ARCHIVES OF ELECTRICAL ENGINEERING VOL. 71(3), pp. 731 –754 (2022)

DOI 10.24425/aee.2022.141682

Identification method for power quality disturbances in distribution network based on transfer learning

PENG HEPING✉, MO WENXIONG, WANG YONG, LUAN LE, XU ZHONG

Guangzhou Power Supply Bureau of Guangdong Power Grid Co., Ltd.
Guangdong, Guangzhou 510620, China
e-mail: 26883526@qq.com

(Received: 01.09.2021, revised: 11.05.2022)

Abstract: For a higher classification accuracy of disturbance signals of power quality, a disturbance classification method for power quality based on gram angle field and multiple transfer learning is proposed in this paper. Firstly, the one-dimensional disturbance signal of power quality is transformed into a Gramian angular field (GAF) coded image by using the gram angle field, and then three ResNet networks are constructed. The disturbance signals with representative signal-to-noise ratios of 0 dB, 20 dB and 40 dB are selected as the input of the sub-model to train the three sub-models, respectively. During this period, the training weights of the sub-models are transferred in turn by using the method of multiple transfer learning. The pre-training weight of the latter model is inherited from the training weight of the previous model, and the weight processing methods of partial freezing and partial fine-tuning are adopted to ensure the optimal training effect of the model. Finally, the features of the three sub-models are fused to train the classifier with a full connection layer, and a disturbance classification model for power quality is obtained. The simulation results show that the method has higher classification accuracy and better anti-noise performance, and the proposed model has good robustness and generalization.

Key words: disturbance identification, distribution network, multiple transfer learning, power quality

1. Introduction

In recent years, with the improvement of national industrialization and the development of science and technology, the living standard of residents and the automation level of manufacturing and production departments are increasing day by day, and people's requirements for higher power

quality are increasing year by year. Providing users with electric energy with quality and stability assurance is of great significance to maintain the safe and stable operation of the power system and ensure the normal power consumption of residents. However, the use of a large amount of new power electronic equipment and the grid connection of various distributed generators makes the power quality problem in the power system increasingly prominent, which greatly affects the normal order of social production and life. In order to improve power quality, it is necessary to evaluate the possible hazards caused by power quality problems and formulate corresponding countermeasures on the premise of full understanding of power quality disturbances. Therefore, it is of great significance to identify and classify disturbance signals of power quality accurately and quickly.

For the identification and classification of power quality disturbance signals, more scholars focus on the combination of disturbance feature extraction and disturbance signal classification. The feature extraction of disturbance signals mainly decomposes and analyzes the time series of disturbance signals of power quality, and extracts the feature value of disturbance signals. Feature extraction methods of disturbance signals used commonly include the fast Fourier transform (FFT) [1], S-transform [2], wavelet transform [3], Hilbert Huang transform (HHT) [4], instantaneous reactive power theory [5], etc. Some classifiers are also used, for example the support vector machine (SVM) [6], K-Neighborhood [7], artificial neural network [8], fuzzy clustering [9], decision tree [10], etc. These methods are based on the traditional machine learning methods [11–14], which need to extract the features of disturbance signals of power quality first, and then design an appropriate classifier to recognize and classify the disturbances. However, the existing feature extraction methods do not have a unified standard, which is easy to produce feature redundancy during extraction, they interfere with the extraction of main features, and then affect the classification accuracy, generalization and anti-noise ability of the whole disturbance recognition and classification system.

Deep learning (DL) is a new data processing and analysis method in the field of machine learning and artificial intelligence. Deep learning network models with excellent performance in the field of data classification include the stacked denoising autoencoder (SDAE) [16], recurrent neural network (RNN) [17], deep belief network (DBN) [18], convolution neural network (CNN) [19], etc. Among these methods, the convolutional neural network (CNN) has achieved great success in the field of two-dimensional image recognition and classification. At the same time, benefiting from the advantages of weight sharing, receiving domain and sub-sampling strategy, the parameters of the convolutional neural network to be optimized are greatly reduced, and their application in various fields is also increasing.

During the identification and classification of power quality disturbance signals, whether one-dimensional signals or traditional two-dimensional images converted into gray scale images [20] are used as inputs of the CNN, accuracy, stability and anti-noise performance cannot be satisfied at the same time [21–36].

To sum up, a review of the existing literature indicates several problems in research of classification of power quality disturbance, which are outlined below:

1. The traditional disturbance classification method for power quality separates feature extraction from feature classification, which is easy to produce feature redundancy in the process of feature extraction, which affects the accuracy and robustness of the classification model.

2. The neural network model is difficult to extract effective features from one-dimensional disturbance signals and obtain good classification results when two-dimensional disturbance images fail to retain all the original signal information.
3. In the absence of samples, the neural network classification model is difficult to achieve satisfactory classification results.
4. The existing disturbance classification method for power quality is generally difficult to provide good classification accuracy and noise resistance, and the recognition rate of these methods for composite disturbances is also not high.

To solve the above problems, this paper presents a disturbance classification model for power quality with accuracy, stability and rapidity. The main contributions and novelty of this paper may be summarised as follows:

1. A method of converting one-dimensional PQD signals into two-dimensional images by using the Gramian angular field is proposed. This method can retain all the information of the original signal and make the corresponding image coding matrix have high sparsity, so as to achieve good anti-noise performance.
2. The concept of multiple transfer learning is proposed, which can greatly improve the classification speed, accuracy and recognition rate of disturbance signals of power quality with noises, and improve the generality of the model in the case of insufficient samples.
3. The feature fusion mechanism is used to fuse the output features of different sub neural networks to make it easier to distinguish the features of disturbance signals of power quality.
4. The ResNet neural network is improved, and a comprehensive model for the disturbance signal classification of power quality based on Gramian angular field and multiple transfer learning is constructed. The model has high precision, high speed, high stability and high noise resistance.

The rest of this paper is organized as follows: Section 2 describes the basic concept of the Gramian angle field and its advantages in encoding two-dimensional images of power quality disturbances. The construction method and basic principles of multiple transfer learning are described in Section 3. In Section 4, the original ResNet neural network is improved and combined with the feature fusion method, a comprehensive disturbances classification model for power quality based on the Gramian angular field and multiple transfer learning is constructed. In order to verify the effectiveness of the proposed method, the simulation comparison method to verify the model from many aspects is used in Section 5. Finally, in Section 6, the conclusion of the work is given.

2. Gramian angular field

The Gramian angular field (GAF) is a matrix reflecting the similarity between vectors of a signal sequence based on the improvement of the Gramian matrix (GM) [22]. The expression of the Gramian matrix $\boldsymbol{G}$ is:

$$\boldsymbol{G} = \begin{pmatrix} \langle \boldsymbol{a_1}, \boldsymbol{b_1} \rangle & \langle \boldsymbol{a_1}, \boldsymbol{b_2} \rangle & & \langle \boldsymbol{a_1}, \boldsymbol{b_n} \rangle \\ \langle \boldsymbol{a_2}, \boldsymbol{b_1} \rangle & \langle \boldsymbol{a_2}, \boldsymbol{b_2} \rangle & & \langle \boldsymbol{a_2}, \boldsymbol{b_n} \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle \boldsymbol{a_n}, \boldsymbol{b_1} \rangle & \langle \boldsymbol{a_n}, \boldsymbol{b_2} \rangle & \cdots & \langle \boldsymbol{a_n}, \boldsymbol{b_n} \rangle \end{pmatrix}, \tag{1}$$

where $\langle \boldsymbol{a_j}, \boldsymbol{b_j} \rangle$ represents the inner product between $\boldsymbol{a_i}$ and $\boldsymbol{b_j}$, and its formula is defined as:

$$\langle \boldsymbol{a_i}, \boldsymbol{b_j} \rangle = \|\boldsymbol{a_i}\| \cdot \left\|\boldsymbol{b_j}\right\| \cdot \cos(\theta), \tag{2}$$

where θ represents the angle between two vectors.

According to the Gramian matrix, the values of each element of the matrix are encoded, and the time-dependent encoded image may be obtained. For the time series such as a disturbance signal, because it is one-dimensional, the inner product of each vector in the corresponding Gramian matrix becomes the point product of each element, and the corresponding calculation formula is given as follows:

$$\langle \boldsymbol{x_i}, \boldsymbol{x_j} \rangle = \boldsymbol{x_i} \cdot \boldsymbol{x_j} \langle \boldsymbol{x_i}, \boldsymbol{x_j} \rangle = \boldsymbol{x_i} \cdot \boldsymbol{x_j}\,, \tag{3}$$

where $\boldsymbol{x_i}$ and $\boldsymbol{x_j}$ represent two elements in the time series, respectively.

Since the process of calculating the inner product in the Gramian matrix is also a process of feature extraction and representation in the time dimension, in order to eliminate the influence of scale differences between features and ensure that each feature is treated equally, the time series $\boldsymbol{X} = [x_1 x_2 \cdots x_n]$ needs to be normalized. For the time series, the formula to scale it to the interval $[-1,\ 1]$ is formulated:

$$\widetilde{x_i} = \frac{(x_i - \max(\boldsymbol{X})) + (x_i - \min(\boldsymbol{X}))}{\max(\boldsymbol{X}) - \min(\boldsymbol{X})}\,, \qquad i = 1, 2, \ldots, n. \tag{4}$$

Using the normalized time series data, the corresponding Gramian matrix may be calculated and encoded, and the Gramian field encoded image of the time series may be obtained. However, the resulting image is blurred and has no degree of recognition. This is because the time series is one-dimensional. Although the corresponding Gramian matrix retains the dependence on time, the values of each element in the matrix show the law of Gaussian distribution, which means the dot product cannot distinguish the information with significant characteristics in the time series data from Gaussian noise, resulting in insufficient sparsity of the matrix. This is unfavorable to the CNN, which is good at dealing with sparse data. In order to preserve all the information in the original data signal when converting the image, it is necessary to expand the dimension of one-dimensional sequence data.

Transforming a time series into a polar coordinate system is a common way to expand the dimension of one-dimensional data. Before the transformation, the time series data to be transformed is scaled by Eq. (4), and then the value of the time series and its corresponding time stamp are encoded and converted into angle value and radius value in polar coordinates, respectively. The specific mathematical expression of coordinate transformation is formulated:

$$\begin{cases} \theta = \arccos(\widetilde{x_i}), & 1 \le \widetilde{x_i} \le 1, \quad \widetilde{x_i} \in \widetilde{\boldsymbol{X}} \\ r = t_i / N, & t_i \in N \end{cases}, \tag{5}$$

where $\widetilde{\boldsymbol{X}}$ is the scaled time series. $\widetilde{x_i}$ and t_i represent the scaling values and corresponding timestamps of each element of the time series respectively. N is the span constant factor of the regularized polar coordinate system. r represents the timestamps after regularization.

It may be seen from Eq. (5) that since the time series data is scaled in the calibration interval $[-1, 1]$ in advance, after coordinate transformation, the corresponding angle in the polar coordinate system is limited in the interval $[0, \pi]$, and because $\cos\theta$ is monotonous at $\theta \in [0, \pi]$. Because the time is monotone, the mapping from the original one-dimensional time series to the two-dimensional polar coordinate system space is both injective and surjective. This mapping method is called bijection. Using the bijection mapping method, the transformed time series data will not lose any information in the original series, but also retain the dependence on time.

Based on the time series in the polar coordinate system, the Gramian angular field is defined as:

$$\boldsymbol{G}_s = \begin{pmatrix} \cos(\theta_1+\theta_1) & \cdots & \cos(\theta_1+\theta_n) \\ \cos(\theta_2+\theta_1) & \cdots & \cos(\theta_2+\theta_n) \\ \vdots & \ddots & \vdots \\ \cos(\theta_n+\theta_1) & \cdots & \cos(\theta_n+\theta_n) \end{pmatrix} = \widetilde{\boldsymbol{X}}^T\widetilde{\boldsymbol{X}} - \sqrt{\boldsymbol{I}-\widetilde{\boldsymbol{X}}^2}^T\sqrt{\boldsymbol{I}-\widetilde{\boldsymbol{X}}^2}, \tag{6}$$

$$\boldsymbol{G}_s = \begin{pmatrix} \sin(\theta_1+\theta_1) & \cdots & \sin(\theta_1+\theta_n) \\ \sin(\theta_2+\theta_1) & \cdots & \sin(\theta_2+\theta_n) \\ \vdots & \ddots & \vdots \\ \sin(\theta_n+\theta_1) & \cdots & \sin(\theta_n+\theta_n) \end{pmatrix} = \sqrt{\boldsymbol{I}-\widetilde{\boldsymbol{X}}^2}^T\cdot\widetilde{\boldsymbol{X}} - \widetilde{\boldsymbol{X}}^T\cdot\sqrt{\boldsymbol{I}-\widetilde{\boldsymbol{X}}^2}, \tag{7}$$

where $\boldsymbol{I} = [1, 1, \cdots, 1]$ is the unit vector, and $\widetilde{\boldsymbol{X}}^T$ represents the transpose of $\widetilde{\boldsymbol{X}}$.

$\boldsymbol{G_S}$ and $\boldsymbol{G_D}$ represents the Gramian summary angular field (GASF) and Gramian difference angular field (GADF), respectively. The difference between $\boldsymbol{G_S}$ and $\boldsymbol{G_D}$ is mainly reflected in the definition of the inner product. In order to use the angle in the polar coordinate system to reflect the relationship between each point itself and each other, both $\boldsymbol{G_S}$ and $\boldsymbol{G_D}$ redefine the inner product [23]:

$$\langle \boldsymbol{x_i}, \boldsymbol{x_j} \rangle_S = \cos(\theta_1+\theta_2) = \boldsymbol{x_i}\cdot\boldsymbol{x_j} - \sqrt{1-\boldsymbol{x_i}^2}\cdot\sqrt{1-\boldsymbol{x_j}^2}, \tag{8}$$

$$\langle \boldsymbol{x_i}, \boldsymbol{x_j} \rangle_D = \sin(\theta_1-\theta_2) = \sqrt{1-\boldsymbol{x_i}^2}\cdot\boldsymbol{x_j} - \boldsymbol{x_i}\sqrt{1-\boldsymbol{x_j}^2}. \tag{9}$$

Compared with the inner product expression of Eq. (3), Eqs. (8) and (9) add additional penalty terms to distinguish valuable data information from Gaussian noise, which is also the key to the high sparsity of the Gramian angular field. Due to the inconsistent definition of the inner product, for different time series, the sparsity of $\boldsymbol{G_S}$ and $\boldsymbol{G_D}$ and the ability to represent the characteristics of the original data signal will be different, which method to choose for image coding needs to be determined according to the actual situation. The two-dimensional coded image reflecting the complete information of the original time series data may be obtained by image coding of the Gramian angular field.

3. Multiple transfer learning based on weight optimization

3.1. Basic principles of transfer learning

Transfer learning is a method to accelerate the training of complex neural networks, and it can also cope with the lack of training samples [29].

Given the source domain, source task, target domain and target task, transfer learning is a method for machine learning that transfers the knowledge obtained from the training of a source learning task in the source domain to the training and learning process of a target task in the target domain. The basic diagram of transfer learning is shown in Fig. 1.

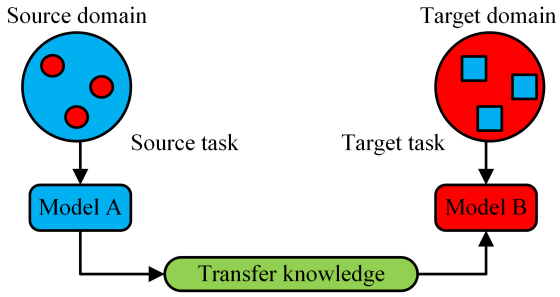


Fig. 1. Process of transfer learning

Transfer learning may be divided into four categories: case-based transfer learning, feature-based transfer learning, weight-based transfer learning and relational knowledge-based transfer learning [29]. The subsequent use of this paper is weight-based transfer learning, that is, the weight parameters of the trained model are transferred to the training and learning of the model under the new target task, which is also the most commonly used learning method in transfer learning.

3.2. Multiple transfer learning

Based on the weight-based transfer learning, multiple transfer learning is to insert one or more similar tasks between the source task and the target task, and transfer the weight layer by layer based on the weight parameters obtained from the model training of the source task. In the target domain, the training weight of the model in the previous intermediate task is used to learn and train the model under the target task, and finally the desired target model is obtained. The schematic diagram of multiple transfer learning is shown in Fig. 2.

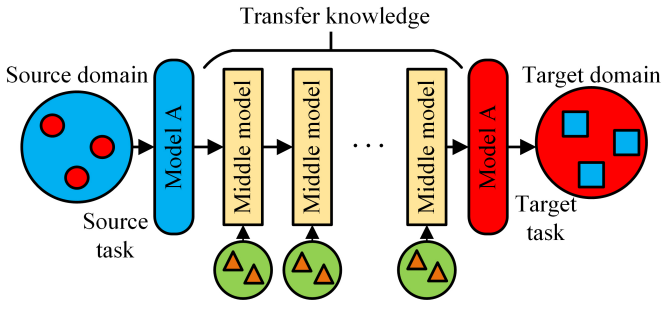


Fig. 2. The process of multiple transfer learning

In Fig. 2, the intermediate task close to the target task needs to be more similar to the target task than the previous tasks. Using multiple transfer learning, the weight parameters of similar tasks may be obtained layer by layer, which can not only improve the accuracy of the target model layer by layer, but also ensure the reliability and stability of the target model. In addition, when the target domain data set is too small, a small amount of data may be used to obtain good training accuracy and improve the universality of the model.

4. Disturbance classification of power quality based on Gramian angular field and multiple transfer learning

The ResNet neural network [28] is selected as the classification neural network in this paper. In order to enhance the convergence speed and classification accuracy of ResNet, this paper improved the basic unit of the residual network by moving the BN layer and rectified linear unit (ReLU) activation layer to the front of the first convolution layer and adding a max pooling layer before the first convolution layer.

Taking the improved ResNet network as the sub-model, this paper selects the disturbance signals of power quality without noise and with signal-to-noise ratios of 20 dB and 40 dB as the input of the sub-model, converts the input signals into GAF pictures and sends them into the sub-model for model training. During the period, multiple transfer learning is used to optimize the model weight, and then the trained sub-model is used to extract the disturbance features, after feature fusion, the extracted features are sent to the classifier to train the full connection layer, and finally a disturbance classification model for power quality with good anti-noise performance and recognition accuracy is obtained. The specific flow chart of the proposed method is shown in Fig. 3.

The specific steps are given as follows:

1. Construct two-dimensional data set. The disturbance signal set of power quality without additional noise and with signal-to-noise ratio of 20 dB, 40 dB are transformed into two-dimensional GAF coded image sets, and the image sets are divided into training set 1 and training set 2 according to their categories.
2. Enhance the training data set. Horizontal flip, random clipping, contrast adjustment and brightness adjustment are used to expand the diversity of sample data, and the data set is standardized.
3. Taking the training set 1 of power quality disturbance without noise and with signal-to-noise ratios of 20 dB and 40 dB as the input of three improved ResNet152 networks. After training sub-model A with pre-training weights trained under the ImageNet data set, the training weights of each sub-model are transferred in turn and the corresponding sub-models are trained with the weight processing method for weight freezing of a shallow network and weight fine-tuning of a deep network, and their respective training weights are retained.
4. Feature fusion. Take the training set 2 of each category as the input of three trained sub-models, load the training weight of each model, and splice the layer of the global pool of the three sub-models to complete the feature fusion.
5. Train the classifier with the full connection layer. The spliced global pooling layer is used as the input of the full connection layer, the weights of the three sub-models are frozen, the full

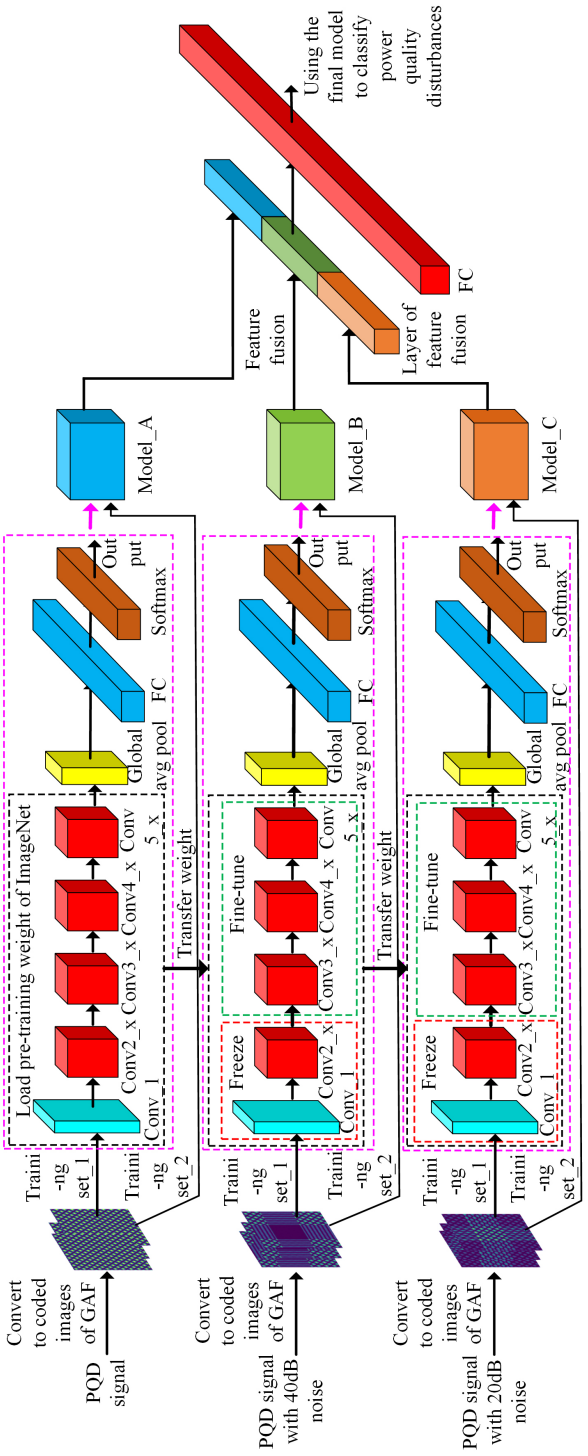


Fig. 3. Flow chart of power quality disturbance classification

connection layer is trained, and finally a disturbance classification model for power quality is obtained.

6. The disturbance signal of power quality under the new random parameters is used as the input of three sub-models to detect the accuracy of the model.

5. Studying cases

5.1. Network parameters and data sets

Under the framework of Pytorch-1.8.0, based on the improved residual basic unit, this paper establishes three ResNet152 network models, which are named model A, model B and model C, respectively. They are used to classify disturbance signals of power quality with different noise. Set the batch size of the model to 32. The optimizer adopts a RAdam optimizer with an adaptive learning rate, and the initial learning rate LR is set to 0.0001. The cross-entropy loss function is selected as the loss function.

In order to fully consider the superposition of disturbances, a total of 17 kinds of disturbance signals of power quality are modeled and adopted in this paper, including the standard signal (C0), voltage sag (C1), voltage swell (C2), voltage interruption (C3), voltage flicker (C4), harmonic (C5), transient pulse (C6), oscillation transient (C7), voltage sag and transient pulse (C8), voltage sag and oscillation transient (C9), harmonic and voltage sag (C10), harmonic and voltage swell (C11), harmonic and voltage flicker (C12), harmonic and transient pulse (C13), harmonic and oscillation transient (C14), harmonic, voltage sag and transient pulse (C15), harmonic, voltage sag and oscillation transient (C16). The detailed model is shown in Table 1. The sampling frequency is 5 000 Hz, and the frequency is 50 Hz of power system frequency. If the sampling interval is 10 cycles (0.2 s), 1 000 data points are sampled in each sampling interval. According to the proposed method, 47 600 disturbance signals of power quality samples with signal-to-noise ratios of 0 dB, 40 dB and 20 dB are generated, respectively. The distribution ratio of training set samples to test set samples is 3:1, and the number of training set samples for each type of disturbance signal is 1 200. In addition, the training set is divided into training set 1 and training set 2, with 20 400 samples each. Training set 1 is used to train three ResNet152 network models, and training set 2 is used to train full connection layer classifiers.

5.2. Generate GAF encoded image

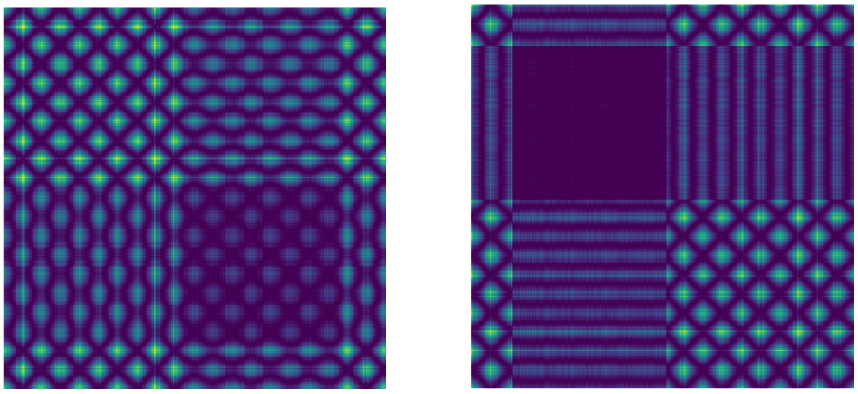
According to the method described in Section 1, the generated disturbance data set of power quality is transformed into a GAF encoded image. For the different redefinition methods of the inner product in the Gramian angular field, the Gramian angular field may be divided into the Gramian angle sum field (GASF) and Gramian angle difference field (GADF), and the corresponding coded images will also be different, which will have a certain impact on the final classification and recognition results. Therefore, in order to achieve the optimal classification effect, this paper first converts the generated disturbance data set of power quality into a GASF encoded image set and GADF encoded image set, respectively, for further selection. The GASF and GADF encoded images obtained by partial conversion are shown in Fig. 4 and Fig. 5, respectively.

Table 1. Mathematical model of disturbance signal of power quality

Disturbance type	Signal model	Parameters
Standard signal	$v(t) = A[1 \pm \alpha(u(t-t_1) - u(t-t_2))]\sin(wt)$	$\alpha \le 0.1, T \le t_2 - t_1 \le 9T$
Voltage sag	$v(t) = A[1 - \alpha(u(t-t_1) - u(t-t_2))]\sin(wt)$	$0.1 \le \alpha \le 0.9, T \le t_2 - t_1 \le 9T$
Voltage swell	$v(t) = A[1 + \alpha(u(t-t_1) - u(t-t_2))]\sin(wt)$	$0.1 \le \alpha \le 0.8, T \le t_2 - t_1 \le 9T$
Voltage interruption	$v(t) = A[1 - \alpha(u(t-t_1) - u(t-t_2))]\sin(wt)$	$0.9 \le \alpha \le 1.0, T \le t_2 - t_1 \le 9T$
Voltage flicker	$v(t) = A[1 + \alpha\sin(\beta wt)]\sin(wt)$	$0.1 \le \alpha \le 0.2,\ 0.1 \le \beta \le 0.5$
Harmonic	$v(t) = A[\alpha_1\sin(wt) + \alpha_2\sin(3wt) + \alpha_3\sin(5wt) + \alpha_4\sin(7wt)]$	$0.05 \le \alpha_2\alpha_3\alpha_4 \le 0.15, \sum \alpha_i^2 = 1$
Transient pulse	$v(t) = \sin(wt) + \alpha[u(t-t_1) - u(t-t_2)]$	$1 \le \alpha \le 3,\ 1\,\text{ms} \le t_2 - t_1 \le 3\,\text{ms}$
Oscillation transient	$v(t) = A[\sin(wt) + \alpha^{-c(t-t_1)/\tau} sinw_n(t-t_1)(u(t_2) - u(t_1))]$	$0.1 \le \alpha \le 0.8, 0.5T \le t_2 - t_1 \le 3T$ $8\,\text{ms} \le \tau \le 40\,\text{ms}\ 300\,\text{Hz} \le f_n \le 900\,\text{Hz}$
Voltage sag and transient pulse	$v(t) = A[1 - \alpha(u(t-t_1) - u(t-t_2))]\sin(wt) +$ $\beta[z(t-t_3) - z(t-t_4)]$	$0.1 \le \alpha \le 0.9, T \le t_2 - t_1 \le 9T$ $1 \le \beta \le 3,\ 1\,\text{ms} \le t_4 - t_3 \le 3\,\text{ms}$
Voltage sag and oscillation transient	$v(t) = A[1 - \alpha(u(t-t_1) - u(t-t_2))]\sin(wt) +$ $\beta^{-c(t-t_3)/\tau} sinw_n(t-t_3)(u(t_4) - u(t_3))$	$0.1 \le \alpha \le 0.9, 0.1 \le \beta \le 0.8$ $T \le t_2 - t_1 \le 9T\ 0.5T \le t_4 - t_3 \le 2T$ $8\,\text{ms} \le \tau \le 40\,\text{ms}\ 300\,\text{Hz} \le f_n \le 900\,\text{Hz}$
Harmonic and voltage sag	$v(t) = A[1 - \alpha(u(t-t_1) - u(t-t_2))] \times$ $[\alpha_1\sin(wt) + \alpha_2\sin(3wt) + \alpha_3\sin(5wt) + \alpha_4\sin(7wt)]$	$0.1 \le \alpha \le 0.9, T \le t_2 - t_1 \le 9T$ $0.05 \le \alpha_2\alpha_3\alpha_4 \le 0.15, \sum \alpha_i^2 = 1$

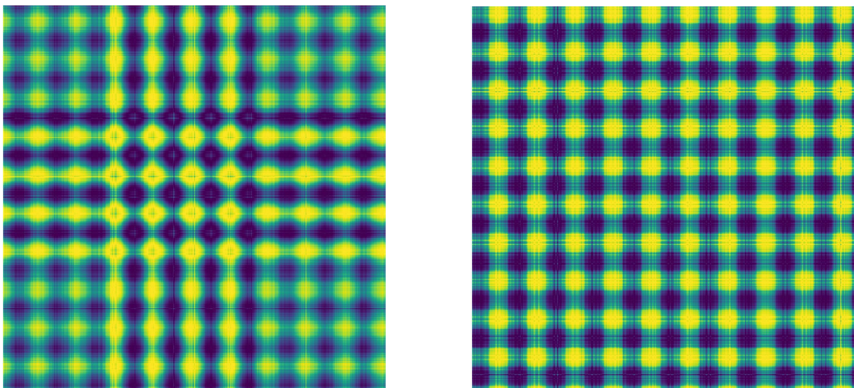
Table 1 [cont.]

Disturbance type	Signal model	Parameters
Harmonic and voltage swell	$v(t) = A[1 + \alpha(u(t-t_1) - u(t-t_2))]\times$ $[\alpha_1 \sin(wt) + \alpha_2 \sin(3wt) + \alpha_3 \sin(5wt) + \alpha_4 \sin(7wt)]$	$0.1 \le \alpha \le 0.8, T \le t_2 - t_1 \le 9T$ $0.05 \le \alpha_2\alpha_3\alpha_4 \le 0.15, \sum \alpha_i^2 = 1$
Harmonic and voltage flicker	$v(t) = A[1 + \alpha \sin(\beta wt)]\times$ $[\alpha_1 \sin(wt) + \alpha_2 \sin(3wt) + \alpha_3 \sin(5wt) + \alpha_4 \sin(7wt)]$	$0.1 \le \alpha \le 0.2, 0.1 \le \beta \le 0.5$ $0.05 \le \alpha_2\alpha_3\alpha_4 \le 0.15, \sum \alpha_i^2 = 1$
Harmonic and transient pulse	$v(t) = A[\alpha_1 \sin(wt) + \alpha_2 \sin(3wt) + \alpha_3 \sin(5wt) + \alpha_4 \sin(7wt)] +$ $\alpha[u(t-t_1) - u(t-t_2)]$	$1 \le \alpha \le 3, 1\ ms \le t_2 - t_1 \le 3$ ms $0.05 \le \alpha_2\alpha_3\alpha_4 \le 0.15, \sum \alpha_i^2 = 1$
Harmonic and oscillation transient	$v(t) = A[\alpha_1 \sin(wt) + \alpha_2 \sin(3wt) + \alpha_3 \sin(5wt) + \alpha_4 \sin(7wt)] +$ $\alpha^{-c(t-t_1)/\tau} sinw_n(t-t_1)(u(t_2) - u(t_1))\}$	$0.1 \le \alpha \le 0.8, 0.5T \le t_2 - t_1 \le 3T$ 8 ms $\le\tau\le$ 40 ms 300 Hz $\le f_n \le$ 900 Hz $0.05 \le \alpha_2\alpha_3\alpha_4 \le 0.15, \sum \alpha_i^2 = 1$
Harmonic, voltage sag and transient pulse	$v(t) = A[1 - \alpha(u(t-t_1) - u(t-t_2))]\times$ $[\alpha_1 \sin(wt) + \alpha_2 \sin(3wt) + \alpha_3 \sin(5wt) + \alpha_4 \sin(7wt)] +$ $\beta[z(t-t_3) - z(t-t_4)]\}$	$0.1 \le \alpha \le 0.9, T \le t_2 - t_1 \le 9T$ $1 \le \beta \le 3, 1$ ms $\le t_4 - t_3 \le 3$ ms $0.05 \le \alpha_2\alpha_3\alpha_4 \le 0.15, \sum \alpha_i^2 = 1$
Harmonic, voltage sag and oscillation transient	$v(t) = A[1 - \alpha(u(t-t_1) - u(t-t_2))]\times$ $[\alpha_1 \sin(wt) + \alpha_2 \sin(3wt) + \alpha_3 \sin(5wt) + \alpha_4 \sin(7wt)] +$ $\beta^{-c(t-t_3)/\tau} sinw_n(t-t_3)(u(t_4) - u(t_3))\}$	$0.1 \le \alpha \le 0.9, T \le t_2 - t_1 \le 9T$ $0.1 \le \beta \le 0.8, 0.5T \le t_4 - t_3 \le 2T$ 8 ms $\le\tau\le$ 40 ms 300 Hz $\le f_n \le$ 900 Hz $0.05 \le \alpha_2\alpha_3\alpha_4 \le 0.15 \sum \alpha_i^2 = 1$



(a) Voltage sag (b) Voltage interruption

Fig. 4. Coded images of power quality disturbance with 20 dB SNR in GASF



(a) Voltage swell (b) Harmonic wave

Fig. 5. Coded images of power quality disturbance with 20 dB SNR in GADF

5.3. Analysis of simulation results

1. Training of sub-models and screening of coded images

Select the training weight of the ResNet152 network obtained from ImageNet dataset training as the pre-training weight input to model A. The GASF encoded image set and GADF encoded image set are used as pre-training weights input to model B and model C, respectively. During the training period, the latter model successively transmits the training weight of the previous model to form multiple transfer learning. The training results are shown in Fig. 6.

It may be seen from the training curve and test curve in Fig. 6 that the training result of the model with the GASF encoded image as input is better than that of the model with the GADF encoded image as input. Therefore, when classifying disturbance signals of power quality,

GASF coding should be used to convert the disturbance signals into two-dimensional images for processing. In the subsequent feature fusion stage of this paper, the sub-model trained with GASF coded images as input will also be used.

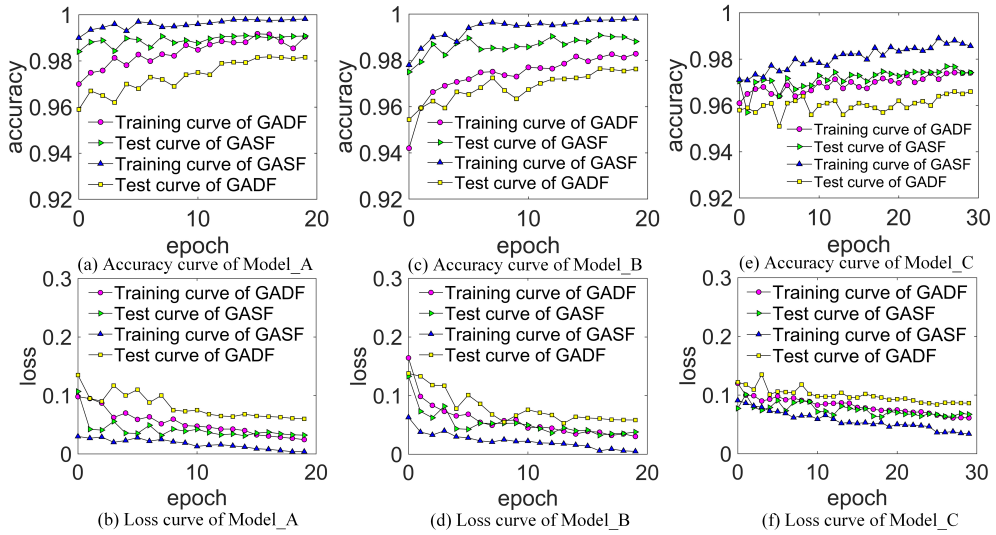


Fig. 6. Accuracy and loss value curve of sub-model under different inputs

In addition, it may be seen from Fig. 6 that due to transfer learning, each model obtains a better training effect after the first iteration. However, the robustness, anti-noise ability and generality of a single sub-model, are all inadequate. Therefore, it is necessary to fuse the models to improve the classification accuracy, stability and generalization of the whole classification system.

2. Feature fusion and result analysis

In order to obtain better classification results, the full connection layer of the trained three sub-models is removed, the global pooling layer of each model is selected, the three global pooling layers from beginning to end are spliced, the deepest features extracted by each model are integrated, and a new full connection layer classifier is constructed. Training set 2 is used to train the full connection layer to obtain the classification model for classification and recognition of disturbance signals of power quality.

New disturbance signals of power quality are randomly selected in batch to verify the model. In order to further prove the effectiveness of the proposed method, this paper compares it with the method proposed in reference [21]. The final results are shown in Table 2.

It may be seen from Table 2 that the method proposed in this paper has significantly improved the overall classification and recognition accuracy compared with the method proposed in literature [21], which can prove that the classification model proposed in this paper not only shows good anti-noise performance but also has a good recognition effect on double disturbance and even triple disturbance of power quality.

Table 2. Classification and recognition results of disturbance signals of power quality with different SNRs

Disturbance type	**Classification accuracy/%**			
	The methods in reference [21]		**The method proposed in this paper**	
	PN4	PN2	PN4	PN2
C0	×	×	100	99.9
C1	98.8	95.8	99.8	99.6
C2	96.2	98.2	99.7	99.6
C3	100	100	100	100
C4	100	100	100	100
C5	100	100	100	100
C6	97.2	96.3	99.4	98.9
C7	96.5	95.7	99.8	99.5
C8	97.8	93.4	99.5	99.0
C9	97.4	94.7	99.7	99.2
C10	×	×	99.6	99.3
C11	×	×	100	99.5
C12	100	97.6	100	99.7
C13	96.3	94.4	99.6	99.2
C14	96.7	94.6	99.5	99.0
C15	96.5	92.1	99.2	98.6
C16	97.0	92.0	99.4	99.0
average value	98.03	95.91	99.72	99.41

PN4: disturbance signals of power quality with noise of 40 dB SNR
PN2: disturbance signals of power quality with noise of 20 dB SNR

5.4. Network parameter and structure analysis

Hyperparameters and network structure are two key factors that affect the performance of neural networks. This section will respectively analyze the impact of different hyperparameters and different network structure on the classification accuracy of the improved ResNet network mentioned above through comparative experiments without using transfer learning, and verify the effectiveness of the proposed network. Experiments are carried out using 17 kinds of disturbance signals of power quality without the additional noise mentioned above, the specific experimental results are shown in Table 3 and Table 4.

For different hyperparameters, it is necessary to find an optimal value to satisfy the requirements of both convergence speed and classification accuracy. As can be seen from the experimental

Table 3. The influence of network parameters on the model recognition rate

Batch Size			Initial learning rate			Epochs		Kernel Size		
V	A (%)	N	V	A (%)	N	V	A (%)	V	A (%)	T (millon)
8	92.58	50	0.00001	93.77	75	40	90.23	3	94.25	60.19
16	93.45	60	0.0001	94.13	80	60	92.44	5	94.67	112.81
32	94.25	80	0.001	92.58	120	80	94.67	6	95.13	148.99
64	94.65	130	0.01	70.45	–	100	93.54	7	94.79	191.75
128	93.87	200	0.1	64.68	–	120	92.21	9	94.92	297.00

V: Parameter value
A: Average classification accuracy
T: Total parameters of network
N: Number of epochs required for model convergence

Table 4. The influence of network structure on the model recognition rate

Neural network structure			Residual unit			Optimizer		
TR	A (%)	N	TU	A (%)	N	TO	A (%)	N
ResNet34	55.86	–	Original residual unit	92.89	95	Adam	93.35	90
ResNet50	90.23	120						
ResNet101	93.87	75	Improved residual unit	94.28	80	RAdam	94.21	80
ResNet152	94.35	80						

TR: Type of ResNet, TU: Type of residual unit, TO: Type of optimizer

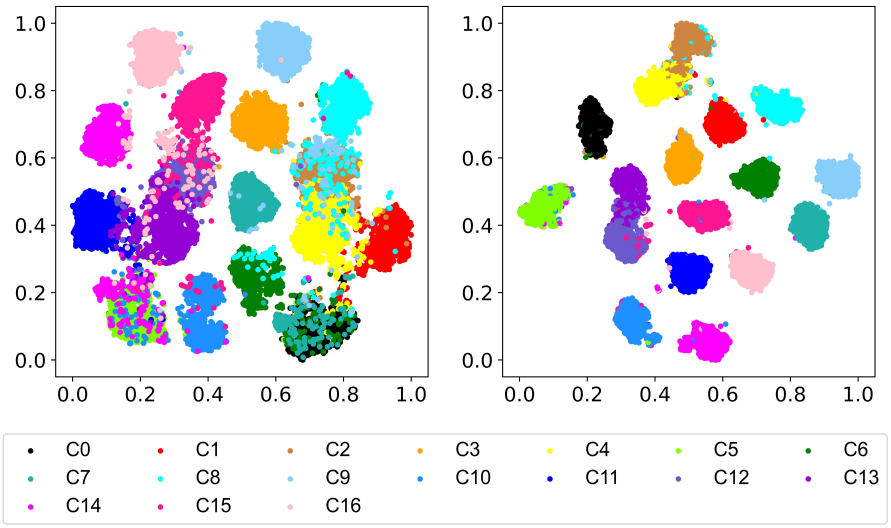
results, when batch size = 32, initial learning rate = 0.0001, epochs = 80, kernel size = 3, the model has the best effect.

In addition, it can be seen from Table 4 that the network structure constructed in this paper also achieves the optimal performance of the model

5.5. Model validity analysis

5.5.1. Effectiveness analysis of feature fusion

In order to show the influence of feature fusion on the feature expression ability of the model in a more intuitive form, this section uses t-SNE visualization technology [37] to conduct comparative experiments to verify the effectiveness of feature fusion. In this paper, 17 kinds of disturbance signals of power quality with a noise of 20 dB SNR values are used as the input of the sub-model and the model obtained after feature fusion in Section 5.1, then, the high-dimensional features extracted from the two models are mapped to two dimensions by t-SNE, respectively, and finally, a visual analysis is carried out. The simulation results are shown in Fig. 7.



(a) Sub-model feature visualization (b) Feature visualization after feature fusion

Fig. 7. Feature visual comparison before and after feature fusion

Figure 7(a) is the feature visualization diagram of the sub-model, Fig. 7(b) is the feature visualization diagram of the model with feature fusion. It may be seen intuitively that compared with the features in the sub-model, the distance between the features representing different signal types after feature fusion is significantly increased, and the positions between the features representing the same signal type are more concentrated. This shows that feature fusion enhances the expression ability of each feature and improves the recognition ability of the model for different types of disturbance signals of power quality. Therefore, this verifies the effectiveness of feature fusion in the model proposed in this paper.

5.5.2. Effectiveness analysis of transfer learning

In order to further verify the effectiveness of transfer learning and multiple transfer learning, taking sub-model C in Section 5.1 as the experimental object, this paper constructs three different sub-models: a sub-model without transfer learning, a sub-model with primary transfer learning and a sub-model with multiple transfer learning. Then, taking the disturbance signal of power quality with a noise of 20 dB SNR as the input, the classification effects of the three models are simulated and analyzed, as shown in Fig. 8.

It may be clearly seen from Fig. 8 that the sub-model with multiple transfer learning is better in classification accuracy and convergence speed than the sub-model with only primary transfer learning, additionally, both of them show a good classification effect at the first epoch. However, the simulation results of the sub-model without transfer learning are far from satisfactory, even until convergence, it does not reach the classification accuracy of the model with transfer learning. Therefore, to sum up, the advantages of the multiple transfer learning proposed in this paper are verified.

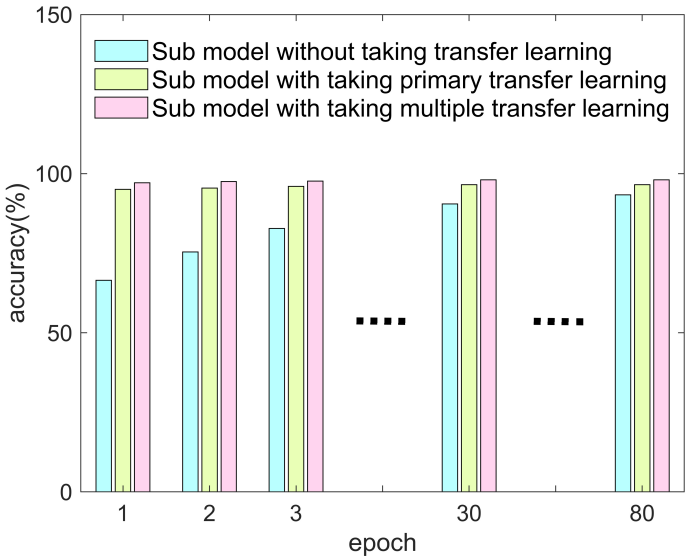


Fig. 8. The influence of transfer learning on the model recognition rate

5.6. Superiority analysis of Gramian angular field

In order to further verify the superiority of the Gramian angular field, this section converts the disturbance signal of power quality into a two-dimensional gray scale image for analysis.

The traditional image coding method [38] is used to encode the one-dimensional sequential disturbance signal of power quality into two-dimensional gray-scale pictures. The coded gray-scale pictures of some power quality disturbances are shown in Fig. 9.

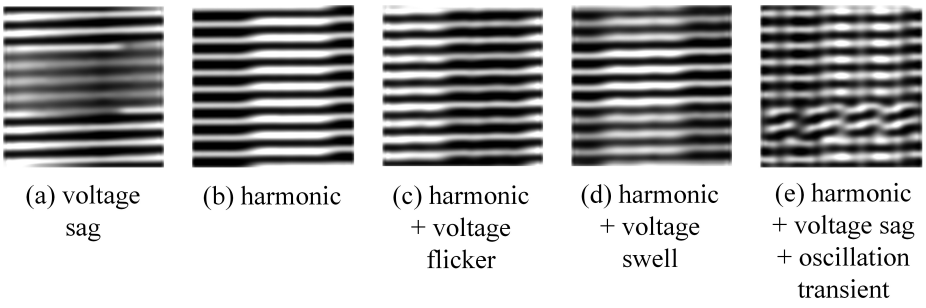


Fig. 9. Gray scale characteristic diagram of power quality disturbance

The encoding method of a gray-scale image is only the process of normalizing the sampling points of a one-dimensional time-series signal and then transforming them into a pixel matrix through simple rules. The sparsity of such a feature matrix is not enough. When Gaussian noise is added, the features of the corresponding picture become blurred, which may be clearly seen from Fig. 10.

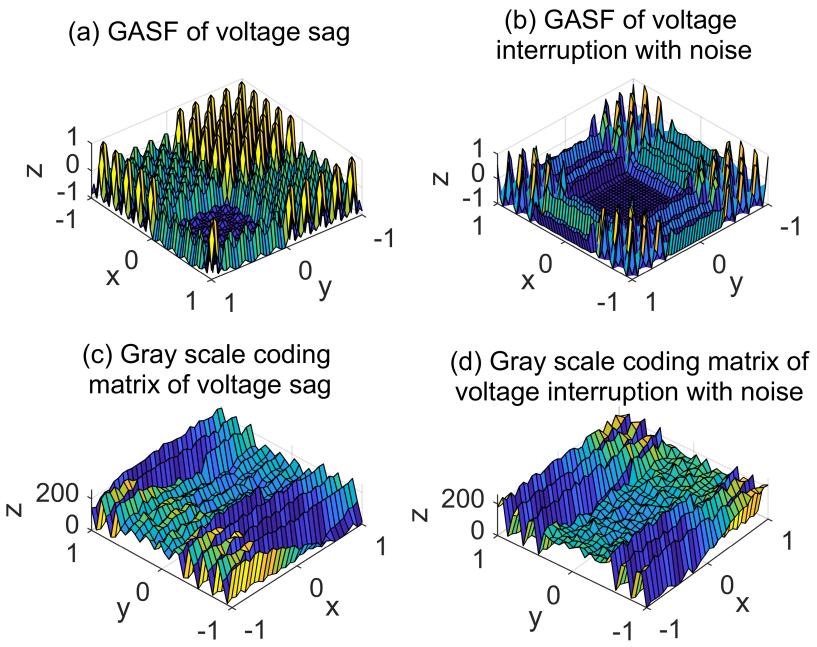


Fig. 10. 3D characteristic diagrams of power quality disturbances under different encoding modes

Figure 10 shows the result of encoding the voltage sag signal and the voltage interruption signal with noise into the Gramian angle sum field and gray scale coding matrix, respectively, and transforming them into a 3D feature graph. It may be seen from Fig. 10 that the gray scale coding matrix of the voltage interruption signal with noise is very similar to that of the voltage sag, which is caused by the insufficient sparsity of the gray scale coding matrix, which makes it impossible to effectively screen out signal features from the noisy signal. However, by comparison, the Gramian angle sum field can distinguish similar disturbance signals well because it retains the characteristic information of the original signal completely. This further verifies the effectiveness of the proposed coding method.

In order to further verify the effectiveness of the proposed method, this paper selects two traditional methods and the classification method using the gray scale images mentioned above, and then carries out comparative simulation combined with the proposed method. The two traditional methods include: the 1D-CNN method and WT-SVM method. The 1D-CNN method is a convolutional neural network-based classifier for one-dimensional disturbance signals of power quality [39], and the WT-SVM method: is a wavelet transform and support vector machine-based classifier for one-dimensional disturbance signals of power quality [40, 41]. The comparison simulation results are shown in Table 5.

As may be seen from Table 5, compared with the traditional method of classifying one-dimensional disturbance signals of power quality directly, the method of converting one-dimensional time-series signals into two-dimensional images and then using a neural network for classification is more effective, which also benefits from the powerful function of the neural network in the

Table 5. Performance comparison of the proposed method with other approaches

Disturbance type	Classification accuracy/%							
	Gray scale picture-ResNet152		The method proposed in this paper		1D-CNN		WT-SVM	
	P	PN2	P	PN2	P	PN2	P	PN2
C0	100	99.6	100	99.9	100	99.7	100	99.2
C1	97.6	95.4	99.8	99.6	97.2	95.0	96.3	93.1
C2	96.2	94.2	99.7	99.6	97.1	95.1	96.4	92.9
C3	98.3	95.3	100	100	97.4	94.2	95.2	92.3
C4	98.7	95.3	100	100	97.1	94.3	95.8	93.4
C5	97.2	93.6	100	100	98.1	93.6	96.8	94.1
C6	95.5	94.1	99.4	98.9	95.3	95.2	94.3	92.4
C7	95.9	95.7	99.8	99.5	95.3	94.7	94.5	92.6
C8	96.5	93.4	99.5	99.0	94.7	94.1	92.3	91.3
C9	96.7	93.1	99.7	99.2	96.2	93.5	93.0	90.5
C10	97.4	94.2	99.6	99.3	96.7	93.3	94.3	92.3
C11	97.8	94.5	100	99.5	96.3	93.1	93.6	92.1
C12	97.5	95.7	100	99.7	96.3	94.8	93.5	92.5
C13	96.5	93.1	99.6	99.2	95.2	94.5	95.3	93.5
C14	96.3	93.2	99.5	99.0	94.4	93.5	94.5	91.5
C15	96.7	93.1	99.2	98.6	95.2	92.7	93.4	91.6
C16	96.8	93.0	99.4	99.0	95.0	93.6	93.1	91.3
Average value	97.2	94.5	99.72	99.41	96.3	94.4	94.8	92.7

P: disturbance signals of power quality without noise

field of machine vision. In addition, similarly to the above analysis, the final classification and recognition effect is not as good as the method proposed in this paper because the characteristics of the original signal cannot be effectively retained after the disturbed signal is encoded as a gray-scale image. These results verify performance of the proposed method.

5.7. Practical analysis

The power quality disturbance data used in the simulation in this paper are obtained by Matlab simulation. Therefore, in order to verify the practicability of the proposed method, the power quality disturbance data obtained from actual measurements in the IEEE PES database [42,43] are

selected to verify the model. Among of these data, some of them are obtained from a measurement for an 8 kV XLPE underground cable, and some of others are one-phase synchronized voltage waveform data measured at the wall outlets (approximately 120 V RMS) in two sites that are 300 km apart. This not only enriches the data source but also further makes the practicality of the method proposed in this paper convincing.

The sampling frequency of the selected data is 256 sampling points in each cycle. In order to keep consistent with the first data, the length of data of each sample is 1 536 sampling points. In addition, it is worth noting that due to the data obtained from actual measurement, there are only 10 types of power quality disturbances, and the sample number of each disturbance is small and different, which is more consistent with the actual situation.

In order to show the superiority of the method proposed in this paper more clearly, the other three methods in Section 5.6 are selected and the actual data mentioned above are used for comparative simulation. Among them, the data used for the neural network method are divided into the training set and the verification set in a ratio of 7:3 for the training and verification of the model. In addition, noise with a SNR of 20 dB is added to the actual data to verify the anti-noise capability of each method. The actual data types and simulation results of power quality disturbance are shown in Table 6.

Table 6. The performance comparison of the proposed method with other approaches for practical signals

Disturbance type	**The total number of samples**	**Classification accuracy/%**							
		Gray scale picture-ResNet152		**The method proposed in this paper**		**1D-CNN**		**WT-SVM**	
		P	PN2	P	PN2	P	PN2	P	PN2
C1	176	90.4	88.2	93.5	92.1	85.7	82.2	82.8	78.8
C2	36	92.0	91.1	96.2	95.8	89.8	84.4	88.5	86.8
C5	95	92.3	90.2	95.3	93.2	90.3	87.6	89.2	87.3
C9	14	88.7	86.3	93.2	91.1	86.5	82.5	82.4	80.1
C10	28	93.5	91.2	96.3	94.1	90.5	84.8	87.1	84.1
C11	12	90.5	89.3	95.1	93.7	90.3	79.8	84.3	78.7
C13	22	92.4	90.5	95.9	93.2	89.4	84.3	87.1	85.9
C14	15	89.0	87.3	92.5	91.3	87.1	81.3	83.9	80.3
C17	13	92.1	90.1	96.1	95.7	90.1	84.3	88.7	86.7
Average value	×	91.2	89.4	94.9	93.4	88.9	83.5	86.0	83.2

C17: Voltage sag and voltage flicker

From the sample distribution of the actual data, voltage sag and harmonics account for a large proportion of the disturbance data, which is also consistent with the actual situation in

the power system. According to the final classification results, the method proposed in this paper can have a good recognition effect even in the case of a small sample number. This is because the multiple transfer learning proposed in this paper enables the model to find the characteristics of the disturbance signal accurately and quickly by pre-learning in the absence of samples, which further verifies the effectiveness of the proposed method. There is a certain gap between the real data and the data obtained by simulation, and it is very costly to label the disturbed data in the actual situation, which also reflects the practicability of the method proposed in this paper.

6. Conclusions

A disturbance classification method for power quality based on the Gramian angle field and multiple transfer learning is proposed in this paper. In the case when a one-dimensional disturbance signal as input cannot give full play to the complete performance of the convolutional neural network, it is proposed to convert a one-dimensional disturbance signal of power quality into a GAF encoded image by using the Gramian angular field. The GAF encoded image can effectively highlight the characteristics of different disturbance types without losing any information of the original signal, and the picture has good sparsity. It is helpful to the classification of convolutional neural networks.

In the case when the classification and recognition effect of a single convolutional neural network is limited, a feature fusion method is proposed. After fusing the features of multiple models, a classifier with a full connection layer is used for classification. This method can effectively improve the classification accuracy and robustness of the model. Considering that the power quality disturbance is greatly affected by noise, in order to further improve the anti-noise performance and generalization of the model, the representative disturbance signals with signal-to-noise ratios of 0d B, 20 dB and 40 dB are used as the input of the feature fusion sub-model, and the features are extracted. At the same time, in order to improve the training speed of the model and further improve the classification accuracy and robustness of the sub-model, a multiple transfer learning method is proposed to transfer the training weights of the sub-model in turn, so that the pre-training weights of the latter model inherit from the training weights of the previous model, and the weight processing method of partial freezing and partial fine-tuning is adopted to ensure the optimal training effect of the proposed model.

Compared with the simulation results of the effectiveness of the Gramian angle field, the effectiveness of the model structure, the anti-noise ability of the model and the universality of the model, it is verified that the method proposed in this paper not only has high classification accuracy, but also has good stability and generalization when dealing with the classification of power quality disturbance signals.

Acknowledgements

This work is supported by Science and technology plan project of Guangzhou Power Supply Bureau of Guangdong Power Grid Co., Ltd. (080037KK52190039/GZHKJXM 20190100).

References

[1] Huang J.M., Qu H.Z., Li X.M., *Classification for hybrid power quality disturbance based on STFT and its spectral kurtosis*, Power System Technology, vol. 40, no. 10, pp. 3184–3191 (2016), DOI: 10.13335/j.1000-3673.pst.2016.10.036.

[2] Yin B.Q., He Y.G., Zhu Y.Q., *Detection and classification of power quality multi-disturbances based on generalized S-transform and fuzzy SOM neural network*, Proceedings of the CSEE, vol. 35, no. 4, pp. 866–872 (2015), DOI: 10.13334/j.0258-8013.pcsee.2015.04.013.

[3] Chandrasekarp K., *Detection and classification of power quality disturbance waveform using MRA based modified wavelet transform and neural networks*, Journal of Electrical Engineering, vol. 61, no. 4, pp. 235–240 (2010), DOI: 10.2478/v10187-010-0033-4.

[4] Li T.Y., Zhao Y., Li N. *et al.*, *A new method for power quality detection based on HHT*, Proceedings of the CSEE, vol. 25, no. 17, pp. 52–56 (2005), DOI: 10.13334/j.0258-8013.pcsee.2005.17.011.

[5] Wei L., Zhang F.S., Geng Z.X. *et al.*, *Detection, location and identification of power quality disturbance based on instantaneous reactive power theory*, Power System Technology, vol. 28, no. 6, pp. 53–58 (2004), DOI: 10.13335/j.1000-3673.pst.2004.06.012.

[6] Yongd D., Bhowmiks M., *An effective power quality classifier using wavelet transform and support vector machines*, Expert Systems with Applications, vol. 15, no. 1, pp. 6075–6081 (2015), DOI: 10.1016/j.eswa.2015.04.002.

[7] Tan H.F., Liu Z.Y., *Multi-label K nearest neighbor algorithm by exploiting label correlation*, Journal of Computer Applications, vol. 35, no. 10, pp. 2761–2765 (2015), DOI: 10.11772/j.issn.1001-9081.2015.10.2761.

[8] Lee I.W.C., Dash P.K., *S-transform-based intelligent system for classification of disturbance signals of power quality*, IEEE Transactions on Industrial Electronics, vol. 50, no. 4, pp. 800–805 (2003), DOI: 10.1109/TIE.2015.2449780.

[9] Dehghanih V., Naghiza D. *et al.*, *Power quality disturbance classification using a statistical and wavelet-based hidden Markov model with Dempster Shafer algorithm*, International Journal of Electrical Power and Energy Systems, vol. 47, no. 5, pp. 368–377 (2013), DOI: 10.1016/j.ijepes.2012.11.005.

[10] Zhao W.J., Shang L.Q., Sun J.F., *Power quality disturbance classification based on time-frequency domain multi-feature and decision tree*, Protection and Control of Modern Power Systems, vol. 4, no. 4, pp. 337–342 (2019), DOI: 10.1186/s41601-019-0139-z.

[11] Yu Z.Y., Zhang W.H., Wang X.K., Huang N.T., Huang X.W., *Feature selection method for power quality disturbance identification based on GA and ELM*, Electrical measurement and instrumentation, vol. 53, no. 23, pp. 62–66 (2016), DOI: 10.3969/j.issn.1001-1390.2016.23.01.

[12] Qu H.Z., Liu H., Li X.M., Huang J.M., *Power quality composite disturbance classification method based on multi label random forest*, Power System Protection and Control, vol. 45, no. 11, pp. 1–7 (2017), DOI: 10.7667/PSPC160899.

[13] Xing J.P., *Research on power quality composite disturbance identification method based on multi feature quantity*, Shandong Electric Power Technology, vol. 44, no. 4, pp. 16–21 (2017).

[14] Qiu X.Y., Li F.L., *Power quality disturbance analysis and identification based on ITD and K-means clustering*, Journal of Power System and Automation, vol. 27, no. 8, pp. 54–59 (2015), DOI: 10.3969/j.issn.1003-8930.2015.08.10.

[15] Hinton G.E., Osindero S., Yee-Whye T., *A Fast Learning Algorithm for Deep Belief Nets*, Neural Computation, vol. 18, no. 7, pp. 1527–1554 (2006), DOI: 10.1162/neco.2006.18.7.1527.

[16] Xiao K., He J.J., Liu C., Chen S.Y., *Electronic circuit classification algorithm based on stack self coding network*, Computer Application Research, vol. 35, no. 9, pp. 2853–2855 (2018), DOI: 10.3969/j.issn.1001-3695.2018.09.069.

[17] Fang R., Meng X., Wang C.Q., Zhong Y.D., *ECG recognition and classification method based on cyclic neural network*, Industrial control computer, vol. 34, no. 7, pp. 122–123 (2021).

[18] Yang W., Yin K.Y., Bao Y.Y., Yin X.G., Xu B., *Power grid fault type identification based on deep confidence network*, Power Engineering Technology, vol. 40, no. 2, pp. 169–177 (2021), DOI: 10.12158/j.2096-3203.2021.02.024.

[19] Wang H., Yang D.S., Zhou B.W., Gao X.T., Pang Y.H., *Fault diagnosis of multi terminal DC transmission line based on parallel convolutional neural network*, Power System Automation, vol. 44, no. 12, pp. 84–92 (2020), DOI: 10.7500/AEPS20191124003.

[20] Jiang H., Zheng Y.H., Wang Z.Z., Chen L., Peng J.C., *Classification of transient power quality disturbances based on digital image processing technology*, Power system protection and control, vol. 43, no. 13, pp. 72–78 (2015), DOI: JournalArticle/5b3bd991c095d70f00940a69.

[21] Chen W., He J.H., Pei X.P., *Classification for power quality disturbance classification based on phase space reconstruction and convolution neural network*, Power System Protection and Control, vol. 46, no. 14, pp. 87–93 (2018), DOI: 10.7667/PSPC171080.

[22] Wang Z., Oates T., *Encoding time series as images for visual inspection and classification using tiled convolutional neural networks*, Workshops at the Twenty-Ninth AAAI Conference on Artificial Intelligence, Austin, Texas, USA, pp. 1–7 (2015).

[23] Wang Z., Oates T., *Imaging time-series to improve classification and imputation*, Twenty-Fourth International Joint Conference on Artificial Intelligence, Michael Wooldridge Buenos Aires, Argentina, pp. 3939–3945 (2015).

[24] Simonyan K., Zisserman A., *Very Deep Convolutional Networks for Large-Scale Image Recognition*, The 3th International Conference on Learning Representations (ICLR 2015), San Diego, CA, USA, pp. 1–9 (2015), DOI: 10.48550/arXiv.1409.1556.

[25] Szegedy C., Liu W., Jia Y. *et al.*, *Going Deeper with Convolutions*, 2015 IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, pp. 1–9 (2015), DOI: 10.1109/CVPR.2015.7298594.

[26] Lecun Y., Bottou L., Bengio Y., Haffner P., *Gradient-based learning applied to document recognition*, Proceedings of the IEEE, vol. 86, no. 11, pp. 2278–2324 (2019), DOI: 10.1109/jproc.2003.813577.

[27] Krizhevsky A., Sutskever I., Hinton G., *ImageNet Classification with Deep Convolutional Neural Networks*, Advances in Neural Information Processing Systems, vol. 25, no. 2, pp. 1–9 (2019), DOI: 10.1016/s0898-1221(96)90463-0.

[28] He K., Zhang X., Ren S. *et al.*, *Deep Residual Learning for Image Recognition*, 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, pp. 770–778 (2016), DOI: 10.1109/CVPR.2016.90.

[29] Tan C., Sun F., Kong T. *et al.*, *A Survey on Deep Transfer Learning*, The 27th International Conference on Artificial Neural Networks, Rhodes, Greece, pp. 270–279 (2018), DOI: 10.48550/arXiv.1808.01974.

[30] Schmidhuber J., *Deep learning in neural networks: an overview*, Neural Networks, vol. 61, pp. 85–117 (2015), DOI: 10.1016/j.neunet.2014.09.003.

[31] Ma P., Fan Y.F., *Power equipment component detection of small sample intelligent substation based on deep transfer learning*, Power Grid Technology, vol. 44, no. 3, pp. 1148–1159 (2020), DOI: 10.13335/j.1000-3673.pst.2018.2793.

[32] He Q., Zhao J.Y., Jiang G.Q., Jia Ch.L., Xie P., *Gearbox fault diagnosis method based on sparse filtering feature fusion of current signal*, Power Grid Technology, vol. 44, no. 5, pp. 1964–1971 (2020), DOI: 10.13335/j.1000-3673.pst.2019.1314.

[33] Lu J.L., Guo L.Y., *Power system transient stability assessment based on improved deep residual shrinkage network*, Journal of Electrotechnics, vol. 36, no. 11, pp. 2233–2244 (2021), DOI: 10.19595/j.cnki.1000-6753.tces.200437.

[34] Wenge C., Guo H., Roehrig C., *Measurement-based harmonic current modeling of mobile storage for power quality study in the distribution system*, Archives of Electrical Engineering, vol. 66, no. 4, pp. 801–814 (2017), DOI: 10.1515/aee-2017-0061.

[35] Khoa N.M., Dai L.V., *Detection and Classification of Power Quality Disturbances in Power System Using Modified-Combination between the Stockwell Transform and Decision Tree Methods*, Energies, vol. 13, no. 14, 3623 (2020), DOI: 10.3390/en13143623.

[36] Velazquez L.M., Troncoso R.J.R., Ruiz G.H., Sotelo D.M., Rios R.A.O., *Smart sensor network for power quality monitoring in electrical installations*, Measurement, vol. 103, pp. 133–142 (2017), DOI: 10.1016/j.measurement.2017.02.032.

[37] Laurens V.D.M., Hinton G., *Visualizing Data using t-SNE*, Journal of Machine Learning Research, vol. 9, pp. 2579−2605 (2008).

[38] Hoang D.T., Kang H.J., *Convolutional Neural Network Based Bearing Fault Diagnosis*, Proceedings of the thirteenth International Conference on Intelligent Computing: Intelligent Computing Theories and Application, Liverpool, UK, pp. 105–111 (2017), DOI: 10.1007/978-3-319-63312-1_9.

[39] Wu A.D., Zhong J.W., Qin M. *et al*., *Research on Power Quality Disturbance Recognition Based on Convolutional Neural Network*, Journal of Hubei Minzu University (Natural Science Edition), vol. 38, no. 3, 318–321 (2020), DOI: 10.13501/j.cnki.42-1908/n.2020.09.015.

[40] Mohammed Y.S., Mahmood T.A., *High impedance fault detection in radial distribution network using discrete wavelet transform technique*, Archives of Electrical Engineering, vol. 70, no. 4, pp. 873–886 (2021), DOI: 10.24425/aee.2021.138267.

[41] Ma J.X., Xu W.N., He F.X. *et al*., *A Novel Power Quality Classification and Identification Method Based on WT and SVM*, Smart Power, vol. 47, no. 3, pp. 16–22 + 37 (2019), DOI: CNKI:SUN:XBDJ.0.2019-03-004.

[42] IEEE Power Engineering Society, *IEEE PES working group P1433 Power Quality Definitions* (2001).

[43] IEEE Power Engineering Society, *IEEE PES working group P3139 Power Quality Data Analytics* (2020).