energies

MDPI

Article

Power Quality Transient Disturbance Diagnosis Based on Dynamic Large Convolution Kernel and Multi-Level Feature Fusion Network

Chen Zheng [1,*], Qionglin Li [1], Shuming Liu [1], Shuangyin Dai [1], Bo Zhang [1] and Yajuan Liu [2]

[1] Electric Power Research Institute, State Grid Henan Electric Power Company, No. 85 Songshan South Road, Erqi District, Zhengzhou 450001, China; 13525571757@139.com (Q.L.); 15286807325@126.com (S.L.); daishuangyin@163.com (S.D.); 13592602687@126.com (B.Z.)

[2] Zhengzhou Power Supply Bureau, State Grid Henan Electric Power Company, Zhengzhou 450001, China; liuyajuan725@163.com

* Correspondence: 20151101009c@alu.cqu.edu.cn

Abstract: Power quality is an important metric for the normal operation of a power system, and the accurate identification of transient signals is of great significance for the improvement of power quality. The diverse types of power system transient signals and strong characteristic coupling brings new challenges to the analysis and identification of power system transient signals. In order to enhance the identification accuracy of transient signals, one method of power system transient signal identification is proposed based on a dynamic large convolution kernel and multilevel feature fusion network. First, the more fine-grained and more informative features of the transient signals are extracted by the dynamic large convolution kernel feature extraction module. Then, the multi-scale local features are adaptively fused by the multilevel feature fusion module. Finally, the fused features are reduced in dimension by the fully connected layer in the classification module and fed into the SoftMax layer for transient signal type detection. The proposed method can effectively improve the small receptive field problem of convolutional neural networks and the lack of ability of Transformer network in extracting local context information. Compared with five other power quality transient disturbance identification models, the experimental results show that the proposed method has better diagnostic accuracy and anti-noise capability.

Keywords: power quality disturbance; transient signal; convolutional neural networks; transformer network; feature fusion strategy

check for updates

Citation: Zheng, C.; Li, Q.; Liu, S.; Dai, S.; Zhang, B.; Liu, Y. Power Quality Transient Disturbance Diagnosis Based on Dynamic Large Convolution Kernel and Multi-Level Feature Fusion Network. *Energies* **2024**, *17*, 3227. https://doi.org/10.3390/en17133227

Academic Editor: Ahmed Abu-Siada

Received: 26 May 2024
Revised: 20 June 2024
Accepted: 27 June 2024
Published: 1 July 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The network structure of power systems is becoming more and more complex, and a variety of non-linear loads are now widely used. This will bring some transient disturbance signals to the power grid, which seriously affects power quality [1,2]. At the same time, the presence of power quality disturbances (PQDs) reduces the lifetime of power semiconductor and solid-state switching devices. In order to ensure the normal operation of various electrical equipment, it is necessary to take appropriate action to improve the quality of electricity.

The effective detection of PQD types is a precondition for the dynamic compensation of power quality. Research on PQD detection mainly focuses on the identification of transient signal types. Transient signals have the characteristics of being non-stationary and having a short duration. Moreover, due to the combined effect of the diversity of disturbance sources, the complexity of the grid structure and other factors, transient signal disturbances in the power grid are not always a single basic disturbance, and most of them are composite PQDs mixed by a variety of basic disturbances. These single disturbances have different disturbance types, different disturbance energies, and different starting and stopping moments. These bring greater challenges to the accurate identification of transient disturbances.

The problem of transient signal PQDs in power systems has become the focus of research among many scholars [3]. Traditional transient disturbance analysis methods mainly rely on signal processing techniques to extract the physical characteristics of transient disturbance signals, and then achieve the classification of disturbance types by classifiers. This approach requires an understanding of power systems and a specific analysis of the equipment characteristics and operating conditions. Such signal processing methods include short time Fourier transform (STFT), wavelet transform (WT), S-transform, empirical modal decomposition (EMD), and variational modal decomposition (VMD).

Today, traditional signal processing methods are well studied in the field of extracting transient disturbance features. Carvalho et al. [4] obtained the time-domain maximum magnitude vector of the disturbance signal by Blackman window STFT, and then input it as a feature vector into the support vector machine to achieve the identification of PQD types. While STFT has a fixed analysis window size, it is not suitable for analyzing the multi-scale characteristics of the signal. WT is suitable for analyzing and identifying non-smooth signals, but it is difficult to choose the wavelet basis function for WT; in addition, it causes some signal distortion when the number of decomposition layers is large in [5]. S-transform is a time-frequency analysis algorithm derived from the combination of STFT and WT. It has good time-frequency characteristics, but the disturbed signal is highly informative after the decomposition of S-transform, and there is no unified form of disturbed eigenvector in [6]. EMD has been widely used in the field of identification of transient signals. Due to the lack of rigorous mathematical proofs for EMD, it leads to the need for strong empirical guidance in its practical application in [7]. VMD is applied as an adaptive and completely non-recursive signal decomposition algorithm for the type detection of transient signals in [8]. However, an inaccurate number of decompositions can easily lead to high noise levels in the signal components, which affects the effectiveness of signal detection. The above methods can effectively identify and classify a single disturbed signal. Due to the feature coupling between the compound transient disturbed signals, it is difficult for traditional signal processing techniques to extract distinguishable features. Therefore, such methods are difficult to cope with composite disturbances in power systems.

The diagnosis of power quality transient disturbances based on artificial intelligence technology is the focus of many scholars' research. This approach does not require an in-depth understanding of the physical operating processes of the power system, but relies on data analysis and processing to identify the characteristics and patterns of PQDs. Shoryu et al. [9] proposed a PQD identification model based on a convolutional network and long short-term memory. The model can efficiently extract salient features from noisy signals and classify them through a SoftMax layer. Hezuo et al. [10] used convolutional neural networks to adaptively learn disturbance features to classify unknown instances. At first, one-dimensional power quality waves are mapped into two-dimensional gray-scale images and disturbance features in the form of quadratic Dgray scale; then, a CNN architecture is constructed for PQD classification based on a LeNet-5 network under different noise conditions. Compared with traditional signal processing techniques, the above methods have achieved some satisfactory results based on the powerful feature extraction capability of CNN. However, when a CNN performs feature extraction, the receptive field of the CNN is limited by the smaller convolutional kernel, which causes it to focus mainly on local regions in the input data, ignoring the global space of the input data.

Inspired by the significant success of the Transformer architecture [11] in the field of natural language processing (NLP), the stacked multi-head attention mechanism captures global features among sequential data, which has great potential for intelligent fault diagnosis. Liang et al. [12] used the STFT to transform a one-dimensional signal into a two-dimensional time-frequency image, and then the time-frequency map was input as a feature map into a ViT network to train for achieving the classification task. Although the Transformer network has better performance in image classification tasks [13], the Transformer lacks some of the inductive biases inherent in CNNs, such as local connectivity, parameter sharing, etc. [14], which makes it possibly unable to extract meaningful features

from a dataset when the amount of training data is insufficient. On the other hand, stacked attention mechanisms often limit the ability of Transformer networks to efficiently extract local contextual information.

In this paper, a transient signal identification method is proposed for power systems based on a dynamic large convolution kernel and multi-level feature fusion network by integrating the advantages of a CNN to extract local features and the Transformer network to capture global features. First, to solve the problem whereby a CNN small receptive field cannot effectively capture the important local features in the time series data, a dynamic large convolution kernel module is designed, which solves the small receptive field problem caused by the small convolution kernel. Meanwhile, unlike the fixed large convolution kernel, the module can make more efficient use of global context information by dynamically adjusting the size of the large convolution kernel to suit features of different sizes. Then, in order to solve the problem of the Transformer network's lack of ability in extracting local contextual information, a dynamic large convolution kernel module is used to replace the multi-head attention mechanism in the transformer network. Finally, a dynamic feature fusion module is designed, which can dynamically find important feature information for adaptive fusion of multi-scale features. Our contributions are presented as follows:

1. In order to increase the receptive field of the convolutional neural networks and enhance the feature extraction capability of the network, a dynamic large convolution kernel module is proposed, which can make more efficient use of global context information by dynamically adjusting the size of the large convolution kernel to adapt to different sizes of features.

2. A multilevel feature fusion module is proposed, which improves on the popular fixed-weight feature fusion strategy by adaptively assigning different fusion weights to different input sequence data.

3. To validate the network framework, experiments were conducted on single and composite PQD datasets under different noise conditions.

The rest of the paper is structured as follows. The proposed dynamic large convolution kernel structure and the multilevel feature fusion module are presented in Section 2. The PQD diagnosis network framework is described in Section 3. Section 4 illustrates the validity of the proposed method through extensive experiments. Finally, Section 5 concludes this paper.

2. Theoretical Background

2.1. Dynamic Large Convolution Kernel Module

To expand the receptive field of CNN, larger convolution kernels were introduced and integrated into the CNN architecture. Ding et al. [15]. proposed a new network architecture called RepLKNet. This network architecture makes some changes on the Swin Transformer [16] architecture by replacing the multi-head self-attention with large deep convolution. Experiments demonstrate that the large convolution kernel is designed to perform well on classification tasks. Large convolution kernel structures have a larger receptive field and are able to capture more discriminative features compared to small convolution kernel structures. However, these networks mainly utilize large kernels of fixed size for feature extraction, which remains slightly difficult for extracting fault features of different sizes. The difficulties include the following:

- Inappropriate large convolution kernel positions may decrease the performance of the network;
- It is difficult to determine the size of the large convolution kernels when the network achieves optimal performance;
- Large convolution kernels will increase the number of parameters and the computational cost in the network.

In order to determine the optimal location and size of the large convolution kernel, a dynamic large convolution kernel (DLCK) structure is proposed to maintain the diagnostic

performance and reduce the number of parameters, as shown in Figure 1. Unlike the parallel aggregation of convolution kernels [17], the dynamic large convolution kernel (DLCK) structure sequentially aggregates multiple large kernels to expand the receptive field.

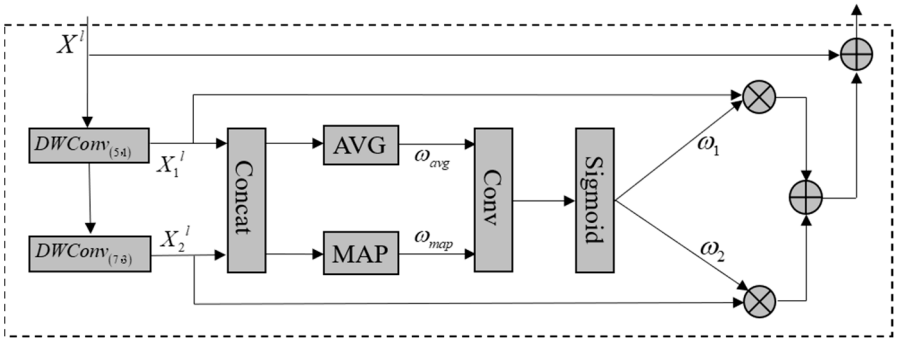


Figure 1. The DLCK structure.

Specifically, the depthwise separable convolution (DWConv) layer uses two large convolution kernels $DWConv_{(5,1)}$ and $DWConv_{(7,3)}$, respectively. $DWConv_{(5,1)}$ denotes a 5×1 kernel with an expansion rate of 1, and $DWConv_{(7,3)}$ denotes a 7×1 kernel with an expansion rate of 3. $DWConv_{(5,1)}$ and $DWConv_{(7,3)}$ are used to convolute the input features of the previous layer.

$$X_1^l = DWConv_{(5,1)}\left(X^l\right) \tag{1}$$

$$X_2^l = DWConv_{(7,3)}\left(X_1^l\right) \tag{2}$$

By cascading these large convolution kernels, the DLCK has the same receptive field as the kernel size 23×1. The global spatial relationships of these local features are then efficiently modeled by fusing the cascaded features $\left[X_1^l; X_2^l\right]$ through the concat operation and by applying average pooling (AVP) and maximum pooling (MAP).

$$\omega_{\text{avg}} = AVG\left(\left[X_1^l; X_2^l\right]\right) \tag{3}$$

$$\omega_{\text{map}} = MAP\left(\left[X_1^l; X_2^l\right]\right) \tag{4}$$

Then, a convolutional layer ($Conv$) with kernel size of 7 is used to interact this information $\left[\omega_{\text{avg}}\,; \omega_{\text{map}}\,\right]$, and a Sigmoid activation function is used to obtain dynamic weight values $[\omega_1; \omega_2]$.

$$[\omega_1; \omega_2] = sigmoid\left(Conv\left(\left[\omega_{\text{avg}}\,; \omega_{\text{map}}\,\right]\right)\right) \tag{5}$$

The features of the different large kernels are adaptively selected by utilizing these weight values for alignment. Finally, a residual join is applied.

$$X^l = \left(\left(\omega_1 \otimes X_1^l\right) \oplus \left(\omega_2 \otimes X_2^l\right)\right) + X^l \tag{6}$$

The DLCK module is constructed by integrating DLCK between two linear layers and adding the GELU activation function, as shown in Figure 2. Residual connections are also applied. Therefore, the output of the l layer in the DLCK module can be calculated as:

$$\hat{X}^l = Conv\left(DLK\left(GELU\left(Conv\left(X^l\right)\right)\right)\right) + X^l \tag{7}$$

In order to handle features of different scales, we construct DLCK block by replacing the multi-head self-attention in hierarchical ViT networks with the proposed DLCK modules. The constructed DLCK block consists of three parts, namely, the DLCK module, the MLP module, and the normalization layer, as shown in Figure 3. DLCK blocks are

constructed similarly to ViT blocks. As a result, two sequential DLCK blocks in Layer l and Layer $l-1$ can be computed as:

$$\hat{X}^l = DLK\left(LN\left(X^{l-1}\right)\right) + X^{l-1} \tag{8}$$

$$X^l = MLP(LN(\hat{X}^l)) + \hat{X}^l \tag{9}$$

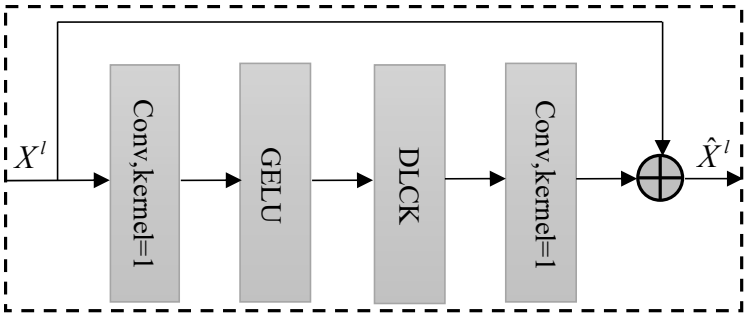


Figure 2. The DLCK Module.

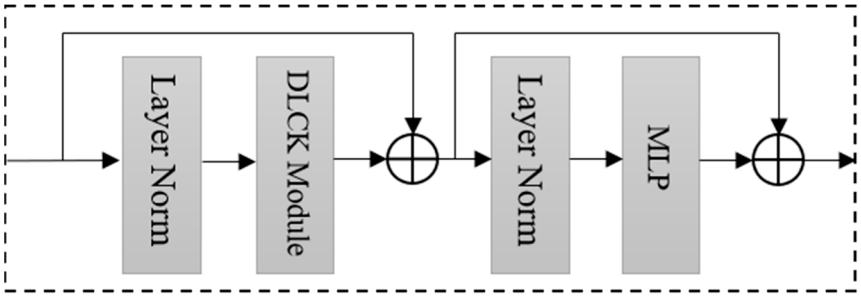


Figure 3. The DLCK block.

2.2. Multilevel Feature Fusion Module

Based on the adaptive fusion of multi-scale local features with global information, we propose a multilevel feature fusion (MLFF) module, as shown in Figure 4. Most of the current popular feature fusion methods use a fixed-weight fusion strategy, which enforces different contexts in a time series to share the same weights, making the common weights for all time series ignore their differences in different local contextual information [18]. To address this problem, we propose a new dynamic feature fusion strategy that adaptively assigns different fusion weights to different input sequence data and locations by a weights learner. The dynamic feature fusion strategy takes into account the differences between feature maps of different locations and help to classify the task more accurately.

Specifically, feature maps F1 and F2 are connected along the channels. To ensure that subsequent blocks can utilize the fused features, the number of channels is reduced to the original number by a channel reduction mechanism. Channel reduction in MLFF is not simply a 1×1 convolution but is guided by the global channel information weight ω_{ch}. Information is extracted to describe the significance of features by cascaded average pooling ($AVGPool$), convolutional layers ($Conv$), and Sigmoid activation.

Fusion features are calibrated by global channel information weights ω_{ch}. At the same time, the use of ω_{ch} helps the convolutional layer retain important features and discard features with less information. Subsequently, the 1×1 convolutional layer ($Conv$) was used to select the appropriate feature maps based on their importance.

$$\omega_{ch} = sigmoid\left(Conv\left(AVGPool\left(\left[F_1^l; F_2^l\right]\right)\right)\right) \tag{10}$$

$$\mathrm{F}^l = Conv\left(\omega_{ch} \otimes \left[F_1^l; F_2^l\right]\right) \tag{11}$$

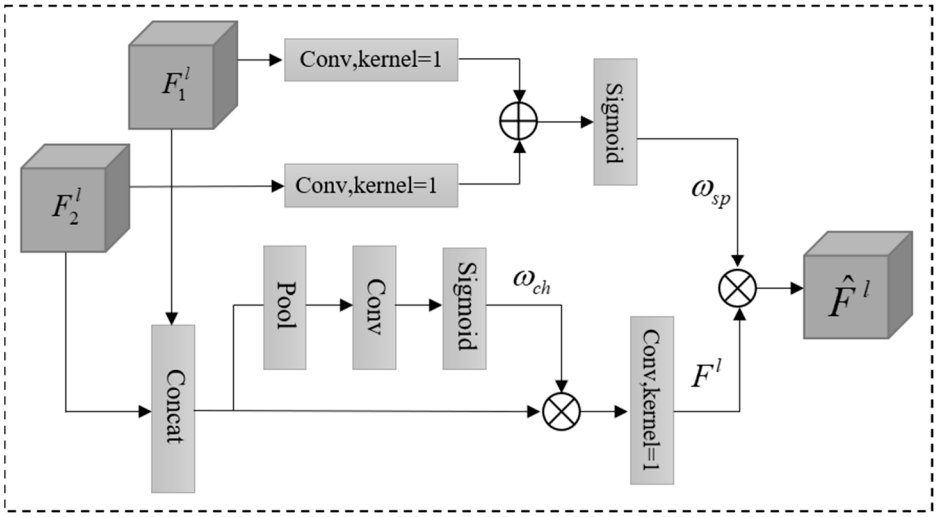


Figure 4. The MLFF module.

To highlight the spatial regions of importance, the global spatial information ω_{sp} is captured by a 1×1 convolutional layer (*Conv*) and Sigmoid activation functions derived from feature maps F1 and F2.

$$\omega_{sp} = sigmoid\left(Conv\left(F_1^l\right) \oplus Conv\left(F_2^l\right)\right) \tag{12}$$

2.3. Classification Module

The classification module is shown in Figure 5. After the multilevel feature fusion layer, the fused features are input to the full connectivity layer for dimensionality reduction, then into the SoftMax classifier for the detection and identification of transient signal faults of power systems [19].

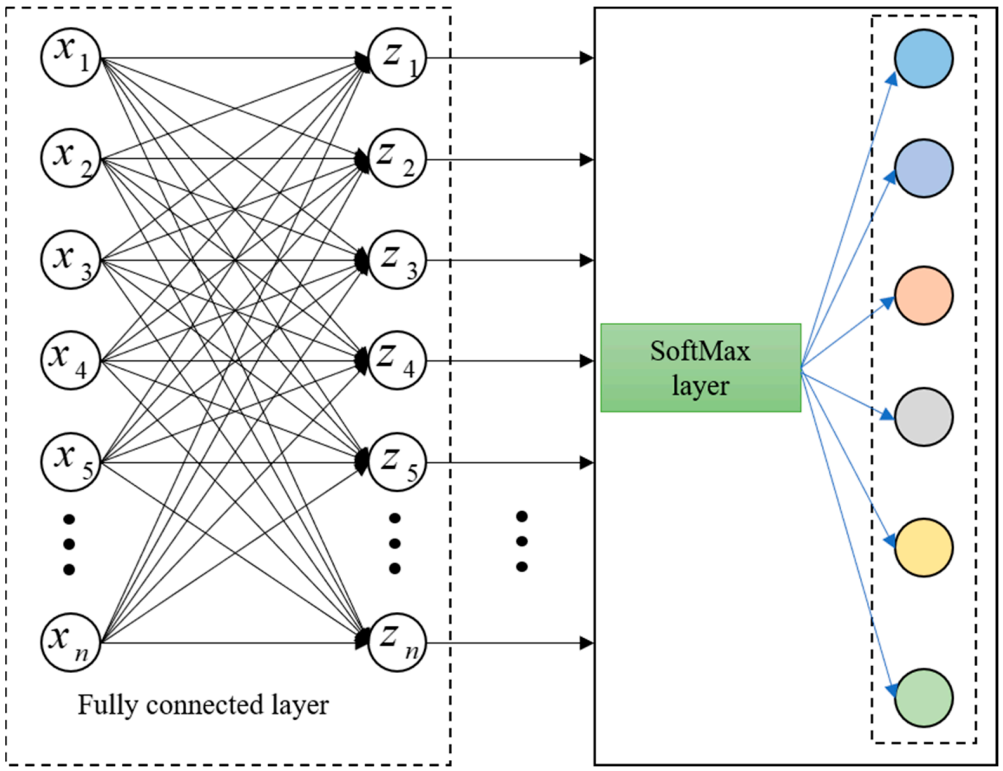


Figure 5. The classification module.

To measure the difference between the predicted and actual results of the model, the cross entropy loss function is used as a metric [20].

$$loss = -\sum_{i=1}^{n} y_i \log(\hat{y}_i) \tag{13}$$

where n denotes the number of fault categories, y_i and $\hat{y}_i$ denote the actual and predicted probability of category i, respectively.

3. The Proposed Network

By using the designed DLCK module, MLFF module and classification module, a framework is proposed for power system transient signal diagnosis and classification in this section, as shown in Figure 6. The framework comprises four hierarchical structures cascaded together, each hierarchy containing three parts: the downsampling layer, the DLCK block, and the MLFF layer.

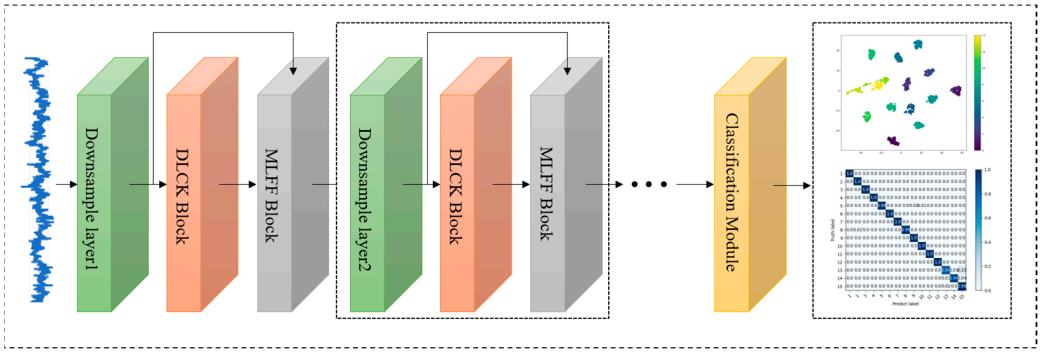


Figure 6. The proposed network.

The downsampling layer is achieved by using convolutional layers with different sized convolution kernels. The downsampling layer facilitates a reduction in the amount of computation and the extraction of the main feature information. The DLCK block cascades the growing kernel sizes and increasing expansion rates to increase the sensory field while allowing dynamic adjustment of the convolutional kernel size, which facilitates the capture of finer and more informative features. The MLFF layer can adaptively assign different fusion weights to different input sequence data through the weights learner. A more accurate classification task is achieved by effectively fusing the local features extracted by the downsampling layer and the global features extracted by the DLCK block. Specific parameter information is listed in Table 1. Finally, the classification module is used to achieve transient signal classification.

Table 1. Specific parameter information of the proposed network.

Structure	Layers	Parameters	Output Size
Input	Input	/	1024 × 1 × 1
Downsample layer1	Convolutional layer	kernel: 7, stride: 2	512 × 64 × 1
DLCK Block	/	/	512 × 64 × 1
MLFF Block	/	/	512 × 64 × 1
Downsample layer2	Convolutional layer	kernel: 2, stride: 2	256 × 128 × 1
DLCK Block	/	/	256 × 128 × 1
MLFF Block	/	/	256 × 128 × 1
Downsample layer3	Convolutional layer	kernel: 2, stride: 2	128 × 256 × 1
DLCK Block	/	/	128 × 256 × 1
MLFF Block	/	/	128 × 256 × 1
Downsample layer4	Convolutional layer	kernel: 2, stride: 2	64 × 512 × 1
DLCK Block	/	/	64 × 512 × 1

Table 1. *Cont.*

Structure	Layers	Parameters	Output Size
MLFF Block	/	/	$64 \times 512 \times 1$
Max Pool	Max Pool	kernel: 64, stride: 1	512×1
FClayer1	Full Connected layer	/	256×1
FClayer2	Full Connected layer	/	Class number

4. Experimental Validation and Analysis

4.1. Experimental Setup and Evaluation Metrics

In all the experiments, the corresponding parameters involved in the proposed model are set as follows: adam optimizer is utilized; initial learning rate is set to 0.001 for stepplr strategy update; batch size is set to 64; maximum epoch is set to 60. The network model is run using PyTorch1.13 and on an Intel i7 and 16 GB GeForce GTX3060 GPU on a PC. In order to qualitatively evaluate and compare the performance of the proposed model with other network methods, a quantitative metric called accuracy was used in the experiments, using classification accuracy [21] as an evaluation metric with the expression:

$$Acc = \frac{TP + TN}{TP + FP + TN + FN} \tag{14}$$

where TP denotes the number of samples that were predicted to be positive and were actually positive, TN denotes the number of samples that were predicted to be negative and were actually negative, FP denotes the number of samples that were predicted to be positive but were actually negative, and FN denotes the number of samples that were predicted to be negative but were actually positive.

4.2. Dataset Description and Preprocessing

Due to the rich types of power quality transient disturbance signals and the random change in disturbance time and location, it is difficult to obtain a large number of actual power quality transient disturbance data. Therefore, in this paper, 15 kinds of transient disturbance mathematical models are simulated by MATLAB 2021a according to the reference [22], which are 7 kinds of single disturbances and 8 kinds of hybrid disturbances. The PQD signal includes harmonics voltage (S1), sag (S2), swell (S3), interruption (S4), flicker (S5), transient oscillation (S6), transient pulse (S7), sag + harmonics (S8), swell + harmonics (S9), interruption + harmonics (S10), flicker + harmonics (S11), sag + transient oscillation (S12), swell + transient oscillation (S13), sag + transient pulse (S14), and swell + transient pulse (S15). The standard parameters of the signal model for the hybrid disturbance voltages are shown in Table 2, and the waveforms of the 15 transient signals simulated based on Table 2 are shown in Figure 7.

In order to eliminate the disturbance of abnormal data on the experimental results, this paper performs the normalization operation on all the data so that the data values are mapped in the range [0, 1], and the normalization formula is as follows:

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \tag{15}$$

where x' denotes normalized data.

Each transient signal generates 500 samples by overlapping samples, so the total number of samples of 15 transient signals is 7500, and the sample data are divided into training set, validation set, and test set according to the ratio of 7:2:1, and then inputted into the model for training.

Table 2. The standard parameters of the simulated signals model.

Signal Type	Signal Models
S1	$V(t) = \sin(\omega t) + a_3 \sin(3\omega t + \varphi_3) + a_5 \sin(5\omega t + \varphi_5) + a_7 \sin(7\omega t + \varphi_7)$
S2	$V(t) = (1 - a(u(t - t_1) - u(t - t_2))) \sin(\omega t)$
S3	$V(t) = (1 + a(u(t - t_1) - u(t - t_2))) \sin(\omega t)$
S4	$V(t) = (1 - a_1(u(t - t_1) - u(t - t_2))) \sin(\omega t)$
S5	$V(t) = \left(1 + a_f \sin(\beta \omega t)\right) \sin(\omega t)$
S6	$V(t) = \sin(\omega t) + a_2 e^{\frac{t-t_3}{\tau}} \sin\{\omega_n(t - t_3)\} \cdot \{u(t - t_3) - u(t - t_4)\}$
S7	$V(t) = \sin(\omega t) + a_4 e^{\frac{t-t_3}{\tau}} \{u(t - t_3) - u(t - t_4)\}$
S8	$V(t) =$ $(1 - a(u(t - t_1) - u(t - t_2))) \sin(\omega t) + \sin(\omega t) + a_3 \sin(3\omega t + \varphi_3) + a_5 \sin(5\omega t + \varphi_5) + a_7 \sin(7\omega t + \varphi_7)$
S9	$V(t) =$ $(1 + a(u(t - t_1) - u(t - t_2))) \sin(\omega t) + \sin(\omega t) + a_3 \sin(3\omega t + \varphi_3) + a_5 \sin(5\omega t + \varphi_5) + a_7 \sin(7\omega t + \varphi_7)$
S10	$V(t) =$ $(1 - a_1(u(t - t_1) - u(t - t_2))) \sin(\omega t) + \sin(\omega t) + a_3 \sin(3\omega t + \varphi_3) + a_5 \sin(5\omega t + \varphi_5) + a_7 \sin(7\omega t + \varphi_7)$
S11	$V(t) = \left(1 + a_f \sin(\beta \omega t)\right) \sin(\omega t) + \sin(\omega t) + a_3 \sin(3\omega t + \varphi_3) + a_5 \sin(5\omega t + \varphi_5) + a_7 \sin(7\omega t + \varphi_7)$
S12	$V(t) = (1 - a(u(t - t_1) - u(t - t_2))) \sin(\omega t) + \sin(\omega t) + a_2 e^{\frac{t-t_3}{\tau}} \sin\{\omega_n(t - t_3)\} \cdot \{u(t - t_3) - u(t - t_4)\}$
S13	$V(t) = (1 + a(u(t - t_1) - u(t - t_2))) \sin(\omega t) + \sin(\omega t) + a_2 e^{\frac{t-t_3}{\tau}} \sin\{\omega_n(t - t_3)\} \cdot \{u(t - t_3) - u(t - t_4)\}$
S14	$V(t) = (1 - a(u(t - t_1) - u(t - t_2))) \sin(\omega t) + \sin(\omega t) + a_4 e^{\frac{t-t_3}{\tau}} \{u(t - t_3) - u(t - t_4)\}$
S15	$V(t) = (1 + a(u(t - t_1) - u(t - t_2))) \sin(\omega t) + \sin(\omega t) + a_4 e^{\frac{t-t_3}{\tau}} \{u(t - t_3) - u(t - t_4)\}$
Parameters	$a = 0.1 \sim 0.9, a_1 = 0.9 \sim 1, a_2 = 0.1 \sim 0.8, a_3 = a_5 = a_7 = 0 \sim 0.15, a_4 = 1 \sim 10, \varphi_3 = \varphi_5 = \varphi_7 = 0 \sim 2\pi,$ $a_f = 0.3 \sim 0.5, t_2 - t_1 = 4T \sim 9T, t_4 - t_3 = 0.05T \sim 3T, \beta = 0.1 \sim 0.4, \tau = 0.008 \sim 0.04, f_n = 300 \sim 900Hz,$

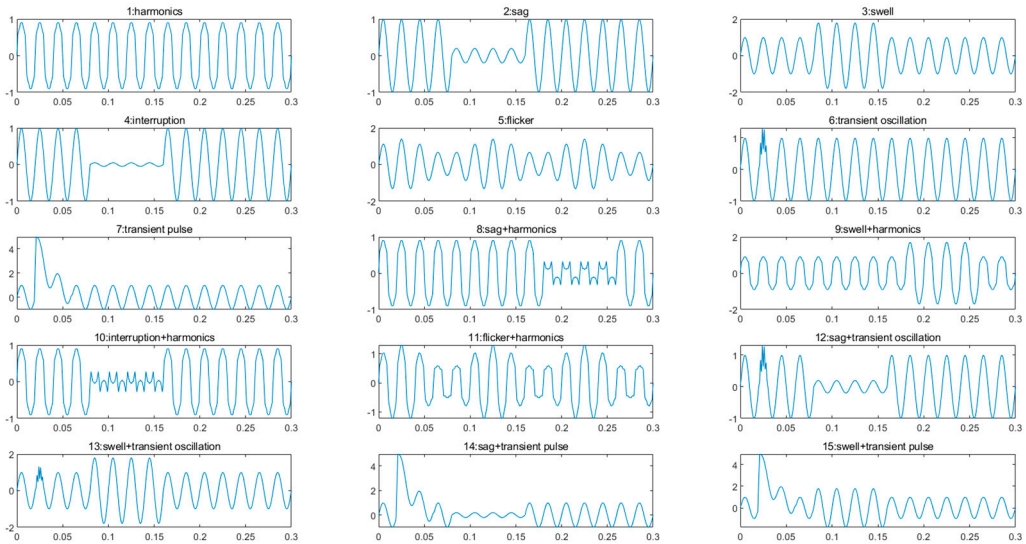


Figure 7. The waveforms of the 15 transient signals simulated.

4.3. Comparative Experimental Results

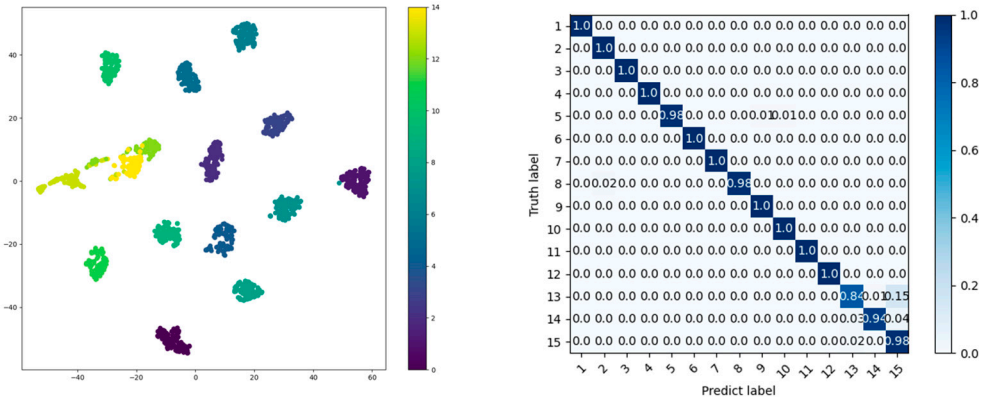
4.3.1. Experimental Results with Noise Data

To ensure that these simulated PQD signals are very close to the real data, white Gaussian noise (WGN) is added to these simulated signals, and the noise simulation environments with an SNR of 5 dB, 15 dB, 25 dB, 35 dB, and 45 dB are constructed. A total of five groups of signals with noise are identified. For these 15 disturbance signals, we

conducted validation experiments. In order to more clearly show the identification accuracy of the proposed model for single disturbance and complex disturbance, we divided the single disturbance and complex disturbance into two categories to show the diagnostic effect respectively. The identification accuracy of various disturbance signals is shown in Table 3. At the SNR of 5 dB, the effect of disturbance signal classification is visualized, as shown in Figure 8.

Table 3. Classification accuracy under different noise conditions (%).

Signal Type	5 dB	15 dB	25 dB	35 dB	45 dB
Single disturbance	99.71	99.80	100	100	100
Compound disturbance	96.75	97.85	99.31	100	100



(a) Visualization of feature classification. (b) Confusion matrix.

Figure 8. Confusion matrix and feature classification visualization at SNR = 5 dB.

As can be seen from Table 3, the proposed model has high recognition accuracy under different SNR values, and the disturbance recognition average accuracy is higher than 97%, indicating that proposed model has good generalization ability. For a single disturbance, when the SNR is greater than 25 dB, the accuracy is as high as 100%. For mixed disturbance, when the SNR is greater than 35 dB, the accuracy rate is as high as 100%, indicating that the model has high anti-noise ability to single disturbance and mixed disturbance.

4.3.2. Experimental Results with Different Models

In order to further verify the superiority of the proposed method, we compare and analyze the proposed method with the existing feature extraction models. The five methods include CNN with small convolution kernels [23], Transformer network [24], CNN–LSTM network [25], CNN–Transformer network [26], and CNN–GRU network [27]. When SNR = 5 dB, the comparison results of each model are as shown in Table 4. Radar charts provide a more visual representation of the performance, as shown in Figure 9.

Table 4. Comparison of different feature extraction methods (%).

Network	5 dB	15 dB	25 dB	35 dB	45 dB	Average
CNN [23]	95.45	95.52	96.82	97.34	97.37	96.50
Transformer [24]	95.57	96.07	96.34	97.86	98.04	96.78
CNN–LSTM [25]	96.61	96.81	97.43	98.41	99.82	97.82
CNN–Transformer [26]	96.27	97.51	97.66	99.32	99.67	98.09
CNN–GRU [27]	97.54	97.18	98.35	99.56	99.83	98.49
OUR	98.13	98.76	99.63	100	100	99.30

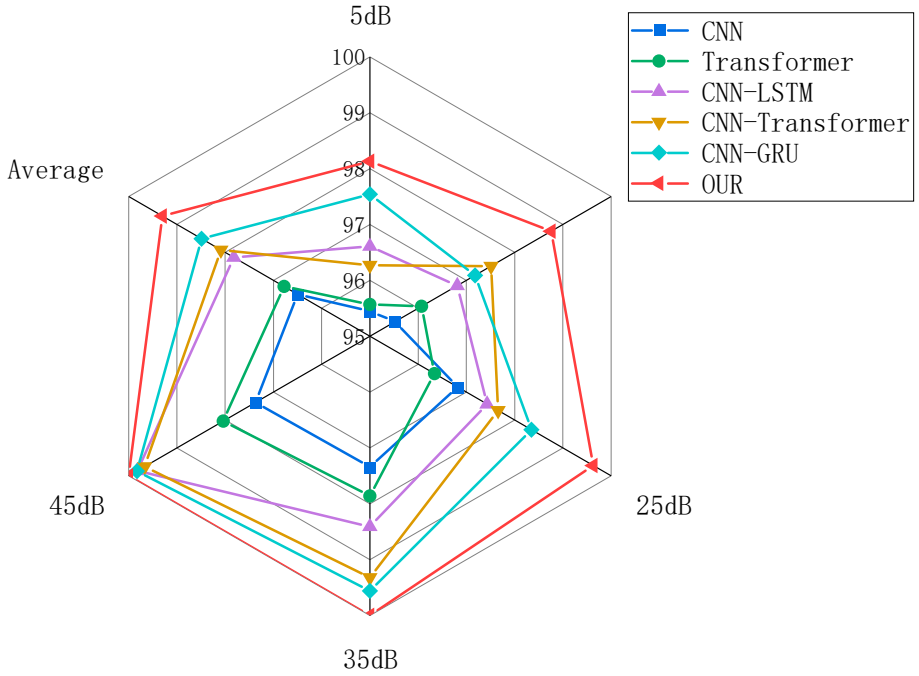


Figure 9. Radar charts displaying the performance of each model.

From the data in Table 4, it can be seen that the CNN with a small convolution kernel [24] has a better feature extraction ability under weak noise conditions; when the noise disturbance is larger, the extracted features are less distinguishable. Compared with the small convolution kernel, the Transformer network [24] with a stacked multi-head attention mechanism can capture global features among the input sequence data, which has great potential in intelligent fault diagnosis. However, it lacks local connectivity and parameter sharing, and is highly dependent on the amount of training data. The average accuracy of its disturbance identification reaches 96.78% under different SNR values. The CNN–LSTM network [25] can achieve multi-level feature extraction, which is helpful to improve the model's expressiveness and generalization ability. However, the CNN–LSTM model has more parameters and the training time may be longer. The average accuracy of its disturbance identification reaches 97.82% under different SNR values. The CNN–Transformer network [26] is achieved by combining the local connectivity of CNN with the global connectivity of the Transformer network. Since it only utilizes a simple concat operation of local and global features to achieve feature fusion, it can only alleviate the feature extraction defects to a certain extent. Compared with the single CNN and Transformer networks, the CNN–Transformer network achieves a higher recognition accuracy of 98.09%. Comparing the CNN–GRU network [27] with the CNN–LSTM network, the gating structure of GRU is relatively simple, and the model interpretation may be slightly higher. Compared with the CNN–LSTM and CNN–Transformer networks, the CNN–GRU network [27] is relatively efficient and has a stronger sequence modeling capability, but the training cost may be higher. The average accuracy of its disturbance identification reaches 98.49% under different SNR values. The proposed network uses a dynamic large convolution kernel instead of a multi-head attention mechanism, which can better meet the needs of different types of features. The use of the dynamic feature fusion module instead of a normal connection fusion operation helps toward fusing better discriminative features. Thus, the proposed network has the highest average accuracy of 99.30% for disturbance recognition.

4.4. Ablation Studies

To validate the reasonability of corresponding modules used in the proposed network, some ablation studies are conducted in this section.

4.4.1. Without DLCK or MLFF

To verify the ability of the proposed model in extracting features, three sets of experiments (whether or not DLCK and MLFF are used) are conducted. For convenience, the network without DLCK is denoted as W/O–DLCK and the network without DFF is denoted as W/O–MLFF. The corresponding experiment results are shown in Table 5.

Table 5. Performance of each module (%).

Network	5 dB	15 dB	25 dB	35 dB	45 dB
W/O–MLFF	96.45	96.57	97.03	98.11	98.50
W/O–DLCK	97.58	97.91	98.45	98.71	99.84
OUR	98.13	99.57	100	100	100

As seen in Table 5, the proposed network has the highest average accuracy, which indicates that the DLCK block and MLFF module have good feature extraction capabilities. One can conclude that the designed modules (DLCK and MLFF) are reasonable and effective.

4.4.2. Network Depth

To further explore the optimal model depth. Six depths are chosen to determine the optimal network depth in this experiment, where depth-1 indicates that the number of hierarchical structures cascaded together is 1. The corresponding results are shown in Figure 10.

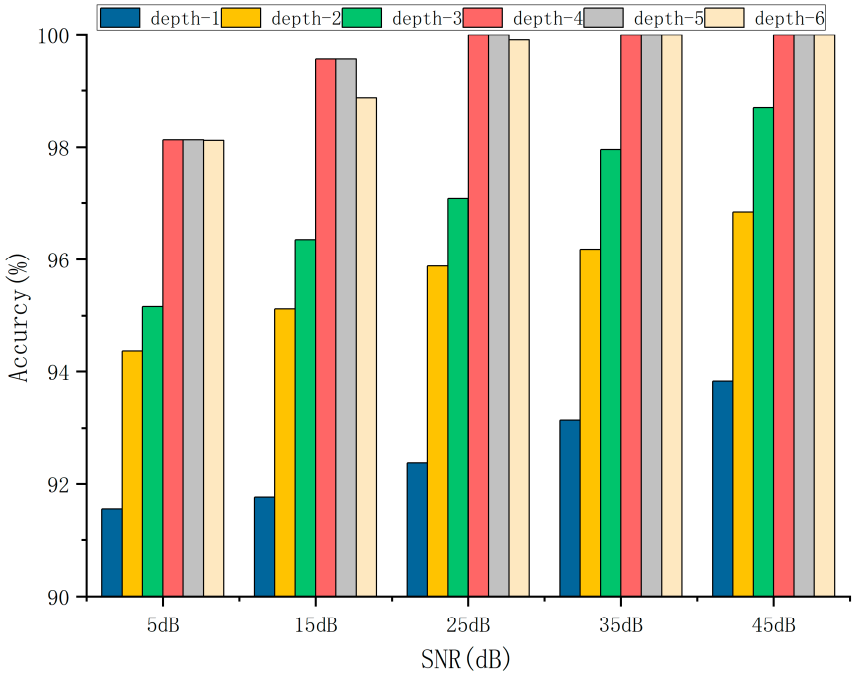


Figure 10. Performance of network depth.

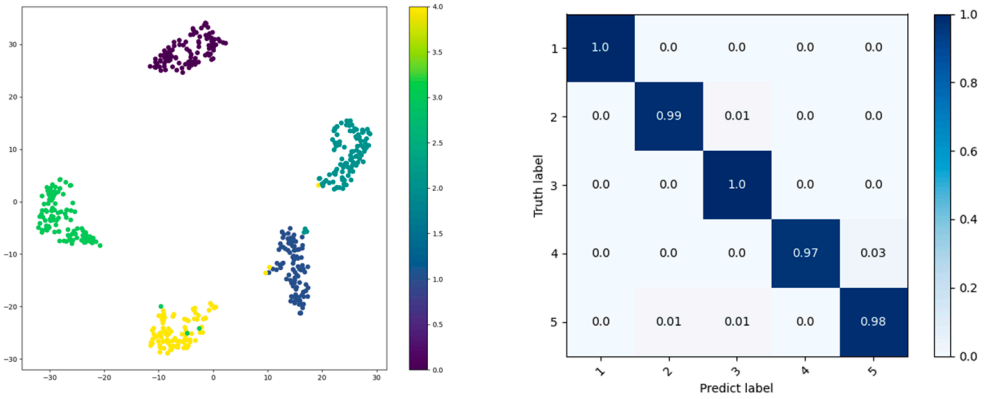
From Table 6, it can be seen that the network achieves the best results when the network depth is set to 4 in the vast majority. In theory, the deeper the network depth is, the better will be the results achieved in feature extraction. However, this is not always true when the number of samples is not sufficient. In fact, it is a difficult task to collect enough samples in practical engineering. This is the reason that the proposed network obtains the best results at the depth of 4. Moreover, with the increase in network depth, the computational cost will increase. Thus, the depth is set to 4 in the proposed network.

Table 6. Performance under different network depths (%).

Depth	5 dB	15 dB	25 dB	35 dB	45 dB
depth-1	91.56	91.77	92.37	93.14	93.83
depth-2	94.36	95.12	95.88	96.18	96.84
depth-3	95.16	96.35	97.08	97.96	98.70
depth-4	98.13	99.57	100	100	100
depth-5	98.57	99.14	99.81	100	100
depth-6	98.12	98.87	99.91	100	100

4.5. Real Data Validation

To further validate the feasibility of the method, this paper validates the method using a set of real data as input. The dataset is provided by the Kaggle public database and consists of five major types of power quality signals, namely, harmonics (S1), sags (S2), transient impulses (S7), sags + harmonics (S8), and sags + oscillations (S12). There are 600 samples for each signal type. The confusion matrix of the classification results is shown in Figure 11.



(a) Visualization of feature classification. (b) Confusion matrix.

Figure 11. Confusion matrix and feature classification visualization of real data.

As can be seen from the confusion matrix, there is still category confusion between S2, S8, and S12. This is because both S8 and S12 contain S2 components. As can be seen from the classification feature visualization graph, the proposed method is still useful for real data, further illustrating the effectiveness of the proposed method.

5. Conclusions

In this paper, a framework is proposed for power system transient signal recognition based on a dynamic large convolution kernel and multilevel feature fusion network. The framework is composed of four parts: the downsampling layer, the dynamic large convolution kernel module, the multilayer feature fusion module, and the classification module. The downsampling layer is designed to reduce the computational effort and extract the main feature information of transient signals. The dynamic large convolution kernel module based on the advantage of dynamically adjusting the size of the convolution kernel, which can effectively solve the difficulty of extracting different types of transient features with a single fixed convolution kernel. The multilayer feature fusion network adapts to the variability in different local contextual information by dynamically adjusting the weights of the fusion strategy, which helps to fuse more discriminative feature information. The SoftMax classification module first performs dimensionality reduction using the fully connected layer, and then calculates the probability value to achieve the classification goal by the SoftMax classifier. In order to verify the effectiveness and generalization of the proposed method, comparative experiments were conducted in five different noise environments. The experimental results show that the model can accurately identify 15 single

and composite disturbance signal types with good noise immunity. It has good superiority compared with five other existing network models. Meanwhile, validation experiments are carried out using real data, and the proposed model still has a better diagnostic effect.

Author Contributions: Q.L.: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Validation, Writing—original draft. C.Z.: Formal analysis, Investigation, Methodology, Validation, Writing—original draft. S.L.: Conceptualization, Funding acquisition. S.D.: Project administration, Resources, Supervision. B.Z.: Conceptualization, Funding acquisition, Investigation, Methodology. Y.L.: Project administration, Supervision, Writing. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the National Natural Science Foundation of China under Grant No. 52277082 and the Science and Technology Program of the State Grid Corporation under Grant No. 5400-202124153A-0-0-00.

Data Availability Statement: Real datasets are available: https://www.kaggle.com/datasets/fykhlef/pqddatasetsarah. accessed on 1 June 2023.

Conflicts of Interest: Authors Qionglin Li, Chen Zheng, Shuming Liu, Shuangyin Dai and Bo Zhang come from Electric Power Research Institute of State Grid Henan Electric Power Company, and the author Yajuan Liu comes from Zhengzhou Power Supply Bureau of State Grid Henan Electric Power Company.

References

1. Veizaga, M.; Delpha, C.; Diallo, D.; Bercu, S.; Bertin, L. Classification of voltage sags causes in industrial power networks using multivariate time-series. *IET Gener. Transm. Distrib.* **2023**, *17*, 1568–1584. [CrossRef]
2. Li, D.; Mei, F.; Zhang, C.; Sha, H.; Zheng, J. Self-Supervised Voltage Sag Source Identification Method Based on CNN. *Energies* **2019**, *12*, 1059. [CrossRef]
3. Raza, A.; Benrabah, A.; Alquthami, T.; Akmal, M. A Review of Fault Diagnosing Methods in Power Transmission Systems. *Appl. Sci.* **2020**, *10*, 1312. [CrossRef]
4. Carvalho, L.; Lucas, G.; Rocha, M.; Fraga, C.; Andreoli, A. Undervoltage Identification in Three Phase Induction Motor Using Low-Cost Piezoelectric Sensors and STFT Technique. *Proceedings* **2020**, *42*, 72.
5. Xiong, S.; Zhang, F.; Zhou, Z.; Zeng, Y. Power data prediction method of new energy system based on wavelet transform and adaptive hybrid optimization. *Int. J. Low-Carbon Technol.* **2024**, *19*, 723–732. [CrossRef]
6. Qiu, W.; Tang, Q.; Liu, J.; Teng, Z.; Yao, W. Power Quality Disturbances Recognition Using Modified S Transform and Parallel Stack Sparse Auto-encoder. *Electr. Power Syst. Res.* **2019**, *174*, 105876.1–105876.10. [CrossRef]
7. Malik, H.; Almutairi, A.; Alotaibi, M.A. Power quality disturbance analysis using data-driven EMD-SVM hybrid approach. *J. Intell. Fuzzy Syst.* **2021**, *42*, 669–678. [CrossRef]
8. Singh, S.; Sharma, A.; Garg, A.R.; Mahela, O.P.; Khan, B.; Boulkaibet, I.; Neji, B.; Ali, A.; Ballester, J.B. Power Quality Detection and Categorization Algorithm Actuated by Multiple Signal Processing Techniques and Rule-Based Decision Tree. *Sustainability* **2023**, *15*, 4317. [CrossRef]
9. Shoryu, T.; Wang, L.; Ma, R. A Deep Neural Network Approach using Convolutional Network and Long Short Term Memory for Text Sentiment Classification. In Proceedings of the 2021 IEEE 24th International Conference on Computer Supported Cooperative Work in Design (CSCWD), Dalian, China, 5–7 May 2021.
10. Qu, H.; Li, X.; Chen, C.; He, L. Classification of power quality disturbances using convolutional neural networks. *Eng. J. Wuhan Univ.* **2018**, *15*, 314.
11. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention Is All You Need. *arXiv* **2017**, arXiv:1706.03762.
12. Liang, P.; Yu, Z.; Wang, B.; Xu, X.; Tian, J. Fault transfer diagnosis of rolling bearings across multiple working conditions via subdomain adaptation and improved vision transformer network. *Adv. Eng. Inform.* **2023**, *57*, 102075. [CrossRef]
13. Bhojanapalli, S.; Chakrabarti, A.; Glasner, D.; Li, D.; Unterthiner, T.; Veit, A. Understanding Robustness of Transformers for Image Classification. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 10–17 October 2021; pp. 10211–10221. [CrossRef]
14. Wu, H.; Xiao, B.; Codella, N.; Liu, M.; Dai, X.; Yuan, L.; Zhang, L. CvT: Introducing Convolutions to Vision Transformers. *arXiv* **2021**, arXiv:2103.15808.
15. Ding, X.; Zhang, X.; Zhou, Y.; Han, J.; Ding, G.; Sun, J. Scaling Up Your Kernels to 31 × 31: Revisiting Large Kernel Design in CNNs. *arXiv* **2022**, arXiv:2203.06717.
16. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin Transformer: Hierarchical Vision Transformer using Shifted Windows. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 10–17 October 2021.

17. Teng, Q.; Tang, Y.; Hu, G. Large Receptive Field Attention: An Innovation in Decomposing Large-Kernel Convolution for Sensor-Based Activity Recognition. *IEEE Sens. J.* **2024**, *24*, 13488–13499. [CrossRef]
18. Liu, H.; Watanabe, H. Feature Transfer Block for Feature Fusion in Lightweight Object Detectors. In Proceedings of the 2023 IEEE 12th Global Conference on Consumer Electronics (GCCE), Nara, Japan, 10–13 October 2023; pp. 302–305. [CrossRef]
19. Pathak, A.K.; Virmani, R.; Garg, A.; Singh, G.; Arya, I.; Chaurasiya, A. Deep Learning based PQD Classification using Time and Frequency Domain Features. In Proceedings of the 2024 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS), Bhopal, India, 24–25 February 2024; pp. 1–6. [CrossRef]
20. Ma, J.; Tang, Q.; He, M.; Peretto, L.; Teng, Z. Complex PQD Classification Using Time– Frequency Analysis and Multiscale Parallel Attention Residual Network. *IEEE Trans. Ind. Electron.* **2024**, *71*, 9658–9667. [CrossRef]
21. Ding, C.; Luktarhan, N.; Lu, B.; Zhang, W. A Hybrid Analysis-Based Approach to Android Malware Family Classification. *Entropy* **2021**, *23*, 1009. [CrossRef] [PubMed]
22. Lee, C.Y.; Shen, Y.X. Optimal Feature Selection for Power-Quality Disturbances Classification. *IEEE Trans. Power Deliv.* **2011**, *26*, 2342–2351. [CrossRef]
23. Husodo, B.Y.; Ramli, K.; Ihsanto, E.; Gunawan, T.S. Real-Time Power Quality Disturbance Classification Using Convolutional neural networkss. In *Recent Trends in Mechatronics Towards Industry 4.0*; Springer: Singapore, 2022. [CrossRef]
24. Chiam, D.H.; Lim, K.H. *Power Quality Disturbance Classification Using Transformer Network*; Springer: Cham, Switzerland, 2022. [CrossRef]
25. Junior, W.L.R.; Borges, F.A.S.; Rabelo, R.D.; De Lima, B.V.; De Alencar, J.E. Classification of Power Quality Disturbances Using Convolutional Network and Long Short-Term Memory Network. In Proceedings of the 2019 International Joint Conference on Neural Networks (IJCNN), Budapest, Hungary, 14–19 July 2019. [CrossRef]
26. Li, B.; Li, K.-C.; Xiao, X.-G. Composite power quality disturbance identification based on multi-scale convolutional fusion time series Transformer. *Grid Technol.* **2024**, *23*, 1–12.
27. Cai, J.; Zhang, K.; Jiang, H. Power Quality Disturbance Classification Based on Parallel Fusion of CNN and GRU. *Energies* **2023**, *16*, 4029. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.