electronics

MDPI

Article

A Wavelet-Embedded Residual Attention Convolutional Neural Network for Fault Location in Distribution Networks

Zhengkai Sun [1,*] and Qian Zhang [2]

[1] Sydney Smart Technology College, Northeastern University at Qinhuangdao, Qinhuangdao 066004, China
[2] School of Electrical Engineering and Automation, Hefei University of Technology, Hefei 230009, China; zhangqian@hfut.edu.cn
* Correspondence: 202319068@stu.neuq.edu.cn

Abstract

Accurate fault location is essential for improving the reliability and service restoration capability of distribution networks. With the increasing penetration of distributed generation, power electronic devices, and flexible loads, fault transient signals become increasingly nonlinear and nonstationary, posing challenges to conventional impedance-based, traveling-wave-based, and feature-engineering-based methods. To improve transient fault feature representation, this paper proposes a wavelet-embedded residual attention convolutional neural network (CNN) for distribution network fault location. The task is formulated as a multi-class classification problem, in which each predefined line section is treated as a candidate fault location class. The proposed method embeds discrete wavelet decomposition into the convolutional feature extraction process, enabling low-frequency trend components and high-frequency transient components to be jointly represented and fused by subsequent trainable network modules. Residual connections improve deep feature propagation, and an attention mechanism enhances fault-sensitive representations. Simulation studies on the IEEE 33-bus distribution system show that the proposed method outperforms multi-layer perceptron (MLP), support vector machine (SVM), standard CNN, ResNet, and Attention-CNN, achieving 98.27% accuracy and a 98.33% F1-score. The class-wise results and robustness tests under different transition resistances, noise levels, and fault types further verify the effectiveness and adaptability of the proposed method.

Keywords: distribution network; fault location; wavelet-embedded convolution; convolutional neural network; residual attention; multi-class classification

1. Introduction

Accurate and fast fault location is essential for improving distribution network reliability and service restoration capability. With the increasing penetration of distributed generation, power electronic devices, and flexible loads, modern distribution networks exhibit more complex operating characteristics, such as bidirectional power flow, variable fault current contribution, and changing topology [1–3]. These factors make fault transient characteristics increasingly nonlinear and nonstationary, thereby increasing the difficulty of locating faults rapidly and accurately under diverse operating conditions.

Traditional fault location approaches can be broadly divided into impedance-based and traveling-wave-based categories. Impedance-based techniques estimate the fault distance or section from measured voltage/current quantities and line impedance parameters. Recent studies have enhanced such strategies by combining impedance models with meta-heuristic optimization, direct load-flow calculation, fault current constraints, and special

Academic Editor: Dorin Petreus

Received: 9 May 2026
Revised: 28 June 2026
Accepted: 2 July 2026
Published: 4 July 2026

grounding-network models [4–7]. These methods are attractive because their physical meaning is clear and their implementation cost is relatively low. However, their location accuracy remains sensitive to fault resistance, load uncertainty, distributed generation output, parameter errors, and equivalent-source modeling assumptions [8]. Traveling-wave-based methods utilize the arrival time, polarity, or waveform characteristics of fault-generated traveling waves to locate faults [9]. Time-matrix modeling, wide-scale time-window operators, and wavefront-distortion compensation have recently been investigated to improve traveling-wave fault location in active or multi-branch distribution networks [10–12]. Although traveling-wave-based methods can achieve high location accuracy, they usually require high-frequency measurement devices, reliable wavefront detection, accurate time synchronization, and precise wave velocity estimation, which may limit their practical deployment in complex distribution networks.

In recent years, artificial intelligence-based methods have been widely investigated for fault diagnosis and location in distribution networks. Machine learning and deep learning models can establish nonlinear mappings between fault measurements and fault location labels, thereby reducing the dependence on explicit system modeling. Learning-based strategies have also been increasingly applied to adaptive operation and control in modern power electronic energy systems, such as electric-vehicle-based fast frequency response with deep reinforcement learning [13]. For example, sparse-meter-based location, sparse overcomplete representation with Bayesian learning, and learning-based identification methods have been used for faulted section location under limited measurements or inverter-interfaced distributed generators [14–16]. Deep convolutional neural networks and wavelet scattering networks have also been introduced to learn discriminative features from synchrophasor or transient signals [17–19]. Nevertheless, many existing intelligent methods still rely on manually designed features or external signal preprocessing. In particular, when wavelet transform is used only as a preprocessing tool, multi-scale time-frequency information is separated from the deep feature learning process, which may weaken the end-to-end representation ability and feature fusion capability of the model. This indicates a specific research gap: conventional "wavelet preprocessing + CNN" schemes usually generate wavelet coefficients before network training, and the subsequent CNN only learns features from the preprocessed outputs. As a result, the interaction between multi-scale decomposition and deep feature extraction is limited, and some fault-sensitive transient information may not be sufficiently fused during model learning. Therefore, it is necessary to integrate wavelet decomposition more closely into the feature-learning process so that multi-scale fault information can be represented and fused within the network.

To address these issues, this paper proposes a wavelet-embedded residual attention convolutional neural network for distribution network fault location. The fault location task is formulated as a multi-class classification problem, where each predefined fault section is regarded as one candidate class. Unlike conventional wavelet-CNN schemes, the proposed method embeds discrete wavelet decomposition into the CNN feature extraction process, allowing low-frequency trend components and high-frequency transient components to be preserved within the network pipeline. In this embedded design, wavelet decomposition is no longer treated as an isolated preprocessing step; instead, it becomes part of the feature extraction pipeline, allowing multi-scale components to be jointly processed, fused, and refined by subsequent trainable convolutional representations. This design helps preserve both global trend information and local transient disturbances, thereby improving the discriminative representation of fault sections. Furthermore, residual connections are introduced to improve the stability of deep feature learning, and an attention mechanism is employed to enhance fault-sensitive channels and frequency components. In this way,

the proposed model can extract more discriminative multi-scale transient features for accurate fault location classification.

The remainder of the paper is organized as follows. Section 2 describes the proposed wavelet-embedded residual attention CNN for distribution network fault location, including the wavelet-embedded convolution layer, wavelet residual attention feature extraction network, and fault location classification objective. Section 3 presents the experimental setup, dataset construction, cross-validation strategy, training settings, and evaluation metrics. Section 4 reports the experimental verification results, including fault location performance comparison, computational-effort analysis, and robustness tests. Finally, Section 5 concludes the paper and discusses future work.

2. The Proposed Method

This section presents the wavelet-embedded residual attention convolutional neural network for distribution network fault location. As shown in Figure 1, the proposed method consists of three main components: (1) the wavelet-embedded convolution layer; (2) the wavelet residual attention feature extraction network; and (3) the fault location classification output and optimization objective. The input fault signal is denoted as X, and the network outputs the probability distribution over K candidate fault location classes.

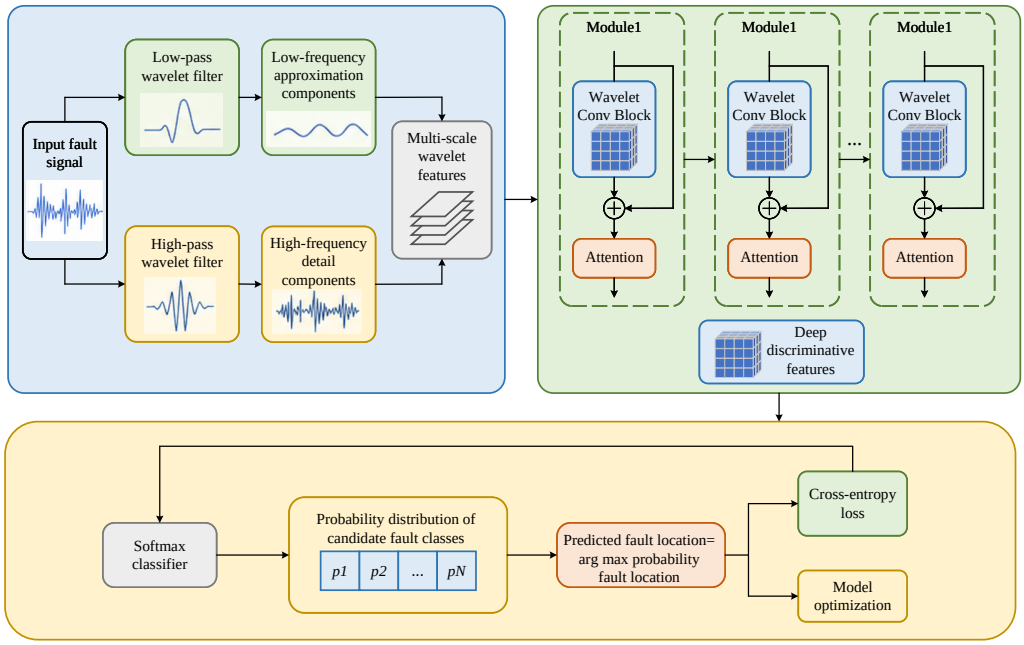

Figure 1. Overall framework of the proposed fault location method.

2.1. Wavelet-Embedded Convolution Layer

Let the input feature of the wavelet-embedded convolution layer be denoted as $A_0 = X$, where X represents the input fault signal or the feature map from the previous layer. Based on discrete wavelet decomposition [20], the input feature is processed by low-pass and high-pass filtering followed by down-sampling. The low-frequency approximation component and high-frequency detail component at the jth decomposition level are calculated as follows:

$$A_j = \downarrow 2(h * A_{j-1}), \tag{1}$$

$$D_j = \downarrow 2(g * A_{j-1}), \tag{2}$$

where A_j denotes the low-frequency approximation component, D_j denotes the high-frequency detail component, h and g are the low-pass and high-pass wavelet filters, re-

spectively, $*$ represents the convolution operation, and $\downarrow 2$ denotes down-sampling by a factor of two. In this study, the wavelet filters h and g are fixed and non-trainable. They are determined by the selected Daubechies 4 (db4) wavelet basis and are not updated during back-propagation. Thus, the wavelet operation provides deterministic multi-scale decomposition, whereas the trainable convolutional, residual, and attention modules perform adaptive feature fusion and representation learning.

The obtained low-frequency and high-frequency components are concatenated along the channel dimension:

$$F_w = Concat(A_j, D_j), \tag{3}$$

where $Concat(\cdot)$ denotes the channel-wise concatenation operation. Then, a trainable convolutional operation is used to fuse the concatenated wavelet features:

$$F_c = \delta(BN(Conv(F_w))), \tag{4}$$

where $Conv(\cdot)$ denotes the convolution operation, $BN(\cdot)$ represents batch normalization, $\delta(\cdot)$ denotes the ReLU activation function, and F_c is the output feature of the wavelet-embedded convolution layer.

2.2. Wavelet Residual Attention Feature Extraction Network

The feature extraction network consists of an initial one-dimensional convolution layer and stacked wavelet residual attention blocks (WRABs). The input fault signal is first mapped into a shallow feature map:

$$F_0 = \delta(BN(Conv1D(X))), \tag{5}$$

where X denotes the input fault signal, $Conv1D(\cdot)$ denotes one-dimensional convolution, $BN(\cdot)$ denotes batch normalization, $\delta(\cdot)$ denotes the ReLU activation function, and F_0 is the initial feature map.

Let the input feature of the lth WRAB be denoted as $F_l^{in} \in \mathbb{R}^{C_l \times T_l}$, where C_l and T_l represent the number of channels and the temporal length, respectively. The wavelet-embedded convolution layer introduced in Section 2.1 is first applied to the input feature:

$$F_l^w = WCL(F_l^{in}), \tag{6}$$

where $WCL(\cdot)$ denotes the wavelet-embedded convolution operation. The wavelet decomposition inside $WCL(\cdot)$ uses fixed db4 filters, whereas the following convolutional operations remain trainable and are optimized together with the whole network.

Next, depthwise separable convolution [21] is applied to F_l^w. The depthwise convolution is performed independently on each input channel:

$$F_l^{dw}(c,t) = \sum_{\tau=1}^{K_s} K_l^{dw}(c,\tau) F_l^w(c, t-\tau), \tag{7}$$

where F_l^{dw} denotes the output of depthwise convolution, K_l^{dw} is the depthwise convolution kernel, K_s is the kernel size, and c denotes the channel index.

Then, pointwise convolution is performed by using a 1×1 convolution kernel:

$$F_l^{pw}(c',t) = \sum_{c=1}^{C_l} K_l^{pw}(c',c) F_l^{dw}(c,t), \tag{8}$$

where F_l^{pw} is the output of pointwise convolution, K_l^{pw} denotes the pointwise convolution kernel, and c' represents the output channel index.

The output of the depthwise separable convolution is then normalized and activated:

$$F_l^c = \delta\Big(BN(F_l^{pw})\Big), \tag{9}$$

where $BN(\cdot)$ denotes batch normalization, and $\delta(\cdot)$ represents the ReLU activation function.

The channel attention operation follows the squeeze-and-excitation structure [22]. For the convolutional feature $F_l^c \in \mathbb{R}^{C_l' \times T_l'}$, global average pooling is first used to obtain the channel-wise descriptor:

$$z_l(c) = \frac{1}{T_l'} \sum_{t=1}^{T_l'} F_l^c(c,t), \tag{10}$$

where $z_l(c)$ represents the global statistical descriptor of the cth channel.

Then, two fully connected layers are applied to generate the channel weight:

$$s_l = \sigma(W_{l,2}\delta(W_{l,1}z_l)), \tag{11}$$

where $W_{l,1}$ and $W_{l,2}$ are learnable weight matrices, $\delta(\cdot)$ denotes the ReLU activation function, and $\sigma(\cdot)$ denotes the Sigmoid activation function. The obtained vector s_l represents the channel attention weight.

The attention-enhanced feature is calculated as follows:

$$F_l^a(c,t) = s_l(c) \cdot F_l^c(c,t), \tag{12}$$

where F_l^a denotes the attention-weighted feature.

The residual branch is defined according to residual learning [23]. The output of the lth WRAB is formulated as follows:

$$F_l^{out} = F_l^a + F_l^s, \tag{13}$$

where F_l^s denotes the residual branch. When the dimensions of the input and output features are the same, the residual branch is directly defined as follows:

$$F_l^s = F_l^{in}. \tag{14}$$

when the dimensions do not match, a projection shortcut with a 1×1 convolution and the corresponding temporal stride is employed for dimension matching:

$$F_l^s = Conv_{1\times1,stride}(F_l^{in}). \tag{15}$$

The complete mapping of the lth WRAB is expressed as follows:

$$F_l^{out} = Attention\Big(DSConv\Big(WCL(F_l^{in})\Big)\Big) + F_l^s, \tag{16}$$

where $DSConv(\cdot)$ denotes the depthwise separable convolution, and $Attention(\cdot)$ denotes the channel attention operation.

Multiple WRABs are then sequentially stacked:

$$F_1 = WRAB_1(F_0), \tag{17}$$

$$F_2 = WRAB_2(F_1), \tag{18}$$

$$F_3 = WRAB_3(F_2). \tag{19}$$

In general, the stacked feature extraction process is written as follows:

$$F_L = WRAB_L(WRAB_{L-1}(\cdots WRAB_1(F_0)\cdots)), \tag{20}$$

where L denotes the number of stacked wavelet residual attention blocks.

Finally, global average pooling is used to convert the final feature map into a feature vector:

$$F = \frac{1}{T_L}\sum_{t=1}^{T_L} F_L(:,t), \tag{21}$$

where F denotes the extracted deep fault feature used for subsequent fault location classification.

Accordingly, the overall feature extraction process is expressed as follows:

$$F = GAP(WRAB_L(\cdots WRAB_2(WRAB_1(\delta(BN(Conv1D(X)))))\cdots)), \tag{22}$$

where $GAP(\cdot)$ denotes global average pooling.

2.3. Fault Location Classification Output and Optimization Objective

Assume that the distribution network is divided into K candidate fault location classes, where each class corresponds to a predefined line section. Based on the feature vector F obtained from the feature extraction network, the classification logits are calculated as follows:

$$O = W_o F + b_o, \tag{23}$$

where $O = [o_1, o_2, \cdots, o_K]$ denotes the output logits of the classification layer, W_o is the weight matrix, and b_o is the bias vector.

The Softmax function converts the logits into the probability distribution of fault location classes:

$$\hat{y}_k = \frac{\exp(o_k)}{\sum_{r=1}^{K}\exp(o_r)}, \quad k = 1, 2, \cdots, K, \tag{24}$$

where $\hat{y}_k$ represents the predicted probability that the fault belongs to the kth candidate location class.

Therefore, the predicted probability vector can be written as follows:

$$\hat{Y} = [\hat{y}_1, \hat{y}_2, \cdots, \hat{y}_K]. \tag{25}$$

The predicted fault location class is determined by the following:

$$\hat{s} = \arg\max_{k\in\{1,2,\cdots,K\}} \hat{y}_k, \tag{26}$$

where $\hat{s}$ denotes the predicted fault location class.

For the true fault location label, one-hot encoding is adopted:

$$Y = [y_1, y_2, \cdots, y_K], \tag{27}$$

where $y_k = 1$ if the input sample belongs to the kth fault location class; otherwise, $y_k = 0$.

The cross-entropy loss for a single input sample is defined as follows:

$$L_{cls} = -\sum_{k=1}^{K} y_k \log(\hat{y}_k). \tag{28}$$

For a mini-batch containing M samples, the average classification loss is expressed as follows:

$$L = \frac{1}{M} \sum_{m=1}^{M} \left[- \sum_{k=1}^{K} y_{m,k} \log(\hat{y}_{m,k}) \right], \tag{29}$$

where $y_{m,k}$ and $\hat{y}_{m,k}$ denote the true label and predicted probability of the mth sample for the kth fault location class, respectively.

The optimization objective of the proposed model can be formulated as follows:

$$\theta^* = \arg\min_{\theta} L(\theta), \tag{30}$$

where θ represents all trainable parameters of the proposed wavelet-embedded residual attention CNN.

3. Experimental Setup

Case studies are conducted on the IEEE 33-bus distribution system, whose topology is shown in Figure 2. The system is a radial distribution feeder with one source bus and 32 load buses, and it is commonly used to verify distribution network protection and fault location methods. Its main feeder and lateral branches provide different electrical distances and branch relationships, making it suitable for examining section-level fault location. In this study, faults are assigned to candidate line sections, and the corresponding transient measurement signals are used as model inputs.

The fault location task is formulated as a multi-class classification problem, where each predefined line section is regarded as one candidate fault location class. Therefore, the output of the model is the faulted section label rather than a continuous distance value, which is consistent with section-level fault isolation and maintenance in distribution networks. This setting also allows the classification result to be directly compared with the actual faulted section, so that both overall accuracy and class-wise location behavior can be analyzed.

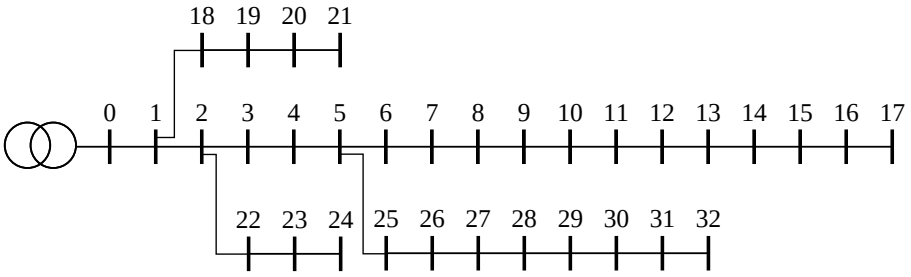

Figure 2. Topology of the IEEE 33-bus distribution system used in the case study.

Before being fed into the network, each input signal is normalized to reduce the influence of amplitude-scale differences among samples. The dataset contains 5000 fault samples. Each sample consists of 33 node signals, and each node signal contains 50 sampling points. The 32 line sections are used as fault location classes. To provide a more complete dataset description, different operating and fault conditions are considered in the simulation. The fault inception angle is varied from 0° to 330° in 30° increments. The load level is set to 0.8, 1.0, and 1.2 times the nominal value to represent light-load, nominal-load, and heavy-load operating conditions, respectively. Distributed generation is modeled as inverter-interfaced sources operating at a fixed power factor, with penetration levels of 0%, 10%, and 20%. The 5000 samples are generated from different combinations of fault

locations, fault types, inception angles, load levels, and distributed-generation conditions, rather than from only one fixed operating condition.

To avoid information leakage across highly similar fault scenarios, the dataset is divided at the scenario level rather than the individual-sample level. Specifically, samples sharing the same fault location, fault type, fault inception angle, load level, and distributed-generation condition are assigned to the same data subset. This prevents nearly identical fault cases from appearing simultaneously in the training and test sets, so that the reported performance better reflects the generalization ability of the model rather than memorization of highly similar samples.

To further evaluate the generalization ability of the proposed model more reliably and reduce the selection bias caused by a single train–test split, stratified 5-fold cross-validation is adopted. In this setting, the whole dataset is divided into five folds while preserving the class distribution of the 32 fault location classes in each fold. The scenario-level grouping described above is also maintained during fold construction. For each validation round, four folds are used for model training, and the remaining fold is used for testing. When validation is required for model selection, it is conducted only within the training folds to avoid any information leakage into the test fold. This process is repeated five times so that each fold is used once as the test set. The final performance is obtained by averaging the results over the five folds.

In the implementation of the proposed wavelet-embedded convolution layer, the Daubechies 4 (db4) wavelet basis is used. The corresponding low-pass and high-pass wavelet filters are fixed and non-trainable during network optimization, rather than being updated by back-propagation. Therefore, the wavelet decomposition provides deterministic multi-scale signal representation, while the subsequent convolutional, residual, and attention modules perform adaptive feature learning and feature fusion. This setting makes the role of the wavelet-embedded layer explicit and clarifies that the performance improvement comes from the joint use of fixed wavelet decomposition and trainable deep feature extraction.

The network is trained using cross-entropy loss, and AdamW is adopted for parameter optimization. To ensure a fair comparison, the same data-processing and evaluation protocol is used for all compared methods. Thus, differences in the final results mainly come from the model structures rather than from inconsistent training settings. The key training parameters of the proposed method are listed in Table 1.

Table 1. Key training parameters of the proposed method.

Parameter	Value
Wavelet basis	Daubechies 4 (db4)
Wavelet filters	Fixed and non-trainable
Number of WRABs	3
Optimizer	AdamW
Loss function	Cross-entropy loss
Initial learning rate	0.001
Batch size	64
Number of epochs	100
Activation function	Rectified linear unit (ReLU)
Output function	Softmax

The performance is evaluated by accuracy, precision, recall, and F1-score. Accuracy reflects the overall proportion of correctly located samples, while precision, recall, and F1-score describe the class-wise recognition quality from different perspectives. Since the fault location task involves multiple candidate sections, relying only on accuracy may obscure

uneven performance among classes. Therefore, precision and recall are calculated in a one-vs-rest manner and then averaged over all classes:

$$Accuracy = \frac{N_c}{N}, \tag{31}$$

$$Precision = \frac{1}{K}\sum_{k=1}^{K}\frac{TP_k}{TP_k + FP_k}, \tag{32}$$

$$Recall = \frac{1}{K}\sum_{k=1}^{K}\frac{TP_k}{TP_k + FN_k}, \tag{33}$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall}, \tag{34}$$

where N_c is the number of correctly located samples, N is the total number of test samples, K is the number of fault location classes, and TP_k, FP_k, and FN_k denote true positives, false positives, and false negatives of the kth class, respectively. For the stratified 5-fold cross-validation, these metrics are first calculated on the test fold of each round and then averaged over the five rounds to obtain the final reported performance.

4. Experimental Verification and Discussion

4.1. Fault Location Performance and Comparison

This subsection first examines the class-wise location behavior of the proposed method and then compares it with MLP, SVM, CNN, ResNet, and Attention-CNN. MLP denotes multi-layer perceptron, SVM denotes support vector machine, and CNN denotes convolutional neural network. MLP and SVM are used as basic classifiers, while CNN, ResNet, and Attention-CNN represent standard convolutional feature learning, residual feature propagation, and attention-based feature enhancement, respectively. These baselines were selected because they are representative and widely used in fault location studies: MLP and SVM serve as classical machine learning baselines; CNN is the basic deep learning baseline; ResNet is used to examine the effect of residual learning; and Attention-CNN is used to assess the effect of attention enhancement. Together, these methods provide a fair and systematic comparison across the main modeling paradigms and allow us to isolate the contribution of each component of the proposed framework under the same data representation and training protocol. The comparison is intended to evaluate whether the joint use of wavelet embedding, residual learning, and attention enhancement improves fault location performance over these representative baseline structures.

The class-wise correct location rate is illustrated in Figure 3. Across the five-fold evaluation, the proposed method achieves an average test accuracy of 98.27% and a class-averaged correct location rate of 98.29%. In the representative test fold used for class-wise visualization, most candidate sections are located with very high accuracy. The lowest class-wise rate appears at F11, where the correct location rate is 90.91%. The remaining non-perfect sections, such as F9, F10, and F16, still maintain correct location rates above 91%. These results show that the proposed model does not rely only on several easily identified sections, but preserves high recognition ability across most candidate fault location classes.

To further examine the location error pattern, Figure 4 gives the misclassification distribution of each class in the representative test fold. The stacked bars show only the misclassification categories with nonzero values, including previous-section errors and other-section errors, while the black curve denotes the total error rate. The misclassified samples are mainly concentrated in a small number of sections, and the maximum class-wise error rate is 9.09% at F11. Most wrong predictions are assigned to the previous adjacent section, whereas non-neighboring errors occur only rarely. In this representative test fold,

no next-section or second-neighbor-section errors are observed. This error pattern indicates that the proposed method seldom produces large location deviations, which is useful for practical fault isolation because localized errors are easier to inspect and correct.

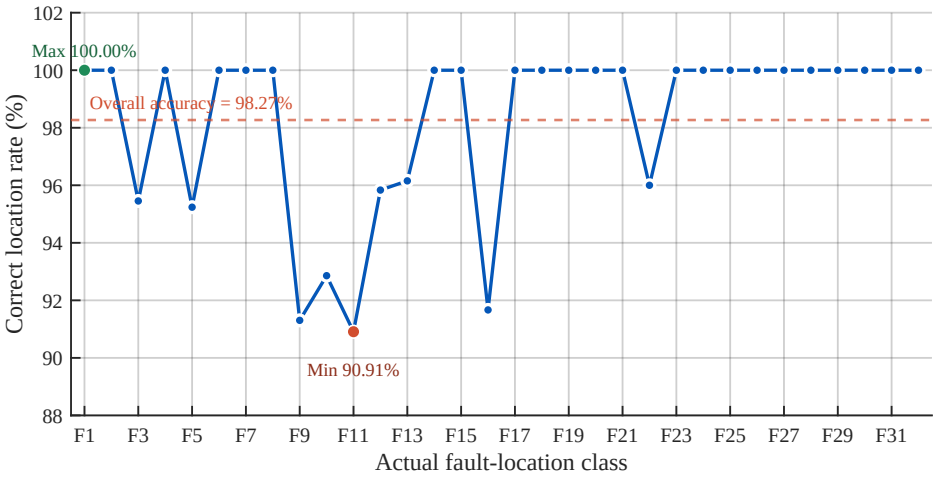

Figure 3. Class-wise correct location rate of the proposed method.

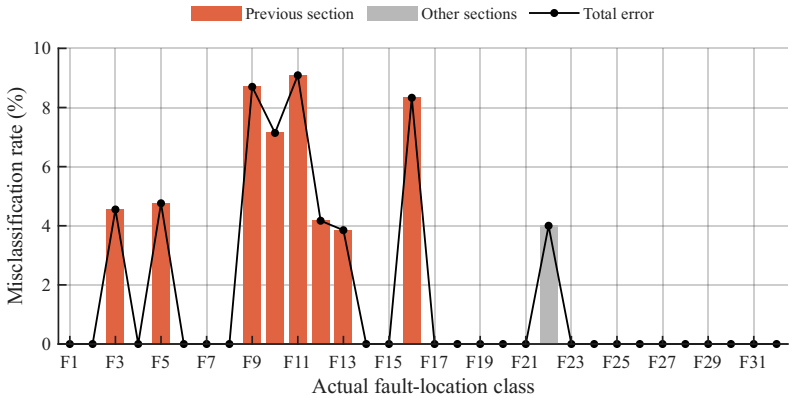

Figure 4. Class-wise misclassification distribution of the proposed method.

The comparison results are summarized in Table 2. MLP is used as a basic nonlinear classifier, and SVM is used as a traditional machine learning baseline. CNN directly uses standard convolution and serves as the basic deep learning benchmark. ResNet introduces residual connections to improve feature propagation, whereas Attention-CNN adds an attention module to emphasize informative features. These methods provide a progressive comparison for evaluating the contribution of the proposed integrated structure.

Table 2. Comparison results of different fault location methods.

Method	Accuracy	Precision	Recall	F1-Score
MLP	90.68%	92.91%	90.51%	91.70%
SVM	94.94%	95.67%	94.92%	95.29%
CNN	96.80%	97.19%	96.86%	97.02%
ResNet	97.60%	97.89%	97.63%	97.76%
Attention-CNN	97.07%	97.36%	97.09%	97.23%
Proposed method	98.27%	98.37%	98.29%	98.33%

Among all compared methods, the proposed method obtains the highest accuracy, precision, recall, and F1-score. Its accuracy is 7.59, 3.33, 1.46, 0.67, and 1.20 percentage points

higher than those of MLP, SVM, CNN, ResNet, and Attention-CNN, respectively. The large gains over MLP and SVM indicate that direct classification based on flattened or pooled features is insufficient to fully capture the spatial–temporal fault patterns. The improvement over CNN shows the benefit of multi-scale wavelet feature extraction. In addition, the gains over ResNet and Attention-CNN suggest that residual propagation and attention weighting become more effective when they are combined with wavelet-based multi-scale representations. Although the numerical accuracy improvement over ResNet and Attention-CNN is moderate, the advantage of the proposed method is not limited to the overall fault-classification metric. Since the classification labels correspond to physical line sections, improved classification performance directly contributes to more reliable section-level fault location. Moreover, the proposed method provides more balanced class-wise recognition, reduces large location deviations, and maintains stable performance under different transition resistances, measurement noise levels, and fault types. Therefore, the proposed model improves not only the overall correctness but also the practical reliability of fault section identification.

To further evaluate the computational effort of different methods, the average inference time was measured on the same evaluation samples under the same hardware environment and batch-size setting. The results are listed in Table 3.

Table 3. Average inference time comparison of different fault location methods.

Method	Average Inference Time
MLP	0.17 ms/sample
SVM	0.29 ms/sample
CNN	0.61 ms/sample
ResNet	0.78 ms/sample
Attention-CNN	0.93 ms/sample
Proposed method	1.05 ms/sample

As shown in Table 3, the proposed method requires a slightly longer inference time than the compared baselines because the wavelet-embedded feature extraction, residual propagation, and attention enhancement introduce additional computational operations. However, the average inference time is still only 1.05 ms/sample, which remains within an acceptable range for online fault location applications. Compared with ResNet and Attention-CNN, the proposed method introduces only modest computational overhead while achieving higher accuracy, more balanced class-wise performance, and stronger robustness. Therefore, the proposed method provides a favorable trade-off between location performance and computational cost. Overall, these results confirm that the proposed method improves both overall correctness and class-wise recognition balance.

4.2. Ablation Study

To further verify the contribution of each component in the proposed framework, an ablation study is conducted. Four model variants are compared: CNN, CNN + Wavelet, CNN + Wavelet + Residual, and CNN + Wavelet + Residual + Attention. The CNN variant is the basic convolutional baseline and is kept identical to the CNN used in the comparison experiment. The CNN + Wavelet variant introduces the wavelet-embedded feature extraction module to evaluate the effect of multi-scale transient representation. The CNN + Wavelet + Residual variant further adds residual connections to examine the benefit of improved feature propagation. Finally, CNN + Wavelet + Residual + Attention corresponds to the complete proposed model, in which the attention mechanism is used to enhance fault-sensitive representations. The ablation results are summarized in Table 4.

Table 4. Ablation study results of different model variants.

Model	Accuracy	Precision	Recall	F1-Score
CNN	96.80%	97.19%	96.86%	97.02%
CNN + Wavelet	97.42%	97.61%	97.35%	97.48%
CNN + Wavelet + Residual	97.98%	98.05%	97.96%	98.00%
CNN + Wavelet + Residual + Attention	98.27%	98.37%	98.29%	98.33%

As shown in Table 4, the performance improves progressively as each module is added. Compared with the basic CNN, CNN + Wavelet improves the accuracy from 96.80% to 97.42%, indicating that the wavelet-embedded module can enrich multi-scale transient features and improve fault location representation. After adding residual connections, the accuracy further increases to 97.98%, which shows that residual learning helps improve feature propagation and stabilize deeper feature extraction. When the attention mechanism is further introduced, the complete model achieves the best performance, with 98.27% accuracy and a 98.33% F1-score. This confirms that the attention mechanism can further enhance fault-sensitive information and improve classification reliability. Overall, the ablation study demonstrates that wavelet embedding, residual learning, and attention enhancement all contribute positively to the final fault location performance.

4.3. Robustness Analysis

The robustness of the proposed method is evaluated under different transition resistances, noise levels, and fault types. These factors are selected because they directly affect the waveform amplitude, transient components, and phase relationships of fault signals. In practical distribution networks, such variations are difficult to avoid, and a practical fault location model should maintain stable performance when signal characteristics change. Therefore, the following tests provide a more comprehensive evaluation of the adaptability of the proposed method.

4.3.1. Influence of Transition Resistance

The influence of transition resistance is investigated first, and the corresponding results are listed in Table 5.

Table 5. Fault location performance under different transition resistances.

Transition Resistance	Accuracy	Precision	Recall	F1-Score
0.01 Ω	98.09%	97.36%	98.64%	98.00%
10 Ω	97.58%	98.21%	96.79%	97.49%
50 Ω	96.66%	95.82%	95.41%	94.61%
100 Ω	91.84%	90.39%	92.88%	90.63%

With the increase in transition resistance, the accuracy decreases from 98.09% at 0.01 Ω to 91.84% at 100 Ω. This trend is expected because a larger transition resistance weakens the fault current and reduces the distinction between different fault sections. In particular, when the transition resistance increases to 100 Ω, the fault current amplitude becomes significantly smaller, and the transient signatures associated with different line sections become less distinguishable. As a result, the feature separability among adjacent fault sections is reduced, which makes the classification task more difficult and leads to a more obvious performance drop. Even under the highest tested resistance, however, the accuracy remains above 90%. Compared with the 0.01 Ω case, the accuracy drop at 100 Ω is 6.25 percentage points. Meanwhile, the F1-score remains above 90.63% in all tested resistance cases, indicating that the proposed method can still preserve effective location

features under high-resistance faults. Nevertheless, high-resistance faults remain more challenging than low-resistance faults because their fault-induced transients are weaker and more easily affected by load variation and measurement noise. Therefore, additional feature enhancement, high-resistance-fault-oriented data augmentation, or adaptive sample reweighting may further improve the location performance under high-resistance fault conditions, which will be considered in future work.

4.3.2. Influence of Measurement Noise

The influence of measurement noise is evaluated by varying the signal-to-noise ratio (SNR), and the results are given in Table 6.

Table 6. Fault location performance under different noise levels.

SNR	Accuracy	Precision	Recall	F1-Score
40 dB	97.96%	98.23%	97.13%	97.84%
30 dB	97.06%	96.38%	97.92%	97.14%
20 dB	95.76%	96.71%	94.76%	95.73%
10 dB	93.86%	92.94%	95.03%	93.97%

When the SNR decreases from 40 dB to 10 dB, the accuracy decreases from 97.96% to 93.86%. The degradation is gradual rather than abrupt, indicating that the model remains stable as measurement interference increases. At 10 dB, the input signal is strongly disturbed, but the F1-score still reaches 93.97%. This result suggests that the proposed model has a degree of noise tolerance, which can be attributed to the combined use of low-frequency trend information and high-frequency transient information.

4.3.3. Influence of Fault Type

Finally, the adaptability of the proposed method to different fault categories is examined. Table 7 lists the results for single-phase-to-ground, phase-to-phase, two-phase-to-ground, and three-phase faults.

Table 7. Fault location performance under different fault types.

Fault Type	Accuracy	Precision	Recall	F1-Score
Single-phase-to-ground	98.31%	98.42%	98.36%	98.39%
Phase-to-phase	98.18%	98.25%	98.12%	98.19%
Two-phase-to-ground	98.07%	98.18%	98.04%	98.11%
Three-phase	98.52%	98.63%	98.64%	98.63%

For the four fault types, the accuracy ranges from 98.07% to 98.52%. Among them, the two-phase-to-ground fault yields the lowest accuracy, while the three-phase fault yields the highest accuracy. This result is consistent with the fact that different fault types produce different transient characteristics and phase couplings. Three-phase faults usually have more pronounced and balanced waveform changes, whereas grounding-related faults may exhibit more asymmetric phase responses and weaker transient differences among phases. Even so, the F1-score remains above 98.11% for all fault categories, showing that the proposed method maintains stable performance when the fault type changes.

5. Conclusions

In this paper, a wavelet-embedded residual attention convolutional neural network has been proposed for fault location in distribution networks. The fault location task is

formulated as a multi-class classification problem, in which each predefined line section corresponds to one candidate class. By embedding fixed discrete wavelet decomposition into the convolutional feature extraction process, the proposed method obtains low-frequency trend information and high-frequency transient information, which are then adaptively fused by trainable convolutional, residual, and attention modules. Residual connections and the attention mechanism are further introduced to improve feature propagation and enhance fault-sensitive representations. Simulation experiments on the IEEE 33-bus distribution system show that the proposed method outperforms representative classical machine learning and deep learning baselines, achieving an average accuracy of 98.27% and an average F1-score of 98.33%. The comparison and ablation results provide a fair and representative evaluation of the proposed method, while the expanded discussion further analyzes the experimental results and their practical implications. The class-wise results are also stable, with most misclassifications limited to adjacent sections rather than large location deviations. In addition, the robustness tests under different transition resistances, noise levels, and fault types indicate that the proposed method has good adaptability to varying fault conditions. The present study is primarily evaluated on the IEEE 33-bus distribution system, and more field-measured data from practical distribution networks should be incorporated in future work. The effects of topology changes, distributed-generation operating modes, load variations, and measurement errors also warrant further analysis. Future research will focus on improving the generalization ability, computational efficiency, and lightweight deployment of the proposed method for practical online fault location.

Author Contributions: Conceptualization, Z.S. and Q.Z.; methodology, Z.S.; software, Z.S.; validation, Z.S. and Q.Z.; formal analysis, Z.S.; investigation, Z.S.; resources, Q.Z.; data curation, Z.S.; writing—original draft preparation, Z.S.; writing—review and editing, Z.S. and Q.Z.; visualization, Z.S.; supervision, Q.Z.; project administration, Q.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to privacy and confidentiality restrictions.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Liu, Q.; Jia, K.; Yang, B.; Zheng, L.; Bi, T. Fault Analysis of Inverter-Interfaced RESs Considering Decoupled Sequence Control. *IEEE Trans. Ind. Electron.* **2023**, *70*, 4820–4830. [CrossRef]
2. Liu, Q.; Jia, K.; Yang, B.; Zheng, L.; Bi, T. Analytical Model of Inverter-Interfaced Renewable Energy Sources for Power System Protection. *IEEE Trans. Power Deliv.* **2023**, *38*, 1064–1073. [CrossRef]
3. Zheng, X.; Chao, C.; Weng, Y.; Ye, H.; Liu, Z.; Gao, P.; Tai, N. High-Frequency Fault Analysis-Based Pilot Protection Scheme for a Distribution Network with High Photovoltaic Penetration. *IEEE Trans. Smart Grid* **2023**, *14*, 302–314. [CrossRef]
4. Pessoa, A.L.d.S.; Oleskovicz, M. Fault Location Algorithm for Distribution Systems with Distributed Generation Based on Impedance and Metaheuristic Methods. *Electr. Power Syst. Res.* **2023**, *225*, 109871. [CrossRef]
5. Arsoniadis, C.G.; Nikolaidis, V.C. Fault Location Method for Overhead Feeders with Distributed Generation Units Based on Direct Load Flow Approach. *J. Mod. Power Syst. Clean Energy* **2024**, *12*, 1135–1146. [CrossRef]
6. Yang, W.J.; Yin, X.Q.; Tao, J.; Zhang, H.Y. Fault Current Constrained Impedance-Based Method for High Resistance Ground Fault Location in Distribution Grid. *Electr. Power Syst. Res.* **2024**, *227*, 109998. [CrossRef]
7. Pang, Q.; Wang, Y.; Wang, Y.; Cao, T. Earth Fault Location for Non-Directly Grounded Distribution Networks. *IEEE Trans. Power Deliv.* **2024**, *39*, 706–717. [CrossRef]
8. Wei, M.; Liu, W.; Shi, F.; Zhang, H.; Jin, Z.; Chen, W. Distortion-Controllable Arc Modeling for High Impedance Arc Fault in the Distribution Network. *IEEE Trans. Power Deliv.* **2021**, *36*, 52–63. [CrossRef]
9. Chen, F.X.; Guo, M.F.; Lin, J.; Zheng, Y.L.; Zeng, X.K.; Hong, Q. Single-Ended Traveling Wave Location for Single-Phase Ground Faults in Distribution Networks Based on Hough Transform. *IEEE Trans. Instrum. Meas.* **2025**, *74*, 9005916. [CrossRef]

10. Cheng, L.; Wang, T.; Wang, Y. A Novel Fault Location Method for Distribution Networks with Distributed Generations Based on the Time Matrix of Traveling-Waves. *Prot. Control Mod. Power Syst.* **2022**, *7*, 46. [CrossRef]
11. Xia, Y.; Li, Z.; Xi, Y.; Feng, Y.; Wu, G.; Liu, G. Distribution Network Fault Location Method Based on Wide Scale Time Window Difference Operator. *IEEE Trans. Ind. Inform.* **2024**, *20*, 3446–3455. [CrossRef]
12. Wang, Y.; Xie, L.; Liu, F.; Yu, K.; Zeng, X.; Bi, L.; Tang, X. Fault Location Method for Distribution Network Considering Distortion of Traveling Wavefronts. *Int. J. Electr. Power Energy Syst.* **2024**, *159*, 110065. [CrossRef]
13. Wan, Y.; Wang, N.; Liu, X.; Wang, Y.; Blaabjerg, F.; Chen, Z. Inertia-Emulation-Based Fast Frequency Response from EVs: A Multi-Level Framework with Game-Theoretic Incentives and DRL. *IEEE Trans. Smart Grid* **2025**, *16*, 5353–5364. [CrossRef]
14. Yang, B.; Jia, K.; Liu, Q.; Zheng, L.; Bi, T. Faulted Line-Section Location in Distribution System with Inverter-Interfaced DGs Using Sparse Meters. *IEEE Trans. Smart Grid* **2023**, *14*, 413–423. [CrossRef]
15. Shan, H.; Wu, Q.H.; Li, C.; Zhang, L. Sparse Overcomplete Representation Fault Location Model in Distribution Networks and Efficient Solution Using FastLaplace Bayesian. *IEEE Trans. Instrum. Meas.* **2024**, *73*, 3517311. [CrossRef]
16. Chhetija, D.; Rather, Z.H.; Doolla, S. Fault Location Identification in Power Islands with Inverter Interfaced Distributed Generators. *IEEE Trans. Instrum. Meas.* **2024**, *73*, 9005310. [CrossRef]
17. Siddique, M.N.I.; Shafiullah, M.; Mekhilef, S.; Pota, H.; Abido, M.A. Fault Classification and Location of a PMU-Equipped Active Distribution Network Using Deep Convolution Neural Network (CNN). *Electr. Power Syst. Res.* **2024**, *229*, 110178. [CrossRef]
18. Yildiz, T.; Abur, A. Convolutional Neural Network-Assisted Fault Detection and Location Using Few PMUs. *Electr. Power Syst. Res.* **2024**, *235*, 110705. [CrossRef]
19. Arsoniadis, C.G.; Nikolaidis, V.C. A Machine Learning Based Fault Location Method for Power Distribution Systems Using Wavelet Scattering Networks. *Sustain. Energy Grids Netw.* **2024**, *40*, 101551. [CrossRef]
20. Mallat, S.G. A Theory for Multiresolution Signal Decomposition: The Wavelet Representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **1989**, *11*, 674–693. [CrossRef]
21. Chollet, F. Xception: Deep Learning with Depthwise Separable Convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1251–1258. [CrossRef]
22. Hu, J.; Shen, L.; Sun, G. Squeeze-and-Excitation Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7132–7141. [CrossRef]
23. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778. [CrossRef]