

中心约束对比学习的自集成卫星异常检测方法

郭国航^{1,2}, 李虎¹, 刘玉荣¹, 胡钰^{1*}

(1. 中国科学院国家空间科学中心, 北京 100190; 2. 中国科学院大学, 北京 100049)

摘要:针对现有异常检测算法对遥测参数特征提取缺乏区分度以及异常决策信息丢失等问题, 提出基于中心约束对比的自集成异常检测方法。融合对比损失和中心损失将正常样本映射到紧凑的特征分布, 并采用多视角、多层次特征集成的方式实现样本的异常检测, 提升了模型对卫星复杂工况的适应性。采用科学卫星真实遥测参数数据和基准数据集进行验证, 结果表明, 所提方法在真实遥测参数上比最优基准方法的 F 值提升 21.8%, 且具有更好的噪声抗干扰性。实验结果验证了方法的可行性, 能够为卫星地面运管提供有效的判读支持。

关键词:科学卫星; 遥测参数; 对比学习; 异常检测

中图分类号: V557+.3 文献标志码: A 文章编号: 1001-2486(2024)06-033-10



论文
拓展

Self-ensemble satellite anomaly detection method for center-constrained contrastive learning

GUO Guohang^{1,2}, LI Hu¹, LIU Yurong¹, HU Tai^{1*}

(1. National Space Science Center, Chinese Academy of Sciences, Beijing 100190, China;

2. University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract: To deal with the problem of the existing telemetry anomaly detection algorithms, such as the poor discrimination capability of the feature, and loss of anomaly decision-making information, a self-ensemble anomaly detection method based on center-constrained contrastive learning was proposed. The method mapped the normal samples to a compact feature distribution by combining contrastive loss and center loss, and a multi-view and multi-level ensemble feature decision method was used to obtain the anomaly detection of the sample. The method improves the adaptability of the model to the complex working conditions of the satellite. The real telemetry parameter data of scientific satellite and benchmark data set are used for verification. The proposed method is robust to noise, and achieves 21.8% improvement of F score than that of the state of the art method. The results of the experiment demonstrate the feasibility of the method, which can provide effective support for satellite operation.

Keywords: scientific satellite; telemetry data; contrastive learning; anomaly detection

受不充分设计验证、生产加工工艺风险、极端空间环境、损伤累积效应以及星上状态动态变化等因素影响, 每颗卫星在整个生命周期内都或多或少会经历不同严重程度的异常事件^[1-2], 严重的异常还会导致任务降级或失败。卫星在轨运行过程中, 及时有效地检测和识别出异常, 并采取有效措施进行维护, 对保证卫星安全、可靠运行与效能发挥有重要意义。

近年来, 随着任务密度的提高以及待监测参

数量的提升, 相关学者提出了数据驱动异常检测方法^[3-6], 利用统计分析、数据挖掘和机器学习方法, 对遥测数据建模, 利用得到的模型发现异常, 具有良好的扩展性和适应性。目前本领域数据驱动的异常检测方法主要分为基于数据分布、基于偏差和基于分类。

基于数据分布的方法通常采用统计或机器学习等对正常数据的分布规律建模, 标记不符合该分布的数据样本为异常。传统的方法包括基于高

收稿日期: 2022-06-28

基金项目: 中国科学院战略性先导科技专项资助项目(XDA15040100)

第一作者: 郭国航(1995—), 男, 河南洛阳人, 博士研究生, E-mail: ggghssc@163.com

*通信作者: 胡钰(1972—), 男, 湖北石首人, 研究员, 博士, 博士生导师, E-mail: hutai@nssc.ac.cn

引用格式: 郭国航, 李虎, 刘玉荣, 等. 中心约束对比学习的自集成卫星异常检测方法[J]. 国防科技大学学报, 2024, 46(6): 33-42.

Citation: GUO G H, LI H, LIU Y R, et al. Self-ensemble satellite anomaly detection method for center-constrained contrastive learning[J]. Journal of National University of Defense Technology, 2024, 46(6): 33-42.

斯混合模型参数化方法^[7]和基于核密度估计的非参数化方法^[8]。这些方法直接在原始数据上建模,而对高维和复杂数据分布的检测效果较差。

基于偏差的方法的基本假设是,每个正常样本都能够使用一组基函数实现准确的重建,相对异常样本会表现出更大的重建误差。传统方法主要采用最近邻^[9]、主成分分析^[10]、 K 均值聚类^[11]以及树模型^[12-13]等方法,这些方法计算复杂度高,不适用于高维数据的异常检测。

基于分类的方法将异常检测看作单类分类问题,在正常数据周围构建决策边界,以排除其他未知样本。传统方法主要使用单类分类支持向量机^[14](one-class support vector machine, OC-SVM)和支持向量描述^[15](support vector data description, SVDD),利用核函数得到分类超平面或超曲面,将正常样本与其他样本分离。

随着深度学习的发展,越来越多学者提出了两阶段的异常检测方法,首先利用深度网络得到样本的特征表示,然后采用基于数据分布^[16-17]、基于偏差^[18-20]以及基于分类^[21]的传统方法构建异常分数以实现异常检测。近年来,基于对比学习的深度学习方法依靠其优秀的特征表示能力,提升了异常检测效果,并成为该领域的研究热点^[22-25]。直接将基于对比学习的异常检测算法用于卫星遥测参数异常检测面临着以下问题:

1) 卫星参数维度高,异常引发遥测参数的不规则变化,正常和异常样本的区分度不明显,且对比学习倾向于将训练样本均匀地映射到特征空间中的单位超球面,这导致映射到特征空间之后正常和异常样本交叠,两者无法有效区分。

2) 异常分数的构建只考虑代理任务中的单视角和单层特征,导致异常决策信息丢失,不能反映卫星的真实状态,且易受噪声数据干扰,异常检测效果较差。

针对上述问题,本文改进对比学习的损失函数和异常分数构建方法,提出一种新的卫星遥测参数异常检测算法。特征学习阶段,提出一种基于中心约束的对比学习(center-constrained contrastive learning, CCL)方法,将对比损失与中心损失结合,减小代理任务中同一视角下样本的方差,得到更加紧凑的特征分布,使特征空间中正常和异常样本更易于区分。异常决策阶段,提出一种自集成决策(self-ensemble decision, SED)方法,综合考虑代理任务中各个视角不同层次的特征,采用马氏距离构建异常分数。并利用 CCL-

SED 方法实现了卫星异常检测系统,在量子科学实验卫星健康管理系统中进行了初步应用。

1 问题定义

1.1 遥测参数数据和系统状态向量表示

为便于描述遥测参数异常检测问题,形式化表示遥测参数。

定义 1: 遥测参数向量 \mathbf{x} , 某分系统或某台设备下行的遥测参数。

$$\mathbf{x} = (d_1, \dots, d_j, \dots, d_m) \quad (1)$$

式中: d_j 为第 j 个参数的取值; m 为参数个数。

定义 2: 某时刻的遥测参数向量 $\mathbf{v}^{(t)}$, 该时刻采集的遥测参数。

$$\mathbf{v}^{(t)} = (t, \mathbf{x}^{(t)}) = (t, d_1^{(t)}, \dots, d_j^{(t)}, \dots, d_m^{(t)}) \quad (2)$$

式中: t 为数据采集时间; $\mathbf{x}^{(t)}$ 为对应时刻的 m 维遥测参数向量。

定义 3: 某分系统或某台设备的健康状态 q , $q \in \{0, 1\}$ 。 $q = 0$ 表示正常, $q = 1$ 表示异常。

定义 4: 状态向量 $\mathbf{u}^{(t)}$, 某分系统或某台设备在某一时刻的健康状态。

$$\mathbf{u}^{(t)} = (t, q^{(t)}) \quad (3)$$

式中: $q^{(t)}$ 为对应时刻设备的健康状态。

1.2 基于遥测参数数据的异常检测问题模型

受空间环境和星上状态动态变化等因素影响,不断有新类型异常出现,无法穷举所有类型的异常,且异常样本较少,而正常状态下的数据样本较容易取得。因此,问题归约为半监督异常检测: 首先基于正常状态下的遥测参数构建模型,然后用该模型检测异常。设 Ω_x 表示遥测参数数据的 m 维空间,则异常检测问题可描述为寻求一个目标函数 ψ , 使其满足:

$$\psi: \Omega_x \rightarrow q \quad (4)$$

给定训练样本集 $\mathbf{X}_{\text{train}} = \{(\mathbf{x}^{(i)}, q^{(i)})\}_{i=1}^N$, 对于每个样本 $(\mathbf{x}^{(i)}, q^{(i)})$, $q^{(i)} = 0$ 。首先根据训练样本集学得特征提取函数 ϕ , 设 Ω_h 为样本的特征空间,则该函数满足:

$$\phi: \Omega_x \rightarrow \Omega_h \quad (5)$$

然后根据样本特征构建异常分数函数 s :

$$s: \Omega_h \rightarrow \mathbf{R} \quad (6)$$

式中: \mathbf{R} 为实数空间。

最后根据异常分数构建判别函数 Δ :

$$\Delta: \mathbf{R} \rightarrow q \quad (7)$$

则给定训练样本集 $\mathbf{X}_{\text{train}}$ 学到的异常检测模型为:

$$\psi(\mathbf{x}^{(i)}) = \Delta \circ s \circ \phi(\mathbf{x}^{(i)}) = q^{(i)} \quad (8)$$

式中:“ \circ ”表示将前后两个函数组合得到复合函数。

2 对比学习和中心损失

本节介绍文中用到的背景知识:对比学习和中心损失。

2.1 对比学习

对比学习是一种自监督学习方法,利用数据自身样本对的相似或相异作为监督信息,以拉近正样本对并推远负样本对之间距离为目标,学习样本集的特征表达。文献[26]将对对比学习归纳为字典查询问题,令 \mathbf{x} 为一个查询, $\{\mathbf{x}_+\}$ 和 $\{\mathbf{x}_-\}$ 为 \mathbf{x} 的正样本和负样本集合,参考文献[24]的表示方法,对比损失函数定义为:

$$L_{\text{cls}}(\mathbf{x}, \{\mathbf{x}_+\}, \{\mathbf{x}_-\}) = -\frac{1}{|\{\mathbf{x}_+\}|} \lg \frac{\sum_{\mathbf{x}' \in \{\mathbf{x}_+\}} \exp(\rho(\mathbf{z}, \mathbf{z}')/\tau)}{\sum_{\mathbf{x}' \in \{\mathbf{x}_+\} \cup \{\mathbf{x}_-\}} \exp(\rho(\mathbf{z}, \mathbf{z}')/\tau)} \quad (9)$$

式中: $|\{\mathbf{x}_+\}|$ 为集合 $\{\mathbf{x}_+\}$ 的基数;超参数 $\tau > 0$ 为温度系数; \mathbf{z} 为样本 \mathbf{x} 的特征表示; \mathbf{z}' 为样本 \mathbf{x}' 的特征表示; $\rho(\mathbf{z}, \mathbf{z}') = \mathbf{z}^T \mathbf{z}' / (\|\mathbf{z}\| \|\mathbf{z}'\|)$ 为特征 \mathbf{z} 和 \mathbf{z}' 之间的余弦距离。

本文采用简单对比学习表示^[27] (simple contrastive learning of representations, SimCLR) 作为对比学习框架。SimCLR 主要包括数据增强器 T 、特征提取器 f_θ 、映射头 g_v 以及对比损失函数。 T 对输入数据 \mathbf{x}_i 做不同的随机变换 $T^1 \sim T$, $T^2 \sim T$, 生成 $\tilde{\mathbf{x}}_i^{(1)} = T^1(\mathbf{x}_i)$ 和 $\tilde{\mathbf{x}}_i^{(2)} = T^2(\mathbf{x}_i)$ 两个增强样本, $\tilde{\mathbf{x}}_i^{(1)}$ 和 $\tilde{\mathbf{x}}_i^{(2)}$ 分别称为 \mathbf{x}_i 的第 1 和第 2 视角, $(\tilde{\mathbf{x}}_i^{(1)}, \tilde{\mathbf{x}}_i^{(2)})$ 和 $(\tilde{\mathbf{x}}_i^{(2)}, \tilde{\mathbf{x}}_i^{(1)})$ 称为正样本对, $\tilde{\mathbf{x}}_i^{(1)}$ 和 $\tilde{\mathbf{x}}_i^{(2)}$ 与其他样本 $\mathbf{x}_j (j \neq i)$ 的增强样本组成负样本对。 $f_\theta(\cdot)$ 是一个基于神经网络的编码器,用于提取数据样本的特征, $\tilde{\mathbf{x}}_i^{(1)}$ 和 $\tilde{\mathbf{x}}_i^{(2)}$ 通过 $f_\theta(\cdot)$ 分别得到 $\mathbf{h}_i^{(1)} = f_\theta(\tilde{\mathbf{x}}_i^{(1)})$ 和 $\mathbf{h}_i^{(2)} = f_\theta(\tilde{\mathbf{x}}_i^{(2)})$ 。 $g_v(\cdot)$ 是一个多层感知机网络,映射 $\mathbf{h}_i^{(1)}$ 和 $\mathbf{h}_i^{(2)}$ 到新特征空间,得到 $\mathbf{z}_i^{(1)} = g_v(\mathbf{h}_i^{(1)})$ 和 $\mathbf{z}_i^{(2)} = g_v(\mathbf{h}_i^{(2)})$,在该空间中应用对比损失函数。对于每次迭代的小批量数据 $\mathbf{B} = \{\mathbf{x}_i\}_{i=1}^n$, SimCLR 的损失函数为:

$$L_{\text{SimCLR}}(\mathbf{B}; T) = \frac{1}{2n} \sum_{i=1}^n L_{\text{cls}}(\tilde{\mathbf{x}}_i^{(1)}, \tilde{\mathbf{x}}_i^{(2)}, \tilde{\mathbf{B}}_{-i}) + L_{\text{cls}}(\tilde{\mathbf{x}}_i^{(2)}, \tilde{\mathbf{x}}_i^{(1)}, \tilde{\mathbf{B}}_{-i}) \quad (10)$$

式中: $\tilde{\mathbf{B}}_{-i} = \{\tilde{\mathbf{x}}_j^{(1)}\}_{j \neq i} \cup \{\tilde{\mathbf{x}}_j^{(2)}\}_{j \neq i}$

2.2 中心损失

中心损失目标是降低同类样本在特征空间的方差,将样本特征尽可能拉向其所属类的中心,以得到更加紧凑的特征分布^[28-30]。中心损失函数定义为:

$$L_{\text{center}} = \frac{1}{2n} \sum_{i=1}^n \|\phi(\mathbf{x}_i) - \mathbf{c}_{y_i}\|^2 \quad (11)$$

式中: $\phi(\mathbf{x}_i)$ 为样本 \mathbf{x}_i 的特征; \mathbf{c}_{y_i} 为第 y_i 类样本特征的中心。 L_{center} 通过惩罚特征与其类中心之间的距离来减小类内方差。

式(11)中的 \mathbf{c}_{y_i} 为一个可学习的参数,每次迭代采用小批量的方式更新:

$$\mathbf{c}_j^{t+1} = \mathbf{c}_j^t - \alpha \frac{\sum_{i=1}^n \delta(y_i = j) \cdot (\mathbf{c}_j - \mathbf{x}_i)}{\epsilon + \sum_{i=1}^n \delta(y_i = j)} \quad (12)$$

式中: $\delta(s)$ 为指示函数,条件 s 满足时为 1,否则为 0;超参数 α 为学习率; ϵ 为一个小的正值,避免分母为 0。

3 中心约束对比学习与自集成异常判别

3.1 异常检测框架

基于问题定义,设计如图 1 所示卫星遥测参数异常检测框架。首先将卫星分包遥测得到的原始数据执行野值剔除、缺失值填补和归一化等预处理操作;然后将其划分为训练数据和测试数据,训练数据中只包含正常数据样本;训练数据用于训练 CCL 模型,得到编码器 f_θ ;测试数据通过 f_θ 映射到特征空间,并以自集成决策函数得到异常分数;最后根据异常分数和阈值 ϵ 判定样本是否异常。

3.2 中心约束对比学习

SimCLR 用于异常检测时,编码器和映射头采用 2.1 节所述的神经网络,数据增强器和对比损失函数如下:

数据增强器 T 采用文献[23]提出的随机仿射变换函数:

$$T(\mathbf{x}) = \boldsymbol{\omega} \mathbf{x} + \mathbf{b} \quad (13)$$

式中: $\boldsymbol{\omega}$ 和 \mathbf{b} 分别为采样自高斯分布的随机变换矩阵和随机向量。

对比损失函数受文献[24]的对比转移实例 (contrasting shifted instances, CSI) 启发,将同一样本的不同增强样本 $\tilde{\mathbf{x}}_i^{(1)}$ 和 $\tilde{\mathbf{x}}_i^{(2)}$ 作为负样本对,不同样本 \mathbf{x}_i 和 \mathbf{x}_j 的同一视角 $\tilde{\mathbf{x}}_i^{(k)}$ 和 $\tilde{\mathbf{x}}_j^{(k)}$ 作为正样本对:

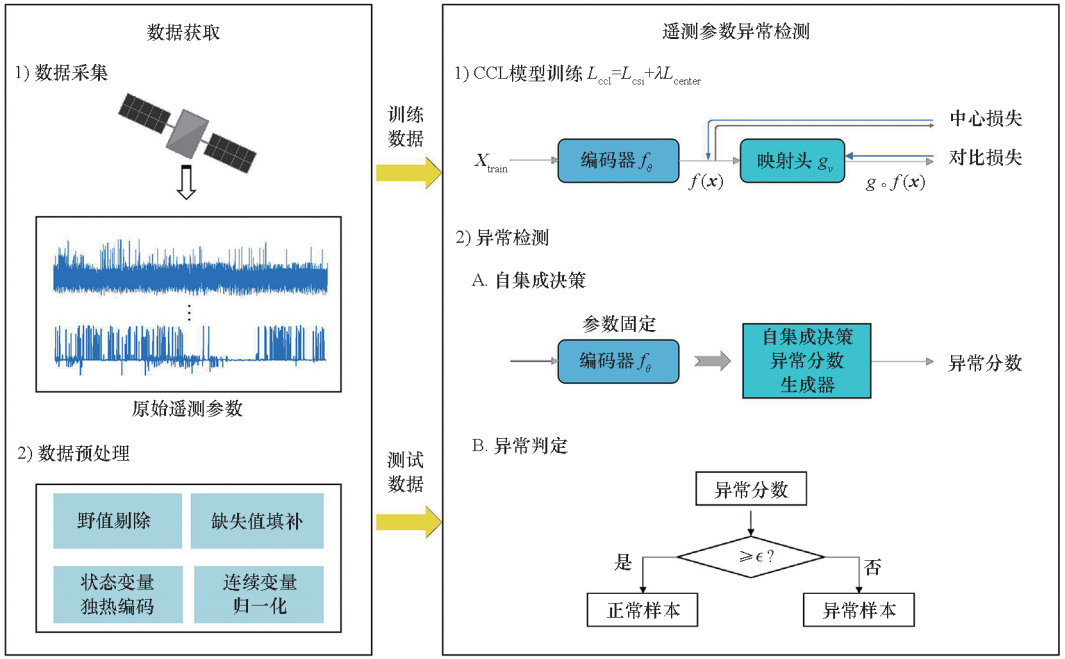


图 1 基于 CCL-SED 的卫星异常检测框架

Fig. 1 Framework for satellite anomaly detection based on CCL-SED

$$L_{csi} = \frac{1}{2n} \sum_{i=1}^n L_{cls}(\bar{\mathbf{x}}_i^{(1)}, \bar{\mathbf{B}}_{-i}^{(1)}, \bar{\mathbf{B}}^{(2)}) + L_{cls}(\bar{\mathbf{x}}_i^{(2)}, \bar{\mathbf{B}}_{-i}^{(2)}, \bar{\mathbf{B}}^{(1)}) \quad (14)$$

式中: $\bar{\mathbf{B}}^{(1)} = \{\bar{\mathbf{x}}_j^{(1)}\}_{j=1}^n$; $\bar{\mathbf{B}}^{(2)} = \{\bar{\mathbf{x}}_j^{(2)}\}_{j=1}^n$; $\bar{\mathbf{B}}_{-i}^{(1)} = \{\bar{\mathbf{B}}_j^{(1)}\}_{j \neq i}$; $\bar{\mathbf{B}}_{-i}^{(2)} = \{\bar{\mathbf{x}}_j^{(2)}\}_{j \neq i}$ 。

异常检测旨在判定样本是否异常,要求学得特征易于区分正常和异常样本。文献[31]证明并实验验证了采用对比损失函数优化网络参数过程中,训练样本逐渐被均匀映射到特征空间的单位超球面上。这意味着异常样本总可以在该超球面上某个邻域内找到一个正常样本,造成正常和异常样本的交叠,不利于正常和异常样本的区分。为了解决该问题,本文引入中心损失函数,通过减小训练样本在特征空间中的方差来增加特征分布的紧凑性,便于特征空间中正常和异常样本的区分。

中心约束对比损失函数定义为:

$$L_{ccl} = L_{csi} + \lambda L_{center} \quad (15)$$

式中:超参数 λ 用于平衡对比损失和中心损失,本文的实验中,根据经验设 $\lambda = 10$ 。

中心约束对比学习如图 2 所示。前向传播时,根据小批量训练数据 \mathbf{B} 中样本的 K 个视角计算对比损失,同时对编码器 f_θ 输出的特征计算中心损失。反向传播时,根据中心约束对比损失的梯度更新网络参数 θ 和 v 。

3.3 自集成决策

自集成决策是一个异常分数函数,根据 CCL 得到的特征,度量样本异常程度。常用的异常分数主要采用基于核密度估计^[17, 25]、基于单类支持向量机^[25]以及基于马氏距离^[32-34]的方法。特征学习阶段旨在将训练数据映射到一个紧凑的特征

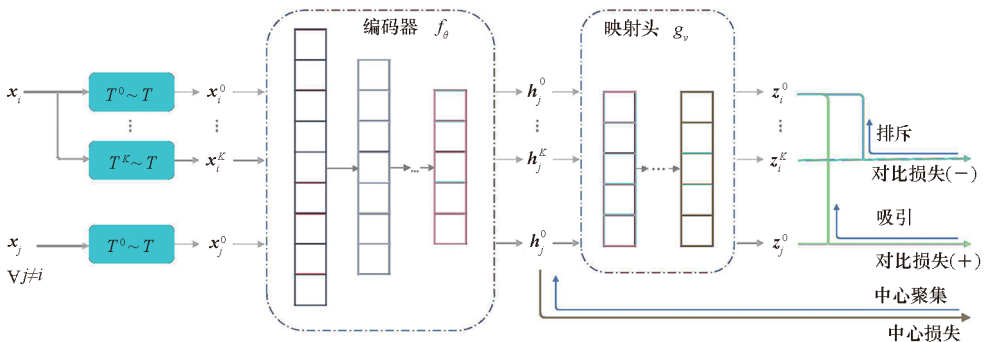


图 2 中心约束对比学习

Fig. 2 Center-constrained contrastive learning

分布,以使得异常样本到所学分布的距离大于正常样本到该分布的距离,从而对正常和异常样本进行区分,因此选用样本到训练数据分布的距离作为样本异常程度的度量。本文以马氏距离作为距离度量,提出了融合多视角和多层特征的自集成决策方法。SED 的主要思想是利用各个神经网络层以及各个视角的信息,提升模型辨别异常样本的能力。

文献[25, 35]表明,映射头过于专注代理任务,输出的特征丢掉了下游任务有用的信息。因此,基于编码器 f_θ 各层的特征构建异常分数函数。

以编码器 f_θ 第 l 层特征计算的增强样本 $\mathbf{x}^{(m)}$ 的异常分数 $s_m(\mathbf{x}, l)$ 为:

$$s_m(\mathbf{x}, l) = -[f_l(\mathbf{x}^{(m)}) - \boldsymbol{\mu}_l^{(m)}]^\top (\boldsymbol{\Sigma}_l^{(m)})^{-1} [f_l(\mathbf{x}^{(m)}) - \boldsymbol{\mu}_l^{(m)}] \quad (16)$$

式中: $\mathbf{x}^{(m)}$ 为样本 \mathbf{x} 的第 m 视角; $f_l(\mathbf{x}^{(m)})$ 为 $\mathbf{x}^{(m)}$ 第 l 层的特征; $\boldsymbol{\mu}_l^{(m)}$ 和 $\boldsymbol{\Sigma}_l^{(m)}$ 分别为 $\{f_l(\mathbf{x}_i^{(m)})\}_{i=1}^N$ 的均值和协方差矩阵, N 为训练样本个数。

自集成决策如图3所示,包括多视角特征集成(multi-view feature ensembled, MVFE)和多层特征集成(multi-layer feature ensembled, MLFE)。

样本 \mathbf{x} 第 l 层的多视角特征集成异常分数为:

$$s_{\text{mvfe}}(\mathbf{x}, l) = F_{\text{agg}}(\{s_i(\mathbf{x}, l)\}_{i=1}^K) \quad (17)$$

式中: F_{agg} 为聚合函数,可选 mean、max 和 min 分别计算集合中元素的均值、最大值和最小值; $\{s_i(\mathbf{x}, l)\}_{i=1}^K$ 为样本 \mathbf{x} 所有视角第 l 层的异常分数集合; K 为视角个数。

样本 \mathbf{x} 的自集成异常分数为:

$$s_{\text{SED}}(\mathbf{x}) = \text{mean}(\{s_{\text{mvfe}}(\mathbf{x}, l_i)\}_{i=1}^W) \quad (18)$$

式中:mean为均值聚合函数; W 为编码器层数。

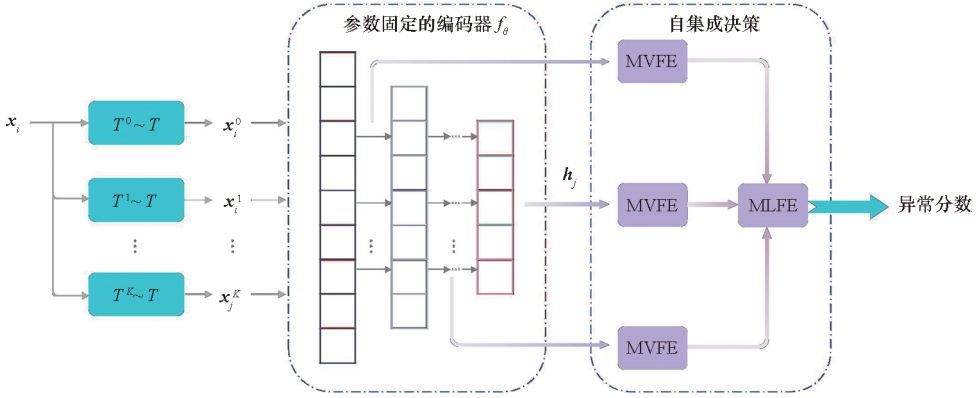


图3 自集成决策

Fig. 3 Self-ensemble decision

4 实验验证与结果分析

4.1 数据来源

本实验采用预处理后的量子科学实验卫星(Micius)^[36]2017年1月7日至2018年2月28日的遥测参数数据,选取了与卫星姿控系统工作状态相关的23个参数。此外,为了验证算法的泛化性,还选取了异常检测领域的公开基准数据集Thyroid和Arrhythmia,两者均来源于传感器数据。数据集相关信息如表1所示。

表1 数据集统计信息

Tab. 1 Statistics of datasets

数据集	维度	样本数	异常样本比例
Micius	23	205 522	0.034
Thyroid	6	3 772	0.025
Arrhythmia	274	452	0.150

4.2 基准算法介绍

选取以下基准算法:经典方法中选择基于分布的核密度估计(kernel density estimation, KDE)算法、基于分类的OC-SVM算法以及基于偏差的孤立森林(isolation forest, iForest)方法,深度方法中选择基于分布的深度自编码高斯混合模型(deep autoencoding Gaussian mixture model, DAGMM)算法、基于分类的Deep SVDD算法以及采用对比学习的基于偏差的通用数据异常检测(generalized outlier anomaly detection, GOAD)算法。基准算法简介如下:

KDE^[8]:一种基于核密度估计的异常检测方法,利用正常数据建立概率分布模型,以该概率模型判定异常。

OC-SVM^[14]:一种基于核函数的异常检测方法,利用核函数将数据映射到高维空间,构建决策边界,以此为判定异常依据。

iForest^[12]:一种基于树模型的异常检测算法,构建集成树模型,基于每个样本到达叶子节点的平均路径长度判定异常。

DAGMM^[16]:一种基于深度自编码器的异常检测方法,包括基于自编码器的压缩网络和基于混合高斯的估计网络。压缩网络利用深度自编码器提取样本特征,估计网络根据特征建立高斯混合模型,以该概率模型判断异常。

Deep SVDD^[21]:一种基于深度神经网络的异常检测方法,利用正常数据学得超球面,以超球面为决策边界判定异常。

GOAD^[23]:一种基于自监督的异常检测方法,将增强样本分类作为代理任务,并引入三元组损失作为对比损失,根据分类器的 Softmax 值检测异常。

4.3 实验性能指标

异常检测可看作对样本的二分类,分类的目标只有两类,即正例(positive)为异常数据,负例(negative)为正常数据^[37]。在对结果分类后,用 F 值和 A 值度量算法性能:

$$F = \frac{2PR}{P+R} \quad (19)$$

$$A = \frac{\sum_{i \in \text{positiveClass}} r_i - \frac{M(1+M)}{2}}{MN} \quad (20)$$

其中: $P = J/(J+U)$, $R = J/(J+V)$; J 为真正例,表示实际为异常、算法检测为异常的样本数; U 为假正例,表示实际为正常、算法检测为异常的样本数; V 为假负例,表示实际为异常、算法检测为正常的样本数;异常分数按升序排序, r_i 表示第 i 条正样本的序号; M 和 N 分别代表正样本和负样本个数。

4.4 实验设置

本实验采用文献[16, 38]中的实验设置方式:将 20 次(run)实验的评价指标的均值和标准差作为异常检测性能度量,单次实验运行 50 轮(epoch);每次实验,从正常数据样本中随机抽取 50% 作为训练集,剩余的 50% 样本以及所有异常样本作为测试集;根据数据集中异常样本比例确定判别阈值 ϵ ,即异常分数升序排序,前 ϵ 的样本判定为异常;自集成决策时,数据集 Micius、Thyroid 和 Arrhythmia 分别选用 mean、min 和 max 作为聚合函数;各个基准算法采用网格搜索法得到最优超参数。

4.5 算法有效性

表 2 展示了 CCL-SED 及其对比算法在 Micius、Thyroid 和 Arrhythmia 数据集上的表现,其中最优化结果加粗标注, S 表示标准差。由表 2 可知,CCL-SED 算法在三个数据集上的 F 值和 A 值大多数情况下优于其他 6 种基准算法,而且较低的标准差表明该算法表现更为稳定。CCL-SED 在 Micius 数据集上 F 值比最优秀的基准算法 DAGMM 提升了 21.8%, A 值比最优秀的基准算法 iForest 提升了 12.5%。CCL-SED 在 Thyroid 数据集上的效果也优于基准算法。在 Arrhythmia 数据集上,CCL-SED 与 OC-SVM 较为接近,前者 F 值优势明显,后者 A 值略优,这是由于 Arrhythmia 数据集的样本数较少,训练样本数 193 小于数据维度 274,深度网络对训练集存在过拟合,对训练集数据分布的估计不足,而 OC-SVM 由于引入了核函数,能够在小数据量上取得较好效果。训练样本个数在大多数情况下大于数据维度,因此 CCL-SED 在另外两个数据集上表现良好,若要提高 CCL-SED 方法在 Arrhythmia 数据集上的 A 值,则需要采用数据增强方法增加训练数据样本。

表 2 CCL-SED 和基准算法的平均 F 和 A

Tab. 2 Average F and A of CCL-SED and benchmark algorithms

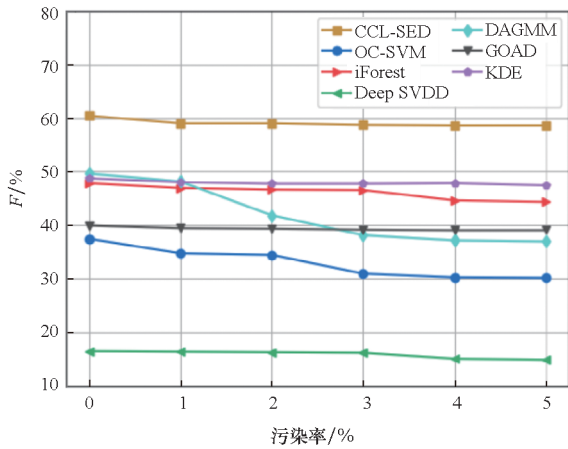
%

算法	数据集					
	Micius		Thyroid		Arrhythmia	
	$F \pm S$	$A \pm S$	$F \pm S$	$A \pm S$	$F \pm S$	$A \pm S$
OC-SVM	37.4 ± 9.4	49.9 ± 9.2	40.0 ± 6.9	93.4 ± 1.1	51.9 ± 1.7	80.6 ± 0.7
iForest	47.8 ± 4.5	68.0 ± 1.9	49.8 ± 2.0	98.6 ± 0.4	29.8 ± 5.3	81.6 ± 1.0
KDE	48.7 ± 3.6	65.6 ± 2.5	73.6 ± 2.7	98.1 ± 0.4	56.5 ± 3.4	79.0 ± 2.3
Deep SVDD	16.4 ± 9.9	48.9 ± 4.4	22.8 ± 7.6	89.0 ± 7.2	44.2 ± 4.7	65.2 ± 4.8
DAGMM	49.6 ± 7.1	33.9 ± 4.3	47.4 ± 5.0	94.7 ± 3.0	35.1 ± 5.3	69.1 ± 3.6
GOAD	39.9 ± 5.5	54.0 ± 6.5	75.8 ± 4.2	95.4 ± 4.0	51.9 ± 3.8	71.1 ± 4.7
CCL-SED	60.4 ± 3.5	76.5 ± 3.0	80.0 ± 1.9	98.7 ± 0.3	58.1 ± 3.4	78.7 ± 2.0

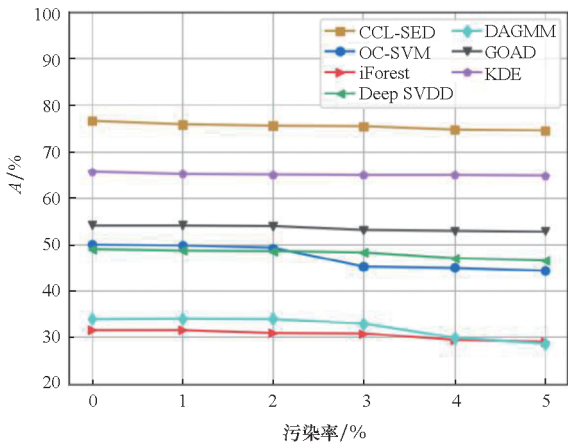
4.6 算法抗干扰性

实际应用中,训练集往往混杂有噪声或异常数据,为了验证算法的抗干扰性,采用不同污染程度的 Micius 数据集对算法进行验证。每次实验从正常数据中随机抽取 50% 的样本,并从异常数据中随机抽取一定比例的异常样本,将两者混合组成训练集。

图 4 描述了 CCL-SED 及其基准算法在 Micius 数据集上 20 轮实验的平均 F 和 A 随污染率的变化情况。



(a) F



(b) A

图 4 Micius 数据集上污染训练集的异常检测结果

Fig. 4 Anomaly detection results of contaminated training data on Micius data set

从图 4 中可以看出,数据污染为各个算法都带来了负面的影响,随着污染率的提升,各个算法的平均 F 值和 A 值都呈下降趋势,但不同算法对污染数据的敏感程度不同。由图 4(a) 可知, DAGMM 对数据污染较为敏感,随着污染率的提升, F 值有较大下降,且算法效果逐渐落后于 iForest 和 GOAD。但由图 4 (b) 可知, DAGMM 在

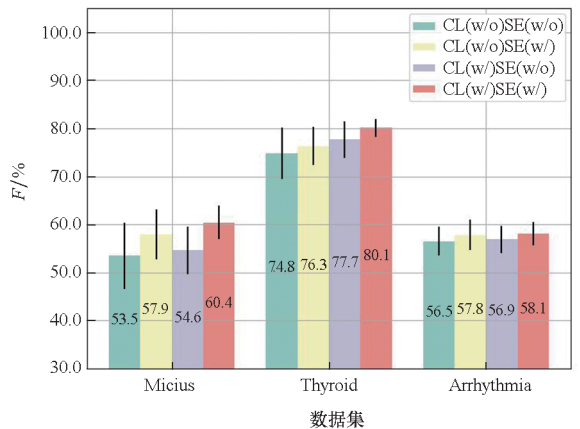
A 值上的下降程度远低于其在 F 值上的下降程度。 A 值下降率较小说明了 DAGMM 异常分数对正常、异常样本的区分度受污染数据影响不大,而 F 值下降率较大则说明了 DAGMM 异常阈值设置不合理,鲁棒性较差。CCL-SED 算法在污染率从 0 增加至 5% 的过程中, F 值和 A 值的下降率较小,且和基准算法相比,均能保持最优秀的表现,表明了 CCL-SED 算法具有良好的噪声抗干扰性。

4.7 消融实验

为了验证 CCL-SED 算法中引入的中心损失和集成策略的有效性,本文进行了消融实验。

图 5 描述了中心损失和自集成策略对 CCL-SED 算法的影响,展示了中心损失和自集成的不同组合方式在数据集 Micius、Thyroid 和 Arrhythmia 上 20 次实验的 F 值和 A 值的均值和标准差。图 5 中:CL 表示中心损失;SE 表示自集成;w/o (without) 表示不采用;w/ (with) 表示采用;CL(w/o)SE(w/o) 表示无中心损失且不采用自集成,依此类推。从图 5 中可以看出,无中心损失且不采用自集成时,算法的性能最差;同时采用中心损失和自集成策略时,算法表现最优;中心损失和自集成策略带来的提升程度不同。

由图 5(a) 中 CCL-SED 算法在 Micius 数据集上的 F 值可知,自集成策略比中心损失更能提升算法的平均 F 值,这是因为自集成策略能够利用更广泛的视角信息和不同深浅程度的特征信息。从图 5 (b) 中 CCL-SED 算法在 Micius 数据集上的 A 值可以看出,中心约束的引入除了能提升算法的 A 值,也能更大程度减小方差,提升算法的稳定性,这是因为施加中心约束能将样本映射到特征空间中更紧凑的特征分布,避免了正常和异常样本在特征空间中的交叠,提升了正常异常样本的辨识度,减小了样本误判的



(a) F

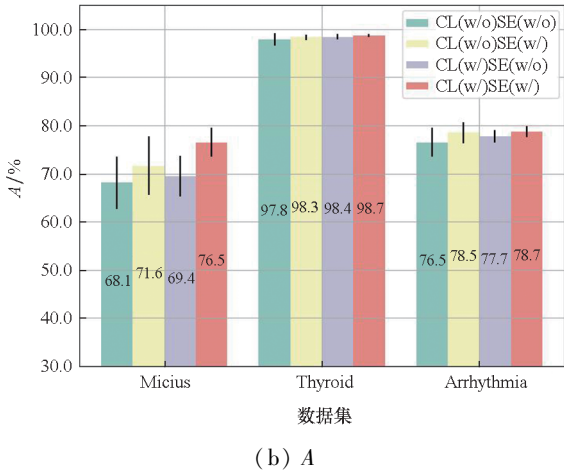


图 5 中心损失和自集成对 CCL-SED 的影响
Fig. 5 Influence of center loss and self-ensemble on CCL-SED

可能。从图 5 可以看出,同时采用自集成和中心损失,能够更进一步提升算法效果并减少误差。

4.8 中心损失对特征分布的影响

为了探究中心损失的引入对特征分布的影响,采用最大均值依赖 (maximum mean discrepancy, MMD) 对样本特征 f_θ 分布的紧凑性进行量化,并对样本特征进行可视化,展示中心损失对特征空间中正常和异常样本的可分性以及特征分布紧凑性的影响。

根据文献[25, 39]中的分析方法,以特征分布和单位球面上采样得到的均匀分布之间的 MMD 作为特征分布紧凑性的度量。MMD 是一个非负值,其越大意味着特征分布越紧凑,反之特征分布越趋近于均匀分布。表 3 展示了三个数据集在采用和不采用中心损失时的 MMD。从表中数据可知,采用中心损失能显著提高所得特征的 MMD,即采用中心损失提高了特征分布的紧凑性。

表 3 特征分布与均匀分布之间的 MMD

Tab. 3 MMD between the feature distribution and the uniform distribution

数据集	MMD	
	CL(w/o)	CL(w/)
Micius	2.616 5	3.118 6
Thyroid	0.557 2	0.997 6
Arrhythmia	0.320 5	0.390 6

中心损失对特征空间的影响如图 6 所示,采

用 T 分布随机近邻嵌入^[40] (T-stochastic neighbor embedding, TSNE) 将特征降至三维并归一化显示。图 6(a)为原始数据经 TSNE 降维后的结果,正常和异常数据分布在整个球面空间,正常和异常数据样本堆叠在一起难以区分;图 6(b)为不采用中心损失的 CCL-SED 得到的特征经 TSNE 降至三维的结果,异常样本周围的正常样本数量相比图 6(a)有所下降,正常样本的紧凑性得到提高,能够在一定程度上防止异常样本映射到特征空间后与正常样本交叠;图 6(c)为采用中心损失的 CCL-SED 得到的特征表示,采用中心损失提升了样本特征的紧凑性,降低了特征空间中正常和异常样本交叠的可能性,增强了正常和异常样本的可分性。

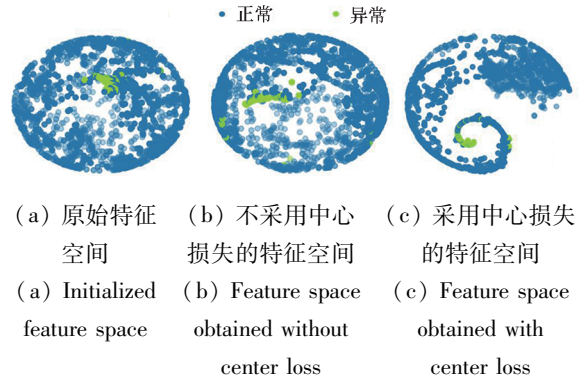


图 6 中心损失对特征空间的影响
Fig. 6 Influence of center loss on feature space

5 结论

本文基于对比学习提出了一种基于中心约束对比的自集成异常检测方法,并用于卫星遥测参数的异常检测。CCL-SED 通过改进 SimCLR 的损失函数,将样本映射到特征空间中更加紧凑的特征分布,然后基于不同视角和不同层次的特征,采用自集成异常评分策略量化样本的异常程度,以实现异常检测。对量子科学实验卫星的真实遥测参数以及文献中广泛使用的异常检测基准数据集进行了对比实验,CCL-SED 基本取得了优于当前主流异常检测算法的效果,并且在受污染数据上也能表现良好,具有良好的噪声抗干扰性,这为卫星地面运管提供了更有效的判读支持。

未来进一步研究的问题:①分析正常和异常样本在特征空间中的分布特点,以得到一种面向异常检测的特征学习方法;②有效利用地面运管人员在日常监控工作中识别出的异常样本,提升异常检测算法的检测精度。

参考文献 (References)

- [1] 庞景月. 基于概率性预测的航天器遥测数据异常检测方法研究[D]. 哈尔滨: 哈尔滨工业大学, 2019.
PANG J Y. Probabilistic prediction based anomaly detection method for spacecraft telemetry data [D]. Harbin: Harbin Institute of Technology, 2019. (in Chinese)
- [2] 杨天社, 金光, 樊恒海, 等. 卫星在轨状态检测与健康管理技术[M]. 北京: 国防工业出版社, 2019.
YANG T S, JIN G, FAN H H, et al. On-orbit satellite state detection and health management technology [M]. Beijing: National Defense Industry Press, 2019. (in Chinese)
- [3] 孙宇豪, 李国通, 张鸽. 距离相关系数融合 GPR 模型的卫星异常检测方法[J]. 北京航空航天大学学报, 2021, 47(4): 844 - 852.
SUN Y H, LI G T, ZHANG G. A satellite anomaly detection method based on distance correlation coefficient and GPR model[J]. Journal of Beijing University of Aeronautics and Astronautics, 2021, 47(4): 844 - 852. (in Chinese)
- [4] 李虎, 郭国航, 胡钰, 等. 遥测参数数据载荷状态判别集成学习方法[J]. 国防科技大学学报, 2021, 43(6): 33 - 40.
LI H, GUO G H, HU T, et al. Ensemble learning for state recognition of payload from telemetry data [J]. Journal of National University of Defense Technology, 2021, 43(6): 33 - 40. (in Chinese)
- [5] 董静怡, 庞景月, 彭宇, 等. 集成 LSTM 的航天器遥测数据异常检测方法[J]. 仪器仪表学报, 2019, 40(7): 22 - 29.
DONG J Y, PANG J Y, PENG Y, et al. Spacecraft telemetry data anomaly detection method based on ensemble LSTM[J]. Chinese Journal of Scientific Instrument, 2019, 40(7): 22 - 29. (in Chinese)
- [6] 杨甲森, 孟新, 陈托, 等. 基于遥测数据相关性的航天器异常检测[J]. 仪器仪表学报, 2018, 39(8): 24 - 33.
YANG J S, MENG X, CHEN T, et al. Anomaly detection of spacecraft based on the telemetry data correlation [J]. Chinese Journal of Scientific Instrument, 2018, 39(8): 24 - 33. (in Chinese)
- [7] LI L S, HANSMAN R J, PALACIOS R, et al. Anomaly detection via a Gaussian mixture model for flight operation and safety monitoring [J]. Transportation Research Part C: Emerging Technologies, 2016, 64: 45 - 57.
- [8] AHMED T. Online anomaly detection using KDE [C]// Proceedings of the 2009 IEEE Global Telecommunications Conference, 2009.
- [9] XIE M, HU J K, HAN S, et al. Scalable hypergrid k-NN-based online anomaly detection in wireless sensor networks[J]. IEEE Transactions on Parallel and Distributed Systems, 2013, 24(8): 1661 - 1670.
- [10] XU H, CARAMANIS C, SANGHAVI S. Robust PCA via outlier pursuit[J]. IEEE Transactions on Information Theory, 2012, 58(5): 3047 - 3064.
- [11] 王婵, 王慧泉, 金仲和. 皮纳卫星遥测数据异常检测聚类分析方法[J]. 哈尔滨工业大学学报, 2018, 50(4): 110 - 116.
WANG C, WANG H Q, JIN Z H. Pico-satellite telemetry anomaly detection through clustering[J]. Journal of Harbin Institute of Technology, 2018, 50(4): 110 - 116. (in Chinese)
- [12] LIU F T, TING K M, ZHOU Z H. Isolation forest [C]// Proceedings of the 8th IEEE International Conference on Data Mining, 2008.
- [13] 张万旋, 薛薇, 张楠. 基于增量式孤立森林的液体火箭发动机异常检测方法[J]. 航空动力学报, 2022, 37(8): 1674 - 1682.
ZHANG W X, XUE W, ZHANG N. Anomaly detection method of liquid rocket engine based on incremental isolation forest [J]. Journal of Aerospace Power, 2022, 37(8): 1674 - 1682. (in Chinese)
- [14] SCHÖLKOPF B, WILLIAMSON R, SMOLA A, et al. Support vector method for novelty detection [C]// Proceedings of the 12th International Conference on Neural Information Processing Systems, 1999.
- [15] 杨健健, 唐至威, 王晓林, 等. 单类学习下基于 VSAPSO-BP 的掘进机异常检测方法[J]. 振动、测试与诊断, 2019, 39(1): 130 - 135, 226.
YANG J J, TANG Z W, WANG X L, et al. Roadheader anomaly detection method based on VSAPSO-BP under the single category learning [J]. Journal of Vibration, Measurement & Diagnosis, 2019, 39(1): 130 - 135, 226. (in Chinese)
- [16] ZONG B, SONG Q, MIN M R, et al. Deep autoencoding Gaussian mixture model for unsupervised anomaly detection [C]// Proceedings of the International Conference on Learning Representations, 2018.
- [17] LYU P, YU Y W, FAN Y Y, et al. Layer-constrained variational autoencoding kernel density estimation model for anomaly detection [J]. Knowledge-Based Systems, 2020, 196: 105753.
- [18] AN J, CHO S. Variational autoencoder based anomaly detection using reconstruction probability [R/OL]. (2015 - 12 - 27) [2022 - 05 - 03]. <http://dm.snu.ac.kr/static/docs/TR/SNUDM-TR-2015-03.pdf>.
- [19] ZENATI H, FOO C S, LECOQUAT B, et al. Efficient GAN-based anomaly detection [EB/OL]. (2019 - 05 - 01) [2022 - 05 - 03]. <https://arxiv.org/abs/1802.06222v2>.
- [20] LI D, CHEN D C, JIN B H, et al. MAD-GAN: multivariate anomaly detection for time series data with generative adversarial networks [M]// Lecture Notes in Computer Science. Cham: Springer International Publishing, 2019: 703 - 716.
- [21] RUFF L, VANDERMEULEN R, GÖRNITZ N, et al. Deep one-class classification [C]// Proceedings of Machine Learning Research, 2018.
- [22] GOLAN I, EL-YANIV R. Deep anomaly detection using geometric transformations [EB/OL]. (2018 - 11 - 09) [2022 - 05 - 04]. <https://arxiv.org/abs/1805.10917>.
- [23] BERGMAN L, HOSHEN Y. Classification-based anomaly detection for general data [EB/OL]. (2020 - 05 - 05) [2022 - 05 - 04]. <https://arxiv.org/abs/2005.02359v1>.
- [24] TACK J, MO S, JEONG J, et al. CSI: novelty detection via contrastive learning on distributionally shifted instances [EB/OL]. (2020 - 10 - 21) [2022 - 05 - 04]. <https://arxiv.org/abs/2007.08176v2>.
- [25] SOHN K, LI C L, YOON J, et al. Learning and evaluating representations for deep one-class classification [EB/OL]. (2021 - 03 - 25) [2022 - 05 - 04]. <https://arxiv.org/abs/2011.02578v2>.
- [26] HE K M, FAN H Q, WU Y X, et al. Momentum contrast for

- unsupervised visual representation learning [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020.
- [27] CHEN T, KORNBLITH S, NOROUZI M, et al. A simple framework for contrastive learning of visual representations [EB/OL]. (2020-07-01) [2022-05-04]. <https://arxiv.org/abs/2002.05709>.
- [28] REISS T, HOSHEN Y. Mean-shifted contrastive loss for anomaly detection [EB/OL]. (2022-11-28) [2022-05-04]. <https://arxiv.org/abs/2106.03844v2>.
- [29] WEN Y D, ZHANG K P, LI Z F, et al. A comprehensive study on center loss for deep face recognition [J]. International Journal of Computer Vision, 2019, 127: 668-683.
- [30] WEN Y D, ZHANG K P, LI Z F, et al. A discriminative feature learning approach for deep face recognition [M]//Lecture Notes in Computer Science. Cham: Springer International Publishing, 2016: 499-515.
- [31] WANG T Z, ISOLA P. Understanding contrastive representation learning through alignment and uniformity on the hypersphere [EB/OL]. (2022-08-15) [2022-05-04]. <https://arxiv.org/abs/2005.10242v10>.
- [32] WINKENS J, BUNEL R, ROY A G, et al. Contrastive training for improved out-of-distribution detection [EB/OL]. (2020-07-10) [2022-05-04]. <https://arxiv.org/abs/2007.05566v1>.
- [33] SEHWAG V, CHIANG M, MITTAL P. SSD: a unified framework for self-supervised outlier detection [EB/OL]. (2021-03-22) [2022-05-07]. <https://arxiv.org/abs/2103.12051v1>.
- [34] CHO H, SEOL J, LEE S G. Masked contrastive learning for anomaly detection [EB/OL]. (2021-03-22) [2022-05-07]. <https://arxiv.org/abs/2105.08793v2>.
- [35] GIDARIS S, SINGH P, KOMODAKIS N. Unsupervised representation learning by predicting image rotations [EB/OL]. (2018-03-21) [2022-05-07]. <https://arxiv.org/abs/1803.07728v1>.
- [36] LIAO S K, CAI W Q, LIU W Y, et al. Satellite-to-ground quantum key distribution [J]. Nature, 2017, 549: 43-47.
- [37] 丁小欧, 于晟健, 王沐贤, 等. 基于相关性分析的工业时序数据异常检测 [J]. 软件学报, 2020, 31(3): 726-747.
- DING X O, YU S J, WANG M X, et al. Anomaly detection on industrial time series based on correlation analysis [J]. Journal of Software, 2020, 31(3): 726-747. (in Chinese)
- [38] ZHAI S F, CHENG Y, LU W N, et al. Deep structured energy based models for anomaly detection [EB/OL]. (2016-07-16) [2022-05-07]. <https://arxiv.org/abs/1605.07717>.
- [39] YOUSEF M, ACKERMANN M, KURUP U, et al. No shifted augmentations (NSA): compact distributions for robust self-supervised anomaly detection [EB/OL]. (2022-03-19) [2022-05-07]. <https://arxiv.org/abs/2203.10344v1>.
- [40] VAN DER MAATEN L, HINTON G. Visualizing data using t-SNE [J]. Journal of Machine Learning Research, 2008, 9: 2579-2605.