Machine Learning Techniques for Intrusion and Leak detection in Gas Pipelines based on Distributed Acoustic Sensing

Nandhakishore C S[1 2] Balaji Srinivasan [1] Balaraman Ravindran [2]

[1]Department of Electrical Engineering, Indian Institute of Technology Madras [2]Wadhwani School of Data Science and Artificial Intelligence, Indian Institute of Technology Madras

Introduction

Distributed fiber Acoustic Sensing (DAS) is an emerging sensing technology that can continuously detect external physical events (vibration and acoustics variations) over long distances with coherent Rayleigh back-scattering of low-noise laser. When light is being sent through the optical fiber placed near the area of interest (locations where physical events are caused by vibrations and acoustics), the phase of the back-scattered light changes. With the back-scattering profile and the data about the laser sent inbound, the events can be recorded in the time-distance domain, with the phase being denoted as intensity. Using this data, machine learning models have been developed. Distributed Acoustic Sensing, as an application is used to monitor perimeters, especially in remote locations. In this poster, various machine learning methods to identify leaks and intrusions in gas pipelines are explored.

Data

- The data for this experiment are the waterfall plots obtained from the DAS experimental setup which contains the intensity of the change in phase plotted in a space time graph. These graphs are then post-processed and saved in 100 dpi with 500 x 500 pixels scale as *.png* files.
- This data is generated from simulations of an optical fiber of length 1500m having a scattering resolution of 1. To mimic real time data, Gaussian and Laplacian noises are added, with events occurring every 50m with value of strains ranging from 0 to 0.0001.

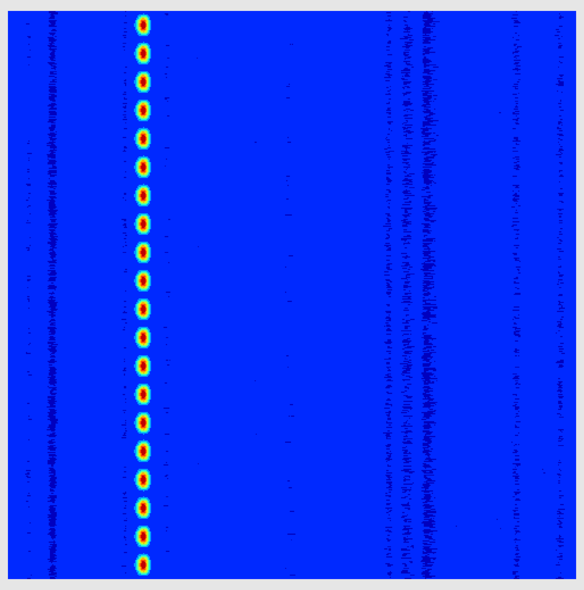

Figure 1. A Model Waterfall plot representing an intrusion at 100m location in a 1500m optical fiber

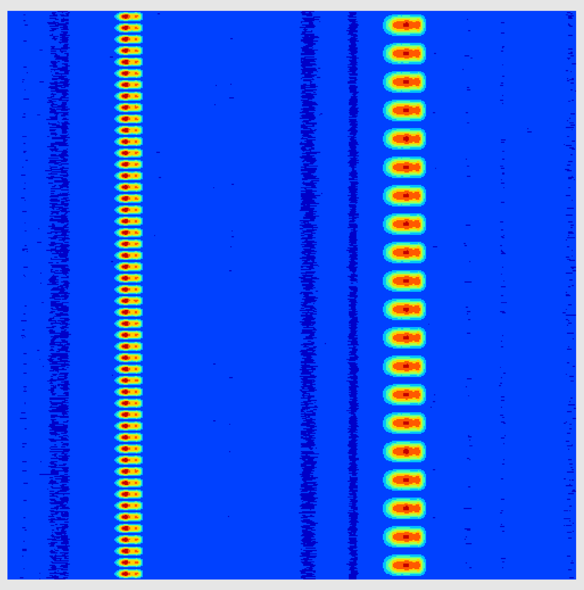

Figure 2. A Model Waterfall plot representing two intrusions (with varying frequencies) at 300m and 1000m location in a 1500m optical fiber

- The data is divided into two classes with labels: *threat* and *no_threat*
- *threat* class contains events which has strain greater than 0.01 and frequency less then 100Hz.
- *no_threat* class contains events which has strain less than 0.01 and frequency greater than 100Hz.
- For feature engineering, the labels of the images in classes *threat* and *no_threat* are one-hot encoded with binary labels (i.e.) *no_threat* is labeled as ′0′ and *threat* is labeled as ′1′.
- *threat* class contains 1442 images and *no_threat* class contains 1090 images.

Conclusion

- The parameter space for the model to train is very large as the graphs represent events with high resolution.
- As the ensemble model has a recal score of 1, the model has labelled the data belonging to the original class as original, but it says nothing about how many items from other classes were incorrectly also labelled as belonging to same class

References

- Shiloh, L., A. Eyal, and R. Giryes (Sept. 2019). "Efficient Processing of Distributed Acoustic Sensing Data Using a Deep Learning Approach". In: *J. Lightwave Technol.* 37.18, pp. 4755–4762. URL: https://opg.optica.org/jlt/abstract.cfm?URI=jlt-37-18-4755.
- Simonyan, K. and A. Zisserman (2015). *Very Deep Convolutional Networks for Large-Scale Image Recognition*. arXiv: 1409.1556 [cs.CV].
- Deng, J., W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei (2009). "ImageNet: A large-scale hierarchical image database". In: *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 248–255. DOI: 10.1109/CVPR.2009.5206848.

Methodology

In a real time environment, the optical fiber is placed parallel to the gas pipelines which run through the length of the pipeline. A laser sequentially connected to EDFA (Erbium-Doped Fiber Amplifier) and AOM (Acousto-Optic Modulator) is connected to the fiber. A detector is then connected to the same side of the laser to acquire data about the back-scattered light .

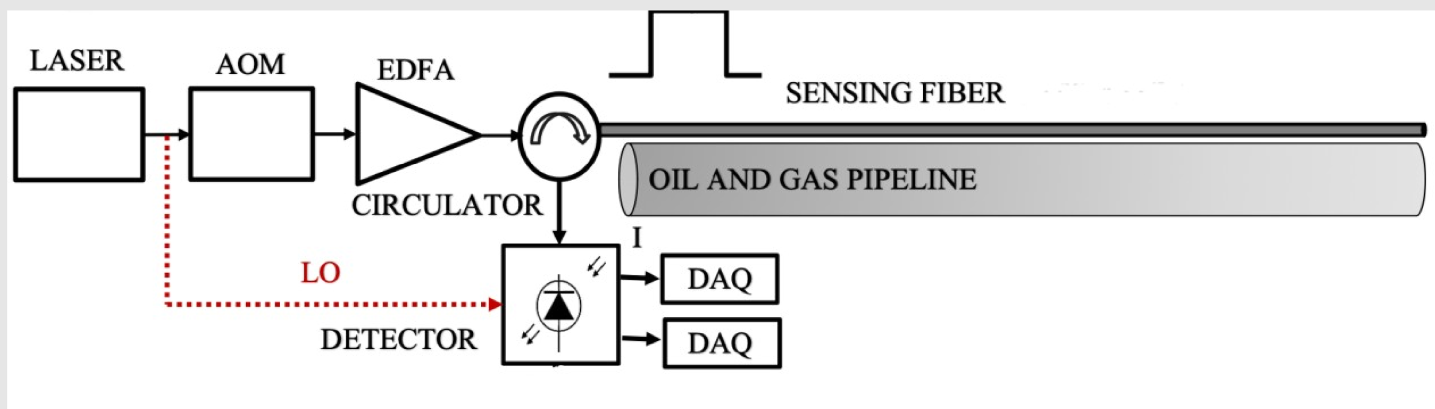

Figure 3. Real Life experimental setup to acquire data using DAS

This setup is simulated using Python scripts and 2532 plots with various events are generated. Image classification models built with CNN (Convolutions Neural Networks) in TensorFlow are used to do binary image classification with Adam optimiser and Binary Cross entropy loss function with a learning rate of 0.001.

- **VGG16_TL** - image classifier with 16 CNN layers with weights from ImageNet by Simonyan and Zisserman 2015, Deng et al. 2009.
- **VGG16_R** - modified VGG16 with weights initialised randomly.
- **FiberNet_TL** - a modified VGG16 classifier with an added CNN layer using weights from ImageNet by Shiloh et al. 2019.
- **FiberNet_R** - a modified FiberNet classifier with weights initialised randomly.
- **Ensemble Model** - an ensemble of custom image classifiers created using Models 1 to 4 with soft voting.

The hyper-parameters used for training the models are as follows:

- **Learning Rate**: 0.001
- **Number of Epochs**: 50
- **Training size**: 70% of dataset
- **Validation size**: 10% of dataset

Results

The ensemble model trained yields an accuracy of 67%. The main findings about the ensemble are as follows:

- The model has a recal score of unity and F1 score of 0.733.
- The models utilised 134304641 parameters occupying a memory of 512.33 MB
- The learning rate of the model is reduced to 0.0001 using ReduceLROnPlatue method once the model's accuracy stops improving.

Figure 4. Accuracy and Loss vs Epochs graphs for the ensemble model

RBCDSAI IITM