Research Article | Vol. 28, No. 19 / 14 September 2020 / Optics Express 27277
Optics EXPRESS

Distributed fiber sensor and machine learning data analytics for pipeline protection against extrinsic intrusions and intrinsic corrosions

ZHAOQIANG PENG,[1,2] JIANAN JIAN,[1] HONGQIAO WEN,[3] ANDREI GRIBOK,[4] MOHAN WANG,[1] HU LIU,[5] SHENG HUANG,[1] ZHI-HONG MAO,[1] AND KEVIN P. CHEN[1,2,*]

[1]*University of Pittsburgh, Department of Electrical and Computer Engineering, 238 Benedum Hall, 3700 O'Hara Street, PA 15261, USA*
[2]*National Energy Technology Laboratory, 626 Cochrans Mill Rd., Pittsburgh, PA 15236, USA*
[3]*Wuhan University of Technology, National Engineering Laboratory for Fiber Optic Sensing Technology, 122 Luoshi Road, Wuhan, Hubei, 430070, China*
[4]*Idaho National Laboratory 2525 N. Fremont Avenue P.O. Box 1625, Idaho Falls, ID 83415, USA*
[5]*Beihang University, School of Instrument Science and Opto-electronic Engineering, No. 37 Xueyuan Road, Beijing, 10091, China*
[*]*pec9@pitt.edu*

Abstract: This paper presents an integrated technical framework to protect pipelines against both malicious intrusions and piping degradation using a distributed fiber sensing technology and artificial intelligence. A distributed acoustic sensing (DAS) system based on phase-sensitive optical time-domain reflectometry (φ-OTDR) was used to detect acoustic wave propagation and scattering along pipeline structures consisting of straight piping and sharp bend elbow. Signal to noise ratio of the DAS system was enhanced by femtosecond induced artificial Rayleigh scattering centers. Data harnessed by the DAS system were analyzed by neural network-based machine learning algorithms. The system identified with over 85% accuracy in various external impact events, and over 94% accuracy for defect identification through supervised learning and 71% accuracy through unsupervised learning.

1. Introduction

Oil, gas, petro-chemical, and energy industries rely on vast networks of pipelines to transport and distribute large quantities of hydrocarbon fuels in liquid and gaseous forms to industrial users and customers. Over the decades, millions of miles of pipeline were built on land, underground, and off shore. These pipeline networks often traverse remote area subject to harsh weather conditions. As key infrastructures, these pipelines are important to national security and economic vitality. It is therefore important to develop a comprehensive monitoring scheme to protect their integrity from external threats such as malicious intrusion, tempering attempts, illegal tapping, construction accidents, and natural disasters, as well as from internal structural degradation, including corrosion/erosions, due to aging and wear from weather.

Over the years, both technology-based and human-based inspection schemes were used to safeguard pipeline networks. Imaging using digital cameras was used to monitor pipelines for various external threats caused by intrusion events and natural disasters. However, conventional camera surveillance systems incur significant installation costs and extensive deployments along pipelines, especially in desolate and harsh environments. The information provided by a camera is limited to line-of-sight field views. Camera systems also cannot effectively monitor pipeline buried underground, deployed off-shore, or used in harsh radioactive environments such as those found in nuclear power plants. Ultrasonic wave-based inspection schemes have been widely used

#397509 https://doi.org/10.1364/OE.397509
Journal © 2020 Received 12 May 2020; revised 17 Aug 2020; accepted 19 Aug 2020; published 1 Sep 2020

Optics EXPRESS

for structure health monitoring. In these applications, unbound or guided acoustic waves are excited by impact events (e.g. pencil breaking) or magnetostriction/piezoelectric actuators [1] to probe various targets. Backscattered acoustic wave ("acoustic echo") can then be detected by ultrasonic sensors as cues to detect and to identify various events (e.g. motion detection) [2,3] or structure changes (e.g. structural defects) [4,5]. The guided acoustic wave approach has been shown in the laboratory setting [4] and in the field [5] to be effective methods for detecting and localizing pipeline anomalies in key sections prone to defects. Although promising results using guided acoustic sensors have been reported in monitoring simple structures such as straight pipes and pulled bends, their abilities are limited when extended to interrogate complex structures in pipeline networks such as tees, welded elbows and diameter changes, which potentially causing overlapping echoes and the discontinuity of acoustic waves [6]. Some unwanted modes persist, along with imperfect direction control, as background noise also strongly affects the detection reliability [7]. The challenge to detect and to identify defects developed in complex pipe structures remain as one of the greatest challenges for this application.

A far more appealing technical approach to pipeline monitoring will be one that adapts to complex structures and owns dual functions, both defending against external threats and performing active or passive measurements to monitor pipeline degradation resulted from aging and weather. Distributed optical fiber sensors are considered promising candidates to perform detections of guided acoustic waves as they are generated by external events or scattered by defects with high spatial resolutions. As a sensing technology, optical fiber sensors are well known for their resilience for harsh environment applications. They are also immune to electromagnetic interference, corrosion-resistant, and highly sensitive for dynamic strain measurements. Fiber Bragg gratings array [8], optical time-domain reflectometry (OTDR) [9], and phase-sensitive OTDR [10,11] have been reported to perform leakage detection or pipeline security.

A unique trait of fiber optical sensors is their capability to perform distributed strain measurements by exploiting various light scattering phenomena in unmodified telecom fibers to achieve high spatial resolution measurements. One of the most utilized in-fiber light scattering process is Rayleigh scattering. Rayleigh is an elastic scattering process while light propagating in fiber cores are scattered due to small inhomogeneity in local refractive index along the optical fiber. Local temperature and strain changes can produce changes on local refractive index variation, thus change the Rayleigh scattering profile of the optical fiber. Using either Optical Time-Domain Reflectometry (OTDR) or Optical Frequency Domain Reflectometry (OFDR), Rayleigh scattering profile can be measured with high spatial resolution and correlated to temperature or strain profiles along the fiber. A great diversity of optical fiber reflectometry has been reported with ability to provide sensing across long distance (up to 175 km [12]).

Phase-sensitive OTDR (φ-OTDR) based on Rayleigh scattering is known as a fully distributed dynamic strain sensor for long-distance and high-sensitivity detection [13]. Fast dynamic strain measurement in distributed optical fiber sensors promises a widely detectable frequency range that harnesses transient mechanical perturbation to optical fiber as an acoustic waveform [14]. Using standard telecom optical fiber, the location of perturbation can be determined thanks to the high spatial resolution. However, the intrinsically low Rayleigh scattering coefficient of optical fiber severely limits the signal-to-noise ratio (SNR) of φ-OTDR system. Insufficient accuracy leads to low probability of detection, eventually giving rise to false alarms.

Data analytics methods based on machine learning have been reported and discussed for pipeline surveillance [15]. However, most of prior works focus on machine learning approaches based on single hidden layer neural networks. These approaches require filtration and data preprocessing of acoustic signal harnessed by fiber sensors and they are insensitive to nonlinear features hidden in data [15]. To mitigate low SNR of acoustic signal, conventional data processing techniques such as moving average and wavelet transform could improve SNR but with significant drawbacks such as long computation time [16,17]. Deep neural networks that use multiple

Optics EXPRESS

hidden layers can outperform single-layer algorithm to address complex nonlinearity. However, the recent reported application of deep neural networks only focused to distinguish events with significantly different acoustic patterns as well [18].

In this paper, we demonstrate a ϕ-OTDR distributed fiber sensor scheme using optical fiber with enhanced Rayleigh backscattering profiles for both pipeline security monitoring and pipeline corrosion defect detection and prognostics. Using femtosecond laser direct writing, artificial Rayleigh scattering points in telecom fibers are fabricated to enhanced SNR by 35 dB. Machine learning data analytics using both supervised and unsupervised neural networks were used to delineate high spatial resolution acoustic signal along the pipeline to locate and deduce the size and shape of defect in pipeline. This paper demonstrates an integrated approach using distributed sensors and artificial intelligence big data analytics to protect large-scale critical infrastructures.

2. Experiment design

2.1. Fabrication of high-sensitivity acoustic sensor

In various applications, strain-dependent Rayleigh backscattering along optical fibers has been a leading approach to perform DAS. A φ-OTDR system was reportedly applied in a 5-km long optical fiber with 20-meter spatial resolution and possessing 80 kHz detectable frequency range for dynamic strain measurement [19]. However, the performance of Rayleigh φ-OTDR is limited by low Rayleigh backscattering signals from standard telecom fibers, which were optimized to be a low scattering loss (< -100 dB/mm) waveguide for long-haul telecommunication using 1550-nm wavelength windows. The low SNR curbs the intensity of Rayleigh backscattering making it usable for interrogating external dynamic perturbations. Significant challenges are posed regarding the interrogation instrument and signal processing.

To improve SNR of φ-OTDR, several approaches were proposed to increase Rayleigh scattering in fiber, such as ultra-violet (UV) exposure of the hydrogen-loaded single mode fiber [20,21] and UV inscription of ultra-weak fiber Bragg gratings (FBGs) [22,23]. Both required complex fabrication process involving photosensitization of standard telecom fiber by using hydrogen loading or having specialty fibers with high germanium concentration in fiber cores. The performance of these fiber sensors are also vulnerable to elevated temperature [24].

Prior works mentioned in Ref. [20–21] are based on weak FBGs to achieve enhancement of SNR. A simpler and more reliable technique named femtosecond ultrafast laser direct-writing can also be used to enhance Rayleigh backscattering. This technique uses focused femtosecond laser to produces permanent nano-gratings in fiber core through fiber polymer coating to induce broadband Rayleigh scattering points. Unlike FBG, these laser-induced Rayleigh scattering points are wavelength independent and high-temperature stable, which is more suitable for harsh environment applications [25–28]. Through optimizations of laser writing parameters, the insertion loss of each nano-reflector can be minimized to 0.0012 dB/point [29].

Figure 1(a) shows the schematic of nano-gratings fabrication. A Coherent MIRA-D Ti-sapphire seed oscillator and a RegA 9000 regenerative amplifier operating at 800 nm with a repetition rate of 250 kHz were used to deliver 270 fs (FWHM) ~160nJ laser pulses. The beam shaping technique consisted of a cylindrical telescope, an 80× aberration-corrected microscope objective and oil immersed objectives (NA 1.25). The waist diameter of the symmetric laser beam was estimated to be around 2 μm. A standard telecommunication single mode fiber (Corning SMF-28e+) was mounted on a three-axis Aerotech ABL2002 motion stage. It underwent real-time interrogation by a commercial optical frequency-domain reflectometry interrogator (LUNA OBR4600). The optical fiber was translated by a custom-built roll-to-roll setup that could ensure the mass producibility of the potential application.

The enhancement of Rayleigh scattering is caused by discontinuities of nano-gratings on a microscopic scale. Figure 1(b) shows the scanning electron microscope image of a typical nano-grating written by fs-laser. The size and damage level of defect in nano-gratings can be

Optics EXPRESS

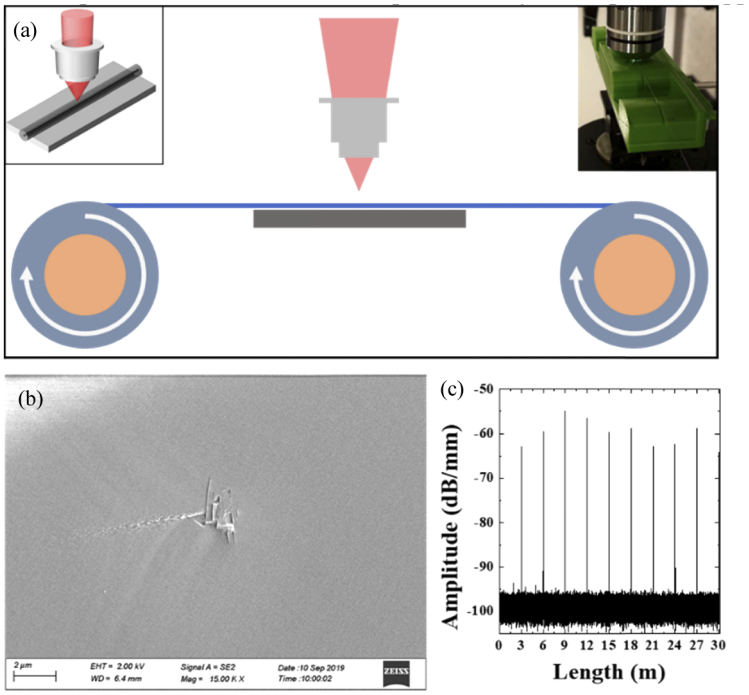

Fig. 1. (a) Schematic of laser direct-writing with roll-to-roll setup. (b) Scanning electron microscope (SEM) image of nano-grating morphology. (c) Enhanced back scattering in standard fibers enabled by fs-laser.

flexibly controlled by manipulating the laser beam. Figure 1(c) shows over 35 dB enhancement of Rayleigh backscattering profiles of nine laser enhanced points along 30-meter standard telecom fiber. By dramatically increasing Rayleigh backscattering, high SNR from those nano-reflectors enables φ-OTDR systems to conduct high-precision measurements.

2.2. DAS in pipeline installation

A homodyne φ-OTDR system based on a 3×3 coupler configuration was designed and implemented to monitor external perturbation induced light phase changes along the fiber inscribed with nano-reflectors to enhanced Rayleigh backscattering signals. The DAS system used for this experiment is schematically shown in Fig. 2. A 1550.12 nm light with 6 kHz linewidth from a distributed feedback fiber laser (NKT Basik C15) was modulated to generate 10 nanosecond pulses using a semiconductor optical amplifier (SOA, Thorlabs, SOA1013SXS). The optical pulse was set to have a 66-kHz repetition rate. An erbium-doped fiber amplifier (EDFA) was used to increase pulse energy to 150-nJ (10 mW output power). A more than 30-meter fiber span consisting of nine nano-reflectors was wrapped around two straight sections of iron pipes connected with a 90-degree sharp elbow as shown in Fig. 2. The pipes tested in this experiment have a 3-inch inner diameter with a 0.5-inch wall thickness. Two straight pipes totaling 5-feet long were tightly screwed together by the sharp iron elbow. 15-meter of optical fiber consisting of 5 nano-reflectors formed 4 acoustic sensors on the surface of longer pipe structure. Another 12-meter optical fiber was wrapped to the shorter section of the straight pipe, which formed 3 acoustic sensors. Figure 2 shows the detailed installation and schematic of homodyne φ-OTDR.

The interrogating pulse produced by the fiber laser, as shown in Fig. 2, was partially reflected by laser-induced nano-reflectors along the fiber conveying transient phase information. The

Optics EXPRESS

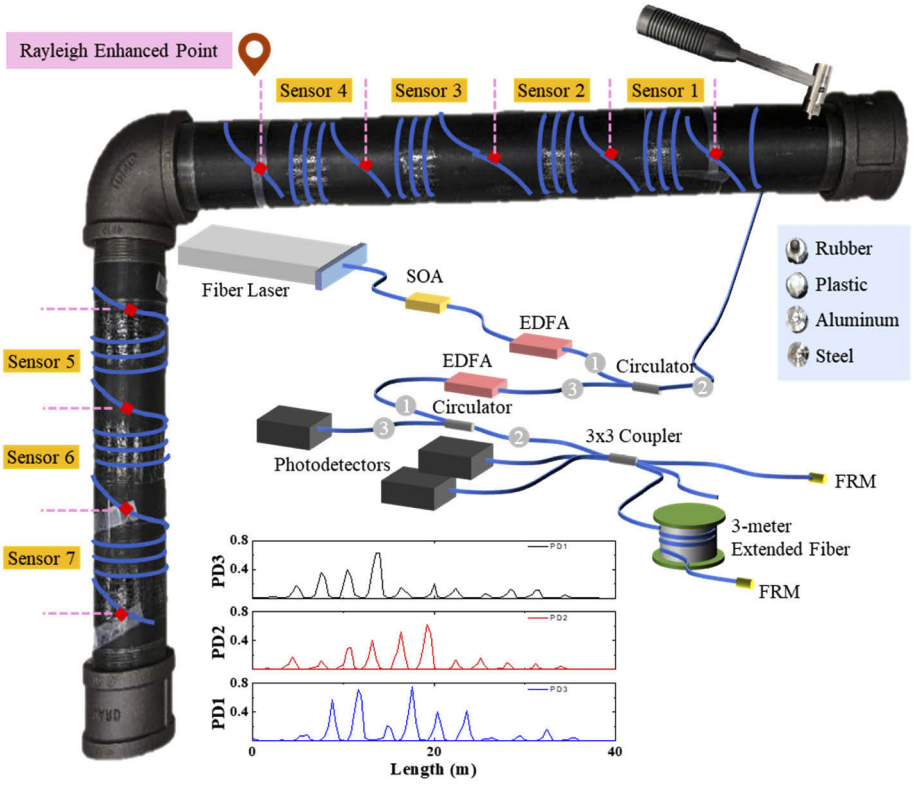

Fig. 2. Schematic diagram of DAS system for pipeline protection and Rayleigh backscattering profile with nine nano-reflectors. SOA: semiconductor optical amplifier; EDFA: erbium-doped fiber amplifier; FRM: Faraday rotator mirror. Temporal profiles of Rayleigh backscattering signal detected by three photodetectors (PD1-PD3) is shown as in inset.

backscattered pulses were further amplified by another erbium-doped fiber amplifier (30 dB gain) and reoriented to a 3×3 coupler via an optical circulator, then divided into three equal outputs. Two of these three outputs were terminated with Faraday rotator mirrors arising from the concept of unbalanced Michelson interferometer in a typical homodyne φ-OTDR [30]. Inset of Fig. 2 show temporal profiles of Rayleigh backscattering signals detected by three photodetectors (PD1-PD3). The measured signals clearly reveal enhanced Rayleigh signals produced by laser-induced scattering points in fibers.

In our φ-OTDR scheme, the optical path difference between those two arms in interferometer was set to be 3 meters—the same distance between two adjacent nano-reflectors in the sensing fiber. This gave rise to homodyne interference of over 35 dB enhanced backscattered light from successive nano-reflectors. External perturbation induced light phase shift within succession of adjacent nano-reflectors can be traced and monitored in real-time. They were continuously recorded by three photodetectors (InGaAs Fixed Gain 200 MHz) and demodulated using the Naval Postgraduate School method [31]. Three outputs of 3×3 coupler have a constant 120° phase difference. Two trigonometric identities can be used to eliminate the DC bias and extract the phase information:

$$\sum_{n=1}^{3} \cos\left[\xi - (n-1)\frac{2}{3}\pi\right] = 0, \tag{1}$$

Optics EXPRESS

$$\sum_{n=1}^{N} \cos^2\left[\xi - (n-1)\frac{2}{3}\pi\right] = \frac{N}{2}, \tag{2}$$

where ξ can be the output voltage of photodetector. The NPS method was developed based on those two trigonometric identities. Each signal from three photodetectors is a function of time containing the transient phase information.

$$V_i = C + Bcos\left[\phi(t) - (i-1)\frac{2\pi}{3}\right],\ i = 1, 2, 3 \tag{3}$$

where $\phi(t)$ is the signal phase, and B is the varied amplitude based on average output intensity C. Through compound operations including differential and integral, the DC component can be removed, leaving only the phase information. The phase can be described by output of symmetric demodulator as:

$$Vout = \sqrt{3}A\phi(t), \tag{4}$$

where A is a known parameter consisting of various gains in operation. The quantity of the transient phase is directly proportional to the output of the deterministic demodulator.

2.3. Experimental setup and analytic methods

Experiments were carried out using DAS systems for both passive and active measurements with optical fiber installed close to a sharp iron elbow. They were designed to test the DAS system using enhanced Rayleigh scattering fibers to detect and identify extrinsic acoustic events against external intrusion threats, as well as inspection tools to detect pipeline corrosions at a specific bend. Both cases require an acoustic excitation source to be classified or analyzed. A specialized hammer was used to generate mechanical vibration as acoustic signals from single tapping as depicted in Fig. 2. This hammer had several heads made of four different materials, including soft rubber, hard plastic, aluminum, and steel. These various hammer heads were used to generate mechanical waves with different frequency characteristics, thereby mimicking external perturbations. Data was acquired by recording temporal phase shift of backscattered light at the intensity peaks of interference, and formed into 333-ms segmented time frame at every second. They were trigged by monitoring the energy variation within the spectrum depicting the predominant aspects of external perturbations. The unrelative noises were truncated and zero-padded during pre-processing. The sampling rate was set to be 66 kSa/s, resulting in a 33-kHz frequency bandwidth for detecting dynamic vibration. In order to give the neural networks direct access to global properties of the signals, fast Fourier transform (FFT) was firstly performed on acoustic waveforms harnessed by sensors. Lowpass filters were applied to frequency spectra of acoustic data to retain low-frequency information within 0–4kHz. This is consistent with frequency characteristics of acoustic signals generated by the hammer (0–2 kHz). The attenuations of forward propagating wave and backward scattering echoes within acoustic waveforms occurring in this frequency range were also investigated through spectrum analysis.

A large amount of Rayleigh backscattering signals was simultaneously harnessed at many locations, producing vast datasets with distinct features. A versatile analytics tool must be used for event identification and feature classifications. For example, the sound generated by excavators at a nearby construction site differ from those produced through intentional sabotage of pipeline—namely, high-pitched noises due to contact with metal. On the other hand, different types of subsurface pipe corrosions all generate similar echoes due to their small size, making it difficult to use a single predictive data analytics scheme to extract features from a wide range of DAS data. A machine learning approach could potentially handle the large amount of data and reveal patterns associated with both external threats and pipeline corrosion/erosion. In this paper, data analytics schemes using neural networks were performed in frequency domains. They were fed into supervised and unsupervised neural networks to explore the possibility of identification.

Optics EXPRESS

3. Experiment results

3.1. Identification of extrinsic acoustic events

Intentional or accidental pipe contact runs the risk of structural damage. Risk-level evaluation serves as the basis for enacting proper repairs based on identifying the types of threats presented at each stage of intrusion. Four types of hammer heads were used to generate extrinsic acoustic signals that mimicked different external impact events and their destructive effects. Impacts of rubber and plastic hammer heads on test pipes produced low-frequency acoustic signals similar to when pipes are struck on their insulation and protective layers, which are made of rubber and fiber glass materials. The direct contact between pipes and metal tools was simulated using aluminum and steel hammerheads. Successful and robust identification of those four extrinsic acoustic events was useful for monitoring the intrusion process and to assess the severity of these types of damages. Figure 3 shows the transient waveforms of acoustic signals produced by four hammerheads mimicking various extrinsic events.

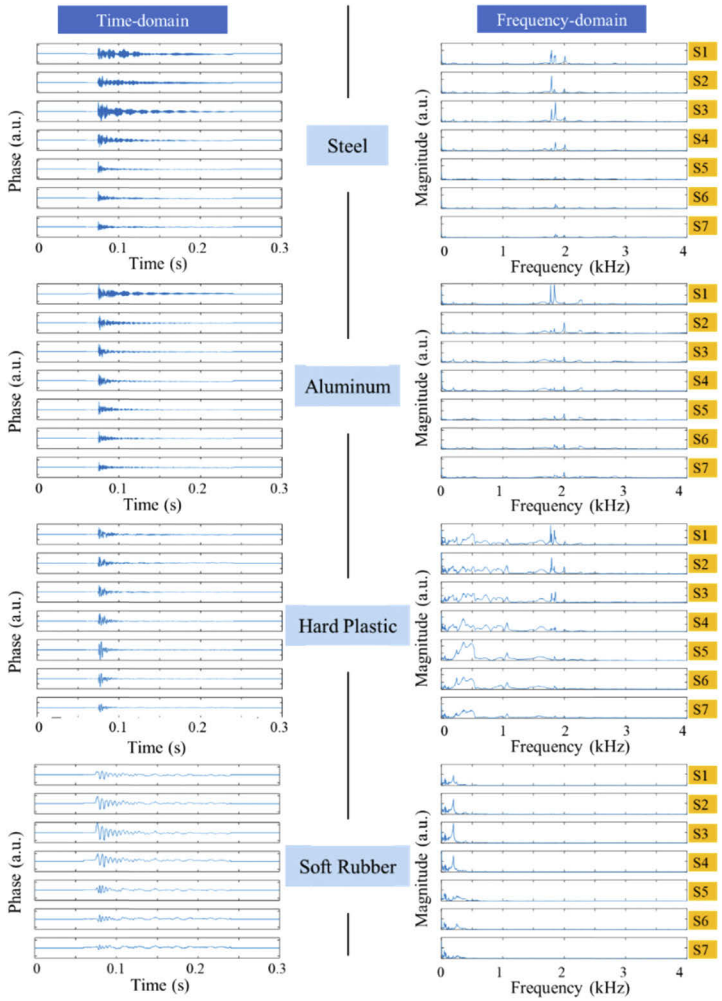

Fig. 3. The time-domain and frequency-domain acoustic waveforms from 7 sensors (S1-S7) by single tapping.

Optics EXPRESS

Tapping the hammer at one end of pipe caused seven optical fiber sensors to capture guided acoustic wave as it propagated and scattered along the pipe structures. The contrast among the acoustic waveforms from those sensors included the information regarding attenuation and reflection shown in Fig. 3. Most acoustic energy produced by aluminum and steel hammerheads registers at relatively high frequencies of 1.5–2 kHz. Soft rubber offers the lowest frequency and weakest intensity among the four kinds of materials. The hard-plastic hammer head generates acoustic signals consisting both high frequencies (1.5–2 kHz) and low frequencies (<1 kHz). The high-frequency components dissipate during forward propagation and lack of amplitude to cross the elbow, but low-frequency components of acoustic signals below 500 Hz can pass the elbow without significant attenuation. Beside the low-pass filtering effects on acoustic signal, the pipe elbow induced strong reflection in the backward direction. The low-frequency echoes generation was prominent in frequency spectrum of hard-plastic hammer. The energy of low-frequency components became higher at S5 when passing through the sharp elbow. This frequency-dependent reflection also contributed to the classification of external acoustic events.

External threats typically have distinct and well-defined acoustic characteristics. These threat events can be simulated to produce training data for neural networks. Therefore, neural networks trained through supervised learning can be an effective approach for event identification and classifications. In this paper, both shallow and deep neural networks are trained through supervised learnings to identify external perturbation events generated by four different hammer heads.

Shallow non-convolutional neural networks have simple structures. With only one hidden layer of neurons, shallow neural networks are easy to implement with low computation burdens. They also have the theoretical capacity to approximate any continuous function on a compact domain. In our application, this is a desired trait since the objective is simple correlation between the acoustic waveforms and the four classes of hammer heads. The architectures of the shallow neural network used in this work are shown in Fig. 4. Each trial obtained the acoustic waveforms harnessed from seven fiber sensors. The maximum energy center was preselected through time-frequency representation with certain thresholds and zero-padded to increase frequency resolution. The first 1200 frequency components (DC to 4 kHz) of acoustic signals were used as input to the neural network. Thus, data collected in each trial were organized in a vector consisting of 8400 elements (1200×7). The shallow neural network took as input a vector formed by vertically stacking the spectral magnitude of the acoustic waveforms from seven fiber sensors. The input layer was then followed by a hidden layer consisting of 10 neurons as well as a softmax output layer for classification among 4 possible events as shown in Fig. 4.

When facing complex nonlinearity within the data, deep neural networks are often better than single-hidden-layer shallow neural networks because of multiple hidden layers. Each layer extracts features of different levels from the data. Take the example of an image, low level features could be edges, corners, and intensity gradients; the mid-level features could be shapes and objects; high level features could be the object combination or arrangement specific to the application.

Convolutional neural networks (CNNs) are popular deep neural network structures. These networks use a series of convolution and pooling operations to reduce the number of parameters making CNNs especially successful at processing large-size inputs with spatial invariance. The pros and cons are not clear enough for neural networks to distinguish acoustic waveforms captured by distributed fiber sensors. To examine the linearity of hidden features, both a single-hidden-layer shallow neural network and CNN were applied to our dataset for supervised learning. CNNs have proven very effective for image processing, thus the acoustic waveforms harnessed by seven fiber sensing regions were arranged into a 1200 (frequency) $\times$ 7(sensors) matrix. The architecture of CNN used for feature extraction is shown in Fig. 4. Three convolution layers were applied including 2×3 or 3×3 convolutional filters and 2×2 max pooling. The use of

Optics EXPRESS

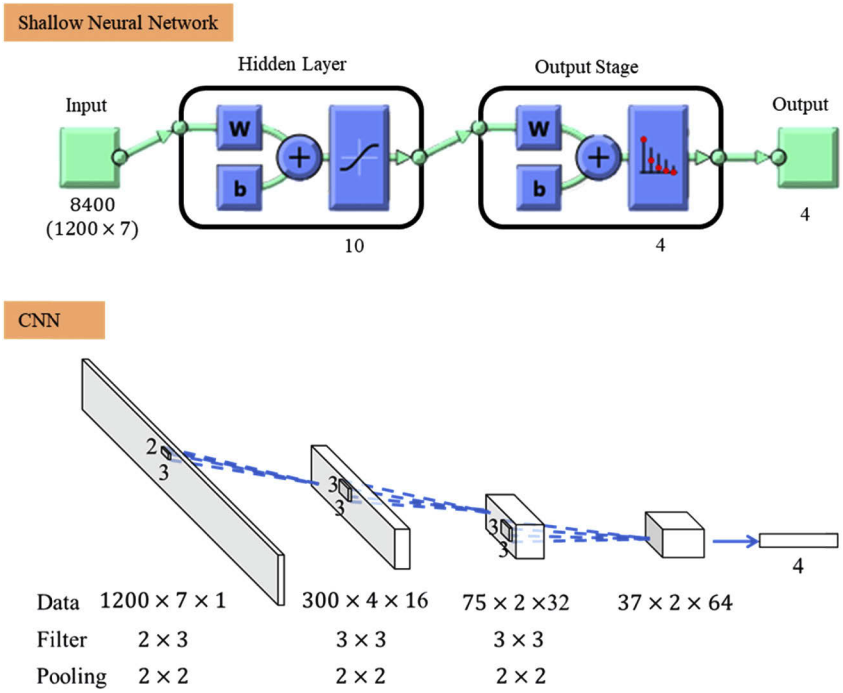

Fig. 4. Architectures of shallow neural network and CNN in the pipeline protection system.

nonlinear activation function (Rectified Linear Unit) created complex mapping between input and output through gradually decreasing the size of the image and increasing the number of extracted features. A fully connected softmax layer was attached at the output stage. Total 68 sets of trial data from four types of hammer single-tapping were harnessed from seven active fiber sensors. For each type of hammer, seventeen sets of trial data were randomly divided to 12 and 5 for training and testing respectively. Each classification was run 10 times to determine uncertainty of data selection and obtain the range of accuracy. Prediction outcome are described using confusion matrices as shown in Fig. 5. Their accuracies were qualified using mean (M) and standard deviation (D). The confusion matrices show that false identifications occurred in classifying signals produced between aluminum and stainless-steel hammer heads. This is not a surprise, Fig. 3 shows that signal patterns produced by stainless steel and aluminum heads are similar in both time and frequency domains. Most of acoustic energy produced by these two hammer heads concentrates around 1.8 kHz. In contrast, signals produced by plastic hammers has both low frequency components in addition to frequency components around 1.8 kHz, while signals produced by the rubber hammer head are predominantly low frequency signals. Overall, both classifiers reached over 80% accuracy every time to identify four types of hammer heads as shown in Fig. 5.

3.2. Identification of intrinsic structural corrosion using unsupervised neural networks

The distributed fiber sensor can also be an important tool for detecting and identifying pipeline corrosions produced at sharp bends by aging and weathering. Nondestructive structural health inspections such as guided acoustic wave measurements have been used in pipeline inspection for many years. In this scheme, sensors are installed at a single location to detect echo backscattered from pipeline defects. However, this approach is ineffective for inspecting pipe networks involving complex shapes and sharp bends. As illustrated in Fig. 3, the 90-degree iron elbow structure

Optics EXPRESS

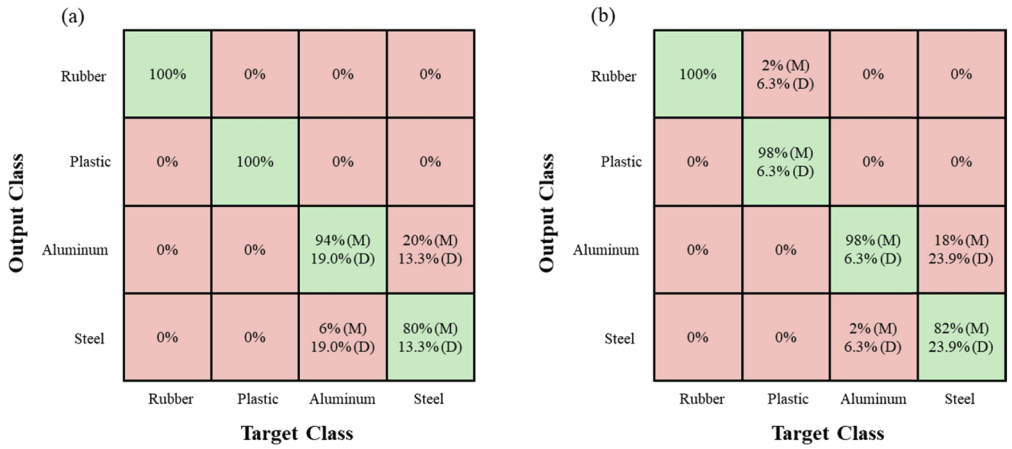

Fig. 5. The confusion matrix of four acoustic events classification including (a) shallow neural network and (b) CNN.

attenuated most high-frequency acoustic signals above 1 kHz. The backscattered acoustic wave produced by a corrosion defect might not have sufficient magnitude or carry enough characteristic information to be recognized by an electrical transducer. In these situations, distributed fiber sensors could have significant advantages over the conventional guided acoustic approach. The high spatial resolution data harnessed by distributed fiber sensors can detect both back-scattered acoustic signal and also forward propagating acoustic signals as it passes through the elbow. Given the weak intensity of backscattered acoustic wave, forward propagating guided acoustic wave captured by distributed fiber sensors can yield additional information for defect detection and classification. If optical fibers are wrap along the pipeline as sensors, distance between sensors and defect locations are also much smaller than the conventional guided wave schemes, which greatly enhance chances of signal detections and subsequent defect classification. This is particularly advantageous in regard to corrosion detection for pipe structures with sharp bends.

Sharp bends or pipe elbows are susceptible to flow-accelerated corrosion induced by turbulent flows that are usually hard to discover. Such corrosion can rapidly perforate metal walls due to the small pitting at the early stage of corrosion. The structural health inspection of complex pipe shapes has been considered as a salient challenge and waiting to be solved. Based on prior studies on external perturbation identification, it was decided that acoustic waveforms for this study would be generated by the single-tapping of a hard-plastic hammer. This produced an acoustic signal with sufficient magnitude and a strong enough characteristic frequency to pass the sharp elbow. The recordings of acoustic waveforms for each time frame were reduced to 120-ms which kept the majority of acoustic waves and eased the computational load of DAS system. Signals were collected by the DAS system and analyzed by neural networks. Figure 6 and Fig. 7 presents defective elbows used to mimic various corrosion. To mimic usual defects occurring at pipe elbows, two simulated scenarios were investigated: pipe insulation failure causing external corrosion (Case 1) and galvanic corrosion at the joint of two metal components (Case 2). External corrosion was carved as a trench in the outer wall of elbow shown in Fig. 6 (Case 1). A loose connection between pipes and elbow was used to simulate galvanic corrosion as depicted in Fig. 6 (Case 2). Propagating acoustic waveforms were harnessed in all scenarios, as well as in a defect-free pipe structure (Case 3).

Four more types of simulated pitting on inner surface were also made as shown in Fig. 7 (Cases 4–7). Case 4 had a groove defect machined into the inner wall of the elbow sized at 1.5 in. (length) $\times$ 0.25 in. (width) $\times$ 0.08 in. (depth). Three holes (0.125 in. in diameter $\times$ 0.2 in. in

Optics EXPRESS

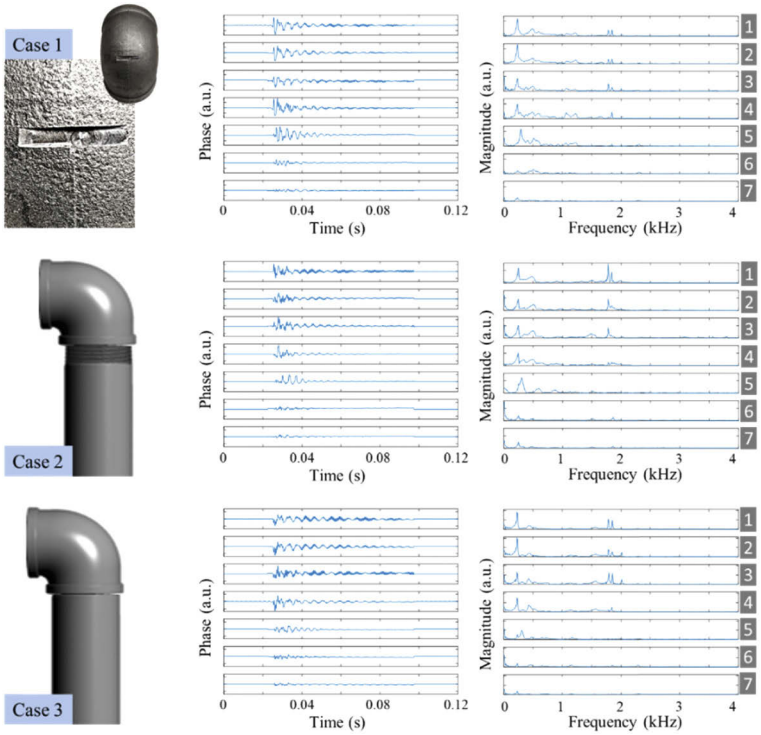

Fig. 6. Three installation scenarios and their waveforms at the time- and frequency-domain by single taping with a hard-plastic hammer. Case 1: One trench at external surface; Case 2: loose connection; Case 3: normal connection.

depth) were drilled inside the groove to mimic accelerated pitting defects in Case 5. Case 6 and 7 doubled the number of defects found in the previous cases to show a much severer situation. Acoustic signals and their frequency spectra generated on Case 1–Case 7 are shown in Fig. 6 and Fig. 7.

Supervised machine learning approaches using the shallow neural network and CNN were first explored to identify and classify acoustic waveforms generated by Cases 1–7. The neural network structures were identical to those used for external threat identifications as shown in Fig. 4. The corrosion classification outcomes using the shallow neural network and the CNN are summarized in Table 1. Although supervised machine learning supported high-accuracy classification of external perturbations and corrosion at the iron elbow, it is not a practical unified approach for analyzing a great diversity of events. Given large variations in the shapes and sizes of pipe networks, in addition to the many different types of defects, sufficient learning data cannot be generated for all scenarios. Prior labelling for each type of defects is burdensome, tedious and even inviable due to their unknown natures.

Instead of examining the relations between the acoustic waveforms and the labels, unsupervised learning aims to find relations within the data themselves. These relations characterize the latent features of the data, which can be related to the classifications of specific corrosions. In addition to supervised neural networks, unsupervised self-organizing map (SOM) and autoencoder were also deployed to investigate the recognizable identities of seven scenarios.

SOM is an effective unsupervised learning technique using neuromorphic principles to sensitize neighboring neurons to similar inputs with much faster training speed. The neurons in a SOM are generally placed on a grid so that there is a well-defined distance between any pair. Their

Optics EXPRESS

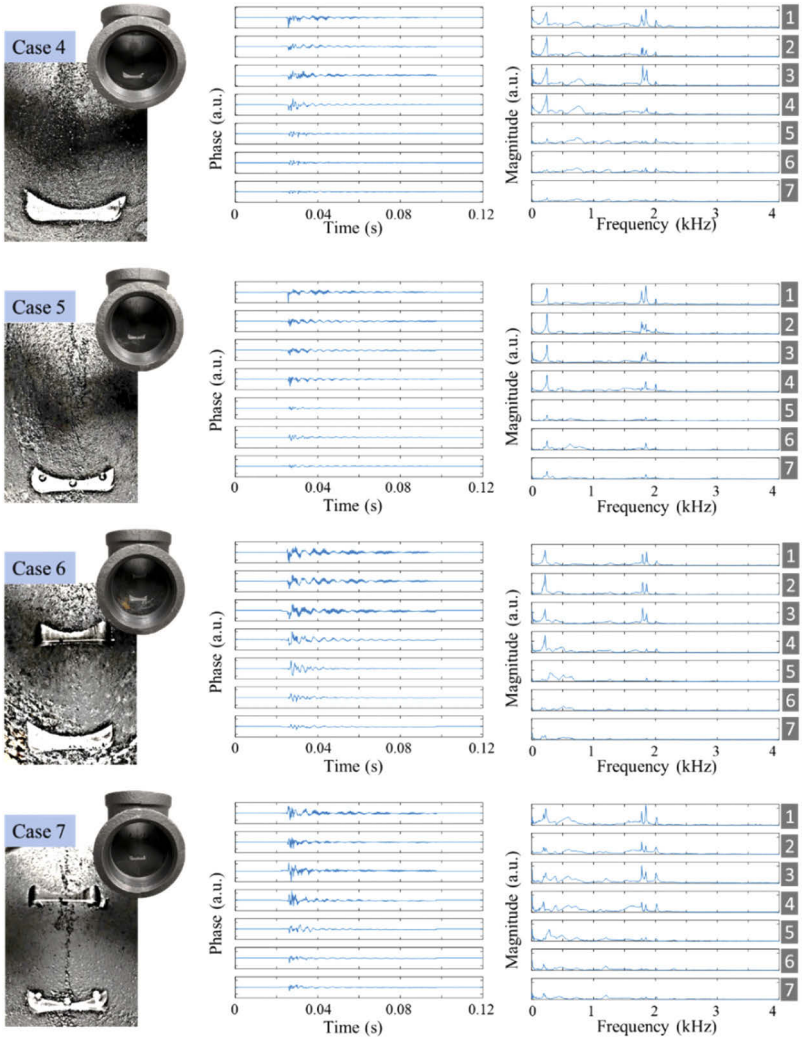

Fig. 7. Another four installation scenarios: Case 4: One cutting groove; Case 5: One cutting groove with three drilling holes; Case 6: Two cutting grooves; Case 7: Two cutting grooves with six drilling holes.

Table 1. Classification result of seven scenarios by single tapping with hard plastic hammer.

Scenario	Shallow	CNN	SOM	SOM + Softmax	Auto-encoder	Autoencoder + Softmax
Normal	97.1%–100%	94.3%–100%	71.4%–83.3%	74.3%–85.7%	73.8%–84.5%	94.3%–100%
Loose						
One inner groove	99.1% (M)	97.7% (M)	78.0 (M)	79.7% (M)	79.8% (M)	96.9% (M)
One inner groove and three holes						
Two inner grooves	1.4% (D)	1.8% (D)	3.8% (D)	3.9% (D)	3.5% (D)	2.8% (D)
Two inner grooves and six holes						
One external trench						

Optics EXPRESS

weights are then randomly initialized. During the training phase, for each input sample, the SOM picks the neuron whose weight is closest to the input data, and this is known as the best matching unit (BMU). Then the weight of each neuron is updated to approach the input data, and the closer to the BMU a neuron, the more significant the update on its weight. Simply put, the update of neuron i's weight is given by:

$$w_i' = w_i + \alpha(d(i, \mathrm{BMU}))(x - w_i) \tag{5}$$

where x is the input data and α is a decreasing function, such as the Gaussian function, of $d(i, \mathrm{BMU})$—the distance between neuron i and the BMU. A trained SOM can be naturally used as a classification network to categorize an input based on its BMU. Figure 8 shows SOM architecture used in this work. The spectral magnitude of first 1200 frequency components from seven acoustic fiber sensors were stacked together to form vector inputs consisting of 8400 elements. Knowing that there are seven types of installation condition at iron elbow, the SOM topology is designed to be a 3-D grid composed of $2 \times 2 \times 2$ (eight) closely located neurons. A total of 119 trials of unlabeled data were used to train the SOM.

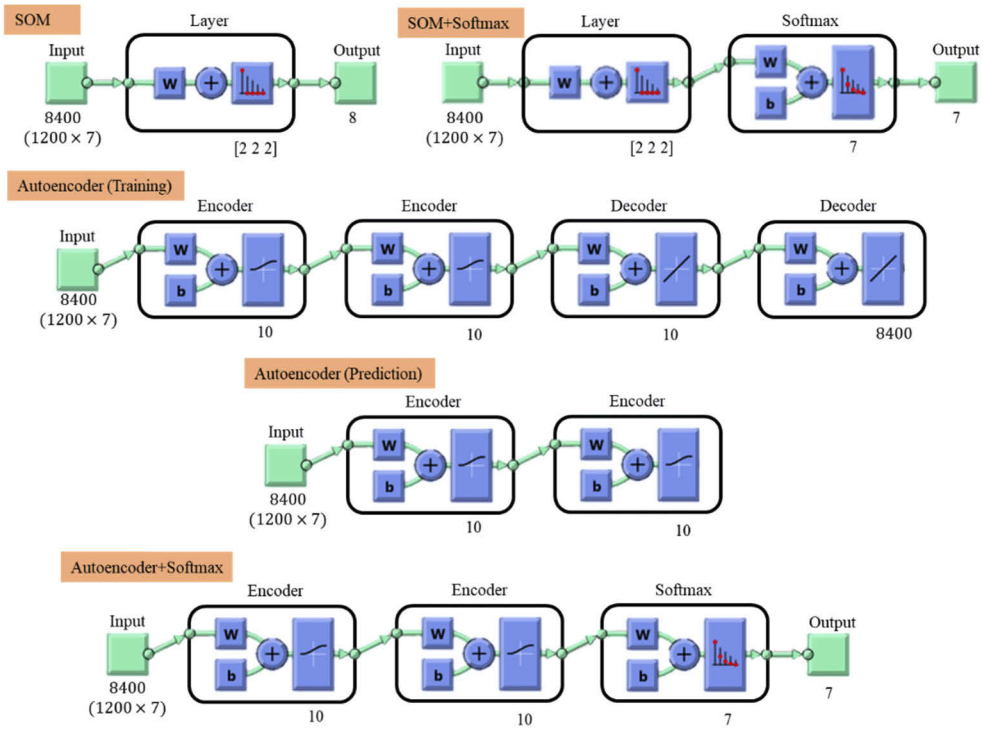

Fig. 8. Unsupervised and supervised neural networks for classification in Cases 1–7.

Results from the unsupervised learning came in the form of unlabeled clusters that were difficult to interpret and identify in association with actual defects. To evaluate the performance of the unsupervised learning, each cluster was manually linked to one scenario among Cases 1–7 or left as unidentifiable. Another way to investigate the relation between unsupervised output clusters and installation scenarios is by training a softmax layer attached at the output stage of the unsupervised network. This small-scale supervised layer can easily learn the association between the actual scenarios and the features that determine the cluster assignment.

Another unsupervised learning scheme adopted to extracted features from acoustic waveforms was the autoencoder. Similar to common data compression processes, the autoencoder aims

Optics EXPRESS

to minimize the distortion as data are compressed onto the feature space. It consists of two neural networks. The first is the encoder network that maps the data space to the feature space $F: \mathbb{R}^n \rightarrow \mathbb{R}^m$. It was followed by the decoder network which mapped the feature space back to the data space $G: \mathbb{R}^m \rightarrow \mathbb{R}^n$. Through gradient-based back propagation, the difference between the original data and the decompressed data was minimized, i.e., for k given input samples $X \in \mathbb{R}^k \times n$, and the trained networks solved the following equation

$$\min_{F,G} ||G(F(x)) - x||. \tag{6}$$

As $m \ll n$ is generally the case, we were able to represent the large size data using a few crucial features. To classify the data based on these features, a regularization term was added in the loss function so that one feature was uniquely presented in each sample. In this case, the encoder output $F(x)$ directly indicates the classification of sample x. This method is known as "sparse autoencoder".

Two cascades of autoencoder were used to accelerate the learning and improve the classification accuracy. The stacked encoder reduces the data size to 10 features. Table 1 shows the classification results by either manually associating unlabeled clusters or deploying a softmax layer at output.

Prediction outcome are described using interval as listed in Table 1. The lower limit of classification accuracy determined false alarm rate of system indicates the worse-case performance or robustness of a classification method and can be used to determine the false alarm rate for the systems. In addition, prediction outcomes are also qualified using mean and standard deviation. Table 1 shows that prediction accuracy using Shallow Neural Network can achieve slightly better outcomes than those used Autoencoder + Softmax. This is probably because the latter might suffer from the intrinsic problem of unsupervised learning originated in the part of unsupervised autoencoder, where underfitting might happen without the feedback from the actual labels. Thus, the algorithm might not distinguish two events with high similarity. In comparison, shallow neural network as supervised learning may abate those problems with a sufficiently large number of neurons in the hidden layer and by adopting appropriate training processing.

Overall, prediction accuracy of 94% was achieved by supervised learning compared to >71% for unsupervised learnings. Most errors for unsupervised learnings resulted from either underfitting (the program fails to distinguish between two defects) or overfitting (the program categorizes a defect into both of two classes). These errors are hard to be corrected without feedback from the actual labels. After adding a softmax layer at the output stage, the overfitting and underfitting errors were able to be compensated. The SOM showed mediocre performance, while the autoencoder achieved over 94% accuracy in conjunction with softmax. It is a rather satisfactory result to distinguish seven scenarios as pertaining to various pipeline corrosion types.

4. Discussion

While this paper shows that the distributed acoustic fiber sensors using Rayleigh enhanced fibers and machine learning data analytics can significantly improve detection efficacies against external intrusion events and internal corrosion, it does have its own limitation. For different pipeline configurations and shapes, it is very likely that neutral networks will have to be re-trained. However, most of sections of pipeline infrastructures are constructed by straight pipes made of identical materials, wall thickness, and diameters. These straight pipes are connected by a few different types of joint structures such as elbow (this paper), T-junction, and etc. Therefore, studies like ours can be useful with relatively wide applicability. A natural extension of research from this paper will be experimental studies of different pipe structures with different topologies and a "digital twin" approach, while data from simulations of different pipe structures are combined with experimental data to train neural network, to accelerate machine learning process to reach better detection and classification accuracy and to expand its applicability.

Optics EXPRESS

In conclusion, this paper presents an integrated approach using advanced fiber sensors, distributed sensor interrogation schemes, and artificial intelligence big data analytics to protect large-scale critical infrastructures. It uses femtosecond laser direct writing schemes to enhance Rayleigh backscattering of sensing fiber to drastically improve SNR for DAS systems. Through high-sensitivity detections of guided acoustic waves in pipeline structures induced by external impact events and internal defects, high spatial resolution sensing data can be classified by artificial intelligence data analytics. The integrated DAS and machine learning algorithm achieved over 85% accuracy to identify various external impact events. Through measurements of both forward-propagating and backward-scattered acoustic wave by defects incurred on pipe elbow structures, deep neural network machine learning algorithm achieved 71% accuracy through unsupervised learning. By manually associating unlabeled output of unsupervised learning results with the corrosion types, a softmax layer served as supervised optimization that increased the accuracies to 94% in the scheme of autoencoder. The capability of distributed measurements to use fiber sensors to perform high spatial resolution measurements integrated into an effective machine learning algorithm could have potential to significantly improve safety and enable condition-based maintenance of pipeline infrastructures for a wide array of industries.

Funding

Office of Nuclear Energy (DE-AC07-05ID14517); Office of Fossil Energy (DE-FE00029063).

Acknowledgments

This work was also funded by the Department of Energy, National Energy Technology Laboratory, an agency of the United States Government, through a support contract (RSS 89243318CFE000003) with Leidos Research Support Team (LRST).

Disclosures

The authors declare no conflicts of interest.

References

1. M. Evans and K. Vine, "Permanently installed transducers for guided wave monitoring of pipelines," in *10th European conference on Non-destructive testing, Moscow, Russia* (2010), pp. 1–6.
2. P. del Hougne, M. F. Imani, T. Sleasman, J. N. Gollub, M. Fink, G. Lerosey, and D. R. Smith, "Dynamic metasurface aperture as smart around-the-corner motion detector," Sci. Rep. **8**(1), 6536 (2018).
3. P. del Hougne, "Robust Position Sensing with Wave Fingerprints in Dynamic Complex Environments," arXiv preprint arXiv:.14513 (2020).
4. W.-B. Na and T. Kundu, "Underwater pipeline inspection using guided waves," Pressure Vessel Technol. **124**(2), 196–200 (2002).
5. M. Klann and T. Beuker, "Pipeline Inspection With the High Resolution EMAT ILI-Tool: Report on Full-Scale Testing and Field Trials," in *2006 International Pipeline Conference* (American Society of Mechanical Engineers Digital Collection, 2006), pp. 235–241.
6. C. Liu, J. Dobson, and P. Cawley, "Efficient generation of receiver operating characteristics for the evaluation of damage detection in practical structural health monitoring applications," Proc. R. Soc. London, Ser. A **473**(2199), 20160736 (2017).
7. P. Cawley, "Practical long range guided wave inspection—Managing complexity," in *AIP Conference Proceedings* (AIP, 2003), pp. 22–40.
8. V. Spirin, M. Shlyagin, S. Miridonov, F. M. Jimenez, and R. L. Gutierrez, "Fiber Bragg grating sensor for petroleum hydrocarbon leak detection," Opt. Lasers Eng. **32**(5), 497–503 (1999).
9. J. Buerck, S. Roth, K. Kraemer, and H. Mathieu, "OTDR fiber-optical chemical sensor system for detection and location of hydrocarbon leakage," J. Hazard. Mater. **102**(1), 13–28 (2003).
10. Y. Shi, H. Feng, Y. An, X. Feng, and Z. Zeng, "Research on wavelet analysis for pipeline pre-warning system based on phase-sensitive optical time domain reflectometry," in *2014 IEEE/ASME International Conference on Advanced Intelligent Mechatronics* (IEEE, 2014), pp. 1177–1182.
11. H. Wu, Z. Sun, Y. Qian, T. Zhang, and Y. Rao, "A hydrostatic leak test for water pipeline by using distributed optical fiber vibration sensing system," in *Fifth Asia-Pacific Optical Sensors Conference* (International Society for Optics and Photonics, 2015), p. 965543.

Optics EXPRESS

12. Z. Wang, J. Zeng, J. Li, F. Peng, L. Zhang, Y. Zhou, H. Wu, and Y. Rao, "175 km phase-sensitive OTDR with hybrid distributed amplification," in *23rd international conference on optical fibre sensors* (International Society for Optics and Photonics, 2014), p. 9157D9155.
13. X. He, S. Xie, F. Liu, S. Cao, L. Gu, X. Zheng, and M. Zhang, "Multi-event waveform-retrieved distributed optical fiber acoustic sensor using dual-pulse heterodyne phase-sensitive OTDR," Opt. Lett. **42**(3), 442–445 (2017).
14. X. Bao and L. Chen, "Recent progress in distributed fiber optic sensors," Sensors **12**(7), 8601–8639 (2012).
15. 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," Appl. Sci. **7**(8), 841 (2017).
16. Y. Lu, T. Zhu, L. Chen, and X. Bao, "Distributed vibration sensor based on coherent detection of phase-OTDR," J. Lightwave Technol. **28**(22), 3243–3249 (2010).
17. Z. Qin, L. Chen, and X. Bao, "Wavelet denoising method for improving detection performance of distributed vibration sensor," IEEE Photonics Technol. Lett. **24**(7), 542–544 (2012).
18. F. Jiang, H. Li, Z. Zhang, and X. Zhang, "An event recognition method for fiber distributed acoustic sensing systems based on the combination of MFCC and CNN," in *2017 International Conference on Optical Instruments and Technology: Advanced Optical Sensors and Applications* (International Society for Optics and Photonics, 2018), p. 1061804.
19. D. Iida, K. Toge, and T. Manabe, "High-frequency distributed acoustic sensing faster than repetition limit with frequency-multiplexed phase-OTDR," in *2016 Optical Fiber Communications Conference and Exhibition (OFC)* (IEEE, 2016), pp. 1–3.
20. S. Loranger, M. Gagné, V. Lambin-Iezzi, and R. Kashyap, "Rayleigh scatter based order of magnitude increase in distributed temperature and strain sensing by simple UV exposure of optical fibre," Sci. Rep. **5**(1), 11177 (2015).
21. J. Jin, H. Zhang, J. Liu, and Y. Li, "Distributed Temperature Sensing Based on Rayleigh Scattering in Irradiated Optical Fiber," IEEE Sens. J. **16**(24), 8928–8935 (2016).
22. F. Zhu, Y. Zhang, L. Xia, X. Wu, and X. Zhang, "Improved Φ-OTDR sensing system for high-precision dynamic strain measurement based on ultra-weak fiber Bragg grating array," J. Lightwave Technol. **33**(23), 4775–4780 (2015).
23. C. Wang, Y. Shang, X.-H. Liu, C. Wang, H.-H. Yu, D.-S. Jiang, and G.-D. Peng, "Distributed OTDR-interferometric sensing network with identical ultra-weak fiber Bragg gratings," Opt. Express **23**(22), 29038–29046 (2015).
24. W. W. Morey, G. Meltz, and J. M. Weiss, "High-temperature capabilities and limitations of fiber grating sensors," in *Tenth International Conference on Optical Fibre Sensors* (International Society for Optics and Photonics, 1994), pp. 234–237.
25. M. Wang, M. A. Zaghloul, S. Huang, A. Yan, S. Li, R. Zou, P. Ohodnicki, M. Buric, M.-J. Li, and D. Carpenter, "Ultrafast Laser Enhanced Rayleigh Backscattering on Silica Fiber for Distributed Sensing under Harsh Environment," in *CLEO: Applications and Technology* (Optical Society of America, 2018), p. ATh3P.4.
26. S. Huang, M. Li, S. M. Garner, M.-J. Li, and K. P. Chen, "Flexible photonic components in glass substrates," Opt. Express **23**(17), 22532–22543 (2015).
27. A. Yan, S. Huang, S. Li, R. Chen, P. Ohodnicki, M. Buric, S. Lee, M.-J. Li, and K. P. Chen, "Distributed Optical Fiber Sensors with Ultrafast Laser Enhanced Rayleigh Backscattering Profiles for Real-Time Monitoring of Solid Oxide Fuel Cell Operations," Sci. Rep. **7**(1), 9360 (2017).
28. Z. Peng, J. Jian, H. Wen, M. Wang, H. Liu, D. Jiang, Z. Mao, and K. P. Chen, "Fiber-optical distributed acoustic sensing signal enhancements using ultrafast laser and artificial intelligence for human movement detection and pipeline monitoring," in *Optical Data Science II* (International Society for Optics and Photonics, 2019), p. 109370J.
29. M. Wang, Y. Yang, S. Huang, J. Wu, K. Zhao, Y. Li, Z. Peng, R. Zou, H. Lan, and P. R. Ohodnicki, "Multiplexable high-temperature stable and low-loss intrinsic Fabry-Perot in-fiber sensors through nanograting engineering," Opt. Express **28**(14), 20225–20235 (2020).
30. C. Wang, C. Wang, Y. Shang, X. Liu, and G. Peng, "Distributed acoustic mapping based on interferometry of phase optical time-domain reflectometry," Opt. Commun. **346**, 172–177 (2015).
31. Z. Zhao, M. S. Demokan, and M. MacAlpine, "Improved demodulation scheme for fiber optic interferometers using an asymmetric 3 (3 coupler)," J. Lightwave Technol. **15**(11), 2059–2068 (1997).