Defect detection method for key area guided transmission line components based on knowledge distillation

Zhenbing Zhao[1,2,3], Xuechun Lv[1,4], Yue Xi[1] and Siyu Miao[1]*

[1]School of Electrical and Electronic Engineering, North China Electric Power University, Baoding, China, [2]Engineering Research Center of Intelligent Computing for Complex Energy Systems, Ministry of Education, North China Electric Power University, Baoding, China, [3]Hebei Key Laboratory of Power Internet of Things Technology, North China Electric Power University, Baoding, China, [4]Institute of Electrical Automation, China Nuclear Power Engineering Co., Ltd., Hebei Branch, Shijiazhuang, China

Introduction: The aim of this paper is to address the problem of the limited number of defect images for both metal tools and insulators, as well as the small range of defect features.

Methods: A defect detection method for key area-guided transmission line components based on knowledge distillation is proposed. First, the PGW (Prediction-Guided Weighting) module is introduced to improve the foreground target distillation region, and the distillation range is precisely concentrated in the position of the first k feature pixels with the highest quality score in the form of a mask. The feature knowledge of defects of hardware and insulators is used as the focus for the teacher network to guide the student network. Then, the GcBlock module is used to capture the relationship between the target defects of the hardware and the transmission lines in the background, and the overall relationship information of the image is used to promote the students' network to learn the teacher's network perception ability of the relationship information. Finally, the classification task mask and regression task mask generated by the PGW module, combined with the overall image relationship loss, form a distillation loss function for network training to improve the accuracy of students' network detection accuracy.

Results and Discussion: The effectiveness of the proposed method is verified by using self-build metal fittings and insulator defect data sets. The experimental results show that the student network mAP_50 (Mean Average Precision at 50) in the Faster R-CNN model with the knowledge distillation algorithm added in this paper increases by 8.44%, and the RetinaNet model increases by 2.6%. The Cascade R-CNN model improved by 5.28%.

KEYWORDS

knowledge distillation, key region guidance, component defects, teacher model, student model

1 Introduction

The transmission line is one of the most important infrastructures of China's energy Internet, and ensuring the reliability of the transmission line is one of the important contents of the construction of the energy Internet. Transmission line components are an important part of mechanical connection, fixing, protection and insulation. However, they are

susceptible to abnormalities and defects such as defects, corrosion and soiling due to the influence of the complex natural environment and harsh climatic conditions (Nguyen et al., 2018). Therefore, regular inspection and maintenance of transmission line fittings, insulators and other important components can effectively reduce a series of safety accidents caused by transmission line faults.

Inspection is a way to guarantee the continuous and stable power supply of the power grid, the purpose of which is to carry out online condition detection and fault diagnosis of components such as shockproof hammers and insulators on the lines (Zhao and Cui, 2018). The current transmission line inspection methods include manual inspection, robot inspection (Toth and Gilpin-Jackson, 2010), helicopter (Pham et al., 2020), unmanned aerial vehicle inspection (Li et al., 2021), and remote sensing satellite inspection (Yang et al., 2021). "Drone inspection is the main focus, supplemented by manual labor" has developed into the main operation and maintenance mode of China's power system (Yang et al., 2020). The construction of intelligent and manual synergistic inspection system is an important initiative to promote the safe operation of the power grid (Du et al., 2022).

The development of deep learning technology provides an effective means for transmission line inspection and can more effectively complete the task of defect detection of transmission components in aerial images. At present, a lot of research work has been done. Literature (Zhai et al., 2023) proposes a transmission line multi-fitting detection method based on implicit spatial knowledge fusion, aiming at the tiny-size and dense occlusion problem in the transmission line multi-fitting detection task. First, in order to mine the implicit spatial knowledge between transmission line fittings to assist the model in detection, the spatial box setting module and the spatial context extraction module are proposed to set the spatial box and extract the spatial context information. Then, the spatial context memory module is designed to filter and remember the spatial context information to assist the location of the multi-fitting detection model. Finally, the post-processing part of the model is improved to further alleviate the low detection accuracy problem caused by dense occlusion fittings. The experimental results show that the proposed model has a promotion effect on the detection of various kinds of fitting. Literature (Li et al., 2023) proposes a metal fittings equipment detection algorithm based on improved YOLOV7. This method adds a CA attention mechanism to the network structure of YOLOV7 to enhance the feature extraction of hardware devices in the network model. At the same time, it reduces the interference of complex backgrounds on the network model to extract features of hardware devices, allowing the network model to extract features in detail, thereby improving the network model's detection generalization for hardware devices. In order to alleviate the problem of misdetection and recheck caused by the lack of context information in various existing hardware and defect detection methods, literature (Zhao et al., 2023) proposes a method of transmission line hardware and defect detection based on context-structure reasoning. First, the image is input into the target detection model; Then, the output result of the detection model is sent to the structural reasoning module, and the output result is sent to the bidirectional gated cycle unit and self-attention for processing. The structural knowledge of transmission line fittings and their defects is used to improve the confidence degree of the correct positive sample and reduce the confidence degree of the

wrong positive sample. Finally, the final output result is obtained through the regressor. To achieve the purpose of improving the average accuracy. Literature (Sun et al., 2023) proposes a two-stage insulator defect detection framework composed of attention-based insulator detection network and defect detection network. Among them, the attention-based insulator detection network is responsible for the location of the insulator, and the defect detection model determines whether the insulator is damaged. The two-stage design of first positioning and then detection avoids the interference of complex background and can realize the high-precision detection of defects. Literature (Li et al., 2023) proposes a multi-scale feature fusion insulator defect detection network for solving the problem of insulator defective regions with little pixel information and varying shapes and sizes. The network used a residual attention network to obtain insulator defect features with different resolutions, and designed a multi-scale feature fusion network based on inverse convolution and multi-branch detection, which gradually fused the deep feature maps with the shallow feature maps. In this way, more abundant image semantic information can be generated for target classification and location regression. In addition, the literature also used Focal loss and Gaussian non-great suppression methods to further enhance the detection effect.

Although these methods improve the accuracy of detection, they inevitably increase the complexity of the model, consume a lot of computing resources and time, and are difficult to deploy on resource-limited equipment. The knowledge distillation algorithm provides a solution to this problem. Literature (Gu et al., 2023) proposes a deep neural network model compression algorithm for knowledge distillation of multi-teacher models, which takes advantage of the integration of multi-teacher models and takes the predictive cross-entropy of each teacher model as the quantitative criterion for screening to select the teacher model with better performance to guide students, and allows the student model to extract information from the feature layer of the teacher model. And give better performing teacher models a greater say in instruction. Literature (Wang et al., 2022) proposes an attention mechanism based on the feature map quality evaluation algorithm (IQE). The knowledge distillation method based on the IQE attention mechanism uses the IQE method to identify important knowledge in the pre-trained SAR target recognition deep neural network. Then in the process of knowledge distillation, the lightweight network is forced to focus on the learning of important knowledge. Through this mechanism, the method proposed in this paper can efficiently transfer the knowledge of the pre-trained SAR target recognition network to the lightweight network, which makes it possible to deploy the SAR target recognition algorithm on the edge computing platform. Literature (Zhao et al., 2022) propose a target detection model distillation (TDMD) framework using feature transition and label registration for remote sensing imagery. A lightweight attention network is designed by ranking the importance of the convolutional feature layers in the teacher network. Multiscale feature transition based on a feature pyramid is utilized to constrain the feature maps of the student network. A label registration procedure is proposed to improve the TDMD model's learning ability of the output distribution of the teacher network.

At present, some researches have applied the knowledge distillation method to the field of electric power. Literature (Yang et al., 2022) proposes a compression and integration application

method based on knowledge distillation. In this method, the Detr model is used to identify the initial target, and the Deformable Detr algorithm is used to compress the Detr model, so that the compression ratio reaches 87.5% and the target detection accuracy is maintained at a high level, and the effective integrated application of the target detection model in the substation inspection robot body is realized. Literature (Zhao et al., 2021) proposes a bolt defect image classification method based on dynamically supervised knowledge distillation, aiming to solve the problem of high computational resource consumption of large models. The method utilizes adaptive weighting and attention transfer techniques to improve the ability of the small model to learn and represent bolt defects, which in turn enhances its classification performance. In addition, literature (Zhang et al., 2022) improved the YOLOv4 model and introduced the PCSA (Positional Contextual Attention Shift) attention module for the problem of anti-vibration hammer small target detection. This method combines pruning and knowledge distillation techniques to tailor and compress the network parameters, and constructs a lightweight anti-vibration hammer detection network model, PCSA-YOLOs, to improve the detection accuracy of small targets in complex backgrounds.

With the intelligent development of power system inspection technology, it is urgent to deploy models on UAV and helicopter aerial photography and online monitoring equipment. Knowledge distillation can help improve the performance of the model with a small number of parameters, but making the student network simulate the teacher's network feature extraction ability without difference cannot achieve the best effect. The focus of this paper is to enable students to learn the effective feature processing ability of teachers' networks. In this paper, the PGW module is first introduced to refine the feature knowledge of foreground object distillation, and the first k most important pixels are extracted to form a feature mask to improve the distillation performance of student network for detecting hardware defects. Then, the GcBlock module is used to capture the relationship between the target defects of the hardware and the transmission lines in the background, and the overall relationship information of the image is used to promote the students' network to learn the teacher's network perception ability of the relationship information. The combination captures information about the relationship between transmission line components and components, and between components and backgrounds, helping to improve target detection accuracy.

2 Research methodology

2.1 Refinement of prospects regional distillation

The application of knowledge distillation algorithms to image target detection has focused on models using feature pyramid networks. Past approaches usually directly used the output of the classification and regression tasks of the teacher network as the target of the student network. With the development, knowledge distillation can guide the training of the student network in a more targeted way to improve the detection accuracy. As shown in the literature (Guo et al., 2021), unlike the general classification task, the classification and regression tasks in

the detection network can be negatively affected if the same objective function is used for both tasks. This is because the two tasks have different preferences for features: classification requires regions with rich semantic information, whereas regression prefers to focus on the edge portion of feature information. Features that produce better classification scores are not accurate enough in predicting bounding boxes (Song et al., 2020). Therefore, the same sensory field does not guarantee optimal performance for both classification and regression tasks. As shown by the images of defects of gold tools and insulators, important feature knowledge exists for defects of the same kinds of gold tools and insulators, and the datasets of defects of gold tools and insulators are much smaller, which should be fully utilized to guide the students' network by taking advantage of the ability of the teacher's network to deal with the feature knowledge of the defects.

Therefore the features are scored to determine whether they are good for classification or regression tasks, in order to reduce the adverse effect of complex background on the detection of defective targets of gildings and insulators, with the help of ground-turth box first decouple each layer of features of FPN (Feature Pyramid Network) whether it belongs to the foreground target region or the background region as shown in Eq. 1:

$$M_{(i,j)} = \begin{cases} 1, if\, f(i,j) \in G \\ 0, if\, f(i,j) \notin G \end{cases} \tag{1}$$

Let (i,j) denote the horizontal and vertical coordinates of the feature pyramid network generating the feature map of the model. If (i,j) is in the ground-truth bounding box, it is determined that this feature belongs to the foreground target region, and the mask is set $M_{(i,j)} = 1$; if (i,j) is not in the ground-truth bounding box, it is determined to be the feature map of the background region, and the mask is set $M_{(i,j)} = 0$.

Amplifying the most meaningful feature distillation signals generated by the teacher network and using them to guide the student network is the purpose of knowledge distillation. For this purpose, we look at the quality of a teacher's bounding box predictions taking both classification and localization into consideration. Formally, the quality score of a box $b_{(i,j)}$ predicted from a position $X_i = (x_i, y_i)$ w.r.t. a ground truth b is as shown in formula (2):
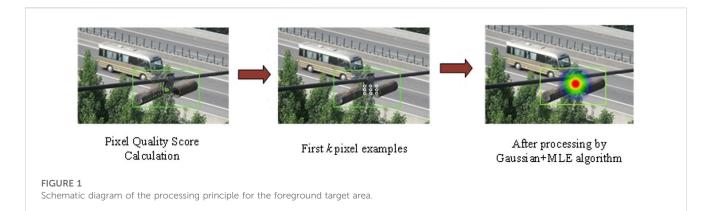
$$q\left(b_{(i,j)}, G\right) = M_{(i,j)} \cdot \left(p_{(i,j)}G\right)^{\lambda} \cdot \left(\text{DIOU}\left(G, b_{(i,j)}\right)\right)^{1-\lambda} \tag{2}$$

where $M_{(i,j)}$ is an indicator function that is 1 if X_i lies inside box b and 0 otherwise; $(p_{(i,j)}G)$ is the classification probability w.r.t. the GT box's category; $\text{DIOU}(G, b_{(i,j)})$ is the DIOU between the predicted and ground-truth box. $b_{(i,j)}$ is a prediction frame and G is a real labeled box, it is the ground-truth box; λ is a hyper-parameter that balances classification and localisation. We calculate the quality score of location X_i as the maximum value of all prediction scores for that particular location, as shown in Eq. 3. Y is used to represent the set of X_i locations as shown in Eq. 4:

$$q_i = \max_{j \in Y} q\left(b_{(i,j)}, G\right) \tag{3}$$

$$Y = \{1, 2, ..., X_i\} \tag{4}$$

Use of mass fraction q_i as an important parameter in determining distillation, these positions are the highest quality predictive ensemble of scores generated by the teacher network.

FIGURE 1
Schematic diagram of the processing principle for the foreground target area.

The purpose of knowledge distillation is to allow the student network to mimic the strong generalization ability of the teacher network. To achieve this, this paper focuses the foreground distillation region on the locations where the teacher network produces predictions with high-quality scores, as these locations contain detection information that combines both classification and regression considerations and represent the excellent performance of the teacher network, helping the student network to improve its ability to detect defective targets in complex contexts.

2.2 Knowledge distillation guided by key area scoring

The features of fitting and insulator defects are very different from those of their intact targets; the features are more consistent between parts of the same species, but the features of each component defect are variable and complex (Zhao et al., 2021), and when labeling the dataset, the size of the labeled box is as close to the target as possible, and the defects are damages that are produced on the component targets, which have a more reduced range of effective features compared to the intact targets. Therefore, we would like to use the knowledge of features of defects in gold tools and insulators as a focus for the instructor network to guide the student network. Therefore, the PGW (Prediction-Guided Weighting) (Yang et al., 2022) module is introduced to improve the prospect distillation region. And the PGW module is precisely concentrated in the first k feature pixels with the highest mass fraction in the prospect region. The effect of each position is then smoothed according to the two-dimensional Gaussian distribution fitted for each ground-truth box by the maximum likelihood estimation method. Finally, only the k position is extracted in the foreground target region, and the weight of the position is assigned by the Gaussian function, the schematic is shown in Figure 1.

We smooth the effects of each position according to the 2D Gaussian distribution fitted by the maximum likelihood estimate (MLE) for each ground-truth box. Finally, foreground distillation is performed only for those k positions, whose weights are assigned by Gaussian. For detection targets with ground-truth O, the quality score q_i for each feature pixel within G is first calculated (Du et al., 2021). The calculation formula of q_i is shown in (3). The k highest scoring pixels are then selected among all the layers of the FPN network. $T^O = \left\{ (X_k^O, L_k^O) | k = 1, ..., K \right\}$ is used to generalize the k highest scoring pixels. Where X_k^O is the absolute coordinate of the

pixel of the detection target O. L_k^O denotes which layer of the FPN where the pixel of the detection target O is located. Both quantities represent the k th pixel. Assuming that the selected pixel description is represented on the image plane as $T_o^k \sim N(\mu, \Sigma|o)$ defined, the maximum likelihood estimation algorithm is used to compute and the two parameters μ and Σ as shown in Eqs 5, 6:

$$\mu = \frac{1}{K} \sum_{k=1}^{K} X_k^o \qquad (5)$$

$$\Sigma = \frac{1}{K} \sum_{k=1}^{K} (X_k^o - \mu)(X_k^o - \mu)^T \qquad (6)$$

Each feature pixel $p_{(i,j)l}$ with absolute $X_{(i,j)}$ coordinates in layer l of the FPN, calculates its importance in the distillation process as expressed in Eq. 7:

$$I^o_{(i,j),l} = \begin{cases} 0 & P_{(i,j)_l} \notin T^o \\ \exp\left(\left(-\frac{1}{2}\left(X_{(i,j)}-\right)\sum^{-1}\left(X_{(i,j)} - \mu\right)^T \right) \right) & P_{(i,j)_l} \in T^o \end{cases} \qquad (7)$$

where $I^o_{(i,j),l}$ denotes that the detection target O is in layer l of the FPN, the importance of the feature pixel $p_{(i,j)l}$ with coordinates (i, j). If $p_{(i,j)l}$ belongs to the highest scoring pixel, calculate the importance value using the formula, and if it does not belong to the pixel, make it equal to zero. If a feature pixel is equally important for more than one object, we use its maximum value $I_{(i,j),l}$, and the formula representation is shown in (8):

$$I_{(i,j),l} = \max_o \left\{ I^o_{(i,j),l} \right\} \qquad (8)$$

l refers to a layer of the FPN layer of size $H_l \times W_l$, by normalizing the importance of the distillation with the number of pixels of that layer that are important and have a non-zero mass fraction so that they are used to assign the distillation weights Q, and the formula is expressed as shown in (9).

$$Q = \frac{I_{(i,j),l}}{\sum_{i=1}^{H_l} \sum_{j=1}^{W_l} M_{(i,j),l}} \qquad (9)$$

H_l and W_l denote the length and width dimensions of the feature map in the l-layer of the FPN, the above process constitutes the Predictive Guidance Weighting (PGW) module, whose output is the

foreground distillation weight Q of all feature levels and pixels, which are used to find out the important feature pixels for use in the teacher's network to guide the student's network.

2.3 Overall distillation loss function

The high-level network of the feature extraction network can output better semantic features, but due to the size of the feature map is too small, the geometric information is not sufficient, which is not conducive to the detection of the target; the shallow network contains more geometric information, but the semantic features of the image are not much, which is not conducive to the classification of the image. Therefore, when the network is trained, the student network is made to learn the ability of the teacher network to extract and process features at each layer of the FPN. By distilling the classification and regression tasks separately, the student network integrally learns the generalization ability of the teacher network for these two tasks, which leads to an increase in its detection accuracy.

Formally, at each feature level in FPN, this paper utilizes the PGW module to generate two distinct foreground distillation masks, $Q_{(i,j)}^{cls}$ is the foreground distillation mask generated specifically for the classification task. $Q_{(i,j)}^{reg}$ is the foreground distillation mask generated specifically for the regression task. In this context, $R^{C \times H \times W}$ represents a feature layer has C channels, and the feature map height and width of each layer H and W. During training, the student model is encouraged to learn the corresponding classification and regression features $F_{n,(i,j)}^{S,cls}, F_{n,(i,j)}^{S,reg} \in R^{C \times H \times W}$ from the teacher network. The classification feature loss function is represented as Eq. 10, and the regression feature loss function is represented as Eq. 11.

$$L_{fea}^{cls} = \sum_{n=1}^{C}\sum_{i=1}^{H}\sum_{j=1}^{W}\left(\alpha Q_{(i,j)}^{cls} + \beta N_{(i,j)}^{cls}\right)\left(F_{n,(i,j)}^{T,cls} - F_{n,(i,j)}^{S,cls}\right)^2 \quad (10)$$

$$L_{fea}^{reg} = \sum_{n=1}^{C}\sum_{i=1}^{H}\sum_{j=1}^{W}\gamma Q_{(i,j)}^{reg}\left(F_{n,(i,j)}^{T,reg} - F_{n,(i,j)}^{S,reg}\right)^2 \quad (11)$$

α, β, and γ are hyperparameters used to balance the loss weights. $N_{(i,j)}^{cls}$ is the normalized mask for the background distillation region. As shown in Eq. 12, when $Q_{(i,j)}^{reg} \neq 0$, $M_{(i,j)}^-$ is 1. The meaning of $\mathcal{M}_{(h,w)}$ is opposite to $M_{(i,j)}$, and if the pixel (h, w) is not inside the region G, it is assigned a value of 1.

$$N_{(i,j)}^{cls} = M_{(i,j)}^- \Big/ \sum_{h=1}^{H}\sum_{w=1}^{W}\mathcal{M}_{(h,w)} \quad (12)$$

In order to distill background information, the GcBlock (Cao et al., 2019) module is utilized to capture the relationship between defects in the hardware target and the transmission lines present in the background. This module leverages the overall relationship information in the image (Park et al., 2019). It encourages the student network to learn the teacher network's ability to perceive relationship information (Hu et al., 2018). The representation of the overall image relationship loss is given by Eq. 13.

$$L_{rela} = \mu \sum \left(R(F^T) - R(F^S)\right)^2 \quad (13)$$

In the equation, μ is a hyperparameter used to balance the loss function, and F^T and F^S are the feature maps generated by the teacher and student models, respectively. The function R represents the relationship information between the hardware target captured by different models and the background region. Its specific formulation is given by Eq. 14.

$$R(F_i) = F_i + W_{v2}Relu\left(LN\left(W_{v1}\left(\sum_{j=1}^{N_P}\frac{e^{W_k F_j}}{\sum_{m=1}^{N_P} e^{W_k F_M}}F_j\right)\right)\right) \quad (14)$$

In this context, W_{v1}, W_{v2} and W_k represent different linear transformation matrices. F_i represents the feature map of an input instance. $\partial_j = \frac{e^{W_k F_j}}{\sum_{m=1}^{N_P} e^{W_k F_M}}$ represents the weights obtained from global attention pooling. LN stands for Layer Normalization. N_P represents the number of positions in the feature layer. In the context of an image, $N_P = H \cdot W$. e represents the natural constant. F_M represents the feature map generated for the foreground region. F_j is any extracted feature from the image. $\delta(\cdot) = W_{v2}Relu(LN(W_{v1}(\cdot)))$ denotes the feature transformation that captures channel dependencies. The GcBlock module consists of two components: global attention pooling for context modeling and bottleneck transformation to capture channel correlations.

The structural diagram of the distillation method in this paper is shown in Figure 2. It consists of two parts: foreground object region distillation and background relationship distillation. For foreground object region distillation, the PGW module is used to calculate the masks for both the classification and regression tasks, which together form the loss function used for training the foreground object region. The background relationship distillation area adopts the GcBlock module to capture relationships in the image. The overall distillation loss function in this paper is represented as Eq. 15.
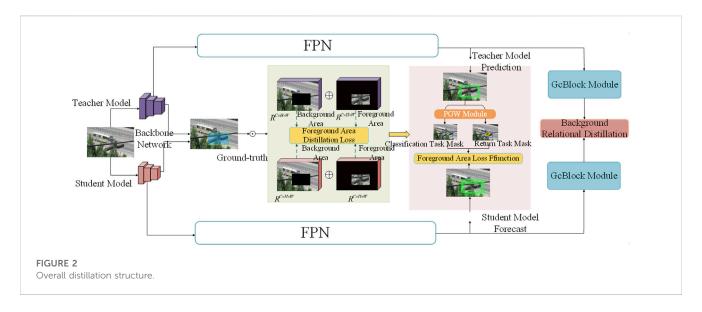
$$L = L_{fea}^{cls} + L_{fea}^{reg} + L_{rela} \quad (15)$$

3 Manuscript experimental results and analysis

3.1 Experiment preparation

This article uses 5 types of fittings and insulators, along with their defects, as the research dataset. Each type of defect has a corresponding normal target feature for comparison, including normal vibration damper, vibration damper cross, vibration damper corrosion, normal single insulator, and insulator drop. There are a total of 2,497 images, with 1997 images in the training set, 250 images in the test set, and 249 images in the evaluation set, with a ratio of 8:1:1. The dataset contains a total of 4,628 objects to be detected.

The model described in this article is trained and tested using the NVIDIA GeForce GTX 1080Ti professional accelerator card. The operating system used is Ubuntu 16.04.6 LTS, with training accelerated using CUDA 10.1. The computer language used is Python 3.7.11, and the network development framework is PyTorch. All programs are executed based on the MMDetection 2.16 toolbox. This article uses the commonly used evaluation metric

FIGURE 2
Overall distillation structure.

in object detection models, mean Average Precision (mAP), to assess the performance of the model. The mAP is calculated by computing the Average Precision (AP) for each class of objects and then taking the mean of all class APs. The resulting mAP is used as the final evaluation metric for the object detection model.

3.2 Comparison experiments with multiple models using distillation algorithm

In the experiments, the teacher network, student network, and the network with the distillation algorithm all use a learning rate of 0.0001. The backbone network for the teacher network is ResNet101, while for the student network, it is ResNet50. The training is conducted for 24 epochs (training rounds), and the batch size used is 50.

The experimental process in this article mainly consists of two steps.

1) Train the object detection models separately using the teacher network (with a larger number of parameters) and the student network (with a smaller number of parameters) on the fittings defect dataset. Calculate and record their respective accuracy.
2) Train the student network, which has a smaller number of parameters, using the knowledge distillation algorithm with the fittings defect dataset. After training, calculate and record its accuracy.

To validate the effectiveness of the proposed knowledge distillation algorithm, a comparative experiment is conducted using the evaluation metrics mentioned earlier. The experimental results are presented in Table 1. Three image detection models, namely, Faster R-CNN, RetinaNet, and Cascade R-CNN, are used in the experiments. Each model is separately trained as a teacher model, a student model, and a model with the knowledge distillation algorithm introduced in this article. The detailed process is to put the same data set into the same model with the backbone model of Resnet101 and Resnet50. The backbone network is identified as the teacher network by Resnet101 and the

backbone network is identified as the student network by Resnet50. The results of two different backbone network training were analyzed. After adding the distillation algorithm in this paper, the student network is trained again, and the difference between the student network with distillation algorithm added and the student network without distillation algorithm added is compared. The control parameters adopted in these processes are consistent, the learning rate is 0.0001, and the epoch of the training rounds is 24. The batch size is 50.

To verify the general applicability of the proposed method, both single-stage and two-stage models are used in the experiments. From the data in Table 1, it can be observed that the detection performance of the student network improves significantly after the knowledge distillation algorithm is applied. This distillation algorithm, as presented in this article, utilizes a decoupling approach between foreground and background information regions. By focusing on the foreground object regions and reducing the interference caused by complex backgrounds, the algorithm enhances the localization ability of the model. By setting the k value to 45 in the foreground object regions, the data indicates that, in most cases, adding the knowledge distillation algorithm proposed in this article improves the detection accuracy of the student network for both normal component targets and defects. Particularly, the improvement in detecting fittings and insulator defects is greater than the improvement in detecting normal targets. In the case of the RetinaNet model, where the teacher network's accuracy is lower than the student network's accuracy in detecting vibration damper corrosion, adding the distillation algorithm does not improve the performance of the student network. This could be because the teacher network's performance is inferior to the student network's performance, which hinders its ability to guide the student network in improving accuracy. In the case of Faster R-CNN, the highest improvement is achieved in detecting vibration damper corrosion, reaching up to 22.6%. For RetinaNet, the highest improvement is seen in detecting normal insulators, with a maximum improvement of 6.6%. Cascade R-CNN shows the highest improvement in detecting vibration damper corrosion, reaching 8.2%.

The table shown in Table 2 compares the Grad-CAM distillation algorithm (decoupling common feature scores) with the distillation

TABLE 1 AP results for different models and after applying the knowledge distillation algorithm.

Method	mAP_50	Normal vibration damper	Vibration damper cross	Vibration damper corrosion	Normal single insulator	Insulator drop
Faster-Res101 (Teacher network)	0.544	0.648	0.312	0.642	0.589	0.530
Faster-Res50 (Student network)	0.419	0.556	0.286	0.370	0.423	0.458
Faster-Res101-KD-Res50	0.503	0.572	0.368	0.596	0.469	0.510
RetinaNet-Res101 (Teacher network)	0.559	0.699	0.069	0.576	0.751	0.701
RetinaNet-Res50 (Student network)	0.485	0.660	0.111	0.549	0.510	0.593
RetinaNet-Res101-KD-Res50	0.511	0.683	0.092	0.567	0.576	0.635
Cascade-101 (Teacher network)	0.617	0.672	0.464	0.695	0.631	0.623
Cascade-50 (Student network)	0.503	0.629	0.394	0.486	0.530	0.474
Cascade-101-KD-50	0.555	0.694	0.510	0.568	0.510	0.495

TABLE 2 Comparison of two foreground knowledge distillation methods.

Detection models	Settings	mAP_50	Normal vibration damper	Vibration damper cross	Vibration damper corrosion	Normal single insulator	Insulator drop
Faster R-CNN	Teacher network	0.544	0.648	0.312	0.642	0.589	0.530
	Student network	0.419	0.556	0.286	0.37	0.423	0.458
	Decoupled scoring	0.499	0.654	0.303	0.463	0.614	0.462
	Proposed method	0.503	0.572	0.368	0.596	0.469	0.510
RetinaNet	Teacher network	0.559	0.699	0.069	0.576	0.751	0.701
	Student network	0.485	0.660	0.111	0.549	0.510	0.593
	Decoupled scoring	0.518	0.723	0.101	0.554	0.582	0.630
	Proposed method	0.516	0.683	0.092	0.567	0.576	0.663

algorithm proposed in this article, using Faster R-CNN and RetinaNet as representatives of two-stage and single-stage detection models, respectively. The Grad-CAM distillation algorithm improves the detection performance of most targets, but for fittings defects, its improvement is slightly inferior to the distillation algorithm proposed in this article. In the case of the Faster R-CNN model, the detection accuracy for normal vibration damper targets and normal single insulator targets is slightly lower in this article's method compared to the Grad-CAM method. However, for vibration damper cross defects, our method

outperforms the Grad-CAM method by 6.5 percentage points, and for vibration damper corrosion and insulator String defects, it outperforms the Grad-CAM method by 13.3% and 4.8%, respectively. The analysis shows that the feature used for defect target detection is more concentrated in the critical pixel regions. Using ground-truth boxes as the range of foreground object regions introduces more noise for defects. This article's method selects the top-k highest-scored pixels to form a mask, which includes essential features for defect detection while avoiding introducing noise from other parts of the foreground object regions. Figure 3
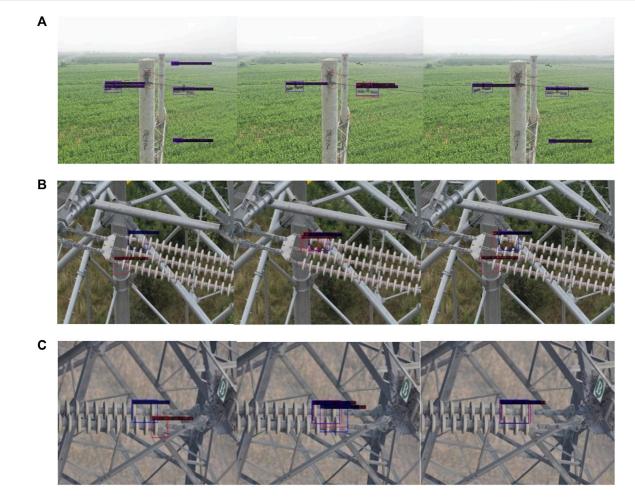
FIGURE 3
Comparison of different network experiment results **(A)** Teacher network **(B)** Student network **(C)** Distilled student.

TABLE 3 Ablation Experiments with Different k Values.

k	1	5	10	15	30	45	60	75
mAP_50	0.189	0.265	0.312	0.458	0.467	0.483	0.479	0.410

shows visualized detection images for different student networks, teacher networks, and networks after applying the distillation method.

3.3 Ablation experiments

To investigate the impact of different k values on improving the accuracy of the student network, ablation experiments were conducted based on the Faster-RCNN model, using ResNet-100 as the teacher network and ResNet-50 as the student network. The mAP_50 values of the student network were observed for different k values, and the results are shown in Table 3. From the results, it can be observed that the student network achieves the optimal mAP_50 value when k is set to 45. It is speculated that if k is too small, it may not capture crucial defect features, while setting k to a large value introduces too much noise from foreground object regions, leading to negative effects.

4 Conclusion

Due to the limited dataset of hardware defects, hardware defects occur as damage to the hardware target. Compared to the hardware target, the effective feature range of hardware defects is smaller. In order to improve the detection accuracy of hardware defects in power transmission lines by the student network, this paper improves the foreground target region distillation. It guides the student network with more refined feature knowledge generated by the teacher network. Considering the influence of the two tasks, classification, and regression, in the foreground target region, pixels are scored, and the top k important pixels' generated masks containing feature knowledge are used to guide the student network. Experimental results show that the proposed method applied to three different single-stage and two-stage detection models, Faster-RCNN, RetinaNet, and Cascade R-CNN, has improved the detection accuracy of hardware and its defects. In Faster R-CNN, after adding the knowledge distillation algorithm in

this paper, mAP_50 has improved by 8.44% for the student network. RetinaNet improved by 2.6%, and Cascade R-CNN improved by 5.28%. This lays a solid foundation for lightweighting the hardware and its defects detection models in power transmission lines.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

Author contributions

ZZ: Conceptualization, Funding acquisition, Investigation, Project administration, Supervision, Writing–review and editing. XL: Conceptualization, Investigation, Methodology, Software, Validation, Visualization, Writing–original draft, Writing–review and editing. YX: Writing–original draft, Writing–review and editing. SM: Formal Analysis, Supervision, Writing–review and editing.

Funding

Acknowledgments

Conflict of interest

Author XL was employed by China Nuclear Power Engineering Co., Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Cao, Y., Xu, J., Lin, S., Wei, F., and Hu, H. (October 2019). "Gcnet: non-local networks meet squeeze-excitation networks and beyond," in Proceedings of the IEEE/CVF international conference on computer vision workshops. Montreal, BC, Canada 0-0

Du, Q., Dong, W., Su, W., and Wang, Q. (September 2022). "UAV inspection technology and application of Transmission Line," in Proceedings of the 2022 IEEE 5th International Conference on Information Systems and Computer Aided Education (ICISCAE), 594–597. Dalian, China. doi:10.1109/iciscae55891.2022.9927674.8

Du, Z., Zhang, R., Chang, M., Zhang, X., Liu, S., Chen, T., et al. (2021). Distilling object detectors with feature richness. Adv. Neural Inf. Process. Syst. 34, 5213–5224. doi:10.48550/arXiv.2111.00674

Gu, M., Ming, R., Qiu, C., and Wang, X. (2023). A multi-teacher knowledge distillation model compression algorithm for deep neural network. Appl. Electron. Tech. 49 (08), 7–12. doi:10.16157/j.issn.0258-7998.233812

Guo, J., Han, K., Wang, Y., Wu, H., Chen, X., Xu, C., et al. (June 2021). "Distilling object detectors via decoupled features," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA 2154–2164.

Hu, H., Gu, J., Zhang, Z., Dai, J., and Wei, Y. (June 2018). "Relation networks for object detection," in Proceedings of the IEEE conference on computer vision and pattern recognition, Las Vegas, NV, USA 3588–3597.

Li, B., Qu, L., Zhu, X., Guo, Z., Tian, Y., Yu, F., et al. (2023). TMT proteomics analysis reveals the mechanism of bleomycin-induced pulmonary fibrosis and effects of Ginseng honeysuckle superfine powdered tea. Trans. China Electrotech. Soc. 38 (01), 60–70. doi:10.1186/s13020-023-00769-x

Li, J., Wang, Q., Hong, S., Fan, L., Chen, X., and Ai, C. (June 2023). "Improved YOLOV7 algorithm for transmission line metal fittings equipment detection: algorithm improvement for integrating attention mechanism," in Proceedings of the 2023 4th International Seminar on Artificial Intelligence, Networking and Information Technology (AINIT), Nanjing, China 496–500.

Li, X., Li, Z., Wang, H., and Li, W. (2021). Unmanned aerial vehicle for transmission line inspection: status, standardization, and perspectives. Front. Energy Res. 9. 713634 doi:10.3389/fenrg.2021.713634

Nguyen, V., Jenssen, R., and Roverso, D. (2018). Automatic autonomous vision-based power line inspection: a review of current status and the potential role of deep learning. Int. J. Electr. Power & Energy Syst. 99, 107–120. doi:10.1016/j.ijepes.2017.12.016

Park, W., Kim, D., Lu, Y., and Cho, M. (June 2019). "Relational knowledge distillation," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, New Orleans, LA, USA. 3967–3976.

Pham, H. C., Ta, Q.-B., Kim, J.-T., Ho, D.-D., Tran, X.-L., and Huynh, T.-C. (2020). Bolt-loosening monitoring framework using an image-based deep learning and graphical model. Sensors 20 (12), 3382. doi:10.3390/s20123382

Song, G., Liu, Y., and Wang, X. (June 2020). "Revisiting the sibling head in object detector," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, New Orleans, LA, USA 11563–11572.

Sun, S., Li, W., Song, C., Cong, X., and Ding, L. (August 2023). "Attention-based insulator defect detection," in Proceedings of the 2023 IEEE International Conference on Mechatronics and Automation (ICMA), 414–419. Harbin, Heilongjiang, China. doi:10.1109/ICMA57826.2023.10215660

Toth, J., and Gilpin-Jackson, A. (October 2010). "Smart view for a smart grid— unmanned Aerial Vehicles for transmission lines," in Proceedings of the 2010 1st International Conference on Applied Robotics for the Power Industry (CARPI 2010), 1–6. Montreal, QC, Canada. doi:10.1109/carpi.2010.5624465

Wang, J., Jiang, T., Cui, Z., Cao, Z., Cao, C., and Li, S. (2022). Seed germination traits and dormancy classification of 27 species from a degraded karst mountain in central Yunnan-Guizhou Plateau: seed mass and moisture content correlate with germination capacity. IGARSS 2022-2022 IEEE Int. Geoscience Remote Sens. Symposium 24, 1043–1056. doi:10.1109/plb.13451

Yang, C., Ochal, M., Storkey, A., and Crowley, E. J. (October 2022). "Prediction-guided distillation for dense object detection," in Proceedings of the European Conference on Computer Vision, Cham: Springer Nature Switzerland. 123–138.

Yang, L., Fan, J., Liu, Y., Li, E., Peng, J., and Liang, Z. (2020). A review on state-of-the-art power line inspection techniques. IEEE Trans. Instrum. Meas. 69 (12), 9350–9365. doi:10.1109/tim.2020.3031194

Yang, Y. (2022). Model compression and integration applications of substation objects detection based on knowledge distillation. Inf. Comput. 34 (13), 50–57.

Yang, Z., Fei, X., Ma, X., Ou, W., Zhao, B., Liu, B., et al. (December 2021). "Design method of core index of remote sensing satellite for power transmission and transformation equipment patrol system, signal and information processing, networking and computers," in Proceedings of the 7th International Conference on

Signal and Information Processing, Networking and Computers (ICSINC), Singapore. 271–279.

Zhai, Y., Guo, C., Wang, Q., Zhao, K., Bai, Y., and Zhang, Y. (2023). Multi-fittings detection method for transmission lines based on implicit spatial knowledge fusion. *J. Graph.* 9, 1–10.

Zhang, K., Yue, Q., Huang, W., Huang, C., Ren, Y., and Hou, J. (2022). Clinical study of factors associated with pregnancy outcomes in pregnant women with systemic lupus erythematosus in the southern China. *Adv. Technol. Electr. Eng. Energy* 41 (11), 59–66. doi:10.46497/ArchRheumatol.2022.8870

Zhao, B., Wang, Q., Wu, Y., Cao, Q., and Ran, Q. (2022). Target detection model distillation using feature transition and label registration for remote sensing imagery. *IEEE J. Sel. Top. Appl. Earth Observations Remote Sens.* 15, 5416–5426. doi:10.1109/jstars.2022.3188252

Zhao, Z., and Cui, Y. (2018). Research progress of visual detection methods for transmission line key components based on deep learning. *Electr. Power Sci. Eng.* 34 (03), 1–6. doi:10.3969/j.ISSN.1672-0792.2018.03.001

Zhao, Z., Jiang, Z., Li, Y., Qi, Y., Zhai, Y., Zhao, W., et al. (2021). Overview of visual defect detection of transmission line components. *J. Image Graph.* 26 (11), 2545–2560. doi:10.11834/jig.200689

Zhao, Z., Jin, C., Qi, Y., Zhang, K., and Kong, Y. (2021). Image classification of transmission line bolt defects based on dynamic supervision knowledge distillation. *High. Volt. Eng.* 47 (02), 406–414. doi:10.13336/j.1003-6520.hve.20200834

Zhao, Z., Xiong, J., Xu, H., and Zhang, L. (2023). Integrating structural reasoning for deep model transmission line fittings and their defects detection. *High. Volt. Eng.* 49 (08), 3346–3353. doi:10.13336/j.1003-6520.hve.20230241