SCIENCE CHINA
Information Sciences

• **Supplementary File** •

Unleashing potentials with deep learning: Decoding the Complex Events for Distributed Fiber Optic Sensing Applications

Yujiao LI[1,2], Liqin HU[1,2] & Kuanglu YU[1,2*]

[1]*Institute of Information Science, School of Computer and Information Technology, Beijing Jiaotong University, Beijing* 100044, *China;*
[2]*Beijing Key Laboratory of Advanced Information Science and Network Technology, Beijing Jiaotong University, Beijing* 100044, *China*

Appendix A Φ-OTDR System

The constructed Φ-OTDR distributed optical fiber sensing system is shown in Figure 1. The signal light generated by NKT E15 laser is amplified through an Erbium Doped Fiber Amplifier (EDFA) and then passes through a filter. The filtered signal light is input into an Acousto-Optic Modulator (AOM) to be modulated into pulses with a width of 400 ns. Then the modulated light is injected through a circulator into a 5 or 10 km single-mode lead fiber (Corning sm28e+) in a soundproof box on the optical platform, with the fiber tail connected to a 100 m armored fiber. The first 50 meters are used for data collection. The reflection is reduced by using an anti-reflection connector (FC/APC) and wrapping the fiber into small radius rings. The sampling rate of data acquisition card is 10 MSamp/s. The system collects intensity signals.

A total of six events, including background noise, digging, knocking, watering, shaking, and walking, are collected and labeled as 0, 1, 2, 3, 4, and 5, respectively. A total of 15,148 samples are collected as shown in Table 1. The dataset is divided into a training set and a testing set with by 8:2, resulting in 12,334 samples in the training set and 3,084 samples in the testing set. For subsequent experiments, the collected data is concatenated and cropped. The data format for all events is unified to 12×10000, that is to say, each event is consisted of 12 adjacent spatial points in the spatial domain and 10000 points in the temporal domain. More details of the dataset can be found in Ref. [1].

Table A1 The numbers of six sample categories

Category	Label	No.of samples
Background noise	0	2946
Digging	1	2512
Knocking	2	2530
Watering	3	2253
Shaking	4	2728
Walking	5	2449
Total		15418

The original signals and the differentiated one-dimensional and two-dimensional signals, as well as the spatio-temporal of the six events are shown in Figure A1. To improve the clarity of visual observation, we performed interpolation in the spatial direction of the two-dimensional image, interpolating 12 points to 120 points. Please note that no interpolation operation was conducted in the subsequent data processing. The background noise observed is a stationary random noise without any intentional interference. Both the digging and knocking signals exhibit a rapid decay in amplitude after reaching their peak. However, the duration of the digging signal is relatively longer due to the backflow of sand, while the knocking signal has a shorter duration and decays more rapidly. The watering signal, being a continuous action, persists for a certain period of time. The variations in amplitude strength arise from the random impact of water droplets falling onto the optical fiber. The shaking signal demonstrates regular oscillations, reflecting the alternating back-and-forth vibration behavior. As for the walking signal, the pulse amplitudes align with the rhythm of human footsteps. In the traditional scheme, differentiation is carried out to better obtain the events' characteristics in the subsequent processing such as FFT, while in our scheme, the original signals are directly fed into the models for further processing with no differentiation, which can shorten the processing process and reduce human intervention.

* Corresponding author (email: klyu@bjtu.edu.cn)

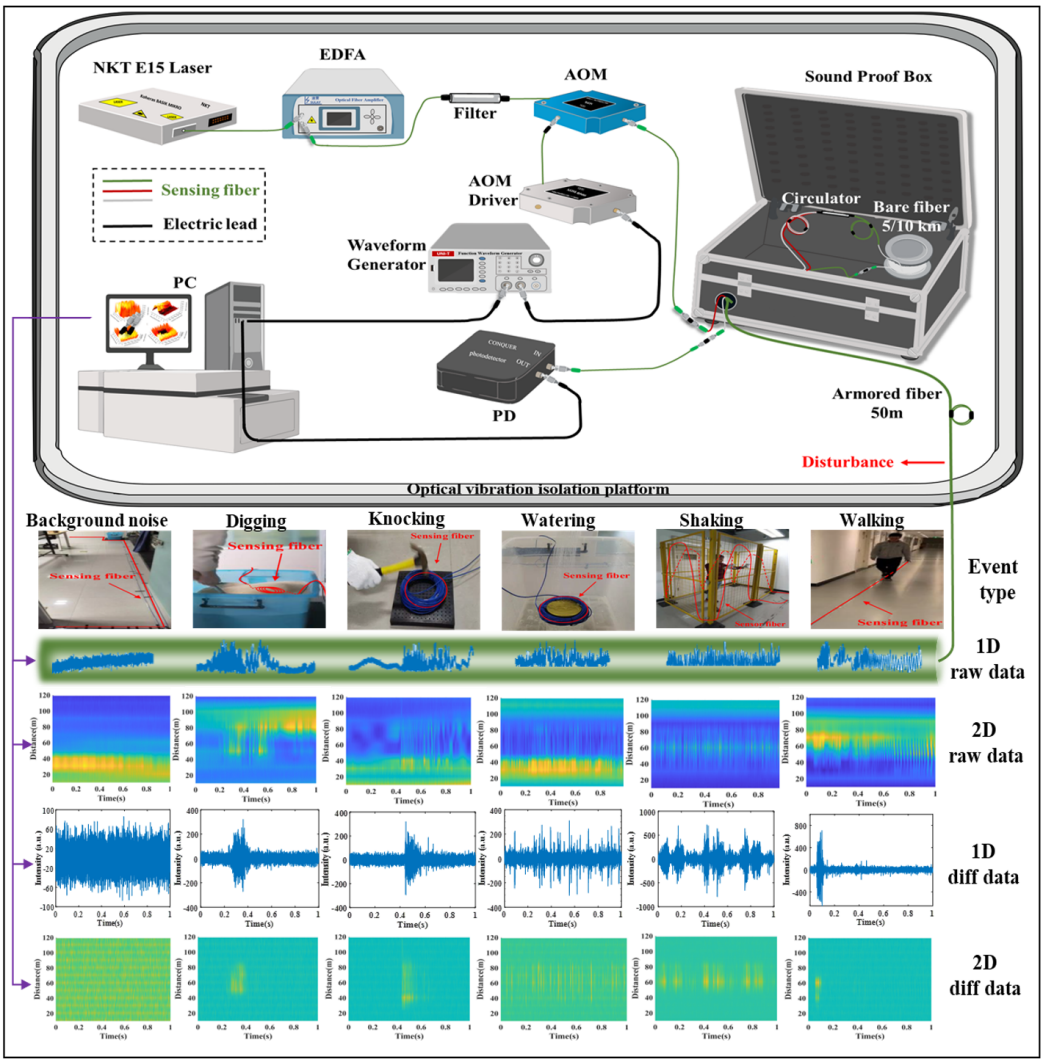

Figure A1 The Φ-OTDR system and collected event signals

Appendix B Methodology

Appendix B.1 Principle of semi-supervised algorithm

To effectively utilize the unlabeled data that is wasted in the traditional schemes, we investigate the three major semi-supervised schemes to improve the performance of Φ-OTDR disturbance classification networks. Semi-supervised learning (SSL) algorithms can be roughly categorized into three types [7] [8]. The first type is based on consistency regularization methods, such as Φ-Model [2], Temporal Ensemble [2], Mean Teacher [3], and UDA [4]. The second type is based on pseudo-labeling methods, also known as Self-training, which includes Pseudo-Labeling [5]. The last category combines the first two methods, such as Mixmatch [6], Fixmatch [7], and FreeMatch [8]. The central idea of the third is that the model should produce similar predictions or identical pseudo-labels for the same unlabeled data with different perturbations added. To comprehensively study SSL algorithms, we select four typical frameworks from all three categories, i.e., Self-training (ST) method based on pseudo-label, MeanTeacher (MT) method based on consistent regularization, FixMatch (XM) and FreeMatch (EM) based on holistic method. Their underlying principles are outlined as follows.

The ACNN-SA-BiLSTM (ACAB) network as shown in Figure B1(a). These models are denoted as ST-ACAB, MT-ACAB, XM-ACAB, and EM-ACAB, as shown in Figure B1(b)-B1(e) respectively.

Self-Training (ST). Self-Training is a semi-supervised method based on pseudo-labeling, and we primarily adopt the Pseudo-Label concept proposed by Lee [5]. Initially, the model is trained on labeled data, and subsequently, the trained model is employed to predict labels for unlabeled data, thereby generating pseudo-labels. Then the labeled data and the pseudo-labeled data are combined to form a new training dataset. The loss of the pseudo-labeling method is:

$$L = L_S + \alpha(t)L_{US} = \frac{1}{n}\sum_{m=1}^{n}\sum_{i=1}^{C} L(y_i^m, f_i^m) + \alpha(t)\frac{1}{n'}\sum_{m=1}^{n'}\sum_{i=1}^{C} L(y_i'^m, f_i'^m), \tag{B1}$$

where n is the number of mini-batch in labeled data, n' is the number of mini-batch in unlabeled data, f_i^m is the output units of m's sample for labeled data, y_i^m is the label of the labeled data, $f_i'^m$ is the output units for unlabeled data, $y_i'^m$ is the pseudo-label

of the unlabeled data. $\alpha(t)$ is a balance coefficient, which is very important for network performance. The model uses annealing algorithm to calculate $\alpha(t)$.

The auxiliary role of pseudo-labels in semi-supervised learning is primarily manifested in introducing additional supervisory signals and leveraging unlabeled data. While pseudo-labels themselves do not directly augment the dataset, the generation of approximate labels for unlabeled data allows incorporating these data along with labeled ones for model training. Thus, pseudo-labels provide additional information to the model, aiding in a better understanding of the data distribution and features, leading to performance enhancement. However, caution is needed when applying pseudo-labels, and measures such as threshold setting and confidence filtering can be employed to ensure their reliability.

Mean Teacher (MT). Mean Teacher is a semi-supervised method based on consistency regularization. It operates under the assumption that the input samples with introduced perturbations should yield similar predictions from the model. Introducing noise parameters contributes to enhancing the robustness of model. The MT framework consists of a teacher network and a student network. The network weights of the teacher model are computed using exponential moving average (EMA) of the student model. The network weights of the student model are updated through gradient descent based on the loss function. The overall loss function comprises an unsupervised consistency loss and a supervised cross entropy loss [3]. By adjusting the scaling coefficient β to balance the importance of both components, the model can effectively leverage both labeled and unlabeled data, thereby improving generalization performance.

$$L = L_S + \beta L_{US} = \sum_{x_i^l \in X_L} CE(f(x_i^l, \theta, \eta), y_i^l) + \beta \sum_{x_i^u \in X_U} MSE(f(x_i^u, \theta, \eta), f(x_i^u, \theta^{'}, \eta^{'})), \tag{B2}$$

where $X_L = \{(x_i^l, y_i^l) : i \in (1, \ldots, N)\}$ is the labeled dataset, while $X_U = \{x_i^u : i \in (1, \ldots, M)\}$ is the unlabeled dataset. $CE(\cdot)$ is cross entropy loss, $MSE(\cdot)$ stands for mean square error loss. β is the scaling coefficient. θ and $\theta^{'}$ are the network parameters of the student model and the teacher model respectively. η and $\eta^{'}$ are the noise parameters added to the student model and the teacher model respectively.

FixMatch (XM) FixMatch is an integrated approach that combines consistency regularization and pseudo-labeling. In this approach, a supervised model is trained using cross-entropy loss on labeled samples. For unlabeled samples, two versions of data are created using both strong and weak data augmentation techniques. Predictions are generated from the weakly augmented version, and a threshold is applied to select high-confidence predictions, which are then used as pseudo-labels. Following this, the model is trained using the strongly augmented data and their corresponding pseudo-labels, optimizing the model through cross-entropy loss. The loss of Φ-OTDR data is calculated as follows [7]:

$$L = L_S + \lambda L_{US} = \frac{1}{B} \sum_{b=1}^{B} CE(y_b, p_\theta(y|\omega(x_b))) + \lambda \frac{1}{\mu B} \sum_{b=1}^{\mu B} \mathbf{1}(max(q_b) > \tau) \cdot CE(\hat{q}_b, p_\theta(y|\boldsymbol{A}(u_b))), \tag{B3}$$

where x_b is the labeled data, y_b represents the corresponding label, and u_b is the unlabeled data. B is the batch size of the labeled data, and μB is the batch size of the unlabeled data. $\omega(\cdot)$ represents the weak augmentation applied to the Φ-OTDR data, and $p_\theta(y|\omega(x_b))$ represents the prediction distribution for the weakly-augmented version of labeled samples. $p_\theta(y|\boldsymbol{A}(u_b))$ represents the prediction distribution for the strongly-augmented version of unlabeled samples. $\boldsymbol{A}(\cdot)$ represents the strong augmentation on Φ-OTDR data. $q_b = p_\theta(y|\omega(\mu_b))$ represents the prediction distribution of the model for the weakly-augmented version of unlabeled samples. $\hat{q}_b = argmax(q_b)$ is the pseudo-label generated by the weakly-augmented Φ-OTDR data, and τ represents the threshold. λ_u is the weight of unsupervised loss.

FreeMatch (EM) The global threshold τ_t is:

$$\tau_t = \begin{cases} \frac{1}{C} & if \quad t = 0 \\ \lambda \tau_{t-1} + (1-\lambda) \frac{1}{\mu B} \sum_{b=1}^{\mu B} max(q_b) & otherwise \end{cases}, \tag{B4}$$

where $\lambda \in (0, 1)$ is the momentum decay of EMA. We initialize τ_t as $\frac{1}{C}$, where C represents the number of classes. We calculate the model's expectations for each class c prediction as follows:

$$\widetilde{p}_t(c) = \begin{cases} \frac{1}{C} & if \quad t = 0 \\ \lambda \widetilde{p}_{t-1}(c) + (1-\lambda) \frac{1}{\mu B} \sum_{b=1}^{\mu B} q_b(c) & otherwise \end{cases}, \tag{B5}$$

where $\widetilde{p}_t = [\widetilde{p}_t(1), \widetilde{p}_t(2), ..., \widetilde{p}_t(C)]$ is the list containing all $\widetilde{p}_t(c)$. Based on the global threshold and local threshold, the self-adaptive threshold is finally calculated as follows:

$$\tau_t(c) = MaxNorm(\widetilde{p}_t(c)) \cdot \tau_t = \frac{\widetilde{p}_t(c)}{max\{\widetilde{p}_t(c) : c \in [C]\}} \cdot \tau_t, \tag{B6}$$

where $MaxNorm$ is the Maximum Normalization ($x^{'} = \frac{x}{max(x)}$).

The unsupervised loss of Φ-OTDR data using adaptive thresholds is corrected as follows [8]:

$$L_{US} = \frac{1}{\mu B} \sum_{b=1}^{\mu B} \mathbf{1}(max(q_b) > \tau_t(\hat{q}_b)) \cdot CE(\hat{q}_b, A_b), \tag{B7}$$

where $\tau_t(n)$ is the adaptive threshold, and the overall loss [8] L is:

$$L = L_S + \lambda L_{US} + \beta L_E, \tag{B8}$$

where λ_u and λ_e represent the weight of unlabeled loss and SAF loss. L_S and L_{US} represent the loss of labeled data and unlabeled data, respectively. L_E represents SAF regularization loss, which normalizes the expectation of probabilities based on the histogram distribution of pseudo-labels. It helps the model produce diverse predictions especially for barely supervised settings.

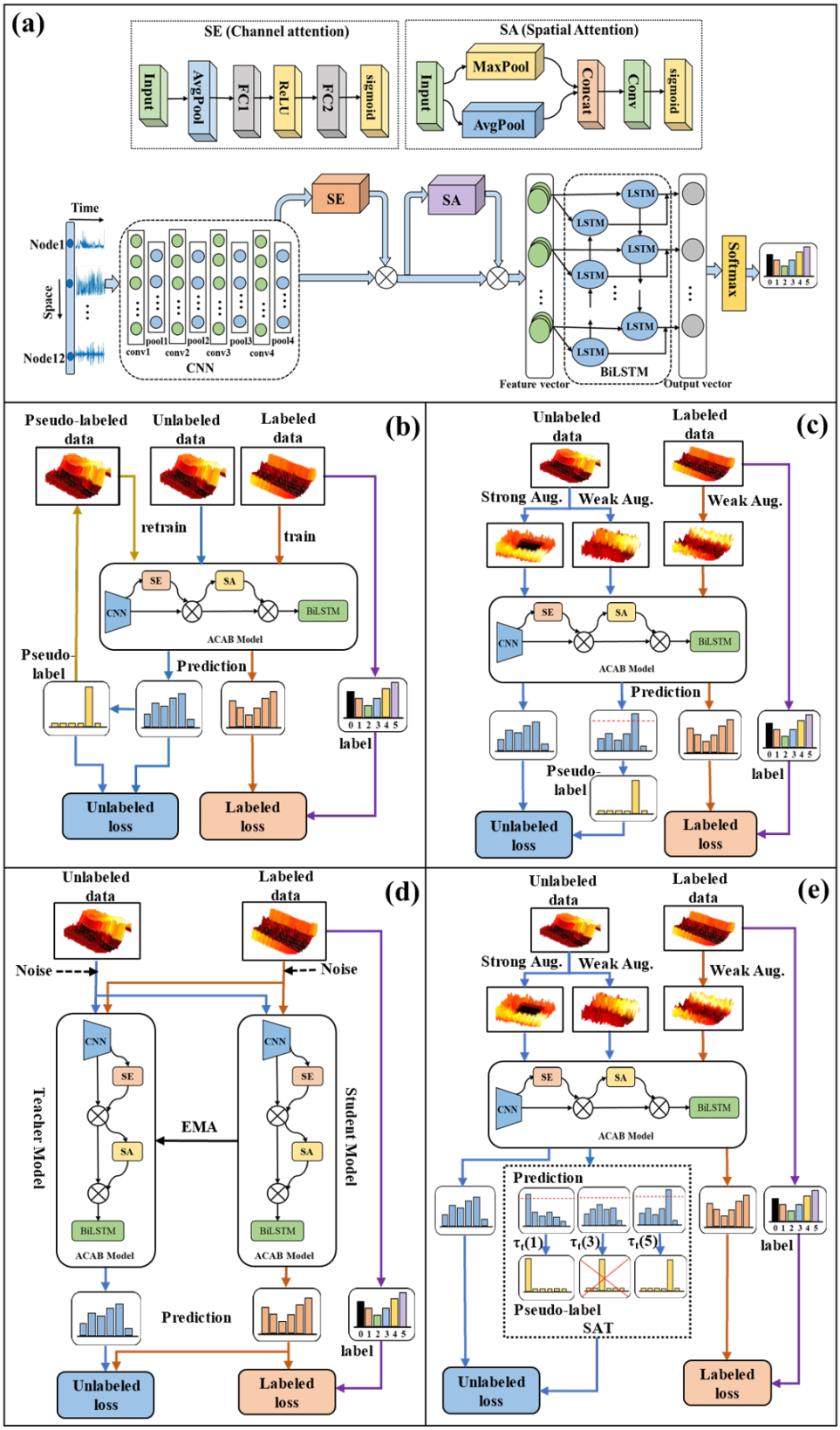

Figure B1 Structure of semi-supervision-ACAB model: (a) ACAB network. (b) ST-ACAB model. (c) XM-ACAB model. (d) MT-ACAB model. (e) EX-ACAB model. (Filtered (masked) samples are marked with red X.)

Appendix B.2 Data Aaugmentation

The data augmentation (DA) techniques employed in XM and EM can assist the model in harnessing information from unlabeled data, thereby enhancing its generalization capability. XM and EM rely on the employment of both weak and strong augmentations. By training the model to disregard the impact of these distortions, it substantially enhances its robustness. Additionally, strong augmentation also aids in enhancing the model's understanding of the task.

Table B1 Strong and weak augmentations of Φ-OTDR data

Weak Augmentation	Strong Augmentation
Gaussian noise	Gaussian noise
	Flip in spatial direction
	Rectangular cutout

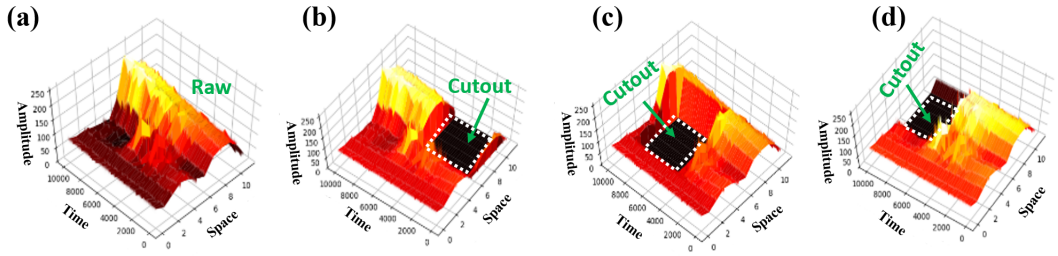

Figure B2 Cutout at random positions on Φ-OTDR samples

The two types of augmentation methods of Φ-OTDR data in our experiment are listed in Table B1. Gaussian noise with a mean of 0 and a variance of 1 is added to the raw sample as a weak augmentation. The strong augmentation involves a combination of three techniques, including Gaussian noise, flipping, and Cutout. On one hand, the signal at each sensing point along the fiber is temporal in nature, thus it is unidirectional and causal correlated; on the other, disturbances impact on both the forward propagating and backscattered light along the fiber, imparting an inherent bidirectional spatial correlation to the signal. Given the above spatiotemporal attributes of the Φ-OTDR signals, the flipping operation is performed only along the spatial direction. Cutout [10] is to randomly mask out specific areas and to set them to zero. It is employed to remove contiguous regions of input images during the training process and to compel the model to rely on other non-occluded regions for learning. As a result, the model becomes more robust and can better generalize local features by enhancing the diversity and quantity of the training data. Although the cutout region in image processing is usually square, taking into consideration the narrow rectangular (12*10,000) size of our Φ-OTDR samples, we need to select a more suitable Cutout region. The size selection of the Cutout region is detailed in the next section. The Cutout operation is applied to the original Φ-OTDR data, resulting in randomly positioned regions as shown in Figure B2.

Appendix C Experiment and results

Appendix C.1 Model hyperparameter settings

The hyperparameters used by EM algorithm in the experiment are shown in Table C1. The specific settings of each main network can be referred to Ref. [9]. The total step size K of the model training is set to 220, and the exponential moving average (EMA) decay is set to 0.999. The SGD optimizer is used for training. The learning rate adopts the cosine attenuation method. The unlabeled loss weight (λ_u) of all experiments is set to 1. The loss weight of Self-Adaptive Fairness Regularization (SAF) (λ_e) is set to 0.01, except that the ACNN-SA-BiLSTM and CNN-BiLSTM models are set to 0.005.

Appendix C.2 Cutout parameter settings

For strong augmentation, gaussian noise, flipping along the spatial domain, and Cutout are jointly applied. The choice of the Cutout size can influence the representation of key information pertaining to various Φ-OTDR events, potentially impacting its classification effectiveness. Therefore, a thorough analysis of the characteristics of Φ-OTDR data is necessary to inform the selection of appropriate Cutout hyperparameters.

Traditional Cutout techniques in computer vision are typically applied to image processing, where a majority of images are square in shape. As a result, the Cutout size is commonly chosen to be square as well. We initially employed the conventional square Cutout approach to test Φ-OTDR signals. Taking the case of 30 labels (5 per class) as an example, as shown in Figure C1(a), the results have revealed an improvement in accuracy compared with the scenario without Cutout. Taking into account of our Φ-OTDR signal's long strip shape of 12 × 10000, smaller square Cutout sizes might have a limited impact on the Φ-OTDR signals. Therefore, we configured and tested the Cutout shape as a narrow rectangle based on the form of the Φ-OTDR data. Again, with 30 labeled samples, we conducted experiments using five different shapes, as shown in Figure C1(b). It is evident that a strip shape yields more significant improvements compared to square shapes. The highest classification accuracy is achieved when the padding area dimensions are 6×5000, resulting in a remarkable 19% enhancement. Consequently, for the subsequent experiments in this paper, the Cutout size in strong augmentation is set at 6×5000.

Table C1 The hyperparameter setting of EM model

	ACNN-SA-BiLSTM	CNN-BiLSTM	2D-CNN	ATCN-SA-BiLSTM	MS 1D-CNN
Unlabeled data to labeled data ratio (μ)	1	1	2	1	1
Learning rate (l_r)	0.01	0.01	0.001	0.01	0.01
Batch size (B)	4	4	16	2	8
SGD momentum (β)	0.9	0.9	0.5	0.5	0.5
Weight decay	1e-5	1e-5	1e-3	1e-3	1e-2
Loss weight (λ_e)	0.005	0.005	0.01	0.01	0.01
Unlabeled loss weight (λ_u)			1		
EMA decay			0.999		
Total training step (K)			2^{20}		

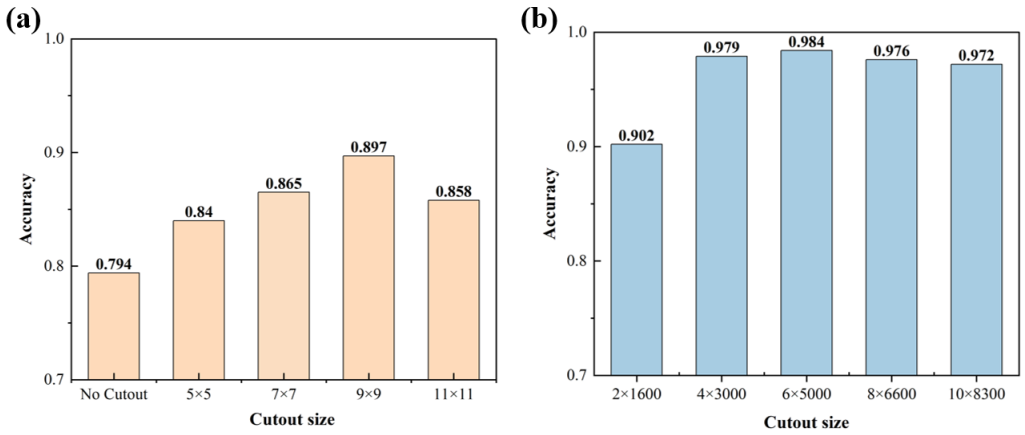

Figure C1 Accuracy of Φ-OTDR signal with different Cutout sizes: (a) square, (b) rectangle

Appendix C.3 Experimental results analysis

To better evaluate the classification performance of Φ-OTDR disturbance signals, we select several commonly used evaluation metrics in this field, including the confusion matrix (C), average accuracy (Acc), nuisance alarm rate (NAR), false negative rate (FNR), precision, recall, and F1 score. The confusion matrix are definded as $C = [d_{ij}],\ i \in [0, 5],\ j \in [0, 5]$, with the rows of the matrix representing the true values, and the columns of the matrix representing the predicted values. The definitions for the other metrics can be expressed as follows:

$$Acc = \frac{\sum_{i=0}^{5} d_{ij}}{\sum_{i=0}^{5}\sum_{j=0}^{5} d_{ij}}, \tag{C1}$$

$$NAR = \frac{\sum_{j=1}^{5} d_{0j}}{\sum_{i=0}^{5}\sum_{j=1}^{5} d_{ij}}, \tag{C2}$$

$$FNR = \frac{\sum_{i=1}^{5} d_{i0}}{\sum_{i=1}^{5}\sum_{j=0}^{5} d_{ij}}, \tag{C3}$$

Experiment 1: Performance comparison of 20 models under 4 major SSL frameworks

In order to fully compare the classification performance of four different semi-supervised frameworks for Φ-OTDR disturbance events, five different primary networks are selected for the four SSL frameworks respectively. The five main networks are ACAB, CNN-BiLSTM (CB), ATCN-SA-BiLSTM (ATAB), 2D-CNN (2DCNN) and MS-1DCNN (MSCNN). In this experiment, the proportion of labeled samples is set to be 10% of the total number of samples during training, that is to say, 1230 labeled samples and 11104 unlabeled samples are used for training. This experiment primarily verifies and compares the enhancement effect of each SSL framework on the classification outcomes of different models.

When a total of 1230 labels are available, the comparison results for a total of the twenty SSL models are presented in Table C2. It is apparent that employing the ACAB network within all four semi-supervised frameworks consistently results in the highest classification accuracy, with accuracy rates exceeding 96%. Therefore, for efficiency considerations, ACAB network is selected as

Table C2 Comparison of the results of five networks under four semi-supervised frameworks*

Framework	Main network	Accuracy	Precision	Recall	F1	NAR	FNR
Self-training (ST)	ACNN-SA-BiLSTM	0.960	0.960	0.957	0.958	0.000	0.009
	CNN-BiLSTM	0.951	0.950	0.949	0.949	0.000	0.007
	2D-CNN	0.920	0.923	0.919	0.917	0.009	0.009
	ATCN-SA-BiLSTM	0.798	0.800	0.794	0.794	0.044	0.065
	MS 1D-CNN	0.715	0.717	0.712	0.707	0.015	0.018
Mean teacher (MT)	ACNN-SA-BiLSTM	0.969	0.969	0.968	0.968	0.000	0.007
	CNN-BiLSTM	0.961	0.962	0.955	0.960	0.001	0.008
	2D-CNN	0.941	0.942	0.940	0.940	0.010	0.010
	ATCN-SA-BiLSTM	0.839	0.833	0.834	0.833	0.007	0.020
	MS 1D-CNN	0.868	0.868	0.866	0.865	0.012	0.020
FixMatch (XM)	ACNN-SA-BiLSTM	**0.999**	**0.999**	**0.999**	**0.999**	**0.000**	**0.000**
	CNN-BiLSTM	**0.999**	**0.999**	**0.999**	**0.999**	**0.000**	0.001
	2D-CNN	0.968	0.967	0.966	0.966	0.002	0.002
	ATCN-SA-BiLSTM	0.878	0.875	0.874	0.863	0.015	0.019
	MS 1D-CNN	0.960	0.962	0.958	0.959	0.005	0.009
FreeMatch (EM)	ACNN-SA-BiLSTM	**0.999**	**0.999**	**0.999**	**0.999**	**0.000**	**0.000**
	CNN-BiLSTM	**0.999**	**0.999**	**0.999**	**0.999**	**0.000**	**0.000**
	2D-CNN	0.967	0.966	0.966	0.966	0.004	0.001
	ATCN-SA-BiLSTM	0.853	0.848	0.846	0.846	0.015	0.029
	MS 1D-CNN	0.961	0.961	0.958	0.959	0.002	0.009

*:1230 labeled samples

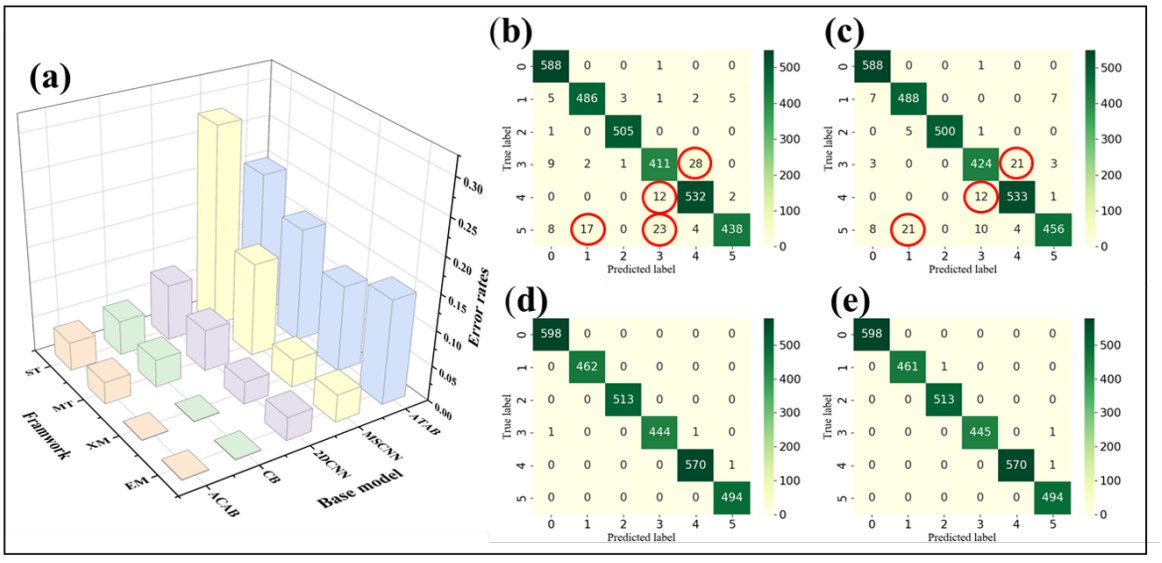

Figure C2 Classification performance of semi-supervised models: (a) Classification error rate comparison of 20 semi-supervised models. Confusion matrices of (b) ST-ACAB, (c) MT-ACAB, (d) XM-ACAB, and (e) EM-ACAB

the primary network for all subsequent experiments in this paper. Furthermore, it can be observed that the adoption of XM and FM frameworks leads to a further improvement in model classification accuracy, with accuracy rates reaching 99.9%. Notably, when the FM framework is used, both NAR and FNR are reduced to 0.

To provide a visual representation of the classification performance of the 20 models, Figure C1(a) illustrates their classification error rates ($Errorrate = 1 - Accuracy$). It is obvious that XM-ACAB, XM-CB, EM-ACAB, and EM-CB exhibit the lowest classification error rates, indicating higher recognition accuracies. In order to more intuitively evaluate the classification performance

of the model for each type of events, the confusion matrices of the four semi-supervised frameworks, ST-ACAB, MT-ACAB, XM-ACAB, EM-ACAB, are presented in Figure C2(b)-C2(e). In the confusion matrices, green cells along the diagonal represent the number of correctly classified instances, while non-diagonal cells show those misclassified. To make the results more obvious, the latter ones with a count > 10 are highlighted with red circles. Additionally, a comparison is conducted on the Precision, Recall, and F1-score performance metrics of these four semi-supervised models in Table C3. It can be found that both MT and ST demonstrate relatively low recognition rates for walking events. We also noted that, when ACAB, 2DCNN, and MSCNN serve as the primary networks, the recognition rates for walking events are also notably low under the XM and EM frameworks. This highlights the complexity of the walking events, as they encompass a variety of walking patterns in our experiments. These patterns include events with different walking and running frequencies, as well as events involving single-foot and double-foot step-ons. Given the intricate nature of these behaviors, the models struggle to accurately learn features with limited labeled samples. Nevertheless, the XM-ACAB and EM-ACAB models remarkably excel in recognizing walking events, boasting a Recall of 100% for this event, which shows that events belonging to walking are rarely misclassified to other categories.

Table C3 Performance Comparison of each event type of ACAB under four SSL frameworks

Event type	FreeMatch(EM)			FixMatch(XM)			Mean Teacher(MT)			Self-training(ST)		
	Precision	Recall	F1	Precision	Recall	F1	Precision	Recall	F1	Precision	Recall	F1
0	**1.000**	**1.000**	**1.000**	0.998	**1.000**	0.999	0.970	0.998	0.984	0.962	0.998	0.980
1	**1.000**	0.998	0.999	**1.000**	**1.000**	**1.000**	0.966	0.972	0.969	0.962	0.968	0.965
2	0.998	**1.000**	0.999	**1.000**	**1.000**	**1.000**	**1.000**	0.988	0.994	0.992	0.998	0.995
3	**1.000**	**0.998**	**0.999**	**1.000**	0.996	0.998	0.946	0.940	0.943	0.917	0.911	0.914
4	**1.000**	**0.998**	**0.999**	0.998	**0.998**	0.998	0.955	0.976	0.966	0.940	0.974	0.957
5	0.996	**1.000**	0.998	**0.998**	**1.000**	**0.999**	0.976	0.931	0.953	0.984	0.894	0.937

Experiment 2: Performance comparison between supervised model and SSL models under different number of labeled samples

To further compare the performance of different semi-supervised methods with limited labeled data, we conduct a comparative experiment by varying the number of labeled samples for supervised baseline model and four semi-supervised models. We set the total number of labeled samples to be 1230, 600, 150, 30, and 12, respectively. The ACAB network performs the best when combined with all semi-supervised frameworks under different sample numbers. Therefore, in this experiment, the primary network for all models is ACAB. The primary objective of this experiment is to simulate the practical field scenario where labeled samples are scarce, but unlabeled samples are abundant, so that the effectiveness of the EM framework with limited labeled samples can be validated.

Table C4 Accuracy for different number of labeled training samples

Number of labels	12334	1230	600	150	30	12
Supervised Model(SM)	99.5%	96.7%	94.6%	89.3%	47.6%	37.9%
Self-training(ST)	-	96.0%	93.6%	40.1%	27.9%	23.6%
Mean Teacher(MT)	-	96.9%	95.2%	89.6%	50.3%	34.8%
FixMatch(XM)	-	**99.9%**	**99.8%**	99.6%	97.1%	91.0%
FreeMatch(EM)	-	**99.9%**	**99.8%**	**99.7%**	**98.5%**	**92.8%**

The comparison of accuracy for the five models is presented in Table C4. It can be observed that with a higher number of labeled samples, such as 1230, both supervised and semi-supervised models achieve an accuracy of over 96%, indicating good recognition performance. Notably, under the XM and EM frameworks, the classification accuracy reaches an impressive 99.9%. With a higher number of labeled samples, models exhibit better generalization and more stable classification outcomes. However, when the number of labeled samples is limited, due to the model's high reliance on the features of the limited labeled data, the model, with a decline in its generalization ability, witnesses a decrease in its recognition performance. To mitigate the impact of randomness when using a small number of labeled samples for training, we conduct three repeated experiments for each model when the number of labeled samples are only 30 and 12, the average accuracy of which is taken as the final accuracy. When the number of labeled samples decreases to only 12 in total, EM still achieves a high average accuracy of 92.8% for the six types of events. The mean and standard deviation of Accuracy and FNR for 30 and 12 labeled samples are shown in Figure C3. EM model demonstrates the highest Accuracy and the lowest FNR, indicating its strong stability and robust generalization ability. Additionally, a further comparison of performance metrics for different models with only 12 labeled samples is illustrated in Figure 1(e) in letter. It can be observed that the EM semi-supervised framework achieves the best results in terms of all Accuracy, Precision, Recall, F1-score, and FNR. Though the NAR metric is slightly inferior to the XM framework, the difference is tiny. With a limited number of labeled samples, the EM semi-supervised framework exhibits a distinct advantage. Notably, with only 25 labeled samples per category (a total of 150), its performance surpasses that of the traditional fully supervised model trained with all 12334 labeled samples.

Experiment 3: Ablation experiments under the EM framework

Due to the superior classification performance of the EM semi-supervised framework, we further conduct ablation experiments to explore and to validate the effectiveness of different components within the EM algorithm. SAT is a threshold-adjusting scheme,

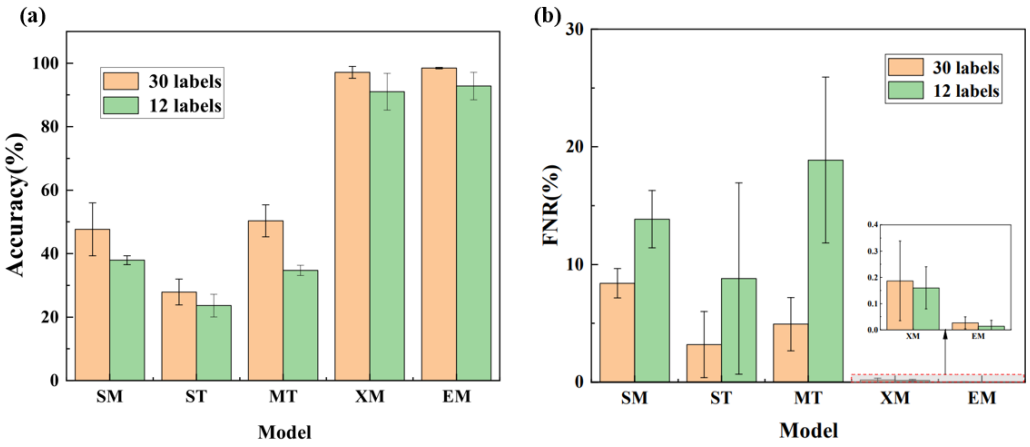

Figure C3 (a) Accuracy and (b) FNR of SSL models with 12 and 30 labeled samples

and SAF encourages diverse predictions. SAF encourages the expectation of the output probability for each mini-batch to be close to a marginal class distribution of the model, after normalized by histogram distribution [8]. We first compare the performances of the EM framework in the conditions with and without SAF. With only 12 labeled samples, we further compare the results of using the XM algorithm with only using Self-Adaptive Thresholding (SAT) threshold and those of using both SAT and SAF methods. As an illustrative example, we select the best-performing set of experiments out of three repetitions for analysis.

Table C5 Performance comparison of ablation experiments under the EM framework

Algorithm	Accuracy	Precision	Recall	F1	NAR	FNR
FixMatch(FT:0.95)	0.951	0.959	0.951	0.949	0.0012	0.0020
FreeMatch(SAT)	0.951	0.960	0.950	0.949	**0.0008**	0.0020
FreeMatch(SAT+SAF)	**0.960**	**0.961**	**0.959**	**0.958**	0.0040	**0.0004**

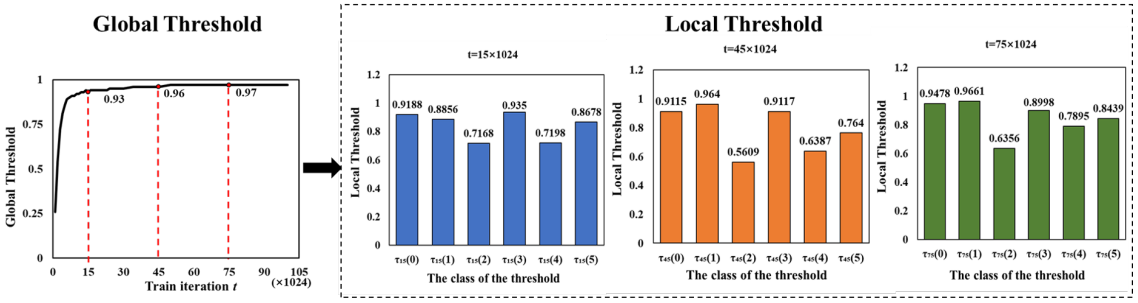

Figure C4 The threshold change of ACAB with EM framework with 12 labeled samples

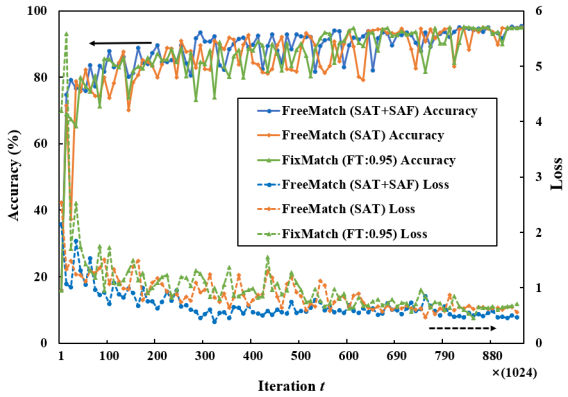

Figure C5 Accuracy and loss curves of ACAB with Semi-supervised framework

During the training process of EM, both the global threshold and the local threshold are shown in Figure C4. On the basis of XM, replacing the fixed threshold of 0.95 with SAT can reduce the NAR of the model for the classification. At the start of training, the threshold is set relatively low, allowing more potentially correct samples to be included in the training process, thus enhancing the utilization of unlabeled data. With the improving of the model's training accuracy, the threshold gradually increases adaptively, filtering out samples that might be incorrect to reduce confirmation bias. On this basis, the incorporation of SAF enables the model to learn more effectively and diversify its understanding, ultimately contributing to the enhancement of classification performance. When there are only 12 labels, the accuracy and loss curves for the three experiments are depicted in Figure C5. The experiments demonstrate that the incorporation of SAF aids the model in generating diverse predictions, especially when the label quantities are extremely limited, as the model exhibits faster convergence and improved generalization.

In practical real-time applications of Φ-OTDR, manually labeling a substantial amount of collected data proves to be challenging. Considering the comprehensive results from the three sets of experiments, it is evident that by utilizing the proposed EM-ACAB framework in this paper, more rapid and accurate classification can be achieved even with an extremely limited labeled samples in the above scenarios. Additionally, the SSL framework exhibits stable performance and strong generalization capabilities. Moreover, the other SSL methods proposed in this paper could also be further explored in future research.

References

1 Cao X, Su Y S, Jin Z Y, et al. An open dataset of Φ-OTDR events with two classification models as baselines. Res Opt, 2023, 10: 100372

2 Laine S, Aila T. Temporal ensembling for semi-supervised learning. In: Proceedings of the International Conference on Learning Representation, Toulon, 2017. 4: 6

3 Tarvainen A, alpola H. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. In: Proceedings of the International Conference on Advances in neural information processing systems, Long Beach, 2017. 1195–1204

4 Xie Q z, Dai Z H, Hovy E, et al. Unsupervised data augmentation for consistency training. In: Proceedings of the International Conference on Advances in Neural Information Processing Systems, online. 2020. 33

5 Lee D H. Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In: Proceedings of Workshop on challenges in representation learning, Atlanta, 2013. 3: 896

6 Berthelot D, Carlini N, Goodfellow I, et al. Mixmatch: A holistic approach to semi-supervised learning. In: Proceedings of the International Conference on Advances in neural information processing systems, Vancouver, 2019. 32

7 Sohn K, Berthelot D, Carlini N, et al. Fixmatch: Simplifying semi-supervised learning with consistency and confidence. In: Proceedings of the International Conference on Advances in neural information processing systems, online, 2020. 33: 596-608

8 Wang Y D, Chen H, Heng Q, et al. Freematch: self-adaptive thresholding for semi-supervised learning. In: Proceedings of the International Conference on Learning Representation, Kigali, 2023

9 Li Y J, Cao X M, Yu K L, et al. A deep learning model enabled multi-event recognition for distributed optical fiber sensing, Sci China Inf Sci, 2024, 67: 132404

10 DeVries T, Taylor G W. Improved regularization of convolutional neural networks with cutout. 2017. ArXiv:1708.04552