ITM Web of Conferences **45**, 01013 (2022) https://doi.org/10.1051/itmconf/20224501013
CSCNS2021

Angle steel tower bolt defect detection based on YOLO-V3

Jingfeng Zhang[1,*], *Yuanwei* Hu[2], and *Shujun* Ji[2]

[1]State Grid Anhui Electric Power Co., Ltd., Hefei, Anhui 230022, China
[2]Department of Applied Mathematics, Wenzhou-Kean University, Wenzhou, Zhejian 325000, China

Abstract. The bolts in the angle steel tower are seriously affected by corrosion and loss. This paper proposes a novel detection system based on YOLO-V3 to avoid the danger of traditional manual detection method for the bolt fault detection of the angle steel tower. A multi-scale convolution module is used to replace the ordinary convolution of original YOLO-V3 so as to obtain the spatial characteristics information of different scales in the image, and enhance the detection accuracy. The experimental results show that mAP of the proposed YOLO-SKIP network is 0.91. Our YOLO-SKIP model has achieved the best detection performance on the defective angle steel tower bolt data.

1 Introduction

The bolt of the angle steel tower is an important insulating component in the overhead transmission network. It is necessary to safely and effectively detect the bolt failure of the angle steel tower. However, the bolt defect detection of angle steel tower is basically in a blank state and can only rely on manual methods. In the detection of defects on power lines, researchers have designed many detection schemes for actual application scenarios. Zhai [1] *et al*. used to obtain the salient characteristics of insulators and combined with adaptive morphological algorithms to detect faulty insulators. Han [2] et al. designed an inspection image fault detection algorithm based on maximum between-class variance and maximum entropy segmentation. With the development of deep learning, fault detection algorithms based on convolutional neural networks have emerged. Yang [3] et al. improved the filter size of the fully convolutional neural network and the fully connected layer. Xian [4] et al. designed a cascaded convolutional neural network, which used a two-stage cascaded detection method to effectively improve the detection accuracy. Huang [5] et al. proposed a convolution feature pyramid detection method, combined with multi-task learning for network training. The model of this algorithm performs well, but the robustness is poor.

There are many types of angle steel tower bolt defects, such as rust, fall off, etc. This paper divides the collected data sets into only two categories, defects and no defects. The current detection algorithms based on deep learning are mainly divided into one-stage and two-stage. The two-stage target detection algorithm first performs feature extraction, and then classifies the extracted features, such as the R-CNN [6] series of networks. The one-

* Corresponding author: ahu0086@163.com

stage models include SSD [7], Retina-Net [8] and YOLO v3 [9], which use an end-to-end approach for direct detection. This paper proposes a multi-scale real-time fault detection model based on the target detection algorithm YOLO v3. This paper designs a multi-scale convolution module to replace the ordinary convolution of the original YOLO v3, and adopts a processing method in which the sampling probability of negative samples is higher than that of positive samples, in response to the imbalance of positive and negative samples in practical applications. Experiments showed that the two strategies enhance the defect detection performance of the inspection model for the angle steel tower of the overhead line.

2 Multi-scale target detection model

YOLO is a target detection and location algorithm based on a deep convolutional neural network. It divides the picture into several grids and predicts the candidate frame at the network scale to finally obtain the category probability and target. The calculation speed of YOLO algorithm is relatively fast, and it can be applied to real-time mark detection system.

2.1 Multi-scale convolution module

This work uses a multi-scale convolution module to replace the single-scale convolution in YOLO. Four convolution kernels are used to extract features from the input feature map, Conv-1x1, Conv-3x3, Conv-5x5, and Con-7x7, so as to obtain an output feature map with rich scale information. For example, the length and width of the convolution kernel of Conv-1x1 are 1, and the step size of all convolutions is 1. Then, the four feature maps of different scales are spliced in the channel direction through the "Concat" operation. Record the convolution functions of the 4 scales as $f_k(\cdot), k = 1,3,5,7$. Remember that the input feature map of the layer i is $\mathrm{M_i}$, the number of channels is $\mathrm{C_i}$, the length is $\mathrm{H_i}$ and the width is $\mathrm{W_i}$, then the calculation formula for the splicing process of the multi-scale convolution is

$$M_{cat} = Concat(f_{1\times1}(M_i),\ f_{3\times3}(M_i),\ f_{5\times5}(M_i),\ f_{7\times7}(M_i)) \tag{1}$$

At this time, the length and width of $\mathrm{M_{cat}}$ are $\mathrm{H_i}$ and $\mathrm{W_i}$, and the number of channels becomes $4 \times \mathrm{C_i}$. Finally, 1x1 convolution is used for feature dimensionality reduction to ensure that the number of output channels is the same as the number of input channels.

$$M_{out} = f_{1\times1}(M_{cat}) \tag{2}$$

2.2 Network feature extraction module

As deep convolutional neural network is getting deeper and deeper, it is more and more difficult to transfer information between networks, which affects the detection accuracy of the network. In order to use multi-layer features, this work adds residual modules and layer jump connections to the basic structure of YOLO, and builds a multi-scale YOLO-SKIP network. The modified network structure can use the features extracted by the multi-layer network, and at the same time, it is easier to train due to the layer jump connection and its detection speed is faster than the original YOLO algorithm, shown in figure 1.

Table 1 shows the specific parameters of the network, where the number represents the repetitions of the layer, “channel” represents the number of output channels of each layer of convolution, “size” is the size and step length of the convolution kernel, and “output” represents the height and width of the output feature map of this layer. The original YOLO v3 contains 53 layers, and the multi-scale improved algorithm proposed in this paper contains a total of 26 layers, in which the residual represents the residual connection [10].

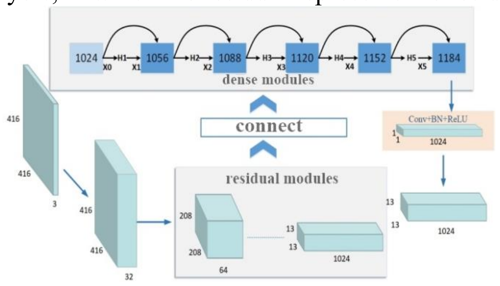


Fig. 1. Schematic diagram of multi-scale YOLO-SKIP network structure.

Table 1. Network structure.

	layer	**channel**	**size**	**output**		**layer**	**channel**	**size**	**output**
	convolutional	32	3x3	256x256		residual			32x32
	convolutional	64	3x3/2	128x128		convolutional	512	3x3/2	16x16
11x	convolutional	32	1x1		**22x**	convolutional	256	1x1	
	convolutional	64	3x3			convolutional	512	3x3	
	residual			128x128		residual			16x16
	convolutional	128	3x3/2	64x64		convolutional	1024	3x3/2	8x8
22x	convolutional	64	1x1		**22x**	convolutional	512	1x1	
	convolutional	128	3x3			convolutional	1024	3x3	
	residual			64x64		residual			8x8
	convolutional	256	3x3/2	32x32		pooling		global	
22x	convolutional	128	1x1			fully-connected		2	
	convolutional	256	3x3						

2.3 Boundary prediction

This work uses the original YOLO v3 bounding box prediction scheme, b_x and b_y represent the center coordinates of the prediction, b_w and b_h represent the width and height of the prediction box. The calculation method is as follows:

$$b_x = \sigma(t_x) + c_x, \; b_y = \sigma(t_y) + c_y, \quad b_w = p_w e^{t_w}, \; b_h = p_h e^{t_h}, \tag{3}$$

where, c_x and c_y represent the coordinates of the upper left corner of the image grid, t_x and t_y represent the predicted position of the network output, t_w and t_h are the width and height of the bounding box predicted by the network, σ represents the Sigmoid activation function, p_w and p_h are the height and width of the a priori box. The loss function is the standard loss provided by YOLO v3, see reference [9] for details.

2.4 Logistic regression

YOLO v3 uses logistic regression to classify the predicted output of the network. The calculation method is as $S(x)=1/(1+e^{-x})$, where x represents the input value, $S(x)$ represents the output value. After the process, the value range is transformed to between 0 and 1, which can represent the predicted probability. If S(x) is greater than 0.5, it means it belongs to a certain category, and if it is less than 0.5, it does not belong to a certain category.

2.5 Unbalanced sample processing

For the bolt images of the angle steel tower of the overhead line under normal conditions, the number of positive samples is much higher than that of the negative samples, that is, there are more normal bolts and fewer faulty bolts. At this time, the model is easy to judge the input data as a positive sample, so as to obtain a higher accuracy rate, but the recall rate of the model is lower. To correctly handle the large difference in the number of positive and negative samples, this work adopts an unbalanced sample processing strategy: in the process of data sampling, ensure that the probability of negative samples being sampled is always higher than the probability of positive samples being sampled. Assuming that the ratio of the number of positive samples and negative samples is m:n, the corresponding sampling probability ratio is n:m. By increasing the sampling probability of negative samples, the model can better perform distributed learning on a small number of samples.

3 Experiment results

Network training adopts stochastic gradient descent (SGD) optimization algorithm, the initial learning rate is 0.001, the learning rate attenuation coefficient is 0.9 (the learning rate is updated every 2000 iterations), the total number of iterations is 70,000, and the batch size is 8. The experimental platform is Win 10, using Intel i7-8700 processor and Nvidia GeForce GTX 1080Ti graphics card.

3.1 Experimental data set and parameter settings

The data set in this paper is a high-quality data set with 1,432 original image data obtained by aerial and field shooting of overhead lines. After removing the problematic images such as shooting angle and shooting distance, a high-quality data set of 1020 images in total is obtained. In the experiment, the original bolt image data set was randomly divided into training data and test data, and the division process was guaranteed to be completely random. The train set contains a total of 820 bolt images with 596 positive samples and 204 negative ones, and the test set contains a total of 200 bolt images with 143 positive samples and 57 negative ones. Some example samples are shown in Figure 2.

Fig. 2 Sample picture of angle steel tower bolt dataset.

3.2 Evaluation index and comparison algorithm

In order to verify the effectiveness of the proposed algorithm, the measures of precision and recall are used to quantitatively analyze the performance of the model. In addition, the mean average precision (mAP) is used to measure the accuracy of the prediction frame, and the calculation method is shown in the literature [11]. In order to verify the effectiveness of the method proposed in this paper, it was compared with other four target detection models, namely SSD [7], Faster R-CNN [6], Retina-net [8] and YOLO v3 [9]. In order to ensure the fairness of the experiment, the training data and test data used by each model in the comparison experiment are exactly the same.

3.3 Quantitative results of fault detection

By comparing the experimental results, it can be found that the performance of SSD are all low, indicating that its performance in detecting faulty bolts is poor, shown in Table 2. Faster R-CNN (F-R-CNN) and Retina-net have similar performance in the three indicators, but they are all lower than YOLO v3 and the algorithm proposed in this article. The three index values of the algorithm proposed in this paper are all higher than YOLO v3, and all are the highest, which proves that the multi-scale improvement strategy of this paper effectively improves the performance of the detection task.

Table 2. Experimental results and comparison.

Models	Precision	Recall	mAP
SSD	84.36	85.42	0.87
F-R-CNN	86.79	90.63	0.89
Retina-net	87.21	90.33	0.86
YOLO v3	88.91	90.82	0.89
My model	90.47	93.56	0.91

3.4 Multi-scale strategy evaluation

Since the introduction of multi-scale strategies will increase the calculation time of the algorithm, in order to verify the real-time performance of the algorithm, frames per second (fps) are used to comprehensively evaluate the multi-scale strategy of the proposed model. For input pictures of the same size, the higher the fps value, the more the number of images processed per second, that is, the better the real-time performance of the algorithm. The following table shows the Precision, Recall, and mAP values obtained by different scale strategies, and the corresponding processing speed index fps value.

From Table 3, it can be seen that the accuracy evaluation index is the best when using a total of four scales of 1+3+5+7, which are all higher than the case of the three scales. When using two scales, the model approximately degenerates to the network structure of YOLO v3, the corresponding evaluation index value is also very close to YOLO v3. Moreover, in terms of speed, the real-time performance of the three scales of 1+3+5 is the best, capable of processing 8.6 images per second, but the values of Precision and mAP are significantly lower than those of the four scales. Therefore, the actual task should be weighed from the two aspects of accuracy and speed, and the appropriate scale should be selected.

Table 3. Experimental results for multi-scale strategy and sampling strategy.

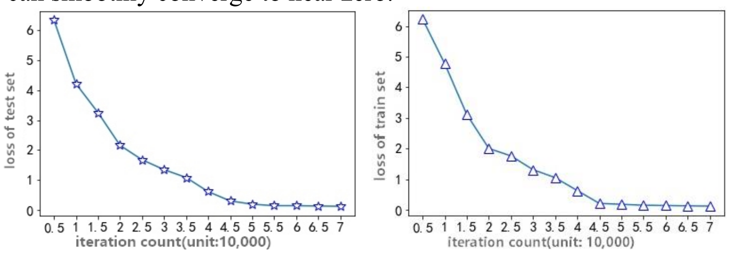
Multi-scale strategy				Sampling strategy				
Scale	**Precision**	**Recall**	**mAP**	**Models**	**Precision**	**Recall**	**mAP**	**Models**
1+3+5	88.36	91.73	0.90	SSD	+1.04	+1.33	+0.87	SSD
1+3+7	88.44	92.74	0.909	F-R-CNN	+0.96	+1.62	+0.85	F-R-CNN
1+5+7	89.14	92.37	0.91	Retina-net	+1.17	+1.25	+0.89	Retina-net
3+5+7	89.76	92.89	0.905	YOLOv3	+1.32	+1.73	+0.92	YOLOv3
1+3+5+7	90.47	93.56	0.913	My model	+1.29	+1.41	+0.97	My model

3.5 Sampling strategy evaluation

For the actual situation where the number of positive and negative samples is unbalanced, this paper adopts a model training strategy that increases the sampling probability of negative samples. In order to seriously the effectiveness of the sampling strategy, the experiment measured the changes in the quantitative evaluation indicators of all algorithms when the sampling strategy is used and when the sampling strategy is not used. Table 3 shows the quantitative experimental results of the unused sampling strategy, and the experimental results of the network after using the sampling strategy, where "-" means decrease and "+" means increase. It can be seen from the results in Table 4 that the processing strategy proposed in the article when the positive and negative samples are imbalanced has the effect of improving network performance.

3.6 Convergence of loss function

The convergence of the loss function is an important evaluation index for the target detection task, which can directly reflect the stability of the network, shown in Figure 3. Due to the complexity of shooting conditions for bolt data of angle steel towers, different shooting angles and weather conditions will cause differences between samples, which may cause certain oscillations in the loss function. For this reason, the target detection network should be sufficiently stable. Ensure its easy training characteristics. The following figure shows the convergence of the proposed algorithm in the training set and the test set. From the convergence trend of the loss value, it can be seen that the stability of the network is good, and it can smoothly converge to near zero.



(a) On testing set (b) On training set

Fig.3. Convergence of loss function.

4 Conclusions

This paper proposes an improved YOLO-SKIP model, based on the deep learning YOLO v3 framework, aiming at the bolt defect detection problem of the angle steel tower in the overhead line, combining the layer jump connection and the multi-scale model. In order to detect the proposed model, this paper uses aerial photography and other methods to collect angle steel bolt image data sets, and mark them as defective and non-defective. The data is expanded by rotation, symmetry and Gaussian noise. The multi-scale fusion convolution module is used to replace the original YOLO v3 convolution module. At the same time, different sampling strategies are used to deal with the problem of sample imbalance, and the recognition accuracy is further improved. The experimental results show that the proposed algorithm can effectively detect the defect data of the angle steel tower bolts, and the average precision of the average value reaches 0.91. At the same time, the detection speed is improved, and the combination of high precision and high speed is realized.

This work was supported by the State Grid Anhui Electric Power Co., Ltd. (No. 52120019007G).

References

1. ZHAI Yongjie, WANG Di, ZHANG Muliu, et al., Fault detection of insulator based on saliency and adaptive morphology. Multimedia Tools and Applications, 2017. 76(9): p. 12051-12064.
2. Han Z, Qiao Y, Sun Y, et al. Research on image recognition based insulator fault detection method for UVA transmission line. Modern Electronics Technique, 2017.
3. YANG Xiao-hui; SHENG Fei; XUE Peng; GU Feng-jie; MI Xin, Defect detection for grid insulator using aerial image based on deep learning, Information Technology (China), 2020(4).
4. XIAN Tao, DAPENG Zhang, ZIHAO Wang, et al., Detection of Power Line Insulator Defects using Aerial Images Analyzed with Convolutional Neural Networks. IEEE Transactions on Systems, Man, and Cybernetics, 2020. 50(4): p. 1486-1498.
5. Huang Ling; Zhao Kai; Li Jidong; Feng Hao; Wang Yanqing; Ma Bihuan, Insulator image detection based on feature pyramid and multi-task learning, Electrical Measurement & Instrumentation (China), 2021, 4: 37-43.
6. SHAOQING Ren, KAIMING He, ROSS Girshick, et al., Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE transactions on pattern analysis and machine intelligence, 2017. 39(6).
7. LIU Wei, ANGUELOV Dragomir, ERHAN Dumitru, et al. SSD: Single Shot MultiBox Detector. in european conference on computer vision. 2016.
8. LIN Tsungyi, GOYAL Priya, GIRSHICK Ross, et al. Focal Loss for Dense Object Detection. in international conference on computer vision. 2017.
9. REDMON Joseph, FARHADI Ali, YOLOv3: An Incremental Improvement. arXiv: Computer Vision and Pattern Recognition, 2018.
10. XIAOHONG Zhang, LIQING Jiang, DONGXU Yang, et al., Urine Sediment Recognition Method Based on Multi-View Deep Residual Learning in Microscopic Image. Journal of medical systems, 2019. 43(11).
11. Zhang H, Li J, Zhang B. Foreign Object Detection on Insulators Based on Improved YOLO v3. Electr. Power, 2020, 53: 49-55.