electronics

MDPI

Article

Using Phase-Sensitive Optical Time Domain Reflectometers to Develop an Alignment-Free End-to-End Multitarget Recognition Model

Nachuan Yang [1,2,*], Yongjun Zhao [1], Fuqiang Wang [2] and Jinyang Chen [3]

[1] Data and Target Engineering Institute, PLA Strategic Support Force Information Engineering University, Zhengzhou 450001, China
[2] Zhengzhou Xinda Institute of Advanced Technology, Zhengzhou 450001, China
[3] Research Institute for National Defense Engineering of Academy of Military Science PLA, Luoyang 471023, China
* Correspondence: paper.robert.young@gmail.com

Abstract: This pattern recognition method can effectively identify vibration signals collected by a phase-sensitive optical time-domain reflectometer (Φ-OTDR) and improve the accuracy of alarms. An alignment-free end-to-end multi-vibration event detection method based on Φ-OTDR is proposed, effectively detecting different vibration events in different frequency bands. The pulse accumulation and pulse cancellers determine the location of vibration events. The local differential detection method demodulates the vibration event time-domain variation signals. After the extraction of the signal time-frequency features by sliding window, the convolution neural network (CNN) further extracts the signal features. It analyzes the temporal relationship of each group of signal features using a bidirectional long short-term memory network (Bi-LSTM). Finally, the connectionist temporal classification (CTC) is used to label the unsegmented sequence data to achieve single detection of multiple vibration targets. Experiments show that using this method to process the collected 8563 data, containing 5 different frequency bands of multi-vibration acoustic sensing signal, the system F1 score is 99.49% with a single detection time of 2.2 ms. The highest frequency response is 1 kHz. It is available to quickly and efficiently identify multiple vibration signals when a single demodulated acoustic sensing signal contains multiple vibration events.

Keywords: distributed acoustic sensing; Φ-OTDR; fiber optic sensing; convolutional neural network; pattern recognition; connectionist temporal classification; Bi-LSTM

check for updates

Citation: Yang, N.; Zhao, Y.; Wang, F.; Chen, J. Using Phase-Sensitive Optical Time Domain Reflectometers to Develop an Alignment-Free End-to-End Multitarget Recognition Model. *Electronics* **2023**, *12*, 1617. https://doi.org/10.3390/electronics12071617

Received: 28 February 2023
Revised: 25 March 2023
Accepted: 28 March 2023
Published: 29 March 2023

1. Introduction

Fiber optic sensing technology can use backscattered light in optical fibers as an information carrier to sense and transmit externally changing physical quantities. Among them, a phase-sensitive optical time-domain reflectometer (Φ-OTDR) is a distributed fiber optic sensing technology based on the detection of Rayleigh backscatter signals (RBSs) [1], which utilize optical fibers to measure sounds or vibrations over long distances [2]. A single device can achieve long-distance [3] and multi-point [4] distributed detection while having a variety of advantages, such as anti-electromagnetic interference, flame and explosion-proof, and corrosion resistance [5]. It is widely used in perimeter security [6], pipeline detection [7], train detection [8,9], etc. According to different implementation methods, Φ-OTDR can be categorized into two detection methods: direct detection scheme and coherent detection scheme [10]. The coherent detection scheme allows the demodulation of the amplitude and phase changes of the Rayleigh scattered light caused by vibrations through coherent optical detection, which can record the dynamic processes of vibrations and obtain high-frequency response [11].

However, two critical problems with coherent detection need to be solved. First, coherent detection requires demodulating the backscattered light amplitude at each position on the fiber, determining the vibration position based on the amplitude change, and then demodulating the phase change at each position, which can result in an excessive amount of computation in the system. Secondly, the high sensitivity and long monitoring distance of Φ-OTDR coherent detection can result in high nuisance-alarm rates (NARs). This causes the subsequent signal demodulation algorithm and pattern recognition processing to be critical factors in improving the overall detection performance of the system. The electrical coherent demodulation technique first acquires the beat-frequency optical signal generated by local and reference light through an optical conversion device. Then, it mixes with an electrical local oscillator (L.O.) to obtain the signal [11,12]. In contrast, the coherent digital technique generates a standard signal with the same frequency as the beat-frequency signal by a computer and finally performs quadrature demodulation [13]. This reduces the thermal and scattering noise of electronic devices and makes signal processing more flexible, but how to reduce the computational effort still needs to be solved. After demodulating the slow time domain signal of the dynamic vibration process, the ability of the system to effectively distinguish between different vibration events is the crucial issue in determining whether the false alarm rate can be reduced [14]. Classification and recognition of slow time domain vibration signals can be divided into traditional time series classification methods and deep learning methods. Traditional time series classification methods first extract time domain signal features based on expert experience and then improve the accuracy of Φ-OTDR for vibration event recognition by applying an appropriate classifier [10]. For instance, after removing the signal noise using a smoothing filtering method, multiple features in the signal dataset's time domain and frequency domain are used to train a random forest classifier [15]. In addition, features such as empirical mode decomposition (EMD) [14], Mel-scale frequency cepstral coefficients (MFCC) [16], wavelet decomposition (W.D.) [17], and wavelet packet decomposition (WPD) [18] can describe the vibration signal well, and, eventually, combined with extreme gradient boosting (XGBoost) [14], k-nearest neighbors (k-NNs) [16], support vector machines (SVMs) [19], Gaussian mixture models (GMMs) [20], and other classifiers can obtain satisfactory classification results. In contrast, deep learning approaches can automatically extract high-dimensional features, avoiding the expert knowledge required for human feature extraction and the dependence on the system's operating environment [21]. The most direct and effective method is to put the original or denoised data into a one-dimensional CNN to extract the signal features, as this is a simple network structure and does not require any transformation or other manual work [22]. Another approach is to convert the 1D time-domain data demodulated by Φ-OTDR into 2D image data [23] then extract depth features using a CNN and implement vibration signal recognition with a classifier. The above deep learning methods do not use the sequential feature information of time sequences. In contrast, a long short-time memory (LSTM) network can distinguish different sequential data based on the relational features of sequences in different time dimensions, which has been studied in many fields, such as speech recognition [24] and medicine [25]. In Φ-OTDR pattern recognition studies, vibration signals can be scanned periodically and then divided into fixed-size frames, which are then fed into a combined detection model of CNN and LSTM for abnormal vibration identification and localization, where the convolutional part extracts the spatial image features and the LSTM analyzes the temporal relationship between signals [26]. As an improved version of LSTM, Bi-directional LSTM (Bi-LSTM) can acquire the context's features more efficiently and perform better sequence detection [27]. However, all of the above methods treat an acquisition sample as a single vibration event, while, for dynamic time series identification problems, the proportion of valid data in a sample to the overall data is not fixed, i.e., the position of the label compared to the valid part of the input sequence is uncertain. To obtain valid data in the sample more accurately, data alignment becomes an urgent problem, which usually has two solutions: the Hidden Markov Model (HMM) and the attention mechanism. Along with extracting local structural features, the

HMM uses statistical methods to mine sequential contextual information in the signal to identify the sequential state evolution of vibration and can explain the development process of the occurring event [28]. Besides combining with GMM, the HMM can also accomplish vibration recognition with depth features extracted by CNN [29]. On the other hand, the HMM has a condition to be satisfied during training, i.e., the corresponding annotation has to be predetermined for each frame in the training data. In order to obtain the annotations, it is necessary to use the existing model to force the alignment between the training data sequence and the annotated sequence, which is time consuming and often has biases in preparing these annotations. LSTMs [30], Bi-LSTMs [31], or CNNs [32] that incorporate an attention mechanism can, without forcing alignment, make the features extracted by the model focus more on locally valid details of the vibration signal rather than on global features, which helps to improve the overall recognition rate.

In the above methods, only the same type of vibration source exists in a single vibration datum. However, various vibration activities may occur together during single datum acquisition for long-distance vibration monitoring environments. Therefore, the Φ-OTDR dynamic time series identification problem with multiple vibration sources in a single detection remains to be discussed as a frequent scenario in Φ-OTDR vibration classification studies. Based on the above analysis, this problem requires the extraction of features of signal frames by deep learning methods, combined with non-forced alignment annotation methods, to form an end-to-end multi-vibration time series classification model. Among them, connectionist temporal classification (CTC) has been applied in automatic speech recognition as a non-forced alignment multi-target annotation method [33]. CTC shows two advantages: it directly labels unsegmented sequences without needing data alignment and one-by-one labeling; and CTC outputs the predicted probabilities of sequences directly, thus, forming label sequences without external post-processing [34] can be applied to end-to-end multi-vibration time series classification.

This paper aims to rapidly demodulate the dynamic process of vibration generation by Φ-OTDR coherent detection and to identify and localize multiple vibration targets in a single detection. The model is trained and evaluated with the collected dataset. The contributions of this study are: (1) the digital coherent detection technique and direct light intensity detection method of Φ-OTDR are combined to achieve fast demodulation of the beat-frequency signal phase; (2) an end-to-end multi-vibration time series classification method is proposed to identify multiple different vibration targets in one detection; (3) the effects of different feature extraction networks on the detection performance of the improved end-to-end model are tested and discussed; (4) the identification method is complemented when multiple vibrational activities are present in one coherent detection vibrational signal. The laborious task of aligning a large number of multiple target sequences is avoided.

This paper is organized as follows: Section 2 describes the hardware system for Φ-OTDR vibration signal acquisition. The variation rules of the original signal after back scattering are analyzed. Section 3 illustrates the overall recognition process of the Φ-OTDR vibration time series and presents the general framework of signal identification. The theoretical principle of demodulating the vibration time series signal from the original signal is introduced. The signal pre-processing method is proposed, followed by refining the recognition model structure for multi-target vibration data. Section 4 focuses on the experimental preparation, including dataset generation methods, the impact of different feature extraction network components on data detection performance, and evaluation metrics. Section 5 discusses the experimental steps and summarizes the experimental results. The effectiveness of fast digital quadrature demodulation of vibration signals is verified. In addition, the performance of the improved end-to-end model is compared with those of other feature extraction networks. The last section summarizes the paper and directions for future work.

2. Sensing Principle

2.1. Φ-OTDR of Direct Detection Combined with Coherent Detection

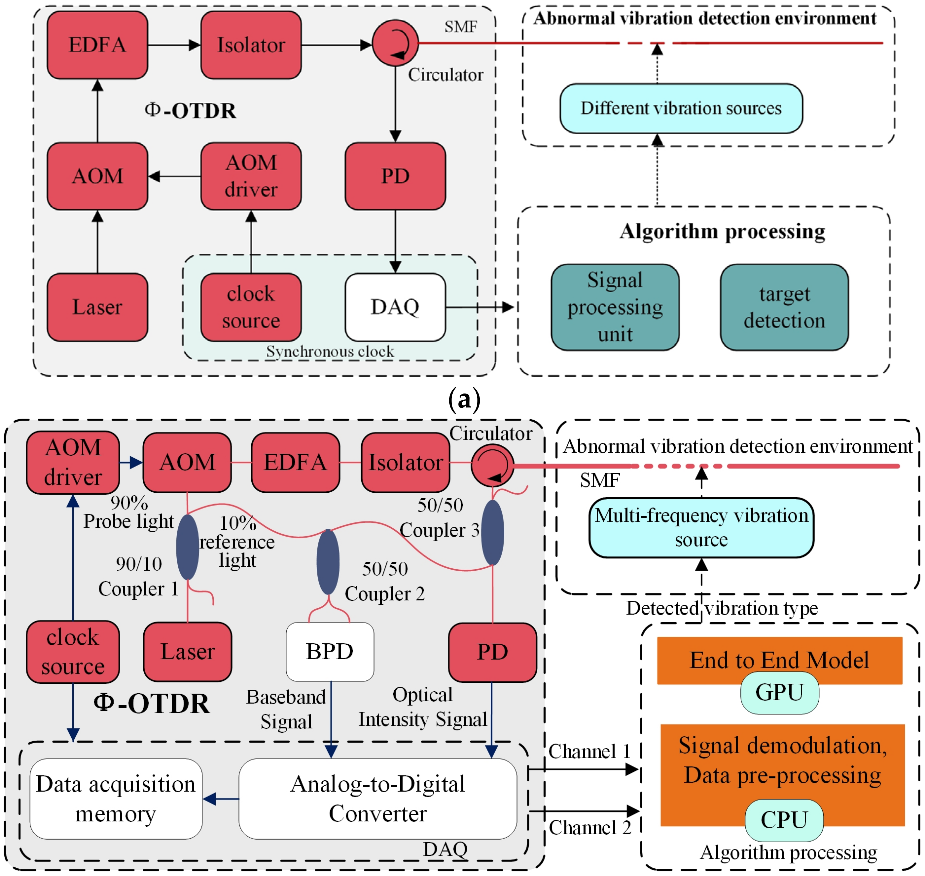
Direct optical intensity detection can effectively determine the location and corresponding time of the vibration signal on the single-mode fiber. The direct optical intensity signal detection technique has the advantages of simple structure, easy implementation, and more mature vibration demodulation algorithms. However, coherent detection can linearly measure external vibration events through phase extraction of coherent signals [35]. This compensates for the direct detection shortage that can only detect backward scattered light intensity and expands the modulable dimension of fiber optic sensing. Moreover, the local light in the outlier detection plays the role of coherent signal amplification, which can improve the signal-to-noise ratio of the detection signal. Therefore, coherent demodulation methods become the first solution to quantify external vibration events. Signal processing of moving differential can be employed to measure the broadband acoustic frequency components generated by pencil break vibrations [11]. The zero-difference detection technique with 90° optical mixing derives the amplitude and phase of Rayleigh scattered light by the I/Q demodulation method and experimentally demonstrates dynamic strain sensing [36].

Nevertheless, there is no interferometer structure in coherent digital detection, the reference signal is generated by digital methods, and the demodulation process is all realized by software algorithms.

Moreover, after the beat-frequency signal is digitized, the digital signal processing method can compensate for dispersion and reduce the impact of laser phase noise on the fiber system and the algorithm is more flexible. However, when demodulating the phase of RBSs using digital coherence, it is necessary first to demodulate the amplitude of all collected signals along the fiber and then determine the location of vibration events based on the amplitude change [37], which results in slow demodulation speed due to excessive computation and thus affects the real-time alarm performance. The system described in this paper combines the two detection methods, taking advantage of their respective advantages, and can achieve rapid demodulation of vibration positions using the direct detection method while using the digital coherence method to obtain the dynamic process of vibration changes and thus achieve multi-vibration sequence identification.

As shown in Figure 1a, in the previous work [9], the algorithmic part of the demodulation of vibration signals using light intensity variations is completed based on the statistical detection principle and enables effective detection of vibration events. The main instruments of the system contain a narrow linewidth laser, an acoustic-optic modulator (AOM), an AOM driver, an erbium-doped fiber amplifier (EDFA), a photodetector (P.D.), and a single-mode fiber (SMF) as a vibration monitoring sensor. The synchronous clock source controls the AOM driver trigger signal and the acquisition moment of the data acquisition module (DAQ). On this basis, the required instruments for coherent detection are introduced to complete the acquisition of coherent signals. To achieve coherent demodulation based on intensity detection, a balanced photodetector (BPD) and three optical couplers are added with the same device setup as above. The DAQ of the dual channel is upgraded. As shown in Figure 1b, the red part is the same optoelectronic element, and the setup is the same compared to Figure 1a. The specific process is as follows:

The narrow linewidth coherent laser source is divided into two paths by a 90:10 coupler 1, with 90% of the light used as the detection light and 10% as the local reference light. The detection light is modulated into pulsed light by AOM with frequency shift. EDFA increases the peak power of the pulsed light. An optical circulator injects the pulsed light into the probe single-mode fiber. Due to the elastic scattering effect, Rayleigh backward scattered light is formed by the interference of pulsed light in the fiber. The backward Rayleigh scattered light carrying vibration information is treated as the signal light, divided into two parts by the 3 dB coupler 3. One part is converted by the photodetector from light intensity to electrical signal, collected by the data acquisition module channel 1. The other part of the signal light is mixed with the local reference light in optical coupler 2 to form a beat-frequency signal. The beat-frequency signal is injected into the BPD. The signal from the BPD output is sampled by the data acquisition module channel 2. Channel 1 and 2 data are fed into the algorithm processing module simultaneously. The vibration position is determined using the light intensity data acquired by channel 1 and the vibration position fast demodulation method described in Section 3.2. Then, the phase of Rayleigh backscattered light near the vibration position is extracted using the beat-frequency signal data acquired by channel 2 and the digital I/Q demodulation method described in Section 3.2. The phase difference change near the vibration position represents the vibration dynamic change process. The demodulated vibration dynamic time domain sequence is fed into the end-to-end deep learning model, which completes the identification of multiple vibration events and label alignment and, finally, outputs the detection results.



(**a**)

(**b**)

Figure 1. Overall system operation structure: (**a**) Φ-OTDR direct detection method; (**b**) Φ-OTDR hybrid detection method.

2.2. Raw Data Acquisition Method

From the analysis in Section 2.1, to reduce the time consumption of vibration position demodulation, all signals in one sampling period are processed using the direct detection method to demodulate the corresponding vibration position. Then, the backward scattered light phase at the vibration position is demodulated by the coherent detection method to extract the dynamic sequence signal at the vibration position quickly.

The optical field of the detection light injected into the single-mode fiber can be expressed as $A(t)\exp(j(2\pi F_L t + \theta(t)))$, where $A(t)$ denotes the local reference light amplitude, F_L indicates the local reference light frequency, and $\theta(t)$ represents the initial phase of the pulse light.

As shown in Figure 2, the backscattered light from the 2N scattering points within the probe pulse light is superposed, and the generated Rayleigh backscattered light can be expressed as:

$$\begin{aligned}\zeta_s(t) &= A_s(t)\exp(j(2\pi F_s t+\theta_s(t)))\sum_{i=-N}^{N}\sqrt{\sigma_i(t)}\exp(-j\theta_i(t)) \\ &= A_s(t)\exp(j(2\pi F_s t+\theta_s(t)))\sqrt{\hat{\sigma}(t)}\exp(-j\hat{\theta}(t)) \\ &= A_s(t)\sqrt{\hat{\sigma}(t)}\exp\big(j\big(2\pi F_s t+\theta_s(t)-\hat{\theta}(t)\big)\big)\end{aligned} \tag{1}$$

where $A_s(t)$ denotes the Rayleigh backscattered light amplitude, F_s indicates the Rayleigh backscattered light frequency, $\theta_s(t)$ represents the phase of the Rayleigh backscattered light, and $\hat{\theta}(t)$ means the superposition echo phase. The photodetector output electrical signal in the fast-time domain can be expressed as $I_d(t) = \rho\zeta_s\zeta_s^* = \rho A_s^2(t)\hat{\sigma}(t)$, where ρ denotes the photoelectric conversion rate. When there is a vibration event around the fiber, the integrated reflection coefficient $\hat{\sigma}(t)$ changes in the slow-time domain. The photoelectric converter outputs this light intensity signal, which is then sampled by the data acquisition module channel 1.

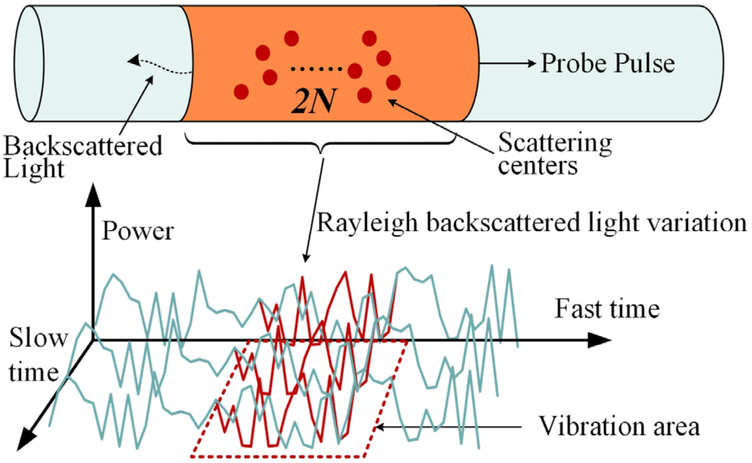


Figure 2. Rayleigh backscattered light variation detected by Φ-OTDR.

Meanwhile, the local reference light $\zeta_L(t) = A_L(t)\exp(j(2\pi F_L t+\theta_L(t)))$ divided by coupler 1 is mixed with the RBSs $\zeta_s(t)$ through 1:1 optical coupler 2, where $A_L(t)$ denotes the reference light amplitude, F_L indicates the reference light frequency, and $\theta_L(t)$ represents the phase of the reference light. The optical power of the two-path interference light output from the coupler is:

$$P_a = \hat{\sigma}(t)A_s^2(t) + A_L^2(t) + 2\sqrt{\hat{\sigma}(t)}A_s(t)A_L(t)cos(\Delta\omega t+\Delta\theta(t)) \tag{2}$$

$$P_b = \hat{\sigma}(t)A_s^2(t) + A_L^2(t) - 2\sqrt{\hat{\sigma}(t)}A_s(t)A_L(t)cos(\Delta\omega t+\Delta\theta(t)) \tag{3}$$

where $\Delta\omega = 2\pi(F_s - F_L) = 2\pi\Delta F$, $\Delta\theta(t) = \theta_L(t) - \theta_s(t) - \hat{\theta}(t)$ denotes the phase difference between the local light and the signal light phase. ΔF is the frequency shift introduced by the AOM. After the two paths of light pass through the BPD, the photocurrent signal obtained after mixing frequency satisfies the following equation [37,38]:

$$I_c(t) \propto 4\sqrt{\hat{\sigma}(t)}A_s(t)A_L(t)cos(\Delta\omega t+\Delta\theta(t)) \tag{4}$$

The echo phase $\hat{\theta}(t)$ is uniformly distributed [12], and it can be assumed that $\hat{\theta}(t) = 0$ according to the law of large numbers. This signal is sampled by the data acquisition module channel 2. The strain of the fiber due to vibration leads to the change of $\Delta\theta(t)$ when a vibration event occurs around the fiber. The dynamic process of vibration events can be detected indirectly by demodulating the phase difference change at the vibration position.

Generally, demodulating $\Delta\theta(t)$ variation only using digital down-conversion requires demodulating the amplitude and phase of the backscattered light at all sampling points of the fiber. The algorithm is computationally intensive. A fast $\Delta\theta(t)$ variation demodulation method is demanded, considering the real-time requirements for Φ-OTDR vibration detection.

3. Methodology

3.1. Proposed Framework

For a long-distance multi-vibration signal detection environment, as shown in Figure 3, to quickly identify multiple vibration sources, the system needs to be processed by three components: (1) First, the Φ-OTDR described in Figure 1 is used for two channels of raw data acquisition. The direct detection signal demodulation algorithm is applied to obtain the vibration position according to the optical pulse detection principle. Then, the phase change of the vibration position is demodulated using the digital coherence technique, i.e., the dynamic time series of vibration change. The demodulated sequence is processed using denoising methods to improve the signal-to-noise ratio and obtain the data to be classified. The system completes the vibration position demodulation and the dynamic sequence signal output at this stage. (2) According to the analysis in Section 1, CNN is the most commonly used and effective signal feature extraction method. The dynamic sequence signal requires pre-processing before using CNN, including windowing, framing, and fast Fourier transform operations, to obtain the time-frequency map to be processed. (3) Finally, all classes of vibration activities are accurately identified by the end-to-end multi-target detection model. The CNN extracts image features frame by frame to form a time–frequency image feature sequence. Bi-LSTM, as a sequence context feature extraction tool, can effectively represent dynamic sequences of vibration signals by employing it in combination with CNN. Eventually, the vibration events are identified by the automatic alignment method. Unlike traditional sequence training [28,29], the end-to-end training and identification process can integrate sequence signal feature extraction, recognition, and decoding, simplifying the model training process.

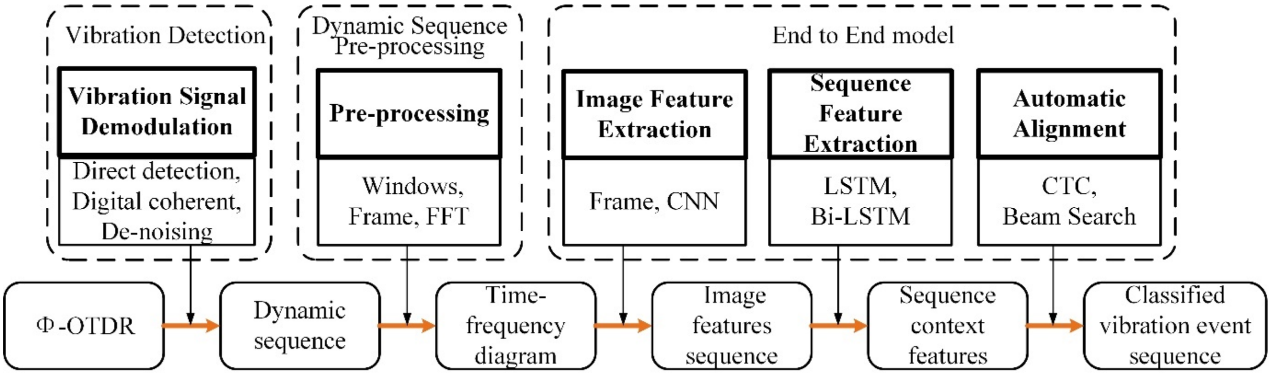


Figure 3. Workflow for end-to-end-based multi-vibration target detection.

In this application, the raw signals acquired by Φ-OTDR are processed by the vibration detection unit to generate a dataset. The dataset is then divided into a training set, a validation set, and a test set. The training and validation sets train and adjust the end-to-end multi-target detection models. The test set is used to test and evaluate the performance of all models and then select the best model. The time–frequency images generated from the real-time demodulated vibration data are transferred to the final selected model to obtain multiple vibration activity types and locations.

3.2. Local Digital I/Q Signal Demodulation

When the Φ-OTDR described in Section 2.1 acquires the signals from both channels, the system first needs to demodulate the vibration signals. This section mainly completes the vibration detection part illustrated in Figure 3, and the two signals will be transformed into the dynamic sequence by this method.

Due to the extensive computational effort of digital coherent demodulation, a local digital I/Q signal demodulation method combining intensity detection and coherence detection is employed to quickly demodulate the variation of $\Delta\theta(t)$ in the slow-time domain to achieve fast demodulation of dynamic vibration sequences. As shown in Figure 4, firstly, the vibration position is demodulated using the data collected by channel 1. Secondly, the phase changes brought by the data acquired by channel 2 are demodulated via digital I/Q. As a result, the dynamic change process of the final vibration event is obtained.

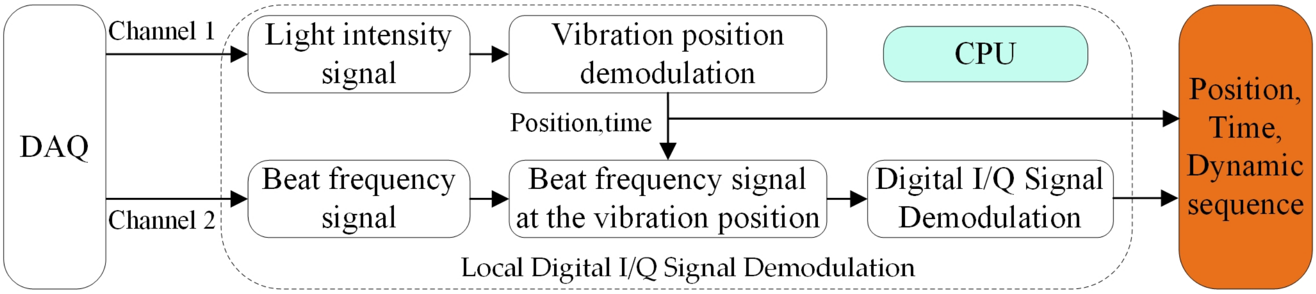

DAQ
Channel 1
Channel 2
Light intensity signal
Vibration position demodulation
CPU
Position,time
Beat frequency signal
Beat frequency signal at the vibration position
Digital I/Q Signal Demodulation
Local Digital I/Q Signal Demodulation
Position, Time, Dynamic sequence

Figure 4. Local digital I/Q signal demodulation process.

To determine the location of the vibration event, measuring the change in the data collected by channel 1 in the slow-time domain is necessary. Therefore, M_p optical pulses are used in one detection scan. The backscattered light is converted to electrical signals by P.D. and then sampled by the data acquisition module to form a two-dimensional data matrix. The rows represent fast-time-domain signals, i.e., the light intensity data of the backscattered light caused by the same pulse light at different fiber locations. The columns represent the slow-time-domain signals, i.e., the information on the variation of the backscattered light intensity obtained by using different optical pulses at the same fiber location. In the slow-time domain, the detection data obtained from one scan at the exact position are $\boldsymbol{X} = \left[x_0, x_1, \cdots x_m, \cdots x_{M_p-1}\right]$. M_c statistics $\boldsymbol{X}' = \left[x_0', x_1', \cdots x_i', \cdots x_{M_c-1}'\right]$ are obtained after M consecutive data noncoherent integration, where $M_c = M_p - M - 1 - M_g$,

$$x_i' = \sum_{m=i}^{M-1+i}\left(x_m - x_{m+M_g}\right), i \in [0, M_c - 1] \tag{5}$$

M_g denotes the pulse canceller interval. By judging the value of $x_{max}' = \max(\boldsymbol{X}')$ and also adjusting the coincidence detection parameter K to improve the reliability of the detection decision, the sampling point n^* corresponding to the location of the vibration can finally be obtained [9].

The above process is the direct detection in Figure 3. On this basis, phase acquisition is achieved by digital I/Q demodulation, i.e., digital coherent, as shown in Figure 5. The N discrete data $Y(n) = E(n)cos(\Delta\omega n + \Delta\theta(n)), n \in [0, N-1]$, collected by channel 2, in one detection are mixed with two quadrature digital reference signals, $\cos(\Delta\omega n + \varphi)$, $\sin(\Delta\omega n + \varphi)$, and a set of quadrature signal, $\cos(\Delta\theta(n) - \varphi)/2$, $-\sin(\Delta\theta(n) - \varphi)/2$, about $\Delta\theta(n)$ is obtained after low pass filter. There is no abrupt change between adjacent phases, and the phase information $\Delta\theta(n) - \varphi$ is solved using the Arctangent algorithm. The value of the variation between the phase difference at the vibration position n^* and the phase difference at the adjacent position is obtained as $(\Delta\theta(n^*) - \varphi) - (\Delta\theta(m) - \varphi) = \Delta\theta(n^*) - \Delta\theta(m)$, which is independent of the phase φ of the quadrature digital reference signal. In one detection scan, M_p sets of discrete data $Y_j(n), j \in \left[0, M_p - 1\right]$ to be demodulated are obtained through M_p detection pulses, while quadrature digital reference signals with fixed phase φ can be generated when the digital I/Q method is used. Therefore, to save the demodulation time of the dynamic vibration sequence, based on the vibration position available through the data of channel 1, the coherent optical phase $\hat{\boldsymbol{\theta}} = (\Delta\theta(n) - \varphi, \cdots, \Delta\theta(n^*) - \varphi, \cdots, \Delta\theta(n + N_p) - \varphi)$ of only N_p neighboring sampling points at the vibration position can be demodulated in a set of detection data using the digital down-conversion method described above, thus the phase difference $\Delta\hat{\theta} = \Delta\theta(n + N_p) - \Delta\theta(n)$ at the vibration position neighbor. The vibrational dynamic sequence $\Delta\hat{\boldsymbol{\theta}} = \left[\Delta\hat{\theta}_0, \Delta\hat{\theta}_1, \cdots \Delta\hat{\theta}_m, \cdots \Delta\hat{\theta}_{M_p-1}\right]$ of length M_p can be obtained in one scan. Finally, the moving average method with a sliding step of M_a is used to achieve the noise reduction process, resulting in $M_p - M_a + 1$ data points $\boldsymbol{\theta}' = \left[\theta'_0, \theta'_1, \cdots \theta'_i, \cdots \theta'_{M_p-M_a}\right]$, where:

$$\theta'_i = \textstyle\sum_{j=i}^{M_a-1+i} \Delta\hat{\theta}_j, i \in [0, M_p - M_a] \tag{6}$$

$\boldsymbol{\theta}'$, the dynamic phase difference change of the Rayleigh backscattered light phase in the fiber at the vibration position, represents the dynamic vibration process and is applied to the subsequent multi-target sequence detection.

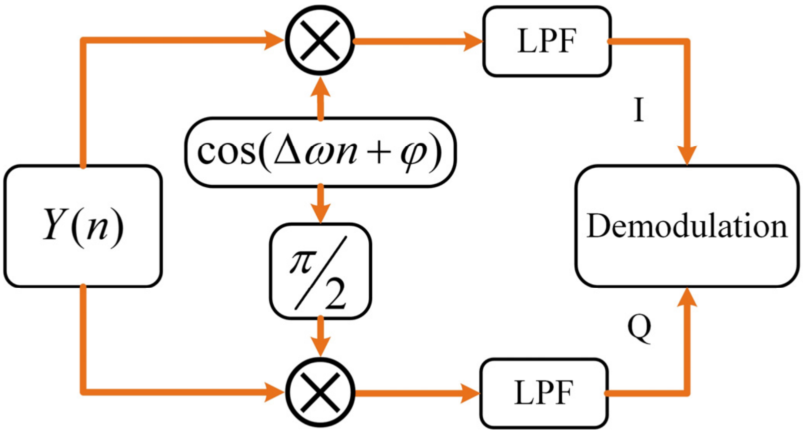


Figure 5. Digital I/Q demodulation process, LPF: low pass filter.

3.3. Data Time-Frequency Conversion

As described in Section 3.2, a set of dynamic vibration data $\boldsymbol{\theta}'$ to be detected is obtained from M_p detection pulses. The vibration features will change over time for non-stationary signals, especially for sequences containing multiple vibration events. The time–frequency analysis reflects the variation of the signal's frequency components and time. The combined information of the time and frequency domains of the non-stationary signal can be initially available by time–frequency analysis. On the other hand, as analyzed in Section 2.2, CNN can automatically extract higher dimensional features. CNN can effectively classify vibration types based on frequency domain features for sequence data that contain only a single vibration mode in one sampling [23]. While recognizing multi-target sequences using CNN, it is necessary to convert the sequence into two-dimensional data in which the features vary with time. Short-time Fourier transform (STFT) is commonly used for time–frequency analysis and can convert non-smooth sequences into multicolor images for CNN training. In addition, multicolor images also help improve feature extraction and target object detection performance.

$$STFT(t, f) = \int_{-\infty}^{\infty} x(\tau) w(t - \tau) \exp(-j2\pi f t) d\tau \tag{7}$$

Equation (7) formulates the STFT [23], including windowing, framing, and fast Fourier transform (FFT). Longer non-smooth signals are slidingly truncated by framing, making each frame nearly smooth. The non-stationary signal is slidingly truncated by frame splitting so that each frame is nearly stationary. The window function can reduce the discontinuity of the signal at the edge of the frame and eliminate high-frequency interference and energy leakage. Each frame is then fast Fourier transformed to generate the time–frequency diagram in real time to train and test the multi-vibration sequence detection model.

3.4. Recognition Model

3.4.1. CLSTM and CTC

After the data processing process in Section 3.3, the raw data are converted into a time–frequency diagram containing multiple vibration events. As described in Figure 3, an end-to-end recognition model needs to recognize the time–frequency diagram to complete the number judgment and type recognition of vibration events. This section focuses on the structure of the end-to-end model and the training process of the model.

Convolutional neural networks (CNNs) and recurrent neural networks (RNNs) are two cornerstones of deep learning research. CNN-based LetNet, AlexNet, and VGGNet have achieved high recognition rates in image recognition and have been applied in Φ-OTDR pattern recognition [29]. Variant networks based on RNN, such as LSTM and Bi-directional LSTM [39], are effective tools for sequence feature extraction. As shown in Figure 6, the CLSTM combines the two networks and is the sequence target feature extraction part of the proposed end-to-end recognition model. The main objectives include detecting the number of targets in the time-frequency diagram, identifying vibration types, and labeling the order of category occurrence. This subsection introduces the training process of the CLSTM sequence detection model and analyzes the factors influencing the recognition performance of the model to provide a theoretical basis for improving the detection performance of the model. The training process of CLSTM can be divided into four parts:

1. The CNN layer focuses on extracting image spatial dimensional features, forming feature maps at different scales employing convolution layers, pooling layers, padding, activation functions, etc., thereby extracting high-dimensional features. Since the convolution layer is computed by sliding from left to right on the original image, it has translation invariance. Therefore, each column of the feature map corresponds to a perceptual field at the relevant position on the original image in the same order. The feature map is a sequence of subframe features obtained by sequential operation after dividing the original time–frequency map into frames.
2. The LSTM layer extracts the temporal dimensional features of the data. Sequences containing contextual features are extracted by memory gates, forgetting gates, cell states, etc., eventually forming the sequence of labels to be aligned. The depth and length of the extracted sequence features depend on the extraction direction of the LSTM, the number of layers, and the number of hidden layers per LSTM.
3. The transcription layer solves the alignment problem between the label and output of the neural network. During the model training, the sequence path probability that matches the label is obtained when the output sequence of LSTM undergoes labeled path search and B transform. The smaller the value of this probability, the greater the overall loss of the system.
4. Once the system loss value is acquired, the backpropagation algorithm updates the parameters of the algorithm module in the first and second steps. So far, an update iteration operation has been completed. The backpropagation algorithm, including learning rate, learning rate update method, batch size, etc., will affect the training speed and final effect. Therefore, to compare different performance metrics of the models, the parameters of the backpropagation algorithm should be consistent when training different models with multiple iterations until convergence.

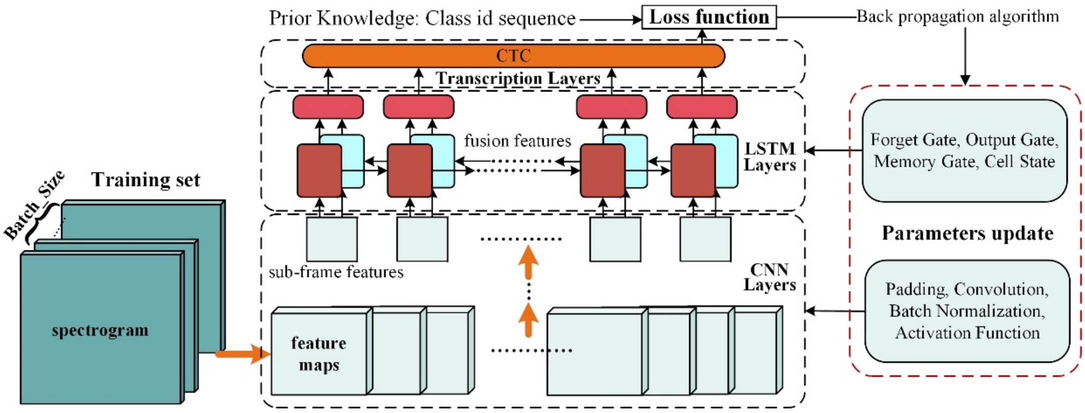

Prior Knowledge: Class id sequence
Loss function
Back propagation algorithm
CTC
Transcription Layers
fusion features
LSTM Layers
Forget Gate, Output Gate, Memory Gate, Cell State
Parameters update
Batch Size
Training set
spectrogram
sub-frame features
CNN Layers
Padding, Convolution, Batch Normalization, Activation Function
feature maps

Figure 6. Multi-target sequence detection meta-architecture training process.

Different network structures of CNN layers directly affect the extraction of spatial features. Suitable network structures can extract features better and achieve different scales of feature propagation to perform feature enhancement, leading to different recognition results. The convolution operation can extract linearly varying features, while the activation function completes the nonlinear transformation of the features.

Similarly, different network structures exist for the LSTM, but the foundation is the LSTM cell structure. The arguments of the LSTM cell will affect the model recognition results. As shown in Figure 7, The LSTM cell model is a combination of the input feature x_t at time t, the cell state C_t that records the information transfer, the memory gate modules $\widetilde{C}_t$ and i_t that record the current state, the forgetting gate f_t that filters the current input information, the output gate o_t, and the hidden layer state h_t at this time, with the expression [40]:

$$f_t = \sigma\left(\boldsymbol{W}_f h_{t-1} + \boldsymbol{U}_f x_t + b_f\right) \tag{8}$$

$$\begin{aligned} i_t &= \sigma(\boldsymbol{W}_i h_{t-1} + \boldsymbol{U}_i x_t + b_i) \\ \widetilde{C}_t &= tanh(\boldsymbol{W}_c h_{t-1} + \boldsymbol{U}_c x_t + b_c) \end{aligned} \tag{9}$$

$$C_t = f_t \odot C_{t-1} + i_t \odot \widetilde{C}_t \tag{10}$$

$$o_t = \sigma(\boldsymbol{W}_o h_{t-1} + \boldsymbol{U}_o x_t + b_o) \tag{11}$$

$$h_t = o_t \odot tanh(C_t) \tag{12}$$

The calculation process of LSTM is to forget the information in the cell state and optionally remember the new information so that the valuable information for the subsequent moment calculation can be passed and the useless information is discarded. The hidden layer state h_t is output at each time step, where h_t is the feature state of the time–frequency map over time. After framing, the hidden layer state can connect the spatial features in the temporal dimension.

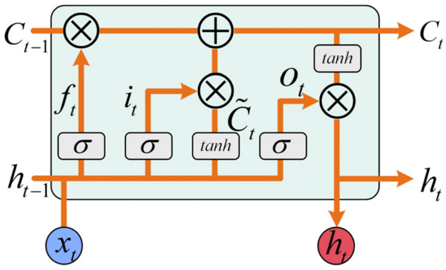


Figure 7. LSTM cell architecture.

In the sampled data, the durations of the vibration signal features in the time–frequency diagram are not consistent with the lengths of the labels, and the locations where the vibrations appear and end are uncertain, which results in the need for automatic alignment of the feature recognition sequences from CLSTM output.

The CTC criterion and temporal neural network combination can be applied directly to end-to-end modeling to solve the above problem. The CTC algorithm achieves alignment and placeholders by introducing a blank symbol. The LSTM outputs different probabilities for all vibration types at each moment. Different output sequences correspond to different feasible search paths. The probability of occurrence of any path is $p(\boldsymbol{\pi}|\boldsymbol{h}) = \prod_{t=1}^{T} r_{\pi_t}^{t}, \forall \boldsymbol{\pi} \in {L'}^{T}$, $r_{\pi_t}^{t}$ represents the probability that the state is π_t at the moment t, and the total length of the sequence is T. By applying the forward-backward, dynamic programming algorithm, all paths matching the labels are found. The summed probability is $p(\boldsymbol{l}|\boldsymbol{h}) = \sum_{\boldsymbol{\pi} \in B^{-1}(l)} p(\boldsymbol{\pi}|\boldsymbol{h})$, where $\boldsymbol{h}$ denotes the output sequence of the LSTM, $\boldsymbol{l}$ denotes the label sequence, and $\boldsymbol{\pi}$ is all paths matching the label $\boldsymbol{l}$ after the B transform. The B transform defines that the blank symbol can self-loop or jump to the next nonblank symbol. Nonblank symbols can jump to blank symbols in addition to self-looping or jumping to nonblank symbols. Using the B transform can transform the LSTM output sequence into the recognition sequence. The recognition sequence can be transformed from the LSTM output sequence using the B transform. The final loss function can be expressed as follows [41]:

$$Loss(S) = -ln\prod_{(\boldsymbol{h},\boldsymbol{l})\in S} p(\boldsymbol{l}|\boldsymbol{h}) = -\sum_{(\boldsymbol{h},\boldsymbol{l})\in S} lnp(\boldsymbol{l}|\boldsymbol{h}) \tag{13}$$

where S denotes a batch size in the training set. $\prod_{(\boldsymbol{h},\boldsymbol{l})\in S} p(\boldsymbol{l}|\boldsymbol{h})$ denotes the product of the probabilities of the correct label output in a given sample S. The loss function quantifies the difference between the recognition sequence output by the recognition system and the prior knowledge. During the model training process, the loss function needs to be minimized, and the smaller the loss function, the higher the probability of the correct label output. The training process is completed by updating all parameters of the convolutional layer and the LSTM layer with the calculated loss values and the backpropagation algorithm with iterative operations until convergence. The training process is completed when the system can achieve a satisfactory detection rate on the test set.

Based on the above analysis, the network structures of the spatial image feature extractor CNN and the time dimensional feature extractor LSTM are crucial to the global detection recognition rate of the system. At the same time, the model's complexity directly impacts the time consumption of detection. In the long-distance fiber optic vibration monitoring, due to the frequent vibration and vibration type changing, the CNN and LSTM structure of the model should be designed to meet the high multi-target recognition rate to reduce the occurrence of false alarms, but also to ensure the real-time requirement of the system recognition.

3.4.2. Beam Search Decoding Function

After the model training is completed, the trained model can be used for data testing. In the testing process, after passing the recognition model without calculating the loss function, the data signal time–frequency diagram forms the unaligned sequence. The Beam Search method can search the maximum probability path of the LSTM output sequence. The final recognition type sequence is obtained after automatic alignment by the B transform. The Beam Search dynamic programming algorithm can be employed for the sequence decoding process [42]. Unlike the greedy decoding strategy, Beam Search does not keep only the single output with the highest probability of the current step in one-step search but keeps the top N categories with the highest probability of the current step as the input state of the next step and iterates to obtain the optimal solution in turn. It is utilized for the decoding process in model testing and validation.

After all the methodological processes described above, the system converts the external vibration information the fiber senses into a time-domain sequence signal through the hybrid demodulation method in Section 3.2 when vibration exists near the fiber. The STFT converts the sequence signal into a time-frequency diagram, and the time-frequency diagram identifies multiple vibration events at once through an end-to-end sequence classification model. The end-to-end sequence classification model identifies multiple vibration events from the time-frequency diagram simultaneously.

4. Experimental Dataset Generation and Preparation

4.1. Data Collection

This study applies the Φ-OTDR system described in Section 2.2 with a narrow linewidth laser (~1.69 kHz), central wavelength of 1550 nm, with a typical output power of 13 dBm. The AOM frequency shift is 200 MHz. The pulse width is 50 ns, corresponding to a resolution of 5 m. The optical pulse repetition frequency is 20 kHz, corresponding to a maximum sensing distance of 5 km. Existing methods have examined the effect of phase difference demodulation at different frequencies [11,13,36,43], and to illustrate the effectiveness of the system in detecting vibration signals of variable duration in different frequency bands, external speakers generate five vibration sources, each consisting of a mixture of two different frequencies, simulating vibrations in different frequency bands. Then, vibration data are collected for model training to verify the model's performance in identifying multi-vibration data. One path of backscattered Rayleigh light is injected into the photodetector, acquired by DAQ channel 1. The other path is mixed with the local reference light in the optical coupler 2 to form a beat-frequency signal, which is injected into the balanced photodetector and acquired by DAQ channel 2 with a sampling rate of 1 Gsample/s. The signal processing part is executed in P.C., including vibration position demodulation and the digital quadrature Rayleigh backscattered light phase demodulation.

Table 1 shows the five types of vibration sources included in the experiment. The system collects a total of 8563 data. Type I data indicates that data contain only one type of vibration source, 1714 in total; Type II data denotes the presence of two vibration sources for one datum, totaling 3426; Type III data represents three types of vibration sources contained in one datum, 3423 in total. Each datum contains 1–3 types of vibration sources. As shown in Figure 8a, '-A-D-' denotes one datum consisting of vibration sources labeled as 1 and 4, '-' denotes empty symbols, and each vibration source has different durations. The data are recorded using the vibration signal detection method described in Section 3.2 to form the dataset finally. The dataset is divided into a training set, a test set, and a validation set in the ratio of 8:1:1, where the training set is used in model training as described in Section 3.4.1. During model training, the model is tested using the validation set along with the decoding method described in Section 3.4.2 to select the model with better performance in different network structures as the final model. Finally, the evaluation metrics of the models are statistics by the test set.

Table 1. Description of the distribution of different vibration types.

Vibration Source	Included Vibration Frequencies	Lasting Time	Number in Type I Data	Number in Type II Data	Number in Type III Data	Label
A	250 hz, 500 hz	0.1–0.3 s	338	1366	2037	1
B	350 hz, 650 hz	0.1–0.3 s	364	1383	2021	2
C	500 hz, 750 hz	0.1–0.3 s	340	1361	2086	3
D	650 hz, 1000 hz	0.1–0.3 s	330	1356	2046	4
E	250 hz, 750 hz	0.1–0.3 s	342	1386	2079	5

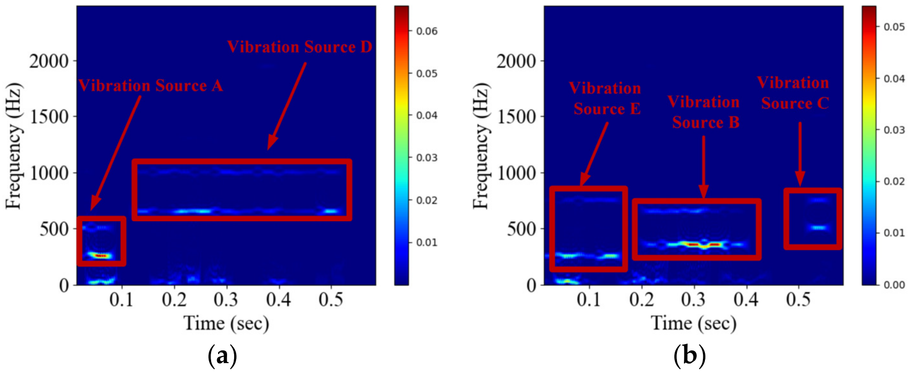


Figure 8. Time–frequency diagrams after image pre-processing: (**a**) data diagram with two vibration types of class A and D; (**b**) data diagram with three vibration types of class E, B, and C.

4.2. Network Construction

According to the CLSTM network structure analysis in Section 3.4, nine end-to-end models with different network structures are implemented, as shown in Table 2. The difference between the models is the structural changes of CNN and LSTM. Generally, the deeper the CNN network depth, the better the recognition performance of the model [44]. However, too many CNN layers also have the problem of degraded recognition performance [45]. The network parameters of LSTM, including the number of hidden layers, the number of layers, and the direction of network propagation, can affect the recognition performance of the model. In addition, a complex model decreases the model's recognition speed, increases the memory occupied by the model parameters, and raises the hardware requirements of the device.

Table 2. Nine end-to-end models are composed of different network architectures.

Model	**CNN**			**LSTM**			**Transcription**
	C2	**C3**	**C6**	**Number of Hidden Layers**	**Number of Layers**	**Bi-LSTM**	**CTC**
CLSTM1			√	32	2	√	√
CLSTM2			√	16	2	√	√
CLSTM3			√	64	2	√	√
CLSTM4			√	64	1	√	√
CLSTM5			√	16	1	√	√
CLSTM6	√			64	2	√	√
CLSTM7		√		64	2	√	√
CLSTM8			√	64	2		√
CLSTM9			√	16	1		√

The improved models require a balance between recognition performance and time consumption. To compare the effects of different model structures on the performance described in Section 4.3, Table 2 lists nine improved models with different network structures. The above nine models are trained respectively using the training set described in Section 4.1 to select the model with the best performance. For the selection of CNNs, three CNN networks are self-designed, wherein the C2 network contains two convolutional layers and one Maxpooling layer, the C3 network contains three convolutional layers and two Maxpooling layers, and the C6 network has six convolutional layers and three Maxpooling layers. The effect of CNN network depth on the model performance can be discussed while keeping the LSTM parameters constant. In addition, by comparing the models with the same CNN network in the list, the impact of LSTM feature extraction in the temporal dimension on the recognition performance of the system can be derived.

4.3. Evaluation Metrics

Six statistical metrics are used to quantify the performance of the proposed model in terms of the time dimension, spatial dimension, and recognition effectiveness, including accuracy, F1 score, recall, precision, memory occupied by model parameters, and inference time.

Concerning the recognition accuracy, the recognition results are divided into four regions: T.P., T.N., F.P., and F.N. T.P. represents the number of positive samples correctly identified; T.N. represents the number of negative samples correctly identified; both of the above are cases of error-free judgment, and the cases for model identification with errors are further divided into F.P. indicates that the model predicts a positive sample, but the actual label is a negative sample, i.e., the number of negative samples misreported. When the label is a positive sample, but the model predicts a negative sample, it is presented by F.N., i.e., the number of underreported positive samples. Equations (14)–(17) provide the metrics to measure the effectiveness of model recognition [46]. Precision expresses the model's reliability in predicting positive cases, recall represents the ability of the model to predict all the positive cases, and the F1 score evaluates the above two metrics together. Next, the portability of the model and the speed of recognition processing are evaluated by the other two metrics.

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

$$Precision = \frac{TP}{TP + FP} \tag{15}$$

$$Recall = \frac{TP}{TP + FN} \tag{16}$$

$$F_1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{17}$$

5. Experimental Procedure

5.1. Experimental Setup

As shown in Figure 9, the effects of different algorithms on various model performance metrics are tested by using fast vibration demodulation methods and different combinations of network structures. Φ-OTDR multi-target sequence recognition system experiments included:

1. The processing time of the raw data is compared between two methods of fast hybrid demodulation and global digital I/Q demodulation [13], using piezoelectric ceramics (PZT) as the vibration source.
2. The effects of different CNN network structures on the model's overall performance are discussed. Nine models consisting of CNNs and LSTMs with different structures are used for training and testing, thus addressing the effect of different network structures on model recognition performance.
3. The validation set observes the training effect, and the model with satisfactory metrics in Section 4.3 is selected as the final model. The test set is used to analyze various performance metrics. In addition, the performance differences between the proposed CNN network and three classical VGG [44] networks are compared.
4. The sequence data collected by Φ-OTDR contain different lengths and different kinds of vibrations. An end-to-end model training method is constructed without using manual forced alignment. The effectiveness of the model for multi-vibration target recognition is verified.

The raw data vibration demodulation process uses Matlab to form a multi-target sequence dataset for model comparison experiments. The end-to-end model training and testing process is performed on the Ubuntu 18.04 operating system with an InterCore i7-11700@2.5 GHz processor and an NVIDIA GTX 3090 GPU. The CLSTM model operates on the PyTorch library (version 1.8.0) and Python (version 3.8).

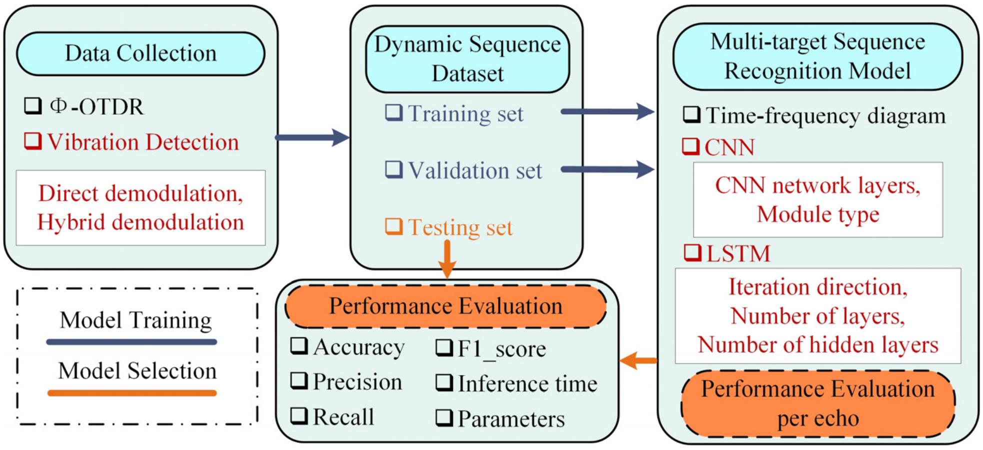
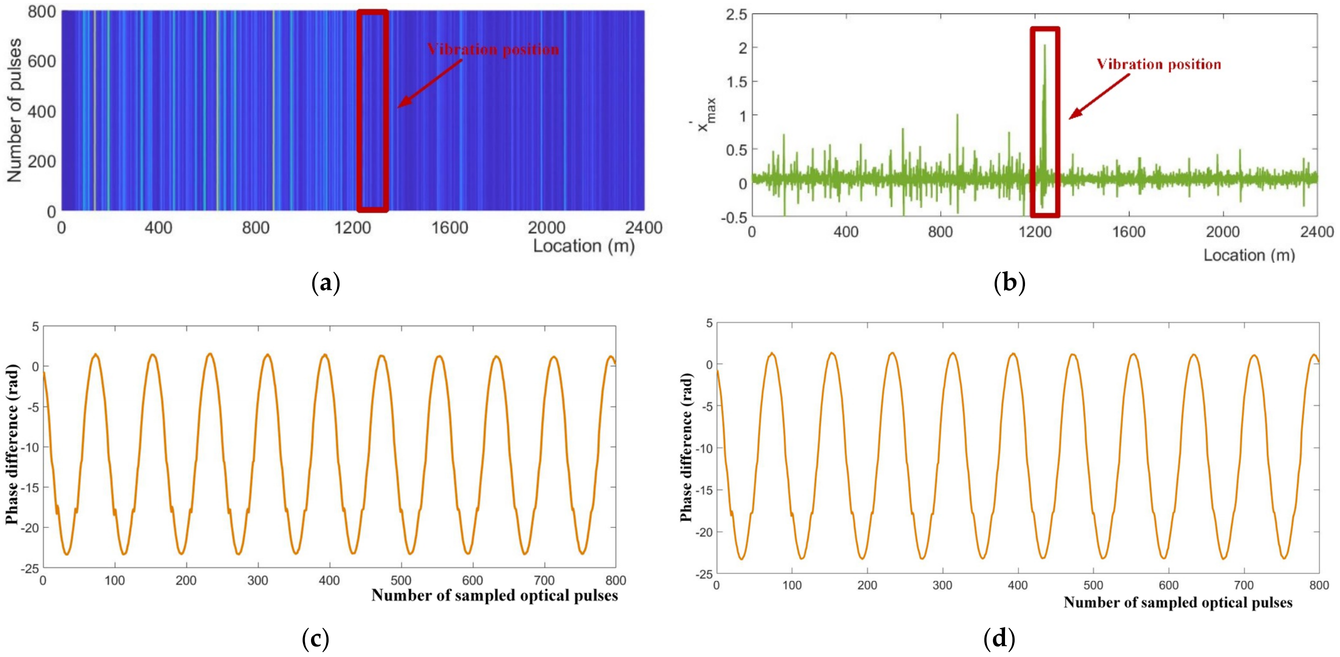


Figure 9. Experimental procedure of the CLSTM-based fiber optic sensing vibration identification method (red sections indicate variables).

5.2. Fast Hybrid Demodulation Method

To verify the time-saving ability and feasibility of the proposed demodulation method, a PZT is used as a vibration source at 1 km from a single-mode fiber with a total length of 4.8 km. A function generator controls the vibration frequency to produce a standard 250 Hz sine signal. Two signals are acquired by Φ-OTDR as described in Section 2.2. Method 1 demodulates the phase change by global digital I/Q demodulation only using the data from channel 2. The proposed method 2, the hybrid demodulation method, first obtains the vibration position by the direct detection method described in Section 3.2 with the light intensity signal collected by channel 1 then performs local optical phase demodulation near the vibration position using the signal collected by channel 2.

Figure 10a demonstrates a waterfall graph of the raw data acquired by channel 1, where all the raw vibration signals are submerged in the background noise. The pulse repetition frequency is 20 kHz, the noncoherent integration samples number M is 30, the number of pulses in a single scan M_p is 800, and the pulse canceller interval M_g is set to 20. Figure 10b illustrates the vibration position detected by the direct detection method. Figure 10d presents the dynamic vibration sequence obtained by local digital I/Q demodulation based on the signal near the vibration position collected by channel 2. The total time from vibration position demodulation to optical phase demodulation is 0.57 s. In contrast, the vibration sequence demodulated using the global digital I/Q demodulation is shown in Figure 10c. The time consumption is 2.35 s, which significantly increases the computational cost. It is found that the hybrid demodulation method reduces the time consumption for demodulating dynamic vibration sequences without affecting the demodulation results. Based on this hybrid demodulation method, a dataset containing five vibration sources is acquired by the data acquisition method described in Section 4.1 for model training, validation, and testing.



(**a**) (**b**)

(**c**) (**d**)

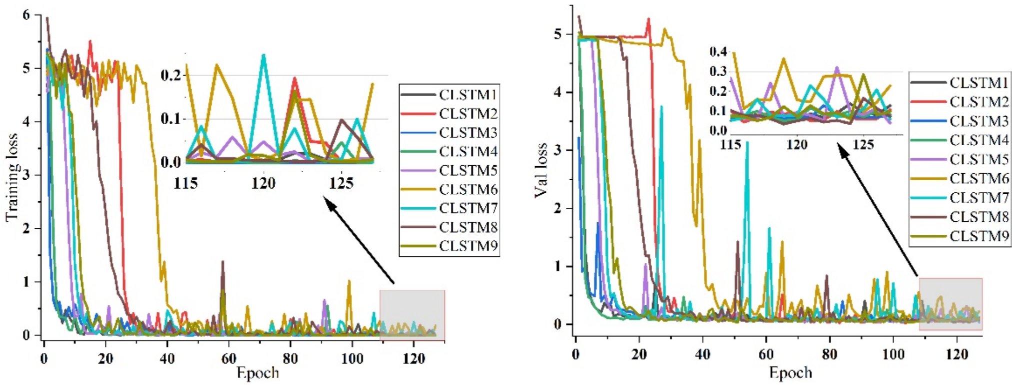
Figure 10. Illustration of hybrid demodulation method: (**a**) original signal waterfall of Rayleigh backscattered light intensity detection; (**b**) result of vibration position demodulation by light intensity variation; (**c**) phase demodulation result of digital I/Q method; (**d**) phase demodulation result of hybrid demodulation method.

5.3. Network Training

After the dataset is generated, CLSTM3 in Section 4.2 is trained as the control group. The performance metrics of different network structures for feature extraction are statistics by varying the CNN and LSTM structures in the model. The training set data described in Section 4.1 and the loss function described in Section 3.4.1 are used for model parameter updates, and the loss function values are recorded. Generally, with a constant training set and loss function definition, the more the loss function converges to zero during the iteration, the better the model's overall performance. Considering the reliability and generalization ability of the models, each CLSTM network is trained three times, running 127 epochs each time until convergence. The model with relatively small loss function values and high overall accuracy in the validation set is selected as the tested model. The test set is then used to count the various model performance metrics as described in Section 4.3.

Figure 11 shows the nine different network structure models' training loss and validation loss value curves after 127 training epochs. As the number of training epochs increases, both the training loss and validation loss of the models gradually decrease, indicating that the network models can all automatically align and efficiently update the network parameters. However, the data provided during training are multi-category mixed vibration data with variable length. To better display the differences in loss values of different models, the loss values from epochs 115 to 127 are selected for detailed comparison. Among the nine models, the training loss values of CLSTM6, CLSTM7, and CLSTM9 are slightly higher than the other models. The difference between the training loss values of the other models is relatively insignificant.

In contrast, CLSTM1, CLSTM3, and CLSTM4 converge faster than others on the validation set and have more stable loss values. Since the validation set is defined as a test set for the training stage, the validation loss values are more indicative of the generalization ability and performance of the models than the training loss values. This implies that CLSTM1, CLSTM3, and CLSTM4 perform better than other models. It also illustrates that there is an effect of different network structures on the recognition effect.



(**a**) (**b**)

Figure 11. Training and validation losses for CLSTM models with nine different network structures: (**a**) training loss values for the nine models; (**b**) validation loss values for the nine models.

5.4. Models Performance

Nine models with different network structures are trained through the training process in Section 5.3. Table 3 counts each model's prediction performance, number of parameters, and inference time for the six evaluation metrics described in Section 4.3. The table also records the number of false and missed alarms on a total of 1947 vibration events for the 907 data points in the test set. Since each experimental data point contains an indefinite length of vibration types, the number of identification types output by the model after each data point's automatic alignment is also indefinite. Therefore, except for the five data types with different vibration frequencies described above, the symbol '0' is specifically added as the blank category. When the predicted sequence length is greater than the label sequence length, i.e., the category is not present in the label but predicted as present, it can be considered false alarm detection. On the contrary, it can be considered as a missed alarm when the predicted sequence length is smaller than the labeled sequence length. The average statistical values of all types in the table are weighted averages to exclude the statistical influence of blank type on other vibration events.

Table 3. Model performance metric statistics.

Model	Type	Precision (%)	Recall (%)	F1 (%)	Accuracy (%)	Number of False Alarms	Number of Missed Alarms	Parameters (M.B.)	Inference (ms)
CLSTM1	All	99.34	99.18	99.26	99.18	8	5	21.6	2.211
	A	98.62	99.44	99.03					
	B	99.75	99.51	99.63					
	C	100	99.23	99.61					
	D	99.75	99.5	99.62					
	E	99.75	99.5	99.62					
CLSTM2	All	99.13	98.46	98.79	98.46	18	5	21.3	2.206
	A	98.61	98.33	98.47					
	B	99.51	99.75	99.63					
	C	99.74	98.72	99.23					
	D	99.75	98.5	99.12					
	E	99.23	98.23	98.73					

Table 3. *Cont.*

Model	Type	Precision (%)	Recall (%)	F1 (%)	Accuracy (%)	Number of False Alarms	Number of Missed Alarms	Parameters (M.B.)	Inference (ms)
CLSTM3	All	99.64	99.33	99.49	99.33	8	2	22.3	2.204
	A	99.17	99.17	99.17					
	B	100	99.51	99.75					
	C	100	99.23	99.61					
	D	99.75	99.5	99.62					
	E	99.75	99.75	99.75					
CLSTM4	All	99.49	99.18	99.33	99.18	9	3	21.9	2.212
	A	98.62	98.89	98.75					
	B	100	99.75	99.88					
	C	100	99.23	99.61					
	D	100	99.5	99.75					
	E	99.49	99.24	99.37					
CLSTM5	All	97.47	97.51	97.48	97.51	17	18	21.3	2.197
	A	96.48	98.89	97.67					
	B	98.76	98.52	98.64					
	C	98.45	98.2	98.33					
	D	99.75	98.5	99.12					
	E	98.22	97.98	98.1					
CLSTM6	All	96.04	96.71	96.37	96.71	16	30	1.55	2.05
	A	95.1	96.94	96.01					
	B	98.05	99.51	98.77					
	C	99.22	97.94	98.58					
	D	96.56	98.5	97.52					
	E	98.47	97.98	98.22					
CLSTM7	All	95.12	96.99	96.04	96.99	7	46	4.93	2.091
	A	93.67	98.61	96.08					
	B	97.11	99.75	98.41					
	C	99.49	99.23	99.36					
	D	97.78	99.25	98.51					
	E	98.5	99.49	98.99					
CLSTM8	All	97.93	98.42	98.17	98.42	6	16	21.9	2.201
	A	96.99	98.33	97.66					
	B	99.02	99.51	99.26					
	C	100	98.97	99.48					
	D	98.03	99.5	98.76					
	E	99.5	99.75	99.62					
CLSTM9	All	99.47	97.28	98.36	97.28	44	1	21.3	2.196
	A	99.15	96.67	97.89					
	B	99.8	98.76	99.25					
	C	99.74	97.43	98.57					
	D	99.75	98	98.86					
	E	99.21	95.7	97.42					

The key performance metrics with the best results are highlighted.

CLSTM3 has the highest overall precision at 99.64%. Although the precision of type A vibration with lower frequencies is the smallest among the five types of vibration at 99.17%, it is also higher than the precision of other models at type A vibration. The precision of the models is generally higher in vibration types B, C, and D than in the other two categories. This may be because the system is less effective at demodulating low-frequency vibrations than high frequency, and both types A and E contain vibrations in that frequency band. From the model structure perspective, CLSTM5 has the lowest precision of 97.47% among all models combining C6 networks. However, the precision of CLSTM5 is still 1.43% and 2.35% higher than that of CLSTM6 and CLSTM7, which contain C2 and C3 network structures, respectively. This indicates that increasing the depth of the CNN network can improve the precision of the model. CLSTM1, CLSTM2, and CLSTM3 only differ in the number of hidden layers, and the precision of the model is improved as the number of hidden layers increases. Similarly, the comparison of the precision of CLSTM4 and CLSTM5 also shows that increasing the number of hidden layers can improve the precision.

The recall column shows that CLSTM3 remains at the highest with 99.33%. The recalls of CLSTM1 and CLSTM4 are both 99.18%, while the recalls of the other models are all below 99%. Likewise, the recalls of the two vibration types, A and E, are generally lower than those of the other three types. Moreover, it can be noticed that the more vibration the model misses, the lower the recall tends to be, indicating that the model tends to identify hard-to-detect vibrations as blanks. It also illustrates that enhancing the recall of the model can effectively reduce the overall missed alarm rate.

The F1 scores can comprehensively evaluate the precision and recall of the dynamic sequence detection model. The F1 scores of CLSTM1, CLSTM3, and CLSTM4 are all above 99%. It is observed that there is no correspondence between the F1 score of the model and the LSTM complexity of the model. For example, compared with CLSTM4, CLSTM1 uses two layers of Bi-LSTM, but the number of hidden layers per layer is half of CLSTM4, and the difference in F1 score between the two is not significant. Similarly, CLSTM9 uses a simpler LSTM network than CLSTM5 and CLSTM8 but has a higher F1 score than the other two models. In contrast, CLSTM3 has more complex LSTM modules than CLSTM1 and CLSTM4, while CLSTM3 has a higher F1 score than CLSTM1 and CLSTM4. From the above observation, the F1 value is not related to the structural complexity of LSTM but to the suitable combination of hidden layers and the number of layers. However, the structural complexity of LSTM is related to the number of false alarms: both CLSTM8 and CLSTM9 use C6 and one-way LSTM, but CLSTM8 has a much higher number of hidden layers and layers than CLSTM9, resulting in 44 false alarms in CLSTM9, which is greater than the 6 false alarms in CLSTM8. Furthermore, CLSTM2 and CLSTM5 have more simple LSTM structures than CLSTM1, CLSTM3, and CLSTM4, which results in more than 17 false alarms for all of them, but less than 10 for the other 3 models. This indicates that a reasonable combination of LSTM arguments can improve the feature extraction ability of the model for the temporal dimension, thus reducing the false alarm rate of the model.

Back to the accuracy, the accuracy of the CLSTM models with the C6 network structure is above 97%, which is significantly higher than CLSTM6 with C2 and CLSTM7 with C3. CLSTM3, which has the same LSTM structure as these two models, even differs in accuracy by up to 2.62%. With the above analysis, it can be concluded that a suitable LSTM can improve the model recognition performance when the sample data contain multiple vibration events. The structure of the CNN has a large impact on the model accuracy. Thus, a reasonable CNN needs to be designed for different situations to extract the features of each image frame.

The lower the number of false and missed alarms, the better the model performance. Overall, although CLSTM3 does not have the lowest number of false and missed alarms, the sum is the best among the nine models.

Finally, the observation combining the two metrics of memory occupied by the model and the recognition time consumption reveals that the smaller the memory occupied by the model, the less time consuming it is to recognize the vibrational dynamic sequences, but only with a millisecond-level impact.

In summary, by analyzing the model structure and comparing the nine models, CLSTM3 can achieve a better balance in each performance metric and identify multi-vibration sequences quickly.

5.5. Comparison of Models Based on Different CNNs

Through the experiments in Section 5.3, the vibrating sequence recognition model is trained on the multi-vibrating sequence training set and achieves 99.33% recognition accuracy. It is analyzed that the CNN structure highly affects the model recognition performance, so the effect of CNN structure on model performance needs further discussion. This subsection compares the performance metrics of the proposed CNN structure with LetNet [47], AlexNet [48], and different VGG networks [44]. Figure 12 shows the accuracy, recall, F1 score, precision, and inference time of six models using the same LSTM structure. The CLSTM3 model containing six convolutional layers has the highest precision score of 99.64%, indicating that this model has the best detection confidence for different vibration frequency classes. The sequence recognition models based on VGG13 and VGG16, which contain 10 and 13 convolutional layers, respectively, have closer scores on precision, 99.38% and 99.33%, respectively, second only to CLSTM3. Both VGG-based models consume more time than CLSTM3 for recognition. The difference between the VGG13-based network and CLSTM3 in the recall, F1 score, and accuracy is insignificant, but the deeper convolutional network's disadvantage is detection speed. Although LetNet and AlexNet are faster than CLSTM3 in recognition speed, the overall performance is not as good as CLSTM3. In addition, excessive structural complexity results in an overfitting-like phenomenon, i.e., the metrics on the test set decrease instead as the model complexity rises. It illustrates that the CNN network structure of CLSTM3 is more appropriate for the identification of multi-vibration target data collected by Φ-OTDR.

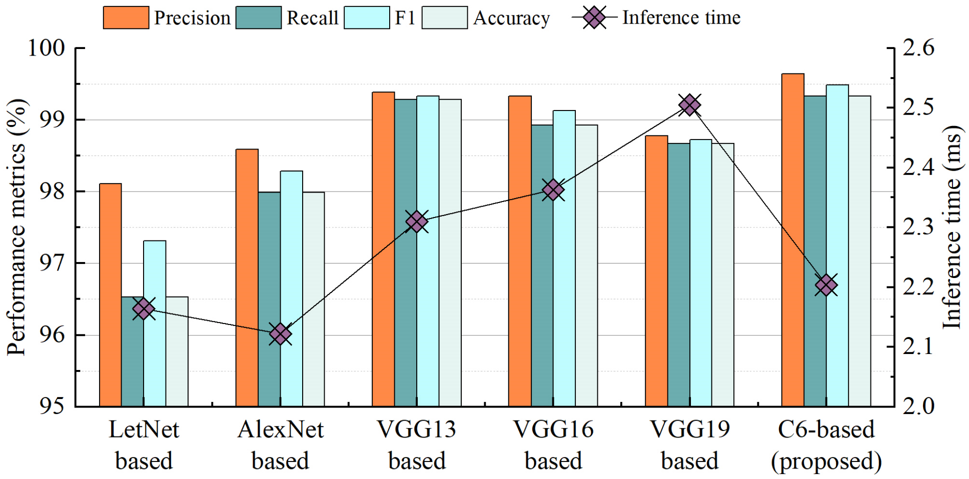


Figure 12. Effect of different CNN-based models on recognition performance.

6. Conclusions and Future Work

This paper proposes a fast digital demodulation method of Φ-OTDR dynamic vibration sequence with single detection and dual acquisition by theoretical analysis of Φ-OTDR back propagation light intensity detection and coherence detection. It is demonstrated that combining the advantages of both methods can rapidly demodulate the dynamic vibration sequences of fiber optic detection. On this basis, variable-length data in different vibration frequency bands are collected and formed into a dataset for sequence recognition model training, validation, and testing. The proposed end-to-end vibration sequence recognition method contains CNN, LSTM, and CTC transcription layers. CTC solves the problem of automatic alignment of multi-target time series. Using the CNN layer effectively extracts the vibration signal time–frequency map features and significantly improves the model detection accuracy, precision, and other metrics. A properly configured LSTM layer can extract contextual sequence features and reduce the occurrence of false alarms. Experiments show that the proposed CLSTM3 model can recognize multiple vibration sequences in different frequency bands. Compared with the other eight CLSTM models using the same framework, it achieves a satisfactory balance in various statistical metrics. It outperforms LetNet, AlexNet, and other classical models containing the VGG structure. In contrast to traditional sequence recognition methods [28,29], the described method achieves alignment-free end-to-end training of multiple target sequences with different vibration durations in the dataset, avoiding laborious manual labeling. The end-to-end vibration sequence recognition method combines the feature extraction advantages of deep learning and the sequence automatic alignment of CTC, which is easy to train and can effectively distinguish vibration sources in different frequency bands.

Future work will focus on updating the hardware system, e.g., by selecting an AOM with a lower operating frequency, it indirectly reduces the data acquisition card sampling rate and, thus, the data processing time, optimizing the signal demodulation algorithm for application to portable detection, optimizing the existing model for collecting more dynamic vibration signals in practical application environments. As well, further validation of the model performance in practical scenarios should be done to improve the robustness and adaptability of the model, aiming to achieve automation of Φ-OTDR vibration recognition and reduce the false alarm rate of the system.

Author Contributions: All authors contributed substantially to the manuscript. Conceptualization, N.Y., Y.Z., F.W. and J.C.; methodology, N.Y. and F.W.; software, N.Y.; validation, N.Y. and F.W.; formal analysis, N.Y. and F.W.; investigation, N.Y., F.W. and J.C.; resources, N.Y., Y.Z., F.W. and J.C.; data curation, N.Y.; writing—original draft preparation, N.Y.; writing—review and editing, Y.Z., F.W. and J.C.; visualization, N.Y.; supervision, Y.Z.; project administration, Y.Z.; funding acquisition, Y.Z., F.W. and J.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the China Postdoctoral Fund (Grant No. 2017M623349).

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Acknowledgments: We thank the reviewers for their valuable feedback and the editors' devoted guidance.

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

Abbreviations

Φ-OTDR	Phase-sensitive Optical Time-domain Reflectometer
RBSs	Rayleigh Backscatter Signals
CNN	Convolution Neural Network
Bi-LSTM	Bidirectional Long Short-term Memory
CTC	Connectionist Temporal Classification
NARs	Nuisance-alarm Rates
L.O.	Local Oscillator
EMD	Empirical Mode Decomposition
MFCC	Mel-scale Frequency Cepstral Coefficients
W.D.	Wavelet Decomposition
WPD	Wavelet Packet Decomposition
XGBoost	Extreme Gradient Boosting
k-NNs	K-nearest Neighbors
SVMs	Support Vector Machines
GMMs	Gaussian Mixture Models
LSTM	Long Short-time memory
HMM	Hidden Markov Model

AOM	Acoustic-optic Modulator
EDFA	Erbium-doped Fiber Amplifier
P.D.	Photodetector
SMF	Single-mode Fiber
DAQ	Data Acquisition Module
BPD	Balanced Photodetector
STFT	Short-time Fourier Transform
FFT	Fast Fourier Transform
RNN	Recurrent Neural Networks
P.C.	Personal Computer
T.P.	True Positive
T.N.	True Negative
F.P.	False Positive
F.N.	False Negative
PZT	Piezoelectric Ceramics

References

1. Shang, Y.; Sun, M.; Wang, C.; Yang, J.; Du, Y.; Yi, J.; Zhao, W.; Wang, Y.; Zhao, Y.; Ni, J. Research Progress in Distributed Acoustic Sensing Techniques. *Sensors* **2022**, *22*, 6060. [CrossRef]
2. Muanenda, Y. Recent Advances in Distributed Acoustic Sensing Based on Phase-Sensitive Optical Time Domain Reflectometry. *J. Sens.* **2018**, *2018*, 3897873. [CrossRef]
3. Wang, Z.N.; Zeng, J.J.; Li, J.; Fan, M.Q.; Wu, H.; Peng, F.; Zhang, L.; Zhou, Y.; Rao, Y.J. Ultra-long phase-sensitive OTDR with hybrid distributed amplification. *Opt. Lett.* **2014**, *39*, 5866–5869. [CrossRef] [PubMed]
4. Wang, Y.; Yuan, H.; Liu, X.; Bai, Q.; Zhang, H.; Gao, Y.; Jin, B. A Comprehensive Study of Optical Fiber Acoustic Sensing. *IEEE Access* **2019**, *7*, 85821–85837. [CrossRef]
5. Bao, X.; Zhou, D.P.; Baker, C.; Chen, L. Recent Development in the Distributed Fiber Optic Acoustic and Ultrasonic Detection. *J. Light. Technol.* **2017**, *35*, 3256–3267. [CrossRef]
6. Huang, X.-D.; Zhang, H.-J.; Liu, K.; Liu, T.-G. Fully modelling based intrusion discrimination in optical fiber perimeter security system. *Opt. Fiber Technol.* **2018**, *45*, 64–70. [CrossRef]
7. Tejedor, J.; Macias-Guarasa, J.; Martins, H.; Pastor-Graells, J.; Martin-Lopez, S.; Corredera, P.; Pauw, G.; Smet, F.; Postvoll, W.; Ahlen, C.; et al. Real Field Deployment of a Smart Fiber Optic Surveillance System for Pipeline Integrity Threat Detection: Architectural Issues and Blind Field Test Results. *J. Light. Technol.* **2017**, *36*, 1052–1062. [CrossRef]
8. Meng, H.; Wang, S.; Gao, C.; Liu, F. Research on Recognition Method of Railway Perimeter Intrusions Based on Φ-OTDR Optical Fiber Sensing Technology. *IEEE Sens. J.* **2021**, *21*, 9852–9859. [CrossRef]
9. Yang, N.; Zhao, Y.; Chen, J. Real-Time Φ-OTDR Vibration Event Recognition Based on Image Target Detection. *Sensors* **2022**, *22*, 1127. [CrossRef]
10. Kandamali, D.F.; Cao, X.; Tian, M.; Jin, Z.; Dong, H.; Yu, K. Machine learning methods for identification and classification of events in ϕ-OTDR systems: A review. *Appl. Opt.* **2022**, *61*, 2975–2997. [CrossRef]
11. Lu, Y.L.; Zhu, T.; Chen, L.A.; Bao, X.Y. Distributed Vibration Sensor Based on Coherent Detection of Phase-OTDR. *J. Light. Technol.* **2010**, *28*, 3243–3249. [CrossRef]
12. Izumita, H.; Koyamada, Y.; Furukawa, S.; Sankawa, I. Stochastic amplitude fluctuation in coherent OTDR and a new technique for its reduction by stimulating synchronous optical frequency hopping. *J. Light. Technol.* **1997**, *15*, 267–278. [CrossRef]
13. Pan, Z.; Liang, K.; Ye, Q.; Cai, H.; Qu, R.; Fang, Z. Phase-sensitive OTDR system based on digital coherent detection. In Proceedings of the Asia Communications and Photonics Conference and Exhibition (ACP), Shanghai, China, 13–16 November 2011; pp. 1–6.
14. Wang, Z.; Lou, S.; Liang, S.; Sheng, X. Multi-class Disturbance Events Recognition Based on EMD and XGBoost in φ-OTDR. *IEEE Access* **2020**, *8*, 63551–63558. [CrossRef]
15. Wang, X.; Liu, Y.; Liang, S.; Zhang, W.; Lou, S. Event identification based on random forest classifier for Φ-OTDR fiber-optic distributed disturbance sensor. *Infrared Phys. Technol.* **2019**, *97*, 319–325. [CrossRef]
16. George, J.; Mary, L.; Riyas, K.S. Vehicle Detection and Classification From Acoustic Signal Using ANN and KNN. In Proceedings of the International Conference on Control Communication and Computing (ICCC), Trivandrum, India, 13–15 December 2013; pp. 436–439.
17. Wang, Y.; Wang, P.; Ding, K.; Li, H.; Zhang, J.; Liu, X.; Bai, Q.; Wang, D.; Jin, B. Pattern Recognition Using Relevant Vector Machine in Optical Fiber Vibration Sensing System. *IEEE Access* **2019**, *7*, 5886–5895. [CrossRef]
18. Wu, H.; Qian, Y.; Zhang, W.; Tang, C. Feature Extraction and Identification in Distributed Optical-Fiber Vibration Sensing System for Oil Pipeline Safety Monitoring. *Photonic Sens.* **2017**, *7*, 305–310. [CrossRef]

19. Tangudu, R.; Sahu, P.K. Rayleigh Φ-OTDR based DIS system design using hybrid features and machine learning algorithms. *Opt. Fiber Technol.* **2021**, *61*, 102405. [CrossRef]
20. Tejedor, J.; Macias-Guarasa, J.; Martins, H.F.; Martin-Lopez, S.; Gonzalez-Herraez, M. A Multi-Position Approach in a Smart Fiber-Optic Surveillance System for Pipeline Integrity Threat Detection. *Electronics* **2021**, *10*, 712. [CrossRef]
21. Chen, J.; Wu, H.; Liu, X.; Xiao, Y.; Wang, M.; Yang, M.; Rao, Y. A real-time distributed deep learning approach for intelligent event recognition in long distance pipeline monitoring with DOFS. In Proceedings of the 10th International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC), Zhengzhou, China, 18–20 October 2018; pp. 290–296.
22. Wu, H.; Chen, J.; Liu, X.; Xiao, Y.; Wang, M.; Zheng, Y.; Rao, Y. One-Dimensional CNN-Based Intelligent Recognition of Vibrations in Pipeline Monitoring With DAS. *J. Light. Technol.* **2019**, *37*, 4359–4366. [CrossRef]
23. Xu, C.; Guan, J.; Bao, M.; Lu, J.; Ye, W. Pattern recognition based on time-frequency analysis and convolutional neural networks for vibrational events in φ-OTDR. *Opt. Eng.* **2018**, *57*, 016103. [CrossRef]
24. Cai, M.; Liu, J. Maxout neurons for deep convolutional and LSTM neural networks in speech recognition. *Speech Commun.* **2016**, *77*, 53–64. [CrossRef]
25. Yildirim, O. A novel wavelet sequence based on deep bidirectional LSTM network model for ECG signal classification. *Comput. Biol. Med.* **2018**, *96*, 189–202. [CrossRef] [PubMed]
26. Li, Z.; Zhang, J.; Wang, M.; Zhong, Y.; Peng, F. Fiber distributed acoustic sensing using convolutional long short-term memory network: A field test on high-speed railway intrusion detection. *Opt. Express* **2020**, *28*, 2925–2938. [CrossRef]
27. Yang, Y.; Li, Y.; Zhang, T.; Zhou, Y.; Zhang, H.; Assoc Advancement Artificial, I. Early Safety Warnings for Long-Distance Pipelines: A Distributed Optical Fiber Sensor Machine Learning Approach. In Proceedings of the 35th AAAI Conference on Artificial Intelligence 33rd Conference on Innovative Applications of Artificial Intelligence 11th Symposium on Educational Advances in Artificial Intelligence, Online, 2–9 February 2021; pp. 14991–14999.
28. Wu, H.; Liu, X.; Xiao, Y.; Rao, Y. A Dynamic Time Sequence Recognition and Knowledge Mining Method Based on the Hidden Markov Models (HMMs) for Pipeline Safety Monitoring With Phi-OTDR. *J. Light. Technol.* **2019**, *37*, 4991–5000. [CrossRef]
29. 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. [CrossRef]
30. Chen, X.; Xu, C. Disturbance pattern recognition based on an ALSTM in a long-distance φ-OTDR sensing system. *Microw. Opt. Technol. Lett.* **2020**, *62*, 168–175. [CrossRef]
31. Pan, Y.; Wen, T.; Ye, W. Time attention analysis method for vibration pattern recognition of distributed optic fiber sensor. *Optik* **2022**, *251*, 168127. [CrossRef]
32. 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. [CrossRef]
33. Lee, W.; Seong, J.J.; Ozlu, B.; Shim, B.S.; Marakhimov, A.; Lee, S. Biosignal Sensors and Deep Learning-Based Speech Recognition: A Review. *Sensors* **2021**, *21*, 1399. [CrossRef]
34. Graves, A.; Fernández, S.; Gomez, F.; Schmidhuber, J. Connectionist temporal classification: Labelling unsegmented sequence data with recurrent neural networks. In Proceedings of the 23rd International Conference on Machine Learning, Pittsburgh, PA, USA, 25–29 June 2006; pp. 369–376.
35. Shao, Y.; Liu, H.; Peng, P.; Pang, F.; Yu, G.; Chen, Z.; Chen, N.; Wang, T. Distributed Vibration Sensor With Laser Phase-Noise Immunity by Phase-Extraction φ-OTDR. *Photonic Sens.* **2019**, *9*, 223–229. [CrossRef]
36. Wang, Z.; Zhang, L.; Wang, S.; Xue, N.; Peng, F.; Fan, M.; Sun, W.; Qian, X.; Rao, J.; Rao, Y. Coherent Φ-OTDR based on I/Q demodulation and homodyne detection. *Opt. Express* **2016**, *24*, 853–858. [CrossRef] [PubMed]
37. Liang, K.; Pan, Z.; Zhou, J.; Ye, Q.; Cai, H.; Qu, R. Multi-parameter vibration detection system based on phase sensitive optical time domain reflectometer. *Chin. J. Lasers* **2012**, *39*, 119–123. [CrossRef]
38. Taylor, M.G. Phase Estimation Methods for Optical Coherent Detection Using Digital Signal Processing. *J. Light. Technol.* **2009**, *27*, 901–914. [CrossRef]
39. Wu, H.; Yang, M.; Yang, S.; Lu, H.; Wang, C.; Rao, Y. A Novel DAS Signal Recognition Method Based on Spatiotemporal Information Extraction With 1DCNNs-BiLSTM Network. *IEEE Access* **2020**, *8*, 119448–119457. [CrossRef]
40. Ma, X.; Hovy, E.H. End-to-end Sequence Labeling via Bi-directional LSTM-CNNs-CRF. *arXiv* **2016**, arXiv:1603.01354.
41. Graves, A. Connectionist Temporal Classification. In *Supervised Sequence Labelling with Recurrent Neural Networks*; Graves, A., Ed.; Springer: Berlin/Heidelberg, Germany, 2012; pp. 61–93.
42. Bacci, T.; Mattia, S.; Ventura, P. The bounded beam search algorithm for the block relocation problem. *Comput. Oper. Res.* **2019**, *103*, 252–264. [CrossRef]
43. Hu, Z.; Tang, Y.; Yang, S.; Pang, T.; Wu, B.; Zhang, Z.; Sun, M. A new phase demodulation method for φ-OTDR based on coherent detection. In Proceedings of the Proc.SPIE, Online, 8 January 2021; p. 116070M. [CrossRef]
44. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv* **2014**, arXiv:1409.1556. [CrossRef]

45. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
46. Yang, N.; Zhao, Y.; Chen, J.; Wang, F. Real-time classification for Φ-OTDR vibration events in the case of small sample size datasets. *Opt. Fiber Technol.* **2023**, *76*, 103217. [CrossRef]
47. Lecun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **1998**, *86*, 2278–2324. [CrossRef]
48. Krizhevsky, A.; Sutskever, I.; Hinton, G. ImageNet Classification with Deep Convolutional Neural Networks. *Neural Process. Syst.* **2012**, *25*, 84–90. [CrossRef]

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