# Type Identification of Substation Equipment Based on Transfer Learning and R-FCN

Xiaoyong Mao [a], Qiang Gao [b]

Wenzhou Polytechnic, Wenzhou 325035, China;

[a]2822708830@qq.com, [b]2020000119@wzpt.edu.cn,

## Abstract

**To address the challenges of diverse equipment types, complex backgrounds, and high annotation costs in substation equipment inspection images, this paper proposes a substation equipment type recognition method based on transfer learning and Region-based Fully Convolutional Networks (R-FCN). Firstly, a ResNet-101-based feature extraction network is employed, utilizing ImageNet pre-trained weights for transfer learning to mitigate overfitting caused by insufficient substation equipment samples. Secondly, a position-sensitive score maps mechanism is introduced to achieve end-to-end equipment localization and classification through a fully convolutional architecture. Finally, a multi-task loss function is designed to optimize network parameters. Experimental results on a self-constructed substation equipment dataset demonstrate that the proposed method achieves a mean Average Precision (mAP) of 91.32%, representing improvements of 2% and 1.4% over the SSD algorithm and R-FCN, respectively. The method also exhibits favorable detection speed, satisfying the real-time requirements for intelligent substation inspection.**

## Keywords

**Substation equipment; object detection; deep learning; intelligent inspection.**

## 1. Introduction

As the critical hub of the power system, the substation's safe and stable operation is directly related to grid reliability. With the advancement of smart grid construction, machine vision-based intelligent inspection technologies for substations have been widely deployed. However, the diversity of substation equipment types—including circuit breakers, disconnectors, current transformers, voltage transformers, surge arresters, and transformers—coupled with significant variations in appearance and complex backgrounds in inspection images, renders conventional image processing methods inadequate for efficient and accurate equipment identification.

In recent years, deep learning technologies have achieved breakthrough progress in the field of object detection. Region proposal-based two-stage detection algorithms, such as R-CNN, Fast R-CNN, Faster R-CNN, and R-FCN, demonstrate high detection accuracy and exhibit promising performance in power equipment recognition. Among these, Region-based Fully Convolutional Networks (R-FCN) significantly improve detection speed while maintaining high precision through the position-sensitive score maps mechanism[1]. Nevertheless, deep neural networks typically require substantial annotated samples for training. The high cost of acquiring substation equipment images and the difficulty of professional annotation often result in insufficient samples, leading to model overfitting. Transfer learning addresses this challenge by migrating knowledge learned from source domains (e.g., ImageNet) to target domains (substation equipment detection), thereby enabling effective model training under small-sample conditions. Existing research predominantly focuses on the application of Faster R-CNN and its variants in power equipment detection, whereas investigations into performance optimization of R-FCN for substation equipment type recognition remain relatively scarce.

To address the aforementioned issues, this paper proposes a substation equipment type recognition method based on transfer learning and R-FCN[2-3]. The main contributions are summarized as follows: (1) construction of an R-FCN detection framework based on ResNet-101, employing ImageNet pre-trained weights for transfer learning; (2) design of feature extraction strategies tailored for small-target detection of substation equipment; and (3) validation of the proposed method's effectiveness on a self-constructed dataset, with comparative analysis against mainstream algorithms.

## 2. R-FCN Algorithm Analysis

The R-FCN algorithm framework primarily comprises three core components: a backbone network, a Region Proposal Network (RPN), and an RoI subnetwork. To achieve the objectives of this section, the specific implementation procedure is elaborated as follows.

Firstly, the acquired infrared images of power equipment are fed into the network as input. The entire infrared image is processed through a convolutional neural network to extract comprehensive convolutional feature maps. The utilization of a deeper backbone network (ResNet-101) enables the extraction of more profound and abstract image features during this process, thereby enhancing recognition accuracy. Subsequently, these feature maps are propagated to the RPN to generate anchors, which are then labeled as foreground or background. High-scoring foreground regions are selected as Region of Interests (RoIs), which are forwarded to the subsequent RoI subnetwork for further training. Each infrared image of power equipment generates 300 proposed windows.

Concurrently, the feature maps from fully convolutional layers are convolved with multi-layer convolutional kernels to generate position-sensitive score maps[4]. The RoIs and score maps are then fed into the subsequent softmax layer for position-aware voting. Classification is performed through the softmax layer, and the RoI with the highest score is ultimately identified as the located and recognized object position and category. The flowchart of the R-FCN algorithm framework is illustrated in Fig. 1.

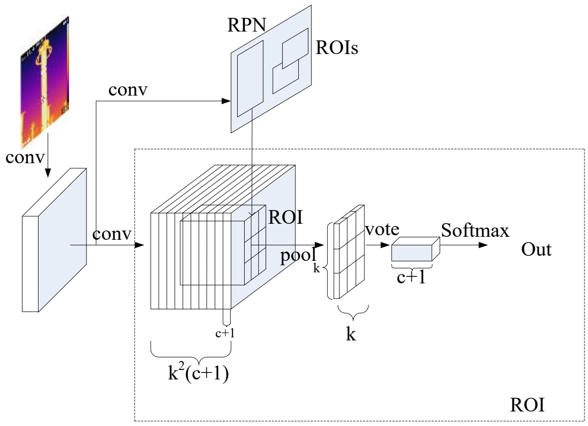

Fig 1 R-FCN algorithm flow chart

The operational principle of the Region Proposal Network (RPN) is fundamentally identical to that employed in the classical Faster R-CNN algorithm. The RPN receives convolutional feature maps as input and generates nine rectangular bounding boxes of varying scales at each spatial position based on the anchor mechanism.During RPN network training, anchors are compared against manually annotated ground-truth regions in the dataset. The anchor exhibiting the maximum overlap ratio with a ground-truth box is labeled as foreground, and anchors with overlap ratios exceeding 0.7 are additionally designated as foreground samples[5]. Conversely, anchors with overlap ratios below 0.3 are labeled as background samples. Positive and negative anchor samples are subsequently selected according to a predetermined proportion. The top 300 RoIs with highest scores are filtered using the Non-Maximum Suppression (NMS) algorithm and related techniques, and these preliminarily screened candidate boxes are propagated to the subsequent RoI subnetwork for further processing. The adoption of the Online Hard Example Mining (OHEM) method within the RPN network can substantially enhance the accuracy of preliminary RoI screening.

## 3. Design of Improved R-FCN Algorithm for Image Recognition

When the depth of convolutional neural networks in deep learning reaches a certain threshold, the problems of gradient vanishing and gradient explosion frequently emerge during training. To address these issues, Kaiming He from Microsoft Research proposed the concept of Residual Network (ResNet), which distinguished itself in the 2015 ImageNet competition[6]. As illustrated in Fig. 2, the residual unit constitutes the fundamental building block of residual networks. It is precisely upon this architectural foundation that the performance of deep networks can be substantially enhanced.

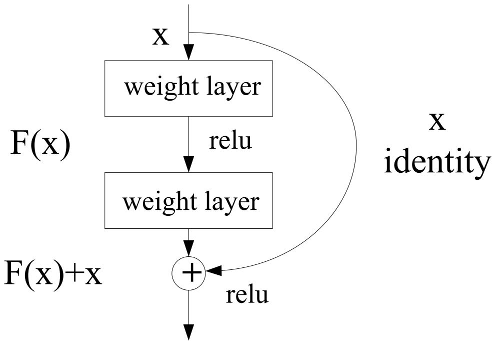

Fig.2 Residual unit

As illustrated in Fig.2, the residual unit essentially approximates the desired mapping residual through these stacked layers. The target mapping to be learned is denoted as $H(x)$, the residual mapping function of the network is represented as $F(x)$ , and the identity mapping is given by $X$ . The concept of residual refers to the discrepancy between the observed value $H(x)$ and the estimated value $X$ . The advantage of the ResNet architecture lies in its utilization of stacked layers to fit $F(x)=H(x)-x$, thereby obtaining the mapping $H(x)=F(x)+x$ . The merit of this formulation is that if the model has already converged to an optimal state, one simply needs to set $F(x)=0$ , which yields $H(x)=x$ without necessitating further parameter updates, thereby circumventing the problems of gradient vanishing and gradient explosion.

$$y_k = h(x_k) + F(x_k, W_k) \tag{1}$$

$$x_{k+1} = \mathrm{f}(y_k) \tag{2}$$

$$x_K = \mathrm{x}_k + \sum_{i-1}^{K-1} F(x_i, W_i) \tag{3}$$

Within the Region Proposal Network, a substantial number of rectangular bounding boxes are generated, among which hundreds to thousands of regions participate in the training for target category and location prediction. The anchor exhibiting the maximum overlap ratio with the ground-truth region is designated as a foreground sample. For the remaining anchors, those with overlap ratios exceeding 0.7 with any ground-truth annotation are additionally labeled as foreground samples, whereas those with overlap ratios below 0.3 with all ground-truth annotations are designated as background samples.

In the dataset constructed for this study, power equipment is typically positioned centrally with relatively small spatial occupancy, resulting in a disproportionately large ratio of background regions to target regions and consequently inducing sample imbalance. Furthermore, the identification of specific power equipment categories encompasses hard examples that are particularly challenging to recognize.

The OHEM methodology posits that the majority of background regions and easily identifiable target regions exhibit high prediction accuracy regarding their categories, yielding relatively small loss values. During training, when region proposals with substantial loss values emerge, these hard examples can be selectively retrained and reclassified, thereby addressing the class imbalance between positive and negative samples and enhancing the recognition accuracy of the power equipment infrared image detection model[7-8]. As illustrated in Fig. 3, hard example training through parameter sharing with preceding training stages enables computational efficiency and accuracy improvement.

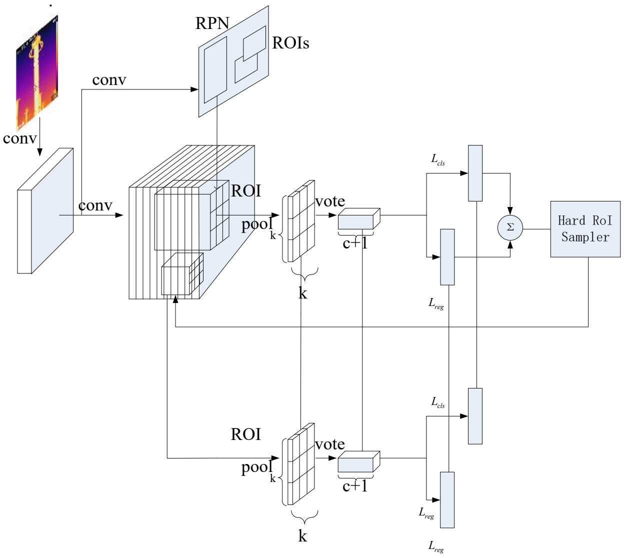

Fig 3 OHEM method architecture diagram

## 4. Experimental results and Conclusion

The experiments conducted in this study utilize a Dell tower workstation as the hardware platform, with the Ubuntu 16.04 operating system. The programming environment is based on Python, equipped with 32 GB of system memory. The graphics processing unit employed is the NVIDIA GeForce GTX 1080 Ti, featuring 11 GB of video memory with a memory clock rate of 11 Gbps. The deep learning framework is implemented using Caffe. The specific software versions for the deep learning development environment are configured as follows: CUDA 9.0, cuDNN 7.0.5, and Python 2.7. Additionally, essential third-party Python libraries are installed, including matplotlib, opencv, numpy, and easydict.

The raw data acquisition process commences with the collection of infrared images of power equipment, followed by rigorous screening to ensure compliance with quality requirements. Data augmentation techniques, specifically horizontal flipping and mirroring operations, are applied to the images to expand the sample size. These augmentation strategies contribute to improved training model accuracy and enhanced robustness of the resulting model.

Target annotation of the images is performed using the LabelImg software, whereby target power equipment within the infrared images are manually annotated in a sequential manner, as illustrated in Fig. 2-8. The detection targets are categorized into four classes: surge arresters, circuit breakers, current transformers, and voltage transformers. For subsequent research purposes, surge arresters are labeled as class 1, circuit breakers as class 2, current transformers as class 3, and voltage transformers as class 4.

Following annotation with LabelImg, corresponding .xml files are generated that maintain one-to-one correspondence with the images, as depicted in Fig. 4 The data annotation process is both time-consuming and labor-intensive, with manual annotation being susceptible to erroneous or missed labels. Furthermore, the dimensions of the annotation bounding boxes directly influence the localization accuracy of the regression boxes; consequently, particular attention is required throughout the annotation procedure to ensure precision and consistency.

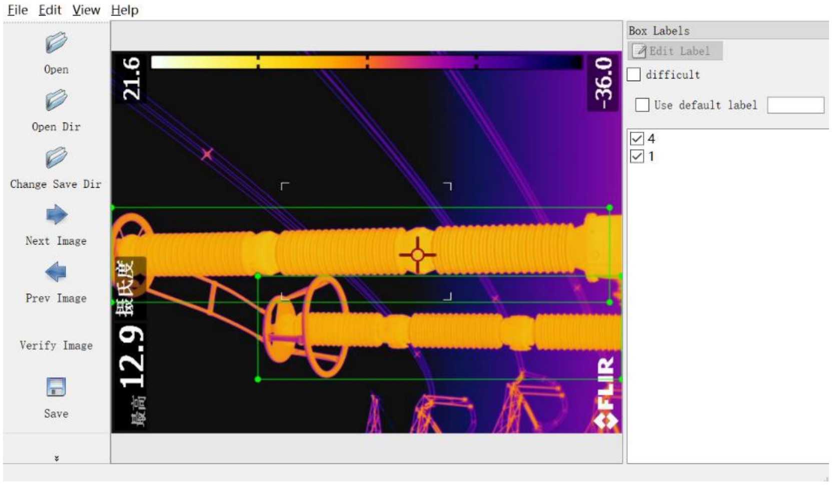

Fig.4 Label the original data with software labelimg

Currently, multiple deep learning object detection models are available for selection, including Faster R-CNN, R-FCN, SSD, and RON. Among these, Faster R-CNN represents the most classical architecture. Both R-FCN and RON constitute improvements based upon the structural foundation of Faster R-CNN. The SSD model similarly incorporates partial structural components from Faster R-CNN. For identical power equipment infrared image datasets, the recognition speed and accuracy of each model on the local GPU are presented in Table 1 below.

Table 1 Algorithm Comparison

| Algorithm | Detection Speed (fps) | mAP （%） |
|---|---|---|
| integrates R-FCN | 51 | 91.32 |
| R-FCN | 68 | 90.18 |
| SSD | 22 | 90.23 |
| RON | 42 | 85.62 |

As illustrated in Table 1, the proposed model algorithm, which integrates R-FCN with the deeper ResNet-101 architecture and the OHEM methodology, achieves superior recognition accuracy for power equipment infrared images compared with conventional target recognition algorithms. Specifically, the power equipment recognition accuracy attains 91.32%, demonstrating significant improvement over baseline approaches.

## Acknowledgements

This research was supported by the Zhejiang Provincial College Students' Science and Technology Innovation Activity Program (Project No.: 2024R470A007).

## References

[1] Liu W , Anguelov D , Erhan D , et al. SSD: Single Shot MultiBox Detector[C]// European Conference on Computer Vision. Springer International Publishing, 2016.

[2] Kong T, Sun F, Yao A, et al. RON: Reverse Connection with Objectness Prior Networks for Object Detection[J]. 2017.

[3] He K , Zhang X , Ren S , et al. Deep Residual Learning for Image Recognition[J]. 2015.

[4] Lin T Y , Dollár, Piotr, Girshick R , et al. Feature Pyramid Networks for Object Detection[J]. 2016.

[5] TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems[J]. 2016.

[6] Kothari DP,Nagrath I J.Power system engineering[M]. Beijing: Tsinghua University Press， 2009

[7] Hinton G, Deng L, Yu D, et al. Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups[J]. IEEE Signal Processing Magazine, 2012, 29(6):82-97.

[8] Aldrich G, Auret L. Unsupervised process monitoring and fault diagnosis with machine learning methods[M].Spring, 2013.