Detection of abnormal dangerous behavior in electrical power grids based on the fusion of contextual features

Xinyu Huang[*], Song Gao, Gang Qiu, Xiao Tan, Nailong Zhang, Jie Chen

State Grid Jiangsu Electric Power Co., Ltd., Electric Power Research Institute, Nanjing 211103, Jiangsu, China

ABSTRACT

Intelligence inspection of the electrical power grid can effectively improve the efficiency of inspection. However, the complex scenes between grid work and the interference of similar objects seriously affect the recognition accuracy of target identification. To address the above problems, this paper proposes the detection of abnormal dangerous behaviors in power grids based on contextual feature fusion. First, the feature extraction backbone network is constructed to obtain image features. Second, the hierarchical contextual attention mechanism is constructed to capture contextual features. Finally, the target detection model with contextual feature fusion is constructed to achieve grid abnormal risk behavior detection. The model proposed in this article is compared with the existing object recognition model in the simulation of the three datasets of Safety helmet (hardhat) wearing detect dataset, Hard Hat Workers dataset, Safety Helmet Detection dataset. A large number of experiments have proved the object recognition algorithm proposed in this paper is effective and outperforms existing algorithms. The average recognition accuracy of the proposed model is 0.948, which is improved by 1.02%.

Keywords: Context information, feature fusion, power grid, abnormal dangerous behaviour, detection algorithm

1. INTRODUCTION

The electric power industry is the basic industry of China's national economy, and the stable operation of the power grid system is an important prerequisite to ensure the steady development of the economy. The country actively promotes the deep integration of artificial intelligence technology and grid construction. The operation method of the electric power industry is undergoing a profound change from manual operation methods to automatic, intelligent and information operation methods. However, there are often irregularities in the behavior of electrical power grid staff. These behaviors are dangerous. Accurate identification can timely avoid the occurrence of accidents, ensure the safety of staff life, and reduce the loss of electrical power grid. Therefore, the detection and identification of dangerous behaviors on the electrical power grid are of great attention to scholars[1].

Electrical power grid hazardous behavior detection aims to detect the presence of hazardous behavior in the electrical power grid[2]. Hazardous behavior detection relies mainly on computer vision. Computer vision technology can process the data automatically and quickly and extract effective information as needed to analyze and recognize the target in the image automatically[3]. The target detection models based on deep learning can be divided into two categories: the first stage model is used to achieve the preliminary detection of the target, and the next stage model is used to achieve the accurate detection of the target. One-stage target detection model has a greater advantage in detection speed, while the two-stage target detection model has the advantage of high detection accuracy[4]. Typical algorithms selected for comparison include YOLO[5], SSD[6], etc. Typical one-stage detection algorithms mainly include R-CNN[7], Fast R-CNN[8], Faster R-CNN[9], etc.

Intelligent monitoring of the electrical power grid can greatly improve the risk resistance of the power system. It can help achieve intelligent inspection and unattended real-time monitoring of the normal operation of electrical equipment. It can also detect hidden faults in time and effectively guarantee the long-term stable and safe operation of the power system. Such as intelligent inspection of electrical transmission lines by drones, staff detection, intelligent grid connection and intelligent dispatching of electricity. However, the working area of the electrical power grid is narrow, and the scene is relatively complex. The existence of a large number of human occlusions in complex scenes undoubtedly increases the

---

[*] jsdw_hxy@126.com

difficulty of model feature extraction. This will affect the recognition accuracy of the model. Modeling the interaction between multiple human targets in a complex scene is also challenging, which will seriously affect the model's understanding of the scene.

To address the above problems, this paper proposes a detection of abnormal dangerous behaviors in power grids based on the contextual feature fusion algorithm, whose recognition results are shown in Figure 1. First, the feature extraction backbone network is constructed to obtain image features. Second, the hierarchical contextual attention mechanism is constructed to capture contextual features. Finally, the target detection model with contextual feature fusion is constructed to achieve grid abnormal risk behavior detection. Its network framework is shown in Figure 2. The accuracy recognition of dangerous behaviors in the electric power grid will help regulate the operation of staff and greatly improve the safety of the electric power system.

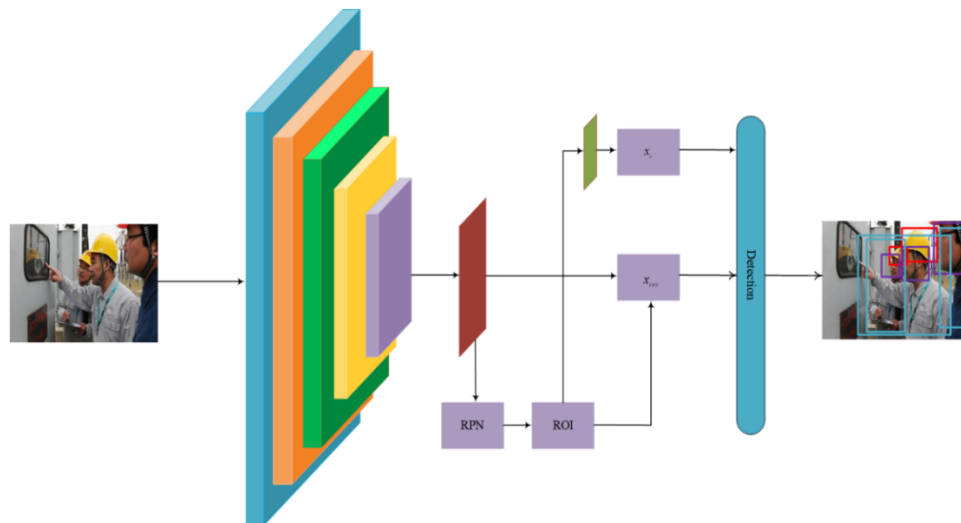Figure 1. The detection results of the algorithm.

Figure 2. The algorithm framework.

The rest of this paper is presented below. Part II describes the related work. Part III presents the principle of the object detection algorithm proposed in this paper. Part IV presents the design of simulation experiments based on the algorithm proposed in this paper and analyzes the experimental results. The fifth part concludes and indicates future research work.

2. RELATED WORK

At present, relevant research institutions and scholars at home and abroad have achieved many remarkable results for the intelligent inspection of power grids, and we will briefly review the relevant contents.

2.1 Intelligent monitoring system

Dutta et al.[10] studied the image processing technology in thermal state monitoring of power equipment. After converting RGB color model to HSI color model, they processed it and extracted different gradient edge descriptors. Edge operators such as Prewitt, Roberts and Sobel detected and recognized infrared thermal images. Finally, the Otsu image segmentation method based on clustering is adopted, which can be used to detect the thermal state of power equipment. Huda et al.[11] applied infrared thermography to thermal anomaly detection of electrical equipment. The imaging characteristics of different thermal anomalies of power equipment are analyzed and the corresponding discriminant methods are given. The effectiveness of this method is verified by experiments. By summarizing the development trend of power equipment fault diagnosis and related fault diagnosis techniques, Heano et al.[12] thoroughly studied the detail enhancement method of infrared images of power equipment, extracted the shape features of target equipment, and verified the feasibility and adaptability of the designed method in power equipment fault diagnosis.

2.2 Contextual mechanisms

The purpose of the context attention mechanism is not only to detect the target and the person in the image but also to infer the relationship between the person and the target. Kato et al.[13] proposed visual composition learning for character interaction detection. In this model, the relationship between objects is decomposed into specific features of objects and verbs, and new interaction samples are formed in feature space by splicing the decomposed features. New interactive samples can be composed in feature space by splicing decomposed features. Xiong et al.[14] proposed a new graph neural network using multidimensional edge features, which uses multidimensional edge feature information to construct the edge matrix in the graph. The implicit relationship between objects is obtained by updating the nodes and edge features. Cheng et al.[15] proposed an end-to-end human interaction detection model. The model uses encoders to generate global memory features, explicitly models the relationships between different target features, and uses multi-layer perception to predict character interactions.

3. ALGORITHM

This paper proposes a target model for detecting abnormal dangerous behavior of power grid based on context feature fusion to improve the ability of intelligent inspection of the power grid. The model mainly includes feature extraction backbone network module, hierarchical context feature extraction module and target detection module.

3.1 Human backbone feature extraction model

The paper uses the VGG-16 backbone to extract depth features from the input images. First, the color image is input to the feature extraction backbone network to generate a feature map $F(H, W, C)$ containing high-level semantic information. All convolution kernels are set to size 3×3 to capture more information. In this paper, a small network is slid on the output convolutional feature map of the last shared convolutional layer. The small network takes as input a $n \times n$ space window of the input convolutional feature mapping. At each sliding window position, we predict multiple region proposals simultaneously. The maximum number of possible proposals for each position is denoted as k. Therefore, the coordinates of k candidate boxes are generated in the reg layer. Also the confidence level of each candidate box proposal object in the cls layer.

3.2 Contextual mechanisms

In this paper, we propose a hierarchical contextual feature extraction network for improving the performance of the model. First, in this paper, we apply the proposal generated by RPN to the feature map X to obtain the local RoI features x_{local}, which are as follows[16]:

$$x_{local} = f_{ROI-Align}(X; h', w') \tag{1}$$

where $f_{ROI-Align}(\cdot)$ represents the ROI-Align operation function. h' and w' respectively represent the height and width of the region of interest.

To better utilize the contextual features, this paper uses the feature map F to obtain global-level contextual RoI features x_{global}, which are as follows.

$$x_{global} = f_{ROI-Align}(X; H, W) \tag{2}$$

where H and W respectively are the height and width of the input image.

Contextual features are obtained by fusing local contextual features with global contextual attention features through a convolutional layer, as shown in the following equation：

$$x_{context} = f_{couv}\left(\left[x_{local} : X_{global}\right]\right) \tag{3}$$

where f_{couv} represents the convolution operation, $[:]$ represents the concatenation, and $x_{context}$ represents the extraction of the contextual information.

To obtain a greater degree of the dynamic relationship between contextual features and targets, and to enhance the role of contextual information for target detection, a more relevant and reliable contextual feature for object detection can be obtained by equation 3.

$$x_c = f_{inr}\left(x_{contest}, x, \Omega[x_{contest}, x], \Omega[x_{contest}, x_{contest}]\right) \tag{4}$$

where $\Omega[x_{contest}, x]$ represents the correlation between contextual features and target features. $\Omega[x_{contest}, x_{contest}]$ represents the correlation between contextual features. f_{inr} represents the dynamic encoding function. x_c represents the more relevant and reliable contextual information extracted by the network.

In this paper, a feature fusion strategy is proposed to achieve the fusion of the above information and the features extracted from the backbone network. The principle is shown in Figure 3.

$$x_F = \left|x_c, x_{FPN}\right| \tag{5}$$

where x_F represents the fused features. $|\cdot|$ represents the feature fusion operation.
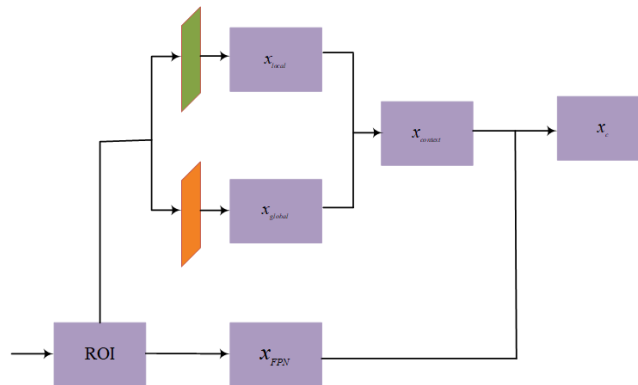

Figure 3. The hierarchical contextual feature extraction module.

3.3 Target detection model

This paper uses the two-layer convolutional layer to generate classification confidence for contextual features, which is fused with the classification confidence generated by ROI as follows:

$$y_F = \left|f_{2fc}(y_c), f_{2fc}(y_{FPN})\right| \tag{6}$$

The fused classification confidence is passed through the soft-max layer to generate new classification confidence. In this paper, we use G to denote the coordinates of the candidate frames. Assuming that the coordinates of the candidate frames are independent, the principle of obtaining the probability distribution can be expressed as:

$$P(\text{G}) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(\text{G}_a - \text{G})^2}{2\sigma^2}} \tag{7}$$

where G_a represents the coordinates of the true bounding box. σ represents the standard deviation, which is used to measure the estimated uncertainty. When $\sigma \to 0$, it means that the candidate box proposed by the identification framework in this paper is worth to be considered.

Not all the candidate frames obtained by the above process are valid. Therefore further screening of them is needed. In this paper, the non-maximum suppression (NMS) method is applied to the proposed region based on the cls scores of the proposed candidate frames. We set the IOU threshold of NMS to 0.5 and use the top n-ranked proposed regions for detection to finally achieve target detection.

4. EXPERIMENT AND ANALYSIS

In order to verify the performance of our design model, this chapter will mainly show a large number of comparative experiments we designed, and carry out targeted analysis of the experimental results. The specific experimental results are as follows.

4.1 Experimental setup

This experiment is run on the server based on Windows 10 system. Run Python3.7. The deep learning platform software used to compile the tests was configured as PyTorch and CUDA V1.0. Hardware configuration: GPU is RTX 1080 Ti. In this paper, 1000 images including insulators, shock hammers and spacers were collected to construct a dataset. These images are scaled to 227×227. This paper will choose four methods to compare with our model. These four methods are SSD, Faster R-CNN and YOLO-v4[17]. The SGD optimizer is used to train all models separately. Models are trained in a batch size of 10. The initial learning rate of the model is set to 0.01.

4.2 Qualitative experimental analysis

To verify that the proposed Detection of Abnormal Dangerous Behavior in Electrical Power Grids Based on the Fusion of Contextual Features model is a solution to the problem of low accuracy of target recognition in complex scenes. To evaluate the performance of the proposed algorithm, using more than 5000 images from the three datasets of Safety Helmet (Hardhat) Wearing Detect dataset, Hard Hat Workers dataset, Safety Helmet Detection dataset, the improved model can well identify the location and classification of people, heads, and helmets.

Figure 4 shows the results of the proposed method in this paper compared with the comparison method under the influence of different backgrounds and different occlusions. From the results, it can be seen that the algorithm proposed in this paper is able to identify the target accurately. Compared with other methods, the corresponding confidence level of the proposed method in this paper is optimal in all experiments. The experimental results show that the method proposed in this paper can accurately identify people, heads and helmets in the grid workspace, and can also identify the targets well in different backgrounds.

4.3 Quantitative experimental analysis

When evaluating the algorithm model through the qualitative experimental results, due to the different observation habits or observation angles of different human individuals, the evaluation criteria of quantitative experiments are too subjective. Therefore, we present the qualitative comparison experimental data of the proposed algorithm and the comparison algorithm in this section. This article will qualitatively analyze the object recognition models of SSD, Faster R-CNN, YOLO-v4 and Ours from the objective data level. This article introduces AP and mAP object recognition evaluation standards to analyze the object recognition results of each object recognition model from objective data. The objective data corresponding to the different scene simulation experiments is shown in Table 1.

The AP and mAP values for the different scene simulation experiments are shown in Table 1. From the data point of view, the algorithm in this paper is the best, and the AP is the highest in the Insulators, Shock Hammers and spacers

experiments. The average recognition accuracy is 94.85% compared with other models, and the average recognition accuracy is improved by 1.02%.
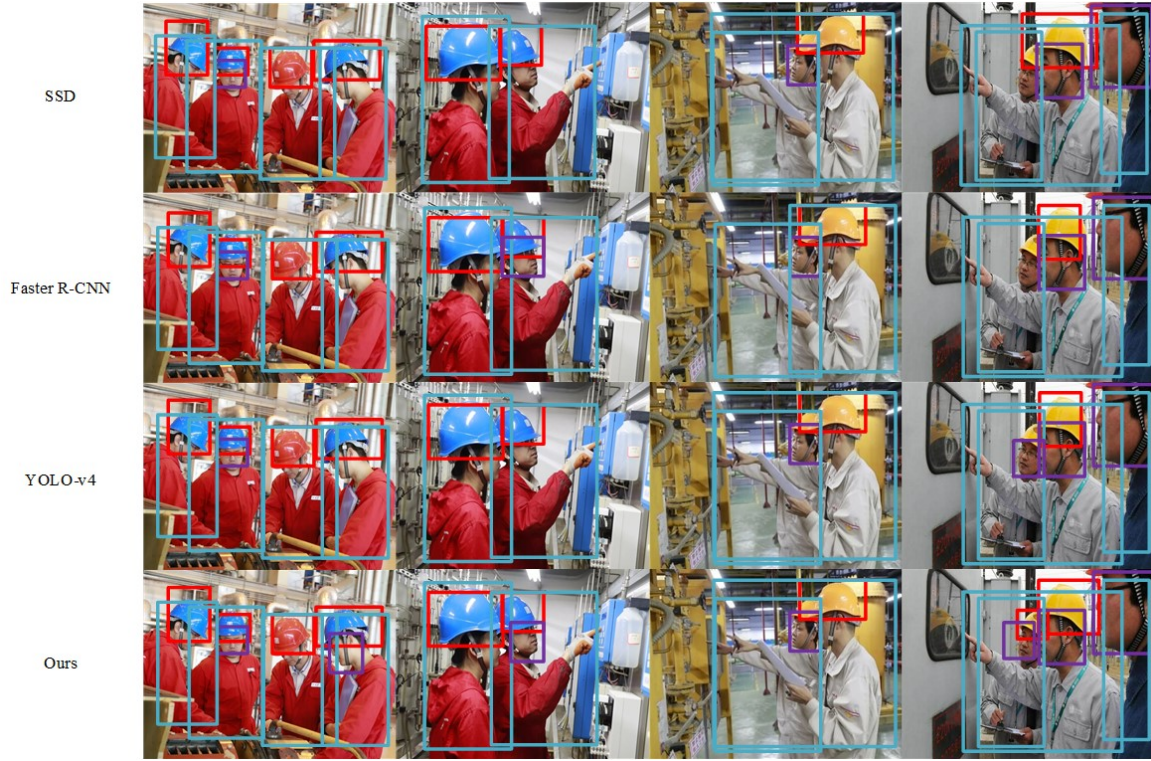

Figure 4. Simulation experiment of target recognition in power network operation room.

Table 1. The objective data corresponding to the different scene simulation experiments.

	AP			mAP
	People	head	helmet	
SSD	85.72	83.19	85.44	85.74
Faster R-CNN	89.24	85.56	87.34	88.26
YOLO-v4	93.85	90.27	93.16	93.72
Ours	95.01	91.49	93.88	94.85

5. CONCLUSION

In this paper, the electrical power grids based on the fusion of contextual features model is proposed. Firstly, the feature extraction backbone network is constructed to obtain image features. Second, the hierarchical contextual attention mechanism is constructed to capture contextual features. Finally, the target detection model with contextual feature fusion is constructed to achieve grid abnormal risk behavior detection. The feasibility of the proposed recognition model is verified by extensive simulation experiments. Regardless of qualitative or quantitative comparisons, the algorithm proposed in this paper has certain advantages. Although the proposed algorithm has some advantages, it cannot be applied in engineering. In future research, underwater object recognition in more complex environments will be further explored.

REFERENCES

[1] Vineetha, C. P. and Babu, C. A. "Smart grid challenges, issues and solutions," *2014 International Conference on Intelligent Green Building and Smart Grid (IGBSG),* 389-400(2014).

[2] Hare, S., Golodetz, S., Saffari, A., Vineet, V., Cheng, M. M., Hicks, S. L. and Torr, P. H., "Struck: Structured output tracking with kernels," *IEEE Transactions on Pattern Analysis & Machine Inteligence*, 38(10), 2096 -2109(2016).

[3] Xu, C. D., Zhao, X. R. and Jin, X., "Exploring categorical regularization for domain adaptive object detection," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 11721-11730(2020).

[4] Zhao, Y., Jia, J. and Liu, D., "He-yolo: Aerial target detection based on improved YOLOv3," *International Journal of Pattern Recognition and Artificial Intelligence*, 35(13), 583-596(2021).

[5] Joseph, R., Santosh, D., Ross, G. and Ali, F., "You only look once: Unified, real-time object detection," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 779-788(2016).

[6] Liu, W., Dragomir, A., Dumitru, and Christian, S., "SSD: Single shot multibox detector," *Proceedings of European Conference on Computer Visio*, 21-37(2016).

[7] Ross, G., Jeff, D., Trevor, D. and Jitendra, M., "Rich feature hierarchies for accurate object detection and semantic segmentation," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 281-290(2014).

[8] Girshick, R., "Fast R-CNN," *IEEE International Conference on Computer Vision*, 2380-7504(2015).

[9] Ren, S. Q., He, K. M. and Girshick, R., "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6), 1147-1149(2017).

[10] Dutta, T., Sil, J. and Chottopadhyay, P., "Condition monitoring of electrical equipment using thermal image processing," *IEEE First International Conference on Control, Measurement and Instrumentation*, 311-315(2016).

[11] Huda, A., Nazmul, A. S. and Taib, S., Application of infrared thermography for predictive/preventive maintenance of thermal defect in electrical equipment. *Applied Thermal Engineering*. 61(2):220-227 (2013).

[12] Heano, H., Capolino, G. A. and Femandez-Cabanas, M., "Trends in fault diagnosis for electrical machines: A review of diagnostic technique," *Industrial Electronics Magazine,* 8(2), 31-42(2014).

[13] Kato, K., Li, Y. and Abhinav, G., "Compositional learning for human object interaction," *Proceedings of European Conference on Computer Vision*, 234-251(2018).

[14] Xiong, C., Li, W. and Liu, Y., "Multi-dimensional edge features graph neural network on few-shot image classification," *IEEE Signal Processing Letters*, 28(99), 573-577(2021).

[15] Chen, Z. M., Jin, X. and Zhao, B., "Hierarchical context embedding for region-based object detection," *IEEE Transactions On Image Processing,* 30(9), 6917-6929(2020)

[16] Mi, J., Tang, S., Deng, Z. and Goerner M., "Object affordance based multimodal fusion for natural human-robot interaction," *Cognitive Systems Research*, 54(5), 128-137(2019).

[17] Cai, Y., Luan, T. and Gao, H., "YOLOv4-5D: An effective and efficient object detector for autonomous driving," *IEEE Transactions on Instrumentation and Measurement,* 70(99), 93-105(2020).