2021 19th International Conference on Optical Communications and Networks (ICOCN) | 978-1-6654-2446-2/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICOCN53177.2021.9563673

# Attention based Temporal convolutional network for Φ-OTDR event classification

*Manling Tian[1,2], Hui Dong[3], Kuanglu Yu[1,2]*

1. Beijing Key Laboratory of Modern Information Science and Network Technology
2. Institute of Information Science, Beijing Jiaotong University, Beijing, 100044, China
3. Signal Processing. RF & Optical Department, Institute for Inforcomm Research, A*Star Research Entities, 138632, Singapore

klyu@bjtu.edu.cn

***Abstract*—We designed a new attention based temporal convolutional network combined with bidirectional long short term memory model named ATCN-BiLSTM for Φ-OTDR event classification, achieving average classification accuracy of 99.6% on three types of events.**

***Keywords: phase-sensitive optical time domain reflectometry; temporal convolution network; channel attention; bidirectional long short term memoryk; event classification***

## I. INTRODUCTION

Phase-sensitive optical time domain reflectometry (Φ-OTDR) has been widely applied in perimeter security, oil and gas pipeline inspection due to advantages such as distributed measurement, wide monitoring range and high sensitivity [1-2]. In order to improve the performance of Φ-OTDR, much work on intelligent event classification method have been carried out.

In recent years, deep learning methods have been widely used in disturbance event classification of Φ-OTDR. 2D images obtained by processing original signal through Fourier transform or spatial-temporal matric were input into two dimensional convolutional neural network (2D CNN) to classify different disturbances [3]. 1D CNN was also used for signal identification [4]. Long short term memory network (LSTM) can also be applied in Φ-OTDR. Chen proposed attention based LSTM (ALSTM) for disturbance classification [5], in which Mel-frequency cepstral coefficient, short time energy and short time zero crossing rate extracted from each frame of disturbance signals were input. To achieve more efficient and accurate identification, Guan proposed multi-scale 1 D CNN (MS 1-D CNN) in disturbance classification of Φ-OTDR by extracting different scale features of signal [6]. In Ref. [7], Wu proposed a new identification method by combining 1D CNNs and bidirectional LSTM, namely 1D CNNs-BiLSTM, extracting temporal and spatial features separately, which is proved to perform better in classification than model based on single domain feature or simultaneous space-time feature.

Some methods based on hand-crafted feature extraction is time-consuming and highly targeted, relying on expert knowledge. However, most end-to-end neural networks that automatically extract temporal or space-time features for classification ignore the causality of time domain signal at each sampling node on fiber. Besides, the influence of different channel features extracted by CNN on classification precision has never been concerned, to the best of our knowledge. Thus, we propose channel attention based temporal convolutional network (ATCN) for temporal feature extraction combined with BiLSTM for spatial feature extraction for Φ-OTDR, namely ATCN-BiLSTM.

## II. METHODOLOGY

Most recognition methods based on temporal feature usually use 1D CNN of which regular convolution ignores causality of time domain signal or LSTM which cannot compute in parallel. The problems can be solved by TCN [8], besides, we improve feature extraction ability of TCN by combining it with channel attention. Bidirectional spatial correlations less concerned in previous studies is taken into account too in our method. The whole event classification process based on ATCN-BiLSTM shown in Fig. 1(a) can be divided into three parts: each time domain signal at one sampling node is normalized and input into ATCN to extract temporal feature; then those temporal features are connected together as spatial sequence and input into BiLSTM to extract spatial relationship in event signal; the final space-time features obtained by BiLSTM are input into softmax classifier to identify the event type.

### A. Temporal feature extraction

Obviously, the signal is time sequential, that is, the signal at the current moment is only affected by the signal at the previous moment. We propose TCN[8] which is a strictly time-constrained model to extract the features of the time domain signal at each sampling node without information "leakage" from future to past.

Different from regular convolution, causal convolution of TCN can calculate in parallel where an output at time $t$ is obtained by convolving only with elements at time $t$ and earlier. The major disadvantage is that we need large kernel or an extremely deep network when processing long time sequence data, which may cause problems such as gradient vanishing or poor network fitting.

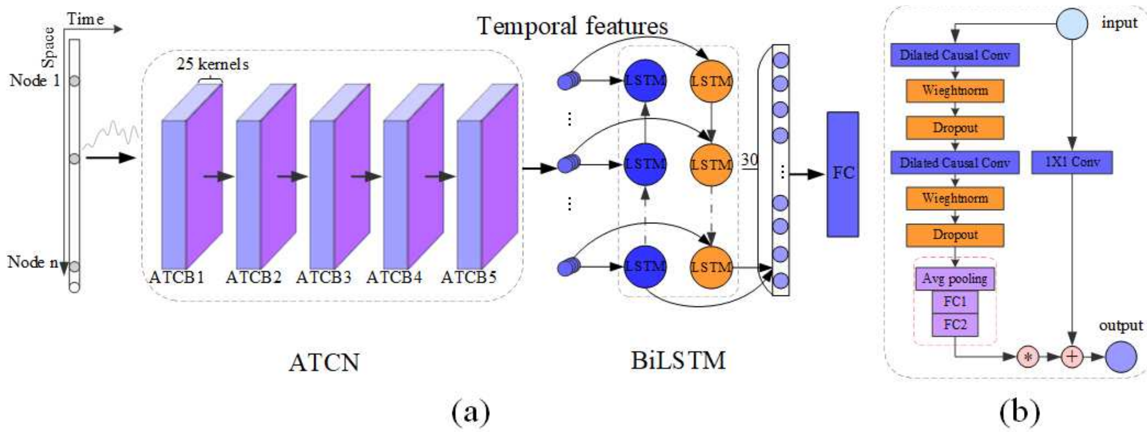


Figure 1. (a) The architecture of ATCN-BiLSTM, (b) The detailed architecture of ATCB (attention mechanism is marked with red dashed line)

Thus, dilated convolution which introduces dilated factor $d$ on the basis of regular convolution is employed to expand receptive field. The principle is to create a hole in the convolution map, setting a fixed step $d$ between two adjacent filter taps. Dilated convolution is equivalent to regular convolution when $d=1$. For 1-D sequence data X, the dilated convolution operator at a certain point $s$ of the sequence is defined as.

$$F(s)= \sum_{i=0}^{k-1} f(i)\cdot X_{s-d\cdot i} \quad (1)$$

where $f$ represents convolutional kernel and $k$ means kernel size. Causal convolution combined with dilated convolution which we call dilated causal convolution is shown in Fig. 2. In order to improve stability of network, basic temporal convolution block of TCN introduces residual connections, weight normalization and dropout which benefit deep neural network. Given kernel size $k$, dilated base $b$ ($d=b^{**}i$, $i \in (1,n)$), input length $l$, we need to stack $n$ temporal convolution blocks to extract temporal features of input $X$ at least. $n$ is calculated as.

$$n=\lceil \log_b(\frac{(l-1)\cdot(b-1)}{(k-1)}+1)\rceil \quad (2)$$

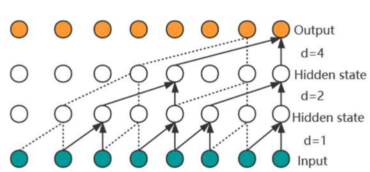


Figure 2. Dilated causal convolution with filter size $k$=2.

## B. Channel attention

The disturbance information contained in the signal at different time is different. Besides, different channel of multi-channel features obtained by temporal convolution block has different effect on final classification results. Thus, we introduce channel attention mechanism which perform dynamic channel-wise feature recalibration. Attention weight vector $s$ is calculated by (3) where $W_1 \in \mathbb{R}^{\frac{C}{r}\times C}$ and $W_2 \in \mathbb{R}^{C\times\frac{C}{r}}$ are the projection parameters to be learned during training. $r$ is the dimensionality reduction ratio. $z$ is channel descriptor generated by using global average pooling on the output of temporal convolution block. Temporal convolution block with channel attention (ATCB) is shown in Fig. 1(b). ATCN is composed of $n$ ATCBs.

$$s= \sigma(W_2\delta(W_1 z)) \quad (3)$$

## C. Spatial relationship extraction

Since disturbance propagates in both directions in space domain, BiLSTM is used to extract bidirectional spatial correlation. The final hidden states of two directions are connected and input into softmax classifier for identification.

# III. EXPERIMENTS AND RESULTS

## A. Data preparation and model construction

A direct Φ-OTDR system was used to collect event signal using the tail 50 meters fiber with a total length of 7.1km at a sampling rate of 10 kHz. Three types of events were recorded, including background without threats, climbing the fence and raining.

Max and min normalization is firstly applied for each time domain signal. Considering a small value range will cause the loss of data precision, the normalization range is set to 0~255, which is equivalent to image grayscale value range. It is worth noting that the dimension of raining event signal is different from other events. Therefore, downsampling is used to keep the input dimension of model consistent. Optical signals at 32 sampling nodes around the disturbance position are collected as each event sample. The total time length and total points of a single sample are 1 second and 4000.

The key parameters of network are demonstrated in Table I. The basic batch size is set as 2 and all of the batches of the training data will be passed though the network at each epoch. The initial learning rate is 0.00015.

TABLE I. PARAMETERS OF ATCN-BILSTM

| Model | Key parameters | |
|---|---|---|
| ATCN | *k* (kernel size) | 1*7*25 |
| | *b* (dilation base) | 4 |
| | *n* (number of temporal blocks) | 5 |
| | dropout | 0.04 |
| BiLSTM | hidden size | 15 |

## B. *Event classification*

The three events, background without threats, climbing the fence and raining are labeled as 0, 1, 2. The number of each type events for training/testing is set as 337/86, 294/72, and 337/87 respectively. We compare our model with ALSTM [5], MS 1-D CNN [6] and 1D CNNs-BiLSTM [7]. However, due to the parameters and network structures are not available in details, they have been adjusted accordingly to make their experimental results on our data as good as possible. Classification accuracy, nuisance alarm rate (NAR), false negative rate (FNR) and the total training time including data preprocessing time of the four methods are respectively computed and compared in Table 2. The confusion matrices of the four methods are illustrated as in Fig. 3.

TABLE II. A COMPARISON OF RECOGNITION PERFORMANCE FOR THE FOUR METHODS.

| Methods | Accuracy (%) | NAR (%) | FNR (%) | Total training time (seconds) |
|---|---|---|---|---|
| **ALSTM** | 96.3 | 1.2 | 4.4 | 6768 |
| **MS 1-D CNN** | 90.6 | 4.7 | 10.1 | 7431 |
| **1D CNNs-BiLSTM** | 95.5 | 0 | 1.9 | 4525 |
| **ATCN-BiLSTM** | **99.6** | **0** | **0.6** | 9114 |

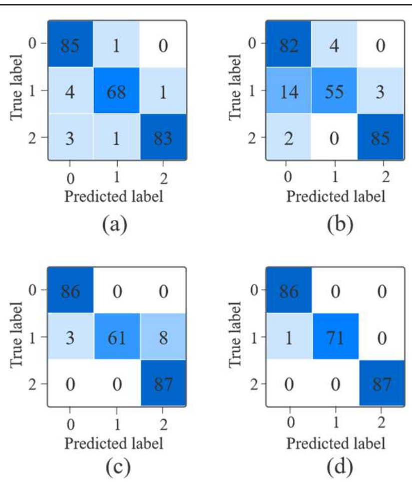


Figure 3. Confusion matrices obtained by the four methods: (a) ALSTM, (b) MS 1-D CNN, (c) 1D CNN-BiLSTM, (d) ATCN- BiLSTM

Firstly, it preliminarily shows that extracting time domain feature and spatial relationship separatly is better than single domain feature extraction since average accuracy of MS 1-D CNN which extract temporal feature only is 90.6%, lower than other three networks, and its NAR is 4.7%, the highest of the four. Secondly, temporal feature will be more disinguishable after considering causality of time sequence. Compared with 1D CNN-BiLSTM, average accuracy of ATCN-BiLSTM has increased by 4% , FNR has decreased by 1.3%, NAR is as low as 0%, which are the best among four. From Fig. 3, proposed ATCN-BiLSTM has the best recognition indicies for each type of events. However, it took longer to train our proposed network because ATCN is more complex than other models.

# IV. CONCLUSION

In this paper, we propose an intelligent event classification method based on ATCN-BiLSTM. All features including attention weight vector are automatically learned. ATCN captures temporal feature where causality of time domain signal is taken into account and key channel features are emphasized through attention. Then, BiLSTM is used to extract bidirectional spatial relationship for final recognition. Experimental results show that our proposed ATCN-BiLSTM model achieves a high classification accuracy of 99.6% with three types events, namely, background without threats, climbing the fence and raining, NAR and FNR are also the best, 0% and 0.6% respectively, better than 1D CNN-BiLSTM based method and other two advanced methods on our dataset.


# ACKNOWLEDGMENT

This work was supported in part by the Fundamental Research Funds for the Central Universities under Grant 2020YJS043 and Grant 2020JBM024, in part by A*STAR GAP Funds under Project ACCL/19-GAP032-R20A, in part by the National Natural Science Foundation of China (NSFC) under Grant 61805008, and in part by Outstanding Chinese and Foreign Youth Exchange Program of China Association of Science and Technology.


# REFERENCES


[1] J. C. Juarez, E. W. Maier, K. Nam, and H. F. Taylor, ''Distributed fiber-optic intrusion sensor system,'' J. Lightw. Technol., vol. 23, no. 6, pp. 2081–2087, Jun. 2005.

[2] J. C. Juarez, H F.Taylor. "Polarization discrimination in a phase-sensitive optical time-domain reflectometer intrusion-sensor system," Optics Letters, vol. 30, no.24, pp. 3284-3286, 2005.

[3] H. Li, Z. Zhang, F. Jiang, and X. Zhang, ''An event recognition method for fiber distributed acoustic sensing systems based on the combination of MFCC and CNN,'' Proc. SPIE, vol. 10618, Jan. 2017, Art. no. 1061804.

[4] J. Chen, H. Wu, X. Liu, "A real-time distributed deep learning approach for intelligent event recognition in long distance pipeline monitoring with DOFS," 2018 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery, pp. 290-296, 2018.

[5] Chen X, Xu C. Disturbance pattern recognition based on an ALSTM in a long-distance phi-OTDR sensing system [J]. Microwave and Optical Technology Letters, 2020, 62(1): 168-175.

[6] Jun W, Luyang G, Ming B, et al. Vibration events recognition of optical fiber based on multi-scale 1-D CNN [J]. Opto-Electronic Engineering, 2019.

[7] Wu H, Yang M, Yang S, et al. A Novel DAS Signal Recognition Method Based on Spatiotemporal Information Extraction With 1DCNNs-BiLSTM Network [J]. IEEE Access, 2020, 8119448-119457.

[8] S. Bai, J.Z. Kolter, V. Koltun, "An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling," Technical Report, 2018.