Pipeline Safety Early Warning Method for Distributed Signal using Bilinear CNN and Lightgbm

Yiyuan Yang[1], Yi Li[1], Haifeng Zhang[2*]

[1]International Graduate School at Shenzhen, Tsinghua University, Shenzhen, China,

[2]Research Institute of Tsinghua, Pearl River Delta, Guangzhou, China

yangyy19@mails.tsinghua.edu.cn, liyi@sz.tsinghua.edu.cn, zhanghf@tsinghua-gd.org

Tsinghua University

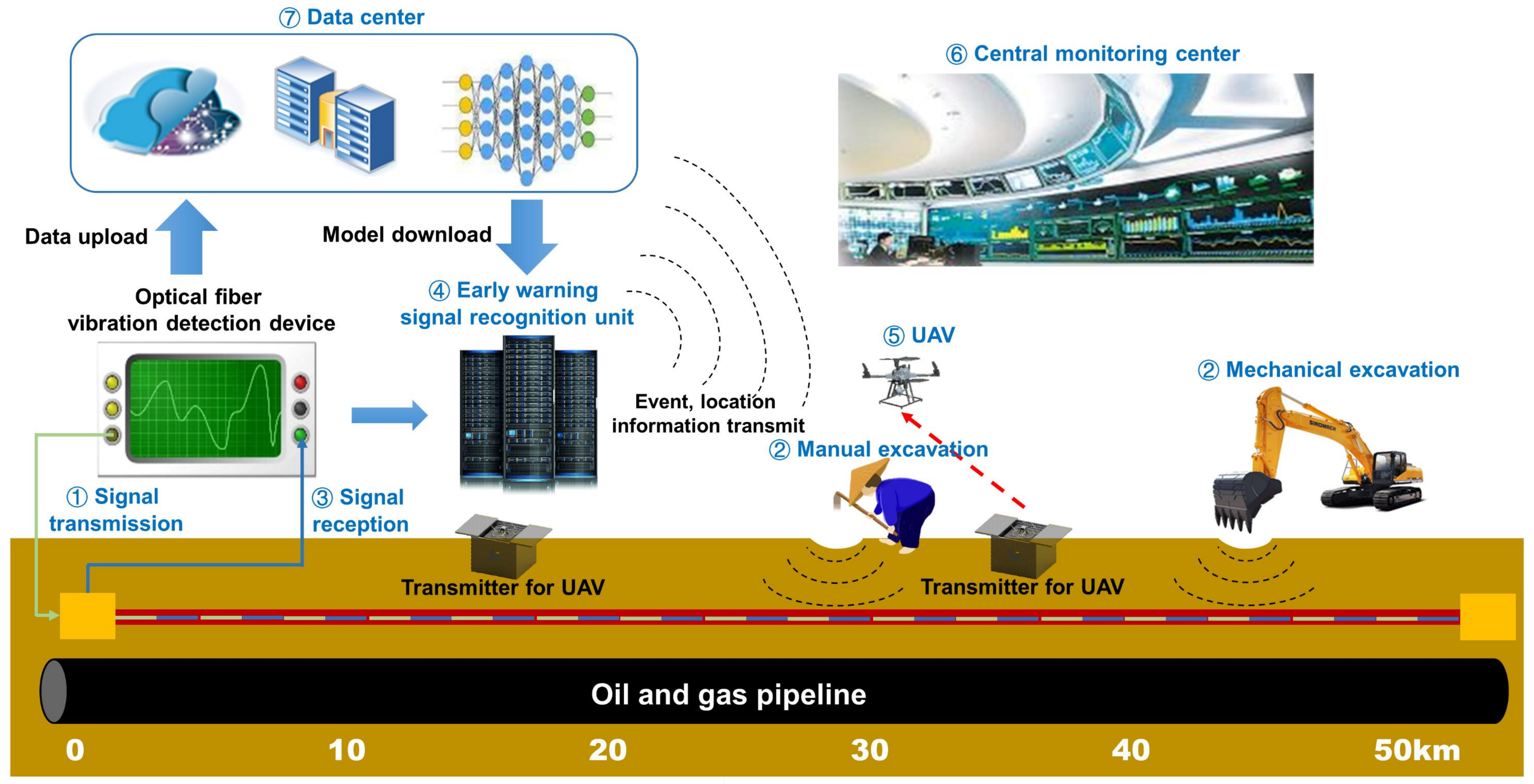
2021 IEEE International Conference on Acoustics, Speech and Signal Processing

ICASSP2021 TORONTO Canada June 6-11, 2021

Metro Toronto Convention Centre

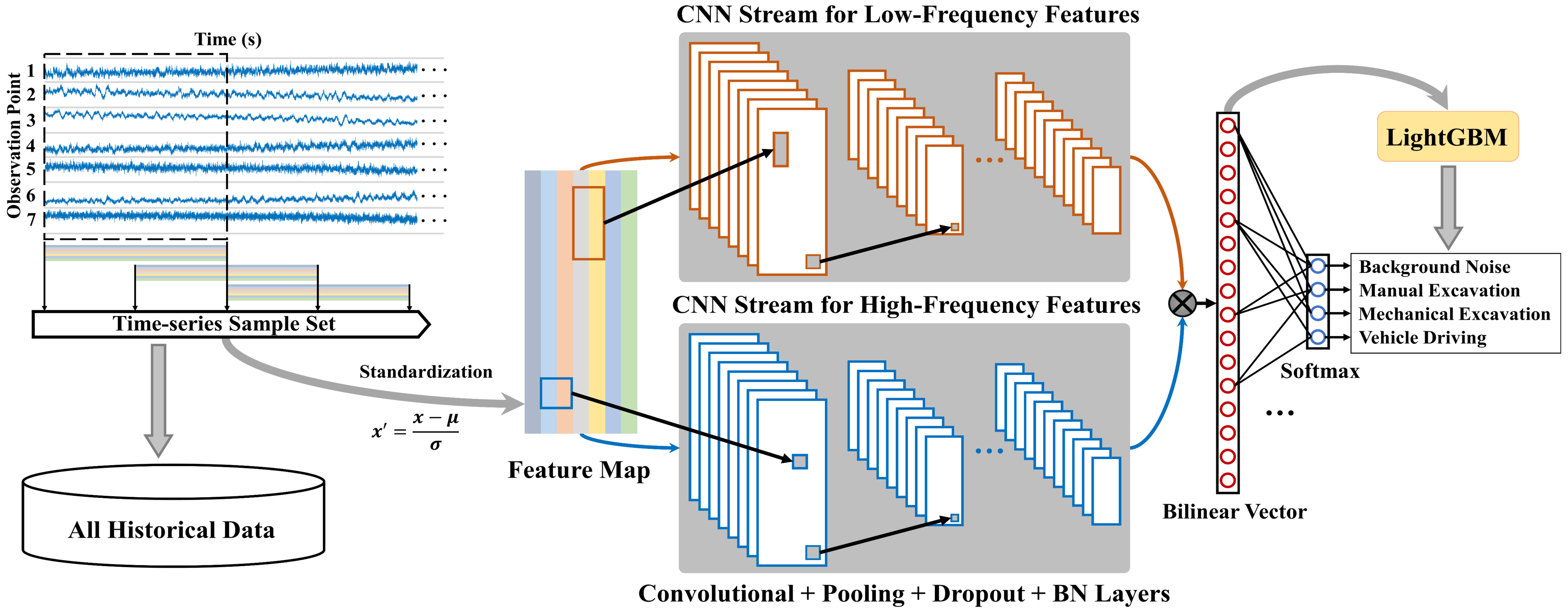
Problem

Oil and gas pipelines are known as the backbone of global energy. Ensuring their safety is related to the energy supply, environmental protection and the stability of the economy and society. Pipeline safety early warning (PSEW) systems aim to automatically identify and locate third-party damage events on energy pipelines and replace traditional, inefficient manual inspections. However, existing PSEW systems cannot achieve universality for various complex environments because they are sensitive to the spatial and temporal stability of the signal obtained by distributed optical fiber sensors at various locations and times. Our research aims to improve the identify and location through ML algorithm based on our novel PSEW system, as shown in the figure below.



Contributions

- An approach that reanalyzes industrial distributed signals in both spatial and temporal domains and obtains excellent location and identification performance.
- We collected a large amount of signal data from long-distance pipelines that are already in service and built a database for model construction and evaluation.
- The Proposed ML model is more adaptable to complex environments and more scalable to hardware than other baselines under the premise of good real-time performance.



Methodology

- Fix the spatial domain, and slide a window in the time domain to generate samples.
- Standardize the above samples separately.
- Input samples into the B-CNN to pre-train and obtain the parameters of the convolutional and fully connected layers.
- Freeze the parameters of the convolutional layer in B-CNN, and the results from the flattened layer are input to the LightGBM for retraining.
- The prediction value represents the result of its central observation point in 4 s of 500-Hz signals or 20 s of 100-Hz.

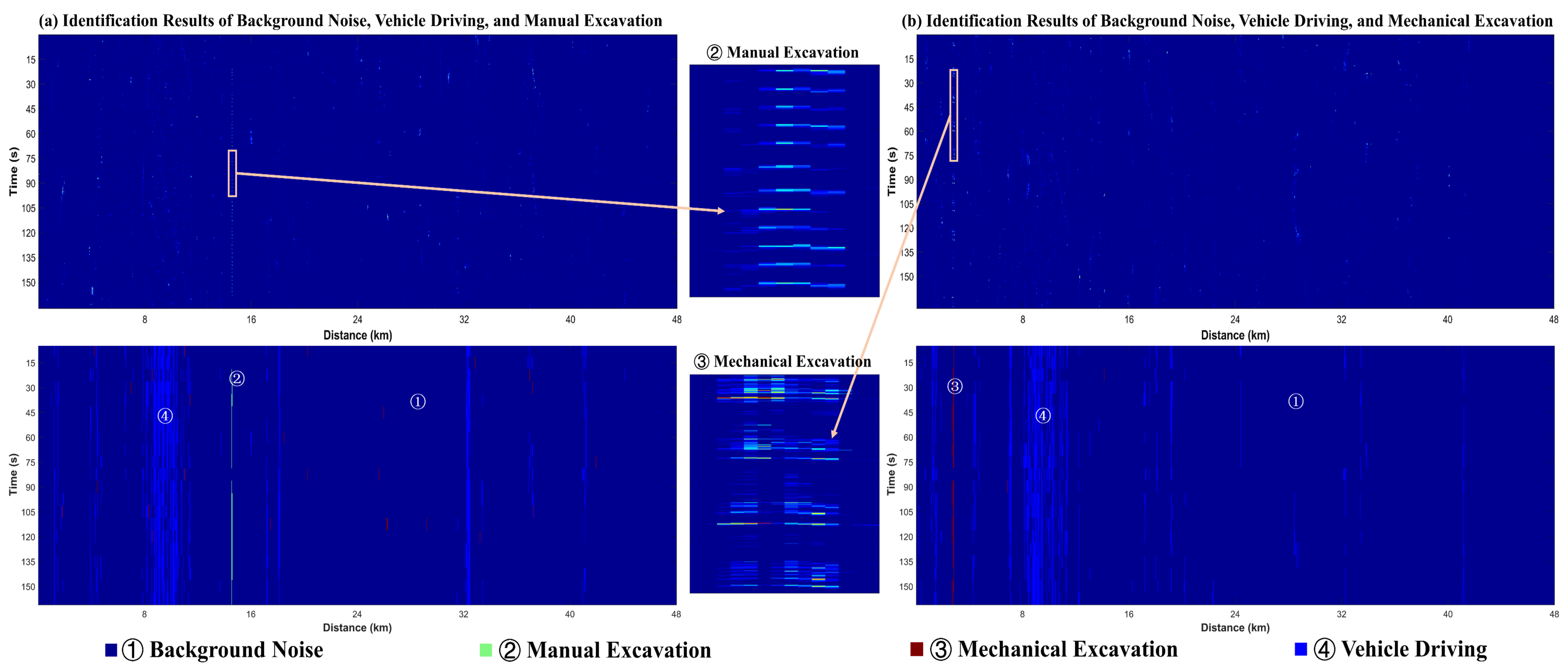
Datasets

The data were gathered at the gas pipe of the West-East Natural Gas Transmission Project Suzhou Section from May 10 to June 2 and from November 19 to December 17. The total data size is approximately 494 GB. In addition, it is approximately 48 km with complex environment under several types of strong noise and weak signal, and signal drift that are unique to long-distance pipelines at real sites. And the spatial resolution is 20 m.

	Model time	Total time	Model size
100 Hz	5.569 s	7.987 s	25.68 MB
500 Hz	15.36 s	19.37 s	

	DNN	1DCNN	2DCNN	B-CNN	B-CNN_LGBM
Background noise					
Precision (%)	87.20/85.33	95.26/96.26	**100.0/100.0**	99.88/**100.0**	**100.0/100.0**
Recall (%)	87.20/86.29	96.58/98.83	**100.0/100.0**	**100.0/100.0**	**100.0/100.0**
F1-score (%)	87.20/85.81	95.92/97.53	**100.0/100.0**	99.94/**100.0**	**100.0/100.0**
AUC	0.873/0.867	0.978/0.989	**1.00/1.00**	0.999/**1.00**	**1.00/1.00**
Manual excavation					
Precision (%)	91.32/89.26	97.75/94.38	**100.0/100.0**	98.98/99.25	**100.0/100.0**
Recall (%)	92.24/90.68	94.83/95.33	91.38/98.06	93.26/98.50	**96.98/99.75**
F1-score (%)	91.78/89.96	96.27/94.85	95.50/99.02	96.03/98.87	**98.47/99.87**
AUC	0.928/0.908	0.972/0.956	0.957/0.990	0.975/0.991	**0.989/0.999**
Mechanical excavation					
Precision (%)	72.66/69.78	81.37/73.59	83.51/75.19	93.57/82.86	**97.25/85.03**
Recall (%)	74.36/75.67	84.32/78.67	95.29/**100.0**	92.88/98.68	**98.67/100.0**
F1-score (%)	73.50/72.61	82.82/76.05	89.01/85.84	93.22/90.08	**97.95/91.91**
AUC	0.753/0.763	0.855/0.808	0.959/0.973	0.967/0.982	**0.988/0.985**
Vehicle driving					
Precision (%)	93.76/82.37	95.23/91.79	97.70/**100.0**	98.33/**100.0**	**98.67/100.0**
Recall (%)	95.28/83.56	97.88/**90.67**	**99.42**/80.05	98.33/85.37	99.12/88.86
F1-score (%)	94.51/82.96	96.54/91.23	98.55/88.92	98.33/92.11	**98.89/94.10**
AUC	0.955/0.848	0.980/0.925	0.992/0.900	0.992/0.950	**0.994/0.968**
Total					
Accuracy (%)	87.02/82.86	92.91/89.87	95.72/93.49	96.89/95.27	**98.83/96.47**

Experiment and Deployment



Conclusion

The described algorithm could identify and locate third-party damage events under the conditions of strong noise, weak signals, and signal drift with accuracies of 96.47 % for 100 Hz data and 98.83 % for 500 Hz data in testing sets. Moreover, our model fully meet industry standards of model size, real time performance, and easy deployment. However, the robust to abnormal and error data need to be improved further.