

基于深度学习 U-Net 架构的设备音频异常检测方法

宋媛媛¹ 李宏滨¹ 汪慧娟¹

SONG Yuanyuan LI Hongbin WANG Huijuan

摘要

随着智能制造和工业互联网的快速发展,设备状态监测和故障检测成为确保设备安全和提升生产效率的关键技术方法。传统的设备异常检测方法大多依赖于人工经验或是基于经典机器学习算法,但这些方法在面对复杂且高维的设备音频数据时往往存在着局限性。针对以上问题,文章提出了一种基于自编码器与 U-Net 架构结合的深度学习模型(deep spectral anomaly net, DSANet),用于设备音频信号的异常检测。具体来说该方法通过梅尔频谱图提取音频信号的时频域特征并结合自编码器和 U-Net 架构对异常进行检测。在训练过程中该模型通过学习正常信号的特征并计算重建误差以此得到阈值,通过比较每帧音频的重建误差与预设阈值判断是否为异常信号。此外,文章在模型中引入了跳跃连接和注意力机制增强了对音频信号局部特征的提取能力,从而提高了异常检测的鲁棒性。

关键词

深度学习; 自编码器; U-Net; 异常检测

doi: 10.3969/j.issn.1672-9528.2025.07.008

0 引言

在工业设备运行中,异常声音通常作为设备故障的早期信号。传统故障检测方法依赖振动信号和温度传感器,但面对设备故障的多样性和复杂性,存在较大的局限性。基于音频信号的故障检测方法逐渐成为新兴研究方向,比如张哲宇等人^[1]提出了一种基于改进(you only look once XS, YOLO XS)算法的视频巡检异常监测方法,结合迁移学习,在电力设备异常检测中显著提升了检测精度与效率,有效解决了智能配网巡检中的多设备和高难度问题。刘毅^[2]提出了基于辅助监督的无监督异常音频检测方法,通过引入自监督信息和类别标签监督信息,有效提取区分正常与异常数据的特征。仇睿等人^[3]提出了一种基于频谱感知音频去噪的无监督机器异常声音检测方法,通过去噪处理和深度可分离卷积分类器提高了在 DCASE Challenge 2021 Task2 数据集上的检测性能。Ghadiya 等人^[4]提出了一种多模态弱监督视频异常检测(WS-VAD)框架,利用跨模态融合适配器(CFA)和双曲洛伦茨图注意力(HLGAtt)来增强音频-视觉特征融合并捕捉层次关系。Shim 等人^[5]提出了一种基于音频信号的感应电动机异常检测模型,结合生成对抗网络(GAN)和集成方法,并通过实验验证了其有效性。Zou 等人^[6]提出了一种基于全球优化层和少样本训练策略(GOL-SFSTS)的多通道音频异常检测框架,并通过 8 通道梅尔频谱图提升了模型的特征捕获能力。

在面临多频率成分、时变模式及环境噪声干扰等挑战,如何有效提取特征并识别异常信号仍是关键问题。针对传统故障检测方法的局限性,本文提出了一种基于自编码器与 U-Net 架构结合的深度学习模型,用于有效提取设备音频信号特征并进行异常检测。

1 技术方法

1.1 梅尔频谱图

为减少计算复杂度并有效提取特征,将设备音频信号转换为梅尔频谱图^[7],进而从频域提取音频信号的频谱特征。梅尔频谱图的计算过程为:给定一段长度为 L 的音频信号 $x(t)$,通过短时傅里叶变换(STFT)将其转化为频域表示:

$$X(t, f) = \sum_{n=-\infty}^{\infty} x(t)w(t-n)e^{-2j\pi ft} \quad (1)$$

式中: $w(t)$ 是窗口函数; t 是时间。后将频谱图映射到梅尔频率尺上,得到梅尔频谱图 $M(t, m)$:

$$M(t, m) = \sum_{f=0}^F |x(t, f)| \cdot H_m(f) \quad (2)$$

式中: F 是频率的总数; $H_m(f)$ 是梅尔滤波器; m 是梅尔频带。通过这种方式,音频信号的频谱特征可以有效地表示设备的运行状态。

在音频信号转换为梅尔频谱图后,对其进行归一化处理,以消除不同音频信号的幅度差异。归一化过程为:

$$\hat{M}(t, m) = \frac{M(t, m) - \min(M)}{\max(M) - \min(M)} \quad (3)$$

式中： $\hat{M}(t, m)$ 表示归一化后的梅尔频谱图。

1.2 U-Net 架构

U-Net^[8]主要特点是采用编码器-解码器结构，来有效恢复图像的空间细节。U-Net 的网络结构包括两部分：编码器用于逐步提取特征，解码器用于重建原始数据的空间分辨率。

1.2.1 编码器

编码器由多个卷积层和最大池化层组成，逐步提取输入数据的高层特征。每个卷积层使用 3×3 卷积核，通过 ReLU 激活函数引入非线性变换公式为：

$$y = \text{ReLU}(W \cdot x + b) \quad (4)$$

式中： W 和 b 是卷积核和偏置项； x 是输入； y 是输出。

1.2.2 解码器部分

解码器则通过转置卷积（Deconvolution）^[9] 逐步恢复图像的空间分辨率。每个解码器层通过上采样将低维特征映射回原始尺寸。通过跳跃连接，解码器直接连接对应的编码器层，从而保留图像的细节信息公式为：

$$y_{\text{decoded}} = \text{ConvTranspose}(x) \quad (5)$$

式中： x 是从编码器中提取的特征图； y_{decoded} 是经过转置卷积处理的输出。

1.3 评价指标

为评估模型的性能，采用 AUC（area under the curve）^[10]、pAUC（partial area under the curve）以及 F_1 -score 等标准评估指标。

AUC：AUC（曲线下面积）是指接收操作特征曲线（ROC 曲线）下面积的大小。ROC 曲线的横坐标是假阳性率（ R_{FP} ），纵坐标是真正率（ R_{TP} ），AUC 值的范围是 $[0, 1]$ ，AUC 值越大，表示模型的分类能力越好。

其中 AUC 的公式：

$$\text{AUC} = \int_0^1 R_{TP}(R_{FP}) dR_{FP} \quad (6)$$

式中： R_{TP} （真正率）和 R_{FP} （假阳性率）分别定义为：

$$R_{TP} = \frac{TP}{TP + FN}, \quad R_{FP} = \frac{FP}{FP + TN} \quad (7)$$

pAUC：pAUC 是 AUC 的一个变种，专注于某一特定假阳性率（ R_{FP} ）区域下的 AUC。可以用于评估模型在假阳性率较低时的性能表现，尤其在数据不平衡的情况下，能够更好地反映模型在低假阳性率区域的有效性。pAUC 的计算公式为：

$$\text{pAUC} = \int_{R_{FP_{\min}}}^{R_{FP_{\max}}} R_{TP}(R_{FP}) dR_{FP} \quad (8)$$

式中： $R_{FP_{\max}}$ 和 $R_{FP_{\min}}$ 是假阳性率的范围，意味着 pAUC 只计算某个特定范围，而非整个 ROC 曲线的面积。

F_1 -score： F_1 -score 是精确率（Precision）和召回率（Recall）的调和均值，用来评估模型的性能，尤其是在数据不平衡的情况下， F_1 -score 是一个重要的评价指标。 F_1 -score 的计算公式为：

$$F_1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (9)$$

2 DSANet 异常检测模型

DSANet（deep spectral anomaly net）模型是本文提出的一种用于设备音频异常检测的深度神经网络结构通过结合自编码器（Autoencoder）^[11]、U-Net 架构和注意力机制来提升模型在复杂异常模式下的鲁棒性。DSANet 模型的设备音频异常检测系统流程图如图 1 所示。

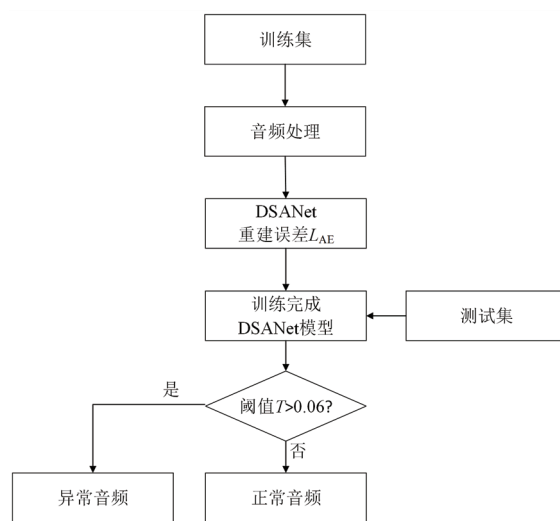


图 1 设备音频异常检测流程图

2.1 数据预处理与特征提取

将音频信号转换为梅尔频谱图（Mel spectrogram），对每段音频信号 $x(t)$ 进行以下处理步骤：

（1）加载与重采样：音频信号首先通过 torchaudio 库加载并通过重采样到统一的采样率（16 kHz）。

（2）短时傅里叶变换（STFT）：通过 STFT 将时域信号转到频域信号，得到频谱图 $X(t, f)$ 。

（3）梅尔滤波器转换：通过梅尔滤波器对频谱进行变换将频率从线性尺度映射到梅尔尺度得到了梅尔频谱图 $M(t, m)$ 。

（4）归一化：为消除不同音频信号之间的幅度差异对梅尔频谱图进行归一化处理。

经过以上这些预处理步骤后将得到的输入特征为归一化后的梅尔频谱图，作为本文所设计 DSANet 模型的输入。

2.2 模型架构

DSANet 模型主要由编码器、解码器、跳跃连接（Skip Connections）和注意力门机制（Attention Gate）^[12] 等部分组成。

所设计模型的目标是通过学习输入信号的正常模式得到重建误差以此来得到阈值判断设备是否出现异常。

具体来说模型采用自编码器结构其中编码器和解码器由卷积块 (ConvBlock) 构成, 每个卷积块包含两个卷积层和一个残差连接。同时在解码器部分加入了跳跃连接和注意力门机制以增强特征提取的能力。

2.3 网络层次结构

DSANet 模型由编码器、中间层、解码器 3 部分组成, 具体表示:

(1) 编码器: 编码器由 4 个卷积块组成且每个卷积块包含有两个卷积层用来逐层提取信号的特征, 同时在每个卷积块后面使用最大池化 (MaxPooling) 层来减小特征图的空间分辨率, 每层卷积块的输出通道数逐渐增加从 32~256。

(2) 中间层 (瓶颈层): 中间层接收编码器的输出特征并进一步提取更深层的特征, 这一层由一个卷积块组成输出通道数为 512。

(3) 解码器: 解码器由 4 个转置卷积层 (deconvolutional layers) 组成每一层解码器的输入和输出通道数依次为 $512 \rightarrow 256$ 、 $256 \rightarrow 128$ 、 $128 \rightarrow 64$ 、 $64 \rightarrow 32$ 。每层的解码器通过转置卷积进行上采样同时, 与编码器对应层的特征通过跳跃连接的方式进行拼接从而保留重要的局部细节信息, 并且在每个转置卷积层引入了注意力机制经过加权拼接特征图提取重要特征并抑制特征不重要的部分最终通过一个 1×1 的卷积层生成重建信号。如图 2 所示。

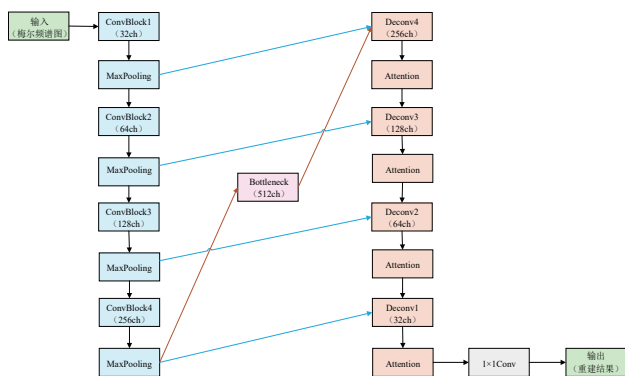


图 2 DSANet 模型图

2.4 异常检测方法

在 DSANet 模型中设备音频异常检测是通过计算输入信号和重建信号之间的重建误差来判断信号是否音频异常。具体步骤如下:

首先, 在测试阶段输入的梅尔频谱图通过编码器和解码器重建为输出信号, 重建误差 L_{AE} 计算公式为:

$$L_{AE} = \|\hat{x} - x\|_2^2 \quad (10)$$

式中: $\hat{x}$ 是重构的梅尔频谱图; x 是原始输入信号。

其次, 通过异常分数与阈值判定具体来说每个测试样本的异常分数 E 是基于重建误差来计算的, 具体来说异常分数是通过计算每个样本的重建误差并得到平均值来计算异常程度的计算公式为:

$$E = \frac{1}{N} \sum_{i=1}^N \|\hat{x}_i - x_i\|_2^2 \quad (11)$$

式中: E 是样本的异常分数 N 为梅尔频谱图的特征点数; $\hat{x}$ 和 x_i 分别是第 i 个特征点的重建值和原始值。

为进行有效的设备音频异常检测, 本文通过设定一个固定的阈值 T 来判断设备是否异常。在训练过程中通过计算得出正常信号的重建误差均值 0.03 (μ_{normal}) 和标准差 0.01 (σ_{normal}), 阈值 T 通常设置为正常信号重建误差的均值加三倍标准差, 这是基于正态分布的统计特性。因为在正态分布中 99.7% 的数据应该落在均值加减三倍标准差的范围内。阈值的计算公式为:

$$T = \mu_{normal} + 3 \cdot \sigma_{normal} = 0.03 + 3 \times 0.01 = 0.06 \quad (12)$$

故阈值 T 被设定为 0.06, 实验验证表明该阈值能够准确区分正常信号和异常信号保证高效的异常检测性能。

3 实验结果与分析

通过在公开的数据集上的实验进行详细分析展示了该模型在设备音频异常检测任务中的有效性。并且通过对比实验结果验证了该模型在异常检测中的性能优势。

3.1 数据集

本文采用了两个公开数据集 MIMII (mechanical impulse noise for industrial instruments) 数据集和 Toy ADMOS (toy anomaly detection for mechanical oscillators) 数据集。

其中 MIMII 数据集包含来自 4 种不同的工业设备包括: 风扇、泵、滑轨和阀门的音频数据, 该数据集涵盖了正常运行和多种故障状态的音频样本共包含 8 000 条音频信号。Toy ADMOS 数据集由两种玩具机械振动装置: 玩具车和玩具传送带的音频信号组成异常信号模拟了机械部件常见故障, 如摩擦和卡滞的异常信号主要用于验证模型在设备故障检测中的表现。

3.2 实验参数

在实验中本文所设计的 DSANet 模型使用梅尔频谱图作为输入特征, 采样率设置为 16 kHz 梅尔频谱图的维度为 128, 步长为 512。采用 Adam 优化器学习率设定为 1×10^{-4} , 并且使用均方误差 (MSE) 作为损失函数。本文将数据集分为: 70% 为训练集、30% 为测试集, 训练过程中的批大小为 32, 训练轮次设置为 100 轮次。异常检测时重建误差阈值 T 设定为 0.06。

3.3 实验对比对象

为验证 DSANet 的性能, 本文将其与 DCASE 2024 Challenge Task 2 中的基线系统 (baseline system, BS) 进行对比, 基线系统使用了一种传统的机器学习方法。该模型首先提取梅尔频谱图和其他音频特征, 然后通过支持向量机 (SVM) 进行异常检测。基线方法的优点在于易于实现并且对于一些简单的异常检测任务表现较好, 但在面对更复杂的异常模式时其效果通常较为有限。本文将 DSANet 与基线系统在 MIMII 和 Toy ADMOS 数据集上的表现进行了比较, 评估了所设计模型的设备音频异常检测的性能。

3.4 实验对比结果

在 MIMII 数据集的实验中, 对 Fan、Pump、Slider 和 Valve4 个设备类别分别进行了训练和测试并在开发集与测试集上计算了 AUC、pAUC 和 F_1 -score 这 3 个评价指标。实验结果表明 DSANet 相较于基线系统 BS 在所有设备类别中均表现出显著的提升, 具体结果如表 1 所示。

表 1 改进方法与基线系统的比较

设备类别	模型	AUC	pAUC	F_1 -score
Fan	DeepSpectralAnomalyNet	0.88	0.80	0.83
	BaselineSystem(BS)	0.78	0.72	0.75
Pump	DeepSpectralAnomalyNet	0.91	0.82	0.85
	BaselineSystem(BS)	0.80	0.75	0.78
Slider	DeepSpectralAnomalyNet	0.81	0.78	0.86
	BaselineSystem(BS)	0.74	0.68	0.70
Valve	DeepSpectralAnomalyNet	0.83	0.84	0.88
	BaselineSystem(BS)	0.76	0.71	0.74

在 Toy ADMOS 数据集的实验中, 对 Toy Car 和 Toy Conveyor 两个设备类别分别进行了训练和测试, 并计算了 AUC、pAUC 和 F_1 -score 这 3 个评价指标。结果表明 DSANet 在 Toy Car 和 Toy Conveyor 的异常检测任务中均优于基线系统在各项指标上都表现出了较高的提升。

表 2 改进方法与基线系统的比较

设备类别	模型	AUC	pAUC	F_1 -score
Toy Car	DeepSpectralAnomalyNet	0.89	0.85	0.88
	BaselineSystem(BS)	0.78	0.73	0.76
Toy Conveyor	DeepSpectralAnomalyNet	0.93	0.86	0.90
	BaselineSystem(BS)	0.80	0.74	0.78

从表 2 结果可以看出, DSANet 在各个设备类别的音频异常检测任务中, 尤其是在 Fan 和 Pump 等复杂设备类别中表现出较高的结果, 表明了该模型在区分正常与异常信号时具有更强的检测能力。同时, 所设计模型的 F_1 -score 的提升也表明该模型能够更好地平衡精确度和召回率, 减少了假阳

性和假阴性的数量。

在 Toy ADMOS 数据集上 DSANet 模型同样展现了优异的性能, 在 Toy Car 和 Toy Conveyor 这两个数据集上, AUC 和 pAUC 均大幅度高于基线系统, F_1 -score 也相应得到了明显提升。这表明了模型不仅能够处理复杂设备的异常检测任务, 也能在较简单的设备故障检测中表现出色。

4 结语

本文提出的 DSANet 模型通过自编码器与 U-Net 架构相结合有效地进行了设备音频信号的异常检测。该模型通过学习正常信号的重建误差成功区分音频的正常与异常信号并在多个数据集上取得了较好的检测性能。在未来的研究中可进一步针对模型的结构层次和针对音频的时序性问题来探索更高效的、更适合于复杂多变的设备音频异常检测的模型, 并且在未来的研究中结合更大规模的数据集以提高模型在复杂环境中的适应性和检测性能。

参考文献:

- [1] 张哲宇, 徐扬, 吴海涛, 等. 基于人工智能算法的 RPA 机器人视频巡检异常监测方法 [J/OL]. 自动化技术与应用, 1-6[2025-02-15].<http://kns.cnki.net/kcms/detail/23.1474.TP.20241223.1342.088.html>.
- [2] 刘毅. 无监督异常音频检测的研究 [D]. 北京: 北京邮电大学, 2023.
- [3] 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.
- [4] GHADIYA A, KAR P, CHUDASAMA V, et al. Cross-modal fusion and attention mechanism for weakly supervised video anomaly detection[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Piscataway: IEEE, 2024:1965-1974.
- [5] SHIM J, SEO J, LEE S, et al. Ensemble-based anomaly detection with comprehensive feature distances for induction motors[J]. IEEE transactions on industry applications, 2024, 60(4): 6033-6044.
- [6] ZOU F Q, LI X M, LI Y Q, et al. GOL-SFSTS based few-shot learning mechanical anomaly detection using multi-channel audio signal[J]. Knowledge-based systems,2024,284:111204.
- [7] USTUBIOGLU B, TAHAOGLU G, ULUTAS G. Detection of audio copy-move-forgery with novel feature matching on Mel spectrogram[J]. Expert systems with applications, 2023, 213: 118963.

拉普拉斯平滑对朴素贝叶斯邮件分类器的影响机制研究

张叔超^{1*} 李廷元¹
ZHANG Shuchao LI Tingyuan

摘要

针对朴素贝叶斯分类器在邮件分类任务中的特征表示和平滑处理问题,文章提出了一种改进的预处理流程和参数优化方案。传统的朴素贝叶斯分类器在处理文本数据时通常采用单一的分割方式进行特征提取,这种方式可能导致部分重要特征的丢失或噪声的干扰,从而影响分类效果。为有效提升特征提取的质量,设计了一种“逗号-空格”双重分割策略,这一策略能够更加细致地处理邮件内容中的结构性信息,从而提升特征的全面性和准确性。同时系统验证了拉普拉斯平滑对模型性能的影响机制,拉普拉斯平滑作为常用的平滑技术,能够有效避免在朴素贝叶斯模型中因某些特征未出现而导致的零概率问题。为验证改进方案的有效性,选取5 574封真实邮件数据集进行实验,实验结果表明,改进后的朴素贝叶斯分类器在准确率方面取得了显著提升,准确率达到97.40%,相比传统实现提升了6.82%。

关键词

垃圾邮件; 正常邮件; 朴素贝叶斯算法; 拉普拉斯平滑算法

doi: 10.3969/j.issn.1672-9528.2025.07.009

0 引言

随着互联网技术的快速发展,电子邮件成为人们日常生活和工作中不可或缺的沟通工具之一。凭借即时传递信息的高效性和低成本优势,打破地域限制,让全球用户得以随时随地互通有无。但与此同时,电子邮件的普及也给不法分子和不良商家提供了可乘之机。利用这一平台大肆开展广告轰炸、网络诈骗等非法活动,给广大用户带来诸多困扰,严重

影响了用户的使用体验和信息安全。

据《中国网民权益保护调查报告 2016》^[1]的数据显示,超过半数的网民认为电子邮件存在个人信息泄露的隐患,尤其是在垃圾邮件泛滥的现状下。2016年上半年,网民们每周平均收到18.9封匿名的垃圾邮件,这些邮件通常是广告推销、钓鱼诈骗、恶意软件传播等形式的邮件,给用户带来了极大的困扰和风险。垃圾邮件的频繁出现,不仅浪费了用户宝贵的时间,还严重影响了其正常的工作与生活。甚至垃圾邮件所含的链接可能隐藏着病毒或恶意程序,用

1. 中国民用航空飞行学院 四川广汉 618307

[8] CAO H, WANG Y Y, CHEN J, et al. Swin-UNet: UNet-like pure transformer for medical image segmentation[C]//Computer Vision-ECCV 2022 Workshops, Berlin: Springer, 2022: 205-218.

[9] MIAO Y H, ZHANG B Y, LIN J, et al. A review on the application of blind deconvolution in machinery fault diagnosis[J]. Mechanical systems and signal processing, 2022, 163: 108202.

[10] HAN Y T, ZHANG J H, JIANG Z H, et al. Is the area under curve appropriate for evaluating the fit of psychometric models?[J]. Educational and psychological measurement, 2023, 83(3): 586-608.

[11] 袁非牛, 章琳, 史劲亭, 等. 自编码神经网络理论及应用综述 [J]. 计算机学报, 2019, 42(1): 203-230.

[12] DENG W J, SHI Q, LI J. Attention-gate-based encoder-decoder network for automatic building extraction[J]. IEEE journal of selected topics in applied earth observations and remote sensing, 2021, 14: 2611-2620.

【作者简介】

宋媛媛(1996—),女,山西晋城人,硕士研究生,研究方向:计算机技术、移动互联应用及开发技术。

李宏滨(1968—),男,山西晋中人,硕士,副教授,研究方向:智能图像处理。

汪慧娟(1999—),女,河南信阳人,硕士研究生,研究方向:计算机技术、移动互联应用及开发技术。

(收稿日期: 2025-02-20 修回日期: 2025-06-30)