



DOI:10.12404/j.issn.1671-1815.2405433

引用格式:章璇,周正康,唐加山.基于注意力和域泛化的异常声音检测算法[J].科学技术与工程,2025,25(15):6310-6317.

Zhang Xuan, Zhou Zhengkang, Tang Jiashan. An anomalous sound detection algorithm based on attention and domain generalization[J]. Science Technology and Engineering, 2025, 25(15): 6310-6317.

机械、仪表工业

基于注意力和域泛化的异常声音检测算法

章璇¹, 周正康², 唐加山^{1*}

(1. 南京邮电大学理学院, 南京 210023; 2. 南京城建隧桥智慧管理有限公司, 南京 211800)

摘要 为了更准确地识别机械设备中的正常和异常声音,提出了一种融合注意力机制和域泛化技术的异常声音检测算法。具体来说,通过联合训练两个神经网络,结合使用子族群余弦自适应损失函数 Sub-Cluster AdaCos (SCAdaCos),并采用高斯混合模型对声音特征进行建模。异常评分使用负对数似然值计算,并通过 90% 分位数设定检测阈值。实验结果显示,该算法在风扇、轴承等 7 类机器上,ROC (receiver operating characteristic curve) 曲线下的面积 AUC (area under curve) 和部分曲线下面积 pAUC (the partial AUC) 的调和平均值分别为 76.69% 和 87.99%,其中阀门数据表现最佳。相比于两种基线系统,AUC 和 pAUC 指标分别提高了 24.08% 和 20.68%。进一步的消融实验表明,高斯混合模型、注意力机制以及 SCAdaCos 损失函数均对性能提升有显著贡献。与同一数据集上的其他 8 种算法相比,新算法在 AUC 和 pAUC 的调和平均值上提升了 4.16%,展现了其在异常声音检测任务中的显著优势。

关键词 声音异常检测; 无监督; 域泛化; 注意力; 深度学习

中图分类号 TH17 TP18; **文献标志码** A

An Anomalous Sound Detection Algorithm Based on Attention and Domain Generalization

ZHANG Xuan¹, ZHOU Zheng-kang², TANG Jia-shan^{1*}

(1. College of Science, Nanjing University of Posts and Telecommunications, Nanjing 210023;

2. Nanjing Urban Construction Tunnel & Bridge Intelligent Management Co., Ltd., Nanjing 211800, China)

[Abstract] A new anomaly sound detection algorithm was studied that combines attention mechanisms and domain generalization techniques to more accurately identify normal and abnormal sounds in mechanical equipment. Specifically, two neural networks were jointly trained using a sub-cluster Adacos loss function, with features modeled by a Gaussian mixture model (GMM). Anomaly scores were calculated using negative log-likelihood values, and a 90th percentile threshold was set for detection. The algorithm demonstrated strong performance across seven types of machines, including fans and bearings, achieving harmonic mean AUC (area under curve) and pAUC values of 76.69% and 87.99%, with the highest performance observed on valve data. Compared to two baseline systems, the algorithm improved AUC and pAUC by 24.08% and 20.68%, respectively. Ablation studies further confirmed the positive impact of the GMM, attention mechanism, and Scadacos loss function. When tested against eight other algorithms on the same dataset, the proposed method showed a 4.16% improvement in the harmonic mean of AUC and pAUC, highlighting its significant advantage in anomaly sound detection tasks.

[Keywords] anomaly sound detection; unsupervised learning; domain generalization; attention mechanism; deep learning

异常声音检测 (anomalous sound detection, ASD) 是工业自动化和智能制造领域的关键技术。在智能制造中,此技术显得尤为重要,因为它能实时监测和诊断生产线上机器的健康状态,及时发现潜在的机械故障^[1-3]。这不仅提高了生产效率和产

品质量,还能显著减少因故障导致的生产停滞和维修成本,从而减少对人工检查的依赖并保障生产流程的顺畅。然而 ASD 的应用范围并不局限于机器故障诊断,在更广泛的智能制造过程中,ASD 还被用于产品质量控制^[4],通过监听生产过程中的声音

收稿日期:2024-07-19 修订日期:2024-11-24

基金项目:南京邮电大学横向科研项目(2023 外 221)

第一作者:章璇(2000—),男,汉族,江苏南京人,硕士研究生。研究方向:无监督学习,声音异常检测。E-mail:1222087506@njupt.edu.cn。

* 通信作者:唐加山(1968—),男,汉族,安徽天长人,博士,教授。研究方向:随机模型,数据科学。E-mail:tangjs@njupt.edu.cn。

来识别不当操作或材料缺陷,以及安全监控,通过识别设备运行中的异常声响来预防可能的安全事故。因此,ASD技术不仅是自动化和智能制造领域的基础,更是提升整个制造业智能化水平的关键因素,这种无侵入式的监控^[5-7]有着极大的优势。

然而,异常声音检测面临的一大挑战是如何准确识别未知的异常声音^[8]。由于实际工作环境中异常事件的罕见性和复杂性^[9],仅依靠有限的正常声音样本来训练模型难以涵盖所有潜在的异常情况^[10]。这一问题的复杂性在于,必须让检测算法能够识别训练阶段未曾见过的异常声音类型。此外,工业应用中存在的域偏移问题,比如机器的运行速度、负荷条件、环境噪声等因素的变化,都可能导致模型的性能受到影响,从而增加误报的风险。域泛化技术应运而生,旨在通过捕获跨不同环境和条件下的共性特征,使模型具备更好的泛化能力,以适应测试阶段可能遇到的各种未见过的情况,而无需针对每一种特定情境重新调整或训练模型^[11-12]。

近年来,对于无监督机器异常声音的研究已经吸引了大量学者的关注和投入。如张晨旭等^[13]利用自编码器的方法对声音进行解码和编码。薛英杰等^[14]利用生成对抗网络的方法,相对于传统的自编码器在对声音进行重构方面做出了改进。仇睿等^[15]提出了一种基于频谱感知的去噪方法,以对声音的降噪为出发点基于MobileNet V2架构进行异常检测。张勇等^[16]利用卷积神经网络与高斯混合模型对水电机组的声音进行特征提取和异常分类。姜慧天^[17]利用了改进的残差卷积神经网络,对原始网络中的全连接网络结构进行了改进。

然而,上述的研究都是以无监督为重点,关于机器声音域转移和域泛化的问题并没有得到解决。随着对这个问题的关注,如Kuroyanagi等^[18]提出了

一种两阶段的ASD系统,由基于异常值暴露的特征提取器和基于内部模型的异常检测器串联组成。由于集成度和复杂度较高,模型的性能可能高度依赖于训练数据的质量和数量,而且在处理新的或不同类型的异常时可能表现不佳。王超等^[19]提出一种基于注意力-跳跃自编码器-生成对抗网络的无监督异常声音检测方法ASAE-GAN(attentional skip-connected auto encoder and generative adversarial network)并引入通道间注意力机制和时间注意力机制进行改进,然而在部分型号的机器上的结果指标提升并不大。章林柯等^[20]在Resnet的基础上结合了坐标注意力机制模块和域混合的方法。毕忠勤等^[21]在Wilkinghoff^[22]提出的双网络特征嵌入的基础上将分类器由高斯混合模型改成了贝叶斯高斯混合模型,但是经过实验贝叶斯高斯混合模型在计算复杂度和时间方面存在较大的弊端,由于先验分布的存在导致模型有着超参数敏感性,在数据集进行推广时准确率会大幅下降。

针对上述问题,提出一种基于注意力机制和域泛化的异常声音检测算法,通过两个网络对原始音频特征进行学习和特征连接,使用了一种子族群损失函数,利用高斯混合模型对特征进行建模,采用了负对数似然函数值来表示异常得分,使用90%分位数作为阈值,对异得分超过阈值的声音判定为异常,并在AUC(area under curve)、pAUC(the partial AUC)评价指标方面对方法进行了评价;还做大量消融实验并与采用相同数据集的算法进行对比。

1 算法模型

本文中采用的模型网络结构由数据预处理、特征提取、计算异常得分并判定异常三个部分组成,如图1所示。

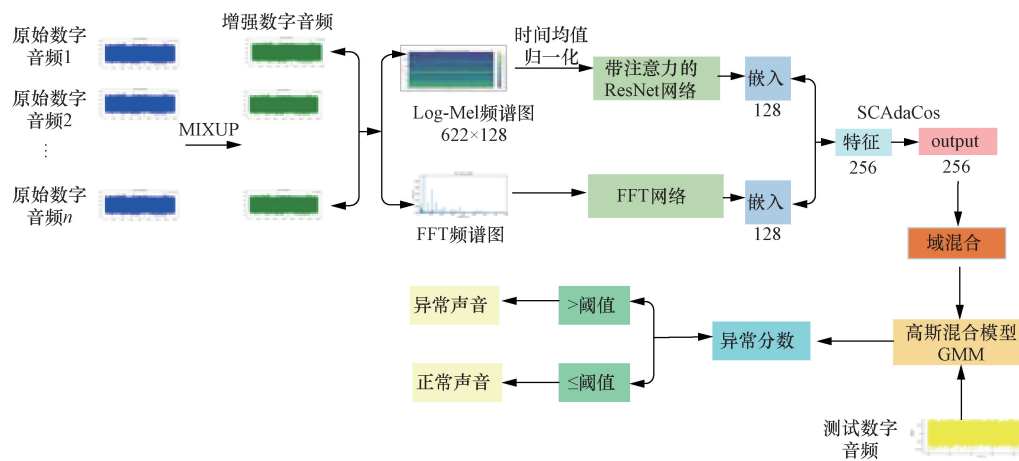


图1 模型结构

Fig. 1 Model structure

1.1 数据预处理

1.1.1 音频信号转换为离散数据

对于一个长度为 10 s, 采样率为 16 000 Hz 的音频, 可以得到一个包含 160 000 个样本点的一维向量。这是因为采样率定义了每秒中采样的次数, 所以在 10 s 内, 就会有 $10 \text{ s} \times 16 \text{ 000 次/s} = 160 \text{ 000}$ 次采样。每个样本点代表了音频信号在特定时间点的幅度, 本文中采用 160 000 Hz 的高采样频率可以捕捉音频信号中非常细微的变化, 从而以极高的保真度重构原始音频信号。这个一维向量将数据从音频转化为了数据的形式, 可以被用于进一步的数据处理, 特征提取或者作为深度学习模型的输入。

1.1.2 数据增强

为了进一步提高模型的泛化能力, 引入样本多样性, 对于向量形式的音频数据进行 MIXUP^[23] 数据增强处理, 公式为

$$\begin{cases} x' = \lambda x_i + (1 - \lambda)x_j \\ y' = \lambda y_i + (1 - \lambda)y_j \end{cases} \quad (1)$$

式(1)中: λ 为从 Beta 分布中采样得到的系数。本文中采用均匀分布, 因此 λ 的取值在 0 ~ 1。本文将每个批次的数据进行逆序排序之后和原顺序进行混合, 如第一个音频数据和最后一个音频数据进行数据增强处理。通过这种方式, 可以确保在进行数据和标签的混合时, 每个样本都与一个不同的样本进行组合, 增加数据的多样性。

1.2 特征提取

用于提取嵌入的神经网络由两个不同子网络 [对数梅尔 (Log-Mel) 网络和快速傅里叶变换 (fast Fourier transform, FFT) 网络] 组成, 并经过训练以联合区分机器类型、机器 ID 和机器的不同属性信息, 从而产生总共 342 个类别。

1.2.1 Log-Mel 网络

Log-Mel 网络结构如图 2 所示, 目的是将音频数据转换为对数梅尔频谱 (Log-Mel spectrogram)。梅尔频谱是音频信号的一种表示, 它考虑了人类听觉的非线性感知特性。具体来说, 梅尔刻度是一个感知刻度, 用来量化声音的感知音高。低频段的部分, 梅尔刻度和正常频率几乎呈线性关系, 而在高频段, 因为人耳的感知变弱, 因此两者呈对数关系。转换为梅尔刻度的频谱能更好地匹配人耳对不同频率的敏感度。梅尔刻度 m 与频率 f 的转换公式为

$$m = 2595 \lg \left(1 + \frac{f}{700} \right) \quad (2)$$

$$f = 700 (10^{m/2595} - 1) \quad (3)$$

将音频转化为对数梅尔频谱具体过程如下:

①首先对信号进行短时傅里叶变换 (short-time Fourier

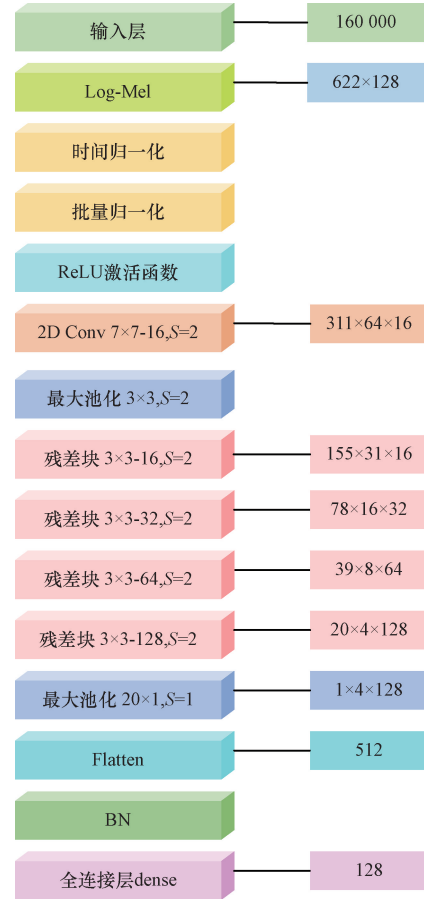


图2 Log-Mel 网络结构

Fig. 2 Log-Mel network structure

transform, STFT), 通过将音频信号分割成小段 (每段被称为帧), 每一段通过傅里叶变换来计算其频谱; ②从 STFT 结果中计算每一帧的功率谱, 计算功率谱是为了获取信号在不同频率上的能量信息; ③应用梅尔滤波器组; ④对每个滤波器的输出取对数, 得到对数梅尔频谱。计算公式分别为

$$X(k, \omega) = \sum_{n=-\infty}^{\infty} x(n)w(n-k)e^{-j\omega n} \quad (4)$$

$$P(k, \omega) = |X(k, \omega)|^2 \quad (5)$$

$$L(k, m) = \lg[S(k, m)] \quad (6)$$

式中: $x(n)$ 为原始音频信号在时间点 n 处的采样值; $w(n-k)$ 为窗函数用于加窗处理; $e^{-j\omega n}$ 为复指数函数用于时频转换; k 为时间帧索引; ω 为频率变量; $P(k, \omega)$ 为在时间帧 k 频率 ω 处的信号能量; $S(k, m)$ 为通过梅尔滤波器组处理后的频谱能量; k 和 ω 分别为时间帧和滤波器的索引。

原始音频长度为 160 000, 通过傅里叶变换窗口大小为 1 024, 帧移为 256 进行转换, 由时间帧数 = (音频长度 - 窗口大小) / 步长 + 1 = 622, 可以得到 622×128 维度的特征。其中时间维度大小是 622, 频率维度大小是 128 (梅尔频带的数量)。

神经网络采用的是 Resnet-18 的架构,精心调整了各参数以达到较好效果。在每个残差块内的残差连接之前添加了通道注意力机制,如图 3 所示^[24],其中 C 为通道、 H 为特征图的高度、 W 为宽度、 r 为通道数的缩减比例。

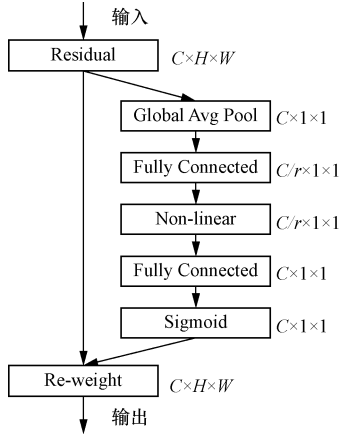


图 3 CA 注意力

Fig. 3 CA attention

1.2.2 FFT 网络

FFT 网络如图 4 所示,主体结构由 3 个 1D 卷积层组成。FFT 是离散傅里叶变换 (discrete Fourier transform, DFT) 的一种算法,其核心思想是分治策略,通过递归地将 DFT 分解为更小的 DFT 来减少计算复杂度。其中最著名的 FFT 算法是 Cooley-Tukey 算法,它的基本思想是如果 N 是一个偶数,则 DFT 可以分解为两个长度为 $N/2$ 的 DFT。传统的 DFT 计算复杂度约为 $O(N^2)$,使用 FFT 算法后,计算复杂度可以降低到 $O(N \lg N)$ 。DFT 的分解公式为

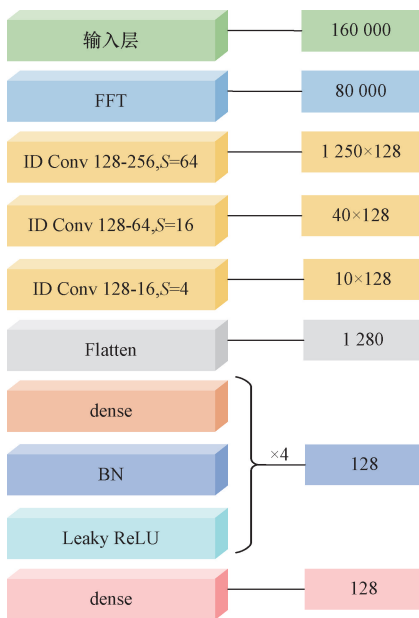


图 4 FFT 网络结构

Fig. 4 FFT network structure

$$X(k) = X_{\text{even}}(k) + e^{-\frac{j2\pi k}{N}} X_{\text{odd}}(k) \quad (7)$$

式(7)中: $X(k)$ 为第 k 个频率分量的频谱值; $X_{\text{even}}(k)$ 和 $X_{\text{odd}}(k)$ 分别为输入信号中偶数索引点和奇数索引点的 FFT 结果; $e^{-\frac{j2\pi k}{N}}$ 为旋转因子用于调整奇数索引点的结果; k 为频率索引; N 为信号的总长度。

将 2 个网络各自得到的 128 维度的嵌入进行合并连接得到 256 维度的特征,再经过一种自适应余弦距离损失函数 (SCAdaCos)^[25] 将输入向量转换为每个类别的概率分布,进而提高分类性能,得到最终的特征向量。

1.3 计算异常分数

1.3.1 域混合

为了提升模型的域泛化能力,结合来自不同域(源域和目标域)的数据样本来生成新的训练样本,公式为

$$x_{\text{mixed}} = \alpha x_{\text{source}} + (1 - \alpha) x_{\text{target}} \quad (8)$$

式(8)中: x_{mixed} 为混合后的样本; x_{source} 为从源域选择的样本; x_{target} 为从目标域随机选择的相对应类别的样本; α 为混合系数,用于控制从源域和目标域中混合样本的比例。 α 设定为 0.5,表示等比例混合源域和目标域的样本。

通过这种方式,模型能够学习到更加泛化的特征表示,从而提高在不同数据分布下的性能和鲁棒性。这种方法尤其适用于那些源域和目标域数据分布存在差异的情况,帮助模型更好地适应目标域的数据分布。

1.3.2 高斯混合模型

高斯混合模型 (Gaussian mixed model, GMM) 假设数据是由多个高斯分布的混合生成的。一个 GMM 可以表示为

$$p(x) = \sum_{k=1}^K \pi_k N(x | \mu_k, \Sigma_k) \quad (9)$$

式(9)中: x 为数据点; K 为高斯分布的数目; N 为正态分布; π_k 为第 k 个高斯分布在混合中的权重(满足 $\sum_{k=1}^K \pi_k = 1, \pi_k \geq 0$); μ_k 为第 k 个高斯分布的均值; Σ_k 为第 k 个高斯分布的协方差矩阵。

通过 GMM 的训练,每个音频最终可以得出属于每一类的概率以及对数似然值。这个值越小,表示样本与模型的适应度越差,因而可以被认为是越“异常”的。以负对数似然值为异常分数,取其 90% 分位数为阈值,高于这个值即判定为异常。

2 实验

2.1 数据集

本文中采用的数据集来自于 DCASE2022-Task2

的开发数据集,共包含7种不同的机器,每种机器由3个不同的部分组成,每条音频长度为10 s,采样率为160 000 Hz,单声道保存。7种机器的数量相同,如表1所示。

表1 开发集数据构成
Table 1 Dataset composition

部分	训练集正常 音频数 (源域和目标域)	测试集正常 音频数 (源域和目标域)	测试集异常 音频数 (源域和目标域)
Section 00	990 + 10	50 + 50	50 + 50
Section 01	990 + 10	50 + 50	50 + 50
Section 02	990 + 10	50 + 50	50 + 50

2.2 消融实验

为了证实各模块的有效性,进行消融实验。以Tensorflow2.6.0为基础框架,在RTX4090(双卡)上进行,采用了文献[26]中的集成策略。更具体的说,系统经过4次训练,每次训练4 × 100个epoch,总共获得了4 × 4 = 16个不同的模型和相应的嵌入集。

评价指标:实验采用ROC(receiver operating characteristic curve)曲线下的面积AUC,部分ROC曲线下的面积pAUC以及它们的调和平均来评估声音异常检测系统的性能,计算公式为

$$AUC_{m,n,d} = \frac{1}{N_d^- N_n^+} \sum_{i=1}^{N_d^-} \sum_{j=1}^{N_n^+} H[A_\theta(x_j^+) - A_\theta(x_i^-)] \quad (10)$$

$$pAUC_{m,n} = \frac{1}{\lfloor pN_n^- \rfloor N_n^+} \sum_{i=1}^{\lfloor pN_n^- \rfloor} \sum_{j=1}^{N_n^+} H[A_\theta(x_j^+) - A_\theta(x_i^-)] \quad (11)$$

式中: m 表示机器类型的索引; n 代表部分(section)的索引; d 表示域,可能是源域或目标域,表示为{source, target}; N_d^- 为特定域 d 中正常测试剪辑的数量; N_n^+ 为部分 n 中异常测试剪辑的数量;函数 $H(x)$ 为阶跃函数,当 $x > 0$ 时返回1,否则返回0; $A_\theta(x_j^+)$ 和 $A_\theta(x_i^-)$ 分别为模型在参数 θ 下对异常剪辑 x_j^+ 和正常剪辑 x_i^- 的异常分数; $AUC_{m,n,d}$ 为特定机器类型 m 、部分 n 和域 d 下模型在区分正常和异常声音剪辑方面的性能; $pAUC_{m,n}$ 则关注在预先指定的假正率(false positive rate, FPR)范围上限 $p = 0.1$ 下,部分 n 中的低FPR条件性能,特别强调在低误报率下的真正率提升,其中 $\lfloor \cdot \rfloor$ 为向下取整函数。

表2为本文多种消融实验的结果,首先进行一项实验,以比较角度边缘损失函数ArcFace与SCAdaCos之间的性能差异,结果验证了SCAdaCos策略的有效

表2 消融实验结果

Table 2 Ablation experiment results

机器	指标	Base-AE	Base-MobV2	GMM + ArcFace	GMM + Scadacos	VBGMM + Scadacos	GMM + Scadacos + SE	GMM + Scadacos + CA	GMM + Scadacos + CBAM
轴承	AUC-S	54.42	60.55	59.36	66.08	71.24	73.85	72.44	75.29
	AUC-T	58.38	60.09	67.15	73.26	74.82	76.06	76.89	78.79
	pAUC	51.98	56.96	54.65	60.66	60.46	63.62	64.37	64.69
风扇	AUC-S	78.59	70.75	93.61	89.98	91.18	95.07	96.41	94.44
	AUC-T	47.18	48.22	75.78	77.37	80.00	84.67	82.74	78.28
	pAUC	57.52	56.94	68.08	73.45	72.43	79.83	78.68	73.22
变速器	AUC-S	68.93	69.19	83.34	83.15	82.95	82.75	84.89	89.06
	AUC-T	62.64	56.23	77.99	82.54	79.01	83.60	84.15	84.19
	pAUC	58.49	56.07	61.74	65.79	65.32	69.45	67.62	68.38
滑轨	AUC-S	77.95	65.05	92.79	92.93	93.74	93.11	93.66	93.39
	AUC-T	47.67	38.40	76.91	84.90	83.89	87.55	88.28	89.35
	pAUC	55.78	54.73	65.49	72.70	71.51	81.28	77.08	80.12
玩具汽车	AUC-S	90.41	58.92	86.35	82.62	86.09	80.66	85.46	83.64
	AUC-T	34.81	51.95	70.75	69.61	66.67	77.17	75.00	71.52
	pAUC	52.74	52.36	58.37	58.58	57.13	58.15	61.33	59.78
玩具火车	AUC-S	76.32	57.57	82.23	86.45	85.87	85.43	85.82	86.14
	AUC-T	23.35	45.79	60.72	63.82	62.74	63.07	61.82	62.08
	pAUC	50.48	51.52	55.69	56.14	53.21	53.95	54.53	55.02
阀门	AUC-S	52.01	67.66	81.63	92.25	92.20	95.04	95.11	94.35
	AUC-T	49.46	57.75	72.71	69.15	62.48	76.98	77.95	78.59
	pAUC	50.36	62.64	64.16	70.34	69.50	78.86	83.00	78.90
调和平均		52.61	56.01	70.07	73.26	72.53	76.36	76.69	76.33

注:AUC-S表明源域AUC,AUC-T表明目标域AUC,pAUC不区分源域和目标域。

性。随后,实验通过将分类器从高斯混合模型(GMM)替换为变分贝叶斯高斯混合模型(variational Bayesian Gaussian mixture model, VBGMM)来探究其对模型性能的影响。不论是在训练时间还是在接收操作特性曲线下面积(AUC)的指标上,VBGMM均未表现出预期的改善,反而展现出较差的性能。进一步地,引入三种不同的注意力机制以评估其对声音异常检测性能的潜在提升。结果表明,所有引入的注意力机制均有效地增强了模型的检测能力,其中通道注意力(coordattention, CA)机制表现最为突出,相较于原始方法,AUC提高了3.43个百分点。与基线系统自编码器(AutoEncoder, AE)和轻量化网络 MobileNetV2 (MobV2)相比,CA 机制分别实现了 24.08% 和 20.68% 的 AUC 提升,从而验证了注意力机制在声音异常检测应用中的有效性和适用性。

表3显示了在最优方法即使用CA注意力搭配Scadacos损失函数和高斯混合模型的算法下,7种机器不同部分的指标。可以看出模型对于不同的部分训练的效果并不都是一致的,在阀门这种机器上的表现最好,在00部分源域AUC达到了99.36,01部分目标域AUC和pAUC达到了百分之百的理想效果,在02部分源域AUC和pAUC分别达到了99.68和98.32,如图5所示条形图。这个结果表明在阀门这个机器上模型的训练效果较好,尤其是在00和02部分的源域以及01部分的目标域上,测试的 $3 \times 100 = 300$ 个数据几乎都正确的判定了。

表3 使用CA注意力的7种机器不同部分的评价指标

Table 3 Evaluation metrics for different parts of 7 machines using CA attention

机器不同部分	AUC	pAUC	AUC (源域)	pAUC (源域)	AUC (目标域)	pAUC (目标域)
轴承_00	59.55	51.68	56.48	60.42	64.64	53.05
轴承_01	86.67	76.74	80.12	67.37	91.92	86.53
轴承_02	83.97	71.05	88.24	75.16	79.48	67.37
风扇_00	97.87	89.89	99.44	97.05	94.40	83.37
风扇_01	87.56	71.11	95.20	92.42	72.00	50.53
风扇_02	89.99	77.00	94.96	89.47	84.76	72.00
变速器_00	87.08	76.89	89.88	81.05	84.48	75.37
变速器_01	80.30	58.89	80.88	66.53	80.08	50.95
变速器_02	82.98	69.26	85.20	77.89	87.20	63.79
滑轨_00	92.20	72.63	97.08	94.95	92.32	68.84
滑轨_01	93.99	80.21	99.76	98.74	93.48	79.16
滑轨_02	82.49	76.42	85.00	82.95	78.96	67.58
玩具汽车_00	73.34	58.21	80.76	65.89	71.92	55.79
玩具汽车_01	74.96	51.84	79.04	50.74	67.20	53.89
玩具汽车_02	94.61	80.74	98.20	92.84	91.68	73.89
玩具火车_00	60.43	52.68	71.08	61.47	56.40	53.05
玩具火车_01	69.67	52.16	92.52	83.79	54.92	51.58
玩具火车_02	80.65	58.95	99.72	98.95	79.88	61.68
阀门_00	93.50	80.58	99.36	96.63	85.88	73.47
阀门_01	91.65	88.05	87.16	75.58	100.00	100.00
阀门_02	86.79	80.42	99.68	98.32	59.00	71.58

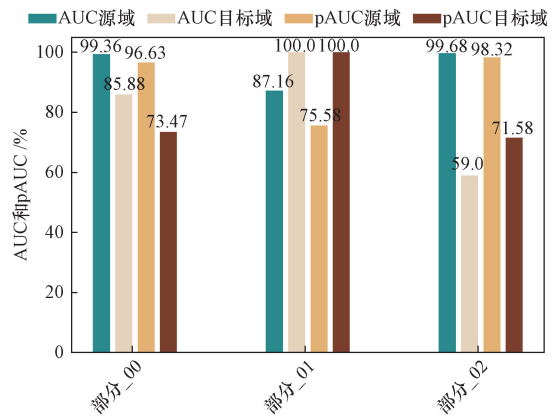


图5 阀门不同部分的指标

Fig. 5 Metrics for different sections of the valve

2.3 对比试验

为了进一步验证本方法的有效性,本文中2022年声学挑战赛中排名前列的技术报告[27-34]的方法进行了对比,结果证实了本方法的优越性。图6为7种机器AUC和pAUC的调和平均前5方法AUC和pAUC的雷达图,图7为调和平均的条形图。

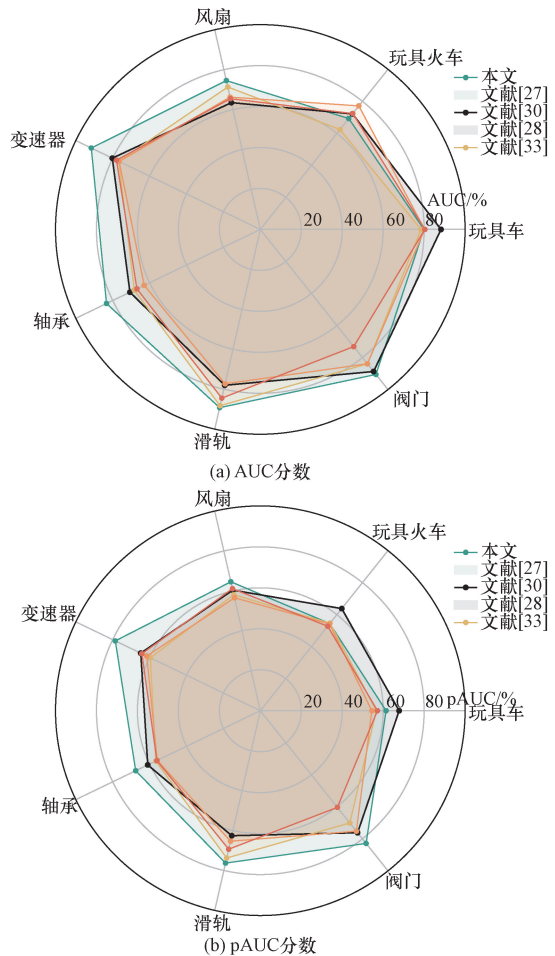


图6 前5名方法的AUC和PAUC雷达图

Fig. 6 Radar chart of AUC and pAUC for the Top 5 methods

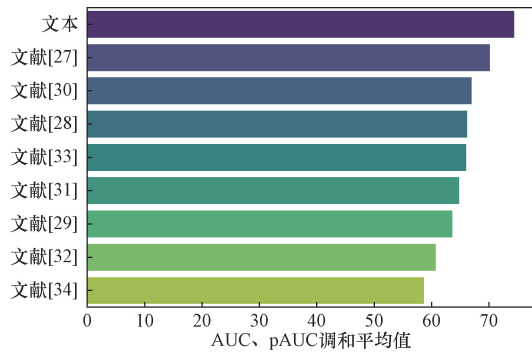


图7 所有机器 AUC 和 pAUC 调和平均值的条形图

Fig.7 Bar chart of the harmonic mean of AUC and pAUC for all machines

从雷达图(图6)和条形图(图7)的对比分析中可以明显看出,本文提出的方法在性能上较之于对比论文中的算法表现更加优越。这一优势具体体现在五种不同类型的机器设备上,包括风扇、变速器、轴承、滑轨和阀门。在这些设备的测试结果中,本方法均显著领先于其他对比方法。特别是在综合指标 AUC 和 pAUC 的总调和平均值上,本文所提出的方法相较于文献中最优方法的性能提升达到了 4.16%。

这一显著的性能提升不仅验证了本方法在单一机器类型上的有效性,而且也展示了其在不同机器类型上的泛化能力。此外,不同机器上性能的差异揭示了本文方法在模型稳定性和预测准确性方面的潜在优化空间,为未来在类似应用领域的研究提供了有力的理论和实践基础。总体来看,这些结果清楚地证明了本文提出的方法在异常声音检测领域的应用潜力和实际价值。

3 结论

提出了一种基于域泛化的声音异常检测系统,首先对数据进行 Mixup 增强,经过带注意力的 LOG-MEL 网络和 FFT 网络,再经由自适应余弦距离损失函数(Scadacos)得到最后的 256 维度的特征向量,将训练集域混合后由高斯混合模型计算异常分数从而进行分类。实验证明了本文方法在域泛化中的有效性,除去网络结构的设计之外,同时揭示了注意力机制在声音异常检测方面的适用性。结果证明本文中提出的系统优于大部分已经提出的系统,在未来的工作中可以得到启示,进一步改进域泛化的能力和声音异常检测的准确度。

参 考 文 献

[1] Zhou H, Wang K P, Li H, et al. Anomaly sound detection of industrial devices by using teacher-student incremental continual

learning [J]. Measurement Science and Technology, 2024, 35: 056107.

[2] 冯伟. 基于自监督学习的异常声音检测[D]. 大庆: 东北石油大学, 2023.

Feng Wei. Abnormal sound detection based on self-supervised learning[D]. Daqing: The Northeast Petroleum University, 2023.

[3] 王圆圆. 基于移动麦克风阵列的设备异常检测方法研究与实现[D]. 重庆: 西南科技大学, 2024.

Wang Yuanyuan. Research and implementation of device anomaly detection method based on mobile microphone array [D]. Chongqing: Southwest University of Science and Technology, 2024.

[4] 王昶, 周思源, 耿红军. 第四次工业革命背景下我国先进制造业发展路径及政策保障研究[J]. 中国科学院院刊, 2024, 39(2): 333-344.

Wang Chang, Zhou Siyuan, Geng Hongjun. Development path and policy guarantee of China's advanced manufacturing industry under background of fourth industrial revolution[J]. Bulletin of Chinese Academy of Sciences, 2024, 39(2): 333-344.

[5] 季坚莞, 胡文军, 王闯. 基于物联网的非侵入式负荷状态监控系统设计[J]. 湖州师范学院学报, 2024, 46(2): 53-64.

Ji Jianwan, Hu Wenjun, Wang Chuan. Design of non-intrusive load condition monitoring system based on internet of things[J]. Journal of Huzhou University, 2024, 46(2): 53-64.

[6] 雷怡琴, 孙兆龙, 叶志浩, 等. 电力系统负荷非侵入式监测方法研究[J]. 电工技术学报, 2021, 36(11): 2288-2297.

Lei Yiqin, Sun Zhaolong, Ye Zhihao, et al. Research on non-invasive load monitoring method in power system[J]. Transactions of China Electrotechnical Society, 2021, 36(11): 2288-2297.

[7] 杨昱, 王千山, 李雪鹏, 等. 一种非侵入式电梯运行状态智能监测方法[J]. 太原理工大学学报, 2021, 52(6): 920-927.

Yang Li, Wang Qianshan, Li Xuepeng, et al. A non-intrusive intelligent monitoring method for elevator state[J]. Journal of Taiyuan University of Technology, 2021, 52(6): 920-927.

[8] Wang M, Mei Q S, Song X Y, et al. A machine anomalous sound detection method using the IMS spectrogram and ES-MobileNetV3 network [J]. Applied Sciences, 2023, 13(23): 12912.

[9] Guan J, Liu Y D, Kong Q Q, et al. Transformer-based autoencoder with ID constraint for unsupervised anomalous sound detection [J]. EURASIP Journal on Audio, Speech, and Music Processing, 2023(1): 1-16.

[10] Jo H K, Kim S H, Kim C, et al. Proposal of a new method for learning of diesel generator sounds and detecting abnormal sounds using an unsupervised deep learning algorithm [J]. Nuclear Engineering and Technology, 2023, 55(2): 506-515.

[11] Obsa G, Jimson M. Subdomain adaptation via correlation alignment with entropy minimization for unsupervised domain adaptation [J]. Pattern Analysis and Applications, 2024, 27(1): 1-17.

[12] Li B X, Song P Y, Zhao C H. Fusing consensus knowledge: a federated learning method for fault diagnosis via privacy-preserving reference under domain shift [J]. Information Fusion, 2024, 106: 102290.

[13] 张晨旭, 李圣辰, 邵曦. 基于自编码器的无监督机器异常声音检测[J]. 复旦学报(自然科学版), 2021, 60(3): 297-302.

Zhang Chengxu, Li Shengchen, Shao Xi. Unsupervised detection of anomalous sounds for machine based on auto-encoder[J]. Journal of Fudan University (Natural Science), 2021, 60(3):

- 297-302.
- [14] 薛英杰, 韩威, 周松斌, 等. 基于生成对抗单分类网络的异常声音检测[J]. 吉林大学学报(理学版), 2021, 59(6): 1517-1524.
Xue Yingjie, Han Wei, Zhou Songbin, et al. Abnormal sound detection based on generative adversarial single classification network [J]. Journal of Jilin University (Science Edition), 2021, 59(6): 1517-1524.
- [15] 仇睿, 张晨旭, 姚瑶, 等. 基于频谱感知音频去噪的无监督机器异常声音检测[J]. 复旦学报(自然科学版), 2022, 61(5): 513-519.
Chou Rui, Zhang Chengxu, Yao Yao, et al. Unsupervised machine abnormal sound detection based on spectrum sensing audio denoising[J]. Journal of Fudan University (Natural Science), 2022, 61(5): 513-519.
- [16] 张勇, 元文智, 段贵金, 等. 基于深度卷积神经网络与高斯混合模型的水电机组异常声音检测[J]. 水电能源科学, 2023, 41(8): 188-191, 130.
Zhang Yong, Yuan Wenzhi, Duan Guijin, et al. Detection of abnormal sound of hydroelectric unit based on combination of deep convolutional neural network and gaussian mixture model[J]. Water Resources and Power, 2023, 41(8): 188-191, 130.
- [17] 姜慧天. 用于机器异常声音检测的自监督学习方法研究[D]. 成都: 电子科技大学, 2023.
Jiang Huitian. Self-supervised detection of anomalous sounds for machine condition monitoring [D]. Chengdu: School of Information and Communication Engineering, 2023.
- [18] Kuroyanagi I, Hayashi T, Takeda K, et al. Improvement of serial approach to anomalous sound detection by incorporating two binary cross-entropies for outlier exposure [C]//30th European Signal Processing Conference. New York: IEEE, 2022: 05929.
- [19] 王超, 李敬兆, 张金伟. 基于注意力机制的无监督异常声音检测方法[J]. 兰州工业学院学报, 2024, 31(1): 1-5.
Wang Chao, Li Jingzhao, Zhang Jinwei. Unsupervised abnormal sound detection method based on attention mechanism[J]. Journal of Lanzhou Institute of Technology, 2024, 31(1): 1-5.
- [20] 章林柯, 许艳武, 余永升. 基于域泛化的无监督机器设备异常声检测[J]. 海军工程大学学报, 2023, 35(6): 52-56, 70.
Zhang Linke, Xu Yanwu, Yu Yongsheng. Anomalous sound detection of unsupervised machines based on domain generalization [J]. Journal of Naval University of Engineering, 2023, 35(6): 52-56, 70.
- [21] 毕忠勤, 李欢峰, 张伟娜, 等. 基于域泛化的工业设备无监督异常声音检测算法[J]. 科学技术与工程, 2024, 24(3): 1091-1099.
Bi Zhongqin, Li Huanfeng, Zhang Weina, et al. Unsupervised anomalous sound detection algorithm of industrial equipment based on domain generalization [J]. Science Technology and Engineering, 2024, 24(3): 1091-1099.
- [22] Wilkinghoff K. An outlier exposed anomalous sound detection system for domain generalization in machine condition monitoring [R]. Tokyo: DCASE2022 Challenge, 2022.
- [23] Zhang H Y, Moustapha C, Yann N, et al. Mixup: Beyond empirical risk minimization [EB/OL]. (2018-04-27) [2024-07-19]. <https://arxiv.org/abs/1710.09412v2>.
- [24] Hou Q B, Zhou D Q, Feng J S, et al. Coordinate attention for efficient mobile network design [EB/OL]. (2021-03-04) [2024-07-19]. <https://arxiv.org/abs/2103.02907>.
- [25] Wilkinghoff K. Sub-cluster adacos: Learning representations for anomalous sound detection, 2021 International Joint Conference on Neural Networks [C]//International Joint Conference on Neural Networks. Shenzhen: IEEE, 2021: 1-8.
- [26] Wilkinghoff K. Combining multiple distributions based on sub-cluster adacos for anomalous sound detection under domain shifted conditions [R]. Tokyo: DCASE2022 Challenge, 2022.
- [27] Tožička J, Bezůšek M, Durkota K, et al. DADAED-double anomaly detector with aediff [R]. Tokyo: DCASE2022 Challenge, 2022.
- [28] Yamashita J, Tanaka R, Ikeda K, et al. Anomaly detection using autoencoder, idnn and u-net using ensemble [R]. Tokyo: DCASE2022 Challenge, 2022.
- [29] Almodévar A, Ortega A, Vicente L, et al. Vision transformer based embeddings extractor for unsupervised anomalous sound detection under domain generalization [R]. Tokyo: DCASE2022 Challenge, 2022.
- [30] Liu G, Liu Y, Cai S F, et al. Unsupervised anomalous sound detection under domain shift conditions based on mobilefacenets and masked autoregressive flow [R]. Tokyo: DCASE2022 Challenge, 2022.
- [31] Gou J C, Shi C, Li H Y. Unsupervised anomalous sound detection using feature extractor and anomaly detector [R]. Tokyo: DCASE2022 Challenge, 2022.
- [32] Peng T, Qiu R, Zhu J Y, et al. Unsupervised abnormal sound detection based on spectral coherence and feature fusion in domain displacement condition [R]. Tokyo: DCASE2022 Challenge, 2022.
- [33] Li X Y, Yang J, Shen H. Anomalous sound detection with ensemble of cnn-based features and autoencoder approaches [R]. Tokyo: DCASE2022 Challenge, 2022.
- [34] Cohen O, Vinokur Y, Arad A, et al. Unsupervised anomalous detection based on riemannian geometry [R]. Tokyo: DCASE2022 Challenge, 2022.