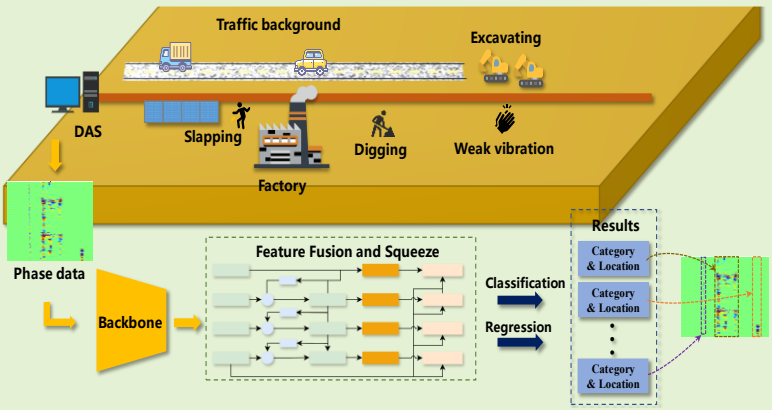
LR-Net for Weak Vibration Event Location and Recognition with DAS

Sen Yan, Ningmu Zou, Ying Shang, Bangwei Liu, Xiao Zhou, Imtiaz Naseeb Awan,

Yixin Zhang, Xuping Zhang**

***Abstract*—Distributed acoustic sensing (DAS) has been deployed across various large-scale infrastructures for safety monitoring and health maintenance operations. Among these applications, the localization and recognition of vibrations are common and critical post-processing tasks. Currently, multi-event vibration localization and recognition present a significant challenge for post-processing algorithms. Moreover, weak vibrations, which are marked by short durations and limited propagation ranges, further exacerbate the difficulty of the accurate location and recognition. Consequently, these challenges contribute to a high false alarm rate and missed detection rate across the DAS system. To address those challenges, this paper proposes a location and recognition convolutional neural network (LR-Net) that can achieves end-to-end and multi-event recognition and localization along the fiber within a single sample. In the model, we propose the location-attention-mechanism feature fusion and squeeze framework (LAMFS) and dynamic matching strategy (DMS) to enable the model to focus on weak vibration and enhance its fitting ability. In the field experiments conducted in three typical scenarios, LR-Net achieves a 99.1% mAP for six types of events with merely average location error of ±1.5m. Moreover, the Nuisance Alarm Rate (NAR) and the Missing Alarm Rate (MAR) were only 1.4% and 1.06% respectively. These results demonstrate superior performance compared with other deep learning models. Above all, the proposed algorithm possesses significant practical value and can be adapted to other scenarios such as pipeline leak detection, perimeter security, and protection of important facilities.**



***Index Terms*—Distributed acoustic sensing, protection of important facilities, location and recognition of weak vibration**

I. Introduction

THE Distributed acoustic sensing (DAS) has been employed to construct the smart city by measuring the acoustic wave and vibration in Internet of Things (IOT)[1],[2].

This work was supported by National Natural Science Foundation of China (NSFC) under Grant (62175100, U2001601), and Fundamental Research Funds for the Central Universities (2024300447, 0213-14380264, 0213-14380265). (Corresponding authors: yixin zhang; xupingzhang)

Sen Yan, Bangwei Liu, Imtiaz Naseeb Awan, Xiao Zhou, Yixin Zhang, Xuping Zhang are with the Key Laboratory of Intelligent Optical Sensing and Manipulation, Ministry of Education, Nanjing University. College of Engineering and Applied Sciences, Nanjing University, Nanjing 210023, China (e-mail: zyixin@nju.edu.cn, e-mail: xpzhang@nju.edu.cn,)

Ningmu Zou is with the Key Laboratory of Intelligent Optical Sensing and Manipulation, Ministry of Education, Nanjing University. School of integrated Circuits, Nanjing University, Nanjing 210023, China.

Ying Shang is with Laser Institute Shandong Key Laboratory of Optical Fiber Sensing Technologies Qilu University of Technology (Shandong Academy of Sciences), Jinan, China. School of Physical Engineering Qufu Normal University Qufu, China.

With the long sensing distance and continuous environment detection way, it can measure the vibration event along the fiber and reconstruct its amplitude and phase with high-fidelity vector information. Therefore, it has been widely applied in perimeter security and power system monitoring [3], [4],[5], oil and gas energy [6], [7] pipeline leak [8], and earthquake [9], which provides a large-scale, full-space, low-cost, high-sensitivity acoustic/vibration dynamic sensing technical basis for fiber-optic IOT.

In the application of DAS, the localization and recognition of vibration signals are two critical tasks that significantly relate its degree of intelligence. However, various invasive means and complex environmental noise lead to a high Nuisance Alarm Rate (NAR) in those tasks, which creates a bottleneck limiting its performance in field application. So far, many researchers have been exhaustively devoted to lower its NAR. Some researchers used 1-D [10], or 2-D convolutional neural networks (CNNs) [11-14], and group CNN [15] to classify types of vibration events for recognition. There are also other deep learning algorithms, such as the long short-term memory

network (LSTM) [16], and Generative Adversarial Network (GAN) [17], as well as some combined models [18], [11]. On the other hand, some studies with the aim of positioning or tracking vibration along optical fibers have been conducted. In those studies, the clustering algorithms have been employed to position the collision vibration [19], and intrusion activities [20]. However, the current works mainly concentrate on either the location task or recognition task. In actuality, it is an urgent need to achieve simultaneous identify event categories and predict the precise spatial range of events occurring along the fibers (recognition and location) within a single network, thereby enhancing the intelligence of DAS and mitigating the dilemma caused by the mismatch of location and recognition results from two algorithms.

Recently, a two-level multitask learning (MTL) approach is proposed to achieve ground event recognition and localization in radial direction [2]. However, this classification algorithm could not distinguish multiple events within a single sample. The novel object detection algorithm Faster-RCNN, YOLO were adopted in some works to tell both the event type, vibration ranges or track [21], [22], [23]. But those works concentrate on one task and perform poorly on another on DAS vibration dataset. Furthermore, these object detection algorithms have not been optimized in accordance with the characteristics of DAS vibration signals, which would generate lots of unnecessary anchors and consume a significant amount of time to process the information of vibration duration. This gives rise to additional computational consumption and affects the performance and efficiency of the algorithm.

Besides, weak vibration detection is also a challenging research issue, because the information of these signals is extremely meager in both temporal and spatial domains and they are prone to be confounded with background noise. Many researchers use specialty fibers or multimode fibers [24],[25], as well as integration of elastomer in Φ-OTDR to detect weak signals [26], [27]. The above schemes can increase the SNR of weak signals, which will increase the quality of the signal and may facilitate the tasks of localization and recognition. But these options are expensive and limited in the long-distance detection. In terms of location and recognition tasks, it is feasible to use only ordinary single-mode optical fiber and more efficient post-processing algorithm to realize those tasks of weak vibration signals.

To solve the two problems, we propose a location and recognition convolutional neural network (LR-Net) model. This end-to-end model can simultaneously locate and classify multiple vibration signals in a single sample. Within this network, a portion of the computational processes and parameters can be shared, which significantly enhances processing efficiency and minimizes the overall computing time and resource consumption. The experimental results show that the proposed model can achieve a high mAP of 99.1% and a fast predict speed of 2.2ms for six types of vibration events with location error of ±1.5 m. The primary points of innovation are as follows.

(1) A novel groups convolutional backbone is designed to extract features from time-space matrix composed of phase data. During the extraction process, the spatial dimensions remain constant while the temporal information is subject to compression gradually. This approach facilitates rapid feature extraction while effectively preserving a greater amount of spatial information for end-to-end prediction.

(2) In order to enable the model to focus on weak vibration and enhance its overall fitting ability, the location-attention-mechanism feature fusion and Squeeze framework (LAMFS) and dynamic matching strategy (DMS) are proposed for the model. Through feature fusion across multiple scales and a dynamic training strategy for weak vibrations, the mAP of the model is 2.26% higher than that of the model without LAMFS and DMS in location and recognition of weak vibrations.

(3) More importantly, a practical and novel one-dimensional prediction feature maps (OPFM) in LAMFS are used to reduce the number of impossible predictions and enhance both the efficiency and prediction speed of the model. the OPFM are obtained by squeezed the two-dimensional feature map. By fixing the vibration duration to match the size of the sample, we reduce the calculation volume and enhance the prediction speed, which is more applicable in practical scenarios.

The proposed model effectively strengthens the intelligence and versatility of DAS and provides a more intelligent scheme for detection applications. The rest of this article is organized as follows: Section II introduces the basic technical principle of DAS. Section III describes the LR-Net based on LAMFS. Section IV presents comparative experiments and results with optimized typical CNN model (Faster-RCNN, YOLO-v11). Finally, the conclusion is included in Section V. Supplementary details about optimization of the LR-Net can be found in Appendix A.

II. Architecture Of Das And Data Collection

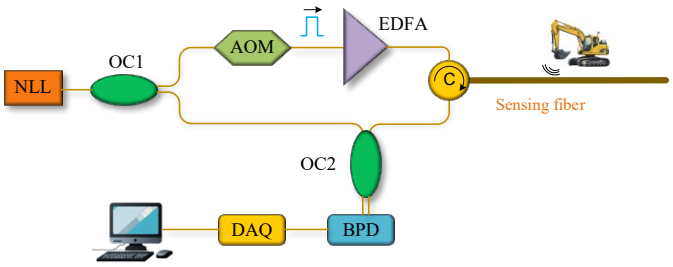


Fig. 1. The structure of distributed optical fiber sensing system.

Fig. 1 illustrates the architecture of a DAS system. The detection process for vibration signals is outlined as follows: When external vibrations are transmitted to the fiber the photoelastic effect induces difference in physical properties of the optical fiber, such as refractive index and length, leading to alterations in the phase of backward Rayleigh scattering light. This phase shift can be demodulated by a demodulator, allowing the vibrations to be sensed and transmitted to computer for data processing. Finally, these external vibrations will be precisely located and identified simultaneously by using a trained neural network in upper computer.

A. The sensing principle and architecture of DAS

DAS is a novel distributed optical fiber sensing technology using coherent detection on the basis of Φ-OTDR. The DAS we designed acquires the phase information of Rayleigh scattering light through spatial differential interferometry technology to

achieve the reconstruction of external vibration signals. The schematic diagram of the system is shown in Fig. 1. The Narrow Linewidth Laser (NLL) with 3kHz frequency width operating at 1550.12 nm was selected as the light source. The output of the NLL was split into two components, at 80% and 20% as the probe light and the local reference light by an optical coupler, respectively. The probe light is modulated by an acousto-optic modulator (AOM) into a probe pulse with a 150 MHz frequency shift. The probe pulse with pulse width of 100 ns and repetition rate of 2 kHz was amplified in an erbium-doped fiber amplifier (EDFA) and then it was injected into the sensing fiber through a circulator. The RBS light reflected from the sensing fiber mixed with the local reference light.

The mixed signal is detected by a balanced photo detector (BPD) with 200 MHz bandwidth. The signals from coupler2 (OC2) enter the PDs. Then, an 8-bit data acquisition card (DAQ) continuously sampled the output data with 1 GHz sampling rat. The phase demodulation was completed by IQ demodulation algorithm. when external vibrations induce fluctuations in the backward Rayleigh light power at specific points along the sensing fiber over time, the variations in the interference signal can be analyzed to reconstruct vibration information. Finally, the 2-D time-space matrix can be obtained by accumulating the temporal responses along the spatial axis

B. DAS Dataset Description

We have conducted field experiments in numerous important application scenarios of urban IOT and gathered data from urban oil and gas pipelines, highways, and industrial facilities. Ordinary armored single-mode optical cable (SMF-28e with a 0.9mm polyethylene protective jacket) is used in the experiment. This data has been systematically summarized to create the DAS dataset. According to field records, we identified and labeled vibration events. The labels including information with the types of events and its spatial positions (including both starting and ending points). The annotation error should be less than 1m, which is consistent with the physical resolution of the optical fiber (each node represents 1m), ensuring that the label reflects the actual spatial range within the hardware limitations.

In the experiments of the above scenarios, we all adopted a DAS with a pulse repetition rate of 2 kHz and set the spatial resolution to 1m. All samples are split into time-space matrices with the size 2000×100, which represents 1s and 100m. Fig. 2 illustrates the amplitude of the time-domain matrices for samples.

In the experiment, six categories of events are collected for model training and validation. Including Background noise (No. 0): the noise generated by vehicles running on the road, the noise of production activities, etc; Non-contact weak vibration (No. I): A person slightly walks near the fiber without touching the fiber, a person walks or talks further away from the fiber, maintaining a distance of 2 to 5 meters between the person and the optical fiber; weak vibrations represent a category of signals that are characterized by their diminutive scale in both temporal and spatial dimensions. Shaking (No. II): a person shakes the optical fiber with his hand; Digging (No. III): A person uses a shovel to dig the ground; Excavating (No. IV): A excavator digs on ground; Slapping (No. V): A person slaps against fence or wall where optical fibers are laid.

TABLE I
NUMBER AND CATEGORY OF SAMPLES

Events	Single /Multiple categories in one sample	Train/Val Samples
Background noise (No.0)	single	310/100
Weak vibration (No. I)	single	210/90
Shaking (No. II)	single	234/63
Digging (No. III)	single	225/90
Excavating (No. IV)	single	327/114
Slapping (No. V)	single	228/105
Multiple categories (No. I-V)	multiple	693/216
Total number	-	2227/778

To enhance the accuracy of the annotation, we complete the annotation through a method of manual annotation by professionals combined with on-site video record verification. With the help of some auxiliary software written in python, professionals first quickly screen out possible signals that exceed three times the average noise level, and then further label them in combination with the on-site video record.

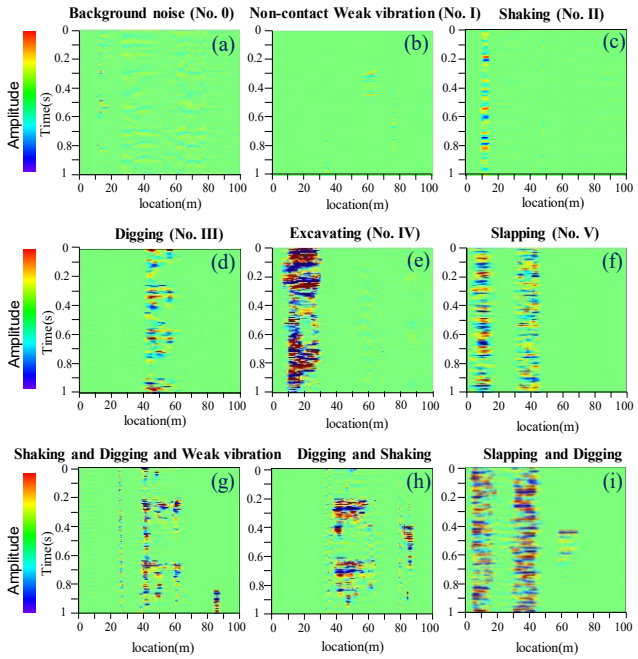


Fig. 2. The amplitude images of the time-space matrix for samples with six kinds of events. (a)-(f) Each sample has only one category of event. (g)-(i) Each sample has multiple events.

All the datasets are divided into validation sets and train sets. Their respective quantities presented in the table I. Data augmentation can expand the scale and diversity of the data set through a series of operations on existing data [28]. Cyclic shift operations of {-10m, -20m, 10m, 20m} are employed in the horizontal direction. During this process, only the samples in which no events have crossed the boundary are retained. This is used to simulate events at various positions for data augmentation, thereby enhancing the generalization capability and robustness of the model. After augmentation, the entire dataset expands to about 4 times its original size.

This article has been accepted for publication in IEEE Sensors Journal. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2025.3616499

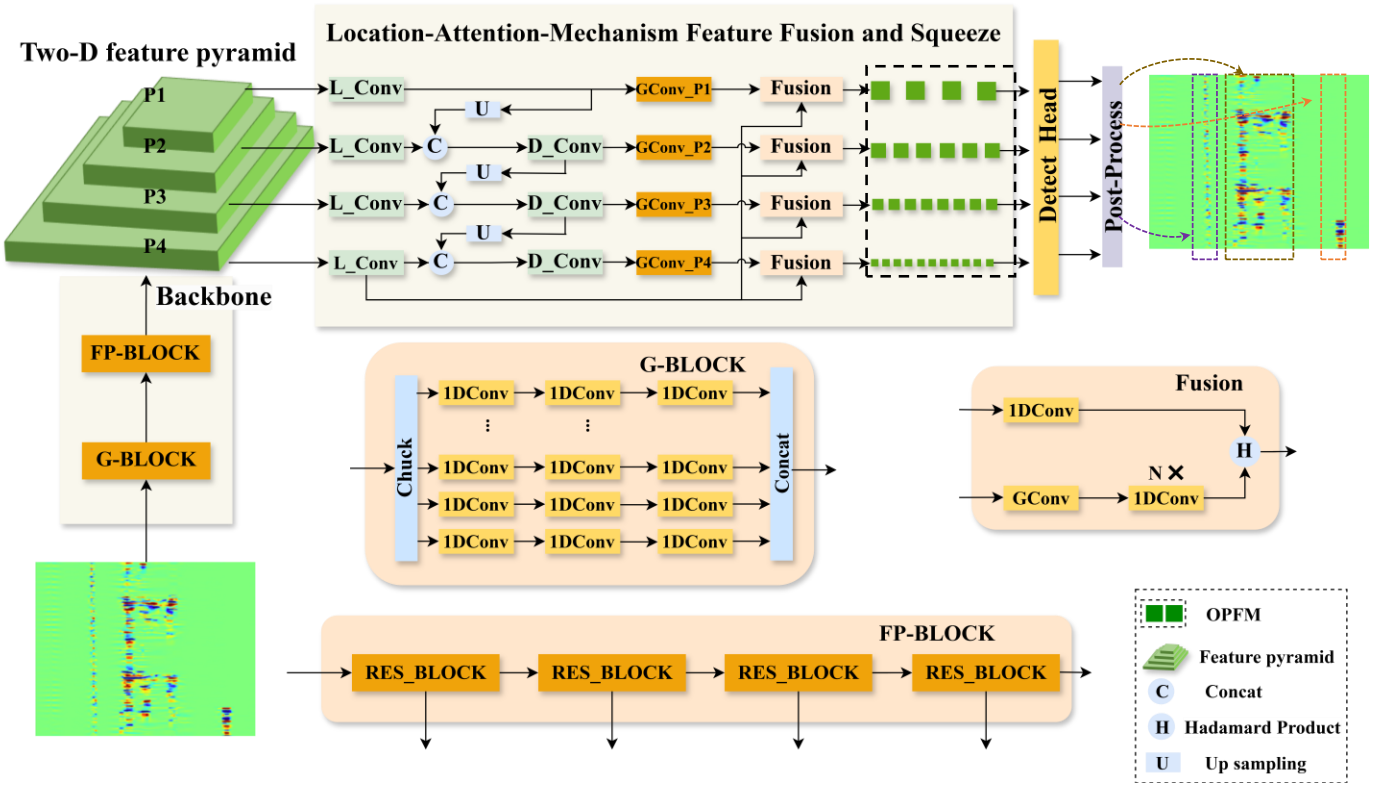


Fig. 3. Overview of LR-Net. The efficient feature extraction and fusion structure transforms the 2-d feature pyramid into one-dimensional prediction feature maps through the LAMFS. Finally, the vibration type and range are obtained through the detection head and post-process

III. Location And Recognition Convolutional Neural Network

We propose an anchor-based LR-Net to classify and locate the vibration event. The primary architecture of the network is shown in Fig. 3.

A. Architecture of LR-Net

In the process of forward propagation, the network generates probability for the presence of each object category and adjustment parameters for the default anchors to match the true vibration ranges. The whole network consists four stages: Backbone, Feature extraction and fusion with attention, Detect Head and Post-process. In the Backbone, G-block contains of three G_ Conv

$$O = f\left(\sum_{i=1,2,\ldots,n} w_i \cdot G_i + b_i\right) \quad (1)$$

The group convolution process can be described as (1). According to the number of sensor nodes in a sample, we divide the input into n groups (G1, G2, G3, ..., Gi, ..., Gn).

Next, the n convolution kernels (C1, C2, C3...Ci ...Cn) are used to extract feature from each group.w_i is the weight of the i-th convolution kernel, b_i is the bias, f represents a nonlinear activation function. The O is the output. Due to the imbalance between the time dimension and the space dimension of vibration data, G-block is employed for initial feature extraction to accelerate the convergence of the model.

This approach enables the model to acquire the vibration characteristics at diverse vibration positions. Then achieve feature extraction in the time dimension while maintaining the space dimension invariant. After G-block, the original non-equilibrium input becomes a balanced feature map with a size of (100, 100).

The group convolution backbone is uniquely tailored to DAS time-space matrices. It aligns with DAS's spatial node independence by partitioning inputs into spatial groups, using dedicated kernels to avoid cross-position interference and preserve critical spatial cues for local features like weak vibrations; Additionally, it balances the imbalanced temporal-spatial dimensions by squeezing redundant temporal information while retaining 100-node spatial resolution for precise localization; Furthermore, the architecture reduces computational complexity via intra-group parameter sharing, enabling efficiency critical for large-scale DAS.

To increase the correlation among each output of group convolution and acquire prediction feature maps with different receptive fields, the FP-Block stacked by four RES-Block are added to produce the 2-D feature pyramid.

The sizes of the feature maps used for predicting signals with diverse vibration ranges vary significantly. The vibration signals characterized by broad vibration ranges are predicted using small feature maps with extensive receptive fields, while those with narrow vibration ranges are forecasted through large feature maps that possess small receptive fields.

In the feature extraction and fusion with attention stage, feature fusion is conducted among the feature pyramid, and location-attention-mechanism feature fusion and Squeeze framework (LAMFS) is proposed to further intensify the focus

of the location task. In feature maps, the semantic information of low-level features is comparatively limited, whereas the location information of targets is abundant. Conversely, the high-level features possess rich semantic information, the localization information for targets remains relatively coarse. Inspired by the FP-Net [29], the features from features pyramid are fused to realize the intended effect.

In this process, the feature maps of each layer are conducted by 2D conv, and then the feature maps of the upper layer are up-sampled and added with the feature maps from the lower layer. Further, in order to reduce the amount of computation, we add g-conv to feature extraction and squeeze the feature maps. Finally, in LAMFS shown in the Fig. 3, the bottom feature map P4 from the feature pyramid serves as positional weight information to facilitate a second feature fusion with the results from the preceding feature fusion. which is fused with high-level features via Hadamard Product to emphasize spatially relevant regions. This ensures spatial information are not diluted during squeezing.

The LAMFS contains g-conv to squeeze the feature maps for original feature map and N+1 1D convolutions to adjust the number of channels. The prediction feature maps are obtained by fusing the two outputs with the Hadamard Product in fusion blocks. After the process of LAMFS, all output feature maps are squeezed from two-dimensional to one-dimensional feature maps along the time dimension, and one-dimensional prediction feature maps (OPFM) are obtained. This operation dynamically emphasizes spatially relevant regions by weighting high-level semantic features with low-level spatial information, while preserving both channel-wise semantic information (for classification) and the full spatial dimension (for localization).

During the stage of Detect Head, the network generates adjustment parameter on default anchor and probability for each event category. Its structure is shown in Fig. 4. Both tasks are accomplished via two 1D Conv with a kernel size of 3. The channels of output for classification task and localization task are 6 and 2 respectively, representing the number of event categories and the number of adjustment parameter. Every element of OPFM is used for predictions on default anchor with three widths. The design of 1D Conv retains the spatial background, reduces model parameters and computational complexity, and is more adaptable to vibration events of different scales.

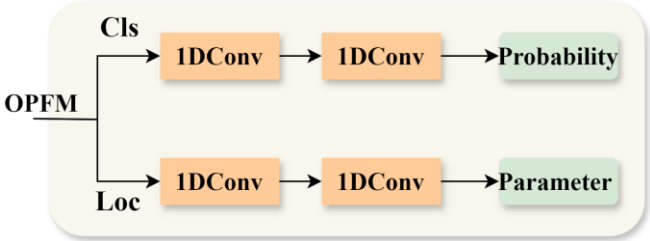


Fig. 4. Detect Head

To simplify the task of location prediction, a set of default widths is initially established at each level of the prediction feature map. During of training, the adjustment parameters generated by the Detect Head are applied to the default anchors, allowing these anchors to be continuously adjusted to better approximate the ground truth over numerous epochs.

B. Dynamic matching strategy to enhance weak vibration detection

During the training phase, it is essential to determine which default anchor aligns with the ground truth for training the network. The DMS effectively select positive and negative samples from default anchor by balancing the IOU (Intersection over Union) between the default anchor and ground truth, as well as the prediction probability. The IOU pertains to the one-dimensional length IOU on the horizontal axis.

$$S = \delta {s_{iou}}^{\alpha} + {s_p}^{\beta} \tag{2}$$

As (2) showed, S_{iou} is the length IOU between the default anchors and ground truth, S_P is the prediction probability of each predicted proposal. α, β and δ is the hyperparameter to balance these two parts. After manual tuning, α, β and δ were respectively set to 1.2, 0.7and 0.3. The topk default anchors with the highest S are selected as positive samples. Since weak signals are hard to fit, the S_P of weak vibration is usually low. By reducing α to enhance the contribution of the S_P part and increasing α to decrease contribution of the s_{iou}, more default anchors matching with weak signals can be dynamically selected as positive samples. In this way, the model can concentrate more on training these weak samples and ultimately enhance the prediction ability of weak vibration samples.

C. Objective Function and Post-process operation

$$Loss = \frac{1}{N}\left(A \cdot L_{cls} + B \cdot L_{loc}\right) \tag{3}$$

The loss function is combined with classification loss L_{cls} and location loss L_{loc}. N is the number of positive samples. A and B is hyperparameter weight for loss function. When no positive sample is matched, the loss set to zero.

The classification loss is the focal loss over multiple classes confidences, which is employed to suppress the issue of quantity imbalance between positive and negative samples.

$$L_{cls} = (1 - p_t)^{\gamma} \cdot \log(p_t) \tag{4}$$

The p_t represents the probability that the model predicts a positive class. γ is hyperparameter, and set to 2.5 in the experiment. y is label with one-hot code.

$$L_{loc} = \sum_{j}^{M} \sum_{i \in pos}^{N} \begin{matrix} smooth_{L1}\left(t_i^x - g_j^x\right) \\ + smooth_{L1}\left(t_i^w - g_j^w\right) \end{matrix} \tag{5}$$

The location loss is the distance of predicted adjustment parameters t_i^x, t_i^w and real adjustment parameters g_j^x, g_j^w. The t_i^x, t_i^w are the predicted adjustment parameters of center and width of i-th default anchor corresponding to the j-th ground truth. And the g_j^x, g_j^w are the adjustment parameters

of center and width of the j-th ground truth, respectively. They are calculated by the formula (5). N represents the quantity of positive samples that are matched, and M represents the number of ground truth. $smooth_{L1}$ function is adopted to assess the distance between the predicted adjustment parameters and the actual adjustment parameters [30]. Only the L_{loc} of positive samples and L_{cls} are computed and propagated backward, while the L_{loc} of negative samples is not calculated.

During the process of model prediction, post-processing primarily conducts four operations. First, the prediction adjustment parameters are applied to the default anchors to obtain the predicted proposals. This process is achieved through the (6).

$$\widehat{g_x} = d_w p_x + d_x, \qquad \widehat{g_w} = d_w \exp(p_w) \tag{6}$$

The center point and width of predicted proposals are denoted as $\widehat{g_x}$ and $\widehat{g_w}$, while p_w and p_x represent the predicted width adjustment parameters and the predicted center point adjustment parameters. d_w and d_x are the center and width of the default anchor. Subsequently, some proposals with low probabilities such as 0.05 are filtered out. Proposals that exceed the samples length will be cropped to no more than the sample length. Finally, the predicted proposals with the same category will be filtered by NMS (Non-Maximum Suppression). When multiple signals are detected, numerous overlapping or similar predicted proposals might be generated. Through the non-maximum suppression algorithm, those proposals with lower prediction probabilities and higher degrees of overlap can be eliminated, and only the most representative ones can be retained.

The model has been optimized to achieve optimal recognition, localization, and real-time performance, as detailed in Appendix A. It is tested using the training and validation sets obtained from the field, as outlined in Section II.

IV. Experiment And Results Analysis

In this section, we carried out the controlled experiment to assess the effectiveness and generalization ability of the proposed algorithm in comparison with the optimized algorithm Faster R-CNN [30] and YOLO-v11 [31]. Considering those algorithms are typical algorithms for Object Detection, we have undertaken specific optimizations for the task of localization and recognition of vibration signal. The major optimization involves the addition of G-conv blocks in the initial layer of the backbone network and the application of OPFM in head detection. These optimizations aim to extract balanced features from imbalanced DAS data and output the category and vibration range of the signal.

A. Performance Comparison

In the experiment, the adaptive moment estimation (Adam) optimizer is used to update the parameters. Throughout the process of the training, we use the validation set to validate and evaluate the trained network after all training iterations in each epoch. It can be considered as successful location and recognition when both the IOU and prediction probability surpass a specified threshold, provided that the predicted label is accurate.

The mean Average Precision (mAP) and loss curves are shown on Fig. 5. The AP represents the area under the precision-recall curve, which provides a comprehensive assessment of both recall and precision for the model.

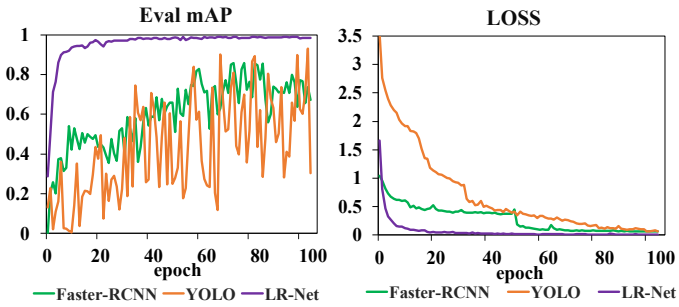


Fig. 5. mAP and loss on validation dataset

The mAP denotes the mean AP value across various threshold of IOU and prediction probability. As can be seen from the Fig. 5, the mAP curve of the proposed LR-Net tends to be stable after about 20 epochs, and the mAP value is higher than that of YOLO and Faster R-CNN. From the loss curve, the loss of LR-Net drops rapidly and remains at a low level continuously.

TABLE II
Validation Map In Various Iou Threshold

Model	mAP^{VAL}(%)	AP_{50}^{VAL}(%)	AP_{75}^{VAL}(%)	AP_{85}^{VAL}(%)
F-RCNN	88.1	89.0	89.0	89.0
YOLO	93.1	96.1	93.4	92.0
LR-NET	**99.1**	**99.5**	**99.3**	**99.2**

The mAP for each model is presented in the table II. It is evident that the mAP of LP-Net achieves a value of 0.991, whereas those of Faster R-CNN and YOLO are comparatively lower at 0.881 and 0.931, respectively. When the IOU thresholds are set at 0.5 and 0.85, the AP for LR-Net reaches up to 0.995 and 0.992, respectively, indicating its performance is relatively high and can meet the needs of the field.

TABLE III
Performance on each Category of samples

Method	AP_{85}^{VAL}(%)				
	I	II	III	IV	V
F-RCNN	80.9	86.1	89.0	91.3	98.1
YOLO	85.6	93.1	94.9	86.6	99.9
LR-NET	**98.1**	**99.9**	**98.2**	**99.9**	**99.9**

In practical applications, the NAR and Missing Alarm Rate (MAR) are critical indicators for model performance. The typical relationship between NAR and MAR with the recall and precision is (7) and (8).

$$NAR = 1 - percision = 1 - \frac{TP}{TP + FP} \tag{7}$$

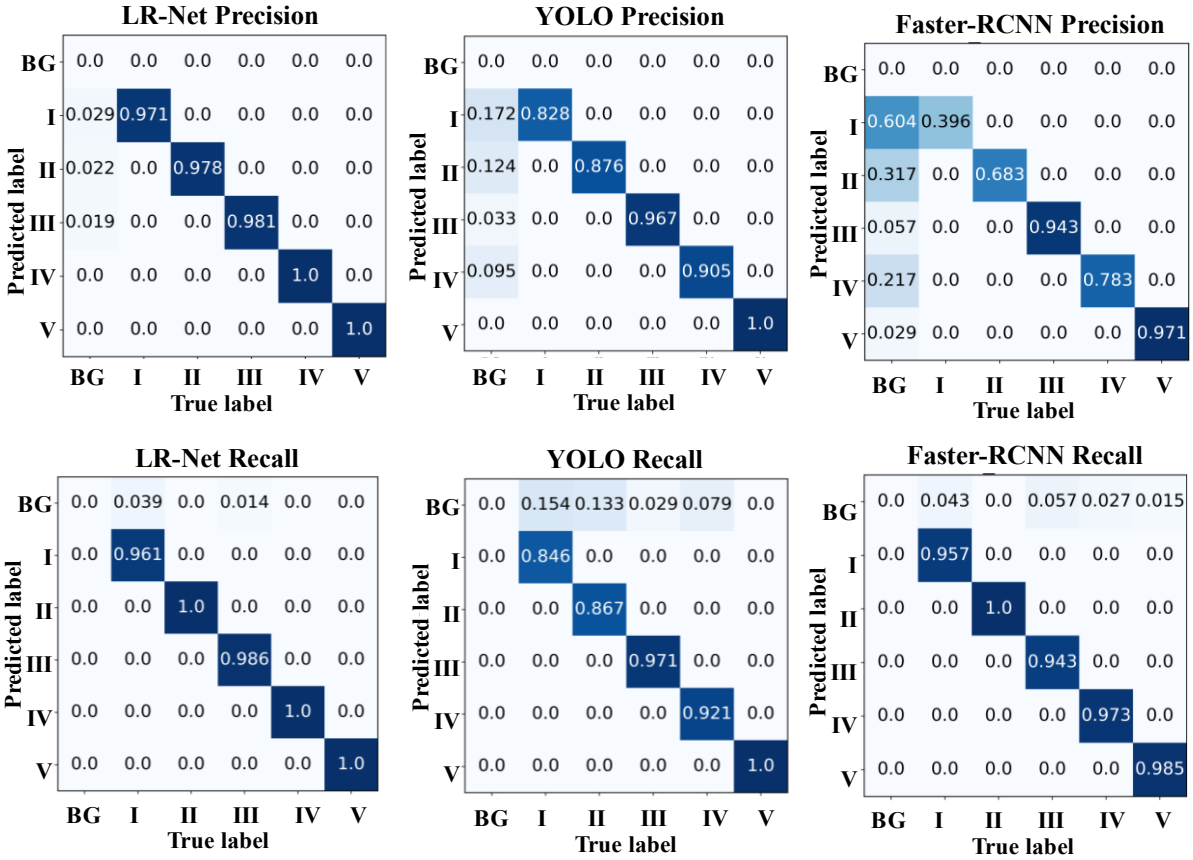


Fig. 6. Precision and Recall confusion matrix with IOU threshold of 0.85 and probability threshold of 0.5

$$MAR = 1 - recall = 1 - \frac{TP}{TP + FN} \tag{8}$$

Among them, TP, FP, and FN are True Positives, False Positives, and False Negatives respectively. Assume that the threshold of probability and IOU are 0.5 and 0.85 respectively, which are Suitable for practical applications. Then the precision and recall confusion matrices of the methods on validation dataset are displayed in Fig. 6. And the AP performance for each category is presented in Table III.

In the precision confusion matrices of LR-Net, YOLO, and Faster-RCNN, the first column represents the probability of redundant detection, while the first row is meaningless here. The diagonal elements represent the probability of correct detection, which corresponds to the precision for each category. Similarly, in the recall confusion matrices the first column lacks meaning, while the first row indicates the probability of missed detections.

It is evident from the Fig. 6 that LP-Net can achieve the highest rate of recall and precision across all sample types. As for weak vibration, LP-Net can reach a recall rate and a precision rate of 96.1% and 97.1%, which is better than 84.6%, 82.8% of YOLO, as well as the 95.7%, 39.6% of Faster-RCNN. In table III, the proposed LR-Net achieved a 98.1% Validation AP for weak vibration, which is 10% and 18% higher than that of YOLO and Faster-RCNN, respectively. That suggests the proposed model can effectively focus on the weak signals, thereby strengthening the capacity to fit the weak signals.

To further compare the performance of models across different IOU thresholds, we fixed the probability threshold at 0.5. Subsequently, the average precision and recall of the three models at IOU values of 0.5, 0.75, and 0.85, along with the corresponding NAR and MAR, are summarized in Table IV.

TABLE IV
AVERAGE PRECISION AND RECALL OF ALL SAMPLE CATEGORIES

Model	F-RCNN	YOLO	**LR-NET**
$\text{Precision}_{50}^{VAL}(\%)$	84.38	94.28	98.88
$\text{Recall}_{50}^{VAL}(\%)$	97.2	95.08	99.22
$\text{Precision}_{75}^{VAL}(\%)$	75.52	92.32	98.7
$\text{Recall}_{75}^{VAL}(\%)$	97.16	92.94	99.02
$\text{Precision}_{85}^{VAL}(\%)$	75.52	92.3	98.6
$\text{Recall}_{85}^{VAL}(\%)$	97.16	92.9	98.94
$\text{NAR}_{\pm 5m}^{VAL}(\%)$	15.62	5.72	1.12
$\text{MAR}_{\pm 5m}^{VAL}(\%)$	2.2	4.92	0.78
$\text{NAR}_{\pm 2.5m}^{VAL}(\%)$	24.48	7.7	1.3
$\text{MAR}_{\pm 2.5m}^{VAL}(\%)$	2.8	7.1	0.98
$\text{NAR}_{\pm 1.5m}^{VAL}(\%)$	24.48	8.48	**1.4**
$\text{MAR}_{\pm 1.5m}^{VAL}(\%)$	2.48	7.9	**1.06**

The precision and recall of all models tend to decrease as the IOU threshold increases. At high IOU of 0.85 threshold, LR-Net still maintains a better average precision and recall rate of

98.6% and 98.94%, respectively. Therefore, the proposed algorithm exhibits high precision and recall not only for ordinary signals but also for weak signals.

This suggests that the algorithm demonstrates a high level of reliability and possesses the capability to detect potential targets in a more comprehensive manner.

The calculation formula for location error E is as follows.

$$E = V_{label} * (1 - IOU_{threshold}) \quad (9)$$

V_{label} refers to the maximum labeled vibration length of the event. The maximum labeled vibration range for the all kinds of vibration signal was 20m. When the IOU threshold was set at 0.85, the maximal location error was 3m (±1.5m).

Therefore, it can be considered that when the threshold of the IOU is 0.85, the overall position error is ±1.5m. As can be seen from the table IV, under the location error of ±1.5m the proposed algorithm LR-NET still had low NAR and MAR, which were 1.4% and 1.06% respectively.

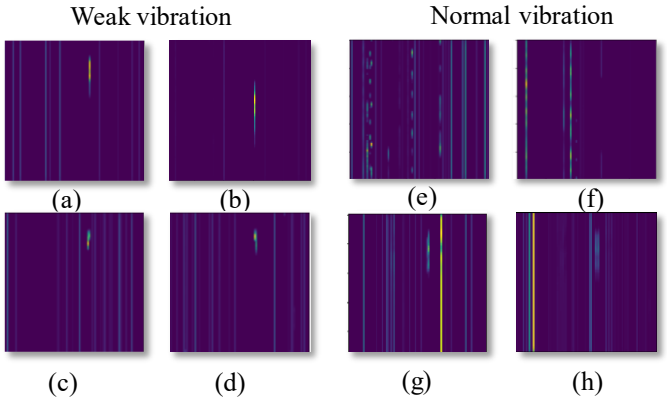


Fig. 7. Weak vibration and other vibration feature maps from Backbone.

To analyze the distribution and characteristics of data, as well as to identify potential patterns and relationships, the representative features extracted from backbone were presented in Fig. 7. The extracted features of weak signals are in the form of points, while the features of ordinary signals are represented as bars. This observation suggests that the group convolutional backbone is proficient in feature extraction. Further, according to the characteristics of feature graphs, it offers valuable insights for subsequent feature squeeze in OPFM.

Finally, the performance of real-time experiments for the localization and recognition of vibration signals is assessed on both GPU and CPU platforms. We use pytorch to implement the model on GPU GeForce 4060 and CPU Intel Core i7-13650. The evaluation of space complexity was conducted by the number of GOPS (Giga Operations per second).

The predicted time refers to the total duration for a sample prediction, which includes both inference time and post-processing time. The results of the comparison are shown in the figure Fig. 8 and table V.

The parameter of proposed LR-Net is 1.9M, which is only 4.5% and 21.8% of the optimized two-stage model Faster-RCNN and one-stage model YOLO-v11.

TABLE V
COMPARISON OF COMPUTATION EFFICIENCY

Model	GOPS	Params (M)	Predict time (GPU)	Predict time (CPU)	Inference time (GPU)	Inference time (CPU)
F-RCNN	14.5	42.1	16.3ms	63.1ms	14.7ms	61.7ms
YOLO	0. 403	8.7	3.2ms	8.4ms	2.0ms	8.2ms
LR-NET	**0.36**	**1.9**	**2.2ms**	**7.3ms**	**1.3ms**	**7.1ms**

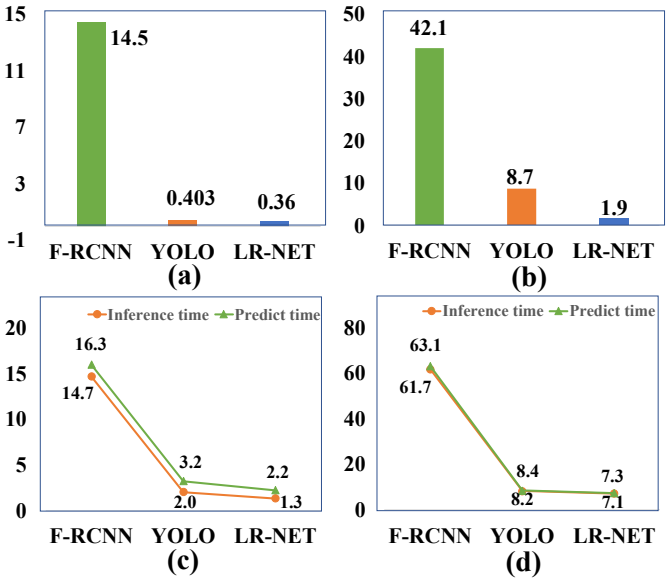


Fig. 8. Computational complexity and speed: (a) GOPS, (b) Number of parameters, (c) Predict time and Inference time on GPU, (d) Predict time and Inference time on CPU.

The computational complexity for the proposed model is 0.36 GOPS. These two performance indicators imply that the proposed model exhibits enhanced efficiency and decreased complexity.

The prediction time of the proposed LR-Net for a single time-space matrix with size of 2000 ×100 was measured at just 2.2ms on GPU and 7.3ms on CPU respectively, which is significantly outperforming YOLO. It is important to emphasize that the input time-space matrix consists of only original phase data, which does not require compression or conversion into images, nor any other pre-processing operations. so that, the end-to-end prediction can be achieved.

While the input time-space matrix size is relatively large, the group-conv backbone efficiently leverages and extracts the data, thereby significantly minimizing both computational resource expenditure and processing time.

Furthermore, the model generates predictions based on OPFM, which substantially decreases the number of anchors and enhances the speed of model. As a result, DAS with tens of thousands of sensor nodes over extensive distances can facilitate location and recognition with greater ease.

V. CONCLUSION

This paper presents an end-to-end scheme for the location and recognition of DAS vibrations events based on LR-Net

model. The groups convolutional backbone, the location-attention-mechanism feature fusion and Squeeze framework (LAMFS) and dynamic matching strategy (DMS) are proposed to enhance the model capabilities of vibration localization and recognition, particularly for weak vibration events. The experimental results demonstrate that the proposed LR-Net achieves a performance of 99.1% mAP and a 2.2ms prediction time for six types of events with a location error of ±1.5 m, which is better than optimized typical CNN models. The proposed algorithm exhibits high versatility and can be widely applied across multiple domains, including perimeter security, vibration tracking, and structural health monitoring. In the future, we intend to incorporate advanced signal enhancement techniques to improve the generalization capability of the network and further expand the scope of our research. Building upon the existing LR-Net as the foundational architecture, our objective is to integrate the radial vibration localization function in order to achieve two-dimensional vibration localization and recognition.

APPENDIX

DETAIL AND OPTIMIZATION OF THE LR-NET MODEL

This section will detail the proposed model and present the optimization of the novel framework introduced in this article. These enhancements may facilitate its implementation and application across relevant domains.

To demonstrate the efficacy of the proposed method, we performed a series of ablation studies and comparative analyses.

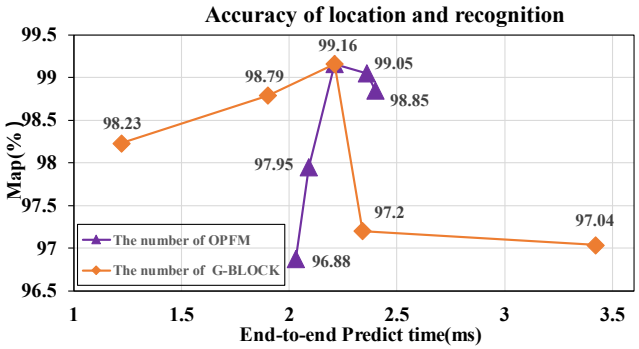


Fig. 9. Validation Accuracy of location and recognition. The triangles (purple lines) from left to right indicate that the numbers of layers of OPFM are 2, 3, 4, 5, and 6 respectively. The squares (orange lines) from left to right indicate that the numbers of G-blocks are 1, 2, 3, 4, and 5 respectively.

TABLE VI
ABLATION STUDY ON LAMFS AND DMS

Mothod	Eval mAP(%)						Predict time (ms)
	I	II	III	IV	V	mAP	
SMS+FPN	95.8	98.33	97.40	98.27	98.63	97.68	2.1
DMS+FPN	96	99.73	95.60	98.63	99.03	97.80	2.1
SMS+LAMFS	97.7	99.86	97.80	99.00	99.86	98.84	2.2
DMS+LAMFS	**98.06**	**99.70**	**98.40**	**99.76**	**99.76**	**99.14**	**2.2**

First, the location and performance with mAP and predict time are shown in Fig. 9. These results are the validation results of the models with different numbers of one-dimensional prediction feature maps (OPFM) and the models with different numbers of G-blocks in backbone. All models are trained and validated on the datasets in Table I in Section II.

It is evident that as the number of modules or layers increases, the mAP initially rises before subsequently declining. The model shows a progression in which its fitting capability gradually rises to the point of overfitting. When the number of prediction feature maps reaches 4 and the count of g blocks is 3, the model achieves its maximum accuracy rate of 99.18%.

In addition, the impact of LAMFS and DMS on model are evaluated through ablation study.

The table II provides a summary of the mAP and predict times for the models with LAMFS and DMS, along with corresponding ablation study.

At first, we employed a static sample-matching strategy and an FPN architecture. To enhance the efficiency of processing the DAS temporal-spatial matrix data and improve the learning ability for weak signals, we developed DMS and LAMFS. From the comparative results in table VI, it can be seen that the model with LAMFS and DMS has the highest mAP, demonstrating an improvement of 1.4% compared to models that do not use those. For the first category of weak vibration signals that are difficult to locate, the model achieved a 98.06% mAP, suggesting that those method effectively enable the model to concentrate on learning weak signals, thereby enhancing its capacity to fit such signals. Furthermore, as the added DMS operates only during the training phase, it does not augment the GOPS nor Slow down predict speed.

REFERENCES

[1] T. He et al., "A Surveillance System for Urban Utility Tunnel Subject to Third-Party Threats Based on Fiber-Optic DAS and FPN-BiLSTM Network," *IEEE Transactions on Instrumentation and Measurement*, vol. 73, pp. 1-9, 2024.

[2] H. Wu et al., "Smart Fiber-Optic Distributed Acoustic Sensing (sDAS) With Multitask Learning for Time-Efficient Ground Listening Applications," *IEEE Internet of Things Journal*, vol. 11, no. 5, pp. 8511-8525, 1 March1, 2024.

[3] H. Wu et al., "Field test of a fully distributed fiber optic intrusion detection system for long-distance security monitoring of national bor derline," in *Proc. 23rd Int. Conf. Opt. Fibre Sens*., 2014, pp. 1285–1288.

[4] T. He et al., "An Efficient Separation and Identification Algorithm for Mixed Threatening Events Applied in Fiber-Optic Distributed Acoustic Sensor," in *IEEE Sensors Journal*, vol. 23, no. 20, pp. 24763-24771.

[5] Y. Zhang, W. Zhao, L. Dong, C. Zhang, G. Peng, Y. Shang, "Intrusion Event Identification Approach for Distributed Vibration Sensing Using Multimodal Fusion," *IEEE Sensors Journal*, vol. 24, no. 22, pp. 37114-37124, 15 Nov.15, 2024.

[6] J. Tejedor et al., "Toward Prevention of Pipeline Integrity Threats Using a Smart Fiber-Optic Surveillance System," *Journal of Lightwave Technology*, vol. 34, no. 19, pp. 4445-4453, 1 Oct.1, 2016.

[7] S. Qu, Y. Xu, S. Huang, M. Sun, C. Wang and Y. Shang, "Recent Advancements in Optical Frequency-Domain Reflectometry: A Review," *IEEE Sensors Journal*, vol. 23, no. 3, pp. 1707-1723, 1 Feb.1, 2023.

[8] H. Wu et al., "One-Dimensional CNN-Based Intelligent Recognition of Vibrations in Pipeline Monitoring With DAS," *Journal of Lightwave Technology*, vol. 37, no. 17, pp. 4359-4366, 1 Sept.1, 2019.

[9] Walter F, Graff D, Lindner F et al. "Distributed acoustic sensing of microseismic sources and wave propagation in glaciated terrain," *Nature Communications*, vol. 11, no. 1, pp. 1-10, 2020.

[10] H. Wu et al., "One-Dimensional CNN-Based Intelligent Recognition of Vibrations in Pipeline Monitoring With DAS," *Journal of Lightwave Technology*, vol. 37, no. 17, pp. 4359-4366, 1 Sept.1, 2019
[11] 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, 1 Aug.1, 2020.
[12] Aktas, Metin , et al. "Deep Learning Based Multi-threat Classification for Phase-OTDR Fiber Optic Distributed Acoustic Sensing Applications." in *Fiber Optic Sensors and Applications XIV*. 2017. pp. 1-18.
[13] Practical multi-class event classification approach for distributed vibration sensing using deep dual path network. *Optics Express*, 2019. 27(17): pp. 23682-23692.
[14]Shi, Y., et al., An Easy Access Method for Event Recognition of Φ-OTDR Sensing System Based on Transfer Learning. J*ournal of Lightwave Technology*, 2021. 39(13): pp. 4548-4555.
[15] S. Yan, Y. Shang, C. Wang, W. Zhao and J. Ni, "Mixed Intrusion Events Recognition Based on Group Convolutional Neural Networks in DAS System," *IEEE Sensors Journal*, vol. 22, no. 1, pp. 678-684, 1 Jan.1, 2022.
[16] Z. Li, J. Zhang, M. Wang, Y. Zhong, and F. Peng, "Fiber distributed acoustic sensing using convolutional long short-term memory network: a field test on high-speed railway intrusion detection," *Optics. Express* vol. 28, no. 3, pp. 2925-2938, 2020.
[17] L. Shiloh, A. Eyal and R. Giryes, "Efficient Processing of Distributed Acoustic Sensing Data Using a Deep Learning Approach," *Journal of Lightwave Technology*, vol. 37, no. 18, pp. 4755-4762, 15 Sept.15, 2019.
[18] H. Wu, M. Yang, S. Yang, H. Lu, C. Wang and Y. Rao, "A Novel DAS Signal Recognition Method Based on Spatiotemporal Information Extraction With 1DCNNs-BiLSTM Network," in IEEE Access, vol. 8, pp. 119448-119457, 2020.
[19] C. Huang, F. Peng and K. Liu, "Pipeline Inspection Gauge Positioning System Based on Optical Fiber Distributed Acoustic Sensing," *IEEE Sensors Journal*, vol. 21, no. 22, pp. 25716-25722, 15 Nov.15, 2021.
[20] S A. Aslangul, "Detecting Tunnels for Border Security based on Fiber Optical Distributed Acoustic Sensor Data using DBSCAN." *International Conference on Sensor Networks*, pp. 78-84, 2020.
[21] W. Xu, S. Liu, F. Yu, and L. Shao, "Disturbance recognition for ϕ-OTDR based on faster-RCNN," in *Proc. 8th Symp. Novel Photoelectron. Detect. Technol. Appl.*, 2022, pp. 1196–1200.
[22] W. Xu et al., "Real-time multi-class disturbance detection for-OTDR based on YOLO algorithm," *Sensors*, vol. 22, no. 5, pp. 1994, 2022.
[23] Z. Sha, H. Feng, X. Rui and Z. Zeng, "PIG Tracking Utilizing Fiber Optic Distributed Vibration Sensor and YOLO," *Journal of Lightwave Technology*, vol. 39, no. 13, pp. 4535-4541, July1, 2021.
[24] C. Wang et al., "Distributed Acoustic Sensor Using Broadband Weak FBG Array for Large Temperature Tolerance," IEEE Sensors Journal, vol. 18, no. 7, pp. 2796-2800, 1 April1, 2018.
[25] J. Yao, B. Han, X. Jiang, S. Cao, Y.Fu, Y. Rao, Z. Ran, W. Wang, H. Guan, J, Long, "Cladding softened fiber for sensitivity enhancement of distributed acoustic sensing," *Optics. Express*, vol. 29, no. 6, pp. 8216-8222, 2021.
[26] Z. Chen, L. Zhang, H. Liu, P. Peng, Z. Liu, S. Shen, N. Chen, S. Zheng, J. Li, and F. Pang, "3D Printing Technique-Improved Phase-Sensitive OTDR for Breakdown Discharge Detection of Gas-Insulated Switchgear," *Sensors*, vol. 29, no. 3, pp. 1045–1556, 2020.
[27] B. Lu, B. Wu, J. Gu, J. Yang, K. Gao, Z. Wang, L. Ye, Q. Ye, R. Qu, X. Chen, and H. Cai, "Distributed optical fiber hydrophone based on Φ-OTDR and its field test," *Optics. Express*, vol. 29, no. 3, pp. 3147–3162, 2021.
[28] S. Yun, D. Han, S. Chun, S. J. Oh, Y. Yoo and J. Choe, "CutMix: Regularization Strategy to Train Strong Classifiers With Localizable Features," *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, Korea (South), 2019, pp. 6022-6031
[29] T. -Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan and S. Belongie, "Feature Pyramid Networks for Object Detection," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, 2017, pp. 936-944.
[30]S. Ren, K. He, R. Girshick and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137-1149, 1 June 2017
[31] Ultralytics, 2025, "YOLOv11", https://github.com/ultralytics/ultralytics.、

Sen Yan received the M.S. degree in Control science and engineering with Qilu University of Technology (Shandong Academy of Sciences), Jinan, China. He is currently pursuing the Ph.D. degree in the College of Engineering and Applied Sciences, Nanjing University.

NingmuZou received the B.S. degree from Nanjing University, Nanjing, China, in 2011, and the M.S. and Ph.D. degrees from Cornell University, Ithaca, NY, USA, in 2014, and 2017, respectively. He is currently currently a Professor with School of integrated Circuits, Nanjing University.

Ying Shang is the director of the Laser Research Institute of Shandong Academy of Sciences. His research interest includes the development and application of fiber seismic sensors and fiber optic distributed sensing.

BangweiLiu is currently pursuing a Master's degree in Optical Engineering at the College of Engineering and Applied Science, Nanjing University.

Imtiaz Naseeb Awan is currently pursuing a Ph.D. degree in Optical Engineering at the College of Engineering and Applied Science, Nanjing University.

Xiao Zhou received the B.S. degree in School of Science from Harbin Institute of Technology, Harbin, China, in 2016. He is currently working toward the Ph.D. degree with College of Engineering and Applied Sciences, Nanjing University, Nanjing, China. His main area of research covers distributed optical fiber sensing and their applications.

Yixin Zhang received the B.S. and Ph.D. degrees from Southeast University, Nanjing, China, in 2006 and 2011, respectively. In 2011, he joined the Nanyang Technological University, Singapore, as a Postdoctoral Fellow. He is currently a Professor with the College of Engineering and Applied Sciences, Nanjing University, Nanjing, China. His research interests include fiber-sensor-based health monitoring technology and digital signal processing.

Xuping Zhang received the B.E., master's, and Ph.D. degrees in electrical engineering from Southeast University, China in 1983, 1986, and 1995, respectively. She was a Research Fellow with the University of Texas at Austin from 2000 to 2002. She worked with ETH-Zurich as a Visiting Professor in 1999. She is the Director of The Key Laboratory of Intelligent Optical Sensing and Manipulation, the Ministry of Education of China, and the Dean of the Institute of Optical Communication Engineering of Nanjing University since 2002. So far, she has authored or co-authored more than 300 articles and 200 patents. Her current research interest focuses on distributed optical fiber sensing and its applications on structure health monitoring