An Identification Method for High Voltage Power Grid Insulator Based on Mobilenet-SSD Network

Xu Tan[1], Fan Yang[1(✉)], Yan Li[2], Jinqiao Du[2], Yong Yi[2], Jie Tian[2], and Zijun Liu[2]

[1] State Key Laboratory of Power Transmission Equipment and System Security and New Technology, School of Electrical Engineering, Chongqing University, Chongqing 400044, China
yangfan@cqu.com

[2] Shenzhen Power Supply Co., Ltd., Shenzhen 518000, China

Abstract. The identification of power equipment using visible image and deep learning methods has become widespread in the power industry. However, current deep learning algorithms often face issues related to large model parameters and high hardware requirements, making it difficult to integrate them into mobile devices. To overcome these challenges, a novel approach has been proposed to identify insulators on overhead transmission lines using UAVs that carry lightweight models. This method utilizes an enhanced lightweight MobileNet-SSD target detection network, enabling accurate classification and location of power equipment. The results demonstrate that this approach can quickly and precisely label power equipment in complex backgrounds. Additionally, it has small model parameters, high efficiency, strong robustness, and an mAP of 82.47%, making it ideal for enhancing patrol accuracy and real-time monitoring of mobile equipment towards the transmission lines.

Keywords: power equipment identification · MobileNet-SSD · insulator · deep learning

1 Introduction

Insulators in overhead transmission lines are essential for withstanding voltage and load. Their integrity is vital for ensuring the stable operation of overhead lines. The patrol inspection of overhead transmission lines has evolved through four stages: manual, robotic, helicopter-based, and unmanned aerial vehicle-based inspections [1, 2]. Among these, UAV-based inspections have gradually become the preferred method for monitoring the status of transmission lines, owing to their efficiency, safety, cost-effectiveness, and other benefits.

Currently, there are two primary methods for using UAVs to obtain images of insulators for fault identification. The first method is the traditional image processing approach, while the second is the deep learning convolutional neural network (CNN) method. The image processing method involves using a sliding window to extract image features, followed by training a classifier to identify faults. However, this method requires manual

K. Hemachandran et al. (Eds.): ICAID 2023, AHIS 9, pp. 160–170, 2023.
https://doi.org/10.2991/978-94-6463-222-4_15

extraction of insulator fault characteristics, which is time-consuming and susceptible to subjective factors. As a result, it often yields low accuracy and efficiency, making it unsuitable for insulator identification.

The identification method based on deep learning CNN can automatically extract features, resulting in high efficiency, accuracy, and robustness. Target detection algorithms, such as R-CNN, SPPNet, Fast R-CNN, YOLO, and SSD, have been proposed. In 2016, Wei Liu et al. proposed the Single Shot MultiBox Detector (SSD), which utilizes the anchor frame mechanism of Fast R-CNN and integrates the regression idea of YOLO, while also using multi-scale convolution layers for prediction. This approach offers the high accuracy of Fast R-CNN and the fast speed of YOLO [3], making it one of the best models for target detection. Researchers such as Du Liqun have used SSD to detect infrared insulator images [4], while Li Ruisheng and others have used an improved SSD to detect pin defects of transmission lines [5]. Gao Jinfeng and colleagues have utilized Fast R-CNN and Full Convolutional Networks (FCN) to identify and segment insulator images [6].

Based on the analysis above, this paper proposes a fast and efficient insulator identification model using the lightweight MobileNet-SSD network. This model employs a lightweight network that can be embedded into a UAV, enabling high-precision and rapid insulator identification. The application process involves training the network model using collected images (obtained via UAV or other means) on a ground workstation, embedding the trained model into the UAV, and using the UAV carrying the model for power grid inspection and diagnosis. This paper primarily focuses on the construction, training, and testing of the network models.

2 Target Detection Network of Lightweight Mobilenet-SSD

2.1 Multi-scale Feature Fusion Target Detection Network

The network structure of SSD adopts a pyramid structure of feature fusion of multiple convolution layers. Its backbone feature extraction network is VGG16 [7]. When connecting, SSD removes the last fully connected layer (FC), classification layer (softmax), and all Dropout layers of VGG16, and replaces FC6 with a 3×3 layer. The FC7 layer is modified to a 1×1 layer. SSD adds eight convolution layers behind VGG16 to enhance the feature extraction capability of the network. In contrast to previous target detection networks, SSD fuses features from six different scales: Conv4_3, FC7, Conv8_2, Conv9_2, Conv10_2, and Conv11_2. In Conv4_3, an L2 regularization term is used to ensure that the characteristics of the lower and higher levels are not significantly different. The network architecture of SSD is shown in Fig. 1.

The backbone feature of SSD extracts detailed information from the input image for network learning, while the eight newly added convolutions delve deeper into the abstract features of the image. These features are used to predict the position coordinates of the default box (DB) and the corresponding category confidence. The prediction box with high overlap rate is then removed using the non-maximum suppression (NMS) algorithm to obtain the final prediction result. SSD adopts a full convolution network, which overcomes the limitation of fixed input image size, making it particularly suitable

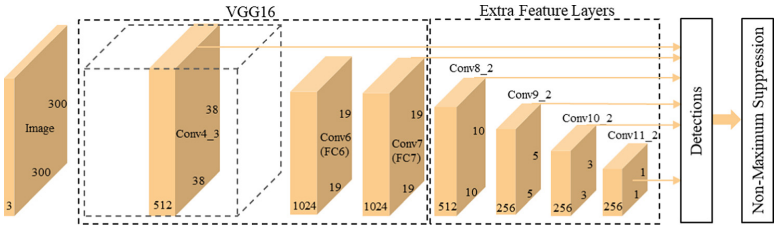

Fig. 1. SSD network architecture

for the current UAV power inspection scenario where the image size is not standardized [8].

2.2 Light Weight Backbone Feature Extraction Network

SSD has many advantages, it uses VGG16 as its backbone network, but the parameter quantity of VGG16 reaches 138 million, the model parameter quantity is too large to run on mobile devices with limited memory. In 2016, lightweight models such as SqueezeNet, ShuffleNet and MobileNet [9–11] appeared successively. MobileNet has a simple streamlined structure, with the advantages of less parameters and low latency. The MobileNet network structure is shown in Fig. 2.

MobileNet has a total of 28 convolution layers, of which 26 are deep separable convolution (DSC). The convolution with stride of 2 is special. It also plays the role of down-sampling while realizing convolution. Finally, the results are output through average pooling layer, FC layer and softmax layer.

One of the advantages of the MobileNet model is that it uses DSC to speed up the operation, so that the central processing unit (CPU) can also meet the real-time requirements. DSC is composed of a deep convolution (DW) with a kernel of 3×3 and a pointwise convolution (PW) with a kernel of 1×1.

The calculation amount of ordinary convolution and DSC is analyzed. $D_K * D_K$ is the size of the kernel, M is the number of input channels, and N is the number of output channels. Assume that the size of input feature map is $D_F * D_F * M$. The output feature

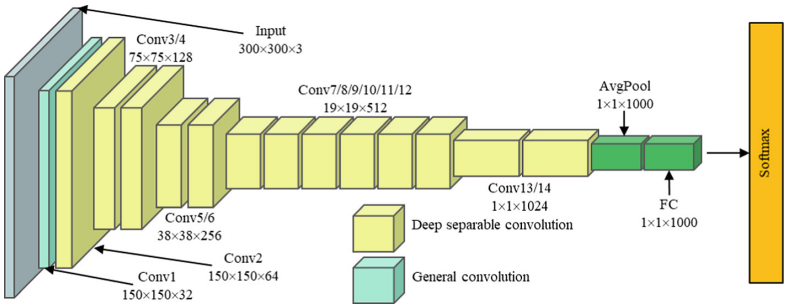

Fig. 2. MobileNet network structure

map size is $D_F * D_F * N$. D_F is the height and width of the feature map, and the size of the input and output feature maps is the same.

For ordinary convolution, the calculation amount is

$$C_1 = D_K D_K M D_F D_F N \tag{1}$$

For DSC, the calculation amount is

$$C_2 = D_K D_K M D_F D_F + M D_F D_F N \tag{2}$$

The ratio of DSC to ordinary convolution computation is

$$\eta = \frac{C_2}{C_1} = \frac{D_K D_K D_F D_F + M D_F D_F N}{D_K D_K M D_F D_F N} = \frac{1}{N} + \frac{1}{D_K^2} \tag{3}$$

Generally, N is relatively large. If the kernel of 3×3 is used, the calculation amount of DSC can be reduced by about 9 times.

Another advantage of MobileNet is the introduction of two Hyperparameters: width coefficient $\alpha \in (0, 1]$ (used to reduce the number of input and output channels) and resolution coefficient $\rho \in (0, 1]$ (not only can adjust the resolution of the input image, but also can reduce the number of model parameters), which can obtain smaller and faster models with minimal changes and will not damage the network structure. The calculation amount of a volume layer after using α is

$$C_3 = D_K D_K \alpha M D_F D_F + \alpha M D_F D_F \alpha N \tag{4}$$

Under the joint action of α and β, the calculation amount of a certain volume layer of MobileNet is

$$C_3 = D_K D_K \alpha M \rho D_F \rho D_F + \alpha M \rho D_F \rho D_F \alpha N \tag{5}$$

2.3 Fused Network of MobileNet and SSD

In order to deploy the deep learning target detection network on embedded mobile devices with limited hardware resources and computing power, this paper replaces the SSD backbone network with MobileNet. The MobileNet-SSD network model after integrating lightweight MobileNet is shown in Fig. 3.

In the MobileNet-SSD network, MobileNetV1 extracts the network for its backbone features. During the fusion, MobileNetV1 deleted the last average pooling layer, FC layer and classification layer, and changed the size of the input image to 300×300, and adopts SSD multi-scale prediction strategy.

After the last convolution layer of MobileNetV1, eight convolution layers with different scales are added, which are Conv15_1, Conv15_2, Conv16_1, Conv16_2, Conv17_1, Conv17_2, Conv18_1 and Conv18_2 in turn. These feature layers decrease in size.

The feature map of the shallow feature layer has high resolution, but the receptive field is small, which is used to detect small target objects; The resolution of the feature

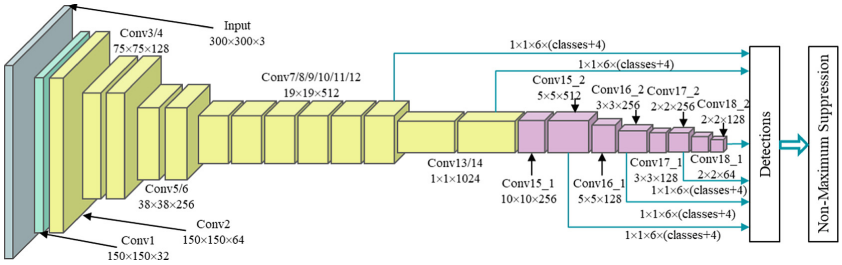

Fig. 3. MobileNet-SSD network structure

map of the deep feature layer is small, but the receptive field is large, which is used to detect large target objects.

MobileNet-SSD extracts six effective feature maps of different scales from the six layers Conv12, Conv14, Conv15_2, Conv16_2, Conv17_2 and Conv18_2 for multi-scale feature prediction, with resolutions of 19×19, 10×10, 5×5, 3×3, 2×2, 1×1.

Assuming that there are m feature maps, the calculation of prediction frame size is shown as follows.

$$S_k = S_{\min} + \frac{S_{\max} - S_{\min}}{m - 1}(k - 1), k \in [1, m] \tag{6}$$

where: S_k is the size of the prediction box; $S_{\min}$ indicates the minimum value of the prediction box, with a value of 0.2; $S_{\max}$ indicates the maximum value of the prediction box, with a value of 0.9; k represents the k^{th} feature map.

The width w and height h of the default box and the coordinates x and y of the center point of the default box are calculated as follows.

$$w = S_k\sqrt{a_r} \tag{7}$$

$$h = \frac{S_k}{\sqrt{a_r}} \tag{8}$$

$$\begin{cases} x = \dfrac{i + 0.5}{|f_k|} \\ y = \dfrac{j + 0.5}{|f_k|} \end{cases} \tag{9}$$

where: a_r is the aspect ratio of the prediction box; i indicates the i^{th} prediction box; j indicates the j^{th} true box; f_k indicates the length or width of the k^{th} feature map.

The loss function of MobileNet-SSD is the weighted sum of category confidence loss and location loss. Assuming that the input sample is defined as x, the total loss function is shown as follows.

$$L(x, c, l, g) = \frac{1}{N}(L_C(x, c) + \beta L_L(x, l, g)) \tag{10}$$

where: N is the number of prediction boxes matched to ground truth (GT); L_C is the loss of classification confidence; L_L is position loss; β is used to adjust the proportion between L_C and L_L, default $\beta = 1$.

c indicates the classification confidence, l is the predicted value of the default box; g is the position information of the real box.

The classification confidence loss function uses cross entropy loss is shown as follows.

$$L_C(x, c) = -\sum_{i \in P}^{M} x_{ij}^p \ln\left(\hat{c}_i^p\right) - \sum_{i \in N} \ln\left(\hat{c}_i^0\right) \tag{11}$$

where

$$\hat{c}_i^p = \frac{\exp\left(c_i^p\right)}{\sum_p \exp\left(c_i^p\right)} \tag{12}$$

where: P is the position of the positive sample boundary box; $x_{ij}^p \ln\left(\hat{c}_i^p\right)$ represents probability prediction, which predicts the matching probability of prediction frame i and real frame j in category p.

$x_{ij}^p \in \{0, 1\}$,When $x_{ij}^p = 1$, it means that the prediction box i matches the real box of the j^{th} in the category p; $\hat{c}_i^p$ indicates the confidence level of the target class, which corresponds to the positive default box containing the target category p.

M is the position of the negative sample boundary box; $\hat{c}_i^0$ indicates the confidence level of the background class, corresponding to the negative default box that does not contain the target object.

The position loss function is shown as follows.

$$L_L(x, l, g) = \sum_{i \in P}^{N} \sum_{m \in \{c_x, c_y, w, h\}} x_{ij}^k L_S\left(l_i^m - \hat{g}_j^m\right) \tag{13}$$

where: x_{ij}^k indicates whether the i^{th} prediction box and the j^{th} real box are the same in the k^{th} category; L_S is to use smooth L1 loss for position error; $\left(c_x, c_y\right)$ represents the center of the bounding box.

w and h represents the width and height of the prediction box; $\hat{g}$ represents the relative offset between the real box and the default box; l_i^m is the prediction box; $\hat{g}_j^m$ is a real box.

The calculation of $L_S(x)$ is shown as follows.

$$L_S(x) = \begin{cases} 0.5x^2, |x| < 1 \\ |x| - 0.5, \text{other} \end{cases} \tag{14}$$

In the prediction stage, after the image passes through MobileNet-SSD, multiple prediction boxes are generated in advance at each location (x, y) of each feature map and category confidence and position regression are performed.

A large number of prediction boxes may contain or overlap with each other. It is necessary to use the non-maximum suppression algorithm for iterative optimization, filter out the prediction boxes with high coincidence, and obtain the final prediction results.

3 Sample Analysis

For this paper, a proprietary insulator dataset was created, consisting of 2004 images of insulators with a resolution of 640 × 480 pixels.

To address the issues of insufficient training sample data and low detection and recognition accuracy, we have performed data augmentation on the original dataset using techniques such as mirroring, contrast adjustment, rotation, cropping, and brightness adjustment, based on variations in drone shooting distances, angles, lighting, and other factors.

This is aimed at improving recognition accuracy that may have been compromised due to inadequate data or low image quality. As a result of data augmentation, we have obtained 13,782 images of insulators.

The open source LabelImg labeling tool was used to label the images, and a dataset was constructed with 90% of the data used for training and 10% for testing.

Currently, most UAVs use CPU processors to simulate hardware environments, and this example analysis was performed on an NVIDIA GeForce RTX 2060 GPU and an AMD Ryzen 7 4800H CPU.

The Windows 11 operating system, along with the Keras deep learning framework and Pycharm compilation environment was used for the analysis.

The model training and parameter settings are as follows: the input image resolution is set to 300 × 300, the initial learning rate (LR) is set to 0.001 and adopts a gradient attenuation method, and the SGD optimizer is used. As the number of samples is small, the features extracted from the backbone network are universal.

Using the method of transfer learning, the model was first pre-trained on a small dataset of 2004 insulation images, and then the pre-trained weights were used as initial training weights for the larger dataset of 13782 insulation images.

The network was fine-tuned in this process. Due to the small sample size, the features extracted by the backbone network were generic, and freezing the training during the fine-tuning stage could speed up the network training and prevent the weights from being

Fig. 4. Single-target, double-target, and multi-target detection results of 110 kV and 500 kV insulators

disrupted in the early stages of training, thereby avoiding the phenomenon of unclear feature extraction caused by overly random weights.

To speed up network training and prevent weight values from being damaged at the initial stage, the method of freezing training is used. The freeze training stage has a batch size of 32. At the thawing stage, the batch size is 16, and the LR is reduced to 0.0001.

To improve the accuracy of target recognition and positioning, firstly, in order to improve the accuracy of insulator detection, the number of default boxes in the first shallow effective feature layer of MobileNet-SSD was reset to 4. After improvement, the number of default boxes in the 6 effective feature layers were adjusted to 4, 6, 6, 6, 6, and 6, respectively.

The size of the default boxes in the effective feature layers was also adjusted from [30, 60, 111, 162, 213, 264, 315] to [21, 45, 99, 153, 207, 261, 315]. By reducing the size of the default boxes in the shallow feature layers, the detection accuracy of the model for defective parts was further improved.

Secondly, in order to improve the accuracy of predicted boxes matching real boxes, the coordinates of the insulator were extracted using code programs, and the aspect ratio of the default boxes was adjusted using statistical analysis.

Through data augmentation, network improvement, and the transfer learning method presented in this paper, the accuracy of insulator detection and the matching degree of detection areas have been improved to a certain extent. The results of insulator tests are shown in Fig. 4, and the training loss curve of MobileNet-SSD is shown in Fig. 5.

The loss value decreases rapidly during the initial and fine-tuning stages of training, and eventually stabilizes at around 11.67. The accuracy-recall (PR) curve and average precision (AP) curve for insulator detection are shown in Fig. 6 and Fig. 7.

The lightweight and simple structure of MobileNet-SSD allows for the shortest training time and fastest detection speed. However, the main advantage of MobileNet-SSD lies in its flexibility for adjusting parameters to match the requirements of platforms with limited resources.

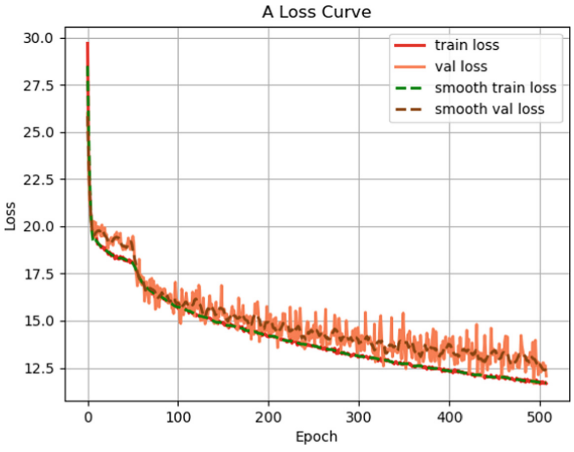

Fig. 5. Network training loss curve

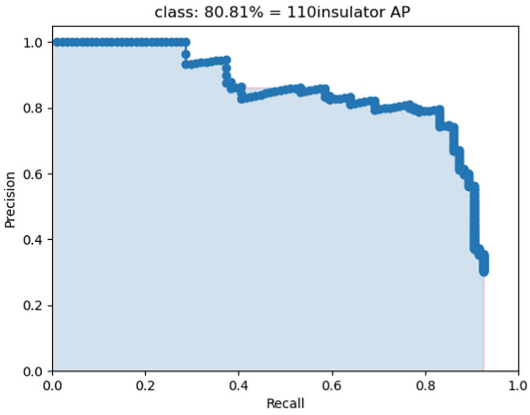

Fig. 6. PR curve of 110 kV insulator

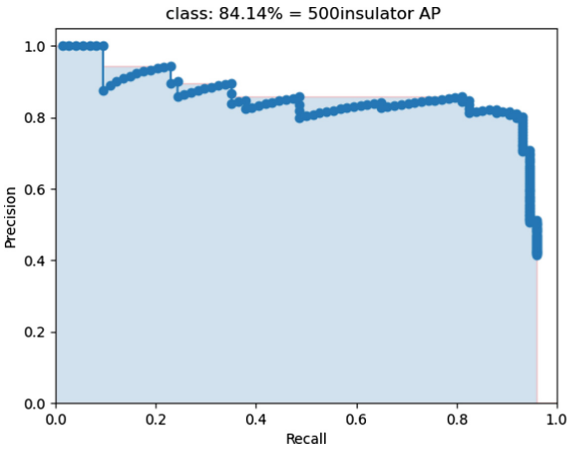

Fig. 7. PR curve of 500 kV insulator

Other models have fixed computation and parameter requirements, and may require other methods to reduce the parameter count. It can be observed that MobileNet-SSD is particularly suitable for embedded mobile devices like UAVs.

4 Conclusion

In response to the application needs of embedded mobile devices like UAVs, a lightweight MobileNet-SSD network method has been developed for identifying insulators. The proposed method achieves an mAP (mean average precision) of 82.47% and can recognize images at a rate of 97.31 FPS.

This method maximizes the advantages of deep learning, delivering high accuracy and efficiency while leveraging the small parameters of the MobileNet-SSD model. It is ideally suited for transplantation onto embedded mobile devices, offering a new approach to intelligent fault diagnosis of power insulators with small devices.

Acknowledgment. The work was supported by the Science and Technology Project of China Southern Power Grid Co., Ltd. (090000KK52220019). The author sincerely thanked China Southern Power Grid Co., Ltd.

References

1. LÜ Zeqing. Intelligent fault recognition of insulators based on drone vision[D]. Beijing, China: Beijing Information Science and Technology University, 2021.
2. SHAO Guiwei, LIU Zhuang, FU Jing, et al. Research progress in unmanned aerial vehicle inspection technology on overhead transmission lines[J]. High Voltage Engineering, 2020, 46(1): 14–22.
3. LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot MultiBox detector [C]//Proceedings of the 14th European Conference on Computer Vision. Amsterdam, the Netherlands: Springer, 2016: 21–37.
4. DU Liqun. Research on insulator detection technology based on SSD[D]. Baoding, China: North China Electric Power University (Baoding), 2019.
5. LI Ruisheng, ZHANG Yanlong, ZHAI Denghui, et al. Pin defect detection of transmission line based on improved SSD[J]. High Voltage Engineering, 2021, 47(11): 3795–3802.
6. GAO Jinfeng, LÜ Yihang. Research on recognition and segmentation of insulator strings in aerial images[J]. Journal of Zhengzhou University (Natural Science Edition), 2019, 51(4): 16–22.
7. SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[C]//Proceedings of the 3rd International Conference on Learning Representations. San Diego, USA: ICLR, 2015.
8. ZHAO Z B, XU G Z, QI Y C, et al. Multi-patch deep features for power line insulator status classification from aerial images[C]//Proceedings of 2016 International Joint Conference on Neural Networks (IJCNN). Vancouver, Canada: IEEE, 2016: 3187–3194.
9. HOWARD A G, ZHU M L, CHEN B, et al. MobileNets: efficient convolutional neural networks for mobile vision applications[J/OL]. arXiv: 1704.04861, 2017. https://arxiv.org/abs/1704.04861.
10. SANDLER M, HOWARD A, ZHU M L, et al. MobileNetV2: inverted residuals and linear bottlenecks[C]||Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE, 2018: 4510–4520.
11. HOWARD A, SANDLER M, CHEN B, et al. Searching for MobileNetV3[C]||Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul, Korea (South): IEEE, 2019: 1314–1324.