Research on automatic location and recognition of insulators in substation based on YOLOv3

Yunpeng Liu1, Xinxin Ji2, Shaotong Pei1, Ziru Ma1, Gonghao Zhang1, Ying Lin2, Yufeng Chen2

1Hebei Provincial Key Laboratory of Power Transmission Equipment Security Defense, North China Electric Power University, Yonghua North Street 619, Baoding City, People’s Republic of China
2State Grid Corp of China Shandong Electric Power Research Institute, No. 1 South Second Ring Road, City District, Jinan, People’s Republic of China

E-mail: ncepu_xinxin@163.com

Abstract: With the development of a smart grid, the automatic location of power equipment is becoming a trend. In this study, a method for automatic location identification and diagnosis of external power insulation equipment based on YOLOv3 is proposed. This deep learning algorithm is used to extract the characteristics of image data under the visible light channel of the insulator. It learns and trains the collected data to realise the rapid location identification and frame selection of the external insulation equipment and extract discharge characteristics of the target box under the ultraviolet channel. According to the number of photons and the spot area information, the operating status of the equipment is determined. The results show that the YOLOv3 algorithm with a training rate of 0.005 achieved a fast convergence of the location recognition model. The average recognition accuracy was 88.7% and the average detection time was 0.0182 s. The combination of visible light path insulator target recognition and ultraviolet light path diagnosis can realise a lean and intelligent diagnosis of power equipment. This method had good real-time performance, accuracy, and robustness to the background. It provides a new concept for intelligent diagnosis and location analysis of power equipment.

1 Introduction

At this stage, the renewed interest in unmanned aerial vehicles, robots and other power inspection measures has realised the cloud-to-cloud connectivity between unstructured data information and power systems. It is a tremendous technological achievement to link the field and dispatching unit, effectively addressing the problems of low efficiency, poor real-time performance and the high risk coefficient of traditional electric power patrol inspection. It also promotes the development of intelligent and automatic electric power equipment detection [1–3].

However, existing image- and video-monitoring systems transmit a large amount of multimedia data to the scheduling unit. The operation and maintenance personnel still need to observe and use their experience in evaluating the images frame by frame, and the efficiency is extremely low [4]. In addition, massive amounts of image data can easily result in a backlog under inefficient screening, which reduces the real-time performance of equipment operations and maintenance, making it difficult to adapt to the development of intelligent power systems [5]. Therefore, methods of effective data screening and the automatic location and recognition of equipment have become a technical bottleneck for the development of power equipment inspection.

Relevant research on the location and identification of power equipment by experts has mainly concentrated on theoretical analysis and feasibility studies. In 1997, Yamada et al. [6] conducted a preliminary attempt at power line detection. However, due to the limitations of image processing technology at that time, the research centred on image filtering and denoising, and the detection effect was general. In 2010, Li et al. [7] focused on aerial imagery, detecting power lines through a pulse-coupled neural filter and an improved Hough transform. The model constructed by this algorithm depended on the environmental background, and it was difficult to identify the power equipment. In 2012, Liu et al. [8] proposed an efficient algorithm for recognising damaged cables based on an improved Freeman rule. However, this method also relied on the environmental background, and the effect was not satisfactory in the evaluation of towers, spacers, and crossover power lines. Li et al. [9] proposed a recognition method based on deep learning and random forest. The recognition rate of this method was good, but the model relied on the completeness of the sample database, and there was still room for improvement. It can be seen that the current research on the location and recognition of power systems still focuses on the macroscopic aspects of power lines and towers and lacks the specific identification and customised diagnosis of power equipment, effective detection means and field application verification.

At present, deep learning algorithms based on computer vision have been constantly changing in the field of target recognition. In 2012, the concept of deep learning gained prominence and achieved good results in the field of image processing, such as facial recognition and signal extraction. This led to some scholars applying and theorising on deep learning in other areas, such as the electric power and energy power fields [10, 11]. Among them, the application of deformable parts models [12], region-based convolutional neural networks (RCNNs) [13], spatial pyramid pooling networks (SPPnets) [14] and fast RCNNs [15] are more popular, and equipment classification and diagnosis can be realised through optimisation algorithms and parameter tuning. In April 2018, ‘You Only Look Once’ Version 3 (YOLOv3) was created, demonstrating unparalleled advantages in training accuracy and speed, and received great attention in the computer field [16]. At present, the algorithm has not been applied and promoted in the power industry.

The research presented in this paper consisted of (i) on-site inspection and testing to obtain a large sample of visible channel data of external insulation equipment; (ii) researching and constructing the ontological feature recognition database of various types of external insulation equipment; (iii) establishing a multi-level evaluation mechanism based on YOLOv3; (iv) unsupervised learning of visible light image characteristics of external insulation equipment; and (v) researching the location and identification method under the visible light channel of external insulation equipment. At the same time, the discharge characteristics in the target frame under the ultraviolet (UV) channel were extracted, and the operation state was determined according to the number of
level of the grid operation and has good prospects for practical application. The average diagnostic accuracy was 88.7%, and the average detection time was 0.0182 s. It can be used to guide the inspection of external equipment, reducing the consumption of manpower and material resources, improving the intelligent detection level of the grid operation and has good prospects for practical application. The combination of visible light path insulator target recognition and UV light path diagnosis can realise lean and intelligent diagnosis of power equipment. At the same time, this is also a new location and recognition method that combines digital image processing, deep learning, and machine learning and has important academic significance.

2 Methodology

2.1 Basic principle of YOLOv3

The YOLOv3 model can be divided into a feature extraction layer and a processing output layer. The feature extraction layer is a combination of the Darknet-53 and ResNet-like networks. The processing output layer is similar to the feature pyramid networks (FPN). The YOLOv3 network structure is shown in Fig. 1.

The YOLOv3 model can also be specifically divided into 106 layers of full convolution architecture, including a conv layer, batch normalisation (BN) layer, shortcut layer, routing layer, up-sample layer, and yolo layer. Among them, the shortcut layer draws lessons from the residual structure of ResNet; the route layer (as the name implies) is the routing layer, indexing to the feature map in front; up-sample is a bilinear up-sampling layer; the yolo layer is segmentation and (FPN). The YOLOv3 network structure is shown in Fig. 1.

The YOLOv3 model can also be specifically divided into 106 layers of full convolution architecture, including a conv layer, batch normalisation (BN) layer, shortcut layer, routing layer, up-sample layer, and yolo layer. Among them, the shortcut layer draws lessons from the residual structure of ResNet; the route layer (as the name implies) is the routing layer, indexing to the feature map in front; up-sample is a bilinear up-sampling layer; the yolo layer is segmentation and (FPN). The YOLOv3 network structure is shown in Fig. 1.

2.2 Network components

The speed and accuracy of the YOLOv3 algorithm in the target recognition process have unparalleled advantages, mainly due to the network characteristics and the methods that use them [19]. The latter are further described below.

2.2.1 End-to-end detection: End-to-end detection is one of the important features to distinguish YOLOv3 from other methods, focusing only on the input and output. For any convolutional network, only through the loss function, an input image, train through the network, and finally output the predicted image to achieve end-to-end processing, thus completing the detection with a significant speed-up effect [17].

2.2.2 Dimension clustering: Traditional algorithms usually use manual frame selection, but this leads to reduced accuracy. In order to better select the previous network, YOLOv3 inherits the YOLOv2 calculation anchor frame selection method and uses the K-means clustering method to train the bounding box. This method uses the IoU score as the final evaluation criterion and selects nine anchor points based on the average IoU to predict the bounding box thus achieving an improvement in precision [20]. The distance function formula used for clustering is

$$d(\text{box, centroid}) = 1 - \text{IoU(box, centroid)}$$

2.2.3 Boundary box calculation: The detection effect of the traditional bounding box prediction method needs to be optimised, so the K-means method in dimensional clustering is used to predict the bounding box [21]. When inputting an image, a target in the network is first selected and then its centre point is determined. The input image is divided into equal-sized cells to calculate the coordinate position of the unit in which the centre point is located. The predicted bounding box is calculated from the coordinates of the centre point. Its coordinate formula is

$$b_i = \delta(t_i) + c_i$$

$$b_i = \delta(t_i) + c_i$$

where $(t_x, t_y, t_w, t_h)$ represents the coordinates of the centre point of the bounding box; $(P_w, P_h)$ represents the width and height of the segmentation and $(c_x, c_y)$ represents the coordinate offset.

2.2.4 Confidence calculation: There are two factors that can help to calculate confidence. The first concerns whether the region has a predicted target. If there is a target to be measured, set it to 1; otherwise, set it to 0. Then, the size of the IoU is calculated [22]. The confidence prediction is as shown in the formula:

$$\text{confidence score} = P_{(object)} \times \text{IoU}_{\text{truth}}$$

2.2.5 Category prediction: Since the forecast box may contain multiple categories, the softmax function puts each forecast box into a category. Therefore, in order to solve the problem that multiple tags may overlap, the binary cross entropy loss prediction class [23] is used instead of using the softmax function as the output. The cross entropy formula is as follows:

$$c = -\frac{1}{n} \sum x (y \log(y) - (1 - y) \log(1 - y))$$
where $c$ represents the value of the cross entropy loss, $n$ represents the number of network layers, $x$ represents the input vector of the network layer, $y$ represents the actual network output value, and $\hat{y}$ represents the network prediction value (Fig. 2).

### 3. Implementation and analysis

This section focuses on the YOLOv3-based insulator location recognition algorithm implementation, training process, test results, and detailed analysis.

#### 3.1 Algorithmic realisation of the insulator location recognition

In this research, 1892 images under visible light channels were obtained by collecting information at the substation. According to different industrial applications, the suspension insulators, pillar insulators, and insulating sleeves were sorted. The ontological feature recognition database of the three types of external insulation equipment was constructed. Through the YOLOv3 algorithm for unsupervised training and learning, the insulator location and recognition model was finally established. The network structure after training is shown in Fig. 3:

![Fig. 3 Structure diagram of insulator location recognition model](image)

The size of the anchor is estimated by $K$-means clustering to determine the prediction frame size of YOLOv3. After 1000 clustering calculations, the optimal anchor parameter values were selected. The clustering boxes were $[12, 138], [29, 24], [23, 116], [92, 107], [42, 156], [75, 266], [78, 39], [37, 61]$. When inputting an image, it was resized and the features were obtained from the feature extraction layer. Then, it entered the processing output layer. Owing to the multi-scale output, the size of the first output at the 82nd layer was $13 \times 13 \times 24$. The operations of convolution, up-sampling, and feature fusion were performed a second time on the 94th layer. The size was $26 \times 26 \times 24$. Similarly, the output size in layer 106 was $52 \times 52 \times 24$. In this algorithm, tensor represents the network tensor [24]. The tensor formula is as follows:

$$\text{tensor} = N \times N[(\text{bounding box}) \times (\text{offset + object + class})]$$

where $N=1$, bounding box $= 3$, offset $= 4$, object $= 1$. Class is the number of output categories, so class $= 3$.

The yolo layers with three different scales parse the feature maps of the previous conv layer outputs and output the predicted frame information ($t_x, t_y, t_w, t_h$) and the prediction confidence objectness through the yolo layer. By comparing the IoU values of the three scales, the matching scale was determined, thus realising the location and recognition of the three types of insulators (Fig. 4).

![Fig. 4 Multi-scale detection structural diagram](image)

#### 3.2 Training process and accuracy analysis

It can be seen from Fig. 5 that different training rates directly affected the accuracy of YOLOv3. When the training rate was 0.005, the model training could achieve fast convergence and reached an accuracy of 99.73%. As the training rate decreased, the convergence speed of the model became slower, the stability of the model became worse and the accuracy curve began to oscillate. When it fell to 0.0001, the convergence failed due to over-fitting, and the accuracy stayed at about 60%. Through the comparison of the five learning rates, the model training effect was optimal when the training rate was selected to be around 0.005. However, this does not mean that increasing the training rate can always optimise the training process. Experiments showed that when the training rate was increased to 0.01 or above because the training rate was...
too high, the model could not converge during the training process and the training failed. In general, a high learning rate meant that more steps and more time were used in the weight updating section, so less time was spent on optimum weight convergence on the model; however, excessive learning rates may cause the training step to span too much to reach the optimal point.

In this research, the YOLOv3 curve used cross entropy loss as the loss function of the insulator classification and recognition model, and the target confidence and category prediction of v3 were logic regression. As shown in Fig. 6, the loss curve rapidly converged and tended to stabilise when the training rate was 0.005, which was appropriate. The lower training rate caused the loss curve to not converge effectively, resulting in a slower convergence rate and a lower model accuracy.

3.3 Application test and effect comparison

Taking the suspension insulator detection as an example, the detection effect of the model field application under different training rates is further explained. As can be seen in Fig. 7, when the training rate was 0.0001 and 0.01, the target insulator could not be detected due to the failure of the model training; when the training rate was 0.0005, the training model could only recognise the three insulators in the target sample, and the recognition was 76.73%, which was a poor result; when the training rate was 0.001, only one insulator was not detected because it was too close to the railing, which caused large background interference, so that the recognition effect needed to be improved. When the training rate was 0.005, all the suspended insulators in the figure were successfully detected; the average recognition degree of the detected insulators was above 88.25%, and the recognition effect was superior.

3.4 Applicable environment and restrictions

3.4.1 Brightness: In this research, the recognition effect of the target image under different brightness levels was tested. The applicable brightness range was calculated by the formula: Brightness = 0.3 x R + 0.6 x G + 0.1 x B. The experiments showed that the algorithm had a good recognition effect when the average image brightness was ≥ 40. As the brightness decreased, the image recognition effect became worse. Figs. 8 and 9 show the location and recognition effect of the insulator under different brightness levels and the trend curve. When the average brightness of the picture was >40, it had a good recognition effect. The recognition degree could reach about 80%, and the recognition degree gradually increased as the brightness increased. When the average brightness of the picture was <40, it was difficult to recognise because the difference between the insulator and the background was not obvious. Therefore, the algorithm in this study is suitable for images taken under conditions of good daylight illumination and is not suitable for night inspection (Figs. 10 and 11).

3.4.2 Degree of cut off: In order to test the anti-interference ability for image interception of the model, the target samples were intercepted by 50, 66.67, and 75%, and tested for recognition, respectively. The results showed that the algorithm had certain robustness for image truncation, and it could recognise normally when the target intercepted 50% of the body and below with the recognition degree maintained above 90%. However, significant miss detection occurred when intercepting >66%. Therefore, the algorithm is suitable for relatively complete target images, and the maximum truncation should not exceed 50% of the body.

3.4.3 Degree of occlusion: An obstacle was artificially placed onto the upper part, the middle part, and the lower part of the target to be identified, respectively, to test the influence of the occlusion degree on the location and recognition model. The test results...
showed that when the occlusion degree was $\leq 33\%$, the target could be recognised normally and when the occlusion degree reached 50%, obvious misdetection and missed detection occurred. The occlusion of the middle part was more disturbing than the occlusions of the upper and lower parts. Therefore, the model has certain robustness for partial occlusion and is suitable for location recognition without obstruction or lighter occlusion.

4 Compared with faster RCNN

In this research, the faster RCNN algorithm proposed by Ross B. Girshick in 2016 is compared with the YOLOv3 algorithm. The pre-training model of faster RCNN is based on the VGG16 training model in Imagenet dataset. The learning step is 0.001 and the training cycle is 20,000.

4.1 Training process

Compared with YOLOv3, faster RCNN has a longer training time, and the convergence of LOSS curve is slower. The trend of the curve is oscillation attenuation. As shown in Fig. 12, when the training reaches 20,000 rounds, the LOSS value still shows an upward trend. However, YOLOv3 can reach the basic convergence state by training to 2500 rounds.

4.2 Detection effect

Fig. 13 shows the results of the YOLOv3 and faster RCNN detection models. It can be seen that with the faster RCNN model, the suspended insulator in the lower right corner of Fig. 13a is not detected. Similarly, the insulating sleeve behind the right side of Fig. 13b has not been successfully detected. The YOLOv3 model accurately detects the target sample and has good performance.
4.3 Mean average precision (mAP)

The mAP under these two models are calculated and compared, and the results are shown in Table 1. Therefore, compared with faster RCNN, the YOLOv3 algorithm has better performance in model convergence rate, detection effect, and mAP.

### Table 1: Comparison of the mAP of faster RCNN and YOLOv3

| Object              | Faster RCNN | YOLOv3  |
|---------------------|-------------|---------|
| suspension insulator| 0.7532      | 0.8953  |
| post insulator      | 0.7649      | 0.9524  |
| insulating sleeve   | 0.8182      | 0.9444  |
| comprehensive performance | 0.7788 | 0.9307 |

5 Insulator state evaluation

The ultimate goal of the location recognition algorithm described above is to aid in device diagnostics. Fig. 14 shows the diagnostic flow chart for an insulator combined with a UV channel. Through two-channel shooting, the target images under visible light and UV light are separately collected. The target box is formed by the visible light channel. At the same time, the discharge condition in the target box under the UV channel is compared, and the number of photons and the spot area in the target box are extracted to see whether the fault threshold is reached, thereby judging the running state of the insulator.

The diagnostic method makes full use of the dual optical path information of the visible light path and the UV light path of the UV imager. Positioning identification provides more accurate and reliable position information for UV diagnosis, while UV diagnosis places the final effect of positioning recognition on the state evaluation of the equipment. The two complement each other to achieve the state evaluation of the external insulation equipment. This method also provides a new concept for the development of intelligent UV imagers.
6 Conclusion
Based on the underlying framework of YOLOv3 and the ontological feature recognition database of external insulation equipment under unsupervised learning, the insulator classification and location recognition model was established. To find the optimal recognition model at 0.005 training rate, the performance tests under five training rates were carried out. The detection accuracy was up to 88.7%, and the average detection time of a single image was 0.0182 s, which can realise real-time detection. Compared with the faster RCNN algorithm, this research showed that the YOLOv3 has better performance in convergence rate, detection effect, and mAP.

The ultimate goal is to examine the external insulation equipment of the power system and combine it with the discharge information under the UV channel to realise intelligent inspection and diagnosis. The use of this method can reduce the consumption of manpower and material resources and improve the intelligent detection level of the grid operation, which has a certain engineering significance.

7 Acknowledgments
This project was supported by the National Natural Science Foundation of China (51577069) and the Fundamental Research Funds for the Central Universities (2015ZD19), and the Headquarters Science and Technology Project of State Grid Corporation of China (SGTYHT/17-JS-199).

8 References
[1] Chen, K., Huang, C., He, J.: ‘Fault detection, classification and location for transmission lines and distribution systems: a review on the methods’, High Volt., 2016, 1, (1), pp. 25–33
[2] Liu, Y., Xu, Z., Li, G., et al.: ‘Review on applications of artificial intelligence driven data analysis technology in condition based maintenance of power transformers’, High Volt. Eng., 2019, 45, (2), pp. 337–348 (in Chinese)
[3] Mitchell, S.D., Welsh, J.S.: ‘Methodology to locate and quantify radial winding deformation in power transformers’, High Volt., 2017, 2, (1), pp. 17–24
[4] Zha, S.: ‘Analysis of integrated management and control of operation, maintenance and maintenance of substation equipment’, Electr. Age, 2015, 10, (10), pp. 60–63 (in Chinese)
[5] Liu, Y., Zhang, K., Fu, W., et al.: ‘Heating characteristics of deteriorated porcelain insulator under different humidities’, High Volt. Eng., 2018, 44, (6), pp. 1741–1749 (in Chinese)
[6] Yamada, Y., Tanaka, Y., Mizushima, Y., et al.: ‘Development of infrared communication controller device’, Sharp Tech. J., 1997, 1, (68), pp. 32–35
[7] Li, Z., Liu, Y., Walker, K., et al.: ‘Towards automatic power line detection for a UAV surveillance system using pulse coupled neural filter and an improved Hough transform’, Mach. Vis. Appl., 2010, 21, (5), pp. 677–686
[8] Liu, K., Wang, B., Chen, X., et al.: ‘Damaged cables recognition based on improved Freeman rule’, J. Mech. Electr. Eng., 2012, 29, (2), pp. 211–214 (in Chinese)
[9] Li, J., Wang, Q., Li, M.: ‘Electric equipment image recognition based on deep learning and random forest’, High Volt. Eng., 2017, 43, (11), pp. 3705–3711 (in Chinese)
[10] Pei, S., Liu, Y., Ji, X., et al.: ‘UV-flashover evaluation of porcelain insulators based on deep learning’, IET Sci. Meas. Technol., 2018, 12, (6), pp. 770–776
[11] Lee, S.L., Zare, M.R.: ‘Biomedical compound figure detection using deep learning and fusion techniques’, IET Image Process., 2018, 12, (6), pp. 1031–1037
[12] Felzenszwalb, P.F., Girshick, R.B., McAllester, D.: ‘Cascade object detection with deformable part models’. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 2010, pp. 2241–2248
[13] He, K., Zhang, X., Ren, S., et al.: ‘Spatial pyramid pooling in deep convolutional networks for visual recognition’, IEEE Trans. Pattern Anal. Mach. Intell., 2015, 37, (9), pp. 1904–1916
[14] Girshick, R., Donahue, J., Darrell, T., et al.: ‘Rich feature hierarchies for accurate object detection and semantic segmentation’. Proc. IEEE Conf. on Computer Vision and Pattern Recognition, Columbus, OH, USA, 2014, pp. 580–587
[15] Girshick, R.: ‘Fast R-CNN’. Proc. IEEE Int. Conf. on Computer Vision, Santiago, Chile, 2015, pp. 1440–1448
[16] Redmon, J., Farhadi, A.: ‘YOLOv3: an incremental improvement’, arXiv preprint arXiv:1804.02767, 2018
[17] Yanan, S., Hui, Z., Li, L., et al.: ‘Rail surface defect detection method based on YOLOv3 deep learning networks’. IEEE 2018 Chinese Automation Congress (CAC), X’An, China, 2018, pp. 1563–1568
[18] Huber-Mork, R., Nolle, M., Oberhauser, A., et al.: ‘Statistical rail surface classification based on 2d and 2112d image analysis’. Int. Conf. on Advanced Concepts for Intelligent Vision Systems, Berlin, Heidelberg, 2010, pp. 50–61
[19] Redmon, J., Divvala, S., Girshick, R., et al.: ‘You only look once: unified real-time object detection’. Proc. IEEE Conf. on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 2016, pp. 779–788
[20] Tumas, P., Serackis, A.: ‘Review on applications of artificial intelligence driven data analysis technology in condition based maintenance of power transformers’, High Volt. Eng., 2019, 45, (2), pp. 337–348 (in Chinese)
[21] Wang, N., Xu, Z., Zhang, J., et al.: ‘Spatial pyramid pooling in deep convolutional networks for visual recognition’, IEEE Trans. Pattern Anal. Mach. Intell., 2015, 37, (9), pp. 1904–1916
[22] Zhao, K., Zhu, X., Jiang, H., et al.: ‘Dynamic loss for one-stage object detectors in computer vision’, Electron. Lett., 2016, 54, (25), pp. 1433–1434
[23] Lee, K., Choi, J., Jeong, J., et al.: ‘Residual features and unified prediction network for single stage detection’, arXiv preprint arXiv:1707.05031, 2017
[24] Redmon, J., Farhadi, A.: ‘YOLOv4: better faster stronger’. Proc. IEEE Conf. on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 2017, pp. 7263–7271