sustainability

MDPI

Article

Anomaly Perception Method of Substation Scene Based on High-Resolution Network and Difficult Sample Mining

Yunhai Song [1], Sen He [1], Liwei Wang [1], Zhenzhen Zhou [1], Yuhao He [1], Yaohui Xiao [1], Yi Zheng [2,*] and Yunfeng Yan [2]

[1] Overhaul and Test Center of UHV Transmission Company of China Southern Power Grid Co., Ltd., Guangzhou 510663, China; songyunhai@im.ehv.csg (Y.S.); hesen@im.ehv.csg (S.H.); wangliwei@im.ehv.csg (L.W.); 13602826956@139.com (Z.Z.); heyuhao@im.ehv.csg (Y.H.); xiaoyaohui@im.ehv.csg (Y.X.)

[2] College of Electrical Engineering, Zhejiang University, Hangzhou 310027, China; 21210004@zju.edu.cn

* Correspondence: zhengyipds@zju.edu.cn

Abstract: The perception of anomalies in power scenarios plays a crucial role in the safe operation and fault prediction of power systems. However, traditional anomaly detection methods face challenges in identifying difficult samples due to the complexity and uneven distribution of power scenarios. This paper proposes a power scene anomaly perception method based on high-resolution networks and difficult sample mining. Firstly, a high-resolution network is introduced as the backbone for feature extraction, enhancing the ability to express fine details in power scenarios and capturing information on small target anomaly regions. Secondly, a strategy for mining difficult samples is employed to focus on learning and handling challenging and hard-to-recognize anomaly samples, thereby improving the overall anomaly detection performance. Lastly, the method incorporates GIOU loss and a flexible non-maximum suppression strategy to better adapt to the varying sizes and dense characteristics of power anomaly targets. This improvement enables higher adaptability in detecting anomalies in power scenarios. Experimental results demonstrate significant improvements in power scene anomaly perception and superior performance in handling challenging samples. This study holds practical value for fault diagnosis and safe operation in power systems.

Keywords: high-resolution network; difficult sample mining; power scene anomaly perception; deep learning

check for updates

Citation: Song, Y.; He, S.; Wang, L.; Zhou, Z.; He, Y.; Xiao, Y.; Zheng, Y.; Yan, Y. Anomaly Perception Method of Substation Scene Based on High-Resolution Network and Difficult Sample Mining. *Sustainability* **2023**, *15*, 13721. https://doi.org/10.3390/su151813721

Academic Editors: Ling Xiang and Xiaoan Yan

Received: 18 July 2023
Revised: 16 August 2023
Accepted: 6 September 2023
Published: 14 September 2023

Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

With the continuous construction and development of power systems in China, ensuring the safety and efficiency of the transmission and transformation process is particularly important for the safe and stable operation of the power grid [1]. China has a vast territory and numerous cities, with substations often distributed in sparsely populated areas and densely populated and complex towns. In addition, the complexity of the substation system itself increases the risk of abnormal transmission lines [2–5]. Traditional power inspection tasks mainly rely on manual inspection, which has low execution efficiency and is easily affected by weather. A significant drawback of manual inspection is that the inspection process requires maintenance personnel to climb to high places and rely on visual and personal subjective experience for inspection [6]. This requires extremely high physical strength and experience for inspection personnel and poses a high risk of safety accidents.

The question of how to ensure the safe inspection of the substation system has become an urgent problem to be solved. Considering the limitations of traditional manual inspection methods, there is an urgent need to introduce innovative methods and technologies to improve the inspection efficiency and accuracy. By applying advanced computer vision and deep learning technology, automated monitoring and anomaly detection of substation equipment can be achieved, thereby improving the safety and efficiency of inspection.

With the rapid development of deep learning in recent years, many universal deep learning algorithms on object detection have emerged [7–13]. Choosing a suitable detection algorithm is crucial in ensuring the normal operation and safety of substation equipment defect detection. The detection algorithm is mainly divided into a one-stage algorithm and a two-stage algorithm. The one-stage detection algorithm has the advantages of real-time performance and simplicity, and it is suitable for scenarios that require a fast response. However, the one-stage detection algorithm may have certain limitations in terms of positioning accuracy and processing small targets. In contrast, two-stage detection algorithms have advantages in identifying accuracy and locating equipment defects. By generating candidate boxes and fine-tuning them, the two-stage algorithm can achieve more accurate defect localization. In practical application scenarios, substations, which require high accuracy, have significant advantages in the high recognition efficiency of two-stage algorithms. This helps to quickly detect and repair equipment defects, ensuring the normal operation and safety of substations. In the detection of substation equipment defects, this paper adopts Faster R-CNN [13] as the basis for targeted modifications.

The feature extraction methods for substation anomaly scenes based on deep learning are flexible and efficient. One type of method is based on encoding and decoding structures, such as U-Net [14] and UNet++ [15]. Another type of method is based on hollow convolutional structures, such as PSPNet [16] and DeeplabV3+ [17]. However, the above methods can easily lead to the loss of spatial details near the boundaries and monitored small targets. Shallow networks are unable to effectively distinguish different features, resulting in the information of small target abnormal areas of some substation equipment or personnel being directly ignored by neural networks. To overcome the above problems, this paper introduces high-resolution representation learning (HRNet) to capture abnormal features of substation equipment and personnel behavior [18]. HRNet achieves multi-scale feature fusion by designing high-resolution and low-resolution sub-networks and connecting them in parallel. The above characteristics can ensure that the proposed algorithm can better capture the details of different scales and small targets in the substation scene image, and this can comprehensively improve the detection and recognition ability of equipment defects.

In the actual monitoring process of substations, the image environment and equipment attributes of the monitored equipment would be constantly affected by external factors (such as weather, lighting conditions), and the trained deep learning image recognition model will experience concept drift leading to model recognition bias, reduced generalization performance, and missed and false detections [19]. Therefore, it is necessary to update the identified network online and adjust the target samples that require focused learning through deviation compensation. The current practical method of detecting defects in substation equipment is OHEM [20], which can help diagnostic models to flexibly deal with difficult samples in substation defect detection by means of online difficult sample mining. Therefore, this paper proposes an improved OHEM algorithm, and experimental results show that it can effectively improve the recognition efficiency of defect detection models.

This paper optimizes the regression loss function for model training by applying the GIOU loss function [21], which can better deal with power equipment defect detection, and uses Soft NMS in the reasoning phase to solve the problem whereby adjacent foreground targets are incorrectly screened [22]. The main contributions of this paper regarding the issue of defect detection in power equipment are as follows.

- The diagnostic network proposed is based on the two-stage model of Faster R-CNN [13], which improves the backbone network with high-resolution representations [18]. A high-frequency feature extractor is designed for power equipment defect images, maintaining high-resolution feature representation throughout detection. This structure supplements low-resolution feature representation through multi-stage parallel subnetworks to improve the feature learning performance for small and medium-sized target objects.
- A difficult sample mining strategy is designed in the second stage of the Faster R-CNN model. By mining difficult samples online and dynamically adjusting the training

samples, the model learns and processes difficult-to-identify defects and abnormal samples, thereby improving the anomaly detection performance.

- GIOU loss is applied as the regression loss function of Faster R-CNN. This loss function can better deal with the situation of different power anomaly targets, which is helpful for model training.
- Soft NMS is adopted to improve the inference strategy of the regression framework, solving the problem of adjacent foreground targets being mistakenly screened and reducing the missed detection rate of abnormal categories in substation scenarios.

2. Related Works

2.1. Object Detection Algorithms Based on Deep Learning

Target detection algorithms can usually be divided into one-stage and two-stage algorithms. Common one-stage algorithms include YOLO [8], SSD [9], and RetinaNet [10]. YOLO proposed the basic framework of a one-stage algorithm to solve the problem of poor real-time performance in object detection. SSD is a remarkable object detection structure after YOLO, which adopts the direct regression boundary box and classification probability methods in YOLO and draws inspiration from the ideas of Faster R-CNN, widely using anchors to improve the recognition accuracy. RetinaNet further proposed the focal loss structure, which improves the robustness and accuracy of single-stage algorithms. Differently, the two-stage algorithm first provides candidate boxes for the monitored targets, and then classifies the targets within the candidate boxes. Typical methods in the two-stage algorithm family include R-CNN [11], Fast R-CNN [12], and Faster R-CNN [13]. Among them, Faster R-CNN proposed an RPN structure, which solved the problem of slow candidate box generation in R-CNN networks and improved the accuracy of the algorithm. In summary, object detection models based on one-stage algorithms have advantages in terms of training speed, while object detection models based on two-stage algorithms have advantages in detection accuracy.

For safety inspection tasks in substation scenarios, the detection speed of the model based on the two-stage algorithm [13] can already meet the requirements of daily inspections. Due to the significant differences in the difficulty of detecting various defects in substations, it is necessary to improve the detection accuracy of difficult samples. Therefore, this study adopts the target detection algorithm framework of the two-stage algorithm, utilizes the RPN module in the one-stage algorithm to calculate candidate regions, and designs a difficult sample mining method based on the two-stage algorithm to improve the detection accuracy of difficult samples.

2.2. Substation Equipment and Behavior Anomaly Detection Based on Deep Learning

In recent years, more and more research has applied deep learning algorithms to power anomaly detection. Fang et al. [23] proposed a Fast R-CNN framework called angle enhancement. They proposed a fine-grained feature extractor to extract features related to substation defects, which can effectively identify small features in local areas of defects. Common feature extractors include Resnet [23], MobileNet [24], Darknet [25], high-resolution networks [18], etc. Resnet utilizes the residual module of neural networks to solve the problem of decreased learning ability after the deep improvement of the network. MobileNets uses deep separable convolution and introduces 1×1 convolution to solve the problem of information non-circulation between channels. A high-resolution network maintains a high-resolution representation throughout the entire forward process and connects multi-resolution subnets to preserve more high-dimensional information during feature extraction. Therefore, this paper improves the feature extractor of Faster R-CNN, transforming the original feature extractor into a high-resolution network to enhance the ability to express details in power scenes and capture small target abnormal area information.

Abdusalomov et al. used the OHEM method for detecting gantry crane tilting, pulling, and hanging, with a focus on mining difficult samples in training [26]. In addition, the

most commonly used method is random sampling. Cascade R-CNN only samples high-quality predictions for training, and the detector is more prone to over-fitting due to the disappearance of positive samples. Cascade regression is proposed to avoid over-fitting during training and quality mismatch during inference [27]. SAPD implements joint sampling on pyramid-shaped feature maps and solves the problem of low efficiency in utilizing positive and negative samples in the model through weighted combination [28]. PISA uses an import-based re-weighting (ISR) sampling strategy to improve the utilization efficiency of positive and negative samples [29]. It is not difficult to find that there are significant differences in the difficulty of identifying different types of substation defects. Therefore, this paper adds an improved difficult sample mining strategy to the Faster R-CNN model, focusing on learning and processing difficult-to-identify abnormal samples to improve the overall anomaly detection performance.

Regarding the post-processing part of the detection algorithm, Chen et al. selected the anchor size to determine the candidate size suitable for the distribution of the substation defect data [30]. The improvement of the post-processing stage can also start with the loss function. Target detection can be regarded as a classification task plus a location task, and both tasks have their own loss functions. The loss functions of classified tasks include CE loss, focal loss [10], GHMC loss [31], etc. Focal loss strengthens the value of hard-to-divide negative samples in the loss function, thus solving the problem that a large number of easy-to-divide negative samples is not conducive to model learning. The loss functions of location tasks include IOU loss [32], GHW-R loss [33], DIOU loss [34], GIOU [21], etc. GIOU loss solves the problem that the loss function value of IOU loss is small when the anchor frames contain each other, and it improves the accuracy of the model. Therefore, this paper proposes a post-processing strategy based on GIOU Soft NMS to address the characteristics of the varying sizes and dense distribution of abnormal targets in power scenarios, resulting in higher adaptability when dealing with abnormal targets in power scenarios.

3. Algorithm for Anomaly Detection of Substation Equipment and Personnel

This paper constructs a two-stage power equipment anomaly detection model based on Faster R-CNN and combines the characteristics of power scene data to design a detection algorithm that is suitable for business scenarios. The overall algorithm is shown in Figure 1, which mainly consists of a high-resolution feature extractor, a region proposal network (RPN) module [13], and a difficult sample mining network.

- Substation power equipment defect feature extraction network: This is used to map images from pixel space to feature space containing positional and semantic information, which is input to subsequent RPN and fully connected layers.
- Region proposal network: Each location on the feature map input to the layer corresponds to an area of the original map. To detect more objects, for each point on the feature map, nine candidate boxes can be obtained in the original image area through three different scale transformations of anchor boxes. Subsequently, each candidate box is binary classified through the Softmax layer to determine whether it belongs to the foreground and labeled. Then, we use regression algorithms to adjust the position and size of the foreground candidate boxes, which can be close to the real foreground area. Finally, we send the adjusted foreground boxes to the next layer of the network.
- Difficult sample mining network: This paper adopts an improved difficult sample mining strategy in the ROI layer and classification detection layer of Faster R-CNN. We calculate difficult samples during each training step and perform gradient updates according to the appropriate proportion of positive and negative samples. Please refer to Section 3.2 for details.

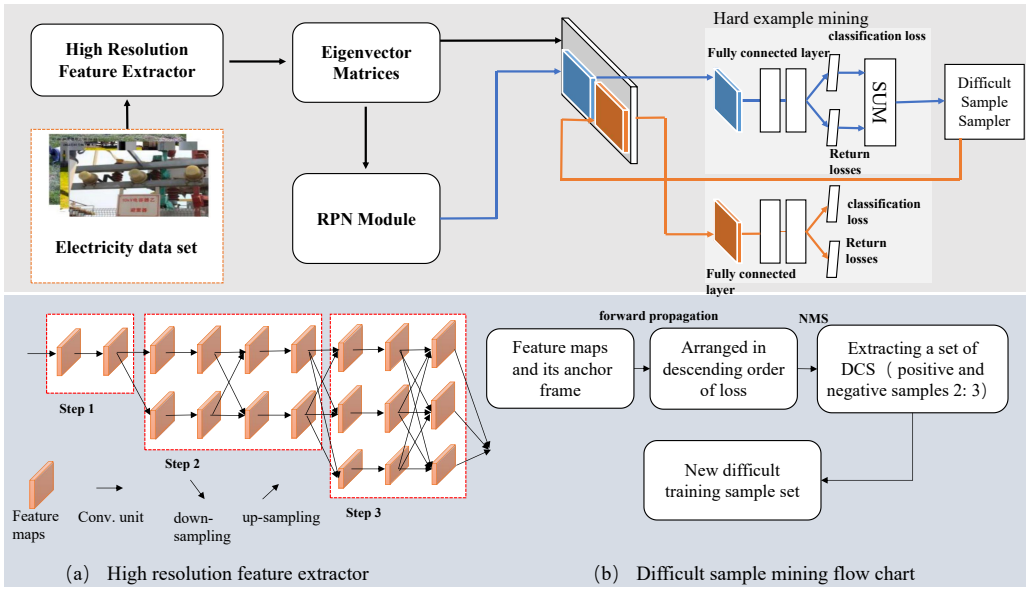


Figure 1. Overall algorithm flowchart. Complex substation image features are captured by the high-resolution feature extractor. The RPN module is used to obtain the candidate boxes of the object. Finally, regression and detection are completed through the strategy of difficult sample mining.

3.1. High-Frequency Defect Feature Extraction Module

In response to the problem of the difficult extraction of abnormal feature points in the dataset of power equipment defect samples, in order to prevent convolution from extracting high-frequency detail information, this paper improves the feature extractor of the original Faster R-CNN model and adopts a high-frequency feature extraction module with a high-resolution network structure. Its advantage is that it can ensure high-level semantic learning while also maintaining the high-resolution information of the feature map.

The structure of the proposed high-frequency feature extraction module is shown in Figure 1a. Based on the actual needs of power scene detection, the proposed high-resolution network feature extraction module adopts parallel and densely interconnected multi-resolution branches to extract features at three different scales through three stages. At each stage, feature fusion is performed between different branches. The down-sampling mainly uses a 3×3 convolutional kernel with a step size of 2. Up-sampling adopts bilinear interpolation operation. This structure can ensure that the high-frequency detail information of the monitored image is not lost. The proposed high-frequency feature detection module utilizes the multi-stage fusion of high-level semantic information at different scales, enhancing the representation ability of features.

3.2. Difficult Defect Sample Mining Network

The proposed difficult defect sample mining network is shown in Figure 1, which consists of region of interest (ROI) pooling and a classification regression layer.

ROI pooling layer: The input of this layer is the feature map corresponding to the foreground candidate boxes of different sizes filtered by the RPN layer. This layer divides feature maps of different sizes into fixed size grids, and then performs pooling operations to obtain fixed size feature maps. This can avoid using cropping and scaling operations, thus effectively preserving the original shape of the image.

Classification regression layer: The feature matrix output by the ROI pooling layer is processed through two fully connected layers (as shown in the fc layer in Figure 1) to obtain a feature representation containing category and position information. We utilize these features to determine the categories of candidate regions and adjust the positions of borders, in order to achieve the accurate localization of target detection.

The proposed diffusion defect sample mining network has two ROI and classification regression layers with the same structure. This structure only performs forward inference

without parameter updates. In one iteration, the forward propagation results of the upper layer are fed into the difficulty sample sampler to obtain a fixed proportion of positive and negative difficulty samples, and then these samples are backpropagated to the lower structure to update the gradient. The sampling strategy of the difficult sample sampler is shown in Figure 1b.

3.3. Post-Processing Optimization

In the inference process of Faster R-CNN, the traditional non-maximum suppression (NMS) algorithm is used. The traditional NMS algorithm first arranges each foreground target in descending order according to category confidence during the calculation process; it then selects the boundary box with the highest confidence as the reserved box, and then removes the remaining boundary boxes with high overlap with the reserved box based on a fixed intersection to union ratio (IOU) threshold. As adjacent targets often appear in substation datasets, the use of traditional NMS algorithms can lead to missed detections. Therefore, this paper introduces Soft NMS [35] to replace the original NMS algorithm. As shown in Equation (1), when the IOU of the remaining box and the maximum confidence box exceeds the threshold, this paper uses a Gaussian weight function for score attenuation.

$$S_i = \begin{cases} S_i & iou(b_{max}, b_i) < N \\ S_i e^{-\frac{iou(b_{max}, b_i)^2}{\sigma}} & iou(b_{max}, b_i) \geqslant N \end{cases} \quad (1)$$

where b_{max} is the detection box with the highest confidence in the boxes, b_i represents the i-th detection box, S_i represents the confidence score corresponding to the i-th type of detection box, and N represents the manually set threshold.

In addition, the traditional benchmark model Faster R-CNN uses smooth $L1$ loss as the loss function. There are two drawbacks to this method of calculating loss based on the coordinates of four vertices of the detection box.

- Not utilizing the correlation between vertices;
- There are targets of different sizes in the substation scene, and vertex coordinate offsets of the same size can cause completely different IOUs for targets of different sizes.

Relatively, GIOU loss can measure the detection effect by combining the shape size and overlapping area of the detection box. As shown in Figure 2, GIOU loss is introduced to replace smooth $L1$ loss as the regression loss function [21]. We find the convex shape C enclosing both the candidate box (represented by A) and the ground truth (represented by B). Then, we calculate the ratio between the area occupied by C excluding A and B and divide it by the total area occupied by C. Finally, GIOU is attained by subtracting this ratio from the IOU value.

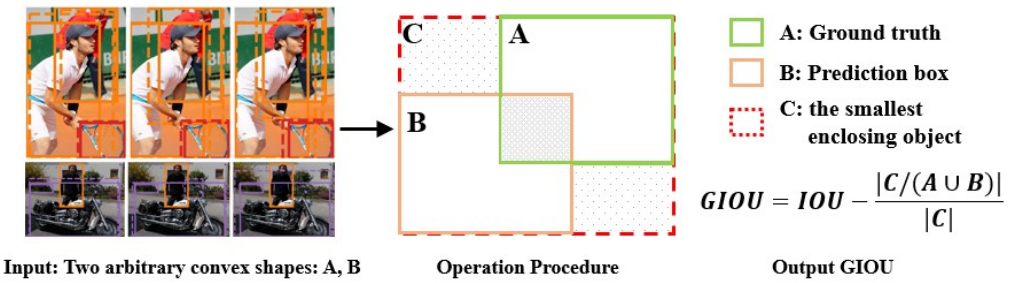


Figure 2. GIOU loss. After the candidate box is obtained by the RPN module, the smallest enclosing convex shape C and the area occupied by C excluding A and B are found. GIOU is attained according to the two values and the IOU value.

3.4. Discussion

In essence, our method is a two-stage detection algorithm that fuses high-resolution feature extraction and difficult sample mining, and it performs well in the anomaly perception of substation scenes. However, the complex model structure also brings great

challenges to the training process. In addition, due to the strategy of difficult sample mining in model training, the detection results of some samples show over-fitting (see Figure 3). Therefore, the question of how to design a more flexible and efficient defect detection algorithm will need further exploration.

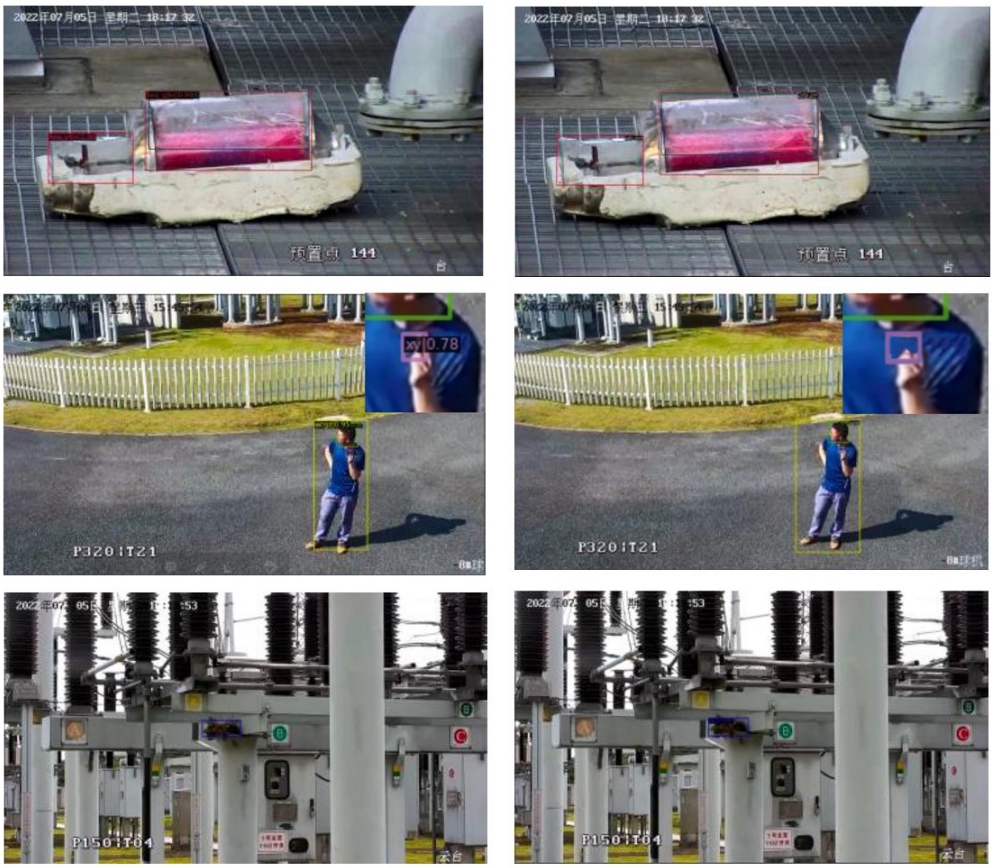


Figure 3. Comparison between model prediction and real labels (the left column shows the model prediction results, and the right column shows the truth labels).

4. Experimental Results and Analysis

4.1. Datasets

The dataset used in this paper comes from real scene inspection records of substations, including three categories: power equipment defects, abnormal personnel behavior, and abnormal equipment status. Among them, the defects of the power equipment include a damaged respirator oil seal, a blurred dial, a damaged dial, a broken insulator, ground oil stains, a damaged silicone tube, abnormal closure of the box door, hanging suspended solids, and a damaged bird's nest cover plate. Abnormal personnel behavior involves smoking, not wearing work clothes, and not wearing safety helmets. Abnormal equipment status involves abnormal oil levels of the respirator oil seal and discoloration of silicone. Each category contains approximately 200 images, totaling 2664. We randomly select 2364 pieces as the training set and the remaining 300 pieces as the test set.

4.2. Evaluation Indicators

The detection precision and recall rate can be used as basic indicators to measure the defect detection performance, and these two indicators can be expressed as

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall = \frac{Tp}{TP + FN} \tag{3}$$

$$MAP = \frac{1}{c}\sum_{1}^{c}\int_{0}^{1} p(r)dr \tag{4}$$

where TP is a positive sample predicted by the model as a positive sample; FP is a negative sample predicted by the model as a positive sample; FN is a positive sample predicted by the model as a negative sample. The map can be obtained by calculating the average accuracy of all classes. The proposed model in this article is a multi-target detection model that requires the evaluation of the classification and localization of multiple targets, using the mean accuracy precision (MAP) as the evaluation indicator.

4.3. Experimental Result Analysis

4.3.1. Ablation experiment

This paper modifies the traditional two-stage detection model in line with the power business scenario. To verify the actual effectiveness of the improved algorithm, ablation experiments are conducted on each improvement step. As shown in Table 1, this paper has improved the backbone network structure, positive and negative sample mining strategy, regression loss function, and detection frame de-duplication strategy in the testing phase.

Table 1. Results of ablation experiments.

Model	Feature Extraction	Sample Mining	GIOU	Soft NMS	MAP/% (IOU = 0.5)	MAP/% (IOU = 0.75)	MAP/% (IOU = 0.95)
Faster R-CNN	–	–	–	–	78.2	62.1	56.2
Test 1	✓	–	–	–	79.0	63.1	56.9
Test 2	✓	✓	–	–	81.3	64.5	57.8
Test 3	✓	✓	✓	–	81.3	66.7	58.6
Test 4	✓	✓	✓	✓	80.7	67.8	59.2

4.3.2. Supplementary Experiment on High-Frequency Feature Extraction Module

This paper selects a high-resolution network as the backbone network, which can better preserve high-frequency feature information during the feature extraction process of images, as the more comprehensive high-frequency features can better represent small target objects and objects with less obvious features. As shown in Table 2, this study conducted supplementary experiments specifically on the detection accuracy of small target areas and the detection accuracy of categories with unclear features. The truth box labeled < 32 represents all labels with a length and width less than 32, while small represents smoking, an anomaly category that only occupies a very small number of pixels. It can be observed that after using high-resolution networks, the detection model achieved an accuracy improvement of 2.8% in small target areas and 4.2% in small target areas.

Table 2. Comparative experiments on backbone networks.

Model	MAP/% (<32)	MAP/% (Small)
Baseline	45.7	10.5
Baseline + high-resolution networks	48.5	14.7

4.3.3. Visualization of Model Output Results

The visualization of the prediction results of the model in several typical samples in this paper is shown in Figure 3.

The images on the first row include three types: silicone tube damage, silicone discoloration, and respirator oil seal damage. Although there is overlap between the target

boxes of silicone tube damage and silicone discoloration, the detection model can still accurately distinguish and locate these two different categories. The images on the second row indicate that the substation staff did not wear safety helmets and smoked. It can be seen that the detection model can better locate the human body area (see the yellow border), head area (see the green border), and cigarette butts (see the purple border). The image on the third row illustrates bird nest detection, where it can be seen that the bird nest is located in a small area in the middle of the image. The detection model can accurately recognize and locate the bird nest.

4.3.4. Feature Visualization

The features extracted through high-frequency feature extraction networks are input into the RPN layer, which is divided into P2–P6 layers according to different resolutions. This paper visualizes these four layers of features, as shown in Figure 4.

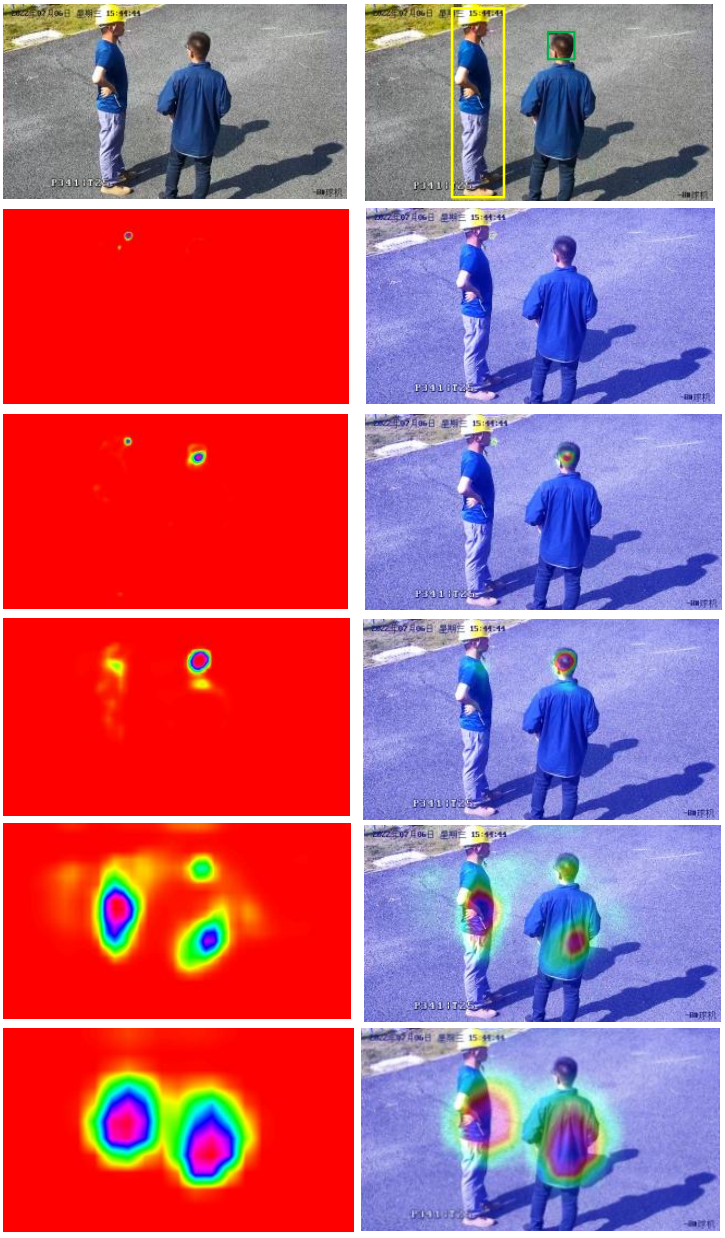


Figure 4. Visualization of high-resolution feature extraction results at different layers. With the deepening of the layers, the perception ability of small target object features is continuously improved. Among them, the blue and green rings represent the feature regions noticed by the model.

The first row is the original image and detection results. In Figure 4, the abnormal categories include individuals without helmets (see green box), smoking (see purple box), and smoking (see yellow box). The second to sixth rows represent the thermal maps of the four different scale feature layers P2–P6. It can be seen that the P2 layer is a high-frequency detail feature layer, which focuses on the small target object of a cigarette butt. The P3 layer focuses on the cigarette butt and safety helmet, while the P4 layer focuses on the human body and safety helmet. The P5 and P6 layers have lost the high-frequency feature information of the small target objects, mainly focusing on the large target object of the human body. The features of the RPN layer depend on the high-frequency feature extraction network based on high-resolution networks proposed in this paper. From the visualization graph, it can be seen that the high-frequency feature extraction network can effectively preserve high-frequency detail information during the feature extraction process, and it can extract information about extremely small target objects—for example, cigarette butts.

4.3.5. Visualization of Truth Box Length–Width Distribution

This paper visualizes the length and width distribution of real labels in the dataset, as shown in Figures 5 and 6. This paper uses K-means clustering to adjust the size settings of the anchor, which are [17, 14], [59, 61], [110, 108], [162, 177], [326, 128], [263, 277], [595, 265], [378, 465], [622, 615].

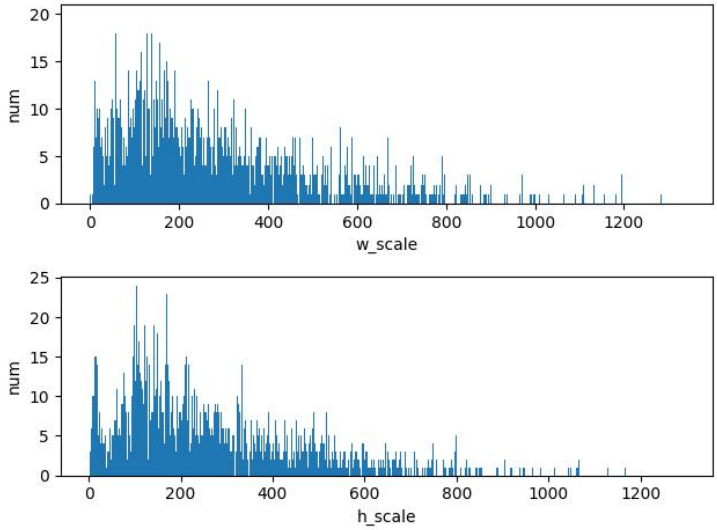


Figure 5. The width and height distribution of real labels in the dataset. Most of the bounding boxes are small objects with sizes of less than 200.

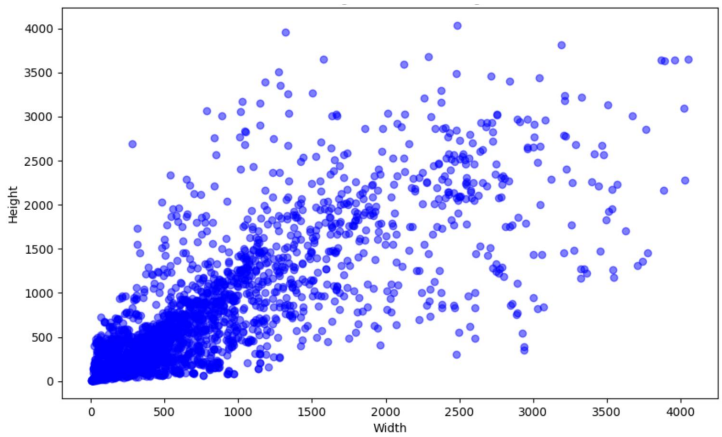


Figure 6. Statistics of length and width distribution. Each point represents a label, and its position in the figure reflects the size of the target object.

5. Conclusions

High-precision scene anomaly detection algorithms play a crucial role in the intelligent inspection process of substations. However, the existing detection of abnormal scenes in substations has problems such as complex actual operation and maintenance environments, imbalanced sample data, and significant differences in detection difficulty. In order to overcome the shortcomings of existing object detection methods in the aforementioned problems, this paper proposes a power scene anomaly perception method based on high-resolution networks and difficult sample mining. Firstly, a high-resolution network is introduced as the backbone network for feature extraction, preserving high-frequency information during the down-sampling process of the module. Secondly, in the second stage of the detection model proposed in this paper, an improved difficult sample mining strategy is adopted to focus on learning difficult samples and maintain a balance between positive and negative samples. Finally, the GIOU loss function and flexible non-maximum suppression strategy are adopted to further improve the detection effect, which are more suitable for electricity business scenarios. The experimental results show that compared to the universal detection model, the proposed method in this paper has significantly improved the average accuracy of substation defect detection compared to the Fast R-CNN method. In particular, the detection accuracy of difficult samples with a width and length less than 32 is improved by 2.8%, and the detection accuracy of very small objects such as smoking is improved by 4.2%.

For future perspectives on abnormal perception tasks in complex power scenes, the research and development of defect detection algorithms with high precision, strong stability, and excellent portability will be a research hotspot. Moreover, with the continuous improvement of detection methods in power scenes, the construction of smart grids will also develop towards the following trends.

- It is expected to develop effective detection algorithms for small-sample electricity data, use a small number of images to realize anomaly perception, and further solve problems such as the difficult labeling of training samples and poor migration ability.
- A large number of positive sample images are accumulated based on UAV inspection, which provides a strong data basis for unsupervised defect recognition. The issue of how to effectively learn the feature distribution of positive samples and effectively solve the problem of many types of defects and few negative samples according to the feature distribution distance between the detection sample and positive samples still needs further exploration.

Author Contributions: Methodology, Y.S. and Y.Y.; software, Y.Z.; validation, Y.H. and L.W.; data curation, Y.X. and Z.Z.; writing—original draft preparation, Y.S.; investigation, S.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded in part by the National Natural Science Foundation of China (Grant No. 6212780029 and 62101490), in part by the Key Technologies R&D Program of Zhejiang Province (Grant No. 2022C01056 and LQ21F030017), in part by the China Southern Power Grid (Grant No. 0120002022030304SJ00015) and in part by Sanya Science and Technology Innovation Project (Grant No. 2022KJCX47).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Ge, L.; Li, Y.; Li, Y.; Yan, J.; Sun, Y. Smart Distribution Network Situation Awareness for High-Quality Operation and Maintenance: A Brief Review. *Energies* **2022**, *15*, 828. [CrossRef]
2. Yan, X.; Liu, T.; Fu, M.; Ye, M.; Jia, M. Bearing fault feature extraction method based on enhanced differential product weighted morphological filtering. *Sensors* **2022**, *22*, 6184. [CrossRef] [PubMed]
3. Chen, H.; Wang, X.b.; Yang, Z.X. Fast robust capsule network with dynamic pruning and multiscale mutual information maximization for compound-fault diagnosis. *IEEE/ASME Trans. Mechatron.* **2022**, *28*, 838–847. [CrossRef]
4. Wang, X.B.; Zhang, X.; Li, Z.; Wu, J. Ensemble extreme learning machines for compound-fault diagnosis of rotating machinery. *Knowl.-Based Syst.* **2020**, *188*, 105012. [CrossRef]
5. Albogamy, F.R. A Hybrid Heuristic Algorithm for Energy Management in Electricity Market with Demand Response and Distributed Generators. *Appl. Sci.* **2023**, *13*, 2552. [CrossRef]
6. Yan, X.; Liu, Y.; Xu, Y.; Jia, M. Multichannel fault diagnosis of wind turbine driving system using multivariate singular spectrum decomposition and improved Kolmogorov complexity. *Renew. Energy* **2021**, *170*, 724–748. [CrossRef]
7. Yan, X.; She, D.; Xu, Y.; Jia, M. Deep regularized variational autoencoder for intelligent fault diagnosis of rotor–bearing system within entire life-cycle process. *Knowl.-Based Syst.* **2021**, *226*, 107142. [CrossRef]
8. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 779–788.
9. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. Ssd: Single shot multibox detector. In Proceedings of the Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, 11–14 October 2016; pp. 21–37.
10. Lin, T.Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal loss for dense object detection. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2980–2988.
11. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 580–587.
12. Girshick, R. Fast R-CNN. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1440–1448.
13. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. *Adv. Neural Inf. Process. Syst.* **2015**, *28*, 91–99. [CrossRef] [PubMed]
14. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, 5–9 October 2015; pp. 234–241.
15. Zhou, Z.; Siddiquee, M.M.R.; Tajbakhsh, N.; Liang, J. Unet++: Redesigning skip connections to exploit multiscale features in image segmentation. *IEEE Trans. Med. Imaging* **2019**, *39*, 1856–1867. [CrossRef] [PubMed]
16. Zhao, H.; Shi, J.; Qi, X.; Wang, X.; Jia, J. Pyramid scene parsing network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2881–2890.
17. Chen, L.C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 801–818.
18. Wang, J.; Sun, K.; Cheng, T.; Jiang, B.; Deng, C.; Zhao, Y.; Liu, D.; Mu, Y.; Tan, M.; Wang, X.; et al. Deep high-resolution representation learning for visual recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *43*, 3349–3364. [CrossRef] [PubMed]
19. Gama, J.; Žliobaitė, I.; Bifet, A.; Pechenizkiy, M.; Bouchachia, A. A survey on concept drift adaptation. *ACM Comput. Surv. (CSUR)* **2014**, *46*, 1–37. [CrossRef]
20. Shi, F.; Qian, H.; Chen, W.; Huang, M.; Wan, Z. A fire monitoring and alarm system based on YOLOv3 with OHEM. In Proceedings of the 2020 39th Chinese Control Conference (CCC), IEEE, Shenyang, China, 27–29 July 2020; pp. 7322–7327.
21. Rezatofighi, H.; Tsoi, N.; Gwak, J.; Sadeghian, A.; Reid, I.; Savarese, S. Generalized intersection over union: A metric and a loss for bounding box regression. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 658–666.
22. Bodla, N.; Singh, B.; Chellappa, R.; Davis, L.S. Soft-NMS–improving object detection with one line of code. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 5561–5569.
23. Hu, L.; Ma, J.; Fang, Y. Defect recognition of insulators on catenary via multi-oriented detection and deep metric learning. In Proceedings of the 2019 Chinese Control Conference (CCC), IEEE, Guangzhou, China, 27–30 July 2019; pp. 7522–7527.
24. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv* **2017**, arXiv:1704.04861.
25. Redmon, J.; Farhadi, A. YOLO9000: better, faster, stronger. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 7263–7271.
26. Abdusalomov, A.; Baratov, N.; Kutlimuratov, A.; Whangbo, T.K. An improvement of the fire detection and classification method using YOLOv3 for surveillance systems. *Sensors* **2021**, *21*, 6519. [CrossRef] [PubMed]
27. Cai, Z.; Vasconcelos, N. Cascade r-cnn: Delving into high quality object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 6154–6162.

28. Zhu, C.; Chen, F.; Shen, Z.; Savvides, M. Soft anchor-point object detection. In Proceedings of the Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, 23–28 August 2020; pp. 91–107.
29. Cao, Y.; Chen, K.; Loy, C.C.; Lin, D. Prime sample attention in object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 11583–11591.
30. Chen, H.; He, Z.; Shi, B.; Zhong, T. Research on recognition method of electrical components based on YOLO V3. *IEEE Access* **2019**, *7*, 157818–157829. [CrossRef]
31. Li, B.; Liu, Y.; Wang, X. Gradient harmonized single-stage detector. In Proceedings of the AAAI Conference on Artificial Intelligence, Honolulu, HI, USA, 29–31 January 2019; Volume 33, pp. 8577–8584.
32. Yu, J.; Jiang, Y.; Wang, Z.; Cao, Z.; Huang, T. Unitbox: An advanced object detection network. In Proceedings of the 24th ACM International Conference on Multimedia, Amsterdam, The Netherlands, 15–19 October 2016; pp. 516–520.
33. Chen, J.; Chen, S.; Chen, X.; Dai, Y.; Yang, Y. CSR-Net: Learning adaptive context structure representation for robust feature correspondence. *IEEE Trans. Image Process.* **2022**, *31*, 3197–3210. [CrossRef] [PubMed]
34. Zheng, Z.; Wang, P.; Liu, W.; Li, J.; Ye, R.; Ren, D. Distance-IoU loss: Faster and better learning for bounding box regression. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; Volume 34, pp. 12993–13000.
35. Wu, J.; Guo, P.; Cheng, Y.; Zhu, H.; Wang, X.-B.; Shao, X. Ensemble generalized multiclass support-vector-machine-based health evaluation of complex degradation systems. *IEEE/ASME Trans. Mechatron.* **2020**, *25*, 2230–2240. [CrossRef]

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