Transmission line image defect diagnosis based on YOLOv3 and SNIPER cascade lifting algorithm

Guangxiu Liu 1,2*, Zhenyu Li 1,2, Xiaobin Sun 3, Yuan Tian 4, Min Li 4

1 State Grid Shandong Electric Power Company Electric Power Research Institute, Jinan, Shandong, 250003, China
2 Shandong Luneng Intelligence Technology Co., Ltd. Jinan, Shandong, 250101, China
3 State Grid Shandong Electric Power Company, Jinan, Shandong, 250001, China
4 State Grid Shandong Electric Power Company Overhaul Company, Jinan, Shandong, 250118, China

*Corresponding author’s e-mail: liuguangxiu16@163.com

Abstract. Deep learning classifies information and combines the underlying features to form more abstract high-level features. Its ability of autonomous learning also provides favorable conditions for industrial application. Aiming at the problem of insufficient performance of current small target detection, this paper proposes a deep learning algorithm using One-stage fusion two-stage. The influence of background noise on network training is reduced. The algorithm combines multi-scale training methods and achieves sample equilibrium adjustment based on sample distribution and stratified sampling. The high efficiency of YOLOv3 method for large target detection is well used, and the multi-scale training advantages of SNIPER algorithm are also played out, which makes the intelligent detection effect of pin defect of transmission line get some performance improvement.

1. Introduction
With the increasing demand for power transmission and the expanding scale of power system in China, the security and stability of power system has become increasingly prominent. The progress of science and technology has greatly affected the development of the electric power industry, in which artificial intelligence technology has been applied in various fields of the electric power industry. There have been many attempts to realize defect diagnosis faster and better for UAV patrol pictures.

Traditional methods such as image segmentation[1], template matching[2] and edge detection[3] were used in the early detection of transmission line defects. However, because the image background complexity, angle randomness, noise diversity and so on make the algorithm research more difficult, its research results have not been able to reach the industrial level of application level.

The recent development of deep learning has brought new vitality to object detection. The more popular deep learning algorithms can be divided into two categories. One type is a two-stage method. Based on the region proposal R-CNN algorithm, you need to use the selective search[4] or CNN network to generate the region proposal, and then perform classification and regression on the region proposal. The other is one-stage algorithm such as YOLO[5], SSD[6], which uses only one CNN network to directly predict the categories and locations of different targets. The first type of method is
more accurate, but slower. The second type of algorithm is faster, but the accuracy is lower. In order to improve the accuracy of the two kinds of algorithms, it is one of the strategies to use image pyramids to improve the detection accuracy of objects. Typical multi-scale lifting algorithms such as SSD[6], FPN[7], SNIPER[8], YOLOv3[9].

The transmission line image studied in this paper is the transmission equipment image collected by UAV, and the pin level defect of high risk is forecasted. Faced with the position diversity of pins and the small proportion of images, the latest algorithm is improved and applied in this paper.

2. YOLO detection algorithm and SNIPER detection algorithm

2.1. YOLO detection algorithm

YOLOv3[9] introduced the idea of anchor boxes used in Faster R-CNN[10]. Anchor boxes are a set of initial and fixed candidate boxes. The selection of initial anchor boxes will directly affect the detection accuracy and speed of the network. YOLOv3 uses the K-means[11] clustering algorithm to cluster the width and height of the target frame in the COCO dataset.

For class probability prediction, in YOLOv2[12], class probability prediction is directly performed using softmax. In YOLOv3, for each class, an independent logistic classifier is used, and the class probability loss is calculated using the cross entropy loss. The reason for using this strategy is that, considering the use of softmax classification, all categories need to be separated, but once there is an inclusion relationship between classes, using softmax is not so appropriate. Therefore, the logistic classifier is combined with multiple tags to deal with such problems.

The border prediction uses the logistic method, and the formula is as follows:

\[
b_x = \sigma(t_x) + c_x \tag{1}
\]

\[
b_y = \sigma(t_y) + c_y \tag{2}
\]

\[
b_w = p_w e^{t_w} \tag{3}
\]

\[
b_h = p_h e^{t_h} \tag{4}
\]

Where \(c_x\) and \(c_y\) are the coordinate offsets of the grid, \(p_w\) and \(p_h\) are the preset length of anchor box. The final coordinate values are \(b_{x,y,w,h}\), and the network learning target is \(t_{x,y,w,h}\), in which the transformation method is sigmoid activation function.

YOLOv3 adopts the method of upsampling and scale fusion, and independently performs detection on the fusion feature maps of multiple scales, and finally the detection effect on small targets is improved significantly.

YOLOv3 uses full convolution on the one hand and residual structure on the other hand[13]. The residual structure can well control the propagation of the gradient, avoiding situations such as gradient disappearance or explosion, which are not conducive to training. This makes the difficulty of training the deep network greatly reduced, so the network can be made to 53 layers, and the accuracy is more obvious. Figure 1 is a schematic diagram of the structure of Darknet-53.
2.2. SNIPER detection algorithm

The most important contribution of the SNIPER[8] strategy is the proposed fixed size chips, which are a subgraph of a scale image in the image gold tower. Chips have positive and negative points, where the positive chip contains many labels of suitable dimensions, while the negative chip does not contain label samples or images that are not suitable for this scale, such as large objects in low-resolution images and low Small size objects in the resolution image. So SNIPER is just a sampling strategy and an alternative to image pyramids. In the article, the author applies SNIPER to Faster R-CNN[10] and Mask R-CNN[14].

According to the idea of the image pyramid, an original input picture will be constructed by constructing image pyramids in the form of multiple scales \{s_1,s_2,\ldots, s_n\} . SNIPER generates n different scale chips on each image, which are \{s_1, s_2,\ldots, s_n\} , and the chips generated for each size are recorded as \(C_i\). For each scale, first resize the image to the corresponding size, use the window, and get all the chips of the current size through the sliding window of step d. The method is as shown in figure 2. Then the total number of chips of the entire image gold tower is calculated as (5).

\[
\sum \frac{W_i}{d} \times \frac{H}{d}
\]

Figure 2. Schematic diagram of Chips generation

For each size of chips, a valid area range R is selected, and ground-truth bounding box within the area limit is recorded as valid for the current size. The chips we select should contain all the objects that meet the area requirement. In a Scale, select the chip that contains the most valid GT box from all
chips as Positive Chip. For the remaining part, the Negative Chip is constructed by choosing more than M recommended chips in the R range.

The proposals generated by the RPN network and the true value box IoU are greater than 0.5, which is marked as a positive example. During training, ground-truth bounding box is not restricted by R, but proposal must fall within a chip in the R range. Throughout the training process, a batch of training samples is randomly selected to train a certain number of chips.

3. Cascade detection algorithm based on YOLOv3 and SNIPER

3.1. Cascading design of predictive architecture

This paper predicts the initial positioning of the target in the sample area using YOLOv3, namely the primary target area detection. Then, the SNIPER-based local area target detection, that is, the second-level target detection. Figure 3 is a cascading design block diagram of the prediction architecture.

![Cascading design block diagram of predictive architecture](image)

There is a problem of uneven distribution of samples for training samples. The paper uses stratified sampling to carry out random expansion of a few samples, and achieves the equilibrium data distribution of the sample. Among them, random expansion methods include image defogging, contrast enhancement, brightness enhancement, chroma enhancement, random clipping and so on.

The YOLOv3 network uses the K-means algorithm to cluster the target frame size of the data set. In the K-means algorithm, Euclidean distance, Manhattan distance, Chebyshev distance, etc. are usually used as distance metrics to calculate the distance between two points. Since the main purpose of setting the a priori box is to make the prediction box and ground truth IOU better. This results in the use of these common distances that do not produce good results on the sample. This article hopes to get a good IOU through the anchor boxes, and the IOU is independent of the size of the box. Equation (6)[15] is the new distance formula used in this paper.

\[ d(\text{box}, \text{centroid}) = 1 - \text{IOU}(\text{box}, \text{centroid}) \]  

Centroid in (6) represents the center of the cluster. Box represents the sample. IOU represents the intersection of the cluster center box and the cluster box. The comparison ratio of IOU indicates the accuracy of the prediction box. The calculation formula is as follows.

\[ \text{IOU}(b_g, b_p) = \frac{b_g \cap b_p}{b_g \cup b_p} \]  

Where \( b_g \) represents the real box and \( b_p \) represents the prediction box.

The original SNIPER algorithm used three scales to predict the graph. The targets predicted by each scale will use the NMS method to remove the overlapping frame operations, and combine the detection targets at the three scales to perform the algorithm detection results. This approach leads to multiple detections of the same target at multiple scales, with redundant frames. Therefore, this paper uses the non-maximum value suppression to filter the above total results. Thereby reducing the redundant frame and achieving the output of the final target frame.

4. Experiments

In this paper, the data collected by UAV patrol inspection are used for experimental work. The detection target of transmission line is small pin defect. Specifically including pin missing, pin installation is not standardized, pin out and so on.
Hardware lab environment: Intel(R) Xeon(R) CPU E5-2683 v3 @ 2.00GHz CPU, NVIDIA GTX 1080TI.
Software experiment environment: Linux Ubuntu 16.04, python 2.7, cuda8.0, gcc 5.4.0.
The YOLOv3 algorithm is used to achieve the primary detection effect diagram as shown in figure 4. The blue frame is the primary detection area.

![Figure 4. Primary test result chart](image)

The SNIPER prediction effect directly using the secondary detection area is as follows.

![Figure 5. Prediction effect chart of secondary detection area using SNIPER directly](image)

In the last layer of SNIPER algorithm, the prediction effect is as follows after adding the maximum suppression algorithm.

![Figure 6. Secondary detection effect after adding maximum suppression](image)

Training experiments based on cascade detection algorithm, YOLOv3 algorithm and SNIPER algorithm are performed for the same data set. The three algorithms are called A algorithm, B algorithm, and C algorithm. Adjust the training parameters in each algorithm, and find the best model.
of the three methods for defect diagnosis and prediction. The three kinds of logical reasoning mechanisms to predict the AP value, false negative rate and false positive rate of defects are shown in Table 1.

| Algorithm   | AP   | Missing rate | False alarm rate | Speed     |
|-------------|------|--------------|------------------|-----------|
| Algorithm A | 46.8%| 60%          | 73.3%            | 7.18s/ frame |
| Algorithm B | 10.1%| 95%          | 94.1%            | 0.187s/ frame |
| Algorithm C | 23.7%| 74%          | 89%              | 7s/ frame |

5. Conclusion
This paper uses the cascading mechanism of YOLOv3 algorithm and SNIPER algorithm. And use the new distance formula in the clustering link in the YOLO algorithm. The maximal suppression is added to the SNIPER three-scale prediction joint result, which significantly reduces the redundant prediction frame. It can be clearly seen from the above experimental results that the cascade detection algorithm based on YOLOv3 and SNIPER algorithm has better detection effect than the simple YOLOv3 and the simple SNIPER algorithm. The use of YOLOv3 for preliminary pinning of significant regional predictions has greatly reduced the false check frame. And the primary detection area eliminates the large background impact. The SNIPER algorithm training can be more targeted and the model can converge faster. Although the method of this paper has improved the detection effect of the pin defect of the transmission line, there is still room for improvement in speed. The next step will be to optimize the work for SNIPER.

Acknowledgments
This work was financially supported by the technology project of State Grid Shandong Electric Power Company (Grand No.2018A-001). The specific project name is “Research on Image Intelligent Analysis Technology of Power Transmission and Transformation Equipment Based on Deep Learning”. Thanks very much.

References
[1] Huang, X. N., Zhang, Z. L. (2010) A method to extract insulator image from aerial image of helicopter patrol. Power System Technology, 34(1): 194-197.
[2] Yao G., Liu Y., Dong F. M., et al. (2010) An Obstacle Detection Approach of Transmission Lines Based on Contour View Synthesis. In: IEEE International Conference on Automation and Logistics. Hong Kong.
[3] Hui, Ali. (2011) Infrared image edge detection of high voltage insulator chain. Applied Mechanics and Materials, 4: 58-60.
[4] Uijlings, J. R., Sande, K. E., Gevers, T., et al. (2013) Selective search for object recognition. International Journal of Computer Vision, 104(2): 154-171.
[5] Redmon, J., Divvala, S., Girshick, R., et al. (2016). You only look once: unified, real-time object detection. In: IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas. pp: 779-788.
[6] Liu W., Anguelov D., Erhan D., et al. (2016) SSD: Single shot multiBox detector. In: European Conference on Computer Vision. Amsterdam. pp: 21-37.
[7] Lin, T. Y., Dollar, P., Girshick, R., et al. (2017). Feature Pyramid Networks for Object Detection. In: IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI.
[8] Singh, B., Najibi, M., & Davis, L. S. (2018). Sniper: efficient multi-scale training. In: Neural Information Processing Systems. Montréal, Canada.
[9] Redmon, J., and Farhadi, A. (2018) YOLOv3: an incremental improvement. arXiv:1804.02767.
[10] Ren, S. Q., He, K. M., Girshick, R., et al. (2017) Faster r-cnn: Towards real-time object detection with region proposal networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(6): 1137-1149.

[11] Hartigan J. A, Wong M. A. (1979) Algorithm AS 136: a k-means clustering algorithm. Journal of the Royal Statistical Society, 28(1):100-108.

[12] Redmon J., Farhadi A. (2017) YOLO9000: Better, Faster, Stronger. In: IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI. pp: 6517-6525.

[13] He, K. M., Zhang, X. Y., Ren, S. Q., et al. (2016) Deep residual learning for image recognition. In: IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas. pp: 770-778.

[14] He, K., Gkioxari, G., Dollar P., et al. (2017) Mask R-CNN. In: IEEE International Conference on Computer Vision. Venice. arXiv:1703.06870.

[15] Zheng Z. Q., Liu Y. Y., Pan C. C., et al. (2019) Application of improved YOLO V3 in aircraft recognition of remote sensing images. Electronics Optics and Control. 26(4): 28-32.