



## OPEN ACCESS

## EDITED BY

Praveen Kumar Donta,  
Vienna University of Technology, Austria

## REVIEWED BY

V. B. Murali Krishna,  
National Institute of Technology, Andhra  
Pradesh, India  
Kenneth E. Okedu,  
Melbourne Institute of Technology,  
Australia

## \*CORRESPONDENCE

Jinheng Li,  
✉ jh.li@st.gxu.edu.cn

RECEIVED 09 September 2023

ACCEPTED 28 September 2023

PUBLISHED 10 October 2023

## CITATION

Qi C, Chen Z, Chen X, Bao Y, He T, Hu S,  
Li J, Liang Y, Tian F and Li M (2023),  
Efficient real-time detection of electrical  
equipment images using a lightweight  
detector model.

*Front. Energy Res.* 11:1291382.

doi: 10.3389/fenrg.2023.1291382

## COPYRIGHT

© 2023 Qi, Chen, Chen, Bao, He, Hu, Li,  
Liang, Tian and Li. This is an open-access  
article distributed under the terms of the  
[Creative Commons Attribution License  
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). 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.

# Efficient real-time detection of electrical equipment images using a lightweight detector model

Chaoliang Qi<sup>1</sup>, Zhigang Chen<sup>1</sup>, Xin Chen<sup>1</sup>, Yuzhe Bao<sup>1</sup>, Tianji He<sup>1</sup>,  
Sijia Hu<sup>2</sup>, Jinheng Li<sup>2\*</sup>, Yanshen Liang<sup>2</sup>, Fenglan Tian<sup>1</sup> and  
Mufeng Li<sup>1</sup>

<sup>1</sup>State Grid Zhengzhou Power Supply Company of Henan Electric Power Company, Zhengzhou, China,  
<sup>2</sup>Key Laboratory of Power System Optimization and Energy Technology, Guangxi University, Nanning,  
China

Infrared technology holds significant importance in the detection of electrical equipment, as it has the capability to swiftly and securely identify electrical apparatus. To simplify the implementation of proficient detection frameworks for electrical equipment within constrained settings (like embedded apparatus), this study presents an enhanced, lightweight model of the single-shot multibox detector (SSD). This model specifically addresses the detection of multiple equipment objects within infrared imagery. The model realized the lightweight of the model by using the network structure characteristics of squeezeNet to modify the backbone network of SSD, and compensated for the impact of the lightweight model on the detection accuracy by adding multiple convolutional layers and connecting branches to enhance the propagation ability and extraction ability of features. To ensure a comprehensive evaluation of the model's detection capabilities, all the models discussed in this study employed the technique of random weight initialization. This approach was utilized to validate the optimal structure of the model and its performance. The experimentation was conducted on both the PASCAL VOC 2007 benchmark dataset and an infrared image dataset encompassing five distinct categories of electrical equipment found within substations. The experimental outcomes indicate that this model offers an efficient approach for achieving lightweight, real-time detection of electrical apparatus.

## KEYWORDS

infrared image, single shot multibox detector (SSD), lightweight model, electrical equipment, real-time detection, object detection

## 1 Introduction

To address the growing need for sustainable energy, there is an accompanying requirement for the expansion and complexity of power systems. The power system needs continuous inspection and preventive maintenance to ensure its normal trouble-free operation (Lu et al., 2017; Song et al., 2018). Among them, detection of substation is very important, because the electrical problems in the substation will not only cause power system outage, local economic losses, and may even lead to casualties. Therefore, real-time and effective detection of substation is extremely crucial to ensure its safety and long-term operation (Menendez et al., 2017; Usamentiaga et al., 2018; Li et al., 2019).

Infrared thermal imaging technology has many advantages over other types of sensors, so it has become a widely accepted condition monitoring technology (Osornio-Rios et al., 2019; He et al., 2017). Infrared thermal image detection is a technology to diagnose whether the equipment is in good running state or not based on the thermal distribution of the equipment. This technology can be non-contact far away from the equipment to be detected, and the temperature measuring range is wide, so technicians can quickly scan and detect electrical equipment (Zou and Huang, 2015; Zhao et al., 2017; Ma et al., 2021).

Due to the substantial image data produced during substation detection, the conventional security inspection process necessitates significant time and extensive human and material resources for performing state assessment on electrical equipment. The extended periods of engagement for technicians in this task can result in fatigue and potential misdiagnosis (Jadin et al., 2015; Zhao et al., 2016; Wang et al., 2020). To circumvent the limitations inherent to manual analysis of infrared images, numerous intelligent methods for detecting electrical equipment have been proposed. Compared with the manual method, intelligent detection methods are more flexible, saves time and resources, and can greatly improve detection efficiency.

A summary of some methods is shown in Table 1. Wu and An (2014) proposed a new region-based ACM algorithm for feature extraction of non-uniform insulators. While this approach effectively addresses the challenge of segmenting images with significantly uneven textures, there is a need for further

enhancement in the adaptability of weight parameter selection. In the work cited as Reddy and Mohanta (2013), an approach was introduced to assess the condition of insulators. This method employs a combination of support vector machines and an adaptive neuro-fuzzy inference system. Additionally, it utilizes features extracted from a discrete orthogonal transformation to enhance its performance. However, the method based on color feature was sensitive to complex background and needed to adjust appropriate threshold parameters. A novel texture segmentation algorithm was proposed in Wu et al. (2012), which could divide complex aviation insulator image into subregions with smooth contours. Nonetheless, there is room for enhancement in refining the adaptive selection process for the weight parameters within this algorithm.

Over the past few years, the surge in computational power has prompted a heightened focus on deep learning (Lecun et al., 2015). Deep learning techniques are progressively finding broader applications in tasks such as image classification, fault detection, and object localization (Lan et al., 2019; Peng et al., 2019; Mei et al., 2020; Xi et al., 2020). The convolutional neural network (CNN) is a special type of deep learning model that has achieved excellent performance in competitions related to computer vision and image processing (Khan et al., 2020). This deep architecture helps learn more complex features from images. AlexNet (Krizhevsky et al., 2012) was one of the pioneering CNN architectures that achieved a breakthrough in image classification tasks. It introduced the concept of using multiple GPU units for training and utilized techniques like

TABLE 1 The summary of research methods.

Detection methods	Main advantages	Main disadvantages
(Wu and An, 2014) A new region-based ACM algorithm for feature extraction of non-uniform insulators	It effectively addresses the challenge of segmenting images with significantly uneven textures	The adaptability of the weight parameter selection of this method needs to be further enhanced
(Reddy and Mohanta, 2013) A method for evaluating insulator conditions using a combination of support vector machines and adaptive neuro-fuzzy inference systems	It utilizes features extracted from a discrete orthogonal transformation to enhance its performance	It based on color feature was sensitive to complex background and needed to adjust appropriate threshold parameters
(Wu et al., 2012) A novel texture segmentation algorithm	It can divide complex aviation insulator image into subregions with smooth contours	There is room for enhancement in refining the adaptive selection process for the weight parameters within this algorithm
(Ren et al., 2017) An approach for both insulator detection and fault identification	The method achieves high accuracy in detecting surface defects on insulators. It can be fine-tuned and improved over time with additional data. What's more, the method can be scaled to handle large datasets and leverage powerful hardware for faster training	The method is primarily designed for detecting surface defects through visual inspection. It may not capture other types of defects or structural issues that can occur in insulators
(Wang et al., 2020) An innovative autokinetic diagnosis technique centered on Mask R-CNN.	It addresses instance segmentation and temperature analysis of infrared images containing insulators	It is specifically designed for analyzing infrared images of insulators and may not be suitable for other types of image data
(He et al., 2020) An automated methodology rooted in YOLOv3	It efficiently recognizes locations and diagnose external power insulation equipment	It cannot achieve the same level of accuracy as more complex detection methods, especially in scenarios where precise insulator detection and recognition are critical
(Redmon et al., 2016) An insulator detection method based on SSD.	It is specially tailored for aerial imagery. It achieves good results with relatively small training datasets compared to some other deep learning models, reducing the need for extensive data collection and annotation	It focuses on object detection and bounding box localization of insulators. It does not provide detailed defect analysis or temperature-based insights, which may be necessary for a comprehensive inspection in some scenarios
(Liu et al., 2016) An inventive detection approach relying on FSSD was introduced for substation insulator infrared imagery	It cleverly integrated a unique module designed to enhance its capability in capturing distinctive attributes within the infrared images of substation insulators	-

ReLU activation functions and dropout to enhance model performance. The structure of CNN includes convolutional layer, pooling layer, and fully connected layer. Instead of using several convolutional layers, Visual Geometry Group (VGG) (Simonyan and Zisserman, 2014a) repeatedly uses a series of smaller  $3 \times 3$  convolutional filters compared to traditional CNN, creating a network with greater depth. There are numerous other advanced CNN architectures beyond VGG and AlexNet, each with its own unique features and design principles. For example, GoogLeNet (Szegedy et al, 2017) introduced the idea of Inception modules, which use multiple filter sizes in parallel to capture features at various scales. Residual Network (ResNet) (He et al, 2016a) employed residual connections to address the vanishing gradient problem, allowing the training of very deep networks. The traditional CNN approach serves as the foundation upon which these advanced architectures are built. They all share the core components of convolutional layers for feature extraction, pooling layers for spatial reduction, and fully connected layers for classification or regression. What distinguishes them are innovations in network depth, architectural design, and regularization techniques, which improve their ability to learn and represent complex patterns in images.

Within this domain, spatial pyramid pooling networks (He et al, 2015), Faster R-CNN (Ren et al, 2017), and Mask R-CNN (He et al, 2020) represent prevalent instances of two-stage object detection architectures. On the other hand, YOLO (Redmon et al, 2016), the single-shot multibox detector (Liu et al.), and CornerNet (Law and Deng, 2018) exemplify widely employed single-stage object detection frameworks. Drawing from the foundations of Faster R-CNN, Kang et al. (2019) introduced an approach for both insulator detection and fault identification. Meanwhile, Wang et al. (2020) put forward an innovative autokinetic diagnosis technique centered on Mask R-CNN. This technique addresses instance segmentation and temperature analysis of infrared images containing insulators. Additionally, Liu et al. (2020) outlined an automated methodology rooted in YOLOv3 to efficiently recognize locations and diagnose external power insulation equipment. The utilization of a deep learning algorithm facilitated the extraction of isolators, thereby allowing the examination of image attributes in the visible light spectrum. Miao et al. (2019) proposed an insulator detection method based on SSD, specially tailored for aerial imagery. In the study outlined by Zheng et al. (2020), an inventive detection approach relying on FSSD was introduced for substation insulator infrared imagery. This model cleverly integrated a unique module designed to enhance its capability in capturing distinctive attributes within the infrared images of substation insulators.

Although these methods have achieved great results in terms of detection accuracy, they have not carried out a good tradeoff study on model size, detection speed and detection accuracy. Given the substantial image data captured amidst complex backgrounds during substation inspections, the pursuit of a real-time detection model exhibiting enhanced precision within confined environments becomes imperative. This research amalgamates the distinctive network structure attributes of squeezenet (andola et al., 2016) to introduce a lightweight SSD detection model. This model is tailored for effecting object detection within infrared images portraying a diverse range of electrical equipment categories. The attainment of a lightweight design for this model is achieved through the

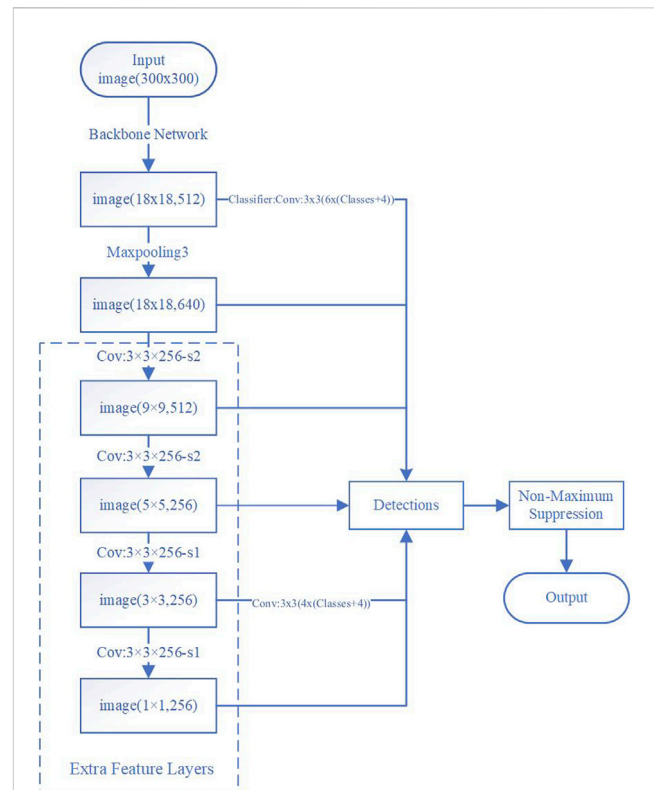
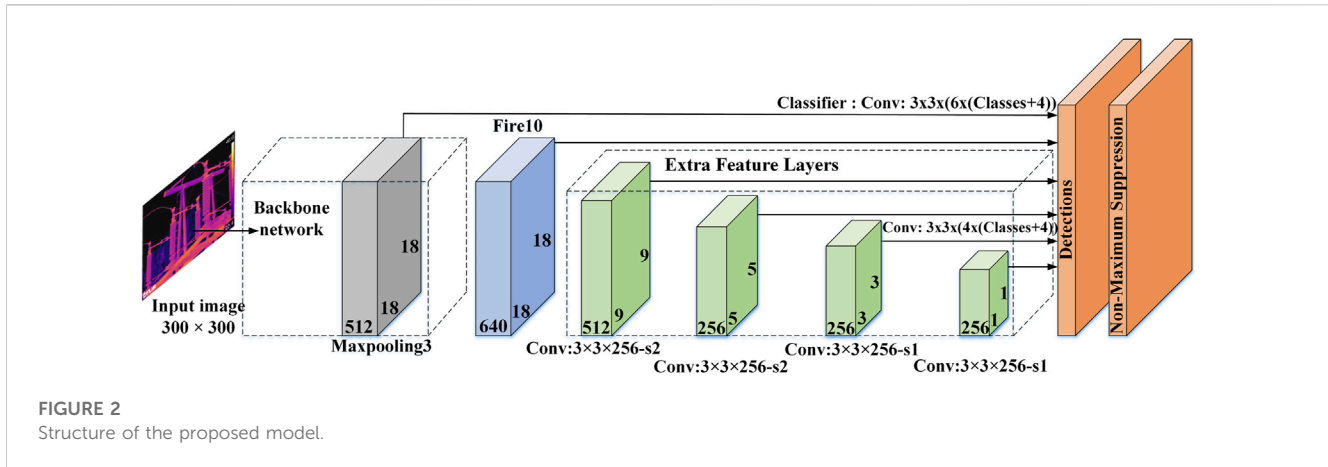


FIGURE 1  
Flowchart of the proposed method.

modification of the SSD's backbone network. However, it's essential to note that reducing the model's weight by lowering its network depth could potentially result in a diminishment of the model's nonlinear expression capabilities. This, in turn, may lead to a decline in the model's efficacy in feature fitting. Hence, to counterbalance the potential effect of the lightweight model on detection accuracy, this approach introduces multiple interconnected branches into the backbone network. This augmentation serves to bolster the feature's propagation capacity. Additionally, following the backbone network, several convolutional layers are incorporated to elevate the model's feature extraction proficiency.

This study presents a novel endeavor that conducts comprehensive experiments and discussions involving the utilization of squeezenet's distinctive structure to achieve the lightweight nature of SSD. To ensure thorough validation of this paper's model performance, all models adopt a randomized weight initialization strategy, excluding transfer learning. Through meticulous control of iteration count and training set proportions, the optimal model is deduced based on analytical assessment of experimental outcomes. Experiments are performed on both the Pascal VOC 2007 benchmark dataset and separate datasets encompassing five categories of electrical equipment. The outcomes reveal that, compared to SSD and other lightweight models, the model established within this paper not only achieves lightweight status but also successfully adheres to the prerequisites of real-time detection, yielding commendable detection efficacy.



## 2 The lightweight detector model

Given the acquisition of numerous infrared images of electrical equipment against intricate backgrounds during substation safety inspections, the enhancement of SSD's lightweight attributes is pursued in this study. This endeavor is aimed at enabling efficient detection of electrical equipment within constrained environments, while concurrently fulfilling the demands of real-time detection. The model's configuration is illustrated in Figures 1, 2.

### 2.1 Improved SSD model

The detection model introduced in this study employs the inherent squeezeNet network characteristics to achieve lightweight functionality. It additionally integrates SSD's multi-scale detection approach to facilitate object detection. The model workflow is shown in Figure 2.

Firstly, conv10 and global maximum pooling layer are deleted on the basis of squeezeNet network, and fire10 is added. Then the modified structure replaces the VGG 16 (Simonyan and Zisserman, 2014b) structure as the backbone network of SSD. In the meantime, in order to make up for the impact of lightweight on detection accuracy, a number of convolutional layers with gradually decreasing scale are added after the backbone network, and then add multiple bypass connections in the backbone network. Feature maps produced through a deeper convNet network have the capacity to embody heightened semantic information coupled with translation invariance. While this confers benefits to object recognition, it may prove less suitable for precise object localization. Due to its enhanced capacity for detailed information, the shallow feature layer proves advantageous for object detection (Li and Zhou, 2017). To optimize the utilization of insights from the model's shallow layers, this study integrates two feature maps within the backbone network to enhance detection capabilities.

### 2.2 The lightweight detector model

#### 2.2.1 Enhanced object detection framework

The cornerstone of this paper's framework encompasses the fusion of the single-stage object detection algorithm SSD and the

compact squeezeNet compression network. The SSD object detection algorithm showcases commendable performance in terms of both speed and accuracy, while additionally exhibiting robust adaptability to shifts in object scale. The SSD model consists of two distinct segments. The primary facet entails the central VGG16 architecture network, while the ensuing aspect entails an integrated convolutional layer structure aimed at augmenting detection precision and orchestrating a gradual diminution in the feature map's dimensions to fashion a feature pyramid structure. Propelled by a feedforward convolutional neural network, the SSD network enhances the algorithm's detection prowess by embracing multi-scale detection. In tandem, the incorporation of 4–6 default boxes, each with distinct proportions, into the feature layers of varying scales substantially bolsters the detection efficacy across objects spanning diverse scales.

In contrast to the AlexNet architecture, the accuracy of squeezeNet experiences a minor reduction, albeit accompanied by a substantially diminished parameter count in comparison to that of AlexNet. To curtail the parameter load within squeezeNet, three distinct strategies are implemented:

- (1) Substituting the  $1 \times 1$  convolution kernel for the  $3 \times 3$  convolution kernel. While the  $3 \times 3$  convolution kernel yields enhanced feature extraction outcomes from the network, squeezeNet employs a  $1 \times 1$  convolution kernel as a replacement. This strategic shift is aimed at reducing the model's parameter count while maintaining the integrity of feature extraction.
- (2) Minimizing Input Channel Count. To maintain parameter efficiency, the squeezeNet architecture takes measures to decrease the number of  $3 \times 3$  convolution kernels, as well as diminishes the input channel count associated with these  $3 \times 3$  convolution kernels.
- (3) Delaying the Pooling Process. The pooling operation plays a crucial role in shaping the ultimate performance of the network. By deferring and minimizing the pooling operation, it becomes possible to cultivate a more expansive feature map within the initial convolution layer of the network. Under equivalent circumstances, a larger feature map is poised to yield enhanced detection outcomes.

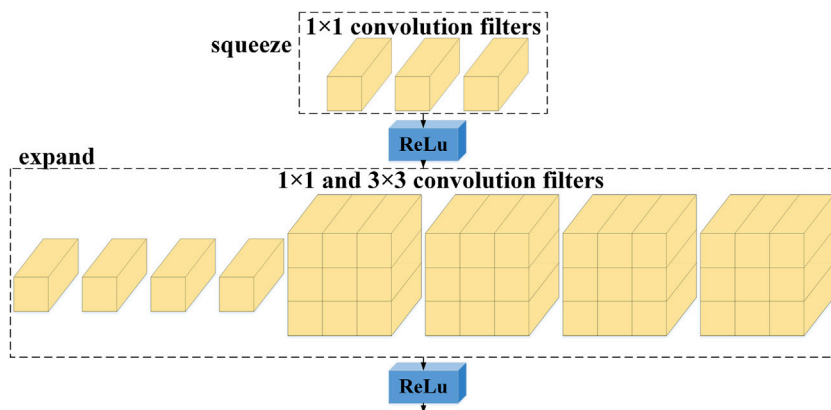


FIGURE 3 Fire module structure.

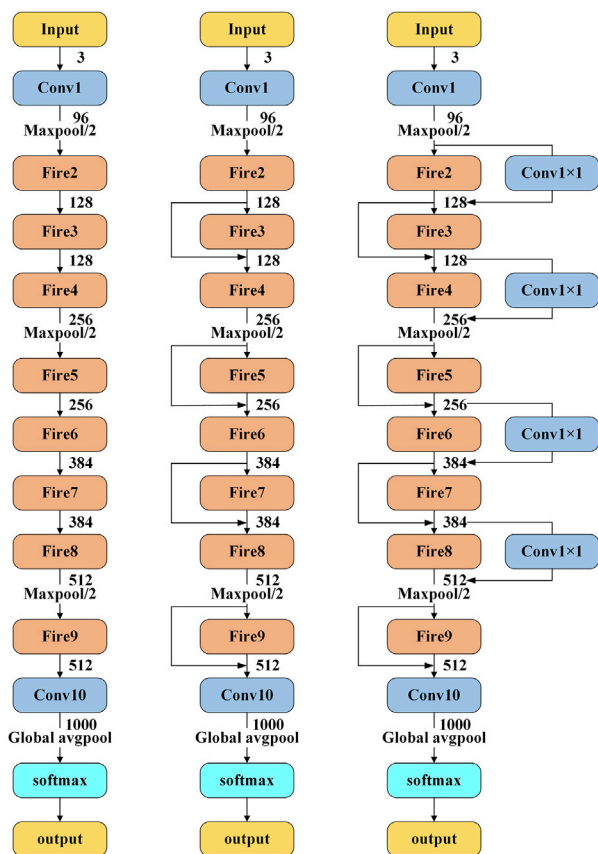


FIGURE 4 Three kinds of squeezeNet1\_0 network structure diagram. On the left: squeezeNet; In the center: squeezeNet featuring a straightforward bypass; On the right: squeezeNet incorporating an intricate bypass configuration.

### 2.2.2 Fire module and bypass connection

The pivotal component within squeezeNet architecture is undoubtedly the fire module. This structural element predominantly consists of two distinct convolutional layers: the

squeeze layer and the expand layer, as visually depicted in Figure 3. The Squeeze layer essentially operates as a convolutional layer comprising multiple  $1 \times 1$  convolution kernels. Conversely, the expand layer serves as a convolutional layer that amalgamates an equivalent quantity of  $1 \times 1$  and  $3 \times 3$  convolution kernels. The fire module encapsulates the initial pair of design strategies inherent to squeezeNet. Integration of this module significantly contributes to the substantial reduction of model parameters.

In Andola et al. (2016), for the purpose of enhancing the model's precision, additional structures with different bypasses are explored by referring to the residual network (He et al, 2016b):

- (1) Some fire modules have simple bypass connections.
- (2) In the first case, complex bypass connections are connected to other fire modules.

In Andola et al. (2016), only squeezeNet1\_0 is discussed, but not squeezeNet1\_1. Three variants of squeezeNet1\_0 are shown in Figure 4.

SqueezeNet1\_1 is further compressed on the basis of 1\_0 to achieve further reduction of parameters by advancing the maximum pooling layer. Within this manuscript, with the focus directed toward detecting a range of electrical apparatus, a comprehensive experimentation was conducted for the inaugural instance, delving into the assessment of diverse configurations of the squeezeNet network on the efficacy of detection. This endeavor sought to ascertain the most suitable lithe, real-time detection model. As revealed by the empirical findings presented herein, the foundational framework of the model put forth within this composition was formulated by drawing inspiration from the structural attributes of squeezeNet1\_0, featuring intricate by-pass pathways.

### 2.2.3 Object detection configuration and optimization

To facilitate the detection of objects displaying varying proportions, the method of feature pyramid multi-scale detection is embraced. The corresponding anchor sizes are established based on the receptive field dimensions of each layer. Varied quantities of

default boxes are generated on the feature layers of distinct scales. Within these layers, the prior boxes affiliated with each feature unit, present within each feature diagram, necessitate computation of scores for  $C$  categories along with 4 positional offsets relative to the original default box. Within every feature unit on the feature map, a total of  $(C + 4) \times k$  prior boxes are formulated. When considering feature maps of dimensions  $m \times n$ , the outcome yields  $m \times n \times (C + 4) \times k$  *a priori* boxes. This process yields an extensive assemblage of prior boxes, which, due to their redundancy, augments computational overhead and diminishes computational efficiency. Consequently, the employment of non-maximum suppression is adopted to address the processing of detection outcomes.

Within the framework of this study, the formula for calculating the default box pertaining to each feature map is depicted as follows:

$$s_k = s_{min} + \frac{s_{max} - s_{min}}{m - 1} (k - 1) \quad k \in [1, m] \quad (1)$$

Herein,  $s_k$  denotes the ratio assigned to the default box within the  $k$ th-level feature map;  $s_{min}$  corresponds to the proportion attributed to the default box within the lowest-level feature map, while  $s_{max}$  pertains to the proportion assigned to the default box within the highest-level feature map. These values are preconfigured as 0.2 and 0.9, respectively. The parameter  $m$  signifies the quantity of feature maps encompassed in this manuscript. Furthermore, the default aspect ratio of the default box for each feature map layer is represented as  $a_y$  and takes on values within the interval  $(\frac{1}{3}, \frac{1}{2}, 1, 2, 3)$ .

During the detection process, each factual box present within the feature map undergoes alignment with a corresponding default box, thereby eliminating excessive redundancy associated with default boxes. In scenarios where a default box aligns with any factual box, the default box boasting the most substantial intersection ratio is classified as a positive instance. Furthermore, if a default box aligns with any factual box and the intersection ratio surpasses 0.5, it also gains classification as a positive example. Given that a considerable majority of the generated default boxes are categorized as negative samples, with their quantity notably exceeding that of the selected positive samples, a course of action is warranted. To achieve this, the default boxes linked with negative samples are arranged in descending order based on their confidence scores. Subsequently, the default box possessing the highest confidence score is elected as the negative sample. This approach is instrumental in maintaining a balanced ratio between positive and negative samples, specifically at a 1:3 proportion.

Amidst the training phase, varying-scale prior boxes are acquired across distinct layers of feature maps. Subsequently, computations are conducted to derive the localization loss and the confidence loss. The overarching objective loss function aggregates these losses, being a weighted amalgamation of the localization and confidence losses. The expression for the calculation is depicted as follows:

$$Loss(m, n, o, p) = \frac{1}{Q} (Loss_{conf}(m, n) + xLoss_{loc}(m, o, p)) \quad (2)$$

among them  $m$  value 0 or 1 indicates the prior box whether matches the ground truth box;  $n$  is confidence of the categories; the actual

information of the prediction box is represented by  $o$ ; the actual information of the ground truth box is represented by  $p$ ;  $Q$  is represented amount of the matched default boxes;  $Loss_{conf}$  is represented the confidence loss,  $Loss_{loc}$  is represented the localization loss,  $x$  is represented the weight of them, we set  $x$  as 1 in the experimental process. The localization loss encompasses a Smooth L1 loss (36) applied to the estimated box ( $o$ ) and the actual ground truth box ( $p$ ) parameters. As for the confidence loss, it entails the softmax loss associated with the confidence values across a spectrum of classes.

### 3 Experiment and analysis

To verify the effectiveness of the method in this paper, the PASCAL VOC 2007 benchmark dataset and an infrared image dataset centered on substation electrical equipment are used to evaluate the model, analyze and compare the experimental results.

#### 3.1 Experimental description

For the purpose of assessing the proposed model's performance, all experiments were conducted using a PC running Ubuntu 16.04LTS, equipped with 64 GB of memory, CPU (Intel Xeon W-2145 3.70 GHz  $\times$  8) and two GPUs (NVIDIA Geforce RTX 2080ti with 11 GB memory size). Within this study, the dimensions of the input image utilized for experimentation encompass 300  $\times$  300 pixels. Each batch is comprised of 16 images. An initial learning rate of 0.001 is designated, complemented by a momentum value of 0.9. Additionally, a weight decay factor of 0.0005 is applied. The optimization scheme employed for refining the model involves the utilization of the stochastic gradient descent algorithm.

The experimental datasets featured in this study encompass the PASCAL VOC 2007 benchmark dataset and an infrared image dataset centered on substation electrical equipment. All images employed in these experiments adhere to the RGB format. Notably, the infrared image dataset concerning electrical equipment originates from the State Grid Corporation of China. Specifically, the Pascal VOC 2007 benchmark dataset encompasses a spectrum of 20 distinct categories. Within this dataset, the training subset includes 5,011 images, while the testing subset comprises 4,952 images. The infrared image dataset of electrical equipment contains 5 categories of electrical equipment, which is surge arrester, circuit breaker, isolating switch, instrument trans-former and insulator. The number of infrared images of these electrical equipment is 412, 456, 1,421, 1,355, and 2,100 respectively, and some infrared images with multiple objects. The acquired dataset of infrared images for electrical equipment is partitioned into two distinct categories. Within this division, a specific proportion is designated for utilization as training samples to facilitate model construction, while the remaining subset is allocated for employment as testing samples. It is a well-known fact that the ratio assigned to training sets significantly impacts the learning process and consequently influences the outcomes of the model. Hence, the present study delves into an examination of how distinct

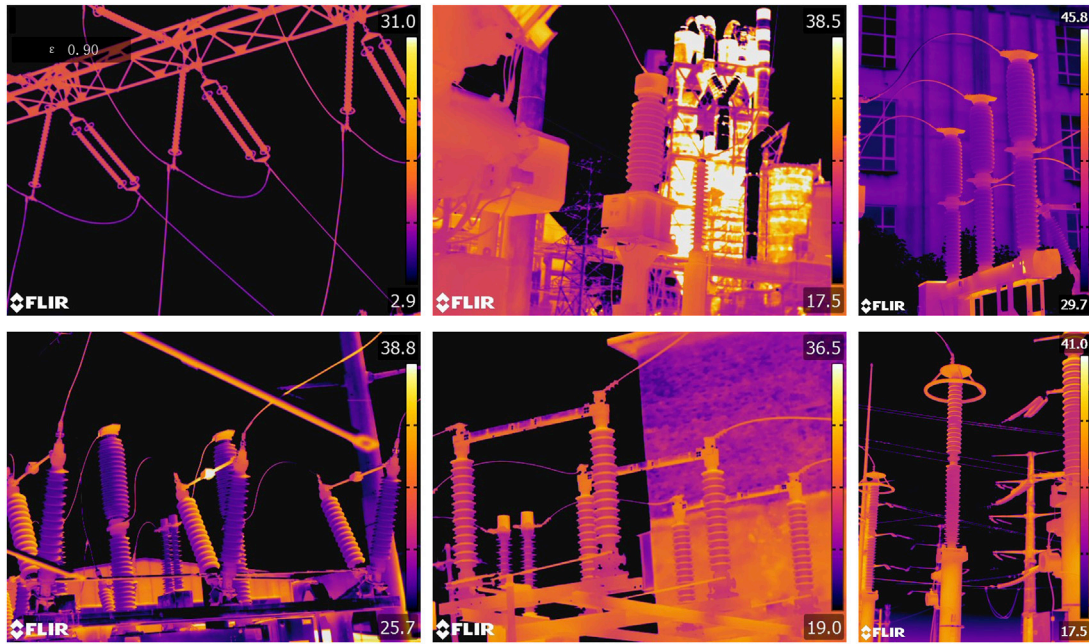


FIGURE 5 Some images of the dataset.

ratios of training samples influence the experimental outcomes within the experimental process. The overarching goal is to discern the most suitable allocation proportion. For visual context, refer to Figure 5, which portrays a collection of infrared images comprising various forms of electrical equipment.

Addressing the challenge of training instability stemming from fluctuations in loss during the training phase, this study adopts a strategy akin to the “warm-up” approach described in Andola et al. (2016). This technique entails commencing with a modest, incremental learning rate during the initial 5 stages. Subsequently, following the “warm-up” stage, the learning rate reverts to a conventional learning strategy. Additionally, the experiment in this paper employs the strategy of random weight initialization for the model. This decision is made to comprehensively assess the model’s performance.

### 3.2 Evaluation metrics

Within this paper, the evaluation indices employed to assess the model encompass the time taken for detecting a singular image, the mean average precision (mAP), and the dimensions of the model itself. These metrics serve as benchmarks for dissecting the experimental findings and facilitating a comprehensive model comparison. During the computation of average precision (AP) and mean average precision (mAP), the predefined intersection over union (IOU) threshold defaults to 0.5. Here, mAP represents the arithmetic mean of the AP values amassed across all categories. The AP value is calculated by determining the area enclosed by the precision-recall (P-R) curve. The subsequent methodology for computing AP, precision rate, and recall rate is illustrated as follows:

$$\begin{cases} AP = \int_0^1 p(r)dr \\ precision = \frac{TP}{TP + FP} \\ recall = \frac{TP}{TP + FN} \end{cases} \quad (3)$$

the precision rate  $p$  signifies the ratio of accurately identified cases, whereas the recall rate  $r$  indicates the proportion of correctly detected actual instances. Within the set of variables,  $TP$  represents the number of cases accurately identified,  $FP$  corresponds to instances incorrectly categorized as positive, and  $FN$  denotes the count of true instances wrongly labeled as false negatives, reflecting the quantity of misclassified objects.

### 3.3 Outcomes from the PASCAL VOC 2007 experimentation

Due to the potential impact of the quantity of training iterations on the outcomes of the model’s detection capabilities, this study investigates the effect of varying training iterations on the training results of the model through empirical investigations. This exploration aims to determine an optimal number of training iterations for conducting experiments involving infrared imagery of electrical apparatus. In this study, to validate the most efficient lightweight model, the lightweight models with different structures proposed in this paper will be compared with SSD, Mobile-ssd, Squeeze-ssd and Tiny-yolov3. Figure 6 displays the outcomes obtained from distinct models across varying training iterations. Among them, Sq1\_0 and Sq1\_1 represent improved models based on squeezeNet1\_0 and squeezeNet1\_1 as the backbone of the

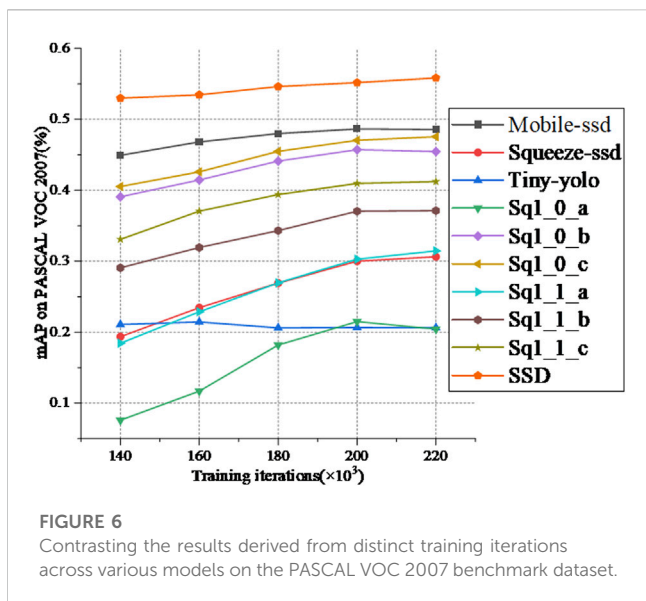


FIGURE 6 Contrasting the results derived from distinct training iterations across various models on the PASCAL VOC 2007 benchmark dataset.

TABLE 2 Detection time and model size of various models on the PASCAL VOC 2007 benchmark dataset.

Method	Model size (MB)	Detection time (ms/image)
①	41.2	8.517
②	22.1	6.878
③	34.9	4.591
④	25.8	6.417
⑤	25.8	6.463
⑥	27.2	6.907
⑦	24.4	6.414
⑧	24.4	6.461
⑨	25.7	6.888
⑩	105.2	30.378

network, respectively, and a, b and c respectively represent the improved models with no bypass, simple by pass and complex bypass.

Observing the data presented in Figure 6, it becomes evident that the detection accuracy of distinct models demonstrates a general inclination towards improvement as the number of iterations increases. Notably, among these, Mobile-SSD, SSD, and Sq1\_0\_c models stand out for their superior detection outcomes. However, it's worth noting that an escalation in the number of iterations corresponds to a substantial rise in the duration of the training process. As a result of a comprehensive deliberation, a pragmatic decision is made, leading to the selection of 20,000 iterations as the optimal count for experimental training iterations dedicated to infrared images of electrical equipment.

Table 2 presents a comparative analysis of detection time and model dimensions among various models, performed on the PASCAL VOC 2007 benchmark dataset. Additionally, Table 3 provides an intricate breakdown of the comprehensive

experimental outcomes yielded by diverse models when subjected to 200 k iterations on the PASCAL VOC 2007 benchmark dataset. The findings from Figure 6 and Tables 2, 3 collectively demonstrate that while the detection precision of Sq1\_0\_c is lesser compared to SSD, its model dimensions are nearly a quarter of SSD's, and its detection velocity is enhanced by a factor of 4.4 in comparison to SSD. The detection accuracy of Sq1\_0\_c is slightly less than that of Mobile-ssd, but the model size of Sq1\_0\_c is 1.5 times smaller than that of Mobile-ssd, and the speed is faster than that of Mobile-ssd. Even though the detection duration for Sq1\_0\_c is marginally extended compared to Tiny-yolo, its model proportions are diminished in contrast to Tiny-yolo. Moreover, the detection precision significantly surpasses that of Tiny-yolo. Compared with the other models, Sq1\_0\_c has higher detection accuracy than the others, although the model size and detection time is slightly increased. Among them, method ①~⑩ respectively represent the top-down models in Figure 6.

### 3.4 Results on infrared images of electrical equipment

In this portion, an investigation is undertaken using a dataset comprising infrared images of electrical equipment within substations. The focus of this phase involves a comparative analysis between the enhanced lightweight models and alternative model counterparts. Insights drawn from Tables 2, 3 reveal a parity in model sizes between models 4 and 5, yet it's evident that the detection accuracy of model 4 markedly lags behind that of model 5. To optimize resource allocation, both in terms of experimental time and costs, the evaluation of Model 4 is omitted concerning the infrared dataset for electrical equipment. In a similar vein, the assessment of Model 7 is likewise omitted.

As the volume of data increases, the model's performance is expected to improve, with the proportion of training samples playing a pivotal role in the model's learning process. Notably, the dataset employed in this study is sourced from the State Grid Corporation of China and differs from the division structure of the PASCAL VOC 2007 benchmark dataset. To forestall any inadvertent influence of training set composition on experimental outcomes and consequent interpretation of model efficacy, a strategy is employed wherein various proportions (ranging from 0.3 to 0.9 in intervals of 0.1) of training samples are selected. Employing uniform experimental protocols, this approach is used to assess the impact of varying sample quantities on the results. In order to mitigate the potential instability of both training and testing sets, the dataset for this experiment is subject to randomized selection. For instance, a ratio of 0.8 is utilized for the training set.

Initially, a random selection process is employed, where 80% of the available infrared images for a specific category of electrical equipment are chosen as the training set for that equipment type. The residual 20% of images are designated as the corresponding testing set. Subsequently, this process is replicated across various equipment categories, culminating in the establishment of a comprehensive 0.8 scale training set through amalgamating the diverse training and testing sets. This overarching dataset encompasses multiple infrared images encompassing various types of electrical equipment objects.



TABLE 3 The results of PASCAL VOC 2007 benchmark dataset (IOU = 0.5, training iterations = 200 k).

Method	mAP	Aero	Bike	Bird	Boat	Bottle	Bus	Car	cat	Chair	Cow	Table	Dog	Horse	mbike	Person	Plant	Sheep	Sofa	Train	Tv
⑩	0.552	0.644	0.689	0.432	0.482	0.135	0.716	0.794	0.548	0.273	0.528	0.569	0.524	0.746	0.707	0.650	0.238	0.465	0.565	0.755	0.579
①	0.487	0.514	0.677	0.313	0.331	0.080	0.655	0.628	0.567	0.278	0.352	0.581	0.560	0.711	0.648	0.561	0.193	0.334	0.551	0.701	0.501
⑥	0.471	0.559	0.616	0.312	0.272	0.063	0.636	0.668	0.541	0.250	0.397	0.540	0.442	0.703	0.644	0.579	0.174	0.350	0.492	0.665	0.509
⑤	0.457	0.543	0.614	0.312	0.266	0.061	0.642	0.653	0.509	0.258	0.365	0.497	0.439	0.687	0.615	0.581	0.147	0.345	0.496	0.650	0.469
⑨	0.410	0.527	0.541	0.278	0.232	0.038	0.607	0.609	0.480	0.185	0.286	0.484	0.400	0.629	0.540	0.536	0.098	0.300	0.411	0.607	0.405
⑧	0.370	0.485	0.496	0.222	0.199	0.034	0.571	0.581	0.416	0.155	0.239	0.425	0.318	0.598	0.505	0.505	0.082	0.259	0.354	0.585	0.379
⑦	0.303	0.421	0.379	0.162	0.125	0.006	0.453	0.571	0.336	0.087	0.197	0.305	0.299	0.580	0.473	0.474	0.021	0.217	0.200	0.454	0.298
②	0.300	0.425	0.357	0.145	0.101	0.003	0.376	0.614	0.356	0.119	0.278	0.243	0.271	0.537	0.436	0.447	0.029	0.309	0.259	0.355	0.339
④	0.215	0.343	0.226	0.076	0.070	0.003	0.332	0.439	0.269	0.055	0.107	0.224	0.171	0.444	0.345	0.338	0.009	0.100	0.154	0.358	0.237
③	0.207	0.288	0.371	0.057	0.138	0.033	0.194	0.458	0.209	0.136	0.187	0.057	0.077	0.305	0.296	0.398	0.038	0.248	0.083	0.298	0.260

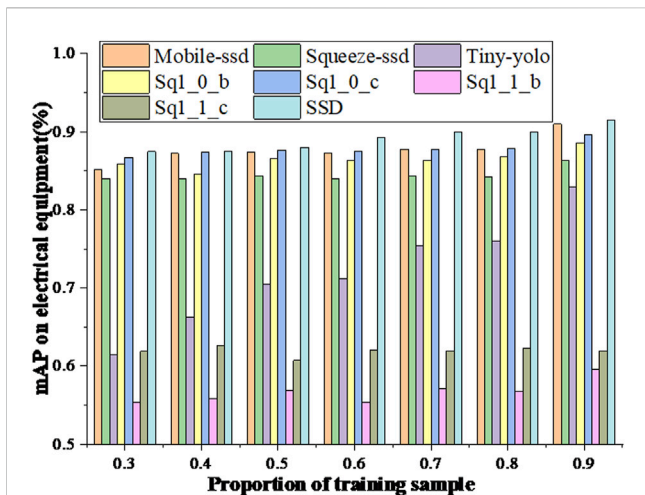


FIGURE 7 Comparison of the detection results on the training set of different proportions of the infrared images of electrical equipment.

TABLE 4 Detection time and model size of different models.

Method	Model size (MB)	Detection time (ms/image)
①	32.6	10.54
②	13.8	8.474
③	34.7	2.723
⑤	18.5	7.252
⑥	19.8	7.459
⑧	17.9	6.814
⑨	19.2	7.025
⑩	97.1	28.47

Depicted in Figure 7 is a comparative analysis of detection outcomes among diverse models across varying training set proportions. As observed in Figure 7, a discernible trend emerges wherein model detection efficacy escalates in tandem with the expansion of the training set. Notably, beyond a training set

proportion of 0.3, Mobile-SSD, SSD, and Sq1\_0\_c consistently exhibit superior detection outcomes. Table 4 shows the detection time and model size of different models on the infrared image dataset of electrical equipment. Presented in Table 5 are the comprehensive outcomes derived from experiments involving various models on the infrared image dataset related to electrical equipment, with a training set ratio of 0.8. The observations drawn from Figure 7 and Tables 4, 5 collectively underscore that the model dimensions of Sq1\_0\_c are approximately one-fifth the size of SSD’s, accompanied by a detection speed that is expedited by a factor of 3.8 in comparison to SSD. Despite exhibiting a lower accuracy than SSD, it still attains a commendable detection precision of 87.82% through the implementation of the model weight random initialization strategy. Sq1\_0\_c features a model size approximately 1.6 times leaner than Mobile-SSD, accompanied by an enhanced detection speed. Moreover, Sq1\_0\_c boasts a detection accuracy surpassing that of Mobile-SSD. In contrast to alternative models, Sq1\_0\_c attains experimental outcomes akin to those witnessed in the context of the PASCAL VOC 2007 benchmark dataset. Therefore, a comprehensive evaluation positions Sq1\_0\_c as the most optimal performer.

Given the need to impartially assess the model’s performance through the testing set, it becomes imperative to deliberate on the adequacy of the testing set’s proportion. Ensuring comprehensive generalization and representative evaluation is challenging when the testing set is inadequately sized. To this end, a judicious decision is made to adopt a final ratio of 0.8, ensuring a balanced compromise.

To demonstrate the advantages of the proposed method, comparisons with the latest SSD variants are also conducted, as shown in Table 6. These results are obtained under the optimal dataset allocation ratios. It can be observed that the proposed method achieves the fastest detection speed and the smallest memory footprint, with detection accuracy only slightly lower than that of the FSSD model.

Figure 8 shows the detection effect of Sq1\_0\_c on some infrared images of electrical equipment. The anticipatory results associated with surge arresters, circuit breakers, isolating switches, instrument transformers, and insulators respectively correspond to the white, cyan, yellow, green, and red bounding boxes delineated within the illustration. The bounding box denotes the location of the object which the model located, and the value that is above this bounding box is the confidence of the model’s prediction category. By

TABLE 5 Comparison of different models.

Method	Arrester	Breaker	Switch	Transformer	Insulator	mAP
①	0.8962	0.7988	0.9031	0.9042	0.8868	0.8778
②	0.7704	0.7777	0.8948	0.8806	0.8899	0.8427
③	0.6848	0.6875	0.7991	0.8295	0.7976	0.7597
⑤	0.8814	0.7880	0.8958	0.8924	0.8812	0.8678
⑥	0.8903	0.8021	0.9050	0.9027	0.8908	0.8782
⑧	0.4236	0.4166	0.6147	0.6588	0.7270	0.5681
⑨	0.4516	0.4783	0.6848	0.7182	0.7823	0.6230
⑩	0.8980	0.8886	0.9065	0.9033	0.9008	0.8994

TABLE 6 Comparison with other methods.

Method	Arrester	Breaker	Switch	Transformer	Insulator	mAP	Detection time (ms/image)	Model size (MB)
FSSD	0.8923	0.9922	0.8997	0.8967	0.8951	0.9152	32.6	130.1
Our method	0.8903	0.8021	0.9050	0.9027	0.8908	0.8782	19.8	7.46

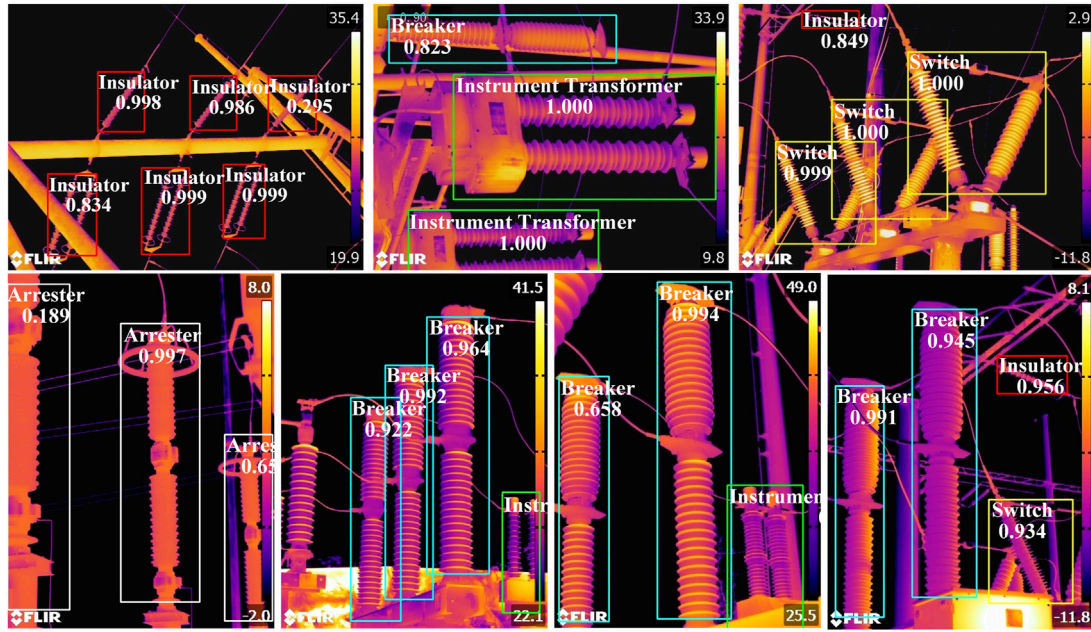


FIGURE 8 Detection results of some infrared images of electrical equipment based on Sq1\_0\_c.

conducting an in-depth analysis of the experiment, this study identifies the optimal lightweight model. This model not only showcases enhanced performance but also demonstrates precision in accurately pinpointing the location of equipment objects within infrared images.

### 4 Discussion

We have successfully developed a lightweight model that exhibits substantial improvements in model size and detection speed compared to traditional SSD and other lightweight models like Mobile-SSD and Tiny-yolov3. This achievement underscores the efficiency of the proposed approach, which can be advantageous in resource-constrained environments. This study demonstrates a careful balance between model accuracy and detection speed. While the optimal lightweight model may not match the accuracy of SSD, it still maintains a high level of detection accuracy, making it suitable for applications where real-time processing is critical. Due to its compact model size, the proposed lightweight model can operate efficiently on hardware with limited computational resources. This makes it suitable for deployment on edge devices, drones, and other platforms where computational power is restricted. The slightly increased cost of the optimal model compared to other improved models is offset by its superior detection accuracy and smaller model

size compared to SSD. This cost-efficiency is important for practical applications where budget constraints may be a concern. The lightweight and efficient model developed in this study has direct applications in scenarios where computational resources are limited, such as autonomous drones, IoT devices, and embedded systems. Real-time object detection is crucial in security and surveillance applications. The improved lightweight model can be deployed in security cameras and surveillance systems for enhanced object recognition and tracking. In industrial settings, where real-time detection of objects or anomalies is essential, the lightweight model can be used in robots, automated quality control systems, and equipment monitoring within substations, as demonstrated in this study. Lightweight models with high-speed detection capabilities can be deployed in intelligent transportation systems for traffic monitoring, pedestrian detection, and vehicle counting. The lightweight model's ability to process images efficiently can be leveraged for environmental monitoring applications, such as wildlife conservation, where quick and accurate detection is crucial.

Considering the promising results obtained in the physical detection of electrical components within substations, we propose a future research direction that extends the application of the approach of this study to off-grid and autonomous load scenarios, particularly in the field of power quality analysis. This direction aims to harness the capabilities of our lightweight and efficient model for the identification and monitoring of critical

power parameters in off-grid and autonomous power generation systems (Krishna et al., 2023; Krishna et al., 2022).

## 5 Conclusion

In this paper, detailed experiment and discussions on the effects of various squeezenet structures on the detection efficacy of SSD model were carried for the first time, and an improved lightweight model was proposed. The model improved the feature extraction ability and model fitting ability by adding multiple different connection branches and convolutional layers, so as to make up for the impact of light-weight model on the detection accuracy. According to the experimental results: 1) Concerning the PASCAL VOC 2007 benchmark dataset, despite the optimal light-weight model exhibiting a detection accuracy lower than that of SSD, it boasts a model size approximately one-fourth that of SSD and a detection speed accelerated by a factor of 4.4 compared to SSD. The precision of the optimal model was marginally lower than that of Mobile-ssd, however, its model dimensions were nearly 1.5 times more compact than those of Mobile-ssd, and its detection speed surpassed that of Mobile-ssd. The model size of the optimal model was smaller than that of Tiny-yolov3, although the detection time of the optimal model was slightly longer than that of Tiny-yolov3, the detection accuracy was much greater than Tiny-yolov3; 2) The identical experimental procedure was conducted on the infrared image dataset related to electrical equipment within a substation. The outcomes of these experiments revealed that the model size of the optimal lightweight model was nearly 5 times less than that of SSD, and the detection speed demonstrated a 3.8-fold improvement over SSD. Although the accuracy was lower than SSD, it still achieves a high detection accuracy of 87.82%. The model size of the optimal model was nearly 1.6 times smaller than that of Mobile-SSD, the detection speed was also improved, and the detection accuracy was higher than that of Mobile-SSD. Compared with other models, the conclusion was consistent with the experimental results on PASCAL VOC 2007 benchmark dataset. Although the model size and detection time of the optimal model were slightly increased compared with other improved models, the detection accuracy was the highest, and the model size was very small compared with SSD, so the slightly increased cost of the optimal model compared with the other improved models could be ignored.

Although this study has a good detection effect while meeting the requirements of lightweight and real-time detection, there are still some limitations. As an illustration, during the pursuit of achieving a lightweight design, it is possible to enhance the model's detection accuracy while still upholding its lightweight

attributes. As a result, our focus in future research must address subsequent tasks aimed at further optimizing the model.

## Data availability statement

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

## Author contributions

CQ: Writing—original draft. ZC: Writing—original draft. XC: Writing—original draft. YB: Writing—original draft. TH: Writing—original draft. SH: Writing—review and editing. JL: Writing—review and editing. YL: Writing—review and editing. FT: Writing—review and editing. ML: Writing—review and editing.

## Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This endeavor received partial support from the National Natural Science Foundation of China, with Grants 52277139 and 51907034, as well as backing from the Specific Research Project of Guangxi for Research Bases and Talents, with Grant 2020AC19010.

## Conflict of interest

Authors CQ, ZC, XC, YB, TH, FT, and ML were employed by State Grid Zhengzhou Power Supply Company of Henan Electric Power Company.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## 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.

## References

- Andola, F. N., Han, S., Moskewicz, M. W., Ashraf, K., Dally, W. J., and Keutzer, K. (2016). SqueezeNet: alexnet-level accuracy with 50x fewer parameters and < 0.5 MB model size. <https://arxiv.org/abs/1602.07360>.
- Girshick, R. (2015) "Fast r-cnn," in Proceedings of the IEEE international conference on computer vision, Santiago, Chile, December 2015, 1440–1448. doi:10.1109/ICCV.2015.169
- He, H. Y., Lee, W., Luo, D. S., and Cao, Y. J. (2017). Insulator infrared image denoising method based on wavelet generic gaussian distribution and map estimation. *IEEE Trans. Ind. Appl.* 53, 3279–3284. doi:10.1109/TIA.2017.2691309
- He, K. M., Gkioxari, G., Dollár, P., and Girshick, R. (2020). Mask r-cnn. *IEEE Trans. Pattern Anal. Mach. Intell.* 42, 386–397. doi:10.1109/TPAMI.2018.2844175
- He, K. M., Zhang, X. Y., Ren, S. Q., and Sun, J. (2016a) "Deep residual learning for image recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition, Las Vegas, NV, USA, June 2016, 770–778. doi:10.1109/CVPR.2016.90
- He, K. M., Zhang, X. Y., Ren, S. Q., and Sun, J. (2015). Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* 37, 1904–1916. doi:10.1109/TPAMI.2015.2389824

- He, K., Zhang, X., Ren, S., and Sun, J. (2016) "Deep residual learning for image recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition, Las Vegas, NV, USA, June 2016, 770–778.
- Jadin, M. S., Taib, S., and Ghazali, K. H. (2015). Finding region of interest in the infrared image of electrical installation. *Infrared Phys. Technol.* 71, 329–338. doi:10.1016/j.infrared.2015.05.010
- Kang, G. Q., Gao, S. B., Yu, L., and Zhang, D. K. (2019). Deep architecture for high-speed railway insulator surface defect detection: denoising autoencoder with multitask learning. *IEEE Trans. Instrum. Meas.* 68, 2679–2690. doi:10.1109/TIM.2018.2868490
- Khan, A., Sohail, A., Zahoora, U., and Qureshi, A. S. (2020). A survey of the recent architectures of deep convolutional neural networks. *Artif. Intell. Rev.* 53 (8), 5455–5516. doi:10.1007/s10462-020-09825-6
- Krishna, V. B. M., Sandeep, V., Narendra, B. K., and Prasad, K. R. K. V. (2023). Experimental study on self-excited induction generator for small-scale isolated rural electricity applications. *Results Eng.* 18, 101182. doi:10.1016/j.rineng.2023.101182
- Krishna, V. B. M., Sandeep, V., Murthy, S. S., and Yadlapati, K. (2022). Experimental investigation on performance comparison of self-excited induction generator and permanent magnet synchronous generator for small-scale renewable energy applications. *Renew. Energy* 195, 431–441. doi:10.1016/j.renene.2022.06.051
- Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Adv. neural Inf. Process. Syst.* 25. doi:10.1145/3065386
- Lan, S., Chen, M., and Chen, D. (2019). A novel HVDC double-terminal non-synchronous fault location method based on convolutional neural network. *IEEE Trans. Power Deliv.* 34, 848–857. doi:10.1109/TPWRD.2019.2901594
- Law, H., and Deng, J. (2018) "CornerNet: detecting objects as paired keypoints," in Proceedings of the European Conference on Computer Vision, Munich, Germany, September 2018, 642–656. doi:10.1007/s11263-019-01204-1
- Lecun, Y., Bengio, Y., and Hinton, G. E. (2015). Deep learning. *Nature* 521, 436–444. doi:10.1038/nature14539
- Li, X., Wang, H. H., Yang, A. J., and Rong, M. Z. (2019). Partial discharge source localization in gas based on image edge detection and support vector machine. *IEEE Trans. Power Deliv.* 34, 1795–1802. doi:10.1109/TPWRD.2019.2925034
- Li, Z. X., and Zhou, F. Q. (2017). FSSD: feature fusion single shot multibox detector. <https://arxiv.org/abs/1712.00960>.
- Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., et al. (2016) "SSD: single shot multibox detector," in Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, October 2016, 21–37. doi:10.1007/978-3-319-46448-0\_2
- Liu, Y. P., Ji, X. X., Pei, S. T., Ma, Z., Zhang, G. H., Lin, Y., et al. (2020). Research on automatic location and recognition of insula-tors in substation based on yolov3. *High. Volt.* 5, 62–68. doi:10.1049/hve.2019.0091
- Lu, S. Y., Zhang, Y., and Su, J. J. (2017). Mobile robot for power substation inspection: a survey. *IEEE/CAA J. Automatica Sinica* 4, 830–847. doi:10.1109/JAS.2017.7510364
- Ma, J. C., Zheng, H. B., Sun, Y. H., Zhang, Z., Wang, X. H., Ding, G. J., et al. (2021). Temperature compensation method for infrared detection of live equipment under the interferences of wind speed and ambient temperature. *IEEE Trans. Instrum. Meas.* 70, 1–9. doi:10.1109/TIM.2021.3052003
- Mei, H. W., Jiang, H. Y., Yin, F. H., Li, L. X., and Wang, L. M. (2020). Detection of small defects in composite insulators using terahertz technique and deconvolution method. *IEEE Trans. Instrum. Meas.* 69, 8146–8155. doi:10.1109/TIM.2020.2987407
- Menendez, O., Cheein, F. A. A., Perez, M., and Kouro, S. (2017). Robotics in power systems: enabling a more reliable and safe grid. *IEEE Ind. Electron. Mag.* 11, 22–34. doi:10.1109/MIE.2017.2686458
- Miao, X. R., Liu, X. Y., Chen, J., Zhuang, S. B., Fan, J. W., and Jiang, H. (2019). Insulator detection in aerial images for transmission line inspection using single shot multibox detector. *IEEE Access* 7, 9945–9956. doi:10.1109/ACCESS.2019.2891123
- Osornio-Rios, R. A., Antonino-Daviu, J. A., and Romero-Troncoso, R. D. J. (2019). Recent industrial applications of infrared thermography: a review. *IEEE Trans. Ind. Inf.* 15, 615–625. doi:10.1109/TII.2018.2884738
- Peng, X. S., Yang, F., Wang, G. J., Wu, Y. J., Li, L., Li, Z. H., et al. (2019). A convolutional neural network-based deep learning methodology for recognition of partial discharge patterns from high-voltage cables. *IEEE Trans. Power Deliv.* 34, 1460–1469. doi:10.1109/TPWRD.2019.2906086
- Reddy, M. J. B., and Mohanta, D. K. (2013). Condition monitoring of 11 kV distribution system insulators incorporating complex imagery using combined DOST-SVM approach. *IEEE Trans. Dielectr. Electr. Insul.* 20, 664–674. doi:10.1109/TDEI.2013.6508770
- Redmon, J., Divvala, S. K., Girshick, R., and Farhadi, A. (2016) "You only look once: unified, real-time object detection," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, June 2016, 779–788. doi:10.1109/CVPR.2016.91
- Ren, S. Q., He, K. M., Girshick, R., and Sun, J. (2017). Faster r-cnn: towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* 39, 1137–1149. doi:10.1109/TPAMI.2016.2577031
- Simonyan, K., and Zisserman, A. (2014a). Very deep convolutional networks for large-scale image recognition. <https://arxiv.org/abs/1409.1556>.
- Simonyan, K., and Zisserman, A. (2014b). Very deep convolutional networks for large-scale image recognition. <https://arxiv.org/abs/1409.1556>.
- Song, Y. F., Wang, H. G., and Zhang, J. W. (2018). A vision-based broken strand detection method for a power-line maintenance robot. *IEEE Trans. Power Deliv.* 33, 2154–2161. doi:10.1109/TPWRD.2018.2328572
- Szegedy, C., Ioffe, S., Vanhoucke, V., and Alemi, A. (2017) "Inception-v4, inception-resnet and the impact of residual connections on learning," in Proceedings of the AAAI conference on artificial intelligence, San Francisco, California, USA, February 2017. doi:10.1609/aaai.v31i1.11231
- Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., et al. (2015) "Going deeper with convolutions," in Proceedings of the IEEE conference on computer vision and pattern recognition, Boston, MA, USA, June 2015, 1–9.
- Usamentiaga, R., Fernandez, M. A., Villan, A. F., and Carus, J. L. (2018). Temperature monitoring for electrical substations using infrared thermography: architecture for industrial internet of things. *IEEE Trans. Ind. Inf.* 14, 5667–5677. doi:10.1109/TII.2018.2868452
- Wang, B., Dong, M., Ren, M., Wu, Z. Y., Guo, C. X., Zhuang, T. X., et al. (2020). Automatic fault diagnosis of infrared insulator images based on image instance segmentation and temperature analysis. *IEEE Trans. Instrum. Meas.* 69, 5345–5355. doi:10.1109/TIM.2020.2965635
- Wu, Q. C., and An, J. B. (2014). An active contour model based on texture distribution for extracting inhomogeneous insulators from aerial images. *IEEE Trans. Geosci. Remote Sens.* 52, 3613–3626. doi:10.1109/TGRS.2013.2274101
- Wu, Q. G., An, J. B., and Lin, B. (2012). A texture segmentation algorithm based on pca and global minimization active contour model for aerial insulator images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 5, 1509–1518. doi:10.1109/JSTARS.2012.2197672
- Xi, L., Yu, L., Xu, Y., Wang, S., and Chen, X. (2020). A novel multi-agent DDQN-AD method-based distributed strategy for automatic generation control of integrated energy systems. *IEEE Trans. Sustain. Energy* 11, 2417–2426. doi:10.1109/TSTE.2019.2958361
- Zhao, Z. B., Fan, X. Q., Xu, G. Z., Zhang, L., Qi, Y. C., and Zhang, K. (2017). Aggregating deep convolutional feature maps for insulator detection in infrared images. *IEEE Access* 5, 21831–21839. doi:10.1109/ACCESS.2017.2757030
- Zhao, Z. B., Xu, G. Z., and Qi, Y. C. (2016). Representation of binary feature pooling for detection of insulator strings in infrared images. *IEEE Trans. Dielectr. Electr. Insul.* 23, 2858–2866. doi:10.1109/tdei.2016.7736846
- Zheng, H. B., Sun, Y. H., Liu, X. H., Djike, C. L. T., Li, J. H., Liu, Y., et al. (2020). Infrared image detection of sub-station insulators using an improved fusion single shot multibox detector. *IEEE Trans. Power Deliv.* 36, 3351–3359. doi:10.1109/TPWRD.2020.3038880
- Zou, H., and Huang, F. Z. (2015). A novel intelligent fault diagnosis method for electrical equipment using infrared thermography. *Infrared Phys. Technol.* 73, 29–35. doi:10.1016/j.infrared.2015.08.019