International Journal of Scientific Engineering and Research (IJSER)
ISSN (Online): 2347-3878
Impact Factor (2022): 7.741

Transmission Line Defective Bolts Detection Based on Fine-grained Recognition and Mutual Learning

Jimin Yu[1], Yilin Ma[1]

[1]North China Electric Power University, School of Control and Computer Engineering, Beinong Road, 2, China
jimin97yu[at]ncepu.edu.cn

Abstract: ***Unmanned Aerial Vehicle (UAV) inspection has replaced most manual tasks to perform transmission line image capture. However, analyzing these images still relies heavily on manual labor, which is time-consuming and prone to errors. Defect bolts detection is an important part of transmission line inspection, but due to the small size of bolts and the difficulty in defects identifying, automated transmission line inspection faces challenges. Therefore, we propose a deep neural network called FG R-CNN (Fine Grained R-CNN) to detect defective bolts in transmission lines. First, to distinguish the bolts that are highly similar between classifications, we introduce the convolutional binary tree branch to carry out fine-grained recognition of RoI features. Secondly, we introduce classification consistency loss to constrain the differences between multiple classifiers of the tree structure to promote mutual feature learning. Finally, we add the feature pyramid network (FPN) to enable our network more suitable for small bolts detection. Our experimental results on a transmission line inspection image dataset show that our model has a 6.98% higher bolts AP and a 12.02% higher defect bolts AP than the commonly used object detection network Faster R-CNN.***

Keywords: detective bolts detection, fine-grained recognition, mutual learning, Faster R-CNN.

1. Introduction

In recent years, more and more unmanned aerial vehicles (UAVs) have been used domestically and internationally to replace manual inspection of transmission lines, effectively reducing labor costs and work hazards, and improving the efficiency and accuracy of inspection results[1]. As a result of the high volume and low value density of aerial images obtained during inspections, relevant working personnel are faced with a significant data amount. This can lead to subjective experience and fatigue affecting their ability to maintain consistent judgment standards, resulting in more errors and potential safety hazards[2]. However, massive data is not difficult for deep learning methods and can make the model more robust. Therefore, using appropriate deep learning algorithms and computer technology to handle defects detect tasks of inspection can help overcome the shortcomings of manual drone inspection. Among the various challenges in transmission line defect detection, detecting defective bolts has emerged as a crucial issue. The small scale of the objects and the difficulty in distinguishing defects pose significant challenges in this area. The goal of this task is to accurately detect the position of each bolt from the inspection image of the transmission line and classify it as a defective bolt. For a more intuitive display, Figure 1 depicts the result of bolts detection, with green bounding boxes representing normal bolts and red ones indicating defective bolts.

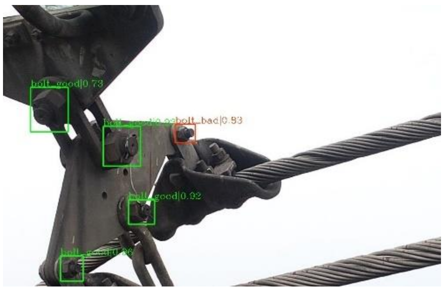


Figure 1: Bolts detection result of an inspection image

Researchers have proposed many methods and systems for identifying and detecting transmission lines and components they contain. Among them, early research mainly relied on traditional image processing and machine learning methods for feature extraction of inspection images. The extracted features mainly include LBP features, HOG features, and Canny operators. Yao et al. [3] applied SVM to classify transmission lines by extracting LBP features of transmission lines and backgrounds to calculate changes in local grayscale values. And locate the transmission line using the image edge features extracted by the Canny operator. Mao et al.[4] obtained the morphological gradient information of transmission lines by extracting HOG features, and used PAC dimensionality reduction to select features with different contributions for transmission line classification. Due to the limitations of the designed feature extractors, these traditional methods can only extract low-level features such as color changes and edge information, making it difficult to fully consider the factors of accurate classification and object localization. When complex backgrounds and other situations arise, the model is prone to failure. In addition, these methods are only suitable for large objects with obvious features such as transmission lines, while their recognition ability for small objects such as bolts is limited.

In recent years, methods based on deep convolutional neural networks have achieved outstanding results in the fields of machine vision and computer vision. Compared with traditional machine learning methods, object detection methods based on deep learning do not require a single manually designed feature extractor, but instead automatically learn and extract task related features through deep neural networks, resulting in higher detection accuracy and stronger generalization performance. Therefore, researchers are gradually applying deep learning methods in transmission line inspection scenarios. Regarding the detection of bolt defects in transmission lines, relevant personnel have found in their research that the main reason for affecting the accuracy of bolt recognition is the representation of features related to classification and positioning tasks. At this point, the deep features of deep convolutional neural networks can lead to a decrease in the distinguishability between bolts and other objects in the background region[5]. Therefore, better object feature learning has become one of the research focuses of bolts defect detection. Xiao et al.[6] used a small-scale convolutional neural network to obtain regions of interest (RoI) in images and automatically selected features that significantly contributed to better bolt recognition to construct a cascade feature pyramid, achieving feature enhancement for bolt recognition. Zhao et al.[7] found that the shooting angle affects the extraction of bolt shape features. Therefore, they proposed an unsupervised clustering method based on bolt shape, which introduces morphological prior information into the network to enhance feature representation. Jiao et al.[8] considered that the position of the bolts on the tower has a certain pattern, so they integrated the characteristics of the bolts and the context of the surrounding background to increase the differentiation between the bolts and the bolts in the background. Although the transmission line defective bolts detection methods based on deep learning have achieved higher accuracy than traditional ones, the feature selection and fusion are greatly affected by hyperparameter, and the feature construction method lacks robustness. Therefore, the actual application effect still needs to be improved.

To obtain a more effective feature representation for bolts recognition, we propose an improved method for detecting defective bolts, called FG R-CNN. This method aims to solve the problem of fine-grained identification within the defect bolts class. Based on the universal object detection network Faster R-CNN[9], we perform fine-grained recognition of RoI features and introduce mutual learning constraints to ensure consistency in multi-classifier feature learning. In addition, the introduction of feature pyramids enables the network to extract multi-scale features, making it more suitable for small object detection. The experimental results show that our method has better recognition performance in detecting transmission line defect bolts compared to several object detection models.

2. Materials and Methods

2.1 Network Framework

Our model FG R-CNN is improved based on the universal object detection network Faster R-CNN, of which the model architecture is shown in Figure 2. It consists of four parts: the backbone network ResNet-50[10] and the connected feature pyramid (FPN[11]), the region proposal network (RPN), the convolutional binary tree branch for object classification and the bounding box regression branch for object location. From the perspective of the network processes, firstly, Due to the small size of bolts in the image, we added the feature pyramid network after the backbone to fuse multi-scale features. This allows the network to integrate shallow features such as texture and edges with deep features, thereby improving the representation of small objects features of bolts. Secondly, the network extracts regional proposals from multi-scale features through RPN. We modified the original RoI pooling method in Faster R-CNN and replaced it with RoI Align[12]. RoI Align extracts more accurate image features using bilinear interpolation technology, which is vital for precise bounding box regression of small objects. Thirdly, we introduce the convolutional binary tree branch to optimize the object classification of RoI features and achieve fine grained recognition within the class. Finally, since the final prediction of the convolutional binary tree is the weighted voting of the leaves as sub-classifiers, we apply the mutual learning strategy to constrain the prediction differences of multiple sub-classifiers. Based on the optimization improvement of the tree structure, by introducing the classification consistency loss, the leaves learn from each other during training, reducing the phenomenon of prediction confrontation and further improving recognition precision of our model.

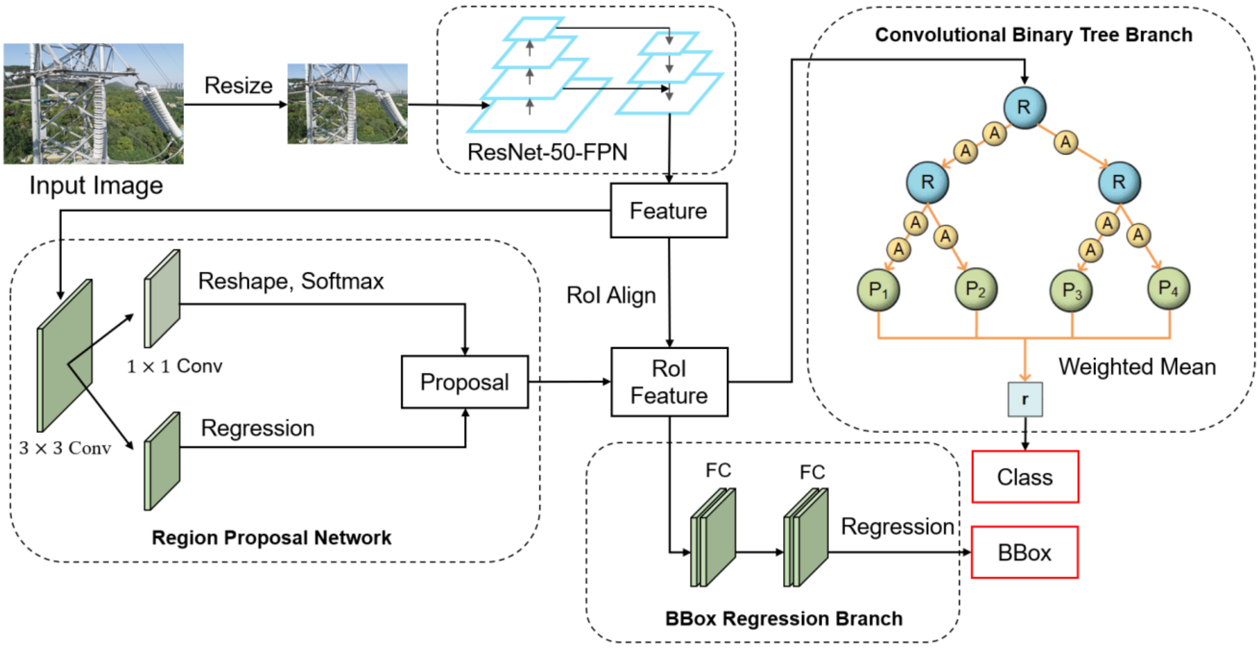


Figure 2: The network framework of FG R-CNN

2.2 Fine-grained RoI Feature Learning

Transmission lines have different types and usages of bolts, some of which are used to support wires and some are used to connect insulators. In addition, different types of bolts may be damaged differently, such as the bolts that support the wires being more susceptible to corrosion and oxidation due to the influence of static electricity, and the bolts that connect the insulators being more susceptible to tension and torque due to greater stress. Therefore, defective bolts have various forms, with the main classification shown in Figure 3. Figures (a-d) represent the defects of single nut bolts, including missing cotter bolt (Figure 3a), corroded bolt (Figure 3b), deformed cotter bolt (Figure 2c), and loose cotter bolt (Figure 3d). The defects of the double nut bolt in sub image e-h include missing cotter bolt (Figure 3e), corroded bolt (Figure 3f), bolt with loose nut (Figure 3g), and bolt with missing nut (Figure 3h). Although these defective bolts are classified into the same class during inspection, there are significant differences in appearance and semantics between these defects. Therefore, we propose a defective bolts detection method based on fine-grained RoI feature learning technology, which can allow the network to learn different subcategories of defect bolts on its own without predicting subcategories. There are significant differences in the appearance and semantics of defective bolts, but the two-layer fully connected layer and Softmax classifier of initial Faster R-CNN simply fitted these different defective samples into the same class, resulting in blurry boundaries for determining whether the object is a defective bolt. The method based on fine-grained RoI feature learning can learn the subtle differences between different types of defective bolts. We introduced sub-classifiers for fine-grained classification, which can better obtain feature representations of different defect types in network learning, thereby improving the accuracy and robustness of detection.

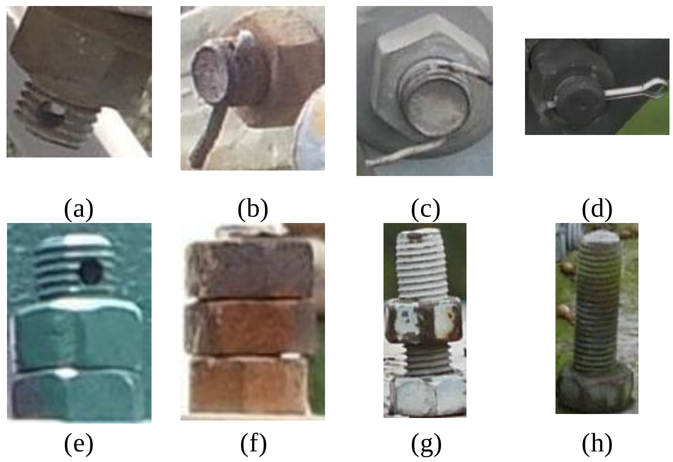


Figure 3: Different forms of defective bolts

Similar to the problem of fine-grained feature learning in defective bolts detection, fine-grained image classification and recognition focus on distinguishing subtle differences between different subcategories of a certain type of research object. Recently, related researchers have applied attention mechanisms to the recognition of fine-grained objects in unstructured areas of images[13][14]. Inspired by Ji et al. [15], we separated the classification and regression branch of Faster R-CNN, and implemented object classification by replacing the original full connection layer with a convolutional binary tree, as shown as in Figure 4.

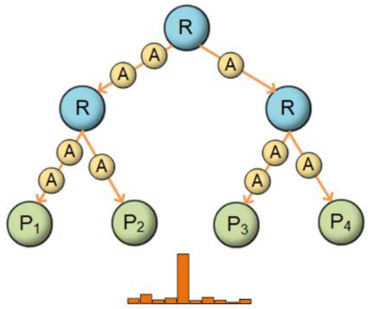


Figure 4: Structure of attention binary tree of FG R-CNN

In our network, the convolutional binary tree structure is generally the same as that proposed by Ji et al., but considering our task different from image classification, the size of RoI features is smaller than that of the whole image in fine-grained classification. To address the issue of increased feature channel numbers due to the feature fusion operation of the feature pyramid network, we made some modifications to the attention module of the attention transformer in the original convolutional binary tree structure. More specifically, we removed the fully connected bypass that caused redundant computing and replaced it with the SE-Node, which extracts channel attention to establish channel information interaction of multi-scale features. Figure 5 shows all the modules in the tree structure.

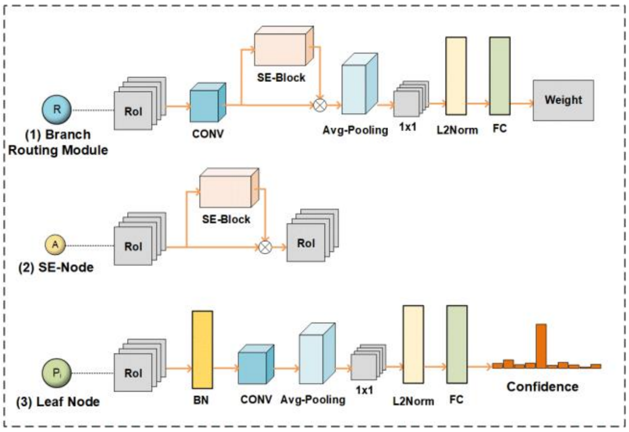


Figure 5: Different modules of convolution binary tree, including Branch Routing Module, SE-Node and Leaf Node.

Among them, the branch routing module determines the weighted calculation path of the class confidence from root to leaf. Different nodes in the same layer share a common parent or ancestor node, but with different classification features. This enables our model to gain the feature learning ability from coarse to fine. The final prediction is calculated by taking the weighted mean of the confidence scores of all leaves, resulting in a maximum vote to determine the final predicted class. As is shown in Figure 6, the extraction of RoI feature channel attention by SE-Node is divided into two steps. The first step is the compression operation, which compresses the features of each channel into a numerical value representing the importance of each channel through global average pooling. The second step is the excitation operation, which converts the compressed value into a channel weight coefficient vector through a fully connected layer, and applies this vector to each channel of the original feature[16]. That is, by channel weighting, the network pays more attention to important channels and suppresses unimportant channels.

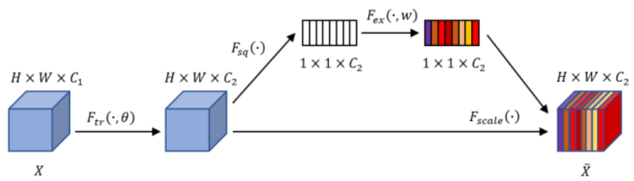


Figure 6: Squeeze and excitation of SE-Node

In the tree structure, the weight of the i-th leaf node or branch routing module of the h-th layer is calculated using Equation 1:

$$r_i^h(x) = \begin{cases} r_{\frac{i}{2}}^{h-1}(x) \cdot w_i^{h-1} & 2|i, h \neq 1 \\ 1 & h = 1 \\ r_{(i+1)/2}^{h-1}(x) \cdot (1 - w_i^{h-1}) & others \end{cases} \quad (1)$$

We define $p_i(x)$ the prediction of the i-th leaf node of RoI x. h is the depth of the convolutional binary tree, of which we find 3 is the best value. $r_i^h(x)$ is the cumulative weight of the RoI x from the root of the tree structure to the i-th node on the h-th layer. $w_i^{h-1}(x)$ is the weight passed from the node to its right child. The final prediction is the weighted mean of all leaves' predictions on confidence, as shown in Equation 2:

$$p(x) = \sum_{i=1}^{2^{h-1}} p_i(x) r_i^h(x) \quad (2)$$

Equation 3 shows the classification loss which is the weighted sum of the final predicted classification loss and classification losses of all leaves:

$$L_{cls}(x) = L_{final}(x) + \lambda L_{leaves}(x) \quad (3)$$

We found that the optimal value of λ is 0.4. Among them, the classification losses of final prediction and all leaves are negative logarithmic likelihood losses, as shown in Equation 4 and Equation 5:

$$L_{final}(x) = -\sum_{n=1}^{N_C} \log p(x_n) \quad (4)$$

$$L_{leaves}(x) = -\sum_{i=1}^{N_L} \sum_{n=1}^{N_C} \log p_i(x_n) \quad (5)$$

In the equations above, N_C is the number of all classes including background. N_L is the number of all leaves, which is 4 in our best model. $p(x_n)$ represents the confidence of the n-th class in the network's final prediction of RoI x. While $p_i(x_n)$ is the confidence on the n-th class of $p(x_n)$. Convolutional binary tree branch decides the final prediction by max voting the results of multiple classifiers of leaves.

2.3 Mutual Learning of Classification Consistency

Convolutional binary tree branch in FG R-CNN introduces multiple sub-classifiers through leaves to carry out fine-grained RoI feature learning. In the process of network optimization, the prediction of classifiers will gradually become different. More specifically, if a particular sub-classifier, denoted as C, achieves higher accuracy in predicting certain classes compared to other classifiers that have lower accuracy for those classes, the final classification results may not be accurately guided by C due to the sum of weights of the other sub-classifiers being greater than the weight of C. Hence, in the training process of sub-classifiers, it is essential not only to enhance their individual classification performance but also to facilitate other sub-classifiers' learning[17]. This view is known as mutual learning in the field of machine learning. Inspired by semi supervised deep learning, we introduced KL divergence and designed classification consistency loss to promote mutual learning among sub-classifiers. KL divergence is a concept in information theory used to measure the distance or difference between two probability distributions. The definition of KL divergence is as follows:

Assuming P and Q are two probability distribution functions, where P represents the true distribution and Q represents the model distribution, then the KL divergence of P relative to Q is shown in Equation 6:

$$D_{KL}(P||Q) = \sum_i P(i) \log \frac{P(i)}{Q(i)} \quad (6)$$

When KL divergence is used as a measure of similarity between classifiers, it can measure the similarity and difference between classifiers, making the interaction between classifiers more targeted and efficient. Consequently, we proposed a classification consistency loss, denoted as L_{cc}, based on KL divergence, to evaluate the consistency of predictions made by each leaf with other leaves., as shown in Figure 7 and Equation 7:

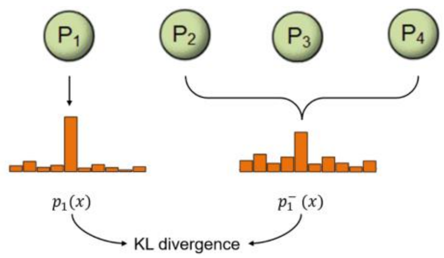


Figure 7: KL divergence between 1-th leaf and others

$$L_{cc}(x) = \sum_{i=1}^{N_L} \sum_{n=1}^{N_C} p_i^-(x_n) \log \frac{p_i^-(x_n)}{p_i(x_n)} \quad (7)$$

In the equation above, $p_i^-(x)$ is the mean of the leaves' predictions on RoI x except for the i-th leaf. While $p_i(x)$ is the prediction of the i-th leaf on RoI x while $p_i^-(x_n)$ represents the mean predicted on the n-th class by other leaves except for the i-th leaf, as shown in formula 8. During network training, the loss of classification consistency is added as an auxiliary loss term to perform end-to-end optimization in the multi task of FG R-CNN, as shown in Equation 9. Among them, $\boldsymbol{w_2 L_{det}}$ represents the sum of losses in the original Faster R-CNN, except for object classification losses. As the consistency loss is computed and summed up over multiple leaf nodes, it is crucial to ensure that there is no significant difference in the magnitudes. To achieve this, the loss weight parameters w_1 and w_2 are set to 0.5 and 1.0, respectively.

$$p_i^{-}(x_n) = \sum_{j=1, j\neq i}^{N_L} \frac{p_j(x_n) r_j^h(x)}{1 - r_i^h(x)} \tag{8}$$

$$L_{total} = L_{cls} + w_1 L_{cc} + \boldsymbol{w_2 L_{det}} \tag{9}$$

3. Results and Discussion

In order to evaluate the detection ability of FG R-CNN's defect bolts, we conducted a series of experiments using an image dataset containing bolts constructed from real transmission line inspection images captured by UAVs.

3.1 Dataset and Evaluation Metrics

Defective bolts are rusted bolts or bolts with loose or missing bolts. As far as we know, there is currently no publicly available image dataset suitable for the task of our study. Therefore, we use numerous typical images captured by UAVs to construct our own image dataset. Figure 8 shows one of the inspection images and zooms in on the area where the bolt is located. It can be seen that the proportion of bolts in the image is very small, so we cropped the original image when constructing the training set, which is conducive to faster network convergence.

Figure 8: An annotated UAV image including defective bolts

To train and test our model, each bolt in the original UAV image is manually annotated with bounding boxes and binary class labels. There are an indefinite number of bolts in each inspection image. The bolts are then classified into two kinds: normal bolts and defective bolts. Usually, a single bolt appears in images taken from multiple different perspectives. Among the 36307 images in the dataset, 64682 samples of normal bolts and 16632 samples correspond to defective bolts. The original size of images captured by UAVs is 5472×3678 pixels. Due to limitations in graphics memory of our computing device, it is not possible to directly train our model using original images. Therefore, we cropped all the images to 1024×1024 pixels. Then we set the ratio of training set, validation set and testing set to 100:1:2.

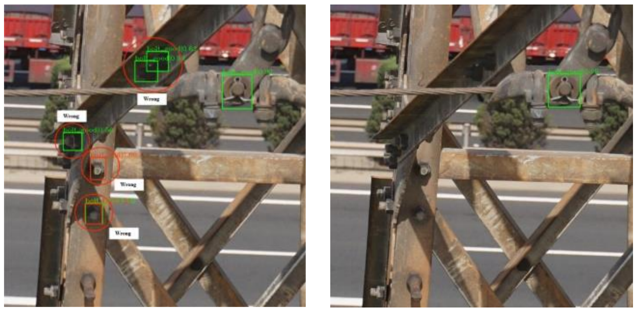
In object detection tasks, average precision (AP) and mean average precision (mAP) are two commonly used evaluation metrics. They were also applied in our study to measure accuracy of our model. The value of AP is the area under the precision-recall curve, where P and R represent precision and recall, respectively. Typically, a better classifier has a higher value of AP[18]. In our task, mAP is the mean AP value across two classes: normal bolts and defective bolts.

3.2 Experiment Setups

The experimental hardware platform utilized in our study is the NVIDIA GeForce GTX TITAN V GPU, equipped with 12GB of graphics memory and operating on the Linux system Ubuntu 18.04. The deep learning framework PyTorch 1.12 is used for training, evaluating and testing of corresponding dataset. All models and experiments were implemented based on the MMDetection[19] object detection toolbox developed by SenseTime to ensure fairness in comparison. We used a unified set of hyperparameters for training in all experiments. The model optimizer was stochastic gradient descent with an initial learning rate of 0.0005, learning rate decay of 0.001 and momentum decay of 0.9. We used only the five-layer feature maps P2 to P6 output by FPN, excluding P6, with P2 to P5 used for training. The batch size was set to 4, and the model was trained and tested on a single GPU for 16 epochs. The following two methods are used to prevent our model from overfitting while training: (1) Up-down and horizontal flipping of images. (2) Model initializing using pre-trained weights from the ImageNet[20] dataset.

3.3 Results and Analysis

Figure 9 presents the detection effect of an image of the test dataset, in which (a) is the result of baseline model Faster R-CNN while (b) is that of our model. From the comparison of results, it can be seen that the model proposed in this article can significantly reduce the confusion between defective bolts and normal bolts.



(a) (b)

Figure 9: Comparison of detection effect between Faster R-CNN and FG R-CNN

To demonstrate the superiority of our model, we compared it with several advanced object detection models, including Faster R-CNN, PANet[21] and YOLOv4[22]. During the training period, all compared models are based on the same training set and validation set. In addition, we perform the same data augmentation for all models. Comparison was conducted in the same test set, on which the experiment results are shown in Table 1. As shown in Table 1, FG R-CNN outperformed the comparative model in all metrics, with an mAP of 85.77%, 6.98% higher than the baseline model Faster R-CNN and 3.68% higher than the suboptimal PANet. The defect bolts AP of Faster R-CNN is only 70.07%, while our model reaches 82.09%, a relative increase of 12 percentage points, which proves that our proposed model significantly improves the accuracy of defective bolts detection and solves the problem of defect bolts detection in transmission line inspection through fine-grained recognition and mutual learning methods.

Table 1: Comparison with other methods of object detection on mAP

Methods	Defective bolts (%)	Normal bolts (%)	mAP (%)
Faster R-CNN	70.07	87.51	78.79
PANet	74.84	89.35	82.09
YOLOv4	58.81	86.92	72.56
FG R-CNN	**82.09**	**89.47**	**85.77**

For further analysis, we calculated the model precision and recall rate on the test set by class based on a confidence threshold of 0.5, as shown in Table 2. Precision is the ratio of true positive samples to the total samples predicted as positive by the model. It is used to measure the prediction accuracy of a classification model, as shown in Equation 10. Recall is the proportion of true positive samples that are correctly predicted as positive, as shown in Equation 11.

$$Precision = \frac{\sum_i^n TP_i}{\sum_i^n (TP_i + FP_i)} \tag{10}$$

$$Recall = \frac{\sum_i^n TP_i}{\sum_i^n (TP_i + FN_i)} \tag{11}$$

Table 2: Comparison with the baseline model in practical applications (classification score>0.5)

Methods	Defective bolts		Normal bolts	
	Precision (%)	Recall (%)	Precision (%)	Recall (%)
Faster R-CNN	57.57	72.48	88.44	87.05
FG R-CNN w/o L_{cc}	79.87	74.96	90.43	89.26
FG R-CNN	**80.10**	**76.14**	**93.05**	**89.91**

Compared with the baseline model Faster R-CNN, the precision of FG R-CNN defect bolts detection reaches 80.1%, and the normal bolts detection reaches 93.05%, which meets the requirements of practical applications. FG R-CNN also improved recall to a certain extent, with defective bolts reaching 76.14%, which can locate the vast majority of defective bolts. The method FG R-CNN w/o L_{cc} in Table 1 stand for the model that only the convolutional binary tree structure is introduced to baseline while the classification consistency loss is not applied, which indicates that the model achieves a detection precision of nearly 80% for defect bolts without introducing mutual learning, but there is a certain gap in recall compared to the final model. This shows that mutual learning can significantly improve recall, which means a reduction in missed detections for bolts in RPN stage and misclassified as background class. This is more important for inspection tasks of transmission lines, as detection is more difficult than classification. However, it can be seen that even if only fine-grained recognition improvement is conducted, the bolts detection precision can be greatly improved, but compared with the final model, the AP of defective bolts is 5.06% lower, which shows that the mutual learning of sub classifiers can improve the final detection precision, proving the effectiveness of mutual learning for the promotion of the tree structure.

3.4 Further Discussion

In this paper, we improve the representation of RoI features by introducing a convolutional binary tree structure to perform fine-grained RoI feature learning without any prior subcategory annotation. However, the spatial attention mechanism of the convolutional binary tree is limited to local RoI areas, which lacks global context information from the whole image. Although we integrated feature pyramids to the backbone to extract multi-scale features, the features are only fused on the channel and lacked spatial correlation, which makes it difficult for the network to recognize bolts with the help of the spatial structure and object semantics in the image and around the RoI. This is crucial for bolts detection, as bolts with similar RoI characteristics usually have different criteria for determining whether they are defective bolts due to their location on transmission lines. Therefore, we plan to integrate the spatial correlation information of the region proposal and other regions of the image into the fine-grained recognition network in the following work, so as to enable the network to obtain more comprehensive information from images during recognition of bolts.

Moreover, there is still significant room for improvement regarding our proposed mutual learning strategy. Specifically, the final prediction confidence of the convolutional binary tree is the weighted mean of the sub-classifiers of leaves, we simply sum the loss terms of all leaves when calculating the classification consistency loss. Although this approach is effective enough, it may not be the optimal strategy. We think that by summing the loss term with different weights and treating these weights as parameters for network learning and testing, it may achieve better results. This approach has been effective in the field of semi-supervised learning, but it has not been explored extensively in the field of supervised learning, which suggests there is a lot of research space to investigate the method further.

4. Conclusion

This article concludes that the identification of defective bolts is not just a simple binary object detection problem through the study of the types of bolt defects. We propose a new bolts detection method based on deep learning, which takes the convolutional binary tree as the object classification branch, and carries out fine-grained recognition of the RoI features of bolts without fine classification of the bolt defects in advance. In addition, we introduce the classification consistency loss, so that the sub-classifiers of leaves can learn from each other during training, which further improves the classification accuracy. We conducted experiments on real transmission line inspection image datasets, and the experimental results showed that our model's multiple evaluation indicators were significantly better than the baseline model and higher than the optimal object detection model. This proves that our proposed model has certain application value in the field of intelligent inspection of transmission lines.

References

[1] Yu, Sui, Ning Pingfan, and Niu Pingjuan, “Review on mounted UAV for transmission line inspection," Power System Technology, pp. 1-15, 2020.

[2] Junfeng, L. I., W. A. N. G. Qinruo, and L. I. Min. “Electric Equipment Image Recognition Based on Deep Learning and Random Forest,” High Voltage Engineering, 43(11), pp. 3705-3711, 2017.

[3] Yao, N., Hong, G., Guo, Y., & Zhang, T., “The Detection of Extra Matters on the Transmission Lines Based on the Filter Response and Appearance,” 2014 Seventh International Symposium on Computational Intelligence and Design, Vol. 2, pp. 542-545, 2014, IEEE.

[4] Mao, T., Ren, L., Yuan, F., Li, C., Zhang, L., Zhang, M., & Chen, Y., “Defect recognition method based on HOG and SVM for drone inspection images of power transmission line,” 2019 international conference on high performance big data and intelligent systems (HPBD&IS), pp. 254-257, 2019, IEEE.

[5] Shouguo, L., Kai, L., Yaohua, Q., Yunqi, L., Yang, S., & Zhenyu, L., “Automatic detection method for small size transmission lines defect based on improved yolov3,” 2020 International Conference on Communications, Information System and Computer Engineering (CISCE), pp.78-81, 2020, IEEE.

[6] Xiao, Y., Li, Z., Zhang, D., & Teng, L., “Detection of pin defects in aerial images based on cascaded convolutional neural network,” IEEE Access: pp. 73071-73082, 2021.

[7] Zhao, Z., Qi, H., Qi, Y., Zhang, K., Zhai, Y., & Zhao, W., “Detection method based on automatic visual shape clustering for pin-missing defect in transmission lines,” IEEE Transactions on Instrumentation and Measurement, pp. 6080-6091, 2020, 69(9).

[8] Jiao, R., Liu, Y., He, H., Ma, X., & Li, Z., “A deep learning model for small-size defective components detection in power transmission tower,” IEEE Transactions on Power Delivery, pp. 2551-2561, 2021, 37(4).

[9] Ren, S., He, K., Girshick, R., & Sun, J., “Faster r-cnn: Towards real-time object detection with region proposal networks,” Advances in neural information processing systems, 28, 2015.

[10] He, K., Zhang, X., Ren, S., & Sun, J., “Deep residual learning for image recognition,” In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770-778, 2016.

[11] Lin, T. Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S., “Feature pyramid networks for object detection,” In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2117-2125, 2017.

[12] He, K., Gkioxari, G., Dollár, P., & Girshick, R., “Mask r-cnn,” In Proceedings of the IEEE international conference on computer vision, pp. 2961-2969, 2017.

[13] Zheng, H., Fu, J., Zha, Z. J., & Luo, J, “Looking for the devil in the details: Learning trilinear attention sampling network for fine-grained image recognition.” In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 5012-5021, 2019.

[14] Zheng, H., Fu, J., Zha, Z. J., Luo, J., & Mei, T, “Learning rich part hierarchies with progressive attention networks for fine-grained image recognition,” IEEE Transactions on Image Processing, 29, pp. 476-488, 2019.

[15] Ji, R., Wen, L., Zhang, L., Du, D., Wu, Y., Zhao, C., ... & Huang, F, “Attention convolutional binary neural tree for fine-grained visual categorization,” In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 10468-10477, 2020.

[16] Hu, J., Shen, L., & Sun, G., “Squeeze-and-excitation networks,” In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 7132-7141, 2018.

[17] Zhang, Y., Xiang, T., Hospedales, T. M., & Lu, H., “Deep mutual learning,” In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4320-4328, 2018.

[18] Margolin, R., Zelnik-Manor, L., & Tal, A., “How to evaluate foreground maps?,” In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 248-255, 2014.

[19] Chen, K., Wang, J., Pang, J., Cao, Y., Xiong, Y., Li, X., ... & Lin, D., “MMDetection: Open mmlab detection toolbox and benchmark,” arXiv preprint arXiv:1906.07155, 2019.

[20] Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L., “Imagenet: A large-scale hierarchical image database,” In 2009 IEEE conference on computer vision and pattern recognition, pp. 248-255, 2019. IEEE.

[21] Wang, K., Liew, J. H., Zou, Y., Zhou, D., & Feng, J., “Panet: Few-shot image semantic segmentation with prototype alignment,” In proceedings of the IEEE/CVF international conference on computer vision, pp. 9197-9206, 2019.

[22] Bochkovskiy, A., Wang, C. Y., & Liao, H. Y. M., “Yolov4: Optimal speed and accuracy of object detection,” arXiv preprint arXiv:2004.10934, 2020.

Author Profile

Jimin Yu received the B.E. degree from North China Electric Power University, Beijing, China, in 2019. He is currently pursuing the M.Sc. degree in Computer Science and Technology with North China Electric Power University. His research interests include Computer Vision, Deep Learning and Computer Graphics.

Yilin Ma received the B.E. degree from North China Electric Power University, Beijing, China, in 2021. She is currently pursuing the M.Eng. degree in Computer Technology Engineering with North China Electric Power University. Her research interests include Computer Graphics, Load Forecast and Computer Vision.