

Article

Novel MOA Fault Detection Technology Based on Small Sample Infrared Image

Baoquan Wei ¹, Yong Zuo ¹, Yande Liu ^{2,*}, Wei Luo ¹, Kaiyun Wen ¹ and Fangming Deng ^{1,*}

¹ School of Electrical and Automation Engineering, East China Jiaotong University, Nanchang 330013, China; 1334@ecjtu.edu.cn (B.W.); 2020028080800007@ecjtu.edu.cn (Y.Z.); 2020028085800007@ecjtu.edu.cn (W.L.); 2018028082302006@ecjtu.edu.cn (K.W.)

² School of Mechatronics and Vehicle Engineering, East China Jiaotong University, Nanchang 330013, China

* Correspondence: jxliuyd@163.com (Y.L.); 2464@ecjtu.edu.cn (F.D.)

Abstract: This paper proposes a novel metal oxide arrester (MOA) fault detection technology based on a small sample infrared image. The research is carried out from the detection process and data enhancement. A lightweight MOA identification and location algorithm is designed at the edge, which can not only reduce the amount of data uploaded, but also reduce the search space of cloud algorithm. In order to improve the accuracy and generalization ability of the defect detection model under the condition of small samples, a multi-model fusion detection algorithm is proposed. Different features of the image are extracted by multiple convolutional neural networks, and then multiple classifiers are trained. Finally, the weighted voting strategy is used for fault diagnosis. In addition, the extended model of fault samples is constructed by transfer learning and deep convolutional generative adversarial networks (DCGAN) to solve the problem of unbalanced training data sets. The experimental results show that the proposed method can realize the accurate location of arrester under the condition of small samples, and after the data expansion, the recognition rate of arrester anomalies can be improved from 83% to 85%, showing high effectiveness and reliability.

Keywords: metal oxide arrester; deep learning; edge computing; condition monitoring



check for updates

Citation: Wei, B.; Zuo, Y.; Liu, Y.; Luo, W.; Wen, K.; Deng, F. Novel MOA Fault Detection Technology Based on Small Sample Infrared Image. *Electronics* **2021**, *10*, 1748. <https://doi.org/10.3390/electronics10151748>

Academic Editor: Davide Astolfi

Received: 13 June 2021

Accepted: 19 July 2021

Published: 21 July 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

A metal oxide arrester (MOA) is widely used as an important protection equipment for safe operation of power transmission and the transformation system. However, due to the impact of lightning overvoltage and switching overvoltage, as well as environmental temperature and humidity, the characteristics of the MOA will change [1]. Therefore, the good operation characteristics of the MOA are particularly important for power transmission and the transformation system.

There are two main methods for MOA condition monitoring. The first one is based on leakage current, which can be divided into three methods, including the full current method [2–4], third harmonic current method [5–7] and the capacitive current compensation method [8,9]. However, these methods need an aging test, which is inefficient and difficult to overcome the interference of harmonic voltage in operation voltage. The second method is based on infrared thermal imaging [10,11]. Because MOA faults will cause a local temperature rise, the grade of heating defect can be judged by comparing the temperature difference of each part of the MOA. Compared with other defect detection methods, infrared detection technology is simple, safe and more efficient. However, the complex environment and less-fault samples are the difficulties of intelligent inspection. Therefore, how to realize automatic fault identification is a hot topic for researchers.

The early research of state detection is mainly based on the traditional image recognition and machine learning methods for image data mining. References [12,13] proposed to use self-organizing mapping (SOM) to analyze the thermal characteristics of an MOA infrared image under working voltage, so as to determine the MOA state; in [14], it is

proposed to use the correlation between thermal image temperature and the third harmonic current to classify the MOA states by multi-layer perceptron (MLP). Zang Xiaochun [15] and others used a back propagation (BP) neural network to extract and classify the temperature characteristics of the MOA and realized the fault identification of the 500 kV MOA. Although all the above methods can realize the fault identification of an MOA, it cannot avoid the shortcomings of an artificial neural network, such as long training time and poor generalization ability, which has proven difficult to deal with the massive data of today's power grid.

Different from traditional shallow learning, deep learning can train deep fault features through feature transformation layer by layer, which makes classification or prediction easier, especially suitable for fault diagnosis of electrical equipment [16]. In [17], a motor-bearing fault diagnosis method based on deep learning is proposed, which uses a three-layer sparse automatic encoder to extract the features of the original data and uses the maximum average error term to minimize the error loss between the features of the training data and the test data. In [18], the improved SSD detection network was used to detect the thermal anomaly area location of infrared image of power equipment in real time, and only 71.54% accuracy was achieved in the test set. However, the existing detection model based on deep learning is a single neural network, which is often limited by the characteristics of the network itself in the face of different background applications.

With the essential improvement of the neural network algorithm, although the performance of classifier based on supervised learning has been greatly improved, it needs a large number of labeled data as the basis. However, in the actual power grid system, the number of some defect samples is scarce, In the case of small samples, there is not enough data for the deep learning algorithm to fully train, so the risk of overfitting is very easy to occur in the model training process, which makes it difficult to train a detection model with good performance [19]. Traditional methods [20,21] mainly use physical methods such as mirror image, rotation and multi-scale scaling to expand the data set. However, due to the global image transformation, it does not change the details of the original image, which may lead to overfitting of the model and can affect the final effect. Goodflow [22] proposed the generation of adversarial networks (GAN) in 2014, which opened a new era of data enhancement. Li Yaxin et al. [23] introduced reinforcement learning and a strategy gradient algorithm into the generative countermeasure network, and proposed a method of generating a transformer oil chromatogram case based on the strategy gradient and generative countermeasure network, which improved the accuracy of transformer fault classification. In order to solve the problem of lack of foreign body intrusion monitoring samples in transmission lines, reference [24] uses a conditional generation countermeasure network to expand the samples of foreign body intrusion images, which can greatly improve the accuracy of model recognition.

In this paper, an infrared thermal fault detection technology for the MOA is proposed, which is based on small sample infrared images. An edge lightweight MOA identification and location algorithm is designed to reduce the amount of data uploaded. A multi-model fusion defect detection method is proposed to improve the prediction accuracy and generalization ability under the condition of small samples.

The organization of this paper is as follows. Section 2 summarizes the defect detection system. The localization identification of the MOA is described in Section 3. In Sections 4 and 5, data expansion and defect detection are described theoretically. The experimental results and analysis are summarized in Section 6. Section 7 introduces the conclusion and future work.

2. System Overview

As shown in Figure 1, a new online monitoring scheme architecture of MOA status is proposed in this paper. The status identification task is completed through the cooperation between the edge intelligent devices (patrol Unmanned Aerial Vehicle (UAV) or patrol smart car, etc.) and the server (cloud server, etc.). The infrared thermal images of the MOA

are collected and stored by edge devices, and then processed quickly on the spot based on the needs of communication and cloud diagnosis and uploaded to the cloud server for fault identification, relying on a high-speed communication network (4G, power wireless private network, etc.). Computing is distributed in the whole system network, including edge intelligent devices and cloud servers, and data is stored in the intelligent devices at the edge of the network. Therefore, the system can meet the construction needs of a low delay, low energy consumption, high-precision power Internet of things.

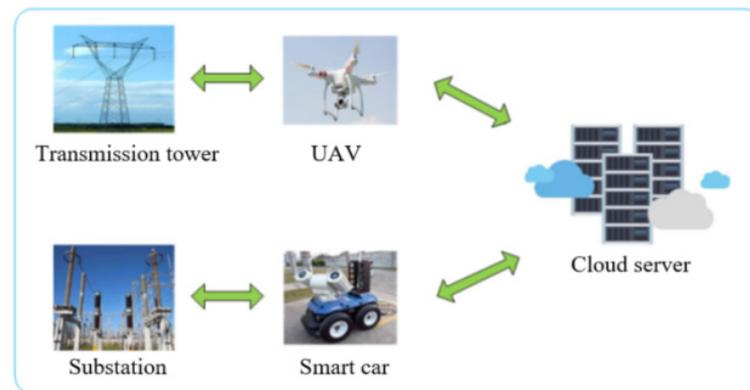


Figure 1. Online monitoring scheme of the metal oxide arrester (MOA).

2.1. MOA Identification and Localization

Aiming at the problem that traditional detection methods are difficult to overcome the complex background interference of a power grid, the key components localization is proposed to localize and extract different types of MOAs in substation or transmission line. In order to apply to edge devices, an improved SSD-MobileNet network that performs well in both speed and scale is adopted.

2.2. Infrared Thermal Fault Detection of the MOA

However, achieving full automation of MOA defect detection is still very challenging due to the visual complexity of defects and the small number of defective MOAs.

- (1) The amount of abnormal MOA data is not enough to train robust classification model.
- (2) The existing fault detection algorithms based on deep learning are usually single neural network, but they are often limited by the characteristics of the network in the face of different background applications.
- (3) The visual complexity of defects makes it difficult, if not impossible, to construct a precise model.

On the one hand, through a transfer learning-generation convolution countermeasure network, the data expansion model is constructed to solve the problem of data imbalance. On the other hand, different infrared features are extracted by multiple neural networks, and multiple classifiers are trained. Finally, the combination strategy is used to fuse the prediction results to improve the accuracy and generalization ability of the detection model.

3. MOA Identification and Localization

A single shot multibox detector (SSD) is a classic one-stage target detection model proposed by Wei Liu in 2016 [25]. As a fast recognition and positing network, SSD is widely used in target detection, and its architecture is shown in Figure 2.

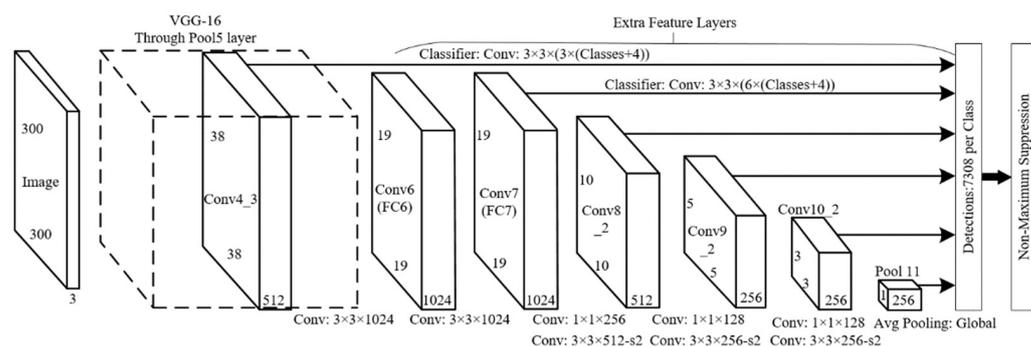


Figure 2. Single shot multibox detector (SSD) Architecture.

In order to cope with the limited computing resources at the edge, in this paper we use the lightweight MobileNet structure to replace the original VGG16 basic network, and cut down the average pooling layer and the full connection layer. MobileNet is a series of lightweight networks proposed by Google [26]. Figure 3 shows the standard convolution and MobileNet structure. MobileNet uses something similar to a deep separable convolution instead of a traditional convolution and decomposes the original standard convolution into deep convolution and point-by-point convolution. Each time, one channel of input data is convoluted, and then convolution is performed by using the convolution core with a channel number of 1×1 input data channel number, thus reducing a large amount of redundant calculation.

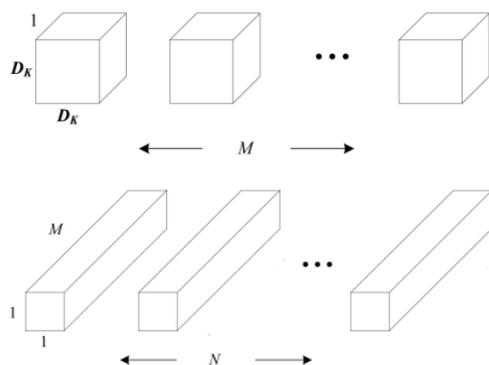


Figure 3. Mobile Architecture.

First, the image size of the MOA is converted to a fixed size of 300×300 . Then, forward propagation is used to extract features through basic network to form the feature map. Finally, the additional feature network is adopted for regression calculation and maximum suppression to generate the prediction of the target object frame and category.

4. Data Expansion

As shown in Figure 4, the proposed data expansion model based on transfer learning and a deep convolution generation adversarial network (TL-DCGAN) is proposed. Firstly, the transfer learning method is used to train a model DCGAN1, which can generate normal samples by using a large number of existing normal MOA images. Then, the weight of dcgan1 is transferred again, and the limited fault data is used to train the data expansion model DCGAN2.

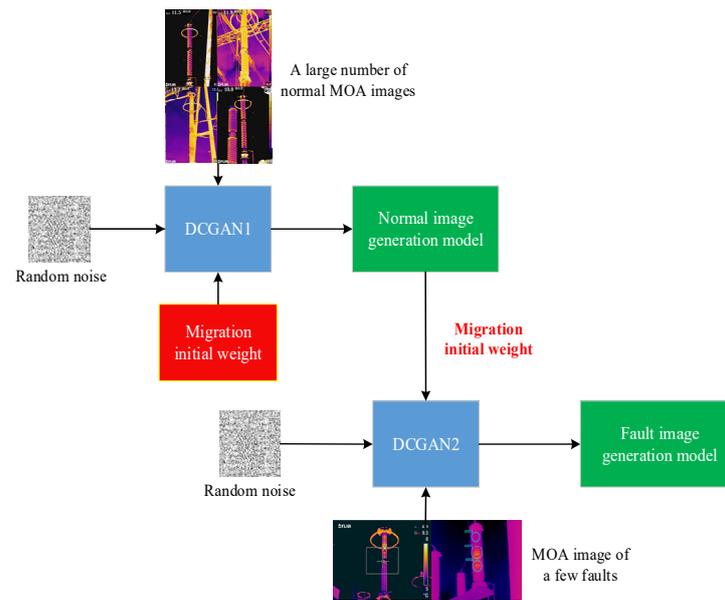


Figure 4. Framework of fault MOA infrared image expansion model.

4.1. Generation Adversarial Network

As shown in Figure 5, the GAN network structure mainly includes the generator (G) and discriminator (D).

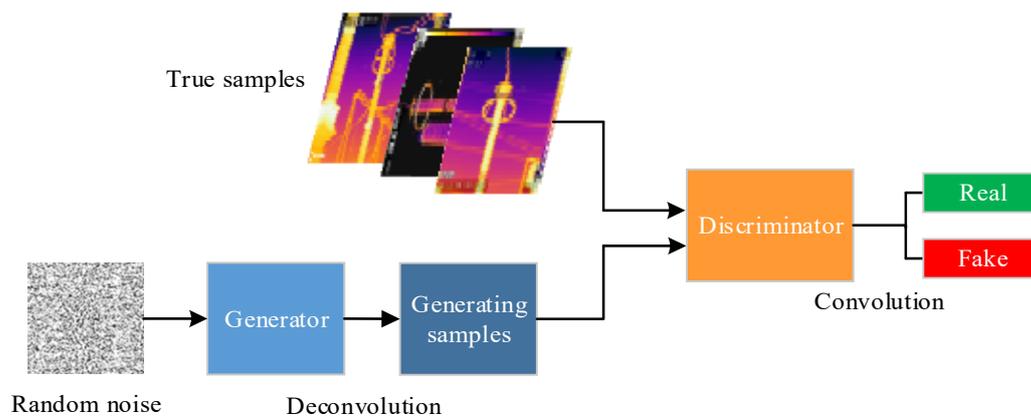


Figure 5. Generation of adversarial networks (GAN) Structure.

The objective function of GAN training is as follows:

$$\min_G \max_D V(G, D) = E_{x \sim p_{data}(x)} [\log D(x)] + E_{z \sim p_z(z)} [\log(1 - D(G(z)))] \quad (1)$$

where $p_{data}(x)$ is the probability distribution of real samples, $p_z(z)$ is the distribution of input random noise and $V(G, D)$ is the cross entropy loss.

The loss function of G is:

$$L_G = E_{x \sim p_g} [\log(1 - p(s | x_{fake}))] \quad (2)$$

The loss function of D is:

$$L_D = -E_{x \sim p_{data}} [\log p(s | x_{real})] - E_{x \sim p_g} [\log(1 - p(s | x_{fake}))] \quad (3)$$

where s is the real sample and x_{fake} is the false sample. The optimization goal of the GAN model is to make the samples generated by G make D unable to distinguish true from false. Therefore, in training, for G , we hope that the larger the $p(s|x_{fake})$ is, the better; that is, $\max_D V(D, G)$ mentioned above. For the D , when the sample comes from the training set x_{real} , the larger the $p(s|x_{real})$ is, the better; when the sample comes from the G , the larger the $p(s|x_{fake})$ is, the better; that is, the $\min_G V(D, G)$ mentioned above.

4.2. Data Expansion Model Based on TL-DCGAN

Compared with GAN, a variant of GAN (deep convolutional GAN (DCGAN)) was proposed in 2016 [27] which uses the mature convolutional neural networks (CNN) instead of MLP and removes the pooling layer, making the overall network model differentiable.

4.2.1. Improved Generator Structure

In order to improve the resolution of the MOA, a cumulus layer is added on the basis of DCGAN. The generator structure of the fault image expansion model of the MOA is shown in Figure 6. In addition, this is conducted in order to make the generated data distribution more close to the real data distribution, prevent the gradient disappearance and improve the network stability.

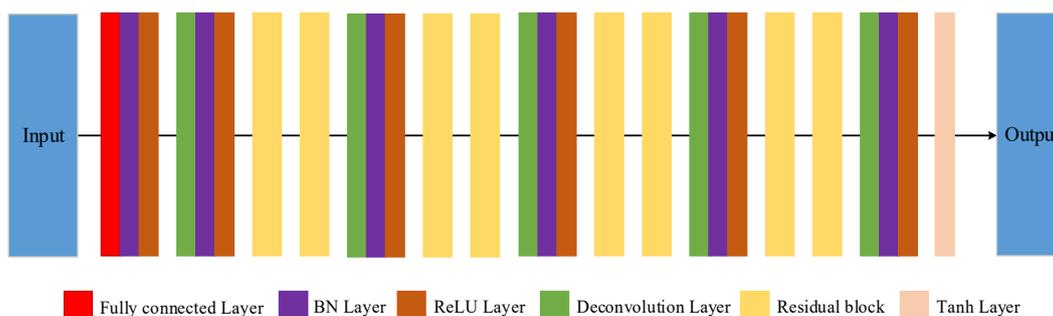


Figure 6. Generator Structure.

The generator is mainly composed of the input layer, full connection layer, convolution layer and residual block, in which the convolution layer is used as fractional step convolution, and the activation function is the ReLU function. Firstly, a set of random noise is inputted, which is uniformly distributed, and is extended to a feature matrix of size $4 \times 4 \times 1024$ through the whole connection layer. Then, through the first convolution layer, deconvolution, batch normalization and activation function operation are performed, and the output characteristic matrix size is $8 \times 8 \times 512$. Then, the output characteristic matrix size is $16 \times 16 \times 256$ through two residual blocks, increasing the network depth and improving the network representation ability. Finally, after much processing of the volume layer and residual block, the pixel is 128×128 MOA image. See Table 1 for the parameters of convolution layer of the producer, where the convolution core size is 3×3 and the step size is 2.

Table 1. Generator convolution layer parameters.

Layer	Kernel Size	Stride	Output
Deconv1	$3 \times 3/512$	2	$8 \times 8/512$
Deconv2	$3 \times 3/256$	2	$16 \times 16/256$
Deconv3	$3 \times 3/128$	2	$32 \times 32/128$
Deconv4	$3 \times 3/64$	2	$64 \times 64/64$
Deconv5	$3 \times 3/3$	2	$128 \times 128/3$

4.2.2. Improved Discriminator Structure

As shown in Figure 7, compared with the original DCGAN network, the discriminator structure designed in this paper adds a layer of convolution network. In addition, in order to improve the network performance, the residual module is constructed similar to the above generator. Among them, the leaky ReLU activation function is used, and the convolution kernel size is 3×3 . After convolution, batch normalization and function activation are performed. The input of the discriminator is the real or generated MOA image with the size of 128×128 . The image size is reduced by sampling under the convolution layer, and then the network is deepened by two residual blocks, and the extracted feature information is transmitted to the deep layer of the network. Finally, through several convolution layers and residual blocks, the image size becomes $4 \times 4 \times 1024$ and is input to the full connection layer to get the result of image discrimination. The parameters of the convolution layer in the discriminator are shown in Table 2, in which the convolution kernel size is 3×3 and the step size is 2.

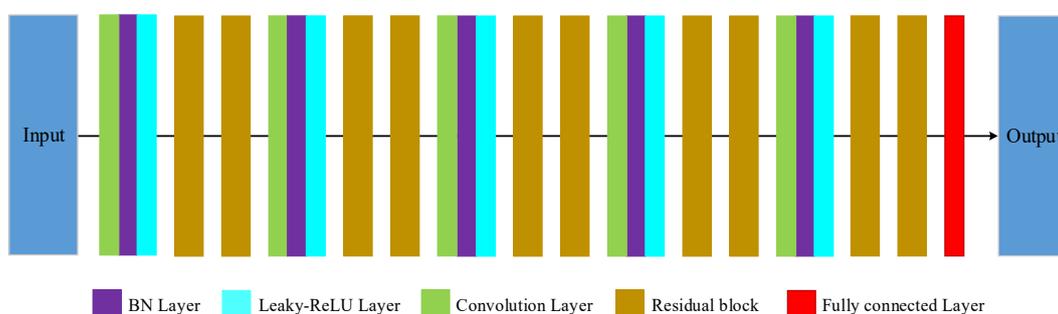


Figure 7. Discrimination Structure.

Table 2. Discriminator convolution layer parameters.

Layer	Kernel Size	Stride	Output
Conv1	$3 \times 3/64$	2	$64 \times 64/64$
Conv2	$3 \times 3/128$	2	$32 \times 32/128$
Conv3	$3 \times 3/256$	2	$16 \times 16/256$
Conv4	$3 \times 3/512$	2	$8 \times 8/512$
Conv5	$3 \times 3/1024$	2	$4 \times 4/1024$

5. Defect Detection

As shown in Figure 8, the proposed MOA infrared state detection framework based on multi-model fusion is proposed. Due to the small number of MOA infrared fault samples, if the traditional single neural network is used to extract the feature vector, it is easy to lead to overfitting of the model in the training process of the classifier. Therefore, this paper uses a variety of convolutional neural networks to extract a variety of MOA fault features, and then selects the relevance vector machine (RVM) as the feature vector classifier to train and generate multiple weak learning machines, and finally uses the combination strategy to fuse them together for defect detection.

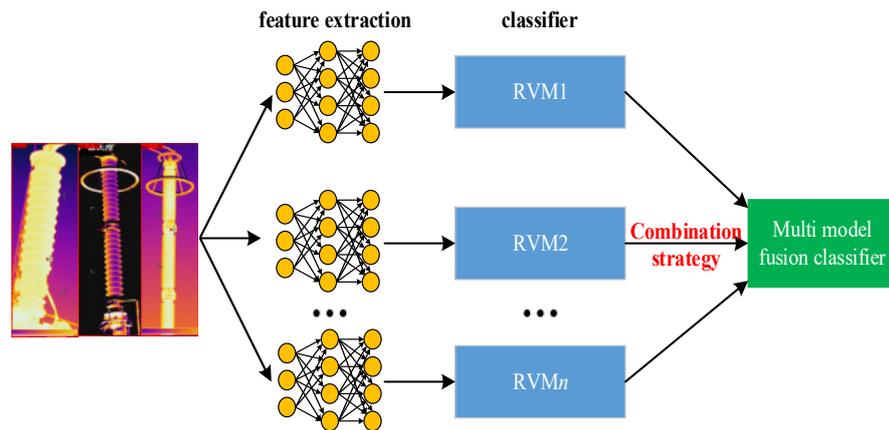


Figure 8. MOA infrared state detection framework based on multi-model fusion.

5.1. Depth Feature Extraction

In recent years, deep learning has developed rapidly; the deep convolution neural network has especially achieved good results in image classification and target recognition, and greatly improved the efficiency. Therefore, this method is used as a feature extractor to identify the infrared thermal fault of the MOA.

In the deep convolution neural network, most of the neurons only connect with the nearby neurons and share the weights, which greatly reduces the network parameters and improves the training speed. As shown in Figure 9, there are three main structures in the deep convolution network: the convolution layer, pooling layer and full connection layer.

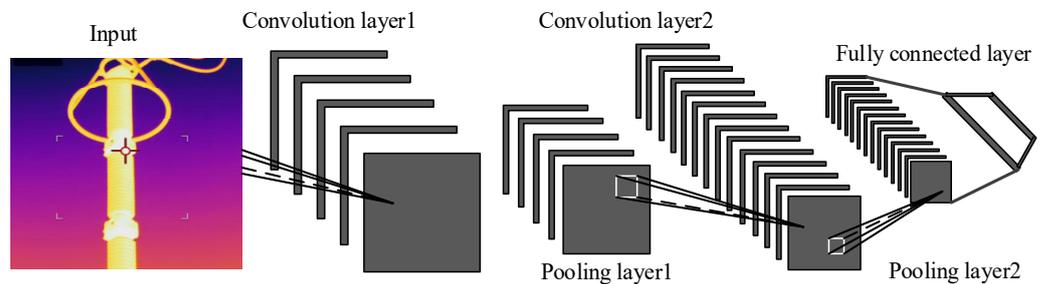


Figure 9. Convolutional neural network structure.

In the convolution layer, after the input data is convoluted with the linear filter, the feature map is obtained through the nonlinear activation function. Each feature map contains one feature and shares the same parameters. Different feature maps use different parameters to extract different features. The convolution formula is:

$$x_{ij}^k = f[(w_{ij}^k \times x_{ij}^{k-1}) + b_j^k] \tag{4}$$

where x_{ij}^k is the k -th layer characteristic graph, i and j are input dimensions and x_{ij}^{k-1} is the input data of the upper layer. The convolution filter of layer k is determined by the weight w_{ij}^k and the bias term b_j^k and f is the nonlinear activation function.

The pooling layer downsamples the feature graph, reduces the dimension of the feature graph and network parameters, makes the feature easier to follow-up processing and reduces the overfitting phenomenon to a certain extent. The pooling formula is:

$$x_{ij}^k = f[\beta_{ij}^k \text{down}(x_{ij}^{k-1}) + b_j^k] \tag{5}$$

where down is the downsampling function, if the downsampling window size is $n \times n$. The output feature map is reduced by N times. β_{ij}^k and b_j^k are multiplicative bias and additive bias parameters, respectively.

The full connection layer is similar to the traditional neural network, in which each neuron is connected to all inputs.

Although the classifier based on supervised learning is very mature, it needs a large number of labeled data to train a classification model with high accuracy and strong generalization. However, in the actual power grid system, the samples of fault MOA infrared data are usually less, and the image background environment is more complex. Therefore, in this paper, different convolution neural networks (AlexNet, GoogLeNet, ResNet, RetinaNet) are used to extract different features of the MOA image, and the image can be comprehensively analyzed from different aspects, so as to obtain more reliable detection results.

5.2. MOA Fault Detection Based on Integrated Learning

In order to get more accurate judgment accuracy and improve the generalization ability of the defect recognition model, this paper proposes a multi-model combination strategy based on weighted voting rule and F_1 score. F_1 score, also known as balanced f score, is a harmonic average of model accuracy (P) and recall (R). Its maximum is 1 and minimum is 0. It is often used to measure the accuracy of the binary classification model.

The F_1 score of M_i of each weak learning machine is calculated, and the formula is as follows:

$$F_1 = \frac{2TP}{2TP + FN + FP} \quad (6)$$

where P is the accuracy rate and R is the recall rate.

Then, according to the performance of RVM classifier in the verification set, the voting weight is calculated by using the following formula to give higher weight to the classifier with high reliability, so as to improve the reliability of the ensemble classifier:

$$w_j = \frac{F_1^j}{\sum_{j=1}^k F_1^j} \quad (7)$$

Finally, according to the prediction result $h(x_i)$ of each weak learning machine and its voting weight w_i , the final model prediction result $H(x)$ is obtained by using weighted voting rule:

$$H(x) = \begin{cases} 1, & \sum_1^n h_i \times w_i \geq 0.5 \\ 0, & \text{otherwise} \end{cases} \quad (8)$$

where n is the number of weak learning machines, that is, the number of depth feature types. According to the prediction result $H(x)$ of the integrated classifier, whether the MOA is abnormal or not can be determined.

6. Experimental Results and Analysis

To evaluate the performance of the proposed MOA defect detection system, we tested it on a MOA image data set of a substation in Jiangxi Province, China. The data acquisition equipment is an advanced pistol thermal imager with 640×480 infrared resolution, and its model is FLIR E98 (Shenzhen Keruijie Technology Co., Ltd., Shenzhen, China). The experiment environment is as follows: Win10, Tensorflow1.3, Anaconda (python3.6), Keras2.1.5, Core i9-9900k and GTX 2080 GPU with 8-GB memory.

6.1. MOA Positioning Experiment

In the experiment, the parameters of the model are initialized by using the weight of the classical network, and then the infrared data of the MOA are divided into training set, verification set and test set according to the ratio of 6:1:3. In this paper, the empirical value is selected as the initial value of the super parameter. Among them, the learning rate is set to 0.0015, batch_size is set to 16 and the epoch is set to 500. Results as shown in Figure 10, the proposed MOA identification and location algorithm can effectively identify and locate different types of MOAs (the rated voltages are 110 kV, 220 kV and 500 kV, respectively).

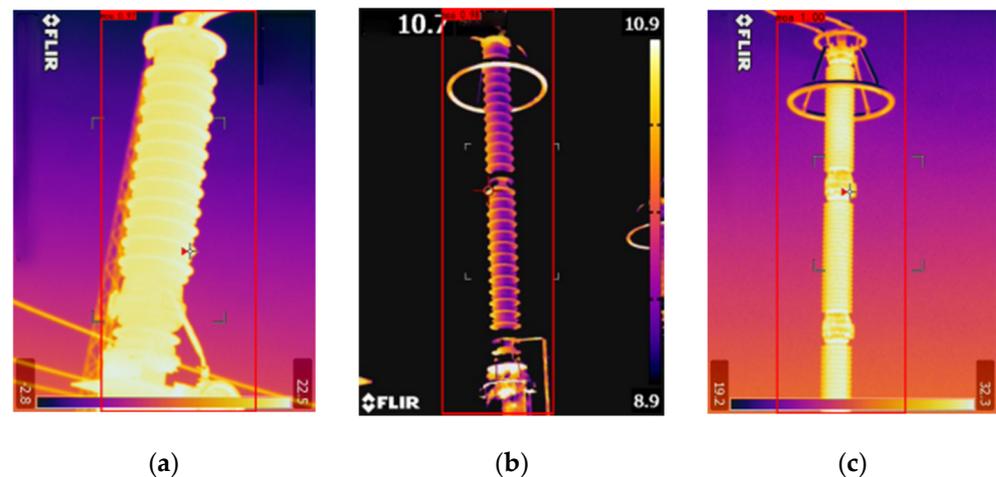


Figure 10. Identification of (a) 110 kV-MOA, (b) 220 kV-MOA, and (c) 500 kV-MOA.

In order to further verify the advantages of the proposed method in MOA identification and location, the proposed algorithm and the commonly used deep learning algorithm are tested and compared on the same data set. The results are shown in Table 3. Different algorithms are compared from map, recognition speed and model training time. It can be seen from the table that a one-stage algorithm is superior to a two-stage algorithm in recognition speed, model size and training time. The recognition accuracy of a two-stage algorithm is significantly higher than that of a one-stage algorithm. Among them, although the proposed algorithm is slightly inferior to the two-stage algorithm in accuracy and slightly slower than the You Only Look Once (YOLO) in speed, it is most suitable to be deployed in the edge end of the embedded device in comprehensive ability, and can realize the MOA fast and with high-precision identification and positioning.

Table 3. Comparison of recognition performance of different algorithms.

Algorithm	MAP	Speed (ms)	Time (h)
YOLO	80.5%	22	12
SSD	93.3%	45	14
Mask-RCNN	99.6%	67	36
Faster-RCNN	98.6%	58	28
SSD-MobileNet	96.5%	30	8

6.2. Data Expansion Experiment

In order to keep the diversity of the samples and enhance the generalization ability of the training model, a total of 2435 infrared images of the MOA in different natural conditions in several areas were obtained from a power grid company, including 1981 normal samples and 454 fault samples. Firstly, the fault samples are expanded to 1696 by the traditional method, and then the original DCGAN model is trained. In the experiment, the parameter optimizer is the Adam optimizer, the learning rate is set to 0.0002, the momentum value is set to 0.5 and the batch_size is set to 64.

It can be seen from the Figure 11a that in the original DCGAN model without transfer learning, some MOA contour information appears only at 100 epoch, and the training is relatively slow, and the complete and usable normal MOA image cannot be generated after 500 epoch training. As shown in Figure 11b, the DCGAN1 model which migrates the weight of classical algorithm can generate the basic features of MOA image at 100 epoch, such as orientation features, target contour, etc., and can generate a more complete image at 500 epoch. On the basis of training the model DCGAN1, we continue to use the idea of transfer learning to train the model. Due to the use of the weight of the model DCGAN1, in Figure 11c, it can be clearly seen that the DCGAN2 has been able to get the basic

characteristics of the MOA at 100 epoch, can learn the characteristics of the fault MOA at 300 epoch and can generate better fault MOA images at 500 epoch.

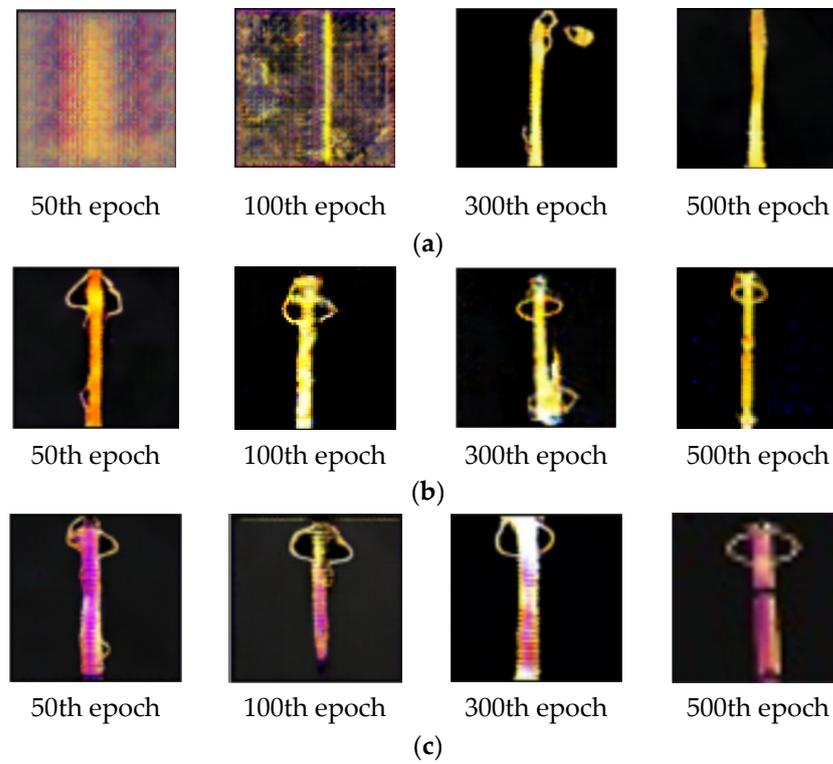


Figure 11. Comparison of images generated by different deep convolutional generative adversarial network (DCGAN) models under different epochs. (a) Images generated by the original DCGAN model, (b) normal images generated by the improved DCGAN1 model and (c) fault images generated by the improved DCGAN2 model.

In order to judge the performance of the improved model more accurately, the discriminator and generator loss rate curves of the original DCGAN model and the improved DCGAN2 model are drawn; respectively, as shown in Figure 12, the x-axis is the different training moments of the model, and the y-axis is the loss function value of the discriminator or generator. In the initial stage of training, the generator training times are less, the extracted MOA features are not comprehensive, the generated image is quite different from the real MOA and the discriminator can easily identify the image “true and false”. Therefore, the generator loss is much larger than the discriminator loss. With the increase in training times, the MOA features obtained by the generator are more and more sufficient, and the generated images are more and more close to the real sample data. Comparing the change of the loss function of the two models, it can be seen that the loss rate of the improved DCGAN2 model can finally converge, which proves that the generated image effect is better.

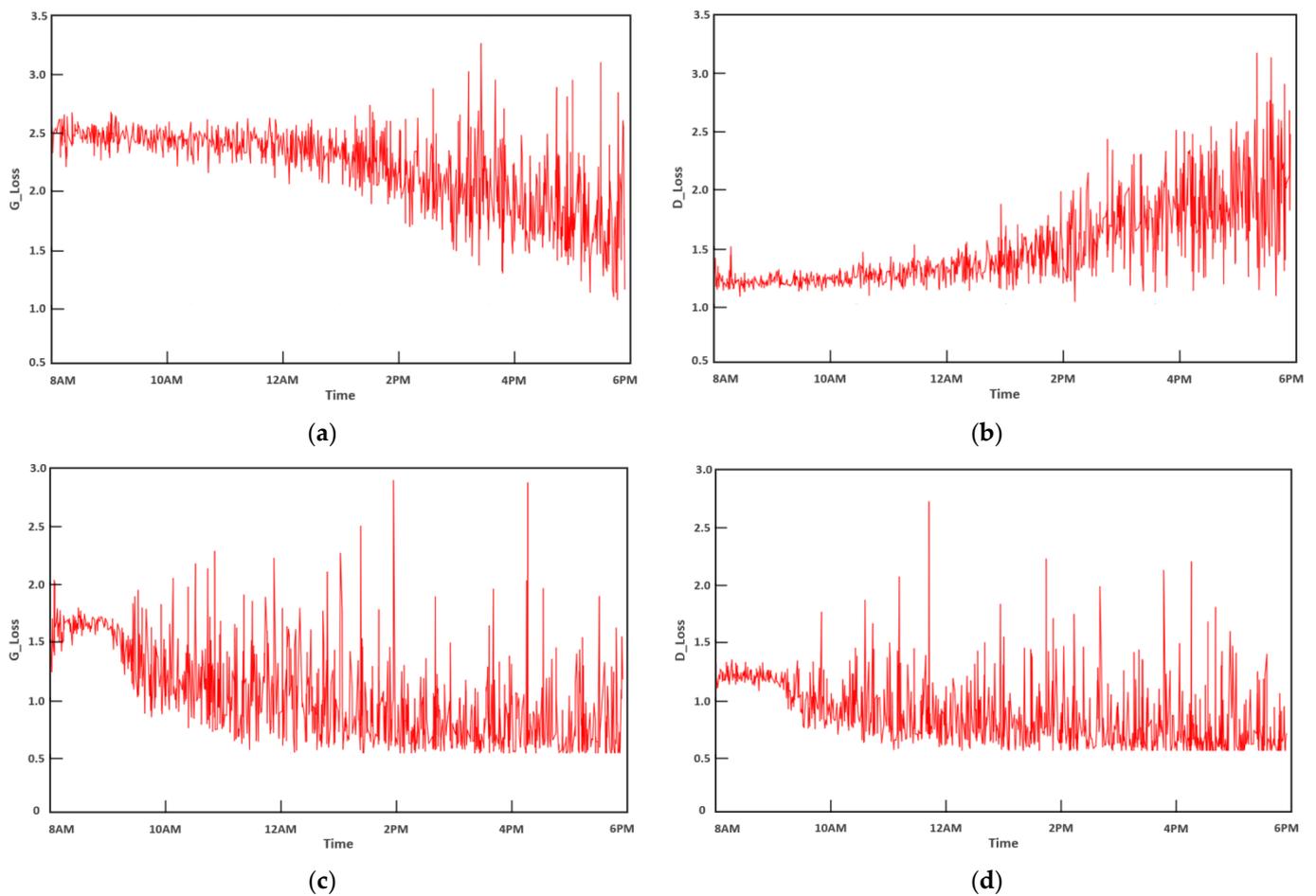


Figure 12. Loss rate of (a) original generator, (b) original discriminator, (c) improved DCGAN2 generator and (d) improved DCGAN2 discriminator.

6.3. MOA Thermal Fault Detection Experiment

In addition to the average accuracy mentioned in the previous chapter, an F_1 score is added to evaluate the performance of the classifier. The fault identification effect diagram of different types of MOA is shown in Figure 13. It can be seen from Table 4 that the highest recognition rate of the model trained by a single neural network is only 76%, while the recognition accuracy of the multi-model fusion classifier proposed in this paper can be improved from 5% to 81%, which can effectively identify the infrared thermal fault of the MOA.

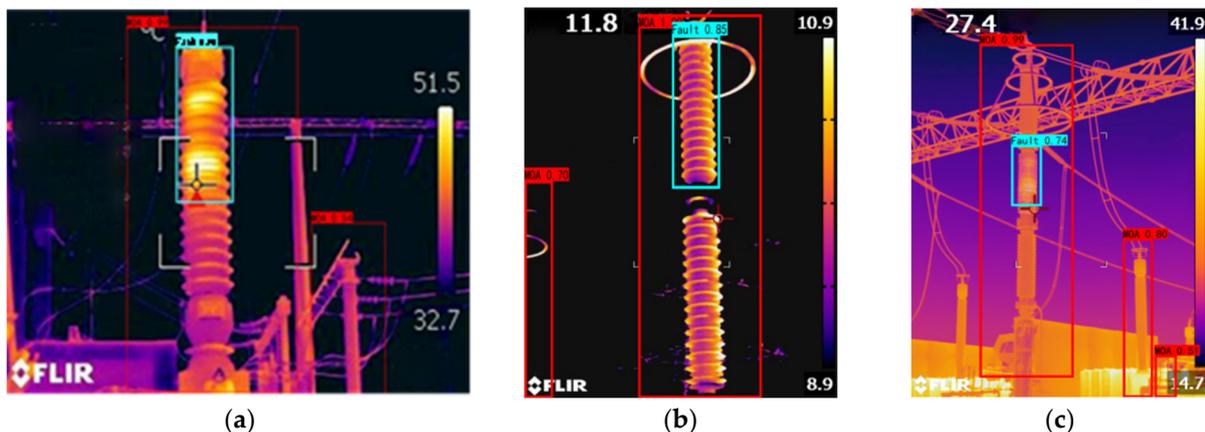
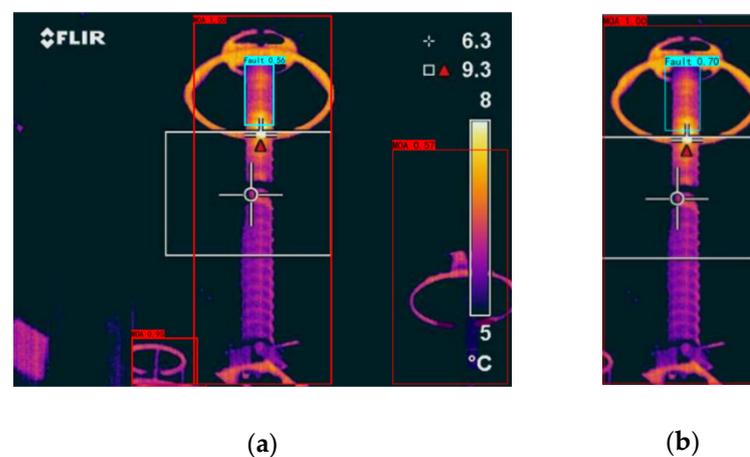


Figure 13. Fault identification of (a) 110 kV-MOA, (b) 220 kV-MOA and (c) 500 kV-MOA.

Table 4. Comparison of fault identification results of different methods.

Location Algorithm	MAP	Recall	F_1
AlexNet + RVM	68%	67%	0.68
ResNet + RVM	67%	66%	0.67
GoogleNet + RVM	76%	75%	0.75
RetinaNet + RVM	73%	73%	0.73
The proposed algorithm	81%	80%	0.81

In addition, the proposed method can identify and locate the MOA before the condition detection, which reduces the search space of the fault detection model. As shown in Figure 14, the fault detection accuracy of the MOA after positioning is significantly higher than the global detection. Therefore, the average accuracy of the final detection increased from 81% to 83%.

**Figure 14.** Fault detection of (a) Global and (b) after positioning.

7. Conclusions

In this paper, an infrared thermal fault detection method for small samples is proposed:

- (1) In order to solve the problem of sample imbalance, transfer learning and deep convolution, generation countermeasure networks are used to expand the data of fault MOAs. Experiments show that the expanded training set can improve the accuracy of a fault detection model by 2%.
- (2) In order to minimize the interference of the background to defect detection, defect detection is divided into two steps: target recognition and state detection. Firstly, the improved SSD algorithm is used to identify and locate the MOA. The experimental results show that the proposed algorithm can accurately locate different types of MOA in different scenarios.
- (3) Through a variety of convolution neural networks to extract a variety of MOA features, then train multiple weak classifiers, and then use the combination strategy to integrate the prediction results, further improving the prediction accuracy and generalization ability of the model.
- (4) The proposed method is based on simulation data and real cases, and many problems need to be further studied: how to combine the fault characteristics of equipment to make the model interpretable and improve the identification accuracy; through the cooperation of edge computing and cloud computing, improving the real-time performance of the detection system to meet the engineering application is the next research direction.

Author Contributions: Conceptualization, Y.L. and F.D.; data curation, W.L. and K.W.; formal analysis, B.W., Y.Z. and K.W.; investigation, B.W. and Y.Z.; methodology, Y.L.; project administration, F.D.; software, Y.Z.; supervision, F.D.; writing—original draft, B.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Natural Science Foundation of China (51767006), Natural Science Foundation of Jiangxi Province (20202ACBL214021, 20202BAB202005), Key Research and Development Plan of Jiangxi Province (20202BBGL73098), Science and Technology Project of Education Department of Jiangxi Province (GJJ190311, GJJ180308).

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to privacy issues.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Tuczek, M.N.; Bröker, M.; Hinrichsen, V.; Göhler, R. Effects of Continuous Operating Voltage Stress and AC Energy Injection on Current Sharing Among Parallel-Connected Metal–Oxide Resistor Columns in Arrester Banks. *IEEE Trans. Power Deliv.* **2015**, *30*, 1331–1337. [[CrossRef](#)]
2. Han, Y.; Li, Z.; Zheng, H.; Guo, W. A Decomposition Method for the Total Leakage Current of MOA Based on Multiple Linear Regression. *IEEE Trans. Power Deliv.* **2016**, *31*, 1422–1428. [[CrossRef](#)]
3. Khodsuz, M.; Mirzaie, M.; Seyyedbarzegar, S. Metal oxide surge arrester condition monitoring based on analysis of leakage current components. *Int. J. Electr. Power Energy Syst.* **2015**, *66*, 188–193. [[CrossRef](#)]
4. Harid, N.; Bogias, A.C.; Griffiths, H.; Robson, S.; Haddad, A. A Wireless System for Monitoring Leakage Current in Electrical Substation Equipment. *IEEE Access* **2016**, *4*, 2965–2975. [[CrossRef](#)]
5. Bokoro, P.; Jandrell, I. Failure Analysis of Metal Oxide Arresters under Harmonic Distortion. *Saiee Afr. Res. J.* **2016**, *107*, 167–176. [[CrossRef](#)]
6. Metwally, I.A.; Eladawy, M.; Feilat, E.A. Online condition monitoring of surge arresters based on third-harmonic analysis of leakage current. *IEEE Trans. Dielectr. Electr. Insul.* **2017**, *24*, 2274–2281. [[CrossRef](#)]
7. Chen, J.; Xu, L.; Wan, Z.; Hu, X.; Zhang, L. Research on monitor technology of MOA state based on third harmonic principle. In Proceedings of the IEEE International Conference on High Voltage Engineering and Application, Chengdu, China, 19–22 September 2016; pp. 1–4.
8. Fu, Z.; Wang, J.; Bretas, A.; Ou, Y.; Zhou, G. Measurement Method for Resistive Current Components of Metal Oxide Surge Arrester in Service. *IEEE Trans. Power Deliv.* **2018**, *33*, 2246–2253. [[CrossRef](#)]
9. Khodsuz, M.; Mirzaie, M. A Modified Time-Delay Addition Method to Extract Resistive Leakage Current of MOSA. *J. Inst. Eng. Ser. B* **2016**, *97*, 445–451. [[CrossRef](#)]
10. Neto, E.W.; Da Costa, E.G.; Maia, M.J.A. Artificial Neural Networks Used for ZnO Arresters Diagnosis. *IEEE Trans. Power Deliv.* **2009**, *24*, 1390–1395. [[CrossRef](#)]
11. Li, S.T.; He, J.Q.; Lin, J.J.; Wang, H.; Liu, W.F.; Liao, Y.L. Electrical-Thermal Failure of Metal–Oxide Arrester by Successive Impulses. *IEEE Trans. Power Deliv.* **2016**, *31*, 2538–2545. [[CrossRef](#)]
12. Lira, G.R.S.; Costa, E.G.; Almeida, C.W.D. Self-organizing maps applied to monitoring and diagnosis of ZnO surge arresters. In Proceedings of the 2010 IEEE/PES Transmission and Distribution Conference and Exposition: Latin America (T&D-LA), Sao Paulo, Brazil, 8–10 November 2010; IEEE: Piscataway, NJ, USA, 2010; pp. 659–664.
13. Jialin, W.; Haoyang, C.; Yong, X. Infrared Image Diagnosis Method of Transformer Substation Equipment Based on SOM Neural Network. *J. Shanghai Univ. Electr. Power* **2016**, *32*, 78–82.
14. Abdul-Malek, Z. Electrical and temperature correlation to monitor fault condition of ZnO surge arrester. In Proceedings of the 2016 3rd International Conference on Information Technology, Computer and Electrical Engineering (ICITACEE), Semarang, Indonesia, 18–20 October 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 182–186.
15. Liao, Z.; Zang, X.; Zhou, K.; Xiao, L.; Mao, Q.; Lan, P. A Detection Method of Low/Zero Fault and Contamination Fault of Insulator String Based on Infrared Image Feature and BP Neural Network. *Insul. Surge Arresters* **2019**, *03*, 204–211.
16. Zhou, N.; Liao, J.; Wang, Q.; Li, C.; Li, J. Analysis and prospect of Deep Learning Application in Smart Grid. *Autom. Electr. Power Syst.* **2019**, *43*, 267–284.
17. Wen, L.; Gao, L.; Li, X. A New Deep Transfer Learning Based on Sparse Auto-Encoder for Fault Diagnosis. *Syst. Man Cybern.* **2019**, *49*, 136–144. [[CrossRef](#)]
18. Xuhong, W.; Hao, L.; Shaosheng, F.; Zhipeng, J. Infrared Image Anomaly Automatic Detection Method for Power Equipment Based on Improved Single Shot Multi Box Detection. *Trans. China Electrotech. Soc.* **2020**, *35*, 302–310.
19. Ren, J.; Huang, X. Defect Detection Using Combined Deep Autoencoder and Classifier for Small Sample Size. In Proceedings of the International Conference on Control Science and Systems Engineering (ICCSSE), Beijing, China, 17–19 July 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 32–35.

20. Jie, J.I.A.N.G.; Changzhen, X.I.O.N.G. Data Augmentation with Multi-Model Ensemble for Fine-Grained Category Classification. *J. Graph.* **2018**, *39*, 244–250.
21. Guo, M.; Xu, H. Hot Spot Defect Detection Based on Infrared Thermal Image and Faster RCNN. *Comput. Syst. Appl.* **2019**, *28*, 265–270.
22. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative Adversarial Networks. *Adv. Neural Inf. Process. Syst.* **2014**, *3*, 2672–2680. [[CrossRef](#)]
23. Li, Y.; Hou, H.; Xu, M.; Li, S.; Sheng, G.; Jiang, X. Oil chromatogram case generation method of transformer based on policy gradient and generative adversarial networks. *Electr. Power Autom. Equip.* **2020**, *40*, 211–218.
24. Yang, J.; Qin, Z.; Pang, X.; He, Z.; Cui, C. Foreign body intrusion monitoring and recognition method based on Dense-YOLOv3 deep learning network. *Power Syst. Prot. Control.* **2021**, *49*, 37–44.
25. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. Ssd: Single shot multibox detector. In Proceedings of the European Conference on Computer Vision (ECCV), Amsterdam, The Netherlands, 8–16 October 2016; pp. 21–37.
26. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv* **2017**, arXiv:1704.04861.
27. Radford, A.; Metz, L.; Chintala, S. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. In Proceedings of the 4th International Conference on Learning Representations, ICLR, San Juan, Puerto Rico, 2–4 May 2016.