IEICE TRANS. INF. & SYST., VOL.E106–D, NO.5 MAY 2023

LETTER *Special Section on Deep Learning Technologies: Architecture, Optimization, Techniques, and Applications*

OPENnet: Object Position Embedding Network for Locating Anti-Bird Thorn of High-Speed Railway

Zhuo WANG[†a)], ***Student Member***, **Junbo LIU**[††b)], **Fan WANG**[††c)], ***and*** **Jun WU**[†d)], ***Nonmembers***

SUMMARY Machine vision-based automatic anti-bird thorn failure inspection, instead of manual identification, remains a great challenge. In this paper, we proposed a novel **O**bject **P**osition **E**mbedding **N**etwork (OPENnet), which can improve the precision of anti-bird thorn localization. OPENnet can simultaneously predict the location boxes of the support device and anti-bird thorn by using the proposed double-head network. And then, OPENnet is optimized using the proposed symbiotic loss function (SymLoss), which embeds the object position into the network. The comprehensive experiments are conducted on the real railway video dataset. OPENnet yields competitive performance on anti-bird thorn localization. Specifically, the localization performance gains +3.65 AP, +2.10 AP50, and +1.22 AP75.

key words: *anti-bird thorn localization, double-head network, symbiotic loss, object position embedding network*

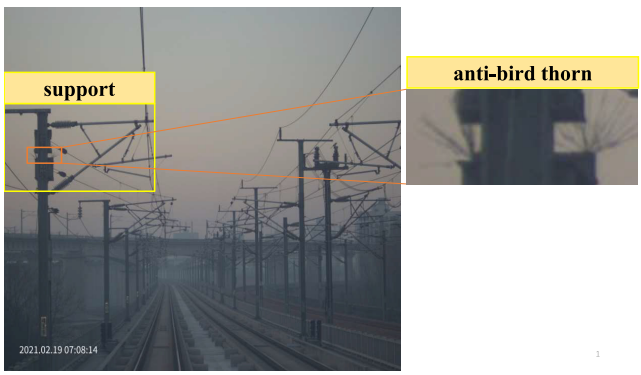


Fig. 1 An example image of high-speed railway.

1. Introduction

Bird activity is a common cause of traction power supply equipment failure of high-speed railway, which seriously menaces the safety of train [1]. In order to reduce the failure rate of traction power supply equipment, we usually install anti-bird thorns on the support device of the catenary tower to dislodge birds. However, severe weather, birds fighting, and nonstandard installation will lead to anti-bird thorn failure, so as lose the dislodge birds function [2]. Therefore, it is necessary to periodically inspect whether the anti-bird thorn is fault.

At present, the anti-bird thorn failure inspection methods based on vehicle video are composed of two parts: anti-bird thorn localization and anti-bird thorn failure identification. In this paper, we mainly focus on some challenges of anti-bird thorn localization task and propose a novel **O**bject **P**osition **E**mbedding **N**etwork (OPENnet), which can fleetly and precisely locate anti-bird thorn.

Figure 1 shows an example image of a high-speed railway collected by the comprehensive inspection train. The visual field of high-speed railway scene image is very open, which has the features of many foreground objects and variable backgrounds. Unfortunately, the anti-bird thorn that needs to be located is very small. Therefore, the traditional visual location methods are difficult to deal with such a complex situation.

Recently, many deep learning-based object detection methods have achieved promising results both in academic and industry. These methods can be divided into one-stage methods and two-stage methods according to different generative mechanisms of object proposal. YOLOv3 [3], SSD [4] and RetinaNet [5] are popular one-stage methods, which greatly improve the detection speed by using the bounding boxes regression method to detect objects. However, these methods will miss the detection for small objects. Two-stage methods have high accuracy of object detection, such as R-CNN [6], Faster R-CNN [7] and Cascade R-CNN [8]. In addition, some scholars have proposed methods to detect coupled objects including VTransE [9] and YOLO+MP [10]. Such two models are devised to predict the relations between coupled objects for deeper interpretation on pixels. Differently, in this work, the support detection (auxiliary task) is used to facilitate the anti-bird thorn detection (main task) based on the dependency between such two coupled objects.

Due to the limited computational resources of the device deployed in high-speed trains and the stringent response requirement of railway inspection, inference time is critical for the task of anti-bird thorn detection. To detect small objects (e.g. anti-bird thorns) at a fast inference time, we first present a double-head network, which can reduce the searching scope of anti-bird thorns. And then, we design a symbiotic loss function to embed the object position

Manuscript received April 1, 2022.
Manuscript revised September 9, 2022.
Manuscript publicized November 14, 2022.
[†]The authors are with School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, China.
[††]The authors are with Infrastructure Inspection Research Institute, China Academy of Railway Sciences Corporation Limited, Beijing 100044, China.
a) E-mail: bjtuwz@bjtu.edu.cn
b) E-mail: liujunbo@rails.cn
c) E-mail: wangfan@rails.cn
d) E-mail: wuj@bjtu.edu.cn (Corresponding author)
DOI: 10.1587/transinf.2022DLL0011

Copyright © 2023 The Institute of Electronics, Information and Communication Engineers

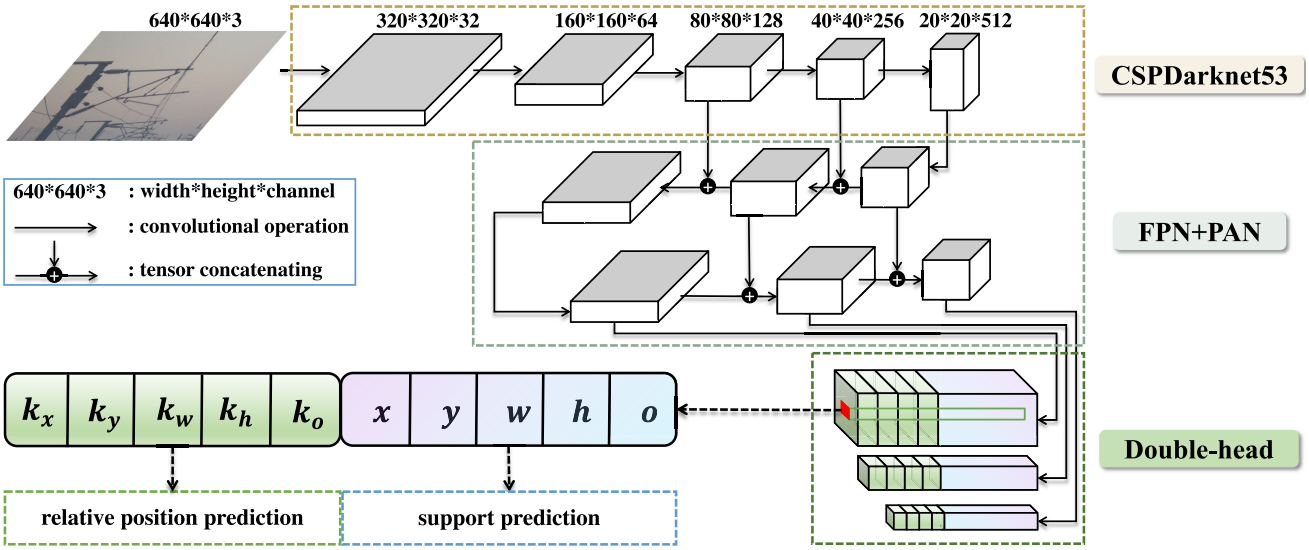


Fig. 2 Overall framework of the proposed OPENnet.

into the network using the symbiotic relationship between the anti-bird thorn and the support device. Comparative experimental results show that our method achieves optimal performance and significantly improves the precision of the anti-bird thorn location box.

2. Object Position Embedding Network

Based on the symbiotic relationship between the support device and the anti-bird thorns, we propose an innovative object position embedding network, which can simultaneously locate the support and the anti-bird thorn shown in Fig. 1. The input of the method is a high-speed railway image, and the output is the coordinates of the support device and the anti-bird thorn. The model is optimized by our proposed symbiotic loss.

2.1 Double-Head Network

To avoid runaway gradients, instabilities, and infinite losses, as same as YOLOv5 [11], our network also predicts offsets instead of coordinates to make it easier to learn. The network predicts four offsets x, y, w, h for the support and a confidence o. The predictions shown in Fig. 3 correspond to:

$$s_x = 2\sigma(x) + c_x - 0.5 \tag{1}$$
$$s_y = 2\sigma(y) + c_y - 0.5 \tag{2}$$
$$s_w = 4p_w\sigma^2(w) \tag{3}$$
$$s_h = 4p_h\sigma^2(h) \tag{4}$$
$$s_o = \sigma(o) = CIoU(T^g, T^p) \tag{5}$$

where p_w, p_h are the bounding box prior's width and height. s_x, s_y, s_w, s_h and s_o are center point coordinates, height, width and confidence of the bounding box of support. The cell is offset from the top left corner of the image by (c_x, c_y).

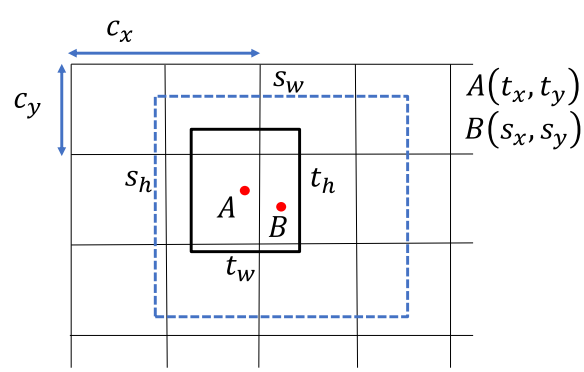


Fig. 3 The relative position relationship between the anti-bird thorn bounding box and the support bounding box. A and B are the center points of support and anti-bird thorn prediction boxes, respectively.

T^g and T^p are ground truth and predict the bounding box of anti-bird thorn. Four main parameters k_x, k_y, k_w, k_h are used to learn the relative position between anti-bird thorn and support shown in Fig. 2. Meanwhile, we also design a function between o and t_o, which is the confidence of the anti-bird thorn predicted bounding box. The position of the anti-bird thorn bounding box is defined as follows:

$$t_x = s_x + s_w(\sigma(k_x) - 0.5) \tag{6}$$
$$t_y = s_y + s_h\left(\sigma\left(k_y\right) - 0.5\right) \tag{7}$$
$$t_w = 2s_w\sigma(k_w) * \min(1 - \sigma(k_x), \sigma(k_x)) \tag{8}$$
$$t_h = 2s_h\sigma(k_h) * \min\left(1 - \sigma\left(k_y\right), \sigma\left(k_y\right)\right) \tag{9}$$
$$\sigma(k_o) = RPIoU(S^p, T^p) \tag{10}$$
$$t_o = \sigma(k_o) * s_o \tag{11}$$

where t_x, t_y, t_w, t_h and t_o are center point coordinates, height, width and confidence of the anti-bird thorn bounding box. $\sigma(\cdot)$ function was used to constrain (t_x, t_y) appear inside the box of support. S^p is predicted support bounding box.

RPIoU is shown in Alg. 1.

2.2 Symbiotic Loss

The loss of object detection is commonly an IoU-based loss function, including IoU loss [12], GIoU loss [13], DIoU loss and CIoU loss [14]. But they are used between ground truth and predict bounding boxes for the same class. They do not consider the symbiotic relationship between different classes. For this reason, based on our dataset, we proposed a symbiotic loss function for the relative positional relationship between the support and the anti-bird thorn. It is defined as follows:

$$\mathcal{L}_{sym} = \mathcal{L}_{CIoU} + \alpha(1 - RPIoU) \tag{12}$$

where α is the coefficient of the second part, which is used to represent the second part importance. RPIoU explained in Alg. 1 is a metric of the extent of the anti-bird thorn inside the support.

Algorithm 1 RPIoU

Input:
Predicted support box S and predicted thorn box T

$$S = (x_1^s, y_1^s, x_2^s, y_2^s), T = (x_1^t, y_1^t, x_2^t, y_2^t)$$

Output: RPIoU

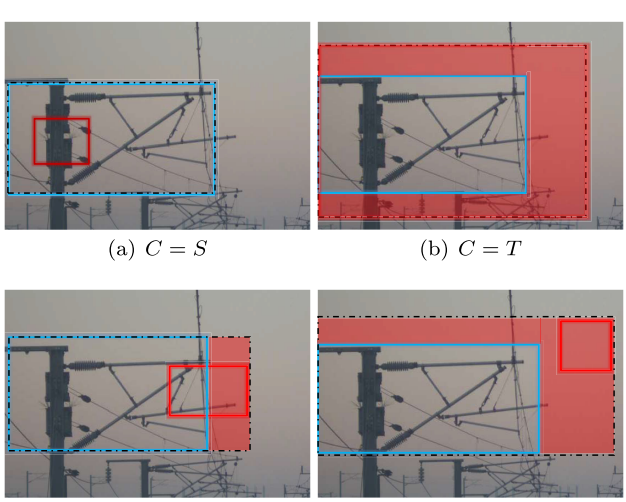
1: For the predicted box S and T, ensuring $x_1^s < x_2^s$, $y_1^s < y_2^s$, $x_1^t < x_2^t$, $y_1^t < y_2^t$.
2: Calculating area of S: $A_s = (x_2^s - x_1^s) \times (y_2^s - y_1^s)$.
3: Calculating area of T: $A_t = (x_2^t - x_1^t) \times (y_2^t - y_1^t)$.
4: Finding the coordinate of smallest enclosing box C;

$$x_1^c = min(x_1^s, x_1^t), x_2^c = max(x_2^s, x_2^t)$$

$$y_1^c = min(y_1^s, y_1^t), y_2^c = max(y_2^s, y_2^t)$$

5: Calculating area of C: $A^c = (x_2^c - x_1^c) \times (y_2^c - y_1^c)$.
6:

$$RPIoU = \frac{A_s}{A^c}$$



(a) $C = S$ (b) $C = T$ (c) $S \cap T \neq \phi$ (d) $S \cap T = \phi$

Fig. 4 There are four different relative positive relationships between the blue predicted bounding box of support S and the red predicted bounding box of anti-bird thorn T. C is the coordinate of the smallest enclosing box of S and T, which is a black dashed box.

To explain RPIoU more intuitively, we list four different kinds of relative position between the ground truth bounding box of support and the predicted bounding box of anti-bird thorn, which are shown in Fig. 4, respectively. The area of the red parts reflects the size of the $1 - RPIoU$. Especially when the predicted anti-bird thorn bounding box is inside the support bounding box, $C = S$, $RPIoU = 1$. In the other cases, $\mathcal{L}_{sym} > \mathcal{L}_{CIoU}$.

3. Experiment and Evaluation

3.1 Dataset and Implement Details

The dataset consists of 2064 training images and 516 testing images at 640 × 640 resolution. The number of anti-bird thorns is 1101, and the number of supports is 3536.

We utilize the stochastic gradient descent with a start learning rate of 0.012 to train the proposed OPENnet. The training set is shuffled, and each mini-batch consists of 16 images. We use the symbiotic loss as the objective function when training our network. For baseline, we keep the original implementation as released by the authors.

To evaluate the performance of our method, We exploit four widely-used metrics in object detection evaluation: Average Precision (AP), Average Precision when intersection over union threshold is 0.5 (AP50), Average Precision when intersection over union threshold is 0.75 (AP75), and Frames Per Second (FPS).

3.2 Results and Discussion

To get the best of hyper-parameter α in Eq. (12), we test the hyper-parameter α in the range of $\{0, 0.1, \cdots, 1.0\}$. The results are shown in Fig. 5. It shows that a value of 0.4 lead to the best AP scores, which is the optimal trade-off between two parts of symbiotic loss in Eq. (12).

Secondly, we compare OPENnet with four object detection methods on our dataset and the corresponding quantitative results are shown in Table 1. The methods are the representative one-stage models, YOLOv3 [3], YOLOv5 [11] and the two-stage models, Faster R-CNN [7],

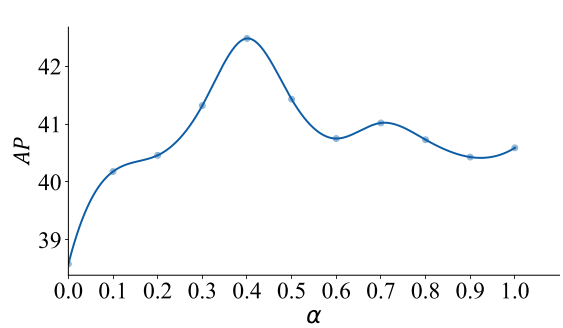


Fig. 5 The performance of YOLOv5 with different alpha

Table 1 Anti-bird thorn testing set results. (OPENnet- represents OPENnet without double-head network, and SymLoss- represents predicting anti-bird thorn directly instead of predicting anti-bird thorn with equations 6-9 indirectly. All experiments were running on Geforce RTX 3090.)

Method	Backbone	Loss function	AP	AP50	AP75	FPS
Faster R-CNN [7]	ResNet-50	Smooth L1 [15]	37.50	80.00	22.30	27.7
Cascade R-CNN [8]	ResNet-50	Smooth L1 [15]	42.30	86.70	30.43	22.5
YOLOv3 [3]	Darknet53	MSE	37.20	85.61	23.21	51.2
YOLOv5 [11]	CSPDarknet53	IoU loss [12]	38.90	85.57	31.38	77.5
YOLOv5 [11]	CSPDarknet53	GIoU loss [13]	39.57	85.69	27.10	76.3
YOLOv5 [11]	CSPDarknet53	DIoU loss [14]	39.67	86.67	24.52	77.5
YOLOv5 [11]	CSPDarknet53	CIoU loss [14]	38.77	86.41	22.80	74.1
OPENnet-	CSPDarknet53	SymLoss-	41.43	87.45	28.36	73.1
OPENnet	CSPDarknet53	SymLoss	**43.22**	**88.80**	**32.60**	71.4

Cascade R-CNN [8]. Specifically, several observations are drawn from the experimental results:

- Overall, Our method (OPENnet) consistently outperforms the other four methods. Such an observation illustrates the importance of symbiosis loss and double-head network for detecting anti-bird thorn.
- Both OPENnet and OPENnet- only drop a few frames per second compare with fastest YOLOv5 in exchange for higher AP, AP50, and AP75.

We further compare OPENnet with its one degenerated variant (OPENnet-) for demonstrating the effectiveness of the double-head network. Both "OPENnet" and "OPENnet-" take the relative position between support and anti-bird thorn into consideration in designing their loss functions. The main difference between them is that "OPENnet" first predicts support and then searches for anti-bird thorn within the bounding box of support, while "OPENnet-" independently predicts both. The ablation experiment shows that, due to reducing the searching scope of anti-bird thorn, "OPENnet" with the double-head network outperforms "OPENnet-", thus proving the effectiveness of the double-head network.

Benefiting from the fact that our model can handle the symbiotic relationship between the anti-bird thorn and the support device, its results are significantly better than those obtained by other comparative models.

4. Conclusion

In this paper, we proposed a novel **O**bject **P**osition **E**mbedding **N**etwork (OPENnet), which includes double-head network and a symbiotic loss function. OPENnet can simultaneously predict the location boxes of the support device and anti-bird thorn and the relative position between them by using proposed the double-head network. The experimental results show that our proposed method improves AP, AP50, and AP75, up to a 9% improvement. In the future, we plan to enhance the image quality to improve the discrimination of anti-bird thorns based on our model.

Acknowledgments

This work was supported in part by the Open Research Fund from the Guangdong Provincial Key Laboratory of Big Data Computing, The Chinese University of Hong Kong, Shenzhen, under Grant B10120210117-OF01 and the key research and development project of the China Academy of Railway Sciences Corporation Limited (2021YJ310).

References

[1] J. Qiu, "Fault analysis of bird transmission in transmission lines," Modern Industrial Economy and Informationization, 2018.

[2] S.H. Wang and Z.Q. Ye, "Analysis of bird damage accidents on overhead transmission lines and prevention techniques," High Voltage Apparatus, vol.47, no.2, pp.61–67, 2011.

[3] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," arXiv preprint arXiv:1804.02767, 2018.

[4] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A.C. Berg, "Ssd: Single shot multibox detector," European conference on computer vision, vol.9905, pp.21–37, Springer, 2016.

[5] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," Proc. IEEE international conference on computer vision, pp.2980–2988, 2017.

[6] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," Proc. IEEE conference on computer vision and pattern recognition, pp.580–587, 2014.

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

[8] Z. Cai and N. Vasconcelos, "Cascade r-cnn: Delving into high quality object detection," Proc. IEEE conference on computer vision and pattern recognition, pp.6154–6162, 2018.

[9] H. Zhang, Z. Kyaw, S.-F. Chang, and T.-S. Chua, "Visual translation embedding network for visual relation detection," 2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017, pp.3107–3115, IEEE Computer Society, 2017.

[10] Y. Liu, L.G. Hafemann, M. Jamieson, and M. Javan, "Detecting and matching related objects with one proposal multiple predictions," IEEE Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2021, virtual, June 19-25, 2021, pp.4520–4527, Computer Vision Foundation / IEEE, 2021.

[11] G. Jocher, A. Stoke, A. Chaurasia, J. Borovec, et al., "ultralytics/yolov5: v6.0 - YOLOv5n 'Nano' models, Roboflow integration, TensorFlow export, OpenCV DNN support," Oct. 2021.

[12] J. Yu, Y. Jiang, Z. Wang, Z. Cao, and T. Huang, "Unitbox: An advanced object detection network," Proc. 24th ACM international conference on Multimedia, pp.516–520, 2016.

[13] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid, and S. Savarese, "Generalized intersection over union: A metric and a loss for bounding box regression," Proc. IEEE/CVF conference on computer vision and pattern recognition, pp.658–666, 2019.

[14] Z. Zheng, P. Wang, W. Liu, J. Li, R. Ye, and D. Ren, "Distance-iou loss: Faster and better learning for bounding box regression," CoRR, vol.abs/1911.08287, 2019.

[15] R. Girshick, "Fast r-cnn," Proc. IEEE international conference on computer vision, pp.1440–1448, 2015.