

Received April 24, 2020, accepted May 4, 2020, date of publication May 20, 2020, date of current version June 5, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2995910

An Improved AlexNet for Power Edge Transmission Line Anomaly Detection

YANPENG GUO¹, ZHENJIANG PANG, (Member, IEEE),
JUN DU, FAN JIANG, AND QILONG HU

Beijing Smartchip Microelectronics Technology Company, Ltd., Beijing 100192, China

Corresponding author: Yanpeng Guo (ypguo1314@163.com)

ABSTRACT Since most outdoor transmission line equipment suffers from harsh disasters, they are prone to wire breakage, tower collapse and insulator flashover. When anomaly occurs, too much time is required for the State Grid Corporation to fix it manually. To reduce the inspection burden, many methods have been presented in the past to diagnose and locate anomaly. In this paper, we propose an improved AlexNet model for anomaly detection. In the aspect of feature extraction, the proposed model extracts the characteristics of transmission line equipment through a deep convolutional neural network (DCNN). In the aspect of the recognition algorithm, by referring to the advantages of the traditional machine learning method and incorporating the advantages of the support vector machine (SVM), an SVM classification method incorporating deep learning is proposed. Finally, the improved AlexNet model and SVM classification method are used to classify images of various types of power equipment. The results show that the proposed methods can be effectively applied to the image recognition of various types of power equipment, and they greatly improve the recognition rate of power equipment images, which has great potential for future real-time transmission line monitoring platform design.

INDEX TERMS Transmission line, edge computing, AlexNet, SVM, anomaly detection.

I. INTRODUCTION

With the rapid development of power electronics systems, anomaly detection and localization are the main concern of research efforts in the ubiquitous power internet of things (UPIoT) [1]–[6]. Because anomaly in electrical power systems cannot be avoided, enough information provided from the anomaly analysis is needed to recognize the cause and interpret the broken-down system. It is also necessary to restore the transfer of power as soon as possible, in addition to learning more about the system to reduce the occurrence of future anomaly if possible. Circuit breakers and other control components are designed to help protect the relay and to take appropriate action and thus minimize power loss and length of power disruption.

While the power load is increasing, which has promoted the scale of power systems to become increasingly large, the number of power transmission lines has also increased. These transmission lines are responsible for power transmission tasks between power plants and load centers and power

exchange between different locations. Therefore, power transmission lines are important for the stable operation of the power system. Since transmission lines are set up over long distances and exposed outdoors long-term, they inevitably suffer from harsh environments such as high temperatures, rain, snow, lightning, and hail. Therefore, the occurrence of failures is inevitable. With the rapid evolution of artificial intelligence and edge computing, it provides cloud computing and cache capabilities to reduce the maintenance cost of power equipment. Therefore, it is a novel architecture which integrates the elements of humans, smart devices, deep learning, and combines the power of mobile edge computing architecture and the large-scale sensing ability.

Transmission line component inspection plays a significant role in industrial processes, especially power system monitoring and protection. In traditional power systems, transmission line component inspection usually depends on human resources. Although these approaches have high inspection accuracy, they result in high costs and even unexpected safety issues. Smart devices, such as sensors, robotics, and unmanned aerial vehicles (UAVs), have made great contributions to enabling the development of an efficient unmanned

The associate editor coordinating the review of this manuscript and approving it for publication was Lu Liu¹.

inspection system for transmission line and tower monitoring applications [7]–[12]. In the terminal level power vision edge intelligence, the vision image of the sensing terminal can be processed by AI chip or small-scale edge computing device to complete the automatic allocation of inspection tasks, as shown in Figure 1. In particular, UAVs have become the most popular device to support automated transmission line inspections with easier operation conditions and acceptable expenses. This technology can obtain real-time videos and digital images for the monitored overhead transmission lines. With recorded inspection images and videos, system operators can apply digital imaging processing techniques, machine learning algorithms, and deep learning algorithms to detect the electrical components and therefore monitor the field more visibly.

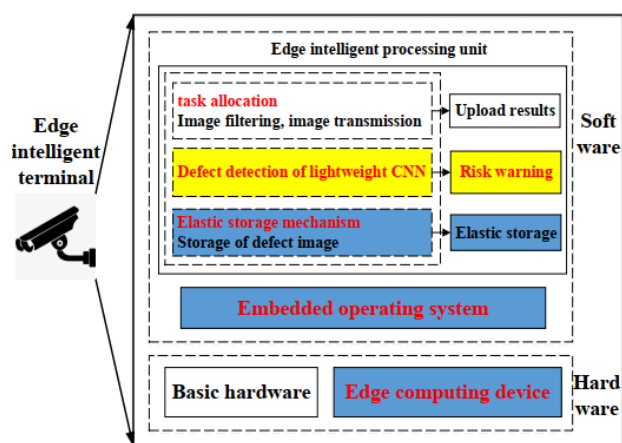


FIGURE 1. Edge computing-based video surveillance system.

Most of the current inspections of transmission lines use aerial photography of drones. Maintenance personnel observe the video of aerial photography on the ground in real time and inspect the transmission lines by the naked eye. With the development of computer vision and pattern recognition, automatic inspection technology for aerial images has been developed [13]–[16]. Combining advanced technologies and methods in the inspection of high-voltage transmission lines, including image acquisition, digital image preprocessing, and final recognition, can not only greatly reduce the number of maintenance personnel required during inspections but also eliminate human involvement caused by subjective factors and improve efficiency to make the assessment of the state of the power grid more accurate and effective.

In recent years, unmanned aerial vehicles (UAVs) with onboard high-definition cameras have developed rapidly, and their applications are very extensive [17]–[20]. Due to their compact model, simple operation, superior human-machine interaction performance, and high resolution of the onboard camera, power researchers are gradually using drones for inspections of transmission lines. Taking advantage of the simplicity and portability of drones, workers began to use drone onboard cameras to inspect transmission lines and capture video images of transmission lines. A large number

of transmission line images taken by drones provide a large quantity of research image data for staff who use image processing methods to study transmission lines. However, the number of video images of power transmission lines taken by aerial drones is extremely large, and the memory occupied is particularly large, which has caused staff to process video images extremely slowly in the later period. How to quickly analyze and process the transmission line images of aerial drones has become one of the important work directions for researchers.

Aerial photography is one of the important sources for detecting transmission line status. The image processing method is used to diagnose anomaly information such as icing status and bird net. The image detection model usually uses image processing to analyze, judge and classify the collected images according to the manually set standards. An online monitoring system can monitor the power transmission lines in real time. However, the online monitoring system cameras are often severely affected by the environment, making the camera easy to cover with ice, blur, and occlusion, resulting in low resolution of the captured images, which seriously affects the judgment accuracy. Therefore, when using image processing methods to judge the information of the transmission line images, the images collected by the online monitoring system cannot be used well. Therefore, how to acquire images suitable for image processing method research is a very valuable research question.

To achieve a fast and accurate overhead transmission line component detection, this work mainly focuses on creating a more profitable inspection image library and improving detecting algorithms. The ImageNet database, an image database organized according to the WordNet hierarchy, is often used as the inputs. for the object detection training model. However, in this database, under the “cable, line, transmission line” syn-set, only a total of 1,290 images can be found. This database cannot provide satisfactory results because the number of images for training is too small and some electrical elements, such as spacer rods, electric power fittings, and some insulators are not included. As a result, the detection accuracy will not be promising. This paper aims to create the first self-learning-based overhead transmission line inspection image library/database. This database contains the ground truth information of various electrical components in the image. They are a benchmark for the training process and can be expanded as more inspection images are included. Another gap is that existing deep learning algorithms still cannot meet the requirement of real-time detection regarding their calculation speed and detection accuracy. This paper improves the structure of the existing AlexNet model in both the convolutional layer and pooling layer to improve the detection speed and accuracy.

Therefore, this paper proposes a transmission line component detection method based on a deep learning framework based on the improved AlexNet model [21], [22] to achieve higher efficiency, effectiveness, and reliability in the line inspection process. In terms of feature extraction,

an improved model to extract the characteristics of power equipment through denoising and normalization is proposed. In terms of the recognition algorithm, after considering the advantages of traditional machine learning methods and the advantages of a support vector machine (SVM), we adopt the SVM classification method combined with deep learning. The improved AlexNet and SVM classification method is used to classify images of various types of power equipment, and the validity of the proposed method is verified. The rest of the paper is organized as follows. Section II briefly reviews some related works, and then we elaborate on the details of our improved AlexNet algorithm in Section III. Moreover, we present the comparative experimental results and analysis in Section IV. Finally, Section V gives the concluding remarks of the paper.

II. RELATED WORK

Anomaly detection in UPIoT is considered a system-level anomaly diagnosis in the dispatch center, where we monitor the status of various types and levels of protection devices and circuit breakers, the voltage and current measurement of electrical quantities [23]–[25]. According to qualitative and quantitative analysis, which are the two main directions for constructing stroke prewarning systems, possible anomaly components and fragile locations are inferred, and the relevant smart decisions are provided to the State Grid. When the grid fails, precise, rapid, automatic emergency maintenance has important significance for the rapid recovery of the power grid. Power systems are prone to faults at times. When anomaly occurs, consumer supply is disrupted, thereby leading to a loss of economy as well. The objective of any electrical power system is to maintain continuity of supply so anomaly is undesirable. Although many anomaly identification methods exist, the use of SVM is almost a novel approach in this concern. A common question arises regarding what the system needs to identify when various signaling systems are already installed. It is just another system for the same purpose, which uses learning techniques to learn anomaly characteristics and apply math to determine whether a new input is faulted or healthy.

The transmission line is the most likely element in the power system to be the location and reason for anomaly, especially when their physical dimension is considered. Intelligent techniques that have their bases on artificial intelligence are under investigation to increase the consistency, rate and accuracy of present digital relays. Traditional digital image processing methods, including Haar wavelet

transformation, the scale invariant feature transform (SIFT) algorithm, and saliency detection and image segmentation, are usually based on image binarization and edge detection [26]–[29]. All the features are predetermined and manually extracted to detect the electrical elements. In reality, it is often hard to obtain an inspection image with satisfactory quality due to the impact of the environment, light, shooting angle, and other uncontrollable conditions. The distortion and mixing of noises in the images will result in reduced processing speed and detection accuracy. As a development, histogram of oriented gradients (HOG) classifiers are used for object recognition by providing them as features to machine learning algorithms such as Bayesian networks and the AdaBoost algorithm. These machine learning algorithms have been widely applied in the electrical component detection process. These methods improve the detection accuracy and reduce processing speed [30], [31]. However, the determination of detection window parameters and possible feature sets rely on prior knowledge and therefore make them hard to apply to any given inspection image set. With the increase in deep learning algorithms, the new trend is to replace the HOG classifiers in machine learning algorithms with more accurate convolutional neural network (CNN)-based classifiers. Several deep learning algorithms have been widely applied for object detection. You only look once (YOLO) and the single-shot multibox Detector (SSD) are the most common single-stage detectors [32]–[36]. For YOLO, detection is a simple regression problem that takes an input image and learns the class probabilities and bounding box coordinates, SSD runs CNN on the input image only once, calculates a feature map and predicts the bounding boxes and classification probability [37]–[41]. Although these detectors have a leading advantage in calculation speed, their recognition performances are not satisfactory, especially when the objects are small or close to each other. The multistage detectors include the regional convolutional neural network (R-CNN), and the improved algorithms, spatial pyramid pooling (SPP-net) and Fast-RCNN. The methodological improvement compared to the single-stage detector is that region proposals are first collected through a selective search (SS) algorithm. From the obtained region proposals, feature extraction and object identification using the CNN model are then accomplished. Furthermore, Faster R-CNN and Mask R-CNN have been designed to improve the detection accuracy as well as reduce the image processing speed. Table 1 compares the advantage and disadvantages of existing CNN-based detecting algorithms.

TABLE 1. The mAP value of the improved AlexNet.

Category	Normal Transmission Line	Abnormal Transmission Line	Bird Nest	Normal Insulator	Abnormal Insulator
1	0.785	0.789	0.872	0.833	0.888
2	0.792	0.794	0.866	0.851	0.891
3	0.783	0.792	0.863	0.848	0.893
4	0.781	0.795	0.856	0.860	0.867
5	0.775	0.798	0.874	0.853	0.875

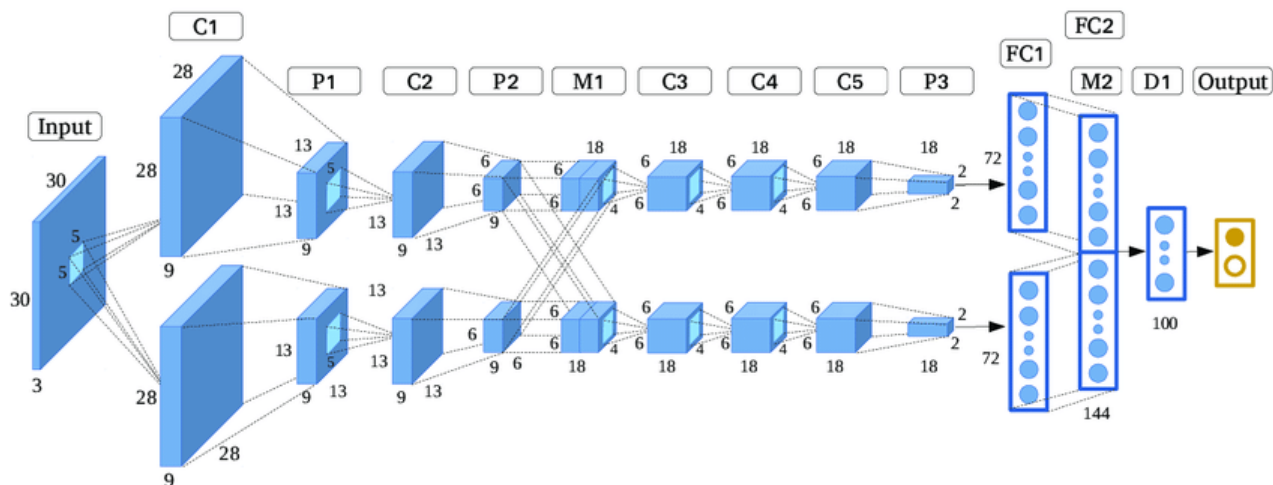


FIGURE 2. AlexNet architecture.

III. AN IMPROVED ALEXNET FOR LINE COMPONENT DETECTION

For an image of a transmission line, the staff can easily determine the content in the image and can clearly analyze whether the conductor is broken, whether the insulator is falling off, and whether there are bird nests. However, in actual aerial inspection work, there are many aerial images at one time. Manually processing these images is very time consuming, and it may delay the repair time for more serious anomaly. Aiming at this problem, this paper uses the popular method of deep learning to process the image of transmission lines, thereby solving the problem of manual workload. Therefore, the staff’s work efficiency is improved to a certain extent, and it has practical application effects. Deep learning is an end-to-end machine learning system that uses raw data as a data input source. Deep learning can automatically learn the deep features of the input data layer by layer. Compared with manual features, deep features are more abstract and expressive. Due to the use of complex models, deep learning can effectively reduce model bias. It can also use a scalable gradient descent algorithm to solve large-scale optimization problems and calculate model parameters through large-scale training data, thereby improving the recognition accuracy. This paper addresses the shortcomings of traditional methods and the practical problem of reducing the workload of workers. It introduces deep learning algorithms to identify the features of transmission line images and uses computers to automatically identify the features of transmission line images to reduce the work burden of personnel. The network used in this article is AlexNet designed and proposed by Krizhevsky *et al.* [21]. The network structure is shown in Figure 2.

A. ALEXNET MODEL

As seen from the AlexNet network structure, Figure 2, the network model is divided into eight layers: the first

five layers are convolution layers, and the next three layers are fully connected layers. Each convolution layer contains the excitation function ReLU and local response normalization (LRN) processing and needs to undergo downsampling (pool) processing. First, a color image is input to the first layer, and this image is extracted by 96 convolution kernels with size specifications $11 \times 11 \times 3$. Each convolution kernel generates a new pixel. After detailed calculation, it finally generates two sets of $55 \times 55 \times 48$ convolved pixel layer data and then uses the ReLU activation function to ensure that the value of the feature map is within a reasonable range. These pixel layer data are processed by the relu1 unit to obtain two sets of active pixel layers. Then, the ReLU activation function is used to ensure that the value of the feature map is within a reasonable range. Then, the generated pixel layer is processed by a maximum pooling operation to obtain $27 \times 27 \times 96$ pixels. Finally, a pixel layer with a size of 5×5 is obtained through normalization processing. The size of the pixel layer formed after the operation of the first convolution layer is $27 \times 27 \times 96$, which corresponds to 96 convolution kernels and is divided into two groups, each of which is operated on an independent GPU. In the second convolutional layer, 256 5×5 filters are used to perform the second feature extraction on the feature map input by the first layer. After going through a process similar to the above, two groups of $13 \times 13 \times 128$ pixel layers are obtained, corresponding to two groups of 128 convolution kernels. Then, the feature map of the second layer is transferred to the third layer and then to the fourth layer until the 4,096-dimensional vector data output by the seventh layer is fully connected with the 1,000 neurons in the eighth layer. After training, the output is the training value. The operations performed by each convolution layer are different, and the specific calculation process is shown in Figure 3.

The process of the AlexNet model can be described as follows: take a sample (X, X_p) from the sample set,

1	2	3	4	5	6	7	8
Cov1	Cov2	Cov3	Cov4	Cov5	Fc6	Fc7	Fc8
Data	norm1	norm2	conv3	conv4	pool5	Fc6	Fc7
conv1	conv2	conv3	conv2	conv2	IP	IP	IP
conv1	conv2	conv3	conv4	conv5	Fc6	Fc7	Fc8
relu	relu	relu	relu	relu	relu	relu	softmax
conv1	conv2	conv3	conv4	conv5	Fc6	Fc7	prob
Conv1	Conv2			Conv5	Fc6	Fc7	prob
pool1	pool2			pool5	dropout	dropout	label
pool1	pool2						
norm1	norm2						

FIGURE 3. The computation diagram of the proposed method.

where X is the input image, X_p is the category of X , and X is also the input value of the whole network, and the output is calculated by

$$O_p = F_n(\dots(F_2(F_1(X_p W(1) + B(1))W(2) + B(2)) \dots W(n) + B(n)) \quad (1)$$

where $W(n)$ ($n = 1, 2, \dots, 8$) is the weight, $B(n)$ ($n = 1, 2, \dots, 8$) is the bias, F_n ($n = 1, 2, \dots, 8$) represents the calculations in each layer, and n is the number of layers of the AlexNet model.

The AlexNet model can calculate the error (the difference between O_p and X_p) and the gradient value of the neural network through backward propagation, and the modified model value can be obtained by adding the calculated gradient to the original model. After obtaining the parameters of the CNN model through training, the test dataset can be used to test the accuracy of the current model.

Due to the particularity of the transmission line image application scenario, the published images can hardly be used directly as the training set of deep convolutional neural networks (DCNNs). Similarly, due to the lack of quantity, the output effect of the DCNN network is not representative or is even wrong. These advantages are greater than those of other networks. By using the AlexNet network, the reliability of the output results can be increased to a certain extent, and the feature of data enhancement also makes the data-poor transmission line image training model more effective.

In this paper, considering the particularity of transmission line feature extraction, the output of the original AlexNet network is improved. The last output layer sigmoid of the original network is removed and dropout, normalized processing and other operations are added to improve the overall performance of the transmission line image feature recognition system. The Figure 2 shows the effect obtained by convolving the image using the AlexNet network. The figure on the left shows that the input image is convolved by 96 convolution kernels. The nucleus has different responses, which are manifested in the light and dark of the black and white color of the feature map. The Figure on the right shows that the same input image is convolved by 256 convolution

kernels. As can be seen, the images have different responses to the 256 convolution kernels. However, after comparison, it can be found that when using 96 convolution kernels, the response effect of the image is better, so this paper uses the output effect of 96 convolution kernels.

The DCNN-based transmission line recognition model proposed in this paper is divided into two phases: the training phase and the testing phase, as shown in Figure 4. The training phase randomly selects 1,000 images from the dataset established in this paper as the training set. The training set contains several types of power transmission line images that have been labeled. This training set is then preprocessed by cropping and segmentation, cropped to a size of 227×227 , and then the cropped image is input to the DCNN to extract depth features. After a series of convolution calculations through deep learning, a 4,096-dimensional feature vector is output. These vectors are actually the probability values of the features in each category of the training image. Then, the feature vectors obtained above are input to the SVM multiclassifier to obtain the final image recognition model of the transmission line. In the test phase, the transmission line feature recognition model obtained above can process images of any size. The smaller the image is, the faster the processing speed, and, in contrast, the larger the image is, the slower the processing speed. Considering the completeness of the image information of the transmission line used for testing, this article chooses to reduce the speed slightly to improve the reliability of the system judgment.

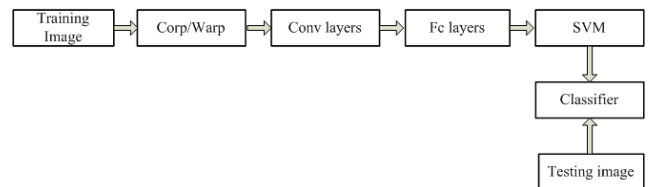


FIGURE 4. The pipeline of the proposed method.

B. FEATURE RECOGNITION FUNCTION

A qualified model should be able to identify the features of the test image based on the features of the training set. In this paper, the features extracted by DCNN are used as the features of the training data set, and then the image retrieval method obtained in the previous work of this article is used to obtain the user preference model.

Assume $U = \{Iu1, Iu2, \dots, Ium\}$ is a manually selected and labeled image set, where Iu represents a labeled image, and then extract the features of the entire training set to obtain a model Γ . For a given image Iu , extract its visual features and retrieve similar images. We use formula (2) to retrieve images with similar features in the joint visual space, where ψ is the adjacent visual space of the marked feature category.

$$S^{(I)} = \{I_i, I_i \in \Gamma \cap I_i \in \Psi\} \quad (2)$$

For each transmission line image used as training one, the same strategy is adopted. m images are dynamically

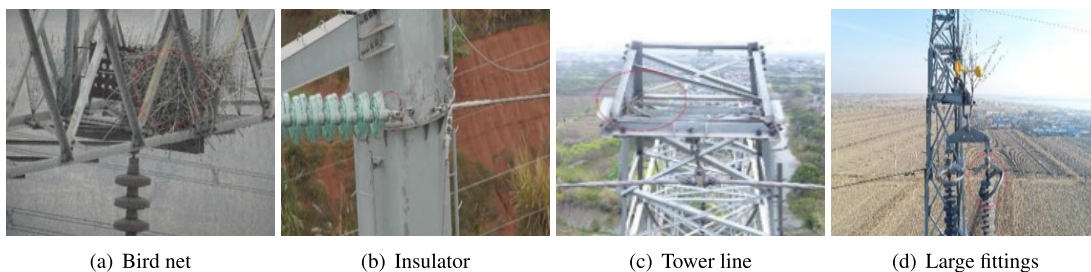


FIGURE 5. The image of tower and its feature maps.

retrieved from Γ with similar features, and these images are connected into a retrieval result S . Once the entire retrieval result is obtained, it can be used to learn a function for transmission line feature recognition. Finally, when an image of a transmission line is again input to the transmission line feature recognition function, it can be automatically determined which category the image belongs to, thus saving staff energy.

During the process of similarity feature retrieval, formula (3) is used to reduce the error in the training phase.

$$f = \begin{cases} \text{ceiling}(I_i), & \forall(I_i, I_j) \in \Gamma, i = j, \\ 0, & \forall(I_i, I_j) \in \Gamma, i \neq j \end{cases} \quad (3)$$

The ceiling function indicates that the retrieved value result jumps to a higher priority. For example, in the labeling process described, a transmission line image contains multiple components, as shown in the bird's nest Figure 5. This figure contains wires, insulators, and a bird's nest. However, when marking, because the wires and insulators are normal and the bird's nest is a kind of anomaly, the bird's nest has higher priority than the wires and insulators in this image, so the features with lower priority should obey the features with higher priority.

IV. EXPERIMENTS

A. DATASETS AND DATA PREPROCESSING

In this section, the overhead transmission line detection test results and analysis are presented. First, the inspection image library visualization is displayed. In addition, parameter determination and selected region proposals are shown. Lastly, detection results using different algorithms are illustrated. All the calculations were conducted on a computer with an Intel, 8-core Xeon processor, a GTX1080 graphics card, and 32 GB of memory. The proposed framework as built and tested on TensorFlow. The test acquired 32,000 transmission line inspection images from the State Grid Corporation of China. UAVs captured all the inspection images with different shooting angles. The original sizes of the inspection images were different from each other. To reduce the calculation time complexity, each image as cropped into the same format as 227×227 . The inspection image library was then built and used for the training and testing procedure.

Thirty thousand images were used as training samples, and the remaining 2,000 were used as testing samples.

Figure 5 displays several inspection image labeling examples in the library. The anchor box info and component label successfully present the ground truth information of all the images. To avoid overfitting in deep learning neural networks, we artificially enlarged the dataset using label-preserving transformations. Transformed images are usually generated from the original images with little computational cost, so these transformed images do not necessarily take up memory space, as shown in Figure 6, 7. In our experiment, these transformed images were produced in Python on a CPU while the GPU was trained on the previous batch of images. Therefore, the data augmentation method is free of computational resources.



FIGURE 6. Data augmentation: deblurring image.

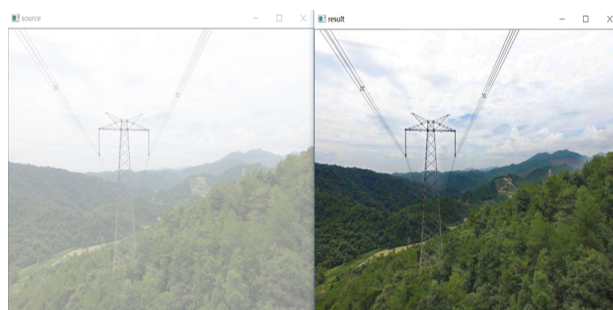


FIGURE 7. Data augmentation: defogging image.

Since the drone airborne cameras currently used are very advanced and the images taken are also very clear, the drone aerial images actually taken are often very large. However,

the oversized image does not meet the requirements of DCNN input and cannot be directly input into DCNN for learning, so the initial input photo needs to be cropped and segmented to meet the DCNN network input image size requirement. The AlexNet network used in this article requires that the size of the input image is 227×227 , so before inputting the DCNN, you need to add the step of compression and cropping. For the convenience of calculation and complete use of the entire image, the input photos were uniformly compressed to $1,135 \times 681$, and then the image was cropped into 227×227 small pieces. This meets the requirements of DCNN input and it can also expand the dataset, and images that do not contain any labeled features after cropping were automatically skipped by the system. Here, it is important to note that the cropped small blocks still retained the original information of the original images, such as the coordinate position information of aerial images. This is very important information. If the detected image is abnormal, then the staff of the maintenance standards need to be notified through the original coordinate position of the imaging equipment. After the image was cropped, these small blocks were stored uniformly as part of the dataset, and then the dataset was labeled with a labeling tool.

B. PERFORMANCE EVALUATION

After several experiments, we trained these datasets for classification. The average precision of classification is shown in Table 1. Table 1 shows the results of several types of images in the datasets. According to the table, the precision of the insulator and the bird's nest were relatively high. The main reason is that the characteristics of the insulator and the bird's nest were relatively obvious and easy to detect compared to the transmission line.

Furthermore, the overall average precision of the transmission line proposed in this paper is calculated as 0.8355, which is a very impressive result in multiclass experiments. Because there is little research on image classification of transmission lines, to illustrate the reliability of the feature recognition algorithm proposed in this paper, we compare it with the other methods that specialize in image classification, as shown in Table 2. Compared with the best detection work, the proposed model has slight limitations. However, compared with similar methods, it still has certain advantages. The classification method in this paper is mainly for the specific images of transmission lines, while the better performance methods of these predecessors' work were aimed at landscapes. There are many studies on landscape-type

TABLE 2. Overall detecting performance of five algorithms.

Algorithms	mAP
Luca et al. [7]	80.56%
Tian et al. [12]	81.68%
Lv et al. [16]	78.24%
Luo et al. [17]	66.79%
Our proposed	83.55%

datasets and improved algorithms. Based on this large number of works, these classification experimental results outperformed our proposed method. To illustrate the reliability of the transmission line feature recognition algorithm in this paper, we attempted to reproduce the algorithm in the literature [7]. After we repeated the experiments using the transmission line dataset instead of the landscape image, the average precision was only 75.56%, 7.99% lower than our algorithm, which demonstrates that the proposed method for transmission line feature recognition is effective and reliable.

V. CONCLUSION

For transmission line feature extraction and anomaly diagnosis, this paper proposes a detection method based on AlexNet as a solution for the diagnosis of anomaly in power transmission components. The experimental results demonstrate that our algorithm has great advantages in the classification performance and wider generalization ability even in a small group of samples in UPIoT. The proposed framework includes the inspection image library collection, data augmentation, improved AlexNet model generation, and performance analysis. The highlights are collecting a transmission line component inspection image datasets and improving the convolutional and pooling layer structure of the model. The proposed improved AlexNet-based framework showed almost 84% mAP includes an insulator, wire, fitting, and tower detection. Through image recognition of the various types of power equipment, the conclusions are as follows:

- The proposed method in this paper can be effectively applied to the image recognition of various types of power equipment, and the obtained accuracy is high.
- The image features extracted by AlexNet have a high abstraction degree and strong expression ability. Compared with the single CNN, our model can obtain richer image features.
- Compared with the other methods, the accuracy of the image recognition by the model classifier is higher.

DATA AVAILABILITY

The data used to support the findings in this study are available from the corresponding author upon request.

CONFLICT OF INTEREST

The authors declare that they have no competing interests.

ACKNOWLEDGMENT

The authors want to thank the authors of the literature cited in this paper for contributing useful ideas to this study. They would like to thank the anonymous reviewers for their helpful insights and suggestions that have substantially improved the content and presentation of this paper.

REFERENCES

- [1] K. Chen, J. Hu, and J. He, "Detection and classification of transmission line faults based on unsupervised feature learning and convolutional sparse autoencoder," *IEEE Trans. Smart Grid*, vol. 9, no. 3, pp. 1748–1758, May 2018.

- [2] S. Han, R. Hao, and J. Lee, "Inspection of insulators on high-voltage power transmission lines," *IEEE Trans. Power Del.*, vol. 24, no. 4, pp. 2319–2327, Oct. 2009.
- [3] Y. Zhao, H. Li, S. Wan, A. Sekuboyina, X. Hu, G. Tetteh, M. Piraud, and B. Menze, "Knowledge-aided convolutional neural network for small organ segmentation," *IEEE J. Biomed. Health Informat.*, vol. 23, no. 4, pp. 1363–1373, Jul. 2019.
- [4] J.-Y. Park, J.-K. Lee, B.-H. Cho, and K.-Y. Oh, "An inspection robot for live-line suspension insulator strings in 345-kV power lines," *IEEE Trans. Power Del.*, vol. 27, no. 2, pp. 632–639, Apr. 2012.
- [5] X. Liu, R. Zhu, B. Jalaian, and Y. Sun, "Dynamic spectrum access algorithm based on game theory in cognitive radio networks," *Mobile Neww. Appl.*, vol. 20, no. 6, pp. 817–827, Dec. 2015.
- [6] K. Muthuvel, S. H. K. Veni, L. P. Suresh, and K. B. Kannan, "ECG signal feature extraction and classification using harr wavelet transform and neural network," in *Proc. Int. Conf. Circuits, Power Comput. Technol. (ICCPCT)*, Mar. 2014, pp. 1396–1399.
- [7] L. Marchesotti, F. Perronnin, D. Larlous, and G. Csurka, "Assessing the aesthetic quality of photographs using generic image descriptors," in *Proc. Int. Conf. Comput. Vis.*, Nov. 2011, pp. 1784–1791.
- [8] G. E. Braga and R. M. R. Junqueira, "Methodology for planning tower leg foundations corrosion maintenance of overhead transmission lines based on GIS," *IEEE Trans. Power Del.*, vol. 31, no. 4, pp. 1601–1608, Aug. 2016.
- [9] S. Wan, Z. Gu, and Q. Ni, "Cognitive computing and wireless communications on the edge for healthcare service robots," *Comput. Commun.*, vol. 149, pp. 99–106, Jan. 2020.
- [10] R. Zhu, X. Zhang, X. Liu, W. Shu, T. Mao, and B. Jalaian, "ERDT: Energy-efficient reliable decision transmission for intelligent cooperative spectrum sensing in industrial IoT," *IEEE Access*, vol. 3, pp. 2366–2378, 2015.
- [11] L. Qi, W. Dou, W. Wang, G. Li, H. Yu, and S. Wan, "Dynamic mobile crowdsourcing selection for electricity load forecasting," *IEEE Access*, vol. 6, pp. 46926–46937, 2018.
- [12] X. Tian, Z. Dong, K. Yang, and T. Mei, "Query-dependent aesthetic model with deep learning for photo quality assessment," *IEEE Trans. Multimedia*, vol. 17, no. 11, pp. 2035–2048, Nov. 2015.
- [13] S. Wan, Y. Xia, L. Qi, Y.-H. Yang, and M. Atiquzzaman, "Automated colorization of a grayscale image with seed points propagation," *IEEE Trans. Multimedia*, early access, Feb. 28, 2020, doi: [10.1109/TMM.2020.2976573](https://doi.org/10.1109/TMM.2020.2976573).
- [14] N. Eltayieb, R. Elhabob, A. Hassan, and F. Li, "An efficient attribute-based online/offline searchable encryption and its application in cloud-based reliable smart grid," *J. Syst. Archit.*, vol. 98, pp. 165–172, Sep. 2019.
- [15] B. Jalaian, R. Zhu, H. Samani, and M. Motani, "An optimal cross-layer framework for cognitive radio network under interference temperature model," *IEEE Syst. J.*, vol. 10, no. 1, pp. 293–301, Mar. 2016.
- [16] P. Lv, M. Wang, Y. Xu, Z. Peng, J. Sun, S. Su, B. Zhou, and M. Xu, "USAR: An interactive user-specific aesthetic ranking framework for images," in *Proc. ACM Multimedia Conf. Multimedia Conf. (MM)*, 2018, pp. 1328–1336.
- [17] Y. Luo and X. Tang, "Photo and video quality evaluation: Focusing on the subject," in *Proc. Eur. Conf. Comput. Vis.* Berlin, Germany: Springer, 2008, pp. 386–399.
- [18] S. Ding, S. Qu, Y. Xi, and S. Wan, "A long video caption generation algorithm for big video data retrieval," *Future Gener. Comput. Syst.*, vol. 93, pp. 583–595, Apr. 2019.
- [19] B. Wang, L. Zhang, H. Ma, H. Wang, and S. Wan, "Parallel LSTM-based regional integrated energy system multienergy source-load information interactive energy prediction," *Complexity*, vol. 2019, pp. 1–13, Nov. 2019.
- [20] X. Xu, C. He, Z. Xu, L. Qi, S. Wan, and M. Z. A. Bhuiyan, "Joint optimization of offloading utility and privacy for edge computing enabled IoT," *IEEE Internet Things J.*, vol. 7, no. 4, pp. 2622–2629, Apr. 2020.
- [21] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, pp. 1097–1105, 2012.
- [22] Y. Xi, Y. Zhang, S. Ding, and S. Wan, "Visual question answering model based on visual relationship detection," *Signal Process., Image Commun.*, vol. 80, Feb. 2020, Art. no. 115648.
- [23] Y. Zhang, G. Cui, S. Deng, F. Chen, Y. Wang, and Q. He, "Efficient query of quality correlation for service composition," *IEEE Trans. Services Comput.*, early access, Apr. 27, 2018, doi: [10.1109/TSC.2018.2830773](https://doi.org/10.1109/TSC.2018.2830773).
- [24] C. Wang, G. Liu, H. Huang, W. Feng, K. Peng, and L. Wang, "MIASec: Enabling data indistinguishability against membership inference attacks in MLaaS," *IEEE Trans. Sustain. Comput.*, early access, Jul. 23, 2019, doi: [10.1109/TSUSC.2019.2930526](https://doi.org/10.1109/TSUSC.2019.2930526).
- [25] P. Zhao, J. Li, F. Zeng, F. Xiao, C. Wang, and H. Jiang, "ILLIA: Enabling k -anonymity-based privacy preserving against location injection attacks in continuous LBS queries," *IEEE Internet Things J.*, vol. 5, no. 2, pp. 1033–1042, Apr. 2018.
- [26] L. Qi, X. Zhang, W. Dou, and Q. Ni, "A distributed locality-sensitive hashing-based approach for cloud service recommendation from multi-source data," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 11, pp. 2616–2624, Nov. 2017.
- [27] X. Xu, R. Mo, F. Dai, W. Lin, S. Wan, and W. Dou, "Dynamic resource provisioning with fault tolerance for data-intensive meteorological workflows in cloud," *IEEE Trans. Ind. Informat.*, early access, Dec. 11, 2019, doi: [10.1109/TII.2019.2959258](https://doi.org/10.1109/TII.2019.2959258).
- [28] C. Chen, J. Hu, T. Qiu, M. Atiquzzaman, and Z. Ren, "CVCG: Cooperative V2 V-aided transmission scheme based on coalitional game for popular content distribution in vehicular ad-hoc networks," *IEEE Trans. Mobile Comput.*, vol. 18, no. 12, pp. 2811–2828, Dec. 2019.
- [29] R. Zhang, P. Xie, C. Wang, G. Liu, and S. Wan, "Classifying transportation mode and speed from trajectory data via deep multi-scale learning," *Comput. Netw.*, vol. 162, Oct. 2019, Art. no. 106861.
- [30] Z. Gao, H.-Z. Xuan, H. Zhang, S. Wan, and K.-K.-R. Choo, "Adaptive fusion and category-level dictionary learning model for multiview human action recognition," *IEEE Internet Things J.*, vol. 6, no. 6, pp. 9280–9293, Dec. 2019.
- [31] S. Wan, Y. Zhao, T. Wang, Z. Gu, Q. H. Abbasi, and K.-K.-R. Choo, "Multi-dimensional data indexing and range query processing via Voronoi diagram for Internet of Things," *Future Gener. Comput. Syst.*, vol. 91, pp. 382–391, Feb. 2019.
- [32] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 779–788.
- [33] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single shot MultiBox detector," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2016, pp. 21–37.
- [34] C. Chen, L. Liu, T. Qiu, D. O. Wu, and Z. Ren, "Delay-aware grid-based geographic routing in urban VANETs: A backbone approach," *IEEE/ACM Trans. Netw.*, vol. 27, no. 6, pp. 2324–2337, Dec. 2019.
- [35] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 7263–7271.
- [36] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018, *arXiv:1804.02767*. [Online]. Available: <http://arxiv.org/abs/1804.02767>
- [37] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 580–587.
- [38] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1440–1448.
- [39] S. Wan and S. Goudos, "Faster R-CNN for multi-class fruit detection using a robotic vision system," *Comput. Netw.*, vol. 168, Feb. 2020, Art. no. 107036.
- [40] J. Wang, J. Luo, X. Liu, Y. Li, S. Liu, R. Zhu, and A. Anjum, "Improved Kalman filter based differentially private streaming data release in cognitive computing," *Future Gener. Comput. Syst.*, vol. 98, pp. 541–549, Sep. 2019.
- [41] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 91–99.



YANPENG GUO was born in Xiaoyi, Shanxi. He received the master's degree. He is currently a Software Engineer with Beijing Smart Core Microelectronics Technology Company, Ltd., Beijing, China. His main research interests include smart grid and deep learning.



ZHENJIANG PANG (Member, IEEE) was born in Harbin, Heilongjiang. He received the master's degree. He is currently a Software Engineer with Beijing Smart Core Microelectronics Technology Company, Ltd., Beijing, China. His main research interests include smart grid and deep learning.



FAN JIANG was born at Chaoyang, Liaoning. He received the bachelor's degree. He is currently a Software Engineer with Beijing Smart Core Microelectronics Technology Company, Ltd., Beijing, China. His main research interests include smart grid and edge computing.



JUN DU was born in Linfen, Shanxi. She received the master's degree. She is currently a Software Engineer with Beijing Smart Core Microelectronics Technology Company, Ltd., Beijing, China. Her main research interests include smart grid and deep learning.



QILONG HU was born in Tai'an, Shandong. He received the bachelor's degree. He is currently a Software Engineer with Beijing Smart Core Microelectronics Technology Company, Ltd., Beijing, China. His main research interests include smart grid and edge computing.

...