scientific reports

Check for updates

OPEN

Researching on insulator defect recognition based on context cluster CenterNet++

Bo Meng

The existing UAV inspection images are faced with many challenges for insulator defect recognition. A new multi-resolution Context Cluster CenterNet++ model is proposed. First, this paper proposes the Context Cluster method to solve the problem of low recognition accuracy caused by non-uniform distribution of targets. The cluster region is used to identify and predict the location of the target, and the improved loss function is used to modify the cluster center. Secondly, this paper uses deformable convolution operator (DCNv2) combined with path aggregation network (PAN) to carry out deformable convolution operation on the image, and accurately predicts the regression box and key point triplet (KP), so as to improve the accurate positioning of the target position of any shape and any scale. The sensitivity of the model to target scale change and deformation is reduced, and the recognition accuracy of the model is improved. Then, Bhattacharyya distance is used to calculate the triplet prediction loss of key points and the target center point offset loss, so as to significantly improve the positioning accuracy of the same target in different frames. Finally, experiments are carried out on the MS-COCO dataset and the National Grid standardized UAV inspection insulator image dataset. Our code is at https://github.com/mengbonannan88/CC-CenterNet.

Keywords Insulator defect recognition, CenterNet++, Context cluster, Deformable convolution, Nonlinear loss function

As a special and very critical insulation control, insulator plays an important role in the transmission and distribution network. The accurate identification and location of insulator defects is of great significance[1,58] in the field of electric power inspection. With the development of unmanned aerial vehicle (UAV) technology, UAV has been widely[2] used in transmission line inspection.

In recent years, more and more researchers at home and abroad have devoted themselves to applying deep learning models to UAV power inspection based on computer vision, and have achieved[3] more results. At present, insulator defect recognition algorithms based on deep learning models in UAV inspection images are mainly divided into two categories: Anchor recognition method and anchor-free recognition method[4].

Region-based Convolutional Neural Network (R-CNN) series algorithms are classical algorithms in object recognition algorithms with detection frames[5]. Since then, many extended and improved methods have been proposed, such as Fast R-CNN[6], Faster R-CNN[7], Mask R-CNN[8], DINO[11], SPD-Conv[12], MR-CNN[13], Cascade R-CNN[14], Fitness-NMS[15], R-FCN[16], TridentNet[17], etc. Han et al.[4] used the residual network ResNet-50 as the backbone network to establish a multi-scale insulator identification model for insulator location, and then used the insulator fault location method based on ROI for insulator fault identification. Chen et al.[2] improved the insulator identification method based on the SSD algorithm. After extracting the rectangular area where the insulator is located, they proposed a method based on K-means and double feature constraints to further improve the accuracy of insulator self-explosion location identification[4]. The above methods pay more attention to the accuracy of identification. Another kind of methods that pay more attention to recognition speed, such as SSD[9], RON[18], YOLO series, RetinaNet[21], DETR[10], FCOS[22], Objects as Points[23], etc., remove the proposal verification phase. Liu Yue et al.[19,20] used YOLOv3 and the improved watershed algorithm to identify and locate insulator burst. Firstly, this paper established a YOLOv3 insulator recognition model to identify the insulator area and accurately locate the target area[20]. Then, rotation cropping and other operations were performed on the identified insulator main image to narrow the processing scope. Finally, the burst area of the insulator is accurately identified by combining the traditional image processing algorithm.

The above target recognition algorithms based on detection boxes need to enumerate all possible target locations and scales, and classify each location to identify the target, resulting in a waste of computing resources.

Northeast Electric Power University, Jilin City, China. email: mengbo_nannan@163.com

In addition, these methods are difficult to perceive objects with special shapes, resulting in low recognition rates for objects with arbitrary shapes, scales, and severe shape changes.

Typical box-free recognition methods include CornerNet[21], CenterNet[24] and CornerNet-Lite[25], which identify the object as diagonal corners. ExtremeNet[26] detects the four extreme points (top, left, bottom and right) of the object. Zhou et al.[27] proposed an object detector without sliding window, which models an object as a key-point (the center point of the object), and regress the size, position and attitude of the object according to the key-point. As the center-based method, it is an end-to-end differentiable recognition method. It is simpler, faster and more accurate than the methods based on detection boxes mentioned above.

Although the method without detection box is not sensitive to the scale and shape of the target, it relies too much on the high-resolution heatmap, which leads to slow inference speed and cannot be used in real time.

Most of the insulator defect images collected by unmanned aerial vehicle (UAV) are small targets with the size less than 30 × 30 pixels, and the problem of small target recognition has always been a tricky and urgent research problem. The following factors pose major challenges to the insulator defect recognition of UAV inspection: (1) During drone inspection, the drone typically approaches the target from a distance, which is reflected in images as a process where the target size decreases from large to small. Smaller targets (with width and height less than 30 pixels)[28] are non-uniformly distributed across the image. Anchor-based detectors perform block-wise prediction across the entire image, but small targets are often absent in most blocks, resulting in computational resource waste and reduced inference efficiency; (2) Moreover, during drone inspections, unavoidable movements such as rotation and vibration lead to phenomena like scale variation and rotational deformation in the captured images. Existing detectors, relying on traditional convolution operations for feature processing during bounding box regression, struggle to effectively recognize targets with irregular edges (e.g., insulators), thereby reducing detection accuracy for deformed images; (3) In drone-captured images, target positions often exhibit nonlinear and random variations. However, current detectors, which predict targets based on linear variation assumptions, are prone to inaccuracies in target localization.

To address the issue of the uneven distribution of small insulator targets in drone-captured images, this paper selects the advanced anchor-free detector CenterNet++[29] as the baseline model and optimizes it. Considering that the process of drone image acquisition often involves scale variations and vibrations, which impair CenterNet++'s ability to effectively extract features from distorted insulator targets, we propose a deformable convolution-based detection module. This module introduces additional offsets during convolution operations, allowing the convolutional kernel to better adapt to distorted insulator targets, thereby improving the model's stability on drone platforms. Furthermore, since insulator targets in drone-captured images are not linearly distributed, the backbone feature extractor of CenterNet++, which utilizes a linear approach for stepwise feature extraction, increases computational overhead and reduces the precision of target localization. To address this, we design a context-clustering-based feature extraction backbone. By merging and filtering similar target regions, sparsely distributed insulator targets are uniformly represented in lower-resolution feature maps. This not only alleviates the computational burden but also enhances the network's focus on nonlinearly distributed targets. Further studies reveal that insulator defects of different types often exhibit similar sizes in images. However, the L1-norm loss used by CenterNet++ to constrain the distance between predictions and ground truths fails to effectively distinguish data with similar means, thereby limiting detection performance. To overcome this, we adopt the variance-sensitive Bhattacharyya distance to improve the original L1-loss, further enhancing the accuracy of the baseline model in insulator fault detection tasks.

In summary, the main contributions of this paper are as follows:

(1) In the backbone module, we propose a method called Context Cluster (CC) to solve the non-uniform distribution and long tail problem. The clustering area is used to identify and predict the location of the target, and the improved loss function is used to modify the feedback of the cluster center, which not only improves the recognition accuracy, but also further improves the computational efficiency.
(2) In the detection module, the deformable convolution operator (DCNv2) combined with Path Aggregation Network (PAN) is used to perform deformable convolution operation on the image, and the regression box and key point triple (KP) are accurately predicted. In order to improve the accurate positioning of the target position of any shape and any scale, the sensitivity of the model to target scale changes and deformation is reduced, and the recognition accuracy of the model is improved.
(3) Bhattacharyya distance is used to calculate the key-point triplet prediction loss and the object center offset loss. The advantage of using the nonlinear distance calculation method instead of the linear distance calculation method is that it is not limited by the change of object scale within frames and object position between frames, and can significantly improve the localization accuracy of the same object between different frames.

Related work

Types of insulator defects

The common defects of insulators have the following appearances, as shown in Fig. 1: (a) self-explosion; (b) damage and crack; (c) corrosion of steel cap; (d) serious fouling; (e) rupture of insulator string; (f) flashover.

The network structure of CenterNet++ identification model

CenterNet++ is an anchor-free single-stage recognition model with a compact parameter, a straightforward architecture and outstanding performance.

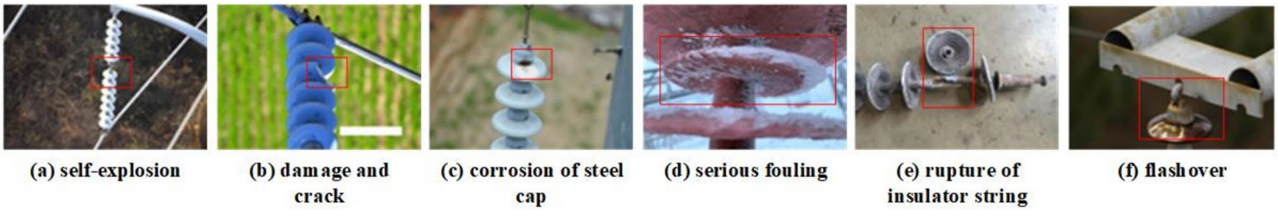
(a) self-explosion (b) damage and crack (c) corrosion of steel cap (d) serious fouling (e) rupture of insulator string (f) flashover

Fig. 1. Insulator defect diagram.

Key-point extraction using multi-resolution CenterNet++ framework

Firstly, the image is input into the ResNet Network as the Backbone Network to extract image features. Then, C3, C4, C5 feature maps are selected from the backbone network as the input[30] of the Feature Pyramid Networks (FPN). FPN outputs the feature maps P3–P7 as the prediction layer of the backbone network.

In each layer of FPN, the results of feature regression map and feature heat map are fused to predict the location of key points. In the prediction based on fusion regression, in order to decouple the top-left and bottom-right prediction results, the Ground Truth (GT) image is divided into four sub-truth boxes (top left, top right, bottom left, bottom right), and the top left sub-truth box (TL GT) and right sub-truth box (BR GT) are selected to supervise the regression results. Taking the regression results of the upper left sub-truth box as an example, some feature points are selected in the TL GT, and three vectors are predicted for each feature point, which are the top left corner key point, the bottom right corner key point and the center key point. Some feature points are also selected in BR GT, and three vectors are predicted for each feature point, which are the top left corner key point, the bottom right corner key point and the center key point. Then, the Intersection of Union (IoU) loss of TL GT and BR GT is obtained. Finally, the loss results of TL GT and BR GT are fused to obtain the prediction of the target position. In the heatmap-based prediction, three light binary heatmaps are predicted to predict the top-left key-point, bottom-right key-point and center key-point. Since the resolution of the heatmap is the same as that of the prediction layer, an additional offset is predicted for each key-point to learn to map the key-point from the heatmap back into the input image.

In the inference process, the regression vector is used as a clue to find the closest key-point on the corresponding feature heatmap, so as to refine the position of the key-point. Then, a bounding box is defined for each pair of key-point pairs that belong to the same class (i.e., TL GT and BR GT belong to the same detection box, and the x and y coordinates of the top left corner point are smaller than the x and y coordinates of the bottom right corner point, respectively). Finally, define a center region for each bounding box and check if the center region contains the two predicted center key-points. If at most one center key-point is detected in its center region, the bounding box will be removed. The score of the bounding box will be replaced with the average score of the points, that is, the top left corner, bottom right corner, and center key-point.

The center region of the target as determined by the triplet key-point

In object recognition, the size of the central region of the detection box restricts the final detection result. If the central region is too small, the recall rate of the detection box is low, and if the central region is too large, the accuracy of detection and recognition will be reduced. In view of the above analysis, the CenterNet++ model adopts a scale-aware center region to adaptively adjust the size of the detection box[35,36].

The above process can be expressed as Eq. (1), where, n is an odd number and determines the size of the target central region j. (tl_x, tl_y) denotes the coordinates of the upper left corner of the detection box, (br_x, br_y) denotes the coordinates of the lower right corner of the detection box, and defines a central region j,(ctl_x, ctl_y) , (cbr_x, cbr_y) denotes the coordinates of the upper left corner and the lower right corner of the central region, respectively.

$$\begin{cases} ctl_x = \frac{(n+1)tl_x+(n-1)br_x}{2n} \\ ctl_y = \frac{(n+1)tl_y+(n-1)br_y}{2n} \\ cbr_x = \frac{(n-1)tl_x+(n+1)br_x}{2n} \\ cbr_y = \frac{(n-1)tl_y+(n+1)br_y}{2n} \end{cases} \tag{1}$$

The multi-resolution CenterNet++ framework uses the function in Eq. (2) to calculate the training loss.

$$L_{multi} = \frac{1}{2}\left(L_{cls}^{tl} + L_{cls}^{br}\right) + \frac{\hat{\alpha}}{2}\left(L_{reg}^{tl} + L_{reg}^{br}\right) + \hat{\beta}\left(L_{kp}^{co} + L_{kp}^{ce}\right) + \hat{\gamma}\left(L_{off}^{co} + L_{off}^{ce}\right) \tag{2}$$

Equation (2) consists of four parts: (1) classification loss, which is calculated by Focal Loss; (2) the regression box size loss, which is obtained by calculating the GIoU loss; (3) the prediction Loss of key-points triple is calculated by Focal Loss[31]; and (4) the target center point bias loss, which is calculated by $L1$ loss[37]. α, β, γ are the corresponding loss weight, which is set to 2, 0.25 and 1, respectively.

Status of development of defect detection tasks

Achieving accurate fault detection can significantly enhance the safety of equipment operation, providing strong support for industrial production[38,59]. By ensuring the healthy operation of various devices, it helps reduce overall energy consumption during the production process[39,4060]. Therefore, conducting research on fault detection holds significant scientific value. With the rapid advancement of deep learning technologies, numerous researchers have developed diverse frameworks tailored to defect detection tasks in various scenarios[41,42]. Ling et al.[43] integrated the concept of image pyramids with deep residual networks for defect detection in trains, achieving promising results. However, due to the inherent computational complexity of image pyramids, the model's processing speed is limited, rendering it unsuitable for real-time applications. To address this issue, Tang et al.[44] proposed a defect detection model for circuit boards based on YOLO, incorporating MobileNetV3 and pruning techniques to achieve both high accuracy and fast processing speed. Similarly, Zhong et al.[45] extended the YOLO framework by introducing attention mechanisms, setting a new benchmark for defect detection on metallic surfaces. Xiao et al.[51] proposed an instance-level defect detection model based on hierarchical features, which advanced the development of freight train detection tasks. Lu et al.[52] designed a defect detection network for powder bed defects based on residual networks, laying the foundation for the automated manufacturing of materials. Wu et al.[53] introduced a two-stage firmware defect detection network, ensuring the stable operation of transportation equipment. Huang et al.[54] designed a train surface defect detection network based on H-CNN, further enhancing the safety of train operations. Zhong et al.[55] proposed a deformable context feature extraction block, which effectively addressed the detection of irregularly shaped defects. Jia et al.[57] proposed a photovoltaic defect detector based on VarifocalNet, which achieves faster speed and higher accuracy compared to similar detectors like YOLO, effectively promoting the development of the photovoltaic power generation industry.

Although anchor-based detectors have demonstrated strong performance in many defect detection tasks, anchor-free detectors have shown superior capabilities when handling dense defects with irregular edges. Zhang et al.[56] designed a lightweight backbone with a global attention mechanism, achieving high-precision anchor-free defect detection. Kim et al. [46] optimized the advanced anchor-free detector CenterNet by incorporating an adversarial architecture, enabling more precise localization of defects and achieving state-of-the-art results in semiconductor defect detection. Moreover, Liu et al.[47] designed a distillation mechanism to further optimize CenterNet, achieving excellent performance in fabric defect detection tasks. Fang et al.[48] enhanced CenterNet through attention mechanisms, achieving optimal performance in power transmission line defect detection. These findings suggest that CenterNet performs robustly in scenarios involving ambiguous boundaries and unevenly distributed defects. Therefore, this study adopts CenterNet as the baseline model and optimizes it to address the specific requirements of insulator defect detection tasks.

Method

The proposed Context Cluster CenterNet++ (CC-CN) is a bounding-box-free object recognition method that predicts the size, pose, and key points of targets based on their center points. Feature extraction is achieved through a fully convolutional network, optimized using a contextual clustering mechanism to mitigate the adverse effects of uneven target distribution on detection performance. After initial feature extraction by the backbone network, a Path Aggregation Network (PAN) module is employed to fuse features across multiple scales, thereby enhancing the network's efficiency in leveraging multi-scale features. Following feature extraction, a heatmap of target centers is generated, and local maxima within the heatmap are identified to locate target centers. Features at the peak positions are subsequently used to infer target size and pose. By leveraging multi-resolution feature maps for target classification, the proposed algorithm accommodates the challenges posed by deformations and scale variations in drone-acquired images. Furthermore, preprocessing steps such as image stabilization and adaptive-scale feature extraction are incorporated to improve the model's accuracy in identifying insulator defects under various environmental conditions.

Figure 2 illustrates the overall workflow of the proposed network. Initially, the input image is processed by the backbone network (CCR), which incorporates a contextual clustering mechanism for feature extraction. The extracted multi-scale features are then fed into a Path Aggregation Network (PAN) for feature fusion. Subsequently, the fused features are passed to a deformable convolution-based detector for insulator bounding box prediction.

Context clustering-based backbone network (CCR)

In order to overcome the UAV image evenly distributed object distribution to the challenges of target recognition, this paper adopts ResNet101[32] as backbone, and designs a new multiresolution based on Context Clustering CentetNet++ recognition model.

Because of the UAV acquisition is insulator image sequences, and as a result, this paper comprehensive clustering result of each frame image, the context of clustering center position. Concrete, after input sequence images, this article will each frame image of each pixel as one who has a color information and location information of (r, g, b, x, y) 5-D data points, converts the input image $I \in \mathbb{R}^{3\times H\times W}$ to the set of pixels $P \in \mathbb{R}^{5\times N}(N = H \times W)$, and analysis method based on point cloud to target visual representation. Among them, this article on every frame image processing in the paper are as follows: (1) to four stages of the image, every phase processing after points cut (Point the Shrink, PS) clustering and context two step, as shown in Fig. 3. In the image space, PS follows the hierarchical clustering method to select some cluster center points, and the points in their 8-neighborhood are spliced and fused into the cluster region by Linear Projection (LP), so as to reduce the number of data points and improve the computational efficiency. CC step is the characteristics of each cluster area of polymerization and the characteristics of dispatch process, similar to the local attention mechanism, in the image into different clustering regions, characteristics of interaction between various regions, polymerization and then dispatch.

Characteristics of aggregation for each cluster area, according to the corresponding points between frames on Euclidean distance, to determine whether the same feature points, for the same dynamic aggregation of feature

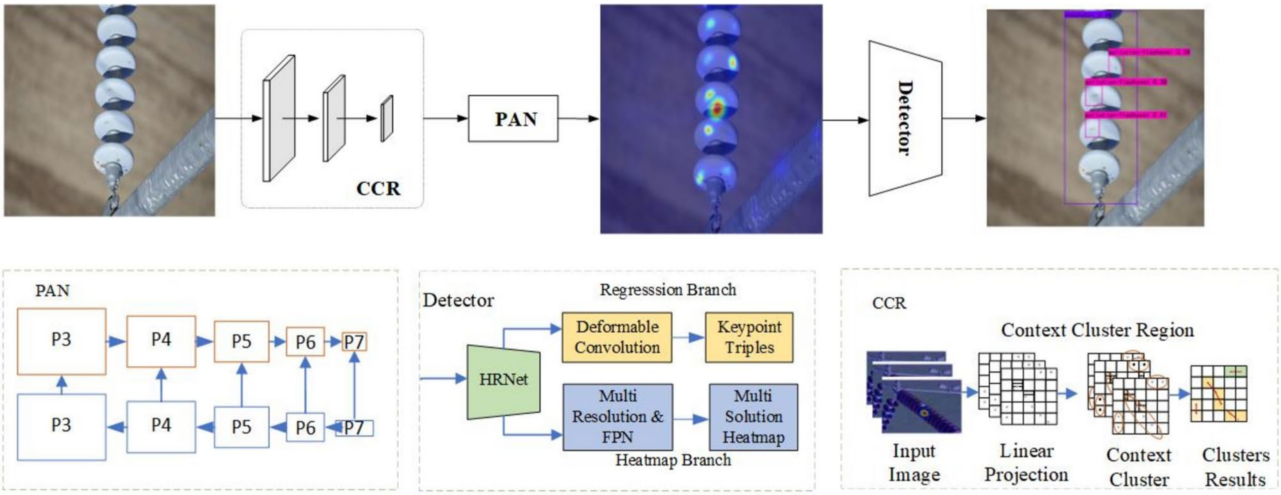

Fig. 2. Context Cluster CenterNet++ framework. The flowchart at the top illustrates the overall structure of the proposed network. The input image is first processed by the CCR backbone for feature extraction, followed by feature fusion using PAN, and finally, the detection results are obtained through the Detector. The three subfigures at the bottom represent the overall structures of PAN, Detector, and CCR, from left to right.

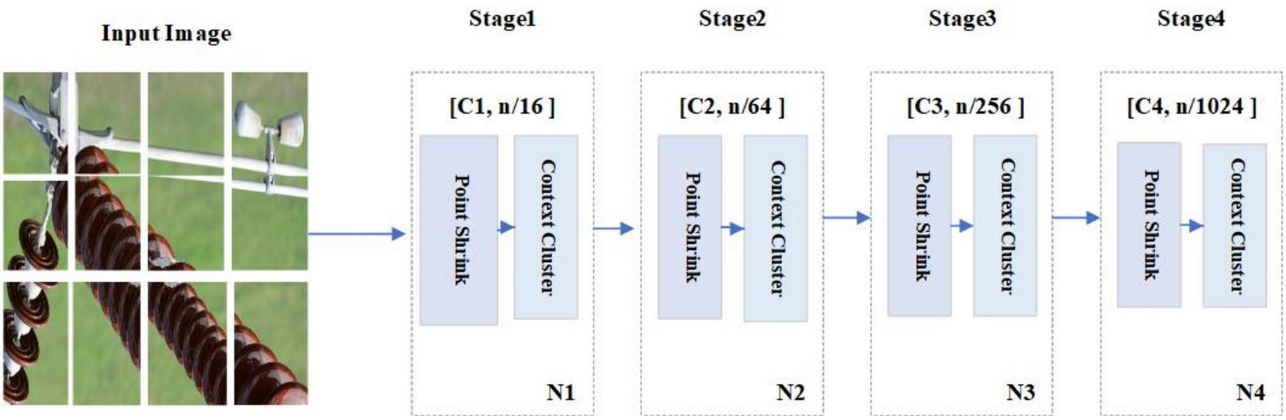

Fig. 3. Context Clustering Process. Where PS denotes Point Shrink and CC denotes Context Cluster.

points, polymerization process, such as formula (3) and (4), as shown in the whole process is the aggregation of points in the group of the characteristics of space projection to a new value.P_v

$$g=\frac{1}{C}\left(v_c+\sum_{i=1}^{m} sig(\alpha s_i+\beta)\times v_i\right) \tag{3}$$

$$C=1+\sum_{i=1}^{m} sig(\alpha s_i+\beta)\times v_i \tag{4}$$

Among them, g said polymerization of target (characteristics), v_c as the center of the cluster area, $s_i \in \mathbb{R}^M$ said the first clustering regions (a total of M cluster area), v_i for the point value space P_v ($\mathrm{P}_v \in \mathbb{R}^{M\times d'}$,$d'$ for clustering region contains the number of points) of the first point i, as the function, α, β scale and offset parameters for learning, C for the normalization factor. i $sig(\cdot)$ is the $sigmoid$ function.

Characteristics of dispatch After getting characteristics after the polymer, according to the characteristics of the dynamic dispatch by the cosine similarity for each point in the cluster, there are:

$$p_i'=p_i+FC(sig(\alpha s_i+\beta)\times g) \tag{5}$$

The FC (Full Connection) is the mean of all connections, mainly from the value space d' to point a space d dimension of feature matching.

After the formation of clustering center, this article selects 10 before the clustering center with larger area, the clustering results to the original input image, and then after CenterNet++ network is used to identify the subsequent reasoning of the identification process.

Deformable convolution-based detector

This paper proposes a backbone network based on contextual clustering for feature extraction from input images. The extracted multi-scale features are fused using a Path Aggregation Network (PAN) layer and then fed into a deformable convolution-based detector for bounding box prediction. The detector comprises two branches: a regression map branch and a heatmap branch. The heatmap branch employs a multi-resolution design, making it more effective in identifying small objects. In the regression map branch, this study adopts the Deformable Convolution v2 (DCNv2) method[33] to replace the fixed kernel convolution approach used in the original model. By introducing an offset to the convolution kernel during training, DCNv2 enables various transformations such as rotation and scale adjustment during the convolution process. This capability effectively addresses challenges associated with small objects and deformations, suppresses background noise, and significantly enhances the precision of small object feature extraction. Furthermore, it improves computational efficiency while introducing only a modest increase in model complexity. Its convolutional process is shown in Fig. 4.

Deformable convolution (DCNv2) can better extract image features and improve the model's ability to adapt to target deformations. By adding an offset to the conventional convolution operation, the sampling position is no longer limited to the traditional rectangular shape, but can be freely deformed to adapt to the geometric deformation target. DCNv2 through modulation mechanism to study the value of each sampling point of key feature set higher weights, to further improve the network model of different attitude, scale and changeable insulator image feature extraction ability.

This paper adopts DCNv2 regression of feature detection module chart branch was improved, the sample point to the target position to offset, to reduce the interference of background. The structure of the proposed deformable convolution-based detector is shown in Fig. 5.

Overall, this study employs parallel computation for the regression branch and the heatmap branch. In the regression branch, predictions are refined using clustering results obtained through context-aware deformable convolutions, along with constrained clustering center offsets, to produce more accurate predictions for object locations and sizes. Additionally, the regression process incorporates the ground truth (GT) to compute clustering-related losses through convolution operations. For the heatmap branch, a Path Aggregation Network (PAN) is utilized to generate multi-resolution heatmaps. Through feature detection calculations, two types of feature maps are derived: one encodes the upper-left and lower-right clustering centers along with key points as triples, which are used to determine the target center and bounding box dimensions; the other is a multi-resolution heatmap. The latter is used to calculate the loss value, as described in Eq. (2), for determining the predicted target location. Finally, the overall loss function, defined in Eq. (6), is applied to optimize and finalize the object localization.

Improved loss function calculation method for nonlinear changes

This article uses the formula (6) the Context Cluster CenterNet++ loss calculation model training.

$$L_{CC-CN} = L_{CC} + \hat{\alpha} L_{reg} + \hat{\beta}\left(L_{co}^{CC} + L_{ce}^{CC}\right) + \hat{\gamma}\left(L_{cooff}^{CC} + L_{ceoff}^{CC}\right) \tag{6}$$

The formula consists of four parts: (1) the clustering loss L_{cc} is calculated by the Focal Loss method; (2) region loss L_{reg} is obtained by calculating GIoU; (3) L_{co}^{CC} and L_{ce}^{CC} represents the prediction loss of the corner and center keypoints in the context clustering region; (4) L_{cooff}^{CC} and L_{ceoff}^{CC} represents the bias loss of the target center point determined by the keypoint triplets in the context clustering region. Due to $L1$ losses in the

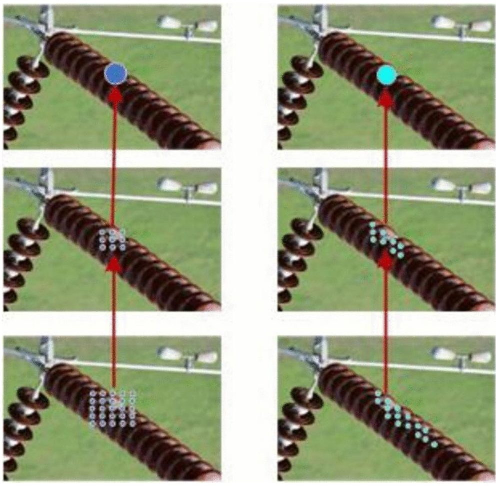

Fig. 4. Comparison between deformable convolution (DCNv2) and standard convolution.

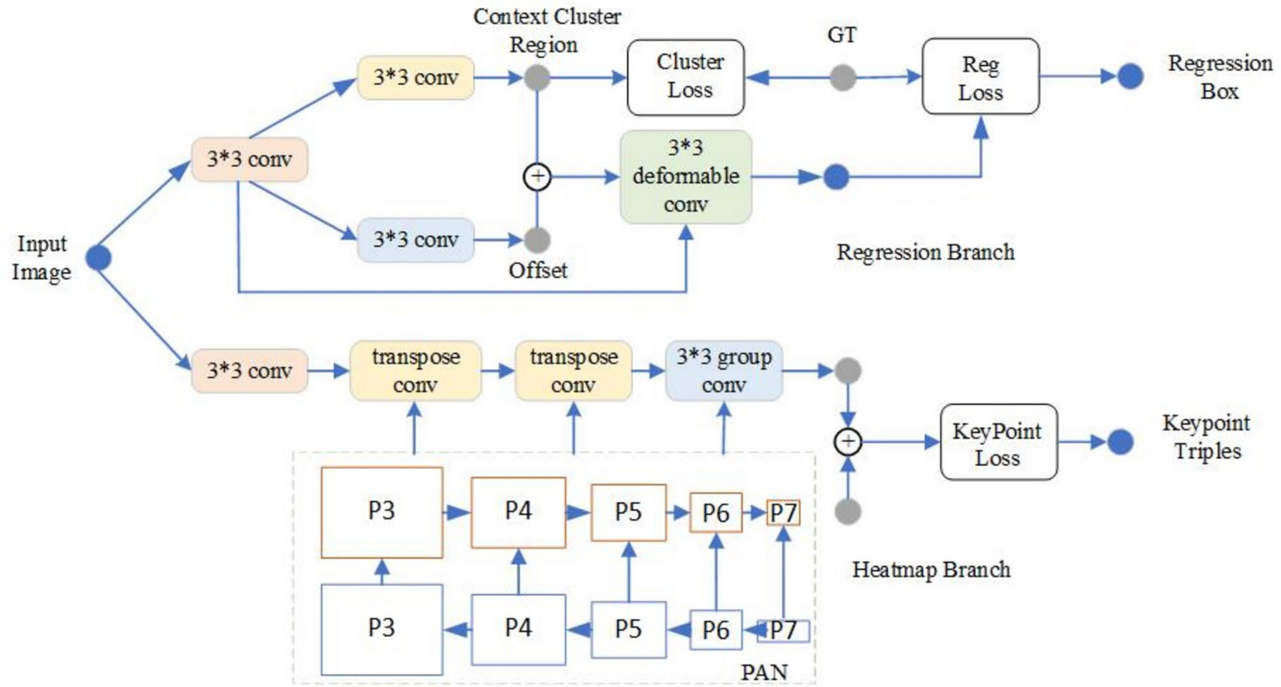

Fig. 5. The Detector module of this paper.

computation of regression is too sensitive to the size of the target change, and neglected to identify the target location and sequence between the correlation between, therefore, of the type (3) and (4) using Bhattacharyya nonlinear distance calculation method proposed in this paper. For the corresponding weight loss, α, β, γ is set to 1, 0.5, and 1 respectively.

In the collected images, target position variations are often characterized by nonlinear random changes. Predicting such variations based solely on linear models frequently results in inaccurate target localization. To address this issue, numerous researchers have explored various optimization algorithms[49,50]. In this study, the Bhattacharyya nonlinear distance calculation method is introduced as a replacement for the loss calculation in Eq. (2), which is employed to compute the center point offset loss in Eq. (6). The Bhattacharyya distance measures the similarity between two discrete probability distributions and is closely related to the Bhattacharyya coefficient, which quantifies the overlap between two discrete sample sets. The definition of the discrete Bhattacharyya coefficient is presented in Eq. (7):

$$c_{Bha} = \sum \sqrt{p(x)q(x)} \tag{7}$$

Among them $p(x)$, said to the target location prediction, $q(x)$ represent the real value of the target location. After getting pap coefficient, pap distance formula (8),

$$d_{Bha} = -\ln\left(c_{Bha}\left[p(x), q(x)\right]\right) \tag{8}$$

Through the type on the computer, can predict box and real box IoU into the similarity between the two distribution. Using nonlinear calculation method of the distance to replace the linear distance calculation method has the advantage of not target dimension in the frame, frame between the limitation of target location changes, can be a significant boost between different frames on the same target positioning accuracy. For small target insulator, in most cases are not the standard rectangular box, therefore, when doing the target location prediction, hard to avoid can appear a lot of the interference of background pixels, makes the forecast does not present a good linear change. Through the calculation of nonlinear distance loss, in order to better describe the nonlinear change of the center of the target.

Experiments and analysis

Used in this section is mainly introduced in this paper, the experimental data set, data preprocessing and data process, evaluation index and the implementation details; It is necessary to develop and then introduces the ablation experiments to verify the proposed innovation part of the backbone of the module, test module and the effectiveness of the improved damage function; Finally, from the Angle of quantitative and qualitative, this article puts forward the model and is the most advanced (the State of the Art, SOTA) model on the basis of the comparison experiment, proved the superiority of the proposed model.

Data collection

Insulator data set

The data set (Table 1) used in this paper is the insulator images collected in the process of standardized UAV inspection by the State Grid Corporation, with a total of 2400 frames and a resolution of 640 × 640. The recognition algorithm is mainly studied for five kinds of defects of insulator: self-explosion, flash, dirty glass, damage and foreign body. The training set consists of 1600 frames, and the total number of tags is 7568, including 2636 self-explosion, 1188 flasher, 1284 dirt, 1140 damage, and 1320 foreign objects. The test set consisted of 800 frames with 503 labels, including 147 self-explosion, 64 flashover, 89 dirty, 123 damaged and 80 foreign objects.

Enhancement of insulator data set

Since the insulator data set collected by UAV is rare, this paper performs data enhancement processing on the images in the training set based on the existing insulator data set: rotation, noise addition, cropping, and color transformation of the images. Finally, the original data set is expanded into 7600 insulator defect data sets containing five types.

Evaluation metrics

In this paper, two evaluation indicators, Class Average Precision (AP) and model Inference Speed (FPS), are used to measure the performance of the proposed model.

AP is the area under the curve of Precision (P) and Recall (R)[34]. The larger the AP, the better the performance of the model. AP calculations for formula shown in (25),

$$AP = \int_0^1 P(R)dR \tag{9}$$

Among them, TP (True Positives), FP (False Positives), FP (False Positives) and FN (False Negatives). $P = \frac{TP}{TP+FP}$ $R = \frac{TP}{TP+FN}$ This paper is chosen as evaluation index, the average precision of all categories, and said in all categories IoU average accuracy of the threshold value is 0.5 and 0.75. AP, AP_{50}, AP_{75} AP AP_{50} AP_{75} In addition, this paper also uses evaluation metrics AP_S, AP_M, AP_L[42] to measure the recognition performance of the model for small, medium and large scale objects.

Model inference speed (frames per second, FPS)

FPS is the number of images detected by the model per second. The larger the FPS value, the better the real-time performance of the algorithm.

The results

This paper respectively in MS COCO data set[43] and standardization of the State grid unmanned aerial vehicle (UAV) inspection insulator image data sets on a wide range of experiment, verify the effectiveness of the proposed model and the most advanced method (State-of-the Art of SOTA). Among them, MS-COCO data set contains 80 categories and more than 1.5 million object instances. A large number of small objects get it to become a widely used dataset. In this paper, train2017 is used as the training data set, test 2017 is used as the test set, and val2017 is used for ablation experiments and visualization experiments.

Ablation experiments

In order to demonstrate the effectiveness of the proposed three methods of innovation, this article described in section 3.1 data set on the ablation experiment. When the innovative method was introduced and the model converged, the comparative results of the ablation experiments were shown in Table 2. Among them, "√" and "×" in the table indicate whether this method is used for innovation, and CC, DCN-KP and Bha in the table represent the three improved methods proposed in this paper.

(a) The influence of the CC. Clustering is proposed in this paper USES the context (CC) method to solve the problem of non-uniform distribution and long tail. From Table 2 row 2 as you can see, when using the method of clustering to return branch image and reduced heat map branch image data points after processing, makes the model improves the IoUs of threshold value is 1 (mAP) and medium target(AP_{50}) big target (AP_{75}),in recognition accuracy (increased by 2.2%, 3.5% and 2.0% respectively). And reasoning speed

Total number of image data		7600	
Training set: 6400		Test set: 1200	
Self-explosion	1170	Self-explosion	307
Flash	1417	flashover	266
Filth	1225	Filth	230
Broken	1631	Broken	340
Foreign bodies	955	Foreign bodies	157

Table 1. All kinds of image data (frames) in the data set.

	CC	DCNKP	Bha	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L	Precision	Recall	FPS
1	x	x	x	51.2	63.5	60.3	32.9	56.4	65.1	62.1	61.0	12.5
2	√	x	x	53.6	65.7	62.4	33.5	57.5	66.4	60.2	65.4	13.6
3	x	√	x	52.3	65.4	62.6	34.8	56.8	64.5	63.5	60.2	12.9
4	x	x	√	52.3	62.3	62.1	35.2	55.3	65.6	62.5	61.8	13.2
5	√	√	x	59.6	68.7	62.5	37.8	56.6	67.8	65.3	63.2	14.2
6	√	x	√	60.3	67.5	60.5	37.3	58.6	68.5	64.1	62.5	13.8
7	x	√	√	61.2	66.5	63.2	36.5	59.5	69.1	65.2	63.5	14.3
8	√	√	√	**60.1**	**75.2**	**63.5**	**38.7**	**60.1**	**69.6**	**68.1**	**64.0**	**13.9**

Table 2. Comparison of ablation experiments. Significant values are in bold.

also had significant increase (11.5%). Is adopted in this paper, and the improved loss function to feedback correction of clustering center, to further improve the precision of recognition (Table 2 rows 5, 6 and 8).

(b) The influence of DCN-KP. In this paper, the deformable convolution operator (DCNv2) combined with Path Aggregation Network (PAN) is used to perform deformable convolution operation on the image, so as to locate the regression box and key point triple (KP) more accurately, so as to improve the accurate positioning of the target position of any shape and any scale, and reduce the sensitivity of the model to the change and deformation of the target scale. To improve the recognition accuracy of the model. From rows 3, 5, 6, and 8 in Table 2, it can be observed that the improved module can improve the overall recognition accuracy of the model, and the recognition accuracy (AP_S) of small targets is particularly effective, with the accuracy increased by 0.5%, 1.5%, 1.1%, and 1.7% respectively.

(c) The influence of the Bha. This article USES the Bha distance to calculate the key points of triples forecast loss and target center offset losses. Using nonlinear calculation method of the distance to replace the linear distance calculation method has the advantage of not target dimension in the frame, frame between the limitation of target location changes, can be a significant boost between different frames on the same target positioning accuracy. From Table 2 rows 4, 6, 7 and 8 lines as you can see, the improvement of the loss function besides can improve the AP_{50} and AP_{75} model and the identification accuracy, on the scale of any target recognition accuracy (AP_S, AP_M, AP_L) increased (0.4%, 0.5%, 1.3%, 1.2%, 1.5%).

In summary, by comparing the results of ablation experiments, it can be seen that when the three improved schemes of CC, DCN-KP and Bha are introduced in turn, the sum will be greatly improved. mAP, AP_{50}, AP_{75} AP_S, AP_M, AP_L. Thus, when the three kinds of improved algorithm introduced gradually, can further improve the accuracy of model identification and efficiency, further verify the feasibility of the scheme.

Model performance comparison experiment

This algorithm with pyramid operation in Faster-RCNN, Mask-RCNN, RetinaNet, YOLOv3, DETR, FOS-DCN with mainstream target identification model of testing box, And CenterNet without testing box, CenterNet++, YOLC recognition model in MC-COCO data set and the collection of data on the contrast experiment of the grid. When each model respectively to achieve convergence. The experimental results as shown in Table 3.

FPN is operated with pyramid, R-101 for ResNet101, R-50 for ResNet50, X-101 for ResNeXt-101, HG-104 for Hourglass-104, and n for the number of target candidate regions to be identified.

Table 3 presents the quantitative comparison with the state-of-the-art method on MS-COCO dataset. The results show that the proposed model performs better than SOTA on all datasets. In general, the method without detection box class is better than the method with box class in recognition accuracy and inference speed. The method proposed in this paper adopts the strategies of context clustering, key point triple and nonlinear loss function, fully considers the influence of adverse factors such as target scale change and deformation on the recognition and detection accuracy, and uses point set instead of region to calculate the target position. Therefore, the recognition efficiency is greatly improved and the calculation amount is greatly reduced under the premise of ensuring the recognition accuracy. In this model it brings great possibilities for landing deployment on downstream tasks such as unmanned aerial systems.

Table 4 presents a comparison of the performance between the proposed method and mainstream detection methods. The table shows a comparison of AP50 results across five categories. By analyzing the results, significant differences in performance across various defect types can be observed, with some methods demonstrating stronger detection capabilities for specific defect types. Among all methods, CC-CN++ (n = 2) excels in all defect categories, achieving the highest AP50 scores. For example, in the dirty detection task, CC-CN++ (n = 2) achieved an AP50 score of 39.8, and for flashover detection, the AP50 score was 35.2, the highest across all methods. This indicates that the CC-CN++ (n = 2) model is particularly effective in handling these defects, owing to its advanced architecture and optimization strategies. Trident Net performs strongly in most defect types, particularly in explosions (34.1) and dirty (34.4), suggesting that this method is effective at detecting large-scale or significant defects. FCOS-DCN also shows good performance across multiple defect types, with an AP50 score of 35.3 for explosions and 36.1 for dirty, close to the performance of CC-CN++ (n = 2). This indicates that FCOS-DCN is a worthy alternative when aiming for high accuracy and speed. In contrast, Yolov3 and RCNN-FPN perform poorly, particularly in the flashover and breakage defect types, with AP50 scores generally below 25. This suggests that these methods may struggle with more complex or subtle defects. The CC-CN++ model

Methods	Backbone	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L	FPS	Speed/ms
Anchor									
RCNN-FPN[6]	R-101	34.9	55.7	20.3	11.6	33.9	54.9	7	142
Mask RCNN[8]	X-101	39.8	62.3	43.4	22.1	43.2	51.2	11	90
RetinaNet[21]	R-101	39.1	59.1	42.3	21.8	42.7	50.2	5.4	185
TridentNet[17]	R-101	48.4	69.7	53.5	31.8	51.3	60.3	23	43
Yolov3[20]	DarkNet	33.0	57.9	34.4	18.3	25.4	41.9	**33.3**	**30**
DETR[10]	R-101	44.9	64.7	47.7	23.7	49.5	62.3	28	35
FCOS-DCN	X-101	50.4	68.9	55.0	33.2	53.0	62.7	27	37
RefineNet	R-101	41.8	62.9	45.7	25.6	45.1	54.1	29	34
Anchor-free									
CenterNet	HG-104	45.1	63.9	49.3	26.6	47.1	57.7	11.9	83
CenterNet++	R-50	46.3	64.7	49.8	26.6	49.6	59.3	13.8	71
CenterNet++ (MS)	R-50	57.1	73.7	62.4	38.7	59.2	71.3	14.8	66
YOLC (k = 1)	HRNet	37.7	61.7	39.0	30.2	48.4	56.4	14	71
YOLC (k = 2)	HRNet	37.8	61.7	39.4	30.5	48.3	55.4	15	67
YOLC (k = 3)	HRNet	38.3	62.3	40.1	31.7	48.0	46.5	12	83
Ours									
CC-CN++ (n = 1)	R-101	59.5	73.9	**65.2**	**39.2**	59.0	**71.2**	15	67
CC-CN++ (n = 2)	R-101	**60.1**	**75.2**	63.5	38.7	**60.1**	69.6	16	62
CC-CN++ (n = 3)	R-101	54.5	74.1	62.9	38.5	59.5	60.1	13.9	71

Table 3. Comparison of the performance of different algorithms. Significant values are in bold.

Methods	explods	flashover	dirty	breakage	foreign
RCNN-FPN[6]	24.3	21.1	24.7	21.2	21.3
Mask RCNN[8]	28.2	24.8	28.6	24.1	25.9
RetinaNet[21]	27.4	24.2	27.8	23.3	25.2
TridentNet[17]	34.1	29.6	34.4	29.2	31.5
Yolov3[20]	23.2	20.5	23.6	19.7	21.3
DETR[10]	31.7	27.9	32.1	26.8	29.1
FCOS-DCN[22]	35.3	31.3	36.1	30.3	32.3
RefineNet	29.4	25.9	29.7	23.7	25.7
CenterNet	25.1	23.9	33.1	19.3	26.6
CenterNet++	26.3	24.7	38.2	29.8	26.6
CC-CN++(n = 1)	29.5	24.9	39.5	35.2	39.2
CC-CN++(n = 2)	30.1	35.2	39.8	33.5	38.7
CC-CN++(n = 3)	**38.7**	**34.1**	**39.2**	**32.9**	**35.5**

Table 4. Comparison of the performance of different defect. The table shows metrics in terms of AP_{50}. Significant values are in bold.

shows varying trends with different values of n. As the n value increases, performance improves. For example, the transition from CC-CN++ (n = 1) to CC-CN++ (n = 2) significantly enhances the detection ability for flashover and dirty, while CC-CN++ (n = 3), although showing some decline in certain defect types such as breakage and foreign, still performs well in most categories. This suggests that as model complexity increases, performance improves, but excessive complexity may lead to performance degradation. Therefore, when choosing the value of **n**, there is a need to balance model complexity and computational cost. Traditional methods like Yolov3 and RCNN-FPN generally show poor performance, especially in detecting complex or subtle defects, with low AP50 scores across the board. This is likely due to limitations in feature extraction capabilities, making it difficult for these methods to detect fine or hard-to-distinguish defects.

Cross-validation

To evaluate the stability of the proposed model, k-fold cross-validation was conducted on the insulator fault detection dataset. Specifically, the dataset was evenly divided into eight subsets, with seven subsets used as the training set and the remaining subset as the test set in each iteration. This process was repeated eight times to ensure comprehensive evaluation. Table 5 summarizes the performance metrics obtained from the cross-

Fold	explods	flashover	dirty	breakage	foreign
Fold 1	38.6	34.3	39.1	33.0	35.4
Fold 2	38.8	34.0	39.3	33.1	35.6
Fold 3	38.7	34.2	39.2	33.0	35.5
Fold 4	38.9	34.1	39.4	32.8	35.3
Fold 5	38.5	34.0	39.3	33.2	35.4
Fold 6	38.8	34.2	39.1	33.0	35.6
Fold 7	38.7	34.3	39.2	32.9	35.5
Fold 8	38.9	34.1	39.3	32.8	35.4

Table 5. Cross-validation metric presentation. The table shows metrics in terms of AP_{50}.

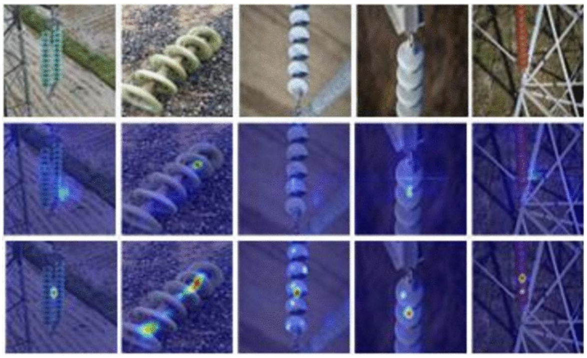

Fig.6. Comparison of attention maps. Top to bottom are the original image, CenterNet and the attention graph generated by the model in this paper respectively.

validation. The results demonstrate that the proposed model exhibited consistent performance across all folds, confirming its reliability and generalization capability.

Attention map comparison

Figure 6 presents a comparison of attention maps generated by different methods. The first row displays the original images, the second row shows the attention maps produced by CenterNet, and the third row illustrates the attention maps generated by the proposed model. The comparison reveals that the attention maps from the proposed model more accurately focus on the fault regions of insulators, demonstrating a clearer distribution of fault features that aligns closely with the fault areas in the original images. In contrast to CenterNet, the proposed model exhibits significant improvements, highlighting its superior feature extraction capabilities and more targeted attention mechanism for fault detection tasks.

Conclusion

This paper analyzes the characteristics of insulator defect recognition from UAV inspection images, and proposes a new multi-resolution context clustering CenterNet++ model to solve the existing problems. (1) The clustering area is used to identify and predict the location of the target, and the improved loss function is used to feedback correct the clustering center, which improves the recognition accuracy and further improves the computational efficiency. (2) The deformable convolution operator combined with the path aggregation network is used to perform deformable convolution operation on the image, and the regression box and key point triples are accurately predicted, so as to improve the accurate positioning of the target position of any shape and any scale, reduce the sensitivity of the model to target scale changes and deformation, and improve the recognition accuracy of the model. (3) Use Bhattacharyya distance to calculate the key point of triples forecast loss and target center offset losses.

This study demonstrates that contextual clustering and deformable convolution play a significant role in enhancing performance for detection tasks involving irregular boundaries. The proposed approach shows great potential for extension to a broader range of fault detection tasks. While the method achieves promising results in insulator detection tasks from a drone perspective, its inference speed remains suboptimal. Future work will focus on optimizing the model's parameter efficiency through techniques such as model pruning and knowledge distillation, aiming to ensure its suitability for more challenging high-resolution input images.

Data availability

All the code and details of the experiment have been uploaded to the web. https://github.com/mengbonannan88/CC-CenterNet.

Received: 5 October 2024; Accepted: 6 January 2025
Published online: 17 January 2025

References

1. Park, K. C., Motai, Y. & Yoon, J. R. Acoustic fault detection technique for high-power insulators. *IEEE Trans. Industr. Electron.* **64**(12), 9699–9708 (2017).
2. Han, J. et al. A method of insulator faults detection in aerial images for high-voltage transmission lines inspection. *Appl. Sci.* **9**(10), 2009 (2019).
3. Pan, C. et al. Research on real-time positioning of insulator strings based on UAV inspection images. *Insul. Surge Arrester* **1**, 234–240 (2020).
4. Chen, H., et al. Top-down meets bottom- up for instance segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* 8573– 8581 (2020).
5. Girshick, R., Donahue, J., Darrell, T. & Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. Vol. 2, 580–587 (2014).
6. Girshick, R. Fast R-CNN. In *Proceedings of the IEEE International Conference on Computer Vision* 1440–1448 (2015).
7. Ren, S., He, K., Ross, G. & Sun, J. Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **39**(6), 1137–1149 (2017).
8. He, K., Gkioxari, G., Dollar, P. & Girshick, R. Mask r-cnn. *IEEE Trans. Pattern Anal. Mach. Intell.* **42**(2), 386–397 (2018).
9. Liu, W. et al. SSD: Single shot multibox detector. In *Proceedings of the Computer Vision–ECCV 2016, Amsterdam, The Netherlands, October, 2016, Proceedings, Part I* vol. 14 21–37 (Springer International Publishing, 2016).
10. Zhu, X. et al. Deformable detr: Deformable transformers for end-to-end object detection. Preprint at http://arxiv.org/abs/2010.04159 (2020).
11. Zhang, H. et al. Dino: Detr with improved denoising anchor boxes for end-to-end object detection. Preprint at http://arxiv.org/abs/2203.03605 (2022).
12. Sunkara, R. & Luo, T. No more strided convolutions or pooling: A new CNN building block for low-resolution images and small objects. In *Proceedings of the Joint European Conference on Machine Learning and Knowledge Discovery in Databases* 443–459 (Springer Nature Switzerland, 2022).
13. Gidaris, S. & Komodakis, N. Object detection via a multi-region and semantic segmentation-aware cnn model. In *Proceedings of the IEEE International Conference on Computer Vision* 1134–1142 (2015).
14. 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* 6154–6162 (2018).
15. Tychsen-Smith, L. & Petersson, L. Improving object localization with fitness nms and bounded iou loss. In *Proceedings of the IEEE conference on computer vision and pattern recognition* 6877–6885 (2018).
16. Dai, J., Li, Y., He, K. & Sun, J. R-fcn: Object detection via region-based fully convolutional networks. *Adv. Neural Inf. Process. Syst.* 379–387 (2016).
17. Li, Y., Chen, Y., Wang, N. & Zhang, Z. Scale-aware trident networks for object detection. In *Proceedings of the IEEE International Conference on Computer Vision* 6054– 6063 (2019).
18. Kong, T. et al. Ron: Reverse connection with objectness prior networks for object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 5936–5944 (2017).
19. Wang, C. Y., Bochkovskiy, A. & Liao, H. Y. M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* 7464–7475 (2023).
20. Redmon, J. & Farhadi, A. Yolov3: An incremental improvement. In *Proceedings of the Computer Vision and Pattern Recognition* 1804–2767 (IEEE, 2018).
21. Law, H. & Deng, J. Cornernet: Detecting objects as paired key-points. In *Proceedings of the European Conference on Computer Vision* 734–750 (2018).
22. Tian, Z., Shen, C., Chen, H. & He, T. FCOS: A simple and strong anchor-free object detector. *IEEE Trans. Pattern Anal. Mach. Intell.* 99 (2020).
23. Zhou, X., Wang, D. & Kr¨ahenb¨uhl, P. Objects as points. Preprint http://arxiv.org/abs/1904.07850 (2019).
24. Duan, K. et al. Centernet: Key-point triplets for object detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* 6569–6578 (2019).
25. Law, H., Teng, Y., Russakovsky, O. & Deng, J. Cornernet-lite: Efficient key-point based object detection. Preprint at http://arxiv.org/abs/1904.08900 (2019).
26. Zhou, X., Zhuo, J. & Krahenbuhl, P. Bottom- up object detection by grouping extreme and center points. In *Proceedings of the IEEE Conference on Computer Vision and Pattern, Recognition* 850–859 (2019).
27. Xu, C. et al. Detecting tiny objects in aerial images: A normalized Wasserstein distance and a new benchmark. *ISPRS J. Photogramm. Remote Sens.* **190**, 79–93 (2022).
28. Dong, Z. et al. Centripetalnet: Pursuing high-quality key-point pairs for object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 10519–10528 (2020).
29. Duan, K. W. et al. CenterNet++ for object detection. In *Proceedings of the European Conference on Computer Vision* 740–755 (Springer, 2022).
30. Lin, T. Y. et al. Feature pyramid networks for object detection. In *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition* 2117–2125 (IEEE, 2017).
31. Lin, T.-Y., Goyal, P., Girshick, R., He, K. & Dollár, P. Focal loss for dense object detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **42**(2), 318–327 (2018).
32. Sun, K. et al. High-resolution representations for labeling pixels and regions. Preprint at http://arxiv.org/abs/1904.04514 (2019).
33. Zhu, X. et al. Deformable convnets v2: More deformable, better results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* 9308–9316 (2019).
34. Zhang, Y. F. et al. Focal and efficient IOU loss for accurate bounding box regression. *Neurocomputing* **506**, 146–157 (2022).
35. Carion, N. End-to-end object detection with transformers. In *European Conference on Computer Vision* vol. 8 213–229 (Springer, 2020).
36. Wang, J. et al. A normalized Gaussian Wasserstein distance for tiny object detection. Preprint at http://arxiv.org/abs/2110.13389 (2021).
37. Zhu, L. et al. BiFormer: Vision transformer with bi-level routing attention. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* 10323–10333 (2023).
38. Duan, K. et al. Location-sensitive visual recognition with cross-iou loss. Preprint at http://arxiv.org/abs/2104.04899 (2022).
39. Zhang, S., Chi, C., Yao, Y., Lei, Z. & Li, S. Z. Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 9759–9768 (2020). 3, 8.
40. Li, Y. et al. Overcoming classifier imbalance for long-tail object detection with balanced group softmax. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* 10991–11000 (2020).

41. Du, B., Huang, Y., Chen, J. & Huang, D. Adaptive sparse convolutional networks with global context enhancement for faster object detection on drone images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* 13435–13444 (2023).
42. Cheng, G. et al. Towards large-scale small object detection: Survey and benchmarks. *IEEE Trans. Pattern Anal. Mach. Intell.* 13467–13488 (2023).
43. Xiao, L., Wu, B. & Hu, Y. Surface defect detection using image pyramid. *IEEE Sens. J.* **20**(13), 7181–7188 (2020).
44. Tang, J. et al. A lightweight surface defect detection framework combined with dual-domain attention mechanism. *Expert Syst. Appl.* **238**, 121726 (2024).
45. Zhong, H. et al. STFE-Net: A multi-stage approach to enhance statistical texture feature for defect detection on metal surfaces. *Adv. Eng. Inform.* **61**, 102437 (2024).
46. Kim, J. et al. Adversarial defect detection in semiconductor manufacturing process. *IEEE Trans. Semicond. Manuf.* **34**(3), 365–371. https://doi.org/10.1109/TSM.2021.3089869 (2021).
47. Liu, Z. et al. D-CenterNet: An anchor-free detector with knowledge distillation for industrial defect detection. *IEEE Trans. Instrum. Meas.* **71**, 1–12. https://doi.org/10.1109/TIM.2022.3204332 (2022).
48. Wang, F. et al. Internal defect detection of overhead aluminum conductor composite core transmission lines with an inspection robot and computer vision. *IEEE Trans. Instrum. Meas.* **72**, 1–16. https://doi.org/10.1109/TIM.2023.3265104 (2023).
49. Mehta, P. et al. MOBBO: A multiobjective brown bear optimization algorithm for solving constrained structural optimization problems. *J. Optim.* **2024**, 5546940.
50. Mashru, Tejani, N. & Ghanshyam, G. Optimal truss design with MOHO: A multi-objective optimization perspective. **19**, 1–37 (2024).
51. Xiao, L., Wu, B., Hu, Y. & Liu, J. A hierarchical features-based model for freight train defect inspection. *IEEE Sens. J.* **20**(5), 2671–2678 (2019).
52. Xiao, L., Lu, M. & Huang, H. Detection of powder bed defects in selective laser sintering using convolutional neural network. *Int. J. Adv. Manuf. Technol.* **107**, 2485–2496 (2020).
53. Xiao, L., Wu, B. & Hu, Y. Missing small fastener detection using deep learning. *IEEE Trans. Instrum. Meas.* **70**, 1–9 (2020).
54. Xiao, L., Huang, T., Wu, B., Hu, Y. & Zhou, J. Surface defect detection using hierarchical features. In *2019 IEEE 15th International Conference on Automation Science and Engineering (CASE)* 1592–1596 (IEEE, 2019).
55. Zhong, H. et al. LiFSO-Net: A lightweight feature screening optimization network for complex-scale flat metal defect detection. *Knowl.-Based Syst.* **304**, 112520 (2024).
56. Zhang, Y., Zhang, H., Huang, Q., Han, Y. & Zhao, M. DsP-YOLO: An anchor-free network with DsPAN for small object detection of multiscale defects. *Expert Syst. Appl.* **241**, 122669 (2024).
57. Jia, Y., Chen, G. & Zhao, L. Defect detection of photovoltaic modules based on improved VarifocalNet. *Sci. Rep.* **14**, 15170 (2024).
58. Meng, X., Lin, L., Li, H., Chen, Y. & Mei, H. Characteristics of streamer discharge along the insulation surface with embedded electrode. *IEEE Trans. Dielectr. Electr. Insul.* **31**(4), 2038–2044 (2024).
59. Yang, J. et al. A hierarchical deep convolutional neural network and gated recurrent unit framework for structural damage detection. *Inf. Sci.* **540**, 117–130 (2020).
60. Deng, X., Zhang, Y., Zhang, Y. & Qi, H. Toward smart multizone HVAC control by combining context-aware system and deep reinforcement learning. *IEEE Internet Things J.* **9**(21), 21010–21024 (2022).

Acknowledgements

This work was supported by National Natural Science Foundation of China (61602108), and the Electric Power Intelligent Robot Collaborative Innovation Group.

Author contributions

Bo Meng did the experiments and wrote the whole manuscript text.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-85630-x.

Correspondence and requests for materials should be addressed to B.M.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.