



HAL
open science

DASatFormer: A Hybrid Multibranch Transformer with Statistical Feature Integration for DAS-Based Pattern Recognitions

Michel Dione, Jerry Lonlac, H el ene Louis, Anthony Fleury, Stephane Lecoeuche

► **To cite this version:**

Michel Dione, Jerry Lonlac, H el ene Louis, Anthony Fleury, Stephane Lecoeuche. DASatFormer: A Hybrid Multibranch Transformer with Statistical Feature Integration for DAS-Based Pattern Recognitions. 2026. <hal-05579121>

HAL Id: hal-05579121

<https://hal.science/hal-05579121v1>

Preprint submitted on 20 May 2026

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destin ee au d ep ot et  a la diffusion de documents scientifiques de niveau recherche, publi es ou non,  emanant des  tablissements d'enseignement et de recherche fran ais ou  trangers, des laboratoires publics ou priv es.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License

DASStatFormer: A Hybrid Multibranch Transformer with Statistical Feature Integration for DAS-Based Pattern Recognitions

Michel Dione¹, Jerry Lonlac¹, H el ene Louis¹, Anthony Fleury¹, and St ephane Lecoecuche²

¹ IMT Nord Europe, Institut Mines-Telecom, Univ. Lille, Centre for Digital Systems
Lille, France

² IMT Mines Ales, Institut Mines-Telecom, Ales, France
`michel.dione@imt-nord-europe.fr`

Abstract. Distributed Acoustic Sensing (DAS) enables large-scale monitoring through optical fibers, but its high dimensionality and complex spatio-temporal patterns make event classification demanding. Existing deep learning approaches—CNNs, recurrent models, and Transformer variants—either fail to capture long-range dependencies or require processing raw DAS matrices at prohibitive cost. We propose *DASStatFormer*, a hybrid multibranch Transformer that combines compact multidomain statistical features with Gated Transformer Networks. Instead of raw signals, we extract 24 ANOVA-selected attributes per channel from the temporal, waveform, and spectral domains, reducing data size by orders of magnitude while preserving discriminative information. Each domain is processed via dedicated step-wise and channel-wise attention branches, fused by an adaptive gating mechanism. Experiments on the open ϕ -OTDR benchmark and a real-scenario DAS dataset show that DASStatFormer achieves up to 99.4% accuracy and near-perfect real-world performance, while using significantly fewer parameters and lower inference cost than models such as DASFormer and DeepViT. These results demonstrate its suitability for scalable, real-time DAS-based monitoring. We release our code at <https://github.com/MichelD-git/DASStatFormer>

Keywords: Distributed Acoustic Sensing · Multi-domain Features · Gated Transformer Network · Pattern Recognition.

1 Introduction

Distributed Acoustic Sensing (DAS) has become a powerful solution for large-scale monitoring of infrastructures, transportation networks, and sensitive perimeters [4]. By converting optical fibers into dense arrays of virtual sensors, DAS enables real-time detection and localization of diverse activities, even in challenging environmental conditions. Deep learning has markedly advanced DAS-based event recognition, with CNNs [11,13], hybrid CNN-BiLSTM models [5], and Transformer-based architectures such as DASFormer [10]. However, these models still face key limitations: they struggle to simultaneously capture fine-grained

temporal patterns and broader statistical trends, and processing raw spatio-temporal DAS matrices remains computationally demanding due to extremely high sampling rates leading to scalability issues in memory usage, training cost, and inference speed. To overcome these challenges, we introduce *DASStatFormer*, a hybrid Transformer-based framework that integrates multidomain statistical feature extraction with a multibranch Gated Transformer Network (GTN) architecture [12]. Instead of operating on raw DAS matrices, the method computes 24 interpretable attributes per channel across three complementary domains temporal, waveform, and spectral following principles from acoustic and vibration analysis. These compact descriptors preserve discriminative information while drastically reducing dimensionality, and are processed by three domain-specific Transformer branches with two-way gating mechanisms that balance step-wise and channel-wise dependencies. This design enables efficient modeling of both local and global structures in DAS data at low computational cost. The proposed framework is evaluated on two datasets: (1) the open ϕ -OTDR dataset [8], and (2) a real-scenario DAS dataset introduced by Tomasov *et al.* [17], which we restructure into a windowed format compatible with learning-based models. Experiments show that *DASStatFormer* achieves accuracy comparable to state-of-the-art approaches while requiring significantly fewer parameters and reduced inference time. The main contributions of this work are:

1. A hybrid Transformer architecture integrating multidomain statistical features with a multibranch GTN-based attention mechanism for DAS event recognition.
2. A compact and interpretable 24-feature representation extracted from temporal, waveform, and spectral domains, offering strong discriminability with minimal data dimensionality.
3. A unified preprocessing pipeline restructuring real DAS recordings into a standardized windowed format suitable for Transformer models.
4. Extensive validation on laboratory and real-world datasets, demonstrating strong accuracy, efficiency, and scalability for real-time DAS monitoring.

The remainder of this paper is organized as follows: Section 2 reviews prior work on DAS event recognition, and Section 3 summarizes the principles of DAS. Section 4 presents the proposed *DASStatFormer* framework, Section 5 reports experimental evaluations, and Section 6 concludes the study.

2 Related Work

Distributed Acoustic Sensing (DAS) has become a key technology for applications such as pipeline monitoring, geotechnical analysis, and infrastructure surveillance [4]. Its growing use for perimeter intrusion detection has stimulated rapid progress in signal processing and machine learning. Early methods based on handcrafted features and classical classifiers provided useful insights but proved insufficient to handle the high dimensionality, noise levels, and strong spatio-temporal correlations that characterize DAS signals. To overcome these

limitations, researchers explored both traditional machine learning models—such as Support Vector Machines (SVM) [3], Random Forests, and gradient-boosted trees and more advanced deep learning architectures. Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), including LSTM variants, have become widely used thanks to their ability to learn discriminative features directly from raw or preprocessed DAS signals [6]. For instance, Cao *et al.* [8] employed a 2D CNN to classify events in a laboratory DAS setup, successfully identifying activities such as digging and walking. Similarly, Wu *et al.* [5] proposed a 1D CNN-BiLSTM hybrid model that jointly captures spatial and temporal dependencies, improving classification accuracy in intrusion scenarios. Other studies reformulated DAS event recognition as a visual classification task by converting spatio-temporal signals into waterfall diagrams [7] or spectrograms [21], enabling the use of computer vision models. However, this transformation often discards subtle signal details, making such approaches suitable only when the temporal and spatial characteristics of the data are relatively uniform. In the context of railway intrusion monitoring with ϕ -OTDR, Meng *et al.* [13] introduced an XGBoost-based identification method, showing the potential of gradient-boosted trees for DAS event recognition. Overall, CNN-based models have become a dominant strategy for processing ϕ -OTDR data, with 1D and 2D variants widely adopted. Despite their success, CNN architectures present intrinsic limitations for DAS analysis: they focus mainly on local patterns, require fixed-size inputs, and struggle to capture global contextual information [12]. Because DAS recordings are extremely high-dimensional and contain long-range spatio-temporal dependencies along the fiber, CNNs are not ideally suited to fully model such complex structures, which ultimately restricts their effectiveness. These constraints have motivated the investigation of alternative architectures capable of modeling long-range interactions.

Transformer-based models, originally introduced for natural language processing [9], have recently been adapted to time-series and sensing tasks thanks to their powerful self-attention mechanism. In the DAS domain, this has inspired works such as DASFormer [10], which directly model spatio-temporal dependencies; and DeepViT [14], building on the success of Vision Transformers (ViT) [20] by applying them to waterfall-diagram representations for intrusion detection. Dione *et al.* [19] introduced *DASViT1D*, a Transformer-based model leveraging time–frequency representations for DAS intrusion recognition. However, existing Transformer-based DAS approaches still face notable challenges: most operate on raw high-dimensional spatio-temporal matrices or on visually transformed images, both of which demand substantial computational resources and may compromise signal fidelity. Furthermore, few studies explicitly exploit the multidomain nature of DAS signals combining temporal, waveform, and spectral information within a unified learning framework. Addressing these gaps, our work introduces *DASStatFormer*, a hybrid multibranch Transformer architecture that integrates domain-specific statistical features with an efficient attention-based modeling strategy. This design achieves a balance between dimensionality reduction, computational scalability, and high discriminative performance, paving

the way for more interpretable and resource-efficient DAS-based monitoring systems.

3 Distributed Acoustic Sensing (DAS) System

Distributed Acoustic Sensing (DAS) exploits Rayleigh backscattering to detect vibrations along optical fibers over long distances [16]. Short optical pulses are injected into the fiber, and external acoustic disturbances induce small strain variations that modulate the phase and intensity of the backscattered light. These variations are sampled at discrete locations, referred to as *channels* or virtual sensors, whose spacing is determined by the system spatial resolution [11]. The DAS interrogator continuously records these measurements, producing dense spatio-temporal strain signals along the monitored fiber.

A DAS system is mainly characterized by its sampling frequency f_s , spatial resolution dx , fiber length fl , and number of channels n . Figure 1 illustrates the general operating principle of a DAS system. A DAS acquisition produces a

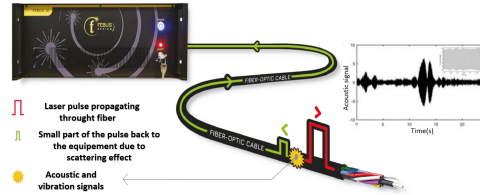


Fig. 1: Distributed Acoustic Sensing Principle [15]

high-dimensional spatio-temporal signal that can be naturally represented as a matrix of strain measurements indexed by time and space. Each DAS recording consists of $m = f_s \times T$ temporal samples acquired along n spatial channels distributed along the optical fiber, where f_s denotes the sampling frequency and T the recording duration. At each time instant t_i , the DAS system measures a strain value str_i^j at channel Ch_j , corresponding to a fixed spatial position along the fiber. This organization results in a dense spatio-temporal representation in which rows correspond to time instants and columns to spatial sensing channels. Such high-resolution DAS data capture the dynamic response of the monitored environment along the entire fiber and form the basis for subsequent feature extraction, signal representation, and event classification.

4 Methodology

DAS recordings are extremely high-dimensional due to the joint effects of high sampling frequency, long acquisition duration, and dense spatial resolution. This results in large spatio-temporal matrices that are costly to process and difficult to scale. The proposed *DASatFormer* framework addresses this challenge through

a two-stage strategy. First, DAS segments are transformed into compact multidomain statistical descriptors that capture the essential temporal, waveform, and spectral characteristics while substantially reducing data volume. Second, the resulting features are processed by a multibranch Transformer architecture specifically designed to exploit cross-domain complementarity and to model both local and global dependencies in DAS signals despite the reduced dimensionality. The following subsections describe the feature extraction process and the *DAsStatFormer* architecture.

4.1 Feature Extraction

Feature extraction aims to convert raw high-dimensional DAS traces into compact, informative representations suitable for learning-based analysis. Rather than processing the full spatio-temporal matrices, we compute 24 statistical attributes distributed across three complementary domains: *time*, *waveform*, and *spectral*. Time-domain descriptors characterize the global amplitude distribution and signal dynamics; waveform-based features capture shape-related and impulsive behaviors; and spectral attributes describe frequency-dependent patterns crucial for distinguishing continuous activities from short transient events. The selected 24 attributes originate from a larger pool of 60 candidate features as proposed by C. Hibert et al. [2]. To ensure objective relevance, we applied an ANOVA-based feature selection, retaining only the descriptors exhibiting high inter-class variance and low intra-class variability across all event categories. This statistically grounded selection guarantees that each feature contributes meaningful discriminative information. By combining temporal, morphological, and spectral cues, the resulting multidomain representation achieves a strong balance between expressiveness and dimensionality reduction.

This compact feature space preserves the essential properties of DAS signals while enabling efficient downstream processing within the proposed Transformer-based framework [18].

Let the raw input signal be denoted as $x = \{x_1, x_2, \dots, x_n\}$ of length n . The complete list of extracted attributes is provided in Table 1.

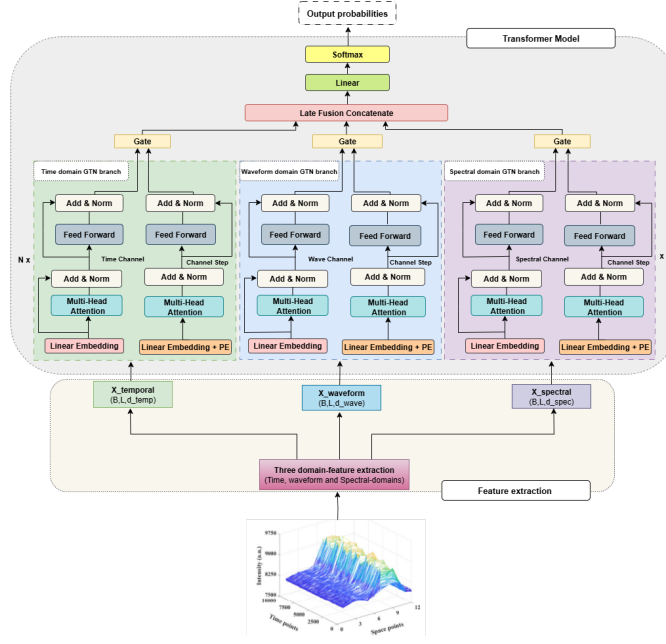
The waveform- and spectrum-based attributes are computed from the raw DAS segment $x(t)$, its Hilbert envelope $e(t)$, and its discrete Fourier transform (DFT) $X(f)$. The main notations are:

- $e(t) = |\mathcal{H}\{x(t)\}|$: Hilbert envelope of the signal.
- $X(f) = \mathcal{F}\{x(t)\}$: discrete Fourier transform.
- μ, σ^2, σ : mean, variance, and standard deviation.
- Skewness: $\frac{1}{n} \sum \left(\frac{x_i - \mu}{\sigma}\right)^3$.
- Kurtosis: $\frac{1}{n} \sum \left(\frac{x_i - \mu}{\sigma}\right)^4$.
- RMS: $\sqrt{\mu^2 + \sigma^2}$.
- Crest factor: $\frac{\max(x)}{\text{RMS}}$.
- Spectral entropy: $H = -\sum p_i \log_2(p_i + \epsilon)$, with $p_i = \frac{|X(f_i)|}{\sum_j |X(f_j)|}$.

Table 1: Selected statistical attributes extracted from DAS signals.

Domain	Attribute	Definition
Time-domain	x_{\max}	Maximum amplitude of the signal
	Peak-to-peak	$x_{\max} - x_{\min}$
	Std. deviation	$\sigma = \sqrt{\frac{1}{n} \sum (x_i - \mu)^2}$
	Skewness	$\frac{1}{n} \sum \left(\frac{x_i - \mu}{\sigma} \right)^3$
	Kurtosis	$\frac{1}{n} \sum \left(\frac{x_i - \mu}{\sigma} \right)^4$
	Zero-crossing rate	Mean rate of sign changes in $x(t)$
	Envelope ratio	$\frac{\max(e(t))}{\text{mean}(e(t))}$, with $e(t) = \mathcal{H}\{x(t)\} $
Waveform-domain	Waveform factor	$\frac{\text{RMS}}{\text{mean}(x_i)}$
	Impulse factor	$\frac{x_{\max}}{\text{mean}(x_i)}$
	Crest factor	$\frac{x_{\max}}{\text{RMS}}$
	Kurtosis factor	$\frac{\frac{1}{n} \sum x_i^4}{\text{RMS}^4}$
	Envelope skewness	Skewness of the Hilbert envelope $e(t)$
	Envelope kurtosis	Kurtosis of the Hilbert envelope $e(t)$
Spectral-domain	Mean spectrum	$\mu_f = \frac{1}{M} \sum X(f_k) $
	Max spectrum	$\max_k X(f_k) $
	Spectral entropy	$H = - \sum p_k \log_2(p_k + \epsilon)$, $p_k = \frac{ X(f_k) }{\sum_j X(f_j) }$
	Spectral centroid	$C = \frac{\sum_k f_k X(f_k) }{\sum_k X(f_k) }$
	Spectral bandwidth	$BW = \sqrt{\frac{\sum_k (f_k - C)^2 X(f_k) }{\sum_k X(f_k) }}$
	Spectral skewness	Third standardized moment of $ X(f_k) $ around C
	Spectral kurtosis	Fourth standardized moment of $ X(f_k) $ around C
	Spectral flatness	$\exp(\text{mean}(\log X(f_k))) / \text{mean}(X(f_k))$
	Spectral slope	Linear regression slope of $ X(f_k) $ as a function of f_k
	Rolloff 85%	Frequency $f_{0.85}$ such that $\sum_{f_k \leq f_{0.85}} X(f_k) = 0.85 \sum_k X(f_k) $
Rolloff 95%	Frequency $f_{0.95}$ such that $\sum_{f_k \leq f_{0.95}} X(f_k) = 0.95 \sum_k X(f_k) $	

A key motivation for using statistical descriptors is the drastic reduction of the temporal dimensionality of DAS recordings without sacrificing event discriminability. High DAS sampling rates generate extremely large spatio-temporal matrices, even for short signals. For example, a raw segment of size 10000×12 can be compactly reduced to a 12×24 representation by replacing each channel with 24 informative attributes, leading to substantial savings in computation and memory. Each statistic is computed independently per DAS channel, yielding ordered spatial feature sequences along the fiber. Although spatial in nature, these sequences can be modeled analogously to temporal signals, making them well suited for Transformer-based architectures. Based on this representation, the next subsection introduces *DASStatFormer*, a multibranch Transformer designed for efficient and accurate DAS event recognition.


 Fig. 2: Overview of the proposed *DAsStatFormer* framework.

4.2 DAsStatFormer model

As described in the previous subsection, the first stage of our framework converts raw distributed acoustic sensing (DAS) traces into multidomain statistical descriptors spanning the temporal, waveform, and spectral domains. These compact yet expressive representations serve as inputs to the attention-based model presented below. Figure 2 provides an overview of the proposed architecture: three parallel branches process feature matrices extracted from the time, waveform, and spectral domains, and their outputs are adaptively fused for final classification.

Input representation Let $d \in \{\text{time, waveform, spectral}\}$ denote one of the three feature domains. The multidomain descriptors for domain d are represented as a third-order tensor $X^{(d)} \in \mathbb{R}^{B \times L \times d_{\text{attr}}^{(d)}}$, where B is the mini-batch size, L is the number of sensing channels along the DAS fibre (interpreted as the sequence length), and $d_{\text{attr}}^{(d)}$ is the number of statistical attributes extracted per channel. We denote by $x_{b,l}^{(d)} \in \mathbb{R}^{d_{\text{attr}}^{(d)}}$ the feature vector corresponding to batch index b and channel l in domain d .

Step-wise encoding The first path within each branch models spatial (inter-channel) correlations along the sensing fibre. Each attribute vector $x_{b,l}^{(d)}$ is projected into a d_{model} -dimensional embedding through a learnable linear transformation:

$$\mathbf{s}_{b,l}^{(d)} = W_{\text{step}}^{(d)} x_{b,l}^{(d)} + b_{\text{step}}^{(d)}, \quad W_{\text{step}}^{(d)} \in \mathbb{R}^{d_{\text{attr}}^{(d)} \times d_{\text{model}}}, \quad (1)$$

where $W_{\text{step}}^{(d)}$ and $b_{\text{step}}^{(d)}$ are trainable parameters. An optional sinusoidal positional encoding $\text{PE} \in \mathbb{R}^{L \times d_{\text{model}}}$ is added to preserve the spatial ordering of channels. The resulting sequence $S^{(d)} = [\mathbf{s}_{b,1}^{(d)}, \dots, \mathbf{s}_{b,L}^{(d)}] \in \mathbb{R}^{B \times L \times d_{\text{model}}}$ is processed by N stacked Transformer encoder layers, each composed of multi-head self-attention (MHA) and a position-wise feed-forward network (FFN) with residual connections and layer normalization.

- **Multi-head self-attention** Given queries Q , keys K , and values V (all equal to $S^{(d)}$), the i -th attention head computes:

$$\mathbf{h}_i = f(W_i^{(q)}Q, W_i^{(k)}K, W_i^{(v)}V), \quad (2)$$

where $W_i^{(q)}, W_i^{(k)}, W_i^{(v)} \in \mathbb{R}^{(d_{\text{model}}/h) \times d_{\text{model}}}$ are learnable projection matrices, h is the number of heads, and f denotes scaled dot-product attention. The outputs from all heads are concatenated and linearly projected:

$$\text{MHA}(S^{(d)}) = W^{(o)}[\mathbf{h}_1; \dots; \mathbf{h}_h], \quad W^{(o)} \in \mathbb{R}^{d_{\text{model}} \times (h d_{\text{model}}/h)}. \quad (3)$$

- **Position-wise feed-forward network** Each position vector $\mathbf{s} \in \mathbb{R}^{d_{\text{model}}}$ is independently transformed by an FFN:

$$\text{FFN}(\mathbf{s}) = \sigma(\mathbf{s}W_1 + b_1)W_2 + b_2, \quad (4)$$

where $W_1 \in \mathbb{R}^{d_{\text{model}} \times d_{\text{hidden}}}$, $W_2 \in \mathbb{R}^{d_{\text{hidden}} \times d_{\text{model}}}$, and σ is a non-linear activation (ReLU in our implementation). Residual connections and layer normalization follow both the attention and FFN sublayers. After N encoder layers, the sequence $S^{(d)}$ is flattened along the channel dimension to yield $\mathbf{s}_{\text{flat}}^{(d)} \in \mathbb{R}^{B \times (L d_{\text{model}})}$.

Channel-wise encoding To capture dependencies among statistical attributes, a second path transposes the input tensor so that attributes form the sequence dimension: $\tilde{X}^{(d)} \in \mathbb{R}^{B \times d_{\text{attr}}^{(d)} \times L}$. Each slice $\tilde{x}_{b,j}^{(d)} \in \mathbb{R}^L$ (with j indexing attributes) is embedded via:

$$\mathbf{c}_{b,j}^{(d)} = W_{\text{chan}}^{(d)} \tilde{x}_{b,j}^{(d)} + b_{\text{chan}}^{(d)}, \quad W_{\text{chan}}^{(d)} \in \mathbb{R}^{L \times d_{\text{model}}}. \quad (5)$$

The resulting sequence $C^{(d)} \in \mathbb{R}^{B \times d_{\text{attr}}^{(d)} \times d_{\text{model}}}$ is passed through the same stack of N Transformer encoders to capture inter-attribute dependencies. It is then flattened to $\mathbf{c}_{\text{flat}}^{(d)} \in \mathbb{R}^{B \times (d_{\text{attr}}^{(d)} d_{\text{model}})}$.

Intra-branch gating and fusion To adaptively balance temporal and attribute-wise information, each branch employs a soft gating mechanism inspired by Gated Transformer Networks. The flattened vectors $\mathbf{s}_{\text{flat}}^{(d)}$ and $\mathbf{c}_{\text{flat}}^{(d)}$ are concatenated and fed to a two-unit softmax gate:

$$\mathbf{g}^{(d)} = \text{softmax}\left(W_{\text{gate}}^{(d)}[\mathbf{s}_{\text{flat}}^{(d)}; \mathbf{c}_{\text{flat}}^{(d)}] + b_{\text{gate}}^{(d)}\right), \quad (6)$$

where $W_{\text{gate}}^{(d)} \in \mathbb{R}^{2 \times (L d_{\text{model}} + d_{\text{attr}}^{(d)} d_{\text{model}})}$. The fused representation for domain d :

$$\text{fused}^{(d)} = [g_1^{(d)} \cdot \mathbf{s}_{\text{flat}}^{(d)}; g_2^{(d)} \cdot \mathbf{c}_{\text{flat}}^{(d)}]. \quad (7)$$

Late fusion and classification The outputs of the three domain-specific branches are concatenated to form a global representation:

$$\text{fused}_{\text{all}} = [\text{fused}^{(\text{time})}; \text{fused}^{(\text{waveform})}; \text{fused}^{(\text{spectral})}] \in \mathbb{R}^{B \times D_{\text{fuse}}}, \quad (8)$$

where D_{fuse} is the total dimensionality of the concatenated vectors. A final linear projection followed by a softmax layer produces the posterior probabilities over event classes:

$$\mathbf{y} = \text{softmax}(W_{\text{out}} \text{fused}_{\text{all}} + b_{\text{out}}), \quad W_{\text{out}} \in \mathbb{R}^{D_{\text{fuse}} \times d_{\text{out}}}. \quad (9)$$

The proposed *DAsStatFormer* architecture combines self-attention ability to model long-range dependencies with interpretable, domain-specific statistical features. The step-wise path captures inter-channel (spatial-temporal) dynamics, while the channel-wise path encodes correlations among statistical attributes. The gating mechanism adaptively re-weights these complementary representations, and the final multidomain fusion leverages cross-domain complementarity between temporal, waveform, and spectral cues—yielding a compact, discriminative, and physically interpretable representation for DAS event classification.

5 Experiments and Results

5.1 Baseline datasets

To evaluate the performance and generalization capability of the proposed *DAsStatFormer* framework, experiments were conducted on two different DAS datasets covering both laboratory and real-world scenarios. Each dataset contains multiple event types representative of typical perimeter monitoring conditions. All recordings were acquired using a phase-sensitive optical time-domain reflectometer (ϕ -OTDR) system. The diversity in data acquisition setups, noise conditions, and event types ensures a robust validation of the proposed approach.

Laboratory ϕ -OTDR Dataset This dataset originates from a controlled laboratory experiment provided and described by Cao et al. in [8], where a ϕ -OTDR system was employed to monitor a buried optical fiber subjected to various mechanical perturbations. The experimental setup simulates a perimeter intrusion monitoring scenario under repeatable and noise-controlled conditions, as showed in Fig 3-1. Six distinct types of events were generated: *background*, *digging*, *knocking*, *watering*, *shaking*, and *walking*. Each event was performed multiple times at different spatial positions along the fiber to ensure sufficient variability and robustness of the dataset. Table 3 summarizes the number of samples per class, along with their train/test split.

This dataset provides a reliable benchmark for DAS event classification, with each sample represented as a spatio-temporal segment of 10000 time samples across 12 adjacent channels. This configuration captures both temporal and spatial dynamics of vibration events, while the controlled experimental setup enables reproducible and consistent comparison of feature extraction methods and learning architectures.

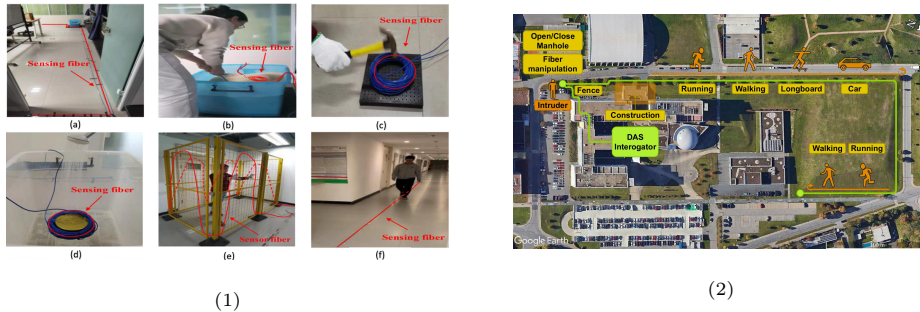


Fig. 3: Overview of the two experimental configurations: **(1)** Laboratory ϕ -OTDR setup illustrating six representative events (background, digging, knocking, watering, shaking, walking), with the sensing fiber highlighted in red [8]. **(2)** Aerial view of the campus DAS deployment, where the buried fiber (green line) monitors activities such as walking, running, longboarding, driving, fence climbing, and fiber manipulation [17].

Table 2: ϕ -OTDR system config [8].

Parameter	Value
Laser wavelength	1550 nm
Pulse width	50 ns
Pulse repetition rate	5 kHz
Sampling frequency (f_s)	100 MHz
Spatial resolution (dx)	5 m
Gauge length (gl)	10 m
Fiber length (fl)	1.2 km
Number of channels (n)	240

Table 3: Class distribution of the six events in the ϕ -OTDR laboratory dataset.

Event type	Number	Label	Train	Test
Background	3094	0	2505	589
Digging	2512	1	2018	494
Knocking	2530	2	2025	505
Watering	2298	3	1853	445
Shaking	2728	4	2183	545
Walking	2450	5	1969	481
Total	15612	-	12553	3059

Real-world scenario dataset The second dataset is derived from a real-world DAS deployment introduced by Tomasov *et al.* [17]. It consists of DAS recordings acquired along an underground optical fiber deployed on a university campus (Fig. 3-2), enabling the monitoring of diverse human and vehicular activities such as walking, running, longboarding, car traffic, as well as security-related events including fence climbing, manhole opening, and fiber manipulation.

Each raw DAS recording is provided as a two-dimensional matrix of size $T \times C$ (time samples \times channels), together with an event bitmap of size $N_w \times C_b$ indicating the temporal and spatial occurrence of events. To ensure compatibility with the laboratory dataset and a unified learning pipeline, the raw data were transformed into a standardized window-based representation. First, the bitmap channels were aligned to the DAS data using nearest-neighbor interpolation. The recordings were then segmented into N_w temporal windows, and each window was assigned an event label based on the corresponding bitmap entries, with non-event windows labeled as background. After this transformation, the dataset is reorganized into uniformly structured spatio-temporal windows of size (2053, 1700), identical to those of the laboratory dataset, allowing seamless integration into the proposed *DAS*StatFormer framework. In total, nine event classes

are obtained. Table 4 reports the resulting class distribution, which remains relatively balanced while reflecting realistic field variability.

Table 4: Label distribution in the transformed real-scenario DAS dataset.

Event type	Number of windows	Label	Train	Test
Car	1094	0	875	219
Construction	494	1	395	99
Fence interaction	571	2	457	114
Longboard	682	3	546	136
Manipulation	529	4	423	106
Open/Close	128	5	102	26
Regular activity	738	6	590	148
Running	1000	7	800	200
Walking	1455	8	1164	291
Total	6691	-	5352	1339

The final dataset structure thus consists of nine labeled categories, each stored in separate directories, allowing for straightforward integration into the preprocessing and feature extraction pipelines of the proposed *DASStatFormer*.

5.2 Results and Discussion

This section presents the experimental results obtained by the proposed *DASStatFormer* on both the open ϕ -OTDR laboratory dataset [8] and the real-scenario dataset from Tomasov *et al.* [17]. The performance is compared with representative approaches from the literature, including classical machine learning, CNN-based, hybrid, and Transformer-based models. The model configuration and the training/evaluation protocol are summarized in Tables 5 and 6, respectively.

Table 5: DASStatFormer configuration.

Component	Setting
Branches	3 (time,waveform,spectral)
Input window	$x \in \mathbb{R}^{B \times L \times F}$, $F=24$
Transformer encoders	$N=8$ blocks per path
Model width	$d_{\text{model}}=256$
FFN hidden dim	$d_{\text{hidden}}=128$
Multi-head attention	$h=8, q=8, v=8$
Dropout	0.2 (MHA/FFN path)
Mask	Enabled (to encoder)
Classifier	Linear \rightarrow Softmax

Table 6: Training protocol settings.

Item	Setting
Loss	Cross-Entropy
Optimizer / LR	Adam, LR= 1×10^{-4}
Batch size / Epochs	32 / 100
Gradient clip	$\ g\ _2 \leq 1.0$
Validation split	10% (random)
Test interval	Every 5 epochs
Seed	30
Device	CUDA:0
PE / Mask	PE: True; Mask: True

Figure 4 shows the confusion matrices of *DASStatFormer* on both datasets. The left matrix corresponds to the laboratory dataset with six event classes (background, digging, knocking, watering, shaking, walking), whereas the right matrix represents the real-scenario dataset containing nine classes (vehicle and human activities recorded on a university campus). In both cases, *DASStatFormer* demonstrates strong discriminative ability with very few misclassifications.

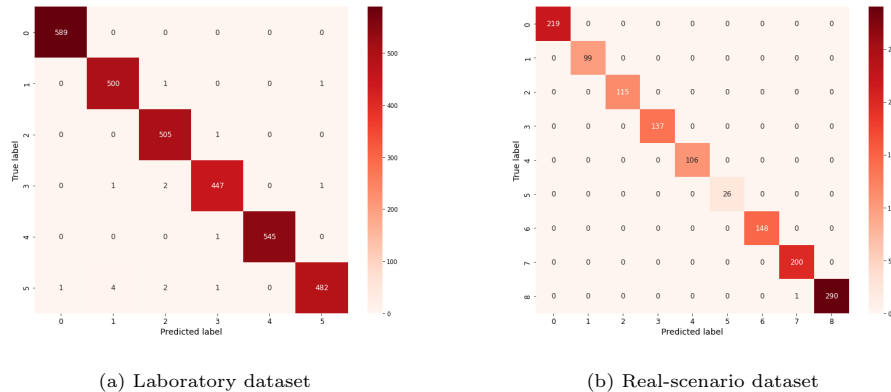


Fig. 4: Confusion matrices of *DASTatFormer* on (a) the laboratory dataset and (b) the real-scenario test sets.

Tables 7 and 8 present a comparative evaluation of *DASTatFormer* against existing models on both datasets. For clarity, the tables are displayed side-by-side to highlight performance differences between laboratory and field conditions.

Table 7: Performance comparison on the open ϕ -OTDR laboratory dataset.

Model	Accuracy \uparrow	NAR \downarrow
SVM [8]	82.6	0.184
2D CNN [8]	94.0	0.0124
CNN-BiLSTM [5]	97.91	0.0042
XGBoost [13]	94.26	0.0074
DeepViT (image-based) [14]	93.10	0.0687
DASFormer [10]	99.60	0.0000
DASTatFormer (ours)	99.48	0.0000

As shown in Table 7, *DASTatFormer* achieves an accuracy of **99.48%** on the open ϕ -OTDR laboratory dataset, closely matching the performance of DASFormer [10] (99.6%), with a nuisance alarm rate (NAR) of 0.00. While the difference in accuracy (0.12%) is negligible, *DASTatFormer* requires lower computational resources. The confusion matrices for the two datasets obtained with *DASTatFormer* are shown in Fig. 4. Unlike DASFormer’s complex dual-tower architecture (over 6 million parameters), *DASTatFormer* leverages compact multidomain statistical embeddings, reducing training time, GPU memory consumption (115 MB allocated), and inference cost by approximately an order of magnitude. These results confirm that *DASTatFormer* maintains high detection precision while achieving higher computational efficiency, making it well suited for real-time DAS event recognition.

Table 8: Performance comparison and inference time on real-field DAS data.

Model	Acc.↑	Prec.↑	Rec.↑	F1↑	Inf. Time↓	NAR ↓
CNN (Tomasov et al., 2025) [17]	91.4	85.7	88.2	86.8	–	–
DASViT1D [19]	93.5	91.93	90.80	91.29	13	0.005
DStatFormer (ours)	99.93	99.94	99.97	99.96	4.13	0.000

Ablation Study on Feature Domains To assess the contribution of each feature domain, an ablation study was conducted on the ϕ -OTDR laboratory dataset. Four configurations were evaluated: using only time-domain features, only spectral-domain features, only waveform-domain features, and finally the full multidomain combination as employed in the proposed *DStatFormer* framework. The goal of this analysis is to quantify how each domain contributes to the classification performance and to verify the complementarity between them.

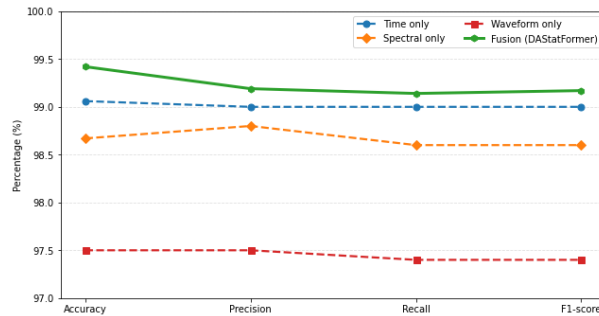


Fig. 5: Performance comparison of **DStatFormer** across three individual domain features (Time, Spectral, and Waveform) and their fusion.

As shown in Fig. 5, the multidomain fusion significantly improves both accuracy and generalization compared to single-domain configurations. While the time-domain features individually achieve high accuracy (99.1%), their combination with spectral and waveform attributes yields the best overall performance (99.48%) with zero false alarms ($\text{NAR} = 0.00$) and a very low false negative rate ($\text{FNR} = 0.0004$). Interestingly, the multidomain configuration also results in a reduced training time (84.37 min), while maintaining a low inference cost (0.94 ms/sample). These results confirm that the three feature domains provide complementary information: time-domain descriptors capture global signal dynamics, waveform features emphasize impulsive patterns, and spectral features encode frequency-dependent behaviors. Their fusion thus enables a more complete and discriminative representation of DAS events.

On the real-field DAS dataset, *DStatFormer* achieves state-of-the-art performance with 99.93% accuracy and a macro F1-score of 99.96%, while producing virtually no false alarms ($\text{NAR} = 0.00$). This represents a substantial

improvement over the CNN baseline of Tomasov *et al.* [17] (91.4% accuracy, 86.8% F1) and the Transformer-based DASViT1D [19] (93.5% accuracy, 91.29% F1). In addition to its superior accuracy, *DASatFormer* is significantly more efficient, requiring only 4.13 ms per sample for inference, compared to 13 ms for DASViT1D. These results highlight the effectiveness of multidomain statistical features combined with gated multibranch attention, enabling robust and scalable DAS event recognition under real-world conditions.

6 Conclusion

This paper presented *DASatFormer*, a hybrid multibranch Transformer framework that combines multidomain statistical feature extraction with gated attention mechanisms for scalable and interpretable DAS-based event recognition. By leveraging compact temporal, waveform, and spectral descriptors, the proposed approach effectively captures both local dynamics and long-range dependencies while drastically reducing data dimensionality and computational cost. Extensive experiments conducted on two DAS datasets covering laboratory, and real-scenario conditions demonstrate strong robustness and generalization, with accuracies exceeding 99.4% and low false-alarm rates. Compared to existing CNN-, hybrid-, and Transformer-based methods, including DASFormer, *DASatFormer* achieves a more favorable trade-off between accuracy, efficiency, and interpretability. Future work will investigate extensions toward multi-sensor configurations, self-supervised learning strategies, and lightweight implementations for real-time and edge-based DAS monitoring systems.

Acknowledgment

This project was funded by the French State as part of France 2030 operated by ADEME, and funded by the European Union – NextGenerationEU.

References

1. Tomasov, A., Zaviska, P., Spurny, V., Dejdar, P., Munster, P., Horvath, T., Klicnik, O.: Enhancing perimeter protection using ϕ -OTDR and CNN for event classification. In: *Optical Fiber Sensors*, pp. W4–39. Optica Publishing Group (2023)
2. Hibert, C., Provost, F., Malet, J.-P., Maggi, A., Stumpf, A., Ferrazzini, V.: Automatic identification of rockfalls and volcano-tectonic earthquakes at the Piton de la Fournaise volcano using a Random Forest algorithm. *Journal of Volcanology and Geothermal Research* **340**, 130–142 (2017)
3. Jia, H., Liang, S., Lou, S., Sheng, X.: A k-Nearest Neighbor Algorithm-Based Near Category Support Vector Machine Method for Event Identification of ϕ -OTDR. *IEEE Sensors Journal* **19**, 3683–3689 (2019)
4. Ghazali, M.F., Mohamad, H., Nasir, M.Y.M., Hamzh, A., Abdullah: State-of-the-art application and challenges of optical fibre distributed acoustic sensing in civil engineering. *Optical Fiber Technology* **87**, 103911 (2024)

5. Wu, H., Yang, M., Yang, S., Lu, H., Wang, C., Rao, Y.: A novel DAS signal recognition method based on spatiotemporal information extraction with 1DCNNs-BiLSTM network. *IEEE Access* **8**, 119448–119457 (2020)
6. Tejedor, J., Macias-Guarasa, J., Martins, H.F., Pastor-Graells, J., Corredera, P., Martin-Lopez, S.: Machine learning methods for pipeline surveillance systems based on distributed acoustic sensing: A review. *Applied Sciences* **7**(8), 841 (2017)
7. Bublin, M.: Event detection for distributed acoustic sensing: Combining knowledge-based, classical machine learning, and deep learning approaches. *Sensors* **21**, 7527 (2021)
8. Cao, X., Su, Y., Jin, Z., Yu, K.: An open dataset of ϕ -OTDR events with two classification models as baselines. *Results in Optics* **10**, 100372 (2023)
9. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. In: *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 30 (2017)
10. Li, Y., Qin, Y., Hu, L., Wu, H., Yu, K.: DASFormer: A long sensing sequence classification and recognition model for phase-sensitive optical time domain reflectometers. *IEEE Sensors Journal* (2025)
11. Jiang, F., Li, H., Zhang, Z., Zhang, X.: An event recognition method for fiber distributed acoustic sensing systems based on the combination of MFCC and CNN. In: *International Conference on Optical Instruments and Technology: Advanced Optical Sensors and Applications*, SPIE **10618**, 15–21 (2018)
12. Liu, M., Ren, S., Ma, S., Jiao, J., Chen, Y., Wang, Z.: Gated transformer networks for multivariate time series classification. *arXiv preprint arXiv:2103.14438* (2021)
13. Meng, H., Wang, S., Gao, C., Liu, F.: Research on recognition method of railway perimeter intrusions based on ϕ -OTDR optical fiber sensing technology. *IEEE Sensors Journal* **21**(8), 9852–9859 (2020)
14. Zhou, D., Kang, B., Jin, X., Yang, L., Lian, X., Hou, Q., Feng, J.: DeepViT: Towards deeper Vision Transformer. *arXiv preprint arXiv:2103.11886* (2021)
15. Jestin, C., Hibert, C., Calbris, G., Lanticq, V.: Integration of machine learning on distributed acoustic sensing surveys. *Copernicus Meetings Technical Report* (2020)
16. Xie, Y., Wang, M., Zhong, Y., Deng, L., Zhang, J.: Label-free anomaly detection using distributed optical fiber acoustic sensing. *Sensors* **23**, 4094 (2023)
17. Tomasov, A., Zaviska, P., Dejdar, P., Klicnik, O., Horvath, T., Munster, P.: Comprehensive Dataset for Event Classification Using Distributed Acoustic Sensing (DAS) Systems. *figshare* (2025)
18. Chen, X., Liu, W., Zhang, J., Song, Z.: A review of feature extraction and selection for vibration-based fault diagnosis of rotating machinery. *Mechanical Systems and Signal Processing* **160**, 107894 (2021)
19. Dione, M., Lonlac, J., Lecoeuche, S., Fleury, A.: Intrusion Pattern Recognition in DAS Using Multi-Domain Features and a Transformer Network. In: *Proceedings of the IEEE 15th International Conference on Pattern Recognition Systems* (2025)
20. Song, D., Liu, C.: A facial expression recognition network using hybrid feature extraction. *PLOS ONE* **20**(1), e0312359 (2025)
21. Guzhov, A., Raue, F., Hees, J., Dengel, A.: ESResNet: Environmental sound classification based on visual domain models. In: *Proceedings of the 25th International Conference on Pattern Recognition (ICPR)*, pp. 4933–4940 (2021)