Automatic Detection Algorithm of Typical Hidden Dangers in Substation Based on Dynamic Convolutional Feature Extraction

Wenbin Wu[1], Manshu Liang[1,*], Xiaojie Wu[2], Weihao Zhang[1] and Ying Ruan[1]

[1] Electric Power Research Institute of State Grid Fujian Electric Power Co., Ltd., Fuzhou, China

[2] State Grid Fujian Electric Power Co., Ltd., Fuzhou, China

INFORMATION

Keywords:
Multi-type hidden danger detection in substations
dynamic snake convolution
self-integrated attention
bounding box regression loss
small object defect detection
irregular defect classification

DOI: 10.23967/j.rimni.2026.10.74630

SCIPEDIA Revista Internacional de Métodos Numéricos para Cálculo y Diseño en Ingeniería

Revista Internacional Métodos numéricos para cálculo y diseño en ingeniería RIMNI

In cooperation with CIMNE

Automatic Detection Algorithm of Typical Hidden Dangers in Substation Based on Dynamic Convolutional Feature Extraction

Wenbin Wu[1], **Manshu Liang**[1,*], **Xiaojie Wu**[2], **Weihao Zhang**[1] **and Ying Ruan**[1]

[1]Electric Power Research Institute of State Grid Fujian Electric Power Co., Ltd., Fuzhou, China

[2]State Grid Fujian Electric Power Co., Ltd., Fuzhou, China

ABSTRACT

With the advancement of smart grid construction, substation equipment, being perpetually exposed to complex environments, is prone to latent hazards such as meter malfunctions and insulator cracks. Additionally, personnel violations may trigger safety incidents, directly jeopardising grid reliability. However, traditional manual inspections suffer from low efficiency and high rates of missed defects, while existing detection methods struggle to accommodate irregular defect classifications and multi-type defect characteristics, failing to meet engineering demands for real-time response and precise identification. To address these challenges, this study proposes a substation hazard detection framework (YOLOv10_DSE) based on an enhanced YOLOv10 (You Only Look Once version 10) architecture, designed to tackle multi-type hazard detection within complex substation scenarios. Firstly, a dynamic feature extraction module (C2fDSC) was designed, employing dynamic snake convolutions to enhance adaptive sampling capabilities for small targets and irregular defects. Secondly, a self-integrated attention module head (SEAMHead) was introduced to decouple localisation and classification tasks, thereby improving multi-type hazard discrimination accuracy. Finally, a bounding box regression loss function (inner_CIoU) was adopted to optimise small target localisation and irregular shape fitting. Experiments demonstrate that on a substation dataset containing 17 defect types, this method achieves mAP@0.5 and mAP@0.95 of 73.3% and 48.2%, respectively, representing improvements of 2.6% and 1.6% over the YOLOv10 baseline. This provides an efficient and reliable solution for multi-type defect detection in substations, holding significant engineering value for ensuring the secure operation of power grids.

OPEN ACCESS

Received: 15/10/2025

Accepted: 17/12/2025

DOI
10.23967/j.rimni.2026.10.74630

Keywords:
Multi-type hidden danger detection in substations
dynamic snake convolution
self-integrated attention
bounding box regression loss
small object defect detection
irregular defect classification

1 Introduction

Substations serve as pivotal nodes for the transmission and distribution of electrical energy, representing a quintessential embodiment of Cyber-Physical Systems (CPS) within the power sector. They

*Correspondence: Manshu Liang (msliang202510@163.com).

encompass not only physical infrastructure such as transformers, switchgear, and metering equipment, but also deeply integrate digital core elements including intelligent sensing, edge computing, and secure communications. The stable operation of physical equipment and the reliable coordination of digital systems collectively form a crucial safeguard for the reliability of the power system [1]. However, substation equipment remains perpetually exposed to complex environments characterised by high temperatures, humidity, and dust. The widespread adoption of semiconductor devices has further diminished power quality in terms of voltage and current. Over extended operational periods, this leads to equipment failures and systemic instability [2]. Such as abnormal meter readings, damaged equipment casings, and broken insulators. Concurrently, non-compliant practices like improper operator procedures or smoking may trigger safety incidents. Traditional manual inspections, reliant on operational staff's experience, suffer from time-consuming and labour-intensive processes [3], high defect detection rates in substations, and inadequate capability to identify concealed hazards like early silicone discolouration or minute cracks. These limitations hinder adaptation to the demands of grid intelligent transformation [4,5].

In recent years, intelligent hazard detection technologies based on computer vision have advanced rapidly. Through deep integration with digital elements within substations, these systems automatically analyse and identify defects via images and video, offering novel solutions for hazard detection and recognition [6,7]. Computer vision technology can complete the automatic processing and analysis of large-scale substation image data in an objective and consistent way, thus improving the accuracy of defect detection. Compared with manual inspection, computer vision technology can significantly improve the efficiency and speed of hidden danger detection. With the help of the automated identification process, areas with hidden dangers can be quickly located and potential hidden dangers can be discovered in time, thereby reducing human resource investment and time costs. It is an important technical means to improve the intelligent and intensive level of power grid operation and maintenance and to promote equipment status operation and maintenance to move towards comprehensiveness, accuracy, and efficiency. Therefore, it is of great practical significance to use computer vision intelligent defect detection technology to carry out hidden danger detection in substations [8].

Computer vision-based intelligent defect detection technology can learn and recognise various types of substation defects, including features such as shape, size, and texture, making it an ideal solution for identifying potential hazards in substations. Within existing smart grid research, Meng [9] proposed an enhanced SD-YOLOv7 model for anomaly detection. This integrates SENet's attention mechanism within the core network backbone while incorporating deformable convolutions via DCNv4 to replace traditional convolutions. This approach significantly improves object detection capabilities across varying angles in factory surveillance scenarios. Yuan et al. [10] developed an unsafe behaviour recognition system for power grid infrastructure sites based on the YOLOv3 algorithm. By optimising the loss function, the system more effectively detects construction personnel violations. Baig et al. [11] proposed a multimodal deep learning framework that effectively integrates the temporal and spectral domains by combining one-dimensional time-series data with corresponding two-dimensional ratio-map images. The proposed model employs parallel one-dimensional and two-dimensional convolutional neural networks (CNNs), each enhanced through an attention mechanism that strengthens feature extraction by focusing on salient information within specific modalities.

However, these studies have primarily focused on single defect types or single technical dimensions, making it challenging to account for multi-category variations. Within the context of substation hazard identification, diverse industrial equipment and complex environments are present, while hazard types encompass physical damage to equipment, abnormal operational states, environmental interference,

and human behaviour. Consequently, more robust object detection models are required to accurately identify defects in such intricate settings. Although existing algorithms have demonstrated significant performance improvements, they still struggle with detecting small and irregularly shaped defects. The reliance on rectangular bounding box overlaps for calculation fails to accurately measure the actual coverage quality of defects, leading to missed detections and misclassifications. These issues directly prevent current methods from meeting the core requirements of substation operations and maintenance: multi-type coverage and precise recognition of small targets.

To address the aforementioned issues, this study proposes an enhanced substation hazard detection framework based on YOLOv10 (You Only Look Once version 10) [12], termed YOLOv10_DSE. This detection framework focuses on multi-type hazard detection scenarios within substations, specifically covering 17 categories of typical hazards, and tackles the challenges of detecting small targets and irregular defects. Its principal contributions are as follows: (1) To enhance detection capabilities for small targets and irregular defects, we propose the Dynamic Feature Extraction Module (C2fDSC). This module employs adaptive sampling via Dynamic Snake Convolution (DySnakeConv) and multi-scale feature fusion within the C2f architecture to strengthen feature representation for small targets and irregular defects;

(2) To address the diverse detection requirements for potential hazards in substation operations and maintenance, we have designed the Self-Integrated Attention Module Head (SEAMHead). This incorporates a Self-Ensembling Attention Mechanism to decouple localisation (regression) from classification tasks, thereby enhancing the accuracy of multi-type hazard classification;

(3) To mitigate false negatives and false positives in multi-class detection, we employed a bounding box regression loss function (inner_CIoU). This optimises small object localisation and irregular defect shape fitting through core region overlap constraints and scale-sensitive aspect ratio penalties. This approach avoids the inaccuracies inherent in rectangular box overlap calculations, which fail to accurately measure the actual coverage quality of defects, thereby preventing missed detections and misclassifications;

This paper is organised as follows. Section 2 describes the related work of this research, which outlines the status of the development of the substation hidden defect detection technology based on manual feature-based machine learning method, deep learning-based substation defect detection method. Section 3 details the proposed YOLOv10_DSE model, encompassing the substation hazard detection framework (YOLOv10_DSE), the dynamic feature extraction module (C2fDSC), the self-integrated attention module head (SEAMHead), and the bounding box regression loss function (inner_CIoU). Section 4 shows the experimental results and analyses, covering the dataset, experimental environment and parameter settings, evaluation metrics, experimental results and their analyses, as well as comparisons with other models and ablation experiments on the YOLOv10_DSE model. Section 5 summarises the whole paper and suggests future research directions.

2 Related Work

Substation hidden danger detection technology can be divided into two stages: traditional manual detection and intelligent visual detection [13,14]. Early detection was mainly manual, relying on the experience of operation and maintenance personnel to observe meter readings and check the status of enclosure integrity equipment, but the efficiency is low and susceptible to subjective factors. With the development of computer vision technology, intelligent detection has gradually become mainstream, the core of which is the automatic identification of defects through image and video

analysis. Substation hidden defect detection can be divided into two categories of equipment defect detection and personnel behaviour detection according to the different detection objects.

2.1 Substation Equipment Condition and Defect Detection

Substation equipment status and defect detection focuses on abnormal state identification of substation meters, insulators, and switchgear hardware equipment. The main methods include two technologies based on traditional machine learning and deep learning: Among them, power equipment defect detection based on traditional machine learning is to extract equipment appearance information through HOG, SIFT design features, and combine Naive Bayes [15], decision tree [16], support vector machine [17,18] and other classifiers to achieve defect identification. Sampedro et al. [19] used the Histogram of Oriented Gradients (HOG) method to extract image information, which increased the detection rate of power towers to 92%, and the classification accuracy of different types of power towers to 93% and 87%, respectively. However, traditional machine learning methods such as HOG, SVM, KNN, and random forest face problems such as weak feature expression ability, high data annotation cost, and insufficient real-time performance in equipment defect detection, making it difficult to adapt to changing environments. In recent years, with advancements in the field of deep learning, deep learning-based algorithms for detecting defects in electrical equipment have optimised the shortcomings of traditional methods. Significant progress has been made in areas such as hybrid architecture design and multimodal fusion. Among these, Cheng and Liu proposed AdIn-DETR [20], which incorporates two adapter modules (GSG-Adapter and LFO-Adapter) to enhance performance, achieving better accuracy and faster inference than the YOLO model. Xie et al. introduced Power-DETR [21], a Transformer-based model combining Swin-large pre-training, contrastive denoising, and hybrid label assignment strategies. It is specifically designed to improve detection of small power line defects. Although Transformer-based architectures achieve high accuracy in detecting large objects, they remain inadequate for detecting small, overlapping insulator defects, requiring a trade-off between timeliness and precision. Consequently, He et al. proposed the hybrid YOLOv8s-SwinT architecture [22], which enhances detection of minute defects by concatenating features extracted via Swin Transformers and CNNs. However, this approach may be constrained when detecting defects such as cracks and contamination. Wu proposed a multimodal insulator defect detection network based on an enhanced YOLOv8 model [23]. By integrating an RGB-D data feature fusion module, it effectively combines RGB images with depth information, enhancing the identification capability of insulator defects. However, this approach necessitates extensive manual annotation for each defect instance and consumes computational resources for generating the depth information.

2.2 Substation Personnel Code of Conduct Testing

Substation personnel behavior code detection focuses on staff operational compliance and violations. The main technologies include personnel and safety protection detection based on traditional algorithms and personnel behavior code detection based on deep learning: among them, personnel and safety protection detection based on traditional algorithms, early safety protection detection mostly focuses on whether the helmet is worn, the color and shape of the helmet, the resolution of the camera, and the change of camera shooting angle and distance are the difficulties of detection [24]. Aiming at this problem, many researchers have conducted in-depth research. Rubaiyat et al. [25] proposed a two-step helmet detection model, which first combines the frequency domain information of the image with HOG features to detect the position of the operator, then fuses the preset color information of the helmet with the circular Hough transform features, and then detects the helmet. Mu et al. [26] used the improved way of local binary pattern LBP (local binary pattern) rotation invariant pattern and circular

LBP operator to make the LBP operator more effectively applied to personnel detection. Dalal and Triggs [27] proposed an operator helmet detection algorithm based on HOG feature. The algorithm proposed that because HOG feature is insensitive to small offset and illumination change, HOG feature is used to describe the local gradient value and direction feature of the image, so as to effectively describe the edge feature of the human body. However, personnel and security protection detection based on traditional algorithms has some disadvantages, such as requiring manual feature extraction, large amount of calculation, low recognition rate, and cumbersome steps. With the rapid development of deep learning, early traditional algorithms have gradually been replaced by algorithms based on deep learning. Wei et al. [28] applied Retina-Net to anomaly detection in personnel wearing insulated gloves. This algorithm effectively mitigates the impact of imbalanced positive and negative samples in the dataset on the training process whilst minimising computational resource usage. However, it fails to account for the influence of varying personnel postures on detection outcomes. Addressing the existing algorithm's inability to accurately recognise actions such as bending over or squatting, He et al. [29] proposed an illegal dress recognition algorithm that introduces CBAM attention mechanism and target key point detection. By adding attention mechanism and adopting human body area positioning strategy, it can adapt to the illegal dress recognition of upright, squatting and other postures, and effectively improve the accuracy of the algorithm. However, in crowded environments, false detections caused by overlapping target boxes persist.

3 Methodology

To address shortcomings in small object recognition, irregular defect classification, and multi-type defect differentiation, this paper designs a substation hazard detection framework (YOLOv10_DSE). First, we conduct an in-depth study of the core design principles and fundamental architecture of the YOLOv10_DSE framework. As shown in Fig. 1, the YOLOv10_DSE model incorporates key components including a Dynamic Feature Extraction Module (C2fDSC), a Self-Integrated Attention Module Head (SEAMHead), and a bounding box regression loss (inner_CIoU). C2fDSC employs dynamic snake convolutions to extract multi-scale dynamic features from input data, generating dynamic feature maps that capture both small-object details and large-object semantics. Subsequently, the SEAMHead employs a dual-branch regression head to independently optimise features processed through the fusion layer, achieving synergistic defect localisation and classification. Finally, guided by ground truth (GT) and constrained by the inner_CIoU bounding box regression loss function to refine localisation accuracy and shape fitting, the model undergoes parameter optimisation through classification loss and regularisation loss. This tripartite collaboration enhances overall detection capability from feature extraction to detection optimisation. Key components such as C2fDSC, SEAMHead, and inner_CIoU enable the model to adapt more effectively to defect shapes like crack orientations and nest-like edges. They focus on critical regions within defect categories while suppressing overlapping areas, allowing YOLOv10_DSE to achieve efficient predictions. Detailed descriptions of C2fDSC, SEAMHead, and inner_CIoU will be provided subsequently.

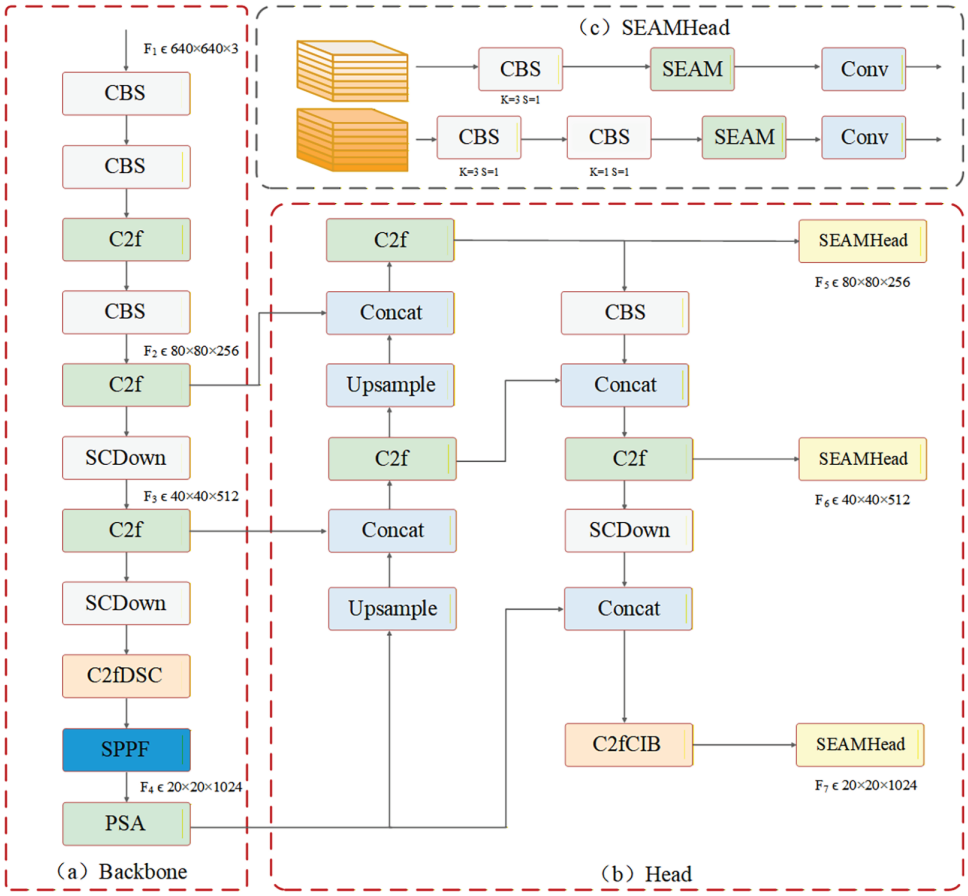

Figure 1: All the overall architecture of the proposed YOLOv10_DSE. (**a**) Backbone feature extraction network consisting of C2f, C2fDSC; (**b**) Head network consisting of key modules such as, SEAMHead, inner_CIoU; (**c**) A detection head employing self-ensemble attention mechanisms (SEAMHead) is applied to both regression heads and classification heads

3.1 Dynamic Feature Extraction Module (C2fDSC)

Aiming at the problem that the fixed sampling mode in traditional convolution is difficult to capture small target details and irregular shapes in substations, this paper proposes C2fDSC module. The C2fDSC module is the core module in the YOLOv10_DSE backbone network. Through the structure of dynamic serpentine convolution (DySnakeConv) [30] and multi-scale feature fusion (C2f), it can retain the global and local information of defects at different scales, providing high-quality dynamic features. Its core structure is shown in Fig. 2a:

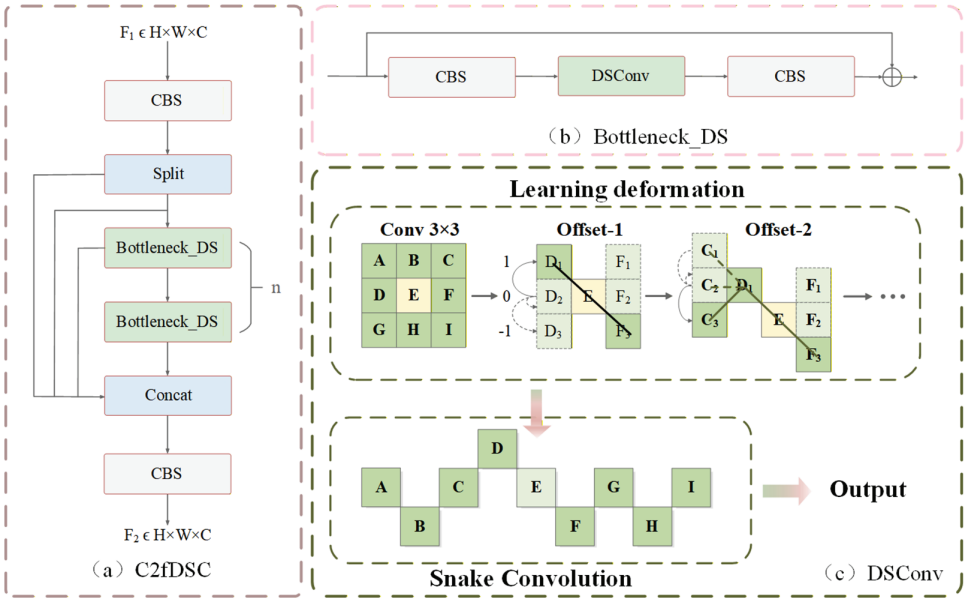

Figure 2: The overall architecture of our proposed C2fDSC. (**a**) C2fDSC module consisting of Bottleneck_DS; (**b**) Dynamic Serpentine Convolution Bottleneck Layer Incorporating DSConv (Bottleneck_DS); (**c**) Schematic diagram of the sampling process for Dynamic Serpentine Convolution (DSConv), featuring offset-learned serpentine convolutions

3.1.1 Dynamic Serpentine Convolution

The core of dynamic snake convolution lies in adjusting the sampling position through learnable offsets, enabling it to adaptively reshape along the target's contour. This enhances its ability to capture local details of small objects like bird nests and irregular defects such as insulator cracks, addressing the challenge that traditional fixed-sampling convolutions struggle to adapt to non-regular defect features. Specifically, DSConv constructs a sampling network based on the centre sampling point of the convolutional kernel, employing a dual mechanism of symmetric expansion and dynamic offset. Taking the centre coordinates $K_i = (x_i, y_i)$ as the origin, it symmetrically expands horizontally and vertically to form a 9-point sampling grid. This grid comprises one central sampling point and eight peripheral sampling points, corresponding to the foundational structure of a 3 × 3 convolutional kernel with a dilation rate of 1. Its mathematical expression is represented as shown in Eq. (1):

$$K = \{(x-1, y-1), (x-1, y), \cdots, (x+1, y+1)\} \tag{1}$$

The position of the relative centre of each grid within K may be expressed as $K_i \pm c = (x_i \pm c, y_i \pm c)$, where the parameter c = [1, 2, 3, 4] denotes the horizontal and vertical distance of the sampling point from the centre.

The grid design with a size of 9 points enables DSConv to capture defect features at different scales simultaneously, and for the grid points near the center (c = 0, 1), it is expanded by small steps (1 pixel), retaining the edge details of small targets. For the edge grid points (c = 3, 4), the global semantic features are extracted by large step expansion (3–4 pixels), covering the overall outline of the large target. Secondly, the dynamic offset is generated based on the center position, Adaptively adjust the shape along the target's contour through cumulative offset from the centre to the periphery.

Starting from the center position K_i, the coordinates of each position $K_i + c$ ($c = 1 \sim 4$) expanding to the left and right are generated by adding a learnable offset Δc to the coordinates of the previous position $K_i + c - 1$, as shown in Eq. (2):

$$K_{i+c} = K_{i+c-1} + (\Delta_{c,x}, \Delta_{c,y}) \tag{2}$$

where, $\Delta_c = (\Delta_{c,x}, \Delta_{c,y})$ denotes the offset value for step c, with $\Delta_{c,x}$ representing the horizontal offset and $\Delta_{c,y}$ representing the vertical offset; Both values are constrained within the range $[-1, 1]$, physically signifying that each offset may only adjust by one pixel horizontally or vertically, thereby preventing excessive displacement that could cause sampling to deviate from the target region. K_i denotes the central sampling point, while K_{i+c-1} represents the coordinate of the previous sampling point. By dynamically sampling through the accumulation of Δ_c, the convolution can adaptively adjust its shape along the target's contour.

During training, we employ backpropagation to adaptively learn the direction of the Δ_c offset, thereby accommodating defect shapes, crack orientations, and the irregular edge profiles of bird's nests. Naturally, to prevent excessive offset accumulation from obscuring critical features, DSConv adopts an iterative strategy for offsets: the total leftward expansion is approximated to be equal to the total rightward expansion. Here we cite the constraint for the horizontal x-axis direction, as per Eq. (3). This constraint ensures the sampling grid consistently centres around the target core region, preventing the omission of critical features (such as crack endpoints or the main body of bird's nests) due to unidirectional offsets, thereby maintaining spatial continuity of attention.

$$K_{i\pm c} = \left\{ \left| \sum_{c=1}^{4} \Delta_{c,x}^{left} - \sum_{c=1}^{4} \Delta_{c,x}^{right} \right| \leq 1 \right. \tag{3}$$

The dynamic sampling grid generated based on the offset field adds the original sampling position to the offset Δ to generate the final sampling position. Through the accumulation and constraint of dynamic offset, the output feature map of dynamic serpentine convolution is generated. Because of dynamic sampling operation, the feature map retains richer local details, and realizes the adaptive extraction of substation defect features. Its module structure design is shown in Fig. 2c.

3.1.2 Multi-Scale Feature Fusion Structure

Within the multi-scale feature fusion module, feature maps are split and multi-branch features are fused, thereby reducing computational load while preserving rich scale information. Specifically, the feature map output from the Dynamic Convolutional Neural Network (DSConv) is split into two components: features from the primary branch are directly passed to the next layer to preserve original characteristics, while features from the secondary branch undergo multi-scale feature extraction through multiple parallel Dynamic Convolutional Neural Network Bottleneck layers (Bottleneck_DS). The structural design of Bottleneck_DS is illustrated in Fig. 2b, in which the dimensionality is reduced by 1×1 kernel size convolution, 3×3 kernel size dynamic serpentine convolution enhances local information, and 1×1 kernel size convolution increases dimension. Finally, it is spliced with the main branch features according to the channel to obtain the final output map. The mosaic multi-scale dynamic feature map retains the edge of small objects in the low channel number part, and the overall outline and global semantics of large objects in the high channel part.

3.2 Self-Integrated Attention Module Head (SEAMHead)

In substation hazard detection scenarios, conventional detection heads typically couple bounding box regression with category classification tasks within the same feature set. This arrangement risks classification heads monopolising detail features essential for localisation, or regression heads compromising localisation accuracy due to classification noise, ultimately causing mutual interference between the two tasks. Furthermore, the small-object characteristics of substation defects—such as bird nests occupying minimal image area, irregular shapes, and complex backgrounds like metallic textures on equipment casings or interlaced wire patterns—exacerbate the performance bottleneck of detection heads. Conventional approaches struggle to simultaneously achieve precise boundary localisation of small targets and classification of irregular defects. To address these challenges, this paper proposes the Self-Integrated Attention Module Head (SEAMHead) for high-precision defect localisation and classification. The SEAMHead architecture is illustrated in Fig. 1c.

Among them, SEAMHead realizes independent optimization of positioning and classification tasks through the collaborative design of dual-branch task decoupling and Self-integrated attention mechanism [31], and enhances key regional features through self-integrated attention. Specifically, the SEAMHead module employs a channel-dimensional quantitative splitting strategy for input features: given three detection scales with channel counts of 128, 256, and 512, respectively, for the input multi-scale features, the regression head selects a low-channel-count subset to preserve edge details of small objects, while the classification head chooses a high-channel-count subset to retain global semantic information of large objects. The edge details of small objects are primarily determined by shallow features from low-channel dimensions, while the semantic information of large objects is mainly determined by deep features from high-channel dimensions. Following this splitting, the regression head and classification head can extract the required features in a targeted manner, fundamentally avoiding task interference while adapting to the detection requirements of coexisting large and small objects in substation scenarios. To further enhance feature response in defect edge regions, both branches employ the SEAM attention module.

The core of the classification head lies in extracting discriminative features relevant to defect categories through a deep convolutional network. Composed of multiple convolutional layers, the input features first undergo a separable convolution with a kernel size of 3 to preserve the original channel dimensions, followed by a 1×1 convolution to uniformly map the channels. We incorporated a SEAM module after the second convolutional layer to enhance key features in the defect edge regions. Within the SEAM module, features are first processed through a 3×3 depthwise separable convolution layer. This extracts local spatial edge and texture features without increasing computational load. Subsequently, a nonlinear transformation is applied to the convolution output, enhancing the model's expressive power while normalising features. This accelerates training and prevents gradient vanishing. The input features are then summed with features processed via pointwise convolution, preserving original information and mitigating gradient vanishing in deep networks. The processing of the separable convolution residual block is illustrated in Eq. (4):

$$x_{\text{out}} = x_{\text{in}} + \text{GELU}\left(\text{BN}\left(\text{PWConv1} \times 1\left(\text{GELU}(\text{DWConv3} \times 3(x_{\text{in}}))\right)\right)\right) \tag{4}$$

Features processed through deep separable convolutional residual blocks first undergo adaptive average pooling with a 1×1 kernel size to extract global contextual information. This eliminates the influence of spatial dimensions, enabling focus on global dependencies between channels. Subsequently, a single-hidden-layer fully connected network generates channel weights $y_{channel}$, where the number of hidden layer neurons is set to one-quarter of the input channel count. The output layer neuron count matches the input channel count. The output is then mapped to the interval [0, 1] via

the Sigmoid function, yielding importance weights for each channel. Finally, the channel weights $y_{channel}$ are element-wise multiplied with the original input features x, thereby enhancing defect-edge channels while suppressing background texture channels. This yields the enhanced features xout, as demonstrated in Eqs. (5) and (6):

$$y_{\text{channel}} = \sigma\left(W_2 \cdot \text{ReLU}(W_1 \cdot x_{\text{pool}})\right) \tag{5}$$

$$x_{\text{out}} = X \cdot y_{\text{channel}} \tag{6}$$

The SEAM module dynamically adjusts the importance of each channel through the channel weight, enhances the channel corresponding to the defect edge, suppresses the channel corresponding to the background texture, and realizes channel attention; At the same time, since the channel weight is the result of global pooling, it implies the context information of the spatial dimension and indirectly realizes self-integrated attention. It can help the detection head better suppress the response of background areas such as flat areas of the equipment housing, and enhance the characteristics of defect edges. The SEAM structure diagram is shown in Fig. 3.

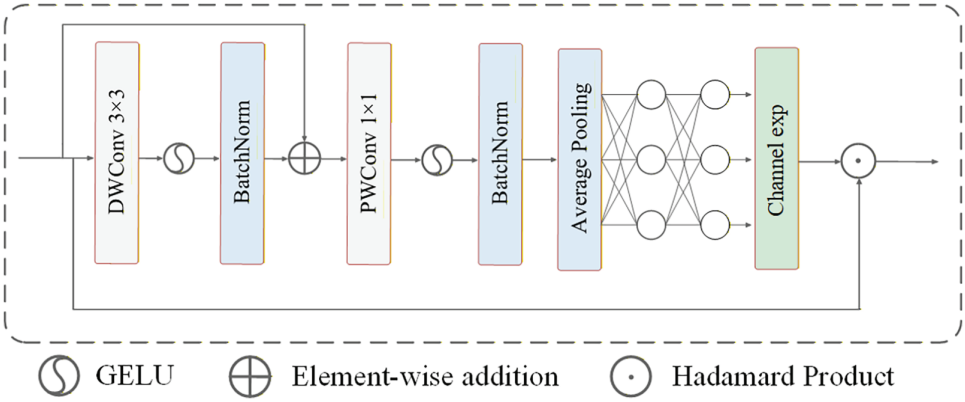

Figure 3: SEAM Structure Diagram

The core of the regression head lies in extracting bounding box regression features through a lightweight convolutional network, while employing SEAM self-ensemble attention to focus on critical regions. To concentrate on critical regions pertinent to defect categories—such as blurred areas in meter readings—we incorporate the SEAM module after a convolutional layer with a kernel size of 3. Employing the same methodology as in the classification head, this generates an attention map for the current feature map and enhances its features. This approach suppresses background interference while amplifying features in category-critical regions. Finally, a convolutional layer with a kernel size of 1 outputs regression features for both boundary coordinates and confidence scores.

3.3 Bounding Box Regression Loss (inner_CIoU)

In the substation hidden danger detection scenario, the traditional bounding box IoU and CIoU regression loss functions have core problems such as fuzzy location of small targets, poor shape fitting of irregular defects, and multi-scale weight mismatch. For example, because the size of a small target 5×5 pixel bird's nest is too small, the traditional loss function only focuses on the overall overlapping area, resulting in false height overlap but core missed detection; Because the shape of irregular defects is non-rectangular, the traditional loss function cannot accurately fit its curved boundary; Multi-scale defects such as device shell damage as small as 5×5 pixels and as large as 500×500 pixels have a large

size span, and the fixed weight of traditional loss function cannot balance the positioning accuracy requirements at different scales. For the above problem, We employ boundary regression loss [32], whose definition is shown by Eqs. (7)–(9).

$$IoU^{\text{inner}} = \frac{inter^{\text{inner}}}{union^{\text{inner}}} \tag{7}$$

$$L_{CIoU} = 1 - IoU + \alpha v \tag{8}$$

$$L_{\text{inner\_CIoU}} = L_{\text{CIoU}} + IoU - IoU^{\text{inner}} \tag{9}$$

Among these, the Intersection over Union (IoU) represents the overall overlap between predicted bounding boxes and ground truth (GT) bounding boxes, serving as a core metric for assessing global coverage quality. The $\text{inter}^{\text{inner}}$ refers to the overlapping region between the auxiliary interior bounding box of the predicted bounding box and that of the actual bounding box, focusing on the matching degree within the core defect area. $\text{union}^{\text{inner}}$ refers to the area of the union between the auxiliary interior bounding box of the predicted bounding box and the actual bounding box's auxiliary interior bounding box, used to normalise the interior intersection. $\text{IoU}^{\text{inner}}$ denotes the intersection-over-union ratio for inner regions, measuring the overlapping area between the core regions of two bounding boxes. It quantifies the overlap between the predicted box and the ground truth box's core regions, addressing the limitation of traditional IoU in insufficiently accounting for the core regions of small targets. L_{CIoU} is the standard loss function, wherein α denotes the aspect ratio penalty weight and v the aspect ratio difference term, serving to balance positioning accuracy with shape fitting. $L_{\text{inner\_CIoU}}$ is the boundary regression loss function proposed herein, which enhances core region constraints by incorporating $\text{IoU}^{\text{inner}}$ while preserving L_{CIoU} shape fitting advantages.

As shown in Fig. 4, the Inner Target Box represents the true auxiliary inner bounding box. This constitutes the core region of the true bounding box scaled proportionally, serving to compute the $\text{inter}^{\text{inner}}$ denoted by the deep green area in the figure and $\text{union}^{\text{inner}}$ metrics. These metrics focus on matching the core defect area. The Target Box denotes the true bounding box, corresponding to the actual defect region within the core substation scenario, such as the annotated defect box. The Inner Anchor Box denotes the predicted auxiliary inner box, representing the scaled core region of the predicted box. It collaborates with the Inner Target Box to compute $\text{IoU}^{\text{inner}}$. The Anchor Box signifies the bounding box predicted by the model, serving as the preliminary localisation result for the defect area. (x_c^{gt}, y_c^{gt}) and (x_c, y_c) denote the centre coordinates of the ground truth box and the predicted box, respectively. w^{gt} and h^{gt} denote the width and height of the ground truth box and predicted box, respectively. w_{inner}^{gt} and h_{inner}^{gt} denote the width and height of the auxiliary inner box for the ground truth box and predicted box, respectively, while w_{inner} and h_{inner} denote the width and height of the auxiliary inner box for the ground truth box and predicted box, respectively.

The figure illustrates a dual mechanism comprising overall box matching and core box constraints. Conventional IoU focuses solely on the overall overlap between the Target Box and Anchor Box, which can lead to core detection failures due to artificially inflated overall overlap rates for small targets. Therefore, Inner Target Boxes and Inner Anchor Boxes are introduced to additionally consider overlap within defect and interest regions, as expressed in Eq. (7) ments of irregular defects.

inner_CIoU constitutes an enhancement to the existing IoU loss metric, addressing its limitations in weak generalisation and slow convergence across diverse detection tasks. It proposes employing auxiliary bounding boxes to compute the loss, thereby accelerating the bounding box regression process. Specifically, inner_CIoU introduces a scaling factor to regulate the size of auxiliary bounding

boxes and utilises auxiliary bounding boxes of varying scales tailored to different datasets and detectors, thereby overcoming the constraints of weak generalisation inherent in existing approaches.

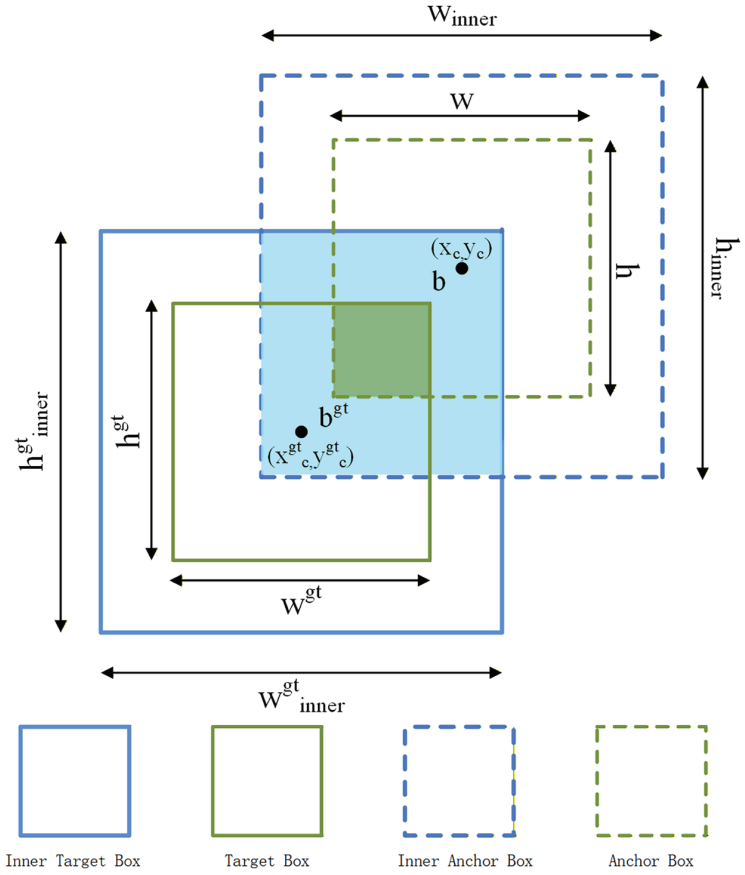

Figure 4: Bounding box regression loss Diagram

4 Experimental Results and Analysis

This section begins by detailing the dataset employed in the experiments, alongside the meticulously designed experimental setup and evaluation criteria. Subsequently, it delves into the practical efficacy of the YOLOv10_DSE architecture for hazard detection tasks within the complex scenarios of substations. Through a series of ablation experiments, the efficacy of each component within the architecture is systematically validated. The performance of the YOLOv10_DSE architecture is compared with other existing methods in terms of metric analysis and model complexity assessment. Furthermore, to visually demonstrate the superior performance of the YOLOv10_DSE architecture, detailed experimental result charts and in-depth analyses are provided.

4.1 Experimental Dataset

In this paper, pictures of 17 typical substation hidden danger scenarios, such as blurred dials and abnormal meter readings, and damaged insulators, are searched, aiming at the image detection data set of hidden dangers of substation equipment and workers' behavior. This dataset contains 8307 annotated images and aims to identify and classify various hidden dangers in substations,

such as equipment damage, abnormal behavior, etc., through machine learning and computer vision techniques. There are 17 types of abnormal hidden dangers classified, such as blurred covering dial, damaged dial, damaged dial shell, broken insulator, oil leakage on the ground, damaged respirator silicone cartridge, abnormal door closure, hanging suspended matter, hanging bird's nest, damaged and missing cover, employees not wearing safety helmets, employees not wearing tooling, employees smoking, abnormal meter reading, abnormal oil level of respirator oil seal, respirator silicone discoloration, and unclosed pressure plate. We adopt an 8:1:1 ratio partition strategy for the dataset to form a training set, a validation set, and a test set. The final distribution of abnormal hidden danger samples is as follows: abnormal meter reading (789), damaged meter shell (523), Existence of the Bird's Nest (883), abnormal door closure (383), pressure plate closed (362), damaged or missing cover plate (654), Unusual suspended object (729), respirator silicone discoloration (1174), meter dial blurred (869), insulator cracked (410), meter dial damaged (723), ground oil stain (833). In order to display the prediction results better, the class labels are coded accordingly. The details of the data set are shown in Table 1 and Fig. 5 provides examples of each hidden danger.

Table 1: Detailed of substation hidden danger dataset

Hidden danger	Label	Standard	Quantities
Meter reading abnormal	bjdsyc	The dial display is blurred, numbers are missing, or the scale is blurred, which makes it impossible to read the data accurately.	789
Damage of meter	bj_wkps	Cracks, breakage or deformation in the meter.	523
Existence of the Bird's Nest	yw_nc	There are foreign objects such as bird nesting and hanging objects around or inside the equipment.	883
Box door closure abnormal	xmbhyc	The equipment box door is not completely closed or the lock fails, which poses a potential safety hazard.	383
Pressure plate state close	kgg_ybh	The opening and closing indication of the switchgear is wrong, the operating mechanism is stuck, etc.	362
Broken or missing cover plate	gbps	The protective cover of the equipment is missing or broken, which may lead to electric shock or foreign object intrusion.	654

(Continued)

Table 1 (continued)

Hidden danger	Label	Standard	Quantities
Unusual suspended object	yw_gkxfw	Hang plastic bags, kite strings and other floating objects on lines or equipment.	729
Respirator silicone discoloration	hxq_gjbs	The silicone inside the respirator changes color (usually from blue to pink) due to moisture.	1174
Watch dial blurred	bj_bpmh	The dial glass is stained, scratched or aged, causing difficulty in reading.	869
Insulator rupture	jyz_pl	Cracks and breakage appear on the surface or inside of the insulator, resulting in degraded insulation performance.	410
Watch dial broken	bj_bpps	The dial is cracked and damaged.	723
Ground oil pollution	sly_dmyu	Oil leakage occurs from transformers and circuit breakers, and oil stains form on the ground.	833
Staff not wearing safety helmets	wcaqm	The workers are not wearing safety helmets and protective equipment.	567
Staff not wearing work clothes	wcgz	Failure of staff to wear uniforms	815
Respirator oil seal oil level abnormal	ywzt_yfyc	Oil level of oil-immersed equipment is too high or too low	331
Breakage of respirator silicone cartridge	hxq_gjtps	Respirator silicone cartridges ruptured or seal failure	106
Smoking	xy	Smoking by inspectors in non-designated areas	607

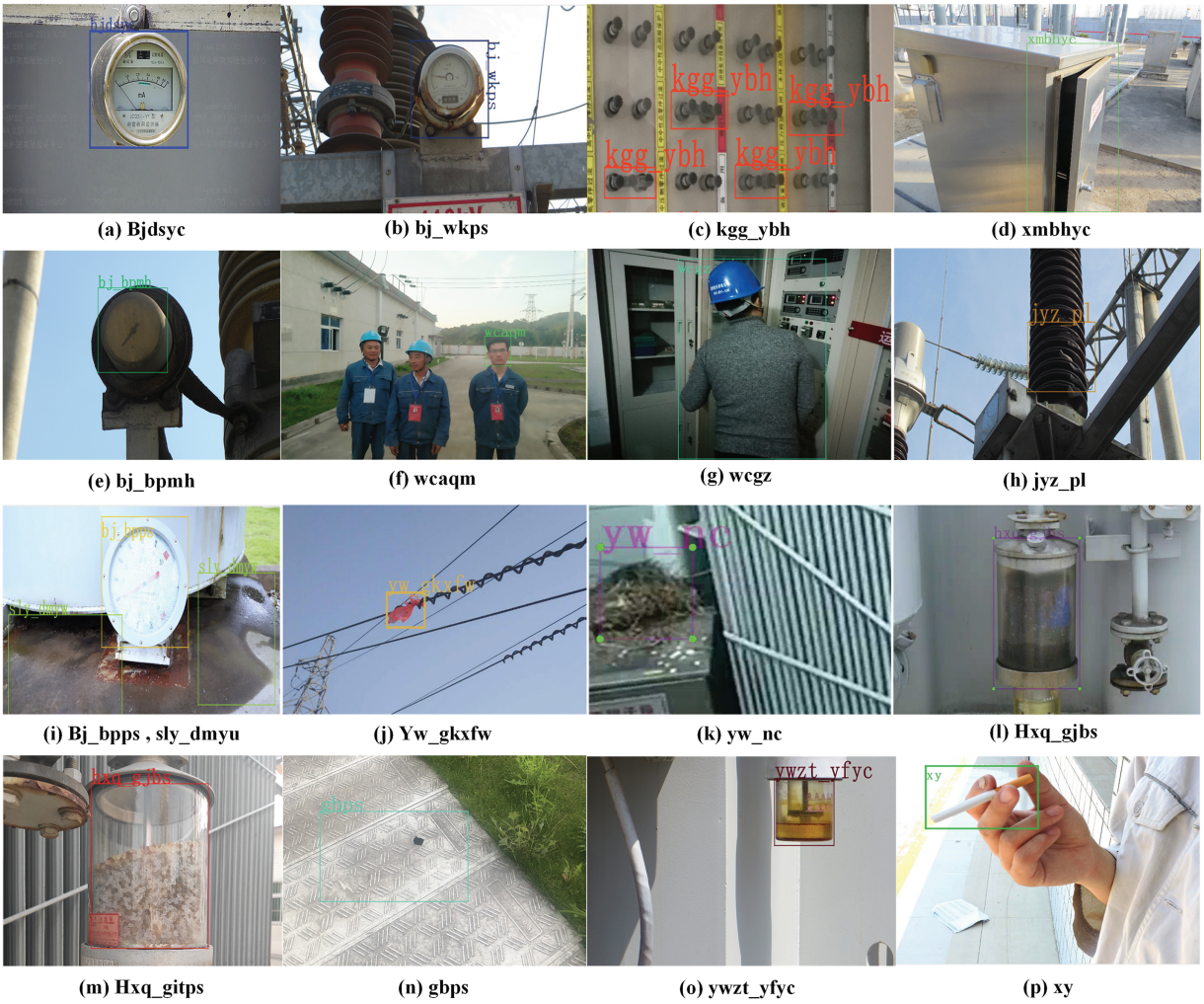

Figure 5: Example diagram of hidden dangers. (**a**) Meter reading is abnormal; (**b**) meter shell is damaged; (**c**) pressure plate is closed; (**d**) box door is closed abnormally; (**e**) meter dial is blurred; (**f**) staff does not wear safety helmet; (**g**) staff does not wear tooling; (**h**) insulator is broken; (**i**) meter dial is damaged and oil on the ground; (**j**) Unusual suspended object; (**k**) Existence of the Bird's Nest; (**l**) respirator silicone discoloration; (**m**) respirator silicone cartridge is damaged; (**n**) cover is damaged or missing; (**o**) respirator oil seal oil level is abnormal; (**p**) Smoking

4.2 Training Environment and Parameter Settings

To prevent variations in experimental environments from influencing results, all experiments in this paper were conducted under identical configuration settings. The experiments were performed on a 64-bit Windows 11 operating system, with hardware comprising an NVIDIA GeForce RTX 3090 (24 GB), a 14-core Intel i7-14600KF processor, and 64 GB of memory. As Python and CUDA versions do not affect experimental outcomes but must remain compatible with both software and hardware, the software environment utilised Python 3.11.11 for deployment. PyTorch 1.13.0 was selected as the deep learning framework, with CUDA 11.8 employed for training acceleration. During experimentation, image input dimensions were set to 640 × 640, batch size was set to 32, the initial

learning rate to 0.001, the optimiser to SGD, the random seed fixed at 0, and momentum set to 0.937. Data augmentation was performed using Albumentations 1.3.1 by default, applying colour space adjustments, set transformations, flipping operations, concatenation augmentation, and random erasure to the images. The process underwent a total of 200 iterations.

4.3 Evaluation Metrics

To evaluate the performance of the YOLOv10_DSE model, we selected average detection accuracy, Recall, number of parameters, GFLOPs, F1-score and inference time as assessment metrics.

Precision: precision measures the proportion of positive samples correctly identified in all positive predictions made by the model. This metric was chosen because it effectively assesses the model's ability to minimise false positives. The precision is calculated as Eq. (10):

$$P = \frac{TP}{TP + FP} \times 100\% \tag{10}$$

where, TP refers to True Positives (true positives, i.e., correctly detected targets) and denotes the number of targets correctly identified by the model correctly detected by the model. FP refers to False Positives and indicates the number of negative instances that the model incorrectly predicted as positive.

Recall (R): recall measures the proportion of true positive samples correctly identified by the model and reflects the sensitivity of the model. It is an important measure of the model's ability to detect all relevant targets and is key to a comprehensive assessment of the model's performance. The recall rate is calculated as Eq. (11):

$$R = \frac{TP}{TP + FN} \times 100\% \tag{11}$$

mAP: mAP measures the average detection accuracy of the model over multiple categories. This metric was chosen because it provides a comprehensive evaluation of the model's overall performance in a multi-category task. mAP@0.95 is a more comprehensive evaluation metric that calculates the mAP values for mAP@0.95 confidence thresholds. mAP@0.95 allows for a better evaluation of the model's robustness and performance. The formula is calculated as Eq. (12):

$$mAP = \frac{1}{n} \sum_{i=1}^{n} AP_i \tag{12}$$

AP_i is the AP of the i-th target class. n is the number of detected images.

F1-score: The F1-score is a statistical metric used to measure the precision of multi-task binary classification models. It simultaneously considers both the accuracy and recall of the classification model. The F1-score can be regarded as a weighted average of the model's accuracy and recall, ranging between 0 and 1, with higher values indicating a superior model. The formula is calculated as Eq. (13):

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{13}$$

Precision refers to accuracy, while Recall denotes recall rate.

4.4 Experimental Results and Analysis

To validate the detection efficacy of the YOLOv10_DSE model in practical applications, we conducted comparative experiments against existing real-time object detection algorithms. Under unified training configurations, model performance was systematically evaluated based on core metrics including mAP@0.5, mAP@0.95, inference time, parameter count, and GFLOPs. A comprehensive comparative analysis was conducted against various mainstream real-time object detection models, Mamba architecture models, and DETR-based detection models. These included YOLOv11n [33], YOLOv8 [34], YOLOv5s [35], YOLOv12 [36], YOLOv13 [37], RT-DETR [38], Faster-RCNN [39], and Mamba-YOLO [40], ensuring comprehensive and impartial comparisons. As shown in Table 2, the proposed YOLOv10_DSE model was compared with other real-time object detection algorithms on the substation hazard dataset.

Table 2: Comparison of the performance of various models on the substation defect dataset

Model	mAP@0.5 (%)	mAP@0.95 (%)	Param (M)	GFLOPs
YOLOv5s	66.4	42.8	2.5	7.1
YOLOv8s	70.3	47.1	3.0	8.1
YOLOv11n	71.0	47.8	2.55	6.3
YOLOv12n	70.9	47.9	2.51	**5.8**
YOLOv13n	71.2	47.4	2.45	6.2
RT-DETR	71.7	48.1	32.0	103.5
Faster-RCNN	60.3	39.6	41.03	83.4
Mamba-YOLO	71.2	47.4	2.45	6.2
YOLOv10_DSE	**73.3**	**48.2**	**2.43**	6.2

Note: Bold text indicates optimal performance.

Compared with the YOLO series models with similar accuracy, the average detection accuracy rates of mAP@0.5(%) and mAP@0.95(%) of YOLOv10_DSE reach 73.3% and 48.2%, respectively. Among them, mAP50 is 2.1% higher than the latest YOLOv13n model and 2.4% higher than the YOLOv12n model, reflecting better feature extraction and target positioning capabilities. It outperforms the latest YOLOv13n model by 0.8% and the RT-DETR model by 0.6%, demonstrating superior detection capabilities for small objects and dense scenes. In terms of parameter quantity, the parameter quantity of YOLOv10_DSE is the same as that of YOLOv13n and YOLOv12n, and the number of parameters of YOLOv10_DSE is 23% fewer than that of YOLOv8s, but it achieves higher precision with superior parameter efficiency, utilising significantly fewer parameters than both Transformer-based and hybrid architecture models such as Mamba-YOLO. In terms of computational complexity, YOLOv10_DSE is 6.2, which is the same as YOLOv13n, but with higher accuracy. Although YOLOv12n accelerates calculations with its unique Flash Attention technology and has lower complexity, YOLOv10_DSE uses self-integrated attention enhancement detection heads to reduce computational complexity. At the same time, it (YOLOv10_DSE) significantly improves performance. Overall, YOLOv10_DSE has successfully achieved the optimal balance of speed, accuracy and efficiency. Its low model complexity and high detection accuracy make it a detection model with great application potential in resource-constrained substation scenarios.

4.5 Ablation Experiments

To evaluate the effectiveness of the innovation in this paper, we sequentially integrated C2fDSC, SEAMHead, inner_CIoU into the baseline model. Our baseline architecture uses YOLOv10 as the backbone architecture. Then, an ablation experiment was performed on the substation defect dataset dataset. And indicators such as Recall, F1-score, number of parameters, GFLOPs, model size, inference time, mAP@0.5 and mAP@0.95 were evaluated. Table 3 presents a comparison of YOLOv10_DSE with other augmentation modules across these metrics. It aims to analyze the specific impact of different enhancement strategies on the model performance, as well as the trade-off between parameter efficiency and computational cost.

Table 3: YOLOv10_DSE model ablation experiments

Exp	C2f DSC	SEAM Head	inner _CIoU	Recall (%)	mAP @0.5 (%)	mAP @0.95 (%)	F1-score (%)	GFL OPs	Inference (ms)	Param (M)
1				66.0	70.7	46.6	68.5	6.5	1.6	2.43
2	✓			65.9	72.0	47.8	69.4	6.7	1.8	2.52
3		✓		68.1	71.9	47.7	68.7	6.0	1.8	2.17
4			✓	68.3	71.4	47.1	69.7	6.4	2.0	2.05
5	✓	✓	✓	68.5	73.3	48.2	71.4	6.2	2.0	2.43

C2fDSC is introduced into substation hidden danger detection, and the sampling method of convolution kernel is redefined by unique dynamic offset accumulation and constraint, Adaptively adjust the shape along the target's contour to achieve high-precision extraction of defect features. Its core advantages lie in dynamic adjustment and multi-scale coverage, which effectively solve the problems of missed detection of small targets, difficult fitting of irregular defects and poor multi-scale robustness. Especially in complex substation scenarios, the YOLOv10_DSE model introducing C2fDSC has a slight increase in the number of parameters and a slight reduction in Recall, but its performance in mAP@0.5 and mAP@0.95 has improved by 1.3% and 1.2%, respectively. The F1-score improved by 0.9%.

The core of SEAMHead is to extract local features through deeply separable convolution residual blocks, adaptively average pooling and aggregate global contexts, generate channel weights in fully connected layers, dynamically adjust the channel and spatial responses of feature maps, effectively suppress background noise, enhance key regional features, and solve the core problem of missed detection of small targets and difficult fitting of irregular defects in substation hidden danger detection. Its integration with the YOLOv10 detection head significantly improves the accuracy and robustness of defect detection in complex scenarios. In this study, compared to the original network, improvements of 2.1%, 1.2% and 1.1% were achieved in Recall, mAP@0.5 and mAP@0.95 performance, respectively.

inner_CIoU controls the scale size of the auxiliary bounding box by introducing the scale factor ratio, and uses auxiliary bounding boxes of different scales for different datasets and detectors, overcoming the limitation of weak generalization in existing methods. Through the dual improvements of core area overlap constraints and scale-aware aspect ratio penalties, the inner_CIoU loss systematically solves the above problems is a key technology to improve positioning and shape accuracy in substation defect detection. In this study, compared to the original network, improvements of 2.3%, 0.7% and 0.5% were achieved in Recall, mAP@0.5 and mAP@0.95 performance, respectively.

Compared with the original network, the YOLOv10_DSE model integrating C2fDSC, SEAM-Head, and inner_CIoU modules improved the performance of mAP@0.5 and mAP@0.95 by 2.6% and 1.6%, Recall and F1-score also improved by 2.5% and 2.9%, respectively. Whilst maintaining the same number of parameters, computational complexity is markedly reduced. Experimental results demonstrate that each component within YOLOv10_DSE significantly enhances detection performance, successfully achieving an optimised balance between speed, accuracy and efficiency. This is characterised by its low model complexity and high detection accuracy.

4.6 Comparison of Different Datasets and Experimental Results

To evaluate the model's generalisation capability, comparative experiments were conducted using the public PTL-AI Furnas dataset [41]. This dataset serves as a public resource for transmission line fault detection, comprising 6295 images with 17,808 labels across 11 defect categories: Insulator in good condition, Insulator defective, Insulator status unknown, Signal device in good condition, Signal device defective, Signal device near good condition, Bird's nest present, Separator in good condition, Separator defective, Bracket bridge normal, Bracket bridge abnormal. The proposed improved YOLOv10_DSE model was trained and tested. Results are presented in Table 4.

Table 4: Performance Comparison of Various Models on the Public Dataset PTL-AI Furnas

Model	P (%)	mAP@0.5 (%)	mAP@0.95 (%)	Param (M)	GFLOPs
YOLOv5s	83.8	80.8	57.9	2.505	7.1
YOLOv8s	85.6	84.7	61.0	3.007	8.1
YOLOv11n	86.6	85.2	60.8	2.584	6.3
YOLOv12n	85.9	84.8	60.4	2.512	**5.8**
YOLOv13n	86.3	84.4	61.4	2.450	6.2
RT-DETR	86.8	**86.2**	**62.2**	32.006	103.5
Faster-RCNN	60.6	67.2	51.2	41.032	83.4
Mamba-YOLO	86.7	85.3	61.5	5.986	11.8
YOLOv10_DSE	**87.1**	86.0	61.9	**2.436**	6.1

Note: Bold text indicates optimal performance.

In benchmark testing on the PTL-AI Furnas dataset, all models demonstrated a certain performance gap relative to YOLOv10_DSE. In terms of accuracy, it outperformed the second-ranked Mamba-YOLO by 0.4 and the optimal model from the mainstream YOLO series by 0.5. This indicates that YOLOv10_DSE achieves a higher proportion of true positives predicted as positive classes in defect detection tasks, effectively reducing misclassification rates and proving crucial for the reliability of defect detection. Its GFLOPS consumption is merely one-thirteenth that of the high-precision model RE-DETR and significantly lower than Mamba-YOLO. Compared to the YOLO series, while surpassing YOLOv12n, improvements have been achieved in both accuracy and mAP@0.5. This demonstrates that YOLOv10_DSE achieves significant accuracy gains with minimal computational overhead, avoiding the computational explosion seen in high-accuracy models like RT-DETR. This makes it well-suited for real-time inference on edge devices. In terms of parameters, it requires only one-tenth of RT-DETR's and fewer than Mamba-YOLO. YOLOv10_DSE demonstrates well-controlled

model complexity; its low parameter count and GFLOPs enable efficient operation on resource-constrained devices, substantially reducing deployment costs and complexity while enhancing the practicality and scalability of detection systems.

4.7 Visualization of Test Results

In order to have a clearer understanding of the detection results of the YOLOv10_DSE detection algorithm proposed in this paper in practical applications with different hidden danger categories, some representative hidden danger images are selected for visualisation. Examples of images and detection results of different hidden danger categories are shown in Fig. 6.

Figure 6: Detection results visualization of the proposed YOLOv10_DSE model for seventeen hidden dangers

As can be seen from the example detection results, the improved model can effectively detect and locate typical hidden defects in substations, such as meter defects, or small target defects such as smoking, broken dials, and Cover plate damaged, etc. The model can still effectively detect and locate all these small-target defects.

5 Conclusion

This paper proposes YOLOv10_DSE, a multi-hazard detection model tailored for complex substation environments. By ingeniously integrating dynamic snake convolutions with multi-scale feature

fusion, it achieves adaptive defect feature extraction through the synergistic design of multi-branch features and a dynamic offset adjustment mechanism. This addresses the limitation of traditional convolutions' fixed sampling patterns, which struggle to balance the fine details of small targets with the global semantics of large ones. Concurrently, the SEAMHead module—featuring task decoupling and self-integrated attention—enables independent optimisation of localisation and classification. This resolves core challenges in substation defect detection: small-object omission and irregular defect fitting difficulties. This paper specifically addresses these challenges through the inner_CIoU loss, featuring core region overlap constraints and scale-aware aspect ratio penalties. By employing progressive feature enhancement and multi-level fusion strategies across these modules, the study effectively resolves the issue of misclassification in segmentation results caused by the diversity of defect morphologies and ambiguous boundaries. Experimental results demonstrate excellent detection performance on a substation defect dataset encompassing 17 common hazards. The core objective of this study lies in exploring and achieving effective enhancements to edge computing performance, with a particular emphasis on model optimisation and algorithmic refinement. The core objective of this study lies in exploring and achieving effective enhancements to edge computing performance, with a particular emphasis on model optimisation and algorithmic refinement. Nevertheless, we fully recognise the significance of statistical validation and deployment real-time testing. Where objective conditions permit, we plan to supplement these aspects in subsequent research to further refine our findings. Furthermore, the YOLOv10_DSE model exhibits certain limitations. Low-illumination environments—such as night-time inspections or indoor equipment with insufficient lighting—represent common complex scenarios in substation inspections. Specific limitations manifest as reduced detection accuracy, increased false positive rates, and diminished inference stability. Consequently, future work will focus on targeted improvements for low-illumination environments, exploring avenues such as image preprocessing and multimodal data fusion.

Acknowledgement: Not applicable.

Funding Statement: This work was supported by the Science and Technology Project of State Grid Fujian Electric Power Co., Ltd.: Research on Mobile Gridded Swarm Inspection Technology Based on Multi-Rotor UAVs under grant 521304250008.

Author Contributions: The authors confirm contribution to the paper as follows: study conception and design: Wenbin Wu, Manshu Liang; data collection: Wenbin Wu, Weihao Zhang; analysis and interpretation of results: Wenbin Wu, Manshu Liang, Xiaojie Wu; draft manuscript preparation: Manshu Liang, Xiaojie Wu, Ying Ruan. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: The data that support the findings of this study are available from the corresponding author, Manshu Liang, upon reasonable request.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Li G, Yu CH, Liu YP, Fan H, Wen FS, Song Y. Fault prediction and health management of power transformers: challenges and outlook. Power Syst Autom. 2017;41:156–67. (In Chinese). doi:10.7500/AEPS20170213002.
2. Baig MAA, Ratyal NI, Amin A, Jamil U, Liaquat S, Khalid HM, et al. An ensemble deep CNN approach for power quality disturbance classification: a technological route towards smart cities using image-based transfer. Future Internet. 2024;16(12):436. doi:10.3390/fi16120436.
3. Liu H, Yang B, Kang F, Li Q, Zhang H. Intelligent recognition algorithm of connection relation of substation secondarywiring drawing based on D-LLE algorithm. Discov Appl Sci. 2025;7(1):1–12. doi:10.1007/s42452-024-06443-7.
4. Wang J, Sun Y, Lin Y, Zhang K. Lightweight substation equipment defect detection algorithm for small targets. Sensors. 2024;24(18):5914. doi:10.3390/s24185914.
5. Yang Q, Ma S, Guo D, Wang P, Lin M, Hu Y. A small object detection method for oil leakage defects in substations based on improved Faster-RCNN. Sensors. 2023;23(17):7390. doi:10.3390/s23177390.
6. Wang C, Liu Z, Du J, Yang W, Wang T. A YOLO-based substation defect detection model with enhanced feature extraction and fusion. Elect Power Syst Res. 2026;252(9):112354. doi:10.1016/j.epsr.2025.112354.
7. Liu Y, Xu Z, Li G, Xia Y, Gao S. Review of AI-driven data analysis technologies in condition-based maintenance of power transformers. High Volt Technol. 2019;45:337–48. (In Chinese). doi:10.13336/j.1003-6520.hve.20190130001.
8. Jiang X, Sheng G. Research and application of big data analysis for power equipment condition monitoring. High Volt Technol. 2018;44:1041–50. (In Chinese). doi:10.13336/j.1003-6520.hve.20180329001.
9. Meng X. Research on industrial safety multi-object detection using an improved YOLOv7 algorithm. Pure Math. 2024;14(7):1–14. (In Chinese). doi:10.12677/pm.2024.147265.
10. Yuan B, Zhang H, Cui M. Deep learning-based safety management system for power infrastructure construction sites. Power Grid and Clean Energy. 2020;36:30–6. (In Chinese). doi:10.3969/j.issn.1674-3814.2020.09.005.
11. Baig MAA, Ratyal NI, Amin A, Jamil U, Khalid HM, Zia MF. A multi-modal deep learning framework for power quality disturbance classification: an integration of 1D time-series signals and 2D scalograms. Comput Electr Eng. 2025;128(1):110716. doi:10.1016/j.compeleceng.2025.110716.
12. Wang A, Chen H, Liu L, Chen K, Lin Z, Han J, et al. YOLOv10: real-time end-to-end object detection. Adv Neural Inf Process Syst. 2024;37:107984–8011. doi:10.52202/079017-3429.
13. Li Z, Yoshie O, Wu H, Mai X, Yang Y, Qu X. Image analysis method of substation equipment status based on cross-modal learning. IEEJ Trans Electr Electron Eng. 2024;19(9):1507–21. doi:10.1002/tee.24111.
14. Lin J, Yan Y, Sheng G, Jiang X, Yang Y, Chen C. Data cleaning for online monitoring of transformers considering time series correlation. Power Syst Technol. 2017;41:3733–40. (In Chinese). doi:10.13335/j.1000-3673.pst.2017.0141.
15. Kang X, Li Z, Hou J, Xu S, Zhang Y, Zhou Z, et al. Variable dimensional bayesian method for identifying depth parameters of substation grounding grid based on pulsed eddy current. Energies. 2025;18(17):4649. doi:10.3390/en18174649.
16. Ding W, Chen Q, Dong Y, Shao N. Fault diagnosis method of intelligent substation protection system based on gradient boosting decision tree. Appl Sci. 2022;12(18):8989. doi:10.3390/app12188989.
17. Zheng H, Liao R, Grzybowski S, Yang L. Fault diagnosis of power transformers using multi-class least square support vector machines classifiers with particle swarm optimisation. IET Electric Power Appl. 2011;5(9):691–6. doi:10.1049/iet-epa.2010.0298.
18. Ashkezari AD, Ma H, Saha TK, Ekanayake C. Application of fuzzy support vector machine for determining the health index of the insulation system of in-service power transformers. IEEE Trans Dielectr Electr Insul. 2013;20(3):965–73. doi:10.1109/TDEI.2013.6518966.

19. Sampedro C, Martinez C, Chauhan A, Campoy P. A supervised approach to electric tower detection and classification for power line inspection. In: International Joint Conference on Neural Networks (IJCNN). Piscataway, NJ, USA: IEEE; 2014. p. 1970–7. doi:10.1109/IJCNN.2014.6889836.
20. Cheng Y, Liu D. AdIn-DETR: adapting detection transformer for end-to-end real-time power line insulator defect detection. IEEE Trans Instrum Meas. 2024;73(1):1–11. doi:10.1109/TIM.2024.3420265.
21. Xie Z, Dong C, Zhang K, Wang J, Xiao Y, Guo X, et al. Power-DETR: end-to-end power line defect components detection based on contrastive denoising and hybrid label assignment. IET Gener Transm Distrib. 2024;18(20):3264–77. doi:10.1049/gtd2.13275.
22. He Z, Yang W, Liu Y, Zheng A, Liu J, Lou T, et al. Insulator defect detection based on YOLOv8s-SwinT. Information. 2024;15(4):206. doi:10.3390/info15040206.
23. Wu JF, Huang YS, Yang SB, Zhang C, Chen JP, Chen ZF. Insulator defect detection algorithm based on RGB-D multimodal images. In: 8th Asia Conference on Energy and Electrical Engineering (ACEEE). Piscataway, NJ, USA: IEEE; 2025. p. 246–50. doi:10.1109/ACEEE66257.2025.11148604.
24. Liang J, Li L, Zhou HL. An enhanced detection method for ship targets based on improved SSD. Navig Position Timing. 2019;6:43–51. (In Chinese). doi:10.19306/j.cnki.2095-8110.2019.05.007.
25. Rubaiyat AH, Toma TT, Kalantari-Khandani M, Rahman SA, Chen L, Ye Y, et al. Automatic detection of helmet uses for construction safety. In: 2016 IEEE/WIC/ACM International Conference on Web Intelligence Workshops (WIW). Piscataway, NJ, USA: IEEE; 2016. p. 135–42. doi:10.1109/WIW.2016.045.
26. Mu YD, Yan SC, Liu Y, Huang T, Zhou BF. Discriminative local binary patterns for human detection in personal album. In: IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE; 2008. p. 1–8. doi:10.1109/CVPR.2008.4587800.
27. Dalal N, Triggs B. Histograms of oriented gradients for human detection. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. 1. Piscataway, NJ, USA: IEEE; 2005. p. 886–93. doi:10.1109/CVPR.2005.177.
28. Wei YH, Wu SJ, Liu TK, Tan X. Detection of smoking behavior in offices using deep learning. Artif Intell Robot Res. 2023;12(02):55–61. (In Chinese). doi:10.12677/AIRR.2023.122008.
29. He G, Qi D, Yan Y. An algorithm for identifying improper attire based on keypoint detection and attention mechanism and its application. Trans Chin Soc Electri Eng. 2022;42:1826–37. (In Chinese). doi:10.13334/j.0258-8013.pcsee.210282.
30. Qi Y, He Y, Qi X, Zhang Y, Yang G. Dynamic snake convolution based on topological geometric constraints for tubular structure segmentation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway, NJ, USA: IEEE; 2023. p. 6070–9. doi:10.1109/ICCV51070.2023.00558.
31. Yu Z, Huang H, Chen W, Su Y, Liu Y, Wang X. YOLO-Facev2: a scale and occlusion aware face detector. Pattern Recognit. 2024;155(10):110714. doi:10.1016/j.patcog.2024.110714.
32. Zhang H, Xu C, Zhang S. Inner-IoU: more effective intersection over union losswith auxiliary bounding box. arXiv:2311.02877. 2023.
33. Khanam R, Hussain M. YOLOv11: an overview of the key architectural enhancements. arXiv:2410.17725. 2024.
34. Sohan M, Sai Ram T, Rami Reddy CV. A review on YOLOv8 and its advancements. In: International Conference on Data Intelligence and Cognitive Informatics. Cham, Switzerland: Springer; 2024. p. 529–45. doi:10.1007/978-981-99-7962-2_39.
35. Jocher G, Stoken A, Borovec J, Changyu L, Hogan A, Diaconu L, et al. Ultralytics/YOLOv5: V3.0. Zenodo. 2020. doi: 10.5281/zenodo.3983579.
36. Tian Y, Ye Q, Doermann D. YOLOv12: attention-centric real-time object detectors. arXiv:2502.12524. 2025.
37. Lei M, Li S, Wu Y, Hu H, Zhou Y, Zheng X, et al. YOLOv13: real-time object detection with hypergraph-enhanced adaptive visual perception. arXiv:2506.17733. 2025.

38. Zhao Y, Lv W, Xu S, Wei J, Wang G, Dang Q, et al. DETRs beat YOLOs on real-time object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE; 2024. p. 16965–74. doi:10.48550/arXiv.2304.08069.
39. Sun X, Wu P, Hoi SC. Face detection using deep learning: an improved faster RCNN approach. Neurocomputing. 2018;299(2):42–50. doi:10.1016/j.neucom.2018.03.030.
40. Wang Z, Li C, Xu H, Zhu X. Mamba YOLO: SSMs-based YOLO for object detection. arXiv.2406.05835. 2024.
41. De Oliveira FS, De Carvalho M, Campos PHT, Soares ADS, Júnior AC, Quirino ACRDS. PTL-AI Furnas dataset: a public dataset for fault detection in power transmission lines using aerial images. In: 2022 35th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI), Vol. 1. Piscataway, NJ, USA: IEEE; 2022. p. 7–12. doi:10.1109/SIBGRAPI55357.2022.9991806.