Research Article | Vol. 29, No. 3 / 1 February 2021 / Optics Express 3269
Optics EXPRESS

Pattern recognition in distributed fiber-optic acoustic sensor using an intensity and phase stacked convolutional neural network with data augmentation

HUAN WU,[1] BIN ZHOU,[2] KUN ZHU,[1,*] CHAO SHANG,[3] HWA-YAW TAM,[4] AND CHAO LU[1,5]

[1]*Department of Electronic and Information Engineering, The Hong Kong Polytechnic University, Kowloon, Hong Kong*
[2]*Guangdong Provincial Key Laboratory of Optical Information Materials and Technology, South China Academy of Advanced Optoelectronics, South China Normal University, Guangzhou, China*
[3]*Key Laboratory of Luminescence and Optical Information, Beijing Jiaotong University, Beijing, China*
[4]*Photonics Research Centre, Department of Electrical Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong*
[5]*The Hong Kong Polytechnic University, Shenzhen Research Institute, 518057, Shenzhen, China*
[*]*kenny.kun.zhu@gmail.com*

Abstract: Distributed acoustic sensors (DASs) have the capability of registering faint vibrations with high spatial resolution along the sensing fiber. Advanced algorithms are important for DAS in many applications since they can help extract and classify the unique signatures of different types of vibration events. Deep convolutional neural networks (CNNs), which have powerful spectro-temporal feature learning capability, are well suited for event classification in DAS. Generally, these data-driven methods are highly dependent on the availability of large quantities of training data for learning a mapping from input to output. In this work, to fully utilize the collected information and maximize the power of CNNs, we propose a method to enlarge the useful dataset for CNNs from two aspects. First, we propose an intensity and phase stacked CNN (IP-CNN) to utilize both the intensity and phase information from a DAS with coherent detection. Second, we propose to use data augmentation to further increase the training dataset size. The influence of different data augmentation methods on the performance of the proposed CNN architecture is thoroughly investigated. The experimental results show that the proposed IP-CNN with data augmentation produces a classification accuracy of 88.2% on our DAS dataset with 1km sensing length. This indicates that the usage of both intensity and phase information together with the enlarged training dataset after data augmentation can greatly improve the classification accuracy, which is useful for DAS pattern recognition in real applications.

© 2021 Optical Society of America under the terms of the OSA Open Access Publishing Agreement

1. Introduction

Distributed acoustic sensors (DASs) are increasingly receiving great attention from both industry and academia in recent years due to their great potential for third-party intrusion detection [1], seismic detection [2], and pipeline monitoring [3]. The working principle of DAS is based on the phase-sensitive OTDR that utilizes the interference effect of Rayleigh backscattering of different scatters within the pulse width. The DAS was realized with the direct detection scheme at the very beginning, which can only obtain the intensity information. For this old-fashioned DAS, the relationship between the strain change and the collected intensity variation is often unpredictable. In recent years, the coherent detection scheme was adopted to retrieve both the intensity and phase information, and the relationship between the recovered phase and strain change is found to be linear. Apart from the advancement of detection schemes, the sensing range

#416537 https://doi.org/10.1364/OE.416537
Journal © 2021 Received 8 Dec 2020; revised 9 Jan 2021; accepted 9 Jan 2021; published 19 Jan 2021

Optics EXPRESS

and spatial resolution have also been improved significantly. The spatial resolution has been reduced to a sub-meter level [4] and the sensing distance has reached hundreds of kilometers [5]. When the spatial resolution is approaching meter level, that is equivalent to interrogating nearly 1,000 interferometers for every kilometer of sensing fiber. Transmitting and processing this amount of data volume with low latency becomes a crucial challenge in DAS. Two-stage vibration detection and recognition method could greatly ease the burden on digital signal processing. The threshold-based technique can be used to quickly locate the vibration position and then only the signals at the vibration positions will be sent for further pattern recognition procedure. We proposed an algorithm to locate the vibration positions with a low false alarm rate in [6]. In this work, we focus on the event classification for the detected vibrations.

Data-driven pattern recognition based on machine learning and deep learning is a recent trend in DAS [7–15]. Q. Sun et al. proposed to use ten features extracted from morphology to classify walking, digging and vehicle passing based on the relevance vector machine (RVM) [7]. To reduce the ambient noises induced false alarms, H. Wu et al. used wavelet decomposition to enhance the signal-to-noise ratio (SNR) [8]. Wavelet coefficients were used as features of an artificial neural network (ANN) to classify no intrusion, human intrusion events, and hand-clapping interferences. A support vector machine (SVM) classifier using three energy-related features were also utilized to classify five events in third-party intrusion scenarios [9]. To reduce the number of data traces in pattern recognition, M. Adeel et al. proposed an impact-based feature extraction method to classify five lab-simulated events with the boosting and bagging method [10]. Some researchers also proposed to use temporal-spatial data with simple data preprocessing as feature vectors [11]. Five events including background, walking, jumping, beating with a shovel and digging with a shovel were classified by the 2D convolutional neural network (2D CNN). To improve computational efficiency, H. Wu et al. proposed to directly use raw or denoised data without any transformation as the feature [12]. Five field test events including background noises, manual diggings, mechanical excavations, traffic noises, and factory noises were classified by the 1D CNN. Though the above works [7–12] all achieved high classification accuracy based on different pattern recognition methods, their vibration events were very limited and most of the events were generated in the lab environment. A field test for high-speed railway intrusion detection was carried out in [13], in which three events were classified by a combination of CNN and long short-term memory network (LSTM) with 69.3% accuracy. J. Tejedor et al. in [14–15] presented fully realistic activities along a pipeline. 8 and 45 machine + activity pairs were classified by Gaussian mixture model (GMM) with short-time Fourier transform (STFT) as the feature vector. Due to the complex environmental perturbations and more events, the classification accuracy was lower than 50%. Almost all these above works only utilize the intensity information or phase information of DAS, which wastes half of the collected information since both the intensity and phase are available in the coherent DAS scheme. Moreover, the classification performances of the pattern recognition methods rely on the quantities and quality of training data to learn a mapping from input feature to output classes. The datasets in [7–15] are not public, which makes it hard to compare the event classification performances of different algorithms.

To accurately classify the vibration event, in this work, we propose a convolutional neural network (CNN) that fully utilizes the precious labeled data. Three main contributions have been done in this work. Firstly, we build a public dataset including both the intensity and phase information collected from a coherent DAS system. The DAS1K (data collected from a DAS system with 1 km sensing fiber) dataset consists of 501 samples in 10 classes with enough pattern variations for each class. Secondly, we propose an intensity and phase stacked CNN (IP-CNN) to fully utilize the collected information to enhance the classification accuracy. Thirdly, we propose to use data augmentation to overcome the problem of data scarcity and explore different types of augmentation methods and their influences on the model's performances in DAS. We show that

Optics EXPRESS

the proposed IP-CNN architecture, in combination with data augmentation yields an accuracy of 88.2% on our 10-class DAS1K dataset.

The organization of this paper is as follows: Section 2 describes the data collection including the DAS experimental setup and the description of DAS1K dataset. Section 3 gives the proposed IP-CNN model with data augmentation method for classification of the DAS1K dataset. Section 4 shows the experimental results and performance comparisons. Finally, the conclusion is drawn in Section 5.

2. Data collection

2.1. Experimental setup of DAS

The experimental setup of the DAS system is shown in Fig. 1. A narrow-linewidth laser (NLL, Connet CoSF-SC-1550-M) with 5 kHz linewidth is used as the optical source. The output of the NLL is split into two branches with a 95:5 coupler. In the upper branch, the continuous wave (CW) is firstly modulated by an acousto-optic modulator (AOM, G&H T-M080-0.4C2J-3-F2S) to generate an optical pulse sequence with an 80 MHz frequency shift. An erbium-doped fiber amplifier (EDFA, Amonics AEDFA-18-M-FA) is used to boost the power of the optical pulses. The amplified spontaneous emission (ASE) noise is filtered out by an optical bandpass filter with 0.8 nm passband. Then the optical pulses are launched into the fiber-under-test (FUT). The Rayleigh backscattering (RBS) traces are further amplified by another EDFA and followed by another 0.8 nm bandpass optical filter. Finally, the RBS light is combined with the optical local oscillator from the lower branch and launched into a balanced photo-detector (BPD, Thorlabs PDB465C). The data are collected by a 14-bit data acquisition card (DAQ) with a sampling rate of 250 MS/s. The intensity and phase information are demodulated by Hilbert transform and down-sampled to 15.625 MS/s, then the distributed phases at all sensing locations are unwrapped independently to restore the acoustic waveforms. The Hilbert transform and the phase unwrapping algorithms are embedded in the field programmable gate array (FPGA) chip (Zynq-7100, Xilinx Inc.) of the DAQ. Benefiting from the parallel computing ability of the FPGA, the distributed acoustic waveforms can be extracted in real-time.

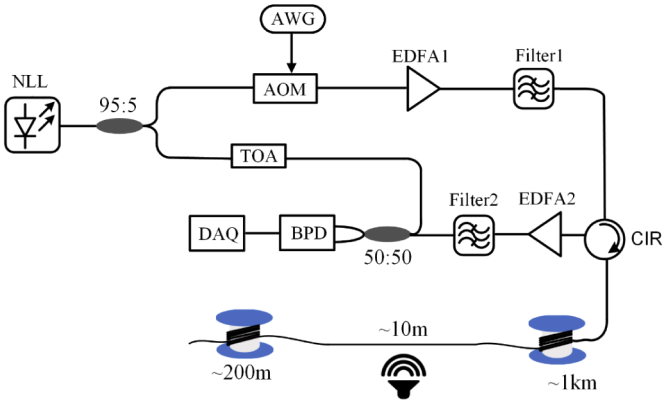

Fig. 1. Experimental setup of the DAS. NLL: narrow-linewidth laser; AOM: acousto-optic modulator; AWG: arbitrary waveform generator; EDFA: erbium-doped fiber amplifier; TOA: tunable optical attenuator; CIR: circulator; BPD: balanced photo-detector; DAQ: data acquisition card.

The sensing fiber is G.652 single-mode fiber (SMF) and the total length is around 1.2 km. The pulse repetition rate is set to 10 kHz and the pulse width is 100 ns. The vibration is applied through a speaker with 10 m fiber stick on it around 1010 m. The stacked differential intensity

Optics EXPRESS

and phase traces are shown in Figs. 2(a1) and 2(b1). The two-dimensional time-space images of 4-second consecutive differential intensity and phase traces along the FUT are shown in Figs. 2(a2) and 2(b2). The vibration position at 1010 m can be identified from both the intensity and phase information.

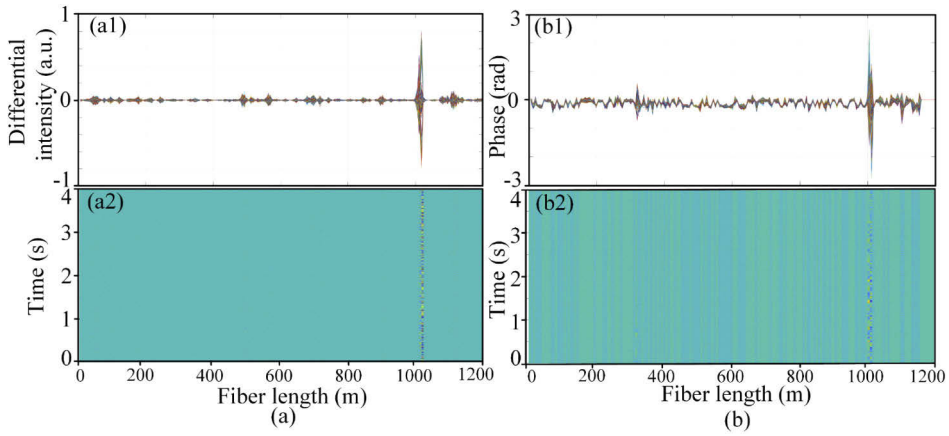

Fig. 2. (a1) Stacked differential intensity traces; (a2) two-dimensional time-space image of 4-second consecutive intensity differential traces along the FUT; (b1) Stacked phase traces; (b2) two-dimensional time-space image of 4-second consecutive phase traces along the FUT.

2.2. DAS1 K dataset description

If the events of training data are created manually, it is labor-intensive and expensive to generate a large signal variation in each class. For example, in the drilling class, the drilling machine should change with different brands and models, different drilling objects may also be considered, and the applied force in drilling should also vary to collect different patterns of drilling. To build a dataset including more classes and have enough signal variations in each class, here we used a speaker to generate the vibration events. The original audios were played by the speaker including 3 top-level groups: mechanical activity, human activity, and nature events. In the mechanical activity, it includes car horn (CH), drilling (DL), jackhammer (JH) and welding (WD). In the human activity, it includes footsteps (FS), hand hammer (HH), hand saw (HS) and shoveling. In nature events, it includes raining (RA) and thunderstorm (TS). To build a dataset including adequate classes with enough pattern variations in each class, the original audio dataset should be field-recordings and the dataset should be sufficiently large and vary in terms of audios and recording conditions. To accomplish this goal, we downloaded the original audios from FreeSound [16], which is an online sound repository under a creative commons license. The original audio dataset contains 501 labeled audio clips (<=5 s) from 10 classes. Then we played the 501 audio clips on the speaker. Both intensity and phase information are collected by the DAS system. Figure 3 and Fig. 4 depict the differential intensity and phase information of 10 representative waveforms from each class of the dataset. The detailed information about the DAS1 K dataset is shown in Table 1. DAS1K dataset includes 10 classes. In each class, it has 47 to 54 samples. The duration of each sample varies from 0.9 s to 4.95 s, with a total time of 1816.6 s.

From a visual inspection, it is trivial to visualize the differences between some of the classes. Particularly, the waveforms for repetitive sounds for footsteps, hand hammer, and handsaw are different in period or shape. However, it is difficult to distinguish among car horn, drilling, jackhammer and thunderstorm from the waveform. There are many signal representations can be

Optics EXPRESS

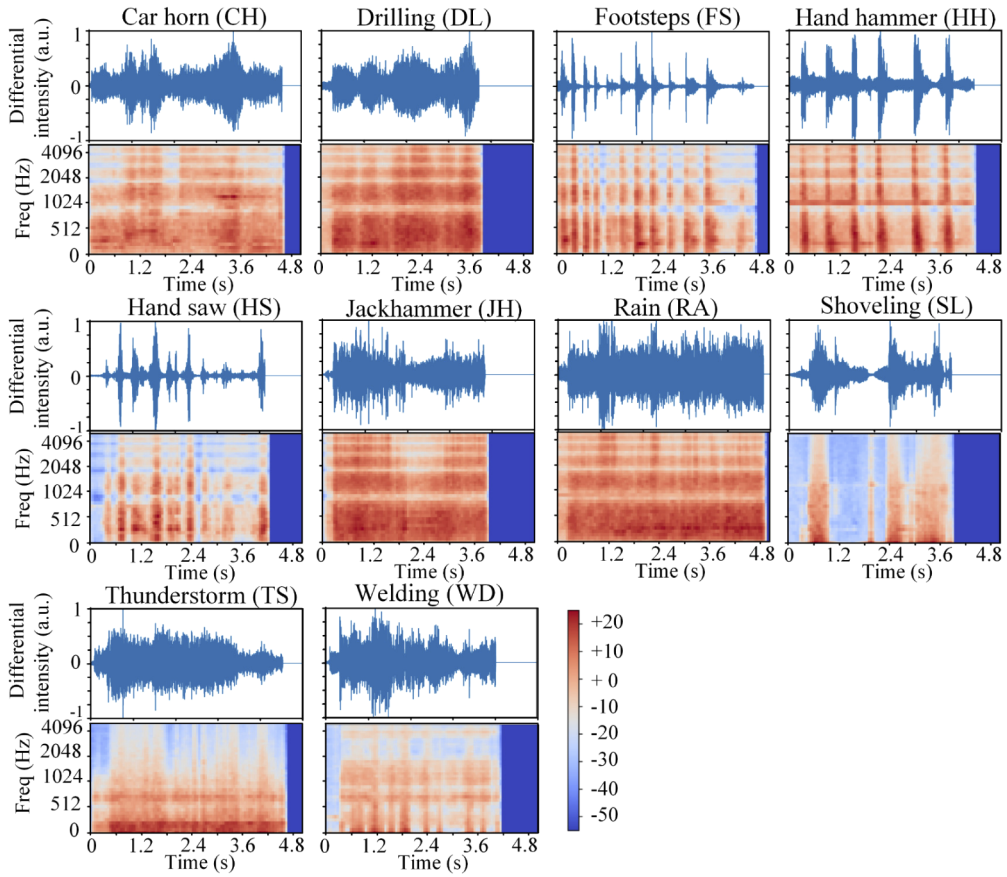

Fig. 3. Differential intensity waveforms at 1010 m and their log-scaled Mel-spectrograms.

Table 1. DAS1 K dataset description.

Number	Class	Abbreviation	Number of samples	Duration of each sample
0	car horn	CH	48	1.85-4.95 s
1	drilling	DL	50	0.90-4.34 s
2	footsteps	FS	50	2.05-4.65 s
3	hand hammer	HH	54	2.55-4.35 s
4	hand saw	HS	53	3.15-4.75 s
5	jackhammer	JH	47	0.90-4.14 s
6	raining	RA	49	3.35-4.83 s
7	shoveling	SL	53	1.15-4.04 s
8	thunderstorm	TS	48	2.85-4.74 s
9	welding	WD	49	3.35-4.04 s
-	Total	-	501	1816.6 s

extracted from the time-series signal, it has been noted that time-frequency representations are especially useful as learning features due to the non-stationary dynamic nature of the waveforms. In this work, we used log-scaled Mel-spectrogram as the feature vector. STFT spectrogram uses a linear spaced frequency scale, whereas log-scaled Mel-spectrogram uses a quasi-logarithmic

Optics EXPRESS

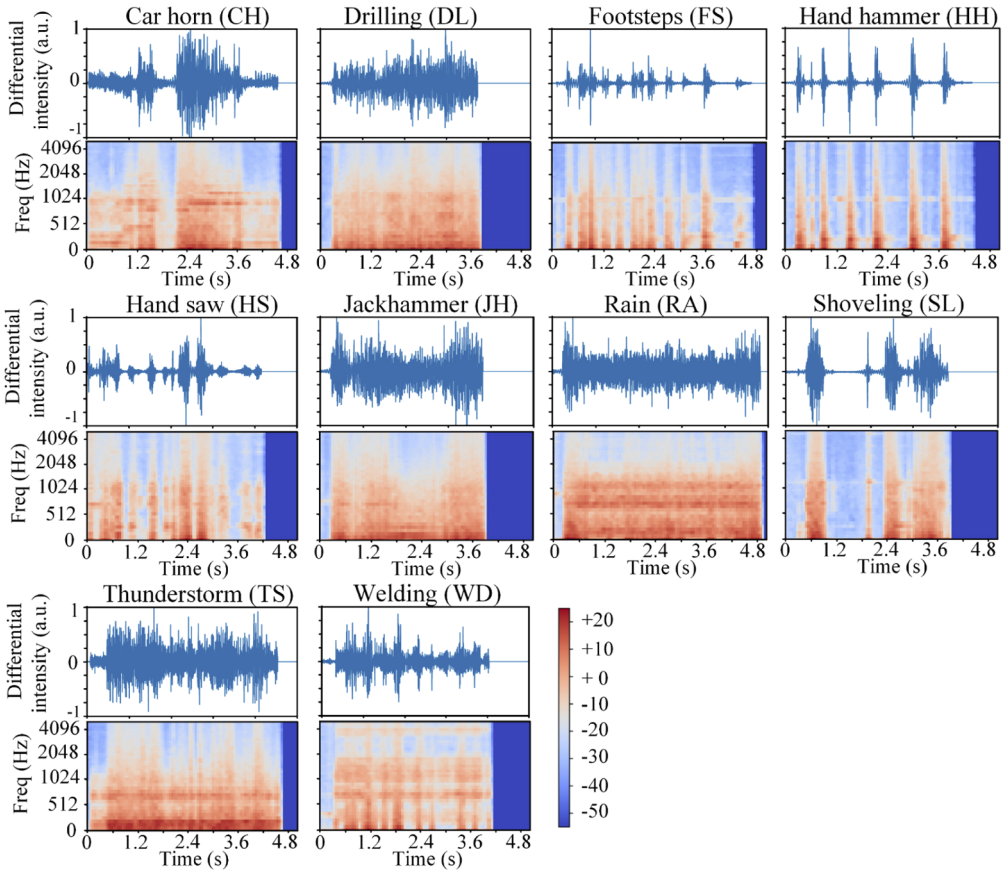

Fig. 4. Phase waveforms at 1010 m and their log-scaled Mel-spectrograms.

spaced frequency scale, which more resembles the human auditory system. Conversion between Hertz (f) and Mel (m) is defined as $m = 2595log_{10}(1 + f/700)$. The log-scaled Mel-spectrogram of each waveform is also shown in Fig. 3 and Fig. 4. We use Librosa [17] to extract log-scaled Mel-spectrograms with 40 components (bands) covering the frequency range (0-5000 Hz), using a window size of 102.4 ms (1024 samples at 10 kHz).

3. Method

3.1. Architecture of IP-CNN

The CNN architecture proposed in this study consists of four 2D convolutional layers (Conv2D) and five pooling layers, followed by a dense layer and Softmax function. Each 2D convolutional layer is followed by a ReLU activation function to provide non-linearity to the model. The convolutional layers extract features automatically from the 2D Mel-spectrograms, then the 2D max pooling layers with shift-invariance property highlight the features from the 2D convolutional layers. Note that we choose a small kernel size (3,3) in Conv2D to allow the CNN to learn localized features. By cascading multiple Conv2D and max pooling layers, spectro-temporal signatures that indicate the presence/absence of different classes are learned. The dense layer maps the feature vector to the final 10 predicted classes with Softmax function. The details of the network are shown in Table 2, where IFM and OFM represent input feature map and output

Optics EXPRESS

feature map of a specific layer, respectively. The total parameter number of CNN is 42442, which is quite lightweight and easy to deploy in a real-time scenario.

Table 2. The architecture of 2D CNN model.

Layer	Name	Kernel size	Number of IFM/OFM	Output shape	Number of parameters
1	Input	-	-	(40,98)	-
2	Conv2D	(3,3)	1/16	(38,96)	160
3	MaxPooling2D	(2,2)	16/16	(19,48)	0
4	Conv2D	(3,3)	16/32	(17,46)	4640
5	MaxPooling2D	(2,2)	32/32	(8,23)	0
6	Conv2D	(3,3)	32/64	(6,21)	18496
7	MaxPooling2D	(2,2)	64/64	(3,10)	0
8	Conv2D	(3,3)	64/64	(3,10)	18496
9	MaxPooling2D	(2,2)	64/64	(1,5)	0
10	Global average pooling 2D	-	-	(1,64)	0
11	Dense	(1,1)	-	(None,10)	650

Since both the intensity signal and phase signal are extracted, we propose an intensity and phase stacked model (IP-CNN) to simultaneously utilize the intensity and phase information. In Fig. 5, the 2D Mel-spectrograms extracted from intensity and phase are processed by I-CNN and P-CNN independently, where the I-CNN and P-CNN have the same architecture as shown in Table 2, but without the final dense layer. The output of each CNN (i.e., a 64-D feature vector) is concatenated to form a 128-D feature vector. This feature vector passes a dense layer and finally be projected to 10 predicted classes. The softmax function $e^{X^T w_j} / \sum_{k=1}^{10} e^{X^T w_k}$ is used to transform the feature vector X into 10 probabilities that summed to 1, where w_j is the weight vector corresponding to class j. As the example illustrated in Fig. 5, the softmax has assigned 90.7% probability to the CH class, this could give the indication that the CNN has high confidence the input should be classified to the CH class. The WD class only has a probability of 0.01%, suggesting that the CNN has high confidence that input does not belong to the WD class. The argmax of the 10 probabilities will be the class label.

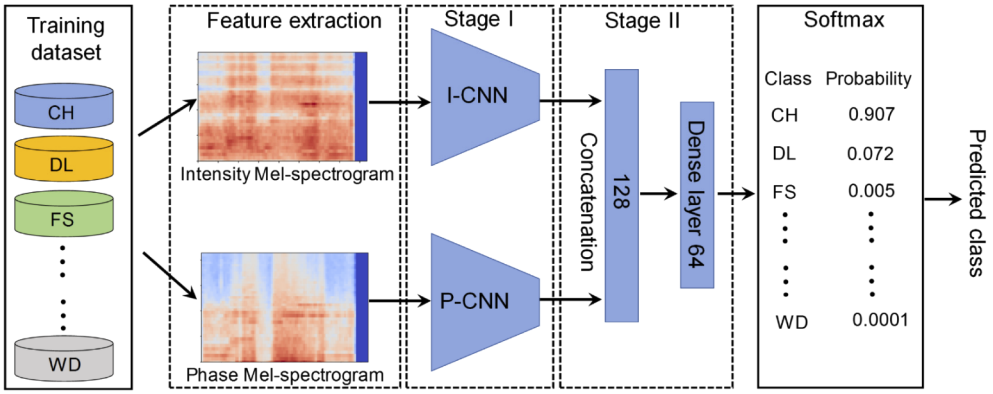

Fig. 5. Training flow of the proposed IP-CNN model.

Optics EXPRESS

For training, the model optimizes categorical cross-entropy loss with L2 regularization via Adam optimizer at 0.001 learning rate. Each batch consists of 128 randomly selected Mel-spectrograms from the training data. Dropout [18] is applied after each of the convolutional layers to reduce overfitting, with a probability of 0.5. For I-CNN and P-CNN, the model is trained for 3000 epochs with early stopping criteria [19]. For IP-CNN, we use a two-stage training strategy. First, we train the I-CNN and P-CNN independently and fix their weights from layer 1 to 10. Then we train the stacked model for another 500 epochs, where only the weights of the final dense layer in IP-CNN are learned in this process. We use Keras and Tensorflow to implement the model on a workstation equipped with Nvidia RTX2070 and AMD Ryzen 7 2700X.

3.2. Data augmentation

CNNs have a high model capacity to learn discriminative spectro-temporal features for classification. However, the scarcity of labeled data impedes the performance of CNNs. Data augmentation is a common strategy to create new training samples by tweaking small factors on the original samples [20]. By changing these small factors, we can increase the number of the dataset by several times. It has been proved that data augmentation is very useful for small datasets, which not only improves the generalization capability of the model but also avoids overfitting [21]. In this work, we performed two different data augmentations, resulting in seven augmentation sets. Each augmentation is directly applied to both the differential intensity and phase waveforms before converting it to the Mel-spectrogram. Two data augmentation methods are described below:

(1) **Stretching** (ST): slow down or speed up both the intensity and phase signals. Each signal is stretched by four factors: {0.8, 0.9, 1.1,1.2}.

(2) **Shifting** (SH): circular shift both the intensity and phase signals by 3 different percentages of the signal length: {0.2, 0.4, 0.6}.

Stretching can change the frequency distribution of the original signal which could provide more variations for each sample in a specific class. Shifting generates more similar samples as the original signal for each class, but with slightly different start and end waveforms. After applying the two data augmentation methods, the whole training dataset is expanded to 8x of its original size.

4. Experimental results and discussions

In this work, we mainly evaluate different CNN models with classification accuracy. The dataset is split into 5 folds evenly and all the models are evaluated with 5-fold cross validation to test their generalization capability. Each time the CNN models are trained with 4 folds and the remaining fold is used for testing. I-CNN and P-CNN without data augmentation are trained first. Figures 6(a) and 6(b) show the training and testing accuracy and the corresponding categorical cross-entropy loss curves of the I-CNN and P-CNN, respectively. For I-CNN, after about 1000 epochs, the loss of the test set converges at 0.91. Although the training loss curve is still decreasing, according to the early stopping criteria [19], the model is overfitted after 1000 epochs. Under this condition, the classification accuracy is about 73% after 1000 epochs. The loss curve of P-CNN is similar to I-CNN, which converges at 0.88 after 1200 epochs, and the classification accuracy converges to 77%.

To evaluate the effectiveness of data augmentation, I-CNN and P-CNN are also trained with different augmentation methods, respectively. For each model with different data augmentations, we save all the model parameters of the last 100 epochs and the evaluation results are summarized in Fig. 7 with boxplot [22]. This approach captures model generalization capability more comprehensively while avoiding unstable results. To the left of the first dashed line, we show

Optics EXPRESS

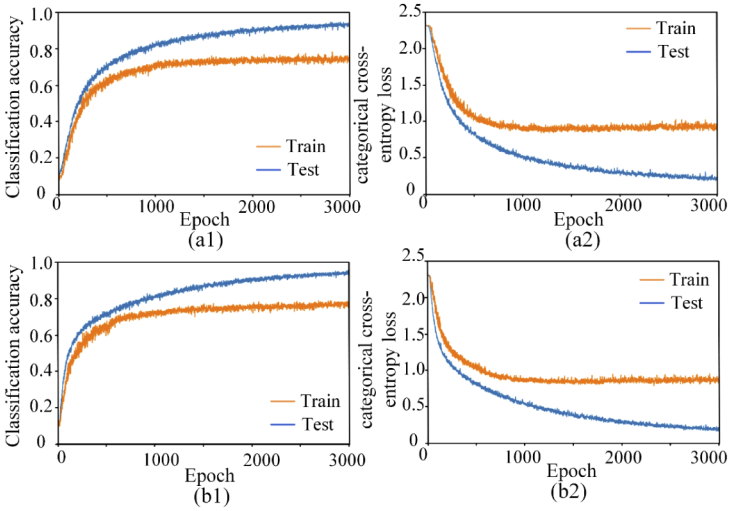

Fig. 6. (a) Train and test (a1) classification accuracy and (a2) categorical cross-entropy loss curves of I-CNN without augmentation; (b) train and test (b1) classification accuracy and (b2) categorical cross-entropy loss curves of P-CNN without augmentation.

the performance of the I-CNN trained on the original dataset without augmentation. We see that the median value of the classification accuracy is 73% (represented in the red line). For comparison, we also provide the performance of I-CNN trained with ST, SH, and ST&SH, with median accuracy improved by 5.4%, 9.4%, and 9.6%, respectively. The performance of P-CNN is also listed in the middle of Fig. 7 between the two dashed lines. The median accuracy of P-CNN without augmentation is 77%. With ST, SH, and ST&SH augmentation, the median accuracy is improved by 3.2%, 6.4%, 7.4%, respectively. This indicates that the single CNN models trained with augmented data can learn the better non-linear mapping function from input to the output and therefore yield higher classification accuracy on unseen data. It can also be observed that I-CNN and P-CNN with both ST&SH augmentation have the best performance, and the models trained with SH augmentation has higher classification accuracy improvements compared with ST augmentation. To gain further insight into the influence of the augmentation methods, accuracy changes of each class are investigated below.

We display the classification accuracy difference of each class in Fig. 8. The accuracy difference is calculated by subtracting the accuracy of I-CNN/P-CNN with data augmentation from that of training with the original dataset. The overall classification accuracy is shown at the bottom of the figure. Except HH w/ST, SL w/SH in I-CNN and HH w/SH&ST, SL w/SH&ST in P-CNN, all the other classes are affected by both types of data augmentations, and the accuracies of most of the classes are improved. In particular, the classification accuracy of jackhammer (JH) is increased by 27.7% for I-CNN with SH augmentation and 19.1% for P-CNN with ST augmentation. In I-CNN, the drilling (DL) and shoveling (SL) are negatively affected by the ST augmentation. In P-CNN, the ST augmentation also deteriorates the classification performance of hand hammer (HH), hand saw (HS), shoveling (SL), and welding (WD). Stretching the waveforms in the time domain changes representative frequencies of the event and this may not be suitable for all the classes. This could be improved by class-specific augmentation during the training phase – different augmentations are selectively applied on different classes. We will investigate this in our future work. Even though augmentations have some negative impacts on some classes, eight classes in I-CNN and six classes in P-CNN benefit from applying any kind of data augmentation. The

Optics EXPRESS

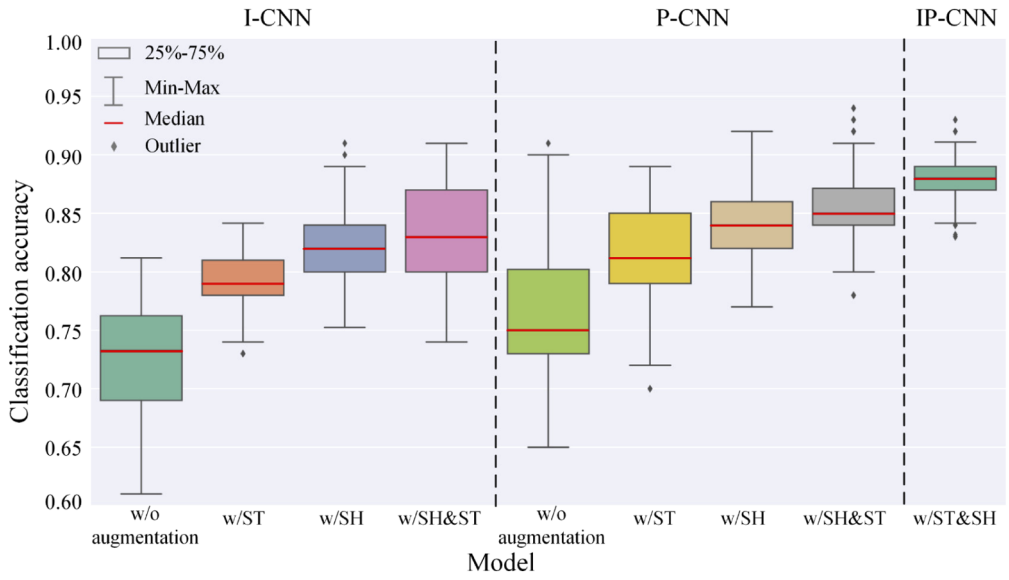

Fig. 7. Boxplot of classification accuracy CNN models.

overall classification accuracies of both I-CNN and P-CNN are enhanced most by a combination of the two augmentations.

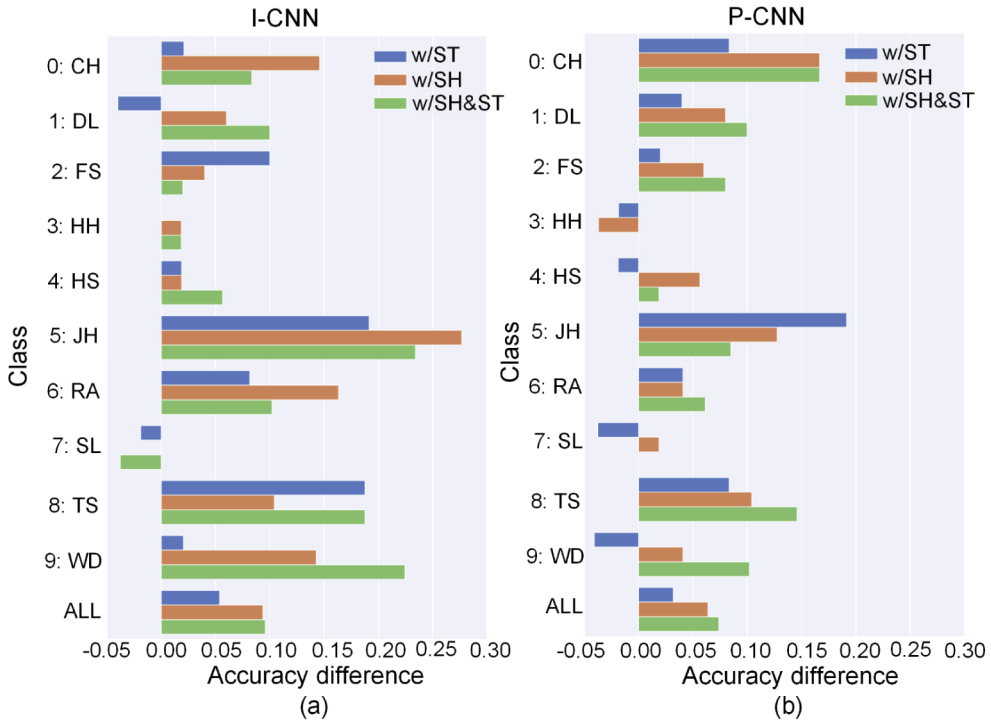

Fig. 8. Accuracy difference for each class with different augmentation applied on (a) I-CNN, (b) P-CNN.

Optics EXPRESS

After investigating the performances of data augmentation, we stack the I-CNN and P-CNN, as described in Fig. 5, and train the IP-CNN with ST&SH augmentation. The performance is shown on the right side of Fig. 7. The median accuracy is further improved to 88.2%, exhibiting accuracy enhancement of 5.2% and 3.4% compared with I-CNN and P-CNN with SH&ST augmentation. This verifies the effectiveness of utilizing both the intensity and phase information for event classification. Since intensity and phase information are used simultaneously in the IP-CNN model, the learned features are more comprehensive to represent each class, which cannot be achieved with a single model like I-CNN or P-CNN.

To give a better idea of what the CNN model is getting right and what types of errors it is making, the confusion matrices are calculated for I-CNN, I-CNN with ST&SH augmentation, P-CNN, P-CNN with ST&SH augmentation, and IP-CNN with ST&SH augmentation, as given in Fig. 9. A confusion matrix $C[M, M]$ is a summary of prediction results on a classification problem. The number of correct and incorrect predictions are summarized with count values and broke down by each class. The element $c_{i,j}$ in the matrix represents the number of samples in class i predicted into class j. Therefore, the values on principal diagonal positions ($i = j$) represent the number of correct predictions while the values on off-principal diagonal positions ($i \neq j$) represent the number of incorrect predictions. The total number of correct predictions divided by the total number of test samples will give classification accuracy. Apart from the overall accuracy, non-zero non-diagonal cells in the matrix contain classification errors, i.e., cases when the true class and predicted class don't match. As shown in Figs. 9(a)–9(d), different classes perform differently with the application of augmentation, but it reduces the confusion in most classes. For example, the number of accurately classified WD samples improves from 31 to 42 in I-CNN. As shown in Figs. 9(a) and 9(b), misclassifying WD to CH and DL are eliminated. The number of misclassified WD samples to RA also reduces from 6 to 2. In P-CNN, the number of correctly classified CH samples increases from 37 to 42 after data augmentation. The P-CNN with data augmentation does not mistake WD for CH anymore. After stacking I-CNN and P-CNN, IP-CNN with data augmentation shows the highest classification accuracy on all classes except HH and HS in I-CNN with ST&SH augmentation.

Both classification accuracy in Fig. 7 and the confusion matrix in Fig. 9 indicate that IP-CNN with SH&ST augmentation has the best performance. To further illustrate the discrimination capability of five CNN models on each class, the features extracted at the final dense layer are embedded into the 2D plane using the t-Distributed Stochastic Neighbor Embedding (t-SNE) toolbox [23], as shown in Fig. 10. t-SNE is an unsupervised, non-linear technique primarily used for visualizing high-dimensional data. It gives an intuition of how the data is arranged in a high dimensional space. Here the main goal of t-SNE is to project the last 64-D feature vector to 2-D visualizable feature vector representation while preserving the information in the initial high-dimensional space. If two points were close in the initial high-dimensional space, they remain close in the resulting projection. If the points were far from each other, they should remain far in the target low-dimensional space, too. Figure 10(a) shows that the learned features in I-CNN are good at recognizing HH(3) and SL(7) since the samples in these two classes are very concentrated. It coincides well with the results shown in Fig. 9(a) that the classification accuracies of HH and SL reach 96.3% and 96.2% in I-CNN. After training the I-CNN with the dataset including SH&ST augmentation, the learned feature representation is illustrated in Fig. 10(b), which shows that the 10 classes are more separated and therefore easier to be classified. Compared with I-CNN, I-CNN with data augmentation achieves higher accuracy on all the classes, as shown in Fig. 9(b). Among these classes, JH(5) and WD(9) have the highest classification accuracy improvement as indicated by Fig. 9(b). The learned feature of P-CNN is shown in Fig. 10(c), from which we see that HH(3) and SL(7) are also very distinctive compared with other classes. This matches well with the results in Fig. 9(c). HH(3) has 98.2% accuracy and SL(7) has 94.3% accuracy in P-CNN. After training the P-CNN with the dataset with SH&ST

Optics EXPRESS

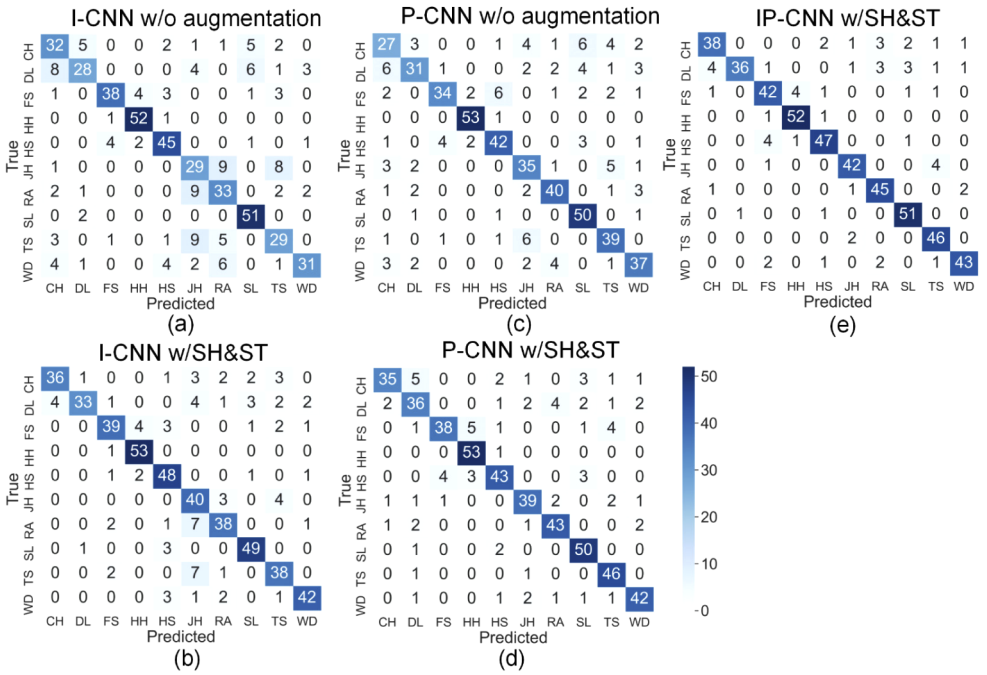

Fig. 9. Confusion matrices for the (a) I-CNN w/o augmentation, (b) I-CNN with SH&ST augmentation, (c) P-CNN w/o augmentation, (d) P-CNN with SH&ST augmentation, and (e) IP-CNN with SH&ST augmentation.

augmentation, the learned features are shown in Fig. 10(d), from which the boundary of each class is clearer than that in Fig. 10(c). Similarly, the classification accuracy of P-CNN with data augmentation has better performance on all the classes as given in Fig. 9(d). From the results shown in Figs. 9(a)–9(d) and Figs. 10(a)–10(d), it is safe to say that the data augmentation can greatly improve the learning capability of the CNN models, which is quite useful in DAS since the labeled data are very expensive and precious. Finally, the learned features of IP-CNN with data augmentation are shown in Fig. 10(e), the samples in each class are more centered and different classes are more separated. In I-CNN with SH&ST augmentation, samples in DL(1), JH(5), RA(6), TS(8), and WD(9) have many overlaps. In P-CNN with SH&ST augmentation, samples in DL(1), RA(6), WD(9) are also very close together. In contrast, the learned features in IP-CNN with SH&ST augmentation construct a map in which the separation among these ten classes is almost perfect. This further validates the effectiveness of utilizing both the intensity and phase information for pattern recognition in DAS.

The data processing time is also an important parameter to evaluate the performance of different models. The processing time including the feature extraction stage and classification stage are shown in Table 3. The most time-consuming parts are CNN model training and training set feature extraction. However, these two processes can be done offline. The testing set feature extraction and CNN model testing time used to classify 100 samples is only 831 ms, which is equivalent to 8.31 ms/sample. This shows that the proposed method is suitable for real-time scenarios.

Optics EXPRESS

Table 3. Processing time of different models.

Model	Processing time			
	feature extraction		CNN model	
	Training set	Testing set	Training	Testing
I-CNN w/o augmentation	1.5 s	380 ms	157.8 s	3.1 ms
I-CNN with SH&ST AUG	12.2 s	380 ms	563.2 s	3.2 ms
P-CNN w/o augmentation	1.5 s	380 ms	158.2 s	3.1 ms
P-CNN with SH&ST AUG	12.2 s	380 ms	562.6 s	3.2 ms
IP-CNN with SH&ST AUG	25.3 s	790 ms	1384.7 s	41 ms

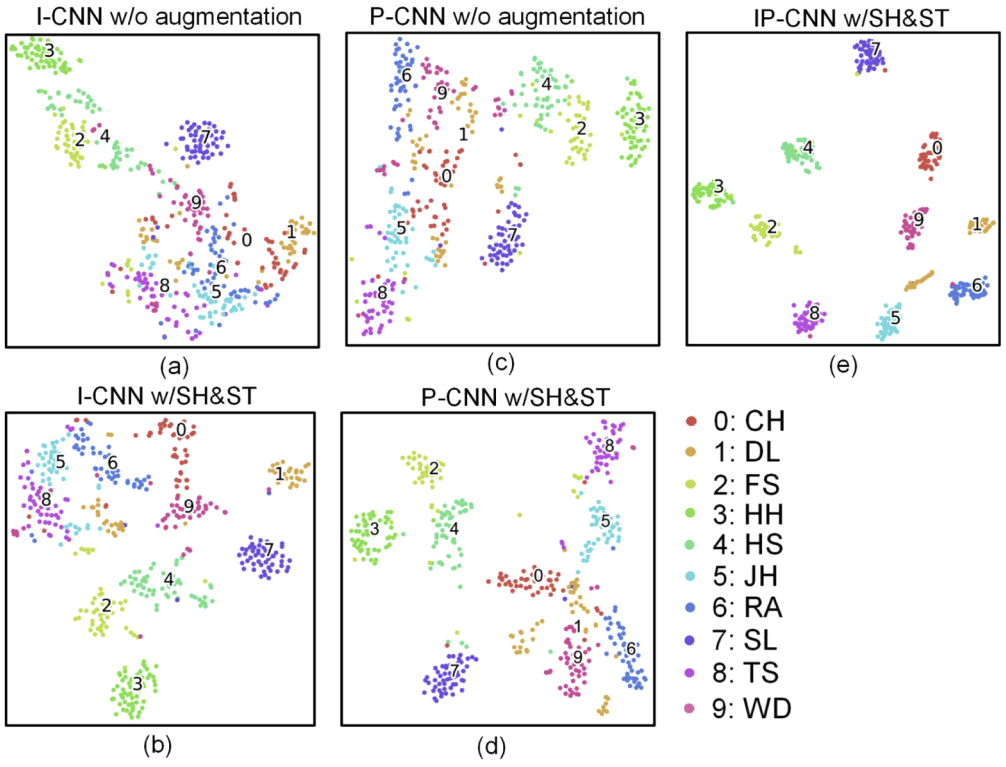

Fig. 10. Feature embedding from (a) I-CNN w/o augmentation, (b) I-CNN with SH&ST augmentation, (c) P-CNN w/o augmentation, (d) P-CNN with SH&ST augmentation, (e) IP-CNN with SH&ST augmentation.

Optics EXPRESS

5. Conclusion

In this work, we propose a real-time event classification method for coherent DAS. Using log-scaled Mel-spectrogram as the feature extractor, the intensity and phase stacked CNN (IP-CNN), in combination with two kinds of data augmentation methods (i.e., stretching and shifting in the time domain) achieved 88.2% classification accuracy on the 10-class DAS1K dataset. We show that the improved accuracy comes from the full utilization of both the intensity and phase information and augmented training dataset. Even though intensity and phase information in DAS convey similar information, different patterns can be learned by the I-CNN and P-CNN models to increase the prediction confidence on unseen data. With the help of data augmentation, the overall classification accuracy gets improved appreciably. Our experimental results suggest that class-specific augmentation may further improve the performance and it will be studied in our future work. The combination of stacking intensity and phase and data augmentation gives 15.2% and 11.2% accuracy improvement over I-CNN and P-CNN with the original training dataset. Therefore, the proposed method can greatly improve pattern recognition accuracy in a coherent DAS.

Funding. Hong Kong Polytechnic University (H-ZG7E, BBWB, ZVGB); National Natural Science Foundation of China (U1701661).

Disclosures. The authors declare no conflicts of interest.

References

1. J. C. Juarez, E. W. Maier, K. N. Choi, and H. F. Taylor, "Distributed fiber-optic intrusion sensor system," J. Lightwave Technol. **23**(6), 2081–2087 (2005).
2. S. Dou, N. Lindsey, A. M. Wagner, T. M. Daley, B. Freifeld, M. Robertson, J. Peterson, C. Ulrich, E. R. Martin, and J. B. Ajo-Franklin, "Distributed acoustic sensing for seismic monitoring of the near surface: A traffic-noise interferometry case study," Sci. Rep. **7**(1), 1–12 (2017).
3. F. Tanimola and D. Hill, "Distributed fiber optic sensors for pipeline protection," J. Nat. Gas Sci. Eng. **1**(4-5), 134–143 (2009).
4. D. Chen, Q. Liu, and Z. He, "High-fidelity distributed fiber-optic acoustic sensor with fading noise suppressed and sub-meter spatial resolution," Opt. Express **26**(13), 16138–16146 (2018).
5. F. Peng, H. Wu, X. Jia, Y. Rao, Z. Wang, and Z. Peng, "Ultra-long high-sensitivity Φ-OTDR for high spatial resolution intrusion detection of pipelines," Opt. Express **22**(11), 13804–13810 (2014).
6. H. Wu, C. Shang, K. Zhu, and C. Lu, "Vibration Detection in Distributed Acoustic Sensor with Threshold-based Technique: a statistical view and analysis," Journal of Lightwave Technology DOI: 10.1109/JLT.2020.3036450.
7. Q. Sun, H. Feng, X. Yan, and Z. Zeng, "Recognition of a phase-sensitivity OTDR sensing system based on morphologic feature extraction," Sensors **15**(7), 15179–15197 (2015).
8. H. J. Wu, S. K. Xiao, X. Y. Wang, Z. N. Wang, J. W. Xu, and Y. J. Rao, "Separation and determination of the disturbing signals in phase-sensitive optical time domain reflectometry (Φ-OTDR)," J. Lightwave Technol. **33**(15), 3156–3162 (2015).
9. C. Cao, X. Fan, Q. W. Liu, and Z. Y. He, "Practical pattern recognition system for distributed optical fiber intrusion monitoring system based on phase-sensitive coherent OTDR," In *Proceedings of the Asia Communications and Photonics Conference*, ASu2.A145.145 (2015).
10. M. Adeel, C. Shang, D. Hu, H. Wu, K. Zhu, A. Raza, and C. Lu, "Impact-Based Feature Extraction Utilizing Differential Signals of Phase-Sensitive OTDR," J. Lightwave Technol. **38**(8), 2539–2546 (2020).
11. Y. Shi, Y. Wang, L. Zhao, and Z. Fan, "An event recognition method for Φ-otdr sensing system based on deep learning," Sensors **19**(15), 3421 (2019).
12. H. J. Wu, J. P. Chen, X. R. Liu, Y. Xiao, M. J. Wang, Y. Zheng, and Y. J. Rao, "One-Dimensional CNN-Based Intelligent Recognition of Vibrations in Pipeline Monitoring With DAS," J. Lightwave Technol. **37**(17), 4359–4366 (2019).
13. Z. Q. Li, J. W. Zhang M, N. Wang, Y. Z. Zhong, and F. Peng, "Fiber distributed acoustic sensing using convolutional long short-term memory network: a field test on high-speed railway intrusion detection," Opt. Express **28**(3), 2925–2938 (2020).
14. J. Tejedor, H. F. Martins, D. Piote, J. M. Guarasa, J. P. Graells, S. M. Lopez, P. C. Guillen, F. D. Smet, and M. G. Herraez, "Toward prevention of pipeline integrity threats using a smart fiber-optic surveillance system," J. Lightwave Technol. **34**(19), 4445–4453 (2016).
15. J. Tejedor, J. Macias-Guarasa, H. F. Martins, J. P. Graells, S. M. Lopez, P. C. Guillen, G. D. Pauw, F. D. Smet, W. Postvoll, C. H. Ahlen, and M. G. Herraez, "Real field deployment of a smart fiber-optic surveillance system for pipeline integrity threat detection: Architectural issues and blind field test results," J. Lightwave Technol. **36**(4), 1052–1062 (2018).

Optics EXPRESS

16. http://www.freesound.org
17. B. McFee, C. Raffel, D. Liang, D. P. Ellis, M. McVicar, E. Battenberg, and O. Nieto, "librosa: Audio and music signal analysis in python," Proc. 14th Python Sci. Conf. **8**, 18–24 (2015).
18. N. Srivastava, G. E. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," J. Machine Learning Res. **15**(1), 1929–1958 (2014).
19. L. Prechelt, "Automatic early stopping using cross validation: quantifying the criteria," Neural Networks **11**(4), 761–767 (1998).
20. P. Luis and J. Wang, "The effectiveness of data augmentation in image classification using deep learning," arXiv preprint arXiv:1712.04621 (2017).
21. C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," J. Big Data **6**(1), 60 (2019).
22. F. Michael, D. C. Hoaglin, and B. Iglewicz, "Some implementations of the boxplot," The Am. Stat. **43**(1), 50–54 (1989).
23. L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," J. Machine Learning Res. **9**, 2579–2605 (2008).

158.132.148.106 {ts '2023-01-03 06:03:02'}