

doi:10.13756/j.gtxyj.2026.260124.

专题:2025全国硕博士光通信菁英论坛

朱程远,蒲岩云,杨楷翔,等.基于多模态融合的长输管道外部安全监测研究[J].光通信研究,2026(5):22-32.

Zhu C Y, Pu Y Y, Yang K X, et al. Research on Safety Monitoring of Long-Distance Pipelines based on Multimodal Fusion [J]. Study on Optical Communications, 2026(5):22-32.

基于多模态融合的长输管道安全监测研究(特邀)

朱程远¹,蒲岩云¹,杨楷翔²,杨秦敏^{1,3}

(1. 浙江大学 控制科学与工程学院 工业控制技术全国重点实验室,杭州 310027; 2. 华南理工大学 计算机科学与工程学院,广州 510006; 3. 浙江省全省数智管理与决策技术重点实验室,杭州 310058)

摘要:【目的】针对长输管道复杂场景下外部振动信号类间相似、类内波动明显以及单一模态特征难以全面表征事件差异的问题,文章提出了一种基于多模态融合的管道外部安全监测方法,以提高分布式声学传感(DAS)系统在实际运行环境中的事件识别性能。【方法】文章所提方法通过省级智慧管网平台DAS系统采集真实场景振动信号,将信号分别构建为原始波形、短时傅里叶变换时频图和格拉姆角场图像3种模态。针对不同模态的数据表达特点,构建多分支特征提取网络,分别利用一维卷积神经网络提取局部时序特征,利用二维卷积神经网络提取时频纹理与能量分布特征,利用视觉Transformer提取图像编码中的全局相关结构特征。在此基础上,设计跨模态注意力融合机制,对不同模态特征进行对齐、交互和融合;同时考虑实际场景中的类别不平衡与模态融合偏差问题,构建跨模态协同焦点损失函数,以增强模型对难分类样本和少数类样本的学习能力。【结果】真实管道外部事件识别数据集上的实验结果表明,文章所提方法训练过程收敛稳定,在管道外部事件识别任务中取得了较好的综合性能,平均精确率、平均召回率和平均F1分数分别达到0.9683、0.9728和0.9705,均优于对比模型;单模态与多模态对比实验进一步表明,多模态输入能够提供更充分的判别信息,明显提升了模型对复杂振动信号的表征能力。【结论】文章所提方法能够有效融合原始时序、时频分布和全局相关结构信息,增强复杂背景下相似事件的可分性,为DAS管道外部安全监测与智能预警提供了一种有效的技术方案。

关键词:光纤传感;多模态;人工智能;模式识别;监测预警

中图分类号:TN929

文献标志码:A

Research on Safety Monitoring of Long-Distance Pipelines based on Multimodal Fusion

ZHU Chengyuan¹, PU Yanyun¹, YANG Kaixiang², YANG Qinmin^{1,3}

(1. State Key Laboratory of Industrial Control Technology, College of Control Science and Engineering, Zhejiang University, Hangzhou 310027, China; 2. College of Computer Science and Engineering, South China University of Technology, Guangzhou 510006, China; 3. Zhejiang Key Laboratory of Decision Intelligence, Hangzhou 310058, China)

Abstract:【Objective】To address the problems of inter-class similarity, significant intra-class variation, and insufficient representation capability of single-modal features for external vibration signals in complex long-distance oil and gas pipeline scenarios, a multimodal fusion-based pipeline external safety monitoring method is proposed to improve the event recognition performance of Distributed Acoustic Sensing (DAS) systems in practical operating environments.【Methods】Real-world vibration signals were collected by a DAS system deployed on a provincial smart pipeline network platform. Each signal sample was constructed into three modalities, namely raw waveform, short-time Fourier transform spectrogram, and Gramian angular field image. According to the data representation characteristics of different modalities, a multi-branch feature extraction network was developed. A one-dimensional convolutional neural network was used to extract the local temporal features. A two-dimensional convolutional neural network was used to extract time-frequency texture and energy distribution features, and a vision Transformer was used to extract global correlation structure features from encoded images. On this basis, a cross-modal attention fusion mechanism was designed to align, interact, and fuse heterogeneous modal features. Meanwhile, considering class imbalance and modal fusion bias in real-world scenarios, a cross-modal collaborative focal loss function was constructed to enhance the learning capability of the model for hard-to-classify and minority-class samples.【Results】Experimental results on a real pipeline external event recognition dataset show that the proposed method achieves stable convergence during training and good overall performance in the pipeline external event recognition task. The average precision, average recall, and average F1-score reach 0.9683, 0.9728, and 0.9705, respectively, outperforming the comparison models. In addition, the comparison between single-modal and multimodal inputs further demonstrates that multimodal input provides more sufficient discriminative information and significantly improves the representation capability of the model for complex vibration signals.【Conclusion】The proposed method can effectively integrate raw temporal information, time-frequency distribution, and global correlation structure, thereby improving the separability of similar events under complex backgrounds. It provides an effective technical solution for DAS-based pipeline external safety monitoring and intelligent early warning.

Key words: optical fiber sensing; multimodal; artificial intelligence; pattern recognition; monitoring and early warning

收稿日期:2026-05-05; 修回日期:2026-07-15; 纸质出版日期:2026-10-10

基金项目:浙江省自然科学基金白马湖实验室联合基金资助项目(LBMHD24F030001);长三角科技创新共同体联合攻关重点揭榜任务资助项目(20259b09050007);工业控制技术全国重点实验室浙大专项资助项目(ICT2026C05)

作者简介:朱程远(1996-),男,山东济宁人。博士,主要研究方向为分布式光纤传感。

通信作者:杨秦敏,教授。E-mail: qmyang@zju.edu.cn

© Editorial Office of Study on Optical Communications. This is an open access article under the CC BY-NC-ND license.

0 引言

长输管道是输送油气等资源的重要基础设施,其安全稳定运行对能源保障和社会生产具有重要意义。随着城市建设活动不断地增加,机械施工、车辆通行及人员活动等外部振动对管道运行安全构成了较大威胁^[1-2]。相比于人工巡检、风险评价等传统方法,分布式光纤传感技术具有监测距离长、抗腐蚀和抗电磁干扰能力强等优点,近年来在周界安防、管道安全预警和轨道振动监测等领域得到了广泛应用^[3-6]。其中,分布式声学传感(Distributed Acoustic Sensing, DAS)系统能够实时获取沿线振动信号,为管道外部事件识别提供了有效手段^[7-8]。

现有 DAS 事件识别方法大致可分为 3 类。第 1 类为基于阈值或规则判别的方法^[9],通常利用振动幅值、能量或频域特征进行报警,方法实现简单,但在长输管道复杂背景噪声下容易出现误报和漏报;第 2 类为基于人工特征和传统机器学习的方法^[10],通过提取时域、频域或时频域特征,并结合支持向量机、随机森林等分类器进行事件识别,该类方法对特征设计依赖性较强,面对多场景、多干扰信号时泛化能力有限;第 3 类为基于深度学习的方法^[11],利用卷积神经网络(Convolution Neural Network, CNN)、循环神经网络或 Transformer 等模型从信号中自动提取特征,近年来在 DAS 事件识别中取得了较好效果。

然而,长输管道沿线环境复杂多变,振动信号会受到土壤介质、传感距离、车辆干扰、施工方式及环境噪声等因素影响,导致不同事件之间存在局部特征相似,同一事件在不同场景下也可能表现出明显差异等情况。已有方法多依赖单一原始时序、时频图或图像编码模态,难以同时刻画振动信号的局部时序动态、时频能量分布和全局相关结构^[12]。针对

上述问题,本文提出了一种基于多模态融合的管道外部安全监测方法,将原始时序信号、短时傅里叶变换(Short-Time Fourier Transform, STFT)时频图和格拉姆角场(Gramian Angular Field, GAF)图像作为联合输入,构建多分支特征提取网络,并设计跨模态注意力融合机制,实现外部事件信号的识别,从而有效提升复杂环境下的识别精度和稳定性。

1 管道外部振动信号分析与问题描述

1.1 管道外部振动信号特征

如图 1 所示, DAS 系统与管道伴行铺设,沿线采集到的外部振动信号各异,具有明显的非平稳性和复杂性,不同类型事件信号在时域波形、频谱分布和结构模式上通常存在一定差异。例如,机械施工类事件往往具有较强冲击性和较宽的频带分布;车辆通行类事件持续时间相对较长;人员活动类事件振动幅值较弱但具有一定节律性;背景噪声则通常表现为随机波动。然而,在实际场景中,这些事件之间并不是完全可分的,不同类别信号在局部波形、主频范围和能量分布上往往存在一定重叠,容易产生混淆^[13]。尤其是在强干扰环境下,部分施工振动与交通干扰在时域幅值和频域能量上可能十分接近,仅依赖单一特征难以实现稳定区分。除类间相似性外,管道振动信号还存在明显的类内差异。由于管道沿线地质条件、埋深状态、周边设施和噪声背景不同,同一类事件在不同监测位置采集到的信号也会表现出不同的响应形式。例如,同样的机械施工事件,在农田、道路或村镇附近传播时,其幅值强弱、衰减速度和频谱形态均可能发生变化。这种“类间易混淆、类内变化大”的特征,使得基于幅值阈值、人工特征或浅层分类器的方法难以兼顾识别精度与鲁棒性。



图 1 长输管道与 DAS 系统监测示意图

Figure 1 Schematic diagram of long-distance pipeline and DAS system monitoring

从信号表征角度看,原始时序信号能够保留振动波形的局部起伏、瞬态变化和节律信息,但对频谱演化和全局结构关系的表达能力有限;时频表示能够反映能量随时间和频率的分布规律^[14],有利于表征不同事件的频带特征,但会弱化部分原始波形细

节;图像编码表示则能进一步挖掘时间序列内部的全局相关性和结构模式^[15],但其判别能力仍受输入形式限制。也就是说,不同表示形式各自保留了信号的部分有效信息,同时也不可避免地存在信息缺失。对于复杂背景下的管道外部事件识别,仅依靠

单一模态往往难以全面描述振动信号特征。因此,如何在复杂环境下充分利用不同表示形式之间的互补信息,是提高管道外部事件识别性能的关键问题。

1.2 问题描述

基于上述分析,本文将管道外部安全监测中的事件识别问题定义为一个分类任务,识别外部背景噪声、车辆行驶、人工作业和挖掘施工事件信号。假设 DAS 系统采集得到长度为 L 的管道外部振动信号样本为 x , x 为原始一维振动时序。由于原始振动序列具有明显的非平稳性,单一表示形式难以同时表征其局部波形、频谱演化及全局相关结构,因此本文对同一样本构造了 3 种互补模态,可表示为

$$X = \{x, S(x), G(x)\}, \quad (1)$$

式中: X 为多模态输入; $S(x)$ 为经 STFT 得到的时频图表示; $G(x)$ 为经 GAF 编码得到的全局结构表示。3 种模态分别对应振动事件的局部动态信息、频带能量分布信息和时序相关结构信息。

针对多模态输入 X , 考虑到不同模态间既包含反映事件本质的共性信息,也包含各自独有的判别信息,本文进一步通过共享-私有特征分解和跨模态注意力交互获得融合表示,并完成事件分类,其目标可表示为

$$\hat{y} = f(X; \theta_{sh}, \theta_{pr}, \theta_{fus}), \quad (2)$$

式中: $\hat{y}$ 为模型输出类别; $f(\cdot)$ 为多模态融合识别网络; θ_{sh} 、 θ_{pr} 和 θ_{fus} 分别为共享特征学习、私有特征学

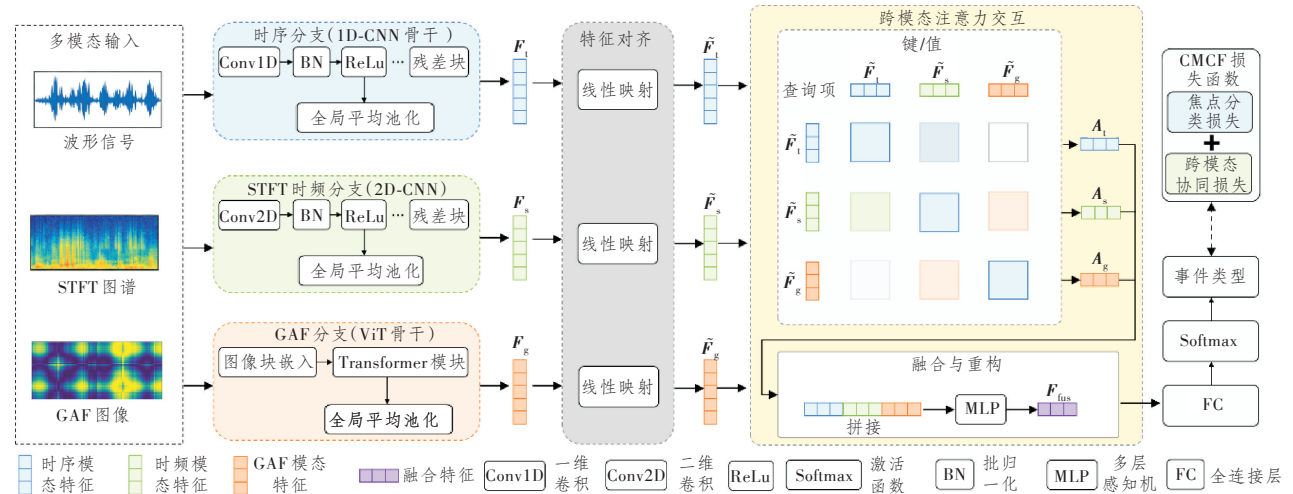
习以及跨模态融合模块的参数。

因此,本文旨在在复杂背景噪声、类间相似性强及类内波动明显的条件下,构建一种能够联合利用原始时序、时频分布和全局结构信息的多模态融合模型,以提高管道外部事件的识别精度和鲁棒性。

2 基于多模态融合的管道外部事件识别方法

2.1 整体框架

本文提出了一种基于多模态融合的管道外部事件识别方法,其整体框架如图 2 所示。该方法以 DAS 振动样本为输入,分别构建原始波形、STFT 时频图和 GAF 图像 3 模态,并采用多分支骨干网络进行特征提取。其中,原始波形分支侧重刻画局部时序动态;STFT 时频图分支侧重提取频谱纹理与能量分布特征;GAF 图像分支侧重建模全局相关结构与长程依赖关系。视觉 Transformer(Vision Transformer, ViT)的基本思想是将图像划分为 patch 序列并通过 Transformer 编码器进行建模,这为 GAF 图像中的全局结构表征提供了合适的实现方式。在此基础上,通过特征对齐与跨模态注意力机制实现 3 种模态之间的信息交互与融合,得到统一判别表示;最后,基于分类头输出事件识别结果,并结合所设计的损失函数完成网络训练。



注: 1D-CNN 和 2D-CNN 分别为 一维卷积神经网络和二维卷积神经网络; CMCF 为跨模态协同焦点; F_t 、 F_s 和 F_g 分别为原始波形、STFT 时频图和 GAF 图像经过多分支骨干网络提取后的高层特征表示; $\tilde{F}_t$ 、 $\tilde{F}_s$ 和 $\tilde{F}_g$ 分别为 3 种模态对齐处理后的特征; A_t 、 A_s 和 A_g 分别为 3 种模态交互后的注意力输出; F_{fus} 为对齐后的模态特征。

图 2 基于多模态融合的管道外部事件识别框架

Figure 2 Framework of the proposed multimodal fusion network for pipeline external event recognition

2.2 多分支特征提取

DAS 振动信号在不同表示空间中具有不同的

判别优势。原始波形能够直接保留振动幅值起伏、脉冲变化和节律信息,对局部动态模式较为敏感;

STFT 时频图能够将振动信号的时间变化与频率分布统一到二维平面上,更有利于表征不同事件在主频范围、能量集中区域及持续时间上的差异;GAF 图像则进一步强化了时间序列不同采样点之间的全局相关性与结构模式。由于 3 种模态在信息表达上各有侧重,单一模态往往只能反映事件特征的一个侧面,而多模态联合建模更有助于提高复杂场景下相似事件的可区分性。

基于上述考虑,本文采用 3 路分支结构对不同模态分别进行特征提取。如表 1 所示,对于原始波形输入 $[1, L]$,采用 1D-CNN 骨干网络进行建模,通过一维卷积、归一化、非线性激活和残差结构块提取局部时序模式,并利用全局平均池化压缩特征维度,以获得稳定的时序表征。对于 STFT 时频图输入 $[1, H, W]$ (其中,1 为通道数, H 和 W 分别为高

度和宽度),采用 2D-CNN 骨干网络进行建模,通过二维卷积和残差结构逐层提取时频纹理、边缘变化和局部能量分布特征,从而增强模型对复杂频谱模式的刻画能力。对于 GAF 图像输入 $[1, H, W]$,采用 ViT 骨干网络进行建模,先将图像划分为若干 patch 并映射为特征序列,再通过建模 patch 之间的长程依赖关系,以提取更强的全局结构特征。ViT 对图像 patch 序列的建模方式,使其在处理具有明显结构关联的编码图像时具有较好的适应性。3 种模态输入 X_t 、 X_s 和 X_g 对应的特征提取骨干网络分别记为 Φ_t 、 Φ_s 和 Φ_g ,则 3 路特征可表示为

$$F_t = \Phi_t(X_t), F_s = \Phi_s(X_s), F_g = \Phi_g(X_g)。 \quad (3)$$

为了保证后续跨模态交互的可行性,本文进一步利用线性映射将 3 路特征投影到统一维度,得到对齐后的模态特征 F_{fus} 。

表 1 3 种分支特征提取网络结构说明

Table 1 Explanation of three branch feature extraction network structures

分支	输入模态	输入尺寸	主要网络结构	输出特征
时序分支	原始波形	$1 \times L$	一维卷积层、批归一化层、ReLU 激活函数、残差卷积模块、全局平均池化层	$F_t \in R^d$
时频分支	STFT 时频图	$1 \times H \times W$	二维卷积层、批归一化层、ReLU 激活函数、残差卷积模块、全局平均池化层	$F_s \in R^d$
图像分支	GAF 图像	$1 \times H \times W$	图像块嵌入层、位置编码、多层 Transformer 编码器、分类标记或全局平均池化层	$F_g \in R^d$

注: R^d 为 d 维实数向量空间。

2.3 跨模态注意力融合机制

由于 3 种模态的表征空间不同,若直接对 3 路特征进行简单拼接,容易引入冗余信息,且难以充分挖掘模态之间的互补关系。为此,本文设计了跨模态注意力融合机制,在统一特征空间中实现 3 种模态之间的深层交互。具体地,先将对齐后的 3 路特征分别映射为查询项和键和值,再利用缩放点积注意力计算模态间相关性。与 Transformer 编码器类似,注意力可通过特征间的相似性自适应分配权重,从而增强对关键模态信息的关注。

设 3 模态对齐特征组成特征集合 $Z = [\tilde{F}_t; \tilde{F}_s; \tilde{F}_g]$,则跨模态交互后的注意力输出 A 可表示为

$$A = \text{Softmax}\left(\frac{QK^T}{\sqrt{d}}\right)V, Q = Z \cdot W_Q,$$

$$K = Z \cdot W_K, V = Z \cdot W_V, \quad (4)$$

式中: W_Q 、 W_K 和 W_V 分别为查询项、键和值的映射矩阵; Q 、 K 和 V 分别为查询键和值的向量; $[*]^T$ 为转置操作; d 为特征维度。这里的 Q 、 K 和 V 并不

是在 3 个模态上相互独立计算,而是基于 3 路模态特征的联合映射共同生成,因此所得注意力权重能够显式表征不同模态之间的相关性与互补性。

在获得跨模态交互结果后,本文进一步采用拼接与前馈映射构建最终融合表示,即将 3 路增强特征进行组合,并通过融合头完成信息压缩与重构,以此输出统一判别表示。由此,模型能够同时保留原始波形中的局部时序模式、STFT 时频图中的频谱分布信息以及 GAF 图像中的全局结构依赖,从而提高复杂背景下外部事件的识别能力。

2.4 损失函数设计

考虑到实际管道场景中不同类别样本数量存在差异,且多模态融合后仍需抑制不同模态之间的表征偏差,本文设计了一种新的损失函数,称为跨模态协同焦点损失(Cross-Modal Collaborative Focal Loss, CMCF Loss)函数。该损失函数由分类焦点项和模态协同约束项两部分组成:前者用于提升模型对难分类样本和少数类样本的关注能力;后者用于约束不同模态增强特征在语义空间中的协调一致

性。焦点损失通过降低易分类样本的损失权重,使模型更加关注难样本,这一思想已被广泛用于处理类别不平衡问题。CMCF Loss 项 L_{CMCF} 可表示为

$$L_{\text{CMCF}} = L_{\text{focal}} + \lambda L_{\text{cm}}, \quad (5)$$

式中: L_{focal} 为焦点损失分类项; L_{cm} 为跨模态协同约束项; λ 为平衡系数。 L_{focal} 和 L_{cm} 可分别表示为

$$L_{\text{focal}} = - \sum_{c=1}^C \alpha_c (1-p_c)^\gamma y_c \ln(p_c), \quad (6)$$

$$L_{\text{cm}} = \frac{1}{N} \sum_{i=1}^N \frac{1}{3} (\|A_i^i - A_s^i\|_2^2 + \|A_i^i - A_g^i\|_2^2 + \|A_s^i - A_g^i\|_2^2), \quad (7)$$

式中: C 为总类别数; α_c 为第 c 类的类别权重; y_c 为第 c 类的真实标签; p_c 为模型对第 c 类的预测概率; γ 为聚焦参数,默认设置为 2,用于降低易分类样本的损失贡献,使模型更加关注难分类样本; N 为样本数。 α_c 采用归一化反频率权重计算,可表示为:

$$\alpha_c = \frac{1/N_c}{\sum_{k=1}^C 1/N_k}, \quad (8)$$

式中, N_c 为第 c 类样本数。焦点损失分类项能够降低易分类样本对训练过程的主导作用,从而提升模型对少数类和难分类样本的学习能力。跨模态协同约束项则用于压缩不同模态增强特征之间的语义偏差,使融合结果在保留互补信息的同时保持较好的表征一致性。通过上述联合优化,模型能够在提高分类性能的同时增强多模态融合过程的稳定性和鲁棒性。

表 2 数据集类别组成与样本规模

Table 2 Dataset category composition and sample size

类别	事件说明	样本数量	占比	训练集数量	验证集数量	测试集数量
背景噪声	无明显外部扰动的环境信号	5 800	53.7%	4 640	580	580
车辆行驶	管道周边道路或场站区域客/货车车辆通行以及施工车辆引起的振动信号	1 350	12.5%	1 080	135	135
人工作业	管道周边人工活动引起的局部振动信号,包括人员行走、人工敲击和人工挖掘等	1 450	13.4%	1 160	145	145
挖掘施工	管道周边机械施工引起的强振动信号,包括挖掘机开挖、破土施工和机械移动等冲击扰动	2 200	20.4%	1 760	220	220

图 3 所示为 4 类管道外部事件的不同模态表示,实验将每段 DAS 振动信号样本分别构建为原始波形、STFT 时频图和 GAF 图像 3 种模态输入。不同事件在 3 种模态下表现出不同特征:原始波形主要反映振动幅值及局部时序变化;STFT 时频图能够展示信号能量在时间和频率维度上的分布;

3 实验研究

为验证所提方法的有效性,本文通过省级智慧管网平台的 DAS 系统在线监测平台采集了真实场景信号。该系统基于沿长输油气管道伴行敷设的光纤实现分布式振动感知,能够连续采集管道沿线外部扰动信号。系统空间分辨率为 10 m,最大探测距离为 50 km/通道,原始采样频率为 2 kHz。所采集信号来自实际管道沿线不同运行环境,包含外部背景噪声、车辆行驶、人工作业和挖掘施工 4 类典型事件,共计 10 800 个样本,各类样本的组成与规模如表 2 所示。其中,背景噪声样本 5 800 个,占比 53.7%;车辆行驶样本 1 350 个,占比 12.5%;人工作业样本 1 450 个,占比 13.4%;挖掘施工样本 2 200 个,占比 20.4%。每个样本取连续 5 s 的 DAS 振动信号片段作为模型输入。实验样本并非来自单一固定采集点,而是从管道沿线不同光纤位置的监测数据中随机抽取,覆盖了不同传感距离下的外部振动信号。对于没有对应目标活动记录、仅在单一或少量防区短暂出现、缺乏持续或相邻防区一致传播模式的偶发扰动,归入背景类别;对于能够与车辆行驶、人工作业或机械施工记录对应的扰动,则标注为相应事件类别。由于实际场景下不同类型振动信号受光纤距离、土壤介质、周边交通和环境噪声等因素影响,同类事件之间可能存在一定波动,不同类别之间也可能出现局部特征相似现象,因此能够较好地反映真实管道监测中的复杂性。

GAF 图像则进一步刻画信号内部的全局相关结构。其中,时域原始波形信号已从确认事件的持续时间内截取连续 5 s,并进行块平均、轻度平滑和标准化处理。需要说明的是,5 s 信号为模型使用的固定长度分类窗口,而非事件从发生到结束的完整记录。对于持续时间超过 5 s 的车辆行驶、人工作业或挖

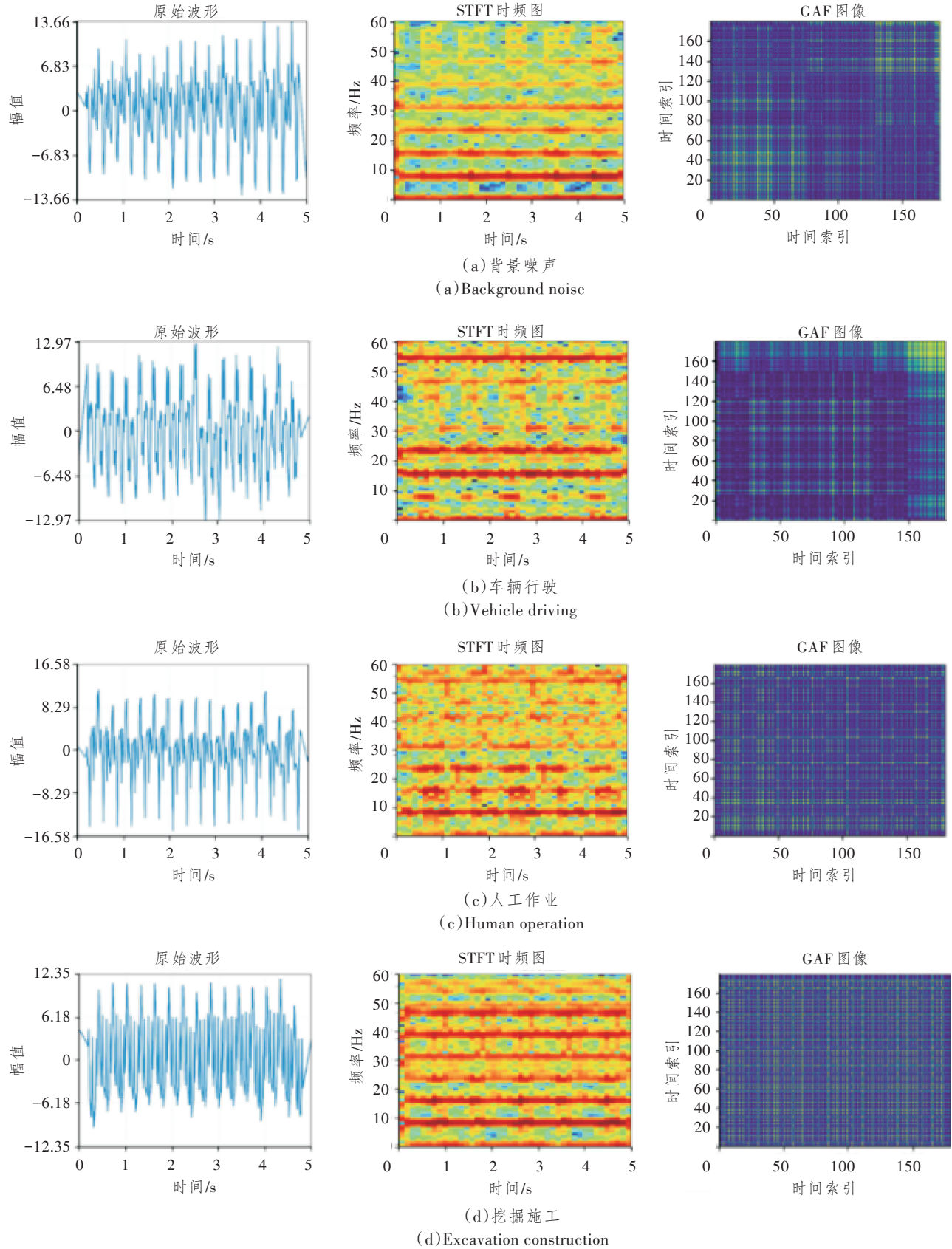


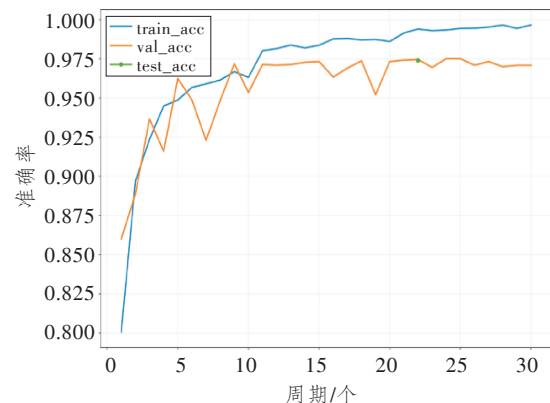
图 3 不同类型数据 3 种模态示意图

Figure 3 Schematic diagram of three modalities for different types of data

掘施工事件,截取窗口可能位于事件活动过程内部,整个窗口均标记为相应事件类别。当周期振动不是理想正弦波,或者其传播过程中受到机械接触、结构响应、介质耦合或反射等影响时,除基频外还会产生整数倍频成分。长输管道沿线可能持续受到道路交通、场站设备、泵机和其他周期性机械活动的共同影响,这些低频振动可经土壤、管道及伴行光缆传播,并作为共同背景叠加在不同事件信号中。不同类别的主要差异体现在瞬态响应的幅值、密度、持续性及频率能量分布。实际机械扰动通常由多种激励共同构成,且存在背景噪声,其频率成分可能随时间变化并分布于较宽的频带,不一定表现为单一稳定主频。同时,车辆轮轴与路面接触、人员脚步及工具敲击、人工挖掘和机械铲斗破土等活动均可能产生单个或成组的短时冲击,并经土壤和光缆耦合传播至传感光纤。例如,车辆行驶具有相对连续或重复的振动响应,人工作业具有较明显的间歇性,挖掘施工则通常表现出更强的机械冲击和更宽的能量分布。因此,图中仍保留了真实监测环境中的非平稳波动、背景干扰及局部瞬态响应,部分相邻点之间的陡峭变化并非信号中断或缺失采样,而是局部瞬态机械响应经块平均后在有限时间分辨率下形成的明显幅值差异。

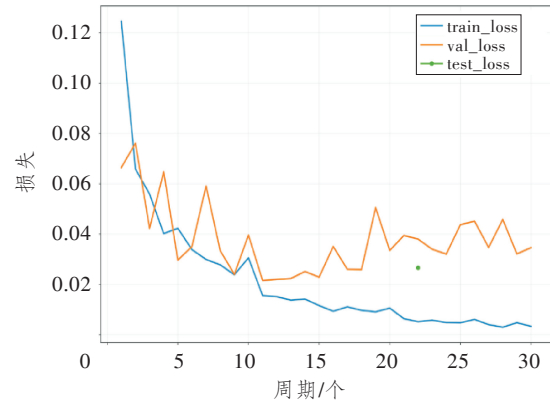
数据集按照训练集、验证集和测试集 8:1:1 的比例划分,使不同距离和不同噪声条件下的样本均参与模型训练与测试。根据归一化反频率权重计算类别权重后,将背景噪声、车辆行驶、人工作业和挖掘施工对应的类别权重 α_c 分别设置为 0.084、0.360、0.335 和 0.221。评价指标选取平均精确率、平均召回率和平均 F1 分数,用于衡量模型的识别性能。其中,精确率表示被预测为某一类别的样本中正确预测样本所占比例,用于衡量模型预测结果的可靠性;召回率表示实际属于某一类别的样本中被正确识别的比例,用于反映模型对该类别样本的识别能力;F1 分数是精确率与召回率的调和平均值,用于综合评价模型在精确率和召回率之间的平衡表现。平均精确率、平均召回率和平均 F1 分数,即分别计算各类别的精确率、召回率和 F1 分数后,对各类别结果进行平均处理。图 4 所示为模型训练过程中准确率和损失函数的变化曲线。其中,准确率表示正确分类样本数占全部样本数的比例,用于描述模型训练过程中训练集和验证集整体分类性能的变化;图中 train_acc、val_acc 和 test_acc 分别为训练集准确率、验证集准确率和测试集准确率,train_loss、

val_loss 和 test_loss 分别为训练集损失、验证集损失和测试集损失。由图 4(a)可知,模型在前几个训练周期内训练集和测试集准确率提升较快,说明所构建的多模态输入和多分支骨干网络能够较快提取有效判别特征。随着训练周期的增加,训练集准确率持续上升,验证集准确率在约 10 个周期后基本收敛,后续仅出现小范围波动,表明模型具有较好的泛化能力。由 4(b)可以知,训练损失在初始阶段下降较快,随后逐渐减小并趋于稳定;验证损失虽然在中后期存在一定波动,但整体保持在较低水平。这说明本文所提模型在训练过程中能够逐步学习到稳定的模态融合表示,所设计的损失函数对模型收敛起到了积极作用。



(a) 模型准确率曲线

(a) Model accuracy curve



(b) 模型损失曲线

(b) Model loss curve

图 4 模型准确率和损失变化曲线

Figure 4 Model accuracy and loss variation curve

为验证本文所提方法的优越性,本文选取残差神经网络(Residual Network, ResNet)、视觉几何组网络(Visual Geometry Group, VGG)和 ViT 作为对比模型,实验结果如表 3 所示。由表可知,ResNet 与 VGG 模型的平均精确率、平均召回率和平均 F1 分数分别为 0.857 1、0.864 2 和 0.860 2 以及 0.860 1、0.855 8 和 0.858 1,说明传统 CNN 能够提

取一定的局部判别特征,但在复杂 DAS 振动信号识别中仍存在局限。其主要原因在于,CNN 更关注局部纹理和邻域模式,对跨时间、跨频率以及全局相关结构的建模能力相对不足,因此在处理类间特征相似的振动事件时容易受到限制。ViT 模型整体优于 CNN,表明基于图像结构建模的方法能够更好地挖掘编码图像中的全局特征。相比之下,本文所提多模态融合方法的平均精确率、平均召回率和平均 F1 分数分别达到 0.968 3、0.972 8 和 0.970 5,均取得最优结果。与 ViT 模型相比,本文所提方法的平均 F1 分数提高了 0.063 4,说明多模态输入与跨模态注意力融合机制能够更充分地利用不同表示空间中的互补信息,从而提高事件识别性能。图 5 所示为不同模型在测试集上的混淆矩阵。通过混淆矩阵可以进一步分析不同模型性能差异的具体来源,由图可知,ResNet 和 VGG 模型虽然能够识别大部分背景噪声和挖掘施工样本,但在背景噪声、车辆行驶和人工作业样本间仍存在一定误判,说明仅

依赖卷积特征难以充分表征 DAS 振动信号的复杂时序变化和全局相关结构。ViT 模型具有更好的整体分类效果,但仍存在一定的类别混淆,说明单一图像模态虽然能够增强全局结构建模能力,但在复杂背景噪声和不同强度外部扰动共存的情况下,仍难以完全区分相似事件。相比之下,本文所提方法的混淆矩阵对角线样本数量最高,非对角线误分类数量明显减少,表明同时利用跨模态注意力机制增强不同模态之间的信息交互,能够有效减少背景噪声、人工活动和机械施工等相似事件之间的误判。

表 3 不同模型在不平衡测试集中的对比实验结果

Table 3 Comparative experimental results of different models on the unbalanced test set

模型	平均精确率	平均召回率	平均 F1 分数
ResNet	0.857 1	0.864 2	0.860 2
VGG	0.860 1	0.855 8	0.858 1
ViT	0.902 1	0.912 4	0.907 1
本文所提	0.968 3	0.972 8	0.970 5

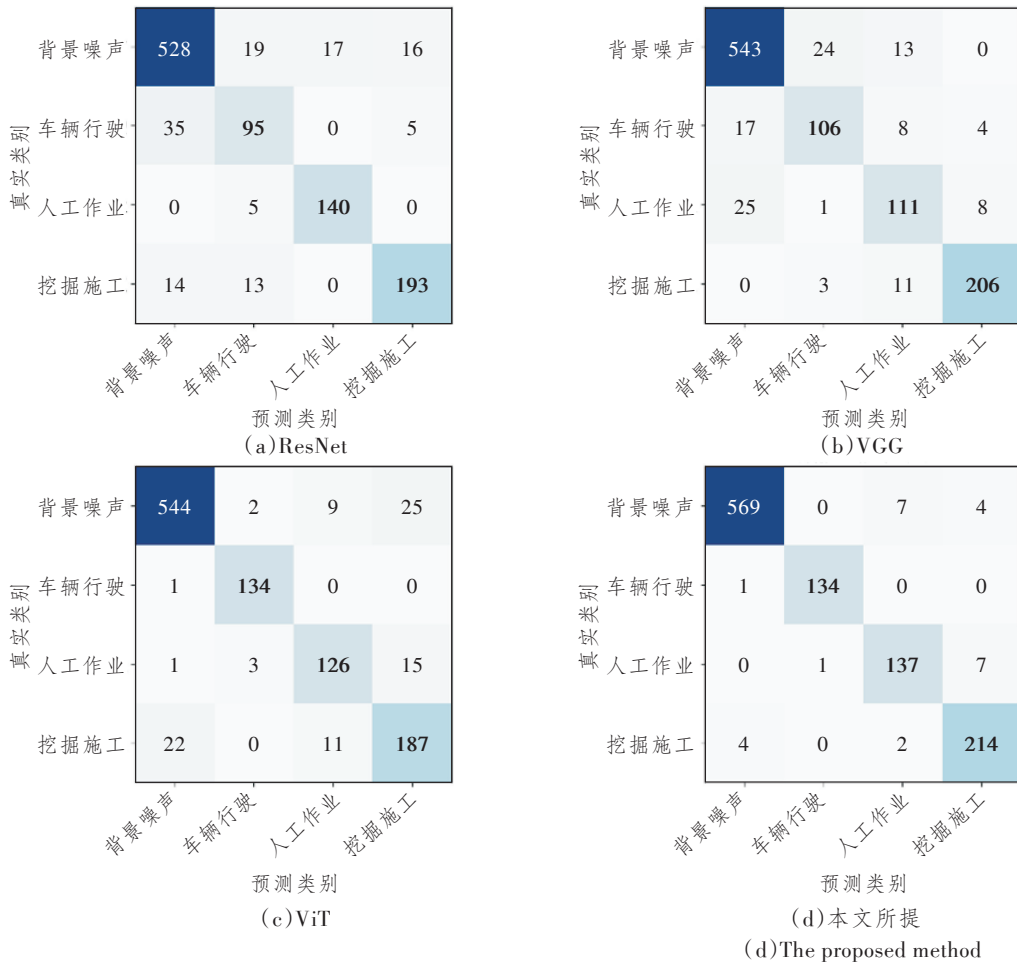


图 5 不同模型在测试集上的混淆矩阵

Figure 5 Confusion matrix of different models in test set

为进一步排除测试集类别分布对模型评价结果的影响,本文在原始测试集基础上构建了类别平衡测试集。从背景噪声、车辆行驶、人工作业和挖掘施工4类测试样本中分别随机抽取135个样本,使各类别均占25%,平衡测试集共包含540个样本。模型训练数据、网络结构和参数设置均保持不变,仅使用训练完成的模型在平衡测试集上进行补充评价,结果如表4所示。在各类别样本数量相同的条件下,所提方法的平均精确率、平均召回率和平均F1分数分别达到0.9757、0.9746和0.9747,仍均优于ResNet、VGG和ViT模型。本文所提方法在平衡测试集的表现与原始测试集接近,且略有提升,表明其识别性能并未因原测试集中背景噪声样本占比较高而被高估。进一步说明,多模态特征提取与跨模态注意力融合能够稳定利用原始波形、STFT时频图和GAF图像之间的互补信息,在不同类别分布条件下均保持较好的事件判别能力和鲁棒性。

表4 不同模型在平衡测试集中的对比实验结果

Table 4 Comparative experimental results of different models on the balanced test set

模型	平均精确率	平均召回率	平均F1分数
ResNet	0.8974	0.8965	0.8968
VGG	0.9144	0.9122	0.9121
ViT	0.9159	0.9136	0.9128
本文所提	0.9757	0.9746	0.9747

图6所示为单模态输入与多模态输入的实验对比结果。其中,单模态输入包括单模态-原始波形、单模态-STFT时频图和单模态-GAF图像。3种单模态实验均采用本文所提方法的单分支分类网络结构和训练设置,仅输入模态不同,以保证实验结果具有可比性。由图可知,单模态输入在平均精确率、平均召回率和平均F1分数上的表现不如多模态输入,说明仅依赖单一表示形式难以完整描述DAS振动信号的复杂特征。对于管道外部事件而言,原始波形能够反映振动幅值变化和局部动态特征,但对频谱分布和长期相关结构表达不足;STFT时频图能够突出不同事件的频率成分和能量演化,但可能弱化部分原始波形细节;GAF图像能够刻画信号内部的全局相关关系,但对局部瞬态响应的表达仍有一定局限。多模态输入在3项评价指标上均表现出明显优势,表明原始波形、STFT时频图和GAF图像之间具有较强互补性。通过多分支特征提取网络,模型能够分别从局部时序动态、时频能量分布和全局相关结构3个角度提取特征;通过跨模态注意

力机制,模型进一步对不同模态特征进行交互和融合,使最终特征表示更加完整和稳定。因此,多模态融合不仅提高了模型对不同类型事件的区分能力,也增强了模型在复杂背景噪声下的鲁棒性。

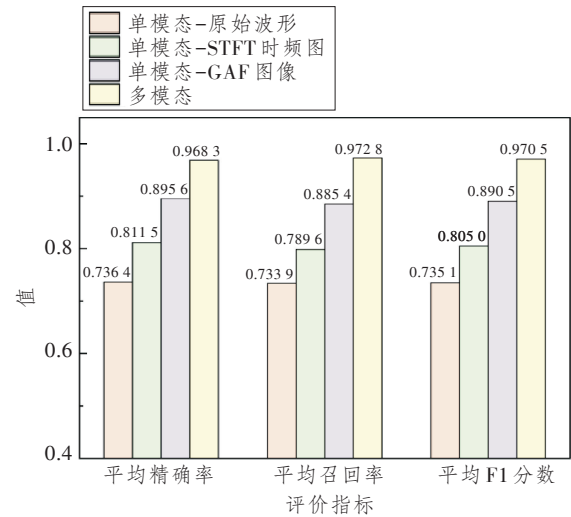


图6 单模态与多模态输入实验对比

Figure 6 Comparison experiment between single-modal and multi-modal input

为验证CMCF Loss中超参数设置的合理性,本文进一步开展了对损失函数中平衡系数 λ 和聚焦参数 γ 的敏感性分析,结果如图7所示。由7(a)可知,固定 $\gamma=2$ 时,随着 λ 由0增加至0.1,模型平均F1分数由0.84逐步提升至0.97,说明跨模态协同约束项能够有效促进不同模态特征之间的语义一致性,提高多模态融合效果。当 λ 继续增大至0.15和0.20时,平均F1分数分别下降至0.94和0.91,说明过强的模态协同约束可能压缩不同模态之间的互补差异,反而影响模型判别能力。由7(b)可知,固定 $\lambda=0.1$ 时, $\gamma=0,1,2,3$ 和4对应的平均F1分数分别为0.89、0.93、0.97、0.88和0.91。当 $\gamma=0$ 时,焦点损失退化为加权交叉熵,模型对难分类样本的关注不足;当 $\gamma=2$ 时,模型取得最高平均F1分数;而当 γ 继续增大时,模型性能下降,说明过大的 γ 可能过度抑制易分类样本的训练贡献,影响整体收敛稳定性。因此,本文实验条件下最优参数设置为 $\lambda=0.1$ 和 $\gamma=2$ 。

综合实验结果可知,本文所提方法能够在真实DAS管道监测数据上取得较好的识别效果。与传统卷积模型相比,本文所提方法能够弥补局部特征建模对全局结构表达不足的问题;与单一ViT模型相比,本文所提方法进一步增强了模型对复杂振动信号的综合表征能力。实验结果验证了本文所提多

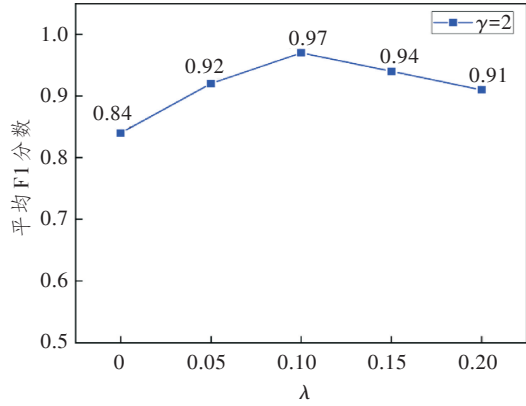
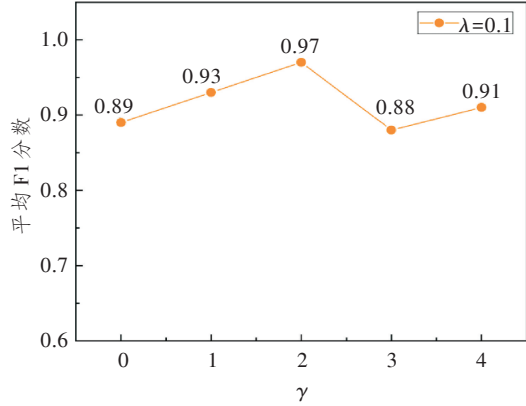
(a) 参数 λ 敏感性分析(a) Sensitivity analysis of parameter λ (b) 参数 γ 敏感性分析(b) Sensitivity analysis of parameter γ

图 7 CMCF Loss 超参数敏感性分析

Figure 7 Sensitivity analysis of CMCF Loss hyperparameters

模态输入构建、多分支特征提取、跨模态注意力融合及损失函数设计的有效性,为实际管道外部安全监测中的智能识别与预警提供了实验依据。

4 结束语

本文针对油气管道场景下 DAS 系统识别外部振动信号单一模态表征不足的问题,提出了一种融合多模态事件的识别方法。实验结果表明,本文所提模型训练过程收敛稳定,在对比实验中平均精确率、平均召回率和平均 F1 分数上达到最优效果,说明多模态输入与跨模态注意力融合能够有效提升事件识别性能。结果进一步表明,原始波形、STFT 时频图和 GAF 图像分别从局部时序动态、频谱分布和全局相关结构角度提供互补信息,相比单模态输入能够获得更具判别性的融合特征。因此,本文所提方法可为基于分布式光纤传感的管道外部安全监测提供一种有效的智能识别方案。

参考文献:

- [1] Zhu C, Pu Y, Yang K, et al. Distributed Optical Fiber Intrusion Detection by Image Encoding and SwinT in Multi-Interference Environment of Long-Distance Pipeline [J]. IEEE Transactions on Instrumentation and Measurement, 2023, 72: 2515012.
- [2] Wu H, Chen J, Liu X, et al. One-Dimensional CNN-based Intelligent Recognition of Vibrations in Pipeline Monitoring with DAS [J]. Journal of Lightwave Technology, 2019, 37(17): 4359–4366.
- [3] 王龙生, 谢欣, 姚猛, 等. 分布式光纤声波传感及其进展研究 [J]. 光通信研究, 2025(5): 240100.
Wang L S, Xie X, Yao M, et al. Distributed Optical Fiber Acoustic Wave Sensing and Its Progress [J]. Study on Optical Communications, 2025(5): 240100.
- [4] Zhu C, Yang Y, Yang K, et al. Monitoring Landslide Disturbances Using Distributed Acoustic Sensing under Extreme Weather Conditions [J]. npj Natural Hazards, 2026, 3: 23.
- [5] Wu D, Liang Q Q, Hu B X, et al. Research Progress of Event Intelligent Perception based on DAS [J]. Sensors, 2025, 25(16): 5052.
- [6] 丁美琪, 桂林, 王子怡, 等. 基于 1D-CNN 的射频强度温度传感研究 [J]. 光通信研究, 2025(2): 240159.
Ding M Q, Gui L, Wang Z Y, et al. Research on RF Intensity Temperature Sensing based on 1D-CNN [J]. Study on Optical Communications, 2025(2): 240159.
- [7] Zhu C, Pu Y, Lyu Z, et al. A Lightweight Pipeline Edge Detection Model based on Heterogeneous Knowledge Distillation [J]. IEEE Transactions on Circuits and Systems II: Express Briefs, 2024, 71(12): 5059–5063.
- [8] 何欣. 基于分布式光纤振动传感的长输埋地管道安全监测技术 [J]. 山西建筑, 2020, 46(7): 197–198.
He X. Safety Monitoring of Long-Distance Buried Pipeline based on Distributed Vibration Fiber Optical Fiber Sensor [J]. Shanxi Architecture, 2020, 46(7): 197–198.
- [9] Zhou X, Wu Z. Research Progress of Pipeline Health Monitoring based on Distributed Fiber Optic Sensing [C]//2024 9th International Conference on Electronic Technology and Information Science (ICETIS). Hangzhou, China: IEEE, 2024: 10593770.
- [10] Zhu C, Pu Y, Lyu Z, et al. Multidimensional Information Fusion and Broad Learning System-based Condition Recognition for Energy Pipeline Safety [J]. Knowledge-Based Systems, 2024, 300: 112259.

- [11] Nie R, Sun Y, Wang S. Research on Underground Pipeline Illegal Intrusion Monitoring based on Distributed Optical Fiber Sensing Technology [C]//First Conference on Distributed Optical Fiber Sensing Technology and Applications (DOFS 2024). Nanjing, China: SPIE, 2025: 136540C.
- [12] 张均伟, 蓝波, 黄嘉庚, 等. 基于分布式光纤传感的光缆防外破监测研究[J]. 光通信研究, 2022, (5): 53–57.
Zhang J W, Lan B, Huang J Q, et al. Monitoring and Early Warning System for Anti-breakage of Optical Cable based on Distributed Optical Fiber Vibration Sensing [J]. Study on Optical Communications, 2022, (5): 53–57.
- [13] Wen C, Hu S, Wang J, et al. Fiber Optic Sensing Technology in Underground Pipeline Health Monitoring: a Comprehensive Review [J]. Structural Health Monitoring, 2026, 25(1): 630–658.
- [14] Zhu C, Yang Y, Yang K, et al. AI-based Energy Transportation Safety: Pipeline Radial Threat Estimation Using Intelligent Sensing System [J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2024, 38(20): 22556–22564.
- [15] Zhu C, Pu Y, Yang K, et al. A Novel Visual Transformer for Long-Distance Pipeline Pattern Recognition in Complex Environment [J]. IEEE Transactions on Artificial Intelligence, 2024, 5(6): 2933–2945.