

Learning Picturized and Time-series Data for Fault Location with Renewable Energy Sources

Yahui Wang; Qiushi Cui, Member, IEEE; Yang Weng, Senior Member, IEEE;
Dongdong Li, Member, IEEE; Wenyuan Li, Fellow, IEEE

Abstract

Transmission lines are heavy assets of power systems. Therefore, the rapid and accurate identification of fault locations is important for power grids' safe and stable operation. Traditional methods based on impedances or traveling waves are facing increasing challenges like uncertain power generation and unknown power electronic device characteristics when modern power systems are transitioning to deeply renewable energy source (RES) penetrated grids. In order to solve the above problems, this paper proposes a high-dimensional time-frequency feature extraction method that does not require expert knowledge of physical features. This paper proposes a fault location framework for learning picturized and time-series data (FltLoc-FPTD) with renewables. The developed loss function is suitable for the classification of faulty lines, considering the challenges of distinguishing faults in adjacent lines. Furthermore, we design an enhanced convolutional neural networks (CNN) subsampling layer with blur kernels to replace the traditional subsampling layer to eliminate the influence of high-frequency noise and improve the robustness against noise. The effectiveness of the method is verified by simulation under two benchmark systems. The average fault location errors with and without environment noises are 0.0189 and 0.0124.

Index Terms

Fault location, artificial intelligence, picturized data, time-series data, renewable energy sources.

I. INTRODUCTION

The modern power system has been involving the integration of heavy assets due to the advancement of smart grid technologies. Therefore, to avoid the damage to heavy assets of utilities, it is becoming increasingly important to identify power system faults before further security issues arise. The State Grid of China's 2022 annual work conference pointed out that the planned investment in power grid construction in 2022 exceeded CNY 500 billion (USD 78.8 billion) for the first time. New energy construction and power grid investment will be the focus in the future [1]. Traditional generators are geographically separated and usually far from the load side. However, the adaptation of RES technologies is changing the geographical tradition. In addition, these technologies are accompanied with uncertain

power generation, unknown power electronics device behaviors, etc., which bring up adverse protection issues especially on fault location, threatening the security of bulk systems.

The purpose of fault location methods is to find the fault occurrence point accurately and speedily on the lines. Traditional fault location methods generally include impedance-based [2]–[5], traveling wave-based [6], [7], and traditional machine learning-based methods [8], [9]. Specifically, impedance-based methods play a significant role in traditional power grids. They usually measure single or multi-terminal voltage and current, calculate line impedance, and then determine the location of the fault according to the distribution parameters of the line. However, in practical operation, due to a series of factors such as the unbalance of power grids, the fluctuation of the load, and the sensitivity to the line parameters, impedance-based methods could result in relatively large errors [10], [11]. Besides, for traditional impedance methods, it is usually assumed that the phase of the faulty branch current is the same as that of the current measured by the relay. Since RESs are connected to the grid relying on many inverter devices, the above assumptions can be invalid [12].

Furthermore, the traveling wave method has the advantages of high accuracy and fast speed. When a fault occurs, the generated traveling wave propagates between the fault point and the measuring device, then the propagation time and wave speed are utilized to calculate the location of the fault point [13]. In [14], the author uses the exact time when the first wave head signal of the traveling wave of the fault point arrives at both ends to accurately locate the fault point. However, this method has high requirements on the accuracy of the measurement devices. In addition, for a traveling wave measurement and positioning at two terminals, time synchronization and communication channels are required [15]. Since the traveling wave method is affected by the branch of the line, and its application is subject to certain restrictions on topology, this leads to another limitation when applying this method.

To resolve the shortcomings of the above traditional impedance and traveling wave methods, many researchers have combined signal processing techniques and artificial intelligence in the application of fault location. In [16]–[19], the authors employ wavelet transform to extract features, and then adopt a series of machine learning methods (such as SVM, SVR or artificial neural network) to classify and accurately locate faults. The application of traditional signal processing and artificial intelligence methods in fault location requires expert knowledge and mathematical skills for methods like wavelet transform or Fourier transform to construct feature engineering. However, when introducing expert knowledge and using the existing mathematical skills to construct feature engineering, the capability of capturing the voltage or current characteristics in large-scale power systems is limited by engineering perception and mathematical frameworks. These kinds of feature engineering are either heuristic or experimental, lacking “the last-mile” improvement on feature exploration.

To bridge the above gap, this paper investigates deep learning methods to design an intelligent fault detection method for grid-connected transmission systems with RESs. It does not require many features to extract the timing characteristics of voltage or current for fault location. With the rapid development of computer hardware capabilities in smart meters, technical support based on data-driven methods is given. Many articles use CNN [20]–[22] to extract features of voltage and current waveforms. The articles abandon the traditional feature extraction method and show better performance in the case of high network complexity. In [20], the authors propose a joint PMU placement algorithm that defines eigenvectors according to substitution theory in the case of limited measurement availability, and transmits them to a CNN network for faulty line identification. However, the article ignores the feature information on the time series and only focuses on the identification of faulty lines, and does not analyze the problem of fault location. In [21] the author proposes multi-task learning, combining two-way LSTM in the time domain and CNN in the frequency domain for simultaneous detection. The large number of parameters of LSTM undoubtedly increases the training burden of the model. In [22], the author image-processes the collected voltage and current signals and transmits them to the CNN network for fault location, which performs well in lightweight power systems. However, the authors imaged the sampled data in the form of a heatmap. When the data difference is small or the RESs are added to the power grid to generate a lot of noise, it is easy to lose some feature information during the image conversion process, resulting in increased errors. In this paper, in order to improve the robustness against noise and the accuracy of fault location, we improve the pooling structure of CNN for feature extraction of picturized data. To avoid the loss of time series information, we use BiGRU for feature extraction of time series data. When the extracted features are passed into the deep neural network for fault line identification, a loss function that is more suitable for fault line identification is adopted. The method relies on voltage and current signals at multiple measurement locations and requires no expertise. Furthermore, it uses deep learning models to extract high-dimensional features. The main contributions of this paper are as follows:

- The design of a novel deep learning framework that extracts high-dimensional fault signals. The framework uses a multi-channel CNN to extract signals features of picturized data and uses BiGRU to strengthen the internal connections of each channel and extract timing signal features.
- A loss function more suitable for fault line identification is designed. By reducing the intra-class distance within a line and increasing the inter-class distance between different lines, the accuracy of fault line identification is further improved.
- According to the principle that the objective function is usually fitted from low frequency to high frequency during training, the fuzzy kernel function is integrated into the subsampling of

traditional convolutional networks. Anti-aliasing is performed on high-frequency noise through blur subsampling, and the translation invariance of CNN is kept as much as possible during subsampling to prevent the model from over-fitting.

The organizational structure of this article is as follows. Section II introduces the specific framework of high-dimensional feature extraction and the application of blur subsampling. In Section III, the fault line identification and fault point location for high-dimensional features are introduced in detail. Section IV establishes a simulation model in PSCAD to evaluate the performance of the method under various conditions. Finally, conclusions are given in Section V.

II. EXTRACTION OF FEATURE VALUES BASED ON IMPROVED CNN AND BiGRU

In this section, we introduce in detail the detection framework of the proposed fault location method. The framework includes three parts: 1) preprocessing of sampled signals, 2) feature extraction of picturized data, and 3) feature extraction of time series data. This framework uses the improved CNN network to extract features of picturized data, and then uses the BiGRU network for serialization learning to obtain the characteristics of the voltage and current time series and strengthen the connection between the CNN pooling layers. The attention mechanism is added to the processing of CNN and BiGRU to calculate the weighted hidden state to complete effective feature screening. The proposed method framework is shown in 1, where (a) is a conceptual diagram and (b) is a flow chart. First, the original signal is acquired and assembled into a numerical matrix, $N@w \times v$. The numerical matrix is then learned through an improved CNN network to extract features and the BiGRU network is used for sequential learning. The learned features are fed into the fully connected neural network, then use an improved softmax loss function (Im-softmaxloss) to identify fault lines, and the mean square error loss function is used to determine fault points.

A. Data Processing from Sampled Signal to Numerical Matrix

Three-phase voltage and current signals directly reflect the state information of the grid before and after faults. Traditional feature engineering relies on expert knowledge to analyze voltage and current signals. However, with the integration of renewable energies, it is difficult for traditional methods to effectively extract voltage and current features for accurate fault location. Inspired by computer vision processing, a method from sampling signal to numerical matrix processing is proposed. The signal is converted from the time domain signal to the high-dimensional domain, providing an analyzable data structure for subsequent feature extraction. The specific steps are as follows:

- i) Collect three-phase voltage and current signals from the selected bus terminal;

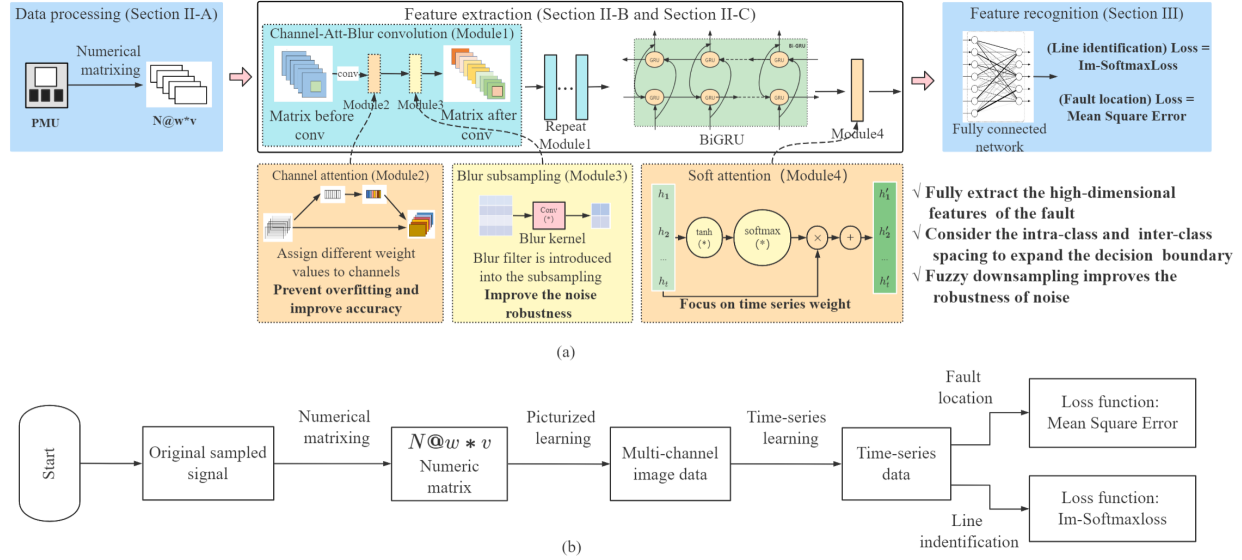


Fig. 1. Functional diagram of the proposed method. (a) The detailed conceptual diagram of the proposed method with specific details, (b) The flowchart of the proposed method.

ii) Perform Z-score standardization processing on the data; $\bar{n} = \frac{n-\mu}{\sigma}$, where μ and σ are the mean and variance of the sampled data, n is the original sampled data, and $\bar{n}$ is the standardized data.

iii) The sampling signals of n measuring devices are constructed into a numerical matrix of $n@6 \times I$, where n is the number of measuring devices on different buses, 6 represents the six channels of three-phase voltage and current, and I is the number of sampling points on the time series. The sampled raw signals are viewed as pictures and sent to the CNN.

B. Extract Features of Signals Picturized Data Based on Improved CNN

With the continuous development of the computer field, CNN has achieved good results in computer vision [23] and natural language processing [24]. The constructed numerical matrix utilizes a CNN for feature extraction of picturized data.

1) Traditional Convolutional Neural Network: CNN is a kind of feedforward neural network with deep structures including convolution calculation, as shown in Fig.2, which has powerful characterization ability. In the convolution process as shown in the Fig.2, the constructed numerical matrix is input to the next hidden layer of the neuron through the convolution of the local receptive field. By moving the local receptive field, all units of the constructed multi-channel numerical matrix are convolved to the next layer. Each hidden layer neuron has a bias and a local perceptual weight connected to it. All neurons in

this layer use the same weight and bias. Therefore, the output of the convolution of each layer on the signal s is:

$$O_i = \sigma \left[\sum_{k=1}^n \sum_{j \in R} w_{ij} s_{ij-k} + b \right], \quad (1)$$

where σ is the activation function, b is the shared bias, and w_{ij} is the array of shared weights. In most cases, the input is multi-channel (n). After the input of the k -th channel passes the global scan of the kernel function w_{ij} , the activation function (σ) is processed after the sum of the channels of each layer.

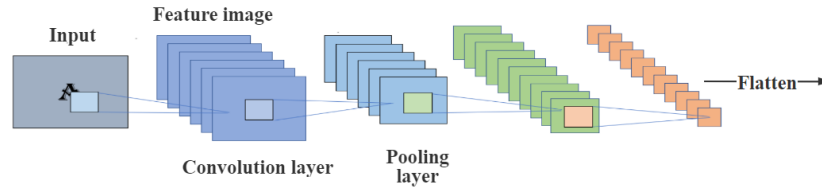


Fig. 2. CNN structure diagram.

After the features of the numerical matrix are obtained through convolution, if the features are directly passed to the next layer of network for training, the dimension is too high, the calculation amount is too large, and it is easy to lead to overfitting. The introduction of the pooling layer reduces the occurrence of over-fitting by reducing the dimensionality of features. To reduce the input dimension of the next layer, a pooling operation is performed $O_i: P = \text{Pooling}(O_i)$. The pooling layer is generally used after the convolutional layer, primarily to prevent overfitting caused by the dimensionality of the output of the convolutional layer which is too large. Each unit of the pooling layer summarizes a small area of the previous layer. A commonly used method is maximum pooling, and the maximum value in the small area of the previous layer is taken as the value of the corresponding pooling layer. After multiple convolution and pooling operations, the final feature vector is obtained.

2) Channel Attention Mechanism Between Different Measuring Devices: Traditional convolution operation basically defaults to fusing all channels of the input feature map, ignoring the relationship within the channel. For the power system, the characteristics of the faults on different lines shown on the measuring device are affected by the line distance and the line structure. We want to focus on the characteristics of the measurement device which is relatively close to the point of failure. Therefore, we introduce an attention mechanism in the channel, and the model can automatically learn the importance of different channel features and assign weight values to different channels, as shown in Fig.3. Only a small amount of data is required to increase, which greatly improves the accuracy of line recognition.

The method mainly consists of three parts:

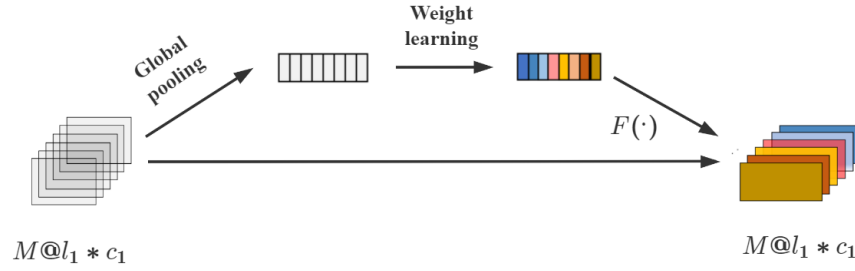


Fig. 3. SENet structure diagram.

(i) Global pooling: encode the entire feature in the channel as a global feature, which adopts the method of combining global pooling and maximum pooling;

$$y_a = \frac{1}{l \times c} \sum_{i=1}^l \sum_{j=1}^c u(i, j), y_m = \max(u(i, j)), y = y_a + y_m, \quad (2)$$

where $u(i, j)$ is the original eigenvalue, h_a is the eigenvalue of global average pooling, h_m is the eigenvalue of global maximum pooling, and h is the final global eigenvalue.

(ii) Weight learning: The neural network is used to learn the nonlinear relationship between the numerical matrix channels, and the sigmoid function is used as the activation function to compress the weight value to between 0 and 1.

$$w_m = \sigma(g(h, W)) = \text{Sigmoid } W_2 \text{ReLU}(W_1 y). \quad (3)$$

(iii) $F(\cdot)$: Conduct a dot product operation on the learned weight value and the feature of picturized data of each corresponding numerical matrix to obtain a new channel feature space.

$$\overline{u(i, j)} = w_m \cdot u(i, j), \quad (4)$$

where $\overline{u(i, j)}$ is the spatial eigenvalue after learning the nonlinear relationship in each channel. After learning, a new numerical matrix is obtained which is assigned according to the degree of feature.

3) Blur subsampling Improves Noise Robustness: Generally, the signal collected from the measuring device will have a certain amount of noise. As the model is trained, the fitting of the noise will be quite high. When the signal of the faulty line is collected and the noise with a signal-to-noise ratio of 40 is added to the signal, the model misjudges that the fault occurs on the line of the adjacent bus. Therefore, in order to improve the robustness of noise, we use blur subsampling in the convolution process to reduce the influence of noise on model prediction. In Fig.2, we use a certain step of the convolution operation, and the activation function adopts the tanh function, so a convolution operation is:

$$O_x = \sum_{i, j \in R} \tanh(w_{ij} \otimes x_{ij} + b). \quad (5)$$

The Fourier transform of $\tanh(w_{ij} \otimes x_{ij} + b)$ is:

$$\tanh(k) = \frac{2\pi i}{w_{ij}} \exp \frac{ibk}{w_{ij}} \frac{1}{\exp - \frac{\pi k}{2w_{ij}} - \exp \frac{\pi k}{2w_{ij}}}. \quad (6)$$

In model training, the general weight value w_{ij} is very small, assuming $\frac{\pi k}{2w_{ij}} > 0$, then:

$$\tanh(k) \approx - \frac{2\pi i}{w_{ij}} \exp \frac{ibk}{w_{ij}} \exp - \frac{\pi k}{2w_{ij}}, \quad (7)$$

$$O_x = \sum_{i,j \in R} - \frac{2\pi i}{w_{ij}} \exp \frac{ibk}{w_{ij}} \exp - \frac{\pi k}{2w_{ij}}. \quad (8)$$

For the difference between the output frequency domain of this convolution and the target frequency domain, we can define the loss function L in the frequency domain:

$$\begin{aligned} L &= \frac{1}{2} \int_{-Z}^Z (L_k)^2 dk \\ &= \frac{1}{2} \int_{-Z}^Z (O(k) - F(k))^2 dk = \frac{1}{2} \int_{-Z}^Z A(k) e^{i\phi(k)}^2 dk, \end{aligned} \quad (9)$$

where $F(k)$ is the Fourier change of the objective function. According to Parseval's theorem:

$$L = \frac{1}{2} \int_{-Z}^Z (O(x) - f(x))^2 dx. \quad (10)$$

The partial derivative of L_k to w_{ij} is:

$$\frac{\partial L(k)}{\partial w_{ij}} \approx A(k) \exp - \pi k / 2w_{ij}, \quad (11)$$

where $A(k) = |O(k) - F(k)|$. It can be seen from the above formula that if $A(k) > 0$ at this time, the gradient does not converge, and the low-frequency signal will be exponentially larger than the high-frequency signal. Currently, the low-frequency signal is dominant, and the direction of the gradient decline is in the direction of low-frequency convergence.

Therefore, in the model training, the signal sampled by the measuring device has high frequency noise. As the training level of the model deepens, the impact of high-frequency noise will be exacerbated, leading to the overfitting of the model. Generally, the signal collected from the measuring device will have a certain amount of noise. Therefore, to improve the robustness of noise, we use blur subsampling in the convolution process to reduce the influence of noise on the model prediction. The purpose is to anti-alias high-frequency noise and reduce the impact of high-frequency noise. To achieve this goal, we design a two-step max subsampling approach. First, the Gaussian kernel is used for convolution operation for anti-aliasing processing, and then maxpooling is performed to reduce the dimensionality of training.

C. BiGRU Based on Attention Mechanism to Extract Features of Time Series Data

To avoid the loss of features of time series signals, recurrent neural grids are often used as feature extraction tools. Long and short-term memory neural network (LSTM) effectively solves the problem of gradient burst or gradient disappearance in recurrent neural network (RNN) and is widely used in the field of speech recognition [25]. However, LSTM contains many parameters, which greatly increases the requirements for model training. The GRU network as an improved version of LSTM as shown in Fig.4. The forget gate and input gate in the original LSTM are replaced with update gates, and the model parameters are greatly reduced. Among them, update gate z_t and reset gate r_t , the output of which are respectively:

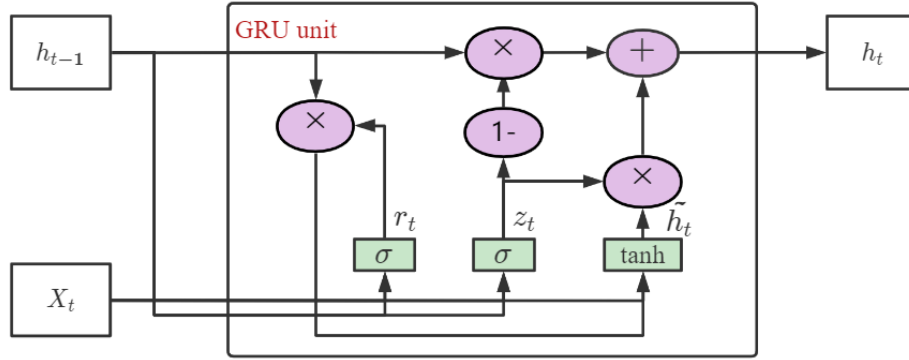


Fig. 4. The structure of GRU.

$$z_t = \sigma W_z \cdot [h_{t-1}, x_t], r_t = \sigma W_r \cdot [h_{t-1}, x_t], \quad (12)$$

$$h_t = \tanh W \cdot [r_t \otimes h_{t-1}, x_t], h_t = (1 - z_t) \otimes h_{t-1} + z_t \otimes h_t, \quad (13)$$

where W_z, W_r, W_t are the weight coefficients of each gate control unit in the network.

In the classic recurrent neural network, the state transmission is generally one-way. However, for some extracted features of picturized data, we hope to obtain more complete information about the current state based on the state before and after the time. Therefore, the two-way GRU can solve this problem and its structure is shown in Fig.5. The output of the BiGRU model at time t is not only related to time $t - 1$, but also related to time $t + 1$. Then the output at time t is:

$$\vec{h}_t = \text{GRU} \ x_t, \vec{h}_{t-1}, \overleftarrow{h}_t = \text{GRU} \ x_t, \overleftarrow{h}_{t-1}, \quad (14)$$

$$h_t = W_t \otimes \vec{h}_t + V_t \otimes \overleftarrow{h}_t + b_t, \quad (15)$$

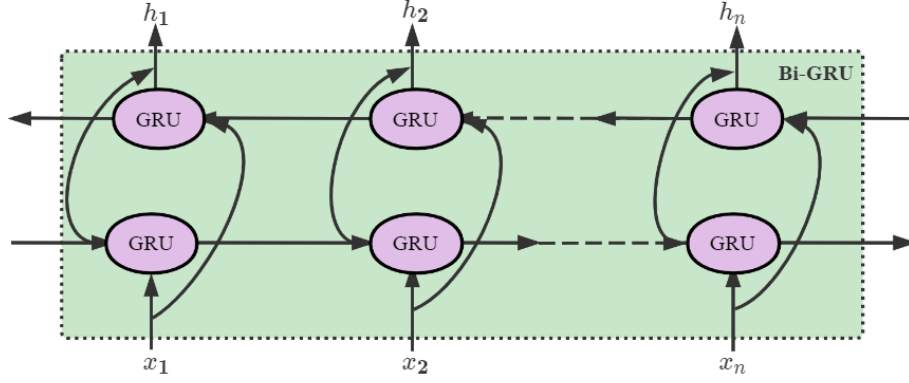


Fig. 5. The structure of BiGRU.

where $\vec{h}_t$ is the forward GRU output at time t , on the contrary, $\overleftarrow{h}_t$ is the reverse GRU output at time t , W_t and V_t are the weight matrix, and bt is the bias vector.

After learning the features of picturized data, the high-dimensional features of the fault signal are obtained. When the waveform changes within a period of time after a failure, we tend to focus more on the useful information. Therefore, we combine the attention mechanism with the high-dimensional features of the output of BiGRU to select features that are more critical to the current state from the many features. The attention mechanism is essentially acquainted with the visual information processing mechanism of the human brain. When people observe specific things, they tend to focus on observing useful information and suppress other useless information. The formula is expressed as:

$$\alpha_{ij} = \tanh(h'_{i-1}, h_j), c_{ij} = \frac{\exp(\alpha_{ij})}{\sum_{j=1}^t \exp(\alpha_{ij})}, h'_i = \sum_{j=1}^t c_{ij} h_j, \quad (16)$$

where h_j is the output of CNN-BiGRU, α_{ij} and c_{ij} are calculated by the tanh activation function and softmax function respectively, and h'_i is the variable of extracted features.

III. HIGH-DIMENSIONAL FAULT LINE CLASSIFICATION AND FAULT LOCATION IDENTIFICATION

There are two steps in this section. First, the neural network classification mechanism is used to find the faulty lines. Second, the regression mechanism is used to locate the fault.

A. Classification of Fault Lines Using Improved Loss Function

In the previous section, the improved CNN and BiGRU are utilized to deeply extract the voltage and current signals. With a time domain and high-dimension nature, the voltage and current signals are sent to a fully connected neural network. For traditional multi-classification problems, a softmax function is

widely used in the most primitive cross-entropy loss function. The traditional cross entropy loss function is as follows:

$$L_1 = -\frac{1}{N} \sum_{i=1}^N \log \frac{e^{W_i^T x_i + b_i}}{\sum_{j=1}^n e^{W_j^T x_i + b_j}}, \quad (17)$$

where W and b are the weight and bias of the model, respectively. At this time, the decision boundary for the classification problem is $(W_1 - W_2)x + b_1 - b_2 = 0$. However, for power systems, the difference of voltage and current at adjacent downstream or upstream lines is relatively small when a fault occurs; the voltage and current difference at both ends of one line and for different types of faults is relatively large when a fault occurs. The traditional cross-entropy loss function can easily lead to a decrease in classification accuracy, therefore, a feature scaling function is introduced to the cross-entropy loss function to improve the accuracy of classification by compressing the intra-class distance and increasing the accumulated distance. So we adopt Im-softmaxloss loss function, the loss function is now:

$$L_2 = -\frac{1}{N} \sum_{i=1}^N \log \frac{e^{s(\cos(\theta_i+m))}}{e^{s(\cos(\theta_i+m))} + \sum_{j=1, j \neq i}^n e^{s \cos \theta_j}}, \quad (18)$$

where $\theta_i = \frac{x_i}{\|x_i\|} \cdot \frac{W_j}{\|W_j\|}$. At this time, the decision boundary for the classification problem becomes $s \cos(\theta_1+m) - \cos \theta_2 = 0$, where $s(\cdot)$ is the feature scaling function. The new loss function fully considers the coupling of different lines within the system and expands the model's decision boundary for different faulty lines, thereby greatly increasing the accuracy of model classification.

B. Identify the Fault Point of The Fault Line

After the fault line is located, the voltage and current signals with high-dimensional features are transmitted to the fully-connected neural network. The fault location point $\hat{y}$ is calculated by the regression mechanism of the neural network:

$$\hat{y} = \sigma(\sigma(w_1 \cdot x)) \cdot w_2, \quad (19)$$

where w_1 and w_2 are the weights of the fully connected network, x is the high-dimensional feature, $\hat{y}$ is the fault location point, and σ is the activation function.

The mean square error loss function is often used in the regression problem of neural networks to train the weight values of neural networks:

$$L_{MSE} = \frac{1}{N} \sum_{i=1}^N y^{(i)} - \hat{y}^{(i)}{}^2, \quad (20)$$

where $y^{(i)}$ is the true value of the sample, and $\hat{y}^{(i)}$ is the predicted value of the sample during training.

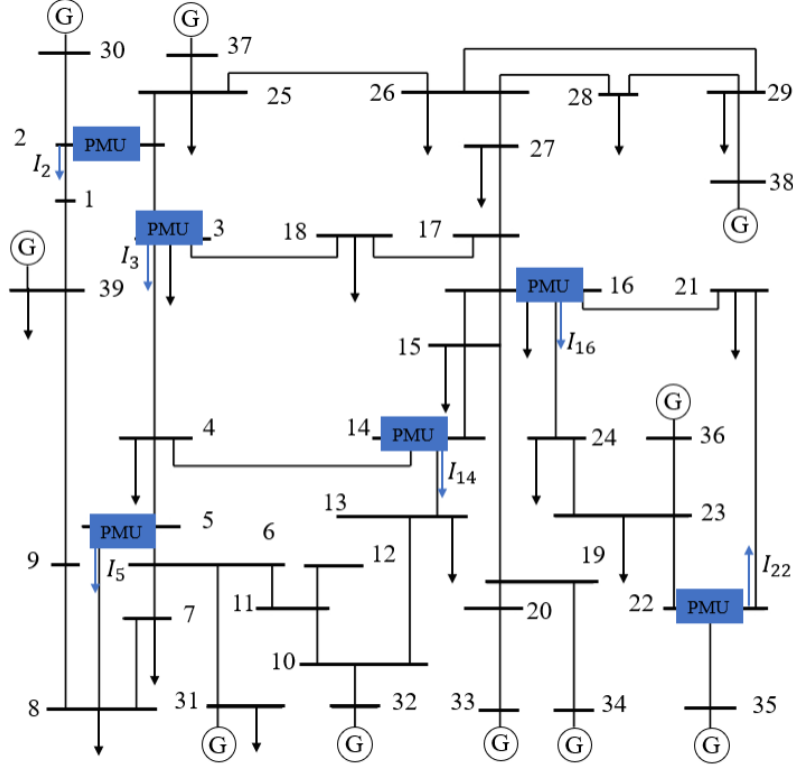


Fig. 6. IEEE 39-bus system with PMU locations and current measurement branches.

IV. NUMERICAL VERIFICATION

Simulations are conducted in PSCAD to verify the proposed fault detection, identification, and location method. Several IEEE benchmark systems are tested. In the IEEE 39-bus system, to be realistic and to show that the proposed framework is functioning without full observability of the whole system, it is assumed that PMUs are only equipped on buses 2, 3, 5, 14, 16, and 22, as shown in Fig.6. The RES adopts the doubly-fed induction machine model. The rated capacity of a single wind turbine is 1.5 MW and the rated voltage is 690 V; it is connected to the 220 kV transmission grid through a step-up transformer [26].

We have set up four types of faults: single-phase to ground fault (a-g), two-phase short-circuit fault (a-b), two-phase short-circuit ground fault (ab-g), and three-phase short-circuit fault (abc-g). The fault type, fault resistance, and fault location parameters are presented in Table I. Seven different fault resistances are tested for each fault type. Therefore, there are 28 situations at each failure point. Besides, each line takes 21 fault points and increases in a step of 5%, so there are $34 \times 28 \times 21 = 19992$ sampled data in total.

Table I: Setting of fault parameters

Settings	Parameter
Fault Line	1-34 lines
Fault Type	a-g, a-b, ab-g, abc-g
Fault Resistance	0.001, 0.1, 1, 5, 15, 25, 50
Fault Location	0%-100% (step=5%)

The performance is evaluated through three aspects:

- Line identification accuracy. The accuracy rate η_1 is used as the evaluation index for the accurate identification of the line. The calculation formula of η_1 is as follows: $\eta_1 = \frac{TP}{TP+FN}$ where TP and FN are the number of correct classifications and the number of misclassifications in the predicted value, respectively.
- Identification accuracy of adjacent lines. When the electrical distance between adjacent lines is small, misclassification may occur when a fault occurs. In order to better analyze the misclassification situation, we set a new accuracy rate. The classification accuracy rate of adjacent lines is η_2 : $\eta_2 = \frac{TP_1}{TP_1+FN_1}$, where TP_1 is the sum of the number of the predicted line is a faulty line or the predicted line is not a faulty line but is adjacent to the faulty line, FN_1 is the number of predicted lines that are not faulty lines and that the predicted line and the faulty line are not adjacent lines.
- Fault location. Mean square error (MSE) and mean absolute error (MAE) are commonly used evaluation indicators for regression problems. For multiple lines, it is needed to calculate the average value of multiple lines. The calculation formulas of MSE and MAE are as follows: $MSE = \frac{1}{m} \sum_{j=1}^m \frac{1}{N} \sum_{i=1}^N y^{(i)} - \hat{y}^{(i)}^2$, $MAE = \frac{1}{m} \sum_{j=1}^m \frac{1}{N} \sum_{i=1}^N |y^{(i)} - \hat{y}^{(i)}|$, where m is the total number of lines, and N is the number of samples.

A. Comparison of FltLoc-FPTD and Traditional Impedance Method

In the IEEE-39 node system shown in Fig.6, we select a line between bus 8 and 30 to compare the performance between the traditional impedance-based fault location method and the data-driven one. The impedance method adopts a single-ended positioning method [27], and an impedance relay is installed on bus 30. To verify the impact of RESs on the traditional method, we replaced the thermal power unit on the original No. 39 bus with a large wind farm in the system. The wind farm consists of 10 wind turbines. The stator winding of the generator is directly connected to the grid, and the rotor winding is connected to the grid side through a back-to-back converter. The rotor-side converter controls the rotor current to

realize the decoupling control of active and reactive power. A thyristor-controlled voltage regulator is installed on the low-voltage side of the wind farm to control the voltage.

The AAE value is deployed to evaluated through the average accumulative error under different fault simulations: $AAE = \frac{1}{N} \sum_{i=1}^N \frac{M_i - A_i}{\text{total line length}}$, where M is the measured value of fault location, A is the actual value of fault occurrence. The results are shown in Table II. For the impedance method, the existence of RESs do affect its location accuracy. The average errors with RESs are all higher than the ones without RESs. However, for the proposed method, the location error does not increase due to the integration of RESs. Through comparison, the proposed method has much lower location error than the impedance method. For example, when there is no RESs, the proposed method's errors reduces more than a half than the impedance method; when there are RESs, the proposed method has errors of almost one third of the impedance method's errors.

Table II: Comparison of traditional impedance method and FltLoc-FPTD on fault location with and without RESs

Method	Fault Type	Error (%)	
		No RESs	Including RESs
Impedance Method	a – g	0.253	0.356
	a – b	0.237	0.412
	ab – g	0.312	0.374
	abc – g	0.264	0.362
FltLoc-FPTD	a – g	0.142	0.132
	a – b	0.113	0.122
	ab – g	0.086	0.115
	abc – g	0.146	0.143

The following analysis of the increase of the error of the traditional impedance method caused by the integration of RESs into the grid. As shown in Fig.7, when the line between the bus bars fails, the single-ended positioning method is used, and the fundamental frequency component is extracted by

Fourier transform. The fault positioning result is:

$$\begin{aligned}
 Z &= \frac{V_{G1}}{I_{G1}} = \frac{Z_{L1}I_{G1} + mZ_{L2}(I_{G1} + I_{G2}) + R_F(I_{G1} + I_{G2} + I_{G3})}{I_{G1}}, \\
 Z &= Z_{L1} + mZ_{L2} + \frac{mZ_{L2}I_{G2}}{I_{G1}} + \frac{R_F(I_{G1} + I_{G2} + I_{G3})}{I_{G1}}, \\
 m &\approx \frac{(Z - Z_{L1})I_{G1}}{Z_{L2}(I_{G1} + I_{G2})},
 \end{aligned} \tag{21}$$

where Z is the calculated impedance of the impedance method, m is the location coefficient of the fault in percent, Z_{L1} and Z_{L2} are the line impedance values, I_{G1} , I_{G2} , I_{G3} are the RESs branch currents, R_F is the fault resistance. Table III compares the difference between the actual fault distance and the calculated fault location (m) between the data-driven method and the traditional impedance method. Two scenarios are considered: with and without RESs. The maximum error of the FItLoc-FPTD is only 0.13%, the minimum error is 0.08%, and the error fluctuation is small. Due to the integration of a large number of power electronic devices and the impact of their different control strategies, the output current of RESs has random fluctuations, which undoubtedly increases the error of the impedance method. When there is no RESs connected to the grid, the error between the actual fault distance and the calculated fault distance is up to 0.26%. However, when renewables are plugged into the power system, the average error ranges from 0.32% to 0.38%. This is due to the randomness and uncertainty of RESs and the integration of a large number of inverters, which greatly increases the error of traditional fault location methods.

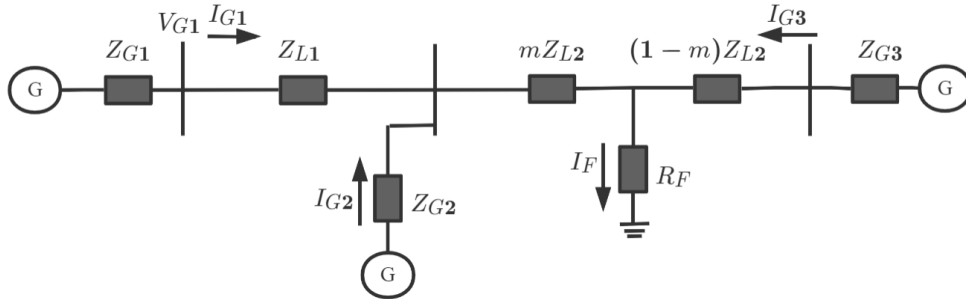


Fig. 7. Permanent fault in the line between bus 30 and bus 8.

B. Performance of Different Deep Neural Networks on Fault Location

We train the sampled data under different deep learning frameworks and use 20% of the aforementioned data as testing in the benchmark system. The data required for training in different frameworks are processed by the numerical matrix in this paper. The line recognition accuracy η , the mean square error (MSE) and the mean absolute error (MAE) of fault point identification are compared in different

Table III: Comparison of fault location error between impedance method and FltLoc-FPTD with and without RESs

Method	Fault location as a percentage of line length	Without RESs			With RESs		
		Measured Impedance (p.u)	Measured the fault location	Error (%)	Measured Impedance (p.u)	Measured the fault location	Error (%)
Impedance Method	0	0.008	0.0024	0.24	0.012	0.0035	0.35
	25%	0.149	0.2478	0.22	0.137	0.2464	0.36
	50%	0.322	0.5020	0.20	0.334	0.5038	0.38
	75%	0.495	0.7526	0.26	0.506	0.7532	0.32
	100%	0.642	0.9977	0.23	0.633	0.9965	0.35
FltLoc-FPTD	0	-	0.0012	0.12	-	0.0009	0.09
	25%	-	0.2487	0.13	-	0.2512	0.12
	50%	-	0.5489	0.11	-	0.5492	0.08
	75%	-	0.7491	0.09	-	0.7488	0.12
	100%	-	0.9988	0.12	-	0.9989	0.11

frameworks to analyze the performance of each method. The results are shown in Table IV. In the test results of CNN and GRU frameworks, the classification accuracy (η_1) and positioning error (MAE) are 0.9520, 0.9556 and 2.42, 2.03, respectively. The detection framework proposed in this paper improves the classification accuracy (η_1) by a maximum of 3%, and reduces the fault location error (MAE) by a maximum of 1.66%. The FltLoc-FPTD proposed in this paper has a line classification accuracy of 0.9815, and its recognition accuracy at the same bus is 1.0. Therefore, even if the faulty line is wrongly judged for the first time, the fault point can be picked up by the next adjacent line. The fault location error is 0.76, which is superior to traditional CNN and GRU in performance.

Table IV: Performance comparison under different frameworks.

Framework	η_1	η_2	MSE(%)	MAE (%)
CNN	0.9520	0.9812	0.155	2.420
GRU	0.9556	0.9872	0.092	2.030
BiGRU	0.9635	0.9970	0.068	1.690
CNN-BiGRU	0.9658	0.9983	0.045	1.456
FltLoc-FPTD	0.9815	1.0000	0.012	0.760

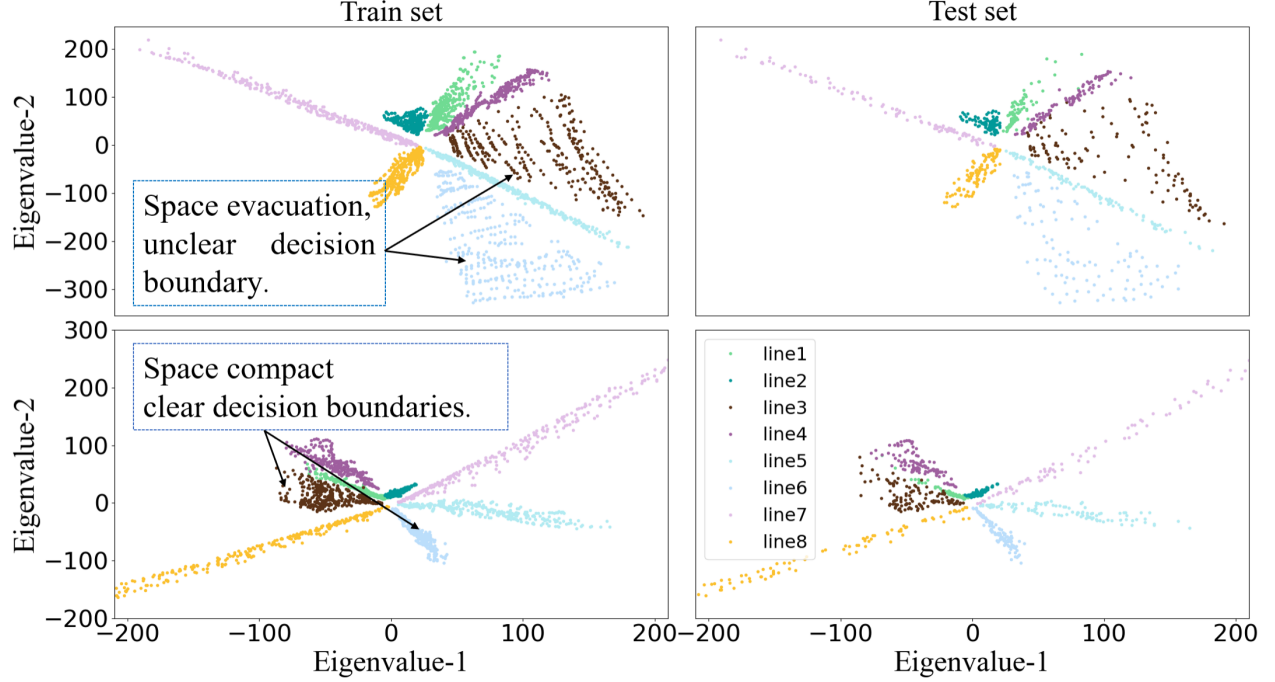


Fig. 8. Two-dimensional visualization of softmax loss on the train set (a) and test set (b).

C. Loss Function More Suitable for Fault Line Classification

In the case of a large number of lines, the coupling of features in space is prevented, and the classification accuracy is further improved. Therefore, we introduce a loss function that is more suitable for high dimensional decoupling. The fault data is extracted into a high-dimensional feature space, and the data in the high-dimensional space is compressed by the introduced large interval cross-entropy loss function. As shown in Fig.8, in the trained model, we select eight different lines to visualize the distinguishability. The two subgraphs in the upper part of the Fig. 8 are the classification results of the training and test sets using the traditional softmax loss. Obviously, the coverage of line 3 and line 6 is loose, and the space utilization rate is greatly reduced. If a new line is added in the space, a coupling relationship will inevitably occur, thereby reducing the accuracy of classification. The two subplots at the bottom of the Fig.8 show the classification results with a large margin cross-entropy loss function. It can be seen the spatial distance within the same line is greatly compressed, and the classification decision boundary between different lines is clearly visible.

D. The Influence of the Introduction of the Attention Mechanism on the Positioning Accuracy

In the 39-node benchmark system, only 6 PMUs are installed for demonstration purposes. For faults at different lines, the distance between the fault point and the measuring device affects the measurement

results. For example, when a fault occurs on line 1-2, the voltage values measured on bus 22 differ from that of bus 2. Since bus 22 is far away from the location of the fault, the voltage fluctuation is almost zero. In addition, the data in time domain contains signals at steady state and the fault period, and the model should pay more attention to the information that is more useful for fault location. As shown in Fig.9, the visualization of the numerical matrix after convolution shows that each channel has a different function in its feature representation. We can see that channels (a) and (d) focus on the waveform characteristics before and after a fault occurs, and channels (b), (c), (e) aim to extract the features at the time of the fault. Therefore, we introduce the attention mechanism into each channel of CNN, corresponding to each measuring device, and the time sequence of the corresponding signal in the sequence output of BiGRU. Through the distribution of weight values, we hope that the model will focus the characteristics of the state change sequence and the characteristics of the failure duration, which reduces the characteristics of the steady state.

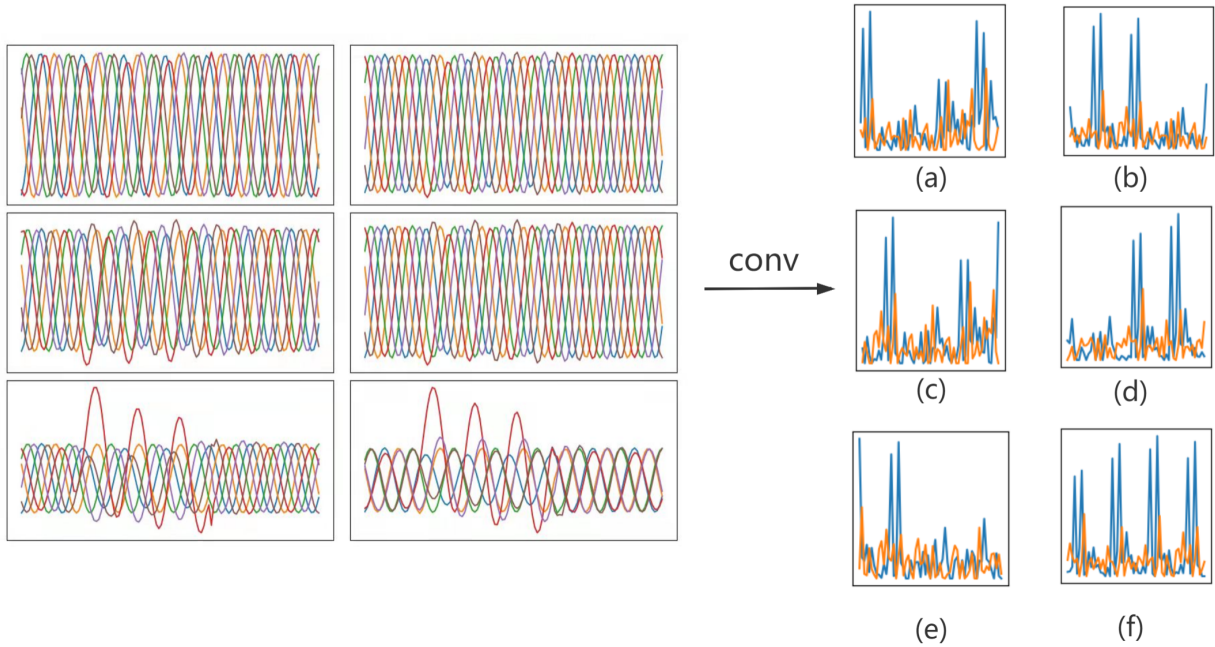


Fig. 9. The visualization result of the numerical matrix after convolution. Channels (a) and (d): waveform characteristics before and after a fault occurs, channels (b), (c) and (e): waveform characteristics when the fault occurred, channel (f): waveform characteristics before and after the fault occurs and when the fault occurs.

As shown in Table V, we compare the classification accuracy (η_1 and η_2) and fault localization errors (MSE and MAE) of three models: 1) CNN BiGRU: twin Deep Neural Networks for Extracting Features of Picturized Data and Features of Time Series Data via CNN and BiGRU. 2) Multi-att CNN BiGRU:

on top of 1), a multi-channel attention mechanism is introduced into CNN to strengthen the connections within each channel. 3) Multi-att CNN BiGRU-att: on top of 2), the attention mechanism is introduced into BiGRU to strengthen the processing of time series features.

In the detection framework using the combination of multi-channel convolution and attention mechanism proposed in this paper, its accuracy η and MSE, MAE have a certain improvement in performance. After the attention mechanism is introduced, the accuracy rate of line classification is higher than 0.99, and the MSE and MAE of fault location are 0.0124 and 0.76 respectively, which are better than the model without attention mechanism. We compare the visualization of the trained models in Table V and the results are shown in Fig. 10. The abscissa is the actual location of the fault point, and the ordinate is the predicted location of the fault point. The visualization results present that, with the introduction of the attention mechanism, the more concentrated the two-dimensional scatter plot is on the diagonal line, the closer the predicted value is to the true value.

Table V: Comparison of the results of introducing the attention mechanism into the model

Model	η_1	η_2	MSE(%)	MAE(%)
CNN BiGRU	0.9832	0.9983	0.0456	1.542
Multi-att CNN BiGRU	0.9931	1.0000	0.0156	0.970
Multi-att CNN BiGRU-att	0.9976	1.0000	0.0124	0.760

E. High Robustness Resulting From Blur Subsampling

Generally, the signals collected from various measuring devices will contain a certain amount of noise. For traditional deep learning models, when fitting data according to the principle of low frequency dominance in the frequency principle, the performance of the model will be greatly reduced if the high frequency components in the signal are dominant. When the noise contains high-frequency signals, there is often a risk of overfitting. The IEEE standard C37.118 [28] does not specify the signal-to-noise ratio (SNR) measured by the PMU and the signal-to-noise ratio of the PMU, as it is different in various regions. We chose Gaussian noise with a signal-to-noise ratio from 40 dB to 100 dB to be added to the data to compare the difference of before and after adding blur subsampling to the model [29]. As shown in Fig.11, the maximum MSE and average MSE of different signal-to-noise ratios on the test set are compared between normal and blur subsampling. It can be seen from Fig. 11 that when blur subsampling is not used, the maximum error of the model is greatly affected by noise, and the error value of fault location fluctuates more obviously. However, after adding the blur subsampling model, the fluctuation

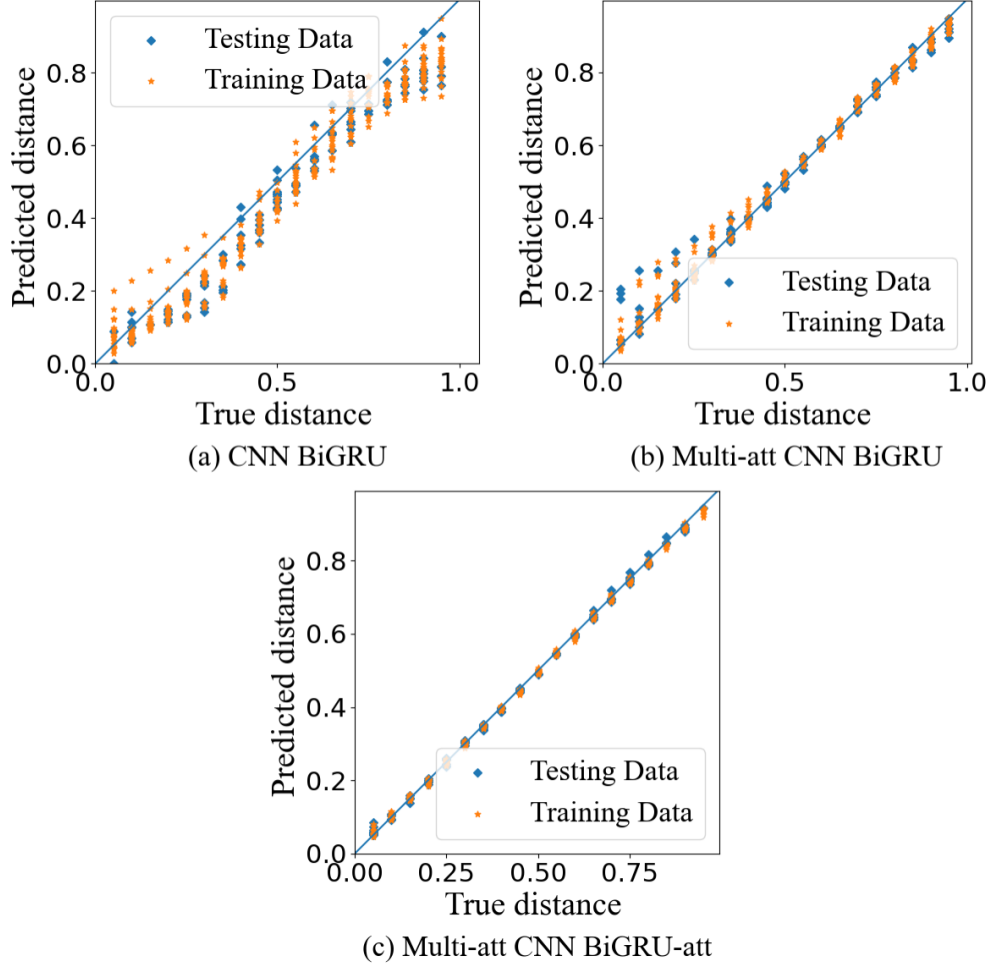


Fig. 10. Visualization of the influence on fault location after adding the attention mechanism.

of the fault location error is significantly reduced. With the increase of signal-to-noise ratio, the average error of fault location of blur subsampling model is lower than that of traditional subsampling. Therefore, we introduce blur subsampling into the model to perform anti-aliasing on high-frequency signals, and the translation invariance of CNN can be maintained as much as possible during blur subsampling to prevent the model from overfitting.

F. Experimental Comparison Under Different Methods

To prove the performance of this method, four machine learning models are selected: Back Propagation Neural Network (BPNN), Support Vector Regression (SVR), Decision Tree (DT), and Random Forest (RF). The same data set is deployed in training and testing. The evaluation indicators are still η_1 , η_2 , MSE and MAE. The results are shown in Table VI. The method proposed in this paper has the highest

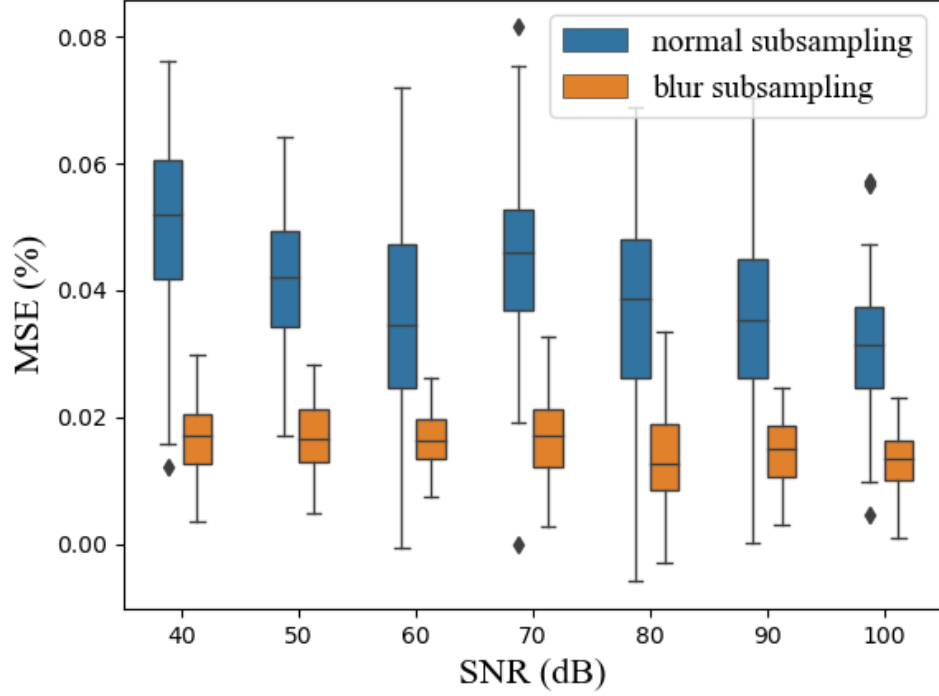


Fig. 11. MSE of blur subsampling under different signal-to-noise ratios.

classification accuracy η_1 of faulty lines under different signal-to-noise ratios, and the lowest is 0.9868. The MAE loss value of fault location is as low as 0.76 and as high as 0.92. Traditional machine learning models cannot achieve such high accuracy. And as the noise of the test data increases, the localization accuracy of the traditional machine learning model fluctuates greatly with the disturbance of the noise. This is due to the complex coupling between lines, and traditional models cannot obtain high-dimensional features. The method proposed in this paper outperforms traditional models in performance and has strong robustness under noise interference. This method can obtain high-dimensional features of fault signals, so it has high fault location accuracy.

G. The Influence of PMU Coverage on Fault Location

In this section, the influence of PMU coverage on the positioning accuracy in the system is discussed. PMU coverage is the ratio of the number of buses with PMU to the total number of buses in the system. The coverage rate of the PMU directly affects the number of channels of the numerical matrix in data preprocessing. Furthermore, a configuration with a high ratio of PMU will get more information about the fault characteristics. As shown in Fig.12, we compare the location performance with different noise levels at a ratio of 15%-30%. The results show that as the proportion of PMU configuration increases,

Table VI: Experimental comparison among different methods.

Method	SNR	η_1	η_2	MSE(%)	MAE(%)
BPNN	0	0.9566	0.9946	0.5230	1.213
	40	0.9446	0.9653	0.8670	1.645
	60	0.9396	0.9787	0.7350	1.455
	80	0.9466	0.9827	0.6480	1.366
SVR	0	0.9433	0.9868	0.6780	1.787
	40	0.9217	0.9655	0.8320	2.122
	60	0.9316	0.9688	0.7890	2.013
	80	0.9423	0.9799	0.7530	1.859
DT	0	0.9215	0.9855	1.0220	2.020
	40	0.9112	0.9732	1.3560	2.350
	60	0.9136	0.9813	1.2440	2.250
	80	0.9215	0.9825	1.1330	2.130
RF	0	0.9326	0.9912	0.9630	1.430
	40	0.9125	0.9789	1.6820	2.120
	60	0.9256	0.9815	1.5730	1.890
	80	0.9306	0.9856	1.2600	1.680
FltLoc-FPTD	0	0.9976	1.0000	0.0124	0.760
	40	0.9868	0.9988	0.0189	0.920
	60	0.9931	1.0000	0.0178	0.850
	80	0.9942	1.0000	0.0160	0.810

the MSE of the model presents a downward trend as a whole. However, as the coverage ratio increases, its investment and operating costs will also increase accordingly.

V. DISCUSSIONS

A. Comparison with other works

The proposed method is compared with other ones in the literature. The results are shown in Table VII. All methods are tested under IEEE 39-bus and IEEE 68-bus benchmark systems. We compare whether the method requires expert knowledge to construct a large amount of feature engineering (feature construction), the difficulty of implementing the method (implementability), the average error with the

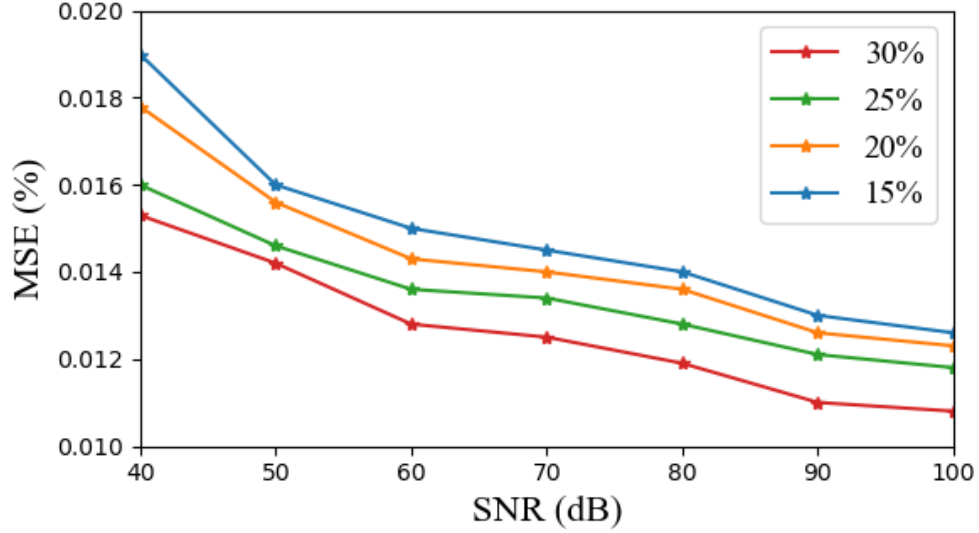


Fig. 12. The average MSE under 15% – 30% PMU coverage against different signal-to-noise ratios.

cases of without and with noises. In order to consider the effectiveness of the method, we further compare the impact of the selection and placement of the PMUs among the proposed method and another two methods in [30] and [31].

In [20], [27], the measured voltage and current signals are artificially processed, while data fusion and data-driven algorithms are utilized to achieve accurate fault location. Although both methods have achieved good results in terms of error, a large number of feature structures undoubtedly increases the difficulty of model implementation. In [29], the author transmits the measured signal value to the BiGRU network without manual participation. The method in [29] is simple to implement, but the error of 0.176% under 40 signal-to-noise ratio is much larger than the method in this paper and the literature [20], [27]. The fault location error in the proposed method is 0.016% in the case of no noise, which is slightly lower than that of the method in [27], however, its implementation difficulty is much lower than that of [27]. Under the premise of referring to the placement of the PMU, the methods in this paper and the compared literature are slightly reduced. It has to be pointed out that this paper does not focus on PMU placement, therefore, the relationship between optimal PMU placement and the accuracy of fault location is not demonstrated here. To sum up, the method proposed has the following advantages: it requires no expert knowledge to construct feature engineering, achieves comparatively simple implementation, and maintains high performance in the error of fault location.

Table VII: Fault location performance comparison.

System	Method	Feature construction	Implementability	Fault location error (%)			
				PMU placement in this paper		PMU placement in [30]	
				With noise	Without noise	With noise	Without noise
IEEE 39-bus	In [20]	Required	Difficulty	0.020	0.032	0.018	0.029
	In [27]	Required	Difficulty	0.014	0.022	0.013	0.021
	In [29]	Not required	Easy	0.087	0.176	0.069	0.163
	FltLoc-FPTD	Not required	Medium	0.015	0.018	0.015	0.019
IEEE 68-bus	In [20]	Required	Difficulty	0.023	0.035	0.021	0.030
	In [27]	Required	Difficulty	0.019	0.026	0.017	0.024
	In [29]	Not required	Easy	0.121	0.193	0.105	0.170
	FltLoc-FPTD	Not required	Medium	0.018	0.021	0.016	0.020

B. Shunt faults and series faults

Power system faults are generally divided into shunt faults and series faults. For traditional transmission line fault location, researchers and engineers usually refer to shunt fault. Shunt faults include line-to-ground faults, line-to-line faults, double-line-to-ground faults, and triple line-to-ground fault. However, series faults, also known as open circuit faults, occur due to the failure of one or more conductors in an electrical system. Series faults generally occur at the terminals of electrical equipment. The most common cause is the faulty connectors on cables and overhead lines [32]. In this paper, a novel deep learning framework is designed to extract high-dimensional features via picturized and time-series learning of sampled data, and a loss function more suitable for power systems is designed to further improve the accuracy. Since this paper relies on a data-driven approach, no additional expert knowledge is required to construct feature engineering. For series faults, the method proposed in this paper can still obtain good accuracy by extracting high-dimensional features of fault signals. However, it depends on two cases. First, series arc faults usually occur at broken conductors, and the arc impedance is so small that the power supply continues to supply power to the load without triggering any relays. Besides, the change of fault current is small. Therefore, the extraction of features will cause great difficulties, and the model is prone to underfitting. Second, for other series faults, by adjusting the configuration and parameters of the model and then preprocessing the series fault signal, the proposed model can have a high accuracy

by uniformly scaling the signal values during the fault to produce large differences. Therefore, the fault location method proposed in this paper improves the fitting capability by adjusting the parameters of the model, making the extraction and preprocessing of fault signals suitable for series faults.

VI. CONCLUSIONS

In this paper, we design a novel deep learning framework, which utilizes a multi-channel attention mechanism CNN to extract features from picturized data and a BiGRU structure to extract features from time-series data. It overcomes the need to rely on a large amount of expert knowledge to construct features for fault detection and location. For the high-dimensional coupling within the power line, a loss function that is more suitable for fault line identification is established, and the decision boundary is expanded for different fault lines to effectively identify the fault line. The fuzzy kernel function is introduced into the down-sampling of the convolutional network to anti-alias the signal in the noise, prevent the over-fitting caused by the noise, and improve the robustness of the noise. Simulation experiments show that this method is superior to traditional impedance-based method and other machine learning methods.

REFERENCES

- [1] S. G. C. of China, "State Grid Corporation Work Conference Report," State Grid Corporation of China, Tech. Rep., 2022.
- [2] R. Krishnathevar and E. E. Ngu, "Generalized impedance-based fault location for distribution systems," *IEEE Transactions on Power Delivery*, vol. 27, no. 1, pp. 449–451, 2012.
- [3] M. Farajollahi, A. Shahsavari, E. M. Stewart, and H. Mohsenian-Rad, "Locating the source of events in power distribution systems using micro-pmu data," *IEEE Transactions on Power Systems*, vol. 33, no. 6, pp. 6343–6354, 2018.
- [4] R. Dashti, M. Daisy, H. R. Shaker, and M. Tahavori, "Impedance-based fault location method for four-wire power distribution networks," *IEEE Access*, vol. 6, pp. 1342–1349, 2018.
- [5] S. Das, S. Santoso, A. Gaikwad, and M. Patel, "Impedance-based fault location in transmission networks: theory and application," *IEEE Access*, vol. 2, pp. 537–557, 2014.
- [6] R. Kumar and D. Saxena, "A traveling wave based method for fault location in multi-lateral distribution network," in *International Conference on Power, Instrumentation, Control and Computing*, 2018, pp. 1–5.
- [7] F. V. Lopes, K. M. Dantas, K. M. Silva, and F. B. Costa, "Accurate two-terminal transmission line fault location using traveling waves," *IEEE Transactions on Power Delivery*, vol. 33, no. 2, pp. 873–880, 2018.
- [8] M. Pourahmadi-Nakhli and A. A. Safavi, "Path characteristic frequency-based fault locating in radial distribution systems using wavelets and neural networks," *IEEE Transactions on Power Delivery*, vol. 26, no. 2, pp. 772–781, 2011.
- [9] H. Livani and C. Y. Evrenosoğlu, "A fault classification and localization method for three-terminal circuits using machine learning," *IEEE Transactions on Power Delivery*, vol. 28, no. 4, pp. 2282–2290, 2013.
- [10] M. Kezunovic, "Smart fault location for smart grids," *IEEE Transactions on Smart Grid*, vol. 2, no. 1, pp. 11–22, 2011.
- [11] M. Shafiullah and M. A. Abido, "A review on distribution grid fault location techniques," *Electric Power Components & Systems*, pp. 1–18, 2017.
- [12] P. Chang, G. Song, J. Hou, and R. Xu, "A single-terminal fault location method for transmission lines integrated by inverter-type source," *IEEE Transactions on Power Delivery*, pp. 1–9, 2021.

- [13] L. Zewen, H. Huanhuan, D. Feng, Z. Xiangjun, and Y. Kun, "Power grid fault traveling wave network location method," in IEEE Industry Applications Society Annual Meeting, 2013, pp. 1–6.
- [14] F. V. Lopes and E. J. S. Leite, "Traveling wave-based solutions for transmission line two-terminal data time synchronization," IEEE Transactions on Power Delivery, vol. 33, no. 6, pp. 3240–3241, 2018.
- [15] A. Fedorov, V. Petrov, O. Afanasieva, and I. Zlobina, "Limitations of traveling wave fault location," in Ural Smart Energy Conference, 2020, pp. 21–25.
- [16] K. Moloi and A. O. Akumu, "Power distribution fault diagnostic method based on machine learning technique," in IEEE PES/IAS PowerAfrica, 2019, pp. 238–242.
- [17] Z. Zhu, "Fault location in power transmission line based on isvr," in The 2nd International Conference on Industrial Mechatronics and Automation, vol. 2, 2010, pp. 514–517.
- [18] P. Ray, D. P. Mishra, and D. D. Panda, "Hybrid technique for fault location of a distribution line," in Annual IEEE India Conference, 2015, pp. 1–6.
- [19] K. Lout and R. K. Aggarwal, "Current transients based phase selection and fault location in active distribution networks with spurs using artificial intelligence," in IEEE Power Energy Society General Meeting, 2013, pp. 1–5.
- [20] W. Li, D. Deka, M. Chertkov, and M. Wang, "Real-time faulted line localization and pmu placement in power systems through convolutional neural networks," IEEE Transactions on Power Systems, vol. 34, no. 6, pp. 4640–4651, 2019.
- [21] P. Tehrani and M. Levorato, "Frequency-based multi task learning with attention mechanism for fault detection in power systems," in IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids, 2020, pp. 1–6.
- [22] Y. Yu, M. Li, T. Ji, and Q. H. Wu, "Fault location in distribution system using convolutional neural network based on domain transformation," CSEE Journal of Power and Energy Systems, vol. 7, no. 3, pp. 472–484, 2021.
- [23] T. Ahmed, P. Das, M. F. Ali, and M.-F. Mahmud, "A comparative study on convolutional neural network based face recognition," in International Conference on Computing, Communication and Networking Technologies, 2020, pp. 1–5.
- [24] A. B. Abdul Qayyum, A. Arefeen, and C. Shahnaz, "Convolutional neural network (cnn) based speech-emotion recognition," in IEEE International Conference on Signal Processing, Information, Communication Systems, 2019, pp. 122–125.
- [25] T. He and J. Droppo, "Exploiting lstm structure in deep neural networks for speech recognition," in IEEE International Conference on Acoustics, Speech and Signal Processing, 2016, pp. 5445–5449.
- [26] M. Kheshti, X. Kang, G. Song, and Z. Jiao, "Modeling and fault analysis of doubly fed induction generators for gansu wind farm application," Canadian Journal of Electrical and Computer Engineering, vol. 38, no. 1, pp. 52–64, 2015.
- [27] Z. Jiao and R. Wu, "A new method to improve fault location accuracy in transmission line based on fuzzy multi-sensor data fusion," IEEE Transactions on Smart Grid, vol. 10, no. 4, pp. 4211–4220, 2019.
- [28] K. E. Martin, "Synchrophasor measurements under the IEEE standard C37.118.1-2011 with amendment C37.118.1a," IEEE Transactions on Power Delivery, vol. 30, no. 3, pp. 1514–1522, 2015.
- [29] F. Zhang, Q. Liu, Y. Liu, N. Tong, S. Chen, and C. Zhang, "Novel fault location method for power systems based on attention mechanism and double structure gru neural network," IEEE Access, vol. 8, pp. 75 237–75 248, 2020.
- [30] Ziad M. Ali and Seyed-Ehsan Razavi and Mohammad Sadegh Javadi and Foad H. Gandoman and Shady H.E. Abdel Aleem, "Dual Enhancement of Power System Monitoring: Improved Probabilistic Multi-Stage PMU Placement with an Increased Search Space & Mathematical Linear Expansion to Consider Zero-Injection Bus," Energies, vol. 11, no. 6, pp. 1–17, June 2018.
- [31] Dehghani, Maryam and Rezaei, Mahsa and Shayanfard, Bitia and Vafamand, Navid and Javadi, Mohammad and Catalađ,

- João P. S., "PMU-Based Power System Stabilizer Design: Optimal Signal Selection and Controller Design," *IEEE Transactions on Industry Applications*, vol. 57, no. 6, pp. 5677–5686, 2021.
- [32] Lutz Hofmann, "Unified fault matrix method for the calculation of shunt and series faults in power systems with complete duality of shunt and series fault matrices," *Electric Power Systems Research*, vol. 189, p. 106556, 2020.