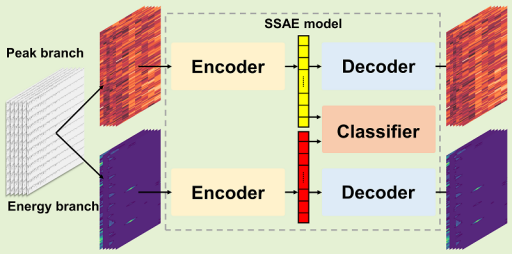
# Long-Distance Pipeline Safety Early Warning: A Distributed Optical Fiber Sensing Semi-Supervised Learning Method

Yiyuan Yang, Haifeng Zhang, and Yi Li

***Abstract*— Pipeline safety early warning (PSEW) systems based on distributed optical fiber sensors are used to recognize and locate third-party events that may damage long-distance energy transportation pipelines and are essential to ensure pipeline safety and energy supply. However, the deployment of PSEW systems in real sites is hindered by the high experimental cost of collecting large real-site data sets for model building and the small percentage of labeled data (typically less than 0.5%). Besides, the optical fiber sensors are sensitive to hardware and the environment, ensuring challenges to directly migrate the old PSEW system for a new deployment. In this study, a novel semi-supervised learning model is proposed to monitor the safety of pipelines in real-time. Concretely, the sparse stacked autoencoder trained with unlabeled data is used to extract more robust features, and the fully-connected network trained with a small amount of labeled data is used for location and identification. Encouraging empirical results on the real-world long-distance energy pipelines of the PipeChina confirm that our method achieves better recognition and localization performance in comparison to the baseline with less labeled data. Further, the model size and recognition latency are reduced by 18.9× and 7.9× of the baseline, respectively. Also, the decoded features have better visualization than the input. This work reduces the cost of PSEW system deployments, improves its performance and portability, and will contribute to the widespread use of PSEW systems in the industry.**




***Index Terms*— Pipeline safety early warning, Distributed optical fiber sensor, Semi-supervised learning, Sparse stacked autoencoder, Pattern recognition**

## I. Introduction

ENERGY pipeline integrity is crucial for safe operation and energy supply. With the recent sharp increase in energy pipeline mileage, pipeline safety early warning (PSEW) systems play a more important role to automatically recognize and locate third-party events that may damage long-distance energy transportation pipelines. They are applied to replace the inefficient and costly manual inspections. Energy pipelines are generally buried to reduce their floor space, which makes it challenging to directly observe their security. Fortunately, when deploying an energy pipeline, a single cable with multiple optical fibers is generally installed close to the pipe at the same time, and only a few fibers in the cable are used for communication. As such, other redundant fibers can be used without any additional hardware cost [1]. Currently, some researchers are using distributed optical fiber sensors (DOFSs) with the phase-sensitive optical time-domain reflectometry ($\phi$-OTDR) technology to detect vibrations that occur on the ground, close to the buried fiber, recognize the threat activities type around the pipeline, and provide early warning in real-time. They have the advantages of weak radiation, good real-time performance, antielectron magnetic interference, and long-distance distributed detection [2].

However, the signals of DOFSs are highly susceptible to the environment and hardware [3], coupled with their inherent characteristics of strong noise, weak signal, and signal drift problem in a time-space domain [4], making it challenging to use a certain method to identify and locate the third-party events under all conditions, and even more challenging to apply the same PSEW algorithm in different pipelines. Therefore, it places higher demands on algorithms that recognize signals. The main challenge in PSEW algorithms can be divided into the following: feature extraction and recognition algorithm [1].

Various features of optical fiber signals can be extracted from the time domain, frequency domain, time-frequency domain, and time-space domain. Tanimola and Hill [5] used transients and tonal features broadband noise in time-domain as the features. However, they did not consider the instability of raw data. Fouda *et al.* [6] utilized frequency-domain power spectral estimation of short-term energy to extract features. Zhang *et al.* [7] obtained the power spectral density

This work was supported by the pipeline branch of the China National Petroleum Corporation (CNPC) under grant number GDGS-2020-JS-24. *(Corresponding Author: Yi Li.)*

Y. Yang is with the Department of Automation, Tsinghua University, Beijing 100084, China (e-mail: yangyy19@mails.tsinghua.edu.cn).

H. Zhang is with the Research Institute of Tsinghua University, Pearl River Delta, Guangzhou 510530, China (e-mail: zhanghf@tsinghua-gd.org).

Y. Li is with the International Graduate School at Shenzhen, Tsinghua University, Shenzhen 518055, China (e-mail: liyi@sz.tsinghua.edu.cn).

and designed an adaptive filter bank to improve the Mel-frequency cepstrum coefficient as feature maps. Nevertheless, the sampling frequency of the signal is rather high, where only computing using the frequency-domain may result in losing the time-domain information of the raw signal. Wang *et al.* [8] used wavelet transform to decompose data and calculate the normalized energy of the reconstruction coefficient. Such time-frequency domain features also include empirical mode decomposition and short-time fast Fourier transform. Although the time-frequency features have effective representations, the high computational complexity hinders the real-time performance of long-distance PSEW systems. Recently, some researchers noticed the spatio-temporal correlation of the distributed signal. As such, they analyzed it together with the time-frequency method. Sun *et al.* [9] applied image processing methods for denoising and extracting features. Moreover, Yang *et al.* [10] proposed two complementary features, which were dimensionless with reliable robustness for various deployments and had good visualization and real-time performance for industrial monitoring.

Considering the recognition algorithm, it is common to use machine learning and deep learning methods (DL) [11]. Sheng *et al.* [12] improved the stochastic configuration network through truncation singular value decomposition. Wang *et al.* [8] and Sun *et al.* [9] applied relevant vector machines, and Xu *et al.* [13] used support vector machines. Besides, Wu *et al.* [14] and Tejedor *et al.* [15] proposed a multi-layer perceptron model. Chen *et al.* [16] and Kong *et al.* [17] used a probabilistic neural network, which does not require back-propagation to optimize parameters. Also, Wang *et al.* [18] and Huang *et al.* [19] used the random forest for classification.

With the development of DL and the advent of the data-driven era, researchers have attempted to fuse feature extraction and event classification into an end-to-end DL model. Many scholars treated DOFS signals as images and used convolutional neural networks (CNNs) or their variants for feature extraction and fully connected networks (FCNs) for classification [20]–[24]. In particular, Yang *et al.* [25] utilized novel bilinear CNNs and LighGBM model, whose extracted features were well visualized. In addition, temporal signals have long-term dependence and variable correlation length bidirectionally because of the $\phi$-OTDR signal scattering and reverse propagation. As such, Bai *et al.* [26], Li *et al.* [27], and Yang *et al.* [10] presented long short-term memory (LSTM), convolutional LSTM, and bi-directional LSTM (BiLSTM) models combined with CNNs. Particularly, Yang *et al.* [10] deployed the model to a real long-distance energy pipeline and achieved over 99% recognition accuracy in real-time.

Currently, there are some PSEW systems already deployed on real pipelines [10], [28], [29]. Nevertheless, they continue to have the following pain points: (1) The spatial migration capability is poor. Since DOFSs are significantly affected by environments, it is challenging to deploy the trained algorithm in a pipeline directly into another pipeline. (2) The algorithm cannot solve the signal time drift well. Due to changes in environmental and hardware conditions over time, such as temperature and soil moisture content, the performance of the method deployed can degrade significantly after a period. (3) The model size and latency may affect the performance of the system. As the distance of monitoring pipelines becomes longer, the model size needs to be reduced to help the deployed algorithm with less latency to ensure real-time monitoring. (4) The cost of collecting data is rather high. As building the DL model requires the collection of large amounts of samples manually, it requires professionally skilled technicians to travel along the pipe to simulate intrusions events, which is manpower-intensive and consumes significant amounts of time [10]. (5) The sample utilization is very small. The DOFS can collect data from every point along the pipe, but experiments are generally undertaken in a small area every time, whose data is labeled, while the signals from other locations are simultaneously unlabeled. For a 50 km pipeline, experiments are conducted in 100 m ranges, the percentage of labeled data is only 0.2%, and the other unlabeled data is not available in the existing works. However, the unlabeled samples contain several signal features, which causes a huge waste.

To address the above problems, with strong industrial demands, semi-supervised learning is employed to improve the utilization of unlabeled data to reduce the experimental cost and address the problem of model migration with smaller model size and latency. The contributions are summarized as follows. (1) A novel semi-supervised learning model is proposed for recognition and spatiotemporal localization of damage events based on sparse stacked autoencoder (SSAE), including CNNs, BiLSTMs, and self-attention methods. (2) It is demonstrated that the proposed method can use unlabeled data with less labeled data to obtain better results in comparison to the baseline, which significantly improves sample utilization and reduces costs. (3) Through the experiments on operating long-distance pipelines, it is proved that the encoding learned from unlabeled data has good spatiotemporal transferability. (4) It is verified that the features decoded by the SSAE model will highlight more useful features for better visualization, which can help supervisors visually monitor the situation for more intuitive analysis. (5) It is confirmed that the model adopted in this study has a smaller size and latency than the baseline, which can match various hardware.

The remainder of this paper is organized as follows. The background and basic algorithms are presented in Section II. The approach of the proposed semi-supervised learning model is described elaborately in Section III. Section IV shows the comprehensive experimental result and discussions. A summary and future direction are provided in Section V.

## II. Background and Basic Algorithms

### A. Energy PSEW Systems

The simplified structure of the energy PSEW system is shown in Fig. 1. The yellow line is the optical fiber cable, which is generally installed within 25 cm away from the underground energy pipe. An ultra-narrow linewidth light (UNLL) source with a center wavelength of 1550 nm is deployed in the monitor center. Subsequent to modulating the continuous light into probe pulses by an acoustic optic modulator (AOM) and compensating for light energy losses by an erbium-doped fiber amplifier (EDFA), the amplified probe pulses are injected into

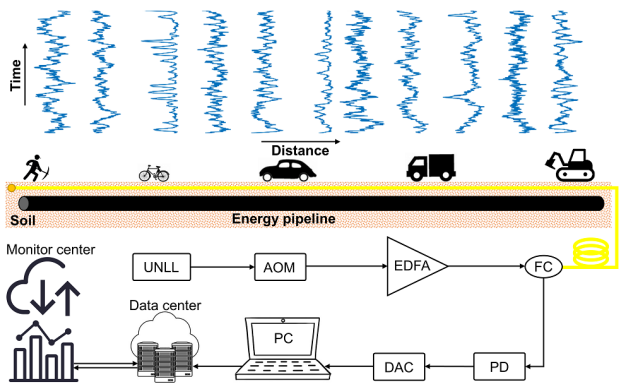



Fig. 1. Structure of the energy PSEW system.

the fiber through a fiber circulator (FC). The vibrations caused by external events, such as cars and excavations, can cause changes to the fiber signal, then the Rayleigh backscattered light containing information of those events is directly routed to a photoelectric detector (PD). Through a data acquisition card (DAC) detecting the intensity evolution over time, the data will be processed by a personal computer (PC) [30]. The PC uploads the raw data to the data center together with the warning information (i.e., the type of event identified and the spatio-temporal coordinates of the warning), via the PSEW algorithm. The data center collates the information and aggregates it to the monitor center. Correspondingly, the monitor center sends the processing results of each alert back to the data center to be stored regularly.

## B. Basic Models and Algorithms

*1) Simple Autoencoder:* Autoencoder (AE) is a popular DL-architecture model in which the raw data is compressed into smaller dimensional encodings and reconstructed at the output. Also, these encodings can represent the data with minimum reconstruction loss. The structure of a simple AE with 2-layers is shown in Fig. 2, where the raw input $X \in R^N$ is compressed to a feature map $Y \in R^F$ and reconstructed into $X' \in R^N$. In detail, the AE in Fig. 2 satisfies

$$\begin{aligned} Y &= \sigma_2(\boldsymbol{W}_2(\sigma_1(\boldsymbol{W}_1\boldsymbol{X} + \boldsymbol{b}_1)) + \boldsymbol{b}_2) \\ X' &= \sigma_1(\boldsymbol{W}_1'(\sigma_2(\boldsymbol{W}_2'\boldsymbol{Y} + \boldsymbol{b}_2')) + \boldsymbol{b}_1') \end{aligned} \tag{1}$$

where $\{\boldsymbol{W}_1, \boldsymbol{W}_2, \boldsymbol{W}_1', \boldsymbol{W}_2'\}$ are the learned weight matrix in the first and second layers of the encoder and decoder, respectively; and $\{\boldsymbol{b}_1, \boldsymbol{b}_2, \boldsymbol{b}_1', \boldsymbol{b}_2'\}$ are the relative bias. In addition, $\{\sigma_1(\cdot), \sigma_2(\cdot)\}$ represent the activation functions, which are generally nonlinear. Then we use an optimization algorithm, such as Adam, to minimize the mean square error in (2), and we obtain the optimal $\boldsymbol{W}_{opt}$ and $\boldsymbol{b}_{opt}$ values.

$$L(\boldsymbol{X}, \boldsymbol{X}') = \frac{1}{N} \left\| \boldsymbol{X} - \boldsymbol{X}' \right\|_2^2 \tag{2}$$

*2) Sparse Autoencoder:* When there are many hidden layer nodes, simple AEs cannot automatically learn important features. As such, some constraints are required to be imposed on the hidden layer nodes. Sparse AE applies some sparsity constraints on the hidden layer nodes, which achieves a sparse effect by suppressing most of the hidden layer neurons. The

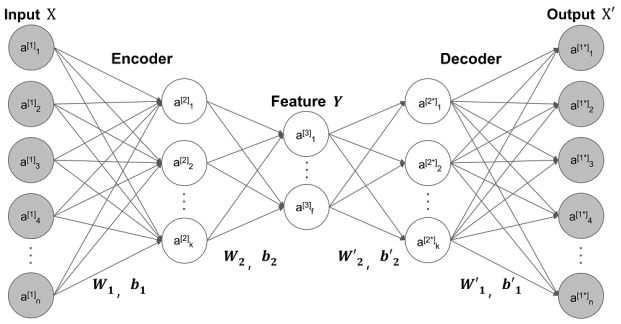



Fig. 2. A simple autoencoder with 2-layers.

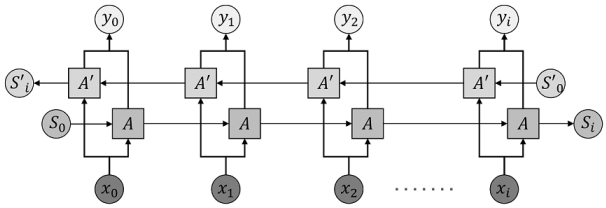


Fig. 3. The structure of BiLSTM model.

loss function of second layer sparse AE is shown in (3)–(5).

$$J_{sparse}(\boldsymbol{W}, \boldsymbol{b}) = \frac{1}{N} \left\| \boldsymbol{X} - \boldsymbol{X}' \right\|_2^2 + \beta \sum_{j=1}^{K} KL(\rho || \hat{\rho}_j) \tag{3}$$

$$KL(\rho || \hat{\rho}_j) = \rho \log \frac{\rho}{\hat{\rho}_j} + (1 - \rho) \log \frac{1 - \rho}{1 - \hat{\rho}_j} \tag{4}$$

$$\hat{\rho}_j = \frac{1}{N} \sum_{i=1}^{N} [a_j^{(2)}(x^i)] \tag{5}$$

where $\beta$ denotes the hyper-parameters that control the weights of the sparsity (i.e., KL divergence). $\rho$ is defined as the expected average activation value, which is generally a smaller value close to 0, and $\rho_j$ is the average activation of the $j$-th node. $a_j^{(2)}(x^i)$ is identified as the activation value of the $j$-th node under the input $x^i$.

*3) Stacked Autoencoder:* Stacked AE is a DL-based model composed of multi-layer AE in which the output of the former AE serves as the input of the later AE. More specifically, stacked AE is trained using the greedy layer-wise pre-training method. For example, the first layer of the stacked AE is trained with the input to obtain the weight $\boldsymbol{W}_1$ and bias $\boldsymbol{b}_1$, and then the weight and bias of the first layer are kept constant, the output of the first layer is used as the input of the second layer, and the weight $\boldsymbol{W}_2$ and bias $\boldsymbol{b}_2$ of the second layer are trained. In this order, all the parameters need to be trained. Subsequent to initialization, all the parameters need to be finetuned.

*4) BiLSTM:* The BiLSTM model is an extension of the basic LSTM model where two LSTMs are used separately for input, as shown in Fig. 3. The input of the first LSTM $A$ is a sequential sequence of the raw data, the reverse form of the input sequence is fed into the second LSTM $A'$, and the forward $S$ and reverse sequences $S'$ are concatenated together to obtain the output $y$. Applying the LSTM model twice aims to obtain long-term dependence, variable correlation length and bidirectional and complex relations appropriately.

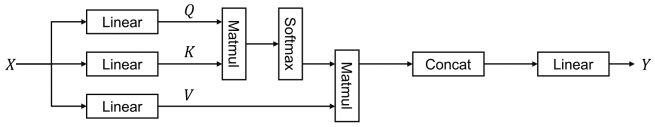



Fig. 4. The structure of self-attention.

*5) Self-attention:* Self-attention is a special case of attention methods that only use one sequence to compute its representation [31], which can obtain the dependencies not only between the source and target, but also between the source or the target itself. In comparison to BiLSTM, self-attention can enhance the ability to obtain long-term dependent information and have less parameters that need to be trained. The specific calculation process is as follows and in Fig. 4. Giving weighting matrices $W_Q, W_K, W_V \in R^{n\times t}$, multiply the input $X \in R^{t\times d}$ with the above three weight matrices, respectively to obtain the corresponding results $Q \in R^{n\times d}$ named query, $K \in R^{n\times d}$ named key, and $V \in R^{n\times d}$ named value, as shown in (6).

$$Q = W_Q X \qquad K = W_K X \qquad V = W_V X \tag{6}$$

where $d$ is the number of hidden units of the model, and the i-th attention value is calculated as (7).

$$Attention_i(Q, K, v_i) = Softmax(\frac{QK^T}{\sqrt{d}})v_i \tag{7}$$

Finally, all the values above are added to obtain the output $y \in R^{n\times d}$, as shown in (8).

$$y = \sum_i Attention_i(Q, K, v_i) \tag{8}$$

## III. Methodology

### A. Signal Compensation and Feature Extraction

When the light source is transmitted in a fiber, its energy shows exponential attenuation over the propagation distance due to deployment reasons, such as bending of the fiber during laying, uneven material, and fusion splicing, and due to inherent consumption, such as Rayleigh scattering [1]. It causes the low signal-to-noise ratio (SNR) at the end of the fiber, and affects the algorithms performance. Recently, engineers have used signal relay amplifiers (SRAs) at approximately every 25 km intervals along the fiber for compensation. However, the signals between SRAs continue to suffer from exponential attenuation. Therefore, each observed point is normalized between two SRAs by dividing the signal value by the average signal value over 12 hours of continuous updating.

Considering the actual industrial use, the extracted features need to be well visualized to help supervisors monitor them more intuitively. In addition, good interpretability, real-time performance, and robustness are also required. As such, the feature generator in the Algorithm of Yang [10] is applied to extract two dimensionless and complementary features (i.e., peak feature with high-frequency and instantaneous information and energy feature with low-frequency and continuous characteristics). Furthermore, sliding window operations are used on the raw distributed signal to ensure that the features can obtain the spatio-temporal information. After repeated experiments, it was found that 7 is the best number of observed point in the space-domain and window size is 100 in the time-domain (i.e., the input size of each branch is 100×7×1, which corresponds to 20 s of 500 Hz signal and 100 s of 100 Hz signal, and 120 m of the pipeline).

TABLE I
Network Specification

| | Layer | Shape | Activation function |
|---|---|---|---|
| Encoder | BN1 ($E^1$) | (100,7,1) | – |
| | Conv1 ($E^1$) | (100,7,16) | Relu |
| | Max-pool1 ($E^1$) | (50,7,16) | Maxpool |
| | Conv2 ($E^2$) | (50,7,32) | Relu |
| | Max-pool2 ($E^2$) | (25,7,32) | Maxpool |
| | Conv3 ($E^3$) | (25,7,64) | Relu |
| | Max-pool3 ($E^3$) | (5,1,64) | Maxpool |
| | Reshape ($E^4$) | (5,64) | – |
| | BiLSTM ($E^4$) | (5,64) | – |
| | Self-attention ($E^4$) | (5,64) | – |
| Decoder | Reshape | (5,1,64) | – |
| | UpSampling1 | (25,7,64) | Bilinear |
| | Conv4 | (25,7,64) | Relu |
| | UpSampling2 | (50,7,64) | Bilinear |
| | Conv5 | (50,7,32) | Relu |
| | UpSampling3 | (100,7,32) | Bilinear |
| | Conv6 | (100,7,16) | Relu |
| | Conv7 | (100,7,1) | Sigmoid |
| Classifier | Concatenate | (10,64) | – |
| | Dropout | (10,64) | – |
| | Flatten | (640) | – |
| | Dense1 | (128) | Relu |
| | BN2 | (128) | – |
| | Dense2 | (4) | Softmax |

### B. SSAE Algorithm for Event Recognization

The SSAE model is applied as the event recognizer. More specifically, the fusion of stacked AE with sparse AE (i.e., SSAE) results in a small model that tries to fit better by layer-wise training and sparsity compression, thus reducing the model size and improving the real-time performance and robustness. Moreover, BiLSTM and self-attention mechanisms are pioneeringly utilized as independent elements in the SSAE model in which BiLSTM can effectively solve the bi-directional dependence problem caused by signal scattering and reflection, and self-attention can obtain the long-term dependence information of the signal in the time-domain and reduce the number of parameters. In short, SSAE can efficiently compress the distributed and spatio-temporal signal through progressive abstraction levels in an unsupervised manner. Subsequent to training, the input can then be reduced to the feature map of the deepest layer for further classification.

Considering the model size, real-time performance, and the ease of continuing end-to-end training, FCNs are applied for location and classification training with a small amount of labeled data based on the feature map extracted by the encoder. Also, batch normalization (BN) is utilized to avoid distributed data bias and keep the data away from saturation zones, which can speed up training and improve accuracy. Besides, max-pooling and dropout strategies aim to simplify the model and alleviate overfitting. The detailed model specification with encoder-decoder and classifier is shown in Table I and the flowchart of the SSAE model is shown in Fig. 5.

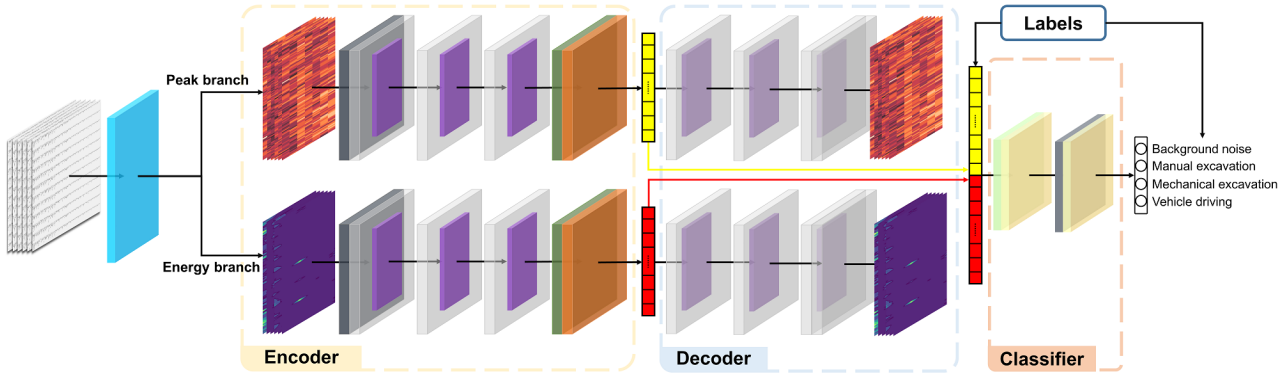



Fig. 5. Flowchart of SSAE model with encoder-decoder and classifier. Blue is the feature extractor from our previous work [10], dark grey is BN, light gray is Conv2d layer, purple is max-pooling, dark green is the BiLSTM layer, orange is self-attention mechanism, light green is dropout, and light yellow is the fully connected layer. The peak features (input of the upper submodel) and energy features (input of the lower submodel) from the feature extractor are expressed by heat maps and contour maps, respectively; and the yellow and red rectangles are their encodings.

**Algorithm 1** SSAE algorithm

**Input:** Peak and energy features of labeled data $L_p, L_e$ with label $T$, peak and energy features of unlabeled data $U_p, U_e$.
**Parameter:** The batch size of encoder-decoder and classifier $m_{ae}, m_c$, the learning rate $\eta$ and decay rate $d$, Adam hyperparameters $\alpha, \mu_{ae}, \mu_c$, AE and classifier iterations per loop $n_{ae}, n_c$, the number of greedy layer-wise pre-training $N$.
**Require:** Initial peak and energy AE $E_p(\cdot;\theta_{p_0^0}), E_e(\cdot;\theta_{e_0^0})$, initial classifier $C(\cdot;\theta_{c_0})$.

**while** $\theta_p, \theta_e$ have not converged **do** {**Part1: Encoder-decoder**}
  **for** each $i \in$ [1,$n_{ae}$] **do**
    **for** each $j \in$ [1,$m_{ae}$] **do**
      Shuffle and sample the unlabeled data $u_p, u_e \sim U_p, U_e$.
      $f_p^2, f_e^2 \leftarrow E_p^1(u_p;\theta_{p^1}), E_e^1(u_e;\theta_{e^1})$
      **for** each $k \in$ [2,$N$] **do**
        $f_p^{k+1}, f_e^{k+1} \leftarrow E_p^k(f_p^k;\theta_{p^k}), E_e^k(f_e^k;\theta_{e^k})$
      **end for**
      $f_p, f_e, u_p', u_e' \leftarrow E_p(u_p;\theta_p), E_e(u_e;\theta_e)$
    **end for**
    $\theta_p \leftarrow \text{Adam}(\nabla_{\theta_p} \sum_{j=1}^{m_{ae}} L(u_{p_j}, u_{p_j}'), \eta, d, \alpha, \mu_{ae})$
    $\theta_e \leftarrow \text{Adam}(\nabla_{\theta_e} \sum_{j=1}^{m_{ae}} L(u_{e_j}, u_{e_j}'), \eta, d, \alpha, \mu_{ae})$
  **end for**
**end while**
**while** $\theta_c$ has not converged **do** {**Part2: Classifier**}
  **for** each $i \in$ [1,$n_c$] **do**
    **for** each $j \in$ [1,$m_c$] **do**
      Shuffle and sample the labeled data $l_p, l_e \sim L_p, L_e$ and their label $t \sim T$, correspondingly.
      $\hat{f}_p, \hat{f}_e \leftarrow E_p(l_p;\theta_p), E_e(l_e;\theta_e)$
      $\hat{t} \leftarrow C([\hat{f}_p, \hat{f}_e];\theta_c)$
    **end for**
    $\theta_c \leftarrow \text{Adam}(\nabla_{\theta_c} \sum_{j=1}^{m_c} L(t, \hat{t}), \eta, d, \alpha, \mu_c)$
  **end for**
**end while**

The complete training process of the SSAE model is shown in Algorithm 1. For the encoder-decoder, unlabeled samples are utilized for training, while the peak and energy features obtained by the feature extractor are fed into two branches of the model after a shuffle. Based on the greedy layer-wise pre-training method, the encoder model is divided into four parts, namely $E^1, E^2, E^3, E^4$, as Table I shows, and are trained as follows. Considering one batch as an example, firstly, the original features are used as the input and $E^1$, (i.e., BN1, Conv1, and Max-pooling1 in Table I) are used as the encoder and trained with a fully symmetric decoder. The model is optimized with the aim that outputs are the same as the input to obtain the optimal parameters of $E^1$. Secondly, freezing the parameters of $E^1$ and its output is then applied as the input of the encoder $E^2$, where $E^2$ includes the Conv2 and Max-pooling2, and is also trained with its fully symmetric decoder. The same is continued till the training of $E^4$, (i.e., BiLSTM and self-attention) is completed. Finally, the complete model is fine-tuned as a whole using the original features to obtain the optimal parameters of the encoder-decoder model. Besides, the above operations are completely independent for the two branches of the model, but the process is identical.

Considering the classifier, a small amount of labeled data is used for training. Based on the optimal encoder-decoder model, all the parameters of its two branches are frozen and the outputs of the two self-attention layers in $E^4$ are concatenated. The same are then used as the input of the classifier. The classifer in Algorithm 1 is then trained for supervised learning, where the ground truth is the label of the raw data.

## IV. Experiments and Results

### A. Experimental Facility and Data Collection

Experimental facilities were deployed in an operational long-distance oil pipeline from the PipeChina Northern Pipeline Company Cangzhou Section. The pipeline tested was from Qingxian Station to Renqiu Station, which was approximately 85 km. Two SRAs were installed at 25 km and 56 km of the pipeline. The real-site deployment and experimental facility of the PSEW system are shown in Fig. 6. More specially, a redundant single-mode optical fiber was only used from the communication cable for monitoring as shown in Fig. 6(a), which was at no additional cost. The environment along the pipeline was complex, crossing several urban areas, rivers, and railway lines, and suffered from multiple types of noise superpositions and low SNR conditions.

1.69 TB of data was collected in the autumn of 2020, of which only 4.12 GB (i.e., 0.24% of total data) had labels. Particularly, two sample frequencies of 500 Hz and 100 Hz were tested. When the frequency is low, the hardware cost of

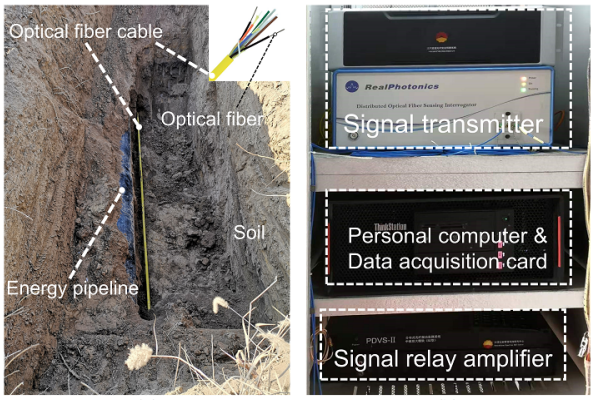



Fig. 6. Real-site deployment and experimental facility of the PSEW system. (a) Relative positions of the underground energy pipeline and optical fiber cable. (b) Deployed facilities.

TABLE II
DESCRIPTIONS OF THE LABELED DATASETS

| Event | Label | 100 Hz (#) | 500 Hz (#) |
|---|---|---|---|
| Background noise | 0 | 13,500 | 13,500 |
| Manual excavation | 1 | 3,920 | 3,600 |
| Mechanical excavation | 2 | 16,600 | 14,400 |
| Vehicle driving | 3 | 10,800 | 9,600 |

the facility is low. However, this results in less information being contained per unit time and the algorithms are more demanding. Balancing signal frequency and algorithm performance is of significant engineering interest. The labeled data collected was divided into four categories with the labels 0, 1, 2, and 3 sequentially, namely *background noise* (no events), *manual excavation* (potential energy theft by pipe drilling), *mechanical excavation* (potential third-party construction damaging the pipe), and *vehicle driving* (potential heavy vehicles rolling over the pipe). In order to verify the spatio-temporal robustness and transferability of the compressed features, the encoder-decoder was trained with unlabeled data from four different conditions, shown in Table III. And for the training of the classifier, the labeled data was divided into a training set, validation set, and testing set in the ratio 3:1:1. All the labeled and unlabeled datasets are shown in Tables II and III.

Further, mean square error was used as the loss of the encoder-decoder and sparse categorical cross-entropy as a classifer loss function, and applied Adam as the optimizer to monitor the loss of the validation set for hyperparameter tuning. Also, the initial learning rate was set at 0.001 and specified it to decay to 95% of the current value after every 10 epochs. The batch size was 128 with 20 epochs for the encoder-decoder model and 100 epochs for the classifier. A PC with an Intel Core i7-8700 CPU running at 3.2 GHz, a RTX 2070Super GPU, and 32 GB of RAM was used for training and verification. Considering the industrial deployment, TensorFlow 2.1.0 was used for model building under Python 3.7.3.

## B. Case Study and Discussion

*1) Visualization of Decoded Features:* Fig. 7 provides the results of the decoded features in comparison to the input

TABLE III
DESCRIPTIONS OF THE UNLABELED DATASETS

| Pipeline info | Duration (min) | Length (km) | Data size (#)[3] |
|---|---|---|---|
| Pipe1 in 2016 summer[1] | 30 | 48 | 418,696 |
| Pipe1 in 2016 winter | 30 | 48 | 418,672 |
| Pipe2 in 2020 spring[2] | 10 | 42 | 122,262 |
| Pipe2 in 2020 autumn | 15 | 85 | 370,685 |

[1] Pipe1: The gas pipe of the West-East Natural Gas Transmission Project Suzhou Section.
[2] Pipe2: The oil pipe of the PipeChina Northern Pipeline Company Cangzhou Section.
[3] For both 100 Hz and 500 Hz dataset.

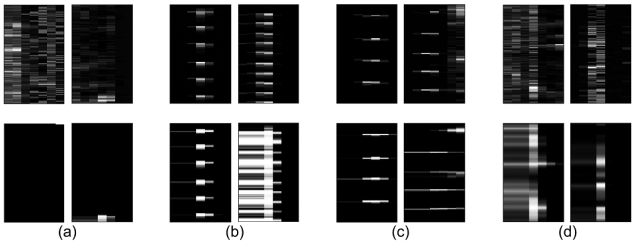



Fig. 7. Decoded features (lower) vs. input features (upper).

features. The upper portion is the input energy features and the lower portion is the corresponding decoded energy features. (a) is the background noise and the results show that the decoded features can filter out various noises of the signal. (b) and (c) are manual and mechanical excavations, respectively; and each light-colored region represents a single action. The results display that the decoded features can highlight more useful features. Different from the distinct features that are already present in manual and mechanical excavations, vehicle driving (d) is harder to distinguish from background noise when the SNR is small. However, the decoded features produce a more significant difference to the background noise. For practical application scenarios, the decoded features can help supervisors visually monitor pipes for more intuitive analysis.

*2) Identification Performance:* To verify the identification performance of the proposed SSAE model and the spatiotemporal transferability of the encoding features, four different unlabeled datasets were utilized in Table III to train the encoder-decoder separately and used all the labeled data in Table II to train the classifier, and the results with an average of 10 repeated experiments are presented in Table IV for 100 Hz and 500 Hz data. For recognition performance, it can be observed that all evaluation indices[1] are above 90% (0.9) based on the different unlabeled datasets and the collection frequency. In particular, the recognition of background noise is rather good as most of the real labels of the unlabeled data are background noise. The recognition of manual and mechanical excavations is slightly better than that of vehicle driving as the former features are inherently more distinctive, while the latter do not have standard periodic or spatiotemporal features. As for spatiotemporal transferability, although the labeled data were collected in pipe2 in the autumn of 2020 and trained with unlabeled data in different conditions, the SSAE model used in this study can also achieve results that meet the requirements of industrial applications, which confirmed that the encoding

[1] Acc is accuracy. Acc = (TN+TP) / (N+P)×100%. Se is sensitivity. Se = TP / (TP+FN)×100%. Sp is specificity. Sp = TN / (TN+FP)×100%. AUC is the area under the receiver operating characteristic curve.

TABLE IV
PERFORMANCE OF THE SSAE ALGORITHM TRAINED WITH DIFFERENT UNLABELED DATASETS

| Data | Data info | Background noise | | | | Manual excavation | | | | Mechanical excavation | | | | Vehicle driving | | | | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Acc(%) | Se(%) | Sp(%) | AUC(#) | Acc(%) | Se(%) | Sp(%) | AUC(#) | Acc(%) | Se(%) | Sp(%) | AUC(#) | Acc(%) | Se(%) | Sp(%) | AUC(#) | Acc(%) |
| 100 Hz | Pipe1 in 2016 summer | 94.81 | 95.33 | 94.28 | 0.951 | 95.23 | 94.67 | 95.78 | 0.956 | 93.65 | 93.93 | 93.36 | 0.939 | 93.28 | 92.68 | 93.87 | 0.935 | 94.05 |
| | Pipe1 in 2016 winter | 97.04 | 96.97 | 97.10 | 0.976 | 94.92 | 95.57 | 94.27 | 0.953 | 93.77 | 92.86 | 94.67 | 0.941 | 93.62 | 93.46 | 93.78 | 0.940 | 94.82 |
| | Pipe2 in 2020 spring | 94.01 | 93.28 | 94.73 | 0.943 | 93.02 | 92.28 | 93.75 | 0.934 | 94.26 | 93.50 | 95.02 | 0.946 | 92.32 | 91.88 | 92.76 | 0.929 | 93.61 |
| | Pipe2 in 2020 autumn | 98.37 | 97.90 | 98.83 | 0.987 | 95.38 | 93.72 | 97.03 | 0.961 | 94.05 | 93.56 | 94.53 | 0.943 | 93.70 | 93.16 | 94.23 | 0.941 | 95.38 |
| 500 Hz | Pipe1 in 2016 summer | 99.22 | 99.18 | 99.26 | 0.998 | 96.27 | 97.23 | 95.30 | 0.966 | 97.95 | 97.93 | 97.97 | 0.982 | 94.58 | 94.73 | 94.43 | 0.949 | 97.43 |
| | Pipe1 in 2016 winter | 99.21 | 99.02 | 99.40 | 0.998 | 96.04 | 96.10 | 95.98 | 0.964 | 96.46 | 95.63 | 97.28 | 0.967 | 94.89 | 95.61 | 94.16 | 0.950 | 96.96 |
| | Pipe2 in 2020 spring | 99.93 | 99.86 | 100.0 | 0.999 | 93.82 | 94.31 | 93.33 | 0.943 | 94.07 | 93.20 | 94.94 | 0.945 | 93.99 | 93.88 | 94.09 | 0.943 | 95.95 |
| | Pipe2 in 2020 autumn | 100.0 | 100.0 | 100.0 | 1.000 | 98.28 | 98.82 | 97.73 | 0.984 | 97.98 | 97.20 | 98.75 | 0.982 | 94.70 | 93.51 | 95.88 | 0.951 | 97.90 |

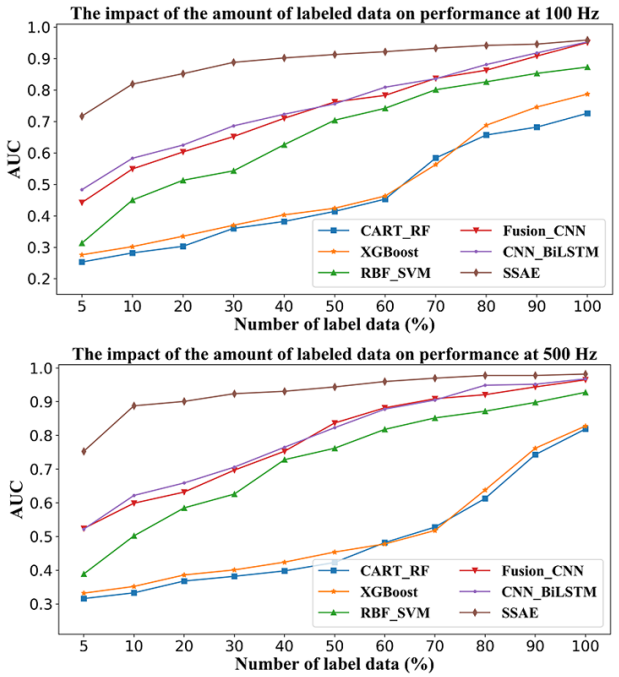



Fig. 8. The impact of the amount of labeled data on performance with various methods at 100 Hz and 500 Hz.

has good spatiotemporal transferability. In addition, the results using 500 Hz data are better than those using 100 Hz data, due to the fact that 500 Hz data has a greater amount of signal fluctuation information per unit of time. It is worth noting that the relatively poor results for pipe2 in the spring of 2020 are due to the fact that the equipment was recently deployed. As such, the signal was less stable.

*3) Impact of the Amount of Labeled Data on Performance:* The impact of the amount of labeled data on performance with various methods considering an average of 10 repeated experiments is presented in Fig. 8. The comparisons include classification and regression tree-based random forest (CART_RF), extreme gradient boosting (XGBoost), radial basis function kernel support vector machine (RBF_SVM), fusion CNN, CNN with BiLSTM, and proposed SSAE method, which are all described in Chapter I. Considering the proposed SSAE model, the unlabeled data of pipe2 in the autumn of 2020 was used to train the encoder-decoder model. It is notable that because the datasets and goals are different, it is impossible to make a direct comparison with the baselines of previous related works. As such, we only apply their methods on the result of feature extractor in Chapter III-A. From the results in Fig. 8, it can be clearly observed that the proposed SSAE model not only achieves a much higher AUC than other baselines when using a small amount of labeled data, but also guarantees a high recognition performance when the labeled data is gradually reduced, both for 100 Hz and 500 Hz data. More specifically, for 100 Hz data, the SSAE model can achieve an AUC value of over 0.9 using only 40% labeled data, and for 500Hz data, this number is 20%. The performance does not degrade significantly as the amount of data gradually reduces to 50%. For other methods, the AUC is reduced by 20% when the number of labeled data is reduced by half. In short, in terms of real applications, the approach adopted in this study can significantly reduce the number of experiments and the costs in time and economy when deployed.

*4) Complete Pipeline Performance:* A complete PSEW case is shown in Fig. 9, which was conducted on 26 November 2020 from 10.30 am to 11 am of pipe2. Continuous mechanical excavation operations were undertaken at 41 km and the model was effective in identification and localization, which can be clearly observed in Fig. 9. In addition, most of the other unknown events could be matched with the results. For example, a large number of vehicles driving past and a small number of excavations were identified in the range of 0–2 km, which is because the pipe passed the Qingxian County and the test was conducted on a weekday morning. As such, the information on vehicles and pedestrians was obvious. Besides, a small number of manual and mechanical excavations were identified at 30–31 km and it was confirmed on-site that there was a road running parallel to the pipe and a factory not far from the pipe whose vertical press machine was used during the test-time, resulting in a large signal being monitored. Moreover, the false alarms were higher at 25–26 km and 55–56 km due to the two SRAs being placed here, resulting in signal instability and even intermittent signal saturation. However, through actual deployment and heuristic debugging, most of these issues have been implemented. In addition, a full day identification results is shown in Fig. 10, which was tested on 25 November 2020 of pipe2. It can be observed that there were significantly more vehicle drivings from 6 am to 9 pm in the 0–2 km (i.e., Qingxian County) and 2–6 km (i.e., highway networks) than the other times of the day, which was consistent with people's activities on weekdays. Also, two experiments with mechanical excavation were carried out respectively at 36 km from 9 am to 11 am and at 57 km from 2 pm to 3 pm, which could be seen in Fig. 10.

*5) Deployment Performance:* The experiments of deployment performance include model latency and size. This was

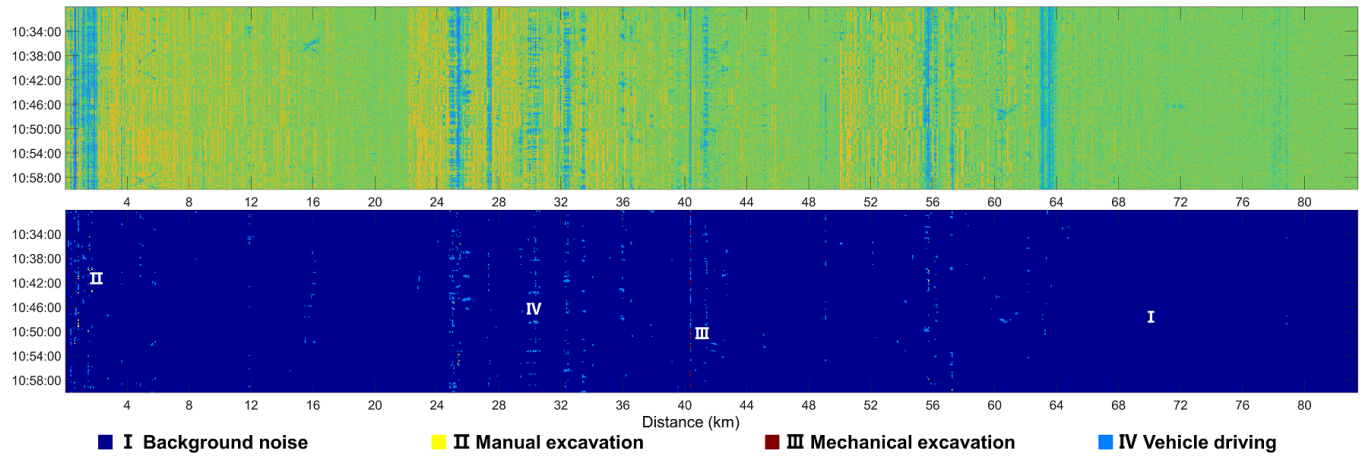



Fig. 9. Peak features and identification results from pipeline2 in the autumn of 2020. The upper portion shows the peak features and the lower portion represents the corresponding identification results. It can be observed that mechanical excavation occurred for 30 mins at around 41 km, which was accompanied by signals of vehicle drivings (i.e., at that moment the engine of excavator was opening and the position was adjusted to find a better angle for operation). Most of the events were background noises and vehicle drivings at several specific mileages.

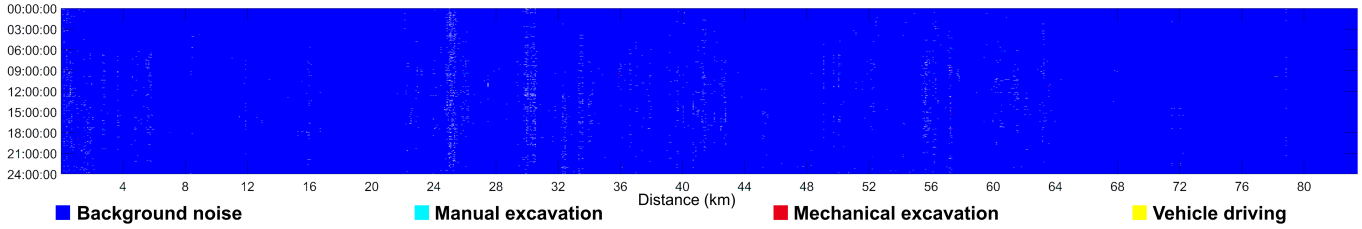



Fig. 10. Full day identification results from the pipeline2 in the autumn of 2020.

TABLE V
DEPLOYMENT PERFORMANCE TESTING

| Data (Hz) | latency (ms) | AE size (KB) | Classifier size (KB) |
|---|---|---|---|
| 100 | 0.68 | 678 | 345 |
| 500 | 1.73 | | |

repeated 10 times and the average is shown in Table V. In detail, the model latency is only approximately 1.73 ms per km for 500 Hz and 0.68 ms per km for 100 Hz with 20 m spatial resolution using GPU. The encoder-decoder size is 678 KB and more specially, the classifier size is only 345 KB allowing it to be deployed in most embedded systems and finetuned with a few labeled datasets. Further, by matrix optimization and heuristic threshold methods, the approach used in this study achieves a relatively good deployment performance and fully meets the industrial requirements.

## V. CONCLUSION

In this paper, a novel semi-supervised learning method is presented based on DOFS to monitor the safety of long-distance energy transportation pipelines in real-time (i.e., recognizing and locating the third-party events). The experiments from long-distance operational energy pipelines indicate that the proposed SSAE model can improve identification and location performance with a significant amount of unlabeled data and a small amount of labeled data with low SNR conditions, which could reduce the cost of data collection and system deployment. The encoding learned from unlabeled data has good spatiotemporal transferability, which can improve the portability of the PSEW system. Besides, the decoded feature shows good visualization and the model has relatively small size and latency. This work provides a new perspective for the practical application of PSEW systems in industrial scenarios.

In the future, few-shot learning is planned to be used and the whole model is expected to be implemented in languages, such C++ instead of TensorFlow's Python interface, which can further reduce the model latency by 15×. In addition, higher spatio-temporal resolutions can be explored by optimizing the sampling frequency and sensing mechanism of DOFS.

## REFERENCES

[1] J. Li, Y. Wang, P. Wang, Q. Bai, Y. Gao, H. Zhang, and B. Jin, "Pattern recognition for distributed optical fiber vibration sensing: A review," *IEEE Sensors Journal*, 2021.
[2] B. M. T. Fouda, B. Yang, D. Han, and B. An, "Pattern recognition of optical fiber vibration signal of the submarine cable for its safety," *IEEE Sensors Journal*, vol. 21, no. 5, pp. 6510–6519, 2021.
[3] K.-S. Lin, K.-H. Yeh, Y.-J. Chiang, and L. Wang, "Fiber-optic perimeter intrusion detection by employing a fiber laser cavity in each defensed zone," *IEEE Sensors Journal*, vol. 18, no. 20, pp. 8352–8360, 2018.
[4] M. A. Soto, J. A. Ramirez, and L. Thevenaz, "Intensifying the response of distributed optical fibre sensors using 2d and 3d image restoration," *Nature communications*, vol. 7, no. 1, pp. 1–11, 2016.
[5] F. Tanimola and D. Hill, "Distributed fibre optic sensors for pipeline protection," *Journal of Natural Gas Science and Engineering*, vol. 1, no. 4-5, pp. 134–143, 2009.
[6] B. M. Tabi Fouda, D. Han, B. An, X. Lu, and Q. Tian, "Events detection and recognition by the fiber vibration system based on power spectrum estimation," *Advances in Mechanical Engineering*, vol. 10, no. 11, p. 1687814018808679, 2018.

[7] Y. Zhang, L. Zhao, Q. Tian, and J. Fan, "Optical fiber intrusion signal recognition based on improved mel frequency cepstrum coefficient," in *2018 11th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*. IEEE, 2018, pp. 1–9.

[8] Y. Wang, P. Wang, K. Ding, H. Li, J. Zhang, X. Liu, Q. Bai, D. Wang, and B. Jin, "Pattern recognition using relevant vector machine in optical fiber vibration sensing system," *IEEE Access*, vol. 7, pp. 5886–5895, 2019.

[9] Q. Sun, H. Feng, X. Yan, and Z. Zeng, "Recognition of a phase-sensitivity otdr sensing system based on morphologic feature extraction," *Sensors*, vol. 15, no. 7, pp. 15 179–15 197, 2015.

[10] Y. Yang, Y. Li, T. Zhang, Y. Zhou, and H. Zhang, "Early safety warnings for long-distance pipelines: A distributed optical fiber sensor machine learning approach," in *The Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021*. AAAI Press, 2021, p. online.

[11] M. He, S. Wang, C. Gao, and L. Feng, "Research on recognition method of railway perimeter intrusions based on $\phi$-otdr optical fiber sensing technology," *IEEE Sensors Journal*, 2020.

[12] Z. Sheng, Z. Zeng, H. Qu, and Y. Zhang, "Optical fiber intrusion signal recognition method based on tsvd-scn," *Optical Fiber Technology*, vol. 48, pp. 270–277, 2019.

[13] C. Xu, J. Guan, M. Bao, J. Lu, and W. Ye, "Pattern recognition based on enhanced multifeature parameters for vibration events in $\varphi$-otdr distributed optical fiber sensing system," *Microwave and Optical Technology Letters*, vol. 59, no. 12, pp. 3134–3141, 2017.

[14] H. Wu, Y. Qian, W. Zhang, and C. Tang, "Feature extraction and identification in distributed optical-fiber vibration sensing system for oil pipeline safety monitoring," *Photonic Sensors*, vol. 7, no. 4, pp. 305–310, 2017.

[15] J. Tejedor, J. Macias-Guarasa, H. F. Martins, D. Piote, J. Pastor-Graells, S. Martin-Lopez, P. Corredera, and M. Gonzalez-Herraez, "A novel fiber optic based surveillance system for prevention of pipeline integrity threats," *Sensors*, vol. 17, no. 2, p. 355, 2017.

[16] P. Chen, C. You, and P. Ding, "Event classification using improved salp swarm algorithm based probabilistic neural network in fiber-optic perimeter intrusion detection system," *Optical Fiber Technology*, vol. 56, p. 102182, 2020.

[17] W. Kong, J. Yu, J. Yang, and T. Tian, "Model building and simulation for intelligent early warning of long-distance oil & gas storage and transportation pipelines based on the probabilistic neural network," in *IOP Conference Series: Earth and Environmental Science*, vol. 546, no. 2. IOP Publishing, 2020, p. 022009.

[18] X. Wang, Y. Liu, S. Liang, W. Zhang, and S. Lou, "Event identification based on random forest classifier for $\phi$-otdr fiber-optic distributed disturbance sensor," *Infrared Physics & Technology*, vol. 97, pp. 319–325, 2019.

[19] X. Huang, B. Wang, K. Liu, and T. Liu, "An event recognition scheme aiming to improve both accuracy and efficiency in optical fiber perimeter security system," *Journal of Lightwave Technology*, vol. 38, no. 20, pp. 5783–5790, 2020.

[20] M. Aktas, T. Akgun, M. U. Demircin, and D. Buyukaydin, "Deep learning based multi-threat classification for phase-otdr fiber optic distributed acoustic sensing applications," in *Fiber Optic Sensors and Applications XIV*, vol. 10208. International Society for Optics and Photonics, 2017, p. 102080G.

[21] C. Lyu, Z. Huo, X. Cheng, J. Jiang, A. Alimasi, and H. Liu, "Distributed optical fiber sensing intrusion pattern recognition based on gaf and cnn," *Journal of Lightwave Technology*, vol. 38, no. 15, pp. 4174–4182, 2020.

[22] Y. Shi, Y. Wang, L. Zhao, and Z. Fan, "An event recognition method for $\phi$-otdr sensing system based on deep learning," *Sensors*, vol. 19, no. 15, p. 3421, 2019.

[23] H. Wu, J. Chen, X. Liu, Y. Xiao, M. Wang, Y. Zheng, and Y. Rao, "One-dimensional cnn-based intelligent recognition of vibrations in pipeline monitoring with das," *Journal of Lightwave Technology*, vol. 37, no. 17, pp. 4359–4366, 2019.

[24] Y. Yang, J. Li, M. Tian, Y. Zhou, L. Dong, and J. x. He, "Signal analysis of distributed optic-fiber sensing used for oil and gas pipeline monitoring," in *Proceedings of the 2019 International Symposium on Signal Processing Systems*, 2019, pp. 21–25.

[25] Y. Yang, Y. Li, and H. Zhang, "Pipeline safety early warning method for distributed signal using bilinear cnn and lightgbm," in *2021 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2021*. IEEE, 2021, p. online.

[26] Y. Bai, J. Xing, F. Xie, S. Liu, and J. Li, "Detection and identification of external intrusion signals from 33 km optical fiber sensing system based on deep learning," *Optical Fiber Technology*, vol. 53, p. 102060, 2019.

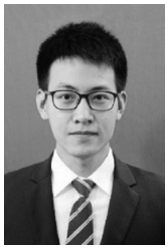
[27] Z. Li, J. Zhang, M. Wang, J. Chai, Y. Wu, and F. Peng, "An anti-noise $\phi$-otdr based distributed acoustic sensing system for high-speed railway intrusion detection," *Laser Physics*, vol. 30, no. 8, p. 085103, 2020.

[28] J. Tejedor, H. F. Martins, D. Piote, J. Macias-Guarasa, J. Pastor-Graells, S. Martin-Lopez, P. C. Guillén, F. De Smet, W. Postvoll, and M. González-Herráez, "Toward prevention of pipeline integrity threats using a smart fiber-optic surveillance system," *Journal of Lightwave Technology*, vol. 34, no. 19, pp. 4445–4453, 2016.

[29] G. Li, J. Zhu, R. Sun, X. Lin, W. Yang, K. Zeng, F. Wang, C. Wang, and B. Zhou, "Pipe line safety monitoring using distributed optical fiber vibration sensor in the china west-east gas pipeline project," in *2018 Asia Communications and Photonics Conference (ACP)*. IEEE, 2018, pp. 1–3.

[30] C. Lyu, Z. Huo, Y. Liu, X. Cheng, J. Jiang, A. A. J. Yang, and H. Su, "Robust intrusion events recognition methodology for distributed optical fiber sensing perimeter security system," *IEEE Transactions on Instrumentation and Measurement*, 2020.

[31] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," *arXiv preprint arXiv:1706.03762*, 2017.

**Yiyuan Yang** was born in Shijiazhuang, Hebei, China, in 1997. He received a B.E. degree at the School of Artificial Intelligence and Automation at Huazhong University of Science and Technology, Wuhan, Hubei, China, in 2019. He is currently pursuing a M.E. degree at the Department of Automation, Tsinghua University, Beijing, China.

His current research interests include time series, optical fiber early warning systems, sensor data mining, signal processing, machine learning, and deep learning.

**Haifeng Zhang** was born in Jiaohe, Jilin, China, in 1986. He received a B.E. degree from the School of Physics at Liaoning University, Shenyang, Liaoning, China, in 2009, and a Ph.D. degree from the China University of Petroleum, Beijing, China, in 2015.

He is currently an Assistant Researcher with the Research Institute of Tsinghua University, Pearl River Delta, Guangzhou, China. His current research interests include optical fiber early warning systems, gas and liquid multiphase flow measurement, and sensor data mining.

**Yi Li** received a B.E. degree from the University of Central Lancashire, Preston, U.K., in 2004, and a Ph.D. degree from the University of Manchester, Manchester, U.K., in 2008.

He is currently an Associate Professor at the Graduate School at Shenzhen, Tsinghua University, Beijing, China. His current research interests include electrical tomography, data fusion, gas and liquid multiphase flow measurement, and image reconstruction.