Electric Tower Target Identification Based on High-resolution SAR Image and Deep Learning

Guishen Tian¹, Song Meng¹, Xuejiao Bai¹, lijuan Liu¹, Yang Zhi², Wenhao Ou², Xiangze Fei², Yuanpeng Tan²

¹State Grid Inner Mongolia Eastern Electric Power Co., Ltd., Hohhot, China
²China Electric Power Research Institute Co., Ltd, Beijing, China

*Corresponding author’s e-mail: tianguishen@md.sgcc.com.cn

Abstract. With regard to the electric tower identification problem in the high-resolution SAR image under complicated background, it is difficult for the existing target detection algorithm to reach balance in identification precision and efficiency. Therefore, this paper combines the advantages of high YOLOv2 calculation efficiency and high VGG classification precision, and proposes a kind of two-stage target detection algorithm of YOLOv2 and VGG cascade connection by virtue of transfer learning. At the Stage-1, the sliding window and the non-maximum suppression algorithms are combined, and the YOLOv2 is utilized to conduct rapid electric power detection to the whole SAR image, thus obtaining the target detection result of high recall rate; at the Stage-2, the VGG classification model is utilized to conduct secondary classification to the target and background with regard to the target detection result at the Stage-1, thus further eliminating the false positive. This algorithm can be used to enhance the accuracy of only using YOLO v2 for target identification, thus reducing the false alarm rate of model effectively. The electric towers in the mountainous belt and the plain belt are used as samples for training, thus enhancing the robustness of algorithm. After conducting algorithm testing to the COSMO image in Zhijiang City, Hubei Province, the result shows that the electric tower recall rate can reach 73.8%. The method adopted in this paper can identify the electric tower target of the whole SAR image more accurately through model transfer and two-stage deep learning, thus being further promoted into actual application scene.

1. Introduction
The Synthetic Aperture Radar image mainly reflects the microwave scattering characteristic of target, and is able to realize all-weather and all-time observation to the ground, without being influenced by climatic conditions such as clod, rain and fog, and has unique application advantages in case of being compared with optical imaging. SAR mainly acquires target information through received echo, and due to natural difference among targets in size, nature, etc., there is relatively large difference[1] among different targets in the back scattering characteristics. The electric tower belongs to man-made metal structure, and it has extremely high dielectric coefficient, so the reflection echo signal is relatively strong, while it shows characteristics [2-3] such as high pixel value and high pixel density in SAR image on the SAR image. Meanwhile, owing to the difference in shape, height and structural characteristic of electric towers as well as the complicated environment around electric towers (e.g., some electric towers are built on a wide plain, while some are on steep mountainous lands), the shape of electric towers on plain is eye-catching, while the shape of those on mountainous land is seriously affected by background and relatively fuzzy, which makes it difficult to identify electrical towers [4-5].
The target detection and identification targeting SAR image is a hot spot of research; for this purpose, Finn proposed the constant false alarm rate (CFAR) based on the Bayesian theory, and utilized it to realize radar signal detection [6-7] in Gaussian background noise, and this method is the most extensive and deep method researched at present. However, when the background environment where the electric tower targets are located is relatively complicated, it will be difficult to extract targets by using fixed threshold value, and the threshold value will be determined via self-adaptation; the GLRT target detection algorithm is proposed based on the CFAR algorithm[8-9], and this algorithm considers statistical characteristics of target, and it is a kind of suboptimal statistical detection method. However, for there are differences in shape, size and direction of actual targets, it is relatively difficult to establish a uniform target distribution model, so it hasn’t been widely applied; Kaplan proposed a kind of target detection algorithm [10] based on extended feature (EF), and this algorithm represents the textural features under different scales through calculating the multi-scale Hurst index in the position of calculation point, and the target detection is conducted based on textural roughness. This algorithm has relatively good detection performance, while it can detect the part similar to the target shape while the gray value is relatively low, with relatively high false rate. In recent years, the deep learning algorithm with the convolutional neural network (CNN) as representative shows outstanding advantages day by day in the aspect of extraction of remote sensing image information [11-12]. As a kind of end-to-end training and forecast algorithm frame, the convolutional neural network can directly use the original pixel of image as input, which reduces the requirement for data pre-processing and avoids complicated characteristic engineering, and compared with artificial design characteristics, it can conduct more radical description to original data. This paper conducts electric tower extraction of SAR image based on deep learning, and is able to extract characteristic information of different scales and different levels of electric towers, and this is the main basis for precise identification and exact positioning to electric tower, and in this way, it can obtain the identification precision higher than that of traditional method.

At present, the target detection algorithm [13] based on deep learning mainly includes algorithms based on regional nomination such as Faster-RCNN and algorithms based on regression such as YOLO; Faster-RCNN [14] transfers the target detection problem into positioning and classification problems. Firstly, the foreground and background are separated via regional nomination; and then, the bounding box regression and the secondary classification are adopted to realize target detection. This method enhances the accuracy of target detection, but it spends a relatively long time during the process of extracting candidate region. The YOLO [15-16] algorithm has no explicit regional nomination process, and its network structure is simple, with rapid detection speed. For the data amount of remote sensing image is huge, and in case of adopting algorithms such as Faster-RCNN, the forward-direction duration to the image of a scene will be much longer than that of YOLO algorithm. Therefore, this paper adopts YOLO v2 [17] as the target detection model, so as to shorten the forward-direction calculation duration. In addition, in order to further reduce the false alarm rate of YOLOv2 model, this paper adds the VGG classification [18] model after YOLO, thus constituting a complete Two-Stage Electric Tower Identification Model.

2. Research Method
This paper will propose a Two-Stage electric power identification model, and through model cascade connection between YOLOv2 and VGG, the advantages including that the high calculation efficiency of YOLOv2 and high classification precision of VGG can be combined fully. As compared with target identification by only utilizing YOLOv2, this method can effectively reduce the false alarm rate of model. The technical route of this Paper is as shown in Figure 1.

---

2
In detail, the technical route in this paper includes the following three parts:

1) **Sample Fabrication.** To obtain the electric tower detection model with stronger robustness, it is necessary to firstly utilize multi-source SAR image to fabricate the training sample, thus enhancing the representativeness of training sample. In this paper, the high-resolution SAR images such as COSMO, RADARSAT-2 and GF-3 are selected for sample fabrication, so as to ensure the samples can cover different terrain conditions and different imaging conditions.

2) **Stage-1 Model Construction.** Select YOLOv2 as the Stage-1 model, and conduct transfer learning based on the weight of VOC2012 pre-training; and then conduct electric tower detection on the whole SAR image based on the sliding window and non-maximum suppression (NMS) algorithms.

3) **Stage-2 Model Construction.** Classify the forecast results of Stage-1 model into correct/wrong, and fabricate Stage-2 training samples; select VGG as the Stage-2 model, and conduct transfer learning based on the ImageNet pre-training model, and then train the FPR model which is used to eliminate false positive.

The technical route combines the YOLOv2 and VGG classification models, which can effectively enhance the accuracy in case of only using YOLOv2; meanwhile, by virtue of transfer learning, the initial performance of model can be enhanced, thus shortening the training period.

### 2.1. Sample Fabrication

This paper involves the Stage-1 target detection model and Stage-2 false positive elimination model. Therefore, it is necessary to construct two sample sets respectively for model training.

#### 2.1.1. Fabrication of Sample of Stage-1 Model

For the purpose of constructing a sample set for training the Stage-1 model (YOLOv2), it is necessary to conduct pre-processing to SAR images such as COSMO, RADARSAT-2 and GF-3 firstly, and then fabricate SAR sample database with multiple sources, multiple dimensions and multiple spatial resolutions, thus covering real objects of different types of electric towers in different terrain environments. Since the electric towers are relatively small and the quantity of pixels is relatively small, in this paper, the samples of electric towers are fabricated under different spatial dimensions and different terrain conditions. Figure 2 shows the examples of electric tower samples of two scales.
The sample standard refers to the Pascal VOC standard format [19] which is internationally universal, and the sample image is 512*512 in size; the corresponding XML document records the basic information of sample image and the relative position of each target in the sample image.

2.1.2. Fabrication of Stage-2 Model Sample
While constructing the Stage-2 model sample, firstly, it is required to utilize the trained Stage-1 model to forecast the SAR image and analyze the identification result, cut out the identified target, and take the correctly detected electric tower target as positive sample, and the remaining background part is used as negative sample. Figure 3 shows the examples of positive and negative samples.

2.2. Construction of Electric Tower Identification Model
According to the designed technical route, this paper involves two deep learning models. The Stage-1 refers to the YOLOv2 model used for target detection, while the Stage-2 refers to the VGG classification model for false positive elimination.

2.2.1. Stage-1 Model Construction
The Stage-1 model refers to YOLOv2, and is produced based on improvement of YOLO, so as to solve problems such as low precision and non-accurate positioning of YOLO network detection. In detail, the advantages of YOLOv2 are shown as below:

a) The YOLOv2 is designed with a new classification network Darknet-19 as the basic model of network, and it only includes convolutional layer and pooling layer. This network uses a large number of 3x3 convolution kernel, with the thought of network in network (NIN) for reference, and uses the global average pooling (GAP) to conduct network forecast, so as to compress characteristics and increase network depth.
b) The model uses the thought of anchor in Faster-RCNN for reference, and adopts the clustering method to conduct dimensional clustering to target borders where data is centralized in, and determines the size and quantity of anchors, thus effectively enhancing the positioning accuracy of target.

c) The multi-scale input is introduced to train the detection network, and the training process can change the scale of input image. Such multi-scale detection training is applied into different input resolutions, and for input image with low resolution, it has relatively high detection rate, while for input image with high resolution, it has relatively high accuracy.

d) During the training, the YOLOv2 will calculate the confidence coefficient to each candidate box. Meanwhile, it will calculate the probabilities of this candidate box belonging to various objects. After multiplying the confidence coefficient of the candidate box and its corresponding forecast type probability, the comprehensive score will be obtained; in addition, the model will conduct non-maximum suppression (NMS) according to the score of candidate box, thus obtaining the final category of the candidate box, and enhancing the accuracy of target detection.

2.2.2. Stage-2 Model Construction

The Stage-2 is mainly used to eliminate false positive of the target body detected by Stage-1, and this paper selects VGG network as the Stage-2 model. The full name of VGG is Visual Geometry Group, and it is a network model with relatively excellent classification performances in the Convolutional Neural Network (CNN); and this paper selects VGG-16 to select target bodies which are detected by Stage-1 for secondary screening, thus enhancing the accuracy of target detection of electric tower.

VGG-16 adopts the combination and overlap of 3×3 filters at all layers, thus obtaining and extracting more tiny characteristics from the input pictures. Through the combination of multiple 3×3 convolution kernels, the calculation effect is the same as that of convolution kernels with template size of 5×5 or 7×7. The input images of VGG-16 convolutional neural network refer to 3-channel 224×224 small images, and the network contains 16 weight layers, consisting of 13 convolutional layers and 3 full-connection layers.

In this paper, the nature of VGG-16 false positive elimination model is a classification model conducting correct/wrong classification to targets, and the model consists of convolutional layer, pooling layer, full-connection layer and Softmax classification layer; the 2-label Softmax classification layer is adopted to replace the Softmax classifier in the original VGG-16 network.

Considering that the pre-trained VGG-16 network is obtained through the training of 1 million images in the ImageNet database, it has relatively strong capability in deep characteristic learning and owns a large number of parameters and weights which have been trained well. To shorten the network training time and enhance the network training efficiency, this paper has “transferred” the pre-trained model parameters into the false positive elimination model in this paper through transfer learning based on VGG-16 network, thus realizing the secondary screening to the Stage-1 result.

2.3. Model Training

The model training process includes two parts: the training of Stage-1 detection model and training of Stage-2 false positive elimination model.

2.3.1. Stage-1 Model Construction

During the process of training Stage-1 (YOLOv2) model, the pre-training weight on the VOC dataset is adopted to initialize the model. Meanwhile, for the forecast needs clearer textural features, the resolution of network input is to be enhanced to 608*608. The training is divided into 3 rounds; the first round has totaling 1,000 epochs, while both the second round and the third round have 6,000 epochs respectively. Each round is set with checkpoint to store the optimal model after Early Stopping. The SGD is selected as the optimizer, with the initial learning rate set as 10^-3.

The Early Stopping mechanism is mainly used to prevent over-fitting during the training process, and its algorithm principle is shown as below. Firstly, through comparison, record the model with the
highest precision verified currently, and set the step length \( N \); if the precisions of \( N \) models after the model with the highest precision are not enhanced, the training is to be ended, and the model with the highest precision will be selected as the final electric tower detection model.

2.3.2. Stage-2 Model Construction
During the process of training Stage-2 (VGG) model, the pre-training weight on the ImageNet dataset is adopted to initialize the model. Modify the full-connection layer of the model, and use the pre-training weight to conduct transfer learning. The training process is easier than that of YOLOv2, and the epochs is set as 1,000, the optimizer refers to Adam, and the initial learning rate is \( 1 \times 10^{-4} \). The same as the YOLOv2 training process, the EarlyStopping mechanism is utilized to prevent the over-fitting phenomenon during the training process.

3. Experimental Result
The electric tower identification model proposed in this paper is adopted for electric tower identification to the testing data. Figure 4 refers to the identification result of electric towers on the plain in single image, and Figure 5 refers to the identification effect picture of electric towers on the mountainous land.

![Identification Result](image1)
Figure 4. Identification Result of Electric Tower on the Plain

![Identification Result](image2)
Figure 5. Identification Result of Electric Tower on the Mountainous Land

The electric identification results in Figure 4 and Figure 5 are given by the light blue bounding box, and the confidence probability that the target is identified as electric tower above the bounding box, and the value of confidence probability shall be taken within \( 0-1 \); the higher the value is, the higher the confidence coefficient that the detected target is electric tower will be. The image on the right side gives the corresponding authentic electric tower position, which is convenient for comparison of performances of electric tower detection algorithms. In case of using two deep learning models trained
well in this paper jointly, the high-precision electric tower target positioning will be realized; each bounding box coincides with the minimum externally connected rectangle of the electric tower basically, i.e. the Intersection of Union (IoU) of the detection result to the authentic value is very high. Meanwhile, the confidence coefficient of electric tower detection is relatively high, and the confidence coefficients of most electric towers are above 0.7, which means that the electric tower detection model has relatively high classification precision. From the detection result of single image, it can be known that the quantity of false alarms is small, which means that the model in this paper has relatively strong ability in false alarm suppression.

After the electric tower identification model trained in this paper is obtained, the detection model on the 608*608 image is promoted to the large-format remote sensing image. Firstly, the sliding window algorithm [20] is adopted to obtain single 608*608 image window, with the overlap rate of adjacent images set as 50%, thus avoiding that the electric tower falls into two windows. Meanwhile, the non-maximum suppression algorithm is adopted to eliminate the overlap detection box, thus obtaining the electric tower detection result on a large-format image. See Figure 6 for the electric tower detection result in Zhijiang City, Hubei Province.

Figure 6 shows that there are totaling 42 electric towers in the large-format image of Zhijiang City, Hubei Province, 31 of which detected, with recall rate of 73.8%. In a word, based on the technical route of this paper, utilizing small sample to train the deep learning model for electric tower detection of SAR image can eliminate the disturbance of complicated background well, with relatively high identification accuracy.

**4. Conclusion**

With regard to the electric tower identification problem of high-resolution SAR image, this paper combines the YOLOv2 target identification model and the VGG classification model, and fabricates
target classification sample database by virtue of YOLOv2 identification result, trains the VGG classification model, and eliminates the false positive in YOLOv2 detection result. The experiment result shows that this method can enhance the target identification accuracy effectively, and the secondary classification can effectively reduce the wrong classification of targets. After combining the sliding window and the NMS algorithm, the target identification can be conducted to the large-format image, thus expanding the scope of application of detection models obtained from the training on the 608*608 image dataset. By virtue of detection model and classification model of training targets of transfer learning, utilize the small dataset and acquire models with relatively high using value, which will effectively over-fitting problem of model and enhance the robustness and generalization ability of model.

Since the background where the electric towers are located is relatively complicated, farmlands, rivers, buildings and mountain forests are criss-cross, and the existing samples cannot cover all of them. Therefore, it is still necessary to expand the sample set, and add operations such as difficult case exploration and characteristic enhancement, thus further enhancing the recall rate and accuracy of electric tower detection based on existing foundation.

References
[1] Jiao Licheng, Liu Fang, et al. Intelligent SAR Image Processing and Interpretation [M]. Science Press, 2008.
[2] Crisp D J. The state-of-the-art in ship detection in synthetic aperture radar imagery[R]. Defence Science and Technology Organization (DSTO), Information Science Laboratory, Edinburgh, Australia, 2004.
[3] Diemunsch J R, Wissinger J. Moving and stationary target acquisition and recognition (MSTAR) model-based automatic target recognition: search technology for a robust ATR[C]. Aerospace/Defense Sensing and Controls. International Society for Optics and Photonics, 1998: 481-492.
[4] WILEY C A. Pioneer award acceptance remarks [J]. IEEE Transactions on Aerospace and Electronic Systems, 1986, 21(3).
[5] Tan Libo, Li Daojing, Wu Yirong, et al. High Resolution SAR Imaging of Moving Ship Targets at Sea [J]. Journal of Electronics & Information Technology, 2006, 28(4):624-627.
[6] G. B. Goldstein. False Alarm Regulation in Log-normal and Weibull Clutter. IEEE Trans. On AES, 1973, 9(1):84-92.
[7] H. Rohling. Radar CFAR Thresholding in Clutter and Multiple Target Situations. IEEE Trans. On AES,1983,19(3):608-621.
[8] C. J. Oliver, S. Quegan. Understanding Synthetic Aperture Radar Images. Artech House: Boston, 1998.
[9] J. Li, E. G. Zelnio. Target Detection with Synthetic Aperture Radar. IEEE Trans. On AES. , 1997, 32(2):613-627.
[10] L. M. Kaplan, Improved SAR Target Detection via Extended Fractal Features. IEEE Trans. On AES. , 2001, 37(2):436-450.
[11] Liu J W, Sun Z K, Luo X L. Review and research development on domain adaptation learning[J]. Acta AutomaticaSinica, 2014, 40(8):1576-1600.
[12] Mohamed A, Dahl G, Hinton G. Deep belief networks for phone recognition[C]. Nips workshop on deep learning for speech recognition and related application, Merano, Italy. 2009, 1(9):39.
[13] Li Junbao, Yang Wenhui, Xu Jianqing, et al. Deep Convolutional Network Based SAR Image Object Detection and Recognition [J]. Navigation Positioning and Timing, 2017, 4(1): 60-66.
[14] Redmon J, Divvala S, Girshick R, et al. You Only Look Once: Unified, Real-Time Object Detection[C]. Computer Vision and Pattern Recognition, Las Vegas, Nevada. 2016:779-788.
[15] Liu W, Anguelov D, Erhan D, et al. Ssd: Single shot multi-box detector[C]. European conference on computer vision. Springer, Cham, 2016: 21-37.
[16] Pan Rong, Sun Wei. Deep Learning Target Detection Based on Pre-segmentation and Regression [J]. Optics and Precision Engineering. 2017, 25(10):221-227.

[17] Joseph R, Ali F. YOLO9000: Better, Faster, Stronger [R]. 2016, Arxiv: [1612.08242v1].

[18] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]. Proceedings of the IEEE conference on computer vision and pattern recognition. Las Vegas, Nevada. 2016: 770-778.

[19] Soheil B, Naveen R, Lukas S, et al. Comparative Study of Deep Learning Software Frameworks [J]. Computer Science, 2015.

[20] Method about Sliding-window-based Weakly Labeled Object Detection [D]. Harbin: Harbin Institute of Technology, 2016.