Deep Learning Based External-force-damage Detection for Power Transmission Line

Junyi Liu*, Hongbing Huang, Yi Zhang, Jia Lou and Jiale He
State Grid Zhejiang Electric Power Co., LTD, China
*E-mail: shshzaa@163.com

Abstract. With the accelerating pace of modernization, engineering construction has frequently threatened the safety of the power transmission lines. The external-force-damage has gradually become a major factor affecting the safety of transmission lines. To overcome the shortcomings of traditional detection techniques, this paper proposes a deep learning based external-force-damage detection method for transmission lines. First, thousands of surveillance images are collected and annotated with bounding boxes for construction objects. Second, a deep learning based object detection model is trained to localize the construction objects. The conducted experiments show that the proposed method can detect the construction objects in surveillance images accurately and is robust to illumination and climate variations.

1. Introduction
With the rapid development of national economy, the demand for electricity in all walks of life grows more vigorous and the layout of power transmission lines becomes denser. However, the increasing engineering construction of railways, highways, quarries, mines and other buildings has inevitably caused a large number of mechanical construction in the protection zone of transmission line. Deconstruction by external forces, such as tower crane hitting the transmission line, would cause large-scale power outage which not only threaten the safety of personal life but also increase the difficulty of maintenance for the power supply department. A lot of manpower and material resources may be spent for the power supply department to patrol the transmission line in order to eliminate potential dangers.

Currently, the common power transmission line inspection methods can be divided into three categories: manual inspection, UAV inspection and video surveillance. Manual inspection requires a lot of manpower and material resources, which is inefficient and cannot adapt to the grid with increasing scale. UAV inspection does not need much human resources, but the endurance of UAV may restrict the inspection scope and we cannot monitor the transmission line for a long time. Video surveillance is a more mature inspection method in recent years. The surveillance camera installed on the power transmission tower can capture the images of surrounding environment in real time and transmit them back to the monitoring center. The staff can judge whether there is a safety hazard on the transmission line by checking the images and thus greatly reduces the manpower and material resources.

Although video surveillance can avoid the inconvenience in manual and UAV inspection methods, it’s still far from practical application as the amount of surveillance data is too large to process if only by human observation. In recent years, the research community has brought up some video analysis algorithms towards automatic external-force-damage detection. Zhang [1] et al. proposed an intelligent
early warning system for preventing external-force-damage. They separated the moving objects by combining the Gaussian mixture model with background subtraction algorithms and localized them with minimum circumscribed bounding boxes. Du [2] et al. designed an early warning software that can perform large mechanical intrusion tracking, mounds of trash detection, extra high tree warning area detection and illegal buildings recognition. Ding [3] et al. proposed a detection system of potential accidents around transmission lines. They detect crane and tower crane in the sky with background modelling method, identify fire by considering colour and texture information and localize engineering vehicles with convolution neural networks. Chen [4] et al. combined moving object segmentation methods and support vector machine (SVM) to recognize crane, they track the lifting arm with Kalman filter and detect its angle with Hough transform so as to report different levels of alarm signals. He [5] et al. proposed a warning system to handle transmission line ice cover, disconnection, engineering vehicle intrusion and other abnormalities in surveillance by combining object extraction algorithm based on adaptive threshold and identification methods based on subtraction and texture characteristics.

Although all the methods mentioned above realize the basic early warning functionality against external-force-damage, most of them still use the traditional image processing and machine learning algorithms which need carefully tuning of the model parameters in order to cope with complex illumination and climate conditions, as well as varied transmission line environments. On the other hand, deep learning algorithms, especially the convolutional neural network (CNN), gradually dominates most of the computer vision and image processing tasks due to the increasing amount of data and computing resources. A typical characteristic of CNNs is that they can learn the feature of objects or images from large amount of data and are more robust to varying scenes and illumination. Currently, the transmission line surveillance systems can produce massive image data every day, making full use of these data to train a CNN-based method for external-force-damage detection becomes applicable.

In this paper, we propose a deep learning based external-force-damage detection method for transmission line. First, we collect about 5700 surveillance images from 223 scenes of transmission line and generate our dataset by eliminating the similar images and annotating the construction objects. Then, we train a Faster RCNN [6] model to detect the construction objects. The results on test dataset shows that our model can detect the targets on transmission line surveillance images effectively and is robust to illumination and climate changes.

2. Constructing dataset

Since the number of surveillance data is too large and images from the same scene usually look similar. We first eliminate the similar images from the same scene to avoid overfitting during model training phase. We collect around 5700 surveillance images of transmission line and divide them into 223 scenes. For each surveillance scene, images are captured from different seasons in a year as well as different time in a day in order to make the detection model learn robust feature against illumination and climate changes. After eliminating the similar ones, we finally generate 2312 surveillance images of transmission line.

We categorize the construction objects that may cause external-force-damage into three classes: crane, tower crane and excavator (we merge bulldozer and excavator into one class due to limited number of samples). An example the annotated surveillance images is illustrated in figure 1.
Figure 1. Example annotations of three construction object classes in transmission line surveillance images

3. Deep learning based detection model

Traditional object detection algorithms rely on hand-crafted features which are vulnerable to the deformation of objects and illumination changes. In recent years, with the growing number of available training data, CNN-based object detection methods have surpassed the traditional methods by a large margin. In our paper, since we have already annotated enough training data, a CNN-based object detector is trained to localize the construction objects in transmission line surveillance images. We choose Faster RCNN with VGG16 [7] for feature extraction. Each module is introduced in the following in detail.

3.1. Feature extraction based on VGG16 network

Simonyan [7] et al. proposed the VGG model and won the ImageNet [8] classification and localization challenges in 2014. VGG network extracts image feature with shallow convolution and pool operations, then fully-connected layers are utilized for image classification. VGG model is the first to replace the large kernel convolution with several stacked small kernel convolutions, which not only reduces the network parameters but also makes it possible to stack more layers to get deeper and better classification networks. The network configurations of VGG model with different depth are illustrated in figure 2. To make a good trade-off between the classification performance and the amount of data for training deep neural networks, we select VGG16 as our feature extraction module for construction object detection model.
3.2. Construction object detection with Faster RCNN

Faster RCNN is a general object detection method proposed by Ren [6] et.al in 2015. It first extracts the image feature with a backbone network, such as the VGG network introduced in section 3.1, then the region proposal network (RPN) predicts region proposals that may contain construction objects. Finally, the region features are generated by RoI pooling and fed to classification and bounding box regression. The overall pipeline is shown in figure 3.

Figure 2. Network configuration of VGG model with different depth

![ConvNet Configuration Table]

| ConvNet Configuration |
|-----------------------|
| **A** | **A-LRN** | **B** | **C** | **D** | **E** |
| 11 weight layers | 11 weight layers | 13 weight layers | 16 weight layers | 19 weight layers |
| conv3-64 | conv3-64 | conv3-64 | conv3-64 | conv3-64 |
| **LRN** | **conv3-64** | conv3-64 | conv3-64 | conv3-64 |
| conv3-128 | conv3-128 | conv3-128 | conv3-128 | conv3-128 | conv3-128 | conv3-128 | conv3-128 | conv3-128 |
| maxpool | maxpool | maxpool | maxpool | maxpool | maxpool |
| conv3-256 | conv3-256 | conv3-256 | conv3-256 | conv3-256 | conv3-256 |
| conv3-256 | conv3-256 | conv3-256 | conv3-256 | conv3-256 |
| maxpool | maxpool | maxpool | maxpool | maxpool |
| conv3-512 | conv3-512 | conv3-512 | conv3-512 | conv3-512 | conv3-512 | conv3-512 | conv3-512 |
| conv3-512 | conv3-512 | conv3-512 | conv3-512 | conv3-512 | conv3-512 | conv3-512 | conv3-512 |
| maxpool | maxpool | maxpool | maxpool | maxpool |
| FC-4096 | FC-4096 | FC-1000 |
| soft-max | |

Figure 3. Illustration of the Faster RCNN architecture
3.2.1. Region proposal network. Region proposal network (RPN) is designed for generating proposals that may contain target objects. Compared to the traditional proposal generation methods, such as EdgeBox [9], RPN could reach higher recall with less proposals, which makes it much more efficient. For an input image, RPN first generates evenly distributed anchor points on the image, each corresponding to several anchor boxes that provide the scale prior of target objects. For each anchor box, RPN predicts its objectiveness, i.e. the probability of containing target objects and the offset of object proposal with regard to this anchor box.

In order to make full use of the training data, we apply K-means clustering algorithm on the training ground-truth objects to generate the anchor boxes. The detailed flow diagram of K-means clustering is illustrated in figure 4.

3.2.2. RoI Pooling module. RoI Pooling module is designed to generate a fixed dimensional feature representation for each region proposal. Specifically, given the input feature map $x$ and a region proposal of size $w \times h$, the RoI pooling module first divides the region into $k \times k$ bins, and the pooled feature $y$ at $(i, j)$-th bin can be represented by

$$y(i, j) = \sum_{p \in A(i, j)} x(p_0 + p) / n_{ij}$$

where $p_0$ is the top-left corner of the $(i, j)$-th bin and $n_{ij}$ is the number of points in it.

After RoI pooling, the input region features of varying size are fixed to $k \times k$, which is then fed into the subsequent region classification and regression module.

3.2.3. Region classification and regression module. Region classification and regression module are responsible to predict the classification probability distribution as well as refine the bounding box for each region proposal. Suppose $p = (p_0, \ldots, p_C)$ represents the predicted probability distribution over $(C+1)$ classes including background (C=3 in our paper), let $u$ denotes the ground-truth class, then the classification loss is computed by

$$L_{cls}(p, u) = -\sum_{i=0}^{C} p(i) \delta_{ij}$$

where $\delta_{ij} = 1$ when $i = u$ and 0 otherwise.
Suppose \((v_x, v_y, v_w, v_h)\) represents the true offset between ground-truth bounding box and region proposal, let \((t_{x_0}^u, t_{y_0}^u, t_{w_0}^u, t_{h_0}^u)\) denotes the predicted offset for class \(u\), then the regression loss for class \(u\) can be written as

\[
L_{loc}(t^u, v) = \sum_{i \in \{x, y, w, h\}} \text{smooth}_{L1}(t^u_i - v_i)
\]

in which

\[
\text{smooth}_{L1}(x) = \begin{cases} 
0.5x^2 & \text{if } |x| < 1 \\
|x| - 0.5 & \text{otherwise}
\end{cases}
\]

Thus the overall multi-task loss for a training RoI would be

\[
L(p, u, t^u, v) = L_{cls}(p, u) + \lambda[u \geq 1]L_{loc}(t^u, v)
\]

where the Iverson bracket indicator function \([u \geq 1] = 1\) when \(u \geq 1\) and 0 otherwise.

### 4. Experiment

#### 4.1. Experiment setting

We divide the annotated 2312 surveillance images into two splits, with 1734 images for training and the rest for testing.

Our model is implemented based on Faster R-CNN in Tensorflow [10]. The input images are resized to 1000x750 before fed into the network. SGD with mini-batch is used for training. The batch size is set to 1, the momentum is 0.9 and the weight decay is 0.0005. All the convolutional layers are initialized with weights trained on the ImageNet dataset, while the rest of layers are initialized with Gaussian distribution.

We conduct our experiments on GeForce GTX 1080Ti. We train our model for 70k iterations, with learning rate \(10^{-3}\) for the first 50k iterations and \(10^{-4}\) for the rest.

#### 4.2. Experiment results

The overall loss in training phase is shown in figure 5.

![Figure 5. loss curve of the model during training phase](image)

From figure 5 we can see that the training loss decreases gradually and stabilizes at last, which indicates that our model has been effectively trained.

After training, we evaluate our model on the test dataset in terms of mean Average Precision (mAP), the same as the standard PASCAL VOC protocol [11], where a detection is considered to be true positive if its IoU (intersection over union) with a ground-truth box is higher than 0.5. The evaluation results are shown in Table 1. As we can see, our deep learning based construction object detection method can reach 90% mAP on the test dataset, which demonstrates the effectiveness of our method. The qualitative results are illustrated in figure 6.
Table 1. Detection AP(%) of three construction object classes

| Crane (%) | Tower crane (%) | Excavator (%) | mAP (%) |
|-----------|-----------------|---------------|---------|
| 91.5      | 90.1            | 88.7          | 90.1    |

Figure 6. Qualitative detection results in the test dataset

From figure 6 we can see that our model could localize the construction objects in transmission line surveillance images accurately. The first row demonstrates the detection results in normal scenes, the second illustrates results on snowy scenes and the last row shows the detection results on foggy day and during dawn time. We can see can our model are robust to the changes of illumination and climate.
5. Conclusion
In this paper we have proposed a deep learning based external-force-damage detection method for transmission line, which can be applied to the video surveillance scenario and save the manpower as well as material resources. We first collect thousands of transmission line surveillance images and annotate the construction objects, then we train a Faster R-CNN based method to localize the construction objects in images. Experiment results on test dataset demonstrates that our model could localize the construction objects accurately and is robust to the changes of illumination and climate.

References
[1] Y. Zhang, X. Huang, X. Chen, et al. Intelligent Early-warning System for Preserving Transmission Line from External Damage[J]. High Voltage Apparatus, 2015, 51(08): 54-61
[2] J. Du. Design and Research on Warning System of Transmission Line under External Force Damage[D]. North China Electric Power University, 2015
[3] J. Ding. Detection System of Potential Accidents around Transmission Line Based on Image Damage[D]. University of Jinan, 2017
[4] W. Chen. Design and Implementation of Large Mechanical Intrusion Video Surveillance Alarm System of Transmission Line[D]. Shanghai Dianji University, 2015
[5] C. He. Research on High Voltage Transmission Line Video Surveillance Technology[D]. Beijing Jiaotong University, 2012
[6] S. Ren, K. He, Girshick R, et al. Faster r-cnn: Towards real-time object detection with region proposal networks[C]//Advances in neural information processing systems. 2015: 91-99.
[7] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv:1409.1556, 2014.
[8] J. Deng, W. Dong, Socher R, et al. Imagenet: A large-scale hierarchical image database[C]//Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. Ieee, 2009: 248-255.
[9] Zitnick C L, Dollár P. Edge boxes: Locating object proposals from edges[C]//European conference on computer vision. Cham: Springer, Cham, 2014: 391-405.
[10] Abadi M, Barham P, J. Chen, et al. TensorFlow: A System for Large-Scale Machine Learning[C]//OSDI. 2016, 16: 265-283
[11] Everingham M, Van Gool L, Williams C K I, et al. The pascal visual object classes (voc) challenge[J]. International journal of computer vision, 2010, 88(2): 303-338.