软件学报 ISSN 1000-9825, CODEN RUXUEW
Journal of Software [doi: 10.13328/j.cnki.jos.007044]

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel: +86-10-62562563

IRCGN: 用于高效多视图离群点检测的生成式网络*

郑 啸[1,2,3], 王权鑫[1,3], 黄 俊[1,2,3]

[1](安徽工业大学 计算机科学与技术学院, 安徽 马鞍山 243032)

[2](合肥综合性国家科学中心 人工智能研究院, 安徽 合肥 230071)

[3](安徽省工业互联网智能应用与安全工程研究中心, 安徽 马鞍山 243032)

通信作者: 郑啸, E-mail: xzheng@ahut.edu.cn

摘 要: 由于多视图数据特征复杂, 多视图离群检测已经成为离群点检测中一个极具挑战性的研究课题. 多视图数据中存在 3 种类型的离群点, 分别为类离群点、属性离群点和类-属性离群点. 早期多视图离群点检测方法大多基于聚类假设, 当数据中没有聚类结构时很难检测出离群点. 近年来, 许多多视图离群点检测方法使用多视图一致的近邻假设来代替聚类假设, 但仍存在新增数据检测效率低的问题. 此外, 大多数现有的多视图离群点检测方法都是无监督的, 在模型学习过程中会受到离群点的影响, 处理高离群率的数据集时效果不佳. 为了解决这些问题, 提出一种用于高效多视图离群点检测的视图内重建和跨视图生成网络来检测 3 种类型的离群点, 所提方法包含视图内重建和跨视图生成两个模块. 通过使用正常数据训练, 所提出方法可以充分捕捉正常数据中每个视图的特征, 并较好地重建和生成相应的视图. 此外, 还提出一个新的离群值计算方法, 为每一个样本计算相应的离群值得分, 从而高效地检测新增数据. 大量的实验结果表明, 所提出的方法明显优于现有的方法. 这是第 1 项将基于生成对抗网络的深度模型应用于多视图离群点检测的工作.

关键词: 离群点检测; 多视图数据; 半监督; 视图内重建; 跨视图生成

中图法分类号: TP18

中文引用格式: 郑啸, 王权鑫, 黄俊. IRCGN: 用于高效多视图离群点检测的生成式网络. 软件学报. http://www.jos.org.cn/1000-9825/7044.htm

英文引用格式: Zheng X, Wang QX, Huang J. IRCGN: Generation Network for Effective Multi-view Outlier Detection. Ruan Jian Xue Bao/Journal of Software (in Chinese). http://www.jos.org.cn/1000-9825/7044.htm

IRCGN: Generation Network for Effective Multi-view Outlier Detection

ZHENG Xiao[1,2,3], WANG Quan-Xin[1,3], HUANG Jun[1,2,3]

[1](School of Computer Science and Technology, Anhui University of Technology, Ma'anshan 243032, China)

[2](Institute of Artificial Intelligence, Hefei Comprehensive National Science Center, Hefei 230071, China)

[3](Anhui Engineering Research Center for Intelligent Applications and Security of Industrial Internet, Ma'anshan 243032, China)

Abstract: Due to the complex features of multi-view data, multi-view outlier detection has become a very challenging research topic in outlier detection. There are three types of outliers in multi-view data, namely class outliers, attribute outliers, and class-attribute outliers. Most of the early multi-view outlier detection methods are based on the assumption of clustering, which makes it difficult to detect outliers when there is no clustering structure in the data. In recent years, many multi-view outlier detection methods use the multi-view consistent nearest neighbor assumption instead of the clustering assumption, but they still suffer from the problem of inefficient detection of new data. In addition, most existing multi-view outlier detection methods are unsupervised, which are affected by outliers during model learning and do not work well when dealing with datasets with high outlier rates. To address these issues, this study proposes an intra-

* 基金项目: 国家自然科学基金 (61806005); 安徽省高校协同创新项目 (GXXT-2019-025, GXXT-2020-012, GXXT-2022-052); CCF-蚂蚁科研基金 (CCF-AFSGRF20210003)
收稿时间: 2023-02-01; 修改时间: 2023-04-20; 采用时间: 2023-08-26; jos 在线出版时间: 2024-01-03

view reconstruction and cross-view generation network for effective multi-view outlier detection to detect the three types of outliers, which consists of two modules: intra-view reconstruction and cross-view generation. By training with normal data, the proposed method can fully capture the features of each view in the normal data and reconstruct and generate the corresponding views better. In addition, a new outlier calculation method is proposed to calculate the corresponding outlier scores for each sample to efficiently detect new data. Extensive experimental results show that the proposed method significantly outperforms existing methods. It is known that this is the first work to apply a deep model based on generative adversarial networks to multi-view outlier detection.

Key words: outlier detection; multi-view data; semi-supervised; intra-view reconstruction; cross-view generation

离群点检测, 也被称为异常检测, 是一种重要的数据分析技术. 其中, 半监督离群点检测的目的是在使用正常标签数据训练的情况下识别给定数据集中的离群点[1]. 离群点检测被广泛应用于各种领域, 如医疗诊断[2]、工业缺陷检测[3]、网络攻击检测[4]等. 然而, 这些方法主要针对单视图数据设计, 无法处理多视图数据的丰富信息. 在许多实际应用场景中, 数据往往来自不同的特征提取器, 相应的每组特征集可被视为对象的一个特定视图, 由此形成了多视图数据. 多视图数据具有一致性和互补性两个特点[5−7], 其中一致性意味着不同视图共享一部分相同的信息, 并表现出一致的行为; 互补性意味着每个视图中除了有共享的信息之外, 还有单个视图特有的信息. 在一致性和互补性给多视图数据带来了更多信息的同时, 也给多视图离群点检测带来了更多的挑战. 现有的多视图离群点检测方法都是通过探索多视图数据的一致性而提出的, 其中在不同视图中表现出不一致行为的数据点可以被视为离群点. 具体来说, 按照文献 [8] 的术语, 多视图数据中的离群点可以分为以下 3 类.

• 类离群点 (class outlier): 在不同视图中表现出不一致行为的实例. 例如, 图 1 中的黄色菱形是一个类离群点, 在每个视图中都表现出正常的行为, 但在多个视图中表现出不一致的离群行为.

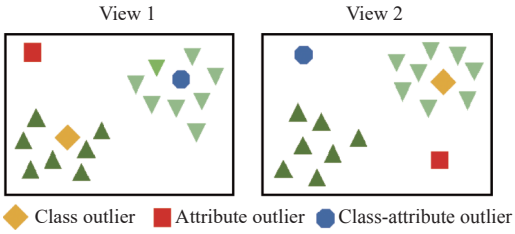

图 1 3 种类型离群点的说明

• 属性离群点 (attribute outlier): 在每个视图中都表现出一致离群行为的实例. 例如, 图 1 中的红色矩形是一个属性离群点, 与正常实例 (不同的绿色三角形) 相比, 在每个视图中都表现出一致的离群行为.

• 类-属性离群点 (class-attribute outlier): 属于类离群点和属性离群点的结合, 既在某些视图中表现出一致的离群行为, 又在某些不同视图中表现出不一致的离群行为. 例如, 图 1 中的蓝色八边形是一个类-属性离群点, 在视图 2 (View 2) 中表现出一致的离群行为, 在视图 1 (View 1) 和视图 2 (View 2) 这两个视图中表现出不一致的离群行为.

近年来, 已有一些方法被提出用于多视图离群点检测[8−16]. 它们在取得各自进展的同时, 仍然存在一些问题需要解决.

• 无法同时检测 3 种类型的离群点 (I1). 3 种类型的离群点分别为类离群点、属性离群点、类-属性离群点. 早期的方法只能检测出一种类型的离群点, 一些现有的方法无法同时检测出 3 种类型的离群点, 且在面对不同离群点比例的数据集时效果不佳.

• 处理没有聚类结构的数据集效果不佳 (I2). 早期的方法大多是基于聚类假设: 正常实例在多个视图中共享一致的聚类结构, 而离群点在不同视图中属于不一致的聚类. 这些方法在面对没有聚类结构的数据集时变得不那么有效.

• 处理高维度的数据集效果不佳 (I3). 数据集中的数据有很多特征. 一些现有的方法在面对高维数据时, 无法从丰富的特征中提取所需要的关键信息, 导致检测性能不尽人意.

• 面对高离群率数据集鲁棒性较低 (I4). 部分数据集中的离群率较高. 绝大部分现有的方法都是无监督的, 在模型学习阶段会受到数据集中离群点比率的影响, 在面对高离群率数据时检测性能不佳.

• 面对新增数据检测效率较低 (I5). 现有的基于聚类或基于近邻的方法在检测新增数据时, 为了保持检测精度, 需要将新增数据添加到原有数据中再进行检测, 导致不必要的重复检测, 检测效率较低.

本文总结了代表性方法在解决上述问题上的能力, 发现仍缺乏能够同时解决上述所有问题的方法 (如表 1 所示, √表示方法可以解决这个问题, 否则为×). 这促使本文找出一个解决方法来同时解决这 5 个问题.

表 1 代表性方法和本文提出的方法在解决 I1–I5 共 5 个问题时的比较

Method	I1	I2	I3	I4	I5
HOAD[9]	×	×	×	×	×
AP[10]	×	×	×	×	×
DMOD[11]	√	×	×	×	×
MLRA[12]	√	×	×	×	×
CRMOD[13]	√	×	×	×	×
LDSR[8]	√	×	×	×	×
MODDIS[14]	√	√	×	×	×
MUVAD[15]	√	√	×	×	×
NCMOD[16]	√	√	√	×	×
IRCGN (proposed method)	√	√	√	√	√

本文提出一种用于高效多视图离群点检测的视图内重建和跨视图生成网络 (intra-view reconstruction and cross-view generation network for effective multi-view outlier detection, IRCGN). 通过使用无离群点的正常数据进行对抗性训练, IRCGN 可以解决上述 5 个问题 (如表 1 所示), 从而有效检测 3 种类型的多视图离群点. 具体来说, 本文的主要贡献归纳如下.

• 本文提出一种新的视图内重建和跨视图生成网络模型. 模型通过对抗性训练学习一个有效的潜空间, 使 IRCGN 能处理高维数据 (I3), 在对抗性训练过程中使用正常数据作为训练集, 使得模型更好捕捉正常数据的分布, 提高模型对高离群率数据的鲁棒性 (I4), 模型训练完成后可以直接对数据进行检测, 不需要对数据集分布进行假设, 从而可以直接处理没有聚类结构的数据集 (I2).

• 本文提出一个新的离群值计算方法. 该方法通过视图内重建得分和跨视图生成得分两部分来有效检测 3 种类型的离群点 (I1). 该离群值计算方法在保持检测精度的同时, 可以直接对单条数据进行检测, 避免不必要的重复检测, 检测效率较高 (I5).

据我们所知, 这是第 1 个使用生成对抗网络 (GAN) 进行多视图离群点检测的方法, 广泛的实验验证了本文方法的优越性. 本文第 1 节介绍了与本文所提方法最相关的两类研究工作, 包括多视图离群点检测和基于生成对抗网络的异常检测. 第 2 节介绍了半监督多视图离群点检测的问题设置, 提出了用于高效多视图离群点检测的视图内重建和跨视图生成网络以及对应的离群值计算方法和优化算法. 第 3 节从多个角度对本文所提出算法进行实验评估和分析. 第 4 节对全文进行总结, 并展望了下一步工作.

1 相关工作

1.1 多视图离群点检测

传统的离群点检测方法只能处理单视图数据, 着重于检测在单个视图中表现出离群行为的离群点[17−19]. 随着多视图学习的发展, 多视图离群点检测成为一个新的研究课题. 早期的方法试图找到具有不一致跨视图聚类成员的样本. 水平异常检测 (HOAD)[9]是这些方法的开创性工作, 通过约束性频谱聚类对每个视图中的样本进行聚类, 将在不同视图中属于不同聚类的样本视为离群点. 基于亲和传播聚类 (AP)[10]和共识聚类[20]的多视图异常检测方法的目的也是通过探索多个视图之间的聚类结果不一致性来检测不一致的离群点, 即类离群点, 但这些方法都只能检测类离群点. MLRA[12]和 DMOD[11]分别将 $l2$, 1-norm 应用于低秩矩阵分析和 K-means 聚类来模拟误差项, 并

提出各自的离群值计算方法来同时检测类离群点和属性离群点. 为了克服成对约束随着视图数量增加造成的严重复杂化, LDSR[8]和 CRMOD[13]分别在 MLRA 和 DMOD 的基础上, 通过学习多个视图的跨视图一致表示来处理两个及两个以上的视图. 然而, 上述方法都依赖于聚类假设, 不能处理没有聚类结构的数据. 为了解决这一问题, 最近的工作大多假设一个正常实例的多个视图具有相似的邻域结构, 通过类离群点在不同视图邻域结构的不一致性来检测没有聚类结构的数据集. MUVAD[15]首先提出了基于近邻的离群值计算方法, 但离群值得分直接使用原始数据计算距离, 在面对高维数据时检测性能较低; NCMOD[16]通过自动编码器将每个视图的数据映射到一个全面的潜空间, 通过捕捉不同视图中不一致的邻域结构来检测多种类型的离群点. 但方法的核心仍然是近邻算法, 在面对新增数据时, 需要将新增数据添加到原有数据中重新进行检测, 导致检测效率较低, 且检测时间随测试样本数量的增加呈指数上升. 上述的算法大多是无监督的, 模型的学习过程同时受到正常数据和离群点的影响, 存在数据鲁棒性问题, 在面对较高离群率的数据集时检测性能不佳.

1.2 基于生成对抗网络的异常检测

使用生成对抗网络进行异常检测一般基于一个直观的假设, 正常样本与生成样本之间的某种形式的残差较小. 模型的训练过程旨在学习生成网络 G 的潜空间, 以便潜空间可以更好捕捉数据的常态性[21]. AnoGAN[22]是首个利用 GAN 进行异常检测的方法. 通过使用正常样本训练, 假设与正常检测样本相比, 异常检测样本与其生成样本之间的残差更大, 其中潜空间被用来捕捉训练数据的基本分布. 然而 AnoGAN 需要通过多次反向传播迭代计算损失函数来找到最接近检测样本的潜空间向量, 导致其计算效率较低. EGBAD[23]建立在双向 GAN (BIGAN)[24]的基础上, 通过一个编码器学习图像空间到潜空间的映射, 显著提高了检测速度. ALAD[25]增加了两个判别器, 将对抗性学习损失和条件熵损失统一为循环一致性损失, 进一步限制潜空间的数据分布. GANomaly[26]则通过修改网络结构和增加损失函数来进一步改进生成器, 提高了异常检测的性能. 之后随着新技术的发展, 基于 GAN 的异常检测被应用到越来越多的领域中[27−29]. 现有的基于 GAN 的异常检测算法, 大多都是为单视图数据设计的, 不能检测出 3 种类型的多视图离群点. 据我们所知, 目前仍缺乏利用生成对抗网络来检测多视图数据的方法.

2 提出的模型

标准的生成对抗网络由两个神经网络组成, 一个是生成器 G, 另一个是判别器 D, 模型旨在通过对抗性训练将一个给定的噪声分布拟合成任何一个想要的分布. 模型的训练过程是一个两人零和博弈过程, 生成器 G 试图生成类似真实样本的生成样本. 判别器 D 试图将真实样本和生成器生成的样本区分开. 综上得到 GAN 的目标函数:

$$\min_G \max_D V(D,G) = \mathbb{E}_{x\sim p_{\text{data}}(x)}[\log D(x)] + \mathbb{E}_{z\sim p_z(z)}[\log(1-D(G(z)))] \tag{1}$$

其中, $p_z(z)$ 表示的是潜空间分布, $p_{\text{data}}(x)$ 表示的是样本空间分布. 经过训练的 GAN 可以被用来拟合任何样本分布, 并可以确定一个给定的新样本是否在分布中.

2.1 问题设置与提出的模型

给定一个有 V 个视图的多视图数据集 $X=\{X^1,X^2,\ldots,X^V\}$, 其中 $X^v=\{X_1^v,X_2^v,\ldots,X_M^v,\ldots,X_N^v\}$, $X^v\in\mathbb{R}^{d_v\times N}$ 是在第 v 个视图中观察到的样本集, N 和 d_v 分别表示第 v 个样本集中样本的数量和特征维度. 多视图数据集中前 M 个样本是已知标签的正常样本, 后 $N-M$ 个样本是未知标签的正常样本或离群点. 半监督的多视图离群点检测方法旨在通过使用正常多视图数据训练完成的模型来检测未知标签的多视图数据. 本文提出的新模型通过视图内重建和跨视图生成两个模块同时检测 3 种类型的离群点. 不失一般性, 本文重点描述两视图数据下的自然假设, 当然该假设可以直接扩展到处理 3 个以上的视图. 对于一个有两个视图的数据集 X, 正常样本表示为 $X_i=\left[X_i^{(1)},X_i^{(2)}\right]$, 属性离群点表示为 $X_j=\left[X_j^{(1)},X_j^{(2)}\right]$, 类离群点表示为 $X_k=\left[X_k^{(1)},X_k^{(2)}\right]$.

首先对于属性离群点 $X_j=\left[X_j^{(1)},X_j^{(2)}\right]$, 本文提出视图内重建假设: 属性离群点的视图内重建误差之和大于正常样本的视图内重建误差之和. 用 $X_i^{(1)}$ 和 $X_i'^{(2)}$ 分别表示正常样本第 1 个和第 2 个视图的视图内重建样本. 这个假设可以被表述为:

$$\left\|X_j^{(1)}-X_j'^{(1)}\right\|+\left\|X_j^{(2)}-X_j'^{(2)}\right\|>\left\|X_i^{(1)}-X_i'^{(1)}\right\|+\left\|X_i^{(2)}-X_i'^{(2)}\right\| \tag{2}$$

其中, $\|\cdot\|$ 代表两个样本之间的某种距离, 例如欧氏距离、马氏距离等. 通过使用正常样本进行对抗性训练, 视图内神经网络可以较好地重建待检测数据集中的正常样本. 但当面对待检测数据集中从未学习过的属性离群点时, 使用其每个视图作为输入重建自身的效果较差, 导致属性离群点的视图内重建误差之和较大. 其次对于类离群点 $X_k=\left[X_k^{(1)},X_k^{(2)}\right]$, 本文提出跨视图生成假设: 类离群点的跨视图生成误差之和大于正常样本的跨视图生成误差之和. 用 $X_i''^{(1)}$ 和 $X_i''^{(2)}$ 分别表示正常样本第 2 个和第 1 个视图的跨视图生成样本. 这个假设可以被表述为:

$$\left\|X_k^{(1)}-X_k''^{(2)}\right\|+\left\|X_k^{(2)}-X_k''^{(1)}\right\|>\left\|X_i^{(1)}-X_i''^{(2)}\right\|+\left\|X_i^{(2)}-X_i''^{(1)}\right\| \tag{3}$$

通过使用正常数据进行对抗性训练, 跨视图神经网络根据待检测数据集中正常样本的每个视图作为输入较好地生成其他视图. 但当面对待检测数据集中不具有多视图一致性的类离群点时, 使用其每个视图作为输入生成其他视图的效果较差, 导致类离群点的跨视图生成误差之和较大. 同理, 第 3 类离群点即类-属性离群点同时具有前两种离群点的特征, 导致视图内重建误差之和与跨视图生成误差之和都很大.

受单视图异常检测算法 GANomaly[26]启发, 本文进一步约束多视图数据的潜空间, 并通过潜空间向量的重建和生成来检测多视图离群点. 根据上述的视图内重建假设和跨视图生成假设, 本文提出了 IRCGN, 它由跨视图生成模块和视图内重建模块组成, 其关键目标是通过对抗性训练来约束多视图数据的潜空间, 该空间可以很好地保留用于多视图数据重建或生成的关键信息. 提出的模型中视图内神经网络自动提取每个视图的内在信息用于重建, 跨视图神经网络则确保所有视图的内在信息是一致的. 针对 3 种类型的多视图离群点, 将离群值计算方法设置为潜空间向量的视图内重建得分和跨视图生成得分的组合.

图 2 给出了本文提出的 IRCGN 网络模型结构, 对于一个有 V 个视图的多视图数据集, 模型在每个视图中都有 $V-1$ 对神经网络组. 在每对神经网络组中, 偏上的为视图内神经网络, 偏下的为跨视图神经网络. 以第 i 个视图为例, 其中第 j $(j\neq i)$ 个跨视图神经网络 $CG^{(i,j)}$ 和视图内神经网络 $IR^{(i)j}$ 分别以第 j 个视图的数据 $X^{(j)}$ 和当前视图的数据 $X^{(i)}$ 作为输入. 使用蓝色指代第 1 个视图的数据以及处理第 1 个视图数据的神经网络. 同理, 绿色、红色和棕色分别指代第 2 个、第 $V-1$ 个和第 V 个视图. 具体来说, 在每对神经网络组中, 各有 4 个子网络. 以第 i 个视图为例, 4 个子网络分别为第 j 个跨视图生成器 $G_c^{(i,j)}$, 跨视图判别器 $D_c^{(i,j)}$, 视图内生成器 $G_a^{(i)j}$ 和视图内判别器 $D_a^{(i)j}$. 其中下标 c 代表跨视图神经网络, a 代表视图内神经网络. 表 2 对视图内神经网络和跨视图神经网络内部不同子网络进行了详细描述, 其中 d_i 是第 i 个视图的特征维度, m 是生成器中潜空间的维度.

跨视图生成器 $G_c^{(i,j)}:\mathbb{R}^{d_j}\to\mathbb{R}^m\to\mathbb{R}^{d_i}\to\mathbb{R}^m$. 对于第 i 个视图中第 j 个跨视图生成器, 主要处理第 j 个视图的数据, 具体由 3 个组件构成, 依次为编码器 $E_{1c}^{(i,j)}$、解码器 $D_{1c}^{(i,j)}$、编码器 $E_{2c}^{(i,j)}$, 其中编码器和解码器均由多个全连接层和 ReLU 层组成. $G_c^{(i,j)}$ 依次将第 j $(j\neq i)$ 个原始视图数据 $X^{(j)}$ 编码成潜空间向量 $z_c^{(i,j)}$, 利用潜空间向量生成当前第 i 个视图的生成数据 $X_c'^{(i,j)}$, 并根据当前视图的生成数据再次编码潜空间向量 $z_c'^{(i,j)}$, 其中生成数据 $X_c'^{(i,j)}$ 可以作为当前跨视图生成器 $G_c^{(i,j)}$ 的输出.

视图内生成器 $G_a^{(i)j}:\mathbb{R}^{d_i}\to\mathbb{R}^m\to\mathbb{R}^{d_i}\to\mathbb{R}^m$. 为了平衡跨视图生成模块和视图内重建模块的权重, 在单个视图中同样构建了相同数量的视图内生成器, 即对于第 i 个视图中第 j 个跨视图生成器, 都有一个相对应的第 j 个视图内生成器. 具体由 3 个组件构成, 依次为编码器 $E_{1a}^{(i)j}$、解码器 $D_{1a}^{(i)j}$、编码器 $E_{2a}^{(i)j}$, 其中编码器和解码器均由多个全连接层和 ReLU 层组成. 在第 i 个视图中的所有视图内生成器都依次将当前视图的原始数据 $X^{(i)}$ 编码成潜空间向量 $z_a^{(i)j}$, 利用潜空间向量重建当前视图的重建数据 $X_a'^{(i)j}$, 并根据当前视图的重建数据再次编码潜空间向量 $z_a'^{(i)j}$, 其中重建数据 $X_a'^{(i)j}$ 可以作为当前视图内生成器 $G_a^{(i)j}$ 的输出.

跨视图判别器 $D_c^{(i,j)}:\mathbb{R}^{d_i}\to\{0,1\}$. 对于每个跨视图生成器, 都有一个专属的跨视图判别器. 每个跨视图判别器都是一个编码器, 由多个全连接层、ReLU 层和一个 Sigmoid 层组成. 对于第 i 个视图中的所有跨视图判别器, 输入原始视图数据 $X^{(i)}$ 和使用各自第 j 个原始视图数据生成的第 i 个视图的生成数据 $X_c'^{(i,j)}$, 可以输出对生成数据 $X_c'^{(i,j)}$ 的真假判断, 即 $D_c^{(i,j)}\left(X_c'^{(i,j)},X^{(i)}\right)\in\{0,1\}$, 其中 1 代表真, 0 代表假. 跨视图判别器的结果也将反馈给相对应的跨

视图生成器, 促使跨视图生成器生成更真实的数据.

视图内判别器 $D_a^{(i)j}:\mathbb{R}^{d_i}\to\{0,1\}$. 同样对于每个视图内生成器, 都有一个专属的视图内判别器. 视图内判别器和跨视图判别器结构相同. 对于第 i 个视图中的所有视图内判别器, 输入原始视图数据 $X^{(i)}$ 和使用原始视图数据重建的重建数据 $X_a'^{(i)j}$, 可以输出对重建数据 $X_a'^{(i)j}$ 的真假判断, 即 $D_a^{(i)j}\left(X_a'^{(i)j},X^{(i)}\right)\in\{0,1\}$, 其中 1 代表真, 0 代表假. 视图内判别器的结果也将反馈给相对应的视图内生成器, 促使视图内生成器重建更真实的数据.

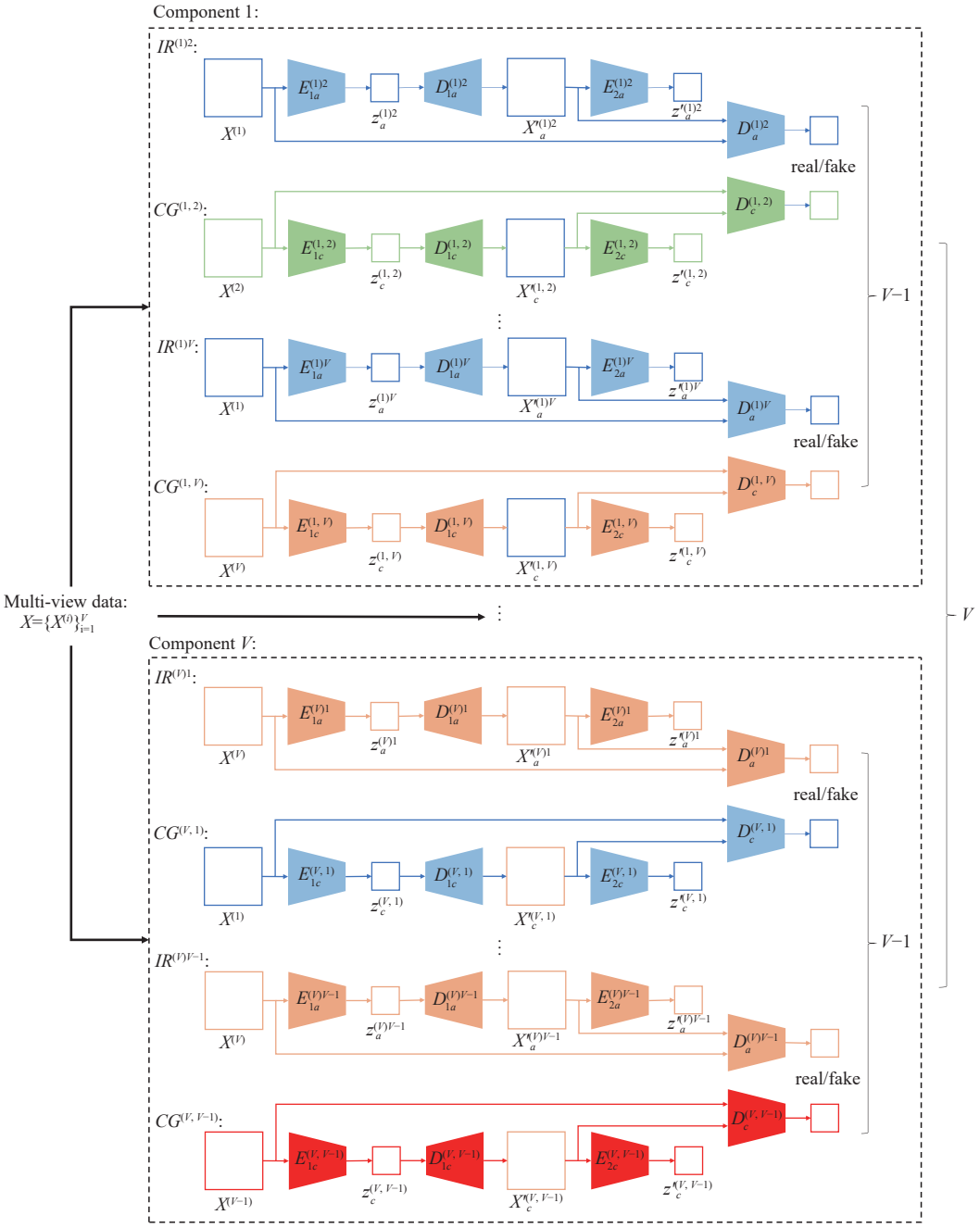

图 2 本文提出的 IRCGN 模型

表 2 模型内部对应于第 i 个视图中第 j 个神经网络组结构概述

Sub-networks	Input shape → Output shape	Components and layer information
跨视图生成器 $G_c^{(i,j)}$	$d_j \to m$ $m \to d_i$ $d_i \to m$	编码器 $E_{1c}^{(i,j)}$, 全连接层、ReLU层 解码器 $D_{1c}^{(i,j)}$, 全连接层、ReLU层 编码器 $E_{2c}^{(i,j)}$, 全连接层、ReLU层
视图内生成器 $G_a^{(i)j}$	$d_i \to m$ $m \to d_i$ $d_i \to m$	编码器 $E_{1a}^{(i)j}$, 全连接层、ReLU层 解码器 $D_{1a}^{(i)j}$, 全连接层、ReLU层 编码器 $E_{2a}^{(i)j}$, 全连接层、ReLU层
跨视图判别器 $D_c^{(i,j)}$	$d_i \to 1$	编码器 $D_c^{(i,j)}$, 全连接层、ReLU层、Sigmoid层
视图内判别器 $D_a^{(i)j}$	$d_i \to 1$	编码器 $D_a^{(i)j}$, 全连接层、ReLU层、Sigmoid层

2.2 跨视图生成和视图内重建

本文的目标函数由两个模块组成, 分别为跨视图生成模块和视图内重建模块.

对于跨视图生成模块来说, 为了保留跨视图低维潜空间中的内在关键信息, 通过 3 种损失函数来约束潜空间.

• 跨视图特征匹配损失 $\mathcal{L}_c^{\text{fml}}$: 为了增加对抗性训练的稳定性, 通过跨视图特征匹配损失来约束潜空间. 模型对于来自第 j 个视图数据分布 $p_X(X^{(j)})$ 的给定输入 $X^{(j)}$, 生成当前视图的生成数据 $\left(X_c'^{(i,j)} = G_c^{(i,j)}(X^{(j)})\right)$, 判别器的最后一层全连接层 $\left(D_{c\text{-}fcl}^{(i,j)}\right)$ 依据当前视图原始数据 $X^{(i)}$ 和生成数据 $X_c'^{(i,j)}$ 输出各自的特征表示. 跨视图特征匹配损失分别计算特征表示之间的 $\mathcal{L}_2$ 距离. 所有视图的跨视图特征匹配损失之和为:

$$\mathcal{L}_c^{\text{fml}} = \sum_{i=1}^{V}\sum_{j=1,j\neq i}^{V}\mathbb{E}_{X^{(j)}\sim p_X\left(X^{(j)}\right)}\left\|D_{c\text{-}fcl}^{(i,j)}\left(X^{(i)}\right) - D_{c\text{-}fcl}^{(i,j)}\left(G_c^{(i,j)}\left(X^{(j)}\right)\right)\right\|_2 \tag{4}$$

• 跨视图一致性损失 $\mathcal{L}_c^{\text{con}}$: 为了使跨视图生成器更好地学习跨视图数据中的一致性, 通过跨视图一致性损失来惩罚一致性信息. 因为使用 $\mathcal{L}_1$ 具有更好的稀疏性, 所以将跨视图一致性损失定义为当前视图的原始数据 $X^{(i)}$ 和使用第 j 个视图数据在第 i 个视图的生成数据 $\left(X_c'^{(i,j)} = G_c^{(i,j)}(X^{(j)})\right)$ 之间的 $\mathcal{L}_1$ 距离. 所有视图的跨视图一致性损失之和被定义为:

$$\mathcal{L}_c^{\text{con}} = \sum\nolimits_{i=1}^{V}\sum\nolimits_{j=1,j\neq i}^{V}\mathbb{E}_{X^{(j)}\sim p_X\left(X^{(j)}\right)}\left\|X^{(i)} - G_c^{(i,j)}\left(X^{(j)}\right)\right\|_1 \tag{5}$$

• 跨视图潜空间损失 $\mathcal{L}_c^{\text{lan}}$: 为了更好地约束数据的跨视图潜空间, 最小化使用第 j 个视图数据生成的第 i 个视图数据的潜空间特征 $\left(z_c^{(i,j)} = E_{1c}^{(i,j)}(X^{(j)})\right)$ 和生成数据的编码特征 $\left(z'^{(i,j)}_c = E_{2c}^{(i,j)}\left(X'^{(i,j)}_c\right)\right)$ 之间的 $\mathcal{L}_2$ 距离. 对于正常数据来说, 原始数据和生成数据彼此接近, 其特征在潜空间中就会保持很小的距离. 因此所有视图的跨视图潜空间损失之和被定义为:

$$\mathcal{L}_c^{\text{lan}} = \sum_{i=1}^{V}\sum_{j=1,j\neq i}^{V}\mathbb{E}_{X^{(j)}\sim p_X\left(X^{(j)}\right)}\left\|E_{1c}^{(i,j)}\left(X^{(j)}\right) - E_{2c}^{(i,j)}\left(G_c^{(i,j)}\left(X^{(j)}\right)\right)\right\|_2 \tag{6}$$

在视图内重建模块中, 仍通过 3 种不同的损失函数约束潜空间, 从而保留视图内低维潜空间中的内在关键信息.

• 视图内特征匹配损失 $\mathcal{L}_a^{\text{fml}}$: 通过视图内特征匹配损失来增加对抗性训练的稳定性. 模型对于来自第 i 个视图数据分布 $p_X(X^{(i)})$ 的原始数据 $X^{(i)}$, 重建当前视图的重建数据 $\left(X_a'^{(i)j} = G_a^{(i)j}(X^{(i)})\right)$. 视图内判别器的最后一层全连接层 $(D_{a\text{-}fcl}^{(i)j})$ 输出第 i 个视图的原始数据和重建数据的特征表示. 使用特征表示之间的 $\mathcal{L}_2$ 距离来表示视图内特征匹配损失. 所有视图的视图内特征匹配损失之和为:

$$\mathcal{L}_a^{\text{fml}} = \sum_{i=1}^{V}\sum_{j=1,j\neq i}^{V}\mathbb{E}_{X^{(i)}\sim p_X\left(X^{(i)}\right)}\left\|D_{a\text{-}fcl}^{(i)j}\left(X^{(i)}\right) - D_{a\text{-}fcl}^{(i)j}\left(G_a^{(i)j}\left(X^{(i)}\right)\right)\right\|_2 \tag{7}$$

• 视图内一致性损失 $\mathcal{L}_a^{\text{con}}$: 为了提高重建数据和原始数据的一致性, 在视图内特征匹配损失的基础上通过视图内一致性损失来惩罚一致性信息. 使用 $\mathcal{L}_1$ 来代替 $\mathcal{L}_2$ 来产生稀疏的权值. 因此, 用当前视图的原始数据 $X^{(i)}$ 和重

建数据 $\left(X_a'^{(i)j}=G_a^{(i)j}(X^{(i)})\right)$ 之间的 $\mathcal{L}_1$ 距离来表示视图内一致性损失. 所有视图的视图内一致性损失之和为:

$$\mathcal{L}_a^{\text{con}}=\sum_{i=1}^{V}\sum_{j=1,j\neq i}^{V}\mathbb{E}_{X^{(i)}\sim pX\left(X^{(i)}\right)}\left\|X^{(i)}-G_a^{(i)j}\left(X^{(i)}\right)\right\|_1 \tag{8}$$

• 视图内潜空间损失 $\mathcal{L}_a^{\text{lan}}$: 最小化第 i 个视图中原始数据 $X^{(i)}$ 的潜空间特征 $\left(z_a^{(i)j}=E_{1a}^{(i)j}(X^{(i)})\right)$ 和重建数据的潜空间特征 $\left(z_a'^{(i)j}=E_{2a}^{(i)j}\left(X_a'^{(i)j}\right)\right)$ 之间的 $\mathcal{L}_2$ 距离来更好地约束数据的视图内潜空间. 因此所有视图的视图内潜空间损失被定义为:

$$\mathcal{L}_a^{\text{lan}}=\sum_{i=1}^{V}\sum_{j=1,j\neq i}^{V}\mathbb{E}_{X^{(i)}\sim pX\left(X^{(i)}\right)}\left\|E_{1a}^{(i)j}\left(X^{(i)}\right)-E_{2a}^{(i)j}\left(G_a^{(i)j}\left(X^{(i)}\right)\right)\right\|_2 \tag{9}$$

结合公式 (4)–公式 (9), IRCGN 中生成器的目标函数 $\mathcal{L}_g$ 为:

$$\mathcal{L}_g=\mathcal{L}_c^{\text{fml}}+\mathcal{L}_a^{\text{fml}}+\mathcal{L}_c^{\text{lan}}+\mathcal{L}_a^{\text{lan}}+\omega_{\text{con}}(\mathcal{L}_c^{\text{con}}+\mathcal{L}_a^{\text{con}}) \tag{10}$$

其中, ω_{con} 是调整使用 $\mathcal{L}_1$ 距离的一致性损失对整体目标函数的权重参数.

2.3 离群值计算方法

多视图数据通过不同视图提供丰富信息的同时, 应该保持视图间信息的一致性. 对于多视图离群点检测任务来说, 一致性是指较小的跨视图生成误差. 利用输入样本和输入样本的重建或生成样本在潜空间学习到的特征表示, 本文提出一个新的离群值计算方法, 可以为每个测试实例 X_t 计算一个离群值得分. 离群值计算方法如下:

$$S_{\text{outlier}}(X_t)=S_c(X_t)+S_a(X_t) \tag{11}$$

其中,

$$S_c(X_t)=\sum_{i=1}^{V}\sum_{j=1,j\neq i}^{V}\left\|E_{1c}^{(i,j)}\left(X_t^{(j)}\right)-E_{2c}^{(i,j)}\left(G_c^{(i,j)}\left(X_t^{(j)}\right)\right)\right\|_2 \tag{12}$$

$$S_a(X_t)=\sum_{i=1}^{V}\sum_{j=1,j\neq i}^{V}\left\|E_{1a}^{(i)j}\left(X_t^{(i)}\right)-E_{2a}^{(i)j}\left(G_a^{(i)j}\left(X_t^{(i)}\right)\right)\right\|_2 \tag{13}$$

分别为跨视图生成得分和视图内重建得分. 这种离群值计算方法有助于同时识别 3 种类型的离群点.

• 对于正常的样本: 由于正常样本的视图内潜空间表征和训练数据是相似的, 重建误差较小. 由于正常样本不同视图具有一致性, 模型可以很好地从一个视图中生成另一个视图. 因此, 正常样本的视图内重建得分 S_a 和跨视图生成得分 S_c 都很小. 因此离群值得分 S_{outlier} 较小.

• 对于属性离群点: 由于属性离群点在每一个视图上都与大多数的正常样本不相似, 视图内神经网络较难提取属性离群点的表征, 导致重建效果不佳, 这就产生了较高的视图内重建得分 S_a . 因此离群值得分 S_{outlier} 较大.

• 对于类离群点: 由于类离群点在多个视图中是不一致的, 跨视图神经网络难以从一个视图中提取关键表征生成另一个视图, 导致了较高的跨视图生成得分 S_c . 因此离群值得分 S_{outlier} 较大.

• 对于类-属性离群点: 由于类-属性离群点同时包含属性离群点和类离群点的特征, 导致其视图内重建得分 S_a 和跨视图生成得分 S_c 都较高. 因此离群值得分 S_{outlier} 较大.

IRCGN 算法的具体算法过程如算法 1.

算法 1. 用于高效多视图离群点检测的视图内重建和跨视图生成网络算法.

Input: multi-view training dataset X_{train} , multi-view test dataset X_{test} , number of generator training iterations k;
Output: X_{test}'s outlier-score.

1. Initialize $\left(G_c^{(i,j)}(\cdot);G_a^{(i)j}(\cdot);D_c^{(i,j)}(\cdot);D_a^{(i)j}(\cdot)\right)_{i=1,j=1,j\neq i}^{V}$;
2. **repeat**
3. **for** n=1 to k **do**

4. **for** i=1 to V **do**
5. **for** j=1 to V **do**
6. **if** i==j **then**
7. continue
8. **end if**
9. Update $\left(G_c^{(i,j)}(\cdot);G_a^{(i)j}(\cdot)\right)_{i=1,j=1,j\neq i}^{V}$ using Equation (10);
10. **end for**
11. **end for**
12. **end for**
13. **for** i=1 to V **do**
14. **for** j=1 to V **do**
15. **if** i==j **then**
16. continue
17. **end if**
18. Update $\left(D_c^{(i,j)}(\cdot);D_a^{(i)j}(\cdot)\right)_{i=1,j=1,j\neq i}^{V}$ using Equation (14);
19. **end for**
20. **end for**
21. **until** convergence
22. **for** i=1 to V **do**
23. **for** j=1 to V **do**
24. **if** i==j **then**
25. continue
26. **end if**
27. Calculate X_{test}'s outlier-score using Equation (11);
28. **end for**
29. **end for**
30. **return** X_{test}'s outlier-score.

2.4 优化算法

本文提出的 IRCGN 的优化方法总结于算法 1. 首先在不含离群点的多视图训练数据上对抗性地训练这些神经网络, 在训练过程中采用自适应矩估计优化器 (Adam) 来优化本文的目标函数. 为了提高训练的稳定性, 在每一轮训练中先训练 k 次生成器之后再训练 1 次判别器. 训练完成后, 将待检测数据依次送入到训练好的神经网络中, 得到各自的离群值得分. 优化工作在更新生成器和更新判别器之间迭代进行.

- 更新生成器网络 $\left(G_a^{(i)j}(\cdot);G_c^{(i,j)}(\cdot)\right)_{i=1,j=1,j\neq i}^{V}$, 当判别器固定时, 生成器的目标函数为公式 (10).
- 更新判别器网络 $\left(D_a^{(i)j}(\cdot);D_c^{(i,j)}(\cdot)\right)_{i=1,j=1,j\neq i}^{V}$, 当生成器固定时, 判别器的目标函数 $\mathcal{L}_d$ 为:

$$\mathcal{L}_d=\mathcal{L}_c^d+\mathcal{L}_a^d \tag{14}$$

$$\mathcal{L}_c^d=\sum_{i=1}^{V}\sum_{j=1,j\neq i}^{V}\max_{D_c^{(i,j)}}\left(\mathbb{E}_{X^{(i)}\sim p_X\left(X^{(i)}\right)}\left[\log D_c^{(i,j)}\left(X^{(i)}\right)\right]+\mathbb{E}_{X^{(j)}\sim p_X\left(X^{(j)}\right)}\left[\log\left(1-D_c^{(i,j)}\left(G_c^{(i,j)}\left(X^{(j)}\right)\right)\right)\right]\right) \tag{15}$$

$$\mathcal{L}_a^d=\sum_{i=1}^{V}\sum_{j=1,j\neq i}^{V}\max_{D_a^{(i)j}}\left(\mathbb{E}_{X^{(i)}\sim p_X\left(X^{(i)}\right)}\left[\log D_a^{(i)j}\left(X^{(i)}\right)\right]+\mathbb{E}_{X^{(i)}\sim p_X\left(X^{(i)}\right)}\left[\log\left(1-D_a^{(i)j}\left(G_a^{(i)j}\left(X^{(i)}\right)\right)\right)\right]\right) \tag{16}$$

3 实 验

3.1 实验设置

3.1.1 数据集准备

公共数据集都是没有多视图离群点的分类数据集, 现有的对多视图离群点检测进行研究的工作都是以一定的比例生成具有 3 种类型离群点的多视图数据集. 本文沿用之前工作 (LDSR[8]、NCMOD[16]) 中的方法对数据集进行处理, 并使用了两个广泛使用的高维基准数据集 MNIST、TTC, 它们的维度分别为 784 和 7 507. 为了得到不同离群点比例的多视图数据集, 首先对于单视图基准数据集, 将原始特征平均分成 V 个特征子集, 其中每个子集对应一个视图. 之后是生成 3 种类型的离群点, 在多类数据集中随机选择 2 个或 3 个类作为“正常类”, 将属于“正常类”的样本提取出来作为原始数据集, 剩下的类在这里被视为“离群类”. 出于实验目的, 本文将原始数据集划分为训练集和测试集, 其中训练集均是正常的样本, 对应真实世界中易于得到的领域专家标记数据集, 这里的 3 种类型离群点都是在测试集中生成的. 对于类离群点, 在“正常类”的两个不同的类别中随机选择两个样本, 在 $\lfloor v/2 \rfloor$ 视图中交换它们的特征向量, 在其余的视图中保持不变. 对于属性离群点, 与 LDSR 中直接使用随机值进行替换不同, 本文采用 NCMOD 的方案, 使用“离群类”中随机抽取的样本来替换属性离群点中所有视图的特征向量. 这是因为随机值与样本特征的区别较大, 导致在检测时太容易被检测出. 对于类-属性离群点, 同样在“正常类”的两个不同的类别中随机选择两个对象, 在 $\lfloor v/2 \rfloor$ 视图中交换它们的特征向量, 在剩余的视图中使用“离群类”中随机样本的剩余视图特征向量进行替换.

对于每个数据集, 按照文献 [16], 考虑构建为 6 个子集: (i) 2% 的属性离群点, 5% 的类离群点, 8% 的类-属性离群点. (ii) 2% 的属性离群点, 8% 的类离群点, 5% 的类-属性离群点. (iii) 5% 的属性离群点, 2% 的类离群点, 8% 的类-属性离群点. (iv) 5% 的属性离群点, 8% 的类离群点, 2% 的类-属性离群点. (v) 8% 的属性离群点, 2% 的类离群点, 5% 的类-属性离群点. (vi) 8% 的属性离群点, 5% 的类离群点, 2% 的类-属性离群点. 本文采用数据集英文名的首字母加上子集编号的方式来表示某个数据集的检测子集, 例如, 用 M-i 来表示 MNIST 数据集的第 1 个子集, 用 T-ii 来表示 TTC 数据集的第 2 个子集, 并以此类推. 通过调整 3 种离群点的比例, 可以对比分析方法能否同时检测到 3 种多视图离群点. 出于实验目的, 对于不同样本数的数据集, 将 MNIST 和 TTC 检测子集的样本数分别设置为 1 000 和 100.

3.1.2 对比方法与评价指标

将 IRCGN 方法与 8 个代表性方法 (HOAD[9]、AP[10]、MLRA[12]、LDSR[8]、MODDIS[14]、NCMOD[16]、IF[30]、GANomaly[26]) 进行比较. 其中 IF、GANomaly 都是用于单视图离群点检测的代表性方法, 将它们用来验证单视图方法在多视图数据上的检测性能. 为了确保单视图离群点检测方法的顺利工作, 将多个视图的数据串联成一个视图作为输入. HOAD、AP、MLRA 是成对约束方法, 无法明确地扩展到 3 个或 3 个以上视图. 因此, 本文在每一对视图中计算它们的离群值得分, 然后将所有视图对的离群值得分相加作为最终的离群值得分. 值得注意的是 MLRA 无法处理不同视图维度不一致的情况. 这些方法都是在一台 2.4 GHz 128 GB 内存的服务器上运行的, 运行环境为 Python 3.7.

对于评价指标, 本文采用广泛使用的 AUC, 即 ROC 曲线下面积来衡量多视图离群点检测算法的性能, 其中 AUC 越高表明检测算法越好. 为了避免数据构建的波动以及检测的随机性, 对同一数据集重复生成 50 次离群点, 并计算所有方法在这 50 个数据集上 AUC 的平均值和标准差.

3.2 对比实验结果

3.2.1 检测性能对比

表 3 和表 4 展示了本文提出的 IRCGN 与其他对比方法在高维多视图数据集上的检测性能 (AUC 值, 平均值±标准差), 最佳的结果加粗表示. 其中, 二视图 TTC 数据集的两个视图特征数分别为 3 753 和 3 754, MLRA 作为一

种无法处理数据集中不同视图维度不一致的检测方法, 其检测结果用符号－来代替. 表 3, 表 4 清楚地显示, IRCGN 在所有的数据集和设置上都大大超过其他所有的方法, 这证明了 IRCGN 在高维多视图离群点检测上的有效性. 但两个数据集间的检测性能仍存在差异, 可能的原因是 TTC 数据集的维度过高, 训练样本较少, 出现欠拟合的情况, 从而导致生成和重建效果较差, 检测性能较低. 此外, 对于同一数据集的 6 种不同比例检测子集, 本文所提出的 IRCGN 的检测性能都很稳定. 这是由于 IRCGN 可以有效捕捉高维数据集上正常数据的特征, 具有不一致特征的 3 种多视图离群点会被很容易地识别出来, 导致其离群值得分较高.

表 3 二视图分割的多视图数据集检测性能

Dataset	IF	GANomaly	HOAD	AP	MLRA	LDSR	MODDIS	NCMOD	IRCGN
M-i	0.628±0.017	0.812±0.020	0.514±0.086	0.935±0.010	0.897±0.010	0.902±0.011	0.806±0.024	0.926±0.014	**0.973±0.005**
M-ii	0.607±0.021	0.824±0.017	0.495±0.060	0.942±0.009	0.902±0.011	0.898±0.011	0.798±0.022	0.913±0.017	**0.978±0.004**
M-iii	0.633±0.018	0.822±0.018	0.499±0.098	0.901±0.013	0.864±0.018	0.883±0.012	0.790±0.025	0.912±0.014	**0.965±0.005**
M-iv	0.613±0.023	0.845±0.014	0.494±0.057	0.916±0.012	0.865±0.013	0.886±0.011	0.781±0.022	0.890±0.019	**0.975±0.004**
M-v	0.631±0.024	0.846±0.020	0.516±0.092	0.858±0.017	0.830±0.019	0.856±0.015	0.758±0.027	0.873±0.020	**0.962±0.004**
M-vi	0.625±0.021	0.854±0.015	0.524±0.085	0.866±0.016	0.827±0.019	0.857±0.015	0.726±0.023	0.865±0.018	**0.968±0.004**
T-i	0.548±0.081	0.508±0.076	0.513±0.105	0.649±0.072	－	0.558±0.074	0.562±0.082	0.603±0.078	**0.728±0.075**
T-ii	0.549±0.065	0.511±0.085	0.520±0.089	0.654±0.070	－	0.550±0.072	0.568±0.068	0.611±0.058	**0.698±0.067**
T-iii	0.568±0.079	0.514±0.082	0.480±0.117	0.665±0.070	－	0.528±0.083	0.597±0.062	0.619±0.077	**0.760±0.070**
T-iv	0.570±0.067	0.491±0.072	0.511±0.087	0.657±0.085	－	0.513±0.093	0.597±0.070	0.627±0.074	**0.714±0.091**
T-v	0.577±0.073	0.489±0.066	0.505±0.084	0.640±0.074	－	0.490±0.080	0.617±0.074	0.646±0.062	**0.721±0.076**
T-vi	0.597±0.080	0.528±0.081	0.500±0.107	0.650±0.086	－	0.502±0.080	0.627±0.084	0.640±0.084	**0.752±0.070**

表 4 三视图分割的多视图数据集检测性能

Dataset	IF	GANomaly	HOAD	AP	LDSR	MODDIS	NCMOD	IRCGN
M-i	0.650±0.020	0.806±0.019	0.513±0.105	0.915±0.008	0.868±0.013	0.774±0.025	0.888±0.026	**0.981±0.004**
M-ii	0.616±0.022	0.786±0.018	0.520±0.089	0.917±0.008	0.859±0.014	0.776±0.024	0.868±0.027	**0.979±0.003**
M-iii	0.653±0.022	0.841±0.018	0.493±0.069	0.881±0.012	0.870±0.014	0.765±0.023	0.899±0.019	**0.980±0.003**
M-iv	0.612±0.023	0.797±0.017	0.501±0.049	0.893±0.009	0.852±0.012	0.739±0.026	0.870±0.021	**0.976±0.004**
M-v	0.643±0.021	0.852±0.018	0.506±0.078	0.847±0.014	0.851±0.016	0.745±0.026	0.878±0.027	**0.976±0.004**
M-vi	0.627±0.022	0.832±0.018	0.503±0.057	0.853±0.015	0.848±0.015	0.726±0.022	0.867±0.024	**0.974±0.004**
T-i	0.555±0.087	0.508±0.073	0.497±0.086	0.627±0.087	0.586±0.076	0.580±0.079	0.612±0.079	**0.714±0.066**
T-ii	0.560±0.066	0.509±0.092	0.458±0.069	0.650±0.076	0.562±0.073	0.586±0.065	0.637±0.055	**0.695±0.063**
T-iii	0.579±0.074	0.520±0.084	0.492±0.088	0.654±0.077	0.584±0.075	0.617±0.069	0.644±0.062	**0.748±0.060**
T-iv	0.578±0.067	0.503±0.070	0.478±0.072	0.648±0.083	0.553±0.086	0.606±0.069	0.640±0.065	**0.725±0.077**
T-v	0.591±0.073	0.497±0.068	0.473±0.072	0.636±0.081	0.548±0.078	0.632±0.066	0.649±0.062	**0.731±0.068**
T-vi	0.605±0.081	0.525±0.084	0.481±0.094	0.649±0.100	0.567±0.082	0.634±0.087	0.650±0.062	**0.757±0.071**

单视图离群点检测方法和早期的多视图离群点检测方法不能有效处理高维数据集, 后来的方法 AP、LDSR、NCMOD 则有相对较好的性能. 具体来说, 由于单视图离群点检测方法无法捕捉多视图数据之间的不一致性, 导致 IF 和 GANomaly 这两个单视图离群点检测方法在多视图数据上检测性能较差, 且性能基本随着属性离群点的比例变化而变化. 对于多视图离群点检测方法来说, AP 在高类离群点比例时取得了较高的得分, 但在高属性离群点比例时效果较差, 原因在于 AP 只能检测出类离群点, 无法检测出属性离群点. MLRA 是一种有监督的多视图离群点检测方法, 在先验聚类类别不可知时变得不那么有效. LDSR 利用多视图子空间聚类方法获得了较好的结果, 但本质上仍是基于聚类的方法, 在面对没有聚类结构的数据集时效果不佳. 与 HOAD、AP、MLRA、LDSR 这些通过不一致的跨视图聚类结果来检测离群点的方法不同, 之后的方法 (MODDIS、NCMOD) 大多使用多视图一致的邻域假设来处理没有聚类结构的数据集. MODDIS 是第 1 个将深度神经网络应用到多视图离群点检测的方法, 它将多视图数据映射到一个潜在的完整空间, 但其直接目标并不是进行离群点检测, 导致检测性能较差. 本文提出的 IRCGN 避免了聚类和近邻假设, 因此可以直接处理没有聚类结构的数据集. NCMOD 是最具代表性的基于神经网络的多视图离群点检测方法, 其使用自动编码器对数据进行降维, 通过学习一个联合的邻域共识空间来检测高维数据的离群点. 但其模型是无监督的, 在学习过程中容易受到离群点的影响, 存在数据鲁棒性问题. 在 IRCGN

中, 通过使用正常数据进行对抗性训练, 模型可以更好捕捉正常数据的分布, 避免受到离群点的影响, 保证了其良好的检测性能.

3.2.2 高离群率鲁棒性对比

为了评估 IRCGN 对高离群率数据的鲁棒性, 本文在更高离群率的数据集上进行实验. 3 种离群点所占的比例为 1:1:1, 每种离群点的离群率从 5% 到 25% 增加, 步长为 5%. 如图 3 所示, IRCGN 在所有离群点比率下表现都优于其他方法. 且在其他方法的检测性能都随着离群率的增加而降低的情况下, IRCGN 仍保持十分稳定. 这证明了所提方法强大的检测能力和对高离群率数据的鲁棒性. 本文所提出的 IRCGN 是半监督的方法, 使用现实世界中易于获得的正常数据集进行训练, 可以避免离群点的污染, 更好捕捉正常数据的特征.

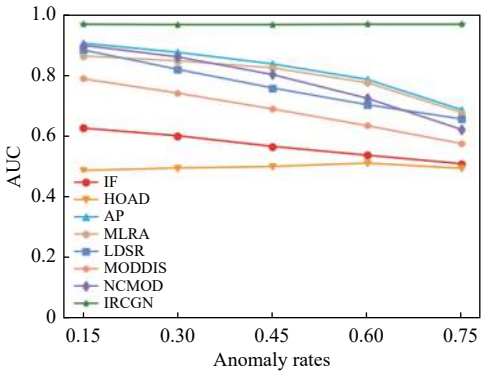

图 3 高离群率鲁棒性对比

3.2.3 新增数据检测效率对比

为了衡量模型对新增数据的检测效率, 本文将检测子集划分为先验数据集和新增数据集, 实验对比在有无先验数据集情况下不同模型对于新增数据集的检测时间 (50 个数据集的检测时间) 和检测精度 (AUC 值, 平均值±标准差). 在 MNIST 数据集上, 将先验数据集和新增数据集的比例划分为 9:1, 在 TTC 数据集上, 将先验数据集和新增数据集的比例划分为 1:1 (比例划分原则是为了保证有足够多的测试样本, 使得检测结果稳定可靠). 表 5 中, 百分数代表检测数据集占总数据集的比例, 100% 代表在检测先验数据集的基础上检测新增数据集, 未到 100% 代表仅检测新增数据集. 如表 5 所示, 之前的方法大多基于多视图一致的聚类或近邻假设, 在没有先验数据集的情况下检测精度较低, 但在使用先验数据集时, 又造成了不必要的重复检测, 导致检测效率较低. 本文所提出的 IRCGN 可以直接对新增数据进行检测, 在保持检测精度的同时检测时间也很短. 此外, 虽然 IRCGN 需要一定的时间进行对抗性训练, 但首先面对新增数据时, IRCGN 不需要重复训练; 其次模型训练是离线的, 可以通过离线训练或重载模型参数避免不必要的训练时间; 最后针对固定视图数的多视图数据, 在面对不同离群比率或不同样本数的检测子集时, IRCGN 只需要训练一次, 检测十分便捷. 在训练时间上, 本文模型的二视图 MNIST 数据集的训练时间为 2 943.972 s, 二视图 TTC 数据集的训练时间为 1 224.846 s.

表 5 二视图分割的新增数据检测效率对比

Dataset	MLRA		LDSR		MODDIS		NCMOD		IRCGN	
	AUC	Time (s)	AUC	Time (s)	AUC	Time (s)	AUC	Time (s)	AUC	Time (s)
M-i (100%)	0.897±0.010	1 133.821	0.902±0.011	1 392.096	0.813±0.022	2 586.228	0.929±0.014	94 608.144	**0.973±0.005**	**47.567**
M-i (10%)	0.820±0.048	141.918	0.789±0.059	68.655	0.716±0.078	2 525.970	0.559±0.109	9 867.783	**0.974±0.021**	**38.693**
M-iv (100%)	0.865±0.013	1 030.633	0.886±0.011	1 379.657	0.782±0.024	2 583.166	0.891±0.018	93 502.872	**0.976±0.004**	**45.308**
M-iv (10%)	0.773±0.057	143.901	0.778±0.069	69.239	0.695±0.084	2 525.532	0.545±0.104	9 880.363	**0.978±0.012**	**38.278**
T-iii (100%)	—	—	0.528±0.083	360.332	0.594±0.064	3 981.012	0.619±0.077	2 291.995	**0.760±0.070**	**54.140**
T-iii (50%)	—	—	0.529±0.012	92.915	0.565±0.121	3 965.244	0.560±0.128	1 195.180	**0.740±0.108**	**48.213**
T-vi (100%)	—	—	0.502±0.079	359.703	0.619±0.083	4 000.608	0.640±0.084	2 287.717	**0.752±0.070**	**53.912**
T-vi (50%)	—	—	0.505±0.010	192.901	0.618±0.111	3 967.437	0.608±0.117	1 198.816	**0.754±0.102**	**48.637**

3.3 参数敏感性分析

3.3.1 潜空间维度分析

本文进一步研究潜空间的维度 m 对模型的影响. 在视图内重建模块, 需要用单个视图的判别潜向量重建自身. 在跨视图生成模块, 需要用样本中一个视图的判别潜向量生成其余视图, 所以样本的潜空间维度十分重要. 面对不同维度的数据集, 需要实时调整模型的潜空间维度. 如图 4 所示, 如果维度较小, 则无法保留足够的判别性信息进行重建或生成; 然而维度较大也会导致模型的性能下降. 基于上述分析, 将潜空间的维度 m 设置为数据集原维度的 5%–10% 之间较为合适.

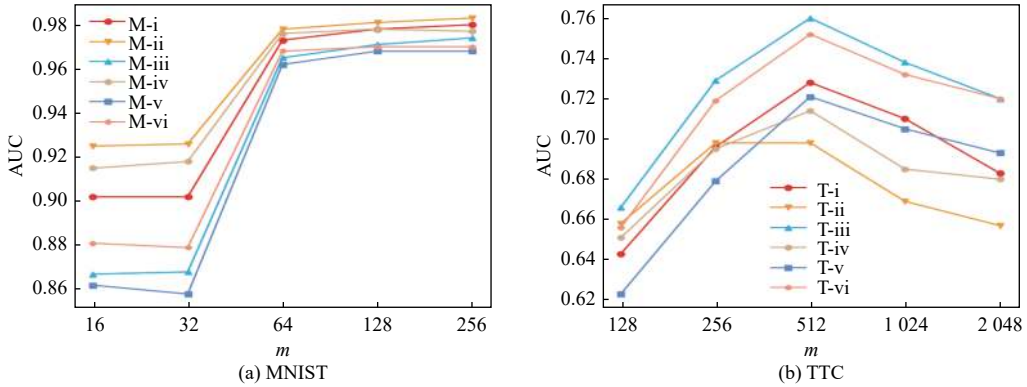

图 4 潜空间维度分析

3.3.2 参数 ω_{con} 和 k 影响分析

提出的 IRCGN 有两个参数, 分别为权重参数 ω_{con} 和生成器迭代次数 k. 其中权重参数 ω_{con} 是为了平衡生成器目标函数中一致性损失项, 生成器迭代次数 k 是为了网络的训练更加稳定. 图 5 显示了 ω_{con} 在{0.001, 0.01, 0.1, 1, 10, 100}和 k 在{1, 5, 10, 15, 20, 25, 30}中不同取值组合下的 AUC. 由图 5 可知, 迭代次数 k 不变时, IRCGN 的检测性能随着 ω_{con} 的增加有下降趋势; 权重参数 ω_{con} 不变时, IRCGN 的检测性能随着 k 值的增大有上升趋势. IRCGN 的检测性能在大部分参数组合下都很稳定, 并且在 ω_{con} =0.01 和 k=25 时, 取得最优性能, 因此可将这组参数值设置为二视图 TTC 数据集参数默认值. 值得注意的是, IRCGN 针对不同的数据集, 其参数都可以通过这种分析检测数据的 AUC 得分分布来快速调整.

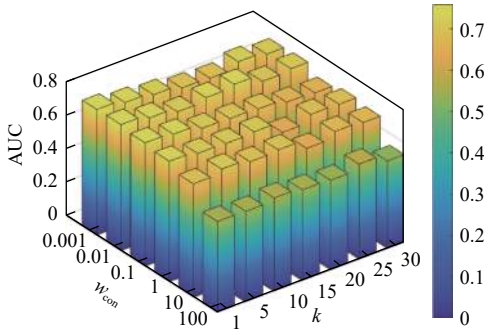

图 5 T-iii 数据集上 k 和 ω_{con} 不同组合的 AUC

3.4 消融实验

3.4.1 视图内重建和跨视图生成

为了研究视图内重建和跨视图生成两个子模块的有效性, 本文对 IRCGN 进行了消融研究. CG 代表只有跨视

图生成模块的模型, IR 代表只有视图内重建模块的模型. 本文在不同离群点比例的数据集上进行了实验, 计算了 AUC 值 (如表 6 所示). 实验表明, IR 的效果最差, CG 的效果次之, 提出的 IRCGN 效果最好. 这是因为 IR 只能检测到属性离群点, 无法捕捉到类离群点不同视图之间的不一致性, 所以它的性能基本随着属性离群点比例的增加而提高. 然而, CG 在可以捕捉类离群点中跨视图不一致性的同时, 还具备一部分检测属性离群点的能力, 因为属性离群点在每个视图都表现出一致的离群行为, 虽然属性离群点的不同视图之间具有一致性, 但跨视图神经网络难以从训练中未见过的属性离群点中编码关键特征并生成其他视图, 导致属性离群点的离群值得分也较大. 当然与完整重建相比, 仅通过编码的 CG 虽然能检测属性离群点, 但检测能力不强, 所以在每个数据集上的检测性能都略低于 IRCGN. 最后, 在所有数据集和设置上 IRCGN 一直优于 CG 和 IR. 因此, 这两个提议的子模块都能有效地提高所提出的模型在多视图离群点检测方面的性能, IRCGN 很好地融合了这两个子模块.

表 6 视图内重建和跨视图生成的消融实验

Dataset	IR	CG	IRCGN
M-i	0.750±0.019	0.959±0.006	**0.973±0.005**
M-iii	0.838±0.015	0.930±0.009	**0.965±0.005**
M-v	0.854±0.013	0.913±0.009	**0.962±0.004**
T-ii	0.571±0.062	0.639±0.069	**0.698±0.067**
T-iv	0.594±0.071	0.657±0.078	**0.714±0.091**
T-vi	0.621±0.087	0.697±0.076	**0.752±0.070**

3.4.2 对抗性训练

本文最后讨论了对抗性训练的有效性. 简单的测试基准是一个没有判别器的生成器模型 (only generator, OG). 如表 7 所示, IRCGN 在所有的数据集中均取得了最佳的 AUC 值. 对于生成器来说, 采用对抗性训练, 判别器可以指导生成器重建或生成更真实的多视图数据, 更有利于进行多视图离群点检测. 总而言之, 对抗性训练可以提高 IRCGN 在多视图离群点检测上的性能.

表 7 对抗性训练的消融实验

Dataset	OG	IRCGN
M-ii	0.972±0.005	**0.978±0.004**
M-iv	0.970±0.005	**0.976±0.004**
M-vi	0.962±0.005	**0.968±0.004**
T-i	0.614±0.090	**0.728±0.070**
T-iii	0.646±0.054	**0.760±0.076**
T-v	0.651±0.068	**0.752±0.070**

4 总 结

本文提出一种新的用于高效多视图离群点检测的视图内重建和跨视图生成网络, 称为 IRCGN. 该方法使用正常数据进行对抗性训练来更好地捕捉正常数据的特征, 避免离群点造成的污染. 它是第 1 个在多视图离群点检测中使用生成对抗网络的方法. 具体来说, 本文通过视图内重建模块和跨视图生成模块来检测 3 种类型的多视图离群点. 大量的实验结果证明了所提出的 IRCGN 的优越性, 其中在高离群率数据和新增数据的检测中表现得更加出色. 但在本文工作中, 尚没有考虑训练网络模型所需要的时间和空间复杂度会随着视图数的增多而增加的问题, 如何利用多视图数据的潜空间共识表征来解决上述问题, 是下一阶段工作的重点.

References:

[1] Chandola V, Banerjee A, Kumar V. Anomaly detection: A survey. ACM Computing Surveys, 2009, 41(3): 15. [doi: 10.1145/1541880.1541882]

[2] Schlegl T, Seeböck P, Waldstein SM, Langs G, Schmidt-Erfurth U. f-AnoGAN: Fast unsupervised anomaly detection with generative adversarial networks. Medical Image Analysis, 2019, 54: 30–44. [doi: 10.1016/j.media.2019.01.010]

[3] Liu K, Li AM, Wen X, Chen HY, Yang P. Steel surface defect detection using GAN and one-class classifier. In: Proc. of the 2019 Int'l Conf. on Automation and Computing. Lancaster: IEEE, 2019. 1–6. [doi: 10.23919/IConAC.2019.8895110]

[4] Qi RB, Rasband C, Zheng J, Longoria R. Detecting cyber attacks in smart grids using semi-supervised anomaly detection and deep representation learning. Information, 2021, 12(8): 328. [doi: 10.3390/info12080328]

[5] Yang L, Yu J, Liu Y, Zhan DC. Research progress on cognitive-oriented multi-source data learning theory and algorithm. Ruan Jian Xue Bao/Journal of Software, 2017, 28(11): 2971–2991 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/5348.htm [doi: 10.13328/j.cnki.jos.005348]

[6] Liu XL, Bai L, Zhao XW, Liang JY. Incomplete multi-view clustering algorithm based on multi-order neighborhood fusion. Ruan Jian Xue Bao/Journal of Software, 2022, 33(4): 1354–1372 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/6471.htm [doi: 10.13328/j.cnki.jos.006471]

[7] Zhang YL, Yang Y, Zhou W, Ouyang XC, Hu J. CMvSC: Knowledge transferring based deep consensus network for multi-view spectral clustering. Ruan Jian Xue Bao/Journal of Software, 2022, 33(4): 1373–1389 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/6474.htm [doi: 10.13328/j.cnki.jos.006474]

[8] Li K, Li S, Ding ZM, Zhang WD, Fu Y. Latent discriminant subspace representations for multi-view outlier detection. In: Proc. of the 32nd AAAI Conf. on Artificial Intelligence, the 30th Innovative Applications of Artificial Intelligence, and the 8th AAAI Symp. on Educational Advances in Artificial Intelligence. New Orleans: AAAI, 2018. 431. [doi: 10.1609/aaai.v32i1.11826]

[9] Gao J, Fan W, Turaga D, Parthasarathy S, Han JW. A spectral framework for detecting inconsistency across multi-source object relationships. In: Proc. of the 11th IEEE Int'l Conf. on Data Mining. Vancouver: IEEE, 2011. 1050–1055. [doi: 10.1109/ICDM.2011.16]

[10] Alvarez AM, Yamada M, Kimura A, Iwata T. Clustering-based anomaly detection in multi-view data. In: Proc. of the 22nd ACM Int'l Conf. on Information & Knowledge Management. San Francisco: ACM, 2013. 1545–1548. [doi: 10.1145/2505515.2507840]

[11] Zhao HD, Fu Y. Dual-regularized multi-view outlier detection. In: Proc. of the 24th Int'l Conf. on Artificial Intelligence. Buenos Aires: AAAI Press, 2015. 4077–4083. [doi: 10.5555/2832747.2832817]

[12] Li S, Shao M, Fu Y. Multi-view low-rank analysis for outlier detection. In: Proc. of the 2015 SIAM Int'l Conf. on Data Mining. Vancouver: SIAM, 2015. 748–756. [doi: 10.1137/1.9781611974010.84]

[13] Zhao HD, Liu HF, Ding ZM, Fu Y. Consensus regularized multi-view outlier detection. IEEE Trans. on Image Processing, 2018, 27(1): 236–248. [doi: 10.1109/TIP.2017.2754942]

[14] Ji YX, Huang L, He HP, Wang CD, Xie GQ, Shi W, Lin KY. Multi-view outlier detection in deep intact space. In: Proc. of the 2019 IEEE Int'l Conf. on Data Mining. Beijing: IEEE, 2019. 1132–1137. [doi: 10.1109/ICDM.2019.00136]

[15] Sheng XR, Zhan DC, Lu S, Jiang Y. Multi-view anomaly detection: Neighborhood in locality matters. In: Proc. of the 33rd AAAI Conf. on Artificial Intelligence. Honolulu: AAAI, 2019. 4894–4901. [doi: 10.1609/aaai.v33i01.33014894]

[16] Cheng L, Wang YJ, Liu XW. Neighborhood consensus networks for unsupervised multi-view outlier detection. In: Proc. of the 35th AAAI Conf. on Artificial Intelligence. AAAI, 2021. 7099–7106. [doi: 10.1609/aaai.v35i8.16873]

[17] Schölkopf B, Williamson R, Smola A, Shawe-Taylor J, Platt J. Support vector method for novelty detection. In: Proc. of the 12th Int'l Conf. on Neural Information Processing Systems. Cambridge: MIT Press, 1999. 582–588.

[18] Zhao Y, Hryniewicki MK. XGBOD: Improving supervised outlier detection with unsupervised representation learning. In: Proc. of the 2018 Int'l Joint Conf. on Neural Networks. Rio de Janeiro: IEEE, 2018. 1–8. [doi: 10.1109/IJCNN.2018.8489605]

[19] Yang XL, Feng S, Yuan Z. Outlier detection based on reversed *k*-nearest neighborhood mst of relative distance measure. Acta Electronica Sinica, 2020, 48(5): 937–945 (in Chinese with English abstract). [doi: 10.3969/j.issn.0372-2112.2020.05.014]

[20] Liu AY, Lam DN. Using consensus clustering for multi-view anomaly detection. In: Proc. of the 2012 IEEE Symp. on Security and Privacy Workshops. San Francisco: IEEE, 2012. 117–124. [doi: 10.1109/SPW.2012.18]

[21] Pang GS, Shen CH, Cao LB, Hengel AVD. Deep learning for anomaly detection: A review. ACM Computing Surveys, 2022, 54(2): 38. [doi: 10.1145/3439950]

[22] Schlegl T, Seeböck P, Waldstein SM, Schmidt-Erfurth U, Langs G. Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. In: Proc. of the 25th Int'l Conf. on Information Processing in Medical Imaging. Boone: Springer, 2017. 146–157. [doi: 10.1007/978-3-319-59050-9_12]

[23] Zenati H, Foo CS, Lecouat B, Manek G, Chandrasekhar VR. Efffficient GAN-based anomaly detection. arXiv:1802.06222, 2018.

[24] Donahue J, Krähenbühl P, Darrell T. Adversarial feature learning. arXiv:1605.09782, 2017.

[25] Zenati H, Romain M, Foo CS, Lecouat B, Chandrasekhar V. Adversarially learned anomaly detection. In: Proc. of the 2018 IEEE Int'l

Conf. on Data Mining. Singapore: IEEE, 2018. 727–736. [doi: 10.1109/ICDM.2018.00088]

[26] Akcay S, Atapour-Abarghouei A, Breckon TP. GANomaly: Semi-supervised anomaly detection via adversarial training. In: Proc. of the 14th Asian Conf. on Computer Vision. Perth: Springer, 2019. 622–637. [doi: 10.1007/978-3-030-20893-6_39]

[27] Jiang T, Xie WY, Li YS, Lei J, Du Q. Weakly supervised discriminative learning with spectral constrained generative adversarial network for hyperspectral anomaly detection. IEEE Trans. on Neural Networks and Learning Systems, 2022, 33(11): 6504–6517. [doi: 10.1109/TNNLS.2021.3082158]

[28] Xia X, Pan XZ, Li N, He X, Ma L, Zhang XG, Ding N. GAN-based anomaly detection: A review. Neurocomputing, 2022, 493: 497–535. [doi: 10.1016/j.neucom.2021.12.093]

[29] Li XR, Ji SL, Wu CM, Liu ZG, Deng SG, Cheng P, Yang M, Kong XW. Survey on deepfakes and detection techniques. Ruan Jian Xue Bao/Journal of Software, 2021, 32(2): 496–518 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/6140.htm [doi: 10.13328/j.cnki.jos.006140]

[30] Liu FT, Ting KM, Zhou ZH. Isolation forest. In: Proc. of the 8th IEEE Int'l Conf. on Data Mining. Pisa: IEEE, 2008. 413–422. [doi: 10.1109/ICDM.2008.17]

附中文参考文献:

[5] 杨柳, 于剑, 刘烨, 詹德川. 面向认知的多源数据学习理论和算法研究进展. 软件学报, 2017, 28(11): 2971–2991. http://www.jos.org.cn/1000-9825/5348.htm [doi: 10.13328/j.cnki.jos.005348]

[6] 刘晓琳, 白亮, 赵兴旺, 梁吉业. 基于多阶近邻融合的不完整多视图聚类算法. 软件学报, 2022, 33(4): 1354–1372. http://www.jos.org.cn/1000-9825/6471.htm [doi: 10.13328/j.cnki.jos.006471]

[7] 张熠玲, 杨燕, 周威, 欧阳小草, 胡节. CMvSC: 知识迁移下的深度一致性多视图谱聚类网络. 软件学报, 2022, 33(4): 1373–1389. http://www.jos.org.cn/1000-9825/6474.htm [doi: 10.13328/j.cnki.jos.006474]

[19] 杨晓玲, 冯山, 袁钟. 基于相对距离的反 k 近邻树离群点检测. 电子学报, 2020, 48(5): 937–945. [doi: 10.3969/j.issn.0372-2112.2020.05.014]

[29] 李旭嵘, 纪守领, 吴春明, 刘振广, 邓水光, 程鹏, 杨珉, 孔祥维. 深度伪造与检测技术综述. 软件学报, 2021, 32(2): 496–518. http://www.jos.org.cn/1000-9825/6140.htm [doi: 10.13328/j.cnki.jos.006140]

郑啸(1975－), 男, 博士, 教授, CCF 高级会员, 主要研究领域为计算机网络, 工业互联网, 云计算与服务计算, 机器学习, 隐私保护.

黄俊(1985－), 男, 博士, 副教授, CCF 专业会员, 主要研究领域为机器学习, 数据挖掘.

王权鑫(1998－), 男, 硕士, 主要研究领域为深度学习, 离群点检测.