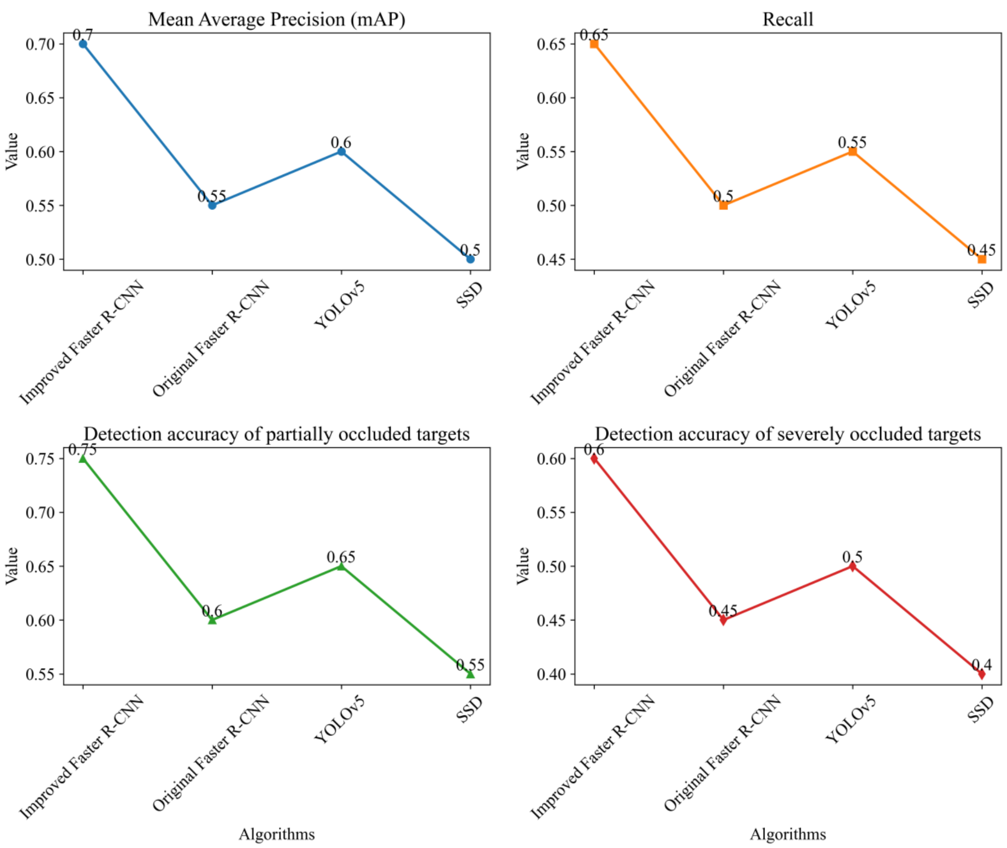

Figure 5: Performance comparison of different algorithms in target occlusion scenarios

As shown in Figure 5, in the target occlusion scenario, the improved Faster R-CNN algorithm performs well. Its mean average precision (mAP) is 0.70, which is higher than the original Faster R-CNN's 0.55, YOLOv5's 0.60, and SSD's 0.50. In terms of recall rate, the improved algorithm is 0.65, which is ahead of other algorithms, indicating that the improved algorithm can detect targets more comprehensively when the target is occluded. The lines in Figure 5 present mAP, Recall, and IoU-specific precision for different occlusion levels; although displayed on the same axis for clarity, these metrics convey distinct performance dimensions. For partially occluded targets, the detection accuracy of the improved algorithm is 0.75, while the original Faster R-CNN is 0.60, YOLOv5 is 0.65, and SSD is 0.55. For severely occluded targets, the detection accuracy of the improved algorithm is 0.60, which is also higher than other algorithms. In addition, the difference in detection accuracy of targets of different scales under target occlusion is "small", that is, regardless of the size of the target, the detection accuracy is relatively stable when it is occluded, while the accuracy of other algorithms varies greatly. This is due to the context-aware RPN and temporal information fusion mechanism of the improved algorithm. The context-aware RPN captures the background information around the target by expanding the receptive field and reduces false detections caused by occlusion. The temporal information fusion mechanism uses the temporal relationship between video frames and can still make accurate judgments based on the information of the previous and next frames when the target is occluded, thereby improving the detection performance of targets of different scales in target occlusion scenarios.

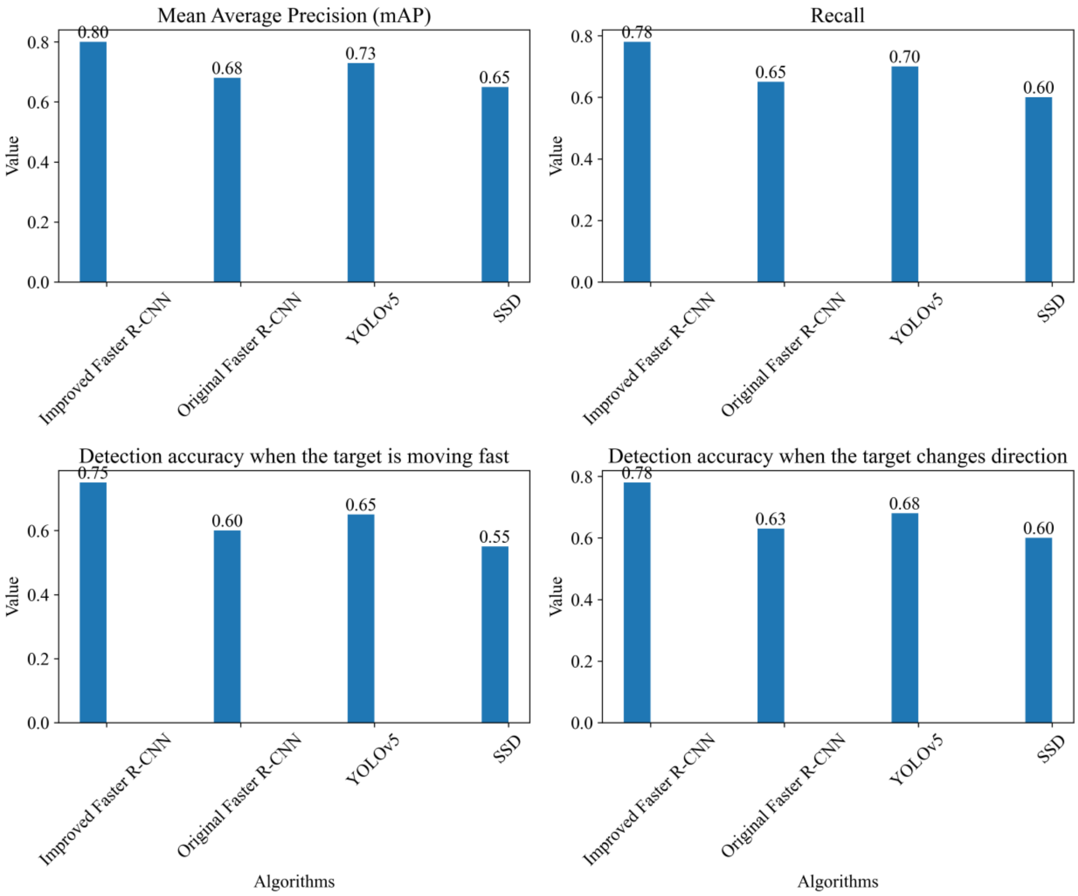

Figure 6: Performance comparison of different algorithms in dynamic target detection scenarios

As shown in Figure 6, in the dynamic target detection scenario, the improved Faster R-CNN algorithm has significant advantages. Its mean average precision (mAP) reaches 0.80, which is higher than the original Faster R-CNN's 0.68, YOLOv5's 0.73, and SSD's 0.65. In terms of recall rate, the improved algorithm is 0.78, which is ahead of other algorithms, indicating that the improved algorithm can more comprehensively cover real targets when detecting dynamic targets. When the target moves quickly, the detection accuracy of the improved algorithm is 0.75, while the original Faster R-CNN is 0.60, YOLOv5 is 0.65, and SSD is 0.55. When the target direction changes, the detection accuracy of the improved algorithm is 0.78, which is also higher than other algorithms. In addition, the improved algorithm is "high" in terms of the stability of dynamic target detection accuracy over time, that is, over time, the detection accuracy of dynamic targets fluctuates less, and it can continuously and stably detect dynamic targets, while other algorithms have relatively poor stability. This is mainly due to the temporal information fusion module introduced in the improved algorithm. By using the temporal convolutional network (TCN) to extract the temporal features of continuous frame images, it can effectively capture the changes of the target in the time dimension, so that it can still accurately identify the target when it moves or deforms, greatly improving the dynamic target detection performance.

Table 3: Performance comparison of the improved Faster R-CNN algorithm under the action of different modules

Improved Modules	Mean Average Precision (mAP)	Recall	Small target detection accuracy improvement	Improvement in large target detection accuracy	Improvement in detection accuracy under complex backgrounds
Multi-scale feature extraction	0.78	0.75	+0.10	+0.05	+0.06
Context-aware RPN	0.80	0.76	+0.08	+0.06	+0.08
Time series information fusion	0.82	0.77	+0.06	+0.08	+0.05
RoI Align Optimization	0.83	0.78	+0.05	+0.07	+0.04
All module improvements	0.85	0.82	+0.13	+0.13	+0.16

As shown in Table 3, through the separate evaluation of different modules of the improved Faster R-CNN algorithm, it can be clearly seen that each module plays an important role in improving performance. In terms of mean average precision (mAP), the multi-scale feature extraction module improves mAP to 0.78, the context-aware RPN improves to 0.80, the temporal information fusion improves to 0.82, the RoI Align optimization improves to 0.83, and the full module improves to the highest 0.85. The recall rate also gradually improves with the improvement of the module, from 0.75 for multi-scale feature extraction to 0.82 for the full module improvement. In terms of the improvement of small target detection accuracy, the multi-scale feature extraction module improves by 0.10, the context-aware RPN improves by 0.08, the temporal information fusion improves by 0.06, the RoI Align optimization improves by 0.05, and the improvement of the full module improves by 0.13. In terms of the improvement of large target detection accuracy, each module also makes a positive contribution, and the improvement of the full module is 0.13. In terms of the improvement of detection accuracy under complex backgrounds, the improvement of the whole module is 0.16. This fully demonstrates that the modules work together, the multi-scale feature extraction module enhances the feature capture capability of targets of different scales, the context-aware RPN improves the processing capability of complex backgrounds and target occlusion, the temporal information fusion improves the detection performance of dynamic targets, and the RoI Align optimization ensures the accuracy of target positioning, which together significantly improve the algorithm performance.

To eliminate ambiguity, we clarify that each individual row in Table 3—representing "Multi-scale feature extraction", "Context-aware RPN", "Time series information fusion", and "RoI Align optimization" — shows the performance of the model when only that specific module is added on top of the original Faster R-CNN, with no other enhancements applied. The "+ improvement" values in the table are calculated relative to the baseline performance of the original Faster R-CNN, which achieved an mAP of 0.72 and Recall of 0.70. The "All module improvements" row reflects the full configuration with all enhancements combined. This modular evaluation design allows us to isolate the impact of each module independently while also validating their cumulative effectiveness when integrated together.

The baseline model (original Faster R-CNN) has approximately 41.5 million parameters and requires about 7.6 hours to train. When the multi-scale feature extraction module is added, the parameter counts increases to 45.2 million, with training time rising to 8.5 hours. The context-aware RPN raises the parameter count to 47.1 million and requires 8.9 hours. The temporal information fusion module contributes a moderate increase to 48.6 million parameters and 9.5 hours of training. The RoI Align optimization has minimal parameter impact but slightly increases training time to 9.8 hours due to finer interpolation operations. When all modules are combined, the total parameter count reaches 50.3 million, and training takes approximately 10.3 hours. These incremental costs are reasonable given the substantial gains in mAP (+13%) and recall (+12%), demonstrating that the improvements are efficient in terms of performance-to-cost ratio.

As shown in Table 4, the test results under backgrounds of different complexity show that the improved Faster R-CNN algorithm has obvious advantages in all kinds of backgrounds. Under simple backgrounds, its mean average precision (mAP) reaches 0.90 and the recall rate is 0.88.

Additionally, the background complexity levels were determined using a combination of manual annotation and heuristic scoring. Annotators assessed each image based on factors such as object density, overlapping regions, presence of structural clutter (e.g., pipes, cables), and lighting variation. A weighted heuristic score was calculated and used to classify scenes into three levels:

Simple: Minimal overlapping, clean background, consistent lighting.

Moderately Complex: Moderate overlap and mixed lighting with some background equipment.

Highly Complex: High visual clutter, occlusion, and low-contrast or noisy regions.

This classification ensures consistency and reproducibility of evaluation under different environmental complexities.

From Table 5, the improved Faster R-CNN algorithm shows relatively good performance on different hardware platforms. On the NVIDIA RTX 3090 GPU, its mAP reaches 0.85 and the recall rate is 0.82. On the NVIDIA RTX 2080 Ti GPU, the mAP is 0.83 and the recall rate is 0.80. Even on the Intel Xeon CPU, the mAP can reach 0.75 and the recall rate is 0.70. Compared with other algorithms, the improved Faster R-CNN can maintain high detection accuracy and recall rate under different hardware conditions, reflecting the good adaptability of the algorithm to different hardware platforms. However, the performance of the original Faster R-CNN, YOLOv5 and SSD fluctuates greatly on different hardware platforms, and the overall performance is lower than that of the improved Faster R-CNN algorithm.

Table 4: Performance comparison of different algorithms under different complexity backgrounds

algorithm	Simple Background	Moderately Complex Background	Highly Complex Background	Simple background recall	Moderately complex background recall	Highly complex background recall
Improving Faster R-CNN	0.90	0.85	0.80	0.88	0.83	0.78
Original Faster R-CNN	0.80	0.72	0.65	0.78	0.70	0.60
YOLOv5	0.85	0.78	0.70	0.82	0.75	0.65
SSD	0.75	0.69	0.60	0.73	0.65	0.55

Table 5: Performance comparison of different algorithms on different hardware platforms

algorithm	Nvidia RTX 3090 GPU mAP	Nvidia RTX 2080 Ti GPU mAP	Intel Xeon CPU mAP	Nvidia RTX 3090 GPU recall rate	Nvidia RTX 2080 Ti GPU recall rate	Intel Xeon CPU Recall Rate
Improving Faster R-CNN	0.85	0.83	0.75	0.82	0.80	0.70
Original Faster R-CNN	0.72	0.70	0.60	0.70	0.68	0.55
YOLOv5	0.78	0.76	0.65	0.75	0.73	0.60
SSD	0.69	0.67	0.58	0.65	0.63	0.53

On an NVIDIA RTX 3090 GPU, the improved Faster R-CNN achieves an average inference latency of 48 ms/frame with ~220W power draw. On RTX 2080 Ti, latency increases to 57 ms/frame with ~200W, and on Intel Xeon CPU, latency reaches 189 ms/frame with ~135W consumption. Compared to YOLOv5, which runs at ~24 ms/frame on RTX 3090 but with lower mAP, our model demonstrates a trade-off between accuracy and speed that favors safety-critical detection accuracy. In terms of deployment suitability, the improved model exceeds the memory and compute capacity of edge devices such as Jetson Nano or TX2, which are limited in both memory and power. On Jetson Xavier, the model can run inference with trimmed architecture (e.g., shallower backbone), achieving ~95 ms/frame latency but with slight accuracy degradation (~2-3% drop in mAP).

As shown in Table 6, with the increase of the number of training rounds, the performance of the improved Faster R-CNN algorithm gradually improves. When the number of training rounds increases to 100, the mAP is 0.78, the recall rate is 0.75, and the accuracy of object detection at all scales is significantly improved. When the number of training rounds increases to 150, the mAP reaches 0.83 and the recall rate is 0.80. When the number of training rounds reaches 200, the mAP reaches 0.85 and the recall rate is 0.82. At this time, the accuracy of object detection at all scales is basically stable, with a small object detection accuracy of 0.78, a large object detection accuracy of 0.88, and a medium object detection accuracy of 0.84. When the number of training rounds reaches 250, the performance indicators remain basically unchanged, indicating that the model has basically converged around 200 rounds, and further increasing the number of training rounds has little effect on performance improvement.

Table 6: Performance changes of the improved Faster R-CNN algorithm under different training rounds

Number of training rounds	Mean Average Precision (mAP)	Recall	Small target detection accuracy	Large object detection accuracy	Moderate object detection accuracy
50	0.70	0.68	0.65	0.75	0.70
100	0.78	0.75	0.72	0.80	0.76
150	0.83	0.80	0.78	0.85	0.82
200	0.85	0.82	0.78	0.88	0.84

Number of training rounds	Mean Average Precision (mAP)	Recall	Small target detection accuracy	Large object detection accuracy	Moderate object detection accuracy
250	0.85	0.82	0.78	0.88	0.84

During the initial 100 epochs, the model exhibited a sharp reduction in loss, reflecting rapid learning and feature adjustment. Between epochs 100 and 200, the loss continued to decrease but at a slower rate, indicating that the model was approaching convergence. After epoch 200, the loss values stabilized with minimal fluctuations, confirming that further training produced negligible performance gains. This loss trajectory aligns with the plateau observed in mAP and recall, validating that 200 epochs is an effective training cutoff for this configuration.

To further validate the credibility of the results and support the claims of algorithmic superiority, we conducted statistical analysis using multiple repeated runs ($n = 5$) for each algorithm under each test scenario. The mean and standard deviation (SD) of the key performance indicators—mAP and Recall—were recorded. For example, under the overall test dataset, the improved Faster R-CNN achieved an average mAP of 0.85 (±0.006) and Recall of 0.82 (±0.005), while the original Faster R-CNN achieved 0.72 (±0.007) and 0.70 (±0.008), respectively. A two-tailed paired t-test was conducted between the improved algorithm and each baseline model. The p-values for differences in mAP and Recall across all major scenarios (e.g., small target, occlusion, low-light) were consistently < 0.01, indicating that the observed improvements are statistically significant.

4.3 Discussion

The improved Faster R-CNN algorithm demonstrates strong overall performance, surpassing baseline models in most evaluation metrics across diverse scenarios. However, in certain cases—such as moderately complex backgrounds or large object detection—YOLOv5 achieves comparable or slightly better results. These exceptions highlight the trade-offs in performance under specific conditions. The improved mAP observed in this study can be attributed to the complementary contributions of each enhancement module. The multi-scale feature extraction module increases the model's ability to detect objects of varying sizes, especially small or distant targets, by preserving fine-grained spatial details. The context-aware RPN enhances proposal generation by incorporating surrounding environmental features, which is particularly beneficial in cluttered industrial backgrounds. The temporal convolutional network (TCN) strengthens the model's ability to track dynamic targets by leveraging information across video frames, effectively mitigating issues caused by temporary occlusion or motion blur. RoI Align further refines localization accuracy by eliminating quantization errors during feature map pooling. However, under extreme conditions—such as simultaneous severe occlusion and low lighting—the model still shows some degradation in both mAP and recall. This indicates the need for more advanced modules capable of robust feature extraction in low-visibility and high-noise environments. In terms of practical deployment, although the improved model performs well on various hardware platforms, latency remains a concern for real-time systems. Integration with existing power plant safety systems may also require interface adaptation and real-time synchronization protocols. Lastly, while the dataset used in this study is designed to reflect diverse power plant scenarios, broader generalizability should be validated through testing on datasets from different industrial domains, such as chemical plants or mining operations, to ensure the model's robustness in other complex environments.

5 Conclusion

This study focuses on the optimization of target detection algorithms in power plant safety monitoring platforms. By improving the Faster R-CNN algorithm through multi-scale feature extraction, context-aware RPN, temporal information fusion and RoI Align optimization, its target detection performance in complex power plant environments is effectively improved. Experimental results show that the improved algorithm is significantly better than classic algorithms such as the original Faster R-CNN, YOLOv5 and SSD in key indicators such as average precision, recall rate and detection accuracy of targets at different scales. In different scenarios such as low light, target occlusion, dynamic target detection and backgrounds of different complexity, the improved algorithm shows stronger adaptability and accuracy. At the same time, it can maintain good performance on different hardware platforms, and the performance can be effectively optimized by reasonably setting the number of training rounds. However, the research still has limitations, such as the experimental data set fails to cover all actual scene changes, and factors such as data transmission delay need to be considered in actual applications. In the future, it is necessary to further explore actual deployment optimization strategies to promote the widespread application of this algorithm in power plant safety monitoring.

While the improved algorithm consistently performs better in terms of overall mAP, small object detection, and robustness under low light or occlusion, some comparative results show near parity with YOLOv5 in large object detection and moderately complex backgrounds. Therefore, the improvements are substantial but not uniformly superior across all metrics.

References

[1] Yu Y, Peng MJ, Wang H, Ma ZG, Cheng SY, Xu RY. A continuous learning monitoring strategy for multi-condition of nuclear power plant. Annals of Nuclear Energy. 2021;164:14. https://doi.org/10.1016/j.anucene.2021.108544

[2] Li YP, Hong F, Tian L, Chen JY, Du H, Liu JZ. Condition monitoring and early fault warning of power plant auxiliary equipment using LSTM-SDAE model with an adaptive threshold. Frontiers in Energy Research. 2023;10:14. https://doi.org/10.3389/fenrg.2022.1037539

[3] Yang ZX, Xu P, Zhang BA, Xu CL, Zhang LM, Xie HY, et al. Nuclear power plant sensor signal reconstruction based on deep learning methods. Annals of Nuclear Energy. 2022;167:12. https://doi.org/10.1016/j.anucene.2021.108765

[4] Ghita M, Siham B, Hicham M, Amine M. HT-TPP: A Hybrid Twin Architecture for Thermal Power Plant Collaborative Condition Monitoring. Energies. 2022;15(15):38. https://doi.org/10.3390/en15155383

[5] Fan ZP, Bai K, Zheng XK. Hybrid GA and Improved CNN Algorithm for Power Plant Transformer Condition Monitoring Model. IEEE Access. 2024;12:60255-63. https://doi.org/10.1109/access.2023.3316251

[6] Dai XY, Ming Y, Wang JP, Xu ZH, Wen HG. Methodology for updating GO-FLOW model to handle scenario changes in nuclear power plants. Frontiers in Energy Research. 2023;10:13. https://doi.org/10.3389/fenrg.2022.1034835

[7] Gao JY, Ma LY, Qing C, Zhao TD, Wang ZP, Geng J, et al. A health monitoring model for circulation water pumps in a nuclear power plant based on graph neural network observer. Sensors. 2024;24(14):24. https://doi.org/10.3390/s24144486

[8] Yu Y, Peng MJ, Wang H, Liu YK, Ma ZG, Cheng SY. Long-term operation monitoring strategy for nuclear power plants based on continuous learning. Annals of Nuclear Energy. 2022;178:13. https://doi.org/10.1016/j.anucene.2022.109323

[9] Lim M, Kim Y, Jin SX, Ha S, Chang SY, Kang HS, et al. Depth-based condition monitoring and contributing factor analysis for anomalies in combined cycle power plant. IEEE Access. 2024;12:73400-12. https://doi.org/10.1109/access.2024.3403864

[10] Gu HX, Liu GJ, Li JX, Xie HY, Wen HG. A framework based on deep learning for predicting multiple safety-critical parameter trends in nuclear power plants. Sustainability. 2023;15(7):15. https://doi.org/10.3390/su15076310

[11] Zhang JY, Xia H, Wang ZC, Zhu YH, Fu Y. Research on unsupervised condition monitoring method of pump-type machinery in nuclear power plant. Nuclear Engineering and Technology. 2024;56(6):2220-38. https://doi.org/10.1016/j.net.2024.01.031

[12] Wang WD, Yu JH, Xu T, Zhao CP, Zhou XZ. On-line abnormal detection of nuclear power plant sensors based on Kullback-Leibler divergence and ConvLSTM. Nuclear Engineering and Design. 2024;428:13. https://doi.org/10.1016/j.nucengdes.2024.113489

[13] Dong FY, Chen S, Demachi K, Yoshikawa M, Seki A, Takaya S. Attention-based time series analysis for data-driven anomaly detection in nuclear power plants. Nuclear Engineering and Design. 2023;404:15. https://doi.org/10.1016/j.nucengdes.2023.112161

[14] Xinyu D, Ming Y, Jipu W, Wei L, Zhihui X. Design of an intelligent operating procedures supervision system of nuclear power plant. Nuclear Technology. 2023;209(5):730-44. https://doi.org/10.1080/00295450.2022.2155023

[15] Gaurav K, Singh BK, Kumar V. Intelligent fault monitoring and reliability analysis in safety-critical systems of nuclear power plants using SIAO-CNN-ORNN. Multimedia Tools and Applications. 2024:25. https://doi.org/10.1007/s11042-023-17707-6

[16] Wei MY, Wang JD, Gao S, Li J, Pang XK. An operating zone model for safety and efficiency monitoring of power generation units in thermal power plants. Control Engineering Practice. 2024;153:10. https://doi.org/10.1016/j.conengprac.2024.106101

[17] Li DY, Zhou FH, Gao YT, Yang K, Gao HM. Power plant turbine power trend prediction based on continuous prediction and online oil monitoring data of deep learning. Tribology International. 2024;191:12. https://doi.org/10.1016/j.triboint.2023.109083

[18] Wu GH, Yuan DP, Yin JY, Xiao YQ, Ji DX. A Framework for Monitoring and Fault Diagnosis in Nuclear Power Plants Based on Signed Directed Graph Methods. Frontiers in Energy Research. 2021;9:11. https://doi.org/10.3389/fenrg.2021.641545

[19] Chaudhary A, Han J, Kim S, Kim A, Choi S.

Anomaly detection and analysis in nuclear power plants. Electronics. 2024;13(22):19. https://doi.org/10.3390/electronics13224428

[20] Lee JC, Choi YR, Yeo D, Moon S. Low-power wireless sensor module for machine learning-based continuous monitoring of nuclear power plants. Sensors. 2024;24(13):16. https://doi.org/10.3390/s24134209

[21] Amin MT, Yao YT, Yu J, Adumene S. Probabilistic monitoring of nuclear plants using R-vine copula. Annals of Nuclear Energy. 2023;190:11. https://doi.org/10.1016/j.anucene.2023.109867