A Deep Learning and Depth Image based Obstacle Detection and Distance Measurement Method for Substation Patrol Robot

XU Hongsheng, CHEN Tianyu, ZHANG Qipei, LU Jixiang, YANG Zhihong
NARI Group Corporation (State Grid Electric power Research Institute)
Nanjing, China
xuhongsheng@sgpri.sgcc.com.cn

Abstract. Recently, substation patrol robot is gradually used to replace the manual inspection in order to improve the inspection efficiency as well as security and automation level of substation maintenance. The research of obstacle avoidance is a hot spot in substation intelligent patrol robot area. The emerging new generation of artificial intelligence (AI) technology provides a new way to solve the obstacle detection and distance measurement problem. To realize accurate, effective and real-time response to the environmental changes, a novel obstacle avoidance method based on deep learning and depth image is proposed. The core of this method is pixel-level instance segmentation between obstacles and roads, along with a pixel-level matching of obstacles’ segmentation mask and depth data. The effectiveness of the proposed method is validated by actual tests in real substation environment.

1. Introduction

A substation is a hub for transforming voltage and transmitting electrical energy in power systems, where serious grid operation failures will be caused once an accident occurs. Therefore, in order to guarantee the safe production and reliable operation of the power grid, regular inspections of substation equipment are essential. However, false investigations and mistakes often occurs due to the labor intensity, professional level and many other complicated factors [1]. The new generation of information and communication technology represented by AI has a significant impact on substation inspections, which has caused profound changes. Among them, the substation intelligent patrol robot is one of the representative products. The substation intelligent patrol robot is a robot that can substitute human workers to inspect the substation equipment in the dangerous environment and severe weather [2]. It significantly improves the efficiency of substation inspection, reduces the work intensity, danger and risk of the front-line workers, and further improves security and automation level of substation maintenance [3].

In the research of substation intelligent patrol robots, autonomous navigation is one of the most critical technology. There are two types of navigation according to the level of the robot’s understanding of the substation environment: environmental information completely known global navigation, and environmental information completely or partially unknown local navigation [4], which includes the obstacle avoidance approach. The obstacle avoidance enables the robot to react to the changing surrounding environment in real time when performing inspection tasks. It should automatically detect the obstacles and distances to avoid the risk of collisions and falls, and continue to follow the planned global path after obstacle avoidance.
Magnetic track guidance is used to guide the robot and to realize the obstacle avoidance in the early time. This method has the advantages of high accuracy, stability and adaptability, but lacks flexibility and requires an expensive deployment cost [5-7]. In addition, it relies on other sensors, such as the radio frequency identification (RFID) sensor, to realize the localization [8], which makes the whole system structure complex and less reliable. With the development of new sensor technologies, laser and vision sensors based navigation system is used to sense the environment and deal with obstacle avoidance. The light detection and ranging (LIDAR) is one of the excellent laser sensors for environment perception and obstacle avoidance for robot [9]. It can achieve high accuracy and resolution, but it suffers from the limited sensing range as mentioned in [4]. Vision based navigation for patrol robots have been implemented by using different methods, like Bumblebee2 [10], motion algorithms [11], and neural network [12]. Compared with the above two systems, the vision system is simple and has wide detection range. But, the existing vision methods has high algorithm complexity. For example, the method based on neural network needs to extract features artificially. And, the algorithm based on binocular stereo vision is much more complex, and more vulnerable to complex illumination.

To the best of our knowledge, it is the first time that instance segmentation and depth image are introduced and utilized in obstacle detection and distance measurement for substation patrol robot. The content of this paper is structured as follows: Section II briefly introduces and compares two classical deep learning algorithms for object recognition. Section III introduces the Mask R-CNN network which is used to fulfill obstacle detection and instance segmentation, and describes the proposed model based on Mask R-CNN and depth image. Section IV provides an experiment in real substation environment to verify the effect of our proposed method. Some results of obstacle detecting and distance measuring are presented. Finally, a conclusion is given in Section V.

2. Related Work
In recent years, great achievements have been made in deep learning [13]. It does not need the manual feature design, and has good feature expression ability and excellent detection accuracy. The deep learning based object recognition algorithms has surpassed the traditional approaches in performance, and has become the mainstream methods. According to different design ideas, they can be divided into two categories: one is two-stage region-based convolutional neural network (R-CNN), the other one is one-stage network.

2.1. Faster R-CNN Network
R-CNN is a classical and fundamental algorithm using region proposal method [14]. On the basis of R-CNN, The improved algorithms Fast R-CNN [15] and Faster R-CNN [16] are proposed. Both R-CNN and Fast R-CNN use traditional region proposal method, such as selective search [17], to generate region proposals, which is very time consuming. Faster R-CNN uses region proposal network (RPN) to replace the previous region proposal method which is relatively less efficient. The core idea of RPN is to use convolutional neural network (CNN) to generate the region of interest (RoIs) directly. Specifically, the convolutional feature maps are shared by the RPN with a region-based detector, for example the Fast R-CNN, to generate RoIs.

The structure of Faster R-CNN is shown in Fig. 1, which can be decomposed into three parts: a CNN extractor, a RPN and a Fast R-CNN detector. The CNN extractor is used to generate the feature maps from the original input image. The RPN is actually a deep CNN, and is used to generates the region proposals. The detector part is actually Fast R-CNN, which utilizes the outputs of the RPN to further determine the position of the bounding-box precisely and the classifications of target objects.

Faster R-CNN presents a fairly good performance, but there is still room for improvement in the capability of small object detection and the calculation speed, which spawns the YOLO network.

![Figure 1. The structure diagram of Faster R-CNN.](image-url)
2.2. YOLO v3 Network
YOLO (You Only Look Once) is a typical one-stage object detection algorithm proposed by Joseph Redmon in 2016 [18]. It ingeniously unifies classification prediction and bounding-box regression into a single regression problem. Compared with region-based approaches, YOLO is faster and more real-time since it passes the input image only once in a fully convolutional network (FCN). Later, in 2017, an improved network YOLO v2 [19] was proposed to overcome the relatively high localization error and low recall by making batch-normalization and higher resolution classifier [20]. Most recently, in 2018, YOLO v3 [21] is released with a higher accuracy through some improvements, like the multi-label classification, the concept of feature pyramid networks (FPN) [22], and new CNN feature extractor named Darknet-53.

![Image of YOLO v3 structure diagram]

**Figure 2.** The structure diagram of YOLO v3.

The structure of YOLO v3 is shown in Fig. 2. First, the input image is divided into a $K \times K$ grid. If the center point of the object’s ground truth fall within a certain grid, the grid is responsible for detecting the object [23]. Each grid cell is validated with a classifier which only predict a fixed number of bounding-boxes, and each box is given a confidence score. Various measures have been adopted in YOLO v3 to obtain image features at different scales, hence greatly improved the detection effect of small targets.

Although Faster R-CNN and YOLO v3 networks perform well in object recognition task, they can only give the rectangular bounding boxes of target objects, lacking the ability of precisely segmenting each instance.

3. Model
As shown in Fig. 3, the framework of our proposed model is based on Mask R-CNN [24], which is a state-of-the-art network for its high performance in instance segmentation. In this section, we introduce the proposed method of obstacle detection and distance measurement in details.

![Image of proposed model framework]

**Figure 3.** The framework of our proposed model.

3.1. Instance Segmentation: Mask R-CNN
Mask R-CNN can not only be used to detect object in an image but also used to generate a segmentation mask for each instance. It extends Faster R-CNN by adding an extra branch of FCN which fulfill the instance segmentation task. As shown in Fig. 3, the segmentation task is in parallel with the identification and localization tasks. Similar to Faster R-CNN, Mask R-CNN also contains three parts. The first part is the FPN backbone playing an important role as feature extractor in Mask R-CNN. The second part is the RPN which receives the feature maps outputted from the backbone network to generate RoIs. The third part is a Fast R-CNN detector and a binary mask prediction branch [25]. When we compare Fig. 1 and Fig. 3, we can see that the ROI pooling layer is replaced by
ROI Align so as to fix the misalignment problem. Then, just like Faster R-CNN, classification and bounding-box regression is done by the Fast R-CNN detector. The only difference is that the FCN branch performs mask prediction and output a binary segmentation mask.

3.2. Obstacle Detection and Distance Measurement Method

The proposed method requires only one depth camera to produce RGB-D depth images. Fig. 4 shows the process diagram of obstacle detection and distance measurement method. In the first step, the video file is converted into single RGB image and depth image frame by frame, and then the pre-processed RGB images are used to train the deep neural network Mask R-CNN. In the second step, the trained Mask R-CNN network is run to perform pixel-level instance segmentation between obstacles and roads, and generate segmentation mask. In the third step, we find the very bottom pixel point on the segmentation mask and the adjacent pixel points. Then, all these pixel points are connected with the depth data which is extracted from the depth image. Grubbs method is used to eliminate the abnormal data, which increases the robustness of the model. Finally, we can measure the distance from the patrol robot to the obstacle through pixel-level matching. Meanwhile, the Mask R-CNN network also outputs the obstacle detection results.

![Figure 4](image.png)

**Figure 4.** The process diagram of obstacle detection and distance measurement method.

4. Experiments

The proposed model is implemented on an embedded AI computing device Nvidia Jetson TX2, which are able to support real-time inference based on relatively low power consumption. This device is installed in a substation patrol robot which is developed and manufactured by us, and a depth camera Intel RealSense D435 is also installed as vision and depth sensor shown in Fig. 5. In order to validate the effectiveness of our proposed method, an experiment is conducted in real substation environment.

![Figure 5](image.png)

**Figure 5.** The patrol robot used for experiments in real substation environment.
4.1. Experiment Methods
The training set totally contains 628 images including 300 road samples of different types of pavement shown as Fig. 6, and 328 obstacle samples of different sizes, shapes and textures. This dataset was collected from images taken by our patrol robot products which have been deployed in more than fifty substations in different provinces in China. We tried to collect samples from different environments and scales to assure the validity of the experiment and to test the extensibility of our proposed method. The training of Mask R-CNN is carried out on a GPU server. The configurations of hardware and software used in this experiment are as follows:
- CPU: Intel Xeon CPU E5-2678 v3 (48 cores, Sky Lake architecture)
- GPU: Nvidia GTX 1080Ti, 11GB GDDR5X
- RAM: 128 GB RAM
- Operating system: Linux (Ubuntu 16.04)
- GPU platform: CUDA 9.0
- Deep learning framework: Pytorch 1.1
- Deep learning pipeline: Mask R-CNN
- Backbone: ResNet101

![Figure 6. Different types of road pavements in substation.](image)

4.2. Results
Results of the tests of Mask R-CNN on obstacle detection and instance segmentation are shown in Fig. 7-9 respectively. In Fig. 7, it is demonstrated that our proposed method performs well in detecting different types of roads, and generate segmentation masks with precise contours. In Fig. 8, we can see that the our proposed model gives reasonable good results of detecting and segmenting different types of obstacles, such as round shapes, cubic shapes, flat objects, and tall objects. In Fig. 9, we presents the results of detecting and segmenting for human body, which is a movable obstacle often appearing in substations. It is worth mentioning that our proposed model performs well even if only part of the body (such as legs and feet) is “seen” by the robot, which is the most common case of robot vision system in practice.
**Figure 7.** Detection and segmentation results of different types of roads.

**Figure 8.** Detection and segmentation results of different types of obstacles.
We also did the experiment of real-time obstacle detecting and distance measuring based on segmentation mask and depth data. Fig. 10 displays a screenshot captured from a video which is recorded during the experiment. The measured distances between the robot and the obstacle are listed in the left column when the robot is approaching the obstacle. From the decreasing values, we can see that our proposed method can measure the distance between the robot and the obstacle, leading the robot to complete the subsequent obstacle avoidance action.

5. Conclusion
In this paper, we have introduced and compared two state-of-the-art algorithms for object detection (Faster R-CNN and YOLO v3). These two networks are capable to complete object recognition task (giving bounding boxes of target objects), but cannot achieve instance segmentation task (providing pixel-level masks). The proposed obstacle detection and distance measurement based on Mask R-CNN and RGB-D depth images have been described in details. The pre-processed RGB images are used to train the Mask R-CNN network, while the depth data is extracted from the depth image. By connecting the pixel-level masks obtained from the trained Mask R-CNN network with the depth data, one can finally get the distance between robot and obstacle through pixel-level matching. Experimental results are presented to validate the effectiveness of our proposed method. Compared with the old methods,
our proposed method has a simple framework with acceptable precision, good reliability, low cost and is valuable in practice.

6. References

[1] L. Liu, J. Peng, R. Zhang, B. Chen, Y. Yang and X. Zhang, “Temporal logic task and motion planning of a smart robot-towards a smart substation environment,” 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Banff, AB, 2017, pp. 1110-1115.

[2] L. Li et al., “A state-of-the-art survey of the robotics applied for the power industry in China,” 2016 4th International Conference on Applied Robotics for the Power Industry (CARPI), Jinan, 2016, pp. 1-5.

[3] H. Wang, J. Li, Y. Zhou, M. Fu and S. Yang, “Research on the Technology of Indoor and Outdoor Integration Robot Inspection in Substation,” 2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), Chengdu, China, 2019, pp. 2366-2369.

[4] Y. Liang, C. Cai, J. Zhao, G. Chang, L. Wang and X. Lv, “Local environments modelling and path planning for patrol robot in the substation,” 2016 IEEE 20th International Conference on Computer Supported Cooperative Work in Design (CSCWD), Nanchang, 2016, pp. 619-624.

[5] Haojie Zhang, Bo Su and Hong Meng, “Development and implementation of a robotic inspection system for power substations,” Industrial Robot, Vol. 44 No. 3, pp. 333-342, 2017.

[6] Han-Shue Tan, J. Guldner, S. Patwardhan, Chieh Chen and B. Bougler, “Development of an automated steering vehicle based on roadway magnets-a case study of mechatronic system design,” in IEEE/ASME Transactions on Mechatronics, vol. 4, no. 3, pp. 258-272, Sept. 1999.

[7] J. Z. Sasiadek and Q. Wang, “Sensor fusion based on fuzzy Kalman filtering for autonomous robot vehicle,” Proceedings 1999 IEEE International Conference on Robotics and Automation, Detroit, MI, USA, 1999, pp. 2970-2975 vol.4.

[8] M. I. Hamzah, and T. Y. Abdall, “Mobile Robot Navigation using Fuzzy Logic and Wavelet Network,” International Journal of Computer Applications, 79(10), pp.4-10, 2014.

[9] L. Shengfang and H. Xingzhe, “Research on the AGV Based Robot System Used in Substation Inspection,” 2006 International Conference on Power System Technology, Chongqing, 2006, pp. 1-4.

[10] X. Z. Xie et al., “Obstacle Detection for Patrol Robot Using Bumblebee2 Stereo Vision System,” Applied Mechanics and Materials, Vols. 48-49, pp. 749-752, 2011.

[11] Chris Harris, Geometry from Visual Motion, Active vision, Eds Blake & Yuille, MIT Press, 1993.

[12] D. A. Pomerleau, “Progress in neural network-based vision for autonomous robot driving,” Proceedings of the Intelligent Vehicles ’92 Symposium, Detroit, MI, USA, 1992, pp. 391-396.

[13] Y. Lecun, Y. Bengio and G. Hinton, “Deep learning,” Nature, Vol.521, No.7553, pp.436–444, 2015.

[14] R. Girshick, J. Donahue, T. Darrell and J. Malik, “Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation,” 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Columbus, OH, 2014, pp. 580-587.

[15] R. Girshick, “Fast R-CNN,” IEEE International Conference on Computer Vision. IEEE, Santiago, Chile, pp.1440–1448, 2015.

[16] S. Ren, K. He, R. Girshick, et al., “Faster R-CNN: Towards real-time object detection with region proposal networks,” Advances in Neural Information Processing Systems, pp.91–99, 2015.

[17] J.R.R. Uijlings, K.E.A. Van De Sande, T. Gevers, et al. “Selective search for object recognition,” International Journal of Computer Vision, Vol.104, No.2, pp.154–171, 2013.

[18] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, “You Only Look Once: Unified, Real-Time Object Detection,” 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, 2016, pp. 779-788.

[19] J. Redmon and A. Farhadi, “YOLO9000: Better, Faster, Stronger,” 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, 2017, pp. 6517-6525.
[20] B. Benjdira, T. Khursheed, A. Koubaa, A. Ammar and K. Ouni, “Car Detection using Unmanned Aerial Vehicles: Comparison between Faster R-CNN and YOLOv3,” 2019 1st International Conference on Unmanned Vehicle Systems-Oman (UVS), Muscat, Oman, 2019, pp. 1-6.

[21] J. Redmon and A. Farhadi, “Yolov3: An incremental improvement,” arXiv:1804.02767, 2018.

[22] T. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan and S. Belongie, “Feature Pyramid Networks for Object Detection,” 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, 2017, pp. 936-944.

[23] F. Wu, G. Jin, M. Gao, Z. HE and Y. Yang, “Helmet Detection Based On Improved YOLO V3 Deep Model,” 2019 IEEE 16th International Conference on Networking, Sensing and Control (ICNSC), Banff, AB, Canada, 2019, pp. 363-368.

[24] K. He, G. Gkioxari, P. Dollár and R. Girshick, “Mask R-CNN,” 2017 IEEE International Conference on Computer Vision (ICCV), Venice, 2017, pp. 2980-2988.

[25] S. Nie, Z. Jiang, H. Zhang, B. Cai and Y. Yao, “Inshore Ship Detection Based on Mask R-CNN,” IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, 2018, pp. 693-696.