SA RCNN for Occluded Damper Detection in the Inspection of Power Transmission Line

Feng Zhou*, Guoyou Wang and Yi Zheng
National Key Laboratory of Science and Technology on Multispectral Information Processing, School of Automation, Huazhong University of Science and Technology, Wuhan, China.
Email: m201872646@hust.edu.cn

Abstract. Dampers can prevent wires from breaking due to vibration, so it plays an important role in ensuring the stable operation of transmission lines. But the accuracy of popular object detection algorithm is still very low, because the occlusion phenomenon of dampers is serious due to the variable shooting angle. In this paper, we introduce our Supervised Attention RCNN (SA RCNN) to improve the detection effect of occluded dampers. SA RCNN mainly consists of two components: Box Supervised Attention Module (BSAM) and One Proposal Multiple Predictions (OPMP). In order to solve the problem of false detection caused by similarity of features, BSAM enhances features by using ground truths as supervised information to guide the generation of attention. In order to solve the problem of missed detection due to the NMS strategy, OPMP predicts a set of instances for one proposal rather than a single one. Experiments show that SA RCNN obtains 4.2% AP improvements on the dataset of occluded dampers compared to FPN.

1. Introduction
In the transmission line inspection, the traditional manual inspection method has high cost due to the complex and changeable environment. So unmanned aerial vehicle (UAV) transmission line inspection [1],[2] is wide-range applied. This is a developing trend to detect the defects of electrical equipment in the transmission line inspection by using the images taken by UAV. In the transmission line inspection, dampers are designed to prevent wires from breaking due to vibration. Therefore, the detection of dampers is particularly important for the stable operation of the power system and it has important practical significance. However, the occlusion phenomenon of dampers is serious due to the variable shooting angle of UAV. There are a large number of occlusion hammers in the dataset. This makes the detection to dampers appear serious misdetection problem.

The current popular detection algorithms have very good detection effect on sparse dampers, but there are still great challenges in the detection of occlusion dampers. The main occlusion of the damper belongs to the in-class occlusion problem. In-class occlusion of the damper is the occlusion between the damper and damper. The difference between the occlusion and other categories lies in that the object features of the same category are very similar, so the detection of in-class occlusion will be seriously disturbed. At the same time, due to the mutual occlusion between dampers, the IOU value of the prediction box exceeds the threshold which will be considered as the same damper by the detector resulting in the phenomenon of missed detection.

To solve the above two problems, this paper proposes a new structure SA RCNN to improve the detection effect of occluded dampers. SA RCNN mainly consists of two components: Box Supervised
Attention Module (BSAM) and One Proposal Multiple Predictions (OPMP). In order to solve false detection caused by the similarity of features, BSAM enhances features by using ground truths as supervised information to guide the generation of attention. In order to resolve the problems of missed detection due to the NMS strategy, OPMP predicts a set of instances for one proposal rather than a single one. At the same time, SA RCNN uses improved NMS to prevent the prediction results of the same proposal from inhibiting each other. SA RCNN is very simple. It only adds the Attentional mechanism and one prediction branch. But SA RCNN has a big improvement. SA RCNN obtains 4.6 AP improvements on the dataset of occlusion dampers compared to FPN[3]. Experimental results show the superiority of our method.

2. Related Work
In the damper detection, the effect of the general target detection algorithm [4],[5],[6] is very poor. The main reason is that dampers usually come in pairs and groups. When UAV shooting, dampers easily cover each other. Therefore, the detection of occlusion object must be studied.

**Advanced NMS.** The traditional NMS[7] algorithm deletes the detection box where the IOU exceeds the threshold. This is easy to cause the problem of missed detection and the setting of IOU greatly affects the result. Soft NMS[8] improves on this basis by reducing the confidence of boxes whose IOU is greater than the threshold rather than directly deleting them. However, the above method relies on the classification score and may leave the bounding boxes with the high score but the inaccurate position. Softer NMS[9] increases the branch of location confidence and uses the KL divergence to improve the loss function. Adaptive NMS[10] is aimed at the crowd scene and the NMS threshold is large in dense places, small in sparse places. However, these methods still have poor effect on highly overlapping objects. Set NMS[11] generates multiple prediction results in a single proposal and the prediction results in the same proposal will not suppress each other. But these methods need additional structure and can’t distinguish badly obscured objects.

**Attention mechanism for occlusion damper detection.** Face attention network[12] uses the attention mechanism to enhance the visible area of faces on the feature map and inserts attention module behind each layer of features to improve the detection effect of occluded faces. MGAN[13] uses the semantic information of real targets to guide the generation of attention modules, so as to improve the quality of ROI features. However, this method requires additional segmentation information. Although attention mechanism can improve the effect of occluded objects, but it needs a lot of extra annotations.

**Loss functions for occlusion damper detection.** Repulsion Loss[14] proposes a rejection Loss function which makes the distance between the prediction box and its ground truths smaller and makes the distance between the prediction box and surrounding objects bigger. OR-CNN[15] proposes the aggregation loss function to make the prediction box as close as possible to ground truths. Although the method of using the loss function can alleviate the occlusion problem, NMS is still needed and it is easy to cause miss detection phenomenon.

So we propose SA RCNN to improve the detection effect of occlusion dampers. SA RCNN uses supervised attention mechanism to enhance features and predicts multiple instances in a single proposal. At the same time, SA RCNN uses improved NMS to prevent the prediction results of the same proposal from inhibiting each other.

3. SA RCNN for Occlusion damper detection
SA RCNN is an extension of region-based convolutional neural networks and mainly consists of two components: Box Supervised Attention Module (BSAM) and One Proposal Multiple Predictions (OPMP). The architecture is shown in figure 1. BSAM enhances features by using ground truths as supervised information to guide the generation of attention. OPMP predicts a set of instances for one proposal rather than a single one. At the same time, SA RCNN uses improved NMS to prevent the prediction results of the same proposal from inhibiting each other. SA RCNN will be introduced as follows:
3.1. Box Supervised Attention Module (BSAM)

The architecture of BSAM is shown in figure 2. The inputs of BSAM are features from backbone. Features go through two 3x3 convolution layers and dimension reduction is carried out by 1x1 convolution layer. Finally the sigmoid function maps the values between [0,1] to get the attention feature map. Also, we produce a BBox Mask by bounding box information and use it to supervise the generation of attention. The area in the ground truth is set to 1 and the area outside the ground truth is set to 0, thus we obtain a rough segmentation information BBox Mask. Feature is multiplied by supervised attention mask.

Attention Loss is designed to measure the difference between supervisory information BBox Mask and spatial Attention:

\[ L_{Attention}(p_{x,y}, g_{x,y}) = (1 - g_{x,y}) \times CELoss(p_{x,y}, g_{x,y}) \] (1)

\( p_{x,y} \) represents the value of spatial attention feature at the coordinate \((x,y)\), \( g_{x,y} \) represents the value of BBox Mask at the coordinate \((x,y)\), \( p_{x,y}, g_{x,y} \in (0,1) \), \( 1 - g_{x,y} \) represents the loss function only calculates the loss of the background area.

3.2. One Proposal Multiple Predictions (OPMP)

OPMP predicts a set of instances rather than a single object for one proposal as shown in figure 3. If we only detect one object in one proposal, it is very difficult to detect all dampers in the occluded area because occluded dampers has similar features. If we predict a set of instances in one proposal, we can detect all dampers easily.

However, if the traditional NMS algorithm is still used, the results predicted by OPMP will also be suppressed due to the high overlap. Therefore, we propose the improved NMS algorithm to replace the traditional NMS algorithm. The core idea is that every time the prediction box starts to suppress
another prediction box, two prediction boxes should be judged whether they come from the same suggested region. If both are the prediction results calculated from the same proposal, the suppression process will be skipped.

![Figure 3](image_url)

Figure 3. The method (a) and our method (b) are used to predict the occluded problem, where the red and blue boxes represent the proposal of the network respectively.

4. Experiments

In this section, we evaluate our SA RCNN on our damper dataset and compare it to different methods for occluded objects.

4.1. Dataset

Our damper dataset is provided by Wuhan NARI Co. Ltd, which is owned by China's State Grid Corp. All images are taken by UAV. When constructing the dataset, images containing dampers are selected. There are a total of 4262 images in the dataset. We divide the images according to the 4:1:1 ratio and obtained 2842 images for training, 710 images for validation and 710 images for testing. Due to the influence of UAV shooting angle, the occlusion phenomenon is serious. We calculate the IOU for each damper and other damper. If the IOU of the damper and other damper is less than 0.35, the damper is considered sparse; if the IOU of the damper and other damper is greater than 0.35, the damper is considered crowded. The statistical results are shown in table 1. The AP value is used to evaluate the results.

| object | total | Sparse | Crowd |
|--------|-------|--------|-------|
| damper | 3562  | 2525   | 1037  |

4.2. Setup

We use FPN as the basic framework. Parameters of SA RCNN are initialized by Xavier method. The backbone network of the detection model selects resnet50. In this paper, we use the pretrained resnet50 model on the ImageNet dataset to initialize the parameters. The training of the model is optimized by SGD. The initial learning rate is set to 0.01, the batch-size is set to 2 and the momentum factor is set to 0.95. A total of 30 epochs are trained. The learning rate is in the 10th epochs, and the 20th epochs is decayed to one tenth of the previous epoch. During the model test, the mean and
variance are based on the training data. The confidence threshold is set to 0.5 and the IOU threshold is set to 0.5. We set the input image size to 1000×600.

4.3 Main Results

As shown in Table 2. Compared with FPN, our method is improved by 4.2% in AP and AP in the dense area is improved especially by 7.2%. Compared with other occlusion algorithms, SA RCNN also performs better. Compared with Repulsion Loss, SA RCNN increases by 1.7% in AP and compared with the Adaptive NMS, SA RCNN increases 1.5% in AP. We can see the superiority of our method.

| Method          | AP (total) | AP (Sparse) | AP (Crowd) |
|-----------------|------------|-------------|------------|
| FPN             | 65.1       | 79.5        | 47.1       |
| Repulsion Loss  | 67.6       | 80.2        | 51.5       |
| Adaptive NMS    | 67.8       | 80.3        | 51.9       |
| SA RCNN         | 69.3       | 80.8        | 54.3       |

4.4 Ablation Study

As shown in Table 3. Box Supervised Attention Module obtains 1.8% improvements in AP and One Proposal Multiple Predictions obtains 3.1% improvements. When two modules are combined, the effect is best and the AP value is increased by 4.2%. This shows the effectiveness of our method.

| Method          | BSAM | OPMP | AP (total) | AP (Sparse) | AP (Crowd) |
|-----------------|------|------|------------|-------------|------------|
| FPN             | √    |      | 65.1       | 79.5        | 47.1       |
|                 |      | √    | 66.9       | 79.9        | 50.4       |
|                 | √    | √    | 68.2       | 80.5        | 52.8       |
|                 |      |      | 69.3       | 80.8        | 54.3       |

4.5 Visual comparison

By comparing the visualization results of Adaptive NMS and SA RCNN, we can find that SA RCNN has better effect.
5. Conclusion
In this paper, SA RCNN is proposed to solve the occlusion problem of dampers. SA RCNN is mainly composed of BSAM and OPMP. BSAM uses ground truths to generate rough segmentation mask to supervise the learning of attention mechanisms, OPMP predicts multiple instances in each proposal region and improved NMS algorithm is used so that the prediction results from the same proposal will not be suppressed. Experimental results show that SA RCNN is better than other occlusion algorithms.

Acknowledgement
This work is supported by Image Processing & Intelligent Key Laboratory in Huazhong University of Science and Technology. Thank my tutor Guoyou Wang for his guidance.

References
[1] Li, Z., Mu, S., Li, J., Wang, W., Liu, Y. (2016) Transmission line intelligent inspection central control and mass data processing system and application based on UAV. In: 2016 4th International Conference on Applied Robotics for the Power Industry (CARPI). pp. 1-5.
[2] Zhang, F., Wang, W., Zhao, Y., Li, P., Lin, Q., Jiang, L. (2016) Automatic diagnosis system of transmission line abnormalities and defects based on UAV. In: 2016 4th International Conference on Applied Robotics for the Power Industry (CARPI). pp. 1-5.
[3] Lin, T., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S. (2017) Feature pyramid networks for object detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2117-2125.
[4] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C., Berg, A.C. (2016) Ssd: Single shot multibox detector. In: European conference on computer vision. pp. 21-37.
[5] Ren, S., He, K., Girshick, R., Sun, J. (2015) Faster r-cnn: Towards real-time object detection with region proposal networks. arXiv preprint arXiv:1506.01497.
[6] Lin, T., Goyal, P., Girshick, R., He, K., Dollár, P. (2017) Focal loss for dense object detection. In: Proceedings of the IEEE international conference on computer vision. pp. 2980-2988.
[7] Neubeck, A., Van Gool, L. (2006) Efficient non-maximum suppression. In: 18th International Conference on Pattern Recognition (ICPR'06). pp. 850-855.
[8] Bodla, N., Singh, B., Chellappa, R., Davis, L.S. (2017) Soft-NMS--improving object detection with one line of code. In: Proceedings of the IEEE international conference on computer vision. pp. 5561-5569.
[9] He, Y., Zhang, X., Savvides, M., Kitani, K. (2018) Softer-nms: Rethinking bounding box regression for accurate object detection. arXiv preprint arXiv:1809.08545. 2,3.
[10] Liu, S., Huang, D., Wang, Y. (2019) Adaptive nms: Refining pedestrian detection in a crowd. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 6459-6468.
[11] Chu, X., Zheng, A., Zhang, X., Sun, J. (2020) Detection in crowded scenes: One proposal, multiple predictions. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 12214-12223.
[12] Wang, J., Yuan, Y., Yu, G. (2017) Face attention network: An effective face detector for the
occluded faces. arXiv preprint arXiv:1711.07246.

[13] Pang, Y., Xie, J., Khan, M.H., Anwer, R.M., Khan, F.S., Shao, L. (2019) Mask-guided attention network for occluded pedestrian detection. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 4967-4975.

[14] Wang, X., Xiao, T., Jiang, Y., Shao, S., Sun, J., Shen, C. (2018) Repulsion loss: Detecting pedestrians in a crowd. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 7774-7783.

[15] Zhang, S., Wen, L., Bian, X., Lei, Z., Li, S.Z. (2018) Occlusion-aware R-CNN: detecting pedestrians in a crowd. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 637-653.