IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

IEEE Power & Energy Society Section

Received 8 July 2024, accepted 28 July 2024, date of publication 31 July 2024, date of current version 8 August 2024.

*Digital Object Identifier 10.1109/ACCESS.2024.3436000*

RESEARCH ARTICLE

# A Defect Detection Method for Substation Equipment Based on Image Data Generation and Deep Learning

**NA ZHANG, GANG YANG, DAWEI WANG, FAN HU, HUA YU, AND JINGJING FAN**
State Grid Shanxi Electric Power Research Institute, Taiyuan 030001, China

Corresponding author: Na Zhang (zhangna19907@163.com)

This work was supported by State Grid Shanxi Electric Power Company Science and Technology Project "Research on Target Detection and Defect Recognition Capability Verification Technology of Substation Equipment Based on Image Generation Technology" under Grant 52053023000X.

**ABSTRACT** The task of detecting surface defects on substation equipment faces several challenges, including a variety of target categories, the scarcity of original defect image data, complex environmental conditions, low accuracy in existing algorithms, as well as notable issues with false alarms and missed detections. Overcoming these obstacles is crucial for the successful implementation of intelligent inspection systems for substations. To address the problem of limited original data, we first employ the method of ADD-GAN to augment the image training set. Furthermore, this paper proposes a target detection model called YOLO-SD to detect various equipment defects in complex real-world scenarios. In order to enhance the network's feature extraction capabilities in the presence of complex backgrounds and to improve detection accuracy, a novel deep perceptual feature extraction module named C3+ was introduced in this research. Furthermore, we incorporates SimAM into the neck network of YOLO-SD. This integration not only bolsters the network's learning capacity but also equips it with the capability to autonomously learn and dynamically fine-tune attention weights to suit different input scenarios. To tackle the challenges posed by variations in size and shapes of different substation equipment defects in images, a novel fusion loss function NWD-CIoU is designed. The improvements enhance the accuracy and robustness of YOLO-SD in defect target detection across different scales. The experiment demonstrated that the YOLO-SD model achieved an *mAP*@0.5 of 90.3% and *mAP*@0.5 : 0.95 of 63.9% in detecting defects in substation equipment. The F1 score reached 81.1%, IoU value was 90.5%. This model realized accurate detection of multi-scale substation defect targets, reaching the state-of-the-art level in substation defects detection.

**INDEX TERMS** Defect detection, substation inspection, image data generation, YOLO, multi-scale targets.

## I. INTRODUCTION

### A. BACKGROUND

Substations serve as vital connections between the primary power grid and the distribution network, acting as the central hub of the smart grid [1]. As the power grid continues to expand, the quantity of equipment housed within substations is on the rise. The significance of equipment maintenance and inspection tasks is progressively growing, accompanied by escalating challenges that demand heightened attention and vigilance. The common equipment defects found within substations include metal corrosion, meter damage, transformer oil leakage, ground oil contamination, damaged breathers, cracked insulators, floating debris, bird nests, etc [2], [3]. During routine inspections of electrical equipment within the substations, any abnormal operational states or equipment defects that are not promptly identified can potentially lead to equipment failures and even result in various levels of power accidents.

### B. CONTEXTUALIZATION

Currently, the inspection of substation equipment primarily depends on inspectors performing routine inspections.

The associate editor coordinating the review of this manuscript and approving it for publication was Ravindra Singh.

Many traditional visual monitoring devices used for these inspections lack the capability to effectively alert of safety hazards or detect sudden faults. In the rapidly evolving power industry, traditional inspection methods have proven insufficient to meet the demands of operational maintenance and safety control in substations. The shift towards intelligent operation and maintenance inspections is increasingly seen as an essential trend for the advancement of substations and the broader power system. With the continuous development of artificial intelligence technologies such as computer vision and deep learning, intelligent fault detection in substations has become a hot research field [4]. By utilizing these advanced technologies, automated and efficient fault detection and diagnosis of substation equipment can be achieved, thereby enhancing the safety and reliability of the equipment.

### C. MOTIVATION AND RELATED WORKS

Nowadays, research on intelligent defect detection of equipment within substations is still in its early stages, focusing primarily on insulators [5], transformer windings [6], transmission lines [7], etc. He et al. utilized insulator infrared images to detect their contamination levels, proposing a method based on statistical probability of various pollution feature component values. By selecting initial hidden center parameters for the radial basis function neural network nodes and incorporating gradient descent algorithm and random control factors to adjust hidden center parameters and weight vectors, they improved the detection accuracy [8]. Sadykova et al. applied data augmentation to aerial images of insulators, using the YOLO [9] detection model to detect insulators under different target resolutions and lighting conditions. They successfully employed various classifiers to assess the condition of insulator surfaces in detected images, determining the presence of anomalies such as ice coverage [10]. In the process of abnormal detection in substations, abnormal samples of substation equipment are scarce. Object-based anomaly detection and other anomaly detection methods find it difficult to effectively train models for all equipment and cannot handle unknown anomalies, resulting in poor robustness. To address this issue, Huang et al. [11] adopted a weakly supervised learning method, detecting anomalies by measuring the distance between the test image and images of normal equipment, making it applicable to all equipment. Aiming at the problems of local temperature increase of cable, heating of insulator porcelain disc, local heating of lightning arresters and transformers in the process of abnormal heat generation fault diagnosis of power equipment, Liu et al. proposes a target detection method that YOLOv4 [12], combined with the infrared image data of power equipment to locate and diagnose the faulty heating area of power equipment [13]. Chander and Abbasi [14] proposed a new hyperbolic fuzzy cross entropy (FCE) measure to identify effective features from generated frequency response trajectories and enhance the interpretive accuracy of such trajectories. This measure was applied to the identification and classification of transformer winding defects within predefined frequency ranges.

The researches mentioned above have made significant progress in defect detection on targets such as insulators and conductors. However, there are several shortcomings that need to be addressed. Firstly, the current methods focus on detecting a single category of defects, limiting their ability to detect multiple types of defects across various substation equipments. Secondly, these methods are tested in relatively simple implementation scenarios with plain image backgrounds, which may not accurately reflect the complexity of on-site substation environments [15]. Thirdly, there is a need to enhance the accuracy of the detection methods to ensure reliable results in real-world applications. Furthermore, Abbasi and Mohammadi [16] presented a comprehensive comparison and assessment between these uncertainty methods based on accuracy, complexity, and simulation time.

With the continuous advancement of machine vision technology and the ongoing enhancement of computer performance, the utilization of deep learning is progressively expanding across diverse domains [17], [18]. Particularly in the realm of automated defect detection, deep learning technology has made remarkable strides [19], [20], [21]. By means of deep learning algorithms, computers are able to grasp and interpret patterns and characteristics within visual data, leading to the realization of more precise and efficient automated detection tasks. With the continuous evolution of technology, target detection models based on deep learning have given rise to two distinct technical branches in network structure: the two-stage detection models and the one-stage detection models [22]. The representative of the two-stage object detection model is the R-CNN series models [23], [24], [25]. By introducing a region proposal network to generate candidate regions, and then classifying and regressing these candidate regions, the accuracy of object detection is enhanced. Although two-stage models excel in accuracy, their reliance on two independent stages of processing leads to a lack of real-time performance. This is due to the need to classify and regress a large number of candidate regions during the testing phase, resulting in a high computational load that hinders real-time application scenarios. In response to the real-time performance limitations of two-stage models, researchers have begun exploring single-stage detection models, exemplified by the SSD [26] and YOLO series models [27].

### D. MAIN CONTRIBUTIONS

In order to achieve precise and real-time detection of various defects on the surfaces of substation equipment in practical scenarios, this study introduces a substation defect detection method based on an improved YOLOv7 model [28] (YOLO-SD). The primary contributions are outlined below:

1) In order to enhance the network's feature extraction capabilities in the presence of complex backgrounds and to improve detection accuracy, a novel deep perceptual feature extraction module named C3+ is introduced in this research. By integrating C3+ into the YOLO-SD backbone network, the network gains improved local deep attention and global processing capabilities, enabling it to accurately capture image features associated with substation defects. This innovative design equips the backbone network to effectively handle feature information across windows, thus expanding the deep receptive field of the network and preserving crucial weight information.
2) This study incorporates SimAM into the neck network of YOLO-SD. This integration not only bolsters the network's learning capacity but also equips it with the capability to autonomously learn and dynamically fine-tune attention weights to suit different input scenarios. These scenarios encompass intricate characteristics of targets, minute details of distant targets, and extensive features of close-range targets. Consequently, the neck network can prioritize capturing the intricate features of defects in electrical equipment while disregarding extraneous data. As a result of enhancing the precision of feature fusion, the overall detection accuracy of YOLO-SD is refined.
3) A novel fusion loss function NWD-CIoU is designed to tackle the challenges posed by variations in size and shapes of different substation equipment defects in images. It achieves this by modeling the predicted bounding box and the actual bounding box as two-dimensional Gaussian distributions. The similarity between these distributions is then measured using the normalized Wasserstein distance. The NWD-CIoU loss function enhances the robustness of defect target detection across different scales, thereby improving the detection accuracy of the YOLO-SD model.

The rest of the article is structured as follows: data set preparation work is presented in Section II and our proposed methods, including the YOLO-SD model and the design details are presented in Section III, experiments and discussion are done in Section IV to verify the method's validity. Finally, conclusions are given in Section V.

## II. IMAGE DATA GENERATION

In this paper, the types of defects to be detected in the substations include a variety of types, such as insulators damage, equipment surface oil contamination, ground oil contamination, metal corrosion, and meter breakage. These defects occur on different equipment within the substation. Due to the complex environment of the substation and the difficulty of inspection, capturing images of the defects poses a significant challenge. We collected images of the above 5 types of defects through monitoring, on-site collection, and online search. We obtained 165 images of equipment surface oil contamination, 186 images of ground oil contamination, 204 images of metal corrosion, 220 images of insulator damage, and 254 images of meter breakage. It is evident that there is an imbalance in the quantity of images depicting various types of defects that we have prepared. Furthermore, this quantity falls significantly short of meeting the requirements for the training set volume of the target detection model based on deep learning.

To enhance the diversity of abnormal samples for defect detection models based on deep learning, in our another work, an ADD-GAN algorithm [29] is proposed for generating defects on substation equipments, as shown in Figure 1. Building upon adversarial deep learning, the algorithm introduces a local region defect generation network and a joint discriminator for overall image and the defect image. This approach allows for segmentation of local regions and generation of relevant defects without altering the global image features. Simultaneously, it pays attention to both the overall image quality and the fusion consistency between global and local defect images. As a result, ADD-GAN can generate high-quality equipment defect images with rich features and high realism. After using the ADD-GAN model for generation, the number of images for each type of defect reached 10000.

## III. METHODOLOGIES

### A. OVERVIEW

Due to the intricate nature of the substation environment, images captured through monitoring or manual collection frequently encounter significant background interference, varying image quality, and inconsistent defect scales, presenting formidable obstacles to target detection. Current methodologies exhibit insufficient robustness in addressing these challenges, leading to diminished accuracy in detecting defects within substations.

To enhance the robustness of the target detection model for detecting defects in substations, we have made improvements to the YOLOv7 model and proposed a substation defect detection model named YOLO-SD. YOLO-SD introduces a novel deep perceptual feature extraction module called C3+ into the backbone, enabling the network to obtain stronger local depth attention and global processing capabilities. This allows the model to accurately capture image features related to substation defects. Further, we integrate SimAM into the neck network of YOLO-SD. This integration not only enhances the network's learning capacity but also equips it with the capability to autonomously learn and dynamically fine-tune attention weights to suit different input scenarios. Finally, a novel fusion loss function NWD-CIoU is designed to tackle the challenges posed by variations in size and shapes of different substation equipment defects in images.

In the following description, we will detail and explain the specific methods.

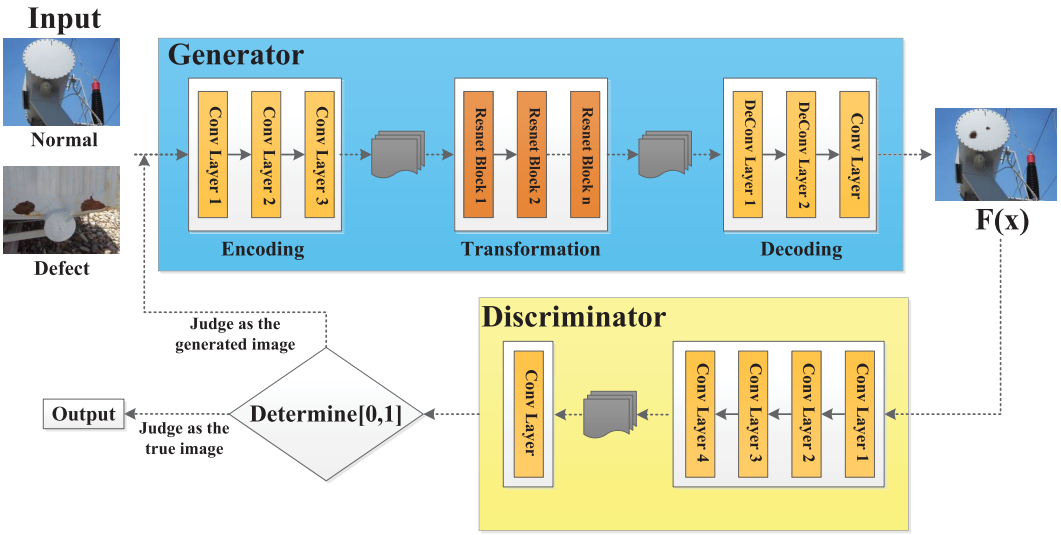


**FIGURE 1.** The ADD-GAN model.

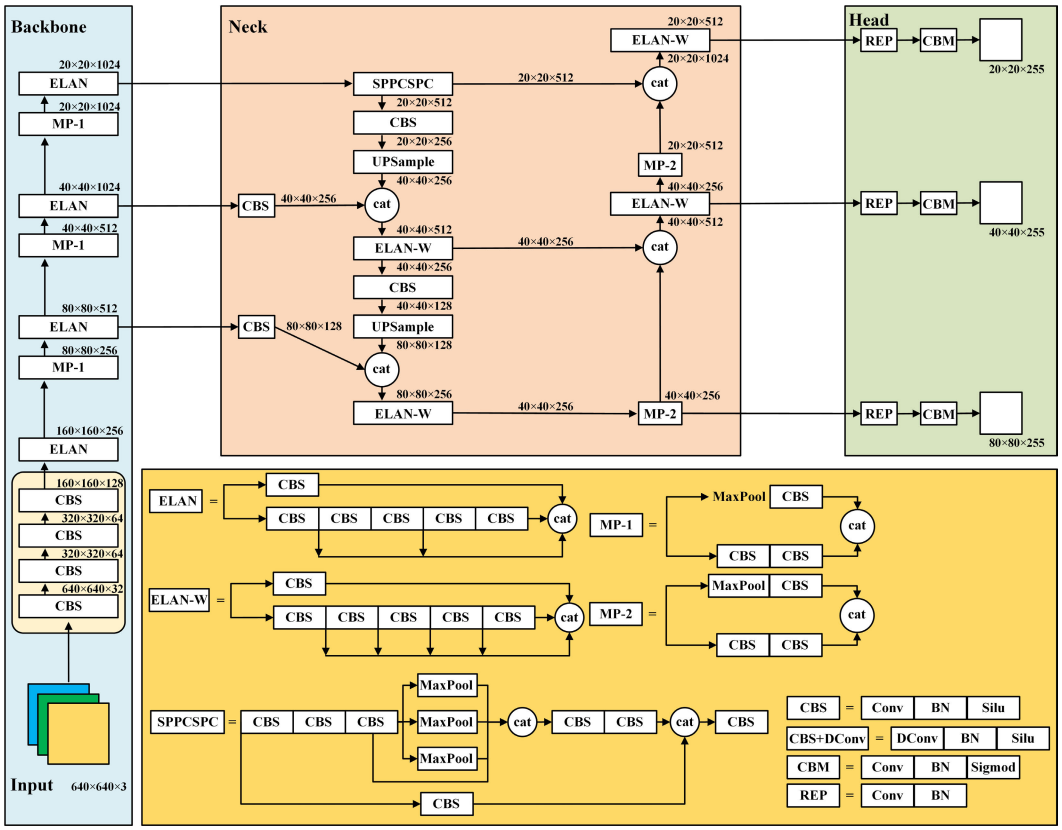


**FIGURE 2.** The YOLOv7 model.

### B. YOLOV7 MODEL

The YOLOv7 object detection framework is made up of three key parts: Input, Backbone, Neck and Head, shown in Figure 2. The Input section pre-processes the input image by employing techniques like image augmentation, resizing, and concatenating. The Backbone segment carries out feature extraction across various scales using downsampling techniques. The Neck portion performs feature inference and produces the detection outcomes. The Backbone part integrates several modules, including CBS, ELAN, and MPConv. The Neck section incorporates CBS, SPPCSPC, UpSampling, MPConv and ELAN-W modules. The Head

part includes REP and CBM modules. To facilitate multi-scale feature fusion for object detection tasks, YOLOv7 establishes a Feature Pyramid Network (FPN) by connecting modules at different layers of the Backbone and Neck.

### C. C3+ MODEL

The C3 module is an important component of the YOLOv5 [30] model, which aims to increase the network depth and receptive field, thus enhancing feature extraction capability. The C3 module consists of three Conv blocks, with the first Conv block having a stride of 2 to reduce the size of the feature map by half. This is done to increase the network's receptive field while reducing computational load. The second and third Conv blocks have a stride of 1, meaning they do not change the size of the feature map. This is to maintain spatial resolution of the feature map, better preserving local object information. Additionally, these two Conv blocks further extract features to increase network depth and receptive field. The Conv blocks in the C3 module use $3 \times 3$ convolutional kernels. Between each Conv block, Batch Normalization (BN) layers and LeakyReLU activation functions are included to enhance model stability and generalization performance.

In order to enhance the depth and receptive field of the YOLOv7 backbone, and to boost the network's feature extraction capabilities, we introduce an improved C3 module, referred to as C3+. This module is seamlessly integrated into the YOLOv7 backbone. The C3+ module harnesses the benefits of the Swin Transformer Block and Switchable Atrous Convolution (SA Conv) [31] to enable seamless inter-window information interaction within the backbone network. It not only preserves intricate depth details and widens the network's scope of perception but also integrates attention mechanisms to promote a more profound exploration of feature information. In essence, the C3+ module's architecture follows a residual network approach and cleverly integrates cutting-edge feature processing modules, thereby amplifying the feature extraction prowess of the YOLO-SD backbone network when processing input image data.

The advanced perceptual feature extraction module C3+ integrates a sliding window mechanism and consists of the CBS basic convolution module, Swin module, SA Conv, as well as Concatenation operations. It is seamlessly combined with lightweight attention mechanisms such as Channel Attention (CA) and the parameter-free attention module SimAM. The core architecture of C3+ is inspired by the design principles of ResNet [32], as shown in Figure 3.

In Figure 3, the Swin module is derived from the essential components of the Swin Transformer [33]. This research incorporates the Patch Partition and Linear Embedding operations from the Swin Transformer model into the Swin module. The structural overview of the Swin module is illustrated in Figure 4(a). Initially, this module employs the Patch Partition operation to segment the input feature map into distinct fixed-size and non-overlapping blocks, enabling individual processing of each block. This segmentation aids in capturing detailed local features more effectively and enhances the computational efficiency of the model. Subsequent to the Patch Partition operation, the Linear Embedding operation transforms the input feature information into a lower-dimensional feature space, reducing feature dimensionality and obtaining more concise feature representations. The processed feature information is then fed into the Swin Transformer block, arranged in a structured manner. This arrangement facilitates information transformation and interaction operations, leading to feature extraction and integration.

The Swin Transformer block incorporates a sliding window mechanism, allowing the model to access cross-window information. This mechanism promotes interaction between different windows, enabling the model to capture both local and global contextual information within the image. By facilitating the exchange of information across windows, the model enhances its overall expressive power and performance.

The Swin Transformer module's architecture incorporates two Window-based Multi-Head Self-Attention (WMSA) blocks as shown in Figure 4(b). WMSA operates with a standard window setup to process features independently for each local window. On the other hand, WMSA utilizes a shifted window configuration, integrating shift operations to facilitate information exchange across different windows, effectively broadening the model's global receptive field.

The SA Conv module comprises Switchable Atrous Convolution, Pre-Global Context, and Post-Global Context modules, along with a Switch Mechanism [31]. This module adjusts the dilation rate and switch value of Atrous Convolution to capture details of targets across different scales, mitigating information weight loss and bolstering the model's capacity to handle image data. The structure of the SA Conv module is depicted in Figure 5. More specifically, this module incorporates a switch mechanism that dynamically selects and balances the ratio between Regular Convolution and Atrous Convolution. The effective fusion of Atrous Convolution with the Switch Mechanism defines the SA Conv component, and the computation formula for transitioning from regular convolution to the SA Conv component is as follows:

$$\begin{aligned} Conv(x, w, 1) &\xrightarrow[to\ SA\ Conv]{Convert} S(x)\cdot Conv(x, w, 1) \\ &+ (1 - S(x))\cdot Conv(x, w + \Delta w, r), \end{aligned} \tag{1}$$

where the hyper-parameter $r$, along with the trainable weight $\Delta w$ play crucial roles in SA Conv. The switch function $S(x)$ is specifically designed to be implemented as an average pooling layer with a $5 \times 5$ kernel, followed by a $1 \times 1$ convolutional layer. This switch function is dependent on both the input data and its location within the network, allowing the backbone model to dynamically adjust and adapt to various scales as required.

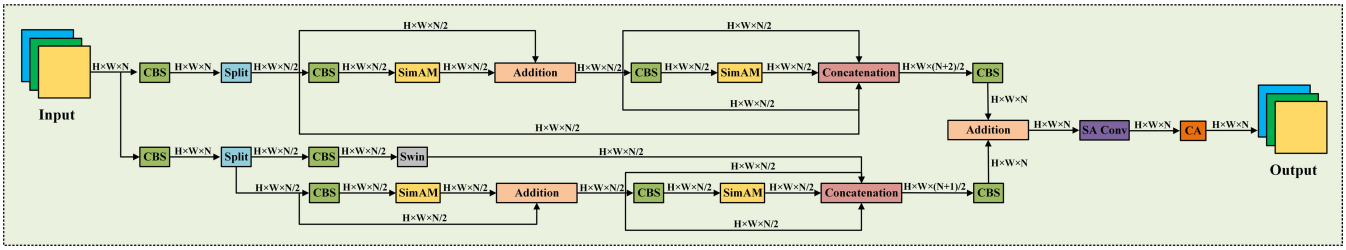


**FIGURE 3.** Structure of the C3+ module.

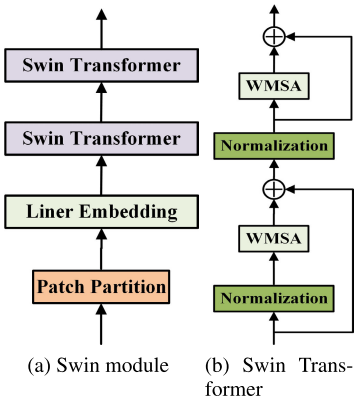


**FIGURE 4.** The structure of the Swin module.

In this paper, the Coordinate Attention (CA) mechanism [34] is integrated into the end of the C3+ module. Its purpose is to reduce the loss of global positional information weights, enhance the focus of the backbone network during feature extraction, and directly input precise feature information to the next layer of the network to avoid retaining unnecessary redundant information. The core idea of Coordinate Attention is to embed positional information into the feature channel dimension, decomposing channel attention into aggregating features along two spatial directions. One direction captures long-range dependency relationships, while the other preserves precise positional information. Finally, the two are complementary fused to learn the importance weights of different channels, thereby enhancing focus on complex target features and suppressing redundant or noisy channels. The specific operations include two steps: coordinate information embedding and coordinate attention generation. The CA structure is shown in Figure 6. In this paper, the CA attention mechanism is incorporated at the conclusion of the C3+ module after SA Conv. The aim is to alleviate the loss of global positional information weight, bolster the focus of the backbone network during the feature extraction process, and feed precise feature information directly into the subsequent layer of the network, thereby preventing the preservation of superfluous redundant data.

### D. INTRODUCING SIMAM IN THE NECK

A Simple, Parameter-Free Attention Module (SimAM) [35] is grounded in neuroscience theory and introduces a novel attention module that aims to enhance the performance of neural networks by optimizing the energy function to evaluate the significance of individual neurons. By doing so, it dynamically adjusts the distribution of attention weights, thereby shaping the focus of the network. One key strength of this module is its ability to achieve these improvements without introducing additional parameters to the existing network architecture. Instead, it effectively utilizes its minimal parameter set to estimate three-dimensional attention weight distributions for feature maps.

Given the simplicity, parameter-free nature, and strong adaptability of the SimAM attention module, this paper integrates SimAM into the Neck network of YOLO-SD. The introduction of SimAM helps enhance the learning capability of the Neck network in YOLO-SD. Additionally, it can automatically learn and dynamically adjust the distribution ratio of attention weights based on various input scenarios, such as complex features of the target, tiny features from distant target scales, and oversized features caused by nearby target scales. This guides the Neck network to focus more on the detailed features of substation defects, filtering out irrelevant information. Therefore, it improves the accuracy of feature fusion in the neck network, enhancing the overall detection accuracy of YOLO-SD.

After the above improvements, the network architecture diagram of the YOLO-SD model is shown in Figure 7.

### E. NWD-CIOU LOSS

To effectively capture distinguishing features, the proper distribution of positive and negative samples is paramount. While multi-scale defect detection is prevalent in real-world substation scenarios, current YOLOv7 object detectors primarily concentrate on targets of standard sizes. The scale-dependent nature of the intersection over union (IoU) metric introduces notable fluctuations in IoU scores across targets of diverse scales. Particularly for smaller targets, even minor shifts in position can yield significant IoU variations, unlike with regular-sized objects where similar positional adjustments have minimal impact on IoU values. This scale sensitivity of the IoU metric manifests as abrupt changes in specific bounding box placements, ultimately compromising the accuracy of predicted bounding box alignment.

In an effort to address this issue, a novel evaluation metric for detecting small objects using Wasserstein Distance

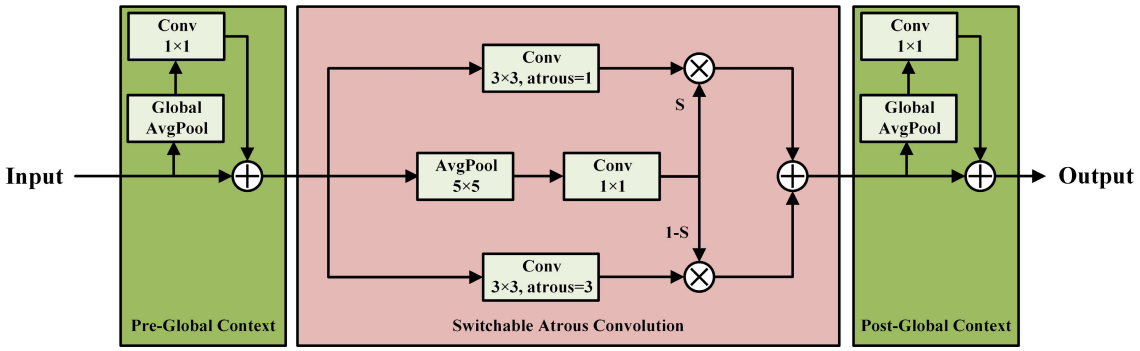


**FIGURE 5.** Structure of the SA Conv module.

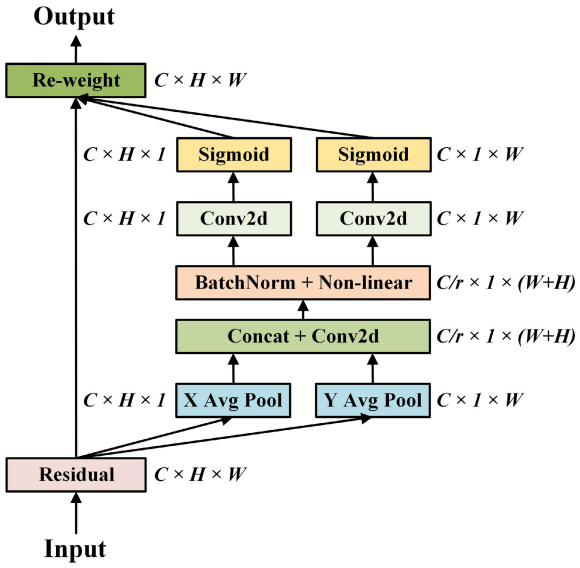


**FIGURE 6.** Structure of the CA module.

has been introduced [36]. Initially, bounding boxes are represented as 2D Gaussian distributions. Subsequently, a new metric known as the Normalized Wasserstein Distance (NWD) is proposed to measure the similarity between them by comparing their respective Gaussian distributions. The NWD metric can be seamlessly integrated into the assignment, non-maximum suppression, and loss functions of any anchor-based detector, serving as a replacement for the traditional IoU metric. NWD offers two key advantages: firstly, it can accurately quantify the similarity in distribution between small targets even in the absence of overlap; secondly, its insensitivity to targets of varying scales renders it particularly well-suited for assessing similarity among small targets.

For small targets, there are always background pixels present within the bounding box, as real objects are unlikely to perfectly fit within a rectangle. Typically, foreground pixels are concentrated towards the center of the bounding box, while background pixels tend to be more prominent towards the edges. To better weight each pixel within the bounding box, one approach is to model the bounding box as a two-dimensional Gaussian distribution $N(\mu, \Sigma)$, as shown below:

$$\mu = \begin{bmatrix} c_x \\ c_y \end{bmatrix}, \Sigma = \begin{bmatrix} \frac{w^2}{4} & 0 \\ 0 & \frac{h^2}{4} \end{bmatrix}, \tag{2}$$

where $c_x$, $c_y$ represents the center coordinates of the bounding box, and $x$ and $h$ represent the width and height of the box. For two 2D Gaussian distributions $\mu_1 = N(m_1, \Sigma_1)$ and $\mu_2 = N(m_2, \Sigma_2)$, their second order Wasserstein Distance is as follows:

$$W_2^2(\mu_1, \mu_2) = ||m_1 - m_2||_2^2 + ||\Sigma_1^{1/2} - \Sigma_2^{1/2}||_F^2, \tag{3}$$

where $||\cdot||$ is the Frobenius norm. By simplifying it using Gaussian distributions $N_g$ and $N_t$, where $N_g$ represents the predicted bounding box $B_g = (cx_g, cy_g, w_g, h_g)$ and $N_t$ represents the ground truth bounding box $B_t = (cx_t, cy_t, w_t, h_t)$, the second order Wasserstein Distance between the two bounding boxes can be simplified as follows:

$$W_2^2(N_g, N_t) = ([cx_g, cy_g, \frac{w_g}{2}, \frac{h_g}{2}]^T, [cx_t, cy_t, \frac{w_t}{2}, \frac{h_t}{2}]^T)_2^2, \tag{4}$$

Since $W_2^2(N_g, N_t)$ is a distance unit, while the threshold representing the similarity of bounding boxes should be a ratio in the $(0, 1)$ interval, it is necessary to normalize $W_2^2(N_g, N_t)$ to obtain the normalized Wasserstein distance (NWD) as follows:

$$NWD(N_g, N_t) = exp[-\frac{W_2^2(N_g, N_t)}{C}]. \tag{5}$$

Furthermore, NWD loss function is as follows.

$$L_{NWD} = 1 - NWD(N_g, N_t). \tag{6}$$

NWD loss function works well for small-scale targets. However, since the images of defects in substations are multi-scale, we have also incorporated CIoU loss [37] in the loss function part. CIoU takes into account the overlap area, center point distance, and aspect ratio, these three geometric

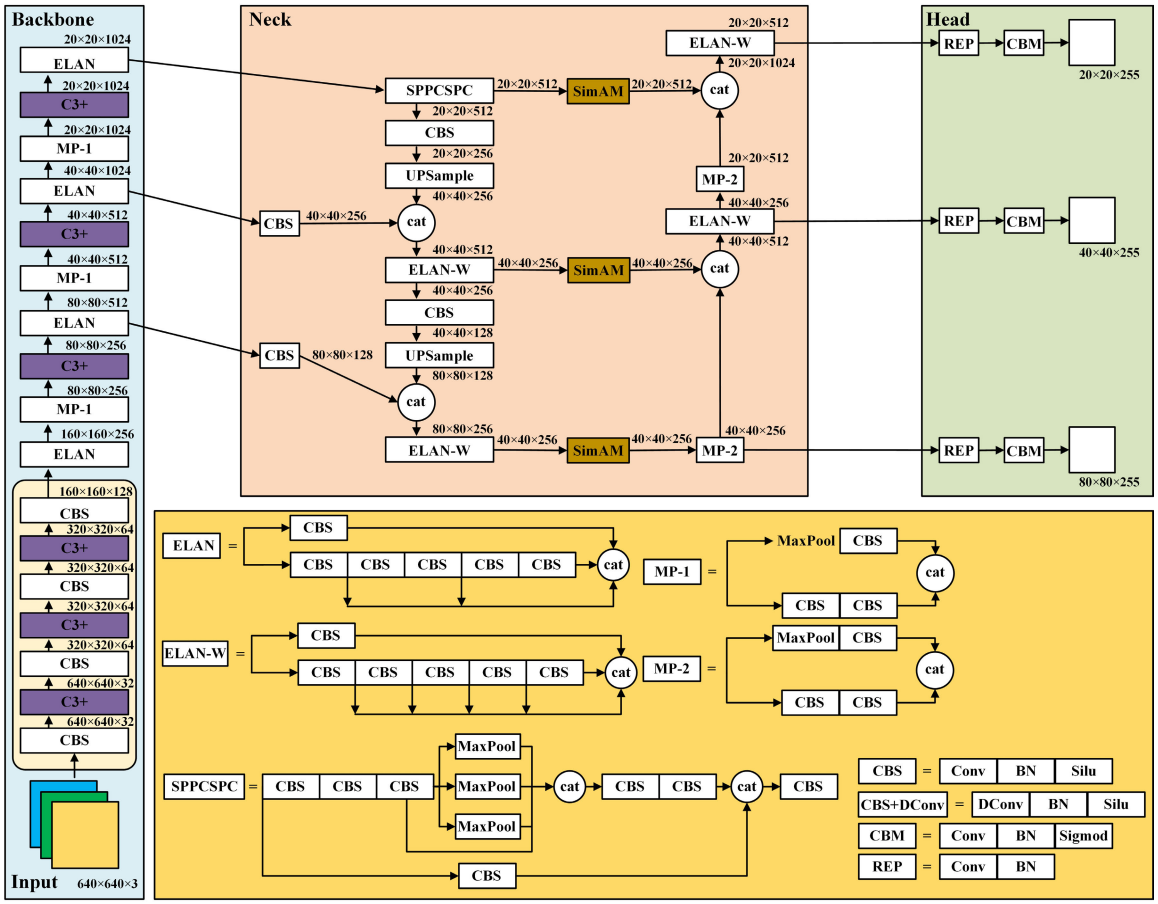


**FIGURE 7.** The YOLO-SD model.

measures, enabling faster and better bounding box regression. The CIoU loss is as follows:

$$IoU = \frac{B_{pre} \cap B_{gt}}{B_{pre} \cup B_{gt}}, \tag{7}$$

$$v = \frac{4}{\pi^2}(arctan\frac{w_{gt}}{h_{gt}} - arctan\frac{w_{pre}}{h_{pre}})^2, \tag{8}$$

$$\alpha = \frac{v}{(1 - IoU) + v}, \tag{9}$$

$$L_{CIoU} = 1 - IoU + \frac{d^2}{c^2} + \alpha v, \tag{10}$$

where $B_{pre}$ represents the predicted bounding box and $B_{gt}$ represents the ground truth. $v$ is a parameter that assesses the aspect ratio consistency between the width and height. $w_{gt}$ stands for the width and $h_{gt}$ for the height of the ground-truth bounding box, while $w_{pre}$ denotes the width and $h_{pre}$ denotes the height of the predicted bounding box. Additionally, $\alpha$ determines the balance between the aspect ratios of the predicted bounding box and the ground-truth bounding box. Furthermore, $d$ represents the Euclidean distance between the center point of the predicted bounding box and the center point of the ground truth, and $c$ signifies the diagonal length of the smallest enclosing rectangle that encompasses both the predicted bounding box and the ground truth.

By assigning appropriate fusion weights $\lambda$ to $L_{NWD}$ and $L_{CIoU}$, the NWD-CIOU loss function is proposed as a metric by combining NWD with CIoU as follows:

$$L_{NWD-CIoU} = \lambda \cdot L_{CIoU} + (1 - \lambda) \cdot L_{NWD}. \tag{11}$$

## IV. EXPERIMENTS AND DISCUSSION

### A. EXPERIMENTAL PREPARATION

In this paper, we selected 5 types of substation equipment defects, namely insulators damage, equipment surface oil contamination, ground oil contamination, metal corrosion, and meter breakage. We augmented the number of images for each type of defect to 10000 using ADD-GAN method, forming the training dataset for the YOLO-SD model.

In this study, we assess the performance of the YOLO-SD model by examining its detection capabilities for 5 types of substation equipment defects. The model is trained and tested on the IW4210-8G server, with detailed specifications provided in Table 1. Initialization of the YOLO-SD model includes the parameters outlined in Table 2. To ensure optimal performance, the input image resolution is set to

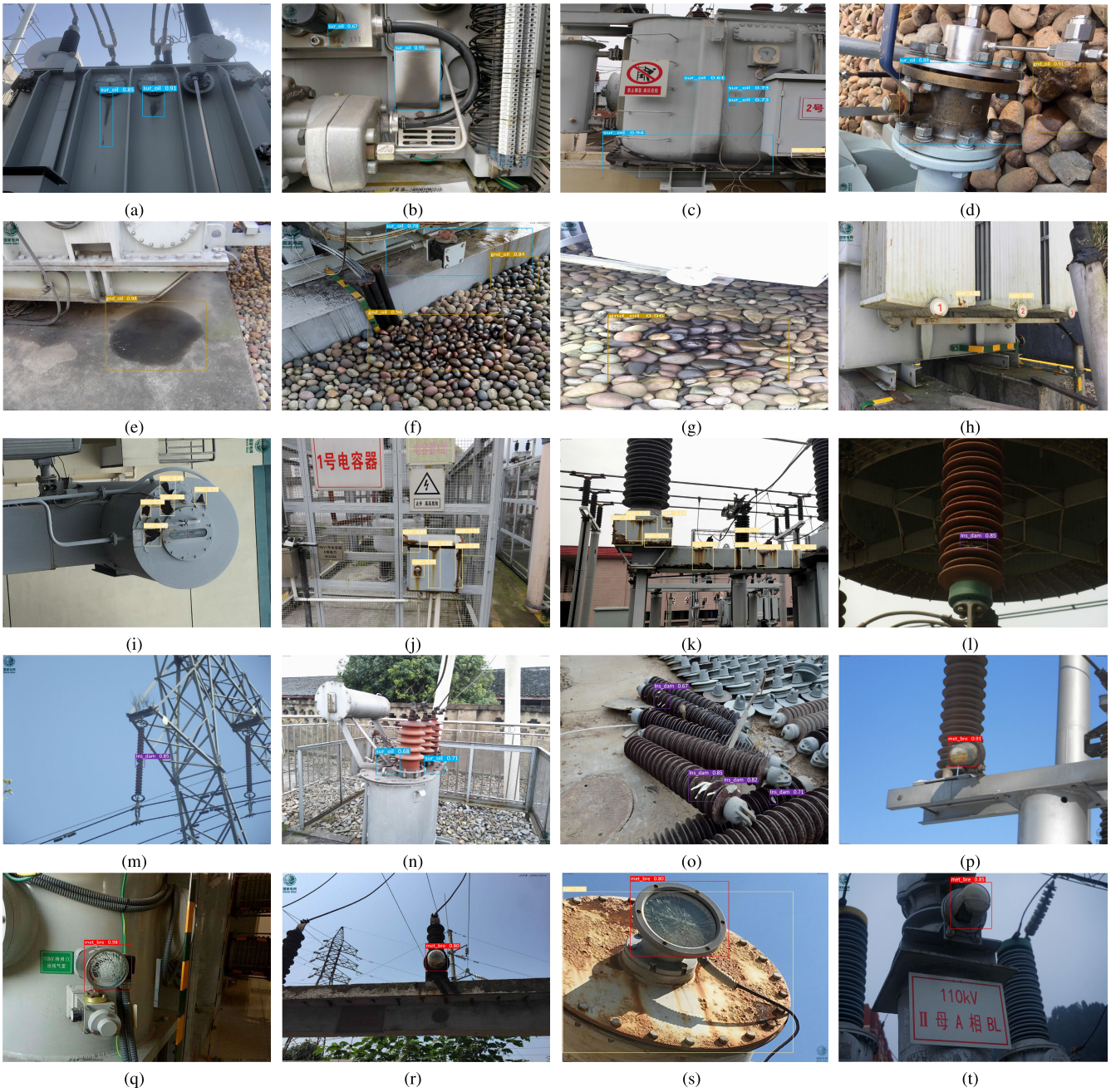


**FIGURE 8.** Detection results of YOLO-SD.

**TABLE 1.** Configurations of the IW4210-8G server.

| | |
|---|---|
| CPU | Intel(R) Xeon(R) Silver 4214 CPU @ 2.20GHz |
| CPU MHz | 1000 |
| CPU cache size | 16896 KB |
| RAM | 257580 GB |
| GPU | NVIDIA GeForce RTX 2080 Ti |
| VRAM | 11 GB GDDR 6 |
| Memory clock | 14000 MHz |
| Core clock | 1350-1545 MHz |

**TABLE 2.** Initialization parameters of YOLO-SD network.

| | |
|---|---|
| Size of input images | 640×640 |
| Batch | 8 |
| Optimizer | Adam |
| Momentum | 0.999 |
| Initial learning rate | 0.001 |
| Decay | 0.0005 |
| Training steps | 3000 |

640 × 640, and a batch size of 8 is used during training. Key parameters such as momentum, initial learning rate, and decay are inherited from the original YOLOv7 model due to their proven effectiveness. To gain a deeper understanding of the training process, we conduct 3000 training steps. The learning rate is dynamically adjusted during training to improve model performance and convergence. By adhering

**TABLE 3.** Detection evaluation parameters of different species of defects.

| Category | mAP@.5:.95(%) | mAP@.5(%) | mAP@.75(%) | mAP@.9(%) | F1 score(%) |
|---|---|---|---|---|---|
| Equipment surface oil contamination | 62.7 | 89.3 | 66.9 | 41.7 | 80.5 |
| Ground oil contamination | 61.5 | 87.1 | 65.4 | 39.9 | 77.5 |
| Metal corrosion | 64.3 | 89.7 | 68.5 | 42.4 | 80.8 |
| Insulator damage | 65.8 | 91.4 | 70.2 | 43.8 | 82.0 |
| Meter breakage | **65.6** | **93.4** | **74.8** | **46.3** | **83.6** |
| Over all | 63.9 | 90.3 | 69.5 | 43.2 | 81.1 |

to these training protocols and parameter configurations, we effectively train the YOLO-SD model and evaluate its efficacy in detecting the specified substation equipment defects.

In the following sections, we will utilize average detection time, mean Average Precision (mAP), F1 score, mAP under various IoU thresholds, and other evaluation metrics.

### B. RESULTS AND ANALYSIS OF YOLO-SD

In this paper, we collected images of the aforementioned 5 types of defects through monitoring, on-site collection, and online search.

We gathered 165 images of equipment surface oil contamination, 186 images of ground oil contamination, 204 images of metal corrosion, 220 images of insulator damage, and 254 images of meter breakage. Further, we use ADD-GAN method to increase the number of images for each class to 10000. We use the aforementioned 50000 images to train the YOLO-SD model, and validate it using a test set composed of new defect images collected from the substation site. Some test results are shown in Figure 8. It is evident that the YOLO-SD model excels in accurately detecting the categories and locations of defects in substations within complex real-world scenarios. Moreover, it has the capability to identify multiple defects within a single image.

Furthermore, we analyze the performance metrics of the YOLO-SD model in detail, as shown in Table 3. It can be observed that the detection results of the YOLO-SD model on the aforementioned 5 types of substation equipment defect test sets achieve an *mAP*@.5 value of 90.3% and an F1 Score value of 81.1%, demonstrating accurate detection of substation equipment defects. Among them, the YOLO-SD model exhibits the highest accuracy in detecting meter wear defects, benefiting from the more prominent features of meter damage defects among the various defects. The detection accuracy of ground oil contamination is the lowest, attributed to the indistinct image features of ground oil contamination and its susceptibility to factors such as shadows and watering.

### C. DISCUSSION OF NWD-CIOU LOSS

In this paper, we introduce a NWD-CIoU loss function for multi-scale object detection tasks, aiming to strike a balance between the NWD loss and CIoU loss through the utilization of a weight parameter $\lambda$. To validate the efficacy of the proposed NWD-CIoU loss function, we systematically varied the $\lambda$ parameter from 0 to 1 during the training of the YOLO-SD model. The performance outcomes of the trained model on the test dataset are detailed in Table 4. The analysis of metrics like mAP and IoU indicates that the model achieves optimal performance when the value of $\lambda$ is approximately 0.6. Moreover, the comparative results highlight that the integration of the NWD-CIoU loss function enhances the YOLO-SD model's ability to accurately detect small defects in images, enabling precise detection of multi-scale defective targets. Therefore, in this paper, we choose $\lambda = 0.6$.

**TABLE 4.** Detection evaluation parameters of different values of $\lambda$.

| $\lambda$ | mAP@.5:.95(%) | mAP@.5(%) | F1 score(%) | IoU(%) |
|---|---|---|---|---|
| 0 | 62.3 | 88.6 | 80.5 | 89.7 |
| 0.2 | 62.9 | 89.0 | 80.6 | 90.1 |
| 0.4 | 63.7 | 89.5 | 80.8 | 90.3 |
| **0.6** | **63.9** | **90.3** | **81.1** | **90.5** |
| 0.8 | 63.4 | 89.6 | 80.9 | 90.2 |
| 1 | 63.1 | 89.0 | 80.6 | 90.1 |

### D. COMPARISON OF DIFFERENT DETECTION MODELS

The YOLO-SD model is benchmarked against SSD512 [26], Faster R-CNN [25], YOLOv7 [28], YOLOv8 [38], and the state-of-the-art substation defect detection model [13], as illustrated in Table 5. Through a comprehensive analysis of the comparison results, it is evident that YOLO-SD outperforms the other advanced detection models in terms of accuracy, especially in detecting multi-scale defect targets. This accomplishment establishes YOLO-SD as the state-of-the-art model in substation defect detection.

Furthermore, an analysis was conducted to understand the underlying factors contributing to the reduced accuracy observed in several other models, each displaying distinctive levels of missed detections in the small target defect detection task. This examination reinforces the effectiveness of the C3+ module integrated into the backbone, the SimAM mechanism implemented in the Neck, and the NWS-CIoU loss function introduced by YOLO-SD. These enhancements have

**TABLE 5.** Comparison results of different detection models.

| Methods | mAP@.5(%) | mAP@.5:.95(%) | F1 score(%) | IoU(%) | Params (MB) | Testing Time (FPS) |
|---|---|---|---|---|---|---|
| SSD512 | 70.3 | 46.7 | 66.4 | 79.1 | 177.2 | 15.8 |
| Faster R-CNN | 79.5 | 54.7 | 73.2 | 85.5 | 535.5 | 4.1 |
| YOLOV7 | 84.6 | 58.3 | 78.4 | 87.5 | 55.7 | 84.9 |
| YOLOV8 | 87.8 | 60.7 | 74.5 | 85.9 | **38.7** | **102.1** |
| YOLO-Liu | 74.8 | 52.6 | 71.6 | 81.7 | 108.9 | 38.3 |
| YOLO-SD | **90.3** | **63.9** | **81.1** | **90.5** | 53.7 | 75 |

**TABLE 6.** Results of the ablation experiments.

| C3+ | SimAM | mAP@.5(%) | F1 score(%) | IoU(%) |
|---|---|---|---|---|
| | | 85.1 | 78.7 | 88.0 |
| ✓ | | 87.8 | 79.5 | 88.7 |
| | ✓ | 85.8 | 79.0 | 89.4 |
| ✓ | ✓ | **90.3** | **81.1** | **90.5** |

notably bolstered the models' capacity to extract features from targets of varying scales, particularly emphasizing small targets, consequently facilitating precise defect detection in substations.

### E. ABLATION EXPERIMENT

In YOLO-SD, two significant components were incorporated: the C3+ module within the feature extraction network and the SimAM mechanism in the neck. A series of ablation experiments were conducted to assess the impact of each functional module on the performance of the YOLO-SD model. The detailed findings and outcomes of these experiments are summarized in Table 6 for reference.

The improvements mentioned indeed play a crucial role in enhancing the accuracy of the model. The incorporation of the C3+ module within the backbone of the YOLO-SD model significantly contributes to enhancing the overall accuracy. This is attributed to the C3+ module's ability to empower the network with enhanced local depth attention and global processing capabilities, facilitating the precise capture of image features associated with substation defects. This innovative design empowers the backbone network to effectively manage feature information across windows, thereby expanding the network's depth receptive field and preserving vital weight information.

On the other hand, the integration of the SimAM module within the Neck component demonstrates a notable improvement in enhancing IoU accuracy. This highlights that the introduction of SimAM not only boosts the network's learning capacity but also equips it with the capability to autonomously learn and dynamically adjust attention weights to suit various input scenarios. These scenarios encompass intricate features of the target, subtle details of distant targets, and extensive features of nearby targets. Consequently, the neck network can selectively capture the complex features of electrical equipment defects while disregarding irrelevant data. By elevating the accuracy of feature fusion, the overall detection accuracy of YOLO-SD experiences a substantial enhancement.

## V. CONCLUSION

This paper is focused on tackling the issue of low detection accuracy in substation defect detection, which is influenced by various factors like the diverse types of defects, complex environmental conditions, and significant variations in defect image scales. By leveraging the YOLOv7 model and implementing several enhancement techniques, the proposed YOLO-SD model aims to achieve improved precision and enhanced robustness in substation defect detection tasks. This paper validates the effectiveness and reliability of the algorithm through comprehensive experiments, highlighting its main contributions and innovations as follows:

1) In the substation defect detection task, The YOLO-SD model achieves an $mAP@0.5$ of 90.3% and $mAP@0.5:0.95$ of 63.9%. The F1 score reaches 81.1%, IoU value is 90.5%, surpassing other models such as YOLOv7 and YOLOv8, reaching the state-of-the-art.
2) To improve the network's ability to extract features in complex backgrounds and enhance detection accuracy, a new deep perceptual feature extraction module called C3+ is introduced in the study. This novel design enhances the backbone network's capability to process feature information across windows, thereby enlarging the network's deep receptive field and preserving essential weight information.
3) In this study, SimAM is integrated into the neck network of YOLO-SD. This incorporation not only strengthens the network's learning capacity but also enables it to autonomously learn and dynamically adjust attention weights to adapt to various input scenarios. Consequently, the neck network can focus on capturing the intricate features of defects in electrical equipment while filtering out irrelevant data. By improving the accuracy of feature fusion, the overall detection accuracy of YOLO-SD is enhanced.
4) A novel fusion loss function, named NWD-CIoU, is specially crafted to address the complexities arising from variations in sizes and shapes of diverse substation equipment defects captured in images. By integrating the NWD-CIoU loss function, the YOLO-SD model's

defect detection capability is fortified to effectively handle targets of different scales, ultimately leading to a significant enhancement in detection accuracy.

However, the model proposed in this paper also encountered challenges such as insufficient data during training, which affected the further improvement of the model's performance.

In future work, we will standardize the evaluation criteria for image generation data to assess the quality of generated images. Building upon this foundation, our focus will be on augmenting the quantity and diversity of the substation defect dataset, laying the groundwork for enhancing the performance of defect detection methods for substation equipments.

## REFERENCES

[1] A. R. Abbasi and D. Baleanu, "Recent developments of energy management strategies in microgrids: An updated and comprehensive review and classification," *Energy Convers. Manage.*, vol. 297, Dec. 2023, Art. no. 117723, doi: 10.1016/j.enconman.2023.117723.

[2] Y. Cui, X. Huang, X. Zhang, J. Ye, and L. Zhong, "A defects detection system for substation based on YOLOX," in *Proc. IEEE 5th Int. Electr. Energy Conf. (CIEEC)*, May 2022, pp. 4703–4707, doi: 10.1109/CIEEC54735.2022.9846606.

[3] A. Abbasi and A. Seifi, "Fast and perfect damping circuit for ferroresonance phenomena in coupling capacitor voltage transformers," *Electric Power Compon. Syst.*, vol. 37, no. 4, pp. 393–402, Mar. 2009, doi: 10.1080/15325000802548780.

[4] J. Liu, M. Hu, J. Dong, and X. Lu, "Summary of insulator defect detection based on deep learning," *Electric Power Syst. Res.*, vol. 224, Nov. 2023, Art. no. 109688, doi: 10.1016/j.epsr.2023.109688.

[5] L. Khalayli, H. Al Sagban, H. Shoman, K. Assaleh, and A. El-Hag, "Automatic inspection of outdoor insulators using image processing and intelligent techniques," in *Proc. IEEE Electr. Insul. Conf. (EIC)*, Jun. 2013, pp. 206–209, doi: 10.1109/EIC.2013.6554234.

[6] A. R. Abbasi and C. P. Gandhi, "A novel hyperbolic fuzzy entropy measure for discrimination and taxonomy of transformer winding faults," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–8, 2022, doi: 10.1109/TIM.2022.3212522.

[7] T. Jabid and Md. Z. Uddin, "Rotation invariant power line insulator detection using local directional pattern and support vector machine," in *Proc. Int. Conf. Innov. Sci., Eng. Technol. (ICISET)*, Oct. 2016, pp. 1–4, doi: 10.1109/ICISET.2016.7856522.

[8] H. He, D. Luo, W.-J. Lee, Z. Zhang, Y. Cao, and T. Lu, "A contactless insulator contamination levels detecting method based on infrared images features and RBFNN," *IEEE Trans. Ind. Appl.*, vol. 55, no. 3, pp. 2455–2463, May 2019, doi: 10.1109/TIA.2018.2889835.

[9] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 779–788, doi: 10.1109/CVPR.2016.91.

[10] D. Sadykova, D. Pernebayeva, M. Bagheri, and A. James, "IN-YOLO: Real-time detection of outdoor high voltage insulators using UAV imaging," *IEEE Trans. Power Del.*, vol. 35, no. 3, pp. 1599–1601, Jun. 2020, doi: 10.1109/TPWRD.2019.2944741.

[11] Z. Huang, W. Xie, W. Liu, C. Song, Y. Zhang, and J. Jing, "TSCDNet+: A highly robust substation anomaly detection method," *Optik*, vol. 246, Nov. 2021, Art. no. 167808, doi: 10.1016/j.ijleo.2021.167808.

[12] A. Bochkovskiy, C.-Y. Wang, and H.-Y. Mark Liao, "YOLOv4: Optimal speed and accuracy of object detection," 2020, *arXiv:2004.10934*.

[13] T. Liu, G. Li, and Y. Gao, "Fault diagnosis method of substation equipment based on you only look once algorithm and infrared imaging," *Energy Rep.*, vol. 8, pp. 171–180, Oct. 2022, doi: 10.1016/j.egyr.2022.05.074.

[14] C. Parkash and A. R. Abbasi, "Transformer's frequency response analysis results interpretation using a novel cross entropy based methodology," *Sci. Rep.*, vol. 13, no. 1, p. 6604, Apr. 2023, doi: 10.1038/s41598-023-33606-0.

[15] K. Rahmani, F. Kavousifard, and A. Abbasi, "Consideration effect of wind farms on the network reconfiguration in the distribution systems in an uncertain environment," *J. Experim. Theor. Artif. Intell.*, vol. 29, no. 5, pp. 995–1009, Sep. 2017, doi: 10.1080/0952813x.2016.1270359.

[16] A. R. Abbasi and M. Mohammadi, "Probabilistic load flow in distribution networks: An updated and comprehensive review with a new classification proposal," *Electric Power Syst. Res.*, vol. 222, Sep. 2023, Art. no. 109497, doi: 10.1016/j.epsr.2023.109497.

[17] N. Sharma, R. Sharma, and N. Jindal, "Machine learning and deep learning applications—A vision," *Global Transitions Proc.*, vol. 2, no. 1, pp. 24–28, Jun. 2021, doi: 10.1016/j.gltp.2021.01.004.

[18] P. P. Shinde and S. Shah, "A review of machine learning and deep learning applications," in *Proc. 4th Int. Conf. Comput. Commun. Control Autom. (ICCUBEA)*, Aug. 2018, pp. 1–6, doi: 10.1109/ICCUBEA.2018.8697857.

[19] M. Fang, T. Sun, and Z. Shao, "Fast helmet-wearing-condition detection based on improved YOLOv2," *Opt. Precis. Eng.*, vol. 27, no. 5, pp. 1196–1205, May 2019, doi: 10.3788/ope.20192705.1196.

[20] W. Zhu, Y. Shu, and S. Liu, "Power grid field violation recognition algorithm based on enhanced YOLOv5," *J. Phys., Conf.*, vol. 2209, no. 1, Feb. 2022, Art. no. 012033.

[21] X. Qian, X. Wang, S. Yang, and J. Lei, "LFF-YOLO: A YOLO algorithm with lightweight feature fusion network for multi-scale defect detection," *IEEE Access*, vol. 10, pp. 130339–130349, 2022, doi: 10.1109/ACCESS.2022.3227205.

[22] X. Bao and S. Wang, "Survey of object detection algorithm based on deep learning," *Transducer Microsyst. Technol.*, vol. 41, no. 4, pp. 5–9, 2022, doi: 10.13873/J.1000-9787(2022)04-0005-05.

[23] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," 2013, *arXiv:1311.2524*.

[24] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1440–1448.

[25] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 28, Jan. 2016, pp. 1–9.

[26] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single shot multibox detector," in *Proc. Eur. Conf. Comput. Vis.* Amsterdam, The Netherlands: Springer, 2016, pp. 21–37.

[27] Y. Shao, D. Zhang, H. Chu, X. Zhang, and Y. Rao, "A review of YOLO object detection based on deep learning," *J. Electr. Inf. Technol.*, vol. 44, no. 10, pp. 3697–3708, Oct. 2022, doi: 10.11999/JEIT210790.

[28] C.-Y. Wang, A. Bochkovskiy, and H.-Y. Mark Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," 2022, *arXiv:2207.02696*.

[29] N. Zhang, G. Yang, F. Hu, H. Yu, J. Fan, and S. Xu, "A novel adversarial deep learning method for substation defect generation," *Preprints*, 2024, Art. no. 2024051016, doi: 10.20944/preprints202405.1016.v1.

[30] G. Jocher. (2020). *YOLOv5*. [Online]. Available: https://github.com/ultralytics/yolov5

[31] S. Qiao, L.-C. Chen, and A. Yuille, "DetectoRS: Detecting objects with recursive feature pyramid and switchable atrous convolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 10208–10219, doi: 10.1109/CVPR46437.2021.01008.

[32] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.

[33] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 9992–10002, doi: 10.1109/ICCV48922.2021.00986.

[34] Q. Hou, D. Zhou, and J. Feng, "Coordinate attention for efficient mobile network design," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 13708–13717, doi: 10.1109/CVPR46437.2021.01350.

[35] L. Yang, R. Zhang, L. Li, and X. Xie, "SimAM: A simple, parameter-free attention module for convolutional neural networks," in *Proc. 38th Int. Conf. Mach. Learn.*, vol. 139, 2021, pp. 11863–11874.

[36] J. Wang, C. Xu, W. Yang, and L. Yu, "A normalized Gaussian Wasserstein distance for tiny object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 1–12.

[37] Z. Zheng, P. Wang, D. Ren, W. Liu, R. Ye, Q. Hu, and W. Zuo, "Enhancing geometric factors in model learning and inference for object detection and instance segmentation," *IEEE Trans. Cybern.*, vol. 52, no. 8, pp. 8574–8586, Aug. 2022, doi: 10.1109/TCYB.2021.3095305.
[38] G. Jocher, A. Chaurasia, and J. Qiu. (2023). *Ultralytics YOLO (Version 8.0.0)*. Computer Software. [Online]. Available: https://github.com/ultralytics/ultralytics

**NA ZHANG** received the M.S. degree from Harbin Institute of Technology, Harbin, China, in 2015. She is currently an Engineer with the State Grid Shanxi Electric Power Research Institute, Taiyuan, China. Her research interest includes image feature recognition technology of electrical equipment.

**GANG YANG** received the Ph.D. degree from Beijing Jiaotong University, Beijing, China, in 2013. He is currently a Senior Engineer with the State Grid Shanxi Electric Power Research Institute, Taiyuan, China. His research interests include image generation and screening technologies of electrical equipment.

**DAWEI WANG** received the M.S. degree from China University of Mining and Technology, Beijing, China, in 2015. He is currently an Engineer with the State Grid Shanxi Electric Power Research Institute, Taiyuan, China. His research interest includes fault diagnosis technology of electrical equipment based on images.

**FAN HU**, photograph and biography not available at the time of publication.

**HUA YU**, photograph and biography not available at the time of publication.

**JINGJING FAN**, photograph and biography not available at the time of publication.