Deep Learning-based Intrusion Detection and Impulsive Event Classification for Distributed Acoustic Sensing across Telecom Networks

Shaobo Han, Ming-Fang Huang, Tingfeng Li, Jian Fang, Zhuocheng Jiang, and Ting Wang

(Invited Paper)

***Abstract*—We introduce two pioneering applications leveraging Distributed Fiber Optic Sensing (DFOS) and Machine Learning (ML) technologies. These innovations offer substantial benefits for fortifying telecom infrastructures and public safety. By harnessing existing telecom cables, our solutions excel in perimeter intrusion detection via buried cables and impulsive event classification through aerial cables. To achieve comprehensive intrusion detection, we introduce a label encoding strategy for multitask learning and systematically evaluate the generalization performance of the proposed approach across various domain shifts. For accurate recognition of impulsive acoustic events, we compare several standard choices of representations for raw waveform data and neural network architectures, including convolutional neural networks (ConvNets) and vision transformers (ViT). We also study the effectiveness of the built-in inductive biases under both high- and low-fidelity sensing conditions and varying amounts of labeled training data. All computations are executed locally through edge computing, ensuring real-time detection capabilities. Furthermore, our proposed system seamlessly integrates with cameras for video analytics, significantly enhancing overall situation awareness of the surrounding environment.**

***Index Terms*—Acoustic event classification, deep learning, distributed optical fiber sensing, gunshot detection, intrusion detection, machine learning, network field experiment.**

I. Introduction

Distributed fiber optic sensing (DFOS) technology, which utilizes the fundamental sensing capabilities of optical fiber with wide area coverage, has been applied in diverse applications. These include earthquake detection [1], pipeline monitoring and leakage detection [2], [3], structure change monitoring [4], road traffic monitoring [5], and railway intrusion detection [6].

Recently, there has been growing interest in applying DFOS technology into the telecommunications sector, given the vast fiber infrastructures that telecom carriers have built over the past 30 years accommodate the growth of internet traffic and the interconnection of 5G and beyond networks among cities, towns, homes, and data centers. While transmission fibers were initially intended solely for data transmission, they are now being explored as potential sensing media [7]–[10]. Meanwhile, there is also an increasing interest in applying machine learning (ML) and AI algorithms to extract actionable information from fiber sensing data and support autonomous decisions in real-time [11]–[17].

Operational telecom fiber networks offer significant potential for optical sensing applications. DFOS technology turns existing communication cables into individual sensing elements located every meter, with all the measurements synchronized [18]. Consequently, this sensing technology can be employed to identify threats both to the fiber infrastructure and the surrounding environment, contributing to community safety [19]. In this paper, we present field results utilizing DFOS and ML/AI technologies for (1) perimeter intrusion detection over buried cables to safeguard the infrastructure, and (2) impulsive acoustic event detection over aerial cables to identify gunshot events and other threatening situations in the surrounding environment.

DFOS boasts a multitude of advantages, including long distance, immunity to electromagnetic interference, and robust in harsh environment [20]. Additionally, its unique feature of integrating sensing and data transmission within the same fiber negates the need for electrical for electrical power in the field. When synergized with ML, DFOS system involves into a covert, instantaneous, and simultaneous classifier of multiple physical intrusions or disturbances. This amalgamation facilitates robust perimeter intrusion detection, offering enhanced security measures for safeguarding national borders, airports, seaports, data centers, power plants, and other critical assets.

When physical intrusions induce vibrations transmitted through subterranean optical fibers along land borders, property boundaries, or facility perimeters, the disturbances trigger discernible changes in light. This paper introduces machine learning-based methodologies can simultaneously classify physical intrusions types and other auxiliary labels and provide valuable information to law enforcement agencies. This includes details such as proximity to the cable and the direction of the intruder's movement. Additionally, we also study the generalization performance with different fractions of training data, across different field conditions (e.g., heavy rain to snow), and information sharing across multiple related tasks through a shared encoder with multiple task heads [21].

Detection system identifying impulsive acoustic events, such as gunshots, within public areas (e.g., cities, schools, hotels, etc.) play a pivotal role in fortifying public safety

Preprint version. (Corresponding author: Shaobo Han.)

The authors are with NEC Laboratories America, Princeton, NJ 08540 USA (e-mail: shaobo@nec-labs.com; mhuang@nec-labs.com; tli@nec-labs. com; jfang@nec-labs.com; zhijiang@nec-labs.com; ting@nec-labs.com).

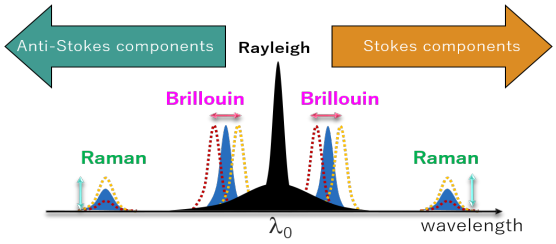

Fig. 1. Schematic of backscattering signals that are exploited in reflectometry-based DFOS.

and swiftly alerting law enforcement to gun-related incidents. Conventional methods relying on electrical microphones necessitate dense installations due to the limited coverage, incurring substantial maintenance costs, encountering power and data transmission issues. Notably, these solutions raise privacy concerns and are susceptible to false alarms triggered by similar impulsive sounds like firecrackers or door slams.

In this paper, which is an extension to [22], we demonstrate that distributed acoustic sensing (DAS) technology can capture rich acoustic characteristics in recorded data, and deep learning-based approaches are capable of classifying multiple types of impulsive acoustic events with high accuracy. Meanwhile, we also empirically study the sufficiency of some dimension-reduced representations of the DAS waveform data that protects privacy while preserving the utility of data for specific downstream machine learning tasks only.

The rest of paper is organized as follows: Section II outlines the principles behind the employed DFOS sensing technology and the field setup. Section III and IV introduces the two aforementioned use cases, providing detailed experimental design and validation. In Section V, we present system implementation, focusing on edge processing and multi-modal integration with video analytics. Section VI concludes the paper. Additional details about data processing and the neural network architectures used in this paper can be found in Appendices A and B.

II. Field Experimental Setup

Figure 1 illustrates a visual aid highlighting the fundamental aspect of DFOS: showcasing in the measurement of nonlinear backscattering signals — Rayleigh, Brillion and Raman phenomena [20], [23]. Our study employed a DAS system, utilizing Rayleigh optical time-domain reflectometer (OTDR) detection [18]. This methodology gauges alterations in Rayleigh scattering intensity via interferometric phase beating. With coherent detection, the DAS system retrieves comprehensive polarization and phase information from the backscattering signals. Our setup involved a 1550-nm laser, operating at a sampling rate of 125 MHz, utilizing brief optical pulses, and on-chip fast processing. These configurations enabled achieving an impressive sensor resolution, as fine as 1 meter.

Figure 2 presents the setup employed during the field trials, showcasing the configuration comprising a DAS system situated at the central office (CO), a 38-km field fiber, and extension fiber deployed beyond the CO perimeter to fortify facility security. The extension fiber consists of three

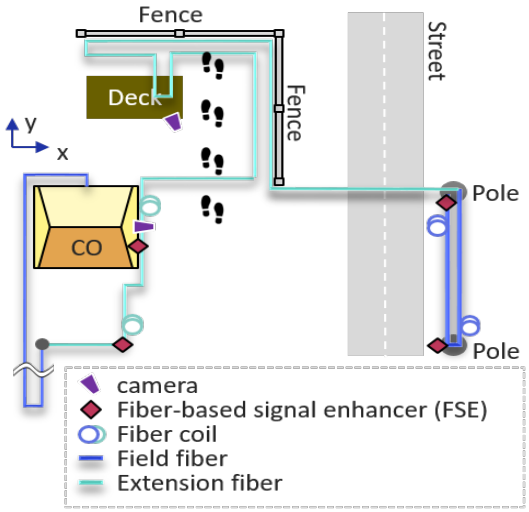

Fig. 2. Experimental Setup with three sections of cable: buried, attached to the fence, and hanging on poles with coils.

segments: buried underground, affixed to the perimeter fence, and suspended on poles with coils. The CO serves as a the core sensing infrastructure, utilizing existing fiber for comprehensive environmental monitoring, such as road traffic monitoring. This central hub then extends connectivity to the extension fibers, facilitating the integration of new sensing branches. These extensions enable diverse applications, such as intrusion detection through buried cables and impulsive acoustic event monitoring via aerial cables. Incorporating fiber coils and fiber-based signal enhancers (FSEs) into the test bed, the system architecture remains straightforward and seamless. Localization and data derived from events is amalgamated with video analytics, offering spatial and temporal insights to track individuals associated with the detected events.

The proposed solution holds the potential to facilitate uninterrupted monitoring of intrusion attempts, gunshots, and various public safety threats across areas in regions with telecommunication fibers. The technique and discoveries outlined in this report offer a substantial stride forward in developing a vibration surveillance and acoustic monitoring system leveraging DFOS technology. This contribution signifies a noteworthy advancement in enhancing security measures through innovative monitoring techniques.

III. Intrusion Detection

While there exists broad categories of intrusion events, such as fence shaking, our specific focus lies on human-movement intrusion scenarios, particularly involving an individual's proximity to a cable and the moving direction. Our solution combines ML techniques with the DAS system. In this setup, fiber optic cables detect vibrations that can signal different types of intrusions. ML algorithms then use pattern recognition to identify unusual behavior patterns, helping to spot potential security threats.

Intrusion events exhibit inherent diversity, characterized by varying elements such as the movement direction, speed, and proximity to the cable. Given this complexity, our focus rests on developing a model adept at concurrently recognizing these

distinct facets. To tackle this challenge, we've devised and compared two distinct approaches, thoroughly evaluated across multiple tasks through experimentation.

Additionally, we conduct experiments to assess the method's performance, examining its ability to sustain accuracy under varying weather conditions and with different types of movements. This comprehensive evaluation spans changes in environmental dynamics, encompassing scenarios like rain and snow, alongside different physical movements like running, walking, or other activities influencing the detection process.

A. Data Collection

Our study involved a structured data collection process, as depicted in Fig. 3. We focused on a segment of cable spanning six locations, each separated by an interval of 5 meters. Each sample collected is a composite of various factors, including weather conditions, moving directions, and activity types. Specifically, the data encompasses: **(1) Weather Conditions**: collections were made on days following rain (83 samples) and snow (88 samples). **(2) Activity Directions and Types**: movements were captured in four directions relative to the cable's orientation:

- Moving parallel to the cable, (127 instances), where 64 *to_right* and 43 *to_left* instances, 64 moving *near* cable (0m-7m) and 63 moving *far* from cable (8m-15m).
- Moving vertically to the cable, either towards or away from it (44 instances in total and 22 each).

Illustrations of typical movement patterns include an orange figure demonstrating vertical walking away from the cable starting at Location 3, and a green figure depicting running from left to right parallel to the cable are shown in Fig. 3.

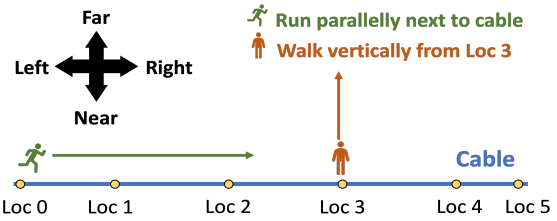

Fig. 3. Data collection setting for intrusion detection

Our dataset construction involved extracting 44×24 patches from each image. With the spatial resolution 1.6m and temporal sampling rate 2 KHZ, each patch covers 38m and 22ms. These patches were labeled based on their overlap with predefined ground-truth bounding boxes (referenced as white boxes in Fig. 4). We randomly splitting these patches into training and testing sets. Detailed statistics of the dataset are presented in Appendix A, Table VII and Table VIII.

B. Experiment Design

a) Generalization: Data collection in the field often encounters constraints due to high cost or practical limitations, making it challenging to encompass all conceivable combinations of field conditions. In most cases, training data collection is confined to a few accessible scenarios, sometimes lacking

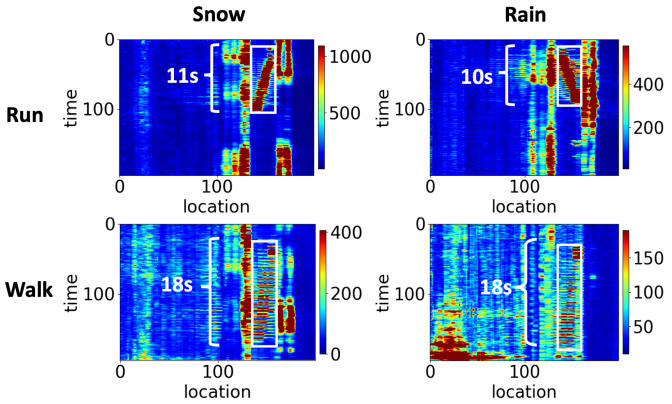

Fig. 4. Waterfall images of different weather conditions and moving activities

data from the target domain during the training phase entirely. Moreover, test environments are often different from the model's training domain. The characteristics of sensing data exhibit variations across diverse weather-ground conditions, signal source types, and time-varying backgrounds [15], [24], [25]. In this work, we study two types of generalization: (1) *Weather conditions: snow & rain.* Data collected post-rainy or snowy day feature distinct ground conditions — wet ground or snow covered — affecting the signal sensed by the cable. As illustrated in Fig. 4, the data intensity shown in color bars collected post-snowy days notably surpasses that after rainy days; (2) *Activity types: run & walk* demonstrates that walking patterns typically exhibit longer duration (18s) compared to running ones (around 10s) for the same distance, as shown in Fig. 4. These exemplary figures aim to succinctly convey the challenges encountered in data collection scenarios while setting the stage for the specific types of generalization studied in the work.

In our approach to address this generalization task, we employed a neural network model designed for binary classification, using patches in sensing waterfall images as the input data. The principal aim of this model revolves around identifying the presence or absence of detectable movement activity within these specified patches.

b) Multitask and fine-grained label: It is interesting and useful to investigate the detailed categories of the intrusion events. For example, from the direction of the event, we may be able to determine if the person is trying to trespass or simply passing by. Also, we can determine if an event is a threat from the vertical distance to the cable. To explore the details of an event, we designed four classification tasks. For events parallel to the cable, there are two tasks: Task1 *to_left* vs. *to_right* to distinguish the event direction, Task2 *near* vs. *far* to distinguish the event distance to the cable. For events vertical to the cable, there is Task3 *go_further* vs. *go_closer* to distinguish the vertical direction. We also consider Task4 *walk* vs. *run* in this experiment.

To simultaneously learn all the classification tasks, there are two formulations:

- *Distributed Label Encoding*. The label can be represented as 4-digit integer as shown in Fig. 5 (a), which shows four label examples for encoding above mentioned four tasks. Each digit is either 0 or 1 (light or dark green) represent-

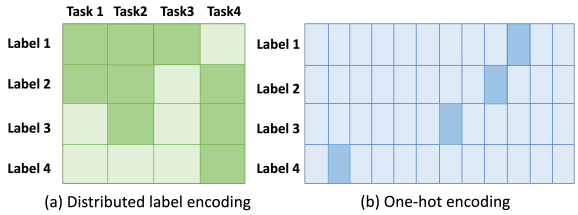

Fig. 5. Label encoding schemes: (a) distributed and (b) one-hot encoding

ing the label for one task. We adopt a ConvNet model with a shared backbone for common feature extraction, and 4 heads for outputs where each head corresponds to one binary classification task. This formulation involves training a model on multiple learning tasks simultaneously. We aim to determine whether the model's learning improves when labels are mapped into four semantically meaningful subspaces.

- *One-hot Label Encoding*. This one-hot encoding, on the other hand, is a case of single-label classification that associates the input samples to a unique target label from a set of disjoint labels. In other words, the labels corresponding to the input samples form a disjoint set. For 4 binary classification tasks, there are 16 possible combinations while 4 are not present in our dataset, resulting only 12 combinations. We use one-hot label coding, where N=12, as shown in Fig. 5 (b). We adopt the same backbone with head architecture as before, except that the head has 12 outputs. In this setting, the model won't have the concept of tasks, all that is asked is to distinguish 12 exclusive classes instead of 2 outputs per task in previous setting.

C. Experimental results

Table I shows the results of the four generalization tasks, where *Train Acc* denotes the accuracy in source domain and *Test Acc* is for target domain. We run ten times per setting and report the average and standard deviation of the accuracies. It can be seen that the model can generalize to unseen weather condition and activity type. It is also interesting to see that the performance differs when source and target domain exchanges. For example, the model of Snow2rain is more accurate than Rain2snow on target domain. One possible reason is that features learned under post-snowy condition may be more robust and transferable to other conditions. In contrast, features learned in post-rainy condition may be more specific to rain-related challenges and may not generalize as well. Moreover, the variability within the training data can also play a role. Further experiments are needed to analyze the reason and we leave them as future work.

TABLE I
GENERALIZATION EXPERIMENT RESULTS

Task	Snow2rain	Rain2snow	Walk2run	Run2walk
Train Acc (%)	99.43±0.29	99.84±0.07	99.17±0.58	98.12±1.12
Test Acc (%)	99.06±0.14	96.66±0.55	98.40±0.33	95.90±0.88

Table II compares the performances of Distributed Label Encoding (DLE) and One-hot Label Encoding (OLE), where *Task Acc* is the accuracy of each task. We also list the accuracies in different ratios of training set size over all samples. The results demonstrate that the DLE consistently outperforms the OLE by a significant margin across all tasks and training set ratios. Furthermore, the OLE model exhibits a greater sensitivity to variations in the training set size, particularly in Task 2. In this case, the accuracy of the OLE model drops by 18% when the training set size ratio decreases from 75% to 20%, whereas the DLE experiences only 6% decrease in accuracy.

TABLE II
COMPARISON OF DISTRIBUTED LABEL ENCODING (DLE) AND ONE-HOT LABEL ENCODING (OLE).

Ratio	Model	75%	50%	30%	20%
Task1	OLE	84.91±4.77	83.43±2.56	77.80±2.61	76.16±3.13
Acc (%)	DLE	94.72±1.40	93.27±0.67	90.43±1.43	90.85±0.67
Task2	OLE	58.57±5.31	55.45±6.01	47.21±4.29	39.55±4.63
Acc (%)	DLE	87.68±3.31	86.36±3.12	84.87±1.62	81.25±2.60
Task3	OLE	73.71±3.06	74.62±2.41	72.02±2.75	69.90±2.27
Acc (%)	DLE	83.90±1.68	83.49±1.74	83.80±0.94	80.12±2.50
Task4	OLE	84.09±2.96	81.52±2.18	76.51±2.51	74.80±2.09
Acc (%)	DLE	89.21±1.67	87.36±2.26	83.16±2.27	80.75±3.40

Distributed label encoding is commonly used in Multi-Label Classification (MLC), which has been successfully applied in computer vision, natural language processing and data mining [26]. For example, a natural image may have multiple objects, it is more practical to associate each image with multiple tags or labels. Thus, developing methods to address MLC problem becomes increasingly crucial for real-world image classification tasks, among which [27], [28] are the first few methods proposing neural networks to solve these problems.

Our experiments have shown that the performance of models is notably enhanced when labels are converted into subspaces that are not only semantically meaningful but also closely related to the specific tasks at hand. This approach helps the development of a more detailed and precise classification system, specifically designed to handle the complexities in real-world data.

In this section, we introduce a machine learning approach for intrusion detection, emphasizing both model generalization capabilities and fine-grained label classification across multiple tasks. The experiments demonstrate the effectiveness of this method in both areas. While the performance of our distributed label encoding approach surpasses that of one-hot label encoding, we observe that it still falls short of the performance achieved by training separate models for each task. This discrepancy may be attributed to the large capacity of the shared backbone, thus there is no interference between tasks. Other factors, such as task interdependencies and the weighting of the training process for each task, could also influence the performance of the multi-task learning (MTL) model, as discussed in [29]. Enhancing the performance of MTL is important, especially with the aim of expanding it to include a wide range of tasks within a single model in future applications. We leave this as future work.

IV. Impulsive Acoustic Event Classification

We propose using DAS and ML techniques to detect gunshot events and distinguish gunshot and gunshot-like sounds, leveraging the high sampling rate provided by DAS. In addition to starter guns, we add several similar, loud, and impulsive sounds, including four kinds of firework-induced false alarms: crackers, cannon, fountain cannon, and high-altitude firework. We also consider two vehicle-related alarm: vehicle door slamming and vehicle alarms, along with background noise. We collected samples from 8 classes over two different days. Detailed information is summarized in Table III.

TABLE III
Dataset details

Sensor	Class	Training	Validation	Test
FSE	Background	84	21	53
	Starter Gun	102	25	59
	Door Slam	44	11	28
	Car Alarm	108	27	72
	Crackers	67	16	16
	Cannon	44	11	28
	Fountain Cannon	166	41	104
	High Altitude Firework	32	8	24
	Total	647	167	384
Fiber Coil	Background	84	21	53
	Starter Gun	51	12	34
	Door Slam	44	11	28
	Car Alarm	73	18	48
	Crackers	33	8	15
	Cannon	22	5	14
	Fountain Cannon	83	20	52
	High Altitude Firework	16	4	12
	Total	406	106	256

Data is collected at a sampling rate of 20 kHz from two types of sensors, (1) Fiber-based signal enhancer (FSE) with a mandrel to improve the sensitivity, (2) Fiber Coil, readily available in the field. The FSEs were made by wrapping the single-mode fiber (SMF) around thin-walled cylindrical transducer made of elastic material (glycol-modified polyethylene terephthalate, PETG, Young's modulus = 2.1 GPa, Possion's ratio = 0.34). The manufacturing process was similar to [30] with additional protection for outdoor use. The dimensions of each cylindrical transducer are approximately 50 mm (outer diameter) x 120 mm (height) x 0.5 mm (thickness). The total fiber wrapped on each FSE is around 30 meters. When sound wave arrives the FSE, the acoustic pressure creates deformation in the cylinder, thereby causing a phase change of the optical fiber wrapped on it. The radial deformation of the cylinder corresponds to a longitudinal change of fiber, resulting in the differential phase measured by DAS. More details on the theoretical equations and calculations for the FSE can be found in [30]. The fiber coils were made by wrapping the SMFs as coils. Each fiber coil has the diameter of around 100 mm. The fiber length of each coil is around 30 meters. The fiber coils were fixed using zip-ties.

The sensing waveform data is cut into segment of 1-second length, with each segment containing a single type of sound. This procedure is automated using a peak-finding algorithm. Exemplary waveform and its Mel spectrogram for each of the 8 type of events are shown in Figure 6. Note that the two sensors have different signal-to-noise ratio (SNR) performance.

In total, we gathered 1198 and 768 data samples from FSEs and fiber coils, respectively. Among them, 384 and 256 were collected from different runs of sound event creation and held out as test data. The remaining data were divided into a training set and a validation set in an $80:20$ ratio. To evaluate the sampling efficiency of different methods, a proportion of training set $\mathcal{T}$ is used for model training, ranging from 10%, 20%, ..., to 100%. Across all experiments, the validation set $\mathcal{V}$ and test set remain the same for model selection and evaluation, respectively. Specifically, hyperparameters are optimized based on empirical loss on the validation set. $\lambda^\star = \arg\min_{\lambda\in\Lambda} \mathbb{E}_{x\sim p(x)}\left[\mathcal{L}(x; A_\lambda(\mathcal{T}))\right] \approx \arg\min_{\lambda\in\Lambda} \frac{1}{|\mathcal{V}|}\sum_{x\in\mathcal{V}}\left[\mathcal{L}(x; A_\lambda(\mathcal{T}))\right]$, where $\lambda^\star$ represent the optimal hyperparameter (e.g., the number of epochs) selected from candidate set Λ, and $A_\lambda(\mathcal{T})$ represent the learning algorithm that maps training data $\mathcal{T}$ to a neural network model fitted under λ. For each setting, 10 Monte Carlo experiments are performed with different random seeds but using the same model selection strategy. For model comparison, the average accuracy on the test set is reported.

A. Baseline methods

We consider five types of data representations that are commonly used in audio signal processing and speech recognition, including

1) DAS waveform signals with high-pass filter at 200 Hz,
2) Short-time Fourier transform (STFT) spectrogram,
3) Mel spectrogram (with amplitude converted to the decibel scale),
4) Mel-frequency cepstral coefficients (MFCCs), and
5) Flatten MFCCs, converting 2D matrix in 1D vector with time-frequency structure destroyed.

These methods are arranged in order based on incremental processing procedures, leading to increasing dimension reduction. The latter four representations provide hand-crafted features with certain characteristics that can be considered as implicit inductive biases. In contrast, the waveform signal representation lacks built-in inductive bias (except for the high-pass filtering), relying instead on automatic feature engineering by the neural network model. It is of interest to evaluate their effectiveness in the context of DAS data under different SNR conditions and training sample sizes.

Alongside each data representation, we considered different machine learning methods, including

1) Convolutional neural networks (ConvNet), 1D convolution are used for waveform data,
2) Vision transformer (ViT) [31], and
3) Random forest.

The combination of data representations and machine learning methods results in seven baseline methods, as summarized in Table IV. The checkmark (✓) and cross mark (✗) denote whether a property is satisfied or not. The waveform represents the original 1D DAS signal in the temporal domain, with only high-pass filtering applied as a preprocessing step. The waveform can be processed into 2D *time-frequency* representations such as Short-Time Fourier Transform (STFT), Mel Spectrogram, and Mel-Frequency Cepstral Coefficients (MFCCs),

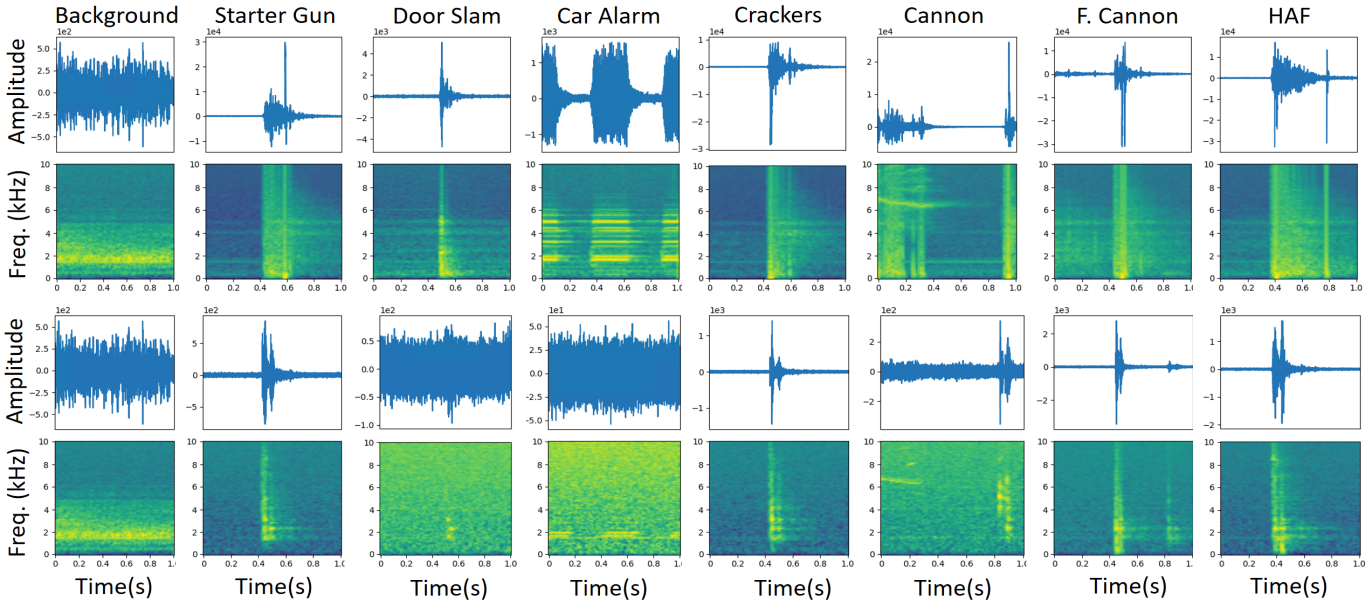

Fig. 6. Fiber-based signal enhancer (FSE) data recorded by DAS: original waveform of of 1 second (first row) and its Mel spectrogram (second row); Fiber coil data recorded by DAS: original waveform of of 1 second (third row) and its Mel spectrogram (fourth row). F. Cannon and HAF stand for Fountain Cannon and High-Altitude Firework, respectively.

TABLE IV
INDUCTIVE BIASES OF DIFFERENT METHODS (SIGNAL REPRESENTATIONS AND ML MODEL CHOICES.)

Method	Hand-craft feature	Time-frequency structure	Human auditory	Cepstral analysis	Group-equivariance
Waveform (ConvNet)	✗	✗	✗	✗	✓
STFT (ConvNet)	✓	✓	✗	✗	✓
STFT (ViT)	✓	✓	✗	✗	✗
Mel Spectrogram (ConvNet)	✓	✓	✓	✗	✓
Mel Spectrogram (ViT)	✓	✓	✓	✗	✗
MFCC (ConvNet)	✓	✓	✓	✓	✓
MFCC (Random Forest)	✓	✗	✓	✓	✗

serving as the feature extraction step before machine learning. The latter two techniques emphasize human auditory system's frequency perception. Among the ML models, ConvNet incorporates built-in group-equivariance in both time and frequency axis. For the sake of reproducibility, key design choices for signal processing procedures and hyperparameters for neural network training are provided in Table IX in the Appendix.

B. Experimental results

One of the major challenges in impulsive sound detection systems is false alarms. To address this issue, we propose the use of high-sampling-rate DAS data and dedicated deep learning approaches. Following customized signal processing procedures, we trained classification models to distinguish multiple impulsive acoustic events based on the chosen data representation of the vibrations captured by the DAS. Figure 7 and Figure 8 show the mean and standard deviation of the test accuracy of different methods with different amount of training data on FSE and fiber coil, respectively.

Short-term power spectrum representation such as STFT, Mel spectrogram, and MFCC, convert the 1D DAS waveform into *time-frequency* representations. ConvNet and ViT models can utilize the time-frequency information by treating the 2D spectrogram data as image patches and processing them using local kernels, with ConvNet assuming additional translation invariance. Given the high sampling rate, the majority of the STFT channel information lies in the high-frequency range. In contrast, Mel spectrogram and MFCC filter banks, inspired by human auditory ability, emphasize lower-frequency channels and the envelope of the short-time power spectrum.

Despite limited training data, our model can accurately recognize various fireworks sounds, including crackers, cannons, fountain cannons, and high-altitude fireworks, along with other safety-related sound events like car alarms, starter guns, and door slams. The results in Figure 7 and Figure 8 demonstrate the extent to which the inductive biases listed in Tabel IV are helpful, from the low-data regime to the high-data regime. Note that in the high SNR setting (FSE), the Mel spectrogram (ConvNet) approach achieves the best performance with a significant margin. Conversely, in the low SNR setting (fiber coil), the waveform (ConvNet) approach without hand-craft features show a slight advantage in the low-data regime. The confusion matrix results of these models are shown in Figure 9 and Figure 10. Using the same Mel spectrogram representation, the results from (supervised) ViT consistently show lower performance compared to ConvNets. This can be explained by the preference for local patterns in this application and the lack of necessity for long-range dependencies. Results from the random forest classifiers based on vectorized MFCC features lead to suboptimal performance, indicating the importance of the time-frequency structure.

Both collecting fiber sensing data and providing labels for

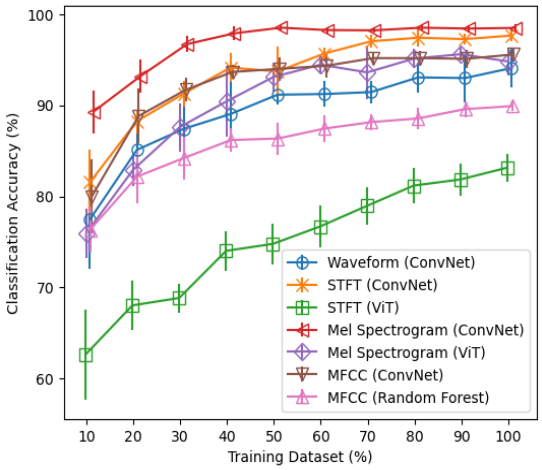

Fig. 7. FSE: Test accuracy vs. training sample size

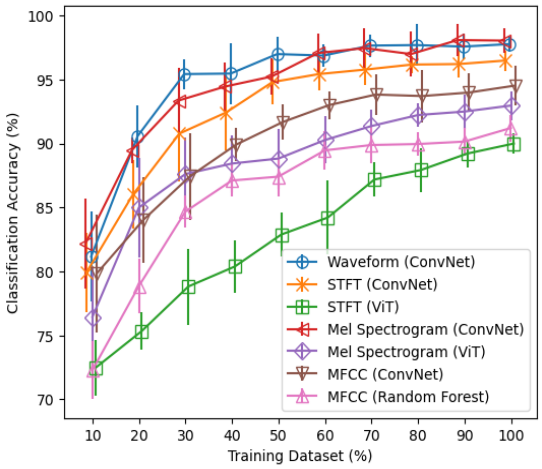

Fig. 8. Fiber coil: Test accuracy vs. training sample size

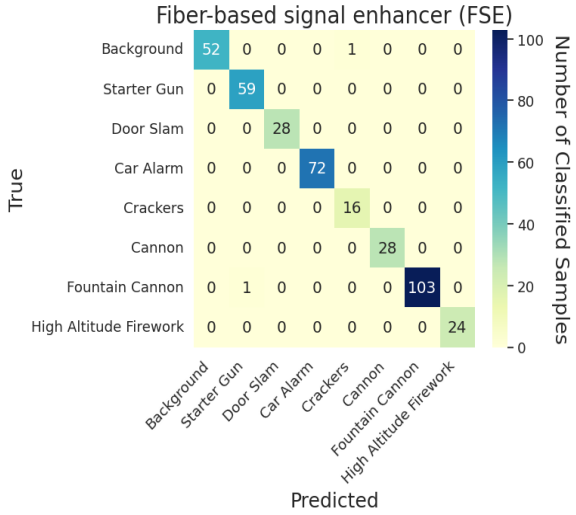

Fig. 9. Confusion matrix on fiber-based signal enhancer (FSE)

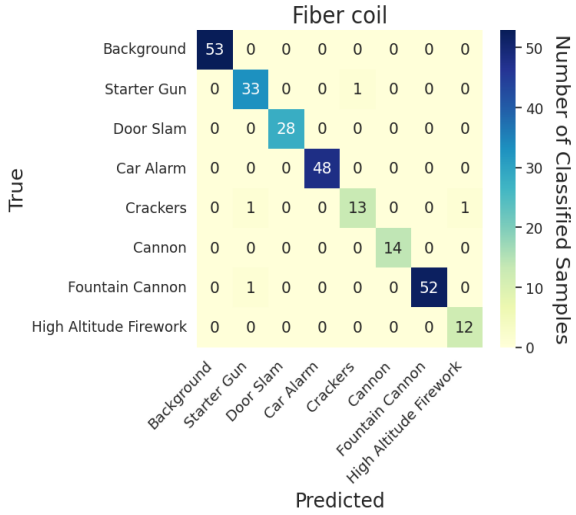

Fig. 10. Confusion matrix on fiber coils.

machine learning training are costly processes. To enhance the data variability, one can also consider generative model-based approaches [13], which have demonstrated success on DFOS data. In the two applications presented in this paper, DFOS data containing precise event occurrences are used, requiring manual labeling along both the time and location dimensions. The annotation effort can be mitigated by adopting attention-based framework for weakly-supervised learning of DFOS data [16].

V. System Implementation

Considering the unique challenges and requirements of DFOS over telecom networks [10], we propose customized AI/ML modules for each distributed fiber sensing application, with additional considerations for (1) edge AI processing, and (2) multimodal sensor fusion. The modules are hosted in an edge AI platform, which can be placed in the central office or terminal of carriers.

A. Edge processing

The massive volume of distributed acoustic sensing data presents a significant challenge to the computing hardware, architecture, and algorithms used in training and inference [32]. Considering the costs associated with data transmission and storage, the stringent latency requirements, and the protection of data privacy, edge AI (or on-premise AI) infrastructures stands out as a more appropriate choice for fiber sensing applications than cloud-based infrastructures [33]. In comparison, cloud computing is not a viable solution to meet these needs, given the difficulty in transporting the sheer volume of sensing data and the double-loop delay caused by moving data to the cloud and waiting for the results to return to the local site.

Through edge processing, ML/AI inference on DFOS data is provided in real-time, enabling timely actions to be taken. Fig. 11 illustrates our multiple-in-one AI system, which is hosted on a platform designed to execute pipelined computations locally. Upon receiving the sensing data, the engine filters out signals under normal conditions, such as road traffic trajectories, before feeding the data into the Fiber-InD (intrusion detection) and Fiber-IAD (impulsive acoustic event detection) modules.

The Fiber-InD module, based on convolutional neural network (ConvNet), is proposed to classify intrusion events, such as human walking or running. Its outputs include event type and auxiliary information such as the location, time stamp, direction along the cable, closeness to the cable, and

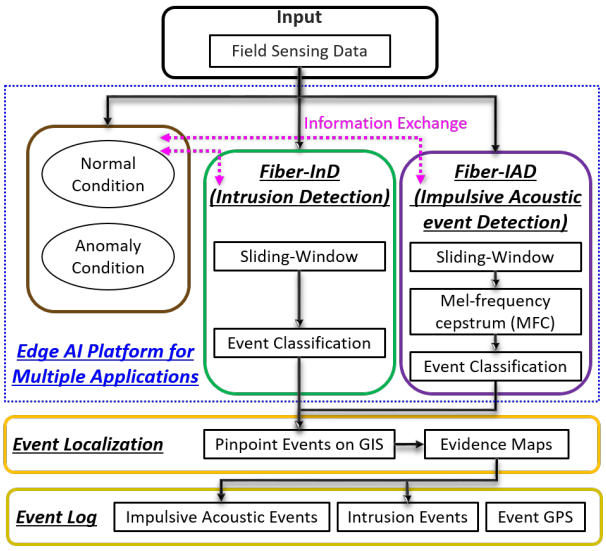

Fig. 11. Flow chart of the AI platform for simultaneous multipurpose sensing.

probability. Additionally, this module can be expanded to include more intrusion events, such as digging, fence shaking, and fence cutting. The Fiber-IAD module reports the category of the detected acoustic events and stores only the processed data after sufficient reduction for re-training. The processing pipeline reduces lengthy time series data into intermediate representations, and a neural network with only a small number of layers is needed for inference. These computations can be done efficiently on modern hardware.

B. Multimodal fusion

We developed a comprehensive multimodal impulsive acoustic detection system incorporating aerial coils, buried fiber, FSEs, and cameras. This advanced system not only detects and localizes impulsive event through DAS, capturing vibration patterns generated by the source but also employs video analytics to track the origin of the sound. In response to DAS triggers, the cameras precisely pinpoint individuals at the sound location, identifying them as potential threat. A distinct alarming boundary box is supermiposed on the visual recording, enhancing threat visibility. The system is equipped with both a regular camera (Camera 1) and a fisheye camera (Camera 2) strategically positioned at varying angles of expansive coverage. To facilitate person identification and tracking, a ConvNet-based object detection model meticulously trained. The model, utilizing ResNet50 as the backbone, incorporates a cascade Region-based ConvNet detector with a shared region proposal network across datasets. The subject's identify and trajectory are established by associating the location coordinates of the detected sound with those of the identified person's boundary box, utilizing an intersection-over-union operation.

As shown in Fig. 12 (a) and (b), the subject was detected and tracked with a red boundary box in both cameras, while pedestrians and vehicles are marked in green and blue boundary boxes, respectively. Additionally, the impulsive sound location was visualized as an augmented heatmap through time difference of arrival (TDoA) analysis on a GIS. Subsequently (Fig. 12(b)), the suspect invaded the protected area, which was in the cameras' blind zones, as illustrated in Fig. 12(c). However, the buried optical fiber clearly detected the suspect's footsteps and tracked his movement even in blind zones of the cameras. The system demonstrated the effectiveness of impulsive sound detection and tracing, particularly in cases where single-modality detection is insufficient.

This system can be utilized for reconstructing crime scenes and continuously monitoring various events, including car alarms triggered by theft, car break-ins, home break-ins, and prohibited fireworks, to enhance public safety in future smart and secure city applications.

VI. Conclusions

Our exploration of fiber optic sensing across telecom networks has encompassed a diverse spectrum of infrastructure protection applications, notably intrusion detection and and impulsive acoustic event monitoring. Central to our approach is the integration of fiber sensing with machine learning techniques, yielding data-driven solutions. The proposed approaches are designed to be label-efficient by maximally taking into account the physical knowledge and are adaptable to changing deployment environments. Powered by an edge AI platform, our system is able to process multiple applications simultaneously and locally with low latency, allowing us to achieve real-time response. Through field tests, our system has showcased exceptional performance in event detection, classification, as well as precise localization and identification of potential threats using advanced video analytics. With continuous innovation and development, we believe that this technology has the potential to significantly enhance public safety and security, for both telecom infrastructure protection and situation awareness of its surrounding environment.

Acknowledgment

The authors would like to thank James M Moore, Jason Cascio, TJ Xia and Glenn Wellbrock from Verizon for their great support to this work.

References

[1] P. D. Hernández, J. A. Ramírez, and M. A. Soto, "Deep-learning-based earthquake detection for fiber-optic distributed acoustic sensing," *Journal of Lightwave Technology*, vol. 40, no. 8, pp. 2639–2650, 2022.

[2] P. Stajanca, S. Chruscicki, T. Homann, S. Seifert, D. Schmidt, and A. Habib, "Detection of leak-induced pipeline vibrations using fiber-optic distributed acoustic sensing," *Sensors*, vol. 18, no. 9, p. 2841, 2018.

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

[4] F. Bastianini, A. Rizzo, N. Galati, U. Deza, and A. Nanni, "Discontinuous Brillouin strain monitoring of small concrete bridges: Comparison between near-to-surface and smart FRP fiber installation techniques," in *Smart Structures and Materials 2005: Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems*, vol. 5765. SPIE, 2005, pp. 612–623.

[5] C. Narisetty, T. Hino, M.-F. Huang, R. Ueda, H. Sakurai, A. Tanaka, T. Otani, and T. Ando, "Overcoming challenges of distributed fiber-optic sensing for highway traffic monitoring," *Transportation Research Record*, vol. 2675, no. 2, pp. 233–242, 2021.

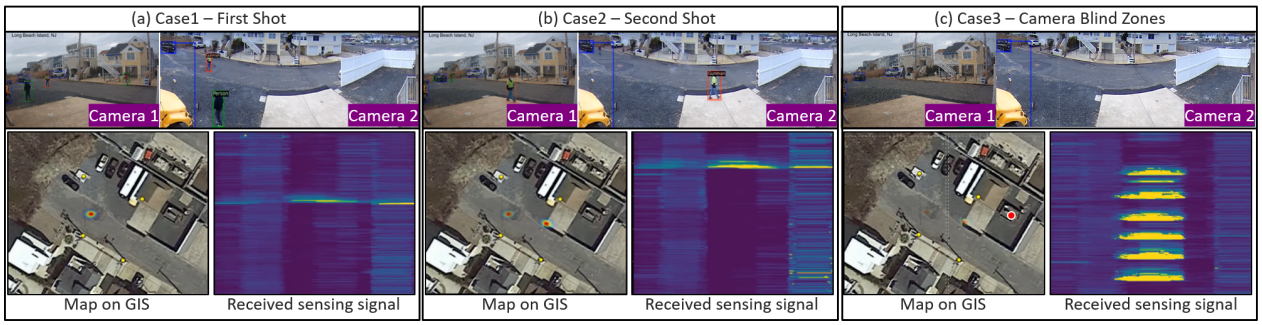

Fig. 12. Sensing fusion field results by integrating video and fiber sensing technologies to locate the subject.

[6] 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.

[7] M.-F. Huang, M. Salemi, Y. Chen, J. Zhao, T. J. Xia, G. A. Wellbrock, Y.-K. Huang, G. Milione, E. Ip, P. Ji *et al.*, "First field trial of distributed fiber optical sensing and high-speed communication over an operational telecom network," *Journal of Lightwave Technology*, vol. 38, no. 1, pp. 75–81, 2019.

[8] G. A. Wellbrock, T. J. Xia, M.-F. Huang, M. Salemi, Y. Li, P. N. Ji, S. Ozharar, Y. Chen, Y. Ding, Y. Tian *et al.*, "Field trial of distributed fiber sensor network using operational telecom fiber cables as sensing media," in *2020 European Conference on Optical Communications (ECOC)*. IEEE, 2020, pp. 1–3.

[9] E. Ip, J. Fang, Y. Li, Q. Wang, M.-F. Huang, M. Salemi, and Y.-K. Huang, "Distributed fiber sensor network using telecom cables as sensing media: technology advancements and applications," *Journal of Optical Communications and Networking*, vol. 14, no. 1, pp. A61–A68, 2022.

[10] G. A. Wellbrock, T. J. Xia, M.-F. Huang, S. Han, Y. Chen, T. Wang, and Y. Aono, "Explore benefits of distributed fiber optic sensing for optical network service providers," *Journal of Lightwave Technology*, vol. 41, no. 12, pp. 3758–3766, 2023.

[11] J. Tejedor, J. Macias-Guarasa, H. F. Martins, J. Pastor-Graells, P. Corredera, and S. Martin-Lopez, "Machine learning methods for pipeline surveillance systems based on distributed acoustic sensing: A review," *Applied Sciences*, vol. 7, no. 8, p. 841, 2017.

[12] H. Liu, J. Ma, T. Xu, W. Yan, L. Ma, and X. Zhang, "Vehicle detection and classification using distributed fiber optic acoustic sensing," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 2, pp. 1363–1374, 2019.

[13] 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, 2019.

[14] Y. Lu, Y. Tian, S. Han, E. Cosatto, S. Ozharar, and Y. Ding, "Automatic fine-grained localization of utility pole landmarks on distributed acoustic sensing traces based on bilinear resnets," in *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2021, pp. 4675–4679.

[15] T. Li, Y. Chen, M.-F. Huang, S. Han, and T. Wang, "Vehicle run-off-road event automatic detection by fiber sensing technology," in *2021 Optical Fiber Communications Conference and Exhibition (OFC)*. IEEE, 2021, pp. 1–3.

[16] A. Bukharin, S. Han, Y. Chen, M.-F. Huang, Y.-K. Huang, Y. Xie, and T. Wang, "Ambient noise-based weakly supervised manhole localization methods over deployed fiber networks," *Optics Express*, vol. 31, no. 6, pp. 9591–9607, 2023.

[17] N. Tonami, S. Mishima, R. Kondo, K. Imoto, and T. Hino, "Event classification with class-level gated unit using large-scale pretrained model for optical fiber sensing," in *Proceedings of the Workshop on Detection and Classification of Acoustic Scenes and Events (DCASE)*, 2023, pp. 196–200.

[18] E. Ip, F. Ravet, H. Martins, M.-F. Huang, T. Okamoto, S. Han, C. Narisetty, J. Fang, Y.-K. Huang, M. Salemi *et al.*, "Using global existing fiber networks for environmental sensing," *Proceedings of the IEEE*, vol. 110, no. 11, pp. 1853–1888, 2022.

[19] M.-F. Huang, S. Han, G. A. Wellbrock, T. J. Xia, C. Narisetty, M. Salemi, Y. Chen, J. M. Moore, P. N. Ji, G. Milione *et al.*, "Field trial of cable safety protection and road traffic monitoring over operational 5G transport network with fiber sensing and on-premise AI technologies," in *Optoelectronics and Communications Conference*. Optica Publishing Group, 2021, pp. T5A–8.

[20] Z. He and Q. Liu, "Optical fiber distributed acoustic sensors: A review," *Journal of Lightwave Technology*, vol. 39, no. 12, pp. 3671–3686, 2021.

[21] Y. Zhang and Q. Yang, "A survey on multi-task learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 34, no. 12, pp. 5586–5609, 2021.

[22] M.-F. Huang, S. Han, and J. Fang, "Real-time intrusion detection and impulsive acoustic event classification with fiber optic sensing and deep learning technologies over telecom networks," in *49th European Conference on Optical Communications (ECOC)*, 2023, pp. Tu.C.7.1:1–4.

[23] S. P. Singh, R. Gangwar, and N. Singh, "Nonlinear scattering effects in optical fibers," *Progress In Electromagnetics Research*, vol. 74, pp. 379–405, 2007.

[24] Z. Feng, S. Han, and S. S. Du, "Provable adaptation across multiway domains via representation learning," in *International Conference on Learning Representations*, 2021.

[25] H. Wu, D. Gan, C. Xu, Y. Liu, X. Liu, Y. Song, and Y. Rao, "Improved generalization in signal identification with unsupervised spiking neuron networks for fiber-optic distributed acoustic sensor," *Journal of Lightwave Technology*, vol. 40, no. 9, pp. 3072–3083, 2022.

[26] W. Liu, H. Wang, X. Shen, and I. W. Tsang, "The emerging trends of multi-label learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 11, pp. 7955–7974, 2021.

[27] C.-K. Yeh, W.-C. Wu, W.-J. Ko, and Y.-C. F. Wang, "Learning deep latent space for multi-label classification," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 31, no. 1, 2017.

[28] M.-L. Zhang and Z.-H. Zhou, "Multilabel neural networks with applications to functional genomics and text categorization," *IEEE transactions on Knowledge and Data Engineering*, vol. 18, no. 10, pp. 1338–1351, 2006.

[29] S. Wu, H. R. Zhang, and C. Ré, "Understanding and improving information transfer in multi-task learning," in *International Conference on Learning Representations*, 2020.

[30] J. Fang, Y. Li, P. N. Ji, and T. Wang, "Drone detection and localization using enhanced fiber-optic acoustic sensor and distributed acoustic sensing technology," *Journal of Lightwave Technology*, vol. 41, no. 3, pp. 822–831, 2022.

[31] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," in *International Conference on Learning Representations*, 2020.

[32] V. Dumont, V. R. Tribaldos, J. Ajo-Franklin, and K. Wu, "Deep learning on real geophysical data: A case study for distributed acoustic sensing research," *arXiv preprint arXiv:2010.07842*, 2020.

[33] T. Wang, M.-F. Huang, S. Han, and C. Narisetty, "Employing fiber sensing and on-premise AI solutions for cable safety protection over telecom infrastructure," in *Optical Fiber Communication Conference*. Optica Publishing Group, 2022, pp. Th3G–1.

APPENDIX A
IMPLEMENTATION DETAILS FOR INTRUSION DETECTION

The implementation details including model architectures and hyperparameters used during training in Section III is provided in Table V and VI.

TABLE V
NEURAL NETWORK ARCHITECTURES

Layer Configuration
Conv2D (1, 32, 3) + ReLU + MaxPool2d (2, 2)
Conv2D (32, 64, 3) + ReLU + MaxPool2d (2, 2)
FC1(4224, 512) + ReLU
DLE: FC-head (512, 2) * 4 heads OLE: FC-head (512, 12)

TABLE VI
HYPERPARAMETERS FOR MODEL TRAINING

Parameters	Choice
Batch size	128
Number of epochs	100
Optimizer	Adam
Learning rate	5e-5
Weight decay	1e-5
Dropout rate	0.1
LR scheduler steps	[40, 70, 90]
LR decay factor	0.1

TABLE VII
NUMBER OF SAMPLES PER SETTING FOR TABLE I

	Snow2wet	Wet2snow	Walk2run	Run2walk
Train set	2200	2075	2100	2175
Test Set	2075	2200	2175	2100

TABLE VIII
NUMBER OF SAMPLES PER SETTING FOR EACH TASK IN TABLE II.

Ratio	Set	Task 1	Task 2	Task 3	Task 4
75%	train	236:240	240:236	82:82	314:326
	test	79:80	80:79	28:28	106:109
50%	train	157:160	160:157	55:55	210:217
	test	158:160	160:158	55:55	210:218
30%	train	94:96	96:94	33:33	126:130
	test	221:224	224:221	77:77	294:305
20%	train	63:64	64:63	22:22	84:87
	test	252:256	256:252	88:88	336:348

Task 1, To_left: To_right; Task 2, near: far; Task 3: go_further: go_closer; Task 4: walk: run

APPENDIX B
IMPLEMENTATION DETAILS FOR IMPULSIVE EVENT CLASSIFICATION

We provide further implementation details of each baseline methods discussed in Section IV. Table. IX lists a few design choices about the signal processing pipeline used and model training hyperparameters. Table. X $\sim$ Table. XIV detail signal processing procedures followed by the neural network architectures. Unless otherwise specified, default parameters from standard Python libraries are used.

TABLE IX
HYPERPARAMETERS FOR ML AND SIGNAL PROCESSING.

Parameters	Choice
Batch size	32
Number of epochs	1024
Optimizer	Adam
Learning rate	1e-4
Weight decay	1e-4
Balanced sampling	False
Dropout rate	0.1
Butterworth filter type	5th order, digital
Cut-off frequency	200 Hz, high-pass
Mel spectrogram nFFT, hop length	1000, 200
STFT/MFCC nFFT, overlap	512, 307
Number of cepstrum	13

TABLE X
IMPLEMENTATION DETAILS: WAVEFORM (CONVNET)

Preprocessing
High pass filtering
Layer Configuration
Conv1D (1, 8, 256) + BN + ReLU + MaxPool1D (4) + Dropout
Conv1D (8, 16, 128) + BN + ReLU + MaxPool1D (4) + Dropout
Conv1D (16, 32, 64) + BN + ReLU + MaxPool1D (4) + Dropout
Conv1D (32, 64, 32) + BN + ReLU + MaxPool1D (4) + Dropout
Conv1D (64, 128, 16) + BN + ReLU + MaxPool1D (4) + Dropout
Conv1D (128, 256, 8) + BN + ReLU + MaxPool1D (4) + Dropout
FC1 (256, 64) + ReLU + Dropout + FC2 (64, 8)

TABLE XI
IMPLEMENTATION DETAILS: STFT (CONVNET)

Preprocessing
Short-time Fourier transform
Layer Configuration
Conv2D (1, 32, 3) + BN + ReLU + LPPool2d (2, 2)
Conv2D (32, 64, 3) + BN + ReLU + LPPool2d (2, 2)
Conv2D (64, 128, 3) + BN + ReLU + LPPool2d (2, 2)
Conv2D (128, 256, 3) + BN + ReLU + LPPool2d (2, 2)
FC1 (12288, 64) + BN + ReLU + FC2 (64, 8)

TABLE XII
IMPLEMENTATION DETAILS: STFT/MEL SPECTROGRAM (VIT)

Preprocessing
Short-time Fourier transform or Log Mel Spectrogram
Layer Configuration
Image Size: 128 x 128 x 1
Patch Size: 64
Embedding Size: 512
Number of Attention Heads: 32
Number of Transformer Layers: 3
Hidden Size: 64

TABLE XIII
IMPLEMENTATION DETAILS: MEL SPECTROGRAM (CONVNET)

Preprocessing
Log Mel Spectrogram
Layer Configuration
Conv2D (1, 32, 3) + BN + ReLU + LPPool2d (2, 2)
Conv2D (32, 64, 3) + BN + ReLU + LPPool2d (2, 2)
Conv2D (64, 128, 3) + BN + ReLU + LPPool2d (2, 2)
Conv2D (128, 256, 3) + BN + ReLU + LPPool2d (2, 2)
FC1(16384, 64) + BN + ReLU + FC2 (64, 8)

TABLE XIV
IMPLEMENTATION DETAILS: MFCC (CONVNET)

Preprocessing
High pass filtering, MFCC
Layer Configuration
Conv2D (1, 6, 2) + BN + ReLU + LPPool2d (2, 2) + Dropout
Conv2D (6, 8, 2) + BN + ReLU + LPPool2D (2, 2) + Dropout
Conv2D (8, 10, 2) + BN + ReLU + LPPool2D (2, 1) + Dropout
FC1(230, 24) + BN + ReLU + Dropout + FC2 (24, 8)