Orchestrating a brighter world

AI Enabled Optical Fiber Anomaly Detection

By: Anugraha Sinha (CTO Office)

Abstract

Fibre Optics cable acts as the backbone for providing last-mile connectivity for growing internet consumption within the masses. Apart from providing long-distance network connectivity, these cables are also used along critical industry establishments to provide surveillance capabilities, for example, pipelines used for natural gas transmission. For data transmission setups, these cables provide low-loss & high bandwidth data transmission and also high sensitivity to surrounding vibrations. However, fiber optics cables are considerably more susceptible to damage due to their inherent architectural setup.

In the industry fiber optics cables are used in multiple types of installation, such as aerial cables, and buried cables, and in some unique cases, the same fiber length is placed in an aerial setup for some distance and buried set up for another piece of distance. Distributed Acoustic System (DAS), is especially well suited for monitoring hundreds of kilometers of fiber length with a single Sensitive Optical Time Domain Reflectometer). Distributed Acoustic Systems (DAS) can be applied in varied installations and assist in converting vibration intensities received over fiber optics cable into digital data. Multiple attempts have been made by the research community to build pattern recognition systems (PRS) using machine [1] learning techniques to analyze, predict anomalies and classify vibrations into different categories, thereby, attempting to avoid damage to fiber optics infrastructure or the targeted industry infrastructure in surveillance centric implementation. However, the majority of the pattern recognition methods used up till now can be grouped under the following categories:

Rule-based methods

- Employing use-case and domain-specific thresholds for classifying anomalies

Discriminative model-based methods

- Support Vector Machine (SVM) - (linear & non-linear kernels)
- Simple neural networks (fully connected), acting as classification models.

Generative model-based methods

- Gaussian Mixture Models (GMM) - (with and without postprocessing)
- Convolutional Neural Network 2 Dimensional (CNN2D) - (single channel) based auto-encoders with regression-based classification

Introduction

Anomaly detection in DAS-based FiberOptics vibrations has been a popular area for research for many years. While a lot of research has been done for achieving this using advanced signal processing-based methods and also using conventional unsupervised and supervised machine learning approaches, the application of deep learning-based methods has been quite limited. There have been sporadic references where deep learning-based methods have been applied to FiberOptics vibrations, however, even that has been from a discriminative modeling approach perspective. The recent advancement in generative algorithms using deep learning has rarely been tried in this domain. This research aimed to evaluate the applicability and efficacy of generative modeling techniques for anomaly detection for DASbased FiberOptics vibrations, which takes the research in this domain to a new level.

While this research evaluates the conventional deep learning generative algorithm (AutoEncoder) for anomaly detection, it also evaluates some of the most advanced and recent development in the area of GAN for anomaly detection. It is worthwhile to mention that architectures like AnoGAN, EGBAD and, Ganomaly were originally conceptualized for anomaly detection on image-based data, however, this research applies the same ideology to fiber optics vibrations. This research hypothesizes the correlation between the 2 domains (image and FiberOptics vibrations) by considering the spatio-temporal nature of fiber optics vibrations to be in line with space and time information expressed in images (time is a static quantity in each image, and in fiber optics vibrations, a specific time window is considered as a single element). Additionally, the frequency element in fiber optics vibrations is related to color channels in image-based data, because, each color channel presents information from the same space-time combination from a different frequency (RGB) domain. Similarly, a timelocation window in fiber optics vibrations could be conceptualized with multiple frequencies over the channel axis.

This research has also been able to draw empirical comparisons between the efficacies of different modeling techniques. Additionally, models have also been compared from a more real-life scenario perspective by using spectrogram-based graphical evaluation technique, and this evaluation also correlated to the conclusion drawn through empirical evaluation.

Rule-Based vs Discriminative model vs Generative model

Rule-based and discriminative model-based methods pose some major challenges

- Getting labeled data.
- These methods can classify vibrations among a pre-defined set of classes, resulting in frequent model training with changing operating conditions.
- Sensitivity to changing environmental conditions beyond the distribution of training data.

Within the scope of generative model-based methods, conventional methods like GMM lack performance when applied to real-world conditions. CNN2D-based autoencoders (Oh & Yun, 2018) have taken one step ahead in implementing complex neural network-based methods for anomaly detection. However, fundamentally, it is known that unconstrained auto-encoders are not the best architectures for building robust lower latent space representations.

The current state-of-art generative models are based on adversarial learning (Goodfellow et al., 2014), also known as GAN (Generative Adversarial Networks). GANbased generative models provide (virtually) unlimited scope for performing modeling in multiple different manners, thereby, providing a wider scope for addressing the problem of pattern recognition for vibrations obtained from DAS.

This research aims at implementing GANbased anomaly detection methods to vibration data obtained from DAS, with the objective for the application of this method to real-world conditions.
Distributed acoustic sensing for fiber optics cable is based on a physical phenomenon called Rayleigh scattering that is naturally seen in fiber optics. Molecular-level imperfections in the fiber act as micromirrors resulting in partial backscattering. This backscattering phenomenon is termed Rayleigh scattering, and the micromirrors are also often called scattering centers (Aktas et al., 2017). The principle of Rayleigh scattering is employed in mainly two different manners for measurement.

- Phase Sensitive Optical Time Domain Reflectometry (OTDR)
- Optical Frequency Domain Reflectometry (OFDR)

This research concentrates on measurement based on OTDR method. DAS vibration sensing within the scope of OTDR based measurement technology has shown promising results for monitoring potential threats for fiber optics cables or vibrations in general. OTDR provides high sensitivity to fiber optics vibrations and also spatial resolution in the order of tens of kilometers and has been extended even over 100kms by using optical amplification. (Tejedor et al., 2017).

Extensive research has been done on the application of DAS+PRS to classify the vibration in one or more classes. Before the application of machine learning techniques to DAS vibration data, historically, simple threshold-based strategies on the energy values received were used for - based data systems (Tejedor et al., 2017).

Discriminative model-based methods have been employed in prior research for anomaly detection and also for vibration classification. Rule-based and statistical generative algorithm model-based approach towards anomaly detection has also been a choice in prior research. While discriminative models are a strong candidate for pattern recognition for vibration data, however, there are some fundamental challenges (such as availability of labelled data, compatibility with varying environmental conditions) around this methodology (Aktas et al., 2017). Similarly, unsupervised learning methods and conventional generative modelling also pose challenges when tackling real-world conditions (Tejedor et al., 2017).

Pioneering work was accomplished by (Oh & Yun, 2018), wherein auto-encoder architectures were used for anomaly detection for machine sounds. However, such an implementation has not been seen for vibration data obtained from DAS for fiber optics cables.

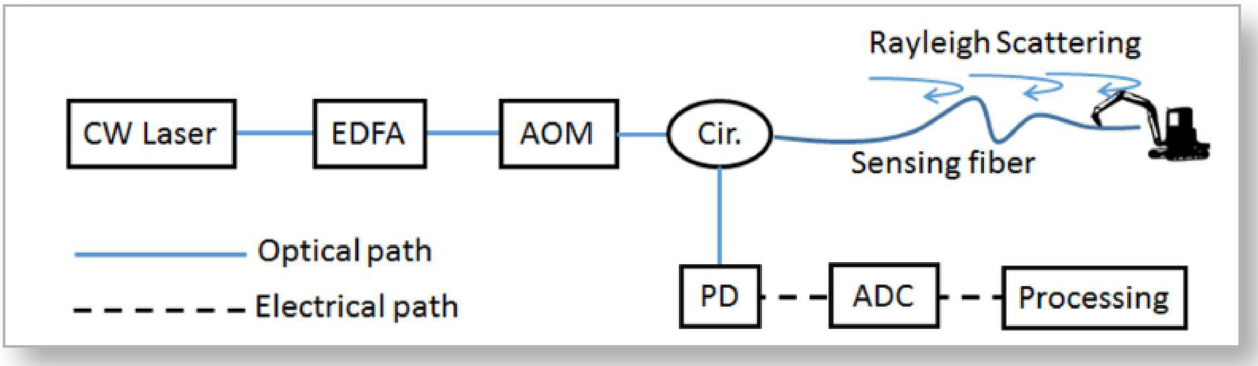
(Goodfellow et al., 2014) proposed a new family of deep learning-based generative algorithms, based on adversarial learning. This family of algorithms is commonly known as GANs (Generative Adversarial Networks). In recent years, GANs have gone through an exponential evolution. (Gui et al., 2020). While GANs have been extensively used in the context of generating data samples that are theoretically inexistent, they have also been extended for anomaly detection. Considerable research has been done in the application of GANs for anomaly detection (Di Mattia et al., 2019), and some of the notable contributions have been around AnoGAN (Schlegl et al., 2017), EGBAD (Efficient GAN Based Anomaly Detection) (Zenati et al., 2018) & GANomaly (Akcay et al., 2019). However, all of the GAN-based anomaly detection research has concentrated on image data but not on acoustic and non-acoustic vibration data.

This research explores GANs and their application for anomaly detection in a detailed manner. It also aims at evaluating the capability of GANs-based anomaly detection algorithms for vibration data obtained from DAS. An extensive literature review has revealed that such an evaluation and experimentation has not been done before, and therefore, this research would like to put forward this as a unique application of GANs with the domain of PRS for DAS.

Fiber optics cables have revolutionized the way a common household uses internet connectivity and also the way surveillance is carried out at high-security and critical industry establishments. In a surveillance application, the high sensitivity of fiber optics cables toward surrounding vibrations enables proximity detection and also the prevention of damage to associated industries or high-security establishments. In the application of data transmission, these cables act as the backbone for the "Age of Data". Such critical applications and usage of fiber optics cable warrant the need for a perpetual protection system that safeguards the physical and operational sanctity of the cables. Potential damage to this infrastructure may pose serious implications for the common public and industry at large. According to the DIRT Report (2019)*, approximately USD 600 million was spent as a direct repair cost for fiber optics infrastructure in the United States.

Considerable research has been done over the past 2.5 decades to build systems that could analyze the vibration picked by fiber optics cable. This study aims at augmenting this area of research with the application of generative modeling for anomaly detection to provide the enhanced capability of ubiquitous monitoring and safeguarding the fiber optics infrastructure.

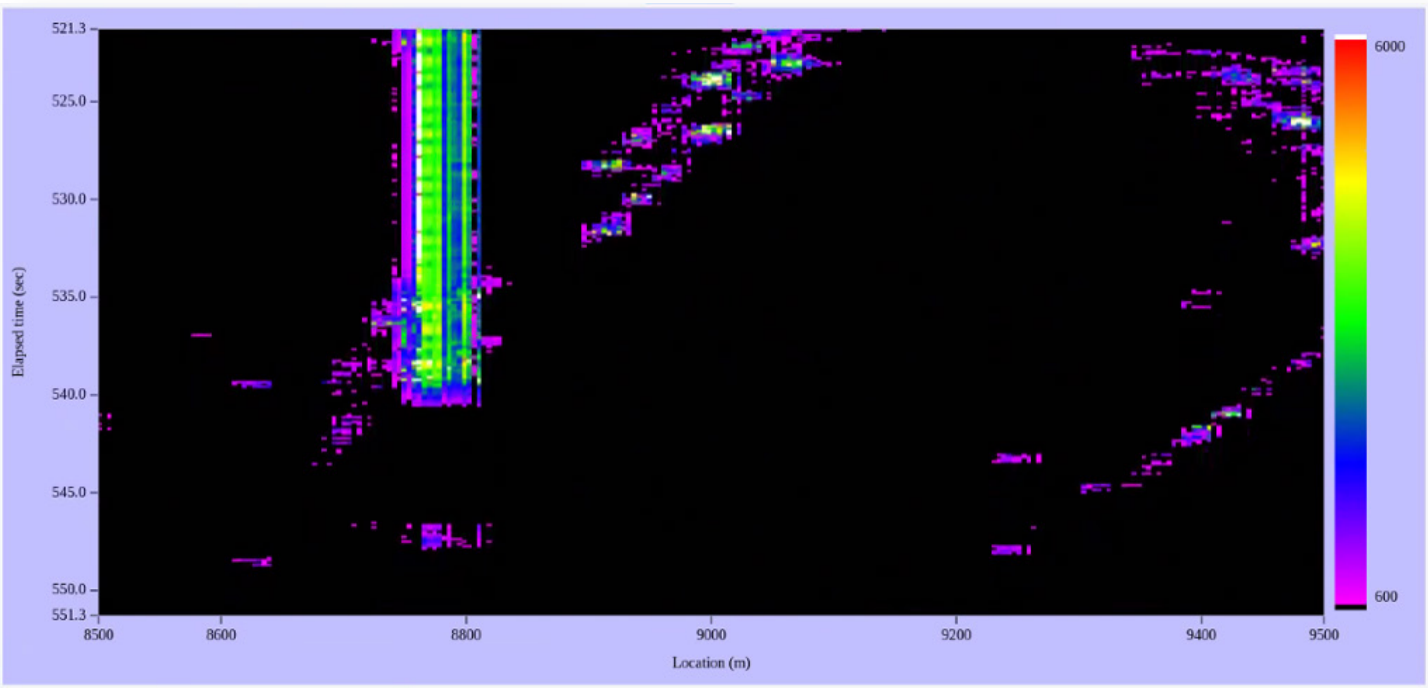
There are multiple different ways (apparatus) in which DAS is employed with fiber optics cable for vibration measurement. While there has been extensive research and experimentation done in this area, this research covers one of the widely used methods for data generation. It is important to mention that the fiber specification mentioned below are specific to the experiment carried out for this research and may differ based on the equipment manufacturer's specifications and supported features.
There are two detection methods for both OTDR and OFDR methodologies, direction detection method and coherent (heterodyne) detection method. Direct detection methods are simpler to implement and it expects optical signals associated with fewer constraints. However, the coherent detection method has higher SNR (signal-to[1]noise ratio) and dynamic range for measurement, provided coherency is preserved. (Aktas et al., 2017)



The block diagram of the DAS system (Aktas et al., 2017)

The figure presents a DAS system as an example that uses a continuous ware with narrow linewidth (<1 kHz) and a (CW) laser having a wavelength of 1550nm. The amplification is provided by an Erbium-doped fiber amplifier (EDFA), thereafter the signal is fed to an acousto-optics modulator (AOM) which generated the required pulses. The generated pulses are fed into the fiber using a circulator. The AOM is also supplied with an electrical impulse with a width of 100ns. This light traveling inside the fiber experiences Rayleigh scattering. External vibrations result in an optical phase shift on the backscattered light. This phase shift is detected by the photodetector (PD). The output of the photodetector becomes the input to the analytical system for classification and anomaly detection. In the direct detection approach, there is no demodulation performed on the backscattered light. It is important to mention that clock synchronization between AOM and ADC is of utmost importance because a mismatch in the clock frequency would result in erroneous harmonics in the digital data obtained from the photodetector.

The data acquisition rate of the sensor is about ~ 370Msamples/sec, however, this is reduced to 10Msamples/sec to avoid signal fading phenomenon. Each data sample is a 16-bit integer [-32678, 32678], therefore, the data input rate equates to ~ 152.6Mbits/secs, which is equivalent to 19MB of data generated every second. Realtime visualization of this data can be seen in a waterfall graph (as shown below). Signal amplitude is color-coded based on the legend provided on the graph. The waterfall graph below shows the visualization of a slice of cable from 8.5km to 9.5km only.



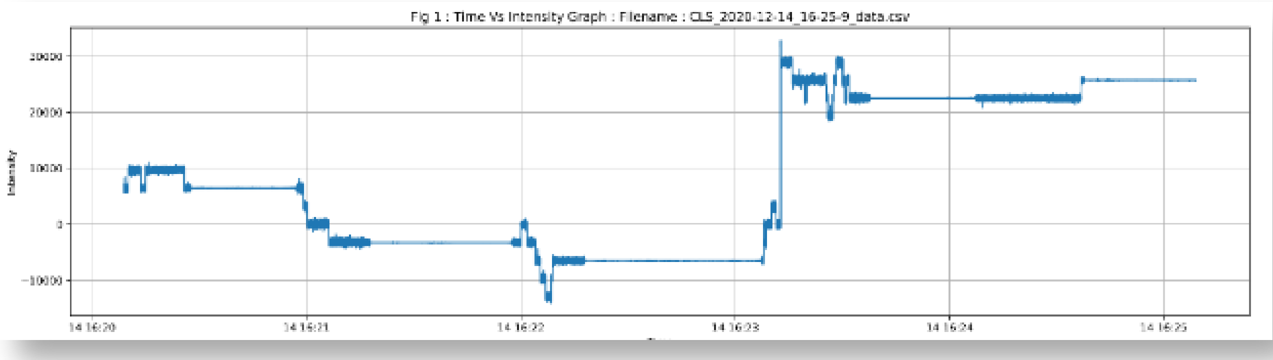
Waterfall diagram of a vibrator machine sensed at ~8.7 km range with the cable buried at a depth of 1 meter.

In φ-OTDR-based systems, the transduction function will not be stationary as it will be heavily affected by environmental conditions, varying along time and location. This is especially relevant when the DAS system is based on amplitude measurements, which inherently imply non-linear behavior, except for very small perturbations. This non-stationary (and in some cases non-linear) response, can be observed by analyzing the signal resulting from the detection of pure vibration frequencies: the φ-OTDR-based recorded signal will include amplitude varying harmonics and sub-harmonics along with the original vibration frequency. The linear transduction mechanism of the sensing systems implies that the acquired signals will have reasonably consistent behavior, thus providing a favorable scenario for the classification task.

The data generation process employed in some of the research did not consider reallife conditions. Sometimes the sensing equipment and sensing area were too close (Madsen et al., 2007). There are cases where the sensed area was also too small (Wu et al., 2014a). It has also been noticed that in some research, simulated data has been used, thereby ignoring the complications and complexities of the real-life environment. (Martins et al., 2013). Some recent work published has addressed prior issues with data generation, where real-life data were used from different fiber lengths, such as 17kms (Papp et al., 2016b), 24kms (Wu et al., 2015), 220kms (Wu et al., 2014b). Some research has also approached the problem from a more realistic viewpoint. In the research presented by (Martins et al., 2015), cross-validation sets from the data were also used to validate the model on data points not seen during the training process. Around the same period, research presented by (Wu et al., 2015) & (Papp et al., 2016b) performs vibration measurement at the same point, therefore, biasing the model training process towards the point of vibration rather than vibration features.

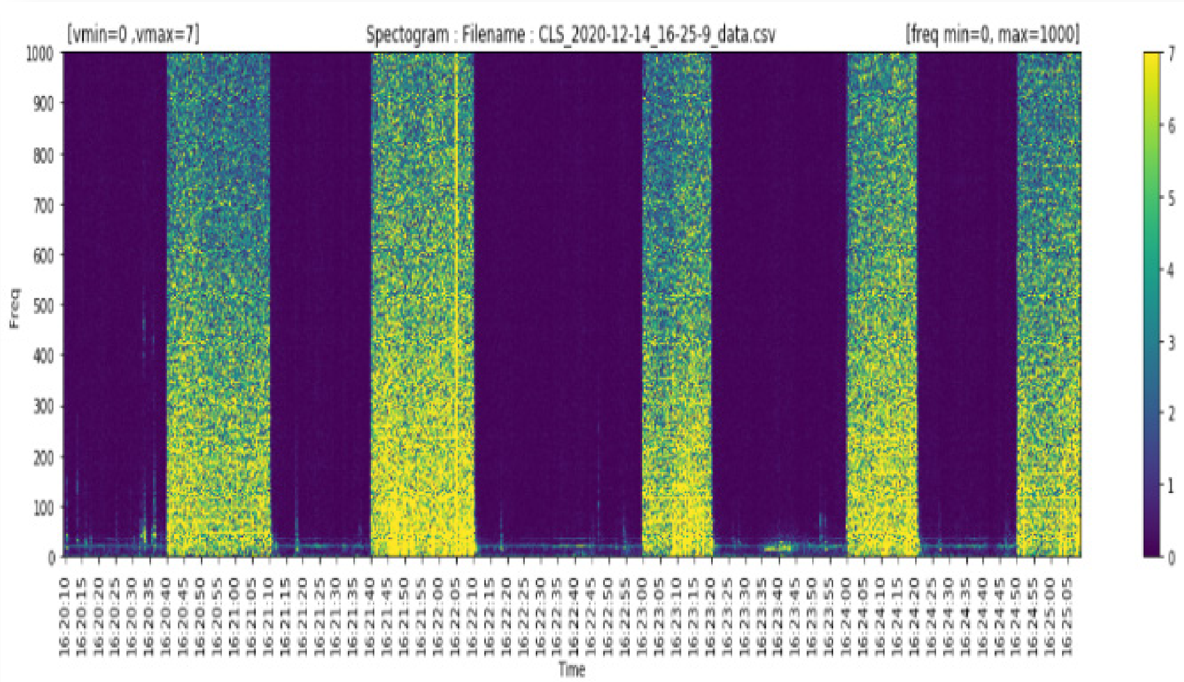
Data generation fundamental and techniques around fiber optics cables have come to an advanced maturity level, and concepts related OTDR and OFDR has become an industry standard. Along with this, HW systems have also been standardized with the capability of providing data based on customizable settings as per \the need of the application. However, feature extraction methods and vibration analysis methodologies are still popular areas of research within the community. The research community has been implementing different machine learning approaches for performing anomaly detection, deep learning-based applications have been few. There has been some pioneering work related to the application of deep learning-based auto-encoders for anomaly detection, however, the recent state-of-the-art approaches for anomaly detection have not been explored for vibration data obtained from fiber optics cables. This research tries to explore the efficacy of the application of recent advancements in the field of GANs for anomaly detection for fiber optics vibration data to perform anomaly detection.

As explained in the previous section, the data is made available in a matrix, where rows signify the temporal axis and columns signify the spatial axis. Considering only 1 of anomalous location for understanding and visualizing data, the raw intensity value visualization graph is shown below.



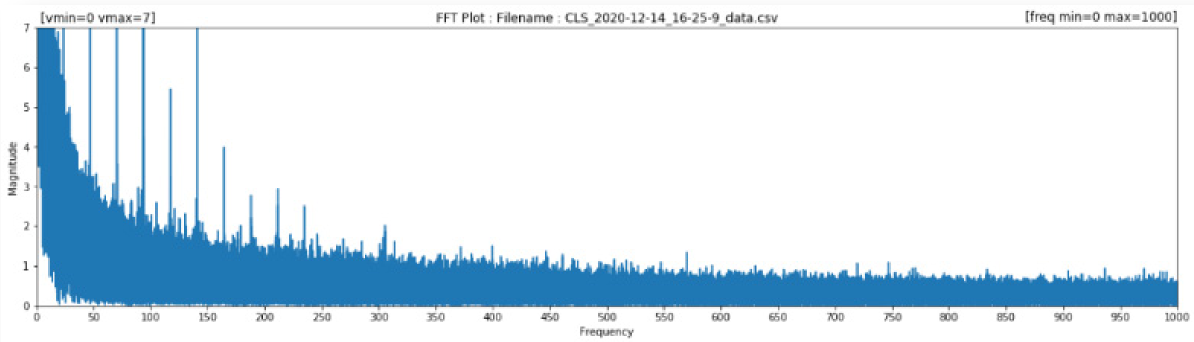
Time Vs Intensity Graph (Raw DAS Data)

A spectrogram plot of the data is shown below



Spectrogram For the data considered for visualization, maximum intensity is present for frequencies <= 500Hz.

A depiction of data from the FFT perspective is shown below

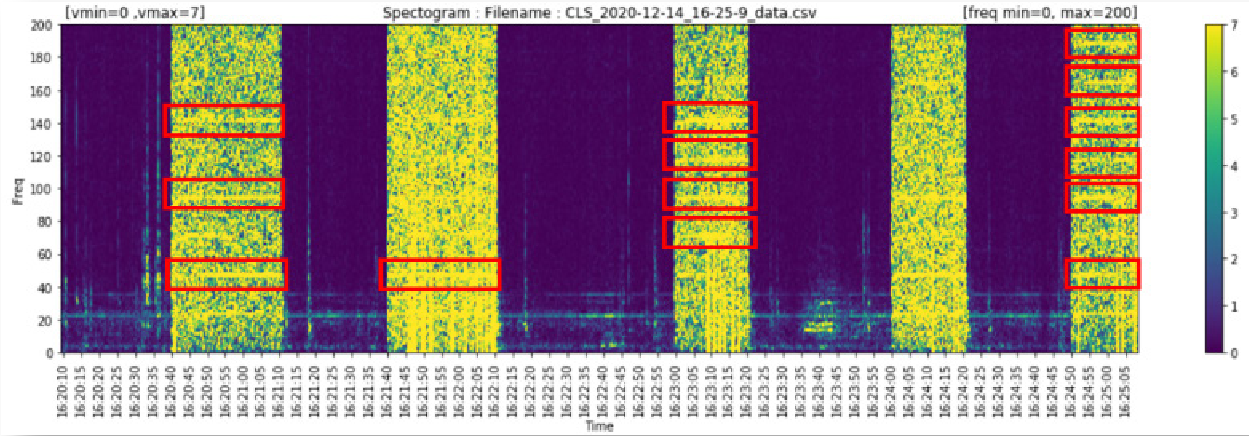


FFT Plot

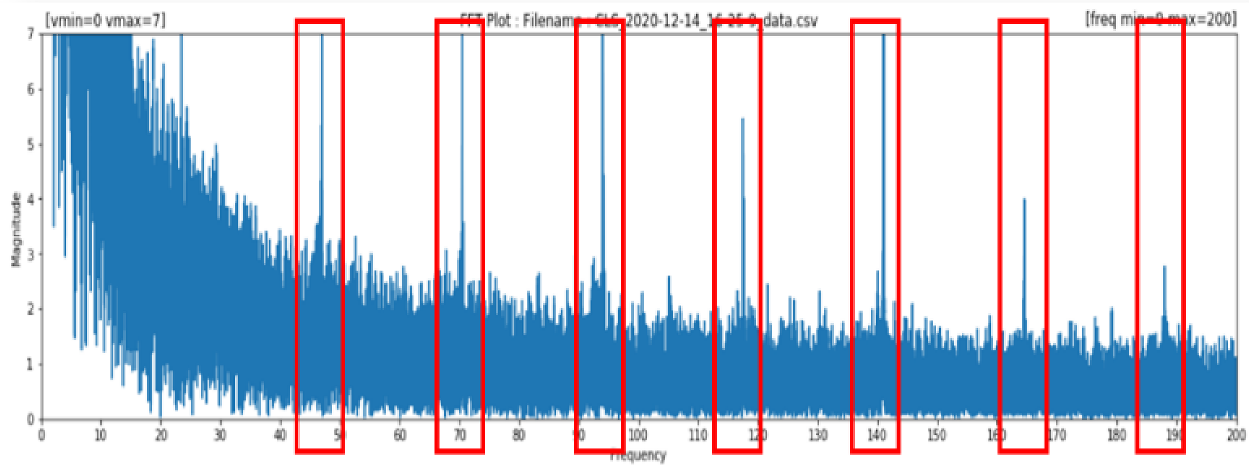
Comparing the spectrogram and FFT plot, the following can be inferred

- The majority of the signal has frequencies <= 500Hz
- 99 percentile of magnitude data <= 7.0

There are some harmonic frequencies present around 45Hz, 70Hz, 95Hz, 118Hz, 140Hz, 165Hz, 188Hz. This can be seen from the spectrogram and FFT plots below:



Spectrogram Plot (Limited Freq ϵ [0,200])



FFT Plot (Limited Freq ϵ [0,200])

Another important point to notice here is that pre-processing based on FFT only provides frequency-domain information. However, from the spectrogram, it can be seen that the harmonic frequencies are present at different time periods. The spectrogram-based pre-processing is similar to time-frequency domain-based data pre-processing. More details about data pre-processing are covered in the next section.

While evaluation metrics provide a good empirical viewpoint for comparing different modeling techniques and respective variations, however, it is also important to evaluate the models from a real-life perspective viewpoint. This research has extensively used spectrogram-based visualization in performing data understanding, data preparation and, data transformation methods. Therefore, it is also important to use the same technique and perform model performance evaluation and judge which model is performing better and where all do pitfalls occur.

To achieve the above evaluation methodology, some specific data files were chosen which could simulate real-world scenarios.

- **Mixed Vibrations (Anomalies and Environmental Vibrations)**
  - In this scenario, there are timesteps where it is expected that the model should detect anomalies as there are high-intensity vibrations at lower frequencies. This is a complex scenario, where the models are expected to have high precision and high recall while performing prediction
- **Maximum Anomaly Vibrations**
  - In this scenario, there are long-term anomaly vibrations and the models are expected to predict a high recall rate.
- **Only Environmental Vibrations**
  - In this scenario, there are only environmental vibrations, and the models are expected to predict them as non-anomalous.

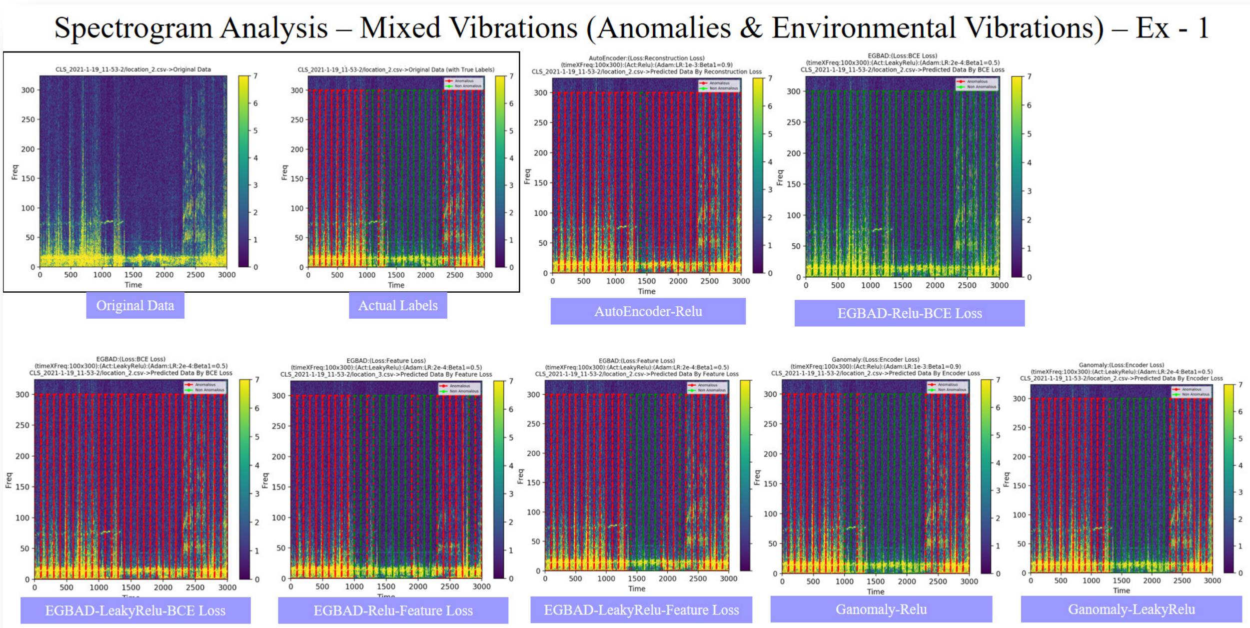
Table: Summary of Visualization of Prediction by Spectrogram

Summary – Visualization of Prediction by Spectrogram

Scenario	Example	AutoEncoder Relu	EGBAD – BCE (Relu)	EGBAD – BCE (LeakyRelu)	EGBAD – Feature (Relu)	EGBAD – Feature (LeakyRelu)	Ganomaly - Relu	Ganomaly – Leaky Relu	Ref Figure
Mixed Vibrations	Example 1								
	Non-Anomalous Prediction	Very Low	High	Very Low	Moderate	Very Low	High	High	
	Anomalous Prediction	High	Very Low	High	High	High	High	High	
	Example 2								
	Non-Anomalous Prediction	Very Low	Very Low	Moderate	High	High	High	High	
	Anomalous Prediction	High	Very Low	High	High	High	High	High	
Maximum Anomaly Vibrations	Example 1								
	Non-Anomalous Prediction	High	Very Low	High	High	High	High	High	
	Anomalous Prediction	High	Very Low	High	High	High	High	High	
Only Environmental Vibrations	Example 1								
	Non-Anomalous Prediction	High	Moderate	Moderate	High	High	High	High	
	Anomalous Prediction	Moderate	Very Low	Very Low	High	High	High	High	
	Example 2								
	Non-Anomalous Prediction	Very Low	Moderate	Very Low	High	High	High	High	
	Anomalous Prediction	Very Low	Moderate	Very Low	High	High	High	High	

1.1.1. Spectrogram Analysis – Mixed Vibrations (Anomalies & Environmental Vibrations)

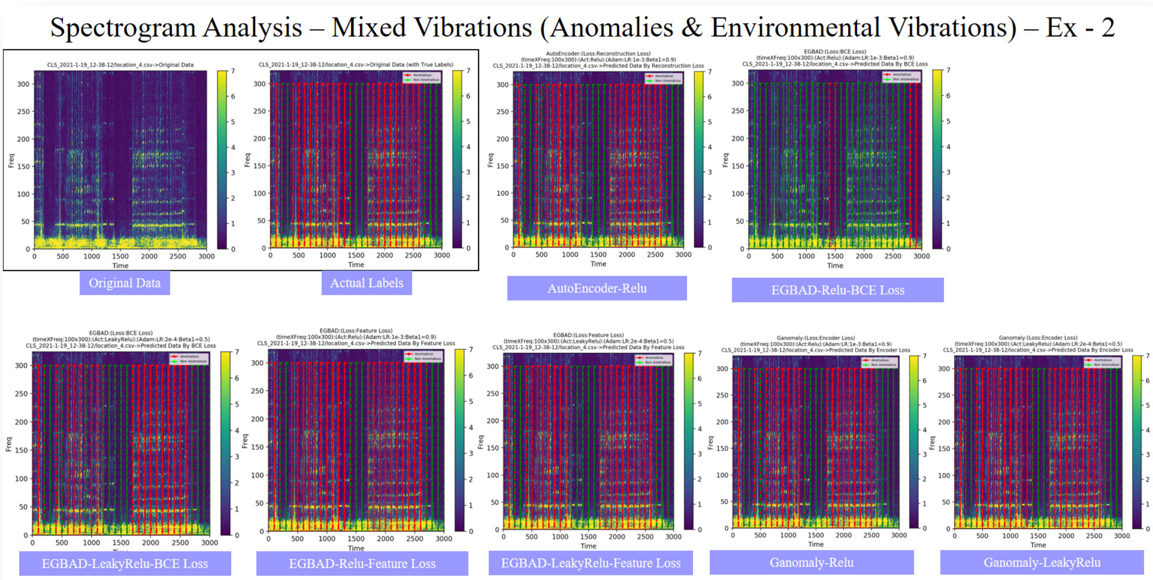
1.1.2. Example – 1



Spectrogram Analysis – Mixed Vibrations – Example-1

Note: Spectrogram plots presented in this research are high-resolution in specifications. The reader is requested to zoom into the prediction's plots for readability & clear viewability.

1.1.3. Spectrogram Analysis – Mixed Vibrations (Anomalies & Environmental Vibrations) – Example - 2



Spectrogram Analysis – Mixed Vibrations – Example – 2

Note: Spectrogram plots presented in this research are high-resolution in specifications. The reader is requested to zoom into the prediction's plots for readability & clear viewability.

1.1.4. Spectrogram Analysis – Maximum Anomaly Vibrations – Example -1

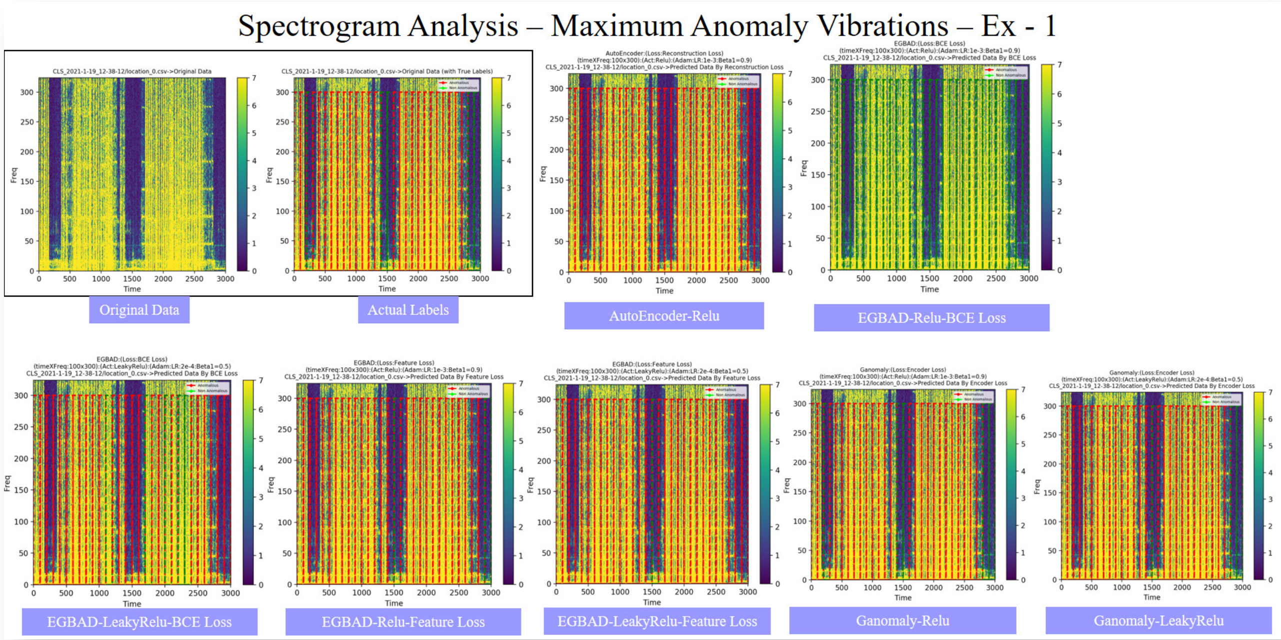


Figure : Spectrogram Analysis – Anomaly Vibrations - Example – 1

Note: Spectrogram plots presented in this research are high-resolution in specifications. The reader is requested to zoom into the prediction's plots for readability & clear viewability.

1.1.5. Spectrogram Analysis – Only Environmental Vibrations – Example – 1

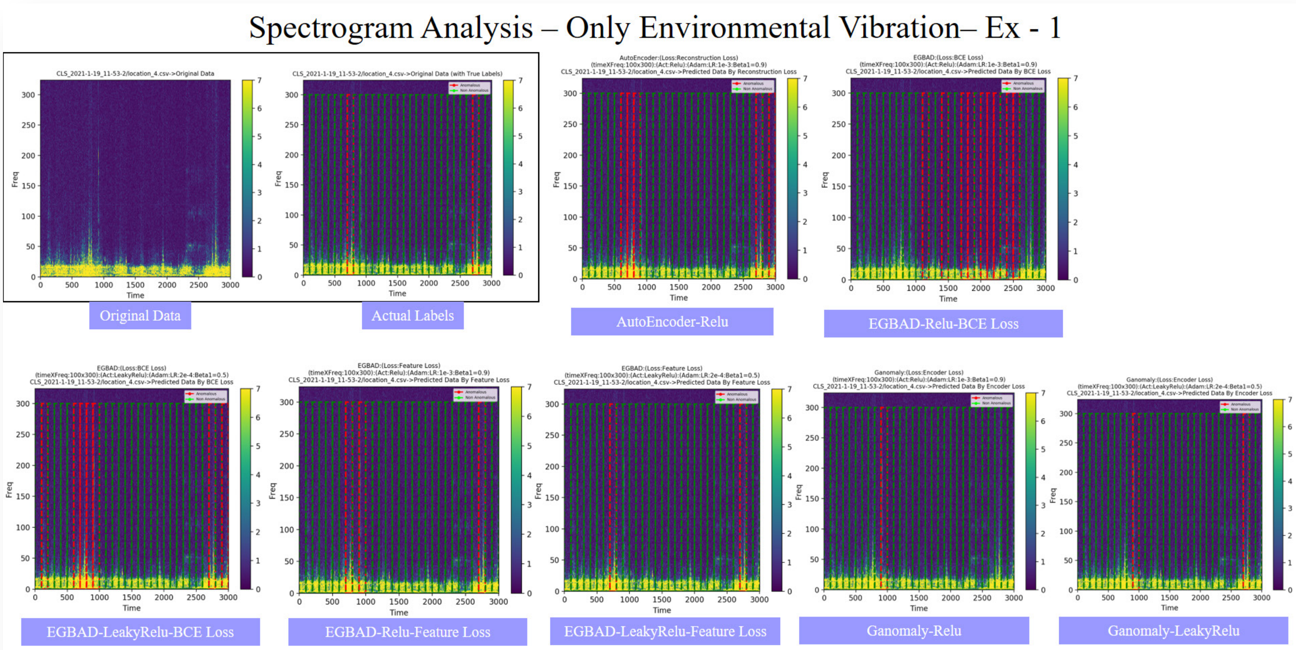


Figure: Spectrogram Analysis – Environmental Vibrations – Example - 1

Note: Spectrogram plots presented in this research are high-resolution in specifications. The reader is requested to zoom into the prediction's plots for readability & clear viewability.

Discussion & Conclusion

One of the key achievements of this research have been to prove the applicability of GAN[1]based anomaly detection algorithms, originally conceptualized for image-based data, to a different domain of DAS-based fiber optics vibrations. Some other key conclusions from this research are categorized and discussed below.

- **Data Processing & Data Transformation Techniques**

Signal processing has been a backbone for analyzing fiber optic vibrations. However, there are multiple ways to process time-series data, starting from simple techniques like moving averages, all the way to Fourier Transform-based processing. DAS systems provide energy domain-based data, which can be converted to the frequency domain using Fourier Transform and also the time[1]frequency-domain using Short Time Fourier Transform. This research uses time-frequency domain-based transformation. Furthermore, this research also evaluates some unique ways to transform time-frequency domain-based data from the STFT algorithm, to build 3D tensors which could be used as input data for deep learning models. It concludes that (Time, Location, Frequency) based 3D tensors align best with Convolution2D-based deep learning models. Such an analysis of data processing and data transformation technique are some of the key contributions of this research.

- **Modeling Techniques**

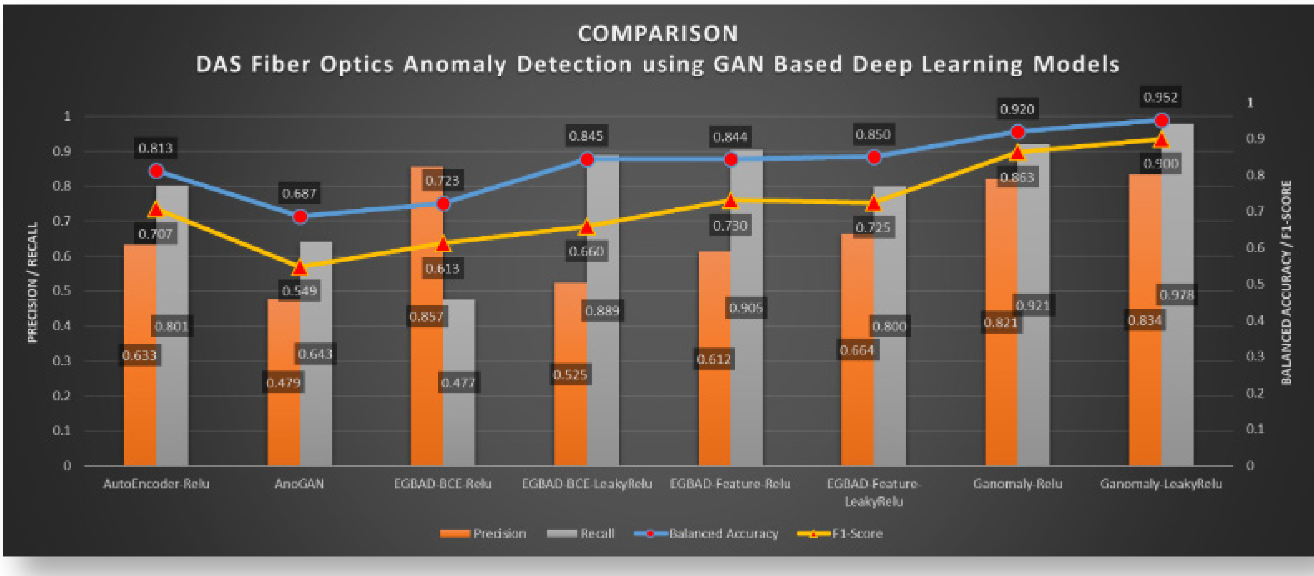
- Among the 4 modeling architectures, evaluation viz-a-viz Auto Encoder, AnoGAN, EGBAD, and, Ganomaly-based anomaly detection technique is providing the best evaluation metrics. This can be validated from the spectrogram-based evaluation graphs as well.

- Auto Encoder-based anomaly detection models also show efficacy in many practical evaluations, however, their sensitivity to environmental noise is quite high, leading to higher False Positive predictions.

- AnoGAN does not have good Precision and cannot be chosen as a practical application methodology. Additionally, their backpropagation-based prediction methodology does not fit the requirement for any practical application.

- Within the scope of EGBAD-based models, binary-cross-entropy loss methods show disappointing spectrogram evaluation results. Conversely, the feature loss[1]based EGBAD method has shown considerable balanced accuracy, however, environmental noise sensitivity leading to False Positives and False Negatives are also seen in multiple cases.

- Ganomaly is the most recent development in GAN-based anomaly detection techniques. One of the key advantages of Ganomaly architecture is the ability to provide lower dimension latent space representation of input data and also provide a generative capability from lower dimension latent space. Ganomaly-based anomaly detection methods have come out to be best within the scope of this research, and provide exceptional balanced accuracy and F1-Score. This can also be seen in the spectrogram-based graphical evaluation methods. The model evaluation methodology for Ganomaly is also one of the simplest leading to the conclusion that it is a strong candidate for practical application as well.



Comparison DAS Fiber Optics Anomaly Detection – GAN Based Models

Spectrogram-Based Model Evaluation Methodology

Machine learning and Artificial Intelligence-based research usually judge the models only based on evaluation metrics, such as accuracy, AUC/ROC, F1-Score, Precision/Recall, etc. This research evaluates the models from a more practical viewpoint as well, by visualizing the predictions provided by the models over spectrograms. These evaluations have been done from multiple different real-life scenario perspectives (mixed vibrations, majority anomaly vibrations & environmental vibrations scenarios), thereby adding value to the way how models should be evaluated. This also adds another dimension to the efficacy of the models from a practical application perspective.

The results presented as part of this research do conclude that Ganomaly-based anomaly detection is the best among the compared modeling technique, however, it is also important to mention that this comparison is based on the set of variations experimented.

Deep Learning provides a wide array of hyper-parameters and virtually an infinite scope of modeling architectures. However, within the scope of experiments executed as part of this research, it can be strongly concluded that Ganomaly-based anomaly detection for DAS-based fiber optics vibrations is the best and shows strong practical application candidature.

Contribution to Knowledge

Analysis of vibrations obtained from fiber optics cables has been an area of strong research for many years. There have been constant contributions from the community in this area and there have been regular advancements in the application of different techniques to analyze these vibrations. While signal processing methods have seen maximum work being done in this field, needless to say, application and experimentation based on machine learning techniques have also picked up in the last few years. However, the application of generative deep learning-based methods has been rare, if not none. This research aims to further push the efforts done in the area of application of generative deep learning -methods for fiber optics vibration analysis, by acting as a cornerstone in concluding that GAN-based anomaly detection methods can be applied to fiber optics vibrations. This research can also result in attracting more members of the community to experiment with different data and model architectures for anomaly detection as well as the application of generative deep learning as a whole for fiber optics vibration analysis.

Some of the key contributions of this research around data preparation, data transformation, and, application of GAN[1]based modeling techniques can be used in various different scenarios and use cases in the domain of fiber optics vibrations. This research hopes to trigger a ripple effect and attraction to generative deep learning in the community working on fiber optic vibrations.

Future Recommendation

While this research does bring forth some very important conclusions related to GAN[1]based anomaly detection for fiber optics vibrations, there are still many more variations that can be evaluated. In this research, certain elements have been kept constant, such as input tensor shape and basic building blocks (Generalized Encoder, Generalized Decoder, and, Generalized Classifier/Discriminator).

The reason to keep these elements constant across different architectures was to provide a fairground for comparison. However, this also limits the search space for the model architecture, activation functions, hyper-parameters, and, optimizer parameters. This research can be taken further with more variations executed for individual modeling techniques.

Additionally, this research is based on a dataset collected from a specific location that was exposed to some specific external vibrations (as anomalies) and some environmental conditions/vibrations. Environmental vibrations should be classified as non-anomalous vibrations and therefore it is important to gather more non[1]anomalous data points for building a robust model. For example, data collected from area which has a high amount of random traffic noise can also be a good addition to this dataset. There can be many kinds of non-anomalous high-intensity random noise vibrations, which could be added to build a comprehensive databank for building models.

References

[1] Reference for DIRT Report (2019)

https://commongroundalliance.com/Portals/0/Library/2020/DIRT%20Reports/2019%20DIRT%20Report%20FINAL.pdf?ver=2020-10-14-185343-180

Scholarly articles:

- Oh & Yun, 2018
- (Di Mattia et al., 2019),
- AnoGAN(Schlegl et al., 2017)
- EGBAD (Zenati et al., 2018)
- GANomaly (Akcay et al., 2019).
- Goodfellow et al., 2014, also known as GAN (Generative Adversarial Networks)
- (Papp et al., 2016b), (Wu et al., 2015),
- (Wu et al., 2014b).
- (Martins et al., 2015)
- (Tejedor et al., 2017)
- (Aktas et al., 2017)

\Orchestrating a brighter world

NEC

Acknowledgement

Sandeep Khurana
Chief Technology Officer
NEC Technologies India Pvt Ltd

Amit Kumar Sahu
Enterprise Architect
NEC Technologies India Pvt Ltd

Contact Us :

marketing@india.nec.com

in.nec.com