2025 International Conference on Intelligent Computing
July 26-29, Ningbo, China
https://www.ic-icc.cn/2025/index.php

MS-ConformerNet: A Multi-Scale Joint Encoding Network for OTDR Signal Analysis

Xueqing Xu[1,2], Zhihui Sun[1,2,*], Jiwen Xu[3] ,Shaodong Jiang[1,2] , Shilei Wei[3] , Faxiang Zhang[1,2] and Xianlong Liu[3]

[1] Key Laboratory of Computing Power Network and Information Security, Ministry of Education, Shandong Computer Science Center (National Supercomputer Center in Jinan), Qilu University of Technology (Shandong Academy of Sciences), Jinan, China
[2] Shandong Provincial Key Laboratory of Industrial Network and Information System Security, Shandong Fundamental Research Center for Computer Science, Jinan, China
[3] Qingdao Novker Communication Technology,Qingdao, China
bitszh@163.com,shdjiang@qlu.edu.cn,zhangfx@semi.ac.cn

Abstract. This paper proposes a multi-task deep neural network architecture for optical time-domain reflectometer (OTDR) signal analysis, enabling end-to-end learning for fiber fault classification and event localization. To address the limitations of traditional methods in complex scenarios, such as insufficient feature representation and conflicts in multi-task optimization, this study designs a multi-scale pooling module to extract cross-scale features, integrates an improved bi-directional feature pyramid network (BiFPN) to enhance multi-resolution feature fusion, and introduces a Conformer hybrid encoding block that combines self-attention and gated convolution to model both global and local features. Additionally, a task-aware dynamic gating mechanism is proposed to mitigate conflicts in multi-objective optimization. Experimental results demonstrate that the proposed model outperforms traditional methods in classification accuracy and fault localization, providing a high-precision, cost-effective solution for optical network monitoring and maintenance.

Keywords: OTDR, Multi-task, Deep Learning, Fault Diagnosis.

1 Introduction

The Optical Time Domain Reflectometer (OTDR) [1], as a core tool for fiber network fault diagnosis, plays a crucial role in modern communication infrastructure. Its working principle involves emitting optical pulses into the fiber and detecting the backscattered signals [2] to generate a reflection loss curve of the fiber link. These curves can accurately locate faults within the fiber, such as breaks, splice losses [3] , and connector degradation, while also providing quantitative information on fiber quality. However, with the continuous expansion and increasing complexity of fiber networks, traditional OTDR analysis methods face challenges such as noise interference, insufficient detection capability for weak signals, and imprecise fault feature extraction. According to the International Telecommunication Union (ITU) report in 2023, global economic losses due to fiber failures exceed \$1.2 billion annually, with approximately 35% of faults failing to trigger timely warnings due to the sensitivity limitations of traditional OTDR systems.

In recent years, machine learning techniques, particularly deep learning models, have demonstrated significant advantages in OTDR signal analysis. For example, Zhimin Yang et al.[4] 2021 developed a high-sensitivity OTDR event detection method based on machine learning. This method preprocesses OTDR signals using n-order differencing and denoising techniques, followed by a machine learning classifier, achieving a 95% detection rate for connection splice events. Khouloud Abdelli et al. [5] 2021 proposed a BiLSTM-CNN-based multitask learning approach for fiber fault diagnosis in OTDR signals, which outperforms traditional methods, especially in low signal-to-noise ratio (SNR) conditions. However, existing methods typically treat fault classification and localization as independent tasks, overlooking their intrinsic relationship. This results in inefficient utilization of model parameters and insufficient feature sharing. Furthermore,most studies [6,7,8,9,10] rely on simulated data for model training and validation, lacking adequate modeling of the complex characteristics of real-world optical fiber links.

Traditional OTDR signal analysis methods [11,12,13] primarily rely on rule-based threshold detection and manual feature extraction techniques. While these methods perform well in detecting prominent faults, their limitations become increasingly evident in complex scenarios. Existing OTDR analysis techniques face three major challenges: first, threshold-based detection methods exhibit instability in complex fiber links, leading to an increased false alarm rate as the number of fault points grows; second, deep models based on single-task learning struggle to balance the conflicting objectives of classification and localization, resulting in inefficient parameter utilization; finally, current methods lack the ability to effectively model the complex characteristics of real-world fiber networks, making it difficult to adapt to diverse deployment environments. For instance, the attenuation characteristics and fault patterns of optical fibers vary significantly between urban underground pipelines and submarine optical cables. However, existing approaches often rely on simulated data for training, limiting their generalization capabilities in real-world applications.

To address these challenges, this paper proposes a Multi-Scale Conformer Network (MS-ConformerNet), which enhances the detection of weak signals and adapts to different fault patterns by extracting multi-scale contextual features using pooling kernels of varying sizes. The traditional feature concatenation approach is improved to enable dynamic fusion of high- and low-level features. By integrating gated convolution and multi-head attention, the model adaptively focuses on important features, improving fault detection accuracy in complex environments.To tackle the issues of low parameter utilization and insufficient feature sharing, a task-adaptive routing mechanism is introduced. This mechanism employs differentiable task gating to achieve gradient coordination between classification and localization, dynamically adjusting the fusion ratio of task-specific and shared features. On our dataset, the proposed network achieved a classification accuracy of 97.12% and a localization error of 0.86.

The remainder of this paper is structured as follows: Section 2 introduces relevant background knowledge, including OTDR signal characteristics, multi-task learning theory, and deep learning applications in OTDR analysis. Section 3 provides a detailed description of the proposed MS-ConformerNet's cascaded architecture and key modules. Section 4 presents experimental results and performance analysis. Finally, Section 5 concludes the study and discusses future research directions.

2 Related Work

This section introduces the basic characteristics of Optical Time-Domain Reflectometer (OTDR) signals, the theoretical frameworks of multi-task learning techniques, and the application of deep learning in OTDR signal analysis, thereby providing a solid theoretical foundation for the proposed multi-task deep neural network architecture.

2.1 OTDR Signal Characteristics

Optical Time Domain Reflectometer (OTDR) is an instrument specifically designed for optical fiber network testing and fault diagnosis. It works by sending light pulses into the optical fiber and analyzing the backscattered and reflected signals generated during the light pulse's transmission through the fiber. The resulting attenuation-distance curve is used to assess the health of the optical fiber link.

OTDR signals typically consist of position points, loss, and reflective power. Under normal conditions, the OTDR signal presents a smooth attenuation curve, while in the case of fiber faults, such as fiber breaks, increased connector loss, or splice defects, the signal exhibits abrupt changes, forming distinct feature points. OTDR can measure key parameters such as the total length of the fiber, loss levels, and reflection loss. However, the identification of specific event points often requires manual or intelligent algorithmic analysis to accurately determine various types of events in the optical fiber line.

OTDR signals are strongly spatially dependent and are influenced by the nonlinear effects of the fiber itself and environmental noise. The fluctuation patterns of signals at different position points vary significantly. Especially in complex network environments, noise interference can blur signal features, increasing the difficulty of fault detection. Therefore, effectively extracting fault features from OTDR signals to improve the automation, accuracy, and robustness of fault diagnosis is a crucial research direction in the field of optical fiber monitoring.

2.2 Multi-Task Learning (MTL)

Multi-Task Learning (MTL) [14] is a machine learning framework that simultaneously learns multiple related tasks by sharing a portion of the model's parameters. It aims to improve the model's generalization ability and learning efficiency through knowledge transfer and feature sharing across tasks. Compared to Single-Task Learning (STL), MTL makes better use of data information, especially in scenarios where tasks are strongly correlated.

In Optical Time-Domain Reflectometer (OTDR) signal analysis, fault classification and event localization are two closely related tasks. Traditional methods often handle these tasks separately, optimizing individual objective functions, which may lead to inefficient parameter utilization and insufficient feature sharing. However, classification and localization tasks share common characteristics, such as fiber attenuation patterns and noise distribution. By adopting MTL, both classification and localization results can be generated simultaneously, meeting real-time monitoring requirements. Through a single forward pass, MTL not only improves fault detection accuracy but also reduces training time and enhances parameter efficiency.

To address issues such as task conflicts and the rationality of feature sharing in MTL, researchers have proposed various improvements, such as task-specific dynamic routing mechanisms [15] and differentiable gating techniques [16]. These methods help optimize gradient conflicts [17] and feature sharing across tasks, further enhancing model performance in complex environments.

2.3 Applications of Deep Learning in OTDR Signal Analysis

The widespread application of OTDR technology in fiber optic communications and the demand for high-precision fault detection and localization have driven research in this field. The successful application of deep learning in signal processing domains, such as image recognition and speech processing, has provided new insights for OTDR signal analysis. Deep learning is primarily used for tasks such as fault classification, fault localization, anomaly detection, and signal denoising.Common deep learning models for OTDR signal analysis include Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and their variants. CNNs, which excel in image processing, are well-suited for OTDR signal feature extraction due to their local receptive fields and weight-sharing properties. For instance, ResNet-based models effectively extract fault features from OTDR signals through deep network structures. Meanwhile, Long Short-Term Memory (LSTM) networks and Bidirectional LSTMs (BiLSTMs) have advantages in processing sequential data, offering new possibilities for addressing the challenges of event detection and high-precision localization.

Furthermore, the recent development of Transformer architectures [18] has introduced new approaches to OTDR signal analysis. The self-attention mechanism can establish global dependencies across the entire OTDR signal sequence, facilitating the capture of long-range relationships and enhancing fault detection robustness. Additionally, integrating a multi-task learning framework enables the collaborative optimization of fault classification and event localization, improving overall model performance.

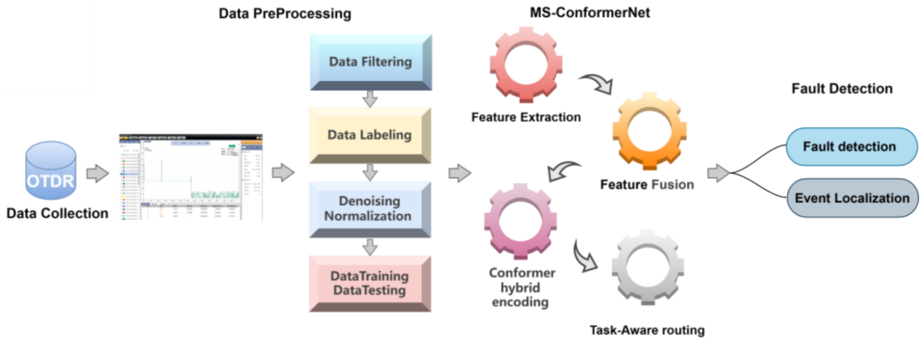

Fig. 1. Data Processing Flowchart with Conformer Hybrid Encoding.

3 Method

This section provides a detailed introduction to the model-based data processing pipeline(**Fig. 1**) and the cascaded architecture of MS-ConformerNet, along with its key modules for optical time-domain reflectometry (OTDR) signal analysis.

3.1 Data Collection

Our dataset is sourced from multiple optical transmission network operators, collected in real-world engineering environments using OTDR tests. It systematically includes signal data generated by different OTDR devices across various fiber link lengths and types, comprehensively covering common reflective events (such as fiber breaks and connector losses) as well as non-reflective events (such as splicing points and fiber bends).

3.2 Data Preprocessing

Due to the large scale of the raw data, which contains a significant number of non-target events and high-noise samples, a specialized annotation tool was developed to filter and label the raw data. This process resulted in the construction of a standardized dataset comprising 9,723 samples, including 3,925 segments with no events, 2,993 segments with non-reflective events, and 2,805 segments with reflective events.

OTDR signals are highly susceptible to system noise and environmental interference. To enhance data quality, denoising, normalization, and downsampling were applied as preprocessing steps. Wavelet denoising was used to smooth the signals, reducing the impact of Rayleigh scattering noise. To eliminate amplitude differences caused by variations in OTDR devices, min-max normalization was applied to both loss and reflective power values, scaling them to the [0,1] range, ensuring feature consistency across different signals.Additionally, the raw OTDR data was originally stored in SOR format, which was converted to a CSV format for standardization. Finally, to evaluate the proposed model's performance, the dataset was randomly split into training (60%), validation (20%), and test sets (20%). The training set was used for model training, the validation set for hyperparameter optimization, and the test set for final evaluation.

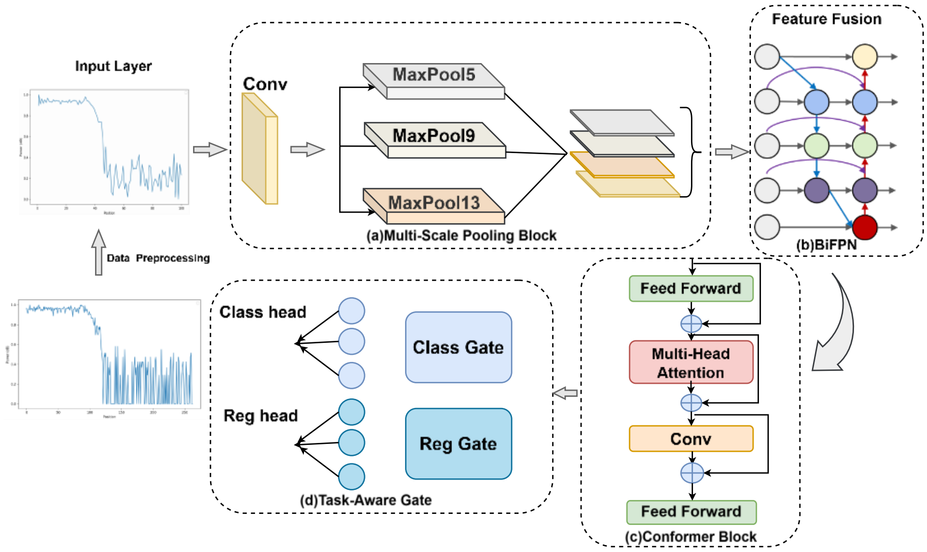

Fig. 2. MS-ConformerNet Model Architecture.

3.3 MS-ConformerNet

As shown in **Fig. 2**, the model utilizes end-to-end learning to simultaneously perform fault type classification and event location regression. Its core architecture consists of four key modules: multi-scale feature extraction, bidirectional feature fusion, hybrid encoding, and task-aware routing.

Multi-Scale Feature Extraction Module. To capture the multi-scale characteristics of reflection/scattering events in OTDR signals, we design a Multi-Scale Pooling module (**Fig. 2** (a)). The Multi-Scale Pooling module[19] aims to extract multi-scale features from the one-dimensional input signal and enhance feature representation through BiFPN-based feature fusion.Upon receiving the one-dimensional input data, the module first applies a Conv1D layer to perform convolution operations, generating an initial feature representation. Then, multiple parallel max-pooling operations with different kernel sizes (5×1, 9×1, and 13×1) are applied. These varying scales of pooling operations capture signal variations over short, medium, and long ranges, enhancing the model's ability to perceive different event patterns.Each pooling layer adopts a stride of 1 and symmetric zero-padding of kernel_size//2 to maintain the spatial alignment of features. Instead of traditional feature concatenation, BiFPN architecture is introduced in the feature fusion stage to enable dynamic weighted feature aggregation, improving the adaptability and expressiveness of the extracted features.

Bidirectional Feature Pyramid Network (BiFPN). The traditional Feature Pyramid Network (FPN)[20] is primarily used in computer vision tasks, enhancing feature representation through top-down and bottom-up information flow. However, in OTDR signal analysis, effectively integrating multi-level features remains a challenge.

To enhance the representation capability of multi-resolution features, an improved BiFPN structure suitable for one-dimensional signals is proposed (**Fig. 2** (b)). After multi-scale pooling, instead of the traditional direct channel concatenation, BiFPN is used for dynamic weighted fusion.Bilinear interpolation is used to achieve spatial alignment between high- and low-level features, while a dual-path information flow, combining top-down and bottom-up processing:

$$F_{td} = W_1 \odot (\mathrm{T}(\mathrm{P}(F_{high}))) + W_2 \odot F_{low} \tag{1}$$

$$F_{out} = C_{3\times1}(\mathrm{BN}(\mathrm{GELU}(C_{3\times1}(F_{td})))) \tag{2}$$

Where W_1 and W_2 are learnable weight parameters,and *$(T(P(F_{high})))$* represents the process of upsampling and transforming the high-level feature F_{high} to match the resolution of the low-level feature F_{low}.By fusing high-level and low-level features, an intermediate feature F_{td} is generated, which contains multi-scale information. The fused feature F_{td} then undergoes further nonlinear transformation and enhancement to produce the final feature representation F_{out}.This structure enhances the model's multi-scale perception capability by integrating features from different abstraction levels.

Conformer Block. Inspired by the field of speech recognition[21], this paper introduces the Conformer hybrid encoding block. This module combines the advantages of Convolutional Neural Networks (CNNs) and the Transformer architecture to design a spatiotemporal joint encoding module (**Fig. 2** (c)).

Each encoding block consists of: (1) Multi-head self-attention mechanism,Utilizing four attention heads to capture global dependencies between different positions within the fused feature information. (2) Gated convolutional unit,Implementing feature selection through GLU activation. (3) Feedforward network,Incorporating Swish activation to enhance nonlinear representation.Each submodule employs residual connections and layer normalization to ensure stable training.

Task-Aware routing. To address the feature conflict issue in multi-task learning, a task-specific gating mechanism is designed (**Fig. 2** (d)). Given shared features, task gating weights are generated through learnable parameters:

$$g_k = \sigma(W_g^k F_s + b_g^k),\ \ \mathrm{k} \in cls, reg \tag{3}$$

$$F_k = g_k \odot (W_p^k F_s) + (1 - g_k) \odot F_s \tag{4}$$

The shared feature F_s is a feature vector obtained from the aforementioned feature extraction module.The gating coefficient g_k dynamically adjusts the fusion ratio between task-specific features and shared features, while represents the task-specific projection matrix. Specifically,W_p^k maps the shared feature *Fs* into a task-specific feature space.

Finally, F_k is the weighted combination of task-specific and shared features. This mechanism enables the model to adaptively allocate feature resources for different tasks.

The network output layer employs a divide-and-conquer strategy: the classification head uses a three-layer MLP to map the features to a probability distribution over 3 classes (normal/reflection event/non-reflection event), while the regression head uses a linear layer to output the event position offset.

The training process adopts a weighted loss function:

$$L = \alpha L_{cls} + \beta L_{reg} \tag{5}$$

The weight coefficients for the cross-entropy loss L_{cls} and the MSE loss L_{reg} were determined through grid search as α=0.5 and β=0.5, respectively.

4 Experiments

This section first introduces the dataset preparation process, followed by an evalua-tion of the model's event classification and localization capabilities. Finally, an abla-tion study is conducted to verify the contribution of each module to the model's per-for-mance. The results indicate that all modules play a crucial role in improving the model's performance.

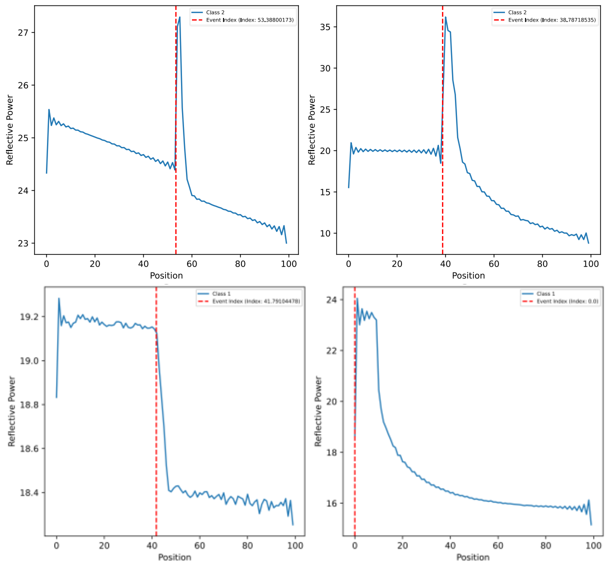

Fig. 3. OTDR reflection events and non-reflection event examples.

4.1 Dataset of OTDR Traces

In this study, a specialized annotation tool was developed for OTDR signal data, enabling the selection and construction of a standardized dataset with 9,723 samples (including 3,925 no-event segments, 2,993 non-reflective event segments, and 2,805 reflective event segments). Data quality was enhanced through wavelet denoising and normalization, and each signal segment was standardized to a fixed sequence length of 100 sampling points.**Fig. 3** illustrates examples of reflective and non-reflective signal segments.The dataset was randomly split into training (60%), validation (20%), and test sets (20%). The model takes reflective power values (Power) as input and outputs the corresponding event category (Classid), where 0 represents no event, 1 represents a non-reflective event, and 2 represents a reflective event. Additionally, the model predicts the position indices of both reflective and non-reflective events within the sequence, facilitating further analysis of the fiber link status.

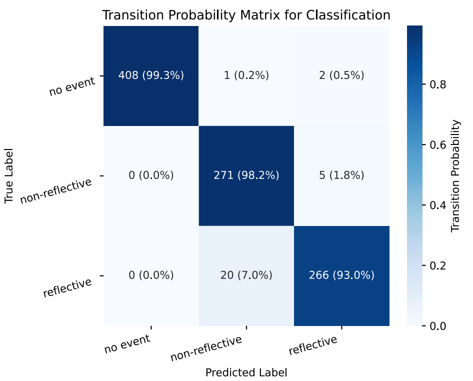

Fig. 4. Confusion Matrix

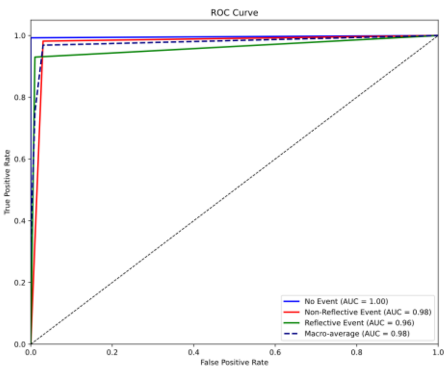

Fig. 5. ROC curve for Fault detection

4.2 Fault detection capability

To evaluate the fault detection capability of the model, we adopted the following classification metrics: Accuracy, Precision, Recall, and F1 Score, which are reported in Table 1.Additionally, as shown in the confusion matrix **Fig. 4**, the model achieves high classification accuracy for no-event and non-reflective events, while its performance on reflective events is slightly lower. This may be due to an imbalanced data distribution, but the overall performance remains strong.Furthermore, ROC curve analysis (**Fig. 5**) confirms the model's robustness, with an Area Under the Curve (AUC) close to 0.98.

4.3 Event localization capability

To comprehensively evaluate the model's event localization ability, we use Mean Squared Error (MSE) and Mean Absolute Error (MAE) as the primary evaluation metrics. MSE reflects the average squared size of position errors in the prediction process, while MAE provides the average deviation between predicted positions and actual event locations. These metrics allow us to clearly quantify the model's performance in different event detection and localization tasks.

The experimental results, as shown in Table 1, indicate that the mean squared error for event localization is around 1 meter, and the model can accurately predict event locations. Additionally,**Fig. 6** illustrates the model's performance on sample curves. After training, the model achieves a low misclassification rate and high event localization accuracy. Although two positions were incorrectly detected as events, overall, the predicted event locations are very close to the actual positions.

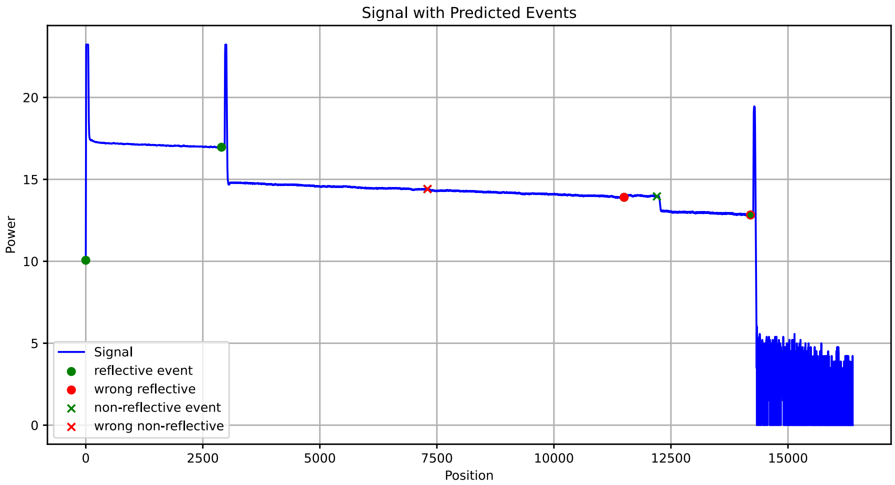

Fig. 6. The Performance of the Model in Event Detection and Localization on Signal Curves.

4.4 Ablation Study

To validate the effectiveness of each module in our model, we conducted an ablation study. Specifically, we evaluated the contributions of the multi-scale feature extraction module, bidirectional feature pyramid network (BiFPN), and Conformer hybrid encoder to the overall model performance. All ablation experiments were performed under identical experimental conditions, with the only variable being the removed module. Specifically, all experiments used the same dataset, identical hyperparameters (learning rate = 0.0001, batch size = 128), and the same training configuration (200 training epochs with an early stopping mechanism, patience = 25). Furthermore, to ensure fairness, we fixed the random seed to maintain consistency in data splitting and model initialization, eliminating potential influences from data distribution variations or random weight initialization. By progressively removing each module, we assessed the impact on the model's performance in fault detection and event localization tasks.

As shown in **Table 1**, BiFPN contributed the most significantly to model performance. Its removal resulted in an approximately 5% drop in fault detection accuracy and a 0.32m increase in event localization error. This indicates that BiFPN plays a crucial role in multi-level feature fusion, particularly in enhancing the model's ability to learn complex OTDR signal patterns. Next in importance were the Conformer hybrid encoder and the multi-scale feature extraction module. The Multi-Scale Pooling module primarily affects event localization accuracy, as its removal led to a noticeable increase in localization error.

In summary, BiFPN, Conformer, and Multi-Scale Pooling work together to enable the model to achieve optimal performance in both fault detection and event localization tasks.

Table 1. Ablation Study.

	Task 1:Fault detection				Task 2:Event localization	
Method	Accurary(%)	Precision(%)	Recall(%)	F1 score(%)	MSE(m)	MAE(m)
MS-ConformerNet	**97.12**	**97.21**	**97.12**	**97.13**	**1.02**	**0.86**
No-MultiScalePool	93.86	93.89	93.86	93.86	1.41	1.15
No-BiFPN	92.83	92.79	92.83	92.79	1.34	1.10
No-Conformer	93.22	93.87	93.22	93.12	1.39	1.13

4.5 Comparison with Existing Methods

To validate the effectiveness of the proposed model, MS-ConformerNet, this study selects three representative temporal modeling methods as comparative baselines: ResNet1D, which is based on residual networks; DeepConvLSTM, which combines convolutional and recurrent neural networks; and the Temporal Convolutional Network (TCN). All comparative experiments were conducted under a unified experimental setup, using the same dataset partitioning and a multi-task learning framework. The tasks include classification of OTDR reflection event types and precise regression-based localization of reflection points.

Table 2 presents the performance of each model in terms of classification accuracy, F1 score, and regression mean squared error (MSE). The results demonstrate that MS-ConformerNet achieves superior performance on both tasks: it outperforms all baseline models in terms of classification accuracy and F1 score, while also achieving the lowest MSE in the regression task. This indicates that MS-ConformerNet possesses stronger capabilities in feature representation and modeling for both event detection and localization.

A detailed analysis of the baseline models reveals the following,ResNet1D benefits from its residual connection structure, showing strong feature extraction capabilities with a classification accuracy of 91.95%. However, due to the lack of an explicit temporal modeling mechanism, its regression MSE (1.63m) remains relatively high. DeepConvLSTM can capture temporal information to a certain extent, but suffers from gradient vanishing when processing long sequences, leading to limited generalization and subpar classification performance.TCN performs robustly in modeling long-range dependencies, but its simplistic inter-layer feature fusion makes it less effective for joint optimization of classification and regression tasks.

In contrast, MS-ConformerNet enhances the integration of local and global information through a multi-scale convolutional module and a BiFPN pyramid structure. The Conformer encoder further captures temporal dynamics, while a task-aware gating mechanism enables adaptive information routing within the shared feature space across tasks. Experimental results show that the model not only improves classification accuracy but also significantly reduces regression error for reflection point localization, demonstrating strong task-coordinated modeling capabilities.

Table 2. Comparison with Existing Methods.

	Task 1:Fault detection		Task 2:Event localization	
Methods	Accurary(%)	F1 score(%)	MSE(m)	MAE(m)
ResNet1D	91.95	91.96	1.63	1.57
DeepConvLSTM	88.66	89.0	1.49	1.21
TCN	90.54	89.96	1.38	1.52
Proposed Model	**97.12**	**97.13**	**1.02**	**0.86**

5 Conclusion and future work

In this paper,we propose a novel multi-task deep neural network architecture, MS-ConformerNet, to address the joint task of fiber fault classification and precise localization. To overcome the limitations of traditional methods, such as insufficient feature extraction in complex scenarios, task optimization conflicts, and limited generalization capability, we innovatively integrate a multimodal feature learning mechanism. Experimental results demonstrate that the proposed model performs exceptionally well on real OTDR datasets, achieving significant improvements in both fault classification accuracy and event localization precision, validating its feasibility and effectiveness in practical applications.In the future, research will further incorporate advanced deep learning algorithms and multimodal data to enhance fault detection accuracy and generalization ability, while exploring real-time sensor fusion and edge computing optimization to improve system efficiency and adaptability, facilitating its deployment in real-world scenarios.

Acknowledgments. This work was supported by Shandong Provincial Science and Technology Small and Medium sized Enterprises Innovation Capability Enhancement Engineering Project (2024TSGC0221，2023TSGC0425)，National Natural Science Foundation of China (No. U23A20378), Shandong Provincial Natural Science Foundation of China (No. ZR2023LLZ009，ZR2021MF127).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Barnoski, M., Rourke, M., Jensen, S., Melville, R.: Optical time domain reflectometer. Applied optics 16(9), 2375--2379 (1977)
2. Lee, W., Myong, S.I., Lee, J.C., Lee, S.: Identification method of non-reflective faults based on index distribution of optical fibers. Optics Express 22(1), 325--337 (2014)
3. Rad, M.M., Fouli, K., Fathallah, H.A., Rusch, L.A., Maier, M.: Passive optical network monitoring: challenges and requirements. IEEE Communications Magazine 49(2), s45--S52 (2011)
4. Yang, Z., Hong, D., Feng, X., Xie, J.: A novel event detection method for otdr trace with high sensitivity based on machine learning. In: 2021 2nd Information Communication Technologies Conference (ICTC). pp. 265--269 (2021). https://doi.org/10.1109/ICTC51749.2021.9441614
5. Abdelli, K., Grießer, H., Tropschug, C., Pachnicke, S.: A bilstm-cnn based multitask learning approach for fiber fault diagnosis. In: 2021 Optical Fiber Communications Conference and Exhibition (OFC). pp.~1--3 (2021)
6. Dou, X., Yin, H., Hao, Y., Yue, H., Qi, X., Jin, Y., Qin, J., Li, L.: Demonstration of chaotic-laser based wdm-pon secure optical communication and real-time online fiber-fault detection and location. In: 2015 Opto-Electronics and Communications Conference (OECC). pp.~1--3. IEEE (2015)
7. Abdelli, K., Azendorf, F., Grießer, H., Tropschug, C., Pachnicke, S.: Gated recurrent unit based autoencoder for optical link fault diagnosis in passive optical networks. In: 2021 European Conference on Optical Communication (ECOC). pp.~1--4. IEEE (2021)
8. Abdelli, K., Cho, J.Y., Azendorf, F., Griesser, H., Tropschug, C., Pachnicke, S.: Machine-learning-based anomaly detection in optical fiber monitoring. Journal of optical communications and networking 14(5), 365--375 (2022)
9. Abdelli, K., Grießer, H., Tropschug, C., Pachnicke, S.: Optical fiber fault detection and localization in a noisy otdr trace based on denoising convolutional autoencoder and bidirectional long short-term memory. Journal of Lightwave Technology 40(8), 2254--2264 (2021)
10. Kaushik, S., Garg, K., Verma, J., Kansal, I., Sharma, P., Khullar, V., Sethi, M.: Enhancing fault detection and classification in optical fiber networks with deep learning algorithms. In: AIP Conference Proceedings. vol.~3209. AIP Publishing (2024)
11. Liu, F., Zarowski, C.J.: Detection and location of connection splice events in fiber optics given noisy otdr data. part ii. r1msde method. IEEE Transactions on instrumentation and measurement 53(2), 546--556 (2004)
12. Bakar, A., Jamaludin, M.Z., Abdullah, F., Yaacob, M., Mahdi, M., Abdullah, M.: A new technique of real-time monitoring of fiber optic cable networks transmission. Optics and Lasers in Engineering 45(1), 126--130 (2007)
13. Qiang, H., Zhihui, Z., Dongsheng, W., Lei, L., Xibao, H.: An otdr event analysis algorithm based on emd-based denoising and wavelet transform. In: 2015 12th IEEE International Conference on Electronic Measurement & Instruments (ICEMI). vol.~2, pp. 1034--1038. IEEE (2015)
14. Caruana, R.: Learning many related tasks at the same time with backpropagation. Advances in neural information processing systems 7 (1994)
15. Rosenbaum, C., Klinger, T., Riemer, M.: Routing networks: Adaptive selection of non-linear functions for multi-task learning. arXiv preprint arXiv:1711.01239 (2017)
16. Hazimeh, H., Zhao, Z., Chowdhery, A., Sathiamoorthy, M., Chen, Y., Mazumder, R., Hong, L., Chi, E.: Dselect-k: Differentiable selection in the mixture of experts with applications to

multi-task learning. Advances in Neural Information Processing Systems 34, 29335--29347 (2021)
17. Yu, T., Kumar, S., Gupta, A., Levine, S., Hausman, K., Finn, C.: Gradient surgery for multi-task learning. Advances in neural information processing systems 33, 5824--5836 (2020)
18. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I.: Attention is all you need. Advances in neural information processing systems 30 (2017)
19. He, K., Zhang, X., Ren, S., Sun, J.: Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence (2015).
20. Ghiasi, G., Lin, T.Y., Le, Q.V.: Nas-fpn: Learning scalable feature pyramid architecture for object detection. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 7036--7045 (2019)
21. Gulati, A., Qin, J., Chiu, C.C., Parmar, N., Zhang, Y., Yu, J., Han, W., Wang, S., Zhang, Z., Wu, Y., et~al.: Conformer: Convolution-augmented transformer for speech recognition. arXiv preprint arXiv:2005.08100 (2020)