Image and Vision Computing 109 (2021) 104159

Contents lists available at ScienceDirect

Image and Vision Computing

journal homepage: www.elsevier.com/locate/imavis

# An efficient foreign objects detection network for power substation

Liang Xu [a], Yongkang Song [b], Weishan Zhang [b,*], Yunyun An [c], Ye Wang [d], Huansheng Ning [a]

[a] College of Computer and Communication Engineering, Beijing University of Science and Technologies,Beijing, China
[b] College of Computer Science and Technology, China University of Petroleum, Qingdao, China
[c] Huangdao District Power Supply Company of Shandong Electric Power Company Qingdao, State Grid, Qingdao, China
[d] PLA 9144, China

ARTICLE INFO

*Article history:*
Received 13 July 2020
Received in revised form 19 December 2020
Accepted 9 March 2021
Available online 16 March 2021

*Keywords:*
Power substation
Deep learning
Foreign objects detection
FODN4PS

ABSTRACT

A power substation is susceptible to intrusions of foreign objects. The intrusions can likely result in failures of power supplies. Therefore, recognizing foreign objects becomes important to ensure constant and stable power supplies. However, existing object recognition methods fail to achieve acceptable accuracy and performance. In this paper, we propose an efficient Foreign Objects Detection Network for Power Substation (FODN4PS) to improve the recognition accuracy with less time. FODN4PS consists of a Moving Object Region Extraction Network (MORE Net) and a classification network, where the MORE Net can get the position of foreign objects, and the classification network can recognize the category of foreign objects. Experimental results show that FODN4PS is faster and more accurate in object recognition than the Fast R-CNN and Mask R-CNN.
© 2021 Elsevier B.V. All rights reserved.

## 1. Introduction

The power substation is an important hub connecting a power plant and users. The safety and stability of its operating environment are of great importance. Once the equipment in the substation is in trouble, the safety of the power system and the stability of the power supply will be significantly affected [1]. However, substation accidents occur frequently due to foreign objects intrusion [2]. The intrusion of foreign objects in the substation will cause an irreversible impact on power equipment, even cause paralysis of the whole substation. Therefore, it is of considerable significance to accurately find foreign objects in time, and then take corresponding measures to remove them.

At present, the inspection of foreign objects in power substations is usually done by manual inspection. This method mainly depends on the subjective sensory qualitative judgment and analysis of inspectors, and it is very much affected by the working experiences of the inspectors. Some abnormal conditions will inevitably be ignored due to the weakness of subjectivity and fatigue of human eyes. At the same time, a substation is a high-risk place, so it is challenging to inspect in poor weather conditions.

Moving target detection is to separate the foreground and background containing the moving target from the stationary or slowly moving background environment, extract and locate the moving target, and prepare for subsequent target tracking, behavior understanding and recognition, etc. The result of moving target detection directly affects the accuracy of the subsequent tasks. ViBe(Visual Background Extractor) [3] is a pixel-level foreground detection algorithm with high real-time performance, low memory occupancy, and high foreground detection accuracy. But when the background is complex, there may be serious "ghosting" and "flashing points" problems in the processing results using Vibe.

Deep learning, esp. convolutional neural networks had achieved great progress in computer vision [4,5] and began to widely used in the patrol inspection system [6]. Researchers are using R-CNN (Region-CNN) [7] and its variations for intrusion detection, such as power transmission lines, airfield pavement, and usually modified networks are used to get better performance. For example, Shi et al. [8] improved the accuracy of the network by expanding data with GAN (Generative Adversarial Networks) [9]. Liu et al. [10] proposed a new network based on Fast R-CNN [11], which significantly improved detection accuracy and speed. However, the current R-CNN series of target detection algorithms are all based on RPN (Region Proposal Network) [12] to generate candidate target boxes. Compared with the original R-CNN that generates target candidate boxes by sliding window, the selective search method filters out a large number of useless anchors. However, after NMS (Non-Maximum Suppression) [13] processing, RPN brings a series of complex calculations and lots of redundant candidate boxes, which is time-consuming. Therefore, we propose the FODN4PS method to reduce time consumption, and redundant boxes are eliminated, which helps to improve detection accuracy and performance for foreign objects detection.

We conducted a detailed evaluation of the proposed network, including the analysis of different data enhancement methods, the

* Corresponding author.
*E-mail addresses:* zhangws@upc.edu.cn (W. Zhang), ninghuansheng@ustb.edu.cn (H. Ning).

https://doi.org/10.1016/j.imavis.2021.104159
0262-8856/© 2021 Elsevier B.V. All rights reserved.

comparison with common classification networks, Faster R-CNN and Mask R-CNN. The contributions for this paper are as follows.

- We propose FODN4PS, which effectively avoid the problems of large amount of computation and redundant candidate boxes caused by RPN. FODN4PS changes the way of generating candidate target boxes and ensure that each target has only one candidate box for subsequent operations, which further reduce time consumption.
- We design a simple but efficient target detection approach composing of extraction of candidate target box and classification implemented by FODN4PS. Compared with the complex two-stage target detection algorithm, the candidate target box after the process of FODN4PS is more accurate, and after the training of classification network in advance, FODN4PS obtains higher accuracy.
- We use image rotating randomly and GAN to expand the number of data sets for improving training accuracy. Through comparative evaluations, we found that the accuracy is improved using these two methods at the same time.
- We make comprehensive comparative evaluations to prove that FODN4PS is more efficient than Faster RCNN and Mask RCNN.

This paper is organized as follows. Section 2 reviews related work. Section 3 introduces FODN4PS structure design, including MORE Net, classification network, and loss function. Section 4 evaluates FODN4PS. Conclusion and future work are given in section 5.

## 2. Related work

In our work, we use a set of methods to improve the performance and accuracy of foreign object detection, including firstly, data enhancement to improve the possible number of data samples to cater for different scenes in a power substation; and also, our work is related the classical computer vision problem of object detection and background modeling, therefore we will review the related work on data enhancement, object detection and background modeling.

### 2.1. Data enhancement

The quality of training data will affect recognition accuracy. In the absence of training data, researches usually use data enhancement to expand the training data or design a method that can achieve good performance in limited training data. Zhao [14] presented an automatic approach for small organ segmentation with limited training data. Data enhancement refers to processing images by keeping their basic categories unchanged, such as cropping, flipping, rotating, zooming, panning, and adding noise. By artificially increasing the number of samples, data enhancement helps to reduce overfitting and improve generalization. Gupta et al. [15] proposed methods to enhance data sets by pasting real-segmented objects into original images [16]. Bochkovskiy et al. [17] used mosaic technology and added random scaling, cropping, flipping, rotating and erasing to achieve data enhancement in YOLOv4, and proved that the method has a good effect through experiments. Some researchers used DCGAN (deep convolutional generative adversarial networks) to increase the number of samples and thus improve the recognition accuracy [18–20]. Wu et al. [21] proposed to use GANs to improve the recognition accuracy of tomato leaf diseases. Wang et al. [22] used night images as input, generated virtual target scenes similar to the daytime environment through DCGAN and combined Faster RCNN target detection to obtain high-precision detection results.

### 2.2. Object detection

Image morphology [23] is a standard method to detect foreign objects. Ke et al. [24] used SIFT algorithm [25] to match the designated area of the template image and patrol image collected in the substation. Based on matching the image, Rosten et al. [26] used corner detection to count the feature points of the designated area. Moreover, foreign object recognition is carried out by comparing the feature points with the threshold set previously. The general process of traditional image morphology detection is as follows, firstly, it uses Gaussian filter [27] [28], median filter [29], or bilateral filter [30] to eliminate noise and improve the accuracy of recognition, then Otsu (maximum inter-class variance) [31] is used to segment the background and foreground of the image. After these steps, the foreign objects will be recognized. Due to the complex background of substation and the influence of weather, it is difficult to determine a threshold value for all data. It can only artificially set the threshold value for each image, which increases the overhead of the whole detection process. Moreover, the efficiency is extremely low.

Deep learning has made great progress in the field of computer vision. CNN (Convolutional Neural Network) is becoming an important method for object detection. Krizhevsky et al. [32] proposed a deep convolution neural network (DCNN), AlexNet, which improves the existing image classification accuracy in the large-scale visual recognition challenge (ILSVRC). Since then, various deep learning methods are designed for general object detection [33] [11] [34]. R-CNN [35] can be considered as the first to use deep learning algorithms for target detection, using a large number of training data to train the convolutional neural network model. Due to the problem of redundant candidate boxes in R-CNN, the subsequent Fast R-CNN [36], Faster R-CNN [37], and Mask R-CNN [38] are all improvements to resolve this problem. These above methods are usually called CNN based two-stage detectors.

Joseph et al. [39] applied the convolutional neural network to the entire image and propose a one-stage network called YOLO, the same idea as SSD (Single Shot MultiBox Detector) [40]. Since then, Joseph proposed the v2 and V3 versions of YOLO [41] [42], which not only guarantees the performance of one-stage target detection, but also obtains higher accuracy. Huang et al. [43] proposed DenseBox to detect targets directly without relying on candidate boxes. After that, some anchor-free target detection methods have emerged. Zhang et al. [44] pointed out that the essential difference between anchor-based and anchor-free detection is actually how to define positive and negative training samples, which leads to the performance gap between them, and proposed an optimization method called activate training sample section. Law et al. [45] got the final bounding boxes by predicting the upper left and lower right corners in Cornernet. Wang et al. [46] proposed an alternative called “Guided Anchoring (GA)” Region Proposal Network, and proves that GA-RPN is very efficient. However, the foreign objects in the substation only account for a very small part of the whole image, therefore, processing the whole image is time-consuming.

### 2.3. Background modeling

Background modeling technology [47] [48] [49] has been applied to moving object detection. Especially, ViBe [3] is a pixel-level background modeling and foreground detection algorithm. The ViBe algorithm not only simplifies the process of background model building, but also deals with the situation of sudden background change. Mao et al. [50] proposed a visual background extracting algorithm based on multi-scale space to solve the problem that there exist ghosts and dynamic background disturbance in the target detection process for the conventional ViBe algorithms. Liu et al. proposed an algorithm based on the Pearson coefficient, which is an improvement on the traditional ViBe algorithm to solve video jitter occurs due to external factors. Bai et al. proposed a improved ViBe algorithm to speed up ghost removal by locating the ghost area and re-initializing the area. Sun et al. [51] and Xia et al. [52] applied the ViBe algorithm to extract the moving vehicles from the video. However, they did not carry out subsequent processing.

## 3. Design of the FODN4PS approach

Currently the two-stage target detection algorithms are all based on RPN, which generates a lot of redundant anchors in the whole image.

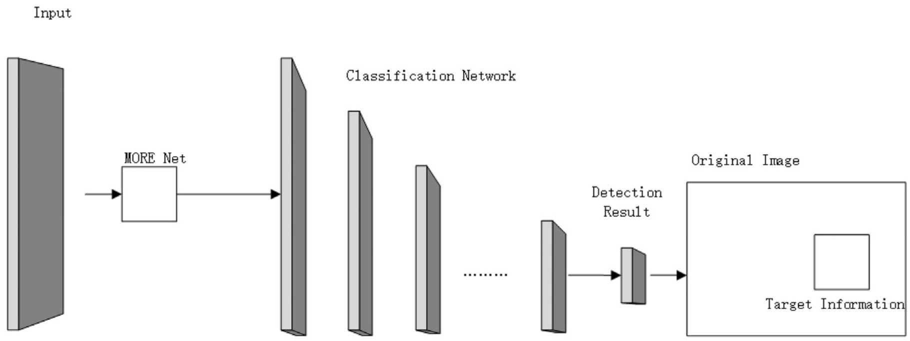



**Fig. 1.** FODN4PS network structure.

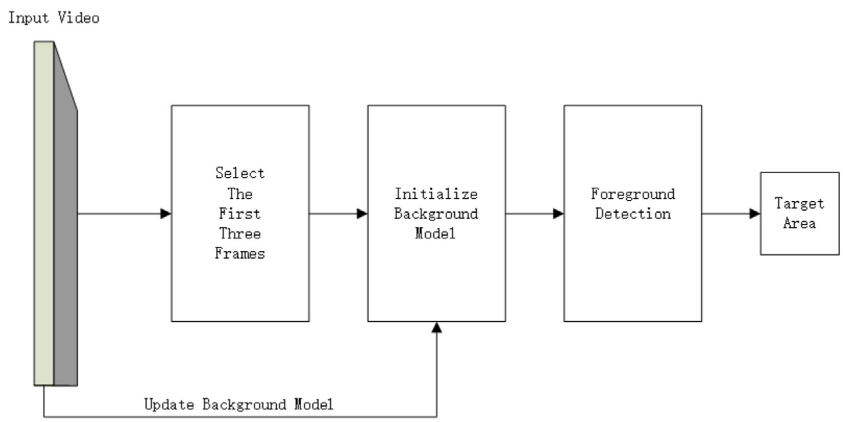



**Fig. 2.** The process of MORE Net.

These algorithms obtain hundreds of regions of interest after the NMS processing, which causes a series of computation and increases overhead of recognition. Compared with the complex two-stage target detection algorithm, we optimize this process by extracting accurate target candidate boxes which do not involve a series of complex calculations. Therefore, improved detection performance by obtaining more accurate candidate target box with higher classification accuracy can be achieved.

Therefore, we design the structure of FODN4PS as shown in Fig. 1, which consists of two parts. The first part is the MORE Net (Moving Object Region Extraction Network) for detecting moving objects, which extracts the candidate region box with foreign objects and only sends one candidate target box of each object to the neural network for feature extraction; The second part is the classification network, whose input is the candidate target box generated by MORE Net. After MORE Net and classification network, FODN4PS can obtain the category and location information of foreign objects. Finally, FODN4PS maps the obtained information to the original image.

### 3.1. MORE net structure

Fig. 2 presents the work flow of the MORE Net. First, it analyzes the input video segment, using the first three frames of the video to construct a background model. When the invasion of a foreign object occurs, the pixel value of a specific range in the background model will change. Therefore, compared with the established background model, MORE Net can obtain the area with foreign objects.

### 3.2. Classification network

Fig. 3 presents the classification network, which is mainly composed of convolution layers and a full connection layer by considering Google Net [53]. The first layer uses $7 \times 7 \times 3$ convolution kernel to filter a $224 \times 224 \times 3$ input image. The step length of the convolution kernel is 2. After Max pooling, the output is $56 \times 56 \times 64$. The second layer convolutes the result of the Max pooling with a $3 \times 3$ kernel. Its output is $28 \times 28 \times 192$. Similarly, the third, forth, fifth layer convolutes the result of the former layer with $1 \times 1$ kernel for dimension reduction and then convolutes with a $3 \times 3$ convolution kernel. After the process of softmax, FODN4PS obtains the object category.

### 3.3. Loss function

Eq. (1) represents the FODN4PS loss function. FODN4PS sets another classifier in the network for auxiliary classification, and adds the loss function of the auxiliary classifier to the total loss function according to the weight λ of 0.3.

$$Loss = L_{cls} + \lambda L_{clsa} \tag{1}$$

$L_{cls}$ is the category of each region predicted by the main classifier, and $L_{clsa}$ is the category of each region predicted by the auxiliary classifier. The category loss function is shown as eq. (2), where $p_i$ is the probability of target prediction, $p_i^*$ indicates whether it is the real target, 1 represents the real target and 0 represents the fake target.

$$L_{clsa} = L_{cls} = \sum_i \left\{ -\log \left[ p_i^* p_i + \left( 1 - pi^* \right) (1 - p_i) \right] \right\} \tag{2}$$

### 3.4. Improved ViBe for candidate box detection

Currently all two-stage target detection algorithms are based on RPN, which generates anchors at the pixel points in the whole image.

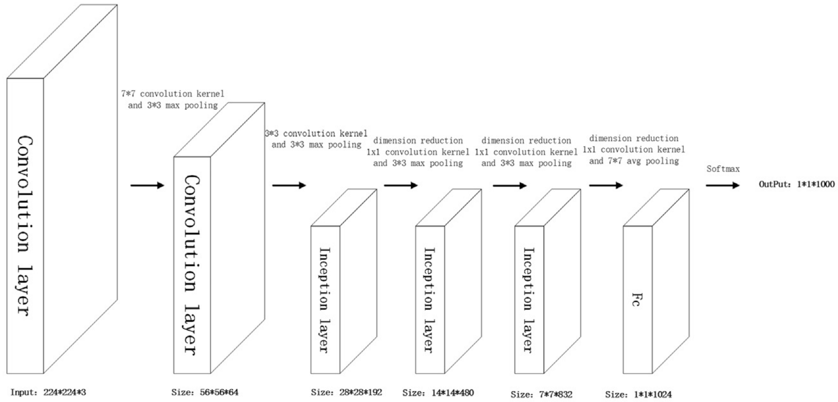
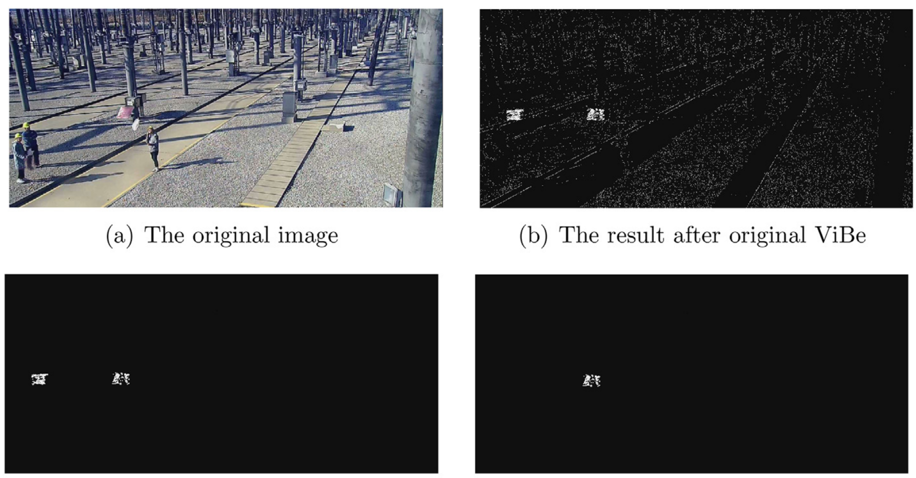



**Fig. 3.** classification network structure.



(a) The original image  (b) The result after original ViBe

(c) The result after eroding and dilating  (d) The result after improved ViBe

**Fig. 4.** The result with imporved ViBe algorithm.

A picture will have 2 K ~ 4 K anchors after the processing with the RPN. However, there are only a very small number of the anchors that have targets, and the rest are redundant anchors which are useless but time-consuming for processing. Instead of RPN, FODN4PS uses the MORE Net to extract candidate boxes based on the improved ViBe algorithm, which is good at extracting moving targets with less time.

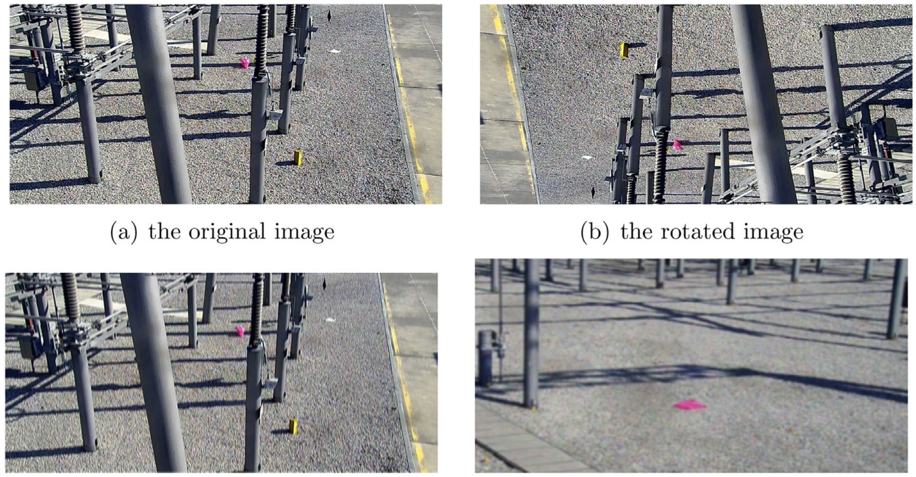
There are some disadvantages in the original ViBe algorithm, which may generate many flashing points and ghosts [54] if the background is complex. Ghosts usually exist in two cases. The first case is that when a background model is established according to the first frame, but the moving object stand there statically at that moment, so the ViBe algorithm thinks that the moving object belongs to background and establishes a wrong background model. Then the seemingly static object moves and ghosts are generated. The other case is that in the middle of the video, the stationary object moves to generate ghosts. In the substation scene, the equipment is stable and it doesn't move suddenly. Therefore we improve the background modeling using more than one frames. MORE Net uses the average pixel value of the first three frames of the video to establish the background model to avoid ghosts, more ever, MORE Net eliminates the flashing points by adding eroding and dilating. Take the pixel of a point as an example, let the pixel point be $x$, and the corresponding pixel value be $V_{(x)}$. Eq. (3) shows the calculation formula, where $n$ represents the number of frames; $i$ represents the R, G and B channels of the color of the image.

$$V_{average}^{C_{(i)}} = \frac{\sum_{k=1}^{n} V_k^{C_i}(x)}{N} \tag{3}$$

Fig. 4 shows the comparisons of the effects before and after improving the ViBe algorithm. Fig. 4(a) is the original image with a foreign object. Fig. 4(b) shows the result after original ViBe algorithm processing. We can see that there are countless flashing points and a ghost area after the original ViBe algorithm. After the operation of eroding and dilating, the flashing points are effectively eliminated, as shown in Fig. 4 (c), and after the process of the improved ViBe algorithm proposed in this paper, the ghosts disappear, and the results are shown in Fig. 4(d).

When there are foreign objects, the whole process of MORE Net is working as follows:

1. MORE Net initializes the background model firstly. For a pixel point, The average of the same neighbor positions in the first three frames is selected as its sample value. The calculation formula is $M_0(x) = v_0\left(\frac{\sum_{i=1}^{n} y_i}{n} | y_i \in N_{G(x)}\right)$. where $n$ is the count of the selected frames, $N_{G(x)}$ is the spatial neighbor of position x.



(a) the original image (b) the rotated image

(c) the blurred image (d) the GAN-generated image

**Fig. 5.** the result of image processing.

**Table 1**
accuracies of data processing methods.

| Methods | True detections | False detections | Accuracy |
|---|---|---|---|
| The original image | 331 | 104 | 76% |
| Image processing | 352 | 77 | 82% |
| GAN | 364 | 74 | 83% |
| GAN and image processing | 371 | 65 | 85% |

2. Then, MORE Net performs foreground detection. In the first step, the background model stores a sample set for each point. When there are a foreign object, MORE Net compares each new pixel value with the sample set to determine whether it belongs to the background by calculating the distance between the new pixel value and each sample value in the sample set. If the distance is less than the distance threshold, the number of approximate sample points increases. If the approximate sample points are greater than the points threshold, the new pixel point is considered as the background. The detection process is mainly affected by three parameters including the number of sample sets of the point$N$ (set to 20), the threshold $m$ (set to 2) for the number of approximate sample points, and the threshold $R$ (set to 20) for similar distance determination.

Because there are lighting effects in the background, MORE Net applies the update of the background model to make it adapt to the continuous changes of the background, such as changes in lighting and changes in background objects. Concerning the influence of lighting, More Net introduces the automatic update of the background model. When a pixel is detected as foreground $N$ times continuously, More Net will update this pixel as a background point.

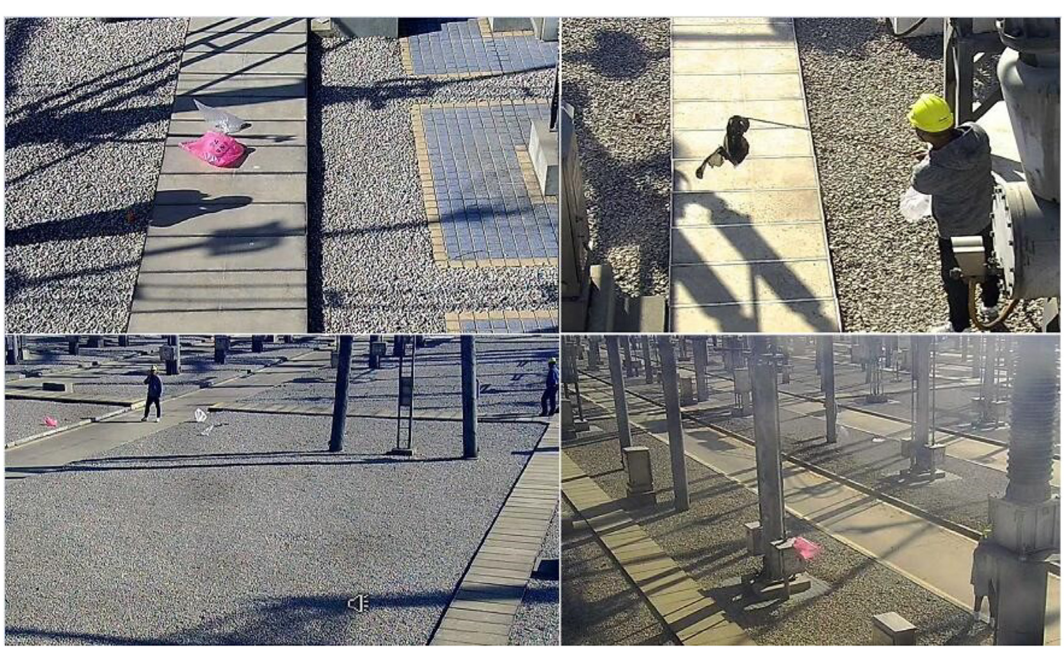


**Fig. 6.** Sample data sets.

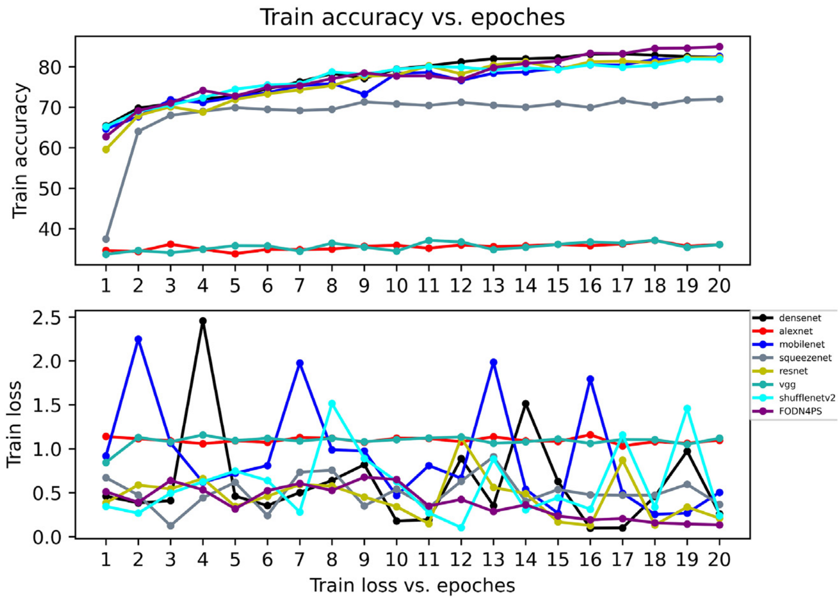
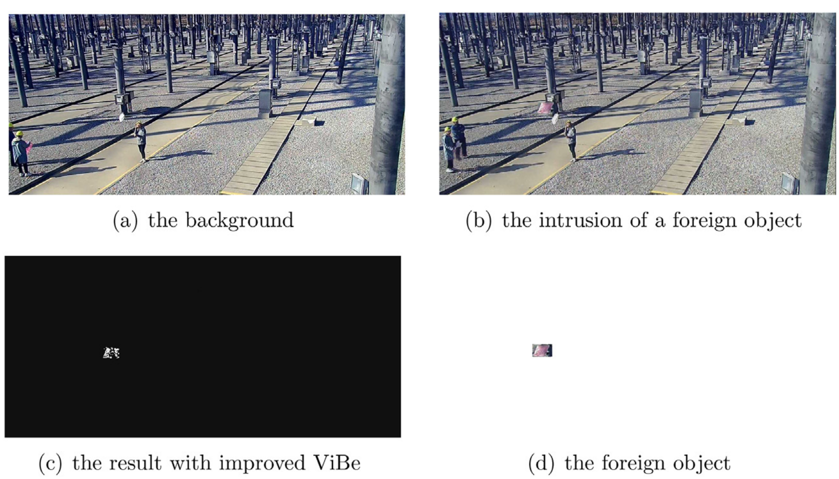



**Fig. 7.** The comparison with other classification networks.



(a) the background

(b) the intrusion of a foreign object

(c) the result with improved ViBe

(d) the foreign object

**Fig. 8.** the results with of MORE Net.

## 4. Evaluation

### 4.1. Hardware setting

The hardware we use for testing includes: the graphic card is NVIDIA GeForce GTX TITAN X with 12G memory, Intel i7 CPU. The software packages include python is 3.6 with CUDA version 10.1.

### 4.2. Data enhancement

The size of training set affects the performance of the trained model. However the data of foreign objects in the substation is hard to collect. To solve the problem of insufficient training data, we adopts GAN [55] method to generate new images and rotates images randomly to enhance the data. Firstly, We expand the data set by rotating the original image. Considering that the weather conditions will increase the difficulty of recognition, we fuzzy the images to make the model friendly for poor weather conditions. The original and the processed image are shown in Fig. 5.

In order to test the effect of the data processing methods, we carried out comparative experiments on original images, pre-processed, GAN generated, GAN and pre-processed images. We collected the following data set, some are with foreign objects. There are 3000 original images, 5000 pre-processed images, and 5000 GAN-generated images. The experimental results are shown in Table 1. we can see that the combination of GAN and image enhancement improves the accuracy of the model more effectively. The accuracy rate is calculated as follows, and the total number of test sets is 500.

Fig. 6 shows the examples of the experimental data.

### 4.3. Performance evaluation

In order to analyze the performance of FODN4PS, we evaluated the MORE Net first, and then the classification network. We compared FODN4PS with other classification networks, such as ResNet50, AlexNet, LeNet, and so on. The result is shown in Fig. 7. We can see that when the network is trained for 20 epochs, most classification networks tends to converge, and FODN4PS has a slightly higher accuracy than other

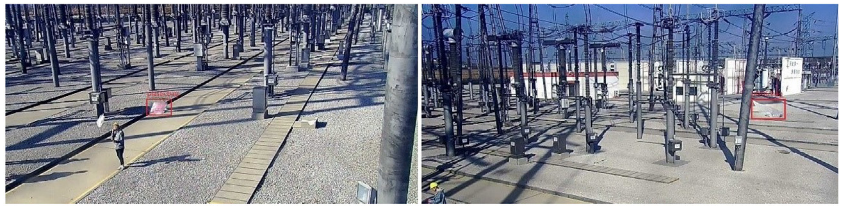


**Fig. 9.** experimental results.

**Table 2**
comparison with Faster R-CNN, Mask R-CNN and YOLOv3.

| Target detection algorithm | Running speed (seconds per frame) | Accuracy |
|---|---|---|
| FODN4PS Net | 0.15 | 85% |
| Faster R-CNN | 0.25 | 78% |
| Mask R-CNN | 0.21 | 81% |
| YOLOv3 | 0.04 | 48% |

networks. Meanwhile, we can find that the loss of FODN4PS gradually tends to be flat, and finally reaches to 0.15, which indicates that FODN4PS is easy to converge with this kind of data.

Fig. 8 shows the process of getting the candidate box with MORE Net. Firstly, MORE Net establishes the background model without foreign objects. Then, when foreign objects invade the substation, MORE Net detects the area with foreign objects. By separating the foreground from the background, we can clearly see the area with foreign objects. Finally, MORE Net determine the minimum circumscribed rectangle of the region according to the connectivity of the region as the result of MORE Net.

After MORE Net, FODN4PS maps the information of category and location to the original image, and the final result is shown in Fig. 9.

### *4.4. Comparisons of performance and accuracy*

To show the effectiveness of FODN4PS, we compared it with the current popular two-stage target detection algorithm, Faster R-CNN and Mask R-CNN and one-stage target detection algorithm, YOLOv3. The results are shown in Table 2. We can see clearly that the detection speed is 0.1 s faster than Faster R-CNN, and 0.06 s faster than Mask R-CNN, and the accuracy is improved by 7% and 4% respectively. Compared with the single-stage target detection network YOLOv3, FODN4PS is slower but it's twice as accurate as it.

## 5. Conclusions and future work

Foreign objects intrusion detection is essential for the safety of power grid operation. However. The existing work usually changes the R-CNN on the basis of retaining RPN components, which generates a lot of redundant anchors that incurs bad performance as computation are wasted on these anchors. We optimize this process by extracting accurate target candidate boxes that do not involve a series of complex calculations. In order to effectively detect foreign objects, this paper proposes a new neural network, FODN4PS, which consists of a region generation network MORE Net and classification network.

To efficiently train the FODN4PS, we increase the number of training samples through GAN and rotating images randomly. We made a series of experiments to evaluate FODN4PS Network. The experimental result shows that FODN4PS has higher accuracy, and faster detection speed.

In the future, we will pay more attention to improve MORE Net and make it extract the target region in a shorter time. We are evaluating more scenes to further refine the FODN4PS network.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgment

The research is supported by the National Natural Science Foundation of China (62072469), National Key R&D Program (2018yfe0116700), Shandong Natural Science Foundation (ZR2019MF049, parallel data-driven fault prediction under online and offline combined cloud computing environment), basic research fund of Central University (2015020031), West Coast artificial intelligence technology innovation center (2019-1-5, 2019-1-6), and the Opening Project of Shanghai Trusted Industrial Control Platform (TICPSH202003015-ZC).

## References

[1] C.G. Saracin, M.O. Popescu, Expert diagnose system and real-time monitoring of an electric power installation, Electroteh. Electron. Automat. 62 (3) (2014) 54.
[2] M. El-Harbawi, F. Al-Mubaddel, Risk of fire and explosion in electrical substations due to the formation of flammable mixtures, Sci. Rep. 10 (1) (2020) 1–9.
[3] O. Barnich, M. Van Droogenbroeck, Vibe: a powerful random technique to estimate the background in video sequences, 2009 IEEE international conference on acoustics, speech and signal processing, IEEE 2009, pp. 945–948.
[4] S. Ding, S. Qu, Y. Xi, S. Wan, Stimulus-driven and concept-driven analysis for image caption generation, Neurocomputing 398 (2020) 520–530, https://doi.org/10.1016/j.neucom.2019.04.095. URL http://www.sciencedirect.com/science/article/pii/S0925231219310367.
[5] S. Wan, Y. Xia, L. Qi, Y. Yang, M. Atiquzzaman, Automated colorization of a grayscale image with seed points propagation, IEEE Trans. Multimed. 22 (7) (2020) 1756–1768, https://doi.org/10.1109/TMM.2020.2976573.
[6] C. Chen, M.-Y. Liu, O. Tuzel, J. Xiao, R-cnn for small object detection, Asian Conference on Computer Vision, Springer 2016, pp. 214–230.
[7] R. Girshick, J. Donahue, T. Darrell, J. Malik, Rich feature hierarchies for accurate object detection and semantic segmentation, The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014.
[8] S. Jintao, G. Chaoyue, S. Hui, S. Jiangang, L. Zhe, Data expansion for foreign object detection in power grid, 2019 IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC), IEEE 2019, pp. 1–4.
[9] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, in: Z. Ghahramani, M. Welling, C. Cortes, N.D. Lawrence, K.Q. Weinberger (Eds.), Advances in Neural Information Processing Systems 27, Curran Associates, Inc. 2014, pp. 2672–2680.
[10] W. Zhang, X. Liu, J. Yuan, L. Xu, H. Sun, J. Zhou, X. Liu, Rcnn-based foreign object detection for securing power transmission lines (rcnn4sptl), Procedia Comp. Sci. 147 (2019) 331–337.
[11] R. Girshick, Fast r-cnn, Proc. IEEE Int. Conf. Comp. Vision (2015) 1440–1448.
[12] T. Kong, A. Yao, Y. Chen, F. Sun, Hypernet: Towards accurate region proposal generation and joint object detection, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2016, pp. 845–853.
[13] A. Neubeck, L.J.V. Gool, Efficient non-maximum suppression, International Conference on Pattern Recognition, 2006.
[14] Y. Zhao, H. Li, S. Wan, A. Sekuboyina, X. Hu, G. Tetteh, M. Piraud, B. Menze, Knowledge-aided convolutional neural network for small organ segmentation, IEEE J. Biomed. Health Inform. 23 (4) (2019) 1363–1373, https://doi.org/10.1109/JBHI.2019.2891526.
[15] A. Gupta, A. Vedaldi, A. Zisserman, Synthetic Data for Text Localisation in Natural Images, 2016.
[16] L. Liu, W. Ouyang, X. Wang, P. Fieguth, M. Pietikāinen, Deep Learning for Generic Object Detection: A Survey, 2018.

[17] A. Bochkovskiy, C.-Y. Wang, H.-Y.M. Liao, Yolov4: Optimal Speed and Accuracy of Object Detection, arXiv Preprint, arXiv:2004.10934 2020.
[18] S.-H. Choi, S.H. Jung, Similarity analysis of actual fake fingerprints and generated fake fingerprints by dcgan, Int. J. Fuzzy Logic Intellig. Syst. 19 (1) (2019) 40–47.
[19] Q. Wu, Y. Chen, J. Meng, Dcgan-based data augmentation for tomato leaf disease identification, IEEE Access 8 (2020) 98716–98728, https://doi.org/10.1109/ACCESS.2020.2997001.
[20] W. Fang, Y. Ding, F. Zhang, J. Sheng, Gesture recognition based on cnn and dcgan for calculation and text output, IEEE Access 7 (2019) 28230–28237, https://doi.org/10.1109/ACCESS.2019.2901930.
[21] Q. Wu, Y. Chen, J. Meng, DCGAN-based data augmentation for tomato leaf disease identification, IEEE Access 8 (2020) 98716–98728 IEEE.
[22] K. Wang, M.Z. Liu, Object recognition at night scene based on dcgan and faster r-cnn, IEEE Access 8 (2020) 193168–193182.
[23] R.M. Haralick, S.R. Sternberg, X. Zhuang, Image analysis using mathematical morphology, IEEE transactions on pattern analysis and machine intelligence, 4 (1987) 532–550.
[24] Y. Ke, R. Sukthankar, Pca-sift: A more distinctive representation for local image descriptors, Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004 , CVPR 2004., Vol. 2, IEEE, 2004, pp. II–II.
[25] D.G. Lowe, Distinctive image features from scale-invariant keypoints, Int. J. Comput. Vis. 60 (2) (2004) 91–110.
[26] E. Rosten, T. Drummond, Machine learning for high-speed corner detection, European Conference on Computer Vision, Springer 2006, pp. 430–443.
[27] B.S. Kumar, Image denoising based on gaussian/bilateral filter and its method noise thresholding, SIViP 7 (6) (2013) 1159–1172.
[28] G. Deng, L. Cahill, An adaptive gaussian filter for noise reduction and edge detection, 1993 IEEE Conference Record Nuclear Science Symposium and Medical Imaging Conference, IEEE 1993, pp. 1615–1619.
[29] V. Jayaraj, D. Ebenezer, A new switching-based median filtering scheme and algorithm for removal of high-density salt and pepper noise in images, EURASIP J. Adv. Sign. Process. 2010 (1) (2010) 690218.
[30] L. Alparone, V. Cappellini, A. Garzelli, A coarse-to-fine algorithm for fast median filtering of image data with a huge number of levels, Signal Process. 39 (1–2) (1994) 33–41.
[31] N. Otsu, A threshold selection method from gray-level histograms, IEEE Trans Syst. Man Cybernet. 9 (1) (1979) 62–66.
[32] A. Krizhevsky, I. Sutskever, G. Hinton, Imagenet classification with deep convolutional neural networks, Adv. Neural Inf. Proces. Syst. 25 (2) (2012).
[33] K. He, X. Zhang, S. Ren, J. Sun, Spatial pyramid pooling in deep convolutional networks for visual recognition, IEEE Trans. Pattern Anal. Mach. Intellig. 37 (9) (2014) 1904–1916.
[34] S. Ren, K. He, R. Girshick, J. Sun, Faster r-cnn: towards real-time object detection with region proposal networks, IEEE Trans. Pattern Anal. Mach. Intellig. 39 (6) (2017) 1137–1149.
[35] R. Girshick, J. Donahue, T. Darrell, J. Malik, Rich feature hierarchies for accurate object detection and semantic segmentation, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2014, pp. 580–587.
[36] C. Eggert, S. Brehm, A. Winschel, D. Zecha, R. Lienhart, A closer look: Small object detection in faster r-cnn, 2017 IEEE international conference on multimedia and expo (ICME), IEEE 2017, pp. 421–426.
[37] S. Ren, K. He, R. Girshick, J. Sun, Faster r-cnn: Towards real-time object detection with region proposal networks, in: C. Cortes, N.D. Lawrence, D.D. Lee, M. Sugiyama, R. Garnett (Eds.), Advances in Neural Information Processing Systems 28, Curran Associates, Inc. 2015, pp. 91–99.
[38] K. He, G. Gkioxari, P. Dollár, R. Girshick, Mask r-cnn, Proceedings of the IEEE International Conference on Computer Vision 2017, pp. 2961–2969.
[39] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, You only look once: unified, real-time object detection, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2016, pp. 779–788.
[40] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, A.C. Berg, Ssd: Single shot multibox detector, European Conference on Computer Vision, Springer 2016, pp. 21–37.
[41] J. Redmon, A. Farhadi, Yolo9000: better, faster, stronger, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2017, pp. 7263–7271.
[42] J. Redmon, A. Farhadi, Yolov3: An incremental improvement, arXiv preprint, arXiv:1804.02767 2018.
[43] L. Huang, Y. Yang, Y. Deng, Y. Yu, Densebox: Unifying landmark localization with end to end object detection, arXiv preprint, arXiv:1509.04874 2015.
[44] S. Zhang, C. Chi, Y. Yao, Z. Lei, S.Z. Li, Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection, 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020.
[45] H. Law, J. Deng, Cornernet: detecting objects as paired keypoints, Proceedings of the European Conference on Computer Vision (ECCV) 2018, pp. 734–750.
[46] J. Wang, K. Chen, S. Yang, C.C. Loy, D. Lin, Region proposal by guided anchoring, 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020.
[47] F.-C. Cheng, S.-C. Huang, S.-J. Ruan, Illumination-sensitive background modeling approach for accurate moving object detection, IEEE Trans. Broadcast. 57 (4) (2011) 794–801.
[48] D. Zhou, H. Zhang, Modified gmm background modeling and optical flow for detection of moving objects, 2005 IEEE international conference on systems, man and cybernetics, Vol. 3, IEEE 2005, pp. 2224–2229.
[49] V. Ramesh, et al., Background modeling and subtraction of dynamic scenes, Proceedings Ninth IEEE International Conference on Computer Vision, IEEE 2003, pp. 1305–1312.
[50] M. Zhengchong, S. Xuesong, Improved vibe algorithm integrated with multiscale transformation, Laser Optoelectron. Progr. 55 (11) (2019) 111501.
[51] W. Sun, H. Du, G. Ma, S. Shi, X. Zhang, Y. Wu, Moving vehicle video detection combining vibe and inter-frame difference, Int. J. Embed. Syst. 12 (3) (2020) 371–379.
[52] X. Xia, X. Lu, Y. Cao, S. Xia, C. Fu, Moving vehicle detection with shadow elimination based on improved vibe algorithm, Journal of Physics: Conference Series, Vol. 1302, IOP Publishing 2019, p. 022080.
[53] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2015, pp. 1–9.
[54] O. Barnich, M. Van Droogenbroeck, Vibe: a universal background subtraction algorithm for video sequences, IEEE Trans. Image Process. 20 (6) (2010) 1709–1724.
[55] A. Radford, L. Metz, S. Chintala, Unsupervised representation learning with deep convolutional generative adversarial networks, arXiv preprint, arXiv:1511.06434 2015.

**Liang Xu** has obtained master degree at China University of Petroleum, majoring in software engineering. Now he is a Ph.D candidate in Beijing University of Science and Technology. His main research interests include deep learning, software engineering and big data processing

**Yongkang Song** is master student at China University of Petroleum, his main research area is computer vision.

**Weishan Zhang** is an professor at China University of Petroleum. His main research areas are deep learning, computer vision, and big data processing.

**Yunyun An** is a researcher at Huangdao District Power Supply Company of Shandong Electric Power Company Qingdao, State Grid, Qingdao. Her main research interest is big data processing. She has obtained master degree at China University of Petroleum.

**Ye Wang** is a researcher at PLA 9144, his research interest is big data processing.

**Huansheng Ning** is a professor in the School of Computer and Communication Engineering, University of Science and Technology Beijing, China. His current research interests include Internet of Things, Cybermatics.