electronics

Article

Detection of Fittings Based on the Dynamic Graph CNN and U-Net Embedded with Bi-Level Routing Attention

Zhihui Xie [1], Min Fu [2,3,*] and Xuefeng Liu [1,4,*]

[1] College of Automation and Electronic Engineering, Qingdao University of Science and Technology, Qingdao 266100, China; 2021040020@mails.qust.edu.cn
[2] College of Electronic Engineering, Ocean University of China, Qingdao 266100, China
[3] Sanya Oceanographic Institution, Ocean University of China, Sanya 572024, China
[4] Shandong Key Laboratory of Autonomous Landing for Deep Space Exploration, Qingdao 266100, China
* Correspondence: fumin@ouc.edu.cn (M.F.); snowclub@qust.edu.cn (X.L.)

Abstract: Accurate detection of power fittings is crucial for identifying defects or faults in these components, which is essential for assessing the safety and stability of the power system. However, the accuracy of fittings detection is affected by a complex background, small target sizes, and overlapping fittings in the images. To address these challenges, a fittings detection method based on the dynamic graph convolutional neural network (DGCNN) and U-shaped network (U-Net) is proposed, which combines three-dimensional detection with two-dimensional object detection. Firstly, the bi-level routing attention mechanism is incorporated into the lightweight U-Net network to enhance feature extraction for detecting the fittings boundary. Secondly, pseudo-point cloud data are synthesized by transforming the depth map generated by the Lite-Mono algorithm and its corresponding RGB fittings image. The DGCNN algorithm is then employed to extract obscured fittings features, contributing to the final refinement of the results. This process helps alleviate the issue of occlusions among targets and further enhances the precision of fittings detection. Finally, the proposed method is evaluated using a custom dataset of fittings, and comparative studies are conducted. The experimental results illustrate the promising potential of the proposed approach in enhancing features and extracting information from fittings images.

Keywords: fittings; automatic inspection; U-Net; DGCNN; attention mechanisms; Lite-Mono

Citation: Xie, Z.; Fu, M.; Liu, X. Detection of Fittings Based on the Dynamic Graph CNN and U-Net Embedded with Bi-Level Routing Attention. *Electronics* **2023**, *12*, 4611. https://doi.org/10.3390/electronics12224611

Academic Editors: Haibin Wu, Aili Wang and Yuji Iwahori

Received: 23 October 2023
Revised: 8 November 2023
Accepted: 8 November 2023
Published: 11 November 2023

1. Introduction

Power line inspection is a crucial aspect of power line management, as it helps in identifying issues, mitigating risks, and ensuring the reliability of electricity production. However, the current approach to inspecting electrical grid facilities heavily relies on manual labor, which poses challenges in terms of time, labor intensity, and safety concerns [1]. Therefore, there is a need to shift towards intelligent inspection methods that are automated and less reliant on manual efforts. In this regard, the use of computer vision and drone operations aligns with the requirements of intelligent and automated power grids in the Industry 4.0 era. Drone line patrol operations exhibit advanced, scientific, and efficient characteristics, making them an ideal solution for collecting transmission line images. This approach reduces labor intensity and costs while providing a safer and more reliable means of inspection [2].

Power fittings serve as metallic attachments used to suspend, secure, and reinforce conductors or towers, thereby ensuring the dependability of power system. Furthermore, fittings target detection is a crucial component of transmission line inspection [3]. However, the accuracy of fittings detection is affected by a complex background, small target sizes, and overlapping fittings in the images. Additionally, the features extracted by many detection algorithms exhibit significant redundancy, impacting the accuracy of intelligent

fittings inspection. Consequently, intelligent fittings detection remains a focal point in smart grid research [4,5].

Convolutional neural networks (CNN) have attained relatively advanced performance across various domains, with particular prominence in computer vision [6]. Deep learning methods are also constantly evolving in the field of the intelligent detection of power fittings [7]. Many researchers have studied this problem using approaches based on CNN. Luo et al. [8] introduced an ultra-compact model for detecting bolt defects based on a CNN, an approach that enables end-to-end detection of bolt defects through a two-stage detection process. In addition, Wan et al. [9] employed a region-based fully CNN to integrate fine-grained features and contextual information among fittings, enhancing the detection accuracy. However, the neural network employed in the above method has a complex structure with many layers, and its scope of application is uncertain.

In the domain of two-dimensional (2D) detection, RGB images are vulnerable to various complicating factors, including occlusion, lighting conditions, and weather effects. In addition, 2D detection cannot determine the three-dimensional (3D) spatial positions of objects, and extracting features from occluded objects remains a challenging task [10]. Consequently, some methods take advantage of the abundant depth information of point clouds and the ability to accurately locate the target, forming a 3D detection method based on 2D data upgrading. Wu et al. [11] introduced a confidence-guided data association method to address challenges such as occlusion and missed detections of distant objects in tracking. This method leverages the geometric, appearance, and motion features of objects in point clouds, associating the predicted and detected states by predicting confidences and aggregating pairwise costs. Chen et al. [12] utilized geometric constraint relationships to construct an equation system for solving object position information by incorporating camera intrinsic parameters with object physical dimensions and orientation information. Wang et al. [13] proposed a 3D multi-object tracking framework, which first employs PointRCNN [14] and recurrent rolling convolution [15] to separately obtain 3D and 2D detections of objects. Then a multi-stage depth association mechanism is devised solely utilizing object motion information to achieve 3D multi-object tracking, focusing on occluded objects.

Through a review of the existing literature, it appears that the method of converting 2D data to 3D for processing fittings images has not been previously employed. To address the challenges in fittings image detection, such as complex image backgrounds and a certain degree of occlusion among multiple objects, a detection method of fittings based on the U-shaped network and dynamic graph convolutional neural network (UD-Net) is herein proposed. The effectiveness of this method is evaluated through several experimental setups. First, a U-shaped network (U-Net) is employed to augment the extraction capability of fittings features. Then, the Lite-Mono algorithm is deployed to generate depth maps for the fittings. Following the fusion of the depth maps with the fittings images, these are fed into a 3D detection network, thereby optimizing 2D object detection through the leverage of 3D detection. The contributions of the paper are as follows:

- A fittings inspection image dataset is constructed: The fittings dataset comprises 2563 inspection images that have been meticulously annotated using the LabelImg tool, encompassing seven distinct fittings component types. This comprehensive dataset, characterized by its diverse scenarios, ensures robust model training;
- The UD-Net detection network is proposed: First, an improved U-Net serves as the backbone for initial extraction of fittings features. Then, incorporating the Lite Mono algorithm and employing the dynamic graph CNN (DGCNN), we aim to detect and extract obscured fittings feature information;
- Enhanced U-Net: First, to improve the computational efficiency, the width of the U-Net is narrowed to reduce the parameter volume. Then, four attention modules are embedded to bolster the model's feature extraction capability in complex backgrounds, addressing the issue of diminished target salience resulting from mutual occlusion among objects;

- Introduction of 3D-detection-driven 2D detection methods into the fittings detection field: First, the Lite mono algorithm is used to generate a depth image of the fittings, and then this depth map is combined with the corresponding RGB images to create a point cloud dataset. Finally, a 3D detection network is employed to capture features that may elude 2D detection algorithms, contributing to the final refinement of the results.

2. Related Works

2.1. U-Net

The U-Net architecture is designed with a symmetrical encoder–decoder structure, distinctively exhibiting a U-shaped topology [16]. The design integrates both encoding and decoding pathways. The encoding pathway, focused on extracting contextual feature information, consists of convolutional blocks, max pooling operations, and ReLU activation functions [17,18]. The ReLU function primarily aids in introducing non-linearity within the model, with the computation given by:

$$f(x) = \max(0, x) \tag{1}$$

where x represents the input value, and $f(x)$ represents the corresponding output value.

After inputting the image into the network, it undergoes four downsampling operations, resulting in feature maps with twice the number of channels. This procedure adeptly extracts high-dimensional features while retaining both global and semantic information. The decoding path parallels the encoding path, featuring convolutional blocks and upsampling operations. Transpose convolutions achieve fourfold upsampling to extract depth information. During the upsampling phase, skip connections merge shallow and deep information from the encoding and decoding pathways, respectively. Finally, in a culmination of this procedure, a 2 × 2 deconvolution block is employed to restore the image resolution, producing the final output.

U-Net [19] is often used in the automatic detection of power system transmission lines. Its symmetrical encoding and decoding structure offers high detection accuracy paired with a simple network topology. For example, He et al. [20] proposed a transmission line and tower segmentation network based on an improved U-Net, which employs a fully connected backbone structure for feature extraction and a hybrid feature extraction module to refine semantic features, thus enabling high-precision segmentation. Han et al. [21] proposed a lightweight U-Net model integrated with GhostNet [22] to enhance the accuracy of transmission line segmentation results. Choi et al. [23] introduced a power line segmentation method based on U-Net. This method involves the combination of visible images and infrared images of transmission lines using a U-Net embedded with attention mechanism, resulting in successful segmentation outcomes.

2.2. DGCNN

In recent years, 3D object detection has seen significant advancements, with PointNet [24] leading the way in combining graph neural networks with point clouds. He et al. [25] proposed sparse voxel-graph attention network (SVGA-Net), which emphasizes advancements in feature extraction and the establishment of a global graph to bolster performance in 3D object detection. Notably, SVGA-Net addresses a pivotal concern overlooked in previous models such as PointNet, ShapeContextNet [26,27], and the PointNet series [28]—the disregarding of inter-point relationships. Wang et al. [29] proposed DGCNN, a network designed for learning using point clouds. DGCNN utilizes edge convolution to extract edge features between points and their neighboring points, effectively capturing the local geometric structure of point clouds. By employing multiple layers of edge convolution, DGCNN generates diverse neighborhood graphs that facilitate the propagation of point information throughout the data. This approach enables the network to select the most suitable neighbors in the feature space, thereby improving its classification performance [30].

Centered around the DGCNN algorithm, Gamal et al. [31] proffered a building segmentation method, which involves the direct segmentation of buildings using light detection and ranging data and employs the DGCNN algorithm to distinguish buildings from vegetation. Xing et al. [32] delineated a technique for extracting geometric features using DGCNN to ascertain a target sphere position within the fully mechanized mining face. Liang et al. [33] introduced a medical image segmentation network based on DGCNN. The approach involves initially employing a dual-path CNN network to segment the boundary of lesion areas in medical images. Subsequently, the preprocessed medical images are reclassified using the DGCNN network, enhancing the segmentation capability of the overall network. The aforementioned methods proposed around DGCNN stand out by dynamically constructing a graph at every layer, eschewing the need for a pre-constructed, static graph. This methodology exhibits superior performance in both classification and segmentation tasks.

3. UD-Net

To enhance the accuracy of fittings target detection, this paper presents a novel fittings detection method based on UD-Net. The architecture of UD-Net is depicted in Figure 1. As the figure shows, the BRA-UNet, which is the U-Net embedded with four bi-level routing attention (BRA) modules, serves as the foundation for extracting fitting features. The Lite-Mono network is then utilized to reconstruct depth maps for fittings, and the information derived from these depth maps is merged with the RGB fittings images to produce pseudo point cloud data. Following this, the preliminary 2D object bounding boxes identified by the BRA-UNet network are converted into 3D object bounding boxes, which are combined with the pseudo point cloud data for fitting objects. Finally, the recognition results are refined using the DGCNN network.

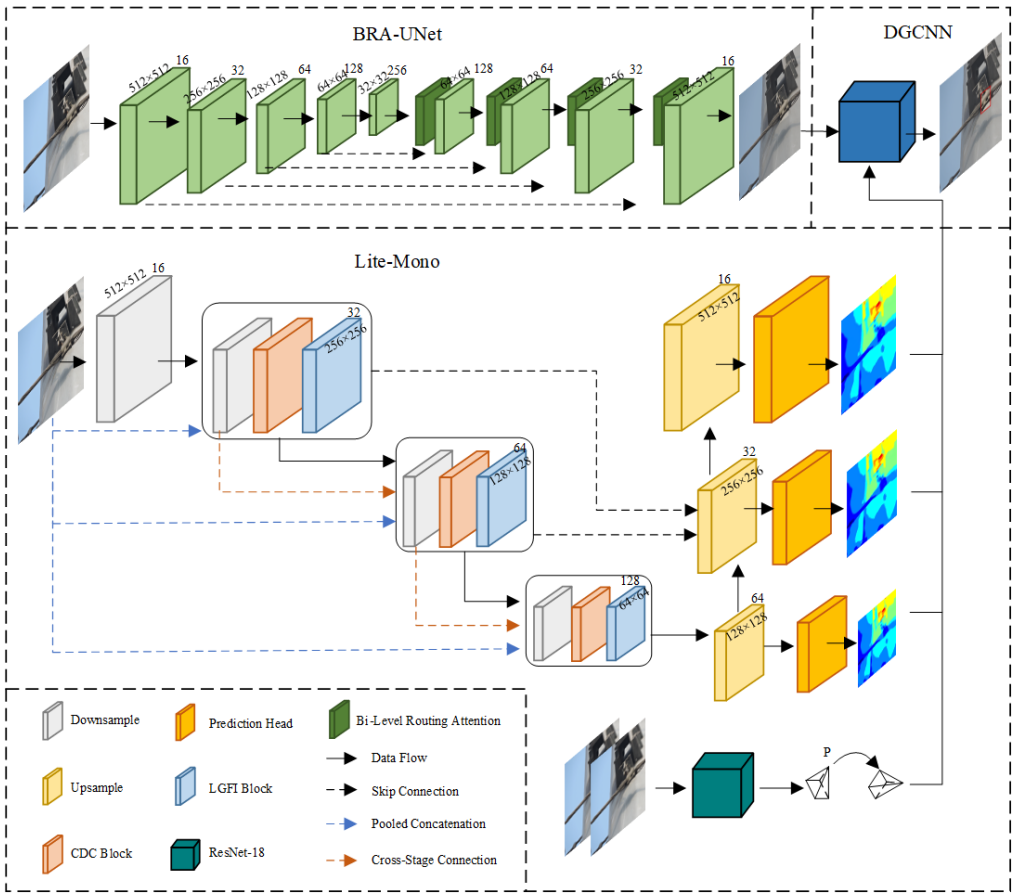

Figure 1. The framework diagram of the UD-Net. The BRA-UNet is used for preliminary feature extraction, and then combined with the Lite Mono algorithm and DGCNN to refine the recognition results.

3.1. BRA-UNet

To augment computational efficiency, the number of encoder blocks is reduced from 5 to 4 in the U-Net model, and the number of convolutional channels in each module is halved. This adjustment balances the increase in parameters resulting from the DGCNN network integration while maintaining an equilibrium between resource usage and performance efficacy.

When employing the U-Net network for feature extraction, it becomes difficult to identify the characteristics and contour details of smaller objects, thereby exacerbating the complexity of fittings detection. Attention mechanisms in deep learning draw inspiration from human visual cognition [34,35]. These mechanisms allow neural networks to autonomously learn and selectively emphasize essential information during input data processing, ultimately bolstering model performance [36]. One such mechanism is the BRA mechanism [37]. Figure 2 depicts the architecture of the BRA mechanism. A feature map is inputted and a query, key, and value are obtained through linear mapping. Then, a directed graph is constructed using an adjacency matrix to find the participation relationship between different key–value pairs. After obtaining the region-to-region routing index matrix, a fine-grained token-to-token attention mechanism is applied. These operations involve GPU-friendly dense matrix multiplications, which are advantageous for accelerating inference on the server-side. Moreover, the BRA mechanism excels at distinguishing between the background and foreground, capturing a wealth of features, and expanding the receptive field and contextual information. This substantially boosts the model's performance. Therefore, in this work, the BRA mechanism is incorporated into the upsampling layer of the U-Net network to enhance its feature extraction capability.

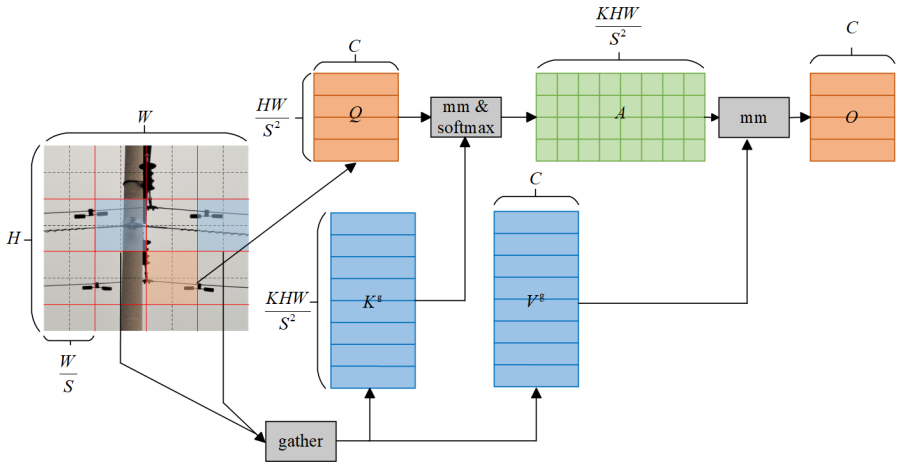

Figure 2. BRA attention mechanism structure. The mechanism aggregates key–value pairs and employs sparse operations to bypass calculations in the less relevant regions, resulting in savings in terms of parameters and computational resources.

3.2. Depth Map Generation

In the pursuit of improving 2D detection outcomes through the use of a 3D detector, a major challenge arises from the lack of a comprehensive and accurate fittings depth dataset. To address this challenge, the Lite-Mono network was introduced [38]. The Lite-Mono network is a cutting-edge framework designed to tackle the complex task of monocular depth estimation. This innovative system combines the computational efficiency of CNN with the sophisticated contextual understanding capabilities of transformer models, all within a self-supervised learning paradigm.

The Lite-Mono network consists of two key components: DepthNet and PoseNet. DepthNet is responsible for estimating multiscale depth maps from input images. Within its encoder section, DepthNet leverages a series of consecutive dilated convolutions (CDC) modules to augment the receptive field of the initial shallow CNN layers. These CDC modules employ dilated convolutions for the extraction of local features at multiple scales. A

suite of dilated convolutions, each with distinct rates of dilation, is strategically embedded along the encoding pathway, facilitating effective multi-scale contextual aggregation. In the decoding phase, DepthNet utilizes bilinear upsampling layers to expand feature map dimensions, thereby improving spatial resolution. Simultaneously, convolutional layers connect features from three encoder stages to ensure seamless information flow to the decoder. Additionally, varying resolutions of output are achieved by attaching a prediction head after to each upsampling, which yields inverse depth maps at assorted scales. PoseNet utilizes a pre-trained ResNet-18 model as its pose encoder, processing pairs of color images for input. It is designed to assess a camera's movement across consecutive frames, culminating in the creation of a reconstructed target image. This methodology transforms the depth estimation problem into an image reconstruction problem. To optimize the model, a loss function is then calculated. The computation process of its loss function is as follows:

$$L_p\left(\hat{I}_t, I_t\right) = \alpha \frac{1 - \mathrm{SSIM}\left(\hat{I}_t, I_t\right)}{2} + (1 - \alpha)\left\|\hat{I}_t - I_t\right\| \tag{2}$$

where I_t is the target image, $\hat{I}_t$ is the reconstructed image, $L_p\left(\hat{I}_t, I_t\right)$ is the loss between the I_t and $\hat{I}_t$, SSIM is the structural similarity index, and α is 0.85. Additionally, the loss of minimum photometric $L_p(I_s, I_t)$ is calculated:

$$L_p(I_s, I_t) = \min_{I_s \in [-1,1]} L_p\left(\hat{I}_t, I_t\right) \tag{3}$$

$$L_r\left(\hat{I}_t, I_t\right) = \mu L_p(I_s, I_t) \tag{4}$$

where μ represents the binary mask parameter and $L_r\left(\hat{I}_t, I_t\right)$ represents the image reconstruction loss. To ameliorate the smoothness of the generated inverse depth maps, an edge-aware smoothness loss, denoted as L_{smooth}, is computed. Subsequent operations are then conducted as follows:

$$L_{\mathrm{smooth}} = \alpha |\partial_x d_t^*| e^{-|\partial_x I_t|} + |\partial_x d_t^*| e^{-\left|\partial_y I_t\right|} \tag{5}$$

$$L = \frac{1}{3} \sum_{s \in \left\{1, \frac{1}{2}, \frac{1}{4}\right\}} \left(L_r + \lambda L_{\mathrm{smooth}}\right) \tag{6}$$

where s represents the various scale outputs produced by the depth decoder and $d_t^* = \frac{d_t}{\hat{d}_t}$ represents the mean-normalized inverse depth. The value of λ is 10^{-3}. Lite-Mono effectively balances network complexity and inference speed. It exhibits strong generalization capabilities and addresses the challenges mentioned above. Hence, in this work, the algorithm is employed to generate depth maps for fittings RGB images, thereby supplementing the missing depth information in fittings images.

3.3. 3D Object Bounding Box Prediction

Relying solely on the pseudo point cloud may not yield optimal detection results; therefore, it is beneficial to map the generated 2D region boxes in the image to their corresponding 3D regions. This process involves converting the 2D coordinate information into 3D coordinate information. It is assumed that the perspective projection of 3D bounding boxes closely aligns with their 2D counterparts. The 3D bounding box is defined using center coordinates $C = [c_x, c_y, c_z]$, dimensions $I = [i_x, i_y, i_z]$, and orientation $O(\theta, \phi, \alpha)$. Given the object's pose (O, C) in the camera coordinate system and the camera's intrinsic parameters, the relationship between the 3D point $X_0 = [X, Y, Z, 1]$ and the projected point $x = [x, y, 1]^T$ in the camera coordinate system is as follows [39]:

$$x = K[O, C]X_0 \tag{7}$$

where K represents the intrinsic matrix.

Presuming that the object's coordinate system origin is located at the center of the 3D bounding box and the object's dimension is known, the eight vertices of the 3D bounding box can be succinctly expressed as $X_1 = [\frac{d_x}{2}, \frac{d_y}{2}, \frac{d_z}{2}]$, $X_2 = [-\frac{d_x}{2}, \frac{d_y}{2}, \frac{d_z}{2}]$, ..., $X_8 = [-\frac{d_x}{2}, -\frac{d_y}{2}, -\frac{d_z}{2}]$.

3.4. Three-Dimensional-Detection-Driven 2D Detection

The implementation of 3D-detection-driven 2D detection mainly relies on the DGCNN network. In this work, DGCNN is employed to train point cloud data related to fittings. Figure 3 delineates the architecture of DGCNN. For each point, the edge convolution (EdgeConv) computes edge features on the layer, and these features are then aggregated for each point to obtain the EdgeConv computation result. EdgeConv utilizes the connecting edges to express the amalgamation of feature information within this pair of inmixed nodes. Following this, a series of non-linear transformations are applied to combine feature information, effectively expressing the local features from the focal node.

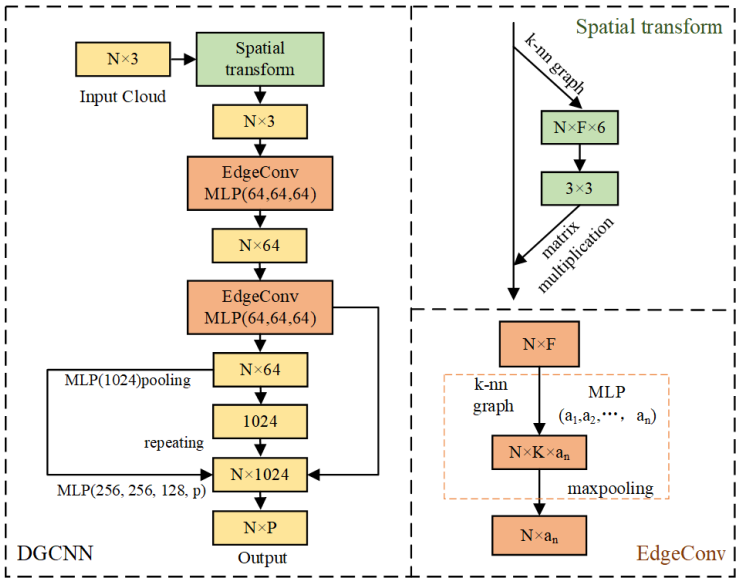

Figure 3. DGCNN architecture. The right-side diagrams represent the spatial transform and EdgeConv, respectively. The spatial transform module utilizes the estimated 3 × 3 matrix to map the input to the canonical space.

As depicted in the EdgeConv module of Figure 3, F stands for the dimensionality of each point, N denotes the total number of points, and $(a_1, a_2, \ldots, a_n)$ within MLP signifies the input and output dimensionality for each layer. K signifies the number of neighboring nodes. Through subsampling the target point cloud, a point cloud with n points and F dimensions is obtained. The function of neurons in each layer is predicated on the preceding layer's output. Subsequently, a directed graph is established to encapsulate the local structure of the point cloud. The edges connecting central points and neighboring points are represented using Equation (8):

$$e_{ij} = h_\theta(x_i, x_j - x_i) \tag{8}$$

where e_{ij} is the edges connecting central points and neighboring points, x_i represents the central point, x_j represents a point adjacent to the central point, θ represents learnable parameters, and h_θ represents an activation function. The output of the central points is as follows:

$$x_i' = pooling_{j:(x,y)} h_\theta(x_i, x_j - x_i) \tag{9}$$

where x_i' is the central points's output.

By stacking numerous network layers, conventional deep neural network models have demonstrated remarkable performance on various problems, due to their potent representational capabilities. In the context of multi-layer EdgeConv in DGCNN, the neighborhood information extracted through edge convolution has the potential to represent distant regions in the original point cloud space.

4. Results and Discussion

4.1. Implementation Details

This study utilized a self-built dataset consisting of 2563 inspection images of power fittings, captured during overhead line inspections. A total of 982 of these images featured rust data and were used for detection. Using the LabelImg annotation tool, seven categories of power fittings were annotated: shackle, eyelink, damper, thimble, suspension clamp, clevis, and ball eyes. To address the challenge of limited training samples for certain fittings, various data augmentation techniques were applied in the object detection task. These techniques encompassed random scaling, flipping, rotation, and the introduction of Gaussian noise. The dataset was divided into training, validating, and testing sets with a ratio of 7:2:1. The final number of power fittings utilized for training is presented in Table 1.

Table 1. Self-built dataset information.

Fittings	Training Dataset	Validating Dataset	Testing Dataset
Shackle	2236	639	320
Ball eyes	1059	303	151
Suspension clamp	1260	360	180
Thimble	1047	299	150
Clevis	1199	342	171
Eyelink	1201	343	172
Damper	1845	338	169

To verify the detection performance of the UD-Net in different scenes, our study also utilized a dataset provided by a power supply company (PSC Dataset), which contains images of thimbles, eyelinks, and shackles and is divided into images under green vegetative scenes and yellow farmland scenes. Among them, there are 726 images in green vegetative scenes and 581 images in yellow farmland scenes. The images in this dataset all have a size of 512 × 512 and are carefully labeled.

Details of the hardware and software utilized in this experimental study are given in Table 2. During the experiments, a training batch-size of 8 was employed, and the training process transpired over a total of 100 epochs.

Table 2. Details of the Hardware and Software used in the Experimental Study.

Computer Systems	Configurations
Hardware	Ubuntu 16.04 operating system NVIDIA GTX2080Ti with 11GB memory
Software	Python 3.7, PyCharm 2020 CUDA 10.1, PyTorch 1.7.0

4.2. Experimental Results

4.2.1. Comparison with State-of-the-Art Models

A sequence of comparative analyses was carried out to assess the efficacy of the UD-Net model. The study compares UD-Net with U-Net, FA-UNet, SSD (single shot multibox detector) [40], Fast R-CNN (fast region-CNN) [41], YOLOv4 [42], and Faster R-CNN [43]. Additionally, it measures UD-Net's performance against lightweight object detection models, including YOLOv3-tiny [44], YOLOX-Nano [45], and YOLOv5s. Table 3 shows

the comparison results. When benchmarked against SSD, UD-Net shows substantial improvements, with a 17.24% increase in Precision, an 11.85% increase in Recall, and a 14.2% increase in mAP (mean average precision) values. Furthermore, when compared with R-CNN series algorithms and U-Net series algorithms, UD-Net consistently exhibits better detection accuracy. When compared with the lightweight models mentioned above, although YOLOX-Nano has the smallest number of parameters, its detection accuracy is lower than that of UD-Net. When comparing UD-Net and YOLOv5s models, although the parameter number of UD-Net is slightly higher than that of YOLOv5s, it performs better in terms of overall accuracy.

Table 3. Comparison between State-of-the-Art Models and UD-Net.

Models	Precision/%	Recall/%	mAP/%	Parameters/Million
SSD	76.01	78.33	75.79	-
Fast R-CNN	78.68	72.77	76.26	-
Faster R-CNN	80.18	78.99	78.56	-
YOLOv3-tiny	72.83	75.69	79.55	8.8
YOLOv4	81.62	82.15	81.08	52.5
YOLOv5s	88.15	89.37	88.26	7.0
YOLOX-Nano	86.59	84.85	85.22	1.8
U-Net	84.98	83.76	86.34	7.7
FA-UNet	90.09	87.81	87.73	19.9
UD-Net	93.25	90.18	89.99	7.2

Figure 4 offers an intuitive representation of the comparison results of various algorithms via a box plot. As the figure shows, YOLOv5s exhibits commendable detection performance among the YOLO (you only look once) series algorithms. However, a noticeable performance disparity exists when juxtaposed with the UD-Net algorithm. The UD-Net model showcases reduced variance across multiple experimental runs, highlighting its consistent and superior performance.

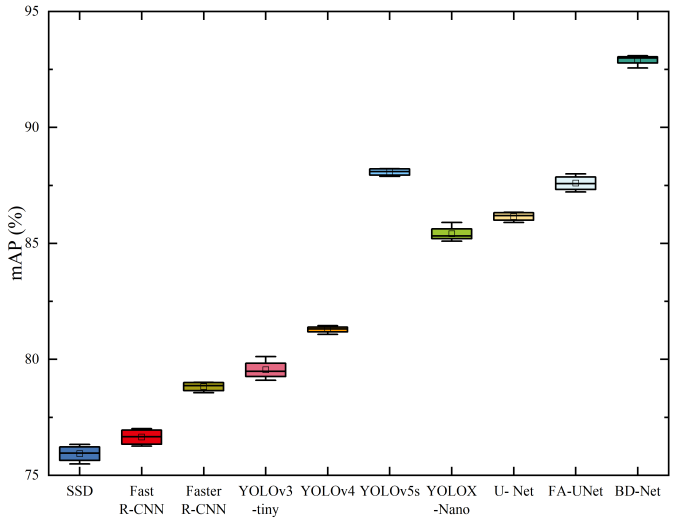

Figure 4. Box plot of comparison results for the different algorithms.

4.2.2. Impact of BRA-UNet

To validate the superiority of embedding the BRA attention module into the UNet network, experiments were conducted by upsampling on the U-Net network, incorporating various attention mechanisms including squeeze-and-excitation networks (SENet) [46], dual multiscale attention network (DMSANet) [47], efficient channel attention networks (ECANet) [48], convolutional block attention module (CBAM) [49], and the BRA attention mechanism. We carried out the experiments four times on the shackle dataset and the results are presented in Figure 5. It is evident that the integration of the BRA mechanism into the network enhances detection accuracy and ensures its stability. The inclusion of the

BRA attention mechanism effectively expands receptive fields and contextual information, thereby substantially enhancing the performance of the U-Net model. As the heatmaps shown in Figure 6 demonstrate, the regions of interest for the target classifier become more pronounced, and the high-response zones in the heatmap are focused on target fittings. These results indicate that the enhanced BRA-UNet effectively focuses on the fittings target. Furthermore, incorporating the BRA attention mechanism lessens the model's reliance on external data and bolsters its ability to discern internal data correlations.

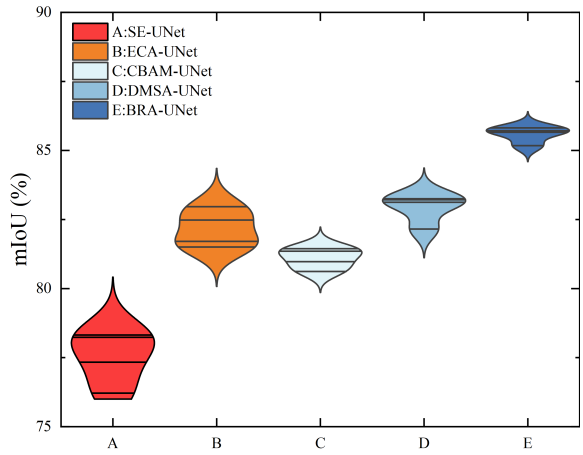

Figure 5. Violin plots of detection results for U-Net networks embedded with different attention mechanisms.

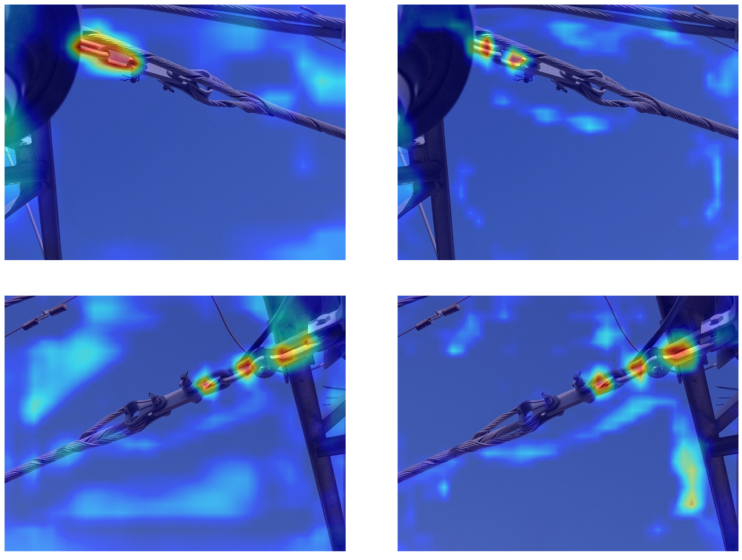

Figure 6. Comparison of heatmap results before and after embedding the BRA attention mechanism in U-Net. (**a**) Represents the heatmap results detected by the original U-Net network, and (**b**) represents the heatmap results detected by the U-Net embedding the BRA attention mechanism.

4.2.3. Ablation Analysis

To assess the validity of the UD-Net model, a series of ablation experiments were conducted on a self-built dataset. Four methods were considered: A, representing the U-Net model, B, representing the U-Net combined with the DGCNN network, C, representing the BRA-UNet network, and D, representing the UD-Net. Table 4 provides insights into the influence of these models on detection performance. By comparing A and C, it is discernible that embedding the attention mechanism yields a degree of improvement in detection accuracy. This indicates that embedding the BRA attention module enhances the model's

feature extraction capability. This enables the U-Net model to better learn the characteristics and patterns of various categories during training, thereby improving the reliability of fittings inspection. Similarly, comparing C and D reveals that the DGCNN algorithm has a positive effect on U-Net detection. When benchmarked against method A, method D shows substantial improvements, with a 1.24% increase in Accuracy, a 9.42% increase in Precision, an 8.27% increase in Recall, and a 6.42% increase in mAP values. This indicates that method D optimizes the outcomes of method A and refines the detection results of the original U-Net. Overall, the enhanced UD-Net model demonstrates the highest detection performance, underscoring the effectiveness of utilizing the DGCNN algorithm to extract feature information from obscured fittings, thereby further improving accuracy.

Table 4. Ablation study results.

Models	Accuracy/%	Precision/%	Recall/%	mAP/%
A	97.88	79.95	84.98	83.76
B	98.41	81.81	85.94	85.23
C	98.46	85.75	87.24	86.37
D	99.12	89.37	93.25	90.18

Figures 7–9 delineate the results of fittings detection under the four different methods, A, B, C, and D, showcasing the superior recognition proficiency of method D, especially in scenarios of occlusion and small target fittings. In Figure 7, all four methods demonstrate the capability to recognize unobstructed and normally sized fittings targets. However, the target boxes in methods A, B, and C show some inaccuracies. Notably, for objects like the eyelink in the first row, methods A and B produce false detections, likely because of the similarity in shape and size between shackles and eyelinks. A shackle is mistakenly labeled as an eyelink in Figure 7. Conversely, method D predicts the target boxes with greater accuracy, underscoring its superior recognition capability. The findings indicate that method D decreases the rates of both missed detections and false detections for fittings.

As depicted in Figure 8, while methods A and B fail to detect the ball eyes among the small targets, methods C and D successfully identify them. For the shackle, method A experiences missed detection issues. With methods B and C, even though the targets are detected, there are inaccuracies in the positioning and dimensions of the target boxes. Relative to the first three methods, method D not only rectifies the missed detection cases in small target fittings but also exhibits superior recognition capabilities. For the thimble that is partially obscured by steel strands, as seen in Figure 9, the detection results in the first row clearly show that method D adeptly identifies thimble images with occlusion challenges. In a similar vein, the image of the damper obscured by the cement pole is uniquely discerned by our proposed method D.

The comparison results of the detection of different fittings are presented in Table 5. It is evident that the UD-Net network exhibits significant enhancements in both Precision and Recall metrics for fittings detection. Notably, detection of the suspension clamp saw an increase of 13.65% in Precision and 2.18% in Recall when compared to the original U-Net network. Because the suspension clamp has significant differences in shape compared to other fittings, the model exhibits superior recognition ability for this type of fitting. In addition, UD-Net significantly outperforms the original U-Net model in recognizing small target fittings, such as dampers. Compared with the U-Net algorithm, the UD-Net algorithm shows substantial improvements, with a 10.29% increase in mIoU, a 26.24% increase in Precision, and a 9.78% increase in Recall. Meanwhile, the enhanced UD-Net algorithm exhibits superior detection accuracy for fittings prone to occlusion, such as the thimble and eyelink. This underscores the algorithm's proficiency in extracting features from occluded objects, thereby optimizing the results. Overall, UD-Net demonstrates superior performance across all four evaluation metrics, Accuracy, mIoU, Precision, and Recall, compared to the U-Net. In terms of algorithm performance, the UD-Net network's average training time increases by a minimum of 12.88% compared to the U-Net. This

suggests that despite the introduction of the DGCNN algorithm, the operation of BRA-UNet within the UD-Net structure—achieved by reducing both the number of encoder blocks and the convolution channels—effectively decreases the model's computational overhead, thereby enhancing its training speed. Furthermore, UD-Net is more lightweight than U-Net. This underscores the efficacy of the proposed algorithm in the domain of electric power fittings detection.

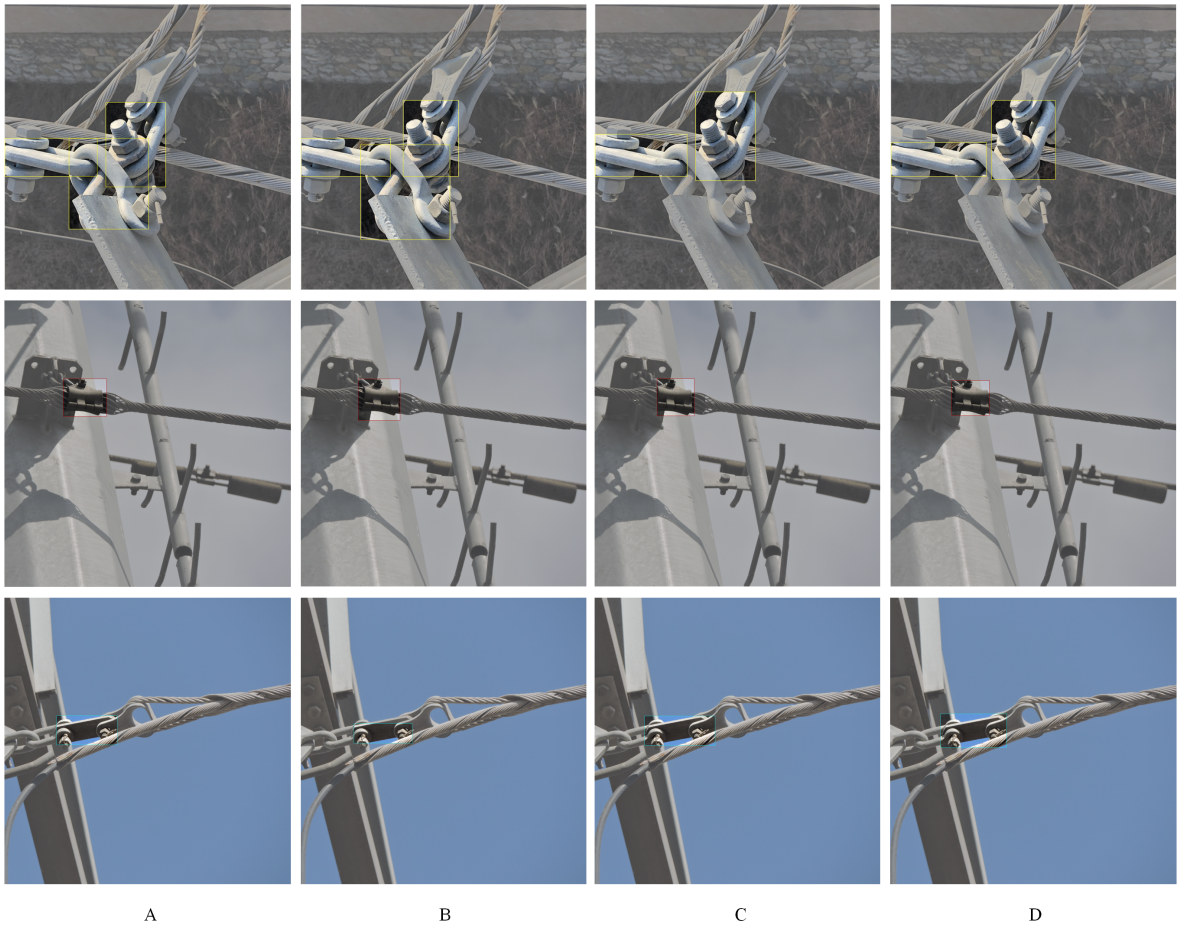

Figure 7. Visual detection results of fittings under various methods. The first row shows images of an eyelink, the second row shows images of a suspension clamp, the third row displays images of a clevis. A is the U-Net, B is the U-Net combined with the DGCNN, C is the BRA-UNet, and D is the UD-Net.

Table 5. Detection results for various fittings.

Category	U-Net					UD-Net				
	Accuracy	**mIoU**	**Precision**	**Recall**	**Average Training Time**	**Accuracy**	**mIoU**	**Precision**	**Recall**	**Average Training Time**
Suspension clamp	97.69	85.92	81.33	94.18	2.86	99.05	93.47	94.59	96.2	3.4
Ball eye	99.12	84.99	83.5	87.26	2.97	99.47	90.52	86.3	95.14	3.53
Clevis	99.02	61.83	57.24	72.95	2.9	99.22	83.77	85.17	94.39	3.47
Shackle	98.12	79.34	64.37	85.44	2.95	98.9	85.9	89.25	87.35	3.33
Damper	99.33	74.5	63.19	83.06	2.94	99.28	84.85	89.43	92.84	3.35
Eyelink	97.85	62.07	48.63	70.28	2.91	98.92	80.18	83.54	86.43	3.44
Thimble	98.16	70.76	57.26	83.57	3.00	98.46	79.29	78.18	88.92	3.52

Figure 8. Visual detection results of small target fittings under various methods. The first and second rows depict the visual inspection results of a ball eye and a shackle, respectively. A is the U-Net, B is the U-Net combined with the DGCNN, C is the BRA-UNet, and D is the UD-Net.

Figure 9. Visual detection results of occluded fittings under various methods. The first row shows images of a thimble obscured by steel strands, the second row shows images of a damper obscured by a cement pole. A is the U-Net, B is the U-Net combined with the DGCNN, C is the BRA-UNet, and D is the UD-Net.

Figure 10 provides a more intuitive representation of the results through bar charts. From the comprehensive results depicted in these three bar charts, the enhanced UD-Net model demonstrates superior detection performance with higher values for mIoU, Precision, and Recall compared to the U-Net model. As can be observed from Figure 10a,b, the UD-Net detection algorithm demonstrates significant enhancements in both Precision and Recall. This improvement is particularly evident for the images of a clevis and an eyelink, suggesting that the algorithm is better equipped to identify targets within fittings

images while significantly reducing the rate of missed detections. Figure 10c reveal that the UD-Net detection algorithm excels in the task of accurately detecting fittings images, while also minimizing the likelihood of misidentifying intricate backgrounds as fittings targets.

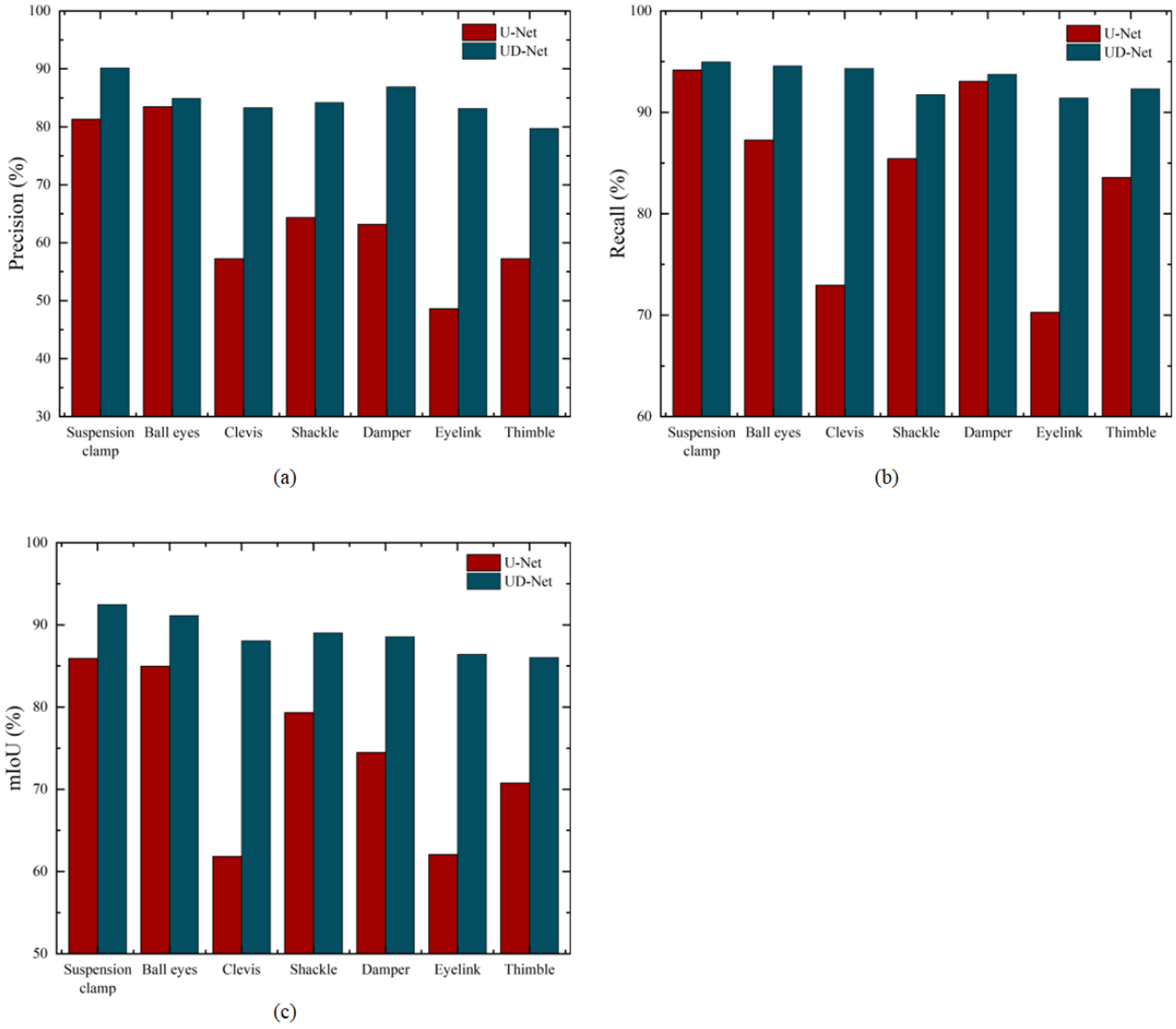

Figure 10. Metrics scores for different fittings. (**a**) Precision score for different fittings. (**b**) Recall score for different fittings. (**c**) mIoU score for different fittings.

4.2.4. Fittings Detection in Different Scenes

To evaluate the generalization ability of UD-Net in different scenes, we conducted a control experiment on the PSC Dataset. Figure 11 shows the mAP values for detecting fittings in green vegetative scenes and yellow farmland scenes. The results indicate that UD-Net effectively detects three distinct types of fittings across these two varied scenes. Specifically, in green vegetative scenes, Figure 11a shows that relative to the U-Net algorithm, the mAP values for the detection of shackles, eyelinks, and thimbles by UD-Net have risen by 6.29%, 5.95%, and 3.84%, respectively. Similarly, in yellow farmland scenes, Figure 11b shows mAP increases of 6.07%, 6.27%, and 3.16% for these fittings, respectively. These results indicate that UD-Net not only exhibits excellent performance on the self-built dataset, but also has good generalization ability when applied to different environments.

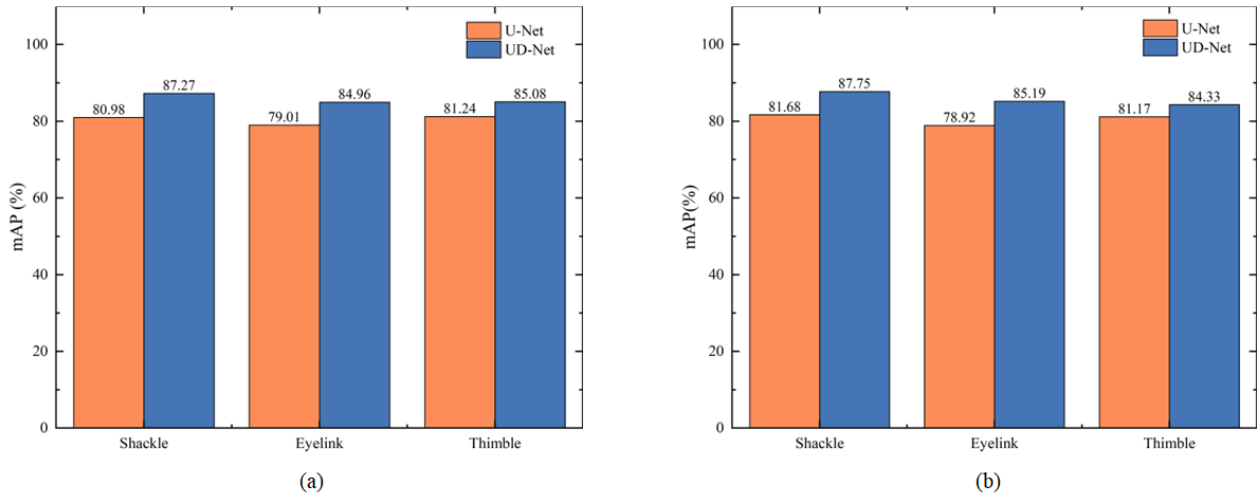

Figure 11. mAP scores for two different scenes. (**a**) Green vegetative scenes. (**b**) Yellow farmland scenes.

Furthermore, to illustrate the training outcomes of UD-Net more vividly, Figures 12 and 13 display the visualized results of fittings detection in green vegetative and yellow farmland scenes, respectively. Figure 12 reveal that in green vegetative environments, the bounding boxes identified by UD-Net are markedly precise. Notably, in the figure's second column, featuring thimble images, UD-Net precisely pinpoints the occluded thimble target, a detail that U-Net overlooks. In Figure 13, UD-Net is able to identify the incomplete shackle below the first column in yellow farmland environments, while U-Net fails to do so. The visualization results further confirm the generalization ability of UD-Net in different environments.

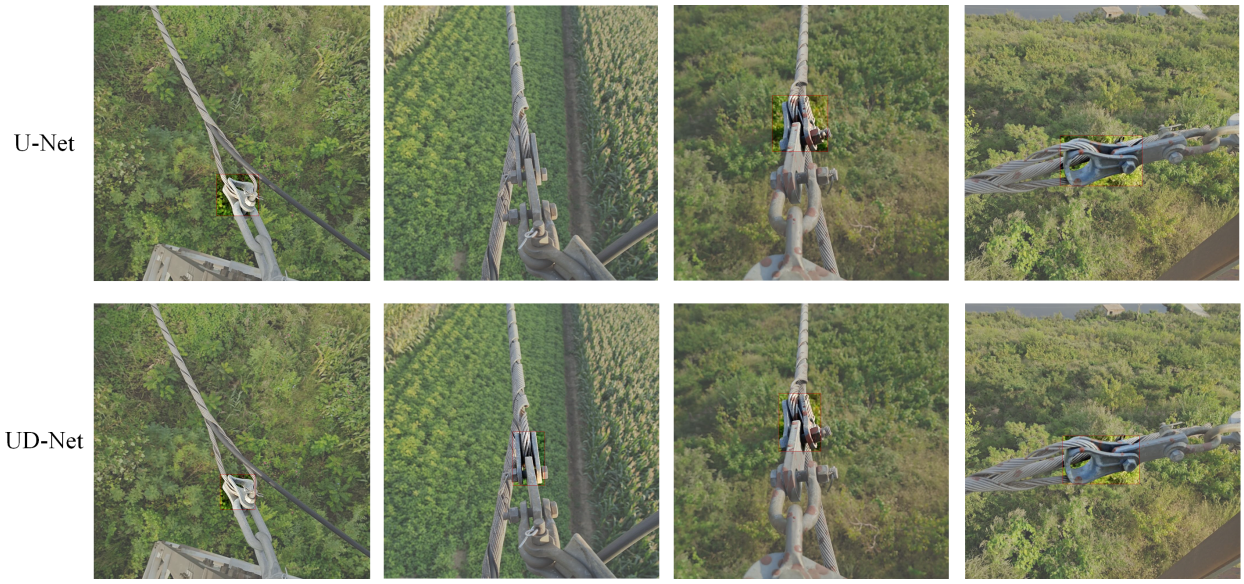

Figure 12. Visual detection results of thimbles under green vegetation scenes. The first row shows the detection results with U-Net and the second row displays the detection results with UD-Net.

4.2.5. The Detection of Rusted Fittings

In real-world electric power environments, fittings can be impacted by several elements, rust being a prime example. Rust can cause changes in the surface color, texture, and shape of the fittings, thus increasing the complexity of their identification. To validate the robustness of the UD-Net algorithm, we performed training and testing on 982 images of rusted fittings from a self-built dataset. Figure 14 provides a visualization of these results,

highlighting rusted fittings with a particular emphasis on rusty ball eyes. Notably, the UD-Net proficiently identifies rusted ball eyes, even amidst complex field backgrounds. This proficiency underscores UD-Net's efficacy in detecting defects in fittings and emphasizes its significant potential for practical applications.

Figure 13. Visual detection results of shackles under yellow farmland scenes. The first row shows the detection results with U-Net and the second row displays the detection results with UD-Net.

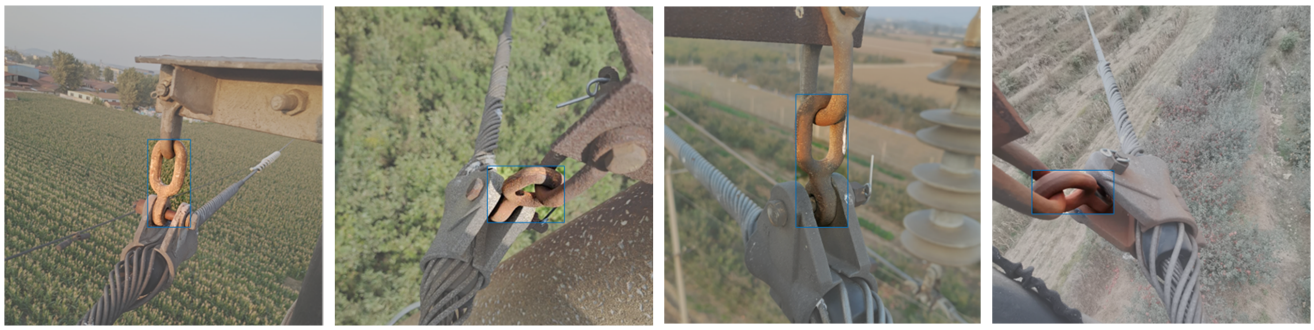

Figure 14. Visual results of detection of rusted fittings with UD-Net.

5. Conclusions

In addressing the inherent challenges associated with power fittings inspection, particularly characterized by intricate backgrounds and mutual occlusions amongst fittings, a detection method based on the novel UD-Net is proposed. First, a U-Net model embedded with BRA mechanisms is used for initial recognition of fittings images to enhance the model's feature extraction capability in complex backgrounds. Then, the Lite-Mono algorithm is utilized for the generation of a depth map for the fittings. This depth map is subsequently combined with the RGB image of the fittings, resulting in the conversion into a point cloud representation. The DGCNN algorithm is then applied to enhance the feature extraction capabilities of the network for fittings targets to fulfill the objective of 3D-detection-driven 2D detection. The simulation results demonstrate that the proposed methodology holds substantial promise, augmenting feature discernibility and facilitating the extraction of more pertinent information from images of fittings compared to other considered methods. The algorithm not only accomplishes the high-precision detection of power fittings but also harbors the potential to be applied to the automatic detection of rusted fittings within images.

Although the current recognition accuracy for shackles and eyelinks satisfies detection requirements, it is imperative to note that the structural similarity between shackles and eyelinks may still exert an influence upon recognition outcomes. Consequently, future endeavors may strategically focus on further optimizing the model for fine-grained object recognition, particularly pertaining to these two categories of fittings. Moreover, constrained by human resources and material availability, this study focuses on the annotation and identification of seven types of fittings. However, many types of fittings exist in reality. Future work can encompass a broader array of fitting types, enhancing the model's detection scope and performance for transmission line fittings.

Author Contributions: Conceptualization and methodology, X.L.; Data curation and formal analysis, X.L. and Z.X.; Funding acquisition, M.F. and X.L.; Investigation and project administration, Z.X. and M.F.; Resources, M.F.; Software, M.F. and Z.X.; Supervision, X.L.; Visualization, Z.X. and X.L.; Validation, Z.X., M.F. and X.L.; Writing—original draft, X.L., M.F. and Z.X.; Writing—review and editing, M.F., Z.X. and X.L. All authors have read and agreed to the published version of the manuscript.

Funding: The National Natural Science Foundation of China, Grant No. 61971244, the Shandong Provincial Natural Science Foundation, Grant No. ZR2020MF011, funded this research.

Data Availability Statement: Data are contained within the article.

Acknowledgments: We express our gratitude to the Editors and Reviewers for their valuable comments.

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

References

1. Luo, Y.; Yu, X.; Yang, D.; Zhou, B. A survey of intelligent transmission line inspection based on unmanned aerial vehicle. *Artif. Intell. Rev.* **2023**, *56*, 173–201. [CrossRef]
2. Ghobakhloo, M. Industry 4.0, digitization, and opportunities for sustainability. *J. Clean. Prod.* **2020**, *252*, 119869. [CrossRef]
3. Yang, L.; Fan, J.; Liu, Y.; Li, E.; Peng, J.; Liang, Z. A review on state-of-the-art power line inspection techniques. *IEEE Trans. Instrum. Meas.* **2020**, *69*, 9350–9365. [CrossRef]
4. Liu, Z.; Wu, G.; He, W.; Fan, F.; Ye, X. Key target and defect detection of high-voltage power transmission lines with deep learning. *Int. J. Electr. Power Energy Syst.* **2022**, *142*, 108277. [CrossRef]
5. Zoph, B.; Le, Q.V. Neural architecture search with reinforcement learning. *arXiv* **2016**, arXiv:1611.01578.
6. Zhu, L.; Ji, D.; Zhu, S.; Gan, W.; Wu, W.; Yan, J. Learning statistical texture for semantic segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Virtual, 19–25 June 2021; pp. 12537–12546.
7. Sharma, K.U.; Thakur, N.V. A review and an approach for object detection in images. *Int. J. Comput. Vis. Robot.* **2017**, *7*, 196–237. [CrossRef]
8. Luo, P.; Wang, B.; Wang, H.; Ma, F.; Ma, H.; Wang, L. An ultrasmall bolt defect detection method for transmission line inspection. *IEEE Trans. Instrum. Meas.* **2023**, *72*, 1–12. [CrossRef]
9. Wan, N.; Tang, X.; Liu, S.; Chen, J.; Guo, K.; Li, L.; Liu, S. Transmission line image object detection method considering fine-grained contexts. In Proceedings of the 2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), Chongqing, China, 12–14 June 2020; IEEE: Piscataway, NJ, USA, 2020; Volume 1, pp. 499–502.
10. Lian, Q.; Li, P.; Chen, X. Monojsg: Joint semantic and geometric cost volume for monocular 3d object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 1070–1079.
11. Wu, H.; Han, W.; Wen, C.; Li, X.; Wang, C. 3D multi-object tracking in point clouds based on prediction confidence-guided data association. *IEEE Trans. Intell. Transp. Syst.* **2021**, *23*, 5668–5677. [CrossRef]
12. Chen, Y.; Tai, L.; Sun, K.; Li, M. Monopair: Monocular 3d object detection using pairwise spatial relationships. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 12093–12102.
13. Wang, X.; Fu, C.; Li, Z.; Lai, Y.; He, J. DeepFusionMOT: A 3D multi-object tracking framework based on camera-LiDAR fusion with deep association. *IEEE Robot. Autom. Lett.* **2022**, *7*, 8260–8267. [CrossRef]
14. Shi, S.; Wang, X.; Li, H. Pointrcnn: 3d object proposal generation and detection from point cloud. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 770–779.
15. Ren, J.; Chen, X.; Liu, J.; Sun, W.; Pang, J.; Yan, Q.; Tai, Y.W.; Xu, L. Accurate single stage detector using recurrent rolling convolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 5420–5428.
16. Liu, Z.; Cao, Y.; Wang, Y.; Wang, W. Computer vision-based concrete crack detection using U-net fully convolutional networks. *Autom. Constr.* **2019**, *104*, 129–139. [CrossRef]
17. Wu, X.; Hong, D.; Chanussot, J. UIU-Net: U-Net in U-Net for infrared small object detection. *IEEE Trans. Image Process.* **2022**, *32*, 364–376. [CrossRef]

18. Yang, X.; Li, X.; Ye, Y.; Lau, R.Y.; Zhang, X.; Huang, X. Road detection and centerline extraction via deep recurrent convolutional neural network U-Net. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 7209–7220. [CrossRef]
19. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention, Proceedings of the MICCAI 2015: 18th International Conference, Munich, Germany, 5–9 October 2015*; Proceedings, Part III 18; Springer: Berlin/Heidelberg, Germany, 2015; pp. 234–241.
20. He, M.; Qin, L.; Deng, X.; Zhou, S.; Liu, H.; Liu, K. Transmission Line Segmentation Solutions for UAV Aerial Photography Based on Improved UNet. *Drones* **2023**, *7*, 274. [CrossRef]
21. Han, G.; Zhang, M.; Li, Q.; Liu, X.; Li, T.; Zhao, L.; Liu, K.; Qin, L. A Lightweight Aerial Power Line Segmentation Algorithm Based on Attention Mechanism. *Machines* **2022**, *10*, 881. [CrossRef]
22. Cao, M.; Fu, H.; Zhu, J.; Cai, C. Lightweight tea bud recognition network integrating GhostNet and YOLOv5. *Math. Biosci. Eng. MBE* **2022**, *19*, 12897–12914. [CrossRef] [PubMed]
23. Choi, H.; Yun, J.P.; Kim, B.J.; Jang, H.; Kim, S.W. Attention-based multimodal image feature fusion module for transmission line detection. *IEEE Trans. Ind. Inform.* **2022**, *18*, 7686–7695. [CrossRef]
24. Shi, W.; Rajkumar, R. Point-gnn: Graph neural network for 3d object detection in a point cloud. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 1711–1719.
25. He, Q.; Wang, Z.; Zeng, H.; Zeng, Y.; Liu, Y. Svga-net: Sparse voxel-graph attention network for 3d object detection from point clouds. In Proceedings of the AAAI Conference on Artificial Intelligence, Virtual, 22 February–1 March 2022; Volume 36, pp. 870–878.
26. Xie, S.; Liu, S.; Chen, Z.; Tu, Z. Attentional shapecontextnet for point cloud recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 4606–4615.
27. Zhou, Y.; Tuzel, O. Voxelnet: End-to-end learning for point cloud based 3d object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 4490–4499.
28. Qi, C.R.; Yi, L.; Su, H.; Guibas, L.J. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 5105–5114.
29. Yue, W.; Yongbin, S.; Ziwei, L.; Sarma, S.E.; Bronstein, M.M.; Solomon, J.M. Dynamic graph cnn for learning on point clouds. *ACM Trans. Graph. (TOG)* **2019**, *38*, 1–12.
30. Wang, Y.; Solomon, J.M. Object dgcnn: 3d object detection using dynamic graphs. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 20745–20758.
31. Gamal, A.; Wibisono, A.; Wicaksono, S.B.; Abyan, M.A.; Hamid, N.; Wisesa, H.A.; Jatmiko, W.; Ardhianto, R. Automatic LIDAR building segmentation based on DGCNN and euclidean clustering. *J. Big Data* **2020**, *7*, 102. [CrossRef]
32. Xing, Z.; Zhao, S.; Guo, W.; Guo, X.; Wang, Y. Processing laser point cloud in fully mechanized mining face based on DGCNN. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 482. [CrossRef]
33. Liang, H.; Lv, J.; Wang, Z.; Xu, X. Medical image mis-segmentation region refinement framework based on dynamic graph convolution. *Biomed. Signal Process. Control* **2023**, *86*, 105064. [CrossRef]
34. Peng, X.; Zhong, R.; Li, Z.; Li, Q. Optical remote sensing image change detection based on attention mechanism and image difference. *IEEE Trans. Geosci. Remote Sens.* **2020**, *59*, 7296–7307. [CrossRef]
35. Chen, F.; Pan, S.; Jiang, J.; Huo, H.; Long, G. DAGCN: Dual attention graph convolutional networks. In Proceedings of the 2019 International Joint Conference on Neural Networks (IJCNN), Budapest, Hungary, 14–19 July 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 1–8.
36. Perreault, H.; Bilodeau, G.A.; Saunier, N.; Héritier, M. Spotnet: Self-attention multi-task network for object detection. In Proceedings of the 2020 17th Conference on Computer and Robot Vision (CRV), Ottawa, ON, Canada, 13–15 May 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 230–237.
37. Zhu, L.; Wang, X.; Ke, Z.; Zhang, W.; Lau, R.W. BiFormer: Vision Transformer with Bi-Level Routing Attention. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 18–22 June 2023; pp. 10323–10333.
38. Zhang, N.; Nex, F.; Vosselman, G.; Kerle, N. Lite-mono: A lightweight cnn and transformer architecture for self-supervised monocular depth estimation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 18–22 June 2023; pp. 18537–18546.
39. Mousavian, A.; Anguelov, D.; Flynn, J.; Kosecka, J. 3d bounding box estimation using deep learning and geometry. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 7074–7082.
40. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. Ssd: Single shot multibox detector. In *Computer Vision, Proceedings of the ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, 11–14 October 2016*; Proceedings, Part I 14; Springer: Berlin/Heidelberg, Germany, 2016; pp. 21–37.
41. Girshick, R. Fast r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1440–1448.
42. Bochkovskiy, A.; Wang, C.Y.; Liao, H.Y.M. Yolov4: Optimal speed and accuracy of object detection. *arXiv* **2020**, arXiv:2004.10934.
43. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. *Adv. Neural Inf. Process. Syst.* **2015**, *28*, 1137–1149. [CrossRef]
44. Redmon, J.; Farhadi, A. Yolov3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767.

45. Ge, Z.; Liu, S.; Wang, F.; Li, Z.; Sun, J. Yolox: Exceeding yolo series in 2021. *arXiv* **2021**, arXiv:2107.08430.
46. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 7132–7141.
47. Sagar, A. Dmsanet: Dual multi scale attention network. In Proceedings of the International Conference on Image Analysis and Processing, Lecce, Italy, 23–27 May 2022; Springer: Berlin/Heidelberg, Germany, 2022; pp. 633–645.
48. Wang, Q.; Wu, B.; Zhu, P.; Li, P.; Zuo, W.; Hu, Q. ECA-Net: Efficient channel attention for deep convolutional neural networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 11534–11542.
49. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. Cbam: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19.