Improve the Negative Effects of Unknown Disturbance on the Event Recognition for Phi-OTDR

Xingwei Chen[1,a], Huaxin Gu[2,b], Jingming Zhang[1,3,c], Weihao Lin[4,d], Feihong Yu[1,e], Deyu Xu[2,f] and Liyang Shao[1,g *]

[1] Department of Electrical and Electronic Engineering, Southern University of Science and Technology, Shenzhen 518055, China

[2] College of Innovation and Entrepreneurship, Southern University of Science and Technology, Shenzhen 518055, China

[3] Department of Electronic and Information Engineering,The Hong Kong Polytechnic University, Hong Kong, China

[4] School of Mechanical Electrical and Information Engineering, Xiamen Institute of Technology, Xiamen 361021, China

[a]12232126@mail.sustech.edu.cn , [b]12233182@mail.sustech.edu.cn , [c]jingming.zhang@connect.polyu.hk , [d]11510630@mail.sustech.edu.cn , [e]11930480@mail.sustech.edu.cn , [f]12233181@mail.sustech.edu.cn , [g]shaoly@sustech.edu.cn

ABSTRACT

This paper presents a new method to reduce the high false alarm rate. When the Phi-OTDR system is applied in complex environment, irrelevant event signals similar to the target event often interfere with the recognition of the target event and leads to false positive in the system. This method includes two key components: one is the automatic labeling mechanism of unknown information, which is used to find unknown events easily confused with target events in real environment; the other is the improved Region of Interest (ROI) module based on contrast clustering and EBMs, which can detect target event signals in real complex environment and reduce the negative impact of unknown events on target event recognition.

Keywords: Distributed Acoustic Sensing, Phi-OTDR, Pattern Recognition, Open-set

1. INTRODUCTION

Distributed Acoustic Sensing (DAS) is an optical fiber sensing technology based on the principle of back-Rayleigh scattering, which realizes the perception of the external environment by measuring the phase change generated when the light pulse propagates in the optical fiber. It offers advantages such as long sensing distance, high sensitivity, and accurate positioning. It is widely used in pipeline monitoring[1], perimeter security[2], transportation[3] and other scenarios. The ultra-long sensing distance of Phi-OTDR generates a large volume of monitoring data. The identification algorithm can be used to analyze the data, and the remote real-time alarm of multi-source risk event signals can be realized. Most traditional methods rely on feature extraction to recognize different types of event signals. Although these methods can achieve a high recognition rate, the process of feature extraction is highly dependent on expert experience, and the calculation process of event signal recognition is complex and time-consuming.

Fourth International Computational Imaging Conference (CITA 2024),
edited by Xiaopeng Shao, Proc. of SPIE Vol. 13542, 1354255
 · doi: 10.1117/12.3057175

With the development of artificial intelligence, machine learning and deep learning have been widely introduced into Phi-OTDR to solve event recognition problems.Machine learning relies on mathematical methods to learn from existing data, build predictive models, and predict and classify data.Support vector machines[4], F-ELM[5], and XGBoost[6] are all used for DAS signal recognition and classification.The advantage of deep learning is that it can automatically learn features from raw data and perform efficient pattern recognition. Neural networks such as CNN[7], SNN[8] and MLPNN[9] have been used for DAS signal pattern recognition.The researchers proposed a two-stage recognition network to improve the problem of the increased false alarm rate of the intrusion detection system caused by animal activities in complex environments[10].

These methods have achieved significant results in event recognition problems of DAS, but the performance of these methods is mostly based on research using static, closed test datasets. However, in the field deployment of DAS equipment, due to the ultra-long monitoring distance of DAS, the environment along the sensing optical cable is extremely complex and variable. As shown in Fig. 1, there are a large number of unknown event interferences outside the training dataset that affect the model's recognition of target events, which have a negative impact on the actual deployment and application of machine learning or deep learning models.

Fig.1 Confusion event interference recognition.

This paper presents a new method for the target event detection of phi-OTDR systems deployed in real environments, which effectively improves the recognition accuracy compared with the baseline. This method includes two key modules: the first is the automatic labeling mechanism of unknown information, which is used to find unknown events easily confused with target events in the data to be detected in an open environment; the second is the improved ROI module based on contrast clustering and EBMs, which can effectively separate the interference of unknown events on the detection results of target events. After training the proposed model and baseline on a closed-set, we test them on a closed-set and an open-set with interference items, respectively. Baseline mean precisions have decreased from 85.6% to 58.65%, and the model performance we proposed is only down from 91.92% to 67.60%. This means that our proposed model has a lower false alarm rate regardless of whether there is interference signal. By effectively processing the confusing unknown events, this model can enhance the reliability of DAS event detection tasks in real environments.

2. METHOD

We use Faster-RCNN as the basic detector, which shares convolutional features with the detection network through the regional proposal network (RPN), and can realize real-time target detection of distributed optical fiber acoustic sensing signals in the closed data set, while maintaining high accuracy[11].

Our improved model consists of three parts: convolutional layers, an unknown-aware region proposal network (UARPN), and an ROI-Head. In the convolutional layer, CNN network is used to extract features and generate feature maps. The generated feature map will be processed in two ways simultaneously. The first is to continue the convolution to produce a feature map of higher dimensions; the second is to input the feature map into the UARPN network to obtain the region proposal and confidence. Then, the firstknown and firstunknown region proposals are retained using non-maximum suppression. The high-dimensional features and the region proposal given by UARPN are simultaneously put into the ROI-Head, and the features are extracted from the high-dimensional features according to the region proposal. Finally, using EBMs combined with contrastive clustering loss, unknown interference classes were excluded, and the target classes in the closed set were classified. At the same time, a bounding box regression was performed to obtain the final accurate location of the detection frame. Our proposed module works on both RPN and ROI components to improve the performance of the basic detector deployed in a real environment. The entire model structure is shown in Fig. 2.

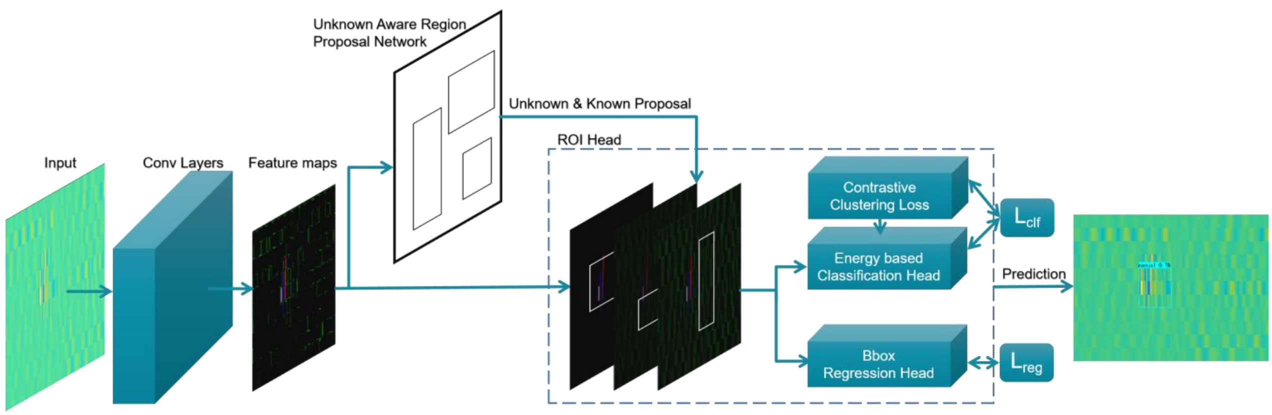

Fig.2 The overall structure of the model proposed.

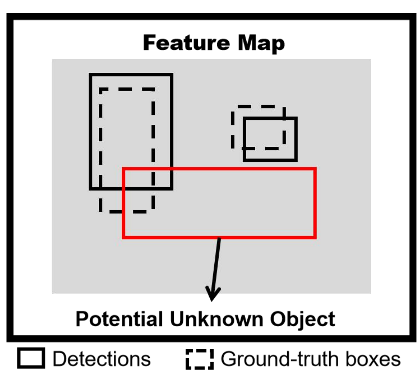

Fig. 3 Auto-labelling unknowns schematics.

It is not possible to manually label all potential unknown disturbance signals in a real environment. As shown in Fig. 3, we rely on the regional proposal network (RPN) being a class agnostic characteristic, given an input image, the RPN generates a set of bounding box predictions for the foreground and background instances, along with the corresponding confidence levels. We mark as potentially unknown those proposals that have a high confidence, but do not overlap with ground-truth boxes. This automatic marking mechanism of unknown information quickly finds out the confusing unknown samples in the recognition process and treats them as "unknown". In order to effectively exclude unknown events, we introduced Energy-based models (EBMs)[12] to improve the classification head. EBMs can estimate the compatibility between the observed variable F and the set of possible output variables L by learning a function that uses a single output scalar. By learning the data within the distribution, EBMs can output a low energy value to the known class sample, and give a high energy value to the out-of-distribution data, so as to realize the separation of known and unknown classes.

In ROI-Head, we model the classification problem under open sets as a contrast clustering problem, where instances of the same type will be clustered together and instances of different types will be further apart. We define the class set of all known classes as $K^t = \{1, 2, \dots N\}$. For each known class $i \in K^t$, we keep a prototype vector p_i. Letbe the feature vector generated by the intermediate layer of the object detector for the class c. We define contrastive loss as follows:

$$\mathcal{L}_{cont}(f_c) = \sum_{i=0}^{c} \ell(f_c, p_i) \quad \text{where}$$

$$\ell(f_c, p_i) = \begin{cases} D(f_c, p_i),\ i = c \\ \max(0, \Delta - D(f_c, p_i)),\ \text{otherwise} \end{cases}$$

where D is any distance function and Δ defines how close a similar and dissimilar item can be. Minimizing this loss would ensure the desired class separation in the sample space.

3. EXPERIMENTS

The DAS system was deployed in the 8.4 km pipeline field to carry out data acquisition experiments. The DAS equipment used was 1000 Hz pulse frequency and 10 m spatial resolution. In the experiment, we collected approximately 1000 samples of each of 4 typical signals, and formed data set A after labeling, including mining signals of rammer, artificial mining signals, pump station operation signals, and vehicle driving signals.

Table 1. Experiment database: the instance number of each type of event.

Event	Experiment database A	Experiment database B
PowerRammer	1000	1000
Manual	1000	1000
Subway	1000	1000
Vehicle	1000	1000
Unknown	0	2000

The instance types and sizes of the dataset are shown in Table 1 below. In order to simulate the field situation faced by the model after deployment in the real scenario, we used the disturbance positioning method to detect signals collected in the non-experimental area, and obtained about 2000 unknown signal samples, which were directly added to dataset A without annotation to form dataset B. Both datasets were predivided into the training set and the test set in a 9:1 ratio.

After the Faster-RCNN and the model proposed in this paper are trained on the train set of dataset A respectively, the recognition effect of the four known classes and the performance degradation caused by unknown disturbance are tested on the test set of dataset A and B. The results are shown in Table 2.

Table 2. Test results of event identification influenced by unknown disturbance

Precisions	Faster-RCNN (baseline)		Our	
	Experiment database A	Experiment database B	Experiment database A	Experiment database B
PowerRammer	97.1%	82.6%	87.0%	70.9%
Manual	78.7%	38.3%	85.9%	45.2%
Subway	99.0%	69.6%	99.0%	82.4%
Vehicle	67.6%	44.1%	95.8%	71.9%
Mean Precisions	85.60%	58.65%	91.92%	67.60%

It can be seen that for the data containing unknown signal samples, the recognition effect of Faster-Rcnn and our proposed model on different categories of events decreases to different degrees. However, the model that applied the unknown sample labeling mechanism and the novel ROI-Head had better resistance to the negative effects of unknown disturbance than the base detector. This means that when deployed in real-world environments, the new model will have better false alarm rate performance.

4. CONCLUSION

This paper proposes a method for identifying disturbances in distributed optical fiber acoustic sensing systems in open environments. The experimental results show that the automatic labeling mechanism of unknown disturbance and the new ROI-Head used in the proposed model weaken the effect of unknown disturbance on the performance of the model trained in the open environment and trained in the closed data set to some extent. It has better performance on open datasets than the baseline, which hopefully improves the high false alarm rate when the model is deployed in the real world.

REFERENCES

[1] Yang Y, Li Y, Zhang T, et al. Early safety warnings for long-distance pipelines: A distributed optical fiber sensor machine learning approach; proceedings of the Proceedings of the AAAI Conference on Artificial Intelligence, F, 2021 [C].

[2] Abufana S A, Dalveren Y, Aghnaiya A, et al. Variational Mode Decomposition-Based Threat Classification for Fiber Optic Distributed Acoustic Sensing[J]. IEEE Access 2020, 8:100152
[3] Wang S, Liu F, Liu B. Research on application of deep convolutional network in high-speed railway track inspection based on distributed fiber acoustic sensing [J]. Optics Communications, 2021, 492: 126981.
[4] Fouda B M T, Yang B, Han D, et al. Pattern Recognition of Optical Fiber Vibration Signal of the Submarine Cable for Its Safety[J]. in IEEE Sensors Journal, 2021, 21(5): 6510-6519.
[5] Jia H Z, Lou S Q, Liang S, et al. Event Identification by F-ELM Model for Phi -OTDR Fiber-Optic Distributed Disturbance Sensor[J]. IEEE Sensors Journal, 2020, 20(3): 1297-1305.
[6] Meng H, Wang S L, Gao C, et al. Research on Recognition Method of Railway Perimeter Intrusions Based on Φ -OTDR Optical Fiber Sensing Technology[J]. IEEE Sensors Journal, 2021,21(8): 9852-9859.
[7] Zhao X, Sun H B, Lin B, et al. Markov Transition Fields and Deep Learning-Based Event-Classification and Vibration-Frequency Measurement for φ-OTDR[J]. IEEE Sensors Journal, 2022,22(4): 3348-3357.
[8] Wu H J, Gan D K, Xu C R, et al. Improved Generalization in Signal Identification With Unsupervised Spiking Neuron Networks for Fiber-Optic Distributed Acoustic Sensor[J]. Journal of Lightwave Technology, 2022,40(9): 3072-3083.
[9] Mi Q S, Yu H D, Xiao Q, et al. Intrusion behavior classification method applied in a perimeter security monitoring system[J]. Opt. Express 2021,29(6): 8592-8605.
[10] He T, Sun Q, Zhang S, et al. A dual-stage-recognition network for distributed optical fiber sensing perimeter security system [J]. Journal of Lightwave Technology, 2022.
[11] Xu W J, Yu F H, Liu S Q, et al. Real-Time Multi-Class Disturbance Detection for Φ-OTDR Based on YOLO Algorithm[J]. Sensors. 2022; 22(5):1994.
[12] Yann LeCun, Sumit Chopra, Raia Hadsell, M Ranzato, and F Huang. A tutorial on energy-based learning. Predicting structured data, 1(0), 2006.