Two-Dimensional CNN Based on Channel and Spatial Attention Mechanism for Φ-OTDR Vibration Recognition in Optical Transport Networks

Xia Gao, Chengliang Zhang, Qian Hu, Yi Ding, Xin Qin, Xiaoli Huo

China Telecom Research Institute, State Key Laboratory of Optical Fiber and Cable Manufacture Technology, Beijing, China.

*gaox12@chinatelecom.cn

Abstract *A 2D-CNN structure based on channel and spatial attention mechanisms (CS-CNN) is proposed for Φ-OTDR vibration recognition in optical transport networks. Field experiments show that our scheme can achieve 2x faster convergence and 11% higher accuracy than traditional 2D-CNN, suggesting promise for improving online monitoring. ©2023 The Author(s)*

1 Introduction

In recent years, distributed optical fibre sensing (DOFS) based on phase-sensitive optical time domain reflectometry (Φ-OTDR) has gained significant attention in coherent optical communication networks for simultaneous data transmission and distributed vibration detection [1,2], owing to its high sensitivity, spatial resolution, and real-time performance over long distances. However, the inherent strong noise, weak signal, and signal drift of Φ-OTDR signal [3] make it an urgent and challenging task to identify target events in real time and with high accuracy.

Artificial intelligence provides an opportunity to enhance smart sensing capabilities. However, traditional machine learning methods such as Artificial Neural Networks (ANN) [4], Support Vector Machines (SVM) [5], Random Forests (RF) [6], and Extreme Gradient Boosting (XG-Boost) [7] suffer from the limitations of expert-dependent feature extraction and slow updating speeds that cannot keep up with the changing patterns of massive sampling points, thereby restricting their application. Deep learning has the advantage of automatically extracting distinguishable features hidden in signals and achieving high recognition accuracy, making it the preferred method for Φ-OTDR event recognition, such as convolutional neural network (CNN) [8], temporal convolutional network (TCN) [9], Long Short-Term Memory (LSTM) [10], and Generative Adversarial Network (GAN) [11]. However, the convolution kernel convolves both channel and spatial information of the signal, leading to a reduction in the significance of its distinctive features and constraining recognition accuracy. Recently, a transfer learning-based architecture was utilized in Φ-OTDR system, achieving event recognition network with high accuracy of 98% [12,13]. However, the network structure is complex and requires high hardware. In addition, Squeeze-and-Excitation Networks were employed to extract important features through global average pooling for enhancing classification accuracy [9, 14]. But global average optimization can only obtain the sub-optimal features on the channel [15].

In this paper, In order to fully utilize the spatiotemporal distribution information of Φ-OTDR and further improve the accuracy of event recognition in optical transmission networks, the channel and spatial attention model is introduced into 2D-CNN, referred to as CS-CNN. In this algorithm, the convolutional layer of 2D-CNN is used to automatically extract the structural features of Φ-OTDR signals, and the channel and spatial convolution attention models extract important features and suppress unnecessary features through average pooling and max pooling. Field experiments collected four types of Φ-OTDR phase information on communication optical cables. The results showed that compared with traditional 2D-CNN, CS-CNN has 2x faster convergence and 11% higher accuracy rate, providing a potential method for online monitoring of Φ-OTDR events in optical transport networks.

2 Φ-OTDR Event Recognition Method Based on CS-CNN

A. Φ-OTDR system and data set preparation

To gather vibration events in communication optical cables, a field experiment was conducted at the China Telecom Research Institute, as depicted in Fig. 1(a). The Φ-OTDR located in building 1 is connected to building 2 through a 1 km buried communication optical cable, passing through 7 tube wells, as shown by the red dots in the Fig. 1(a). The Φ-OTDR system based on the I/Q demodulator [16] is shown in Fig. 1(b). The 200 ns optical pulse is generated by the AOM and amplified by the EDFA before entering the sensing fibre. Then, the Rayleigh backscattering signal of the pulse passes through the circulator and couples with the intrinsic light in the hybrid module to generate the IQ signal. After the signal is detected by the PD and collected by the ADC, the phase information of the vibration events is

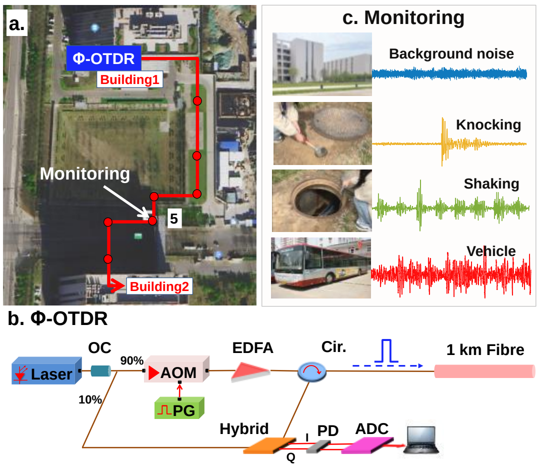

Fig. 1: (a) A field experiment using Φ-OTDR; (b) The Φ-OTDR system based on the I/Q demodulator. OC: optical coupler, AOM: acoustic optic modulator, EDFA: erbium-doped fiber amplifier, Cir.: circulator, PG: pulse generator, PD: photodetector, ADC: analog-to-digital converter; (c) Monitoring of four types of events.

recovered through digital signal processing. Four types of vibration events are monitored at the fifth pipe well near the road, including knocking on the manhole cover, shaking optical cables, driving vehicles and background noise, as shown in Fig. 1(c).

Tab. 1: DATABASE CONSTRUCTION WITH REAL FIELD DATA.

Events type	Training/Testing data set size	Label
Background	265/108	1
Knocking	328/133	2
Shaking	352/141	3
vehicles	372/149	4

Each type of sample matrix is 19199*10, where the rows of the matrix represent time domain information of 1.92 ms, with one point sampled every 0.1 ms, and the columns of the matrix represent spatial domain information of 200 m, with one point sampled every 20 m. The four types of samples are divided into training set and test set in a ratio of 5:2, as shown in Tab. 1. It should be noted that in order to reduce the influence of singular samples on the convergence speed, the sample points are normalized to 0 ~ 255, and then enter the neural network for training.

B. The network structure of CS-CNN

The network structure of CS-2DCNN proposed in this paper is shown in Fig. 2. In the scheme, 2D-CNN is used as the basic architecture to build a classification model, including convolutional layers, pooling layers, and fully connected layers. In order to further improve the recognition ability of features and the convergence speed of the network, the channel and spatial attention mechanism are introduced into the CNN network [15].

The channel attention mechanism enables screening and enhancement of unique features by attaching the spatial attention matrix to the original feature map at a low computational cost. The structure is shown in the green dashed box, which can be described as: 1) Global max-pooling and global average-pooling are used to aggregate the spatial information of the input feature map $F \in \mathbb{R}^{C\times H\times W}$, and two different spatial descriptors F_C^M and F_C^A are generated. 2) The two descriptors are forwarded to a shared multi-layer perceptron (MLP). The weights of the MLP are $W_1 \in \mathbb{R}^{C/r\times C}$ and $W_2 \in \mathbb{R}^{C\times C/r}$, where r represents the reduction ratio. 3) The output vectors of the two perceptron channels are summed, and then multiplied by the input feature map F. The above channel attention process can be described mathematically as:

$$F^C = \sigma\big(W_2\gamma(W_1F_C^M) + W_2\gamma(W_1F_C^A)\big) \otimes F \qquad (1)$$

Where $\otimes$ denotes element-wise multiplication, σ is sigmoid function, γ is RELU activation function.

The spatial attention and the channel attention mechanism are complementary and determine the concentrated feature locations. The structure is shown in the purple dashed box, which can be described as: 1) Using average-pooling and max-pooling along the channel axis to highlight information regions and generate two 2D vectors F_S^M and F_S^A. 2) F_S^M and F_S^A are concatenated to generate a valid feature descriptor. 3) A standard convolutional layer follows and then multiplies the features of input feature map F^C. The above spatial attention

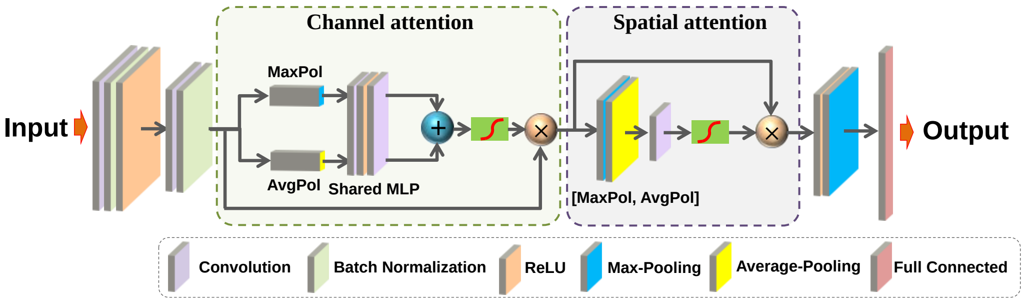

Fig. 2: The network structure of CS-CNN.

process can be described as:

$$F^S = \sigma(f^{7\times7}[F_S^M; F_S^A]) \otimes F^C \quad (2)$$

where $f^{7\times7}$ denotes a convolution operation with a filter size of 7×7.

It should be noted that batch normalization is performed after the CNN convolution kernel to keep the data away from the saturation area, avoid distributed data deviation, and improve the accuracy of the model [17].

C. Results and discussion

To evaluate the performance of our algorithm, we compare it with traditional 2D-CNN. First, the initial learning rate is set to 0.0001 and the batch size is set to 50 with 80 epochs. In this case, the accuracy curves of 2D-CNN and CS-CNN are compared, as shown in Fig. 3. It can be clearly observed that after 10 epochs of training, the accuracy rate of CS-CNN on the test set reaches about 93%, while the accuracy rate of 2D-CNN is only about 80%. Moreover, CS-CNN has a fast learning ability and basically reaches the plateau of accuracy growth at 30 epochs, while 2D-CNN tends to be stable at 60 epochs. Therefore, the CS-CNN-based training model is more suitable for high-precision online monitoring of Φ-OTDR events in optical transport networks.

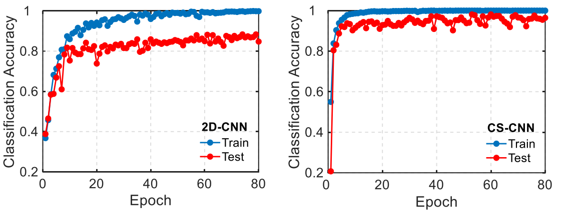

Fig. 3: The accuracy curves of (a) 2D-CNN and (b) CS-CNN.

Then, in order to quantify the performance of our algorithm, Fig. 4 further shows the confusion matrix of the two methods. According to the TP (True Positive), TN (True Negative), FP (False Positive) and FN (False Negative) parameters obtained from the confusion matrix statistics in Fig. 4, the commonly used accuracy and precision, false alarm rate (NAR) and F-score are calculated [18], as shown in Tab. 2. It can be found that the CS-CNN always performs best in terms of precision, recall and f-score compared to the 2D-CNN. Its average accuracy rate is 96%, which is better than 85% of 2D-CNN.

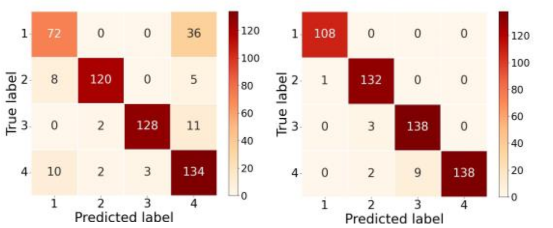

Fig. 4: The confusion matrix of (a) 2D-CNN and (b) CS-CNN.

Tab. 2: A COMPARISON OF RECOGNITION RESULTS FOR THE TWO METHODS.

Classifier	Event Types	Precision	NAR	F-score
2D-CNN	1	1.000	0.056	0.971
average	2	0.950	0.008	0.971
accuracy=0.96	3	0.952	0.021	0.965
	4	0.966	0.060	0.952
CS-CNN	1	0.859	0.491	0.640
average	2	0.914	0.038	0.938
accuracy=0.85	3	0.903	0.007	0.946
	4	0.738	0.148	0.791

Finally, in order to observe the discriminability of event features more intuitively, all feature vectors generated by the optimal models of 2D-CNN and CS-CNN are mapped into the three-dimensional feature space through linear discriminant analysis [19], as shown in Fig. 5. Obviously, there are 4 cluster centers with clear boundaries in Fig. 5 (b), which proves the effectiveness of CS-CNN to extract features.

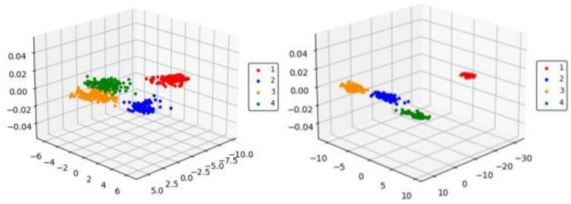

Fig. 5: The feature vectors of (a) 2D-CNN and (b) CS-CNN.

3 Conclusion

In this paper, a 2D-CNN structure based on channel and spatial attention model is proposed for the classification of Φ-OTDR vibration events. Field experiment results show that an average of 96% recognition accuracy can be achieved based on our proposal, which is 11% higher than the traditional 2D-CNN. In addition, due to the presence of attention models, the convergence speed of CS-CNN is faster, twice that of traditional 2D-CNN, which means that CS-CNN provides a potential method for online monitoring of Φ-OTDR events in optical transport networks. For example, CS-CNN-based Φ-OTDR technology is used to monitor events around communication cables to reduce the risk of man-made damage to optical cables.

4 References

[1] H. He, L. Jiang, Y. Pan, A. Yi, X. Zou, W. Pan, A. E. Willner, X. Yan, Z. He, L. Yan, “Integrated sensing and communication in an optical fibre”, *Light: Science & Applications*, vol. 12, no. 25, 2023.

DOI: 10.1038/s41377-022-01067-1.

[2] S. Guerrier, K. Benyahya, C. Dorize, E. Awwad, H. Mardoyan and J. Renaudier, “Vibration detection and

localization in buried fiber cable after 80km of SSMF using digital coherent sensing system with co-propagating 600Gb/s WDM channels", *In 2022 Optical Fiber Communications Conference and Exhibition (OFC)*, San Diego, CA, USA, 2022.

DOI: 10.1364/OFC.2022.M2F.3

[3] Y. Yang, Y. Li, T. Zhang, Y. Zhou and H. Zhang, "Early safety warnings for long-distance pipelines: A distributed optical fiber sensor machine learning approach", *In Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 35, No. 17, pp. 14991-14999, 2021.

DOI: 10.1609/aaai.v35i17.17759

[4] 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," *Applied Sciences*, Vol. *7, No.* 8, pp. 841, 2017.

DOI: 10.3390/app7080841

[5] Y. Tang, "Deep learning using linear support vector machines," *International Conference on Machine Learning 2013: Challenges in Representation Learning Workshop*, Atlanta, Georgia, USA.

DOI: 10.48550/arXiv.1306.0239

[6] M. Zhang, Y. Li, J. Chen, Y. Song, J. Zhang, and M. Wang, "Event detection method comparison for distributed acoustic sensors using ϕ-OTDR," *Optical Fiber Technol.* Vol. 52, pp. 101980, 2019.

DOI: 10.1016/J.YOFTE.2019.101980

[7] A. V. Timofeev and D. I. Groznov, "Classification of seismoacoustic emission sources in fiber optic systems for monitoring extended objects," *Optoelectron. Instrum. Data Process*. Vol. 56, pp. 50-60, 2020.

DOI: 10.3103/s8756699020010070

[8] Y. C. Manie, J. W. Li, P. C. Peng, R. K. Shiu, Y. Y. Chen, and Y. T. Hsu, "Using a machine learning algorithm integrated with data de-noising techniques to optimize the multipoint sensor network," *Sensors (Switzerland),* Vol. 20, pp. 1070, 2020.

DOI: 10.3390/s20041070

[9] M. Tian, H. Dong, X. Cao, and K. Yu, "Temporal convolution network with a dual attention mechanism for φ-OTDR event classification," *Applied optics,* Vol. 61, No. 20, pp. 5951-5956, 2022.

DOI: 10.1364/ao.458736

[10] Ö. Yildirim, "A novel wavelet sequences based on deep bidirectional LSTM network model for ECG signal classification," *Computer Science,* Vol. 96, pp. 189–202, 2018.

DOI: 10.1016/j.compbiomed.2018.03.016

[11] J. Yoon, D. Jarrett, and M. van der Schaar, "Time-series generative adversarial networks," *in Advances in Neural Information Processing Systems*, Vol. 32, pp. 1–11, 2019.

[12] Y. Shi, Y. Li, Y. Zhang, Z. Zhuang, and T. Jiang, "An Easy Access Method for Event Recognition of Φ-OTDR Sensing System Based on Transfer Learning," *Journal of Lightwave Technology,* Vol. 39, pp. 4548-4555, 2021.

DOI: 10.1109/JLT.2021.3070583

[13] H. Zhang, J. Gao and B. Hong, "Φ-OTDR Signal Identification Method Based on Multimodal Fusion", *Sensors*, Vol. 22, No, 22, pp. 8795, 2022.

DOI: 10.3390/s22228795

[14] Y. Shi, J. Chen, S. Dai, X. Liu, and C. Wei, "A Multi-Signal Feature Fusion Method with Attention Mechanism for Φ-OTDR Event Recognition System," *Optics express*, Vol. 30, No. 23, pp. 42086-42096, 2022.

DOI: 10.1364/oe.472794

[15] S. Woo, J. Park, J. Lee, and I. Kweon, "CBAM: Convolutional Block Attention Module," *European Conference on Computer Vision*, 2018.

DOI: 10.1007/978-3-030-01234-2_1

[16] Z. Wang, L. Zhang, S. Wang, N. Xue, F. Peng, M. Fan, W. Sun, X. Qian, J. Rao, and Y. Rao, "Coherent Φ-OTDR based on I/Q demodulation and homodyne detection," *Optics Express,* Vol. 24, No. 2, pp. 853–858, 2016.

DOI: 10.1364/OE.24.000853

[17] S. Ioffe, and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," *in International Conference on Machine Learning*, 2015.

[18] X. Cao, Y. Su, Z. Jin, and K. Yu, "An open dataset of φ-OTDR events with two classification models as baselines," *Results in Optics,* 2023.

DOI: 10.1016/j.rio.2023.100372

[19] F. Anowar, S. Sadaoui, and B. Selim, "Conceptual and empirical comparison of dimensionality reduction algorithms (PCA, KPCA, LDA, MDS, SVD, LLE, ISOMAP, LE, ICA, t-SNE)," *Computer Science Review,* Vol. 40, pp. 100378, 2021.

DOI: 10.1016/J.COSREV.2021.100378