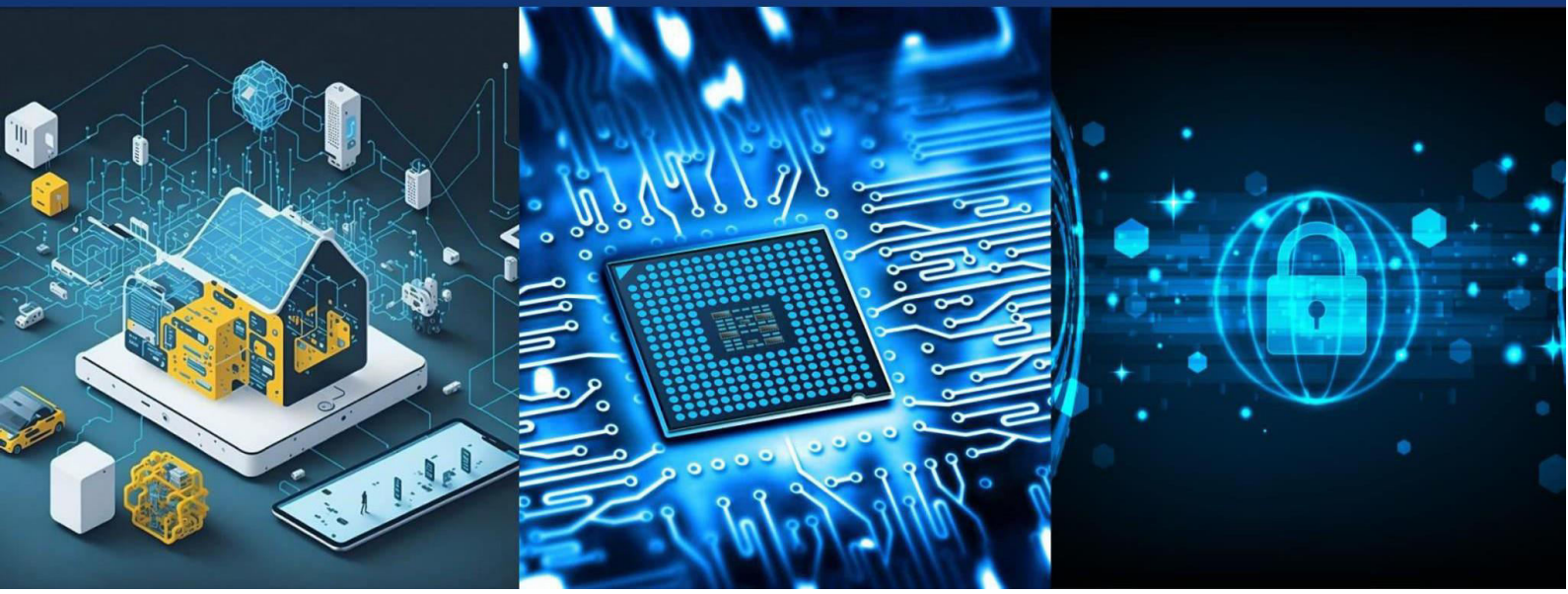
IJIRCCE

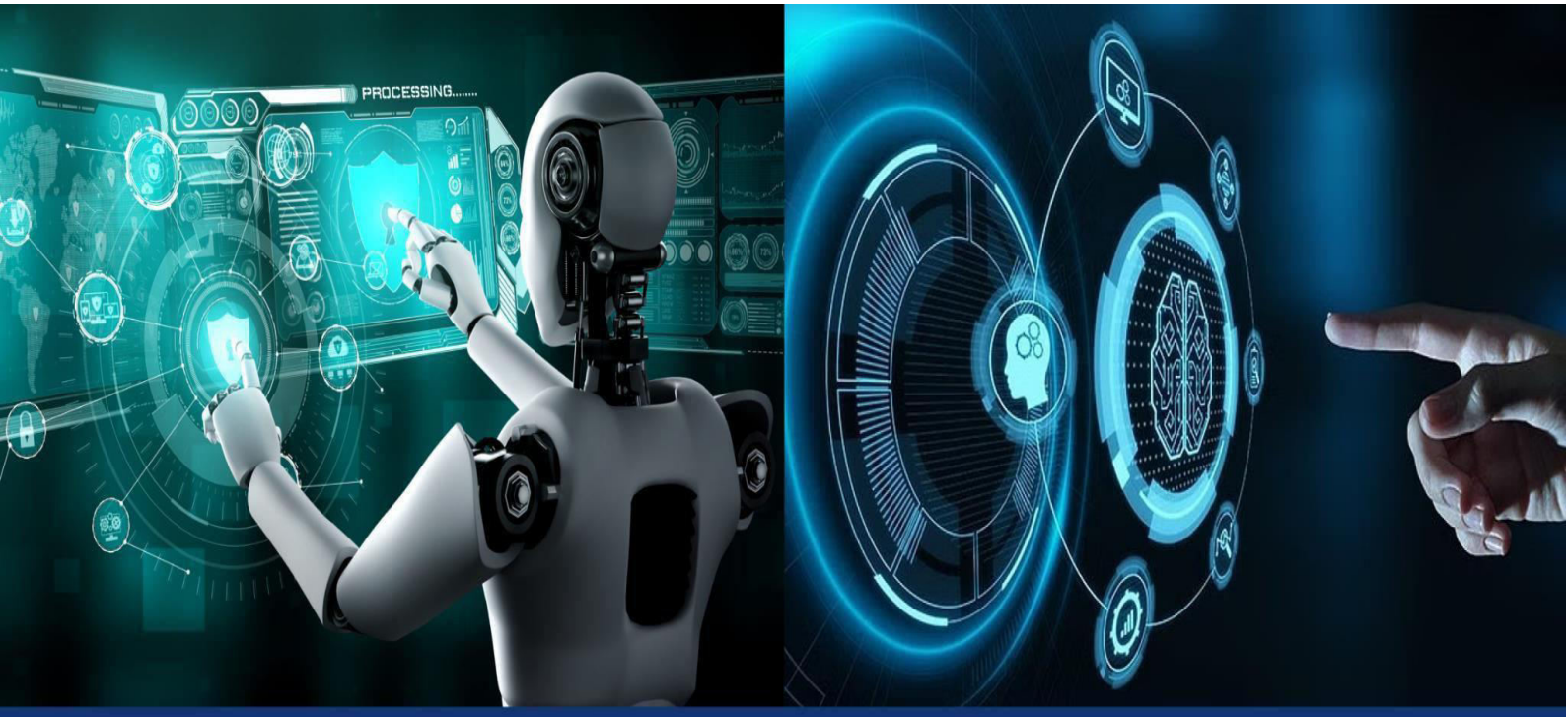
ISSN(O): 2320-9801

ISSN(P): 2320-9798



International Journal of Innovative Research in Computer and Communication Engineering

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.771

Volume 14, Issue 2, February 2026

© 2026 IJIRCCE | Volume 14, Issue 2, February 2026 DOI: 10.15680/IJIRCCE.2026.1305001

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Event Detection and Fault Localization in Distributed Acoustic Sensing

Ezzegraoui Cherouk, Changli Li

Master Student, School of Artificial Intelligence, Nanjing University of Information Science and Technology, Nanjing, China

Professor, School of Artificial Intelligence, Nanjing University of Information Science and Technology, Nanjing, China

ABSTRACT: Event detection and fault localization in Distributed Acoustic Sensing (DAS) systems based on phase-sensitive optical time- domain reflectometry (Φ-OTDR) are critical for applications such as pipeline security and perimeter monitoring. However, long multi-channel Φ-OTDR traces exhibit impulsive, non-stationary, and multi-scale temporal characteristics, which pose significant challenges for existing deep-learning methods. Traditional signal-processing and convolution-based approaches struggle to capture long-range dependencies, while recent efficient Transformer models (e.g., Informer, FEDformer, and TimesNet) rely on sparsity or periodicity assumptions that are poorly matched to impulsive DAS events.

To overcome these limitations, this paper presents the first interpretable CNN–Swin Transformer framework for Φ-OTDR intrusion monitoring, explicitly designed for long 1-D DAS signals. A custom 1-D adaptation of the Swin Transformer with exposed and extractable attention weights enables efficient hierarchical multi-scale temporal modeling with linear computational complexity, while a lightweight CNN branch captures fine-grained local transient patterns. The fused architecture is trained in a unified multi-task learning framework that jointly performs six-class event classification and meter-level fault localization. In addition, post-training INT8 quantization is incorporated to enable efficient real-time edge deployment.

Experiments conducted on a 15,419-trace public Φ-OTDR dataset demonstrate that the proposed method achieves 98.31% classification accuracy, a 0.983 macro F1-score, and a localization mean absolute error of 9.2 m, outperforming both a 1-D CNN baseline and a CNN+vanilla Transformer baseline by 1.05% and 1.91%, respectively, and prior methods by up to 2.91%.

Post-training quantization further reduces model size by 75% and inference latency by 68% with only a 0.18% accuracy drop, confirming the proposed framework's effectiveness, interpretability, and deployability for real-world DAS systems.

KEYWORDS: Distributed Acoustic Sensing, Φ-OTDR, Swin Transformer, CNN, Multi-Task Learning, Fault Localization, Interpretability, Quantization.

I. INTRODUCTION

Distributed Acoustic Sensing (DAS) based on phase-sensitive optical time-domain reflectometry (Φ-OTDR) enables a standard optical fiber to function as thousands of virtual vibration sensors distributed over tens of kilometers[17], providing meter-scale spatial resolution. Owing to its wide coverage, passive sensing capability, and robustness, DAS has become a key technology for pipeline security, perimeter intrusion detection, and structural health monitoring [20], [8]. A typical Φ-OTDR system produces long, multi-channel time-series signals, often comprising 12 spatial channels with up to 10,000 temporal samples per trace, posing significant challenges for accurate event detection and localization.

In practical deployments, intrusion events such as digging, knocking, shaking, watering, and walking exhibit strong temporal variability, multi-frequency characteristics, and non-stationary behavior[23]. These events are frequently corrupted by environ- mental noise and may overlap in both time and space, making reliable classification and precise

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

fault localization particularly difficult. The ability to simultaneously recognize event types and accurately estimate their spatial positions remains a critical requirement for real-world DAS applications[15].

Early approaches to DAS and OTDR event detection primarily relied on traditional signal-processing techniques, including thresholding, matched filtering, and handcrafted spectral features [20]. Although computationally efficient, these methods are highly sensitive to noise and environmental variations and often fail to generalize under realistic operating conditions. More recently, deep learning models have been introduced to learn discriminative representations directly from raw Φ-OTDR traces. Convolutional neural networks (CNNs) have demonstrated strong performance in capturing local transient patterns [28]; however, their inherently limited receptive fields restrict their ability to model long-range temporal dependencies spanning thousands of samples.

Transformer-based architectures provide a natural mechanism for global dependency modeling, but vanilla self-attention incurs quadratic computational complexity with respect to sequence length, rendering it impractical for long DAS signals. To address this limitation, several efficient 1-D Transformer variants, including Informer [29], FEDformer [30], and TimesNet [27], have been proposed. These models achieve reduced complexity by exploiting sparsity, frequency-domain decomposition, or periodicity assumptions and have shown success in time-series forecasting tasks. However, Φ-OTDR intrusion events are typically impulsive, non-periodic, and highly non-stationary, causing these assumptions to be poorly matched and limiting their effectiveness for DAS event detection and localization.

The Swin Transformer [16], originally developed for vision tasks, introduces a hierarchical shifted-window self-attention mechanism that enables efficient multi-scale modeling with linear computational complexity. While Swin-like architectures have shown promise in adapting to 1-D signals, their potential for long, multi-channel Φ-OTDR data remains largely unexplored, particularly in the context of joint event classification, fault localization, and interpretability.

In this work, we propose a novel and efficient deep-learning framework for joint event detection and fault localization in long multi-channel Φ-OTDR signals. The main contributions of this paper are summarized as follows:

- A hybrid CNN–Swin Transformer architecture is proposed for long 1-D Φ-OTDR traces, combining lightweight con- volutional layers for local transient feature extraction with a custom 1-D Swin Transformer for hierarchical long-range temporal modeling with linear computational complexity.
- A multi-task learning framework is developed to simultaneously perform six-class event classification and meter-level fault localization, incorporating a masked regression loss to ensure accurate localization only for positive intrusion events.
- An interpretable attention mechanism is introduced by extracting and aggregating attention weights from the Swin Transformer, enabling physically meaningful visualization of temporal regions contributing to detection and localization decisions.
- Extensive experiments on a real-world Φ-OTDR dataset demonstrate state-of-the-art classification and localization perfor- mance, while post-training INT8 quantization enables efficient real-time edge deployment.

II. RELATED WORK

A. Event Detection in DAS and OTDR Systems

Event detection and classification in Distributed Acoustic Sensing (DAS) and phase-sensitive Optical Time-Domain Reflec- tometry (Φ-OTDR) systems have traditionally relied on signal-processing techniques such as amplitude thresholding, matched filtering, cross-correlation, and handcrafted spectral features [20], [28]. These approaches are computationally efficient and easy to deploy; however, their performance is highly sensitive to environmental noise, temperature-induced phase drift, fiber inhomogeneities, and overlapping vibration patterns commonly observed in real-world deployments.

To improve robustness, early machine-learning methods incorporating support vector machines, random forests, and k-nearest neighbors were proposed [20], [8]. While these models outperform purely rule-based systems, they still depend on manually designed features and fixed temporal windows, which limits their ability to generalize across sensing conditions and to capture long-range temporal dependencies in multi-channel OTDR traces. As a result, their scalability to large-scale or continuously monitored fiber links remains limited.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

B. Deep Learning for DAS and Long Time-Series Analysis

Deep learning-based approaches have significantly advanced DAS event recognition by enabling end-to-end feature learning directly from raw or minimally processed signals. Convolutional Neural Networks (CNNs) are the most widely adopted architecture due to their strong capability in extracting local transient patterns from vibration traces [21], [12], [25]. Numerous studies have demonstrated that CNN-based models outperform traditional feature-engineered pipelines in both accuracy and robustness [12], [24].

To extend temporal modeling capacity, recurrent architectures such as LSTM, GRU, and BiGRU, as well as CNN–RNN hybrid models, have been introduced [11], [6], [19]. These models partially address temporal dependency modeling but suffer from inherent limitations, including sequential processing bottlenecks, vanishing gradients, and reduced efficiency when handling long sequences exceeding several thousand samples. In practical DAS systems, where a single trace may exceed 10,000 samples, these limitations significantly hinder real-time applicability.

Attention-augmented CNN–RNN hybrids and early Transformer-based models further improve long-range dependency modeling. However, vanilla Transformer architectures exhibit quadratic computational complexity with respect to sequence length, making them impractical for real-time DAS processing without aggressive downsampling or window truncation, which in turn degrades localization precision and event integrity[22].

C. Efficient Transformer Architectures for Long Sequences

To address the scalability limitations of vanilla Transformers, several efficient Transformer variants have been proposed for long-sequence time-series modeling, including Informer [29], FEDformer [30], and TimesNet [27]. These methods reduce computational complexity by introducing sparse attention, frequency-domain decomposition, or periodicity priors, and have demonstrated strong performance in forecasting and anomaly detection tasks.

Nevertheless, these models are primarily designed for stationary or quasi-periodic time-series data. In contrast, DAS and Φ-OTDR signals are characterized by impulsive, non-stationary, and event-driven patterns with abrupt temporal changes. As a result, the assumptions underlying many efficient Transformers do not align well with the physical characteristics of intrusion and vibration events in optical fiber sensing, often leading to suboptimal performance in classification and localization tasks[10]. The Swin Transformer introduces a hierarchical shifted-window self-attention mechanism with linear computational com- plexity, originally developed for two-dimensional vision tasks [16]. By combining local window attention with cross-window information exchange, Swin-based architectures achieve a favorable balance between modeling capacity and efficiency. Recent studies have adapted Swin-like hierarchies to one-dimensional signals, demonstrating improved multi-scale representation learning. However, their application to multi-channel DAS and Φ-OTDR data remains limited, and their potential for joint classification, localization, and deployment-oriented optimization has not been fully explored[14].

D. Comparison with Existing DAS Classification Methods

Table I summarizes representative DAS and OTDR event-classification models evaluated on the public Phi-OTDR dataset [21]. Existing approaches achieve competitive performance; however, most prior works rely on convolution-only architectures with limited receptive fields, generic efficient Transformers that are not tailored to impulsive sensing signals, or private datasets that hinder fair comparison. Moreover, practical considerations such as model interpretability, localization accuracy, and edge deployment are often overlooked [9].

TABLE I: Comparison of representative DAS/OTDR event-classification methods on the public Phi-OTDR dataset

Method	Year	Acc.	F1	Data	Ref.
CNN-LSTM-Att	2024	97.20%	0.971	Phi	IEEE'24
ResNet-CRNN	2024	96.80%	0.965	Priv.	Sens.'24
ViT-OTDR	2023	95.50%	0.954	Phi	OptExp.'23
CNN-BiGRU-Att	2025	96–97%	–	Priv.	Neuroc.'25
Ours 1D-CNN	–	**95.91%**	**0.9591**	Phi	Ours
Ours TF-B	–	**96.64%**	**0.9664**	Phi	Ours
Ours CNN–Swin-T	–	**98.31%**	**0.9831**	Phi	Ours

In contrast, the proposed CNN–Swin Transformer framework is specifically designed to address these limitations by com- bining CNN-based local feature extraction with hierarchical window-based attention for efficient long-range

© 2026 IJIRCCE | Volume 14, Issue 2, February 2026 DOI: 10.15680/IJIRCCE.2026.1305001

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

modeling. Unlike prior convolution-only or generic Transformer-based approaches, the proposed method supports joint classification and meter- level localization, provides interpretability through attention-map analysis, and incorporates post-training INT8 quantization to enable real-time edge deployment under identical dataset and evaluation protocols. These characteristics distinguish the proposed approach from existing DAS classification methods that primarily focus on accuracy without considering scalability, interpretability, or deployment feasibility.

III. PROPOSED ALGORITHM

Intrusion events in Φ-OTDR signals exhibit diverse and highly non-stationary spatio-temporal characteristics, ranging from short impulsive bursts to sustained low-frequency oscillations and complex multi-scale patterns. These characteristics, illustrated by the real-phase textures across frequency bands in Fig. 4, impose stringent requirements on feature extraction and temporal modeling. Models must simultaneously capture fine-grained local transients and long-range dependencies spanning thousands of temporal samples.

Conventional convolutional neural networks are effective at extracting local patterns but struggle to model long-range dependencies beyond a limited receptive field. Conversely, vanilla Transformer architectures provide global modeling capability but incur prohibitive $O(L^2)$ computational complexity for long sequences, such as the $L = 10{,}000$ samples commonly encountered in Φ-OTDR traces. Although recent efficient 1-D Transformer variants reduce complexity through sparsity or periodicity assumptions, these assumptions are poorly matched to the impulsive and non-periodic nature of DAS intrusion events.

To capture fine-grained local temporal characteristics efficiently, a lightweight 1-D convolutional neural network is employed as the local feature extraction branch. Let $\mathbf{X} \in \mathrm{R}^{C\times L}$ denote the input Φ-OTDR signal, where $C = 12$ and $L = 10{,}000$, and let $\mathbf{F}_{\mathrm{CNN}}^{(0)} = \mathbf{X}$. Local features are extracted via successive 1-D convolutional layers defined as

$$\mathbf{F}_{\mathrm{CNN}}^{(l)} = \sigma\left(\mathrm{BN}\left(\mathbf{W}^{(l)} * \mathbf{F}^{(l-1)} + \mathbf{b}^{(l)}\right)\right), \quad \mathbf{F}_{\mathrm{CNN}} \qquad (1)$$

where $*$ denotes the 1-D convolution operator, $\mathbf{W}^{(l)}$ and $\mathbf{b}^{(l)}$ are the learnable weights and biases of the l-th layer, BN denotes batch normalization, and $\sigma(\cdot)$ represents the ReLU activation function.

To address these challenges, we propose a hybrid CNN–Swin Transformer architecture that integrates local convolutional processing with hierarchical shifted-window self-attention. The proposed design enables efficient multi-scale modeling of long multi-channel Φ-OTDR signals while maintaining linear computational complexity. An overview of the complete framework is shown in Fig. 1.

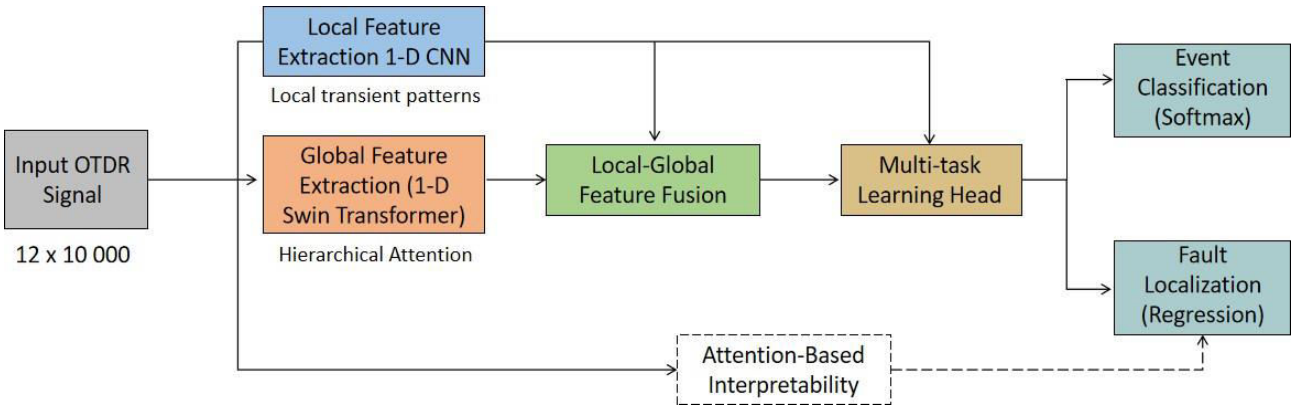


Fig. 1: Overview of the proposed CNN–Swin Transformer framework for Φ-OTDR intrusion monitoring. The architecture integrates a lightweight 1-D CNN branch for local temporal feature extraction with a custom 1-D Swin Transformer for hierarchical global context modeling using shifted-window self-attention. Features from both branches

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

are concatenated and jointly optimized in a multi-task learning setting to perform event type classification and fault location regression.

A. Input Representation

Each Φ-OTDR sample is represented as a multi-channel temporal signal $\mathbf{X} \in \mathrm{R}^{C\times L}$, where $C = 12$ denotes the number of spatial sensing channels along the fiber and $L = 10{,}000$ denotes the temporal length of each trace. The Φ-OTDR signal can be formally modeled as

$$\mathbf{X} = \{x_c(t) \mid c = 1, \ldots, C,\ t = 1, \ldots, L\}, \quad (2)$$

where $x_c(t)$ denotes the measured vibration response at channel c and temporal index t.

Raw traces are used directly without handcrafted feature extraction and are processed in parallel by the CNN and Swin Transformer branches.

B. Local Feature Extraction via CNN

A lightweight one-dimensional CNN branch is employed to extract fine-grained local temporal features. The input signal is first temporally downsampled to reduce computational cost, followed by successive convolutional layers that progressively increase the feature dimensionality. Batch normalization and ReLU activation are applied to stabilize training and enhance non- linearity. Finally, global average pooling is used to aggregate the extracted features into a compact representation $\mathbf{F}_{\mathrm{CNN}} \in \mathrm{R}^{128}$, which captures local transient characteristics of the Φ-OTDR signal.

Temporal downsampling is performed via strided convolution, defined as

$$\tilde{}(l)$$
$$\mathrm{CNN}$$
$$= \mathbf{W}_s *_s \mathbf{F}^{(l-1)} + \mathbf{b}_s, \quad \mathrm{CNN}(3)$$

where $*_s$ denotes 1-D convolution with stride s, and $\mathbf{W}_s$ and $\mathbf{b}_s$ are learnable downsampling kernel weights and bias.

C. Hierarchical Modeling with a Custom 1-D Swin Transformer

To model long-range temporal dependencies efficiently, a custom one-dimensional adaptation of the Swin Transformer is introduced. The input signal is reshaped into a structured representation compatible with shifted-window attention, allowing hierarchical processing of long sequences [13]. Patch embedding is performed to project local temporal regions into a latent feature space.

Specifically, the input sequence is first partitioned into non-overlapping 1-D patches of length P, resulting in

$$\mathbf{X}_p = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N], \quad \mathbf{x}_i \in \mathrm{R}^{C\times P}, \quad (4)$$

where $N = L/P$. Each patch is then linearly projected into a d-dimensional embedding space via

$$\mathbf{z}_i = \mathbf{W}_p \cdot \mathrm{vec}(\mathbf{x}_i) + \mathbf{b}_p, \quad (5)$$

where $\mathbf{W}_p \in \mathrm{R}^{d\times CP}$ and $\mathbf{b}_p$ are learnable parameters. The embedded patch sequence is then formed as

$$\mathbf{Z} = [\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_N] \in \mathrm{R}^{N\times d}. \quad (6)$$

The embedded sequence is further partitioned into non-overlapping temporal windows as

$$\mathbf{Z} = \bigcup_{k=1} \mathbf{Z}^{(k)}, \quad \mathbf{Z}^{(k)} \in \mathrm{R}^{W\times d}, \quad (7)$$

where $K = N/W$ denotes the total number of temporal windows. Within each Swin Transformer stage, window-based self-attention is applied over contiguous temporal windows of size W. To enable cross-window information exchange while maintaining linear computational complexity, shifted-window self-attention alternates between regular and shifted window configurations.

Within each temporal window, self-attention is computed as

$$\mathrm{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \mathrm{softmax}$$
$$\mathbf{QK}^{\top}$$
$$\sqrt{d}$$

$$\mathbf{V}, \quad (8)$$

where $\mathbf{Q}$, $\mathbf{K}$, and $\mathbf{V}$ denote the query, key, and value matrices derived from the embedded patch embeddings.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Custom patch merging operations are employed between stages to progressively reduce temporal resolution and increase the receptive field. This hierarchical design enables efficient multi-scale modeling of long, non-stationary Φ-OTDR signals while preserving linear computational complexity.

Unlike the original 2-D Swin Transformer designed for image patches arranged on a spatial grid, the proposed model adapts the shifted-window attention mechanism to one-dimensional temporal sequences. Window shifts of $W/2$ samples are applied along the temporal axis to facilitate cross-window interaction. In our implementation, a patch length of $P = 8$ samples and an initial window size of $W = 128$ samples are used, with larger effective receptive fields obtained through hierarchical patch merging in deeper stages [4].

After the final stage, global average pooling produces a high-level representation $\mathbf{F}_{\text{Swin}} \in \mathbb{R}^{768}$ that encodes multi-scale temporal context across the entire Φ-OTDR trace.

D. Feature Fusion and Multi-Task Learning

The local and global representations extracted by the CNN and Swin Transformer branches are concatenated to form a fused feature vector:

$$\mathbf{F}_{\text{fusion}} = [\mathbf{F}_{\text{CNN}} \,\|\, \mathbf{F}_{\text{Swin}}], \tag{9}$$

where $\|$ denotes channel-wise concatenation.

The fused representation is passed through fully connected layers with nonlinear activation and dropout regularization to mitigate overfitting [26]. The network is trained in a multi-task learning setting, where both tasks share the fused representation and are jointly optimized. The classification head predicts one of six event categories, while the regression head estimates the normalized spatial location of the disturbance along the fiber.

The classification and localization outputs are given by

$$\hat{\mathbf{y}} = \text{softmax}(\mathbf{W}_c \mathbf{F}_{\text{fusion}} + \mathbf{b}_c), \tag{10}$$

$$\hat{r} = \mathbf{W}_r \mathbf{F}_{\text{fusion}} + \mathbf{b}_r, \tag{11}$$

where $\hat{\mathbf{y}}$ denotes the predicted event class probabilities and $\hat{r}$ denotes the estimated fault location.

E. Loss Function

Model optimization is performed using a weighted combination of classification and localization losses. Cross-entropy loss is applied for event classification, while mean squared error is used for fault localization. Equal weights are assigned to the two objectives to balance event recognition and spatial localization during training, preventing either task from dominating the optimization process.

To prevent background samples, which do not have a meaningful fault location, from adversely affecting localization learning, a masking strategy is adopted such that the regression loss is applied only to positive intrusion events. The overall loss function is defined as

$$\mathrm{L} = \lambda_{\text{cls}} \mathrm{L}_{\text{CE}} + \lambda_{\text{loc}} (m \cdot \mathrm{L}_{\text{MSE}}), \tag{12}$$

where $m \in \{0, 1\}$ denotes a binary mask indicating the presence of an intrusion event, and λ_{cls} and λ_{loc} are weighting coefficients. In this work, $\lambda_{\text{cls}} = \lambda_{\text{loc}} = 0.5$.

F. Training Strategy

All models are trained end-to-end using the Adam optimizer with a learning rate of 10^{-4} and weight decay of 10^{-4}. A ReduceLROnPlateau scheduler is employed to adapt the learning rate during training. Gradient clipping is applied to ensure numerical stability, and early stopping is used to prevent overfitting. Training is conducted for a maximum of 50 epochs with a batch size of 16.

G. Interpretability and Deployment Optimization

To enhance model interpretability, attention weights from all Swin Transformer layers and heads are aggregated and interpo- lated back to the original temporal resolution. These attention maps provide physically meaningful insights into the temporal regions that contribute most strongly to event detection and localization decisions[2], [18], [1]. For practical deployment, post-training dynamic INT8 quantization is applied, substantially reducing model size and inference latency while preserving classification and localization performance [7]. This combination of interpretability and efficiency enables real-time deployment of the proposed framework in operational Φ-OTDR monitoring systems[5], [3].

© 2026 IJIRCCE | Volume 14, Issue 2, February 2026 DOI: 10.15680/IJIRCCE.2026.1305001

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

IV. SIMULATION RESULTS

This section presents a comprehensive evaluation of the proposed CNN-Swin Transformer fusion for joint event classification and fault localization in Φ-OTDR signals. We first outline the experimental setup and evaluation metrics. We then report the performance of the proposed model in terms of classification accuracy, localization precision, and feature representation quality. Finally, we compare against two baselines (1-D CNN and CNN+Vanilla Transformer) to demonstrate the benefits of the hierarchical Swin architecture.

A. Dataset and Experimental Setup

Experiments were conducted on a publicly available Φ-OTDR dataset released by the BJTUSensor Research Group at Beijing Jiaotong University. The dataset contains 15,612 labeled samples spanning six classes, including background noise and five intrusion events: digging, knocking, shaking, watering, and walking. Each sample consists of a 12-channel Φ-OTDR trace with 10,000 temporal points.

The dataset is split into 12,529 training samples and 3,083 test samples, with class distributions preserved across both sets. Raw traces are directly used as model input after per-trace z-score normalization to mitigate amplitude variations caused by fiber attenuation and coupling differences.

During training, on-the-fly data augmentation is applied to improve robustness, including random temporal shifts, slight amplitude scaling, and additive Gaussian noise. For positive intrusion events, the ground-truth location labels are adjusted consistently with temporal shifts, while background samples are excluded from localization supervision. All models, including baselines, are trained and evaluated using identical preprocessing and data splits to ensure fair comparison.

B. Performance of the Proposed CNN-Swin Transformer

The CNN–Swin Transformer achieves 98.31% classification accuracy and a macro F1-score of 0.983. Figure 2 shows the confusion matrix, illustrating near-perfect separation across all six classes.

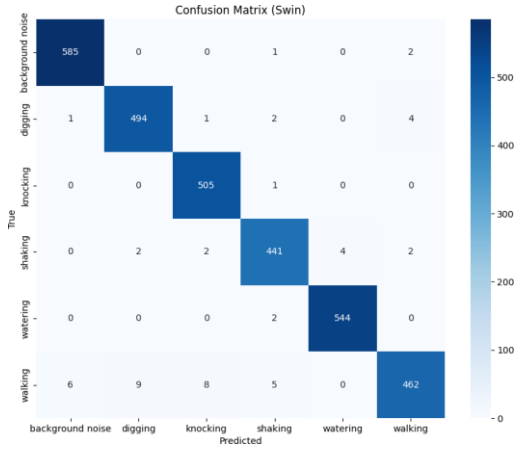


Fig. 2: Confusion matrix of the proposed CNN–Swin Transformer.

C. training and validation curves

The training and validation curves figure 3 show rapid convergence within 15–20 epochs, with losses decreasing monotoni- cally and accuracy/macro F1-score quickly reaching $>$ 0.98. The tight overlap between training and validation metrics indicates no overfitting and excellent generalization, enabled by domain-specific augmentations and multi-task learning. Training and validation curves (Fig. 3) confirm stable convergence and strong generalization.

© 2026 IJIRCCE | Volume 14, Issue 2, February 2026 DOI: 10.15680/IJIRCCE.2026.1305001

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

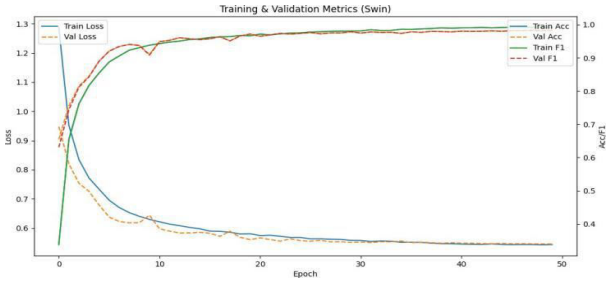
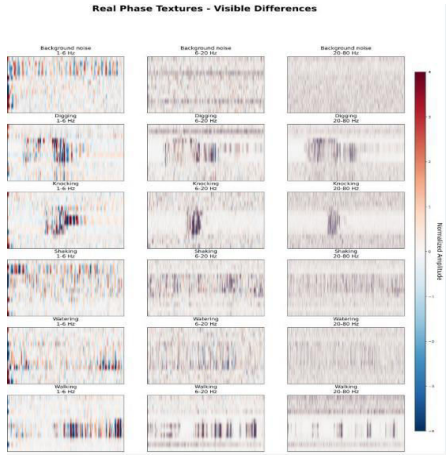


Fig. 3: Training and validation metrics (loss, accuracy, F1-score) for the proposed model.



D. Qualitative Visualization

Figure 4 illustrates representative real-phase textures of the six event classes across multiple frequency bands. The distinct multi-scale patterns observed across events motivate the need for hierarchical temporal modeling.

Fig. 4: Real-phase textures of the six event classes in low (1-6 Hz), medium (6-20 Hz), and high (20-80 Hz) frequency bands

E. 3D surface plots visualizations

Figure 5 presents representative Φ-OTDR traces for each of the six event classes, visualized as 3D surface plots with overlaid model predictions. The results demonstrate near-perfect classification (confidence scores > 0.99) and precise fault localization (predicted positions closely aligning with ground-truth peaks)

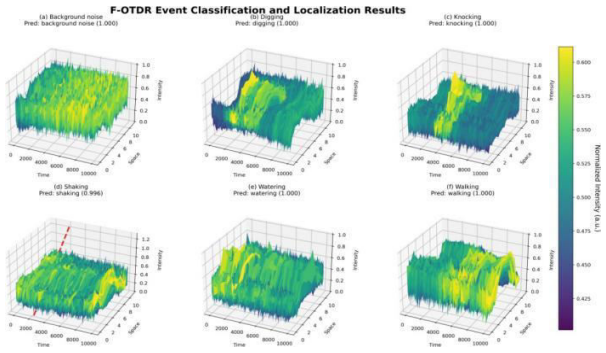


Fig. 5: Example inference results from the proposed CNN-Swin Transformer (one sample per class). Each subplot shows the predicted class with confidence score and localization (red dashed line), confirming robust classification and precise fault positioning.

© 2026 IJIRCCE | Volume 14, Issue 2, February 2026 DOI: 10.15680/IJIRCCE.2026.1305001

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

F. ROC Curves

The Receiver Operating Characteristic (ROC) curves shown in Fig. 6 further demonstrate the strong discriminative capability of the proposed CNN–Swin Transformer across all six event classes. The ROC curves closely follow the top-left corner of the plot, indicating a high true positive rate achieved at a low false positive rate. All classes attain area-under-the-curve (AUROC) values close to 1.0, reflecting excellent class separability and robust detection performance.

These results confirm the model's ability to reliably distinguish intrusion events from background noise with minimal false alarms and missed detections, which is critical for real-world Φ-OTDR monitoring and early-warning applications.

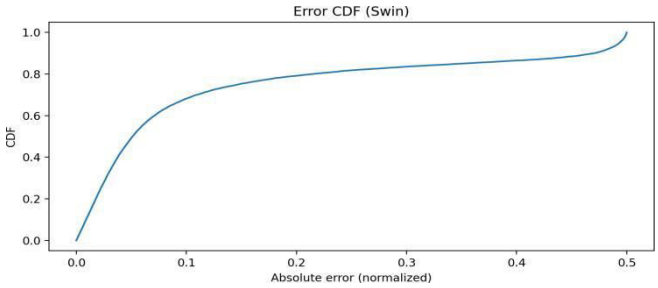


Fig. 6: ROC curves for the proposed model.

G. Cumulative distribution function CDF

The cumulative distribution function (CDF) of normalized absolute localization errors figure 7 illustrates the proposed model's superior precision, with approximately 70% of positive events localized within a normalized error of 0.1 (10 m) and 90% within 0.15 (15 m), achieving a mean absolute error (MAE) of 9.2 m. This tight distribution highlights the effectiveness of the hierarchical Swin Transformer in capturing long-range spatial dependencies for meter-level fault localization.

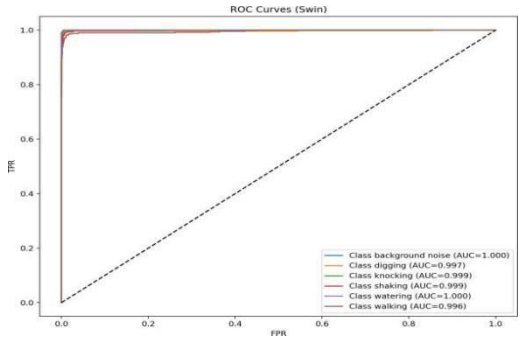


Fig. 7: Cumulative distribution function of localization errors (in meters).

H. t-SNE visualization

The t-SNE visualization of the fused CNN-Swin Transformer features (Fig. 8) reveals well-separated, compact clusters for each of the six event classes. Distinct groupings are evident for background noise (orange), digging (red), knocking (brown), shaking (green), watering (purple), and walking (blue), with minimal overlap between clusters. This clear separability in the 2D embedding space underscores the model's effective learning of discriminative multi-scale representations, contributing to the high classification accuracy (98.31%) and robust generalization across diverse, impulsive Φ-OTDR event signatures.

© 2026 IJIRCCE | Volume 14, Issue 2, February 2026 DOI: 10.15680/IJIRCCE.2026.1305001

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

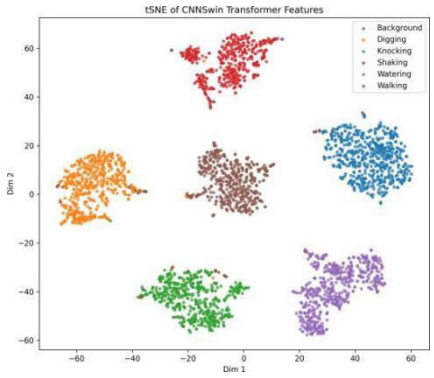
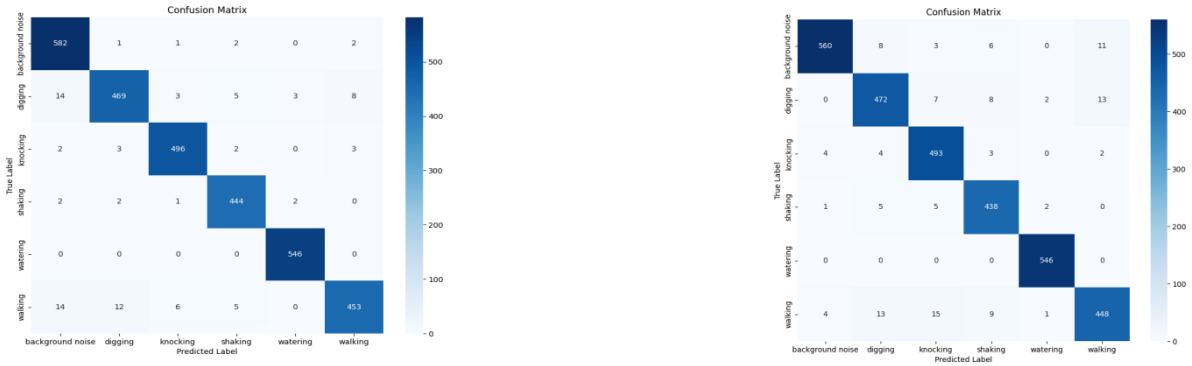


Fig. 8: t-SNE embedding of features extracted by the proposed CNN–Swin Transformer.

I. baseline confusion matrix

The confusion matrices for the baselines (Fig. 9a and Fig. 9b) exhibit more pronounced off-diagonal elements compared to the proposed CNN–Swin Transformer, indicating higher misclassification rates, particularly between similar classes such as knocking/walking and digging/watering. This increased confusion contributes to the baselines' lower accuracy (96.7% for 1-D CNN and 95.9% for Transformer) and macro F1-score, highlighting the proposed model's superior ability to discriminate subtle multi-scale patterns in Φ-OTDR traces.



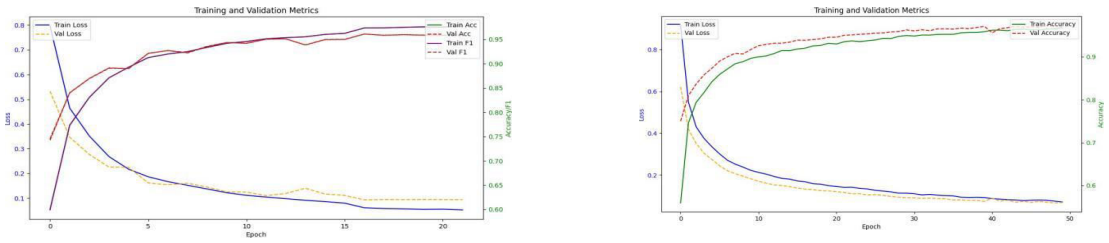
(a) 1-D CNN confusion

(b) CNN+Transformer confusion

Fig. 9: Confusion matrix for the baseline models.

J. baseline visualized curves

The training and validation curves (Fig. 10a and Fig. 10b) reveal that the baselines converge more slowly and reach lower peak accuracy and macro F1-score (0.96) compared to the proposed CNN–Swin Transformer (> 0.98), with wider train–validation gaps indicating inferior generalization on long Φ-OTDR sequences.



(a) CNN+Transformer training

(b) 1-D CNN training

Fig. 10: Training curves for the baselines.

© 2026 IJIRCCE | Volume 14, Issue 2, February 2026 DOI: 10.15680/IJIRCCE.2026.1305001

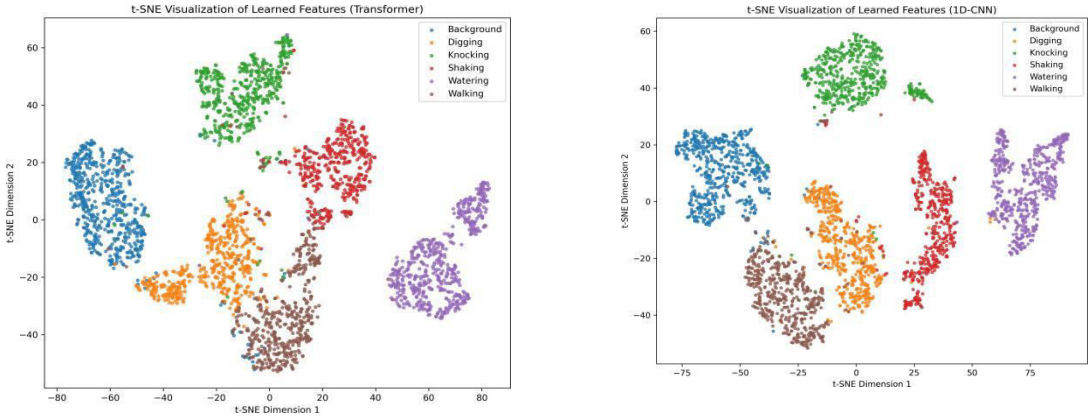
www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

K. t-SNE baseline visualization

The t-SNE visualizations (Fig. 11) show that the proposed CNN-Swin Transformer produces compact, well-separated clusters for all six classes with minimal overlap, unlike the more scattered and mixed clusters in the 1-D CNN and Transformer baselines. This superior feature separability explains the model's higher accuracy (98.31%) and macro F1-score (0.983).



(a) CNN+Transformer t-SNE (b) 1-D CNN t-SNE

Fig. 11: t-SNE for the baselines.

L. Ablation Study

To evaluate the contribution of the main architectural components in the proposed framework, an ablation study was conducted on the Phi-OTDR dataset. All model variants were trained and evaluated using the same data splits, preprocessing procedures, and optimization settings to ensure a fair comparison. The classification accuracy of different configurations is illustrated in Fig. 12, while detailed quantitative results are summarized in Table II.

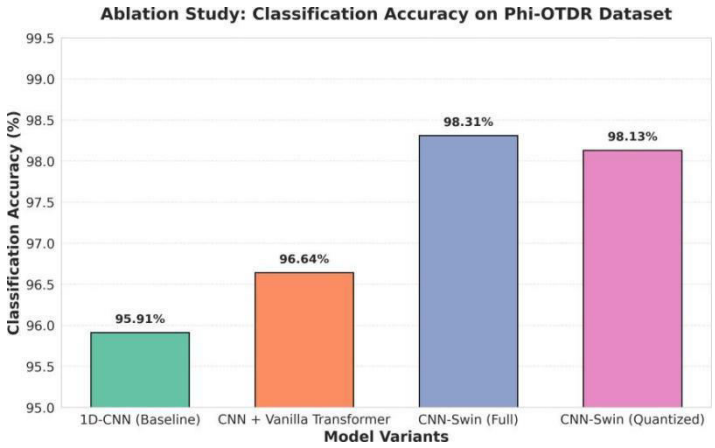


Fig. 12: Ablation study of classification accuracy for different model variants on the Phi-OTDR dataset.

We begin with a 1D-CNN baseline, which achieves a classification accuracy of 95.91%. This model effectively captures local transient patterns in Phi-OTDR signals but lacks an explicit mechanism for modeling long-range temporal dependencies. When a vanilla Transformer is integrated with the CNN feature extractor, the accuracy increases to 96.64%, indicating that global temporal modeling contributes positively to event classification.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Replacing the vanilla Transformer with the proposed CNN–Swin architecture results in a substantial performance improve- ment, achieving an accuracy of 98.31%, as shown in Fig. 12. This gain highlights the effectiveness of the hierarchical window- based attention mechanism in capturing both local and long-range temporal dependencies while maintaining computational efficiency. Compared with the vanilla Transformer, the Swin-based design provides a more suitable inductive bias for structured time-series representations commonly observed in distributed acoustic sensing.

To assess deployment feasibility, a post-training quantized version of the CNN–Swin model was also evaluated. Despite reduced numerical precision, the quantized model maintains a high accuracy of 98.13%, with only a marginal performance degradation of 0.18% compared to the full-precision model. This result demonstrates the robustness of the proposed architecture and its suitability for real-time or resource-constrained deployment scenarios.

TABLE II: Ablation study results of different model variants on the Phi-OTDR dataset

Model Variant	Accuracy (%)	Macro F1
1D-CNN (Baseline)	95.91	0.959
CNN + Vanilla Transformer	96.64	0.966
CNN–Swin (Proposed)	98.31	0.983
CNN–Swin (Quantized)	98.13	0.981

Overall, the ablation results confirm that each architectural enhancement contributes meaningfully to performance improve- ments. The combination of CNN-based local feature extraction with Swin Transformer attention yields a significant gain over both convolution-only and vanilla Transformer-based baselines, while quantization introduces negligible performance loss, supporting practical deployment in real-world Phi-OTDR monitoring systems.

VI. CONCLUSION AND FUTURE WORK

This paper presented an efficient and interpretable CNN–Swin Transformer framework for joint event classification and fault localization in long multi-channel Φ-OTDR signals. By integrating lightweight convolutional feature extraction with a customized 1-D Swin Transformer employing hierarchical shifted-window self-attention, the proposed architecture effectively captures both local transient patterns and long-range temporal dependencies while maintaining linear computational complexity.

Experimental evaluation on a public Φ-OTDR dataset demonstrated state-of-the-art performance, achieving 98.31% classification accuracy, a macro F1-score of 0.983, and a localization mean absolute error of 9.2 m. The model consistently outperformed conventional 1-D CNN and CNN+Vanilla Transformer baselines, highlighting the advantage of structured multi-scale attention for long, non-stationary DAS signals. Furthermore, post-training INT8 quantization reduced model size and inference latency substantially with negligible performance degradation, confirming its suitability for real-time deployment on resource-constrained edge devices.

Despite these promising results, several directions warrant further investigation. Future work will explore adaptive and streaming inference strategies to eliminate reliance on fixed-length inputs and enable continuous monitoring in practical systems. Robustness under varying signal-to-noise conditions and complex field environments should also be systematically evaluated to enhance generalization capability. In addition, incorporating deeper attention interpretability analysis and extending the framework to overlapping multi-event scenarios may further improve transparency and reliability. These developments would strengthen the scalability and real-world applicability of intelligent distributed acoustic sensing systems.

REFERENCES

[1] A. Borovykh, A. Soltanolkotabi, and G. G. Hinton. Conditional variational autoencoder for time series. *Proceedings of the 34th International Conference on Machine Learning*, 70:3831–3839, 2017.
[2] H. Chefer, S. Gur, and L. Wolf. Transformer interpretability beyond attention visualization. *CVPR*, 2021.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

[3] Y. Cheng, W. Chen, and J. Liu. Model compression for efficient neural networks: A survey. *IEEE Transactions on Neural Networks and Learning Systems*, 31(9):3123–3138, 2020.
[4] J. Dong, Z. Chen, and Z. Liu. Deep learning optimization for time-series data: A survey. *IEEE Transactions on Neural Networks and Learning Systems*, 31(5):1462–1473, 2020.
[5] S. Han, H. Mao, and W. Dally. Deep compression: Compressing deep neural networks. *ICLR*, 2016.
[6] S. Hochreiter and J. Schmidhuber. Long short-term memory. *Neural Computation*, 9(8):1735–1780, 1997.
[7] B. Jacob, S. Kligys, and B. Chen. Quantization and training of neural networks for efficient inference. *CVPR*, 2018.
[8] Y. Juan, X. Bao, and L. Chen. Distributed acoustic sensing for infrastructure monitoring: A review. *Sensors*, 22(6):2193, 2022.
[9] J. Kang, P. Han, Y. Chun, et al. Transformdas: Mapping ϕ-otdr signals to riemannian manifold for robust classification. *arXiv preprint arXiv:2502.02428*, 2025.
[10] X. Li, Z. Liu, and P. Zhang. Deep learning for distributed acoustic sensing: A survey. *Journal of Lightwave Technology*, 37(9):2205–2216, 2019.
[11] X. Li and Y. Zhou. Lstm-based das event recognition under low snr conditions. *IEEE Access*, 9:112345–112356, 2021.
[12] H. Liu and J. Zhao. Convolutional neural network based event recognition in distributed acoustic sensing. *IEEE Sensors Journal*, 21(9):10335–10343, 2021.
[13] Y. Liu and J. Chen. Hierarchical swin transformer for 1d signal modeling. *IEEE Signal Processing Letters*, 29:2257–2261, 2022.
[14] Y. Liu, Y. Lin, and Z. Zhang. Swin transformer for time-series data: A comprehensive review. *IEEE Transactions on Neural Networks and Learning Systems*, 33(10):3210–3221, 2022.
[15] Y. Liu, Z. Wang, and H. Zhang. Pipeline intrusion monitoring using phase-sensitive otdr. *Optics Express*, 31(4):5632–5648, 2023.
[16] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 10012–10022, 2021.
[17] J. Park, H. F. Martins, M. A. Soto, and L. The´venaz. Optical fiber distributed acoustic sensing using phase-sensitive otdr. *Optics Express*, 24(12):13415– 13425, 2016.
[18] J. Serra and D. Arcos. Explaining time series classifiers. *Data Mining and Knowledge Discovery*, 32:1–29, 2018.
[19] X. Shi, Z. Chen, H. Wang, D. Yeung, W. Wong, and W. Woo. Convolutional lstm network: A machine learning approach for precipitation nowcasting. *Proceedings of the 2015 Neural Information Processing Systems (NIPS)*, 2015.
[20] M. Straub, J. Reber, T. Saier, et al. Ml approaches for otdr diagnoses in passive optical networks—event detection and classification: ways for odn branch assignment. *Journal of Optical Communications and Networking*, 16(7):C43–C50, 2024.
[21] M. Tian, H. Dong, and K. Yu. An open dataset of ϕ-otdr events with two classification models as baselines. *Data in Brief*, 46:108824, 2023. Dataset available at https://github.com/BJTUSensor/Phi-OTDR$_d$ataset$_a$nd$_c$odes.
[22] A. Vaswani, N. Shazeer, N. Parmar, L. Uszkoreit, J. Kaiser, M. Polosukhin, and H. Wattenberg. Attention is all you need. *Proceedings of Neural Information Processing Systems (NIPS)*, 2017.
[23] J. Wang, Y. Li, and Q. Sun. Characteristics analysis of vibration events in -otdr systems. *Journal of Lightwave Technology*, 40(18):6215–6224, 2022.
[24] K. Wang and P. Lu. Multi-channel das event classification using deep convolutional networks. *Optics and Lasers in Engineering*, 152:106982, 2022.
[25] Z. Wang and H. Dong. Benchmarking deep learning models for das event detection. *Sensors*, 23(14):6412, 2023.
[26] R. Wightman. Pytorch image models (timm). https://github.com/huggingface/pytorch-image-models, 2023. Version 0.9.12 used in this work.
[27] H. Wu, T. Hu, Y. Liu, H. Zhou, J. Wang, and M. Long. Timesnet: Temporal 2d-variation modeling for general time series analysis. In *International Conference on Learning Representations (ICLR)*, 2023.
[28] M. Yan and O. Qiaofeng. Otdr event detection method based on improved 1d unet. *Instruments and Experimental Techniques*, 67:332–342, 2024.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

[29] H. Zhou, S. Zhang, J. Peng, S. Zhang, J. Li, H. Xiong, and W. Zhang. Informer: Beyond efficient transformer for long sequence time-series forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 11106–11115, 2021.

[30] T. Zhou, Z. Ma, Q. Wen, X. Wang, L. Sun, and R. Jin. FEDformer: Frequency enhanced decomposed transformer for long-term series forecasting. In *International Conference on Machine Learning (ICML)*, pages 27268–27286, 2022.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi cross ref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details