计 算 机 研 究 与 发 展
Journal of Computer Research and Development
https://crad.ict.ac.cn
63(10): 2650−2665, 2026

基于双重异常感知的时间序列异常检测

王志鸿[1] 黄 维[1,2,3] 王德贤[4] 李天瑞[5]

[1](福州大学计算机与大数据学院 福州 350108)

[2](大数据智能教育部工程研究中心(福州大学) 福州 350108)

[3](福建省网络计算与智能信息处理重点实验室(福州大学) 福州 350108)

[4](成都中医药大学智能医学学院 成都 611137)

[5](西南交通大学计算机与人工智能学院 成都 611756)

(wzh1071429405@163.com)

Time Series Anomaly Detection Based on Dual Anomaly-Aware

Wang Zhihong[1], Huang Wei[1,2,3], Wang Dexian[4], and Li Tianrui[5]

[1](*College of Computer and Data Science*, *Fuzhou University*, *Fuzhou* 350108)

[2](*Engineering Research Center of Big Data Intelligence* (*Fuzhou University*), *Ministry of Education*, *Fuzhou* 350108)

[3](*Fujian Key Laboratory of Network Computing and Intelligent Information Processing* (*Fuzhou University*), *Fuzhou* 350108)

[4](*School of Intelligent Medicine*, *Chengdu University of Traditional Chinese Medicine*, *Chengdu* 611137)

[5](*School of Computing and Artificial Intelligence*, *Southwest Jiaotong University*, *Chengdu* 611756)

Abstract Anomaly detection for time series holds practical value in diverse fields such as financial risk control and industrial fault diagnosis, where the core challenge lies in constructing a representation space that effectively distinguishes between normal and anomalous patterns. However, the inherent characteristics of temporal data, coupled with the extreme scarcity of labeled anomalies, make it highly difficult to learn a representative mapping function with strong discriminative power, often leading to limited generalization capabilities. Meanwhile, contrastive representation learning constructs a discriminative feature space that clearly differentiates between instances, offering a more natural and promising solution for temporal anomaly detection. To address these challenges, we propose a dual anomaly-aware Transformer (DAT) model. By incorporating anomaly injection and anomaly awareness mechanisms, DAT is specifically enabled to capture the complex features of potential anomalous samples during the training phase. Additionally, a contextual contrastive loss function is adopted to guide a dual-branch structure for extracting more intrinsic and discriminative semantic features between normal and anomalous samples. Comprehensive experiments conducted on five public datasets demonstrate the superiority of DAT over 21 baseline models. The results show that DAT not only achieves competitive performance across all benchmark datasets but also achieves state-of-the-art (SOTA) performance on the MSL and PSM datasets.

Key words time series; contrastive representation learning; anomaly detection; Transformer; anomaly-aware

摘 要 时间序列的异常检测在金融风控、故障诊断等领域中有着重要的应用价值,其关键在于构建能够有效区分正常与异常模式的表征空间。然而时间序列内在的特点以及异常的稀缺性,导致学习具有强判别能力的表征映射函数颇为困难,且泛化能力受限。与此同时,对比表示学习通过构建清晰可区分的样

收稿日期: 2025-08-19; **修回日期**: 2026-05-12

基金项目: 国家自然科学基金青年科学基金项目(62406070, 62506049)

This work was supported by the National Natural Science Foundation of China for Young Scientists (62406070, 62506049).

通信作者: 黄维(hwfj@fzu.edu.cn)

本实例的表征空间，为时序异常检测提供了更为自然且富有前景的解决方案。鉴于此，设计了基于双重异常感知的 Transformer 模型（DAT），通过异常注入和异常感知的手段，使 DAT 提前捕获异常样本的复杂特征。此外，通过结合上下文对比损失函数指导双分支学习正常样本与异常样本之间更本质、更具区分度的语义特征。实验在 5 个公开数据集上进行验证，并与 21 种基线模型进行对比。结果表明 DAT 在公共数据集上优于基线模型的平均性能，甚至在 MSL 和 PSM 数据集上能够达到 SOTA。

关键词 时间序列；对比表示学习；异常检测；Transformer；异常感知

中图法分类号 TP391

DOI: 10.7544/issn1000-1239.202550613 **CSTR:** 32373.14.issn1000-1239.202550613

目前，时间序列的异常检测在现实中有极为广泛的应用，包括故障诊断、金融欺诈检测以及健康诊断等[1-3]。时间序列异常检测是从高维、非线性且具有复杂时序依赖的大规模数据中，构建具有检测异常点、异常子序列及异常模式的算法框架。

但是，相关技术的发展仍面临着巨大挑战。首先，异常的具体形式尚待确认中。根据实际观测到的情况，所采集的数据可能是无规则、不一致、有缺失的。其次，是罕见的异常，对于数据的异常标注，需要专业人士参与且标注成本高昂，导致大部分的数据集并没有提供真实的标注。这意味着大多数监督学习方法的性能可能会显著下降。最后，时间序列本身的时间依赖性、多维性和非平稳性特征蕴含在异常检测的过程当中。这些特征相互耦合，使得时间序列检测模型难以综合考虑。

传统的异常检测方法往往忽略时间顺序的重要性，难以捕捉长期的依赖关系[4]。此外，监督学习方法对标注数据具有强依赖性，在异常标注有限的时间序列上其也面临着泛化能力受限的挑战[5]。无监督学习方法可以有效利用无标注数据，因此在时间序列异常检测领域逐渐成为研究热点。同时，以对比表示学习在内的自监督学习，因多样的架构设计以及在多个领域中的显著表现而备受青睐[6]。然而，对比表示学习的性能受到正负样本的强烈影响，因此该技术在时间序列的异常检测的有效性依旧需要进一步的探索。

因此，本文基于对比表示学习和 Transformer 架构，设计了面向时间序列的双重异常感知模型 DAT。受自然语言处理（natural language processing, NLP）领域 BERT（bidirectional encoder representations from Transformers）[7] 的启发，DAT 向输入数据中随机注入人工合成异常数据并训练模型来找出这些异常数据，从而模仿掩码语言建模。异常注入机制作为一种启发式策略，无需模拟所有场景的异常数据，其目的在于引导模型捕捉正常与异常样本的差异化表征[8]。此外，DAT 模型基于时序数据增强一致性的假设，即同一时间序列经不同增强方法变换后，其潜在空间特征表示应保持相似，设计了双分支的数据增强以及对比学习。本方法的主要贡献可归纳如下：

1）通过混合多种数据增强方式，并引入双分支结构指导生成语义一致但多样化的增强视图，增强时间序列数据的特征表达，有利于模型提取数据中更为广泛、更具鲁棒性的表征，从而提高泛化能力。

2）采用可控的异常注入方式，缓解时序标注数据的不足。并结合异常感知，指导模型捕捉正常数据和异常数据在潜在表示空间上的本质差异，间接达到异常检测的目的。

3）基于上下文对比损失函数引导模型在潜在空间中对齐具有相似语义的跨视图表征。同时，联合优化异常感知损失函数，在保证语义连贯的情况下，提升异常特征的区分度。

1 相关工作

1.1 时间序列异常检测方法

近年来，学术界已经设计了多种异常检测的方法。由于异常的稀缺性，研究人员逐渐采用无监督的深度学习方法进行时间序列异常检测。

深度学习在时间序列异常检测领域应用了不同的学习策略。其中，基于一类分类器的方法通过最大化间隔超平面和构建最小超球体边界实现高维空间中的异常点分离[9-10]。基于数据重构的方法通过编码器-解码器架构学习正常数据特征并依据重构误差识别异常，例如 LSTM-VAE[5]，OmniAnomaly[11]，BeatGAN[12]，InterFusion[13] 等。更进一步，Donut 模型在此基础上优化了 ELBO 损失函数，利用 MCMC 机制减少异常值对重建概率的影响[14]；MSDMM 利用双记忆模块，提取时间序列的全局通用模式和局部共

性特征,抑制异常的泛化能力,保证对正常数据的重构[15]; NPSR 引入量化点异常的点模型和量化区间异常的序列模型,权衡点和区间异常的检测权重[16]。基于预测的方法通过训练预测模型预报未来的观测值,根据真实值与预测值的显著差值,进行异常识别,如 LSTM[17], CL-MPPCA[18]。此外,也有从特征融合、窗口优化及空间拓展等方面进一步改进重构机制。例如,FLAD 通过融合学习同时建模时序的局部与全局特征[19]; SELAD 利用多窗口划分与大语言模型集成策略捕捉复杂的周期模式[20]; MMGAN 引入多模态对抗学习,实现时频双域分布的联合度量[21]。

1.2 Transformer 模型及其变体

Transformer 使用多头注意力机制构建编码器-解码器对,其中每个子层包含若干自注意力模块[22]。该技术迅速渗透到各类 NLP 任务。例如,BERT 采用完全 Transformer 编码器的架构,对大规模无标注数据集进行掩码语言建模预训练,实现上下文语义理解[7]; XLNet 对自然语言进行重排实现自编码与自回归模型的融合[23]; T5 则构建了统一的文本到文本的框架,充分释放编码器-解码器结构的表征潜力[24]。LongNet 采用膨胀注意力机制,将序列建模能力扩展至 10 亿级 token 规模[25]; DeepSeek-V2 引入多头潜在注意力机制,通过构造低秩矩阵,实现键值高效压缩,兼顾了模型的性能和推理效率[26]。

1.3 对比表示学习

早期的对比学习通过对比相似与非相似的图片,实现数据潜在特征的学习[27]。目前,更多研究开始倾向使用对比学习提取时间序列的表征服务下游任务,其中包括异常检测。例如,CL-TAD 利用重构方法生成正样本,再与原数据进行对比学习[28]; PACdetector 通过双重注意力机制与对比学习,构建序列周期内波动与周期间趋势的联合表征空间,结合重构误差实现异常评分[29]; AMSL 借鉴卷积自动编码器结构,提出自适应记忆网络与自监督学习的异常检测方法[30]; MultiRC 采用多尺度架构和多尺度拼贴的方式,利用可控的生成策略,生成硬负样本,进行对比学习和重构的联合训练[31]。DCdetector 划分块内和块外 2 种视图,构建正样本对进行对比学习[32]。TF-ECL 利用时频增强构建双视图对比框架,通过协同检测机制实现更精准的异常定位[33]。

1.4 异常分析与注入实践

在时间序列异常检测中,异常类型的理论建模与分类一直是研究的核心问题之一。早期的研究[34]基于 ARIMA 模型,将异常类型系统地划分为以下 4 类:加性异常(additive outlier, AO);局部瞬时扰动,通常表现为单个时间点的剧烈偏离;水平跃迁(level shift, LS),序列均值在某一时间点发生突变并维持;暂态变化(temporary change, TC),短时间偏离后迅速回归正常趋势;创新异常(innovational outlier, IO),对序列生成过程中的创新项施加扰动,影响具有累积性。

后续的研究指出上述类型未能涵盖对序列结构本身的异常,如周期性、相位等变化。因此提出了季节性水平跃迁(seasonal level shift, SLS)的概念,用于刻画周期结构突变、相位漂移等复杂结构性异常,从而将“结构异常”纳入更完整的理论体系[35]。

随后,大量异常检测模型与数据增强方法均参考了上述理论,将尖峰型(spike)异常、均值漂移型(level shift)异常和结构形变型(pattern shift)异常 3 类形态作为基础异常原型。例如,GutenTAG 中构建了通用异常生成器,明确定义了极值(extremum)扰动、均值(mean)变化和结构扰动(pattern shift)等合成异常[36]; CARLA[37], NCAD[8], COUTA[38] 等方法也将点(point)异常、趋势(trend)异常与模式(pattern)异常作为注入与建模的核心形态,并指出这些异常具有良好的代表性与覆盖性。

2 双重异常感知 Transformer 检测模型

2.1 问题描述

时间序列异常检测任务的定义如下:给定长度为T的训练序列$\boldsymbol{X}=(\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_T)$(其中每个时间点$\boldsymbol{x}_t\in\mathbb{R}^d$),模型从中学习正常模式。随后对长度为$T'$的未知测试序列$\boldsymbol{X}_{\text{test}}=(\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_{T'})$,需要预测其对应的异常标签序列$\boldsymbol{Y}_{\text{test}}=(y_1,y_2,\cdots,y_{T'})$,其中$y_t\in\{0,1\}$是二进制标签,0 表示正常数据点,1 表示异常数据点。

DAT 采用滑动窗口的输入方式,根据指定窗口长度$L(L\leqslant T)$和移动步长s将长度为T的时间序列划分为$n=\lfloor(T-L)/s\rfloor+1$个子序列$S=\{\boldsymbol{S}_1,\boldsymbol{S}_2,\cdots,\boldsymbol{S}_n\}$。其中$\boldsymbol{S}_t=(\boldsymbol{x}_t,\boldsymbol{x}_{t+1},\cdots,\boldsymbol{x}_{t+L-1})$,表示窗口$t$对应的子序列所包含的时间点。

2.2 模型概述

图 1(a)展示了 DAT 的总体架构。其核心思想是结合多角度数据增强、异常注入与感知、上下文对比学习等模块,来学习时间序列表征。具体地,时间序列经滑动窗口处理后输入双分支数据增强模块,得到 2 个增强视图。接着 2 个视图同步注入预设比例的异常样本并保持异常位置和类型一致,经 Transformer 编码器提取特征后输出异常得分,与真

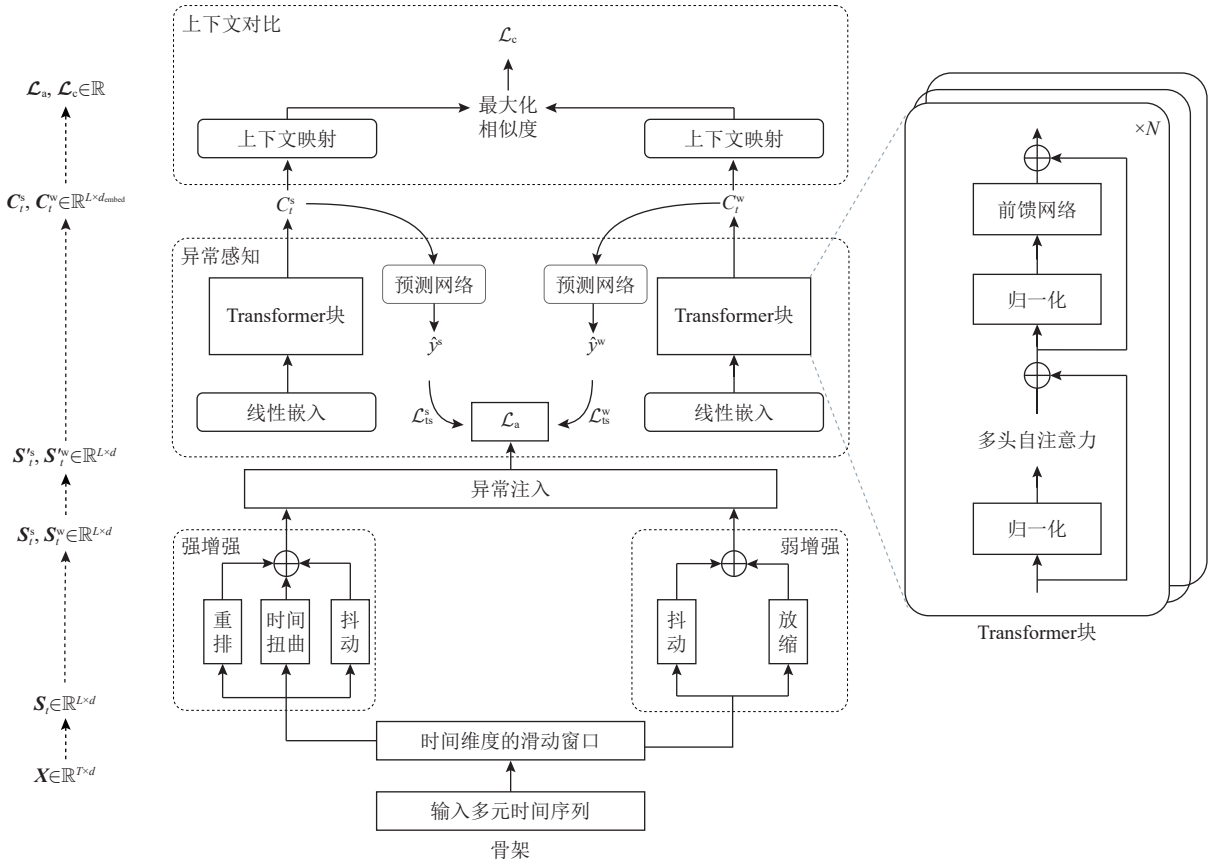


Fig. 1 DAT model

图 1 DAT 模型

实值对比计算异常感知损失。同时，2 个视图表征通过对比损失函数计算上下文差异损失，最终联合优化 2 类损失实现异常检测。

为了更好地分析 DAT 在时序异常检测中的工作内容，我们将 DAT 与 Anomaly Transformer[4] 和 DCdetector[32] 进行方法上的比较。作为显式关联建模的代表，Anomaly Transformer 通过高斯核构建先验差异，并通过 Transformer 学习关联差异，最终结合重构损失与 MinMax 策略进行优化。但是，该方法存在 2 点局限：其一，高斯核的学习过程复杂；其二，单一尺度的高斯先验关联难以适应多类型异常。相较之下，DAT 将 MinMax 策略替换为双分支的上下文对比损失函数。在避免梯度冲突和训练震荡的同时强化样本在潜在空间的区分度。同时，DAT 通过合理定义正负样本，有效避免了 DCdetector 因完全依赖正样本而可能引发的模型崩溃风险。此外，Anomaly Transformer 和 DCdetector 都是基于正常样本进行关联性建模，且认为训练数据均是正常样本，这在实际中往往难以满足。DAT 的异常注入通过扰动原始数据，模拟异常，结合异常感知对不同试图的正常和异常样本进行建模。后续的理论分析与实验结果均证明了 DAT 采用的注入策略与真实的异常分布接近，能够较为完备地模拟有效的异常。并且辅以双分支的上下文对比损失，强化正样本间的关联，弱化负样本间的联系。总之，DAT 的实现方式更为简洁、直观。

2.3 数据增强

在对比学习框架下，针对标签缺失的挑战，DAT 采用基于双分支数据增强的策略予以解决。基于“增强前后语义一致”的假设，DAT 将数据增强模块划分为“强增强”与“弱增强”模块，为对比学习提供视角多样但语义一致的增强视图。考虑到时间序列语义的复杂性及传统增强方式的局限性，DAT 的每个增强模块不是 1 种增强方式，而是包含若干候选方式的增强族 Φ。其中，每个增强族 Φ 各自包含了一系列的增强方式 f，如抖动、放缩等。Φ 中的所有 f 通过加权平均的方式得到对应的增强算子 $\mathcal{A}$：

$$\mathcal{A}(\cdot) = \frac{1}{|\Phi|} \sum_{i=1}^{|\Phi|} a_i \times f_i(\cdot) \,, \tag{1}$$

其中 $f(\cdot)$ 中的“$\cdot$”为输入的时间序列，$|\Phi|$ 表示 Φ 内的增强方法的个数，a_i 表示增强方法 f_i 的权重，是超参数，且 $\sum a_i = 1$。由此，增强模块可通过调整对应权

重,综合不同增强方式的特点,以保证序列前后语义特征相对一致。

DAT一共涉及到4种数据增强方式:1)抖动,向序列添加满足高斯分布的随机噪声;2)放缩,通过将序列与满足高斯分布的缩放因子相乘;3)时间扭曲,将序列以最大速度比率10.5实现83%~120%的扰动,然后插值实现时间轴的形变;4)重排,打乱序列内部的观测值。强增强和弱增强分别对应着各自的增强算子$\mathcal{A}_{\mathrm{s}}$,$\mathcal{A}_{\mathrm{w}}$。$\mathcal{A}_{\mathrm{s}}$包含重排、时间扭曲以及抖动,通过大幅度改变数据形态,迫使模型关注整体模式而非局部噪声;$\mathcal{A}_{\mathrm{w}}$包含抖动以及放缩的增强方式,仅做细微的调整,保留主要特征,为模型提供可靠的学习锚点。

2.4 异常注入与异常感知

为了解决时间序列标注数据的不足,DAT将掩码语言建模应用到时间序列上,通过人工合成异常的方式实现类似的操作,结合Transformer架构,将正常与异常样本进行建模[7]。在训练阶段,异常注入模块通过预先定义的异常合成方法以及异常比例r,生成人工合成的异常数据,替换掉输入序列的随机位置,异常注入方式如图2所示。对于序列$\boldsymbol{S}_t=(\boldsymbol{x}_{t_0},\boldsymbol{x}_{t_0+1},\cdots,\boldsymbol{x}_{t_1})$,采用如下合成方式。

1)混合替换。选择子区间$[t_0',t_1']\subseteq[t_0,t_1]$,将该区间的观测值替换为窗口外随机的相同大小区间数据与其自身的算术平均值,并对替换后的数据沿横轴、纵轴或二者联合的翻转操作。

2)均值覆盖。在序列的取值范围内进行1次随机采样,得到随机常数$c\in[\min(\boldsymbol{x}_i),\max(\boldsymbol{x}_j)]$,将区间$[t_0',t_1']\subseteq[t_0,t_1]$替换为$c$。

3)峰值噪声。随机选取$r\times|\boldsymbol{S}_t|$个时间点,将观测值$x_{t_i}$替换为人工合成的异常值,合成规则为

$$x'_{t_i}=x_{t_i}+(1-2\times x_{t_i})\times\epsilon, \tag{2}$$

其中$\epsilon\sim Uniform[0,1)$是均匀分布随机变量。噪声项的幅度和方向由$x_{t_i}$动态调节:当$x_{t_i}>0.5$时,噪声项为

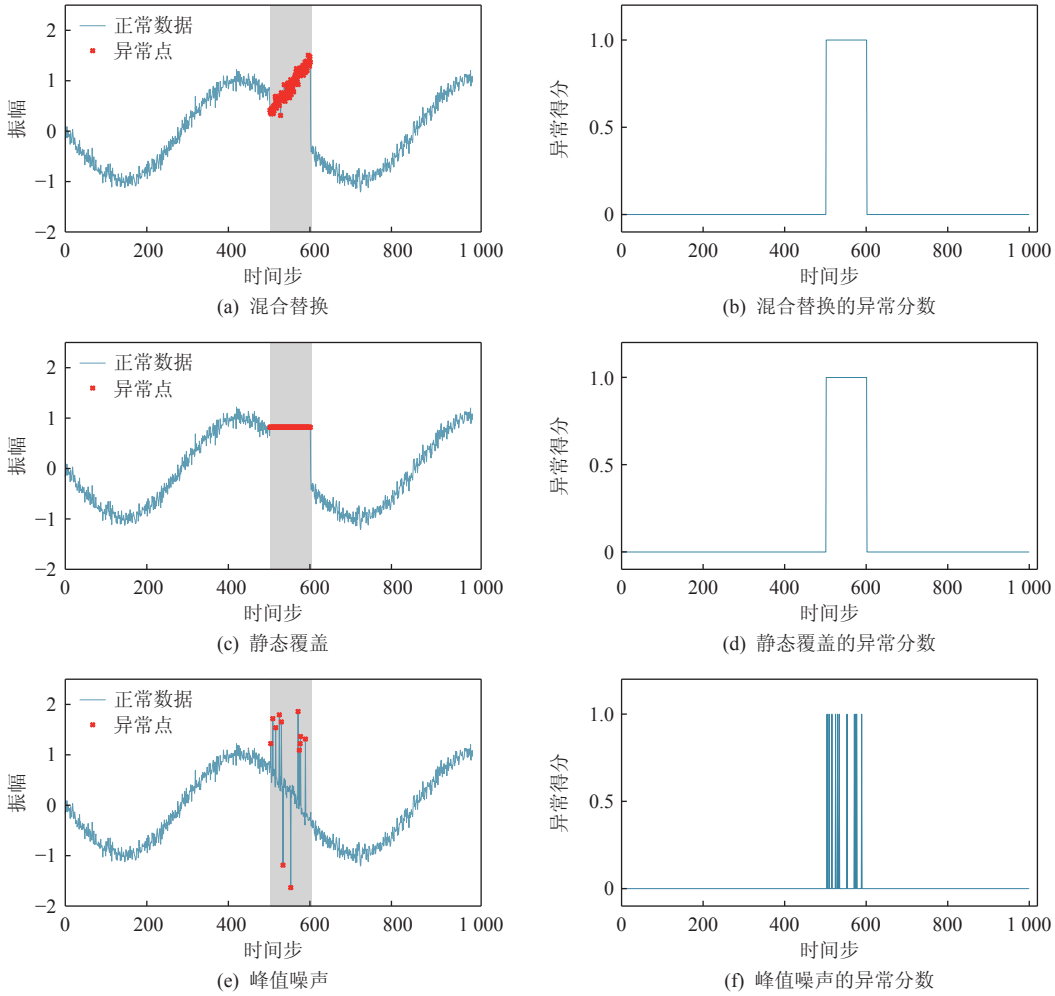


Fig. 2 Synthetic anomaly

图2 合成异常

负向扰动；当$x_{t_i}<0.5$时，噪声项为正向扰动。

之后，得到输入数据$\boldsymbol{S}_t$的真实标注$y=\{0,1\}^L$。2个分支的视图在相同位置采用同样的合成方式注入异常。

如图1所示，异常注入后2个视图会通过一个简单的线性映射，从原来的d维映射到嵌入空间d_{embed}，接着通过各分支上的Transformer块以及简单的全连接网络进行异常感知。如图1右侧所示，Transformer块由N个Transformer编码器堆叠而成。DAT采用“预规范化”的Transformer编码器，为训练提供更为稳定的梯度[39]，且2个分支共享权重。此外，Transformer块在每个注意力矩阵中添加1维的相对位置偏置记录窗口内特征的相对位置[40]。

对于强增强视图$\boldsymbol{S}_t^{\text{s}}\in\mathbb{R}^{L\times d_{\text{embed}}}$，在强增强的分支Transformer块中，每个自注意力头的查询、键以及值矩阵按如下公式初始化：

$$\boldsymbol{\mathcal{Q}}_i^{\text{s}}=\boldsymbol{\mathcal{W}}_{\boldsymbol{\mathcal{Q}}_i}\boldsymbol{S}_i^{\text{s}},\ \boldsymbol{\mathcal{K}}_i^{\text{s}}=\boldsymbol{\mathcal{W}}_{\boldsymbol{\mathcal{K}}_i}\boldsymbol{S}_i^{\text{s}},\ \boldsymbol{\mathcal{V}}_i^{\text{s}}=\boldsymbol{\mathcal{W}}_{\mathcal{V}_i}\boldsymbol{S}_i^{\text{s}},\ 1\leqslant i\leqslant H,\tag{3}$$

其中$\boldsymbol{\mathcal{Q}}_i^{\text{s}},\boldsymbol{\mathcal{K}}_i^{\text{s}},\boldsymbol{\mathcal{V}}_i^{\text{s}}\in\mathbb{R}^{L\times\frac{d_{\text{embed}}}{H}}$表示强增强分支的查询、键和值矩阵，$\boldsymbol{\mathcal{W}}_{\mathcal{Q}_i},\boldsymbol{\mathcal{W}}_{\mathcal{K}_i},\boldsymbol{\mathcal{W}}_{\mathcal{V}_i}$分别是$\boldsymbol{\mathcal{Q}}_i^{\text{s}},\boldsymbol{\mathcal{K}}_i^{\text{s}},\boldsymbol{\mathcal{V}}_i^{\text{s}}\in\mathbb{R}^{\frac{d_{\text{embed}}}{H}\times\frac{d_{\text{embed}}}{H}}$的可学习矩阵，$H$表示自注意力头个数。计算每个自注意力头的注意力得分：

$$Atten_i^{\text{s}}\left(\boldsymbol{\mathcal{Q}}_i^{\text{s}},\boldsymbol{\mathcal{K}}_i^{\text{s}},\boldsymbol{\mathcal{V}}_i^{\text{s}}\right)=Softmax\left(\frac{\boldsymbol{\mathcal{Q}}_i^{\text{s}}\boldsymbol{\mathcal{K}}_i^{\text{s}\text{T}}}{\sqrt{d_{\text{embed}}}}+\boldsymbol{B}_i\right)\boldsymbol{\mathcal{V}}_i^{\text{s}},\tag{4}$$

其中$Softmax(\cdot)$函数用于归一化注意力权重，$\boldsymbol{B}_i$表示自注意力头i中输入特征的偏移量，d_{embed}表示输入序列嵌入空间的维度数。最后，再将每个自注意力头的输出拼接起来并乘以所有自注意力头共享的全局投影矩阵$\boldsymbol{\mathcal{W}}^{\text{output}}$。

类似地，对于弱增强视图$\boldsymbol{S}_t^{\text{w}}\in\mathbb{R}^{L\times d_{\text{embed}}}$，在弱增强分支中Transformer块中采取与强增强视图相同的操作。此外，$\boldsymbol{\mathcal{W}}_{\boldsymbol{\mathcal{Q}}},\boldsymbol{\mathcal{W}}_{\boldsymbol{\mathcal{K}}},\boldsymbol{\mathcal{W}}_{\mathcal{V}},\boldsymbol{\mathcal{W}}^{\text{output}}$是共享权重的。

随后，将Transformer块提取的表征$(\boldsymbol{C}_t^{\text{s}},\boldsymbol{C}_t^{\text{w}})$，输入对应的预测网络，得到双分支异常感知得分$\hat{\boldsymbol{y}}^{\text{s}},\hat{\boldsymbol{y}}^{\text{w}}$。通过均方误差损失函数，计算与真实$\boldsymbol{y}$的损失：

$$\mathcal{L}_{\text{ts}}^{\text{s}}=\|\hat{\boldsymbol{y}}^{\text{s}}-\boldsymbol{y}\|^2,\ \mathcal{L}_{\text{ts}}^{\text{w}}=\|\hat{\boldsymbol{y}}^{\text{w}}-\boldsymbol{y}\|^2。\tag{5}$$

加权平均得到异常感知损失$\mathcal{L}_{\text{a}}$：

$$\mathcal{L}_{\text{a}}=\omega_{\text{s}}\times\mathcal{L}_{\text{ts}}^{\text{s}}+\omega_{\text{w}}\times\mathcal{L}_{\text{ts}}^{\text{w}},\tag{6}$$

其中$\boldsymbol{y}\in\mathbb{R}^L$是异常注入后的真实标注，$\hat{\boldsymbol{y}}^{\text{s}}\in\mathbb{R}^L$和$\hat{\boldsymbol{y}}^{\text{w}}\in\mathbb{R}^L$分别是强增强分支和弱增强分支输出的预测异常分数，$\omega_{\text{s}},\omega_{\text{w}}$分别是双分支的权重。

2.5 基于STL的异常统一建模

我们基于STL（seasonal-trend decomposition using loess）时间序列分解思想，引入异常扰动建模框架，统一刻画不同类型的异常行为。STL将时间序列$x(t)$分解为趋势$Tre(t)$、季节$Sea(t)$与残差$Res(t)$ 3部分[41]：

$$x(t)=Tre(t)+Sea(t)+Res(t)。\tag{7}$$

在正常情况下，序列的观测行为可以由式(7)充分解释。当异常发生时，序列将偏离其正常结构。为刻画这种偏离，本文引入异常扰动项$Ano(t)$，构建通用的异常类型：

$$x^*(t)=Tre(t)+Sea(t)+Res(t)+Ano(t),\tag{8}$$

其中当t处于正常状态时$Ano(t)=0$，当t处于异常状态时$Ano(t)\neq0$。式(8)表明，不同异常类型可视为扰动函数$Ano(t)$的不同结构形式，从而在统一框架下进行刻画。

2.5.1 典型异常原型的公式化表达

我们将异常注入的3类异常与典型的异常原型进行关联。

1)混合替换。混合替换表现为子序列结构的改变，例如周期模式破坏、频率变化或形态替换。其扰动函数可定义为

$$Ano_{\text{mixed}}(t)=h(t)\times\mathbb{1}_{\{t_0\leqslant t<t_1\}},\tag{9}$$

其中$h(t)$为异常区间内的偏离函数，用以描述替换后的模式结构，$\mathbb{1}_{\{\cdot\}}$为指示函数，当且仅当条件成立取值为1。当$h(t)$主要改变季节结构时，该异常可视为季节异常的特例；当其主要改变趋势结构时，则可视为暂态变化/水平跃迁的扩展形式。

2)均值覆盖。均值覆盖表现为序列均值在某一时刻发生结构性变化，可建模为阶跃函数

$$Ano_{\text{constant}}(t)=c\times1_{\{t\geqslant t_0\}},\tag{10}$$

其中c表示均值覆盖中采样的随机参数。式(10)刻画了从t_0开始持续存在的水平漂移，对应于经典水平漂移异常。当漂移仅在有限区间内存在时，可进一步表示为2个阶跃函数之差，从而对应于短暂变化异常。

3)峰值噪声。峰值噪声表现为序列在孤立时间点上的瞬时突变。其扰动函数可建模为脉冲函数

$$Ano_{\text{spike}}(t)=\lambda\times\mathbb{1}_{\{t=t_0\}},\tag{11}$$

其中λ表示异常幅值，t_0为异常发生时刻，对应于经典的加性异常。

基于上述分析，本文将上述视为3类基本异常原型，记为异常基函数集合：

$$Ano=\{\text{mixed},\text{constant},\text{spike}\}。\tag{12}$$

任意一段含有异常的序列均可表示为这些基异常的组合形式：

$$Ano_{\text{syn}}(t)=\sum_{k}\lambda_k Ano_k(t),\ Ano_k(t)\in Ano, \tag{13}$$

其中λ_k为各类异常的占比权重。

2.5.2 真实异常与合成异常的近似关系

真实异常往往具有更复杂的结构特征，难以完全由有限数量的异常原型精确描述。因此，本文假设真实异常$Ano_{\text{real}}(t)$可由合成异常$Ano_{\text{syn}}(t)$近似表示：

$$Ano_{\text{real}}(t)\approx Ano_{\text{syn}}(t)+\varepsilon(t), \tag{14}$$

其中$\varepsilon(t)$表示模型残差，反映真实异常与合成异常之间的差异。该残差项揭示了合成异常的“覆盖误差”，当$\varepsilon(t)$较小时，合成异常能够较为正确地刻画真实异常；当$\varepsilon(t)$较大时，则说明真实异常包含未被异常原型捕捉的结构特征。由于$\varepsilon(t)$难以在时域中直接度量，本文在实验部分通过表示空间中的距离间接评估真实异常与合成异常之间的分布差异。

2.6 上下文对比学习

为强化双视图语义一致性，DAT引入上下文对比模块。该模块通过跨视图的同点对比学习，协同数据增强共同提取语义敏感且噪声鲁棒的判别性表征。

由Transformer块所提取的表征$(\boldsymbol{C}_t^{\text{s}},\boldsymbol{C}_t^{\text{w}})$输入到上下文对比学习模块。对于有$N$个输入样本的序列，每个样本都有2个来自不同增强视图的上下文表征$(\boldsymbol{C}_t^{\text{s}},\boldsymbol{C}_t^{\text{w}})$，因此一共有$2N$个上下文。给定上下文$\boldsymbol{c}_t^i$，其正样本$\boldsymbol{c}^{i+}$取自另一个视图中的相同位置，从而$(\boldsymbol{c}_t^i,\boldsymbol{c}_t^{i+})$构成一个正样本对。同时，同一批次中的其他$2N-2$个上下文可以被视为$\boldsymbol{c}_t^i$的负样本，即$\boldsymbol{c}_t^i$可以与其负样本形成$2N-2$个负样本对。参考InfoNCE对比损失[27,42]，得到一个形式上与之相近的上下文对比损失$\mathcal{L}_{\text{c}}$，以最大化正样本对的相似性，最小化负样本对的相似性：

$$\mathcal{L}_{\text{c}}=-\sum_{i=1}^{N}\log\frac{\exp\left(\dfrac{sim\left(\boldsymbol{c}_t^i,\boldsymbol{c}_t^{i+}\right)}{\tau}\right)}{\displaystyle\sum_{m=1}^{2N}\mathbb{1}_{\{m\neq i\}}\exp\left(\dfrac{sim\left(\boldsymbol{c}_t^i,\boldsymbol{c}_t^m\right)}{\tau}\right)}, \tag{15}$$

其中$\mathbb{1}\in\{0,1\}$是一个指示函数，当且仅当$i\neq m$时取值为1，$sim(\boldsymbol{u},\boldsymbol{v})=\dfrac{\boldsymbol{u}^{\text{T}}\boldsymbol{v}}{\|\boldsymbol{u}\|\|\boldsymbol{v}\|}$表示2个归一化的向量$\boldsymbol{u}$和$\boldsymbol{v}$之间的点积(即余弦相似度)，$\tau$是一个温度参数，控制模型对负样本的区分程度。

优化上下文对比损失函数的过程本质上是优化每对正样本之间的互信息，且相关研究也已经证明了如下关系[42]：

$$I(\boldsymbol{c}_t,\boldsymbol{c}_t^{+})\geqslant\log N-\mathcal{L}_{\text{c}}, \tag{16}$$

其中$I(\boldsymbol{c}_t,\boldsymbol{c}_t^{+})$表示批次内正样本对的互信息，$N$表示输入批次的大小。最小化$\mathcal{L}_{\text{c}}$的过程实际上是最大化$I(\boldsymbol{c}_t,\boldsymbol{c}_t^{+})$的下界，随着$\mathcal{L}_{\text{c}}$的减小，下界会逐渐增大，间接提高了$I(\boldsymbol{c}_t,\boldsymbol{c}_t^{+})$。

最终，模型结合异常感知损失$\mathcal{L}_{\text{a}}$和对比损失$\mathcal{L}_{\text{c}}$完成联合学习，得到最终损失$\mathcal{L}$：

$$\mathcal{L}=\lambda\mathcal{L}_{\text{a}}+(1-\lambda)\mathcal{L}_{\text{c}}, \tag{17}$$

其中λ用于调节对比损失和异常感知损失的权重，异常感知损失$\mathcal{L}_{\text{a}}$是2种增强视图在异常注入和异常感知之后计算的损失的加权平均。

2.7 异常检测

异常检测的具体步骤如图3所示。在异常检测的过程中，测试数据$\boldsymbol{X}$经过滑动窗口处理之后，同步输入到强、弱增强分支生成对应视图。随后由Transformer及预测网络提取特征并计算对应异常得分$\hat{\boldsymbol{y}}$。如果$\hat{y}_t$大于预定义的阈值δ，则该时间点被视为异常点，反之为正常点。

$$\hat{y}_t=\begin{cases}1, & \text{if } \hat{y}_t\geqslant\delta,\\ 0, & \text{其他}。\end{cases} \tag{18}$$

与训练不同的是，窗口的滑动步长等于窗口的长度，即每次输入的是互不重叠的子序列。

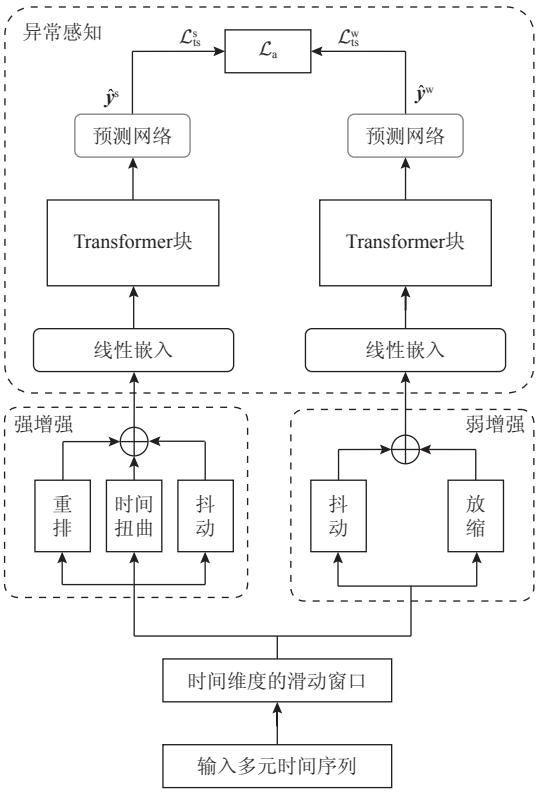


Fig. 3 Anomaly detection

图3 异常检测

3 实　　验

3.1 数据集

实验一共涉及 5 种数据。如表 1 所示，数据集涵盖航天、地球科学、信息技术、环境工程等领域。每个数据集被划分为训练数据、测试数据以及测试数据的异常标签。

Table 1　Benchmark Datasets of Time Series

表 1　时间序列基准数据集

数据集	维度	训练数据条数	测试数据条数	异常率/%
MSL[17]	55	58 317	73 729	10.5
SMAP[17]	25	135 183	427 617	12.8
PSM[43]	25	132 481	87 841	27.8
SMD[11]	38	708 405	708 420	4.2
SWaT[44]	51	495 000	449 919	12.1

3.2 基线模型与评价指标

将 DAT 与 21 种基线模型进行综合比较，包括基于密度的估计方法：LOF[45], MMPCACD[46], DAGMM[47]；基于聚类的方法：ITAD[48], THOC[49], Deep-SVDD[50]；传统方法：IForest[51], OCSVM[10]；基于重构的方法：Anomaly Transformer （AnomalyTrans）[4], LSTM-VAE[5], Omni-Anomaly[11], BeatGAN[12], InterFusion[13], NPSR[16]；基于预测的方法：LSTM[17], CL-MPPCA[18]；以及基于对比学习的方法：CL-TAD[28], PACdetector[29], AWSL[30], MultiRC[31], DCdetector[32]。

DAT 在推理时对输入序列的异常得分，由阈值进一步调整为{0,1}的离散序列。因此，异常检测被视为一个二分类问题。其常见的评估指标包括精确率（P）、召回率（R）以及 $F1$ 分数（$F1$）。同时，还可以采用点调整（point adjustment, PA）方法检查异常[14]。实验将以 PA 策略计算 P、R 和 $F1$。

然而，最近的一些研究指明了 PA 策略的缺陷。因此，实验额外采用其他一些广泛使用的指标作为补充，如关联精确率（*Aff-P*）和关联召回率（*Aff-R*）。*Aff-P* 和 *Aff-R* 是基于真实事件和预测事件之间的距离计算的[52]。

3.3 实验设置

实验超参数设置如下：强增强算子 $\mathcal{A}_s$ 由抖动、时间扭曲和重排构成，权重分别是 0.3，0.6，0.1；弱增强算子 $\mathcal{A}_w$ 则包含权重均为 0.5 的抖动和放缩。异常注入采用混合替换、均值覆盖和峰值噪声 3 种方式，其发生概率分别设定为 0.5, 0.15, 0.25。数据在进入 Transformer 模块前均被映射到了 $d_{\text{embed}}=256$ 维的空间上。Transformer 块由 $N=3$ 个编码器组成，且每个编码器包含 $H=8$ 个自注意力头。在异常感知阶段，强增强分支的得分与弱增强分支的得分权重分别是 $\omega_s=0.6$，$\omega_w=0.4$。对比损失温度系数 τ 参考相关研究[53] 取值 0.1；$\mathcal{L}_c$ 与 $\mathcal{L}_a$ 在联合学习的权重为 $\lambda=0.8$。在阈值选择方面，以 10^{-5} 作为风险水平，训练数据异常得分的 99% 分位数作为初始阈值，利用 SPOT 算法[54] 确定最终阈值。

3.4 结果和分析

表 2 展示了 DAT 与基线模型在 PA 策略的 P, R 以及 $F1$。观察发现，在所有的基准数据集中，DAT 采用 PA 策略获得的 $F1$ 达到了均值以上，接近最优水平，甚至在 MSL 和 PSM 上达到 SOTA 水平。然而，DAT 在 MSL 和 SMAP 上的精确率较低（86.88% 和 85.89%），表明 DAT 虽能广泛覆盖潜在异常，但在领域内特异性的正常模式建模能力稍显不足。

鉴于 DCdetector 在基线模型中的突出表现，我们在多指标比较中将 DAT 与其进行对比，结果见表 3。结果表明，在 P、R 和 $F1$ 相当的情况下，DAT 的关联性指标性能同样出色。

以 SWaT 和 PSM 数据集为例，图 4 较为详细地展示 DAT 在处理含有未知异常样本的时间序列数据时的异常感知能力。“异常片段”展示输入序列的真实状况以及所需感知的异常成分；“数据分布”是原始数据经过 t-SNE[55] 降维处理后的 2 维可视化结果；“预测分数”则是 DAT 模型对该片段进行异常感知之后的预测得分，以及学习到的潜在向量在 2 维空间的投影。

可以发现，经过异常注入和异常感知的训练后，正常点和异常点的分布呈现出明显的分离趋势，并且开始各自聚集，这表明 DAT 一定程度上学习到了正常点和异常点的分布差异，有效地提取了区分正常与异常状态的关键特征。同时，异常点普遍获得了相对高的预测分数，而正常点得到了相对降低的分数。这种预测结果进一步证实了 DAT 在异常检测任务中的有效性。

为验证上下文对比损失对提取判别性特征的作用，本文以 MSL 数据集为例，可视化了学习过程中双分支特征在 2 维空间的投影演变。

如图 5 所示，随着训练轮次的迭代，在对比损失函数的引导下，双分支学习的特征逐渐呈现出聚类的趋势：簇内正样本对相互靠近，加强了语义的一致

Table 2 Performance Comparison of Anomaly Detection Algorithms

表 2 异常检测算法的性能对比 %

模型	MSL			PSM			SWaT			SMAP			SMD		
	P	*R*	*F*1	*R*	*R*	*F*1	*P*	*R*	*F*1	*P*	*R*	*F*1	*P*	*R*	*F*1
LOF[45]	88.87	59.96	71.61	57.89	90.49	70.61	69.25	62.33	65.61	87.97	52.03	65.39	56.34	39.86	46.68
OCSVM[10]	36.57	56.35	44.35	58.64	77.95	66.93	46.77	49.30	48.00	50.56	61.27	55.42	43.64	65.82	52.48
IForest[51]	80.29	62.23	70.12	84.64	88.83	86.68	28.66	94.45	43.97	85.20	56.51	67.95	42.70	75.99	54.68
DAGMM[47]	89.60	63.93	74.62	93.49	70.03	80.08	89.92	57.84	70.40	86.45	56.73	68.51	67.30	49.89	57.30
ITAD[48]	69.44	84.09	76.07	72.80	64.02	68.13	63.13	52.08	57.08	82.42	66.89	73.85	86.22	73.71	79.48
MMPCACD[46]	81.42	61.31	69.95	76.26	78.35	77.29	82.52	68.29	74.73	88.61	75.84	81.73	71.20	79.28	75.02
CL-MPPCA[18]	73.71	88.54	80.44	56.02	**99.93**	71.80	76.78	81.50	79.07	86.13	63.16	72.88	82.36	76.07	79.09
Deep-SVDD[50]	91.92	76.63	83.58	95.41	86.49	90.73	80.42	84.45	82.39	89.93	56.02	69.04	78.54	79.67	79.10
LSTM-VAE[5]	85.49	79.94	82.62	73.62	89.92	80.96	76.00	89.50	82.20	92.20	67.75	78.10	75.76	90.08	82.30
BeatGAN[12]	89.75	85.42	87.53	90.30	93.84	92.04	64.01	87.46	73.92	92.38	55.85	69.61	72.90	84.09	78.10
LSTM[17]	85.45	82.50	83.95	76.93	89.64	82.80	86.15	83.27	84.69	89.41	78.13	83.39	78.55	85.28	81.78
OmniAnomaly[11]	78.48	**99.98**	87.94	85.77	75.32	80.21	97.60	69.57	81.23	**97.60**	69.57	81.23	88.78	**99.85**	93.99
InterFusion[13]	81.28	92.70	86.62	83.61	83.45	83.52	80.59	85.58	83.01	89.77	88.52	89.14	87.02	85.43	86.22
THOC[49]	88.45	90.97	89.69	88.14	90.99	89.54	83.94	86.36	85.13	92.06	89.34	90.68	79.76	90.95	84.99
AnomalyTrans[4]	85.37	97.98	91.24	94.14	90.88	92.48	**99.77**	68.79	81.43	80.43	98.82	88.68	**90.72**	99.73	**95.01**
DCdetector[32]	91.34	93.53	92.42	94.52	98.94	96.68	90.13	**99.96**	**94.79**	92.31	96.27	**94.25**	82.89	88.58	85.64
MultiRC[31]	86.56	84.01	85.27	93.62	87.95	90.70	72.58	85.58	78.55	95.80	65.50	77.80	79.17	90.19	84.32
NPSR[16]	87.65	75.74	81.26	64.85	91.01	75.73	89.10	90.02	89.56	65.48	85.97	74.34	65.46	80.84	72.34
AWSL[30]	74.44	73.56	74.00	78.91	85.24	81.95	75.52	85.55	80.22	85.93	90.38	88.10	72.17	86.88	78.85
PACdetector[29]	**95.18**	86.26	90.50	97.22	82.04	88.98	78.96	87.12	82.84	96.99	81.04	88.30	70.03	81.64	75.39
CL-TAD[28]	89.04	56.74	69.31	95.95	81.14	87.93	95.99	70.37	81.21	95.72	69.16	80.30	84.25	65.88	73.94
DAT（本文）	86.88	99.27	**92.66**	**97.47**	96.71	**97.09**	90.90	91.18	91.04	85.89	**99.85**	92.35	88.90	92.72	90.77

注：黑体数值为指标最优值，下划线数值为指标次优值。

Table 3 Comparison of Multi-Metric Performance

表 3 多指标性能对比 %

数据集	算法	评价指标				
		P	*R*	*F*1	*Aff-P*	*Aff-R*
MSL	DCdetector	**91.34**	93.53	92.42	50.28	**93.19**
	DAT（本文）	86.88	**99.27**	**92.66**	**54.09**	87.63
PSM	DCdetector	94.52	**98.94**	96.68	51.64	**94.51**
	DAT（本文）	**97.47**	96.71	**97.09**	**61.26**	89.45
SWaT	DCdetector	90.13	**99.96**	**94.79**	51.93	97.04
	DAT（本文）	**90.90**	91.18	91.04	**53.11**	**97.57**
SMAP	DCdetector	**92.31**	96.27	**94.25**	49.52	96.47
	DAT（本文）	85.89	**99.85**	92.35	**50.82**	**97.60**

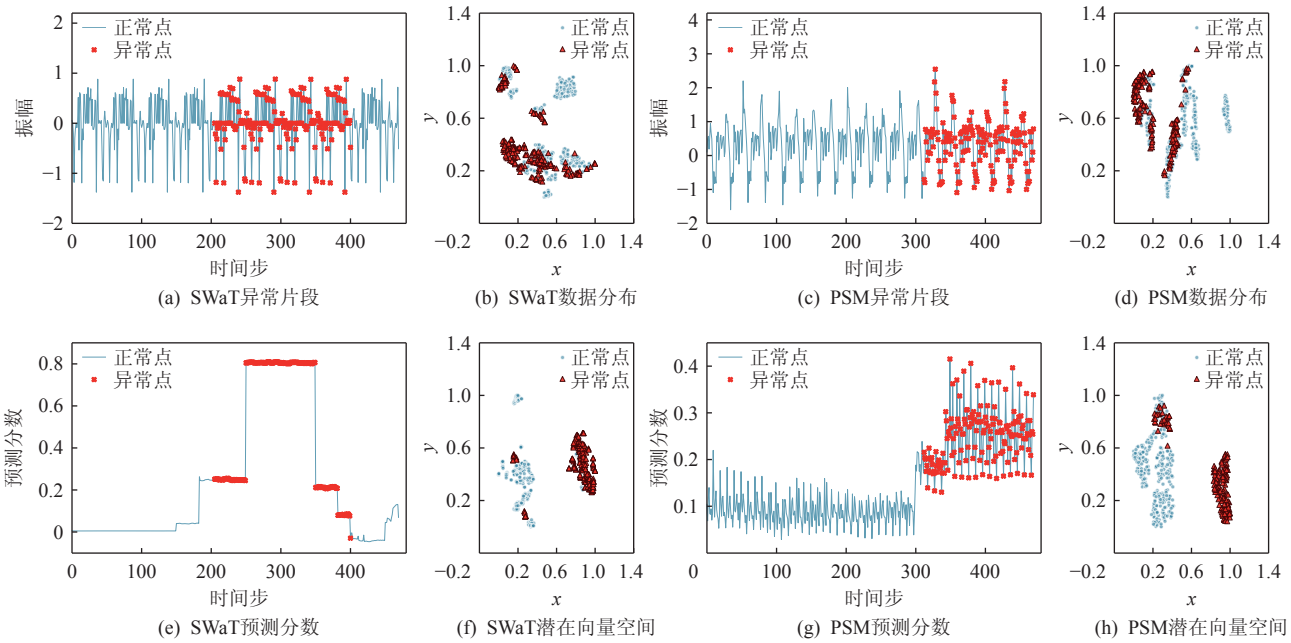
注：黑体数值为指标最优值。

性；而簇间负样本对则相互远离，提升了判别性。为了更准确地分析这一现象，我们进行了定量分析，计算了双表征中正样本对和负样本对的平均余弦相似度、平均欧氏距离和平均 KL 散度，结果如表 4 所示。随着训练轮次的增加，正样本对之间的相似度在稳定提高，而负样本对的相似度也有下降的趋势。

3.5 消融分析

本节进行消融实验分析，探究数据增强模块、异常感知和异常注入模块、上下文对比学习模块对 DAT 性能的影响。DAT-Aug 是没有数据增强模块的模型；DAT-Ano 是没有异常注入与异常感知模块的模型；DAT-Con 是没有上下文对比学习模块的模型。图 6 展示了 3 类简化模型在 MSL, PSM, SMAP, SWaT 数据集上的表现。

观察发现，3 类模型的 *F*1 均有不同程度的下降：DAT-Aug 平均下降了 17%，DAT-Ano 平均下降了约 23%，DAT-Con 平均下降了约 25%。说明引入多角度的数据增强模块能够在扩充训练数据的多样性的同时，增强序列原本的特征。而异常注入模块与异常感知模块的结合，能够有效提升 DAT 对异常数据的识别能力。上下文对比学习模块的引入，有助于模

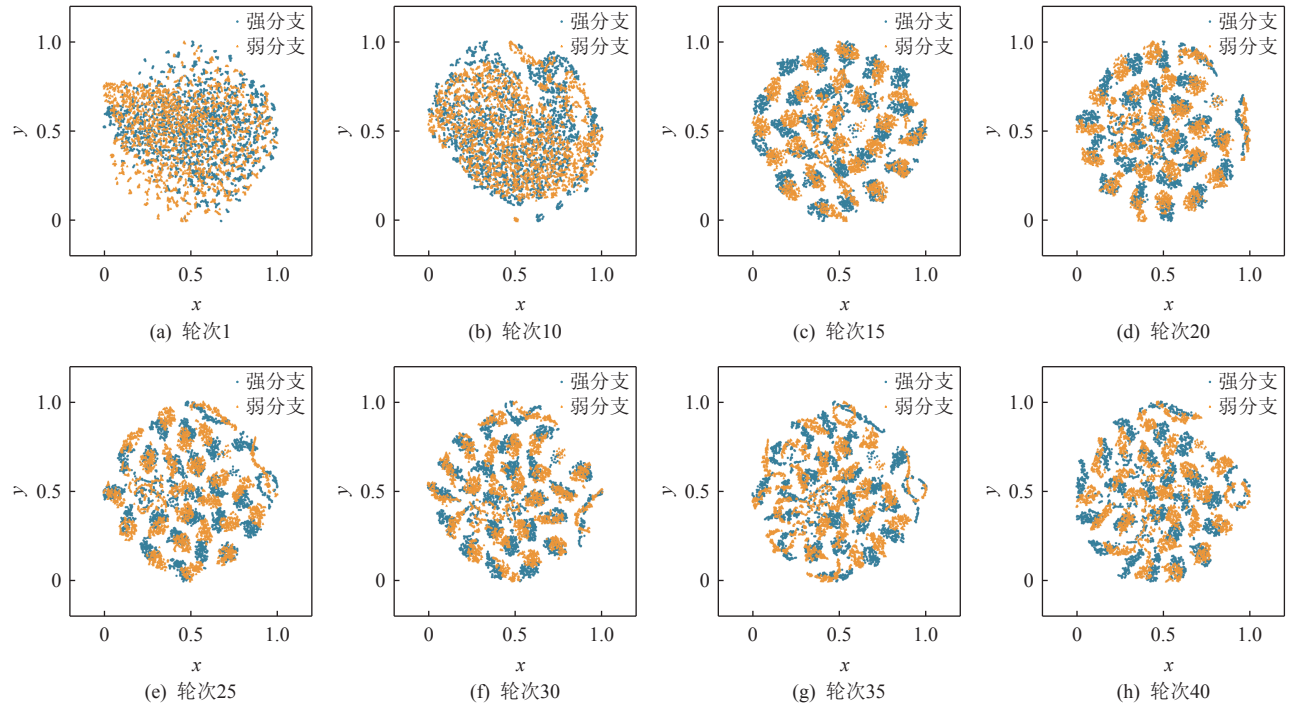


(a) SWaT异常片段 (b) SWaT数据分布 (c) PSM异常片段 (d) PSM数据分布

(e) SWaT预测分数 (f) SWaT潜在向量空间 (g) PSM预测分数 (h) PSM潜在向量空间

Fig. 4 Visualization of anomaly awareness for SWaT and PSM

图 4 SWaT 及 PSM 的异常感知可视化



(a) 轮次1 (b) 轮次10 (c) 轮次15 (d) 轮次20

(e) 轮次25 (f) 轮次30 (g) 轮次35 (h) 轮次40

Fig. 5 The distributions change in dual branches representation

图 5 双分支表征分布变化

型更好地理解数据的上下文信息，学习可辨识的语义特征。

同时，为评估 3 种异常合成方式的有效性，我们进行了消融实验。表 5 展示了异常合成方式在 MSL, PSM, SWaT, SMAP 数据集上对 $F1$ 的影响，已有研究将序列异常划分为 5 种基于行为特征的分类体系：全局异常、上下文异常、形状异常、季节异常和趋势异常[56]。基于此的异常合成方式需要具有对原数据的先验知识。

但是，表 5 显示，“混合替换”“均值覆盖”“峰值噪声”3 种方法在无需依赖先验知识的情况下，仍能有效提升模型的异常感知能力。这表明异常注入的

Table 4 Quantitative Changes of the Distribution in Dual Branches Representation

表 4 双分支表征分布的定量变化

轮次	余弦相似度		欧氏距离		KL 散度	
	正样本	负样本	正样本	负样本	正样本	负样本
1	0.997-	0.997-	2.325-	2.600-	0.023-	0.016-
10	0.978↓	0.923↓	2.647↑	5.323↓	0.166↑	0.754↑
15	0.981↑	0.089↓	0.113↓	0.998↑	1.433↑	9.346↑
20	0.987↑	0.077↓	0.092↓	0.780↓	1.927↑	9.500↑
25	0.990↑	0.061↓	0.086↓	0.842↑	1.394↓	9.374↓
30	0.989↑	0.022↓	0.099↑	1.011↑	0.716↓	8.928↓
35	0.985↓	0.005↓	0.118↑	1.099↑	0.121↓	8.962↑
40	0.992↑	0.000↓	0.101↓	1.156↑	0.521↑	8.969↑

注: "-"表示较前一期数值持平, "↓"表示较前一期数值下降, "↑"表示较前一期数值上升。

关键并非完全复现异常的具体形态, 而是通过对数据部分分量施加扰动, 引导模型学习区分正常与异常表征的特征。结合多种异常合成方式能够从不同角度扰动原始表征空间, 形成互补。

3.6 参数敏感性分析

如图 7 所示, 通过调整 DAT 的 4 个主要参数进行敏感性分析, 讨论在不同参数的取值下 DAT 在 MSL, PSM, SMAP 数据集上表现。选取的窗口大小是{0, 50, 100, 150, 200}, 块大小为{0, 2, 4, 6, 8, 10}, 注入的异常比例为{0, 0.2, 0.4, 0.6, 0.8}, 每次滑动的步长为{0, 10, 20, 30, 40}。

Table 5 Impact of Synthetic Anomaly on *F*1-Socre

表 5 合成异常对 *F*1 分数的影响 %

合成异常			数据集			
混合替换	均值覆盖	峰值噪声	MSL	PSM	SWaT	SMAP
○	○	○	92.66	97.07	91.04	92.53
○	○	×	84.88	91.60	86.77	78.56
○	×	○	90.22	96.58	88.63	81.56
×	○	○	88.89	95.98	86.18	90.55
○	×	×	83.55	84.52	88.20	89.48
×	○	×	85.45	82.88	90.20	85.38
×	×	○	74.88	81.27	69.98	73.76

注: ○为使用这类合成异常, ×为不使用这类合成异常。

观察发现, *P*、*R* 和 *F*1 在参数的变化的过程中大致呈现出先上升后下降的趋势。而 SMAP 在这三者之中相对稳定, 呈现出较低的敏感性。并且, 当异常注入比例在 [0.2, 0.4] 区间内往往可以得到较好的性能。这可能是因为在这个范围内, 异常数据足以提供有效的异常用于检测, 又不会过多地干扰正常数据的分布, 导致过拟合, 从而使得模型可以更精准地

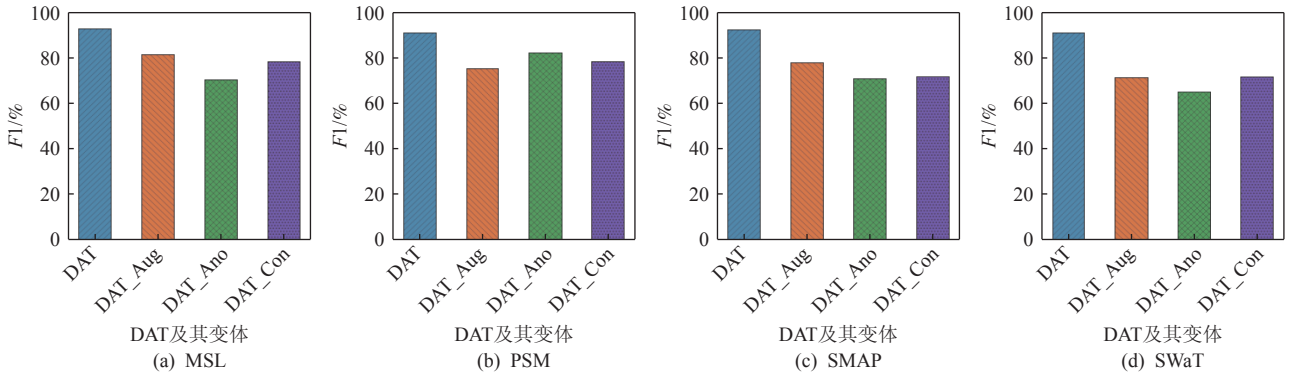


Fig. 6 Results of ablation analysis

图 6 消融分析结果

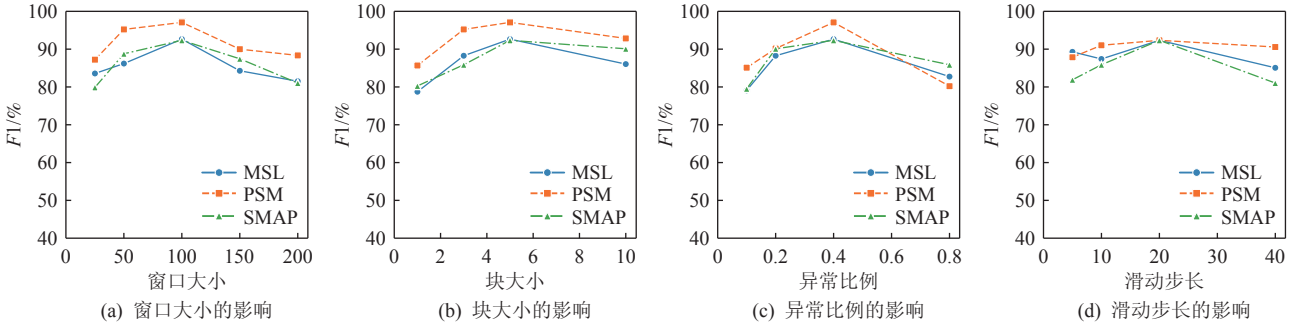


Fig. 7 Parameter sensitivity analysis results on different datasets

图 7 在不同数据集的参数敏感性分析结果

识别异常。

同时，我们通过改变λ的大小，探究异常感知损失和对比损失的联合学习对性能的影响。如图 8 所示，$F1$ 随着λ的增大出现先上升后下降的趋势。当$\lambda \leqslant 0.6$时，对比损失主导，模型过度关注正常样本的表征区分，削弱对异常模式的敏感度；当$\lambda \geqslant 0.9$时，异常感知损失主导，导致特征空间过拟合人工合成异常，减弱对真实异常分布的泛化能力。因此，当$\lambda \in (0.6, 0.9)$时，P 与 R 达到一个平衡，DAT 能够有较高的 $F1$。在 MSL, PSM, SMAP 上 P 与λ呈现显著的正相关，与 R 有负相关；而 SWaT 在$\lambda \in [0.3, 0.9]$时有较好的鲁棒性。λ向异常感知模块的倾斜，体现了 DAT 对检测任务的聚焦。此时，异常感知损失主导最终性能，上下文对比损失则协同优化特征分布，起辅助增强作用。

为了探究强增强算子与弱增强算子内不同增强方式对模型性能的影响，我们分别对其权重分配进行了敏感性分析，如图 9 所示。实验结果表明，模型在权重配置为 [0.3, 0.6, 0.1] 时在所有数据集上均达到最优性能。这说明时间扭曲在强增强中起主导作用；而过高的重排权重会破坏时间序列的局部连续性，导致性能下降。对于弱增强（抖动与放缩），均衡配置为 [0.5, 0.5] 时表现最为稳定。

我们进一步分析了强分支和弱分支对异常感知损失的贡献，结果如图 10 所示。为了探究双分支在协同优化过程中的重要性权衡，我们将强增强分支的权重系数ω_s设置为从 0.1 逐步增加至 0.9（对应的弱增强权重$\omega_w = 1 - \omega_s$）。根据实验结果，随着 ω_s的增加，模型性能呈现出先上升后下降的趋势。具体而言，当权重配置为 [0.6, 0.4]（即$\omega_s = 0.6$）时，DAT 在绝大多数的数据集上均达到了性能峰值。

3.7 误差分析

在 2.5.2 节中，真实的异常与人造异常覆盖误差 $\varepsilon(t)$难以直接度量，为了量化二者在潜空间的分布差异，我们采用最大均值差异（maximum mean discrepancy, MMD）作为分布距离指标，在本次实验的 5 个数据集上比较了 3 类单一合成异常（峰值噪声、均值覆盖、混合替换）及复合异常与真实异常之间的分布差异，结果如表 6 所示。并且我们对 SMD 数据集进行了分布可视化分析，如图 11 所示。

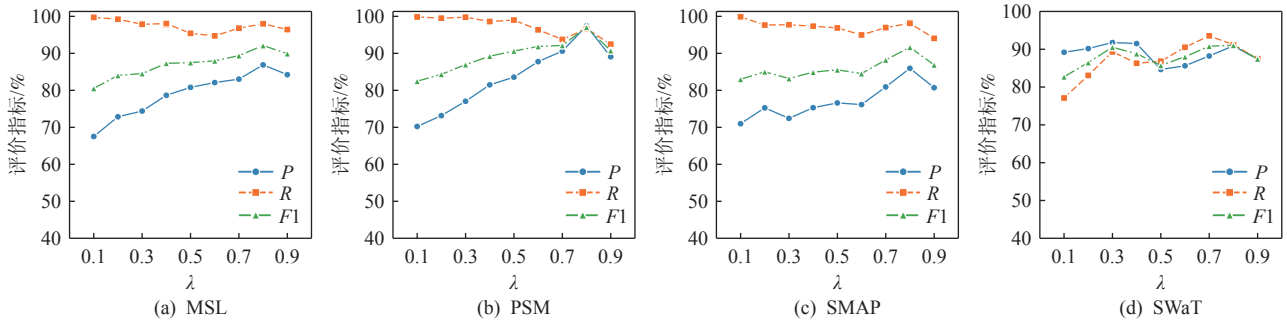


Fig. 8 Impact of λ on the performance of different datasets

图 8 λ 对各数据集性能的影响

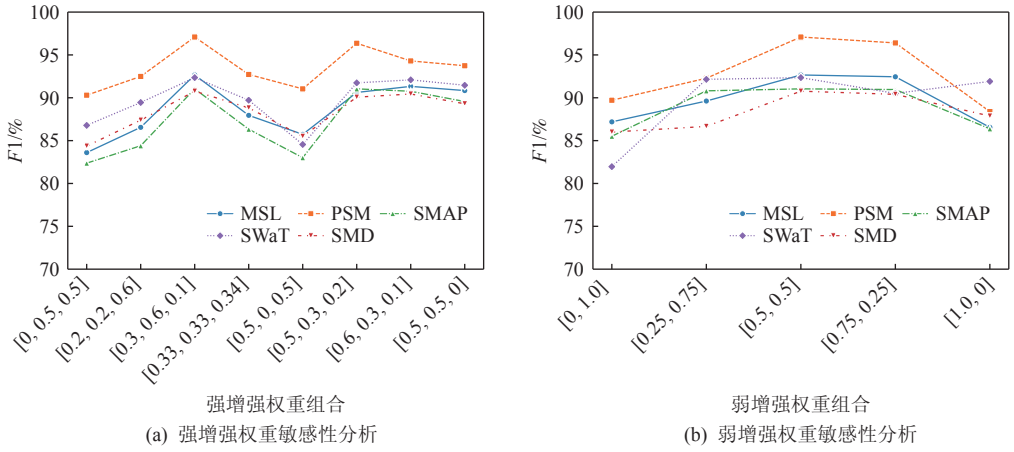


Fig. 9 Sensitivity of strong and weak augmentation weights across different datasets

图 9 强弱增强权重在不同数据集上的敏感性

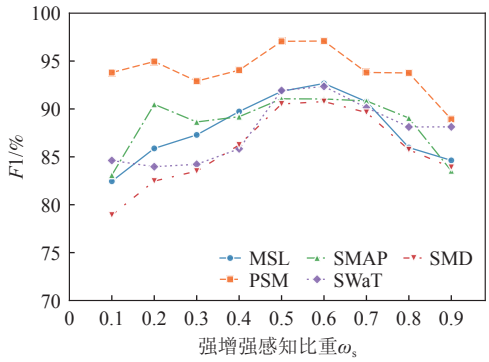


Fig. 10 Parameter sensitivity analysis of dual-branch anomaly-aware weights across different datasets

图 10 双分支异常感知权重在不同数据集上的参数敏感性分析

其中复合异常是指三者按照实验设置的概率{0.25, 0.15, 0.5}组合而成的异常片段。实验结果表明，不同合成异常与真实异常之间的分布差异具有显著的类型与数据集依赖性。均值覆盖与混合替换异常在潜空间中更接近真实异常分布，而复合异常在多数数据集上取得最小或次小的 MMD 值，表明混合多种不同的异常模式能够更好地逼近真实异常的潜在分布。这说明真实异常分布着由多种不同的异常产生机制，单一异常形式难以充分刻画序列的异常特性。

基于此，我们通过量化 3 类异常在各数据集上的

Table 6 Distribution Differences Between Real Anomaly and Each Type of Synthetic Anomaly

表 6 真实异常与各类合成异常的分布差异 %

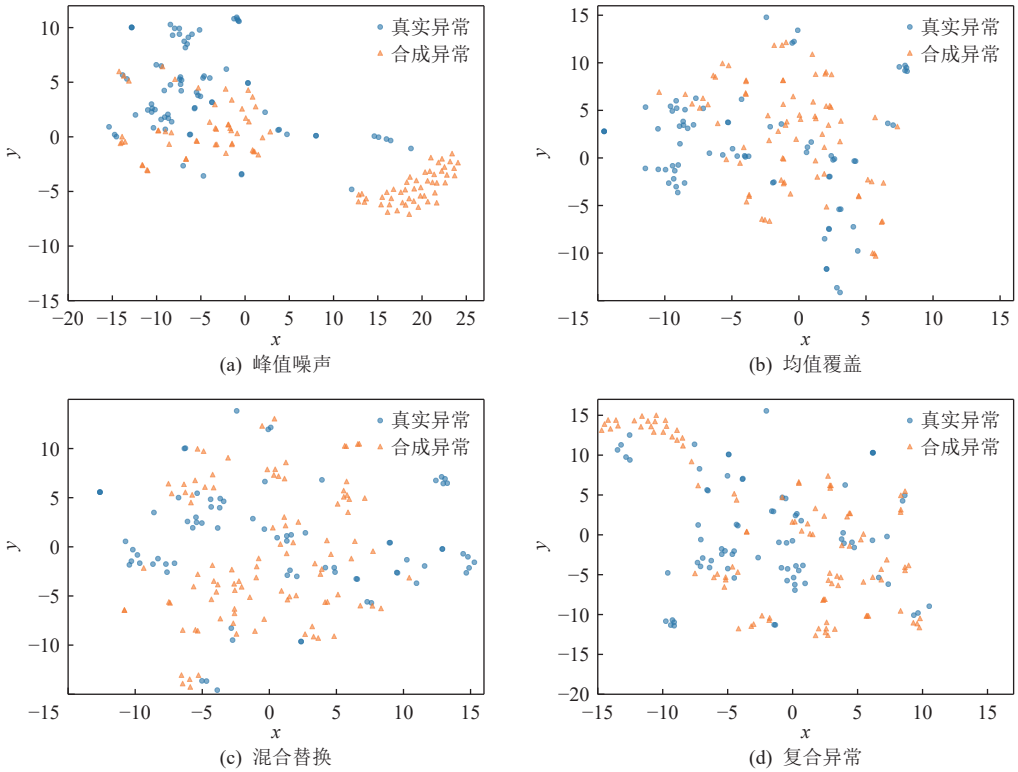
数据集	峰值噪声	均值覆盖	混合替换	复合异常
MSL	2.05±6.96	2.26±0.88	2.01±2.02	2.11±1.23
PSM	9.48±4.20	9.82±2.25	9.52±3.07	8.95±1.08
SMAP	45.41±8.98	16.42±9.51	2.47±5.85	7.22±2.55
SWaT	20.06±6.22	21.65±7.27	21.20±4.88	12.82±3.58
SMD	10.13±4.62	4.53±4.25	5.55±3.95	3.24±1.54

注：数值为最大均值差异（均值±标准差），表示每种人造异常与对应数据集的真实异常之间的分布差异。

相对表现确定其生成比例。按照 MMD 值从小到大排序，我们将排名第 1~3 的异常类型分别赋值为 2 分，1 分，0 分。经统计，峰值噪声、均值覆盖与混合替换在 5 个数据集上的累积得分比为 4∶3∶8。结合实验结果的方差特性，我们在实际注入过程中对该比例进行了微调，最终将 3 类生成概率分别配置为 4∶3∶8≈0.25∶0.15∶0.5。

3.8 鲁棒性分析

在实际应用中，训练数据往往难以保证绝对纯净，通常含有一定比例的未知异常点。为验证 DAT 模型在现实环境下的性能稳定性及实用价值，我们通过向原始训练集中注入不同比例的异常噪声(5%~



(a) 峰值噪声 (b) 均值覆盖

(c) 混合替换 (d) 复合异常

Fig. 11 Distribution of real and synthetic anomalies on SMD dataset

图 11 SMD 数据中真实和合成异常的分布

40%)进行鲁棒性实验。图 12 记录了 MSL, PSM, SWaT 数据集在不同污染程度下的模型性能演变趋势。在训练集污染比例较低(5%~10%)时，DAT 的性能几乎不受影响，表现出较强的鲁棒性。当污染比例超过 20% 后，性能产生显著下降。

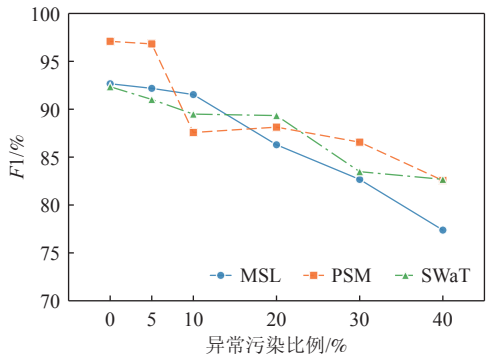


Fig. 12 Robustness analysis across different datasets under varying contamination ratios

图 12 不同数据集在不同污染比例下的鲁棒性分析

4 结 论

本文设计了基于双重异常感知的 Transformer 模型 DAT。通过引入不同视角的数据增强方式让模型从多个视角学习时间序列的潜在表征。并且，建立了异常注入和异常感知机制，引导模型在训练过程中学习正常数据与异常数据表征上的不同，间接完成异常检测的任务。同时，利用上下文的对比表示学习，确保模型在多个视角下的学习可辨别、更本质的表征。最后，我们还对 DAT 进行了多方面的实验。结果表明 DAT 在 PA 策略下能够优于平均水平，甚至在个别公共数据集上能达到最优水平。

但是，目前不同的数据集的增强方式需要研究人员通过试错的方式进行选择，无法根据下游任务自动增强。与图像不同，时间序列的内在模式难以从人类的视角理解，主流的序列增强方式能否保持前后语义一致仍然需要探究。并且 DAT 用的仍是传统的 Transformer 架构，其多头注意力层通过学习序列的关联关系，构建、维护概率矩阵，无意增加了模型的时空开销。未来工作将沿 3 个方向推进：1)将时序数据的增强策略建模为元优化问题，设计基于任务或数据驱动的增强方法；2)为提高模型的在大规模数据上的效率，探索时间序列特化的稀疏注意力机制，或者探究 NLP 领域中相关解决方案在时间序列中的可行性；3)构建包含多粒度异常标注的评估基准，缓解当前指标对点异常检测的局限性。

作者贡献声明：王志鸿负责方案整体设计并撰写论文；黄维负责部分实验思路和修改论文；王德贤、李天瑞负责提出指导意见。

参 考 文 献

[1] Yang Yiyuan, Li Yi , Zhang Taojia, et al. Early safety warnings for long-distance pipelines: A distributed optical fiber sensor machine learning approach[C]//Proc of the 35th AAAI Conf on Artificial Intelligence. Palo Alto, CA: AAAI, 2021: 14991−14999

[2] Archana A, Senthil K, Alexander S, et al. Anomaly detection in finance: Editors' introduction[C]//Proc of the 2017 Workshop on Anomaly Detection in Finance. New York: ACM, 2018: 1−7

[3] Ren Hansheng, Xu Bixiong, Wang Yujing , et al. Time-series anomaly detection service at microsoft[C]//Proc of the 25th ACM SIGKDD Int Conf on Knowledge Discovery & Data Mining. New York: ACM, 2019: 3009−3017

[4] Xu Jiehui, Wu Haixu, Wang Jianmin, et al. Anomaly Transformer: Time series anomaly detection with association discrepancy[C/OL]// Proc of the 10th Int Conf on Learning Representations. 2022[2025-03-01]. https://openreview.net/forum?id=LzQQ89U1qm_

[5] Park D, Hoshi Y, Kemp C C. A multimodal anomaly detector for robot-assisted feeding using an LSTM-based variational autoencoder[J]. IEEE Robotics and Automation Letters, 2018, 3(3): 1544−1551

[6] Liu Xiao, Zhang Fanjin, Hou Zhenyu, et al. Self-supervised learning: Generative or contrastive[J]. IEEE Transactions on Knowledge and Data Engineering, 2021, 35(1): 857−876

[7] Devlin Jacob, Chang Mingwei, Kenton Lee, et al. BERT: Pre-training of deep bidirectional Transformers for language understanding[C]//Proc of the 17th Conf of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Stroudsburg, PA : ACL, 2019: 4171−4186

[8] Carmona C U, Aubet F X, Flunkert V, et al. Neural contextual anomaly detection for time series[C]//Proc of the 31st Int Joint Conf on Artificial Intelligence. San Francisco, CA: Morgan Kaufmann, 2022: 2843−2851

[9] Schölkopf B, Platt J C, Shawe-Taylor J, et al. Estimating the support of a high-dimensional distribution[J]. Neural Computation, 2001, 13(7): 1443−1471

[10] Tax D M, Duin R P. Support vector data description[J]. Machine Learning, 2004, 54(1): 45−66

[11] Su Ya, Zhao Youjian, Niu Chenhao, et al. Robust anomaly detection for multivariate time series through stochastic recurrent neural network[C]//Proc of the 25th ACM SIGKDD Int Conf on Knowledge Discovery & Data Mining. New York: ACM, 2019: 2828−2837

[12] Zhou Bin, Liu Shenhua, Hooi Bryan, et al. BeatGAN: Anomalous rhythm detection using adversarially generated time series[C]//Proc of the 28th Int Joint Conf on Artificial Intelligence. San Francisco, CA: Morgan Kaufmann, 2019: 4433−4439

[13] Li Zhihan, Zhao Youjian, Han Jiaqi, et al. Multivariate time series

anomaly detection and interpretation using hierarchical inter-metric and temporal embedding[C]//Proc of the 27th ACM SIGKDD Conf on Knowledge Discovery & Data Mining. New York: ACM, 2021: 3220−3230

[14] Xu Haowen, Chen Wenxiao, Zhao Nengwen, et al. Unsupervised anomaly detection via variational auto-encoder for seasonal KPIs in web applications[C]//Proc of the 27th World Wide Web Conf. New York: ACM, 2018: 187−196

[15] Xue Bing, Gao Xin, Li Baofeng, et al. A robust multi-scale feature extraction framework with dual memory module for multivariate time series anomaly detection[J]. Neural Networks, 2024, 177: 106395

[16] Lai C Y A, Sun F K, Gao Z, et al. Nominality score conditioned time series anomaly detection by point/sequential reconstruction[C]//Proc of the 36th Annual Conf on Neural Information Processing Systems. Cambridge, MA: MIT, 2023: 76637−76655

[17] Hundman K, Constantinou V, Laporte C, et al. Detecting spacecraft anomalies using LSTMs and nonparametric dynamic thresholding[C]//Proc of the 24th ACM SIGKDD Int Conf on Knowledge Discovery & Data Mining. New York: ACM, 2018: 387−395

[18] Tariq S, Lee S, Shin Y, et al. Detecting anomalies in space using multivariate convolutional LSTM with mixtures of probabilistic PCA[C]//Proc of the 25th ACM SIGKDD Int Conf on Knowledge Discovery & Data Mining. New York: ACM, 2019: 2123−2133

[19] Zhou Xiaohui, Wang Yijie, Xu Hongzuo, et al. Fusion learning based unsupervised anomaly detection for multi-dimensional time series[J]. Journal of Computer Research and Development, 2023, 60(3): 496−508 (in Chinese)
(周小晖, 王意洁, 徐鸿祚, 等. 基于融合学习的无监督多维时间序列异常检测[J]. 计算机研究与发展, 2023, 60(3): 496−508)

[20] Wang Zenan, Wang Yijie, Zhou Xiaohui, et al. Anomaly detection for multivariate time series based on multi-window segmentation ensemble learning[J]. Journal of Computer Research and Development, 2025, 62(12): 2962−2973 (in Chinese)
(王泽南, 王意洁, 周小晖, 等. 基于多窗口划分集成学习的多维时间序列异常检测[J]. 计算机研究与发展, 2025, 62(12): 2962−2973)

[21] Huang Xunhua, Zhang Fengbin, Fan Haoyi, et al. Multimodal adversarial learning based unsupervised time series anomaly detection[J]. Journal of Computer Research and Development, 2021, 58(8): 1655−1667 (in Chinese)
(黄训华, 张凤斌, 樊好义, 等. 基于多模态对抗学习的无监督时间序列异常检测[J]. 计算机研究与发展, 2021, 58(8): 1655−1667)

[22] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C]//Proc of the 31st Int Conf on Neural Information Processing Systems. Cambridge, MA: MIT, 2017: 6000–6010

[23] Yang Zhilin, Dai Zihang, Yang Yang, et al. XLNet: Generalized autoregressive pretraining for language understanding[C]//Proc of the 33rd Conf on Neural Information Processing Systems. Cambridge, MA: MIT, 2019: 5753–5763

[24] Raffel C, Shazeer N, Roberts A, et al. Exploring the limits of transfer learning with a unified text-to-text transformer[J]. Journal of Machine Learning Research, 2020, 21(140): 1−67

[25] Ding Jiayu, Ma Shuming, Dong Li, et al. LongNet: Scaling transformers to 1, 000, 000, 000 tokens[J]. arXiv preprint, arXiv: 2307.02486, 2023

[26] Liu Aixin, Feng Bei, Wang Bin, et al. DeepSeek-v2: A strong, economical, and efficient mixture-of-experts language model[J]. arXiv preprint, arXiv: 2405.04434, 2024

[27] He Kaiming, Fan Haoqi, Wu Yuxin, et al. Momentum contrast for unsupervised visual representation learning[C]//Proc of the 33rd IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2020: 9729−9738

[28] Ngu H C V, Lee K M. CL-TAD: A contrastive-learning-based method for time series anomaly detection[J]. Applied Sciences, 2023, 13(21): 11938

[29] Xu Jiawei, Zhu Zhaowei, Ye Weiwei, et al. Periodicity association based contrastive learning for time series anomaly detection[C]//Proc of the 34th Int Joint Conf on Neural Networks. Piscataway, NJ: IEEE, 2024: 1−8

[30] Zhang Yuxin, Wang Jindong, Chen Yiqiang, et al. Adaptive memory networks with self-supervised learning for unsupervised anomaly detection[J]. IEEE Transactions on Knowledge and Data Engineering, 2022, 35(12): 12068−12080

[31] Hu Shiyan, Zhao Kai, Qiu Xiangfei, et al. MultiRC: Joint learning for time series anomaly prediction and detection with multi-scale reconstructive contrast[J]. arXiv preprint, arXiv: 2410.15997, 2024

[32] Yang Yiyuan, Zhang Chaoli, Zhou Tian, et al. DCdetector: Dual attention contrastive representation learning for time series anomaly detection[C]//Proc of the 29th ACM SIGKDD Conf on Knowledge Discovery and Data Mining. New York: ACM, 2023: 3033−3045

[33] Ge Lun, Wang Yijie, Zhang Ying, et al. Time-series anomaly detection method based on time-frequency enhanced contrastive learning[J]. Journal of Computer Research and Development, 2026, 63(3): 615−626 (in Chinese)
(葛伦, 王意洁, 张莹, 等. 基于时频增强的对比学习时序异常检测方法[J]. 计算机研究与发展, 2026, 63(3): 615−626)

[34] Fox A J. Outliers in time series[J]. Journal of the Royal Statistical Society: Series B (Methodological), 1972, 34(3): 350−363

[35] Kaiser R, Maravall A. Seasonal Outliers in Time Series[M]. Madrid: Banco de España, 1999: 6−14

[36] Schmidl S, Wenig P, Papenbrock T. Anomaly detection in time series: A comprehensive evaluation[J]. Proceedings of the VLDB Endowment, 2022, 15(9): 1779−1797

[37] Darban Z Z, Webb G I, Pan S, et al. CARLA: A self-supervised contrastive representation learning for time series anomaly detection[J]. Pattern Recognition, 2025, 157: 110874−110875

[38] Xu Hongzuo, Wang Yijie, Jian Songlei, et al. Calibrated one-class classification for unsupervised time series anomaly detection[J]. IEEE Transactions on Knowledge and Data Engineering, 2024, 36(11): 5723−5736

[39] Wang Qiang, Li Bei, Xiao Tong, et al. Learning deep Transformer models for machine translation[C]//Proc of the 57th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA: ACL, 2019: 1810−1822

[40] Liu Ze, Lin Yutong, Cao Yue, et al. Swin Transformer: Hierarchical vision Transformer using shifted windows[C]//Proc of the 18th IEEE/CVF Int Conf on Computer Vision. Piscataway, NJ: IEEE,

2021: 10012−10022

[41] Cleveland R B, Cleveland W S. STL: A seasonal-trend decomposition procedure based on loess[J]. Journal of Official Statistics, 1990, 6(1): 6−33

[42] Oord A, Li Yazhe, Vinyals O. Representation learning with contrastive predictive coding[J]. arXiv preprint, arXiv: 1807.03748, 2018

[43] Abdulaal A, Liu Zhuanghua, Lancewicki T. Practical approach to asynchronous multivariate time series anomaly detection and localization[C]//Proc of the 27th ACM SIGKDD Conf on Knowledge Discovery & Data Mining. New York: ACM, 2021: 2485−2494

[44] Mathur A P, Tippenhauer N O. SWaT: A water treatment test bed for research and training on ICS security[C]//Proc of the 2nd Int Workshop on Cyber-physical Systems for Smart Water Networks (CySWater). Piscataway, NJ: IEEE, 2016: 31−36

[45] Breunig M M, Kriegel H P, Ng R T, et al. LOF: Identifying density-based local outliers[C]//Proc of the 30th ACM SIGMOD Int Conf on Management of Data. New York: ACM, 2000: 93−104

[46] Yairi T, Takeishi N, Oda T, et al. A data-driven health monitoring method for satellite housekeeping data based on probabilistic clustering and dimensionality reduction[J]. IEEE Transactions on Aerospace and Electronic Systems, 2017, 53(3): 1384−1401

[47] Zong Bo, Song Qi, Min M R, et al. Deep autoencoding gaussian mixture model for unsupervised anomaly detection[C/OL]//Proc of the 6th Int Conf on Learning Representations. 2018[2025-03-01]. https://openreview.net/forum?id=BJJLHbb0-

[48] Shin Y, Lee S, Tariq S, et al. ITAD: Integrative tensor-based anomaly detection system for reducing false positives of satellite systems[C]//Proc of the 29th ACM Int Conf on Information & Knowledge Management. New York: ACM, 2020: 2733−2740

[49] Shen Lifeng, Li Zhuocong, Kwok J. Timeseries anomaly detection using temporal hierarchical one-class network[C]//Proc of the 34th Int Conf on Neural Information Processing Systems. Cambridge, UK: MIT, 2020: 13016−13026

[50] Ruff L, Vandermeulen R, Goernitz N, et al. Deep one-class classification[C]//Proc of the 6th Int Conf on Machine Learning. New York: ACM, 2018: 4393−4402

[51] Liu F T, Ting Kaiming, Zhou Zhihua. Isolation forest[C]//Proc of the 8th IEEE Int Conf on Data Mining. Piscataway, NJ: IEEE, 2008: 413−422

[52] Huet A, Navarro J M, Rossi D. Local evaluation of time series anomaly detection algorithms[C]//Proc of the 28th ACM SIGKDD Conf on Knowledge Discovery and Data Mining. New York: ACM, 2022: 635−645

[53] Chen Ting, Kornblith S, Norouzi M, et al. A simple framework for contrastive learning of visual representations[C]//Proc of the 37th Int Conf on Machine Learning. New York: ACM, 2020: 1597−1607

[54] Siffer A, Fouque P A, Termier A, et al. Anomaly detection in streams with extreme value theory[C]//Proc of the 23rd ACM SIGKDD Int Conf on Knowledge Discovery and Data Mining. New York: ACM, 2017: 1067−1075

[55] Van der Maaten L, Hinton G. Visualizing data using t-SNE[J]. Journal of Machine Learning Research, 2008, 9(86): 2579−2605

[56] Lai K H, Zha Daochen, Xu Junjie, et al. Revisiting time series outlier detection: Definitions and benchmarks[C]//Proc of the 35th Conf on Neural Information Processing Systems Datasets and Benchmarks Track (round 1). Cambridge, MA: MIT, 2021. https://datasets-benchmarks-proceedings.neurips.cc/paper_files/paper/2021/hash/ec5decca5ed3d6b8079e2e7e7bacc9f2-Abstract-round1.html



Wang Zhihong, born in 2003. Master candidate. His main research interests include time series analysis and data mining.

王志鸿, 2003 年生。硕士研究生。主要研究方向为时间序列分析、数据挖掘。



Huang Wei, born in 1994. PhD, associate professor. Her main research interests include federated learning and urban computing.

黄　维, 1994 年生。博士，副教授。主要研究方向为联邦学习、城市计算。



Wang Dexian, born in 1993. PhD, associate professor. His main research interests include machine learning and data mining.

王德贤, 1993 年生。博士，副教授。主要研究方向为机器学习、数据挖掘。



Li Tianrui, born in 1969. PhD, professor. His main research interests include data mining and knowledge discovery.

李天瑞, 1969 年生。博士，教授。主要研究方向为数据挖掘、知识发现。