

Research Article

Fault Line Selection Method Based on Transfer Learning Depthwise Separable Convolutional Neural Network

Haixia Zhang¹ and Wena Cheng²

¹Henan Industry and Trade Vocational College, Zhengzhou 450000, Henan, China

²Henan Electric Power Survey & Design Institute, Zhengzhou 450000, Henan, China

Correspondence should be addressed to Haixia Zhang; zhanghaixia@hngm.edu.cn

Received 6 September 2021; Revised 25 October 2021; Accepted 29 October 2021; Published 10 November 2021

Academic Editor: Yang Li

Copyright © 2021 Haixia Zhang and Wena Cheng. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the continuous development of artificial intelligence technology, the value of massive power data has been widely considered. Aiming at the problem of single-phase-to-ground fault line selection in resonant grounding system, a fault line selection method based on transfer learning depthwise separable convolutional neural network (DSCNN) is proposed. The proposed method uses two pixel-level image fusions to transform the three-phase current of each feeder into the RGB color image, which is used as the input of DSCNN. After DSCNN self-feature extraction, the fault line selection is completed. With the consideration that not all of power distribution systems can obtain a large amount of data in practical applications, the transfer learning strategy is adopted to transplant the trained line selection model. The smaller number of DSCNN parameters increases the portability of the model. The test results show that not only does the proposed method extract obvious features, but also the line selection accuracy can reach 99.76%. It also has good adaptability under different sampling frequencies, different noise environments, and different distribution network topologies; the line selection accuracy can reach more than 97.43%.

1. Introduction

The structure of the power distribution system is becoming more and more complex. The transmission line faults occur from time to time, such as single-line faults, line-to-ground faults, two-line-to-ground faults, and three-line faults [1]. Arc suppression coils are widely utilized in medium voltage distribution systems. They reduce the fault capacitive current in the case of single-phase-to-ground fault. Therefore, the three-phase line voltage remains symmetrical and maintains power supply to users for a short time [2]. However, operating with the single-phase-to-ground fault will pose the threat to the insulation of the power distribution system. The faulty feeder must be detected in time and accurately, and the faulty feeder must be isolated to prevent further expansion of the fault [3, 4]. When a single-phase-to-ground fault occurs, the fault characteristics are weak due to the adoption of the arc suppression coil. This has increased the difficulty of the fault feeder detection [5, 6].

A series of fault detection methods are proposed for single-phase-to-ground faults, which are mainly divided into three categories: (1) fault transient signal based methods [7, 8], (2) fault steady-state signal based approaches [9–11], and (3) methods based on artificial intelligence technology [7, 12]. The first two methods have certain adoption limitations such as network structure, fault type, and signal noise. With the popularization and application of smart substations and distribution automation equipment, the value of massive power data has been widely considered. Artificial intelligence technology has been widely used in single-phase-to-ground faults detection because of its efficient data processing capabilities, learning efficiency, and superior performance [13–17]. In [13], a method of time-series imaging using three-phase current and three-phase voltage is proposed, and the image is used as the input of the self-attention convolutional neural network for fault identification. In [14], an approach of using continuous wavelet transform to process transient zero-sequence current signals

to obtain time-frequency grayscale images is proposed, and the image is used as input to convolutional neural networks for fault detection. In [15], a method for processing current through wavelet packet transform, extracting energy and entropy to construct a feature matrix, and using support vector machine for fault identification is proposed. In [16], a fault feeder identification method based on discrete wavelet transform and Bayesian selectivity technique is proposed. In [17], an approach of decomposing the three-phase fault current waveform by discrete wavelet transform, extracting features such as standard deviation and energy value, and then using a naive Bayes classifier for fault identification is proposed. The features are extracted from the obtained voltage and current signals by wavelet transform or wavelet packet transform. The artificial neural network is trained by a large amount of data to achieve the goal of effectively identifying the fault feeder. Although the above methods use artificial intelligence algorithms, they extract the feature manually as the input. The choice of wavelet transform scale relies on the research experience of researchers in related fields. The effectiveness of feature extraction is easily impacted by human factors; and these methods only intercept the fault characteristics of the most obvious fault before the one-cycle fault and after the one-to-two-cycle fault signal processing. In the actual resonant grounding system, the transient component is generally very small after the fault occurs for five to six cycles, so the electromagnetic transient process is basically completed. The duration of the transient process varies with types of fault, system wiring, and so forth. Therefore, it is not universal to intercept only two to three periodic signals. When the system structure changes, it will cause a large amount of effective information to be lost.

With the development of artificial intelligence technology, deep learning algorithms are driven by massive amounts of data by constructing multilayer deep neural network models. They have received wide attention for their self-learning characteristics and fast calculation speed [18]. Especially in image recognition, fault diagnosis and other fields have achieved remarkable results [19]. In [20], an attention-based graph residual network, a novel structure of graph convolutional neural network (GCN), was presented to detect human motor intents from raw EEG signals, where the topological structure of EEG electrodes was built as a graph. In [21], a novel deep learning framework based on the graph convolutional neural networks (GCNs) was presented to enhance the decoding performance of raw EEG signals during different types of motor imagery (MI) tasks while cooperating with the functional topological relationship of electrodes. In [22], a long short-term memory (LSTM) based assessment model is built through learning the time dependencies from the post-disturbance system dynamics. It has gradually become a research trend to directly convert the original signal into an image for processing. In the field of motor bearing fault diagnosis, the deep learning algorithm is used to diagnose faults after the vibration signal is converted into a vibration image. The accuracy rate is increased from 80% of the traditional method to about 98% [23–26]. At present, in the field of small current ground fault line selection, although there have been related studies on converting the line selection problem into an

image processing problem, there is no way to directly use the fault current information and completely retain the electromagnetic transient process information combined with deep learning algorithms for the study report of the line selection.

Due to the rapid development of the computer vision algorithm in the field of image classification, the idea of representing the three-phase signals into visual clues is inspired. The signal is represented as a two-dimensional image, so that the model can intuitively classify faults, in order to reveal high-level fault features not found in one-dimensional time series [27]. How to convert the three-phase current signal into a two-dimensional image is a key issue. If you directly project the three-phase current three-dimensional space images of each feeder under the same working condition, three images can be obtained from one feeder at this time. If you directly use convolutional neural networks (CNN) for learning, too many images will inevitably lead to an increase in the numbers of layers and parameters of the CNN network, especially for power distribution systems with a large number of feeders. This increases the time and difficulty of network training. Therefore, the image fusion technology will be used to fuse the projections of each feeder under the same working condition into an RGB color image [28, 29]. In order to explore the deeper features from images, more discriminative artificial intelligence algorithms are needed. In this light, CNN is introduced, which learns advanced features with its deep structure and ensures better classification performance [30]. The multilayer structure of CNN reveals highly discriminatory features while deepening the network. However, in order to improve the performance of the traditional CNN structure, enhance the portability of the model, and reduce the dependence on the hardware platform, this research will replace the standard convolutional layer with a depthwise separable convolutional layer [31] and propose a depthwise separable convolutional neural network (DSCNN) for fault line selection tasks. DSCNN greatly reduces the amount of calculation and increases the portability of the model [32]. In order to cope with the task of small sample data sets, this research also proposes a method to transplant the model by using transfer learning strategy [33].

The main contributions of this research are as follows.

The current signals of one cycle before the fault and nine cycles after the fault are selected for processing to ensure the complete end of the electromagnetic transient process. The projection of each feeder is combined into a full RGB color image that retains the fault information by image fusion.

By replacing the traditional convolutional layer with a depthwise separable convolutional layer, the amount of model parameters and calculations can be reduced to achieve the purpose of network structure lightweight. It can accurately select lines under different signal noises, different sampling frequencies, and different system structures with a strong generalization ability.

Using the transfer learning strategy, the trained DSCNN model is applied to different power distribution system structures, which solves the problem of selecting lines for small sample data sets in actual situations.

Using MATLAB/Simulink software to realize the automatic batch simulation of various failure conditions, the need for the massive data samples required for deep learning is solved. It plays an important reference role for practical engineering applications.

The rest of the paper is organized as follows. The process of generating images based on the fault data under different working conditions and the method of fault line selection are reported in Section 2. The designed modeling of the power distribution system and the specific structure of the DSCNN line selection model are reported in Section 3. The performance of the proposed model is evaluated from many aspects and the method proposed in this paper is compared and analyzed with other existing methods in Section 4. The paper is concluded in Section 5.

2. Propose the Method of Fault Line Selection

In this section, a single-phase-to-ground fault detection method based on DSCNN in resonant distribution systems is proposed, which realizes fault feature extraction and fault feeder detection. The flow chart of the proposed algorithm is shown in Figure 1. When the instantaneous value of the zero-sequence voltage u_0 is higher than the set threshold $K_u U_N$, K_u is the starting reliability factor and U_N is the rated phase voltage of the system. The algorithm of fault feeder detection is started.

The specific implementation process is as follows:

- (1) With the sampling frequency of 12.8 kHz, the three-phase current signals of one cycle before the fault and nine cycles after the fault are recorded to draw a three-dimensional image.
- (2) The three-dimensional image is, respectively, projected on the three projection surfaces of AOB, BOC, and AOC; then three projection images are obtained for each feeder.
- (3) The weighted average image fusion algorithm is used to perform pixel-level image fusion on the three projection images of each feeder; and each feeder obtains a fused image.
- (4) According to the number of system feeders, the fusion image obtained by each feeder in step (3) is combined again using the weighted average image fusion algorithm. Finally, an RGB color image containing the complete fault information is obtained under a condition. The image is used as the input of DSCNN.
- (5) The resulting fusion image is input into the DSCNN fault detection model; it enables automatic fault detection.

2.1. A Two-Dimensional Image Fusion Method for a Three-Dimensional Signal. The three-phase current signal has a strong periodicity. Under normal operation, the images of each period in the three-dimensional space completely overlap into an ellipse. When a single-phase-to-ground fault occurs, the

three-phase current signal in the previous period of the fault is an ellipse in the three-dimensional space, but the three-phase current signal after the fault shows obvious disturbance around the elliptic curve. In order to convert the three-dimensional current signal into a two-dimensional image, the three-dimensional image is projected to three projection surfaces, respectively. Because the deep learning algorithm can automatically focus attention on the salient area, which has a great relationship with the underlying vision, such as color, edge, contour, and other information, the different projected images of different feeders are drawn in different colors; for example, a system of six feeders uses eighteen colors. Since the fault phase is unknown, the significance of the three projected images is consistent. The weighted average fusion method is adopted; namely, the weight of each projected image is given to $1/3$, where i is the feeder label:

$$C'_i(x, y) = \frac{1}{3} (C_{i1}(x, y) + C_{i2}(x, y) + C_{i3}(x, y)). \quad (1)$$

As the fault feeder is also unknown, the significance of each feeder is consistent in the same fault state. The weighted average is still used for secondary fusion. The number of fault feeders is n , so the weight of the fusion image is $1/n$:

$$C'(x, y) = \frac{1}{n} (C'_1(x, y) + C'_2(x, y) + \dots + C'_n(x, y)). \quad (2)$$

The specific process of image fusion is shown in Figure 2. The advantages of transforming 3D signals into 2D images through image fusion are more obvious than simply splicing the signals into 1D signal analysis. The 1D signal can only be analyzed on a single time scale, which usually learns cycle by cycle. It is not good to dig out the characteristics of periodic signals. The two-dimensional fusion image concentrates the fault characteristics of each feeder at different angles around the elliptic curve in the fault state. This improves the correlation between the periodic data, which is conducive to fully mining the characteristics of different fault states.

2.2. Depthwise Separable Convolution. The DSCNN model is composed of multiple maximum pooling layers and depthwise separable convolutional layers. Traditional convolutional neural networks often use large-size convolution kernels for calculations and need to consider all channels in the image area at the same time. Although there are many features obtained in this way, the parameters of the model are very large. Especially when the model has a large number of layers, the amount of calculation is too large and the hardware equipment requirements are high. The model is not suitable for large-scale practical applications. The use of depthwise separable convolution instead of traditional convolution to build a neural network can make it run efficiently on the mobile terminal.

Depthwise separable convolution proposes completing the traditional convolution operation in 2 steps. The first step is to consider the regional characteristics and perform spatial convolution on each input channel separately. This process is called depthwise convolution. The second step is to consider the channel characteristics; the output channels

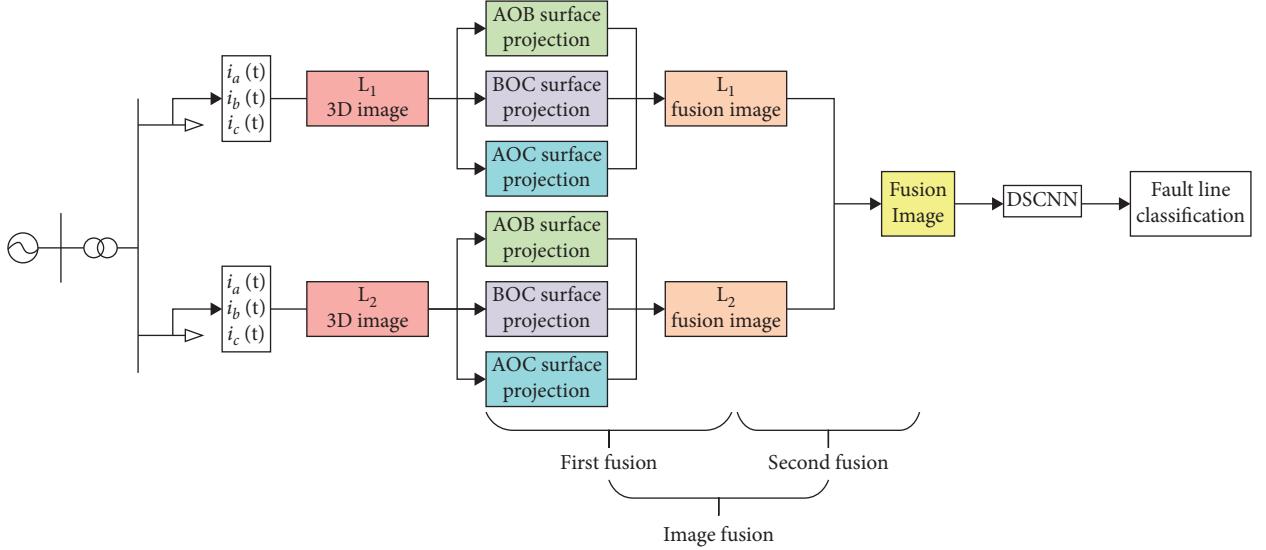


FIGURE 1: Flow chart of the proposed single-phase-to-ground fault detection in power distribution systems.

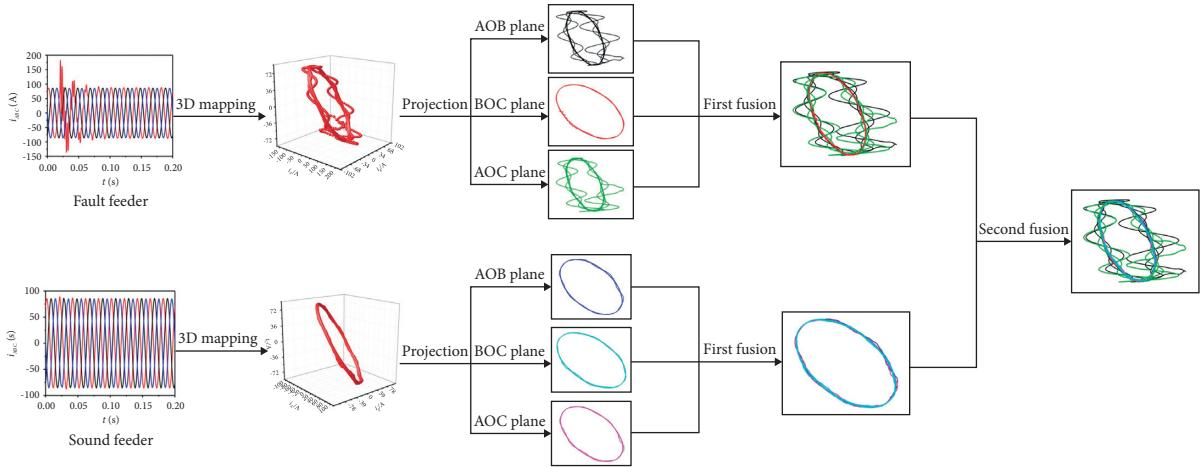


FIGURE 2: The specific process of image fusion.

are mixed through point-by-point convolution (1×1 convolution). This process is called pointwise convolution. Compared with the standard convolution module, the depthwise separable convolution module is shown in Figure 3.

The standard convolution operation is shown in Figure 4. When the input is $N^*H^*W^*C$, the n^*n convolution kernel is used for the operation, and the output is $N^*H^*W^*K$. The calculation amount is $H^*W^*C^*K^*n^*n$.

For the depthwise separable convolutional layer, the depthwise convolution operation is shown in Figure 5. When the input is $N^*H^*W^*C$, n^*n convolution is performed on each channel to extract the spatial characteristics of channel, and the output is $N^*H^*W^*C$. The calculation amount is $H^*W^*C^*n^*n$.

The pointwise convolution operation is shown in Figure 6. The depthwise convolution output Map1 is subjected to 1×1 convolution operation to extract channel features. The final output result is $N^*H^*W^*K$, which is the same as the

standard convolution process result. The calculated amount is $H^*W^*C^*k$.

The compression ratio P of DSCNN to the amount of calculation is shown in the following formula:

$$P = \frac{\text{Depth wise + Point wise}}{\text{convolution}} = \frac{1}{k} + \frac{1}{n * n} \quad (3)$$

It can be seen from the above formula that the reduction of DSCNN calculation is related to the number of output channels k and the size of the convolution kernel. If a 3×3 size convolution kernel is used and the output channel is 128, the calculation amount of DSCNN is only 0.12 times that of standard CNN.

2.3. Transfer Learning Strategy. Deep learning models require massive data training to achieve the best results. In practical applications, the structure of the power distribution system and the number of feeders are different. If you collect

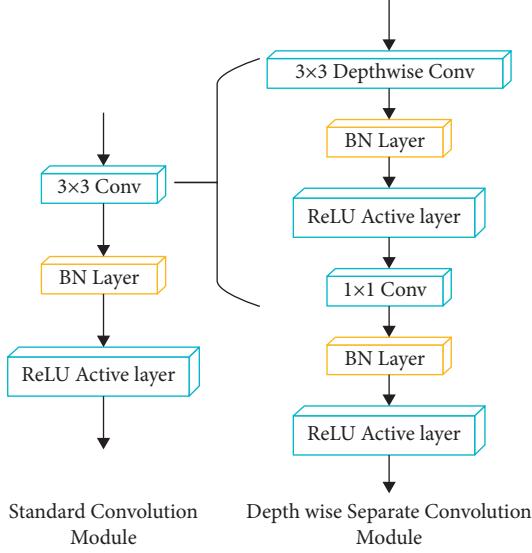


FIGURE 3: Standard convolution and depthwise separable convolution module.

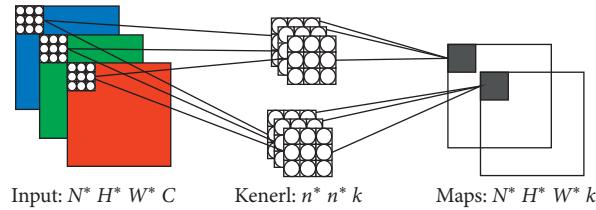


FIGURE 4: Standard convolution operation.

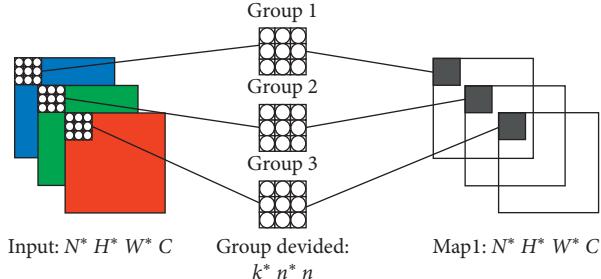


FIGURE 5: Depthwise convolution operation.

a large amount of sample data for each power distribution system and retrain the selection model, it will undoubtedly be a huge waste of manpower, material, and financial resources. There is a very effective way to solve this problem. First, the pretrained network model uses a large data set. Then, a transfer learning strategy is used to apply the pretrained model to a small sample data set different from the original task.

In view of the problem of single-phase-to-ground line selection of resonant grounding system, the characteristics of the fusion image obtained by the different power distribution systems learned by the DSCNN model are similar, but the task of line selection is different. For example, the four-feeder and six-feeder power

distribution systems are divided into four categories questions and six classification questions. The prelearned characteristics of this model are an important advantage of deep learning over shallow learning networks for portability between different tasks; this makes deep learning models very effective for small data problems.

Compared with the standard CNN model, DSCNN has fewer parameters, which is more portable. The implementation of the transfer learning strategy is shown in Figure 7. The feature map extracted by the separable convolutional pooling layer in the DSCNN model represents the existence of common concepts in the image. This feature map is universal in the face of the same type of image classification problem. However, the representation learned by the classification module must be for

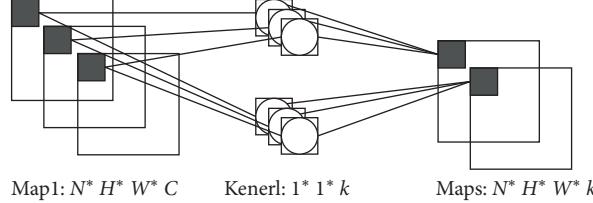


FIGURE 6: Pointwise convolution operation.

a specific task category. The representation of the densely connected layer no longer contains the location information of the target in the input image, so the extracted features are no longer universal for the line selection problem. In summary, this article chooses to freeze the separable convolutional pooling module called feature extraction layer and replace the trained classification module with a new classification module.

The specific implementation steps are as follows.

The classification module in the pretrained DSCNN network is replaced with a new classification module, which includes flatten layer, dense layer, and output layer.

The parameters of the separable convolutional pool layer are frozen, which include four layers of SeparableConv2D and MaxPooling2D layers.

The new classification modules are trained on small data sets to adapt them to new line selection tasks.

3. System Modeling with Data Preparation

3.1. Simulation Model and Parameters. Combined with the actual operating conditions, this article uses MATLAB/Simulink to establish a typical 10 kV resonance grounding system model shown in Figure 8. O_L and C_L denote the length of the overhead feeder and the cable feeder. The feeder parameters of the simulation model are shown in Table 1. When the arc suppression coil is adjusted according to 10% overcompensation, the calculation can get $R_L = 22.51 \Omega$ and $L = 1.434H$.

The fault conditions designed in this paper include different fault phases, different initial phase angles, different fault lines, and different fault distances. The sampling frequency is 12.8 kHz, the sampling time is 0.2 s, and the fault is applied after 0.02 s in the normal operation of the system. Each simulation samples 2560 points of the line current. The model has six feeders and each feeder has three phase lines. The sample size is $3*6*2560$. The fault phase is randomly selected among the three phases A, B, and C. The initial phase angle of the fault is randomly selected from 0° to 360° . The fault location is randomly selected on the overhead line; and the ground impedance is randomly selected from 0.1Ω to 2000Ω . Each feeder generates 6000 fault samples. The entire data set has a total of 36000 samples. Among them, 21600 samples are used as the training sample set matrix $G_{\text{train}} \in R^{21600 \times (2560 \times 6 \times 3)}$, 7200 samples are used as the validation sample set matrix $G_{\text{validation}} \in R^{7200 \times (2560 \times 6 \times 3)}$, the remaining 7200 samples are used as the test sample set matrix $G_{\text{test}} \in R^{7200 \times (2560 \times 6 \times 3)}$, and the specific process is shown in Table 2.

3.2. Two-Dimensional Fusion Image Representation of Three-phase Current. According to Section 2.1, the three-dimensional image projection of the sample data is fused. This paper selects 10 cycles of signals, and the number of sampling points for each sample is 2560. After two pixel-level image fusions, each fault state gets a 150×150 RGB color image. Considering the influence of the initial phase angle of the fault, the fault location, and the fault resistance, the fusion image of the three-phase current under different fault states is shown in Figure 9.

When a single-phase-to-ground fault occurs, the projection colors of each feeder phase are different in the fusion image. The significant characteristics of the fault transient part are very obvious, which are higher than the rest of the fault characteristics. Comparing Figures 9(a) and 9(b), it can be seen that, with the continuous increase of the initial phase angle of the fault, the amplitude of the fault transient current gradually increases, and the amplitude of the disturbance around the ellipse of the fusion image continues to increase. Comparing Figures 9(c) and 9(d), although the fault location is getting closer and closer to the bus, the fusion image only slightly increases the perturbation amplitude at the periphery of the ellipse, and the outline of the ellipse does not change much. This indicates that the fault location has limited influence on the fusion image. As shown in Figures 9(e) and 9(f), with the increase of the fault resistance, the disturbance amplitude and contour around the ellipse of the fused image have changed significantly. In the case of high-resistance grounding, the traditional line selection method has a low line selection accuracy, but the fusion image obtained by the method proposed in this paper still retains more obvious fault characteristics compared with the normal operation state. The elliptic curves cannot be completely overlapped, and the peripheral disturbance still exists.

Through the above analysis, when a single-phase-to-ground fault occurs, the transient current of the fault appears as a disturbance on the periphery of the ellipse in the two-dimensional fusion image. According to the different conditions such as the initial phase angle of the fault, the fault location, and the fault resistance, the amplitude, position, contour, and color of the disturbance are also different. This characteristic information is the basis for further fault line selection.

3.3. DSCNN Model Structure and Visual Analysis. This experiment uses Dell R730 server as the hardware platform, the CPU model is Intel Xeon E5-2630 v4, the memory is 64G,

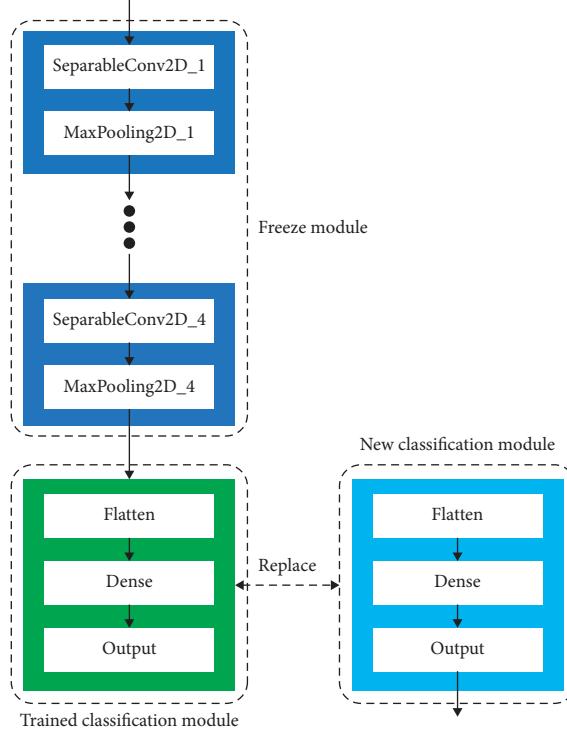


FIGURE 7: Transfer learning strategy implementation diagram.

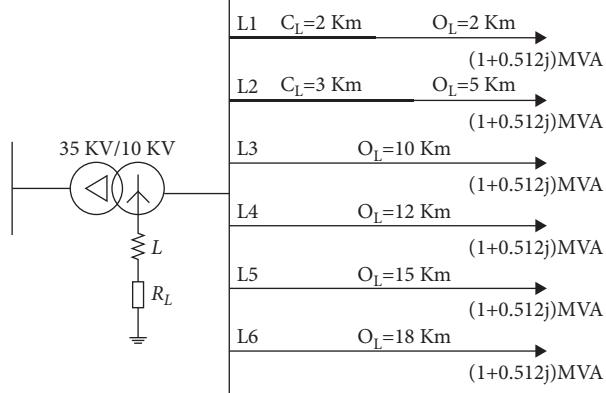


FIGURE 8: Simulation model of 10 kV distribution system.

TABLE 1: Parameters of feeders.

Feeder types	Phase-sequence	R (Ω/km)	L (mH/km)	C ($\mu\text{F/km}$)
Cable feeder	Positive-sequence	0.27	0.255	0.339
	Zero-sequence	2.7	1.02	0.28
Overhead feeder	Positive-sequence	0.178	1.21	0.015
	Zero-sequence	0.25	5.54	0.012

the GPU model is NVIDIA TITAN Xp, and the operating system is Windows Server R2012 R2. The sample processing and model training use *Python* language. The model is implemented with TensorFlow as the backend under the Keras deep learning framework. The typical structure of DSCNN is shown in Figure 10. The parameter configuration

of each layer of DSCNN is shown in Table 3. After replacing the standard convolutional layer with a separable convolutional layer, the parameter comparison is shown in Table 4. P represents the parameter compression ratio. It can be seen from Table 4 that the parameter compression ratio conforms to the theoretical analysis in Section 2.2.

TABLE 2: Automated algorithm simulation process.

Algorithm	Fault simulation automation based on MATLAB/Simulink
1	Call the Simpower System Toolbox to build a 10 kV power distribution system model
2	Configure fault parameters: different fault lines, phases, impedances, and locations
3	Start the simulation and apply a fault when running for 0.02 s
4	Stop the simulation after running for 0.2 s
5	Read 3*6 channel current simulation data (3*6*2560) and save it as a mat file
6	Repeat steps 2–5 to generate 6000 samples required for each faulty feeder
7	Repeat step 6 to get 6 types of fault sample data, for a total of 36000 samples

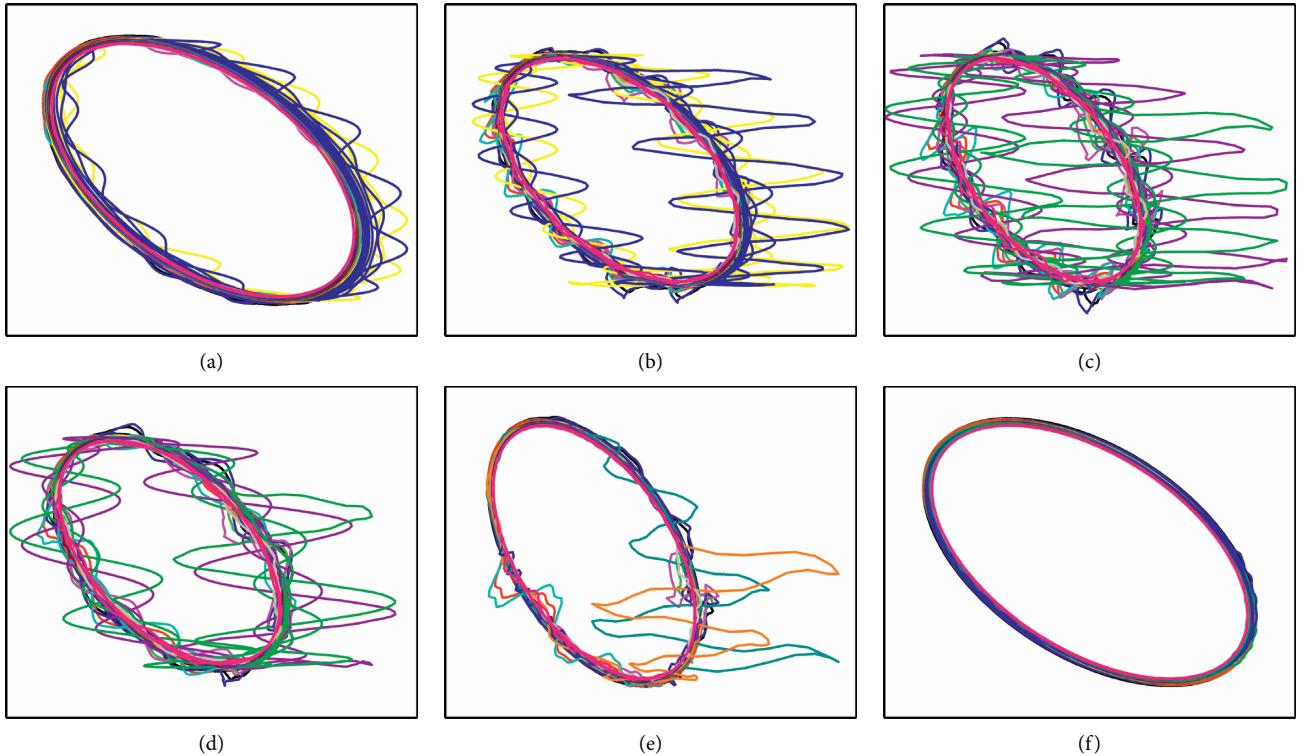


FIGURE 9: Fusion image of three-phase current under different fault conditions. (a) Fault line L_3 , grounding point transition resistance 0.1Ω , fault location 3 km, and failure initial phase angle 0° . (b) Fault line L_3 , grounding point transition resistance 0.1Ω , fault location 3 km, and failure initial phase angle 30° . (c) Fault line L_4 , grounding point transition resistance 0.1Ω , fault location 4 km, and failure initial phase angle 60° . (d) Fault line L_4 , grounding point transition resistance 0.1Ω , fault location 10 km, and failure initial phase angle 60° . (e) Fault line L_5 , grounding point transition resistance 10Ω , fault location 6 km, and failure initial phase angle 90° . (f) Fault line L_6 , grounding point transition resistance 1500Ω , fault location 6 km, and failure initial phase angle 90° .

Considering the influence of learning rate on network model performance, this article uses Adam algorithm to realize automatic adjustment of learning rate. DSCNN automatically extracts features from input samples, the authors conduct a visual analysis of the first convolutional layer. Its feature map size is 148×148 . The figure contains 32 channels, and some of the channel visualization images are shown in Figure 11. Different channels have different sensitivity to different features. The first channel shown in Figure 11(a) is an “elliptical edge” detector, which is sensitive to perturbations on the periphery of an elliptic curve. The 16th channel shown in Figure 11(b) is the “ellipse area” detector, which is sensitive to the overlapped part of the elliptic curve. The 20th channel shown in Figure 11(c) is an “ellipse shape” detector, which is more sensitive to finding

the position and size of the ellipse. The 28th channel shown in Figure 11(d) is the “overall contour” detector, which is sensitive to the overall contour of the image. The first convolutional layer is a collection of various edge detectors. The activation of this layer retains almost all the information of the original image.

As the number of layers deepens, activations will become more and more abstract and difficult to understand intuitively. They filter out irrelevant information and refine useful information. They are abstract features that are more representative of the original image. In order to visually show the ability of DSCNN to extract features, the authors use principal components analysis (PCA) to reduce the data dimension and use t -distributed stochastic neighbor embedding (t -SNE) to convert PCA. The obtained

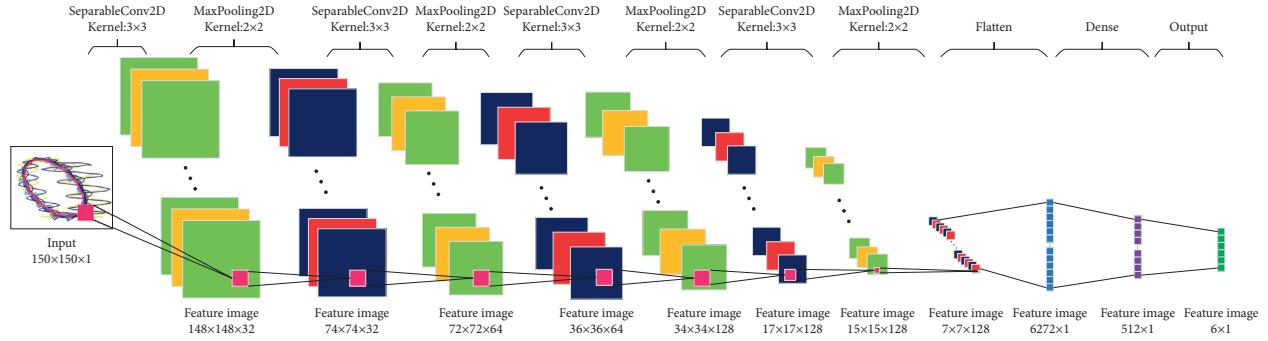


FIGURE 10: Structure of DSCNN algorithm.

TABLE 3: The specific structural parameters of DSCNN.

Layer	Convolution kernel/sampling window size	Output feature map size
Input	---	150 × 150
SeparableConv2D_1	3 × 3	148 × 148 × 32
MaxPooling2D_1	2 × 2	74 × 74 × 32
SeparableConv2D_2	3 × 3	72 × 72 × 64
MaxPooling2D_2	2 × 2	36 × 36 × 64
SeparableConv2D_3	3 × 3	34 × 34 × 128
MaxPooling2D_3	2 × 2	17 × 17 × 128
SeparableConv2D_4	3 × 3	15 × 15 × 128
MaxPooling2D_4	2 × 2	7 × 7 × 128
Flatten	---	6272 × 1
Dense	---	512 × 1
Output	---	6 × 1

dimensionality reduction result is transformed into a two-dimensional image. The visualization of the extracted features of each layer is shown in Figure 12. It can be seen from Figure 12(a) that the features extracted by the first convolutional layer are completely intertwined and cannot be distinguished. After multiple convolutional pooling layers, 128 feature maps are obtained, which represent the 128 abstract features of the original data. The layer obtains a feature vector with a dimension of 6272×1 , and the dimension is reduced to a feature vector of 512×1 through a fully connected layer. As can be seen from Figure 12(e), the feature has a clear dividing line at this time. As the output of the softmax classifier, the output vector of 6×1 is finally obtained. It can be seen from Figure 12(f) that similar features can be effectively clustered, but different features can be clearly distinguished. In summary, through the DSCNN model in this paper, the fault features of the fused image can be effectively separated.

In the figure, 0 represents the first feeder, 1 represents the second feeder, 2 represents the third feeder, 3 represents the fourth feeder, 4 represents the fifth feeder, and 5 represents the sixth feeder.

4. Results

4.1. Comparison of Line Selection Effects under Different Sampling Frequencies. Taking into account the differences in equipment sampling frequencies in actual situations, exploring the line selection effects of the DSCNN model under

different sampling frequencies has important reference significance for practical applications. The authors conduct experiments to obtain sample data by setting five different sampling frequencies as 12.8 kHz, 6.4 kHz, 3.2 kHz, 1.6 kHz, and 0.8 kHz. In order to avoid the interference of random factors and ensure the accuracy of the experimental results, experiments were repeated ten times at each sampling frequency. The averages of accuracy and loss are shown in Table 5.

It can be seen from Table 5 that the average accuracy of line selection decreases with the decrease of sampling frequency, but the average loss is higher and higher. The average accuracy of line selection can reach 99.76% under the sampling frequency of 12.8 kHz. The average accuracy of line selection is only 97.43% when the sampling frequency is 0.8 kHz. Because, in the process of data sampling, some key fault feature information is lost, it is difficult for the model to extract complete fault features under low sampling frequency. The higher sampling frequency can ensure the accuracy of model selection.

4.2. Noise-Resistant Performance Analysis. In order to explore the line selection ability of the DSCNN model proposed in this article under different noise interference conditions, the authors extract the characteristics of the three-phase current through wavelet packet transform (WPT) and input them into the support vector machine (SVM), decision tree (DT), and random forest (RF). Three

TABLE 4: Comparison of CNN and DSNN parameters.

Convolutional layer	Number of channels	CNN parameter amount	DSCNN parameter amount	P
1	32	1108	155	0.14
2	64	18496	2400	0.13
3	128	73856	8896	0.12
4	128	147584	17664	0.12

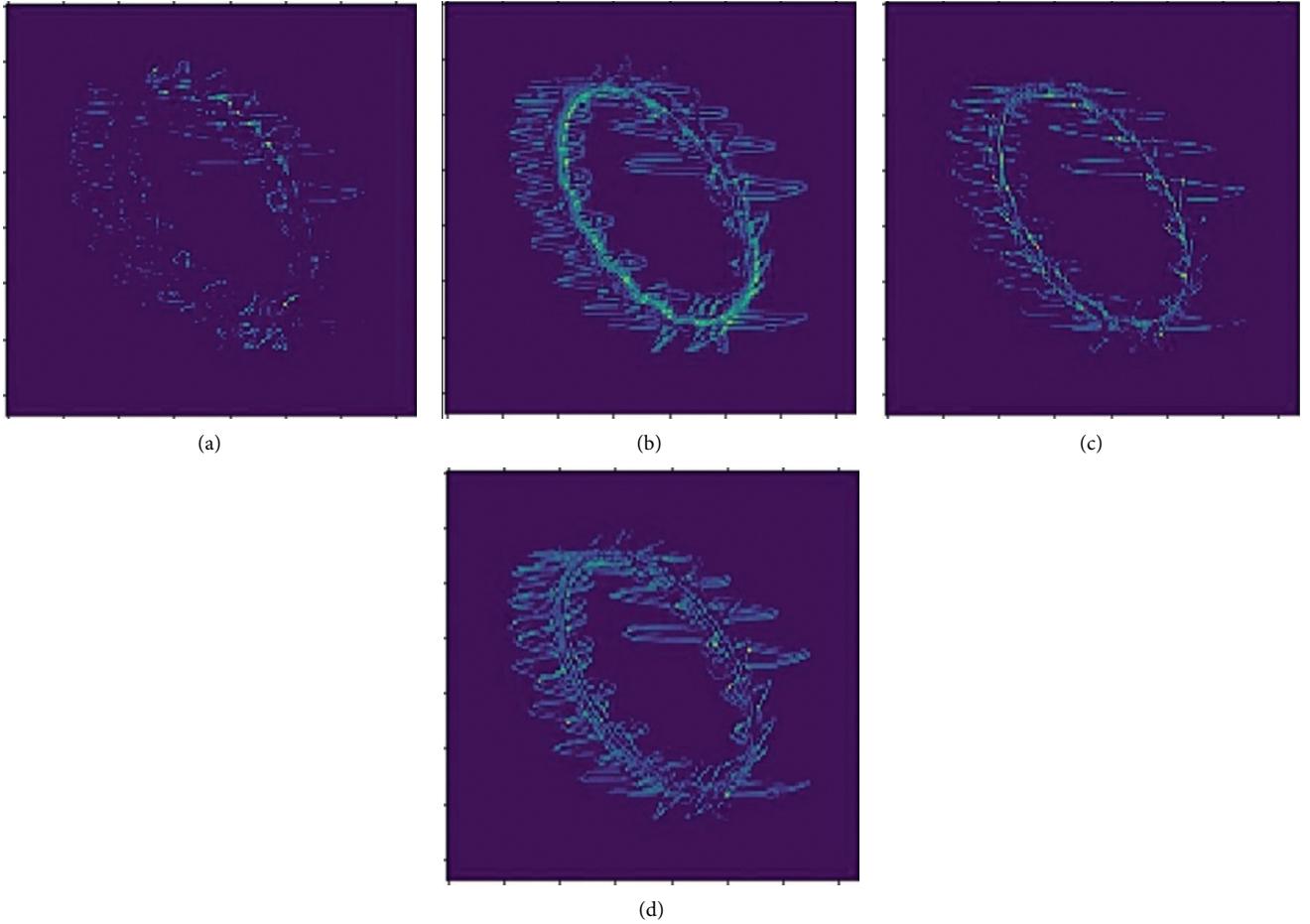


FIGURE 11: Separable convolution part channel visualization. (a) Channel 1 visualization. (b) Channel 16 visualization. (c) Channel 20 visualization. (d) Channel 28 visualization.

commonly used machine learning algorithms are compared with the method proposed in this article. The line selection accuracy of each model under different noise conditions is shown in Table 6.

It can be seen from Table 6 that the DSCNN model has always maintained accuracy above 99.20% by recognizing the fusion image. However, the antinoise ability of the line selection combined with the other three types of machine learning algorithms after manual extraction of features by WPT is not good. The accuracy rate is less than 90%. It can be seen that the method of fusion image combined with DSCNN for line selection proposed in this paper has better antinoise performance than the line selection method that relies on manual extraction of features.

4.3. Time Test of Line Selection Algorithm. In order to further verify that the image after the complete fusion is better than the original projection image of each feeder and the fusion image of each feeder obtained by fusion only once, combined with DSCNN, the performance is better.

In this paper, the original projection image is combined with DSCNN for learning (Original-Model), the primary image is combined with DSCNN for learning (Primary-Model), and the fully fused image is combined with DSCNN for learning (Fully-Model).

The test time and classification accuracy of a test sample of the three methods are calculated. The above experiments are all carried out in the same computer environment. The results are shown in Table 7.

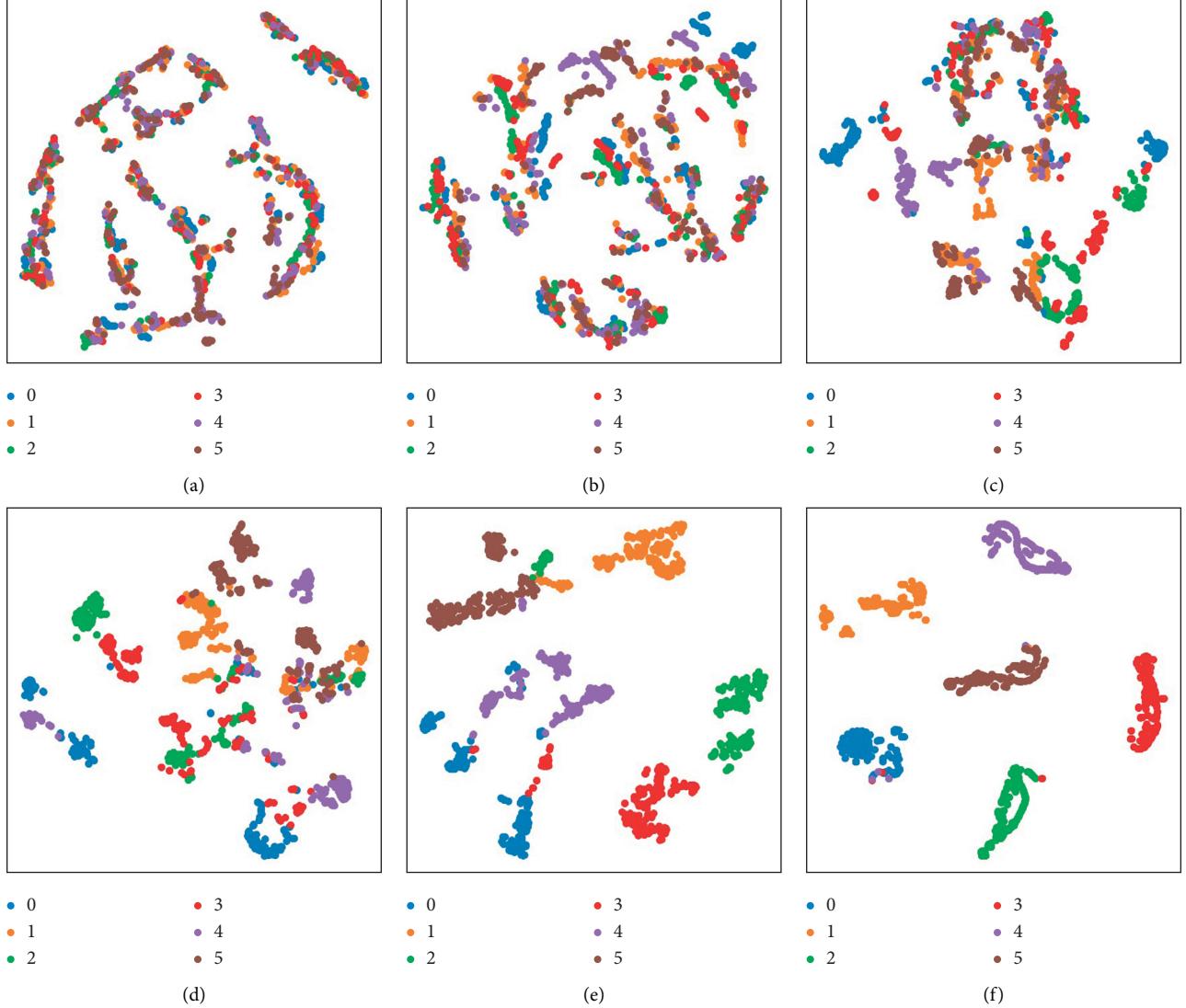


FIGURE 12: DSCNN extract feature visualization. (a) SeparableConv2D_1 visualization. (b) SeparableConv2D_2 visualization. (c) SeparableConv2D_3 visualization. (d) SeparableConv2D_4 visualization. (e) Dense visualization. (f) Output visualization.

TABLE 5: Test sample accuracy and loss average.

Line selection effect parameters	Sampling frequency (kHz)				
	12.8	6.4	3.2	1.6	0.8
Accuracy	0.9976	0.9955	0.9923	0.9843	0.9743
Loss	0.0015	0.0028	0.0081	0.0131	0.0165

It can be seen from Table 7 that Original-Model and Primary-Model take more time and their classification accuracies are lower than that of Fully-Model. This is because the data processing volume of the first two methods is several times that of Fully-Model and contains multiple repetitive feature information, which greatly increases the difficulty of network training and the risk of overfitting. However, the test time for a test sample using fully fused images for deep learning is only 80 ms, which is faster than the human response speed (100–400 ms), and the classification accuracy rate can reach 99.76%.

4.4. Adaptability Analysis of Line Selection Method under Transfer Learning Strategy. In order to explore the adaptability of the DSCNN line selection method using the transfer learning strategy, this paper adds four new types of topological structures as shown in Figure 13 for analysis. The four newly added types of structures generate data according to the method described in Section 3.1. Each type of structure data set has a total of 1000 samples. Among them, 600 samples are used as the training sample set matrix $G_{\text{train}} \in R^{600 \times (2560 \times 6 \times 3)}$, 200 samples are used as the validation sample set matrix $G_{\text{validation}} \in R^{200 \times (2560 \times 6 \times 3)}$, and the

TABLE 6: Classification accuracy of each model under different noise conditions.

Line selection model	SNR/dB			
	30	35	40	45
WPT + SVM	84.21	85.03	85.43	86.68
WPT + DT	86.21	87.96	89.35	89.45
WPT + RF	87.06	88.78	89.84	90.16
DSCNN	99.20	99.28	99.28	99.33

TABLE 7: Test time and classification accuracy of each method.

Method	Testing time (ms)	Accuracy (%)
Original-Model	545	97.45
Primary-Model	326	98.32
Fully-Model	80	99.76

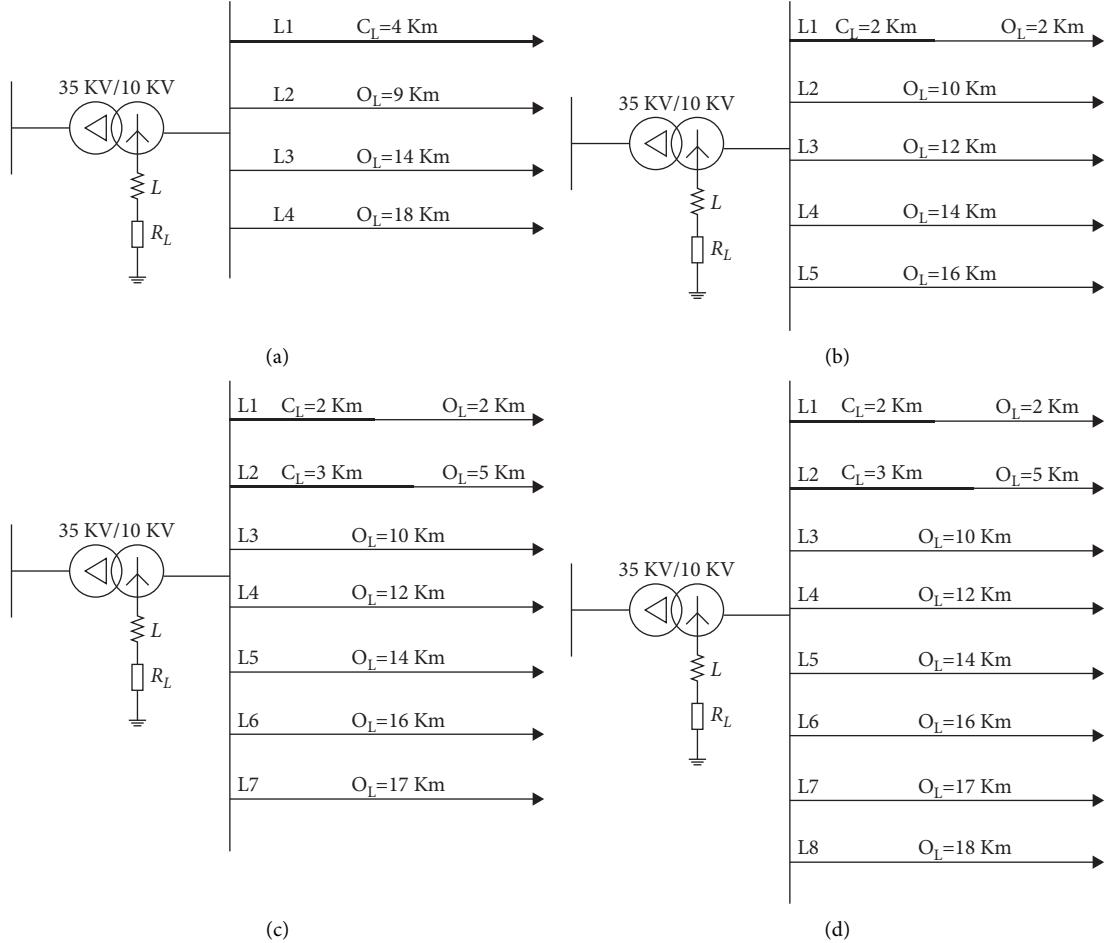


FIGURE 13: New network topology. (a) Model_1: four-feeder power distribution system. (b) Model_2: five-feeder power distribution system. (c) Model_3: seven-feeder distribution system. (d) Model_4: eight-feeder power distribution system.

remaining 200 samples are used as the test sample set matrix $G_{\text{test}} \in R^{200 \times (2560 \times 6 \times 3)}$. The line selection accuracy rate of various structures is shown in Table 8.

It can be seen from Table 8 that the DSCNN line selection method using the transfer learning strategy has good adaptability to change in the network topology. The new power distribution system training samples and test

samples are only 1/36 of the pretraining network. Compared with the original distribution system, the Model_1 structure has improved the line selection accuracy to 99.92%, while the line selection accuracy of the power distribution system with Model_4 structure of eight feeders has been slightly reduced, but it remains at 99.15%.

TABLE 8: Line selection accuracy of different network topologies.

Network structure	Training/validation/test sample size	Accuracy (%)
Model_1	600/200/200	99.92
Model_2	600/200/200	99.81
Model_3	600/200/200	99.28
Model_4	600/200/200	99.15

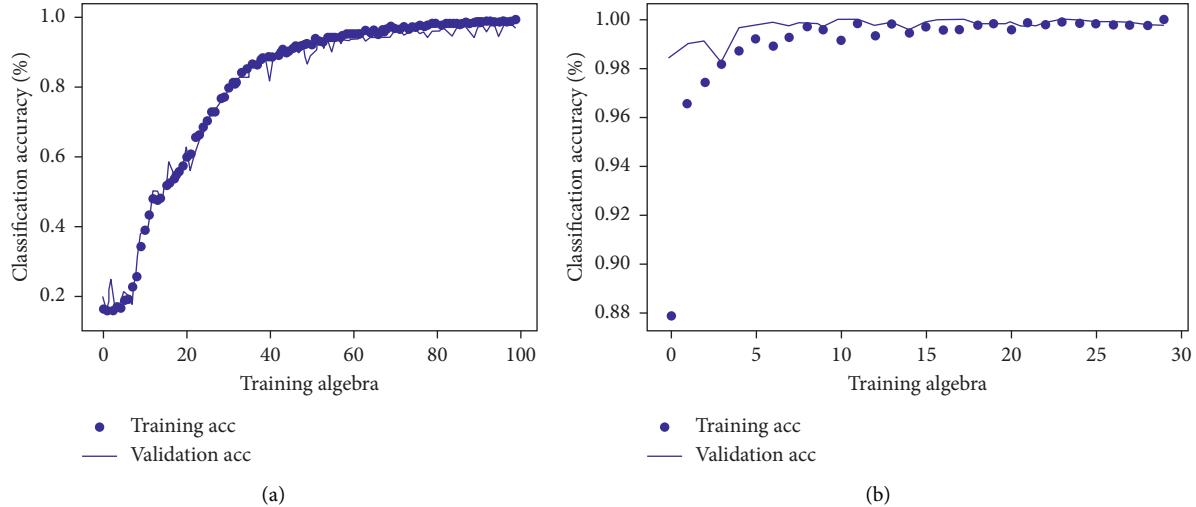


FIGURE 14: Line selection training process. (a) DSCNN training process. (b) The transfer DSCNN training process.

The training process of the four-outlet distribution network structure with the best transfer learning effect of the pretrained DSCNN and the transfer DSCNN models is shown in Figure 14.

After the pretrained DSCNN model fine-tuned training, the transfer DSCNN model only needs 5 iterations of training to achieve a line selection accuracy of about 99%. As the number of iterations increases, the line selection accuracy remains at 99.92% after 15 iterations. It shows that if the number of feeders used by the pretraining network is higher than the number of feeders in the power distribution system after the transfer, the line selection accuracy will be higher.

5. Conclusions

A new method of single-phase-to-ground fault line selection in resonant grounding system is proposed. This method integrates the three-phase current signals of each feeder through two pixel-level image fusions. It can completely retain all fault information. Each fault state obtains an RGB color image, which provides abundant raw materials for effective training of the DSCNN model. By replacing the standard convolutional layer with a separable convolutional layer, the parameter amount is at least 0.14 times lower than the original one, which achieves the goal of lightening the line selection model. The model's dependence on hardware conditions is reduced, and the portability of the model is increased. Under different sampling frequencies, the line selection accuracy can reach up to 99.76%. In different noise environments, the line selection accuracy can reach 99.2% at the lowest. The test results show that the line selection

method proposed in this paper can achieve good performance. At the same time, the transfer learning strategy is used to solve the problem of multiple collection of massive data when the structure of the power distribution system is different, which causes a waste of resources. The source domain line selection network is migrated to the target domain model, and the line selection accuracy can still be maintained above 90%. The test results show that the method proposed in this paper can effectively adapt to the change of the distribution network topology, further verifying the method's performance.

The method in this paper uses simulation data as experimental samples and achieves a high accuracy of line selection. But, for single-phase high resistance, the fault line selection effect when the initial phase angle of the fault is low needs to be further improved. At the same time, considering the difference between the actual distribution network data and the simulation data, the next research direction will use the image generation and fusion method proposed in this paper to adopt a multitype input and multinetwork fusion model. The fault data of the actual power grid are added to the training samples to further improve the line selection model and improve the practicability of the method.

Data Availability

The data in this paper were generated by Matlab/Simulink simulation, and the specific model parameters are derived from Dr. Tang Tao's graduation thesis of Hunan University in 2018: "Research on Single-Phase Grounding Fault Protection Method of Distribution Network."

Conflicts of Interest

The authors declare no conflicts of interest.

Authors' Contributions

Haixia Zhang and Wenao Cheng conceptualized the study; Haixia Zhang developed the methodology; Haixia Zhang provided the software; Haixia Zhang and Wenao Cheng performed validation; Haixia Zhang performed formal analysis; Haixia Zhang performed investigation; Wenao Cheng provided the resources; Wenao Cheng performed data curation; Haixia Zhang and Wenao Cheng wrote the original draft; Haixia Zhang and Wenao Cheng reviewed and edited the article; Haixia Zhang and Wenao Cheng performed visualization; Haixia Zhang supervised the study; Haixia Zhang performed project administration; Haixia Zhang was responsible for funding acquisition. The authors have read and agreed to the published version of the manuscript.

Acknowledgments

The study was funded by Key R&D and extension projects in Henan Province (202102210177).

References

- [1] D. P. Mishra and P. Ray, "Fault detection, location and classification of a transmission line," *Neural Computing & Applications*, vol. 30, no. 5, pp. 1377–1424, 2018.
- [2] J.-A. Jiang, C.-L. Chuang, Y.-C. Wang et al., "A hybrid framework for fault detection, classification, and location-Part I: concept, structure, and methodology," *IEEE Transactions on Power Delivery*, vol. 26, no. 3, pp. 1988–1998, 2011.
- [3] Z. Zhixia, L. Xiao, and P. Zailin, "Fault line detection in neutral point ineffectively grounding power system based on phase-locked loop," *IET Generation, Transmission & Distribution*, vol. 8, no. 2, pp. 273–280, 2014.
- [4] Y. Wang, J. Zhou, Z. Li, Z. Dong, and Y. Xu, "Discriminant-analysis-based single-phase earth fault protection using improved PCA in distribution systems," *IEEE Transactions on Power Delivery*, vol. 30, no. 4, pp. 1974–1982, 2015.
- [5] N. I. Elkalashy, N. G. Tarhuni, and M. Lehtonen, "Simplified probabilistic selectivity technique for earth fault detection in unearthed MV networks," *IET Generation, Transmission & Distribution*, vol. 3, no. 2, pp. 145–153, 2009.
- [6] A. Ghaderi, H. A. Mohammadpour, H. L. Ginn, and Y.-J. Shin, "High-impedance fault detection in the distribution network using the time-frequency-based algorithm," *IEEE Transactions on Power Delivery*, vol. 30, no. 3, pp. 1260–1268, 2015.
- [7] V. Malathi, N. S. Marimuthu, S. Baskar, and K. Ramar, "Application of extreme learning machine for series compensated transmission line protection," *Engineering Applications of Artificial Intelligence*, vol. 24, no. 5, pp. 880–887, 2011.
- [8] P. S. Bhowmik, P. Purkait, and K. Bhattacharya, "A novel wavelet transform aided neural network based transmission line fault analysis method," *International Journal of Electrical Power & Energy Systems*, vol. 31, no. 5, pp. 213–219, 2009.
- [9] K. Krishnanand and P. Dash, "A new real-time fast discrete s-transform for cross-differential protection of shunt-compensated power systems," *IEEE Transactions on Power Delivery*, vol. 28, no. 1, pp. 402–410, 2012.
- [10] S. Devi, N. K. Swarnkar, S. R. Ola, and O. P. Mahela, "Detection of transmission line faults using discrete wavelet transform," in *Proceedings of the 2016 Conference on Advances in Signal Processing (CASP)*, pp. 133–138, IEEE, Pune, India, June 2016.
- [11] A. H. Al-Mohammed and M. A. Abido, "An adaptive fault location algorithm for power system networks based on synchrophasor measurements," *Electric Power Systems Research*, vol. 108, pp. 153–163, 2014.
- [12] M. Jamil, R. Singh, and S. K. Sharma, "Fault identification in electrical power distribution system using combined discrete wavelet transform and fuzzy logic," *Journal of Electrical Systems and Information Technology*, vol. 2, no. 2, pp. 257–267, 2015.
- [13] S. R. Fahim, Y. Sarker, S. K. Sarker, M. R. I. Sheikh, and S. K. Das, "Self attention convolutional neural network with time series imaging based feature extraction for transmission line fault detection and classification," *Electric Power Systems Research*, vol. 187, Article ID 106437, 2020.
- [14] M. F. Guo, X. D. Zeng, D. Y. Chen, and N.-C. Yang, "Deep-learning-based earth fault detection using continuous wavelet transform and convolutional neural network in resonant grounding distribution systems," *IEEE Sensors Journal*, vol. 18, no. 3, pp. 1291–1300, 2017.
- [15] P. Ray and D. P. Mishra, "Support vector machine based fault classification and location of a long transmission line," *Engineering science and technology, an international journal*, vol. 19, no. 3, pp. 1368–1380, 2016.
- [16] N. I. Elkalashy, A. M. Elhaffar, T. A. Kawady, N. G. Tarhuni, and M. Lehtonen, "Bayesian selectivity technique for earth fault protection in medium-voltage networks," *IEEE Transactions on Power Delivery*, vol. 25, no. 4, pp. 2234–2245, 2010.
- [17] E. Aker, M. L. Othman, V. Veerasamy, I. b. Aris, N. I. A. Wahab, and H. Hizam, "fault detection and classification of shunt compensated transmission line using discrete wavelet transform and naive Bayes classifier," *Energies*, vol. 13, no. 1, p. 243, 2020.
- [18] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [19] A. Krizhevsky, I. I. Sutskever, and G. Hinton, "Imagenet classification with deep convolutional neural network-s," in *Proceedings of the Advances in Neural Information Processing Systems*, pp. 1097–1105, Lake Tahoe, USA, December 2012.
- [20] S. Jia, Y. Hou, Y. Shi, and Y. Li, "Attention-based graph ResNet for motor intent detection from raw EEG signals," 2020, <https://arxiv.org/abs/2007.13484>.
- [21] X. Lun, S. Jia, Y. Hou, Y. Shi, and Y. Li, "GCNs-net: a graph convolutional neural network approach for decoding time-resolved eeg motor imagery signals," 2020, <https://arxiv.org/abs/2006.08924>.
- [22] M. Zhang, J. Li, Y. Li, and R. Xu, "Deep learning for short-term voltage stability assessment of power systems," *IEEE Access*, vol. 9, pp. 29711–29718, 2021.
- [23] W. Zhang, G. Peng, C. Li, Y. Chen, and Z. Zhang, "A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals," *Sensors*, vol. 17, no. 2, p. 425, 2017.
- [24] G. Li, C. Deng, J. Wu, X. Xu, X. Shao, and Y. Wang, "Sensor data-driven bearing fault diagnosis based on deep convolutional neural networks and S-transform," *Sensors*, vol. 19, no. 12, p. 2750, 2019.

- [25] C. Lu, Z. Wang, and B. Zhou, *Intelligent Fault Diagnosis of Rolling Bearing Using Hierarchical Convolutional Network Based Health State classification*, Elsevier Science Publishers B. V, Amsterdam, Netherlands, 2017.
- [26] J. W. Oh and J. Jeong, "Convolutional neural network and 2-D image based fault diagnosis of bearing without retraining," in *Proceedings of the 2019 3rd International Conference*, Tokyo Japan, December 2019.
- [27] C.-L. Yang, Z.-X. Chen, and C.-Y. Yang, "Sensor classification using convolutional neural network by encoding multivariate time series as two-dimensional colored images," *Sensors*, vol. 20, no. 1, p. 168, 2020.
- [28] S. Li, X. Kang, L. Fang, J. Hu, and H. Yin, "Pixel-level image fusion: a survey of the state of the art," *Information Fusion*, vol. 33, pp. 100–112, 2017.
- [29] Y. Liu, X. Chen, Z. Wang, Z. J. Wang, R. K. Ward, and X. Wang, "Deep learning for pixel-level image fusion: recent advances and future prospects," *Information Fusion*, vol. 42, pp. 158–173, 2018.
- [30] S. Mei, J. Ji, J. Hou, X. Li, and Q. Du, "Learning sensor-specific spatial-spectral features of hyperspectral images via convolutional neural networks," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 8, pp. 4520–4533, 2017.
- [31] F. Chollet, "Xception: deep learning with depthwise separable convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1251–1258, Honolulu, HI, USA, July 2017.
- [32] A. G. Howard, M. Zhu, B. Chen et al., "Mobilenets: efficient convolutional neural networks for mobile vision applications," 2017, <https://arxiv.org/abs/1704.04861>.
- [33] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, 2010.