RESEARCH ON SUBSTATION EQUIPMENT DETECTION METHOD BASED ON MULTI-SENSOR FUSION

Chentao LI [1*], Zifan ZHAO[2], Ce XU[3]

To address the inefficiency and safety risks in substation equipment inspection, a multi-sensor fusion-based detection method is proposed. We extend YOLOv8 model by integrating the CBAM (Convolutional Block Attention Module) to boost the capacity for feature extraction, while the ATFL (Adaptive Threshold Focal Loss) loss function is adopted to mitigate complex background interference. Compared to the baseline YOLOv8, the improved model achieves increases of 1.9% in precision, 3.5% in recall, and 2.2% in mAP. For point cloud processing, multi-sensor joint calibration enables image-point cloud fusion. After preprocessing with voxel and pass-through filtering, Euclidean clustering is applied to segment substation equipment point clouds within detection regions. The results obtained from experiments reveal that the proposed method achieves accurate equipment identification and reliable pose estimation, meeting practical inspection requirements.

Keywords: substation equipment, lidar, camera, YOLOv8, attention mechanism, loss function

1. Introduction

In 2009, the State Grid Corporation of China proposed the development plan for a strong smart grid, requiring the use of cutting-edge communication, information, and control technologies to build a smart grid characterized by informatization, automation, digitalization, and interactivity, aiming to reach international advanced levels, reflect independent innovation capabilities, and demonstrate Chinese characteristics [1]. Among these, substation equipment is a vital component of substations, and its normal operation is the premise for ensuring the stable operation of the substation. However, most substation equipment operates in open-air outdoor environments, making it highly susceptible to the impacts of adverse climate and extreme weather, which seriously jeopardizes the safe and stable operation of substations [2]. Therefore, substations regularly conduct inspection work on equipment. The first step of this inspection is to collect information such as image data and spatial pose of the substation equipment, followed by automatic identification and classification. Due to variations in the

[1]* Guangdong Power Grid Co., Ltd. Guangzhou Power Supply Bureau, Guangzhou 510620, China, corresponding author, E-mail：lichentao@guangzhou.csg.cn

[2] Guangdong Power Grid Co., Ltd. Guangzhou Power Supply Bureau, Guangzhou 510620, China E-mail：448200515@qq.com

3 Guangdong Power Grid Co., Ltd. Guangzhou Power Supply Bureau, Guangzhou 510620, China

professional skill levels of personnel, traditional methods represented by manual inspection are prone to problems like missed detections and false detections. Furthermore, traditional manual inspection methods typically involve close-range contact checks, posing significant electrical shock safety hazards. More importantly, when dealing with massive amounts of data, the efficiency of manual inspection methods drops dramatically. Thus, there is an urgent need for a long-range, non-contact method for detecting substation equipment.

Substations contain a wide variety of equipment. To identify and classify these numerous types of substation equipment, visible light cameras can be utilized in conjunction with computer vision object detection and classification techniques. To achieve long-range, non-contact detection, it is necessary to first obtain the spatial distance to the equipment. This distance measurement can be achieved using point cloud data generated by LiDAR.

Traditional methods for detecting substation equipment largely rely on specific features of the equipment and exhibit poor scalability. Baker et al. [3] utilized the Hough transform to map straight lines in images to intersection points in parameter space, effectively extracting power lines from images. This algorithm possesses strong anti-interference capabilities but requires prior conversion between parameter space and image space, leading to high computational complexity and significant memory consumption. Zhang Congxin et al. [4] first applied Canny edge detection to the image and then successfully extracted power lines from complex background images using the Hough transform, reducing computational costs and resource consumption; however, this method is sensitive to noise. Fu et al. [5] designed a power line detection method based on the LSD (Line Segment Detector) algorithm, which can quickly extract power lines in images with cluttered backgrounds, but it cannot distinguish between power lines and other straight-line objects with similar appearances.

Over the past few years, computer vision technology has been advancing at a swift pace, and deep learning-based algorithms for detecting substation equipment have received widespread attention. Yetgin et al. [6] employed VGG-19 and ResNet-50 network architectures to build a power line identification system, achieving binary classification detection. However, this method can only determine the existence or non-existence of power lines and cannot accurately extract key information such as their spatial position and orientation. Tao et al. [7] made specific improvements to the YOLOv8 algorithm for infrared images in the power equipment domain. By introducing FasterNet, PConv, GSConv, and VoVGSCSP modules, and integrating the EMA (Efficient Multi-scale Attention) attention mechanism, they improved the identification accuracy of substation equipment while reducing computational complexity. However, this method is only applicable to infrared images of substation equipment.

To reduce misdetections and missed detections of substation equipment caused by human factors, while simultaneously improving inspection efficiency when dealing with large volumes of data and mitigating electrical shock safety hazards, this paper utilizes machine vision and deep learning techniques to fuse image and point cloud data temporally and spatially. It achieves the identification and classification of two types of substation equipment – busbars and switches – and can obtain their spatial pose information. This holds significant importance for simplifying the inspection process for substation equipment and reducing the workload of operators.

2. LiDAR and Visible Light Camera Data Fusion

Visible light cameras can capture rich color information about the surrounding environment and objects, but due to the lack of depth information, they cannot obtain accurate three-dimensional spatial information and are susceptible to environmental factors such as weather and lighting conditions. LiDAR, conversely, possesses strong anti-interference capabilities and can accurately acquire the spatial data of objects, but it cannot obtain object category information. Therefore, fusing the perception information from visible light cameras and LiDAR allows their advantages to complement each other, leveraging the strengths of both. This fusion enables the assignment of depth values to the pixels of the camera's two-dimensional images, thereby achieving a more ideal environmental perception effect [8].

Additionally, image data is referenced to the camera coordinate system, while point cloud data is referenced to the LiDAR coordinate system. Since their operating positions differ, joint calibration must be performed to determine the extrinsic parameters (including the rotation matrix R and translation vector T) needed to spatially synchronize the image and point cloud data [9].

Fig. 1 illustrates the transformation relationships between the various coordinate systems.

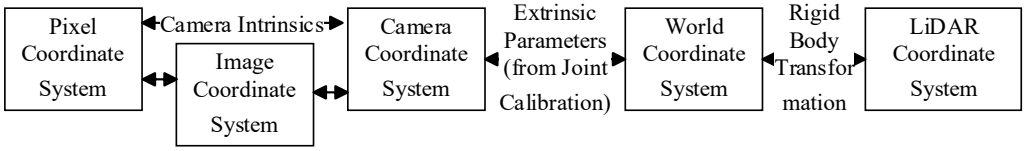

Fig. 1. Coordinate system transformation relationship

2.1 Camera Intrinsic Calibration

In a visual imaging system, a rigid body transformation is first used to map the coordinates in three-dimensional space to the camera coordinate system, and then a projection transformation is applied to convert them into coordinates on a two-dimensional image plane. Fig. 2 is a schematic diagram of the various

coordinate systems; the distance from the origin O_C of the camera coordinate system $O_cX_cY_cZ_c$ to the origin O_i of the image coordinate system O_ixy is the focal length f.

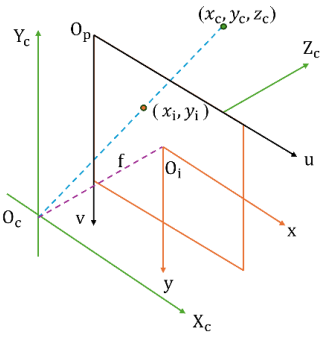

Fig. 2. Schematic diagrams of various coordinate systems

Camera calibration is completed using the Camera_Calibration toolkit based on the ROS (Robot Operating System), with its theoretical foundation being Zhang's calibration method [10]. The basic principle can be described as follows: by capturing multiple images of a checkerboard calibration pattern from different poses, corner features are extracted, and the homography matrix for each image is solved to calculate the initial intrinsic camera parameters. These are then refined through non-linear optimization to obtain the final intrinsic parameters [11]. Specifically, the camera is fixed, and an operator holds a calibration board with specifications of 7x9 squares, each square having a side length of 108 millimeters, approaching the camera from different spatial poses to capture images. Fig. 3 shows some of the captured sample images of the calibration board.

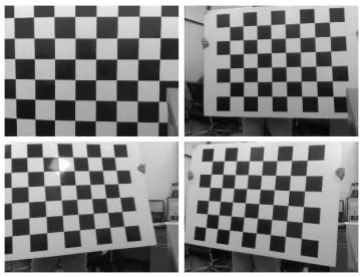

Fig. 3. Checkerboard calibration pattern

The finally calculated camera intrinsic matrix M is:

$$M = \begin{bmatrix} 879.12 & 0 & 637.38 \\ 0 & 879.55 & 387.89 \\ 0 & 0 & 1 \end{bmatrix} \tag{1}$$

2.2 LiDAR and Visible Light Camera Joint Calibration

The key to fusing image and point cloud data lies in joint calibration. Joint calibration primarily addresses the issue of spatio-temporal synchronization: in the spatial dimension, it establishes the transformation relationship between the coordinate systems by solving for the extrinsic parameters between the LiDAR and the camera; in the temporal dimension, it ensures that the data frames collected by the two sensors have temporal consistency or approximate synchronicity, thereby improving the accuracy of multi-source data fusion.

2.2.1 Spatial Synchronization

The primary task in achieving spatial synchronization of camera and LiDAR data is to solve for the transformation matrix between their respective coordinate systems. Based on the positional arrangement of the two sensors in space, corresponding three-dimensional coordinate systems are established, with their spatial relationship shown in Fig. 4. Here, $O_lX_lY_lZ_l$ is the LiDAR coordinate system, and $O_cX_cY_cZ_c$ is the camera coordinate system.

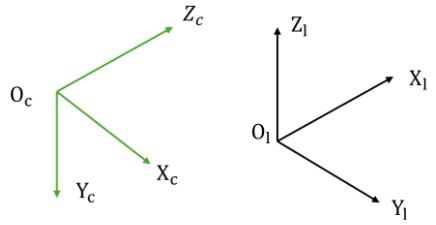

Fig. 4. The coordinate systems of lidar and camera

Since both the camera coordinate system and the LiDAR coordinate system follow the right-hand rule, they can be transformed into each other using a rotation matrix R and a translation vector T [12-13]. Let an arbitrary point P_l in the LiDAR coordinate system correspond to point P_c in the camera coordinate system. Then, P_c and P_l satisfy the following relationship:

$$P_c = \begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} = R \begin{bmatrix} X_l \\ Y_l \\ Z_l \end{bmatrix} + T = RP_l + T \tag{2}$$

That is:

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} = \begin{bmatrix} R & T \\ \mathbf{0}^T & 1 \end{bmatrix} \begin{bmatrix} X_l \\ Y_l \\ Z_l \\ 1 \end{bmatrix} = K \begin{bmatrix} X_l \\ Y_l \\ Z_l \\ 1 \end{bmatrix} \tag{3}$$

The joint calibration of the LiDAR and camera is performed using the open-source livox_camera_calib tool [14] from The University of Hong Kong. The

finally solved rotation matrix R and translation vector t are shown in Equations 4 and 5, respectively. Fig. 5 displays the outcome of the joint calibration.

$$R = \begin{bmatrix} -0.0117 & -0.9999 & -0.0099 \\ -0.0007 & 0.0099 & -0.9999 \\ 0.9999 & -0.0117 & -0.0008 \end{bmatrix} \tag{4}$$

$$T = \begin{bmatrix} 0.0395 \\ -0.0493 \\ -0.0253 \end{bmatrix} \tag{5}$$

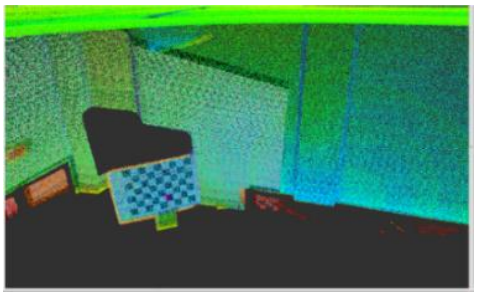

Fig. 5. Lidar and camera joint calibration

2.2.2 Temporal Synchronization

Different sampling frequencies between the LiDAR and the camera can lead to deviations in spatio-temporal information matching. This manifests as the inability to precisely align the timestamps of point cloud data and image frames, thereby reducing the accuracy of multi-sensor data fusion. To achieve accurate matching of multi-sensor data, a reliable time synchronization mechanism must be established.

By subscribing to the LiDAR's point cloud topic and the camera's image topic separately using the message_filters library, the ApproximateTime policy is employed to synchronize the point cloud and image data. The principle of time synchronization is illustrated in Fig. 6.

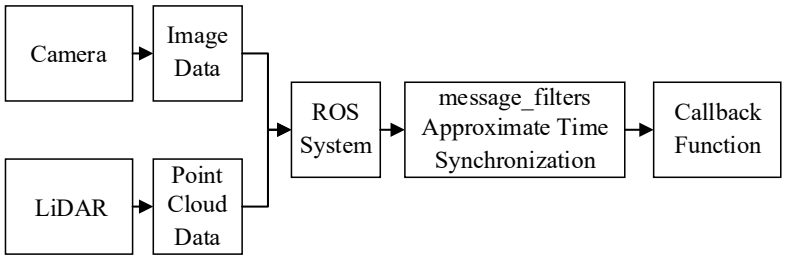

Fig. 6. Schematic diagram of time synchronization

3. LiDAR and Camera Fusion Detection Algorithm

3.1 YOLOv8 Substation Equipment Visual Detection Algorithm

3.1.1 YOLOv8 Network

YOLOv8 is a detection model open-sourced by Ultralytics in early 2023, capable of accurately completing object detection tasks while ensuring real-time performance. The YOLOv8 network architecture is made up of three sections: Backbone, Neck, and Head. Its specific structure is shown in Fig. 7.

YOLOv8's architecture design fully incorporates the advantageous experiences of previous YOLO series models. Compared to YOLOv5, this version features optimized improvements in the feature extraction network part, employing the C2f module to replace the original C3 module, effectively reducing the model's parameter count. In the prediction network design, the decoupled head structure from YOLOX [15] is introduced. By decoupling the classification task from the detection task, it improves the model's adaptability as well as its detection accuracy. These enhancements enable the network to attain superior computational efficiency while preserving high performance.

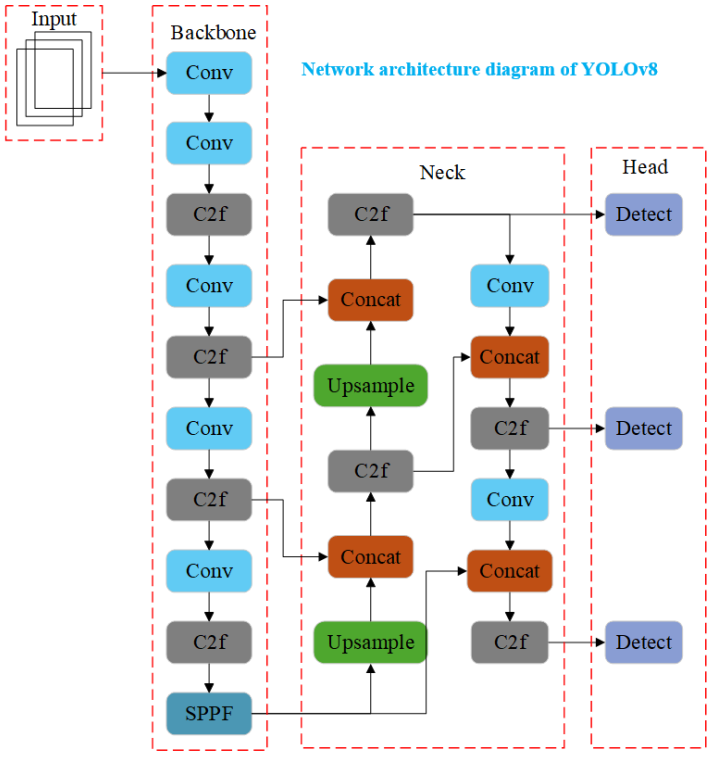

Fig. 7. The network architecture diagram of YOLOv8

3.1.2 Improved YOLOv8 Model

1) In deep learning, models need to process massive amounts of input data, yet only a small portion of this data substantially contributes to the model's training. To make the YOLOv8 model focus more on the substation equipment data itself and ignore meaningless data, the CBAM attention mechanism is introduced. Woo [16] proposed a lightweight CBAM that sequentially applies both a Channel Attention Module (CAM) and a Spatial Attention Module (SAM). Sun [17] introduced a novel hybrid architecture (T-CBAM) that combines the global modeling capability of Transformer with the local feature refinement ability of the CBAM. CBAM integrates both channel attention and spatial attention mechanisms, and it is very lightweight, allowing for plug-and-play integration. The overall process of CBAM is illustrated in Fig. 8.

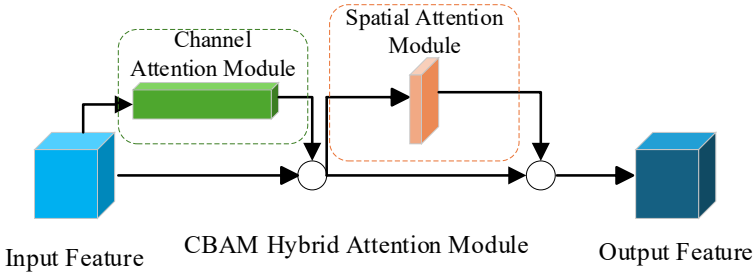

Fig. 8. CBAM module

In the processing flow of the attention mechanism, the intermediate feature map is first fed into the channel attention module for feature optimization. This module evaluates the importance of each channel and calculates corresponding weights, then performs weighted optimization on the input feature map. Subsequently, the features enhanced by channel attention are passed to the spatial attention module. This module further calculates the saliency distribution of the feature map in the spatial dimension and applies the resulting spatial weights to the feature map, thereby achieving focus on key spatial regions. The finally improved YOLOv8 network model is illustrated in Fig. 9.

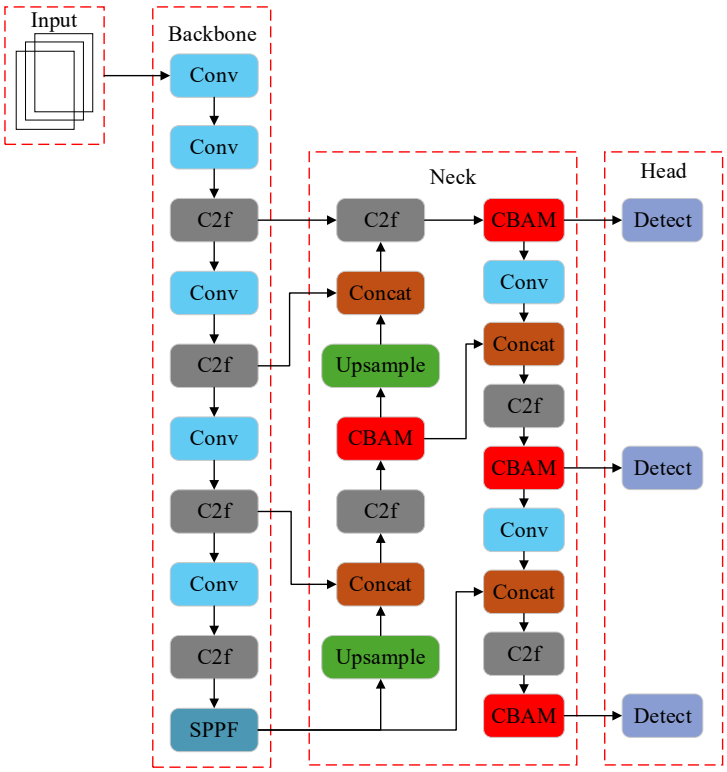

Fig. 9. Improved YOLOv8 model structure diagram

2) In the captured images, substation equipment often occupies only a small part of the entire image, meaning there is a considerable disparity between the detection target and the background. This may lead the model to pay more attention to the background instead of the target features, leading to a decrease in the detection accuracy for substation equipment. To address this issue, the ATFL (Adaptive Threshold Focal Loss) function is introduced. ATFL [18] decouples the target and background. Zhang et al. [19] proposed YOLO-ATFL, an enhanced YOLO framework that integrates ATFL to improve detection accuracy in domain-shift scenarios. Wang et al. [20] proposed ATFL module that automatically re-weights multi-scale features during transfer learning, enhancing detection performance for edge devices with limited resource. Based on sample features and model output results, it dynamically alters the weight distribution of the loss function, effectively reducing the impact of easy samples on model training while enhancing the learning effect on difficult samples, thereby boosting the overall performance in cases of target-background imbalance.

In the original YOLOv8, the loss function for the classification task is the Binary Cross-Entropy (BCE) loss for N targets, as shown in Equation 6.

$$L = \frac{1}{N}\sum_{i} -[y_i \cdot log(p_i) + (1-y_i) \cdot log(1-p_i)] \tag{6}$$

Where N is the entire quantity of samples, y_i is the actual label for the i-th sample (with a value of 0 or 1, indicating whether the target exists), and p_i indicates the predicted probability of the i-th sample being classified as positive (i.e., the target exists).

The main issue with the BCE loss function is its inability to mitigate the effects of sample imbalance, which led to the development of the Focal Loss function. Focal Loss [21] introduces an adjustable focusing parameter γ to reduce the loss weight for easily classified samples and increase the loss weight for hard-to-classify samples.

$$\mathrm{FL}(p_t) = -\alpha_t (1-p_t)^{\gamma} \log(p_t) \tag{7}$$

Where p_t represents the predicted probability for the positive class. To effectively handle sample imbalance, a balancing factor α_t is introduced to effectively adjust the contribution of positive and negative samples to the loss function. γ is the focusing parameter used to adjust the weights of easy and hard samples.

However, while Focal Loss reduces weight loss for easy samples, it can also have a similar effect on normal samples. To address this issue, the ATFL loss function is used instead of the Focal Loss function. ATFL dynamically adjusts the threshold based on the difference between the prediction results and the true labels, causing the model to focus more on difficult-to-classify samples during training.

$$\mathrm{ATFL} = \begin{cases} -(\lambda - p_t)^{\eta} \log(p_t) & p_t <= 0.5 \\ -(1-p_t)^{\eta} \log(p_t) & p_t > 0.5 \end{cases} \tag{8}$$

$$\hat{p}_c = 0.05 \times \frac{1}{t-1}\sum_{i=0}^{t-1} \overline{p}_i + 0.95 \times p_t \tag{9}$$

Where p_t signifies the present mean predicted probability value, $\hat{p}_c$ indicates the forecast for the subsequent epoch, and $\overline{p}_i$ denotes the mean predicted probability value across each training epoch.

3.2 Substation Equipment Detection Algorithm Based on 3D Point Clouds

3.2.1 Point Cloud Data Preprocessing

The initial point cloud data collected by LiDAR is voluminous and contains many non-substation equipment point clouds, placing very high demands on processor performance. To improve efficiency and avoid interference from noise points, it is necessary to preprocess the raw point cloud data [22-23].

1) To address the excessive density of the raw point cloud, voxel grid filtering downsampling [24] is employed to reduce the density of the original point cloud and speed up computation. Each fixed-size three-dimensional cube in space is defined as a voxel. First, the raw point cloud is divided into grids based on the voxel size. Then, the average coordinates of all points within each grid cell are calculated, and the point corresponding to this average value is used to represent all points within that grid cell. This significantly reduces the number of points while preserving the structure of the original point cloud, thus achieving downsampling. The point cloud after voxel filtering is shown in Fig. 10.

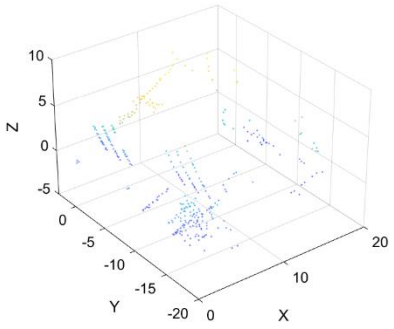

Fig. 10. Voxel-filtered point cloud

2) To deal with invalid non-substation equipment point clouds present in the data, passthrough filtering is used for removal. By setting coordinate axis ranges, only the point clouds within these ranges are retained, thereby extracting the region of interest. Considering that subsequent point cloud clustering operations are based on the point clouds projected within the scope of the image detection bounding box, and the image detection box already imposes limits on the point cloud range along the Y and Z axes, it is only necessary to set the passthrough filter range for the X-axis. Since the distance from the LiDAR to the substation equipment is typically 2-4 meters, the filtering range for the X-axis is set to [2 m, 4 m]. The point cloud after passthrough filtering is shown in Fig. 11.

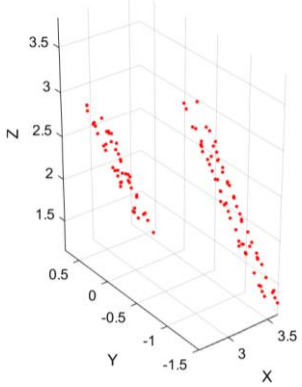

Fig. 11. Pass-through filtered point cloud

3.2.2 Point Cloud Euclidean Clustering

After obtaining the point cloud of the region of interest through preprocessing, it is still necessary to accurately extract the point cloud corresponding to the substation equipment from the regional point cloud to acquire the true pose information of the equipment in three-dimensional space. Euclidean clustering [25] is a segmentation algorithm based on Euclidean distance. For any two points P1 and P2 in space, the Euclidean distance between them can be calculated by the following formula.

$$d(P_1, P_2) = \sqrt{(x_{P_1} - x_{P_2})^2 + (y_{P_1} - y_{P_2})^2 + (z_{P_1} - z_{P_2})^2} \tag{10}$$

For a point P in space, a KD-Tree [26] nearest neighbor search is used to find all points whose distance to point P is less than a preset threshold . These points are added to a set Q. The same operation is performed for each unmarked point within Q until no new points can be found. When the size of set Q no longer changes, a new set is created, and the aforementioned operations are repeated until all points have been marked. Finally, the number of points in each set is compared; if it is less than a preset number threshold N_th, that set is discarded.

4. Experimental Results and Analysis

4.1 YOLOv8 Substation Equipment Visual Detection Experimental Results

By collecting images on-site at a substation, a YOLO format dataset containing 536 images of substation equipment was constructed and manually annotated to ensure annotation quality. To meet the needs of model training and evaluation, the dataset was proportionally split into two subsets: 80% as training samples and 20% for the validation process. Fig. 12 displays representative image samples from this dataset.

Fig. 12. Dataset images sample

Model training and evaluation were conducted under the experimental environment shown in Table 1. The number of iterations was set to 500, the batch

size was 32, the optimizer used was SGD, and both the initial and final learning rates were set to 0.01, with a weight decay of 0.0005.

Table 1

Experimental configuration

Configuration Name	Parameter
Operating System	Ubuntu22.04 LTS
CPU	i7-10700
GPU	RTX 3090
Python	3.10.0

To evaluate the performance of the improved model, a comparison was made between the original YOLOv8 model, a model with only the CBAM attention mechanism added, a model with only the ATFL loss function improved, and a model incorporating both modifications. Precision (P), Recall (R), and mean Average Precision (mAP@50) were selected as the metrics for evaluating model performance [27]. The performance comparison results before and after improvement can be referenced in Table 2.

Table 2

Ablation experiment results

Model Name	*P*/%	*R*/%	*mAP@50*/%
YOLOv8	92.0	84.8	93.0
YOLOv8+CBAM	89.4	89.6	93.6
YOLOv8+ATFL	90.8	86.7	93.9
YOLOv8+CBAM+ATFL	93.9	88.3	95.2

According to Table 2, all metrics of the improved model showed improvement, with Precision increasing by 1.9%, Recall rising by 3.5%, and mean Average Precision growing by 2.2%. This is because the introduced CBAM attention mechanism enhanced the model's focus on key features, and the ATFL loss function suppressed the negative impacts caused by class imbalance, leading to an enhancement in the model's overall detection capabilities. Using the improved model, detections were performed separately for two types of substation equipment: busbars and switches. The detection results are shown in Fig. 13.

Fig. 13. Substation equipment detection result

As it can be seen from Fig. 13, the trained YOLOv8 model identifies busbars with red detection boxes and switches with pink detection boxes. The detection accuracy for busbars is higher than that for switches, possibly because the slightly tilted posture of the switches affected the detection accuracy.

4.2. Experimental Results of Substation Equipment Detection Based on 3D Point Clouds

A total of 36 GB of three-dimensional point cloud data for substation equipment was collected on-site at the substation using the ROS system's rosbag toolkit. The point cloud processing algorithm designed in this paper was tested. Bounding boxes were used to enclose the substation equipment point clouds segmented by Euclidean clustering. The spatial coordinates of the centroid point and its distance to the LiDAR were calculated, and the centroid point of the substation equipment was displayed as a red dot. Fig. 14 presents the detection results.

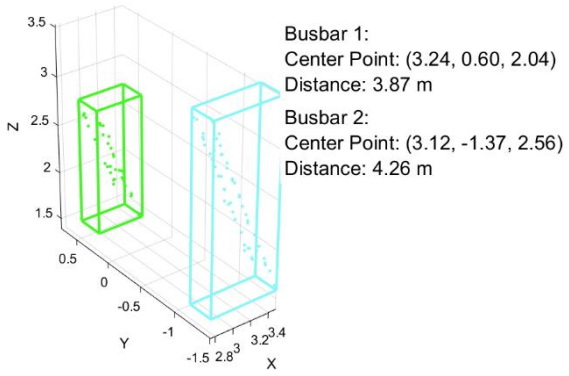

Fig. 14. Point cloud detection performance

As shown in Fig. 14, the bounding box detects the center of Busbar 1 at coordinates (3.24, 0.60, 2.04), with a distance of 3.87 meters to LiDAR. Similarly, the center of Busbar 2 is detected at (3.12, -1.37, 2.56), located 4.26 meters away to the LiDAR. This paper resolves the issue of irrelevant point cloud inclusion in rectangular detection boxes through a region-based Euclidean clustering segmentation method. By restricting clustering operations to only the point clouds within visual detection boundaries, we reduced the computational complexity and improved the accuracy in substation equipment spatial data acquisition. Furthermore, the substation equipment detection algorithm based on three-dimensional point clouds designed in this paper accurately detects substation equipment in three-dimensional space and can calculate the spatial pose information of the equipment.

5. Conclusion

Targeting the substation environment, this paper proposes a substation equipment detection method based on multi-sensor fusion. It utilizes the YOLOv8 object detection algorithm to achieve the identification and classification of substation equipment. By improving the YOLOv8 network model (introducing the CBAM attention mechanism and the ATFL loss function), the model shows improvements in three performance metrics: Precision, Recall, and mean Average Precision, thereby enhancing the algorithm's detection precision. By using voxel filtering, passthrough filtering, and Euclidean clustering, three-dimensional spatial detection of substation equipment is realized, successfully obtaining the pose information of the equipment. This provides a certain reference for simplifying the data collection process in substation equipment inspection work. However, the algorithm proposed in this paper still has some shortcomings, such as lower detection accuracy for rotated targets. Future work will focus on addressing this issue.

Acknowledgement

Science and Technology Innovation Project of Guangdong Power Grid Co., Ltd. Guangzhou Power Supply Bureau (030100KC23110038).

R E F E R E N C E S

[1] Zhang Wenliang, Liu Zhuangzhi, Wang Mingjun, et al. Research status and development trend of smart grid. Power System Technology, 2009, 33(13): 1-11.

[2] Zhao Zhenbing, Wang Fanfan, Liu Liangshuai, et al. Transmission line image fitting detection method based on attention feature fusion YOLOv5 model. Electrical Measurement & Instrumentation, 2023, 60(3): 145-152.

[3] Baker L, Mills S, Langlotz T, et al. Power line detection using Hough transform and line tracing techniques//2016 international Conference on Image and Vision Computing New Zealand(IVCNZ). Palmerston North, New Zealand: IEEE, 2016: 16.

[4] Zhang Congxin, Zhao Le, Wang Xianpei. Fast extraction algorithm for power lines in complex terrain backgrounds. Journal of Wuhan University: Engineering Edition, 2018, 51(8): 732-739.

[5] Fu L, Lu S T. 2011. Obstacle detection algorithms for aviation//Proceedings of 2011 International Conference on Computer Science and Automation Engineering. Shanghai, China: IEEE: 710-714.

[6] Ömer Emre Yetgin, Burak Benligiray, Ömer Nezih Gerek. Power line recognition from aerial images with deep learning. IEEE Trans. Aerospace and Electronic Systems, 2019, 55(5): 2241-2252.

[7] Tao H, Paul A, Wu Z. Infrared Image Detection and Recognition of Substation Electrical Equipment Based on Improved YOLOv8. Applied Sciences, 2024, 15(1): 328-328.

[8] Sun X, Jiang Y, Qin H, et al. Camera-Radar fusion with radar channel extension and dual-CBAM-FPN for object detection. Sensors, 2024, 24(16): 5317-5317.

[9] Li Lifan, Cao Pengbin, Du Bing, et al. Pedestrian detection method based on fusion of lidar and depth camera. Laser & Infrared, 2024, 54(10): 1547-1553.

[10] Zhang Z. A flexible new technique for camera calibration. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(11): 1330-1334.
[11] Zhang H, Che W, Fan Q, et al. Study on calibration method for binocular colour camera. Optics and Lasers in Engineering, 2025, 188108894-108894.
[12] Lin Yulong, Li Bing, Bai Zhengdong, et al. Application of combined laser radar technology in transmission line measurement. Electrical Measurement & Instrumentation, 2023, 60(2): 32-36.
[13] Ren F, Liu H, Wang H. A LiDAR-Camera joint calibration algorithm based on deep learning. Sensors (Basel, Switzerland), 2024, 24(18): 6033-6033.
[14] Yuan C J, Liu X Y, Hong X P, et al. Pixel-level extrinsic self calibration of high resolution lidar and camera in targetless environments. IEEE Robotics and Automation Letters, 2021, 6(4): 7517-7524.
[15] Ge Z, Liu S T, Wang F, et al. YOLOX: Exceeding YOLO series in 2021. arXiv 2021, arXiv:2107.08430.
[16] Woo S, Park J, Lee J, et al. CBAM: Convolutional block attention module. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2018, 11211: 3-19.
[17] Sun Z, Wang C. A Transformer Network Combing CBAM for Low-Light Image Enhancement. Computers, Materials & Continua, 2025, 82(3): 5205-5220.
[18] Yang B, Zhang X, Zhang J, et al. EFLNet: Enhancing feature learning network for infrared small target detection. IEEE Transactions on Geoscience and Remote Sensing, 2024, 62: 1-11.
[19] Zhang X, Li Y, Wang J. YOLO-ATFL: Adaptive Transfer Feature Learning for Real-Time Object Detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(6): 1024-1032.
[20] Wang L, Chen H, Li M. ATFL-YOLO: A Lightweight Object Detector with Adaptive Feature Fusion for Edge Devices. IEEE Transactions on Neural Networks and Learning Systems, 2024, 35(3): 1120-1130.
[21] Mukhoti J, Kulharia V, Sanyal A, et al. Calibrating deep neural networks using focal loss. Advances in Neural Information Processing Systems, 2020, 33: 15288-15299.
[22] Tao Shuqing, Liu Xiaoqiang, Li Baiyan, et al. Denoising method for scanned 3D point cloud based on density clustering and majority voting. Application Research of Computers, 2018, 35(2): 619-623.
[23] Qian Jianguo, Wei Li, Li You, et al. Research on the classification and outlier removal algorithm of transmission lines based on 3D point cloud. Applied Laser, 2022, 42(11): 104-112.
[24] Liu C, Li X, Lv S, et al. Map construction algorithm based on dense point cloud. IET Image Processing, 2024, 18(14): 4492-4502.
[25] Cao Y, Wang Y C, Xue Y F, et al. FEC: fast Euclidean clustering for point cloud segmentation. Drones, 2022, 6(11): 325-342.
[26] Zheng Y, Liu W, Zhang Y, et al. Laser in-situ measurement in robotic machining of large-area complex parts. Measurement, 2025, 241115718-115718.
[27] Yu Hong, Gong Zeweiyi, Zhang Haitao. Research on substation equipment identification and defect detection technology based on Faster R-CNN algorithm. Electrical Measurement & Instrumentation, 2024, 61(3): 153-159.