Academic Journal of Science and Technology
ISSN: 2771-3032 | Vol. 23, No. 3, 2026

Single-Terminal Fault Location Method for HVDC Transmission Lines Based on Propagation-Aligned Interpolation Augmentation and Time–Frequency Gated Fusion

Tianyu Li[1, *], Ruixi Yang[1]

[1]Sichuan University Jinjiang College, Meishan, China
[*] Corresponding author

Abstract: Aiming at the difficulty of accurately characterizing the mapping between single-terminal fault transients and fault distance in HVDC transmission lines, this paper proposes a fault location method based on propagation-aligned interpolation augmentation and a time–frequency gated fusion network, combining the propagation-delay differences induced by distance variation with the time–frequency features of transients. To alleviate the insufficient distance coverage of training samples and the inconsistent wavefront arrival times caused by direct interpolation, adjacent waveforms with the same fault type and transition resistance are aligned and interpolated according to propagation delays calibrated on the training set, and an augmented training set is constructed together with end-zone densification and source quality screening. To promote time–frequency feature fusion, a dual-branch time–frequency gated fusion network is built, in which a CNN–BiLSTM branch and a Conv1D–SE branch extract transient temporal features and S-transform magnitude-spectrum features respectively, and feature-wise gated fusion then performs fault distance regression. Experimental results show that the proposed method achieves a mean absolute error of 0.285 km and a P95 error of 0.752 km, reducing the mean absolute error by about 48.6% compared with the equal-budget linear interpolation augmentation configuration. The proposed method helps improve single-terminal fault location accuracy under the experimental conditions of this paper.

Keywords: HVDC transmission; fault location; propagation-aligned interpolation augmentation; time–frequency feature fusion; gated network.

1. Introduction

High-voltage direct current (HVDC) transmission features large delivery capacity and long transmission distance, and rapid fault location after line faults provides a positional basis for patrol inspection and maintenance. Fault location is generally realized with transient traveling waves, line electrical models, or the mapping between fault signals and fault positions [1]. Among these approaches, single-terminal location requires only local voltage and current measurements and depends on neither two-terminal communication nor clock synchronization, which offers good application convenience, but it must identify the fault position from limited measurement information. How to make full use of single-terminal transient signals to obtain accurate distance estimates under different fault positions and transition resistances is a problem that such methods need to solve.

Traditional fault location mainly establishes the position relationship from line models and traveling-wave propagation laws. Suonan et al. [2] estimated the fault position using two-terminal voltages, currents, and a distributed-parameter model, providing a route for electrical-quantity-based DC line location. In comparison, single-terminal traveling-wave methods locate faults through propagation delays in local transient signals, and their accuracy depends on wave velocity estimation and characteristic wavefront identification. Regarding these issues, Dardengo et al. [3] studied propagation velocity estimation in single-terminal location; Huai et al. [4] combined optimized variational mode decomposition with the Teager energy operator to extract fault traveling waves in multi-terminal DC systems; and Reis et al. [5] improved traveling-wave delay estimation using local correlation in AC lines. These studies improve the location process in terms of propagation parameters and signal processing, and also indicate that wave velocity, reflected-wave discrimination, and wavefront extraction jointly affect location results.

With the development of signal processing and deep learning, data-driven location has become another research path. Such methods learn the mapping between fault signals and distance, reducing the dependence on explicit wavefront discrimination rules. Hao et al. [6] applied Hilbert–Huang transform features to VSC-HVDC fault location; Wang et al. [7] combined VMD-TEO, convolutional section classification, and LSTM distance regression for single-ended hybrid DC line location; Inwumoh et al. [8] built a BiLSTM method for fault type and position estimation using rectifier-side data. Beyond time-domain sequence modeling, Wang et al. [9] adopted wavelet-packet energy spectra with an SSA-optimized GRU to extract fault features, showing that time–frequency representations can also establish distance mappings. Related studies have further extended to two-terminal information fusion [10] and pre-classification of field-recorded traveling waves [11], reflecting the roles of input information, feature representation, and recording quality in location models. In addition, Li and He [12] established a position relationship according to the phase–frequency variation of fault traveling waves, providing another idea for single-terminal location using frequency-domain features.

The above studies lay a foundation for data-driven location,

but the distance estimation capability of a model remains closely related to training sample coverage and feature utilization. On the one hand, fault positions are continuously distributed along the line, whereas training samples usually cover only a limited number of distances, so the model must learn a continuous distance mapping from waveforms at discrete positions. When samples between adjacent training positions are lacking, the fitting accuracy at available positions can hardly directly reflect the prediction capability for intermediate positions. Increasing simulation positions extends coverage but also increases the cost of data generation. It is therefore necessary to study methods that construct intermediate-distance samples from existing same-condition waveforms, and to separately examine the fidelity of the derived waveforms and their location benefit.

On the other hand, single-terminal transient waveforms contain features such as wavefront arrival, subsequent oscillations, and spectral distribution. Time-domain sequences and time–frequency representations describe these variations from different perspectives, but their roles in distance regression need to be coordinated through a fusion structure. Direct concatenation preserves branch features yet lacks explicit regulation of the relative contributions of the two representations. Therefore, while extending training distance coverage, it is necessary to design a location-oriented time–frequency fusion mechanism so that both branches jointly participate in distance prediction.

Based on the above considerations, this paper studies fault location under rectifier-side single-terminal recording from two aspects: training sample construction and time–frequency feature fusion:

1) An augmentation scheme combining propagation-aligned interpolation with source quality screening is constructed. According to propagation delays calibrated on the training set, adjacent training waveforms with the same fault type and transition resistance are aligned to the target distance and then interpolated; together with end-zone densification and source quality screening, this forms an augmented training set with denser distance coverage. The effects of propagation alignment and quality screening are evaluated separately through fidelity comparison against real waveforms and an equal-budget comparison with a fixed network structure.

2) A dual-branch time–frequency gated fusion network is constructed. CNN–BiLSTM and Conv1D–SE are used to extract transient temporal features and S-transform magnitude-spectrum features respectively, and feature-wise gated fusion performs fault distance regression. The location performance of single-branch, concatenation fusion, and gated fusion is compared under the same augmentation condition.

2. Simulation System and Dataset

2.1. Simulation Model and Fault Settings

A bipolar DC transmission system with a rated voltage of ±500 kV and a line length of 500 km is built on the PSCAD/EMTDC platform. Both ends employ 12-pulse converters equipped with smoothing reactors and DC filters, and the transmission line adopts the frequency-dependent phase model. The rectifier-side line terminal is taken as the origin of distance: the fault point is d from this terminal, and the two line sections on either side are d and 500−d km respectively. Positive- and negative-pole voltages and currents are measured on the line side downstream of the smoothing reactor in the rectifier station, forming the input signals for single-terminal location. The system structure and measurement positions are shown in Fig. 1.

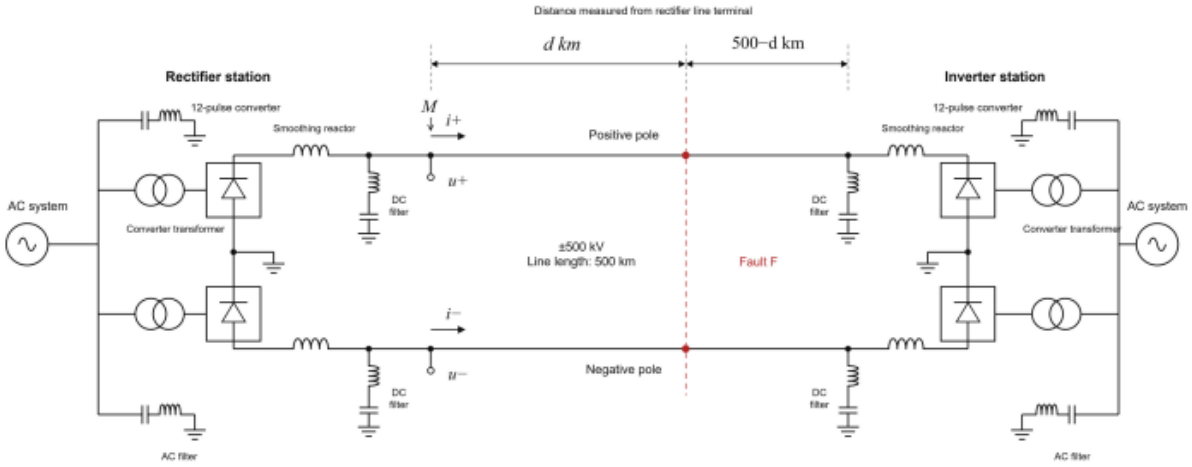


Figure 1. Structure of the bipolar HVDC transmission system

Fault conditions cover three fault types—bipolar short circuit (BIP), positive-pole-to-ground (POS), and negative-pole-to-ground (NEG)—with transition resistances of 0.1, 10, 50, 100, 200, 300, and 500 Ω to examine the location performance under different fault conditions. Faults are triggered at 0.5 s, and 300-point waveforms of 3 ms duration after triggering are captured at a sampling rate of 100 kHz. All samples use the unified fault-inception time as the window start, so the arrival position of the wavefront in the window varies with fault distance, preserving timing information for distance estimation.

2.2. Karrenbauer Modal Decomposition

The positive and negative poles of a bipolar DC line are electromagnetically coupled, so that fault transient voltages and currents interact. The Karrenbauer transform decomposes the two-pole electrical quantities into zero-mode and line-mode components, decoupling the interacting quantities and facilitating the extraction of fault transient features in each mode. The transform is:

$$\begin{bmatrix} u_0 \\ u_1 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} u_+ \\ u_- \end{bmatrix}, \qquad \begin{bmatrix} i_0 \\ i_1 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} i_+ \\ i_- \end{bmatrix} \quad (1)$$

The transform yields the zero-mode voltage and current (u_0,i_0) and the line-mode voltage and current (u_1,i_1). The time-domain input order is fixed as $[u_1,i_1,u_0,i_0]$, with the current reference direction following the original engineering measurement convention. The four components are arranged in the order $[u_1,i_1,u_0,i_0]$ to form a 300×4 modal waveform matrix as the input of the time-domain branch, while the line-mode voltage and current are used to construct the logarithmic magnitude spectrum of the spectral branch.

The dimensions and amplitudes differ across channels, so each channel is z-score normalized to reduce the influence of these differences on network training. After per-channel z-score normalization, each sample forms the four-channel modal waveform $X \in R^{300\times4}$ as the time-domain modal input:

$$\tilde{x}^{(c)} = \frac{x^{(c)} - \mu^{(c)}}{\sigma^{(c)}}, \quad c \in \{u_1,i_1,u_0,i_0\} \quad (2)$$

where $\mu^{(c)}$ and $\sigma^{(c)}$ are the mean and standard deviation of all samples and sampling points of the real training set on channel c; the validation set, test set, and augmented samples uniformly reuse these statistics.

2.3. Dataset Partition and Test Distance Settings

The original simulation samples are partitioned into a training set, a validation set, and a test set; the dataset composition is given in Table 1. The training set covers 115 fault distances, with the two line ends 1–10 km and 490–499 km arranged at 1 km intervals and the middle section 15–485 km at 5 km intervals. Each distance includes 3 fault types and 7 transition-resistance levels, forming 2415 training samples in total.

Table 1. Dataset composition

Dataset	Samples	Distance points	Distance arrangement	Purpose
Original simulation training set	2415	115	1 km intervals at both ends, 5 km in the middle	Model training
Interpolation augmentation set	8442	402	1 km interpolation in the middle, 0.5 km densification at the ends	Model training
Validation set	504	24	10 on-grid and 14 off-grid points	Model selection and early stopping
Test set	504	24	All outside the original training grid	Location performance evaluation

Between the original training distances, same-condition waveform interpolation supplements intermediate positions, with further densification at the line ends, constructing 8442 candidate augmented samples and extending the distance coverage. Merged with the 2415 original simulation samples, this forms an augmented training set of 10857 samples; the construction method is described in Section 2. The validation and test sets each contain 24 fault distances and 504 samples. In the validation set, 10 distances coincide with the training distances and the other 14 lie outside the original training grid, corresponding to 294 samples; the 24 test distances all lie outside the original training grid and are disjoint from the validation distances. The validation set is used for model selection and early stopping, and the test set evaluates the location performance of the fixed model.

3. Interpolation Augmentation Based on Propagation Alignment and Source Quality Screening

3.1. Same-Condition Augmentation Targets and End-Zone Densification

Original training samples cover only a limited number of fault positions. To increase the samples between adjacent training positions, this paper draws on the convex-combination idea of mixup [13], selecting adjacent waveforms with the same fault type and transition resistance and constructing intermediate-position samples through distance weighting.

Let the adjacent training positions be d_1 and d_2, with the corresponding four-channel normalized modal waveforms $w(d_1)$ and $w(d_2)$. The derived waveform and its distance label are:

$$w_{aug} = (1-\alpha)\, w(d_1) + \alpha\, w(d_2), \quad d_{aug} = d_1 + \alpha(d_2 - d_1) \quad (3)$$

where $\alpha = (d_{aug} - d_1)/(d_2 - d_1)$, and $0 < \alpha < 1$. The interpolation is performed pointwise at the same sampling time on the same modal channel, the four channels share the same weights, and the derived samples inherit the common fault type and transition-resistance labels of the two sources.

The propagation delay is calibrated from real training waveforms. For each waveform, the per-channel first-order difference energy is computed, and the sampling position where the cumulative energy reaches 2% of the total is taken as the arrival-time descriptor; the descriptors of the line-mode voltage and current are summarized over distances and their medians taken. The delay slope s is estimated as the median of pairwise slopes between different distances, and the intercept b as the median of the residuals. On the full training set, s = 0.33404 samples/km and b = 0.33191 samples, corresponding to an effective propagation velocity of about 2.99×10^5 km/s.

The two source waveforms are shifted and then mixed by distance weights, computed as

$$\tilde{w}_1 = S(w(d_1), (d_{aug} - d_1)s), \tilde{w}_2 = S(w(d_2), (d_{aug} - d_2)s) \quad (4)$$

$$w_{aug} = (1 - \alpha)\,\tilde{w}_1 + \alpha\,\tilde{w}_2 \tag{5}$$

where $S(\cdot,\delta)$ denotes a shift of δ samples along the time axis, with positive values indicating delay. The four modal channels use the same displacement; non-integer displacements are implemented by linear resampling, and values outside the window boundary are replicated from the endpoints. Taking the target distance as the time reference, this operation alleviates the misalignment of the arrival times of the two-source transients.

Augmentation targets follow a unified grid: within each adjacent 5 km interval, $\alpha \in \{0.2, 0.4, 0.6, 0.8\}$ is taken, forming 8064 targets; within the eighteen 1 km intervals of 1–10 km and 490–499 km, a 0.5 km midpoint is inserted in each, forming 378 end-zone targets, for a total of 8442. Derived waveforms inherit the common fault type and resistance of the sources, and the distance is determined by the interpolation weights. During training, the weight of real samples is 1 and that of derived samples is 0.5.

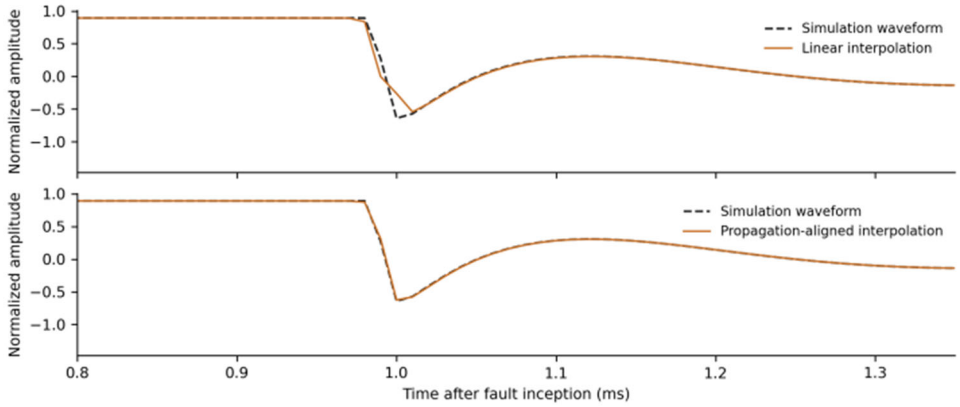


Figure 2. Comparison of the line-mode voltage near the wavefront for a 297 km POS fault at 10 Ω

3.2. Augmentation Quality Screening Based on Source Consistency

Propagation alignment reduces the arrival-time difference between adjacent waveforms, but the transient shape and amplitude differences between the two sources still affect the interpolation results. Therefore, before generating augmented samples, source pairs with pointwise-identical inputs at different distances are first excluded, and the consistency of the remaining sources is evaluated from three aspects: shape, amplitude, and timing.

Source-consistency comparison adopts a unified time reference: the two source waveforms are first shifted to the distance midpoint, and a neighborhood of 16 samples before and after the predicted arrival time at that position is extracted. After subtracting the first recorded value of each record, the four-channel waveforms in the comparison neighborhood are denoted v_1 and v_2, and the relative shape difference and the logarithmic amplitude ratio are defined as:

$$e_s = \frac{\|v_1 - v_2\|_2}{\max\left[\sqrt{\frac{\|v_1\|_2^{\,2} + \|v_2\|_2^{\,2}}{2}}, \varepsilon\right]} \tag{6}$$

$$e_a = \left|\ln \frac{\max[\|v_1\|_2, \varepsilon]}{\max[\|v_2\|_2, \varepsilon]}\right| \tag{7}$$

where e_s measures the shape difference between the two aligned sources, e_a measures their transient amplitude difference, and ε is a small quantity preventing division by zero. Beyond these local waveform features, the arrival timing of the two sources is further checked for consistency with distance variation. Let the arrival-time descriptors defined in Section 2.1 be t_1 and t_2; the timing residual is

$$e_t = \frac{|(t_2 - t_1) - s(d_2 - d_1)|}{\max[s(d_2 - d_1), 1]} \tag{8}$$

This indicator reflects the deviation of the actual arrival-time difference from the expected propagation difference, with all times measured in samples.

To determine the scales and the acceptance threshold of the three indicators, distance-held-out calibration within the training set is used. Each time, all operating-condition samples at one distance are held out, sources are selected from the other training distances on both sides, and the propagation-parameter calibration also excludes that distance. The 113 distances with two-sided sources form 2373 calibration records in total. From the valid source pairs among them, the 90th quantile of each indicator, q_k, is computed, and the composite score is constructed as:

$$Q = \max_{k \in \{s,\,a,\,t\}} \left(\frac{e_k}{q_k}\right) \tag{9}$$

The 95th quantile of the composite scores over the calibration records is taken as the acceptance threshold, giving $\tau = 3.04759$. A source pair is accepted when $Q \le \tau$; otherwise the corresponding derived sample is not generated. The screening rule is fixed before validation-set evaluation, and the formal augmentation uses propagation parameters calibrated on the full training set.

After screening, 8265 of the 8442 augmentation targets generate derived samples; the remaining 177 are not generated, of which 12 correspond to pointwise-identical source inputs and 165 to source pairs whose composite score exceeds the threshold. Finally, 7972 intermediate-distance samples and 293 end-zone samples are retained; merged with the 2415 original real training samples, they form 10680 training samples.

3.3. Fidelity Evaluation of Interpolated Waveforms

The fidelity of augmented waveforms is assessed against the 294 real waveforms at 14 off-grid distances in the validation set, comparing the construction results of plain linear interpolation and propagation-aligned interpolation, with the four-channel full-window root-mean-square error (RMSE) measuring the waveform difference:

$$\mathrm{RMSE}=\sqrt{\frac{1}{4N}\sum_{c=1}^{4}\sum_{n=1}^{N}\left(\tilde{w}_{aug}^{(c)}(n)-\tilde{w}_{true}^{(c)}(n)\right)^{2}} \qquad (10)$$

where N = 300. The RMSE is computed per case and then averaged; the results are given in Table 2. The mean RMSE of propagation-aligned interpolation decreases from 0.05001 to 0.04225, a reduction of about 15.5%. On the 288 common targets retained by quality screening, the mean RMSE values of the two methods are 0.04714 and 0.03933 respectively. Fig. 3 shows an example comparing a real waveform with the augmented waveforms, and the location benefit is further evaluated in Section 4.3.

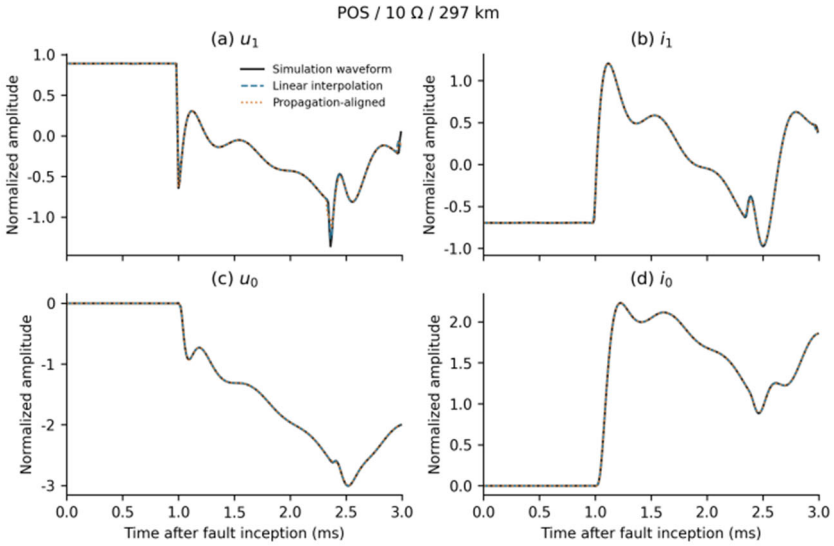


Figure 3. Comparison of real and interpolation-augmented waveforms (297 km, POS, 10 Ω)

Table 2. Waveform fidelity comparison on the validation set (294 cases)

Scheme	Construction	RMSE over all 294 cases
Linear interpolation	In-place mixing with distance weights	0.05001
Propagation-aligned interpolation	Two sources aligned to the target time before mixing	0.04225

4. HVDC Line Location Network Based on Time–Frequency Gated Fusion

4.1. Overall Network Architecture

The structure of the time–frequency gated fusion network TFNet-F is shown in Fig. 4. A single rectifier-side recording window generates a 300×4 normalized modal waveform and a 300×96 spectral sequence, the latter consisting of logarithmic magnitude features at 48 frequency positions each for the line-mode voltage and current. The two branches extract 128-dimensional features respectively, which are fused by feature-wise gating, and a regression layer outputs the fault distance.

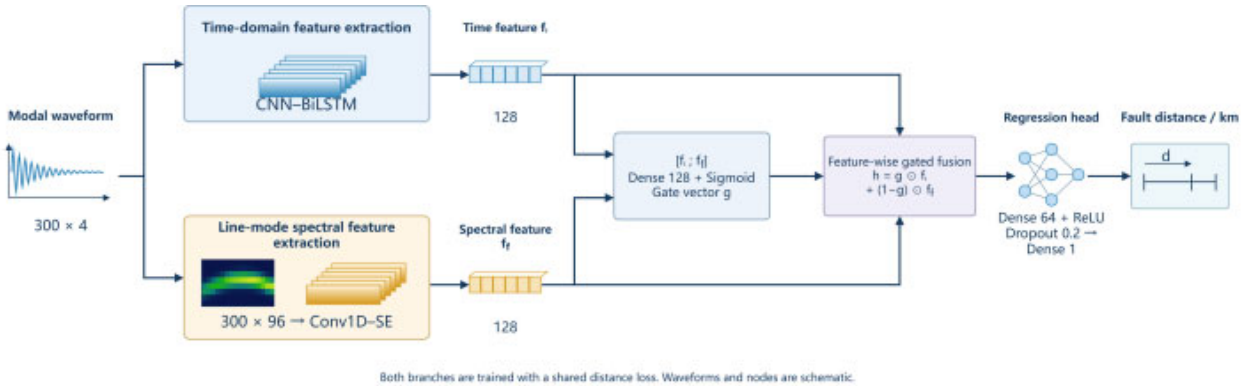


Figure 4. Overall structure of the gated time–frequency fusion network

The two branches encode the time-domain and spectral inputs respectively and are trained jointly through the same distance loss. The fusion layer generates feature weights according to the input, allowing both representations to participate in distance regression.

4.2. Time–Frequency Feature Extraction and Gated Fusion

Single-terminal fault location requires extracting distance-dependent features from transients, while the arrival timing, local shape, and spectral distribution of the waveforms are also affected by operating conditions such as transition resistance. In view of these differences in representation, a CNN–BiLSTM time-domain branch learns the local waveform patterns and the connections between preceding and following instants, a Conv1D–SE spectral branch learns the temporal patterns of the magnitude spectrum, and gated fusion combines the two representations for distance regression.

The time-domain branch takes the 300×4 normalized modal waveform as input and extracts abrupt-change, peak, and local-oscillation features through three one-dimensional convolution layers sliding along the time axis. The kernel lengths of the three layers are 7, 5, and 3, with output channels of 32, 32, and 64, each followed by batch normalization and ReLU activation. To preserve local responses while reducing the computational cost of subsequent sequence modeling, 2× max pooling is applied after the first two layers, compressing the sequence length from 300 to 150 and then 75. The pooled features are fed in temporal order into a BiLSTM with 64 hidden units per direction [14], which combines the information of preceding and following instants within the window; the representation is then mapped through global average pooling and a 128-unit dense layer with ReLU activation to the time-domain feature f_t. The branch structure is shown in Fig. 5.

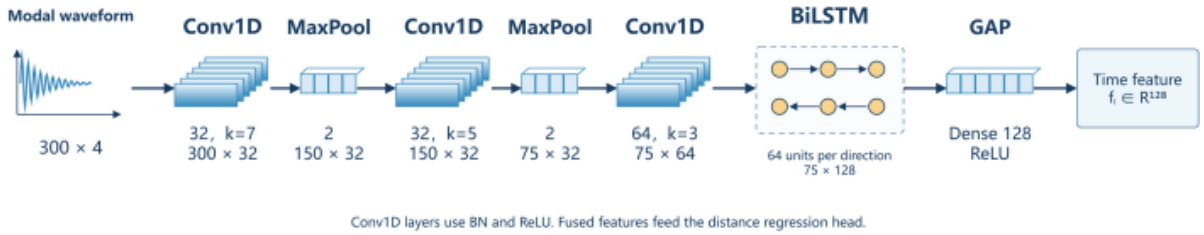


Figure 5. Time-domain feature extraction branch

The time distribution of different frequency components provides another transient representation, so the spectral branch converts the line-mode voltage and current into an S-transform magnitude spectrum and then applies convolutions to extract features. This transform adopts the frequency Gaussian window associated with the S transform [15], yielding the band sequence:

$$F_k[n] = \mathrm{IDFT}\left\{X[j]\exp\left[\frac{-2\pi^2\Delta(j,k)^2}{k^2}\right]\right\}[n],\quad k\geq 1,\quad F_0[n]=0 \qquad (11)$$

where $X[j]$ is the discrete Fourier coefficient of the 300-point normalized waveform, and both j and n range from 0 to 299; the DFT uses a negative exponent without scaling, and the IDFT uses a positive exponent and divides by 300. The circular signed frequency offset in the frequency Gaussian window is defined as follows, where the result of mod lies in 0 to 299. After filtering with this window the carrier phase is preserved, and taking the magnitude yields a discrete S-transform magnitude representation consistent with the common demodulated form.

$$\Delta(j,k) = ((j - k + 150) \bmod 300) - 150 \qquad (12)$$

The transform results of the line-mode voltage and current are each compressed by $\log(1+|F_k|)$ to reduce the numerical range of the spectral magnitudes, retaining 48 positions k = 0, …, 47. Among them, k = 0 is a zero placeholder, and the frequency interval is $\frac{100000}{300}$ Hz. The two groups of magnitude features at each time step are then interleaved in frequency-first order to form the 300×96 spectral input, so that time remains the sequence axis while the magnitudes at all frequency positions jointly constitute the input features at each instant, allowing the subsequent convolutions to extract the temporal variation of the spectral magnitudes.

Based on this serialized spectral input, three Conv1D layers extract the local variations of the spectral magnitudes along the time axis, with channel maps combining information across frequency positions. The kernel lengths are 7, 5, and 3, with output channels of 32, 32, and 64, each followed by BN and ReLU, and 2× pooling after the second layer compresses the time length to 150. Considering that different hidden channels respond differently to the current sample, an SE module [16] is introduced after the third convolution layer to adjust the weights of the 64 convolutional hidden channels according to the input:

$$s_c=\sigma(W_2\,\mathrm{ReLU}(W_1\,\mathrm{GAP}(X)))_c,\quad \widetilde{X}_c=s_c\cdot X_c \qquad (13)$$

The SE module summarizes the response of each hidden channel over the time axis through global average pooling, then passes through a 16-unit dense layer with ReLU activation, a 64-unit dense layer, and a sigmoid activation to map the summary into weights s corresponding to the 64 hidden channels. After channel-wise scaling with these weights, the contribution of each hidden channel to the current sample is adjusted; the weighted features are then mapped through global average pooling and a 128-unit dense layer with ReLU activation to the spectral feature f_f of the same dimension as f_t, facilitating the subsequent feature-wise fusion. The spectral branch and the SE module are shown in Fig. 6; the weights act on the hidden channels learned by convolution and do not directly correspond to the original frequency bins.

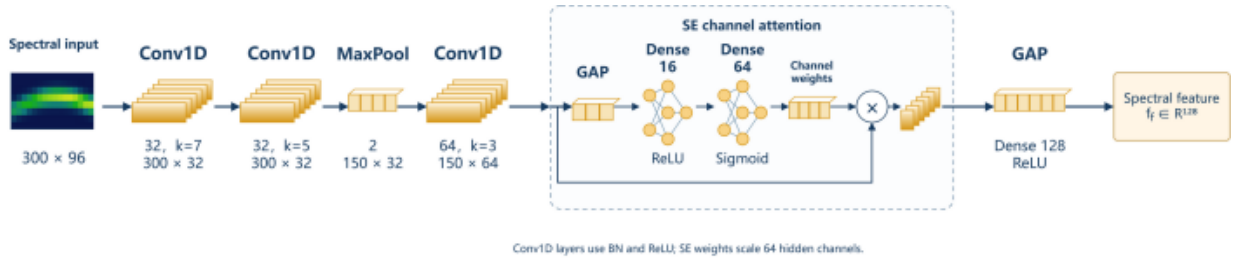


Figure 6. Spectral feature extraction branch and SE module

Corresponding to the spectral input, Fig. 7 shows the line-mode spectral magnitude distributions of the three fault types in the validation set, with the horizontal and vertical axes corresponding to time and frequency, and the color indicating the logarithmic magnitude.

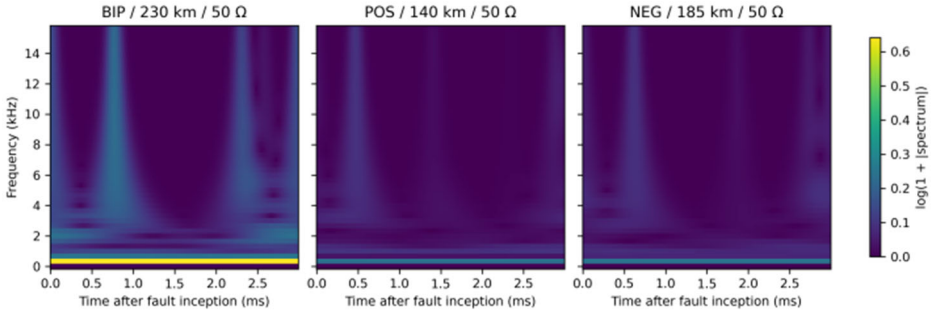


Figure 7. Line-mode spectral magnitude examples in the validation set (unified color scale)

Finally, after the encoding of the two branches, the same transient process yields the time-domain feature f_t and the spectral feature f_f. Since the responses of the two feature types differ across fault samples, the fusion layer dynamically generates weights from the input and combines the two 128-dimensional vectors component-wise. f_t and

$$g = \sigma(W_g[f_t;f_f] + b_g) \tag{14}$$

$$h = g \odot f_t + (1 - g) \odot f_f \tag{15}$$

where each element of g lies in [0, 1] and $\odot$ denotes element-wise multiplication; thus the closer a gating component is to 1, the larger the weight of the corresponding time-domain feature, and the closer to 0, the larger the weight of the corresponding spectral feature. Through weights jointly determined by the features of both branches, the fusion result adapts the combination of the two representations to each sample. The equivalent form $h = f_f + g \odot (f_t - f_f)$ is used in computation.

The fused 128-dimensional feature is further combined by Dense 64 and ReLU, regularized by Dropout 0.2 during training, and finally a linear Dense 1 outputs the distance estimate in km.

4.3. Loss Function and Training Strategy

Distance regression adopts the Huber loss [17], which maintains fine fitting in the small-error region while reducing the influence of large-error samples on the optimization; the threshold δ is set to 0.5 km. Let the predicted and true distances of sample i be $\hat{d}_i$ and d_i, with the regression error $e_i = \hat{d}_i - d_i$; the loss function is:

$$L_{Huber}(e) = \begin{cases} \frac{1}{2}e^2, & |e| \le \delta \\ \delta(|e| - \frac{1}{2}\delta), & |e| > \delta \end{cases} \tag{16}$$

The loss weights of the original and augmented samples are set to 1 and 0.5 respectively, controlling the influence of the derived waveforms on the parameter updates. The weighted Huber loss of each batch, summed and divided by the batch size, is used for optimization:

$$L = \frac{1}{N_B}\sum_{i=1}^{N_B} \omega_i L_{Huber}(e_i) \tag{17}$$

where N_B is the batch size and ω_i is the sample loss weight. Model selection uses the unweighted distance MAE on the real validation samples.

The network uses the Adam optimizer with an initial learning rate of 3×10^{-4}, decaying with the optimizer update steps in a cosine schedule down to 10^{-5}:

$$\eta_t = \eta_{min} + \frac{1}{2}(\eta_0 - \eta_{min})\left(1 + \cos\frac{t}{T}\pi\right) \tag{18}$$

where t is the number of executed optimizer updates and T is the product of the batches per epoch and the 200 epochs, which determines the learning-rate schedule period. Incomplete batches at the end of each epoch still participate in the updates, and this schedule period is retained under early stopping, keeping the same basis for the learning-rate schedule.

Two training settings are used. For the augmentation-density and network-structure comparisons, the batch size is 16, the random seed is 42, and the training cap is 200 epochs;

the distance MAE over all validation samples serves as the early-stopping metric—if no improvement occurs for 30 consecutive epochs, training stops and the parameters with the lowest validation MAE are retained. The augmentation-construction comparison uses the same gated structure and initial weights; each group trains for 200 epochs, using the 2415 original samples and 8442 augmented-sample draws per epoch. For the schemes with fewer samples after screening, uniform resampling makes up the training volume so that all groups share the same training budget; the model of each group is selected by the distance MAE on the 294 off-grid validation samples for the subsequent test-set evaluation.

5. Experimental Results and Analysis

5.1. Experimental Setup and Evaluation Metrics

The experiments use the dataset constructed in Section 1. The test set contains 24 off-grid fault distances and 504 real simulation samples. The following analysis examines in turn the overall location performance, the effects of data augmentation and network structure, and the influence of noise, amplitude variation, and input-channel occlusion on the model outputs.

The augmentation-construction comparison configures four schemes: plain linear interpolation (A), quality-screened linear interpolation (B), propagation-aligned interpolation (C), and quality-screened propagation-aligned interpolation (D). Each group uses the unified training settings of Section 3.3; for groups B and D with fewer samples after screening, uniform resampling equalizes the per-epoch training volume with the groups A and C.

All location results are obtained by a single forward pass of the trained models, without test-time augmentation. The interpolation-density and network-structure comparisons use the early-stopping setting, and the A–D augmentation-construction comparison uses the fixed-epoch setting, separately analyzing the effects of distance coverage, feature fusion, and augmentation construction. Location performance is evaluated by the mean absolute error (MAE), the median absolute error (P50), the 95th-percentile absolute error (P95), and the maximum absolute error, reflecting both the overall accuracy and the behavior of the large-error samples. Let the absolute error of test sample i be a_i; the metrics are defined as:

$$a_i = \left|\hat{d}_i - d_i\right|, \quad \mathrm{MAE} = \frac{1}{N_{test}} \sum_{i=1}^{N_{test}} a_i \tag{19}$$

$$\mathrm{P50} = \mathrm{Quantile}_{0.50}(\{a_i\}), \quad \mathrm{P95} = \mathrm{Quantile}_{0.95}(\{a_i\}) \tag{20}$$

$$E_{max} = \max_{1 \le i \le N_{test}} a_i \tag{21}$$

where N_{test} is the number of test samples and Quantile denotes the quantiles computed by linear interpolation. The test set contains 504 samples. For the multi-output reference model, the fault-type accuracy, the transition-resistance $\log_{10}$ MAE , and the within-half-decade proportion are additionally reported; the resistance error is computed in the $\log_{10}\left(\frac{R}{1\ \Omega}\right)$ domain, and the within-half-decade proportion is the fraction of samples whose predicted-to-true resistance ratio lies between $\frac{1}{\sqrt{10}}$ to $\sqrt{10}$.

5.2. Data Augmentation and Network Structure Comparison

First, under the same linear interpolation with end-zone densification, time-only, spectrum-only, concatenation fusion, and gated fusion are compared to examine how the time–frequency features participate in distance regression; the results are given in Table 3 and the right panel of Fig. 8. A multi-output task-division model TFNet-MT is also included as a complete-configuration reference, in which the time-domain branch predicts the distance and the spectral branch predicts the fault type and resistance, with the spectral input additionally containing the zero-mode complex spectrum; its results serve to compare different complete configurations.

Table 3. Test results of each network under the same linear augmentation (km)

Network structure	MAE	P50	P95	Max error
Time-domain branch only	0.601	0.447	1.987	3.122
Spectral branch only	2.742	1.857	8.283	15.985
Dual-branch concatenation fusion	1.386	1.101	4.154	9.581
Dual-branch gated fusion	0.491	0.334	1.588	3.635
Multi-output task-division model TFNet-MT	0.527	0.435	1.285	1.713

As shown in Table 3, the MAE of gated fusion is 0.491 km, about 18.4% lower than the 0.601 km of the time-domain-only branch, and the P95 decreases from 1.987 km to 1.588 km, while the maximum error increases from 3.122 km to 3.635 km. Concatenation fusion yields an MAE of 1.386 km without a comparable fusion gain. The MAE of TFNet-MT is 0.527 km—lower average accuracy than gated fusion—but its P95 and maximum error are 1.285 and 1.713 km, illustrating the difference between the average error and the tail error. Considering the average location error under this condition, the subsequent augmentation comparison adopts the gated network with its structure unchanged.

Under the same 2.5 km augmentation density, the two-dimensional spectral fusion and the one-dimensional spectral gating schemes implemented in this paper achieve MAEs of 1.651 and 0.860 km respectively; the one-dimensional spectral gating has the lower average error in this configuration.

After determining the gated structure, the effect of data augmentation is examined from two aspects: distance coverage and waveform construction. First, the linear interpolation density is increased step by step to compare the test errors of the complete training configurations; then plain linear interpolation (A), quality-screened linear interpolation (B), propagation-aligned interpolation (C), and quality-screened propagation-aligned interpolation (D) are configured under the same training budget to distinguish the effects of quality screening and propagation alignment. The

results are given in Tables 4 and 5.

Table 4. Test MAE under different linear interpolation densities

Augmentation configuration	Training samples	MAE (km)	Reduction vs. previous
No augmentation	2415	1.257	—
2.5 km midpoint interpolation	4431	0.860	31.6%
1 km interpolation	10479	0.638	25.7%
1 km interpolation with 0.5 km end densification	10857	0.491	23.1%

As the training distance grid densifies, the MAE decreases from 1.257 km without augmentation to 0.491 km. Compared with the 1 km interpolation configuration, adding the 378 end-zone samples increases the training volume by about 3.6% and reduces the MAE by about 23.1%. All density configurations use the same early-stopping rule, but the actual number of updates varies with the sample volume and the stopping epoch; the result therefore reflects the joint change brought by the enlarged samples and the corresponding training process.

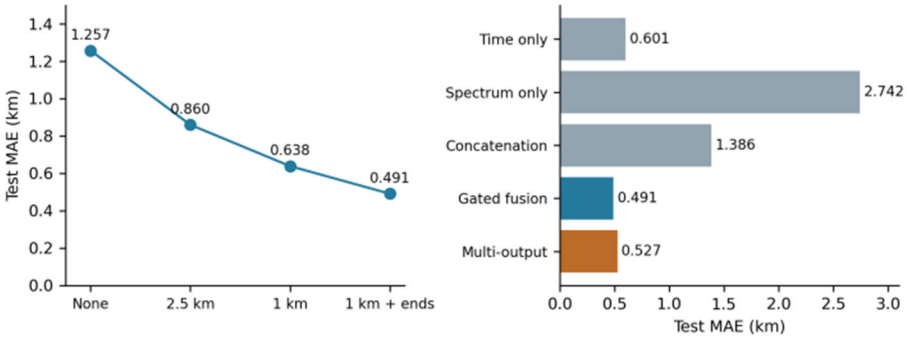


Figure 8. Test MAE comparison: interpolation densities (left) and network structures (right)

Table 5. Equal-budget test results of different augmentation constructions (km)

Augmentation scheme	Augmented samples	MAE	P50	P95	Max error
Plain linear interpolation (A)	8442	0.555	0.447	1.291	3.863
Quality screening + linear interpolation (B)	8265	0.406	0.382	0.939	1.919
Propagation-aligned interpolation (C)	8442	0.298	0.217	0.718	8.142
Quality screening + propagation alignment (D)	8265	0.285	0.209	0.752	2.181

In Table 5, the MAE of the groups B, C, and D is reduced by 26.9%, 46.3%, and 48.6% relative to group A. Quality-screened linear interpolation lowers the MAE from 0.555 km to 0.406 km, and propagation-aligned interpolation to 0.298 km, showing that both treatments improve the test performance of plain linear interpolation in this round of comparison.

Adding quality screening on top of propagation alignment, group D attains an MAE of 0.285 km, 12.4 m lower than group C, and the maximum error drops from 8.142 km to 2.181 km, while the P95 increases from 0.718 km to 0.752 km. The two most distinct cases are end-zone low-resistance bipolar short circuits at 9.5 km and 12 km, predicted by group C as 17.6 km and 17.8 km and by group D as 11.7 km and 13.4 km. A paired bootstrap over the 24 test distances gives a 95% interval of the MAE difference of [−75.8, +51.2] m, which is insufficient to confirm a stable average-accuracy advantage of group D. Considering both the test average error and the extreme errors, group D is adopted as the main reported scheme; Table 5 retains the complete results of all four groups.

5.3. Fault Location Performance Analysis

Based on the above structure and augmentation comparisons, the final location method combines the quality-screened propagation-aligned augmentation with the gated network. Its results on the 504 test samples are given in Table 6, with an MAE of 0.285 km, a P50 of 0.209 km, a P95 of 0.752 km, and a maximum absolute error of 2.181 km. Fig. 9 further shows the per-case errors and their distribution; the mean error at each of the 24 test distances is computed from the 21 samples at that position.

Table 6. Location results of the proposed method on the test set (km)

Model	MAE	P50	P95	Max error
Proposed method (D)	0.285	0.209	0.752	2.181

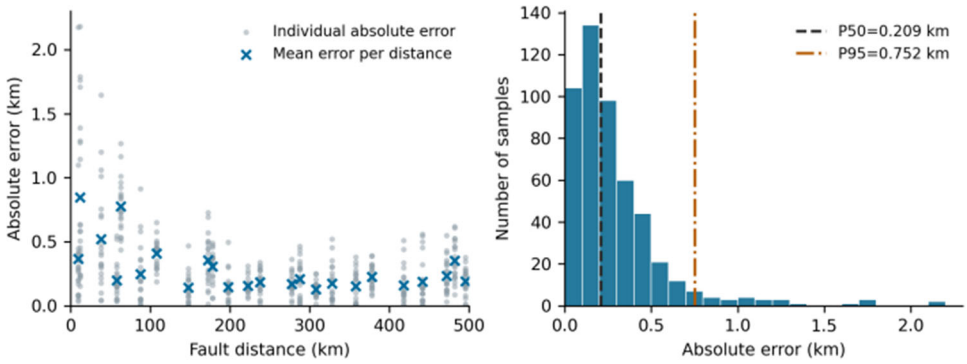


Figure 9. Test absolute errors and distribution of the proposed method (504 cases)

Beyond the overall evaluation, the same test set is grouped by fault type, transition resistance, and distance interval, as shown in Fig. 10. The three fault types contain 168 cases each and each resistance level 72 cases; with 160 km and 330 km as the boundaries, the near, middle, and far sections contain 168, 189, and 147 cases. Each grouping covers all 504 test samples.

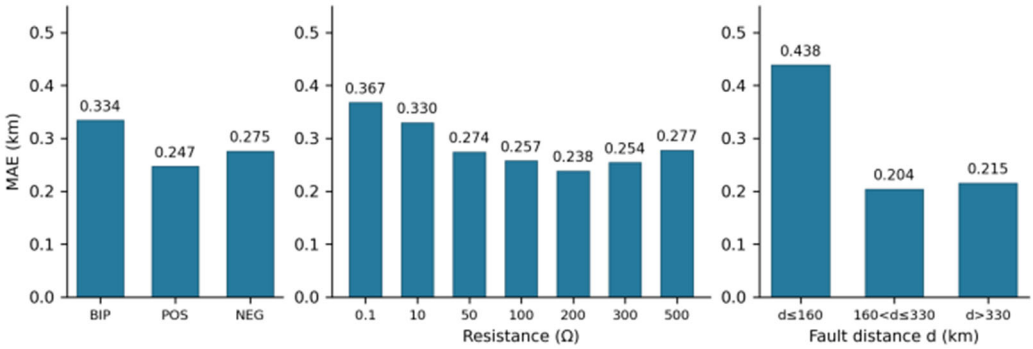


Figure 10. Test MAE of the proposed method by fault type, transition resistance, and distance interval

As shown in Fig. 10, the MAE of bipolar short circuits is 0.334 km, higher than the 0.247 and 0.275 km of the positive-pole-to-ground and negative-pole-to-ground faults. The MAEs at 0.1 Ω and 10 Ω are 0.367 and 0.330 km, and the other five levels range from 0.238 to 0.277 km without sustained growth with the resistance. The near-section MAE is 0.438 km, higher than the 0.204 and 0.215 km of the middle and far sections, indicating that the current errors concentrate in the near-section and low-resistance conditions.

5.4. Effects of Noise and Amplitude Perturbation

Since measurement noise and amplitude variation alter the transient information received by the model, additive Gaussian white noise at various SNRs is injected into the normalized test waveforms, the spectral features are recomputed, and the results are fed into the trained gated model; ±5% amplitude scaling is also applied, and a joint condition of 30 dB noise with +5% scaling is included, forming the noise and amplitude perturbation experiments. The model parameters remain fixed in all tests, and the noise standard deviation is determined by the SNR:

$$\sigma = \sqrt{\frac{P_s}{10^{\frac{SNR}{10}}}}, P_s = \frac{1}{4MN}\sum_{i=1}^{M}\sum_{c=1}^{4}\sum_{n=1}^{N}\left(\tilde{w}_i^{(c)}(n)\right)^2 \quad (22)$$

where P_s is the common average power of the 504 test waveforms over the four channels, all samples use the same noise standard deviation, N = 300, and the random seed is 42. Amplitude scaling multiplies the normalized waveforms by 1.05 or 0.95, and the joint condition scales before adding noise, evaluating the influence of the amplitude variation in the normalized domain on the location results.

Table 7. Noise and amplitude perturbation results of the proposed method (n = 504)

Condition	Distance MAE (km)	P95 (km)	Max error (km)
Clean	0.285	0.752	2.181
AWGN 40 dB	0.291	0.800	2.181
AWGN 35 dB	0.297	0.876	2.172
AWGN 30 dB	0.354	0.956	4.077
AWGN 25 dB	0.525	1.567	8.948
AWGN 20 dB	0.815	2.384	10.127
Gain +5%	0.387	0.901	2.246
Gain −5%	1.672	3.440	4.588
30 dB + gain +5%	0.402	1.131	5.206

As shown in Table 7, at the SNRs of 40, 35, and 30 dB the MAEs are 0.291, 0.297, and 0.354 km, increasing step by step from the noise-free case; when the SNR further drops to 25 and 20 dB, the MAE rises to 0.525 and 0.815 km, with maximum errors of 8.948 and 10.127 km. Relative to the noise-free baseline, the MAE at 30 dB increases by about 24.2%, indicating that stronger noise degrades the location accuracy, with a particularly marked effect on the large-error samples.

Amplitude scaling affects the location asymmetrically.

Scaling the waveforms up by 5% gives an MAE of 0.387 km, whereas scaling down by 5% raises the MAE to 1.672 km with a P95 of 3.440 km. The joint application of +5% scaling and 30 dB noise gives an MAE of 0.402 km. Within the tested range, the negative amplitude variation is the more sensitive perturbation condition.

5.5. Input Feature Sensitivity Analysis

To verify the dependence of the proposed method on different inputs, whole-window mean replacement is applied to the time-domain modal channels and the spectral input channels separately—one channel at a time, with the remaining inputs unchanged. The time-domain replacement values are computed from the original training samples, and the spectral replacement values from the training and augmentation feature pools, both averaged over the samples and time.

Table 8. Time-domain channel occlusion results of the proposed method (km)

Occluded channel	BIP	POS	NEG	All
Baseline (no occlusion)	0.334	0.247	0.275	0.285
Occlude u_1 (line-mode voltage)	221.2	221.4	221.4	221.3
Occlude i_1 (line-mode current)	221.0	221.5	221.5	221.3
Occlude u_0 (zero-mode voltage)	0.334	0.259	0.299	0.297
Occlude i_0 (zero-mode current)	0.334	0.245	0.239	0.273

The time-domain replacement results are given in Table 8. With the spectral branch input unchanged, replacing either the line-mode voltage or the line-mode current raises the MAE to about 221.3 km, whereas replacing the zero-mode voltage and current gives MAEs of 0.297 and 0.273 km. Compared with the 0.285 km without replacement, the whole-window replacement of the line-mode channels causes far larger errors than that of the zero-mode channels, indicating that the model is relatively sensitive to the time-domain line-mode information.

With the time-domain input unchanged, replacing each of the 96 spectral channels by its mean in turn yields a maximum MAE of 0.331 km at the 0.333 kHz line-mode current channel, about 46 m above the no-replacement value; the 0.333 kHz line-mode voltage and the 1.000 kHz line-mode current channels give MAEs of 0.325 and 0.316 km. The single spectral-channel replacement has a smaller effect than the whole-window time-domain line-mode replacement, reflecting the different responses of the model to the two types of input perturbation. Mean replacement alters the inter-channel correlations, so these results are used to evaluate the feature sensitivity and are analyzed separately from the noise and amplitude perturbations in Section 4.4.

6. Conclusion

Aiming at the problem that the single-terminal fault transients of HVDC transmission lines are jointly affected by fault distance, transition resistance, and other factors, making the distance-related features difficult to extract accurately, this paper proposes a fault location method combining propagation-aligned interpolation augmentation with a time–frequency gated fusion network. Augmented samples are constructed according to the transient propagation timing; a dual-branch network extracts the temporal and spectral features; and feature-wise gated fusion achieves the fault distance estimation.

On the 504 test samples of a ±500 kV, 500 km bipolar DC transmission system, the proposed method achieves a mean absolute error of 0.285 km and a maximum absolute error of 2.181 km. With the network structure and training budget fixed, the combination of propagation alignment and quality screening reduces the mean absolute error by about 48.6% relative to plain linear interpolation; under the same linear augmentation, the gated fusion reduces it by about 18.4% relative to the time-domain-only branch. These results show that improving the construction of augmented waveforms and properly fusing time–frequency features both help improve the single-terminal fault location accuracy. In the perturbation tests, the MAE under 30 dB noise is 0.354 km, while strong noise and negative amplitude scaling cause marked error growth.

Therefore, the proposed method improves the single-terminal fault location accuracy through propagation-aligned augmentation and time–frequency feature fusion, but obvious error growth remains under strong noise and negative amplitude deviation. Future work will study the robust extraction of distance-related features under measurement disturbances, and further verify the applicability of the method with different line parameters and field fault recordings.

Acknowledgements

This work was supported by the Youth Science Research Fund of Sichuan University Jinjiang College – Research on HVDC Transmission Line Fault Location Technology Based on Multimodal Features and Time-Frequency Dual-Branch Network (QNJJ-2025-B10).

References

[1] Lei, Z., Hao, L., Dai, J., et al. (2022). A review of fault location research for HVDC transmission lines. Power System Protection and Control, 50(11), 178–187. https://doi.org/10.19783/j.cnki.pspc.211228. (in Chinese)

[2] Suonan, J., Gao, S., Song, G., et al. (2010). A novel fault-location method for HVDC transmission lines. IEEE Transactions on Power Delivery, 25(2), 1203–1209. https://doi.org/10.1109/TPWRD.2009.2033078.

[3] Dardengo, V. P., Fardin, J. F., & De Almeida, M. C. (2021). Single-terminal fault location in HVDC lines with accurate wave velocity estimation. Electric Power Systems Research, 194, 107057. https://doi.org/10.1016/j.epsr.2021.107057.

[4] Huai, Q., Liu, K., Hooshyar, A., et al. (2021). Single-ended line fault location method for multi-terminal HVDC system based on optimized variational mode decomposition. Electric Power Systems Research, 194, 107054. https://doi.org/10.1016/j.epsr.2021.107054.

[5] Reis, R. L. A., Lopes, F. V., Neves, W. L. A., et al. (2021). An improved single-ended correlation-based fault location technique using traveling waves. International Journal of Electrical Power & Energy Systems, 132, 107167. https://doi.org/10.1016/j.ijepes.2021.107167.

[6] Hao, Y., Wang, Q., Li, Y., et al. (2018). An intelligent algorithm for fault location on VSC-HVDC system. International Journal of Electrical Power & Energy Systems, 94, 116–123. https://doi.org/10.1016/j.ijepes.2017.06.030.

[7] Wang, L., He, Y., & Li, L. (2021). A single-terminal fault location method for HVDC transmission lines based on a hybrid deep network. Electronics, 10(3), 255. https://doi.org/10.3390/electronics10030255.

[8] Inwumoh, J., Baguley, C., Madawala, U., et al. (2023). A novel fault location strategy based on Bi-LSTM for MMC-HVDC systems. The Journal of Engineering, 2023(10), e12310. https://doi.org/10.1049/tje2.12310.

[9] Wang, X., Zhang, D., Li, M., et al. (2023). Fault location method for hybrid DC transmission systems based on wavelet energy spectrum and SSA-GRU. Power System Protection and Control, 51(12), 14–24. https://doi.org/10.19783/j.cnki.pspc.221509. (in Chinese)

[10] Su, C., Yang, Q., Wu, X., et al. (2023). A two-terminal fault location fusion model of transmission line based on CNN-Multi-Head-LSTM with an attention module. Energies, 16(4), 1827. https://doi.org/10.3390/en16041827.

[11] Tang, Y., Shu, H., Liu, H., et al. (2025). A pre-classification method of field-recorded traveling waves for online fault location based on CABFAM-Transformer. Transactions of China Electrotechnical Society, 40(5), 1455–1470. https://doi.org/10.19595/j.cnki.1000-6753.tces.240326. (in Chinese)

[12] Li, C., & He, P. (2018). Fault-location method for HVDC transmission lines based on phase frequency characteristics. IET Generation, Transmission & Distribution, 12(4), 912–916. https://doi.org/10.1049/iet-gtd.2017.0967.

[13] Zhang, H., Cisse, M., Dauphin, Y. N., et al. (2018). Mixup: Beyond empirical risk minimization. International Conference on Learning Representations. https://openreview.net/forum?id=r1Ddp1-Rb.

[14] Graves, A., & Schmidhuber, J. (2005). Framewise phoneme classification with bidirectional LSTM and other neural network architectures. Neural Networks, 18(5–6), 602–610. https://doi.org/10.1016/j.neunet.2005.06.042.

[15] Stockwell, R. G., Mansinha, L., & Lowe, R. P. (1996). Localization of the complex spectrum: The S transform. IEEE Transactions on Signal Processing, 44(4), 998–1001. https://doi.org/10.1109/78.492555.

[16] Hu, J., Shen, L., & Sun, G. (2018). Squeeze-and-excitation networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 7132–7141. https://doi.org/10.1109/CVPR.2018.00745.

[17] Huber, P. J. (1964). Robust estimation of a location parameter. The Annals of Mathematical Statistics, 35(1), 73–101. https://doi.org/10.1214/aoms/1177703732.