

Article

Disturbance-Event Recognition Model for Terrestrial Optical Cables Based on CNN-SVM

Xiaorui Qiao ^{1,2}, Junhua Zhang ¹ and Xichen Wang ^{1,*}

¹ Naval University of Engineering, Wuhan 430033, China; 0909042004@nue.edu.cn (X.Q.); m23385401@nue.edu.cn (J.Z.)

² School of Telecommunication Engineering, Xidian University, Xi'an 710071, China

* Correspondence: 1509021020@nue.edu.cn

Abstract

Distinguishing between human-made interferences and natural background disturbances is of great significance for the safe operation of terrestrial optical cables because human-caused damage can be halted through timely intervention. To address the problem of small-sample disturbance recognition in distributed acoustic sensing (DAS) systems, this paper proposes a fused CNN–SVM classification model based on hybrid features. A convolutional neural network is employed to extract the high-level spatiotemporal features of disturbance signals, which are subsequently fused with statistical features and fed into a support vector machine for classification. Evaluated on open-source data, the proposed model achieves accuracy improvements of 9.1%, 8.7%, and 2.7% over the conventional CNN, the statistical-feature-based SVM, and the conventional CNN-SVM model, respectively. Furthermore, based on field-measured data, a dataset comprising 5664 samples was constructed, covering four typical disturbance-event types: background noise, drilling, knocking, and digging. The field classification results demonstrate that the three-layer convolutional structure of the model achieves a recognition accuracy of 98.5%. Both the ROC curves and multiple evaluation metrics indicate that the proposed three-layer fused CNN–SVM model delivers better classification performance and more balanced category recognition, offering a feasible reference for similar fiber disturbance engineering tasks.

Keywords: pattern recognition; distributed acoustic sensing (DAS); convolutional neural network (CNN); support vector machine (SVM)

1. Introduction

Natural disturbances (e.g., geological subsidence, wind and rain erosion) affecting long-distance terrestrial optical cables are generally uncontrollable, whereas anthropogenic disturbances (e.g., construction excavation, mechanical drilling) can be prevented and managed once detected in a timely manner. Therefore, accurately distinguishing anthropogenic disturbances from natural background noise is critical for ensuring the safe and stable operation of optical cable networks. Distributed acoustic sensing (DAS) technology, leveraging its advantages of long-distance coverage and high sensitivity, is widely used for optical cable condition monitoring. But the high sensitivity of DAS equipment leads to severe superimposition of ambient environmental noise and the equipment's inherent thermal noise, while the signal characteristics of different types of external force disturbances exhibit severe aliasing. These factors collectively prevent existing algorithms from



Received: 20 May 2026

Revised: 20 June 2026

Accepted: 23 June 2026

Published: 26 June 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

effectively identifying anthropogenic disturbances, thereby limiting the reliability of early warning and protection.

To address the above issues, recent studies have attempted to introduce machine learning and deep learning methods for intelligent recognition of fiber optic disturbance signals. Mahmoud et al. achieved the classification of intrusion events in fiber optic perimeter security based on level-crossing features and an artificial neural network [1]. Wu et al. applied a one-dimensional convolutional neural network to pipeline intrusion recognition using distributed fiber optic sensing, achieving better performance than several traditional machine learning algorithms, and further proposed a multi-scale CNN structure to enhance feature representation capability [2,3]. Hu et al. combined Gramian Angular Fields with a two-dimensional CNN to realize multi-class signal recognition in perimeter security scenarios [4]. Li et al. proposed a fault prediction method based on time–frequency fused features using a combination of CNN and SVM, achieving a fault diagnosis accuracy of 97.21% and a recall of 96.88% [5]. Liu proposed an attention-enhanced ResNet (ResNet+CBAM) for DAS signal recognition in complicated urban environments. With preprocessing feature selection, the model achieves an average accuracy of 99.014% in field testing and 91.08% under challenging conditions, with a time cost of only 3.3 ms per sample [6].

However, the above studies mostly focus on using fiber optic sensing systems for perimeter security or intrusion detection, with less attention paid to the targeted identification of disturbance events that directly threaten the integrity of optical cables from the perspective of cable self-protection. This limitation, to some extent, restricts the applicability of optical cable safety monitoring in real-world scenarios.

Several researchers have attempted to identify several types of cable disturbances. Zhang et al. constructed a multi-feature classification framework based on statistical features and used a support vector machine to recognize disturbances such as knocking and watering [7], but this method is sensitive to noise and has limited generalization ability. Cao et al. comparatively analyzed the classification performance of CNN and SVM using simulated ϕ -OTDR data; yet the experimental data deviated from actual working conditions, and the model structure was relatively simple [8]. Xing et al. employed a bi-long short-term memory network for disturbance localization in terrestrial cables [9]. Although it achieved high recognition accuracy, their approach relies heavily on long temporal sequences and large-scale data, resulting in high computational costs and limited practical application efficiency.

Based on the research status outlined above, this paper proposes a fused CNN-SVM model for terrestrial optical cable disturbance identification under real-world operating conditions. To address the limitations of single models in disturbance recognition, the proposed model employs a convolutional neural network to automatically extract spatiotemporal features from disturbance signals, while introducing a support vector machine to replace the traditional Softmax classifier, thereby enhancing the ability to characterize classification boundaries under small-sample conditions. Furthermore, the model fuses the deep features extracted by the CNN with manually designed time-domain and frequency-domain statistical features, effectively alleviating the severe signal feature aliasing problem among different disturbance sources (e.g., construction excavation, mechanical drilling). Through this design, the model not only overcomes the superimposed interference of ambient environmental noise and equipment thermal noise in actual optical cable operation but also maintains high recognition accuracy under limited labeled samples. Compared with single deep learning models or traditional signal processing methods, the proposed hybrid model achieves efficient and accurate identification of multiple disturbance-event types, providing a more reliable decision-making basis for early warning and targeted protection of optical cable lines.

2. Convolutional Neural Network and Support Vector Machine

2.1. Structure of Convolutional Neural Network

The convolutional neural network (CNN), serving as a key feature extraction module of the proposed model, is capable of automatically learning hierarchical feature representations of signals via local receptive fields and weight sharing. Its basic architecture comprises an input layer, convolutional layers, pooling layers, fully connected layers, and an output layer [10].

2.1.1. Convolutional Layer

The convolutional layer extracts features by scanning input data through a sliding window in a local region, selecting the maximum or average pooling in the window for downsampling through a pooling layer to reduce data dimensionality and computation, improve computational efficiency, and extract more stable features, and then integrating and classifying.

2.1.2. Pooling Layer

In the CNN structure, the features are extracted and feature mapping is performed on the input data by the convolution kernel in the convolutional layer; the convolution operation of the convolution kernel is given in Equation (1):

$$y^l = f(\sum_k x_k^{l-1} * w_k^{(l)} + b^{(l)}) \quad (1)$$

where $x^{(l-1)}$ denotes the input feature map of the previous layer, $w^{(l)}$ represents the convolution kernel, $b^{(l)}$ is the bias term, $f(\cdot)$ is the activation function, and $*$ denotes the convolution operation [11].

Pooling layers are introduced after convolutional layers to reduce feature map size while preserving salient information. In this study, max pooling is adopted, which can be expressed as

$$y^l = \max(x^{(l-1)}) \quad (2)$$

2.1.3. Activation Function and Fully Connected Layer

To improve nonlinear representation capability and training efficiency, the Rectified Linear Unit (ReLU) is employed as the activation function. Compared with sigmoid and tanh functions, ReLU effectively alleviates gradient saturation and accelerates convergence [11]. The ReLU function is defined as

$$\text{ReLU} = \begin{cases} x & x > 0 \\ 0 & x < 0 \end{cases} \quad (3)$$

2.2. Support Vector Machine

The support vector machine (SVM), serving as the final classifier, is mostly used for pattern recognition, classification, and regression analysis, and is particularly suitable for small-sample and nonlinear classification problems.

For a binary classification problem with a training sample set $S\{(x_i, y_i), i = 1, 2, \dots, n\}$, $x_i \in \mathbb{R}$, $y \in (-1, 1)$, the SVM maps the training samples into a high-dimensional space via a kernel function and constructs an optimal classification hyperplane to achieve event classification [12,13].

The decision function of SVM can be expressed as

$$f(x) = \omega^T \Phi(x) + b \quad (4)$$

where ω represents the weight vector; $\Phi(x)$ denotes the nonlinear mapping; and b is the bias term.

The SVM classifier determines an optimal separating hyperplane by solving a soft-margin optimization problem:

$$\begin{cases} \min_{\omega, b, \xi} \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^N \xi_i \\ \text{s.t. } y_i(\omega^T \Phi(x) + b) \geq \varepsilon - \xi_i, \xi_i \geq 0 \end{cases} \quad (5)$$

where C is the penalty factor, ε indicates the tolerance error, and ξ_i is the slack variable [14].

3. Pattern Recognition Model Based on Fused CNN-SVM

To address the problem of insufficient recognition accuracy of cable disturbance events under small-sample conditions, this paper proposes a fused CNN-SVM disturbance recognition framework, as illustrated in Figure 1. The proposed method fully leverages the advantages of convolutional neural networks in spatiotemporal feature extraction and the strong generalization capability of support vector machines in high-dimensional, small-sample classification tasks.

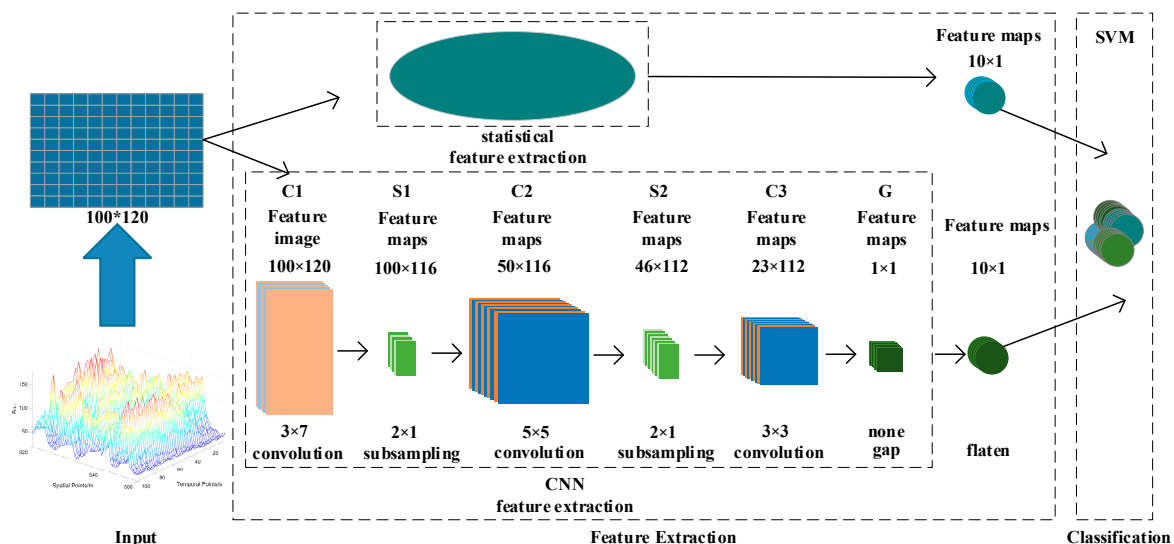


Figure 1. Structure of the fused CNN-SVM model.

Figure 1 illustrates the static architecture of the fused CNN-SVM model, which uses a dual-path strategy to fuse deep CNN features with handcrafted statistical features. Figure 2 then details the step-by-step workflow of model construction—dataset segmentation, feature extraction and normalization, SVM parameter tuning, and pattern recognition—providing a clear roadmap for model development. The fused CNN-SVM model mainly consists of two parts, namely feature extraction and pattern recognition, and involves four stages:

1. CNN feature extraction based on a convolutional neural network;
2. Statistical feature extraction based on statistical formulas;
3. Combination of CNN features and statistical features;
4. SVM-based classification of disturbance events.

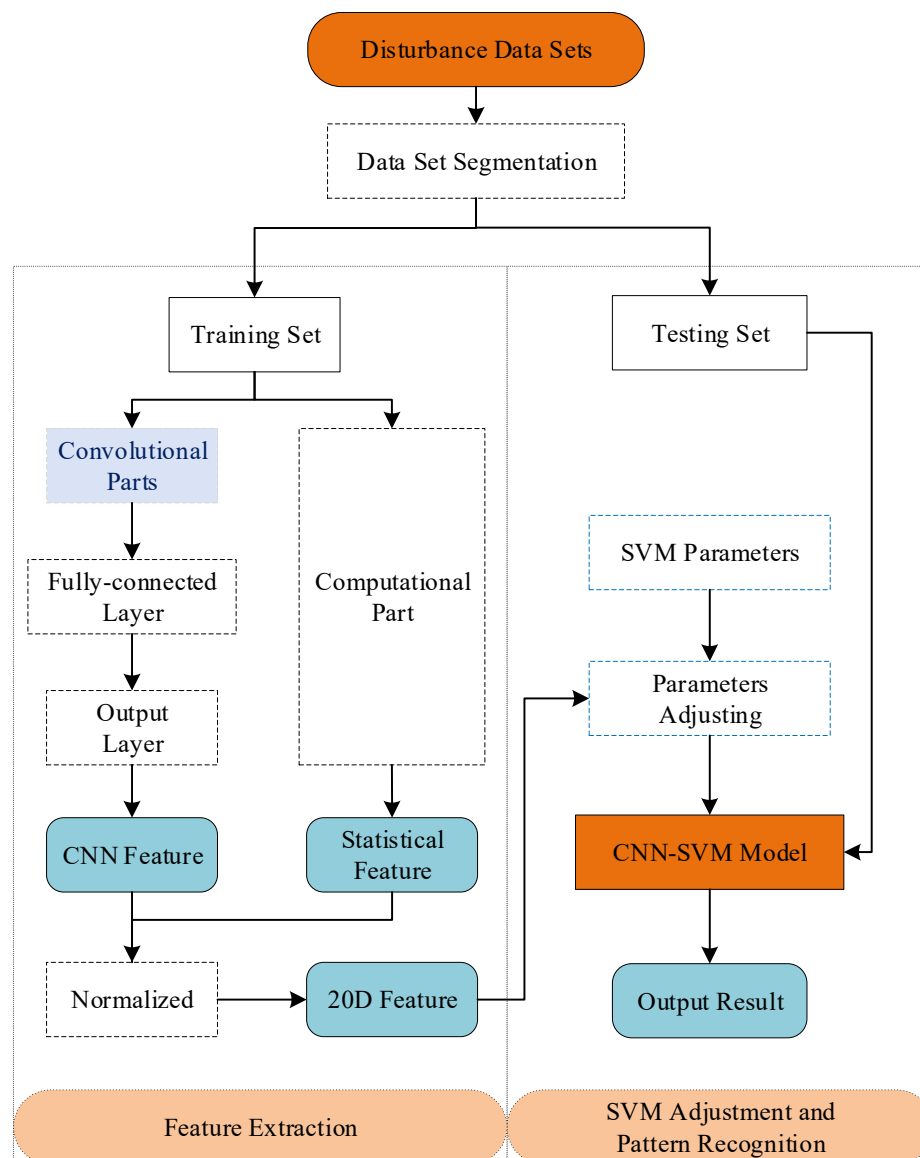


Figure 2. Flowchart of the fused CNN-SVM model.

3.1. Baseline Model Testing

Baseline simulation tests were performed using an open-source dataset [8], which comprises six types of disturbance events: background noise, digging, knocking, watering, shaking, and walking. The dataset was split into training and test sets at a ratio of 8:2. Each sample was formatted with 10,000 temporal sampling points and 12 spatial sampling points (as shown in Figure 3, warmer colors (red) indicate higher values, while cooler colors (purple) indicate lower values.). The quantity and label settings for each disturbance event are listed in Table 1.

Table 1. Data of the six types of events.

Event Types	Number	Label
Background Noise	3094	0
Digging	2512	1
Knocking	2530	2
Watering	2298	3
Shaking	2728	4
Walking	2450	5

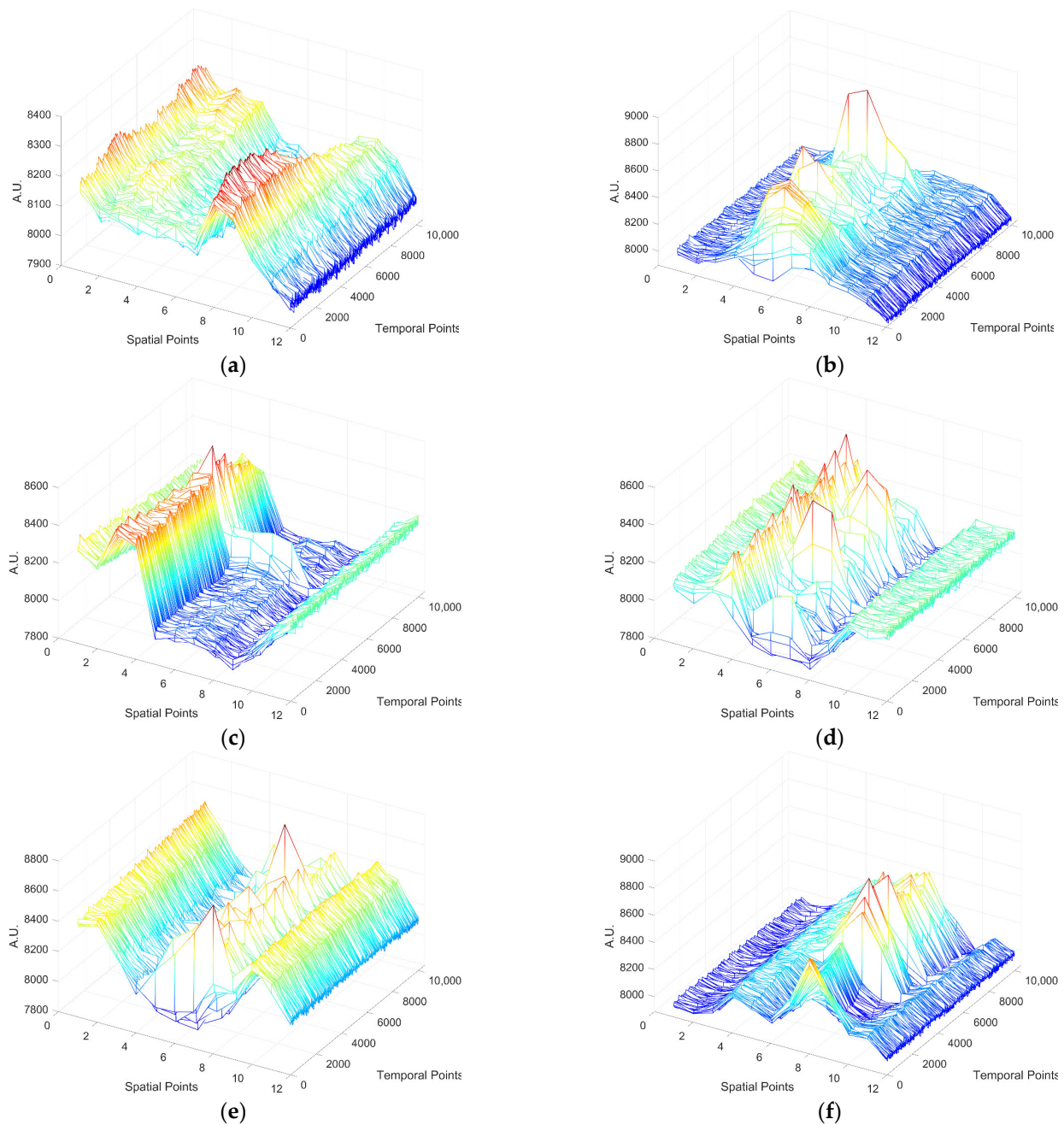


Figure 3. Temporal–spatial samples of the six events. (a) Background noise; (b) digging; (c) knocking; (d) watering; (e) shaking; (f) walking.

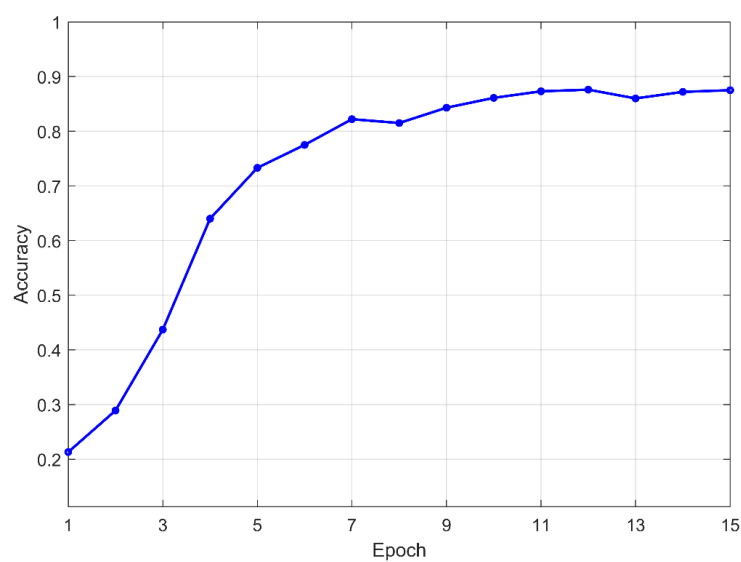
(1) Standard CNN model (Softmax as classifier)

To comply with the experimental conditions of Reference [8], the CNN architecture in this section is configured identically to that in the cited paper. The adopted network structure is presented in Table 2.

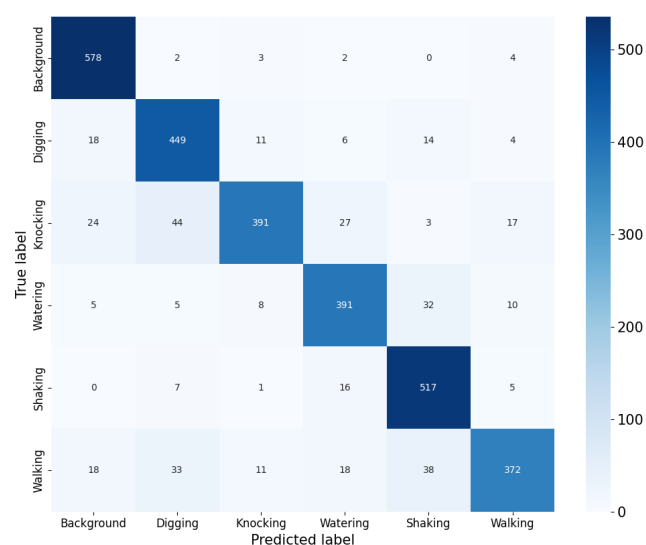
The variation in training accuracy with the number of epochs is shown in Figure 4. The accuracy was only 20% at the 1st epoch; it reached 87% at the 11th epoch, then tended to stabilize, and stabilized at 87.5% after the 15th epoch.

Table 2. Initialization parameters of the traditional CNN.

Layers	Kernel Size/Stride/Padding	Input Size	Activation
Conv1	$200 \times 3/50 \times 1/1$	$1 \times 10,000 \times 12$	ReLU
Pool1	$2 \times 2/2 \times 2/1$	$5 \times 197 \times 12$	/
Conv2	$20 \times 2/4 \times 1/1$	$5 \times 99 \times 7$	ReLU
Pool2	$2 \times 2/2 \times 2/0$	$10 \times 21 \times 8$	/
Flatten	/	$10 \times 10 \times 4$	/
Linear	Classifier: 1×6	400	/

**Figure 4.** Relationship between accuracy and training epochs.

The confusion matrix is shown in Figure 5, yielding a calculated recognition accuracy of 87.5%. It can be visually observed that after 15 epochs, a larger number of recognized instances for each event corresponds to a darker color in the confusion matrix. Accordingly, Event 0 and Event 4 exhibit relatively high accuracy, while considerable cross-confusion occurs, such as Event 2 being misclassified as Event 1 and Event 3 being misclassified as Event 4.

**Figure 5.** Confusion matrix of basic CNN.

(2) SVM model based solely on statistical features

The statistical features adopted in this paper are calculated as follows. In the time domain, the peak-to-peak value peak, variance σ^2 , standard deviation σ , skewness Skew, and crest factor pk_k are used, as shown in Equations (6)–(10). According to [15], the frequency feature extraction equations are as follows. The time-domain signal is transformed into the frequency domain via Fast Fourier Transform (FFT), and the statistical features of its f spectrum are the center frequency f_{CG} , root mean square frequency f_{rms} , frequency standard deviation f_{std} , spectral kurtosis f_{SK} , and information entropy entropy, as given in Equations (11)–(15). These features exhibit good discriminability in both the time and frequency domains of the DAS signal, which facilitates event classification and recognition.

$$\text{peak} = \max(x_i) - \min(x_i) \quad (6)$$

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2 \quad (7)$$

$$\sigma = \sqrt{\sigma^2} \quad (8)$$

$$\text{Skew} = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \mu)^3}{\left(\frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2\right)^{3/2}} \quad (9)$$

$$\text{pk_k} = \frac{1}{n} \sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma} \right)^4 - 3 \quad (10)$$

$$f_{CG} = \frac{\sum_{i=1}^N f_i \cdot P_f(i)}{\sum_{f=1}^N P_f(i)} \quad (11)$$

$$f_{rms} = \sqrt{\frac{\sum_{i=1}^L f_i^2 \cdot P_f(i)}{\sum_{i=1}^L P_f(i)}} \quad (12)$$

$$f_{std} = \sqrt{\frac{\sum_{i=1}^L (f_i - f_c)^2 \cdot P_f(i)}{\sum_{i=1}^L P_f(i)}} \quad (13)$$

$$f_{SK} = \frac{\frac{1}{N} \sum_{i=1}^N P_f^4(i)}{\left(\frac{1}{N} \sum_{i=1}^N P_f^2(i)\right)^2} - 3 \quad (14)$$

$$\text{entropy} = - \sum_{i=1}^n [p(i) \cdot \log_2 p(i)] \quad (15)$$

where x_i denotes the i -th data point, $\mu = \sum_{i=1}^n x_i / n$ is the mean value in the time domain, $P_f(i)$ is the amplitude at the i -th frequency point, f_i is the corresponding frequency, f_c is the center frequency, and N is the length of the frequency spectrum, $p(i) = \log_2[P_f(i) / \sum_{i=1}^N P_f(i)]$.

The SVM parameter settings were kept consistent with Reference [8] (i.e., the RBF kernel was adopted, the penalty factor C was set to 1, and the kernel parameters γ were configured as 'auto'). The classification results are shown in Figure 6.

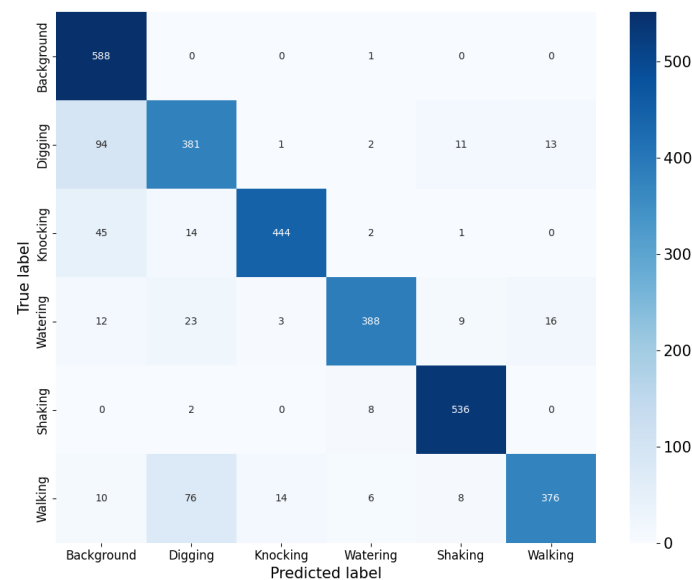


Figure 6. Confusion matrix of conventional SVM.

The confusion matrix is shown in Figure 6. According to the calculation, the conventional SVM model based on statistical features alone achieves an average accuracy of 87.9%. It can be clearly observed from Figure 6 that the SVM significantly outperforms the CNN in classifying Event 0. This is because statistical features such as σ , peak, and others have small values for Event 0, which differ substantially from those of other events, thereby enabling the SVM to classify this event more effectively. However, for the other events, the differences in their statistical values are rather small, leading to unsatisfactory classification performance.

(3) SVM model based solely on CNN features (conventional CNN-SVM model)

In addition to the two baseline models commonly used in Reference [8], there exists a directly end-to-end model architecture based on CNN and SVM, where the features extracted by the CNN are directly fed into the SVM for classification and recognition.

The confusion matrix of conventional CNN-SVM is presented in Figure 7. The conventional CNN-SVM model based on CNN features alone achieves an average accuracy of 93.9%. It can be observed from Figure 7 that the model recognizes Events 0, 2, 3, and 4 with relatively high accuracy, whereas Event 1 and 5 exhibit mutual confusion, and Event 5 suffers from a relatively high number of missed detections. Compared with the standard CNN, this model not only improves the accuracy but also reduces the training time. Compared with the SVM based on statistical features, it effectively enhances the recognition performance for Event 1, circumvents the limitations of statistical features, and improves the recognition balance of Event 1. A limitation, however, is that the decision boundary between Events 1 and 5 remains insufficiently clear, preventing effective separation of the two classes. Since this model relies solely on CNN features, it is prone to producing similar classification errors due to the inherent limitations of these features, as discussed above.

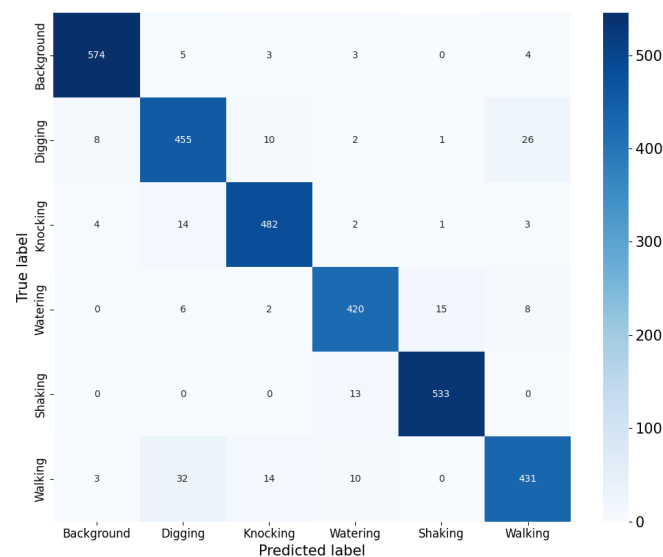


Figure 7. Confusion matrix of conventional CNN-SVM.

3.2. The Fused CNN-SVM Model

CNN features were extracted using the fused CNN-SVM model proposed in this paper (Figure 2) with the CNN parameters listed in Table 3, and were subsequently concatenated with statistical features before being fed into the SVM. The SVM was configured with the RBF kernel, the penalty parameter C of 1.0, and kernel parameters γ of 'auto' following Reference [8] to eliminate potential bias introduced by manual empirical tuning. The recognition results are presented in Figure 8.

Table 3. Initialization parameters of fused CNN-SVM structure.

Layers	Kernel Size/Stride/Padding	Input Size	Activation
Conv1	$15 \times 3/1 \times 1/7 \times 1$	$1 \times 10,000 \times 12$	ReLU
Pool1	$2 \times 1/1 \times 1/0$	$5 \times 197 \times 12$	/
Conv2	$20 \times 5/1 \times 1/1 \times 1$	$5 \times 99 \times 7$	ReLU
Pool2	$2 \times 1/1 \times 1/0$	$10 \times 21 \times 8$	/
Conv3	$30 \times 7/1/7 \times 1$	$96 \times 2500 \times 6$	/
Flatten	/	$96 \times 1 \times 1$	/
Linear	Classifier: 1×6	96	/

The fused CNN-SVM model achieved an accuracy of 96.6%. As is evident from the confusion matrix, the diagonal entries dominate overwhelmingly, with only a small number of off-diagonal misclassified samples, indicating excellent feature discrimination. Consequently, the classification accuracy is substantially improved, and misclassification across different events is effectively reduced.

To quantitatively evaluate the practical performance of the proposed model in disturbance recognition, we construct a confusion matrix based on the classification results on the test set. The diagonal elements of the matrix represent correctly classified samples, while the off-diagonal elements represent various misclassified samples. Following the definitions in [16], we compute the accuracy, precision, recall, and F1-score using the numbers of true positives (TPs), false positives (FPs), false negatives (FNs), and true negatives (TNs)

derived from the matrix. The specific formulas are presented in Equations (16)–(19). The specific performance results of the four models described above are presented in Table 4.

$$\text{accuracy} = \frac{\text{TPs} + \text{TNs}}{\text{TPs} + \text{FPs} + \text{TNs} + \text{FNs}} \quad (16)$$

$$\text{precision} = \frac{\text{TPs}}{\text{TPs} + \text{FPs}} \quad (17)$$

$$\text{Recall} = \frac{\text{TPs}}{\text{TPs} + \text{FNs}} \quad (18)$$

$$\text{F1 - score} = \frac{2 \times \text{TPs}}{2 \times \text{TPs} + \text{FPs} + \text{FNs}} \quad (19)$$

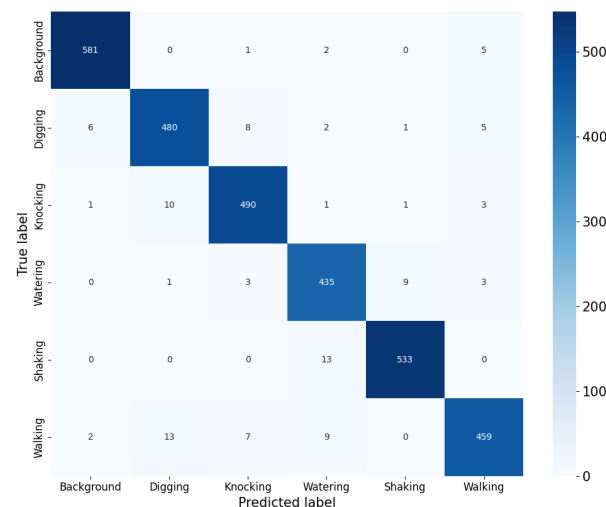


Figure 8. Confusion matrix of fused CNN-SVM.

As shown in Table 4, the proposed fused CNN-SVM model based on hybrid features significantly outperforms all baseline models in terms of overall recognition performance, achieving a classification accuracy of 96.6%. This corresponds to improvements of 9.1%, 8.7%, and 2.7% over the conventional CNN, the SVM based on statistical features, and the SVM based on CNN features, respectively. At the per-type level, the proposed model attains the optimal F1-score for all six event types and exhibits the smallest inter-type performance variance, indicating superior classification consistency and generalization capability across different events. The minimum F1-score among all types is 0.951, and the performance fluctuation is markedly lower than that of the other compared models. Further analysis of precision and recall reveals that the fused-feature model effectively mitigates the imbalance between precision and recall commonly encountered in single-feature models, achieving simultaneous improvements in both metrics for multiple event types, such as Events 0, 2, and 5. These experimental results demonstrate that the fusion strategy combining statistical and deep features yields substantial performance gains: the statistical features supply obvious differences, whereas the CNN-extracted deep features model latent abstract representations within the data. Together, they form a complementary feature type, ultimately delivering optimal overall performance and balanced category recognition among the compared models, with notably enhanced discriminability for easily confused event types.

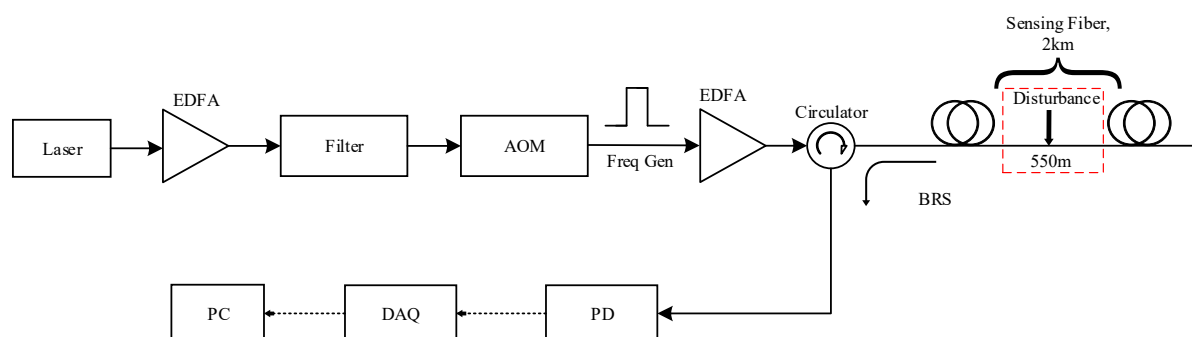
Table 4. A comparison of recognition performance.

Model	Label	Event Type	Precision	Recall	F1-Score
Traditional CNN (accuracy = 87.5%)	0	Background Noise	0.899	0.981	0.938
	1	Digging	0.832	0.894	0.862
	2	Knocking	0.920	0.771	0.861
	3	Watering	0.850	0.867	0.858
	4	Shaking	0.856	0.947	0.899
	5	Walking	0.903	0.759	0.825
SVM based on statistical features (accuracy = 87.9%)	0	Background Noise	0.785	0.998	0.879
	1	Digging	0.768	0.759	0.764
	2	Knocking	0.961	0.877	0.917
	3	Watering	0.953	0.860	0.905
	4	Shaking	0.949	0.982	0.965
	5	Walking	0.928	0.767	0.840
SVM based on CNN features (accuracy = 93.9%)	0	Background Noise	0.985	0.975	0.975
	1	Digging	0.889	0.906	0.897
	2	Knocking	0.943	0.953	0.948
	3	Watering	0.933	0.931	0.932
	4	Shaking	0.969	0.976	0.973
	5	Walking	0.913	0.880	0.896
SVM based on fused features (accuracy = 96.6%)	0	Background Noise	0.985	0.986	0.986
	1	Digging	0.952	0.956	0.954
	2	Knocking	0.963	0.968	0.966
	3	Watering	0.942	0.965	0.953
	4	Shaking	0.980	0.976	0.978
	5	Walking	0.966	0.937	0.951

4. Experimental Design and Data Preprocessing

4.1. Experimental System and Data Acquisition

To validate the effectiveness of the proposed CNN–SVM model under real-world operating conditions, field experiments were conducted based on an armored terrestrial optical cable and DAS. The structure of the experimental DAS system is shown in Figure 9.

**Figure 9.** The simple structure of the DAS system.

In the DAS system, a narrow-linewidth laser serves as the light source to generate a continuous optical signal. The signal is amplified by an erbium-doped fiber amplifier (EDFA), filtered by a dense wavelength-division multiplexer (DWDM), and modulated into optical pulses by an acousto-optic modulator (AOM). The pulsed light is injected into

a single-mode sensing fiber via an optical circulator, generating Rayleigh backscattering signals along the fiber. The backscattered light returns through the circulator and is received by a photodetector (PD), then sampled by a data acquisition card (DAQ) and stored in a computer for subsequent analysis.

The parameters of the DAS system used in the experiments are listed in Table 5. Specifically, an armored terrestrial optical cable with a length of 2 km and a burial depth of 0.3 m was laid. The corresponding environmental parameters in the three days are provided in Table 6.

Table 5. The system parameter settings.

Parameter Types	Settings
Output Wavelength	1550 nm
Output Power	3–30 mW
Fiber Length	2 km
Pulse Width	100 ns
Sampling Frequency	100–3000 Hz

Table 6. Parameter of experiment environment.

Event Type	Time	Noisy Level	Temperature	Humidity
Drilling	First Day	45.6 dB	35°	49%
Knocking	Second Day	45 dB	36°	55%
Digging	Third Day	61 dB	37°	49%
Background Noise	Randomly extracted from the undisturbed part of the above three events.			

4.2. Dataset Construction and Analysis

Four representative types of disturbance events are selected in this paper, namely background noise, drilling, knocking, and digging, which are labeled as Classes 1 to 4, respectively. These events cover normal operating conditions and typical anthropogenic disturbance scenarios and can effectively reflect the main types of interference that threaten the safety of terrestrial optical cables.

Drilling: Operating an electric drill near the burial location at a distance of approximately 0.3 m from the optical cable.

Knocking: Manually knocking on the armored optical cable with a hammer at the 550 m location.

Digging: Striking the ground near the cable route with the bucket of an excavator. Considering that the initial impact during excavation has the most significant effect on the cable, only the initial impact stage is selected for excavation event recognition.

Background noise: Signals acquired from fiber segments other than the disturbance location, representing normal operating conditions without obvious disturbances.

The field experimental scenarios of the four disturbance types are shown in Figure 10.

Due to differences in the duration of different disturbance events, the raw collected data contain a large number of blank or invalid signal segments. To ensure consistency of model inputs, the raw signals are cropped. Each disturbance sample consists of 100 time sampling points and 120 adjacent spatial sampling points, forming a unified spatiotemporal feature matrix. As shown in Figure 11, it presents examples of typical spatiotemporal distribution characteristics of the four types of disturbance events, where the x-axis and y-axis represent time and spatial sampling points, respectively, and the z-axis represents signal intensity. It can be observed from Figure 11 that the background noise exhibits uniform and gentle random fluctuations across the entire spatiotemporal domain without

significant energy concentration. The drilling event, due to the instantaneous vibration generated by a strong impact, manifests as intermittent temporal distribution with a wide spatial coverage. The knocking event, resulting from manual impacts, shows relatively fixed frequency characteristics and a more localized spatial distribution. The digging event has a wide range of influence, forming an energy envelope spanning a certain temporal duration.

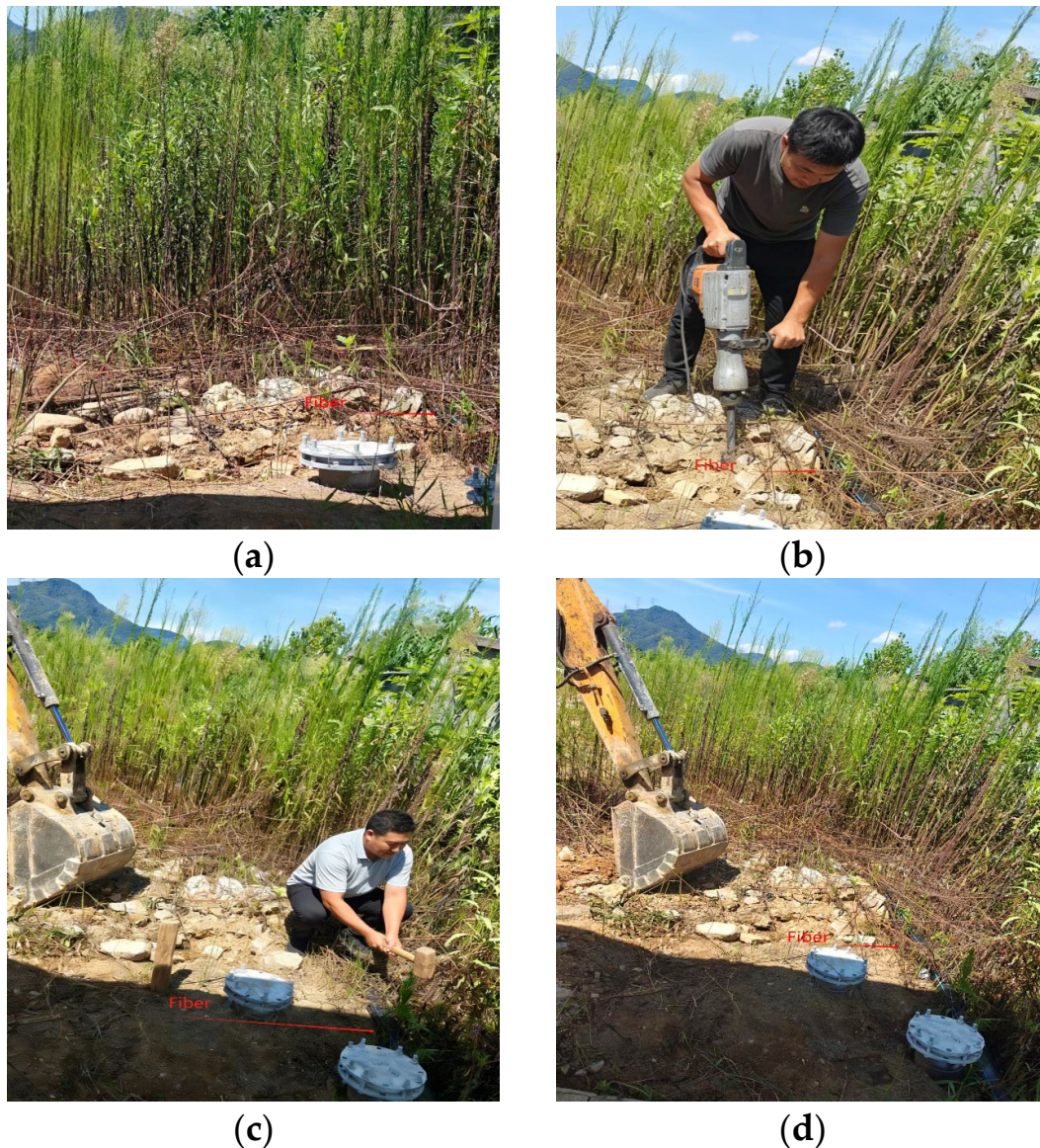


Figure 10. Field scenarios of four typical events. (a) Background noise; (b) drilling; (c) knocking; (d) digging.

After non-overlapping cropping, the final constructed dataset comprises 5664 independent samples. The sample counts for each disturbance-event type are listed in Table 7. The dataset was split into training, validation, and test sets at a ratio of 6:1:3, with the validation set used for SVM hyperparameter tuning. To prevent data leakage, the partitioning was performed at the event-recording level, ensuring that all cropped segments from the same continuous acquisition were assigned exclusively to a single subset. Stratified sampling was applied to preserve class proportions across all splits, with a fixed random seed for reproducibility. The test set (30%) was strictly withheld from all training and tuning procedures, and the final performance metrics are reported solely on this unseen set. For SVM

hyperparameter optimization, we employed a sequential search on the validation set—fixing two parameters while adjusting the third iteratively—to balance tuning effectiveness with computational efficiency.

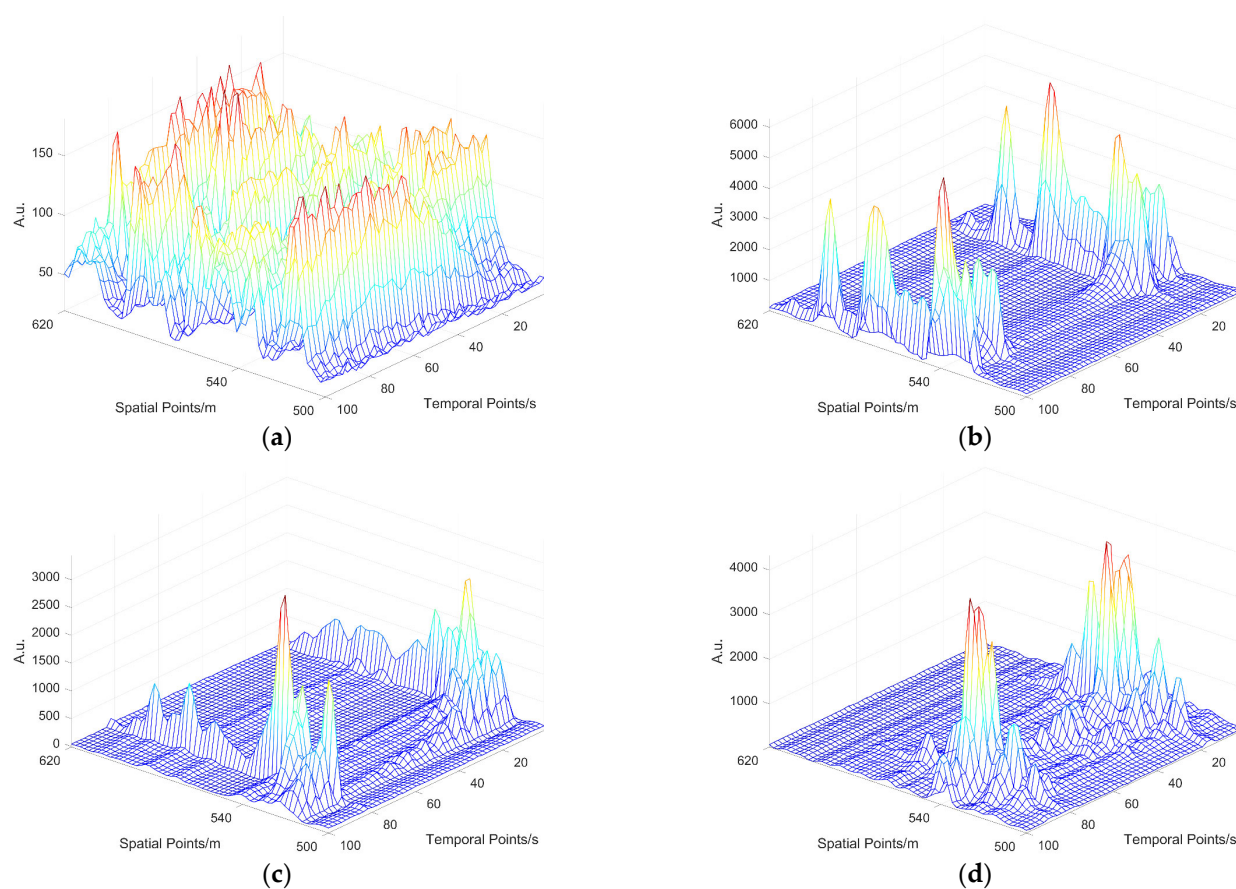


Figure 11. Temporal–spatial samples of the four typical events. (a) Background noise; (b) drilling; (c) knocking; (d) digging.

Table 7. The total number of each of four types of disturbances.

Event Type	Number	Label
Background Noise	1149	1
Drilling	1620	2
Knocking	1487	3
Digging	1408	4

5. Setting of the Model

5.1. SVM Parameter Configuration

To determine the optimal combination of SVM parameters, controlled variable tests were conducted for the penalty factor C , kernel function, and the parameter γ . To improve experimental efficiency, based on existing research [17], the kernel function was first fixed as RBF and γ was set to auto, while the optimal penalty factor C was sought. After C was determined, the optimal combination of kernel function and γ was then evaluated.

(1) Sensitivity of the penalty factor C

With the kernel function fixed as RBF and γ set to ‘auto’, the model accuracy was tested for C values ranging from 1 to 15. The results are shown in Figure 12. As can be observed, the highest accuracy of 98.2% was achieved when $C = 10$. For $C < 10$, the accuracy

ranged between 96.8% and 97.8%, and for $C > 10$, it fluctuated between 97% and 97.9%. Accordingly, $C = 10$ was adopted in the subsequent experiments.

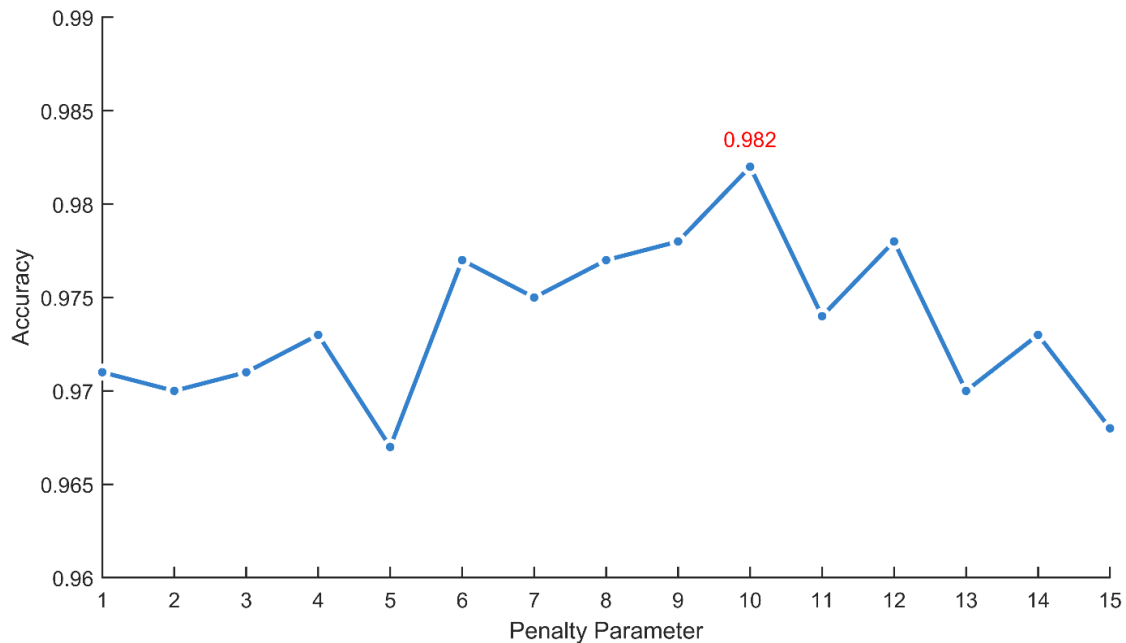


Figure 12. Accuracy variation with different penalty factors.

(2) Sensitivity of the parameter γ

With C fixed at 10 and the RBF kernel retained, the recognition accuracy was tested for γ values of 0.01, 0.02, 0.05, 0.1, 0.2, 0.5, 1.0, and the scale setting. The results are presented in Figure 13. The highest accuracy of 98.2% was obtained with $\gamma = \text{scale}$. When $\gamma = 0.1$, the accuracy was 97.8%; at $\gamma = 0.5$, it dropped to 96.3%; and at $\gamma = 1.0$, it decreased sharply to 89.6%. Therefore, $\gamma = \text{'auto'}$ was adopted in this study.

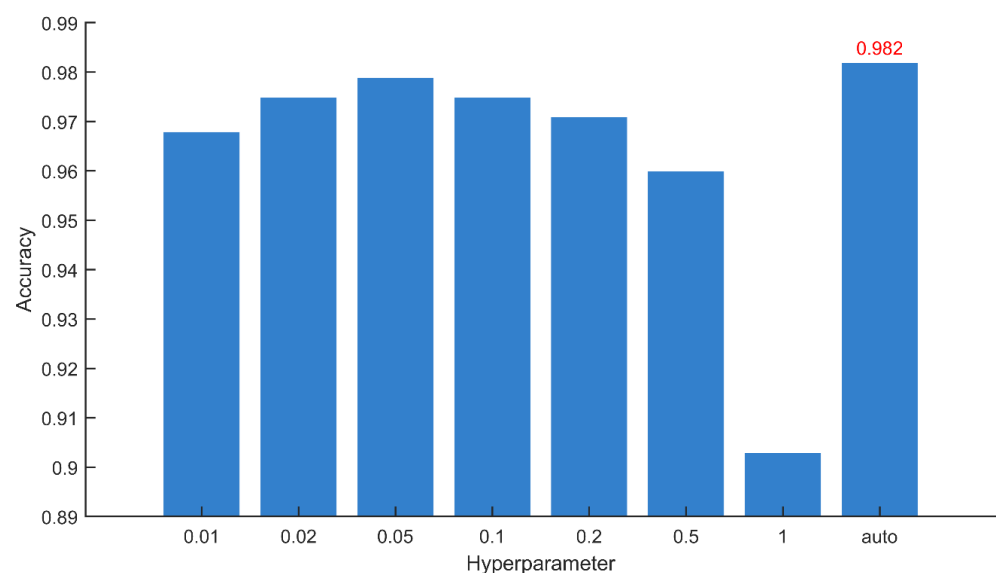


Figure 13. Accuracy variation with different kernel parameters.

(3) Comparison of kernel functions

With C fixed at 10 and γ at scale, the recognition accuracies of four kernel functions—sigmoid, Poly, linear, and RBF—were compared, as shown in Figure 14. The RBF kernel

achieved the highest accuracy (98.2%), followed by the Poly kernel (97.0%), the linear kernel (96.9%), and the sigmoid kernel (88.4%). The results indicate that the spatiotemporal features of DAS signals exhibit a nonlinear distribution, and the RBF kernel can more effectively map the samples into a high-dimensional space to achieve linear separability. Hence, the RBF kernel was adopted in this paper.

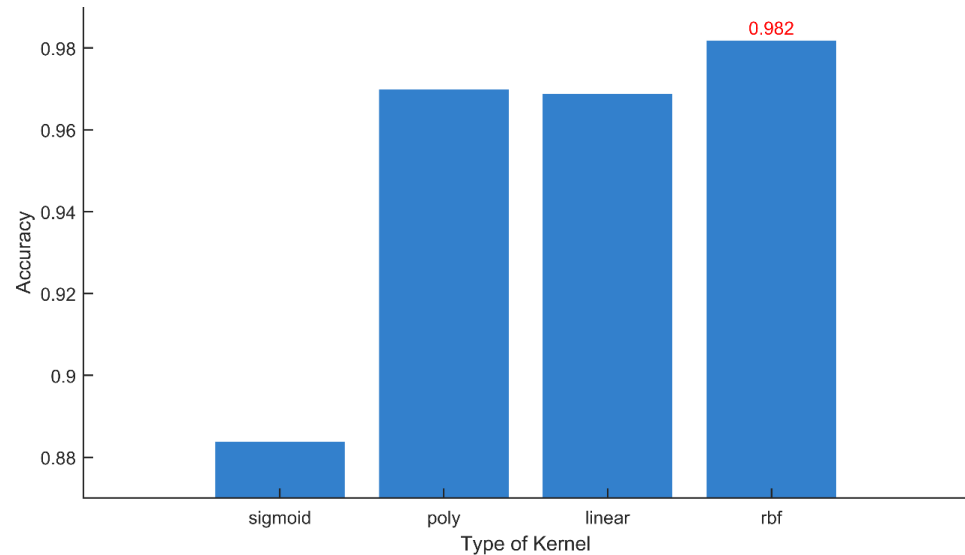


Figure 14. Accuracy variation with different types of kernel.

In conclusion, the penalty factor C was set to 10, the kernel parameter γ was set to 'auto', and the kernel function chosen was RBF, whose form is as follows:

$$k(x, x_i) = \exp\left(-\frac{\|x - x_i\|^2}{2\sigma^2}\right) \quad (20)$$

where σ represents the width parameter of the kernel function; x and x_i denote the two sample points in the input space [18].

5.2. CNN Structure Configuration

Similarly to as described above, the CNN employs the ReLU activation function and a dropout rate of 0.1 is applied to suppress overfitting [19]. The specific configurations of the convolutional and pooling layers are presented in Table 8. The CNN configured as specified in the table is then used for feature extraction in this part.

Table 8. Initialization parameters of 3-layer fused CNN-SVM.

Layer	Name	Activation	Stride, Padding
1	Input	$1 \times 100 \times 120$	/
2	Conv_1	$5 \times 3 \times 7$	5, 1
3	Maxpool	2×2	/
4	Conv_2	$10 \times 5 \times 5$	5, 0
5	Conv_3	$10 \times 3 \times 3$	1, 1
6	Maxpool	2×2	/
7	Dropout	$10 \times 3 \times 3$	/
8	Fully Connected	$10 \times 1 \times 1$	/

For comparison, a simple one-layer fused CNN-SVM architecture was constructed in this paper to verify the influence of the network structure on classification accuracy. The specific network settings are shown in the Table 9.

Table 9. Initialization parameters of 1-layer fused CNN-SVM.

Layer	Name	Activation	Stride, Padding
1	Input	$1 \times 100 \times 120$	/
2	Conv	$10 \times 50 \times 60$	$50 \times 60, 0$
3	Maxpool	2×2	/
4	Dropout	$10 \times 3 \times 3$	/
5	Fully Connected	$10 \times 1 \times 1$	/
6	Output	$10 \times 1 \times 1$	/

6. Results of Testing

6.1. Normalization Processing

Based on the above analysis, disturbance signals can be effectively recognized through strongly distinctive features. Prior to model training, all samples were normalized to the range [0, 1] to improve numerical stability and enhance model generalization [12]. Min–max normalization was applied as follows:

$$x_i^* = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \quad (21)$$

where x_i^* represents the outcome of normalization, x_i denotes the sample data, $x_{\min}$ indicates the sample's lowest value, $x_{\max}$ refers to the sample's highest value.

6.2. Feature Visualization

After normalization, the 20-dimensional fused features extracted by the three-layer and one-layer fused CNN-SVM models are visualized in Figures 15 and 16, respectively.

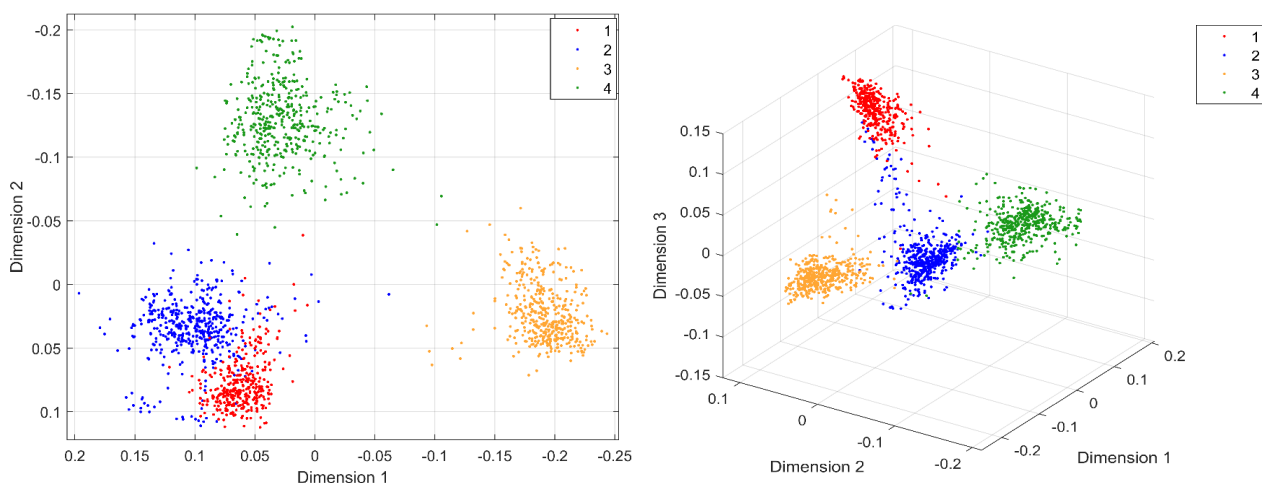


Figure 15. Feature visualization of 1-layer fused CNN-SVM model.

It can be clearly observed from the normalized feature distributions that in the features extracted by the one-layer fused CNN-SVM the features of the drilling event (Event 2) exhibit significant overlap with those of background noise (Event 1) and the digging event (Event 4), resulting in blurred class boundaries. Only the features of the knocking event (Event 3) form an independent distribution region. By contrast, in the features extracted by the three-layer fused CNN-SVM, the features of all four disturbance events are more

compact, with clear inter-class boundaries, and the previously overlapping regions have essentially disappeared.

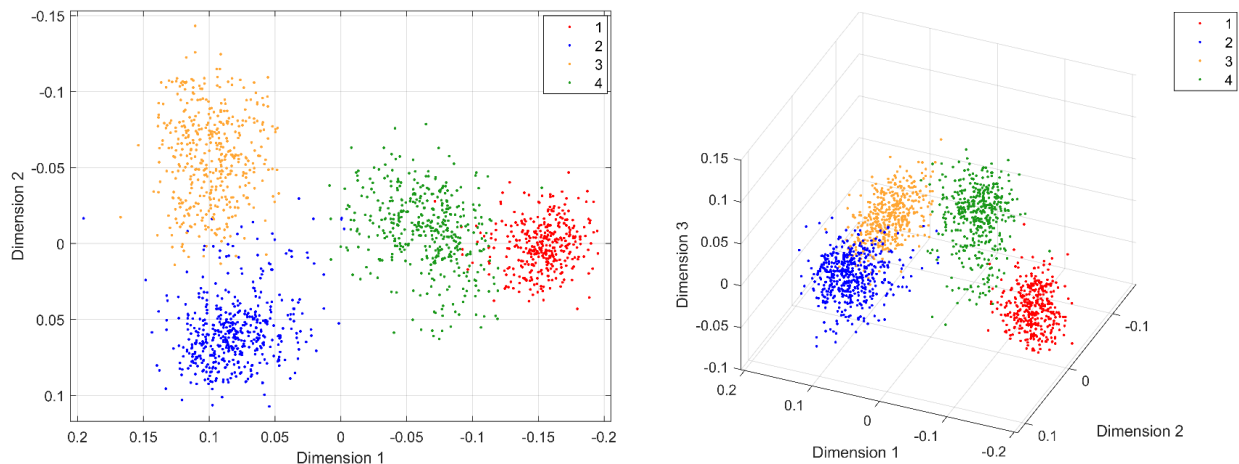


Figure 16. Feature visualization of 3-layer fused CNN-SVM model.

6.3. Comparison of Recognition Results

Furthermore, we use the receiver operating characteristic (ROC) curve to verify model stability and event recognition authenticity, with the false positive rate (FPR) on the horizontal axis and the true positive rate (TPR) on the vertical axis [20].

(1) One-layer fused CNN-SVM model

The detailed recognition results of one-layer fused CNN-SVM model are shown in Figures 17 and 18. The model achieves a recognition accuracy of 96.7%. It can be observed from the confusion matrix that background noise, knock, and digging events are classified with relatively high accuracy and without notable misclassification. However, a considerable number of drilling events are misclassified as background noise, indicating that the features of the drilling event overlap with those of the noise event. This observation is consistent with the visualization in Figure 15, where the feature quality of this event is relatively low, causing blurred decision boundaries and making it difficult for the SVM classifier to distinguish between their similar features. This issue stems fundamentally from the limited receptive field of the one-layer convolutional network, which is incapable of capturing high-order features of the disturbance signals. As a result, the feature discrimination between the drilling event and other event types is insufficient, constituting a structural bottleneck in the model's generalization performance and limiting the overall recognition performance at a relatively low level.

The computed evaluation metrics are presented in Table 10, and the data therein further corroborate the above analysis. The one-layer fused CNN-SVM model demonstrates sound overall performance; however, an imbalance between precision and recall is observed across different event types: background noise achieves saturated recall yet insufficient precision, whereas the drilling event exhibits the opposite pattern. The immediate cause is that the classifier tends to assign ambiguous samples to background noise and digging events, sacrificing the recall of the drilling event in favor of higher accuracy for the majority classes. The deeper cause lies in the excessively shallow receptive field of the single-layer CNN, which is incapable of capturing complete structural features, thereby leading to a series of classification deficiencies.

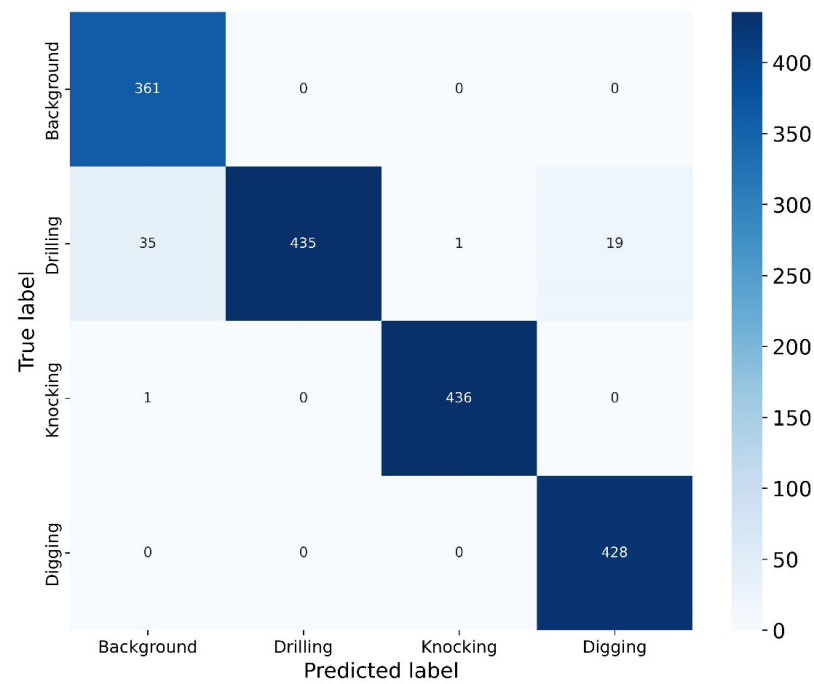


Figure 17. The confusion matrices of the 1-layer fused CNN-SVM.

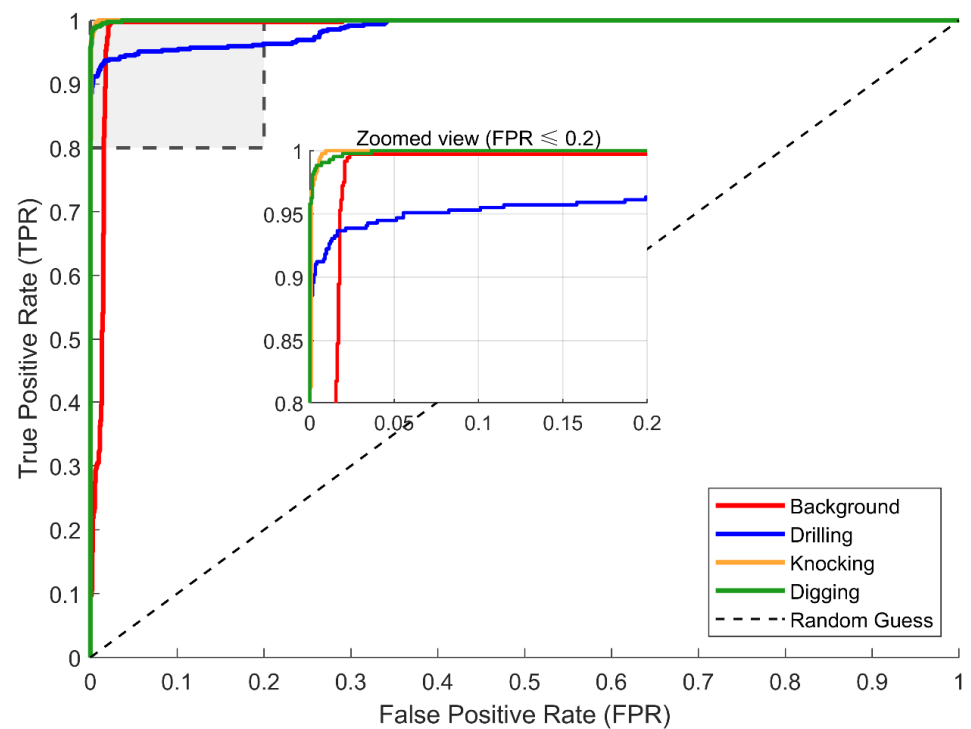


Figure 18. The ROC of the 1-layer fused CNN-SVM.

Table 10. The performance of 1-layer fused CNN-SVM.

Classifier	Event Type	Precision	Recall	F1-Score
1-Layer fused CNN-SVM	Background	0.909	1.000	0.953
	Drilling	1.000	0.888	0.941
	Knocking	0.997	0.954	0.997
	Digging	0.958	1.000	0.978

(2) Three-layer fused CNN-SVM model

The recognition results of the three-layer fused CNN-SVM model are presented in Figures 19 and 20. The model achieves a recognition accuracy of 98.5%. As can be observed from the confusion matrix, among the four event types in this experiment, only a small number of background noise samples are misclassified as digging events, which is consistent with the feature distribution shown in Figure 16. Moreover, the ROC curves converge rapidly, effectively compensating for the insufficient representation of the drilling event by the one-layer CNN structure, thereby demonstrating the effectiveness of the three-layer CNN architecture in recognizing complex overlapping features.

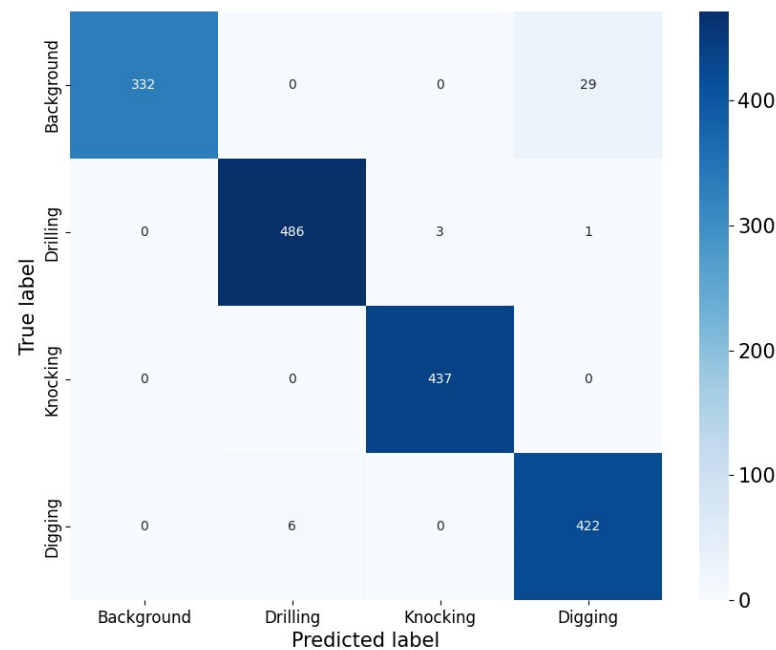


Figure 19. The confusion matrices of the 3-layer fused CNN-SVM.

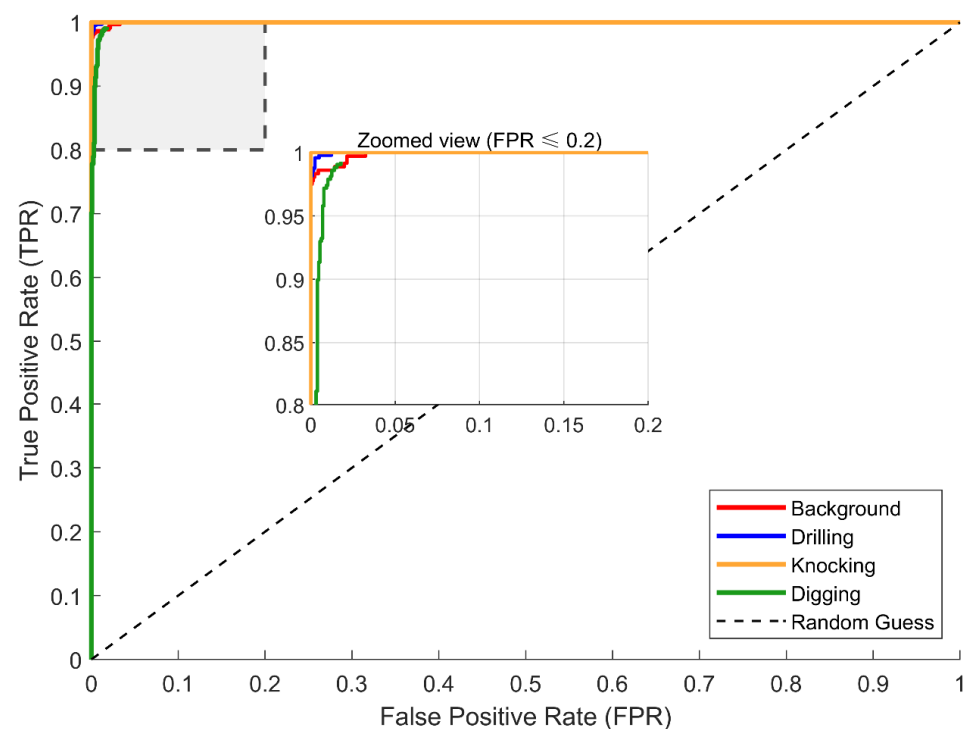


Figure 20. The ROC of the 3-layer fused CNN-SVM.

As can be readily observed from Table 11, the three-layer fused CNN-SVM achieves a substantial overall improvement over the one-layer model, with markedly enhanced balance between precision and recall across event types. The recall of the drilling event increases significantly without an appreciable decline in precision, indicating that the deepened network expands the receptive field, thereby capturing the long-term structural characteristics of drilling events and enhancing sample compactness. This, in turn, reduces the boundary shift in the SVM under class imbalance. The precision of background noise rises to saturation, while its recall undergoes a minor adjustment and the false alarms presented in the previous one-layer model are eliminated. The precision and recall of the digging event exhibit slight reciprocal changes, reflecting decision-boundary refinement driven by the enhanced discriminability of deep features. Both metrics for the knocking event remain saturated. Overall, the three-layer structure resolves the principal structural deficiencies of the one-layer model by improving the representation quality, resulting in more balanced recognition metrics across all event types and yielding stronger overall classification stability and generalization capability.

Table 11. The performance of the 3-layer fused CNN-SVM.

Classifier	Event Type	Precision	Recall	F1-Score
3-Layer fused CNN-SVM	Background	1	0.945	0.972
	Drilling	0.994	0.996	0.995
	Knocking	0.995	1	0.998
	Digging	0.955	0.993	0.974

Comparing the two models shows that by using three-layer fused CNN-SVM to extract deep features rather than relying on one-layer fused CNN-SVM the three-layer structure of model better characterizes the distinct features of different cable disturbance events. It comprehensively improves the recognition accuracy and overall F1-score for all event types, especially addressing the problem of missed detection in drilling events. The proposed model offers enhanced classification balance and practical engineering applicability, confirming the effectiveness and superiority of combining deep features with traditional machine learning methods for disturbance signal recognition.

7. Conclusions

This paper proposes a fused CNN-SVM model that combines high-dimensional features extracted by a CNN with handcrafted statistical features, which are then fed into an SVM for recognition. Simulations were performed on open-source DAS data containing six event types (e.g., shaking, watering). The results show that compared with the conventional CNN, the SVM relying on a single feature type, and the conventional CNN-SVM, the proposed model achieves markedly improved accuracy, with a maximum gain of 9.1%. Subsequently, field tests were conducted on a 2 km armored optical cable, and a dataset comprising 5664 samples was constructed, covering four typical disturbance types: background noise, drilling, knocking, and digging. The dataset was split into training, validation, and test sets at a 6:1:3 ratio with strict event-level partitioning and stratified sampling; the test set was completely withheld from all training and tuning procedures to ensure unbiased generalization evaluation. Using a sequential search on the validation set, we first determined the SVM configuration (RBF kernel, $C = 10$, $\gamma = \text{'auto'}$) and then evaluated the network depth by comparing one-layer and three-layer CNN structures. This stepwise approach validated our model configuration with controlled computational cost. The experimental results demonstrate that the three-layer fused CNN-SVM model attains a

recognition accuracy of 98.5%, with high precision, recall, and F1-score maintained across all event types, exhibiting well-balanced overall classification performance.

Furthermore, the findings confirm that the hybrid feature combination strategy can effectively enhance the feature discriminability of DAS signals and that SVM can construct robust decision boundaries under limited training samples. Given the high cost of data acquisition and annotation in optical cable monitoring scenarios, as well as the need for real-time and accurate identification of anthropogenic disturbances, the proposed method is well-suited for such engineering applications. Despite these promising results, several limitations remain: systematic ablation studies on kernel sizes, dropout rates, and the contribution of individual statistical features are yet to be conducted; and the SVM parameter search, while effective, is heuristic rather than globally optimized. These aspects will be addressed in future work. Future work will extend the applicability of the model to more complex disturbance conditions and explore online learning and incremental learning strategies to improve the model's environmental adaptability under complex dynamic operating conditions. In addition, comprehensive benchmarking against state-of-the-art deep learning models (e.g., ResNet, LSTM, and Transformer-based classifiers) will be conducted to further validate the performance of the proposed method.

Author Contributions: Conceptualization, J.Z., X.Q. and X.W.; methodology, J.Z.; software, J.Z. and X.Q.; validation, J.Z.; data curation, J.Z. and X.Q.; writing—original draft preparation, J.Z.; writing—review and editing, J.Z. and X.Q.; supervision, X.Q. and X.W.; project administration, X.W.; funding acquisition, X.W. 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 supporting the conclusions of this study will be available on request from the corresponding author.

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

References

1. Mahmoud, S.; Katsifolis, J. Robust event classification for a fiber optic perimeter intrusion detection system using level crossing features and artificial neural networks. *Fiber Opt. Sens. Appl.* **2010**, *7677*, 767708. [[CrossRef](#)]
2. Wu, H.; Liu, X.; Xiao, Y.; Rao, Y.-J. A dynamic time sequence recognition and knowledge mining method based on the hidden Markov models (HMMs) for pipeline safety monitoring with Φ -OTDR. *J. Light. Technol.* **2019**, *37*, 4991–5000. [[CrossRef](#)]
3. Wu, H.; Yang, S.; Liu, X.; Xu, C.; Lu, H.; Wang, C.; Qin, K.; Wang, Z.; Rao, Y.; Olaribigbe, A.O. Simultaneous extraction of multi-scale structural features and the sequential information with an end-to-end mCNN-HMM combined model for fiber distributed acoustic sensor. *J. Light. Technol.* **2021**, *39*, 6606–6616.
4. Hu, S.; Hu, X.; Li, S. Research on Φ -OTDR perimeter security monitoring based on GAF-HorNet. *Laser Optoelectron. Prog.* **2024**, *61*, 0906005.
5. Li, X.; Wang, T.; Yuan, C. Fault prediction of pumping well based on CNN-SVM. *J. Phys. Conf. Ser.* **2025**, *3059*, 012003. [[CrossRef](#)]
6. Liu, X.; Wu, H.; Wang, Y.; Tu, Y.; Sun, Y.; Liu, L.; Song, Y.; Wu, Y.; Yan, G. A fast accurate attention enhanced ResNet model for fiber-optic distributed acoustic sensor (DAS) signal recognition in complicated urban environments. *Photonics* **2022**, *9*, 677.
7. Zhang, J.; Lou, S.; Liang, S. Study on pattern recognition of φ -OTDR distributed optical fiber disturbance sensing system based on SVM algorithm. *Infrared Laser Eng.* **2017**, *46*, 219–225. [[CrossRef](#)]
8. Cao, X.; Su, Y.; Jin, Z.; Yu, K. An open dataset of φ -OTDR events with two classification models as baselines. *Results Opt.* **2023**, *10*, 100372. [[CrossRef](#)]
9. Xing, W.; Zhao, Z.; Xing, M.; Gan, W.; Li, G. Experimental Research on Submarine Cable Disturbance Detection with Optical Fiber Sensing. *Opt. Precis. Eng.* **2023**, *31*, 1432–1442. [[CrossRef](#)]
10. Ge, Z. Research and Application of Distributed Fiber Optic Vibration Event Identification Based on Time-Space Signals Analysis. Master's Thesis, Huazhong University of Science and Technology, Wuhan, China, 2022.
11. Wu, M.; Shen, Y.; Chen, Q.; Wang, D.; Li, S.; Xie, S.; Yin, J.; Xu, Y. Pattern recognition method of distributed optical fiber sensing signal based on multi-classification Gaussian SVM. *Laser Technol.* **2025**, *49*, 128–134.
12. Zhou, L.; Yan, P.; Li, X.; Liu, T.; Liu, Z.; Jia, W. Research on prediction model of high geothermal tunnels temperature based on CNN-SVM. *Energy Build.* **2025**, *348*, 116505.

13. Sheng, X.; Wang, Y.; Wan, L.; Wu, Z.; Shi, B.; Li, D. Underwater target interrupted track association and connection algorithm. *Acta Acust.* **2026**, *51*, 243–254.
14. Shi, Y.; Wang, Y.; Wang, L.; Zhao, L.; Fan, Z. Multi-event classification for Φ -OTDR distributed optical fiber sensing system using deep learning and support vector machine. *Optik* **2020**, *221*, 165373.
15. Qian, J.; Guo, J. Overview of Feature Extraction and Recognition Methods for Fiber Optic Vibration Signals. *Study Opt. Commun.* **2024**, *50*, 23011601–23011609.
16. Zhan, Y.; Liu, L.; Li, K. Application of machine learning for signal recognition in distributed fibre optic acoustic sensing technology. *IET Optoelectron.* **2024**, *18*, 81–95. [[CrossRef](#)]
17. Song, X.; Jiang, X.; Luo, J.; Yao, J. Analysis of the Inter-class Distance-based Kernel Parameter Evaluating Method for RBF-SVM. *Acta Armamentarii* **2012**, *33*, 203–208.
18. Franceschi, L.; Donini, M.; Perrone, V.; Klein, A.; Archambeau, C.; Seeger, M.; Pontil, M.; Frasconi, P. Hyperparameter Optimization in Machine Learning. *Found. Trends Mach. Learn.* **2025**, *18*, 1054–1201. [[CrossRef](#)]
19. Peng, R.; Zhou, Y.; Yuan, M.; Zheng, C.; Li, X. Deep neural network-based vibration signal classification in distributed coherent Reyleigh fiber sensor security pre-warning systems. *J. Appl. Acoust.* **2023**, *42*, 833–843.
20. Rajarathinam, G.; Rajakannan, R.; Chinnasamy, R.; Karthikeyan, K.; Thangavel, G.; Balasubramanian, M.G.; Srinivasan, B.; Radhakrishnan, G.; Ranjith. A Hybrid CNN-BiLSTM Model for Real-Time Activity Classification in Distributed Acoustic Sensing Systems. *Trait. Signal* **2026**, *43*, 641–647.

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