Hierarchical annotation method for metal corrosion detection of power equipment

Zhang Baili[1]　　Cao Yong[1]　　Zhang Pei[2,3]　　Zhang Zhao[2,3]　　He Yina[1]　　Zhong Mingjun[4]

([1] School of Computer Science and Engineering, Southeast University, Nanjing 211189, China)
([2] State Key Laboratory of Smart Grid Protection and Control, State Grid Electric Power Research Institute, Nanjing 211106, China)
([3] NARI Group Corporation, Nanjing 211106, China)
([4] Department of Computing Science, University of Aberdeen, Aberdeen AB24 3UE, UK)

Abstract: To solve the ambiguity and uncertainty in the labeling process of power equipment corrosion datasets, a novel hierarchical annotation method (HAM) is proposed. Firstly, large boxes are used to label a large area covering the range of corrosion, provided that the area is visually continuous and adjacent to corrosion that cannot be clearly divided. Secondly, in each labeling box established in the first step, regions with distinct corrosion and relative independence are labeled to form a second layer of nested boxes. Finally, a series of comparative experiments are conducted with other common annotation methods to validate the effectiveness of HAM. The experimental results show that, with the help of HAM, the recall of YOLOv5 increases from 50. 79% to 59. 41% ; the recall of Faster R-CNN + VGG16 increases from 66. 50% to 78. 94% ; the recall of Faster R-CNN + Res101 increases from 78. 32% to 84. 61% . Therefore, HAM can effectively improve the detection ability of mainstream models in detecting metal corrosion.

Key words: deep learning; Faster R-CNN; YOLOv5; object detection; hierarchical annotation

DOI: 10. 3969/j. issn. 1003 − 7985. 2021. 04. 002

Corrosion is the result of deterioration caused by the reaction of metallic materials' surface and internal micro-structure with corrosive environments[1-2]. With long-term exposure to the external environment, corrosion has become a common defect of power equipment. Corrosion can substantially shorten the power equipment's life span and cause economic losses if not tackled promptly. Hence, it is important to regularly detect and eliminate corrosion in time.

Currently, non-destructive methods are usually applied to detect corrosion. These include X-rays[3], local wave-number[4], infrared thermography[5-6], magneto-optic imaging[7], and cameras. However, all these methods have strict requirements related to the testing environment, testing equipment, and the professional level of testing personnel. At present, convolutional neural networks (CNNs)[8] are employed to perform corrosion detection using collected RGB images.

Since the first appearance of AlexNet[9] in ImageNet competition, methods using CNNs for feature extraction and image classification have rapidly developed and are now widely used in image recognition and object detection across various fields. Among the current CNN-based object detection models, prominent architectures include Faster R-CNN[10] and YOLO[11-12].

However, due to the irregular shape and detachable nature of corrosion, these detection models cannot directly achieve satisfying results[13]. The main reason is that traditional annotation methods entail ambiguity and uncertainty during the labeling process, making it difficult for this process to be unified. As a result, inconsistent annotation causes problems for detection models intending to effectively learn the inherent features of metal corrosion during the training process. At present, three main data annotation approaches are used in object detection tasks: two-dimensional (2D) bounding box, three-dimensional (3D) cuboids, and polygonal segmentation.

While the 2D bounding-box approach is relatively simple and widely used, the rectangular area usually contains non-target objects. Particularly relevant to the present context is that metal corrosion exhibits obvious irregularities, resulting in a large area of non-corrosion in the labeled box. Ref. [14] used a sliding window to intercept a small area so as to reduce the non-corroded area contained in the window. Subsequently, a small CNN was used to judge whether each window contained a corroded area. However, this method divides an image into many tiny windows, so the model only focuses on local pixel data in each window and ignores global pixel data. Thus, the features learned by the model are too partial and one-sided. Meanwhile, due to the size of the window, the designed CNN should be small, which limits the CNN's learning ability. 3D cuboids can display the depth of the

target, but this is rarely required for metal corrosion detection. Moreover, polygonal segmentation can fit the target shape well and resolve the shortcomings of the bounding-box method. However, this annotation method is highly time-consuming and costly and hence is rarely used on a large scale in practice.

This paper proposes a novel hierarchical annotation method (HAM). Firstly, large boxes were used to label a large area covering the range of corrosion, provided that the area is visually continuous and adjacent to corrosion that cannot be clearly divided. Secondly, within each box established in the first step, regions with distinct corrosion and relative independence were labeled to create a second layer of nested boxes. This method is simple and can easily produce unified and unambiguous annotation results. In addition, it highlights the corrosion features and increases the number of ground truth, which enables data augmentation to be achieved and makes it easier for detection models to learn the inherent features of corrosion during the training process.

In summary, this work makes three main contributions: Firstly, a novel method named HAM is proposed to accurately detect metal corrosion in power equipment. Secondly, a detection box merging algorithm is applied to merge intersecting boxes. Finally, novel definitions of precision and recall are proposed in view of corrosion features.

1 HAM

1.1 Labeling of training samples

Due to the diversity of equipment and ways in which corrosion spreads, the shape and size of corroded areas tend to be highly irregular. Meanwhile, corrosion is often detachable and can thus be regarded as either a whole area or several small independent areas. In view of the characteristics of metal corrosion, two methods are generally used in applying the 2D bounding-box approach to label corrosion areas:

1) Fine-grained annotation method (FAM). Using the smallest possible box to label the corrosion area

2) Coarse-grained annotation method (CAM). Using the largest possible box to label the corrosion area

FAM can minimize the area containing non-target objects within a rectangular box. However, the tendency of metal corrosion to spread will cause significant ambiguity in the labeling process, yielding inconsistent annotation results. One simple solution is to ignore corrosion spreading areas. Fig. 1 (a) shows some unlabeled corrosion spreading areas near the labeled corrosion. However, corrosion spreading is such a highly important feature of metal corrosion where ignorance will significantly reduce the annotation quality of the dataset.

CAM adopts a large box to label a dataset, specifically one large enough to encompass all areas of spreading
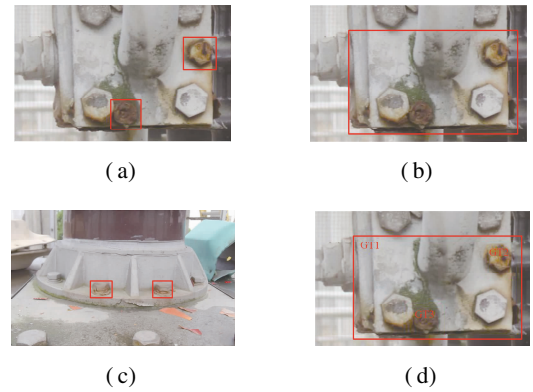

(a)   (b)

(c)   (d)

Fig. 1 Performance of the labeling process using different methods. (a) The annotation result by FAM; (b) The annotation result by CAM; (c) The detachability of metal corrosion; (d) The annotation result by HAM

corrosion. Thus, the problem encountered by FAM can be avoided during the labeling process. The annotation results for this method are shown in Fig. 1(b). However, CAM also has certain shortcomings: Firstly, the number of non-target objects contained in the labeling box greatly increases. Moreover, different objects may have the same annotation results, whereas similar objects may have different annotations. For example, this method causes the normal nut in the lower-right corner and the corroded nut in the top-right corner in Fig. 1(b) to be labeled as a part of the labeling box. Meanwhile, the corroded nut in Fig. 1(b) is labeled as a part of the corrosion area, but another similar corroded nut in Fig. 1(c) is labeled as an independent corrosion area. Hence, CAM results in the instability of detection model training.

The HAM proposed in this paper is adopted to solve the above problems. Firstly, a large-enough box labels a large area covering the range of corrosion, as long as the area is visually continuous and adjacent to corrosion that cannot be clearly divided; an example is shown in GT1 in Fig. 1(d). Then, based on the previous labeling box, areas with distinct features and relative independence are labeled twice to form a second layer of nested labeling, as shown in GT2 and GT3 in Fig. 1(d).

Therefore, the first operation of HAM provides a solution to the problem of labeling omission in FAM using a large box to annotate metal corrosion, and the second operation of HAM reannotates independent corrosion in the large box to refrain the problem of labeling ambiguity in CAM. At the same time, the number of ground truth per sample is increased after using HAM. Thus, data augmentation is achieved to a certain extent, which will counteract the negative impact of the rectangular box containing non-target objects.

1.2 Detection box merging algorithm

Although HAM can effectively solve labeling omission and labeling ambiguity, there must be labeling

nested boxes in the outer labeling boxes. Therefore, in the final detection results, the models trained by a hierarchical labeling dataset will also output nesting or intersection detection boxes, as shown in Fig. 2 in detection boxes A and B. Although this condition does not affect actual detection results, it interferes with the visualization of results and the calculation of the detection models' precision and recall. Accordingly, we transformed the detection results in this study. The intersection and nesting detection boxes were merged to form the final annotation result. That is, in Fig. 2, orange box C is adopted to replace A and B as the final annotation result.
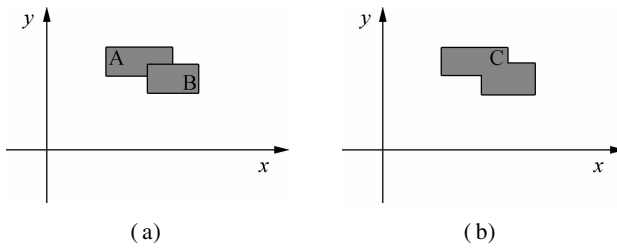

Fig. 2 Diagram of merging boxes. (a) Detection results before merging; (b) Detection results after merging

1.3 Evaluation criteria

Based on the above discussion, one corrosion sample may obtain different annotation results due to the use of different annotation methods. Consequently, the test sets labeled by different methods are also different. Thus, the performance of models trained by the three annotation methods cannot be meaningfully compared. To solve this problem, in this study, the test set was labeled uniformly using polygonal segmentation to create a unified test set. Some of the labeled samples are shown in Fig. 3.
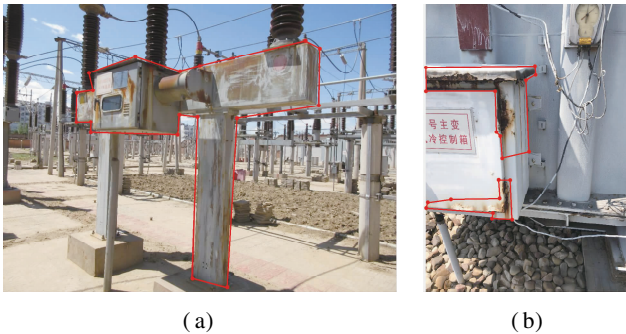

Fig. 3 Two annotation results using polygons. (a) One corrosion area labeled; (b) Two corrosion areas labeled

In the experiment, precision and recall were adopted as the evaluation criteria. According to the positive and negative of predicted results and actual results, the classification results can be divided into four categories, as shown in Tab. 1.

Tab. 1 Confusion matrix for binary classification

Actual results	Predicted results	
	Positive	Negative
Positive	TP	FN
Negative	FP	TN

For object detection, the intersection over union (IoU) is typically used to evaluate whether or not the detection result is correct. The IoU calculates the ratio of the intersection and union of the "detection box" and "ground truth." Usually, the threshold value of the IoU is set to 0.5 in actual projects. When the value of the IoU is larger than 0.5, the target has been detected. However, in this study, the test set was labeled via polygonal segmentation. As a result, the value of the IoU of the detection box and ground truth is often very low. As shown in Fig. 4, the value of the IoU is lower than 0.5; thus, it is not suitable to directly use the IoU to judge whether one predicted box is correct.

Fig. 4 Large non-corroded area

Accordingly, the precision and recall of the models were evaluated using the following expressions.

As shown in Fig 4, even if the model effectively detects the corrosion area, several non-corroded areas are present in the results. Therefore, if the ratio between the area sum of correct detection areas and the area sum of detection areas is used as the precision, then the calculated precision is often very low, and the performance of the models cannot be correctly evaluated. Hence, in this study, the precision was calculated with reference to the number of detection areas. The expression for calculating the precision is

$$P = \frac{N_{ca}}{N_{da}}$$

where P is the precision of the models; N_{ca} is the number of correct detection areas; and N_{da} is the number of detection areas.

To reduce the influence of image size, all image sizes were normalized before calculation. The expression for calculating the recall is

$$R = \frac{S_{ca}}{S_{ta}}$$

where R is the recall of the models; S_{ca} is the area sum of the correct detection area; S_{ta} is the area sum of the true corrosion areas.

Therefore, the unified test set was used to ensure that the three annotation methods can be compared correctly through the use of the two calculation expressions presented above. Finally, the performance of different models can be evaluated realistically.

2 Experimental Results and Analysis

2.1 Experimental environment

The experimental software and hardware environments are presented in Tab. 2.

Tab. 2 Experimental software and hardware environments

Configuration	Description
CPU	I7-9800X/8@4.4GHz
RAM	32GB
GPU	Nvidia RTX2080Ti@11G
OS	Centos7
Deep learning framework	PyTorch
Object detection model	Faster R-CNN, YOLOv5[15]
Dataset	Collected from actual substations by Nari Corp; 1 180 training set, 199 test set
Evaluation indicators	Precision, Recall

2.2 Experimental procedure

Firstly, this study used polygonal segmentation to label 199 images as the test set. Secondly, 1 180 images were labeled by three annotation methods for training Faster R-CNN and YOLOv5 models. Subsequently, VGG16 or Res101[16] (as the backbone networks) and SGD[17] (as the optimizer) were applied to train Faster R-CNN, whereas DarkNet53[11] (as the backbone network) and SGD (as the optimizer) were applied to train YOLOv5. Finally, after the detection box merging algorithm was used to merge the intersecting boxes of each image, the precision and recall of the detection models were calculated. The experimental results are shown in Tab. 3.

Tab. 3 Experimental results %

Model	Annotation method	Precision	Recall
YOLOv5	FAM	89.58	51.20
	CAM	93.84	50.79
	HAM	91.14	59.41
Faster R-CNN + VGG16	FAM	84.72	66.50
	CAM	84.69	70.51
	HAM	82.95	78.94
Faster R-CNN + Res101	FAM	80.49	78.32
	CAM	82.31	82.69
	HAM	81.48	84.61

2.3 Discussion

In the experiment, Faster R-CNN + Res101 based on HAM had the best detection result with the highest recall, as shown in Fig. 5. In addition, Faster R-CNN achieved a higher recall than YOLOv5, although its precision was lower. For Faster R-CNN, using Res101 as the backbone network resulted in a higher recall than that using VGG16, and its precision was slightly reduced. With HAM, YOLOv5 and Faster R-CNN + VGG16 exhibited great improvement in recall and kept slight fluctuations in precision. Furthermore, the recall of Faster R-CNN + Res101 reached the maximum out of all the experiments.
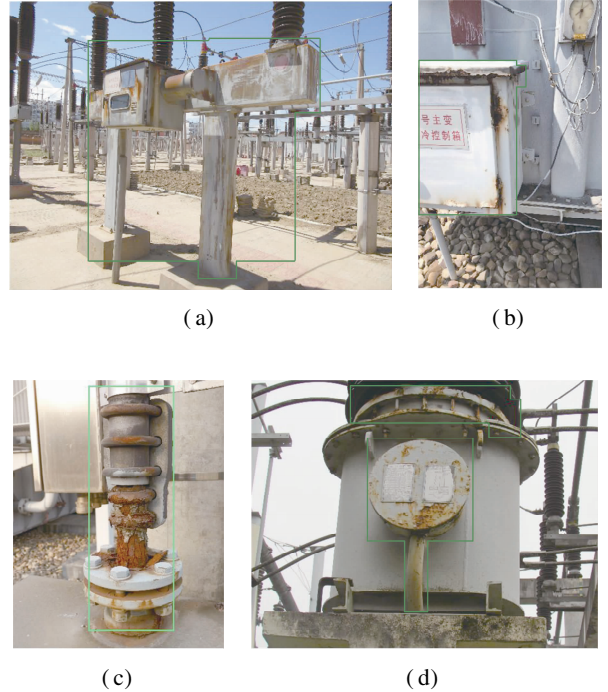

Fig. 5 Some outputs of Faster R-CNN + Res101 using HAM. (a) One detection result by merging three boxes; (b) One detection result by merging two boxes; (c) One detection result without merging; (d) Two detection results by merging four boxes

According to the actual requirements of corrosion detection and processing procedures, further analysis of the experimental results shows the following findings:

1) Although YOLOv5 has the highest precision, it also has the lowest recall. In practical applications, it is often necessary to manually review the detection results of a model. Therefore, compared with higher precision, it is preferable to have a higher recall for corrosion detection models to avoid the omission of corrosion to the greatest extent possible.

2) In Faster R-CNN, Res101 has a higher recall than VGG16, and its precision is slightly lower. Because the structure of Res101 is more complicated than that of VGG16, Res101 has more convolution layers than VGG16 and can thus learn more features. At the same time, Res101 uses batch normalization[18] and rectified linear unit activation function[19] to ensure that the gradient can achieve good back propagation and speed up convergence. Simultaneously, Res101 enables the model to be fully trained and not degrade through the residual module. Therefore, its recall is higher.

3) After using HAM, the recall of each model clearly increases as compared with those of the other two meth-

ods, as shown in Tab. 4. The models used in the experiment are all object detection models based on region proposal; that is, the object detection models detect objects with different sizes using predefined anchors with different sizes. Therefore, compared with FAM, HAM increases the number of larger predicted boxes. Although this does not affect the detection ability of small-size anchors, it improves that of large-size ones, thereby strengthening the models' comprehensive detection ability. The analysis of CAM reveals the same results. Therefore, through the use of HAM, the detection effect of models will be improved.

Tab. 4　Recall of models with different methods　　%

Model	FAM	CAM	HAM
Res101	78.32	82.69	84.61
VGG16	66.50	70.51	78.94
YOLOv5	51.20	50.79	59.41

3　Conclusions

1) The training effect of object detection models is highly dependent on the availability of sufficient and high-quality datasets. Traditional annotation methods encounter ambiguity and uncertainty because of the irregularity and detachability of metal corrosion. Therefore, when Faster R-CNN and YOLOv5 are applied to directly detect corrosion in power equipment, it is difficult to standardize the labeling process of training samples and keep annotation results consistent.

2) In this paper, a novel HAM that utilizes the characteristics of corrosion is proposed. Meanwhile, a detection box merging algorithm is applied to solve the problem of nested boxes.

3) Ultimately, according to experimental findings, the corrosion detection ability of YOLOv5 and Faster-R-CNN models is found to be greatly improved after adopting HAM. The models further obtain better generalization ability, which can be popularized and widely applied in practice.

References

[1] Roberge P R. *Handbook of corrosion engineering*[M]. New York, USA: McGraw-Hill, 2019: 25−61.

[2] Volkan C. *Corrosion engineering*[M]. Beverly: Scrivener Publishing LLC, 2014: 1−19.

[3] Dunn W L, Yacout A M. Corrosion detection in aircraft by X-ray backscatter methods[J]. *Applied Radiation and Isotopes*, 2000, **53**(4): 625−632. DOI: 10.1016/S0969-8043(00)00240-2.

[4] Gao T, Sun H, Hong Y. Hidden corrosion detection using laser ultrasonic guided waves with multi-frequency local wavenumber estimation[J]. *Ultrasonics*, 2020, **108**: 106182. DOI: 10.1016/j.ultras.2020.106182.

[5] Doshvarpassand S, Wu C, Wang X. An overview of corrosion defect characterization using active infrared thermography[J]. *Infrared Physics & Technology*, 2019, **96**: 366−389. DOI: 10.1016/j.infrared.2018.12.006.

[6] Wicker M, Alduse B P, Jung S. Detection of hidden corrosion in metal roofing shingles utilizing infrared thermography[J]. *Journal of Building Engineering*, 2018, **20**: 201−207. DOI: 10.1016/j.jobe.2018.07.018.

[7] Dudziak M J, Chervonenkis A Y, Chinarov V. Nondestructive evaluation for crack, corrosion, and stress detection for metal assemblies and structures[C]//*Nondestructive Evaluation of Aging Aircraft, Airports, and Aerospace Hardware* Ⅲ. International Society for Optics and Photonics, Newport Beach, CA, USA, 1999, **3586**: 20−31. DOI: 10.1117/12.339888.

[8] LeCun Y, Boser B, Denker J S. Backpropagation applied to handwritten zip code recognition[J]. *Neural computation*, 1989, **1**(4): 541−551. DOI: 10.1162/neco.1989.1.4.541.

[9] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks [J]. *Communications of the ACM*, 2017, **60**: 84−90. DOI: 10.1145/3065386.

[10] Ren S, He K, Girshick R. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. *IEEE transactions on pattern analysis and machine intelligence*, 2016, **39**(6): 1137−1149. DOI: 10.1109/TPAMI.2016.2577031.

[11] Redmon J, Divvala S, Girshick R. You only look once: Unified, real-time object detection[C]// *IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas, NV, USA, 2016: 779−788. DOI: 10.1109/CVPR.2016.91.

[12] Redmon J, Farhadi A. YOLO9000: Better, faster, stronger[C]// *IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu, Hawaii, US, 2017: 6517−6525, doi: 10.1109/CVPR.2017.690.

[13] Fan H B, Hu X X, Liu Y M. Application of deep learning in corrosion detection of power equipment[J]. *Guangdong Electric Power*, 2020, **33**(9): 154−165. DOI: 10.3969/j.issn.1007−290X.2020.009.020. (in Chinese)

[14] Yao Y, Yang Y, Wang Y. Artificial intelligence-based hull structural plate corrosion damage detection and recognition using convolutional neural network[J]. *Applied Ocean Research*, 2019, **90**: 101823. DOI: 10.1016/j.apor.2019.05.008.

[15] Glenn J, Alex S, Jirka B. Ultralytics/yolov5: v3.1-bug fixes and performance improvements[EB/OL]. (2020-10-29) [2021-02-01]. https://zenodo.org/record/4154370#.YT4F4nbJ3Us. DOI: 10.5281/zenodo.4154370.

[16] He K, Zhang X, Ren S. Deep residual learning for image recognition[C]// *IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas, NV, USA, 2016: 770−778.

[17] Sutskever I, Martens J, Dahl G. On the importance of initialization and momentum in deep learning[C]//*International Conference on Machine Learning*. Atlanta, USA, 2013: 1139−1147.

[18] Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift[C]//*International Conference on Machine Learning*. Stroudsburg, PA, USA, 2015, **37**: 448−456. DOI: 10.5555/3045118.3045167.

[19] Nair V, Hinton G E. Rectified linear units improve restricted boltzmann machines[C]// *International Conference on International Conference on Machine Learning*. Haifa, Israel, 2010: 807–814.

面向变电设备金属锈蚀检测的分层嵌套标注方法

张柏礼[1]　曹　勇[1]　张　沛[2,3]　张　昭[2,3]　贺依娜[1]　钟明军[4]

([1] 东南大学计算机科学与工程学院, 南京 211189)
([2] 国网电力科学研究院智能电网保护和运行控制国家重点实验室, 南京 211106)
([3] 南瑞集团有限公司, 南京 211106)
([4] Department of Computing Science, University of Aberdeen, Aberdeen AB24 3UE, UK)

摘要: 为解决变电设备锈蚀数据集标注过程中经常遭遇的模糊和不确定问题, 提出了一种基于分层嵌套的训练样本标注方法. 首先, 采用较大的矩形框对锈蚀区域进行大面积标注, 将视觉上连续的、相邻的或不能清晰划分的锈蚀区域用一个矩形框标注; 然后, 在标注框内对特征明显并具有相对独立性的区域进行二次标注, 形成第 2 层内部嵌套标注. 为了验证分层嵌套方法的有效性, 与常用标记方法进行对比实验. 结果表明, 采用分层嵌套标注方法后, YOLOv5 模型的召回率由 50.79% 提升至 59.41%, Faster R-CNN + VGG16 模型的召回率由 66.50% 提升至 78.94%, Faster R-CNN + Res101 模型的召回率由 78.32% 提升至 84.61%. 由此可见, 通过分层嵌套标注可以有效提升主流模型在金属锈蚀方面的检测能力.

关键词: 深度学习; Faster R-CNN; YOLOv5; 目标检测; 分层嵌套

中图分类号: TP181