



Security Monitoring Strategy of Distribution Community Operation Site Based on Intelligent Image Processing Method

Zhiqing Sun, Yi Xuan*, Libo Fan, Rongjie Han, Yongwei Tu, Jiabin Huang, Xinyu Xiang, Yi Wang and Xiang Fang

State Grid Zhejiang Electric Power Co., Ltd. Hangzhou Power Supply Company, Hangzhou, China

OPEN ACCESS

Edited by:

Bin Zhou,
Hunan University, China

Reviewed by:

Xu Xu,
Hong Kong Polytechnic University,
Hong Kong SAR, China
Zhekang Dong,
Hangzhou Dianzi University, China

*Correspondence:

Yi Xuan
991006321@qq.com

Specialty section:

This article was submitted to
Process and Energy Systems
Engineering,
a section of the journal
Frontiers in Energy Research

Received: 29 April 2022

Accepted: 27 May 2022

Published: 27 June 2022

Citation:

Sun Z, Xuan Y, Fan L, Han R, Tu Y,
Huang J, Xiang X, Wang Y and Fang X
(2022) Security Monitoring Strategy of
Distribution Community Operation Site
Based on Intelligent Image
Processing Method.
Front. Energy Res. 10:931515.
doi: 10.3389/fenrg.2022.931515

With the continuous development of energy infrastructure, a large number of distributed new energy, distributed energy storage, and various power electronic equipment are connected to distribution communities. The access to a large amount of equipment not only increases the workload of power operators, but also leads to a complex field operation environment, which will threaten the security of field operators. Traditional monitoring strategy for distribution community operation site relies on manual operation, which is tedious, labor-intensive, and error-prone. To solve this problem, this paper proposes a security monitoring strategy for distribution community operation sites based on an intelligent image processing method. Firstly, a power image enhancement method based on multi-filtering algorithm is proposed. The filter operators are used to smooth and denoise the image, enhance the edge and detail information, and expand the security monitoring data set of operation sites by merging the filtered images. Secondly, an object detection method for personnel and security protection tools based on the improved Faster R-CNN algorithm is proposed, and a pyramid structure is constructed to improve the detection and localization accuracy of small objects such as safety helmets and safety belts. To simplify the object detection task and the complexity of the labeling system, it is disassembled into four single tasks: operators, non-operators, safety helmets, and safety belts. Finally, the detection frame labels are fused by calculating the overlap of the object detection boxes to describe the security of the operator. Experiments show that the method proposed in this paper can effectively locate the object and accurately describe the security of the operator.

Keywords: energy infrastructure, distribution community, distributed energy resources, machine vision, object detection

INTRODUCTION

With the continuous development of energy infrastructure, distributed energy resources (Lin et al., 2020), distributed energy storage (Li et al., 2021), and electric vehicle support equipment (Al-Hanahi et al., 2022) are widely connected to distribution communities. The community power operators not only need to maintain the transformers, distribution cabinets, and other equipment in the community but also regularly inspect and maintain the connected photovoltaic

equipment, energy storage equipment, and power electronic equipment. Due to technical limitations, most of these operations require operators to operate on-site. However, lots of equipment leads to a cluttered operation environment, which will threaten the security of field operators. Besides, various equipment comes from different manufacturers or is in charge of different entities, on-site operations may involve different maintenance by different management entities (Ge et al., 2021). When there are multiple operation site tasks, security supervisors cannot effectively obtain the working status and location of all personnel at each operation site. In the limitation of field operation information, potential security hazards cannot be eliminated and resolved promptly. It is necessary to use an effective on-site security monitoring strategy to control the behavior of on-site operators, reduce potential security hazards and avoid the occurrence of security accidents.

In recent years, machine vision technology based on deep learning algorithms (Amiri et al., 2019) has been widely used in operation site monitoring to improve the efficiency of operation site security supervision. The literature (Huang et al., 2022) roughly estimates the proportions of safety helmets, tops and pants in the human body region image based on the human body structure, and then counts the histogram of gradient (HOG) features and the histogram of color (HOC) feature, and finally use radial basis neural network to identify violations. The literature (Liu and Jiang., 2021) judges whether to wear operated clothes according to the distribution of black and white pixels in the binarized image. The above algorithm manually designs a feature extraction method based on the characteristics of the target structure, which has poor adaptability. The literature (Tang et al., 2021) is based on the SmallerVGGNet model, uses one-hot encoding and Sigmoid loss function for multi-label classification, and judges whether to wear safety helmets and operate clothes at the same time. The identification of electrical intelligent security monitoring based on YOLO (Chen et al., 2019) is a method for identifying insecurity behaviors related to construction, including wearing safety helmets, operating clothes, and insulating gloves. Cheng et al. proposed a CNN-based method to detect whether practitioners wear safety belts in high-altitude operations (Cheng et al., 2020). The literature (Yang and Lei, 2021) is based on the YOLO model as the main body, and the multi-scale features are fused to judge and identify the position of the safety helmet. The recognition results it is proved that the method improves the detection speed and accuracy.

Although deep learning-based methods improve the accuracy, the detection performance of deep learning models is usually affected by the number and quality of images. In reality, there are no available datasets for security monitoring of specific operation sites. The way to generate new images by rotating, cropping, zooming, and panning images of power equipment is a common image augmentation method. However, due to the change of the position of the insulator relative to the image during the data expansion process, the image annotation needs to be re-completed, resulting in a lot of repetitive labor and a lot of manpower (Wu et al., 2020).

This paper proposes a security detection for operation sites in distribution community based on an intelligent image processing method. The main contributions of this paper are summarized as follows:

- 1) Aiming at the problem that the object detection tasks based on deep learning algorithms are limited by the quantity and quality of images, a power image enhancement method based on multi-filtering algorithms is proposed. The method first uses Gaussian filter operators to smooth and denoise images, enhance edge and detail information, and improve image quality. On this basis, the images obtained based on the four filtering operators are merged. Since the position of the object in the image does not change after filtering, the annotation file of the original image can be reused, thereby greatly expanding the object detection image data set.
- 2) A security monitoring method for operating at the operation site based on the improved Faster R-CNN algorithm is proposed. In the process of minimizing the localization loss function, the method uses a deep convolutional neural network to adaptively adjust the image feature categories, extract the optimal features, and complete the category recognition while determining the object detection. By constructing a feature pyramid structure and fusing the multi-scale features in the original network, based on retaining the deep semantics, the shallow features are fully utilized to improve the recognition accuracy of small objects such as safety helmets and safety belts.
- 3) To reduce the complexity of the object detection task, the security monitoring task of operating at operation site is disassembled into four single task identification of safety helmets, safety belts, operating at heights and non-operating at heights. On this basis, a label fusion method based on the overlap of object detection boxes is proposed. By measuring the overlapping relationship of object detection boxes, the method fuses the identification result labels of a single task to describe the security of operators.

POWER IMAGE ENHANCEMENT BASED ON MULTI-FILTER ALGORITHM

Gaussian Smoothing Filtering Algorithm

The Gaussian filter is a type of linear smoothing filter that selects weights according to the shape of the Gaussian function (normal distribution function). The Gaussian smooth filter can effectively remove the noise of obeying the normal distribution. When an image is smoothed by a Gaussian method (Kowalski and Smyk., 2018), the kernel is a Gaussian function:

$$G_c(x) = \frac{1}{4\pi c^2} e^{-\frac{|x|^2}{4c^2}} \quad (1)$$

Where, c is the standard deviation of G_c , which is usually small in Gaussian smoothing methods. In image processing, a two-dimensional zero-mean discrete Gaussian function is often used as a smoothing filter:

$$G(x, y) = Ae^{-\frac{x^2+y^2}{2\sigma^2}} = Ae^{-\frac{r^2}{2\sigma^2}} \quad (2)$$

The following two conditions should be considered when selecting the Gaussian smoothing method.

Condition 1: The image method noise of convolution with Gaussian kernel G_c is:

$$v - G_c * v = -c^2 \Delta v + o(c^2) \quad (3)$$

In addition, at a reference pixel $j = 0$, the effect of Gaussian smoothing on noise can be estimated by the following formula:

$$G_c * m(0) = \sum_{j \in J} \int_{P_j} G_c(x) m(x) dx = \sum_{j \in J} \eta^2 G_c(j) m_j \quad (4)$$

Where, J refers to the entire area of an image, P_j is a square block with j as the center pixel, η^2 is the size of the square P_j , $G_c(j)$ refers to the average value of the function G_c at the pixel point j , and m_j is the pixel value of the noise m in the local area j . The noise reduction effect using the Gaussian smoothing method can be expressed by the following conditions.

Condition 2: Let $m(x)$ be a piecewise constant white noise with $m(x) = m_j$ at each square pixel j . Assume that all m_j are independent and identically distributed random variables with mean zero and variance σ^2 . Then the “noise residual” after Gaussian convolution of m with G_c satisfies:

$$\begin{aligned} \text{Var}(G_c * m(0)) &= \sum_j \eta^4 G_c(j)^2 \sigma^2 \\ &\approx \sigma^2 \eta^2 \int G_c(x)^2 dx = \frac{\eta^2 \sigma^2}{8\pi c^2} \end{aligned} \quad (5)$$

In conclusion, Gaussian convolution can remove the noise in the relatively flat areas of the image very well. However, the denoising effect of the singular parts in the image is relatively poor, especially the residual noise still exists in the edge and texture regions of the image.

Laplace Operator

Laplacian (Isabelle., 2008) denoising is equivalent to applying a filter to the imaging data, which can be expressed as:

$$H'(X) = L\{H(X)\} = \Delta H(X) \quad (6)$$

Where, L represents the Laplacian filter. X represents the data space. $H(X)$ represents the reverse time migration imaging data. $H'(X)$ represents the filtered imaging data. The Laplacian operator Δ can represent:

$$\Delta = \nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \quad (7)$$

The Fourier transform of the above formula into the wavenumber domain can be expressed as:

$$\nabla^2 \xrightarrow{\text{FFT}} -\left(k_x^2 + k_y^2 + k_z^2\right) = -|K_I|^2 \quad (8)$$

Where, K_I is a vector in the imaging wavenumber domain, and has the following relationship:

$$K_I = K_r - K_s \quad (9)$$

Where, K_r and K_s are the wavenumber vectors of the wavefield at the receiver point and the wavefield of the shot point respectively. Applying the string theorem to it, we get:

$$\begin{aligned} |K_I|^2 &= |K_r|^2 + |K_s|^2 - 2|K_r||K_s|\cos\theta \\ \cos(\pi - 2\theta) &= \frac{4\omega^2}{v^2} \cos^2 \theta \end{aligned} \quad (10)$$

Substitute into (8) again to get:

$$\nabla^2 \xrightarrow{\text{FFT}} -\left(k_x^2 + k_y^2 + k_z^2\right) = -\frac{4\omega^2}{v^2} \cos^2 \theta \quad (11)$$

Where, θ is the incident angle. Combining **formulas (6) and (11)**, it is not difficult to see that the larger the incident angle θ is in the range of $0-90^\circ$, the greater the numerical attenuation. When the incident angle is 90° , the numerical attenuation is 0. It has been known from the analysis of the mechanism of noise generation that the low-frequency noise of reverse time migration mainly occurs when the incident angle is equal to or close to 90° , Laplacian filtering is equivalent to angular domain attenuation of reverse time migration imaging data, but Laplacian filtering is only a linear operation and does not need to output common imaging gather in the angular domain. Therefore, the operation is simple and the efficiency is relatively high.

LOG Operator

The LOG operator (He and Yan., 2011), the Gauss-Laplace operator, first uses a Gaussian filter to smooth the grayscale image, and then uses the Laplace operator to transform the smoothed image. Detect the edges of the image based on the zero-crossings of the second derivative.

First, assuming that the original image is $h(x, y)$, the Laplace operator is ∇^2 , and the output image is $g(x, y)$, the LOG operator is derived as follows:

$$\begin{aligned} g(x, y) &= \nabla^2 (G(x, y) * h(x, y)) \\ \nabla^2 &= \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \end{aligned} \quad (12)$$

Where, $*$ is the convolution symbol. From the commutativity of convolution we get:

$$g(x, y) = \nabla^2 G(x, y) * h(x, y) \quad (13)$$

Where, $\nabla^2 G(x, y)$ is the LOG operator:

$$\text{LOG}(x, y) = \nabla^2 G(x, y) = \frac{1}{\pi \delta^4} \left[\frac{x^2 + y^2}{2\delta^2} - 1 \right] e^{-\frac{x^2+y^2}{2\delta^2}} \quad (14)$$

Unsharp Masking

Unsharp masking (Ngo and Kang, 2019) is a commonly used method for image edge and detail enhancement. The unsharp mask is to preprocess the original image through blurring such as spatial smoothing filtering to obtain a blurred image, subtracting the blurred image from the original image to obtain the high-frequency information in the image, and then multiplying it by a

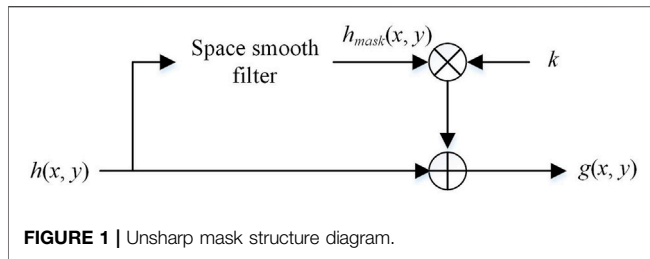


FIGURE 1 | Unsharp mask structure diagram.

correction factor to adjust the obtained height. The modified high-frequency information is superimposed on the original image, to achieve the effect of enhancing the image outline and sharpening the image. Due to the long wavelength of infrared light, the infrared image is susceptible to atmospheric attenuation and long propagation distance, and has the characteristics of large spatial correlation and blurred visual effect. Therefore, the details of the image can be enhanced by the method of unsharp masking.

The formula of the unsharp mask algorithm is expressed as:

$$g(x, y) = h(x, y) + k * h_{mask}(x, y) \quad (15)$$

Its block diagram is shown in Figure 1. Where, $h(x, y)$ represents the original image, $h(x, y)$ is obtained by spatial smoothing filtering $\bar{h}(x, y)$ and the unsharp mask is:

$$h_{mask}(x, y) = h(x, y) - \bar{h}(x, y) \quad (16)$$

The mask part is the high-frequency information of the image, including the detail layer in the image, the existing noise, and the overshoot and undershoots caused by the blurring of sharp edges. Commonly used spatial smoothing filters include Gaussian filtering, mean filtering, etc. k ($k \geq 0$) is the high-frequency information gain coefficient, which controls the degree of detail enhancement. The larger the k , the more obvious the details of $g(x, y)$. The smaller the k , the closer the $g(x, y)$ is to the original image. There are two types of unsharp masks, linear and non-linear. The linear unsharp mask means that the gain coefficient k of the high-frequency component of the image is a constant, and the non-linear unsharp mask means that the gain coefficient k is no longer a constant, but a variable $k'(x, y)$. Therefore, the non-linear unsharp mask is also an adaptive method.

Image Information Optimization and Expansion Based on Filtering Algorithm

The accuracy of intelligent image processing methods based on deep learning is often limited by the image quality and the number of images, and in practical applications, it is often difficult to collect available datasets for specific tasks. In addition, to ensure that the deep learning model has sufficient generalization ability and superior detection effect, it is often necessary to train the model based on large data sets. The premise of this goal is to manually label a large number of images, which is time-consuming and labor-intensive.

Aiming at the above problems, this paper proposes an image information optimization and data set expansion method based on a filtering algorithm. After the image is filtered, noise is removed, pixels are smoothed, edges and details are enhanced. During the process, some pixel values are changed, for the algorithm based on the convolutional neural network, the result of the sliding window calculation of the convolution kernel on the image is not the same as the original. Therefore, for the convolutional neural network, the filtered image is quite different from the original image. The two images contain different pixel information, but they contain the same object, which can be used for training the object detection model. The flow of this method is shown in Figure 2.

OPERATION SITE OBJECT DETECTION BASED ON FASTER R-CNN ALGORITHM

Principle of Faster R-CNN Algorithm

The Faster R-CNN(Liu et al., 2017) algorithm mainly consists of three parts: a feature extraction network CNN(convolutional neural network), a region proposal network (RPN) and a region-of-interest (ROI)-based classifier. The Faster R-CNN algorithm shows in Figure 3.

The image will first go through a convolutional neural network (CNN) to extract feature maps. This network consists of convolutional layers (Conv layers), ReLU activation function layers, and pooling layers. The Conv layer and ReLU layer do not change the size of the image, while the pooling layer reduces the size of the input image. In the convolutional layer, kernel size = 3 means that the size of the convolution kernel is 3×3 ; pad = 1 means that the edge is expanded by 1 pixel to ensure that the image size remains unchanged after convolution; stride = 1 means that the convolution kernel step size is 1. ReLU is an activation function, and its main purpose is to avoid gradient disappearance and increase the sparsity of the network to reduce the occurrence of overfitting. In the pooling layer, kernel size = 2, which means that each 2×2 area is converted into a pixel, that is, the length and width of the output become 1/2 of the input; stride = 2, which means that each operation moves 2 pixels point. The purpose of the pooling operation is to reduce the dimension of the feature vector output by the convolutional layer while improving the result.

After the feature map obtained by the feature extraction network is input into the RPN network, the feature map is traversed by sliding window convolution. In the original image area corresponding to each point, a certain number of anchor boxes will be generated according to preset parameters, and the degree of coincidence between the image in the box and the object will be calculated with the anchor box as the boundary. After filtering out some anchor boxes, the network will mark the remaining anchor boxes as positive samples, negative samples or useless samples, and calculate the offset between each anchor box and the marked box of the training image, the network will learn through the offset, so that the RPN network can predict the region of the object after training, and output the region proposal

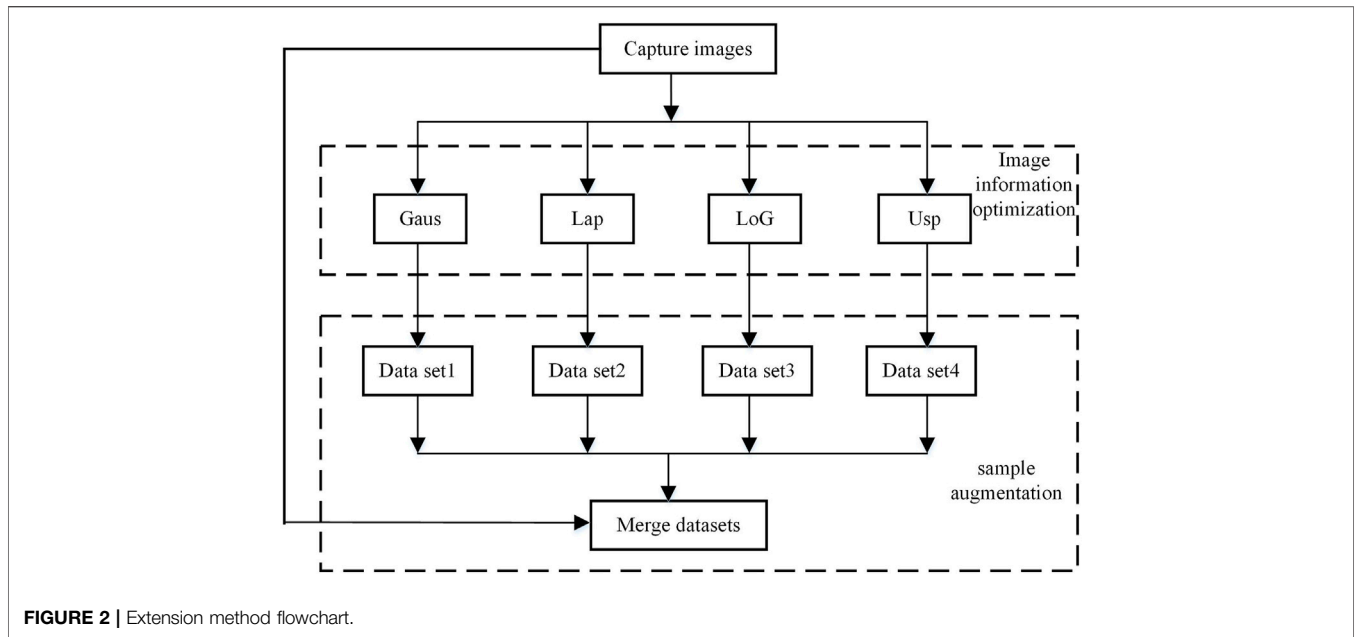


FIGURE 2 | Extension method flowchart.

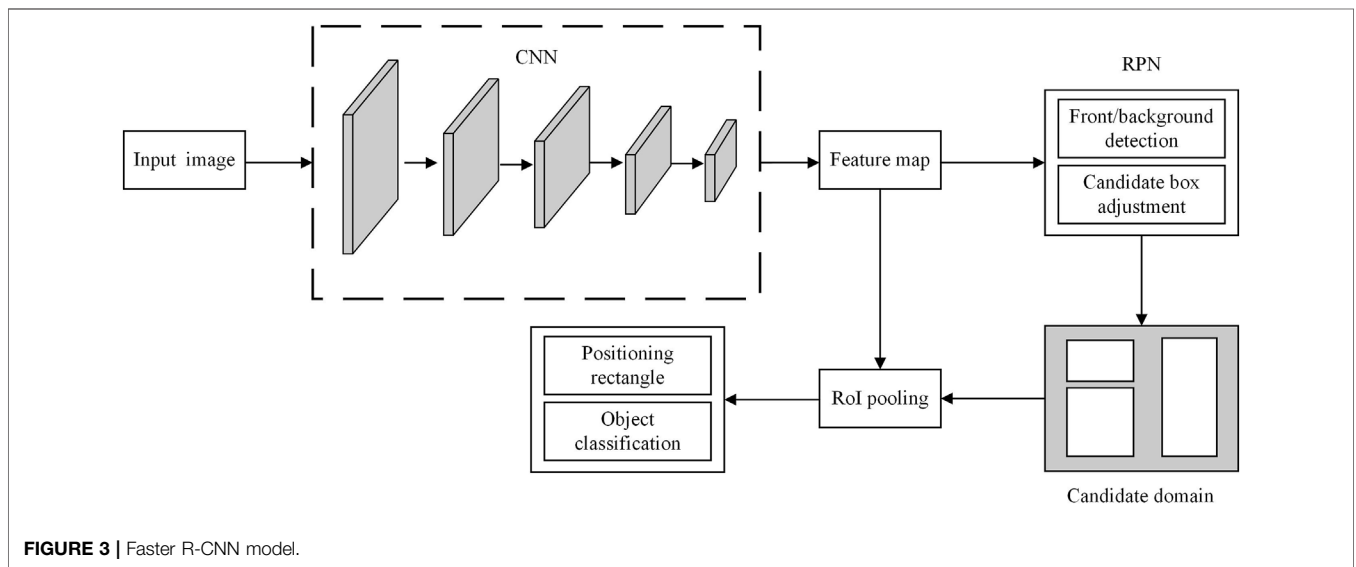


FIGURE 3 | Faster R-CNN model.

box. The RPN network is back-propagated with the gradient descent algorithm, and the loss function is as follows:

$$L(\{p_i\}, \{t_i\}) = \frac{1}{N_{cls}} \sum_i L_{cls}(p_i, p_i^*) + \lambda \frac{1}{N_{reg}} \sum_i p_i^* L_{reg}(t_i, t_i^*) \tag{17}$$

$$L(t_i, t_i^*) = R(t_i - t_i^*) \tag{18}$$

Where, i is the sequence number of the anchor frame. p_i represents the probability that the anchor box is a positive sample. p_i^* indicates the type of anchor box, positive sample is 1, negative sample is 0. t_i represents a 4-parameter vector of the anchor box. t_i^* is a 4-parameter vector representing the marker box. The classification function L_{cls} uses a logarithmic loss

function for binary classification. The regression loss L_{reg} is calculated by the **formula (18)**, R is the robust loss function. $\{p_i\}$ and $\{t_i\}$ are the outputs of the classification and regression layers respectively. N_{cls} and N_{reg} are used to normalize the outputs. λ is the balance weight.

Subsequently, the proposal box output by the RPN is mapped to the convolutional feature map of the last layer of the CNN to obtain the region of interest (ROI), then, the size of each proposal frame is fixed through the ROI pooling layer, and the low-dimensional feature vector of the ROI is obtained through the fully connected layer. Finally, it is sent to the Softmax classifier, the Softmax loss function and the Smooth L1 loss function are used to further adjust the object classification confidence and locate the rectangular box position to complete the specific

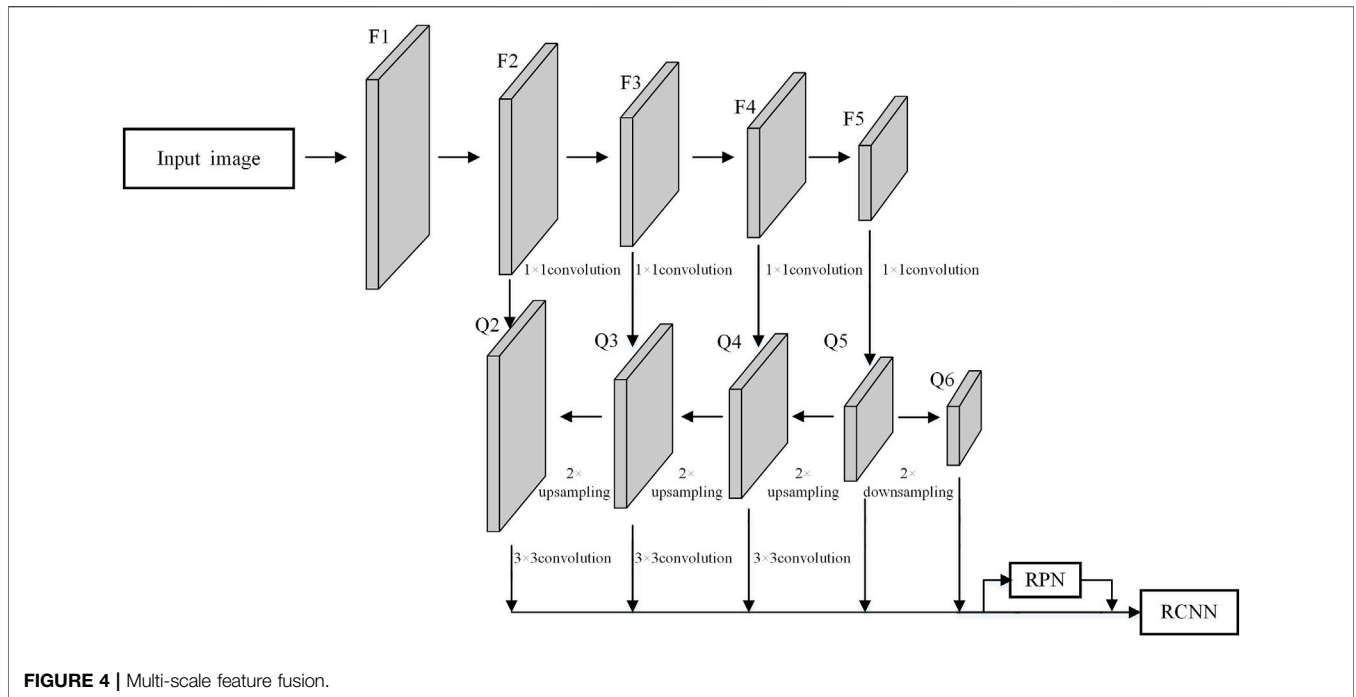


FIGURE 4 | Multi-scale feature fusion.

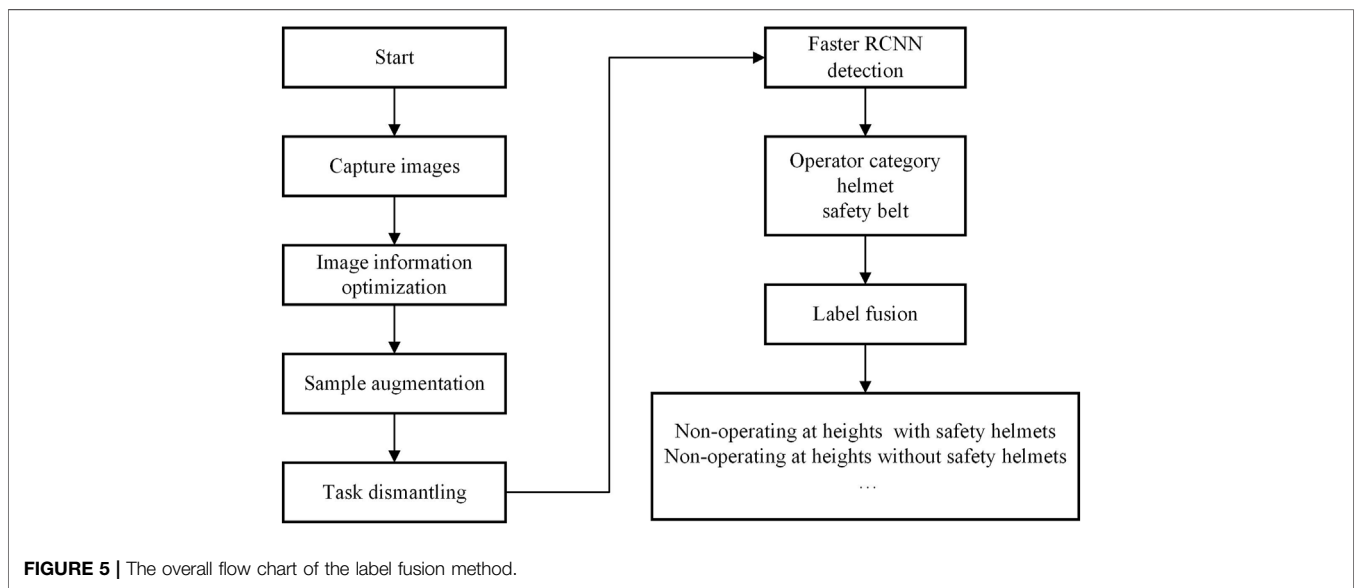


FIGURE 5 | The overall flow chart of the label fusion method.

category judgment of the proposed area and accurate border regression. The resulting output is the final recognition result.

Softmax Loss Function

$$p_i = \frac{e^{z_i}}{\sum_{j=1}^k e^{z_j}} \quad (19)$$

Smooth L1 Loss Function

$$p_{smoothL1}(x) = \begin{cases} 0.5x^2, & |x| < 1 \\ |x| - 0.5, & else \end{cases} \quad (20)$$

Improved Faster R-CNN Algorithm

The original Faster R-CNN network only extracts candidate region features on the last layer of feature maps. Deep features have rich semantic information, but ignore a lot of detailed features, which is not conducive to detecting small objects. In this paper, a feature pyramid (Wang and Zhong, 2021) is constructed to fuse multi-scale features to improve the recognition of small objects. For the convenience of discussion, the feature map before pooling in the pre-feature extraction network is called F1, and the feature map of the last layer after each pooling is called F2, F3, F4, and F5 in turn. The

TABLE 1 | Wearing of personnel safety tools in different situations.

Situation	Wearing Safety Tools
Operating at heights	With safety helmet/With safety belt With safety helmet/Without safety belt Without safety helmet/With safety belt Without safety helmet/Without safety belt
Not operating at heights	With safety helmet Without safety helmet

multi-scale feature fusion process is shown in **Figure 4**. First, a 1×1 convolution kernel is used for F5 to reduce the number of feature maps, and a low-resolution feature map Q5 is generated. Then do 2 times nearest neighbor upsampling for Q5, and use a 1×1 convolution kernel to extract low-resolution feature maps for F4. Both have the same scale, and add fusion features element by element to generate the required fusion feature map Q4. By analogy, the multi-scale fusion feature atlas {Q2, Q3, Q4, Q5} is obtained. Q6 is obtained after downsampling Q5 by a factor of 2 to enhance robustness. To reduce the aliasing effect of upsampling, a 3×3 convolution kernel is applied to Q2, Q3, and Q4 respectively. Since only simple convolution and element-by-element addition operations are involved in the feature fusion process, the increased computational load is not large. Set different anchor boxes for the multi-scale fusion feature atlas {Q2, Q3, Q4, Q5, Q6}, and optimize the initial anchor box parameters through the K-means clustering algorithm. Finally, the multi-scale fusion feature atlas {Q2, Q3, Q4, Q5, Q6} is sent to the region proposal network RPN to obtain candidate regions, and then sent to the object detection network RCNN for category and location prediction. In the RCNN network, the feature maps Pd of different scales are selected for the candidate regions of different scales as the input of the ROI pooling layer, and the coefficient d is defined as:

$$d = d_0 + \log_2 \frac{\sqrt{wh}}{224} \quad (21)$$

Where, the constant 224 is the standard input size of the image. d_0 defaults to 5, which represents the feature map of Q5. w and h represent the length and width of the candidate region respectively.

The multi-scale fusion feature map contains different levels of semantic information and detailed features from the bottom to the top, and has strong generalization. Based on retaining deep semantics, more shallow features are extracted, which is helpful for the recognition of small objects.

OBJECT DETECTION TASK DISASSEMBLY AND RECOGNITION RESULT LABEL FUSION

Object Detection Task Dismantling

The tasks of security monitoring at operation sites studied in this paper are shown in **Table 1**, which can be divided into two types of operators and six types of safety tools and equipment wearing

conditions monitoring and identification. If the operator category and the wearing of safety tools are taken as a whole task, the labeling system for manual labeling of the data set is complex, and the detection and identification of the operator category and safety tools need to be completed simultaneously in object detection (Wu et al., 2019), which increases the complexity of the task. To simplify the labeling system, reduce the complexity of the object to be identified, and improve the accuracy of object detection. In this paper, the task is disassembled into single object detection tasks for safety helmets, safety belts, operating at heights, and not operating at heights.

Label Fusion Method Based on the Calculation of the Overlap of Detection Box

Since the label of the object obtained after task disassembly is only partial description information, the ultimate goal of operation site security monitoring is to describe human security, therefore, it is necessary to fuse the tags of a single object to jointly describe the security situation of field operators.

The principle of the label fusion method proposed in this paper is to calculate the overlap degree of two object detection boxes, if one of the detection boxes is basically inside the other detection box, then these two detection boxes belong to one detection object, and the labels corresponding to the detection boxes can be used to jointly describe the same object. For not operating at heights, the safety belt label is not detected, and it is only necessary to detect whether there is a “safety helmet” detection box in the “operator category” detection box. If so, the fusion label is “not operating at heights with a safety helmet”. Otherwise, the fusion label is “not operating at heights without safety helmets”. The tag fusion process for operators at heights is similar. The overall flow of the method proposed in this paper is shown in **Figure 5**.

SIMULATION

This paper builds a simulation model based on the TensorFlow deep learning framework, and the simulation computer is configured as Windows 10, x64 operating system, 8 cores, RTX 2080Ti, Tensorflow-gpu = 1.14.0, Keras = 2.24, development environment CUDA = 10.0.130, cuDNN = 7.3.1, Python = 3.6.5.

Comparison of AP Values of Training Results

The original images contain four types of objections: safety helmets, safety belts, operating at heights and non-operating at heights. Among them, 400 are used for training and 200 are used for testing. The object of operating at heights mainly refers to the people far from the ground in the image. To verify the effectiveness of the method of optimizing image information based on filtering algorithms, different filtering operators are used to preprocess the original images, and the images preprocessed by different filtering algorithms are compared with the original

TABLE 2 | Identification results of different filter operators.

Sample	Safety Helmet	Safety Belt	Operating at Operation Site	Not Operating at Operation Site
Unfiltered	1.0000	0.6840	0.9091	1.0000
Gaus	1.0000	0.7273	0.9091	1.0000
Lap	0.5455	0.4513	0.7972	0.9930
LoG	1.0000	0.523	0.8182	1.0000
Usp	1.0000	0.5227	0.9091	1.0000

TABLE 3 | Recognition effects after different processing of the dataset.

Sample	Safety Helmet	Safety Belt	Operating at Operation Site	Not Operating at Operation Site
Unfiltered	1.0000	0.6840	0.9091	1.0000
Gaus	1.0000	0.7273	0.9091	1.0000
Merge datasets	1.0000	0.8364	0.9273	1.0000

images for training, and the training results are evaluated. **Table 2** shows the effect of various types of behavior recognition after preprocessing the training set with different algorithms. The corresponding relationship between the training set name and the operator used is as follows: Unfiltered means that the sample has not been processed. Sample Gaus means that the sample is filtered by the Gaussian operator. Sample Lap means that the sample is processed by the Laplacian operator. The sample LoG indicates that it is processed by the LoG edge detection operator. The sample Usp indicates that the image is processed by the unsharp contrast enhancement filter.

From the results in **Table 2**, it can be found that comparing the original samples without processing, after the training set is preprocessed by Gaussian filtering, the AP value of the safety belt with the worst recognition effect in the pictures is increased by 6.3%. It can be seen from the results that compared with the other three types of model recognition effects, the ability of the model to recognize safety belts is weaker. The reason may be that the shape structure of safety belts is smaller than the other three types, and the proportion in the image is smaller, and because the safety belt occluded situation is more complex and diverse, the model cannot be trained to obtain sufficient generalization performance for recognizing this category in the case of fewer samples. However, it can be seen from the results in **Table 2** that the recognition effect of the model on the safety belt is significantly improved after the preprocessing of Gaussian filtering on the training set. This shows that the difference between various sample conditions of the safety belt is reduced after the Gaussian filtering process, which improves the generalization ability of the model, so that the recognition of such objects can be more accurate. At the same time, the results in the table also reflect the inappropriateness of other filter operators for the content to be identified in this paper, so they are not used for the time being.

It can also be seen from the experimental comparison results that when the original samples are extremely scarce or limited in number, which may result in poor model training results,

different filtering operators can be used to process the original images to expand the data set and verify the method.

After mixing the original sample data without any processing with the samples preprocessed by Gaussian filtering to form a merged dataset, use the merged dataset for training and evaluate the effect. The training steps are the same as the above-mentioned filtering comparison experiments. **Table 3** shows the recognition results obtained by using the merged data and training model. From the results, it can be seen that the recognition effect after merging the data sets is significantly improved than the recognition effect of the model obtained by using the samples processed by Gaussian filtering. Among the four types of behavioral activities, the AP value of the safety belt category with the worst recognition effect increased by 6.3%. The experimental results show that when the original samples and the Gaussian filtered samples are combined as the training set, the model has the best recognition effect on the object category in the image.

Comparison of Results of Different Object Detection Algorithms

As can be seen from **Table 4**, this paper uses the method of a self-built training sample library to enhance the recognition effect of the recognition model after performing certain preprocessing on the samples, and improves the Faster R-CNN model by introducing multi-scale feature fusion into the model. After multi-scale fusion, different levels of semantic information and detailed features from the bottom to the top are extracted, which improves the generalization. Based on retaining the deep semantics, more shallow features are extracted, which improve the detection accuracy of small objects such as safety helmets and safety belts in the image. Different deep learning algorithms are used to train the dataset. Compared with models such as SSD (Chen et al., 2019), YOLO (Sadykova et al., 2020; Peng et al., 2021; Xiaolu et al., 2021), and Faster R-CNN, the improved Faster

TABLE 4 | Identification results of different algorithms.

Algorithm	Safety Helmet	Safety Belt	Operating at Operation Site	Not Operating at Operation Site
SSD	0.8954	0.8509	0.7313	0.7903
YOLO	0.7621	0.8024	0.7041	0.7214
Faster R-CNN	0.8865	0.8425	0.7241	0.7825
Improved Faster R-CNN	0.9051	0.8674	0.7617	0.8594

TABLE 5 | Results of the task without disassembly based on the improved Faster R-CNN.

Category	AP
Operators not at operation site without safety helmets	0.83
Operators not at operation site with safety helmets	0.83
Operators at operation site with safety helmets and safety belts	0.72
Operators at operation site wear safety helmets without safety belts	0.78
Operators at operation site wear safety helmets without safety belts	0.72
Operators at operation site do not wear safety helmets or safety belts	0.83

TABLE 6 | Effects of different execution methods.

Category	Accuracy
Not dismantled tasks	0.91
Label fusion after dismantling task	0.96

R-CNN in this paper further improves the detection accuracy of the model.

Task Disassembly Validity Verification

From **Table 5**, it can be seen that there are differences in the identification of clothing in different situations in field operations. When there is more standardized clothing content to be identified, the identification effect will be lower. When the recognition effect of the operators not operating at heights in the image is better than that of operating at heights, the possible reason is that the operators at heights account for a smaller proportion of the image, resulting in the recognition effect of the operator's safety helmet and safety belt is also weaker. The recognition effect in complex cases is not good, so this paper proposes to decompose and label the task, and then judge the label and fuse the labels after model recognition. **Table 6** shows the comparison of the recognition effect between the original recognition method and the task disassembly method.

The comparison results in **Table 6** show that although the improved Faster R-CNN algorithm proposed in this paper directly detects and recognizes the objects in the image, the accuracy is improved compared with other algorithms, but in the actual complex operation situation, the object such as safety helmets and safety belts needs to be recognized may not be recognized due to the long-distance or many objects to be recognized. However, if the image recognition task is performed on various objects individually and then labels are performed, after the recognition is completed, the labels of the corresponding images are fused, and output the final recognition result to help improve the accuracy. The identification results of indecomposable tasks and dismantled tasks prove that the method of

label fusion after task dismantling can improve the accuracy of the identification of personnel wearing on the operation site.

CONCLUSION

A large number of energy infrastructures such as distributed energy resources, distributed energy storage and electric vehicle support equipment are connected to distribution communities, making the operation site environment complex. A large amount of equipment maintenance in the complex environment brings security risks to on-site operators. To solve the problem, this paper proposes a security monitoring strategy for distribution community operation sites based on an intelligent image processing method. The following conclusions are drawn:

- 1) Gaussian filtering operators can effectively smooth and denoise the field operation image, improve the image quality, and provide a high-quality information source for the parameter learning process of the deep learning algorithm, thereby improving the detection of the object in the image by the deep learning algorithm to realize on-site monitoring of distribution community operations.
- 2) By building a feature pyramid structure and fusing the multi-scale features in the original Faster R-CNN algorithm network, it is possible to make full use of shallow features based on retaining deep semantics, and improving the detection accuracy of small objects such as safety belts and safety helmets.
- 3) By disassembling the object detection task, the complexity of the detection task label system and the difficulty of learning the parameters of the deep model can be effectively reduced, obtaining an object detection model for the category of the operator and the wearing of safety tools with better detection effect.
- 4) The method based on detection boxes overlap detection can effectively fuse the labels of multiple detection boxes, and use these labels to generate an accurate description of the security of the operator.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

Data curation, ZS and YX; formal analysis, LF and RH; methodology, YT and JH; supervision, XX and YW; writing (original draft), XF. All authors have read and agreed to the published version of the manuscript.

REFERENCES

- Al-Hanahi, B., Ahmad, I., Habibi, D., Pradhan, P., and Masoum, M. A. S. (2022). An Optimal Charging Solution for Commercial Electric Vehicles. *IEEE Access* 10, 46162–46175. doi:10.1109/ACCESS.2022.3171048
- Amiri, S., Salimzadeh, S., and Belloum, A. S. Z. (2019). “A Survey of Scalable Deep Learning Frameworks,” 2019 Proceedings of the 15th International Conference on eScience (eScience), 24–27 Sept. 2019, 2019, pp. 650–651. doi:10.1109/eScience.2019.00102
- Chen, Z., Wu, K., Li, Y., Wang, M., and Li, W. (2019). SSD-MSN: An Improved Multi-Scale Object Detection Network Based on SSD. *IEEE Access* 7, 80622–80632. doi:10.1109/ACCESS.2019.2923016
- Cheng, H.-C., Chang, C.-C., and Wang, W.-J. (2020). “An Effective Seat Belt Detection System on the Bus” in IEEE International Conference on Consumer Electronics, September 15–September 17, 2021 (Taoyuan, Taiwan: ICCE), Taiwan, 1–2. doi:10.1109/ICCE-Taiwan49838.2020.9258080An
- Ge, S., Liu, W., Li, X., and Ding, D. (2021). Coupling Characteristics of Electromagnetic Disturbance of On-Site Electronic Device Power Port in Substations and its Suppression. *IEEE Trans. Electromagn. Compat.* 63 (5), 1584–1592. doi:10.1109/TEMC.2021.3064154
- He, Qiang., and Yan, Li. (2011). Edge Detection Algorithm Based on LOG and Canny Operator. *Comput. Eng.* 37 (3), 210–212.
- Huang, R., Chen, Y., Yin, T., Huang, Q., Tan, J., Yu, W., et al. (2022). Learning and Fast Adaptation for Grid Emergency Control via Deep Meta Reinforcement Learning. *IEEE Trans. Power Syst.*, 1. doi:10.1109/TPWRS.2022.3155117
- Isabelle, L. (2008). Resolution and Illumination Analyses in PSDM: a Ray-Based Approach. *Lead. Edge* 27 (5), 650–663.
- Kowalski, P., and Smyk, R. (2018). Review and Comparison of Smoothing Algorithms for One-Dimensional Data Noise Reduction, Proceedings of the International Interdisciplinary PhD Workshop, 15–17 May 2019, 2018 (Swinoujscie, Poland: IIPHDW), 277–281. doi:10.1109/IIPHDW.2018.8388373
- Li, X., Wang, L., Yan, N., and Ma, R. (2021). Cooperative Dispatch of Distributed Energy Storage in Distribution Network with PV Generation Systems. *IEEE Trans. Appl. Supercond.* 31 (8), 1–4. Nov. 2021 Art no.0604304. doi:10.1109/TASC.2021.3117750
- Lin, D., Li, X., Ding, S., and Du, Y. (2020). Strategy Comparison of Power Ramp Rate Control for Photovoltaic Systems. *Cps Tpea* 5 (4), 329–341. doi:10.24295/CPSSTPEA.2020.00027
- Liu, B., Zhao, W., and Sun, Q. (2017). Study of Object Detection Based on Faster R-CNN, 2017 Chinese Automation Congress (CAC), October 20–October 22, 2017, Jinan, China (Jinan, China: IEEE), 6233–6236. doi:10.1109/CAC.2017.8243900
- Liu, Y., and Jiang, W. (2021). Detection of Wearing Safety Helmet for Workers Based on YOLOv4, Proceedings of the International Conference on Computer Engineering and Artificial Intelligence, August 29 2021, Shanghai, China, 2021 (Shanghai, China: ICCEAI), 83–87. doi:10.1109/ICCEAI52939.2021.00016
- Ngo, D., and Kang, B. (2019). Image Detail Enhancement via Constant-Time Unsharp Masking,” 2019 Proceedings of the IEEE 21st Electronics Packaging Technology Conference (EPTC), 4–6 Dec. 2019, 2019, pp. 743–746. doi:10.1109/EPTC47984.2019.9026580

FUNDING

This study received funding from The Science and Technology Project of State Grid Zhejiang Electric Power Co., Ltd. (5211HZ190151).

- Peng, Q., Wang, X., Kuang, Y., Wang, Y., Zhao, H., and Wang, Z. (2021). Hybrid Energy Sharing Mechanism for Integrated Energy Systems Based on the Stackelberg Game. *Csee Jps* 7 (5), 911–921. Sept. 2021. doi:10.17775/CSEEJPES.2020.06500
- Sadykova, D., Pernebayeva, D., Bagheri, M., and James, A. (2020). IN-YOLO: Real-Time Detection of Outdoor High Voltage Insulators Using UAV Imaging. *IEEE Trans. Power Deliv.* 35 (3), 1599–1601. doi:10.1109/TPWRD.2019.2944741
- Tang, J., Gong, Z., Wu, H., and Tao, B. (2021). RFID-based Pose Estimation for Moving Objects Using Classification and Phase-Position Transformation. *IEEE Sensors J.* 21 (18), 20606–20615. doi:10.1109/JSEN.2021.3098314
- Wang, C., and Zhong, C. (2021). Adaptive Feature Pyramid Networks for Object Detection. *IEEE Access* 9, 107024–107032. doi:10.1109/ACCESS.2021.3100369
- Wu, Q., Chen, P., and Zhou, Y. (2019). A Scalable System to Synthesize Data for Natural Scene Text Localization and Recognition, Proceedings of the IEEE International Conference on Real-time Computing and Robotics, July 15–July 19, 2021 (Irkutsk, Russia: RCAR), 59–64. doi:10.1109/RCAR47638.2019.9043965
- Wu, S., Hu, W., Lu, Z., Gu, Y., Tian, B., and Li, H. (2020). Power System Flow Adjustment and Sample Generation Based on Deep Reinforcement Learning. *J. Mod. Power Syst. Clean Energy* 8 (6), 1115–1127. doi:10.35833/MPCE.2020.000240
- Xiaolu, C., Chao, G., Xiaoqi, Y., Rong, S., Xin, S., and Hao, R. (2021). Electric Intelligent Safety Monitoring Identification Based on YOLOv4, Proceedings of the IEEE International Conference on Electrical Engineering and Mechatronics Technology, Chiang Mai, Thailand, 2021 (Qingdao, China: ICEEMT), 688–691. doi:10.1109/ICEEMT52412.2021.9602009
- Yang, G., and Lei, Q. (2021). The System of Detecting Safety Helmets Based on YOLOv5, Proceedings of the International Conference on Electronic Information Engineering and Computer Science, September 23–September 26, 2021, 2021 (Changchun, China: EIECS), 750–755. doi:10.1109/EIECS53707.2021.9587902

Conflict of Interest: ZS, YX, LF, RH, YT, JH, XX, YW, and XF were employed by the State Grid Zhejiang Electric Power Co., Ltd. Hangzhou Power Supply Company.

Publisher’s Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Sun, Xuan, Fan, Han, Tu, Huang, Xiang, Wang and Fang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.