

面向工业环境气体泄漏检测的多模态融合模型*

王泓森¹ 王建国² 杨建东³ 冯 勇¹

(1. 云南省计算机技术应用计算机重点实验室 昆明 650500; 2. 中国铜业有限公司 昆明 650051;

3. 昆明信息港传媒有限责任公司 昆明 650032)

摘要:在有色金属冶金等重工业环境中,危险气体泄漏的检测对保障工作人员安全和维持生产稳定至关重要。传统单模态的检测方法无法有效应对复杂工业环境中的干扰,在噪声环境下面临准确性降低的问题。针对上述问题,提出了一种面向工业环境下的多模态气体泄漏检测模型,该模型融合了烟雾传感器数据和红外图像数据,利用各数据源的互补优势,提高检测的准确性和鲁棒性。首先针对不同模态数据的特性,使用门控多层感知机(gMLP)捕捉传感器数据中的复杂模式;同时采用 Swin-Transformer 提取红外图像中的局部特征和全局特征。之后,利用基于多头注意力的融合策略,有效融合不同模态数据之间的潜在表示并获得有害气体的识别结果。通过在正常环境下和噪声环境下的多模态气体数据集上进行实验,该模型取得了 97.92% 的识别准确率,结果表明,相较于单一模态,多模态的方法可以有效提升检测的准确性和鲁棒性,提升了在复杂工业场景下气体泄漏检测的性能。

关键词:工业气体泄漏检测多模态;多传感器融合;注意力机制;Transformer

中图分类号: TP274; TN919.5 **文献标识码:** A **国家标准学科分类代码:** 510.4

Multi-modal fusion model for industry environment gas leakage detection

Wang Hongsen¹ Wang Jianguo² Yang Jiandong² Feng Yong¹

(1. Yunnan Province Key Laboratory of Computer Technology Application, Kunming 650500, China; 2. China Copper Corporation Limited, Kunming 650051, China; 3. Kunming Information Hub Media Co., Ltd., Kunming 650032, China)

Abstract: In heavy industries such as non-ferrous metal metallurgy, the detection of hazardous gas leaks is crucial for ensuring employee safety and maintaining stable production. Traditional single-modal detection methods often struggle with reduced accuracy in complex industrial environments due to their limited ability to handle disturbances, particularly in noisy conditions. To address these challenges, this paper introduces a multimodal gas leak detection model designed for industrial environments. This model integrates smoke sensor data and infrared image data, leveraging the complementary strengths of each data source to enhance detection accuracy and robustness. Initially, the gMLP architecture is utilized to capture complex patterns in sensor data; concurrently, the Swin-Transformer is employed to extract local and global features from infrared images. Subsequently, a fusion strategy based on multi-head attention effectively combines the latent representations of different modal data to achieve hazardous gas detection. Experiments conducted on multimodal gas datasets in both normal and noisy environments demonstrate that the model achieves a detection accuracy of 97.92%. The results indicate that, compared to single-modal methods, the multimodal approach significantly improves detection accuracy and robustness, enhancing performance in complex industrial gas leak detection scenarios.

Keywords: industrial gas leak detection; multimodal; multi-sensor fusion; attention mechanism; Transformer

0 引言

在工厂的生产环境中,由于设备老化和操作失误等原因,可能会导致有害气体泄漏。这不仅会造成资源浪费和环境污染,还可能对工作人员的健康和安全构成严重威胁^[1]。此外,工厂设备的稳定运行是确保生产稳定性和产品质量可靠性的关键。因此,快速、准确地检测和处理有害气体泄漏,对于保障工厂的安全生产、提升经济效益以及实现环境保护具有重要的现实意义。

传统气体泄漏检测方法主要包括人工检查、声音监测^[2]、流量检测^[3]。这些方法依赖于数据采集的准确性和所使用的数学模型^[4]。此外,这些方法不仅需要大量的人力、财力和时间投入同时还可能对管道、储罐以及周围环境造成破坏^[5]。而且,随着管道长度和工厂结构复杂性的增加,传统手动气体检测技术的有效性进一步降低^[6]。

为了克服这些方法的局限性,在气体泄漏检测领域已经发表了很多研究工作。其中一类工作主要利用气体传感器检测泄漏^[7-9]。例如,Yang 等^[7]提出了一种基于压力传感器数据和双向长短期记忆网络(bidirectional long short-term memory, BiLSTM)的流体管道泄漏检测与扰动辅助定位方法,实现了对漏源位置的准确定位;Salameh 等^[8]通过设计集成气体传感器的无线传感器网络系统,实现了对液化石油气泄漏的早期检测;张立豪等^[9]提出了一种基于迁移学习的无监督工况气体泄漏智能检测技术,解决多工况干扰与数据标签稀缺的问题。

得益于气体传感器的气敏特性和高灵敏度,这种方法能够迅速检测到特定气体的存在,在相对简单的监测环境中表现良好^[10]。然而,在复杂的混合气体环境中,传感器对不同气体的灵敏度存在差异,这可能导致只能检测到某些特定类型的气体,从而引发误报或漏报的情况^[11]。

另一类工作主要通过图像检测气体泄漏^[12-14]。例如,彭道刚等^[12]等通过将残差网络(residual networks, ResNet)和卷积注意力模块引入 UNet,实现了对电厂高压蒸汽泄漏的检测;彭敏等^[13]研究了基于可见光图像的工业过程气体泄漏检测方法,通过结合帧间差分法与天津法提出了一种自适应检测方法,提高了对运动目标的检测速度;李泉成等^[14]提出了光流增强的红外成像气体泄漏检测方法,通过光流网络提取视频运动特征,实现了高精度的实时检测。

此类基于图像的方法通过视觉信息综合判断泄漏的范围和位置,且不依赖于特定类型的传感器,具有更高的灵活性和适应性^[15]。虽然这种方法能够检测泄漏的存在和位置,但由于图像数据无法提供关于气体种类的直

接信息,因此无法准确确定泄漏气体的类型,进而影响泄漏源的识别^[16]。

为了结合两种模态的优势,也有一类工作同时利用传感器数据和图像信息来检测烟雾泄漏^[17-18]。如 Narkhede 等^[17]提出了一种基于卷积神经网络(convolutional neural network, CNN)和长短期记忆网络(long short-term memory, LSTM)的多模态气体检测方法,通过融合气体传感器阵列和热成像数据,提高了检测的准确率;Faleh 等^[18]提出了一种基于 CNN 和线性判别分析(linear discriminant analysis, LDA)的混合模型,用于污染气体的检测与分类,证明了混合模型在分类精度方面具有一定的可靠性和优势。

Transformer^[19]架构以其强大的并行计算能力和注意力机制,已经在自然语言处理和计算机视觉领域中取得了显著成果。Wang 等^[20]提出了一种渐进式金字塔结构的视觉 Transformer(pyramid vision transformer, PvT),有效降低了特征图的计算成本,同时提升了性能;Huang 等^[21]提出了一种用于处理表格数据的 Transformer(tabular data modeling transformer, Tab-Transformer),通过生成上下文嵌入,有效建模了表格数据信息。

也有研究人员尝试将 Transformer 应用于多模态数据融合和分析中^[22-23]。例如,唐永旺等^[22]提出了一种基于 Transformer 的时序多模态融合特征的异常网络流量检测方法,通过多模态注意力编码器融合分组特征实现异常流量的检测;Zhang 等^[23]提出了一种基于 Transformer 的方法实现了多模态皮肤病变诊断中图像模态与非图像数据的有效融合,提升了诊断准确性。

在制造业工厂等环境中,气体数据通常存在多种模态,如何从这些数据中提取有效特征并融合,准确检测气体泄漏,成为了关键问题。然而,目前很少有研究将 Transformer 架构应用到多模态气体泄漏检测任务当中。针对上述问题,本文从特征编码器和融合策略两个角度出发,提出了一种面向工业环境气体泄漏检测的多模态融合模型。具体来说,使用移位窗口 Transformer(shifted window transformer, Swin-Transformer)^[24]提取红外图像特征,通过层级结构和窗口自注意力机制,有效捕捉图像的局部和全局特征;同时采用门控多层感知机(gated multilayer perceptron, gMLP)^[25]提取传感器数据特征,通过多层感知机(multilayer perceptron, MLP)和门控机制,高效地处理序列数据,并捕捉其中的复杂模式;利用多头注意力机制融合多模态特征,同时关注不同特征子空间的信息,提高特征融合的效果。通过上述方法,不仅能够准确识别气体泄漏,还具备一定的鲁棒性。

1 模型结构

本文提出的模型结构如图 1 所示,该模型由特征提

取网络、特征融合网络和分类器 3 个部分组成。特征提取网络主要包括两个编码器分支,分别处理输入的传感器数据和红外图像数据(图 1(a));这些特征经过提取

和细化之后,在特征融合网络中进行融合(图 1(b));最后通过分类器检测是否发生气体泄漏(图 1(c))。

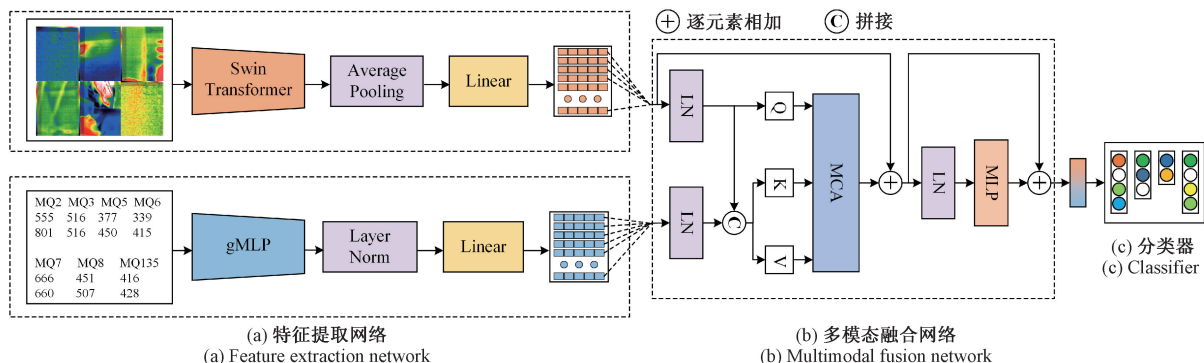


图 1 模型整体结构

Fig. 1 Overall model structure

1.1 特征提取网络

对于输入的传感器数据,采用 gMLP 作为特征编码器。gMLP 不仅具有较高的特征提取能力,而且其模型结构相对简单,计算效率高。gMLP 通过多层感知器结合门控机制,不仅能够有效捕捉传感器数据中的复杂模式和关系,提高特征提取的准确性;还能够动态调节特征的重要性,增强模型的泛化能力和鲁棒性。

如图 2 所示,每个 gMLP 编码器由 4 层相同的结构组成,每层的核心是空间门控单元(spatial gating unit, SGU)。在 SGU 中,每个特征沿通道维度分为两部分,一部分通过空间投影层投影,并与另一部分相乘。空间投影使得提取的特征在全局范围内共享信息,而逐元素乘法保留了前面模块提取的局部特征。因此,SGU 能够根据全局和局部信息调整提取的特征。此外,在模块的开头和结尾使用通道投影层来混合每个特征的不同分量,使得学习到的特征更加灵活。

在提取传感器数据特征时,这些数据会被映射为长度为 L 、维数为 d 的高级嵌入序列 $E_s \in \mathbf{R}^{L \times d}$ 。然后这些嵌入序列 E_s 会被输入到堆叠 4 层的 gMLP 块中进行特征提取并得到传感器数据特征 F_s 。

每个 gMLP 块的定义如式(1)~(3)所示。

$$Z = \sigma(XU) \quad (1)$$

$$\tilde{Z} = s(Z) \quad (2)$$

$$Y = \tilde{Z}V \quad (3)$$

式中: X 为输入序列; U 和 V 表示沿着通道维度的线性投影; σ 表示激活函数; s 表示沿着空间维度的线性投影; Y 表示 gMLP 的输出。

对于输入的红外图像,使用 Swin-Transformer 作为红外图像特征编码器。Swin-Transformer 通过结合移位窗

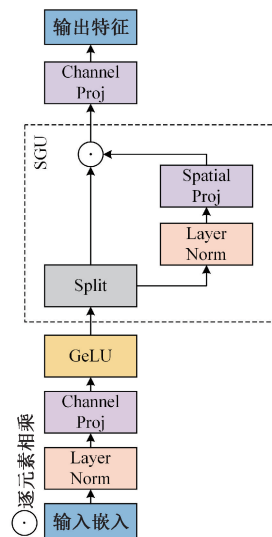


图 2 gMLP 结构

Fig. 2 gMLP structure

口注意力机制和分层特征表示,能够有效捕捉图像中的局部和全局特征,具有较强的特征提取能力。这种特性使其能够同时关注红外图像中的细节和整体结构信息,进而更全面地理解图像内容。

如图 3 所示,每个 Swin-Transformer 块由层归一化层(layer norm, LN)、多头自注意力(multi-head self-attention, MSA)模块、MLP 层和残差连接组成。两个连续 Swin-Transformer 块的 MSA 模块分别由基于窗口的 MSA(window multi self-attention, W-MSA)模块和基于移位窗口的 MSA(shifted window multi self-attention, SW-MSA)模块组成。

同样地,输入的红外图像会被映射为长度为 L 、维度为 d 的高级嵌入序列 $E_i \in \mathbf{R}^{L \times d}$ 。之后将嵌入序列 E_i 输

入堆叠 4 层的 Swin-Transformer 块中得到提取的红外图像特征 F_i 。

连续 Swin-Transformer 的输入和输出如式 (4) ~ (7) 所示。

$$\hat{f}^l = W - \text{MSA}(\text{LN}(f^{l-1}) + f^{l-1}) \quad (4)$$

$$f^l = \text{MLP}(\text{LN}(\hat{f}^l)) + \hat{f}^l \quad (5)$$

$$\hat{f}^{l+1} = \text{SW} - \text{MSA}(\text{LN}(f^l) + f^l) \quad (6)$$

$$f^{l+1} = \text{MLP}(\text{LN}(\hat{f}^{l+1})) + \hat{f}^{l+1} \quad (7)$$

式中: $\hat{f}^l$ 表示第 l 层 W-MSA 和 SW-MSA 的输出; f^l 表示 l 层中 MLP 的输出。

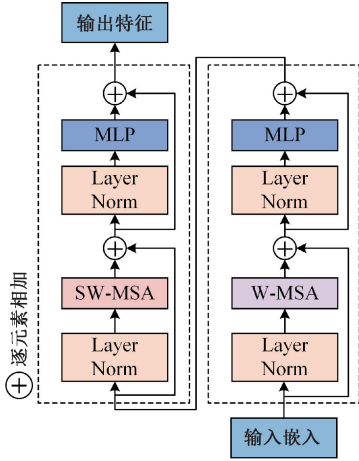


图 3 Swin-Transformer 结构

Fig. 3 Swin-Transformer structure

1.2 特征融合网络

考虑到将多模态信息融合为单一特征表示可能会忽视各模态的特异性。于是,本文引入了一种基于注意力机制的多模态特征融合网络。其是一种单分支多模态 Transformer,旨在融合红外图像特征与传感器特征。它利用当前模态中的特征分布及其他模态中的语义信息来计算注意力,从而实现有效的多模态特征融合。

整个融合过程如图 4 所示,首先,经过线性层 (linear) 的传感器特征 F_s 和红外图像特征 F_i 通过 LN 层进行归一化处理。然后,以 F_s 作为查询矩阵,拼接 F_i 与 F_i 作为键矩阵和值矩阵输入到 MCA 层中计算并得到

3 个注意力权重矩阵 $\tilde{Q}$ 、 $\tilde{K}$ 、 $\tilde{V}$ 。接着,根据计算得到的 3 个注意力权重矩阵 $\tilde{Q}$ 、 $\tilde{K}$ 、 $\tilde{V}$ 与 F_i 相加得到初步融合的特征 F_{mca} 。最后,将输出 F_{mca} 送入具有残差连接的 MLP 层便得到融合后的特征 F_{fus} 。

该过程如式 (8) ~ (12) 所示。

$$\tilde{Q} = \text{LN}(F_s) W^{\tilde{Q}} \quad (8)$$

$$\tilde{K} = \text{Concat}(\text{LN}(F_s), \text{LN}(\text{Concat}(F_i))) W^{\tilde{K}} \quad (9)$$

$$\tilde{V} = \text{Concat}(\text{LN}(F_s), \text{LN}(\text{Concat}(F_i))) W^{\tilde{V}} \quad (10)$$

$$F_{mca} = \text{MCA}(\tilde{Q}, \tilde{K}, \tilde{V}) + F_i \quad (11)$$

$$F_{fus} = \text{MLP}(\text{LN}(F_{mca})) + F_{mca} \quad (12)$$

式中: F_s 和 F_i 分别线性层的传感器特征和红外图像特征; $W^{\tilde{Q}}$ 、 $W^{\tilde{K}}$ 、 $W^{\tilde{V}}$ 表示 3 个权重矩阵; $\tilde{Q}$ 、 $\tilde{K}$ 、 $\tilde{V}$ 表示重新计算后的 3 个矩阵; F_{mca} 表示经过多头注意力层初步融合的特征; F_{fus} 表示融合后的特征。

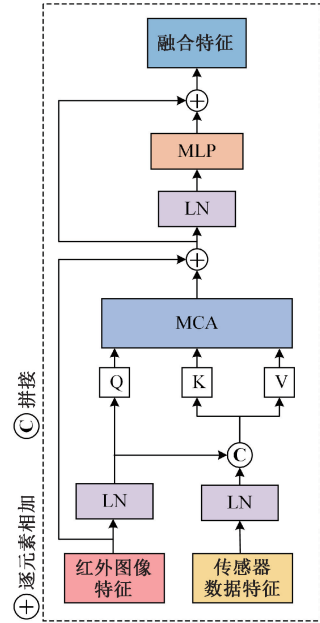


图 4 多模态融合网络结构

Fig. 4 Multimodal fusion network structure

1.3 分类器

最后将融合后的特征 F_{fus} 输入由线性层和激活函数 Softmax 组成的分类器中即可获得各类别的识别结果 P_n 。

该过程如式 (13) 所示。

$$P_n = \text{Softmax}(\text{Linear}(F_{fus})) \quad (13)$$

2 实验设置

2.1 实验环境

本文所有实验均基于 Ubuntu 18.04.6、NVIDIA GeForce RTX 3090、Intel (R) Xeon (R) Gold 6326 CPU、256GRAM 平台;运行环境基于 Python 3.11.7、PyTorch 2.1.2、Cuda 11.8。在训练过程中,批量大小设置为 128,采用 AdamW 优化器更新网络参数,初始学习率为 $1 \times$

10^{-4} ,权重衰减设置为 5×10^{-2} ,每次实验训练 200 个 epoch 取平均测试指标最高的模型保存权重参数以供测试。

2.2 数据集

为了验证所提出的模型,采用公开的多模态气体数据集 MultimodalGasData^[26]进行实验。

MultimodalGasData 数据集为烟雾泄漏检测任务提供了两种模态,收集了 4 种气体类型,即烟雾 (smoke),混合气体 (mixture),香水 (perfume) 和无气体 (nogas)。每种类型共 1 600 个样本,每个样本包括 7 种气体传感器 (MQ2、MQ3、MQ5、MQ6、MQ7、MQ8 和 MQ135) 收集到的标量气体数据和红外相机拍摄到的红外图像数据。

在实验中,对于红外图像数据,统一将图片大小设置为 224×224 。为防止过拟合,采用随机翻转和随机裁剪的方式增强红外图像数据集。对于传感器数据,按照 7 种气体传感器的顺序组织数据,并确保每一条传感器数据与对应的红外图像一一匹配。

此外,为了验证模型的鲁棒性,在确保种子相同的情况下在测试集中的红外图像中随机加入高斯噪声、椒盐噪声和线状噪声,并随机移除或修改传感器的某些数据。

表 1 为正常的传感器数据样本;表 2 为模拟噪声环境下的传感器数据样本;各标签的红外图像样本如图 5(a)~(d)所示,图 5A 为正常数据,图 5B 为模拟噪声环境下的数据。

表 1 传感器数据样本
Table 1 Sensor data sample

类型	MQ2	MQ3	MQ5	MQ6	MQ7	MQ8	MQ135
NoGas	785	530	448	442	570	672	516
Perfume	652	524	398	386	637	576	436
Smoke	561	353	313	355	576	555	285
Mixture	562	417	404	379	458	335	417

表 2 传感器噪声数据样本
Table 2 Sensor noise data sample

类型	MQ2	MQ3	MQ5	MQ6	MQ7	MQ8	MQ135
NoGas	1 774	0	1 703	1 888	570	672	516
Perfume	652	1 333	398	386	637	576	0
Smoke	0	353	0	888	314	555	285
Mixture	562	0	404	379	0	0	417

参考文献[17-18],两种模态的数据均按照 7:3 的比例随机划分为训练集和验证集,分别用于模型的训练和测试。数据集的分布情况如表 3 所示。

2.3 损失函数和评价指标

在训练时主要采用交叉熵为损失函数,损失函数如式(14)所示。

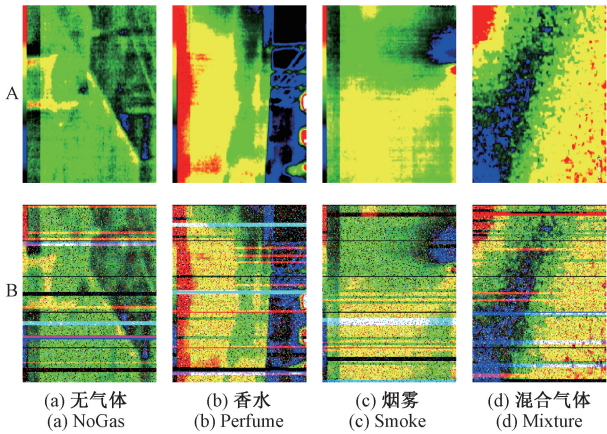


图 5 红外图像样本
Fig. 5 IR image sample

表 3 MultimodalGasData 数据集样本分布
Table 3 MultimodalGasData dataset sample distribution

类型	训练样本	测试样本	标签	总计
NoGas	1 120	480	0	1 600
Perfume	1 120	480	1	1 600
Smoke	1 120	480	2	1 600
Mixture	1 120	480	3	1 600
总计	4 480	1 920	-	6 400

$$Loss = \frac{1}{N} \sum_{i=1}^N CE(p_i, y_i)$$
 (14)

式中: y 为多标签的真实值; p_i 和 y_i 分别表示第 i 个标签的预测值和真实值; N 代表标签的数量; CE 代表交叉熵损失。

参考文献[17-18],主要采用准确率、精确率、召回率和 F1-得分作为评价指标

评价指标如式(15)~(18)所示。

准确率 = $\frac{TP + TN}{TN + FP + FN + TP}$ (15)

精确率 = $\frac{TP}{TP + FP}$ (16)

召回率 = $\frac{TP}{TP + FN}$ (17)

F1-得分 = $\frac{2 \times TP}{2 \times TP + FP + FN}$ (18)

式中: TP 代表被正确分类为正例的样本数量; FP 代表被错误分类为正例的负例样本数量; TN 代表被正确分类为负例的样本数量; FN 代表被错误分类为负例的正例样本数量。

3 实验分析

3.1 模型对比分析

为了验证本文模型的有效性,在确保配置相同的情

况下在 MultimodalGasData 数据集上进行了全面的实验,并对比了 CNN-LSTM^[17]、LDA-CNN^[18]、PvT-TabTransformer^[20-21]、ResNet50-BiLSTM^[27]和本文模型的性能。除了 LDA-CNN 的结果参考文献[18],其余模型的结果均为实验所得。

实验结果如表 4 所示,本文模型在准确率、召回率和 F1-得分上均达到了 97.92%,精确度为 97.93%。相比于文献[17-18]的方法,本文方法的性能提升约为 2%~5%。此外,与 ResNet50-BiLSTM 和 PvT-TabTransformer 模型相比,本文模型在准确率等指标上也展现出一定的优势。

表 4 对比实验结果				
Table 4	Comparative experimental results			
方法	准确率	精确率	召回率	F1-得分
CNN-LSTM ^[17]	95.21	95.25	95.21	95.19
LDA-CNN ^[18]	93.00	93.25	93.00	93.25
Resnet50-BiLSTM	97.29	97.32	97.29	97.29
Pvt-TabTransformer	97.76	97.79	97.76	97.76
本文	97.92	97.93	97.92	97.92

3.2 模型结构分析

为了探讨不同模型结构对于红外图像和传感器数据的影响,以准确率和精确率作为主要指标,分别对比了 LSTM、BiLSTM、Tab-Transformer 和 gMLP 在传感器数据和噪声数据上的表现。以及 CNN、ResNet50、PvT 和 Swin-Transformer 在红外图像数据和噪声数据上的表现。

单模态的传感器数据实验结果如图 6 (a) 和 (b) 所示。对于单模态的传感器数据,gMLP 分别达到了 91.35% 和 91.39% 的准确率和精确率,均优于其他模型。与 Tab-Transformer 相比,两个指标分别提高了 3.38% 和 3.34%。而传统的 LSTM 和 BiLSTM 在这两项指标上均略低。在噪声干扰的情况下,gMLP 依然保持了较高的稳定性,其准确率和精确率分别为 46.72% 和 48.48%,远超其他模型。这表明,gMLP 在处理此类传感器数据时具有一定的优势和鲁棒性。

单模态的红外图像实验结果如图 7(a) 和 (b) 所示。实验结果表明,对于单模态的红外图像数据,Swin-Transformer 取得了 96.61% 的准确率和 96.60% 的精确率,略高于 ResNet50 的 96.25% 和 96.33%。在噪声干扰的情况下,Swin-Transformer 依然展现出较强的鲁棒性,其准确率和精确率分别为 43.54% 和 70.95%,显著优于其他对比模型。实验表明,Swin-Transformer 在处理红外图像数据方面具有一定的优势和鲁棒性。

3.3 多模态融合分析

为了探讨多模态融合的深层意义,仅使用 gMLP 和 Swin-Transformer 在传感器数据集和红外图像数据集上进行实验;使用多模态融合的方式融合传感器数据和红

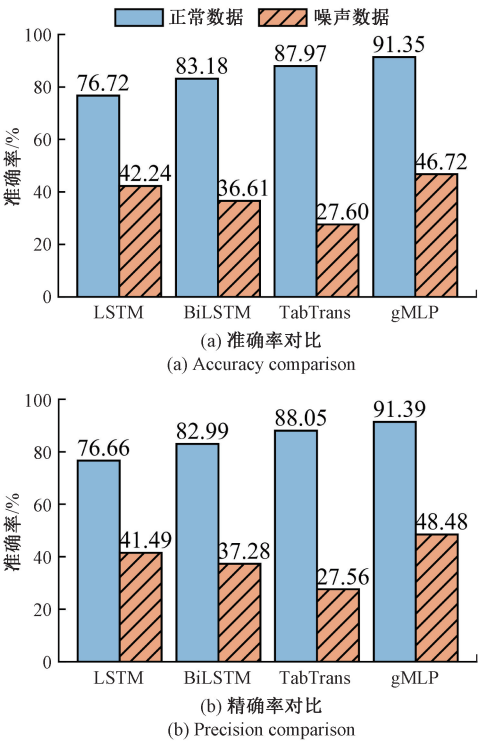


图 6 单模态传感器数据对比
Fig. 6 Single modal sensor data comparison

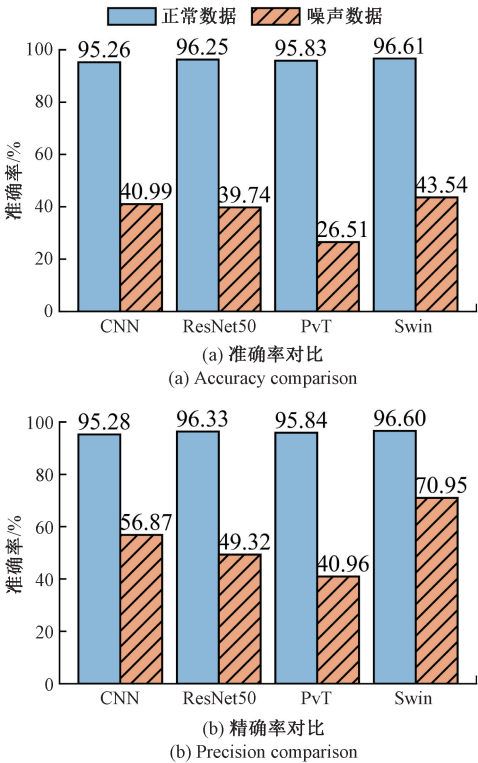


图 7 单模态红外图像对比
Fig. 7 Single modal IR image comparison

外图像进行实验,柱状图如图 8 所示,混淆矩阵和 t-SNE 图以探讨多模态融合所带来的影响。

由图 8(a)和(b)可以看出,在正常数据的环境下,多模态融合在准确率和精确率方面均有所提升,取得了 97.92%的准确率和 97.93%的精确率,优于单一模态下的表现。然而,在噪声环境下,准确率为 56.09% 优于前两种模态,但精确率为 61.71%与单一红外图像模态相比出现了一定程度的下降。

混淆矩阵如图 9 所示,可以看出,在正常环境下,如图 9(a)~(c)所示,多模态融合充分结合了两数据模态的优势,在各类标签的预测中均有所提升。然而,在噪声环境下,如图 9(d)~(f)所示,某些标签的预测性能受到了一定影响,如混合气体这类标签,其预测效果有所下降。

t-SNE 图如图 10 所示进一步揭示了多模态融合对特征分布的影响。在正常环境下,如图 10(a)~(c)所示,相较于单一模态的特征,多模态融合的特征分布更加集中,类别分离明显,各类别点云之间的间隔较大。然而,在噪声环境下,如图 10(d)~(f)所示,相较于单一模态,多模态融合虽能有效区分圆形、上三角、和下三角这 3 类特征,但对于方形特征,仍无法有效区分。

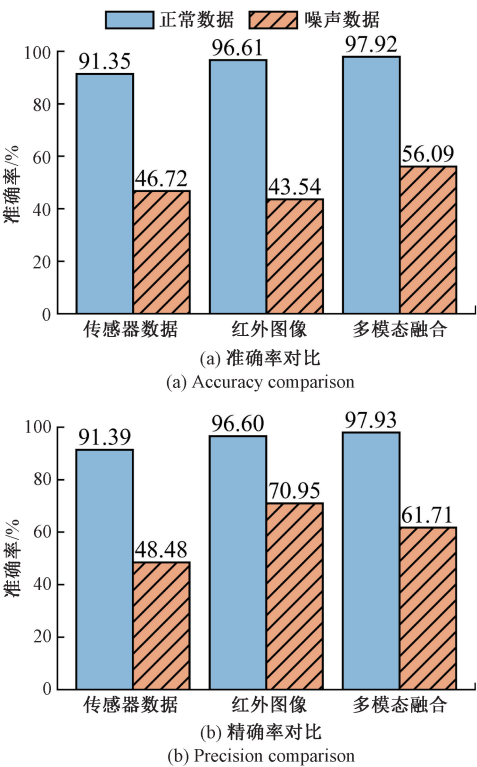


图 8 多模态融合分析柱状图

Fig. 8 Multi modal fusion analysis bar chart

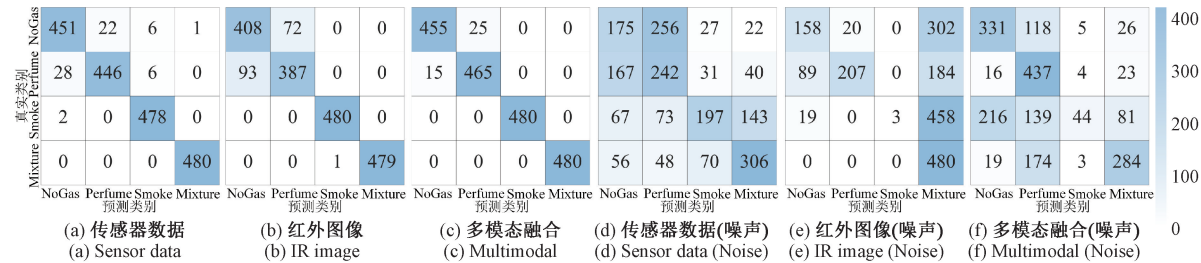


图 9 多模态融合分析混淆矩阵

Fig. 9 Multimodal fusion analysis confusion matrix

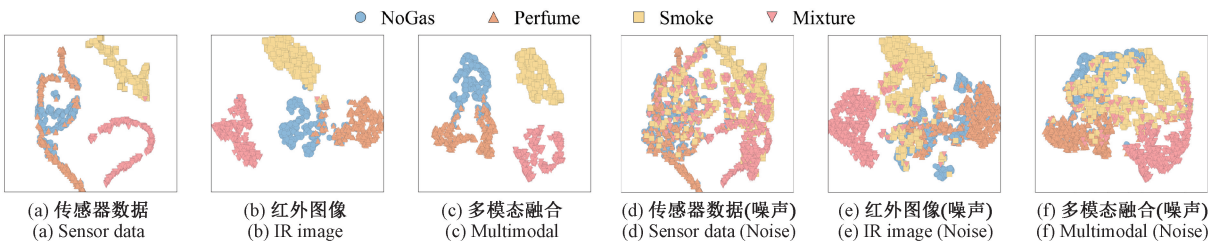


图 10 多模态融合分析 t-SNE 图

Fig. 10 Multimodal fusion analysis t-SNE graph

多模态融合实验分析的结果表明,在正常环境下,多模态融合能够提升准确率和类别区分能力;但在噪声较大的环境中,多模态融合虽在一定程度上增强了识别的

准确率和鲁棒性,但由于同时融合了来自两种不同模态数据的噪声,从而影响某类气体识别的性能。

4 结 论

本文从不同模态的气体数据特性和特征融合策略两个角度出发提出了一种多模态融合模型以解决复杂工业环境下的有害气体泄漏检测的问题。整体框架采用双分支编码器,以 gMLP 提取传感器数据特征, Swin-Transformer 提取红外图像特征,同时,利用基于多头注意力的融合方式融合这些特征表示。在正常环境和噪声环境下的实验结果表明,本文模型在准确率和鲁棒性方面表现优越。

但在实验过程中发现,在噪声干扰的环境下,多模态融合虽可以提高检测的准确性,但由于融合了来自不同模态数据的噪声,导致部分样本在识别时受到影响。未来的研究工作将致力于优化融合策略以抑制这些噪声,并采用更多数据集,进一步优化模型的泛化能力,探索其不同工业场景中的应用。

参考文献

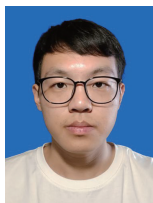
- [1] MERIBOUT M. Gas leak-detection and measurement systems: Prospects and future trends [J]. IEEE Transactions on Instrumentation and Measurement, 2021, 70: 1-13.
- [2] ADNAN N F, GHAZALI M F, AMIN M M, et al. Leak detection in gas pipeline by acoustic and signal processing-A review [C]. IOP Conference Series: Materials Science and Engineering, 2015, 100(1): 012013.
- [3] 宋勇, 苏海波, 董亚娟, 等. 基于变流量测试的天然气输送管道泄漏检测新方法[J]. 油气田地面工程, 2015, 34(7): 4-6.
SONG Y, SU H B, DONG Y J, et al. A new method for natural gas pipeline leak detection based on variable flow rate testing [J]. Oil-Gas Field Surface Engineering, 2015, 34(7): 4-6.
- [4] DOSHMANZIARI R, KHALOOZADEH H, NIKOOFARD A. Gas pipeline leakage detection based on sensor fusion under model-based fault detection framework[J]. Journal of Petroleum Science and Engineering, 2020, 184: 106581.
- [5] ABDULSHAHEED A, MUSTAPHA F, GHAVAMIAN A. A pressure-based method for monitoring leaks in a pipe distribution system: A Review[J]. Renewable and Sustainable Energy Reviews, 2017, 69: 902-911.
- [6] KOPBAYEV A, KHAN F, YANG M, et al. Gas leakage detection using spatial and temporal neural network model[J]. Process Safety and Environmental Protection, 2022, 160: 968-975.
- [7] YANG L, ZHAO Q. A BiLSTM based pipeline leak detection and disturbance assisted localization method[J]. IEEE Sensors Journal, 2021, 22(1): 611-620.
- [8] SALAMEH H A B, DHAINAT M F, BENKHELIFA E. An end-to-end early warning system based on wireless sensor network for gas leakage detection in industrial facilities[J]. IEEE Systems Journal, 2020, 15(4): 5135-5143.
- [9] 张立豪, 李鹏, 刘轩宇. 基于深度迁移学习的变工况气体泄漏检测[J]. 仪器仪表学报, 2023, 44(6): 177-187.
ZHANG L H, LI P, LIU X Y. Gas leak detection for variable conditions based on deep transfer learning[J]. Chinese Journal of Scientific Instrument, 2023, 44(6): 177-187.
- [10] MERIBOUT M, KHEZZAR L, AZZI A, et al. Leak detection systems in oil and gas fields: Present trends and future prospects [J]. Flow Measurement and Instrumentation, 2020, 75: 101772.
- [11] ALDHAFEERI T, TRAN M K, VROLYK R, et al. A review of methane gas detection sensors: Recent developments and future perspectives [J]. Inventions, 2020, 5(3): 28.
- [12] 彭道刚, 刘薇薇, 威尔江, 等. 基于 CBAM-Res_UNet 电厂高压蒸汽泄漏检测研究[J]. 电子测量与仪器学报, 2021, 35(12): 206-214.
PENG D G, LIU W W, QI ER J, et al. Research on high-pressure steam leakage detection in power plants based on CBAM-Res_UNet [J]. Journal of Electronic Measurement and Instrumentation, 2021, 35(12): 206-214.
- [13] 彭敏, 马宁, 王少军, 等. 基于可见光图像的工业过程气体泄漏检测研究[J]. 电子测量与仪器学报, 2019, 33(6): 1-8.
PENG M, MA N, WANG SH J, et al. Gas leakage detection of industrial process based on visible image[J]. Journal of Electronic Measurement and Instrumentation, 2019, 33(6): 1-8.
- [14] 李泉成, 曹江涛, 姬晓飞. 光流增强的红外成像气体泄漏检测方法[J]. 电子测量与仪器学报, 2023, 37(3): 50-56.
LI Q CH, CAO J T, JI X F. Gas leak detection method enhanced by optical flow in infrared imaging[J]. Journal of Electronic Measurement and Instrumentation, 2023, 37(3): 50-56.
- [15] WANG J, TCHAPMI L P, RAVIKUMAR A P, et al. Machine vision for natural gas methane emissions detection using an infrared camera[J]. Applied Energy, 2020, 257: 113998.

- [16] GIBSON G M, SUN B, EDGAR M P, et al. Real-time imaging of methane gas leaks using a single-pixel camera[J]. Optics Express, 2017, 25(4): 2998-3005.
- [17] NARKHEDE P, WALAMBE R, MANDAOKAR S, et al. Gas detection and identification using multimodal artificial intelligence based sensor fusion[J]. Applied System Innovation, 2021, 4(1): 3.
- [18] FALEH R, KACHOURI A. A hybrid deep convolutional neural network-based electronic nose for pollution detection purposes[J]. Chemometrics and Intelligent Laboratory Systems, 2023, 237: 104825.
- [19] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]. Advances in Neural Information Processing Systems, 2017: 5998-6008.
- [20] WANG W H, XIE EN Z, LI X, et al. Pyramid vision transformer: A versatile backbone for dense prediction without convolutions[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 568-578.
- [21] HUANG X, KHETAN A, CVITKOVIC M, et al. Tabtransformer: Tabular data modeling using contextual embeddings[J]. ArXiv preprint arXiv:2012.06678, 2020.
- [22] 唐永旺, 刘会景, 靳彦青, 等. 基于 Transformer 的时序多模态融合特征的异常网络流量检测方法[J]. 信息工程大学学报, 2023, 24(4): 468-474.
TANG Y W, LIU H J, JIN Y Q, et al. Network traffic anomaly detection based on transformer and sequential-multimodal fusion feature[J]. Journal of Information Engineering University, 2023, 24(4): 468-474.
- [23] ZHANG Y L, XIE F Y, CHEN J Q. TFormer: A throughout fusion transformer for multi-modal skin lesion diagnosis[J]. Computers in Biology and Medicine, 2023, 157: 106712.
- [24] LIU Z, LIN Y T, CAO Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 10012-10022.
- [25] LIU H X, DAI Z H, SO D, et al. Pay attention to mlp[C]. Advances in Neural Information Processing Systems, 2021, 34: 9204-9215.
- [26] NARKHEDE P, WALAMBE R, CHANDEL P, et al.

MultimodalGasData: Multimodal dataset for gas detection and classification[J]. Data, 2022, 7(8): 112.

- [27] HE K M, ZHANG X Y, REN SH Q, et al. Deep residual learning for image recognition[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770-778.

作者简介



王泓森, 2023 年于郑州工程技术学院获得学士学位, 现为昆明理工大学硕士研究生, 主要研究方向为计算机视觉。

E-mail: 519235199@qq.com

Wang Hongsen received his B. Sc. degree from Zhengzhou University of Technology in 2023. He is currently a M. Sc. candidate at Kunming University of Science and Technology. His main research interest includes computer vision.



王建国, 博士研究生, 现为中国铜业有限公司高级工程师, 主要研究方向为企业信息化、智能建造。

E-mail: wangjg@chnccopper.com

Wang Jianguo a Ph. D. candidate. He is currently a Senior Engineer at China Copper Corporation Limited. His main research interests include enterprise informatization and intelligent construction.



杨建东, 现为昆明信息港传媒有限责任公司高级工程师, 主要研究方向为网络安全。

E-mail: 429699197@qq.com

Yang Jiandong is currently a Senior Engineer at Kunming Information Hub Media Co., Ltd. His main research interest includes network security.



冯勇(通信作者), 2011 年于电子科技大学获得博士学位, 现为昆明理工大学教授、博士生导师, 主要研究方向为物联网、区块链。

E-mail: fybraver@163.com

Feng Yong (Corresponding author) received his Ph. D. degree from the University of Electronic Science and Technology in 2011. He is currently a Professor and Ph. D. supervisor at Kunming University of Electronic Science and Technology of China. His main research interests include the Internet of Things and blockchain.