



研究与开发

基于钉板分布稀疏变分自编码器的异常检测算法研究

陈华华, 陈哲

(杭州电子科技大学通信工程学院, 浙江 杭州 310018)

摘要: 异常检测由于其广泛的应用一直是数据挖掘中一个重要的研究分支, 它有助于研究人员获得重要的信息进而对数据做出更好的决策。提出了一种基于钉板分布稀疏变分自编码器的异常检测模型。首先, 使用离散-连续混合模型钉板分布作为变分自编码器的先验, 模拟隐变量所在空间的稀疏性, 得到数据特征的稀疏表示; 其次, 以所提出的自编码器构建深度支持向量网络, 对特征空间进行压缩, 并采用最优超球体区分正常数据和异常数据; 再次, 以数据特征和超球体中心之间的欧氏距离完成异常检测; 最后, 在基准数据集 MNIST (modified national institute of standards and technology database) 和 Fashion-MNIST 上的实验评估表明, 与现存的异常检测算法相比, 本文所提出的算法具有更好的检测效果。

关键词: 异常检测; 变分自编码器; 钉板分布; 深度支持向量网络

中图分类号: TP393

文献标志码: A

doi: 10.11959/j.issn.1000-0801.2022238

Research on anomaly detection algorithm based on sparse variational autoencoder using spike and slab prior

CHEN Huahua, CHEN Zhe

School of Communication Engineering, Hangzhou Dianzi University, Hangzhou 310018, China

Abstract: Anomaly detection remains to be an essential and extensive research branch in data mining due to its widespread use in a wide range of applications. It helps researchers to obtain vital information and make better decisions about data by detecting abnormal data. Considering that sparse coding can get more powerful features and improve the performance of other tasks, an anomaly detection model based on sparse variational autoencoder was proposed. Firstly, the discrete mixed model spike and slab distribution was used as the prior of variational autoencoder, simulated the sparsity of the space where the hidden variables were located, and obtained the sparse representation of data characteristics. Secondly, combined with the deep support vector network, the feature space was compressed, and the optimal hypersphere was found to discriminate normal data and abnormal data. And then, the abnormal fraction of the data was measured by the Euclidean distance from the data feature to the center of the hypersphere, and then the abnormal detection was carried out. Finally, the algorithm was evaluated on the benchmark datasets MNIST and Fashion-MNIST, and the experimental results show that the proposed algorithm achieves better effects than the state-of-the-art methods.

Key words: anomaly detection, variational autoencoder, spike and slab distribution, deep support vector network

收稿日期: 2022-03-10; 修回日期: 2022-08-11

通信作者: 陈华华, iseealv@hdu.edu.cn

基金项目: 浙江省“领雁”研发攻关计划项目 (No.2022C03065)

Foundation Item: The “Leading Goose” Technologies Research and Development Program of Zhejiang Province (No.2022C03065)



0 引言

异常数据在统计上具有远离大多数数据点的特点^[1], 因此也被称为离群值。异常检测具有广泛的应用领域, 如互联网金融^[2]、网络入侵检测^[3]、医疗图像辅助病症诊断^[4]、工业设备运行监控^[5]、智能安防^[6]、自动驾驶^[7]等。随着数据规模和维度的增长, 研究有效的异常检测算法对众多领域而言具有重要的现实意义, 一直是国内外学者研究的焦点。

早期异常检测方法通常采用基于特征空间中数据的局部密度来确定离群点, 其中比较经典的方法是局部离群因子 (local outlier factor, LOF) 算法^[8], 该方法根据数据点与其邻域内数据点在可达距离、局部可达密度等因素计算局部离群因子从而判断数据是否异常, 这种方法的缺点是算法复杂度较高。Liu 等^[9]提出了基于集成学习的孤立森林 (isolation forest, IF) 算法, 利用二叉搜索树结构建立孤立样本, 对样本点的孤立计算异常值, 异常点由于数量少且具有与大部分样本的疏离性往往会被更早地孤立出来, 但是该方法只能检测全局异常的情况, 并且不适用于特别高维的数据。

近年来, 随着大数据技术的发展, 数据的维度越来越高, 而深度神经网络通过多层结构提取数据的特征表示, 在大规模高维数据中越来越体现出其优于传统机器学习的良好性能。得益于深度学习的广泛应用和快速发展, 越来越多的研究者利用深度网络进行异常检测。Erfani 等^[10]提出了一种混合模型, 将无监督深度置信网络 (deep belief network, DBN) 训练为降维算法, 将数据转换为低维特征集, 并从 DBN 学到的特征中训练出单类支持向量机 (one-class support vector machine, OC-SVM)。但是 DBN 采用贪婪算法逐层训练参数, 可能导致解非全局最优, 于是研究者开始转向采用自编码器和生成对抗网络等深度网

络应用于异常检测。Li 等^[11]利用胶囊网络构建自编码器和分类器网络, 根据预测概率和重建误差检测异常。Zimmerer 等^[12]提出了一种基于变分自编码器的异常检测方法, 并使用模型内潜在表示偏差和重建误差进行异常检测。Inderjeet 将 6 层自编码网络的第 2、3 层改用胶囊网络结构对数据进行编码, 利用重建误差检测异常^[13]。编码器可以将复杂的高维数据映射到低维, 但是其潜在空间往往会产生许多不携带任何信息的潜在维度, 导致获取捕捉观测数据复杂性的编码产生问题^[14]。Schlegl 等^[15]提出一种深度卷积生成对抗网络获取医学图像的特征, 并使用一种基于从图像空间到潜在空间映射的异常评分方案检测图像异常。但是, 有学者指出在有较少异常数量的情况下深度生成模型的表现没有 K 近邻等传统方法表现出色^[16]。

目前大多数的异常检测算法通过深度神经网络从数据中提取特征, 然后使用重建误差或概率等启发式方法检测异常。因此, 设计从大规模数据中提取判别特征的深度网络是至关重要的。考虑到稀疏编码可以提高特征表示的质量, 本文提出了一种基于稀疏变分自编码器的异常检测模型, 该模型优化了钉板分布 (spike and slab distribution) 先验。钉板分布可以灵活地模拟^[17]稀疏分布阵列, 使它们适合于许多类型的数据, 因为它由一个二进制分量和一个连续分量组成, 通过将两个分量相乘产生一个随机变量, 其值恰好为零或从控制连续分量的分布中得出。本文提出使用变分自编码器的钉板分布先验模拟数据潜在的存在或缺失, 该模型将深度支持向量网络与钉板先验相结合, 为异常检测提供了一个框架。

1 异常检测模型

1.1 钉板分布

钉板分布^[17]是一种统计学领域上的离散混合分布, 主要针对数据是否为零进行建模, 其概率

模型具有稀疏性, 因此常被用于贝叶斯稀疏推断模型^[18]。钉板分布通过一个二元指示变量 s 对变量 x 进行建模, 二元指示变量 $s \in \{0, 1\}$ 是变量是否等于零的标志, 其分布符合参数为 α 的伯努利分布, 如式 (1) 所示。

$$f(s) = \text{Ber}(s | \alpha) = \alpha^s (1 - \alpha)^{1-s} \quad (1)$$

变量 x 依赖于二元指示变量 s 的值, $s = 0$ 表示变量 x 为零, 此时变量的概率密度函数为以零为中心的狄拉克函数; $s = 1$ 表示变量 x 不为零, 此时变量的概率密度函数为标准高斯分布, 如式 (2) 所示。

$$f(x|s) = \begin{cases} \mathcal{N}(x|0, 1), & s = 1 \\ \delta(x), & s = 0 \end{cases} \quad (2)$$

其中, $\mathcal{N}(x|0, 1)$ 表示均值为 0、方差为 1 的标准高斯分布的概率密度函数, 根据式 (1) 和式 (2), 二元指示变量 s 和变量 x 的联合概率分布如式 (3) 所示。

$$f(x, s) = f(x|s)f(s) = \alpha^s (1 - \alpha)^{1-s} \mathcal{N}(x|0, 1)^s \delta(x)^{1-s} \quad (3)$$

通过对 s 的积分, 可以得到变量 x 的先验概率, 如式 (4) 所示。

$$f(x) = \sum_s f(x, s) = f(x, s = 0) + f(x, s = 1) = (1 - \alpha)\delta(x) + \alpha\mathcal{N}(x|0, 1) \quad (4)$$

可以看到钉板分布实质上是一个混合分布, 第一个分量其概率密度函数相对尖锐, 形状与钉子相似, 第二个分量其概率密度函数相对扁平, 形状与板相似。由式 (4) 可知, 如果 α 值很小, 意味着钉板分布的稀疏度很高, 此时 x 很大概率为 0。钉板分布通过二元指示变量 s 直接对变量 x 进行建模, 更有利于层次化建模, 从而使得建模更加灵活。

1.2 稀疏变分自编码器

变分自编码器^[19]采用变分推理解决模型优化

问题, 通过将自编码网络与概率统计结合, 学习数据的特征分布, 是目前深度学习领域的一种重要生成模型。变分自编码器通过优化目标函数的下界来达到训练模型的目的, 它的目标函数变分下界为:

$$\log p_\theta(\mathbf{x}) = D_{\text{KL}}[q_\phi(\mathbf{z} | \mathbf{x}) \| p_\theta(\mathbf{z} | \mathbf{x})] + L(\theta, \phi; \mathbf{x}) \geq L(\theta, \phi; \mathbf{x}) \quad (5)$$

式 (5) 可进一步改写为:

$$L(\theta, \phi; \mathbf{x}) = -D_{\text{KL}}[q_\phi(\mathbf{z} | \mathbf{x}) \| p_\theta(\mathbf{z})] + \mathbb{E}_{q_\phi(\mathbf{z} | \mathbf{x})}[\log p_\theta(\mathbf{x} | \mathbf{z})] \quad (6)$$

变分下界由两项组成: 第一项是先验项, 它鼓励最小化编码分布和先验之间的 KL 差异; 第二项是重建项, 它在识别函数下最大化对数据可能性的期望。

本文引入钉板分布先验以增加隐变量的稀疏性, 可以得到隐变量 $\mathbf{z}$ 的先验概率密度:

$$p_\theta(\mathbf{z}) = \prod_{j=1}^J (\alpha \mathcal{N}(z_j; 0, 1) + (1 - \alpha)\delta(z_j)) \quad (7)$$

其中, $\delta(\cdot)$ 表示以零为中心的狄拉克 δ 函数, z_j 是隐变量 $\mathbf{z}$ 的第 j 个元素, J 是隐变量 $\mathbf{z}$ 的维数。其近似后验 $q_\phi(\mathbf{z} | \mathbf{x})$ 定义为:

$$q_\phi(\mathbf{z} | \mathbf{x}) = \prod_{j=1}^J (\gamma_j \mathcal{N}(z_j; \mu_j, \sigma_j^2) + (1 - \gamma_j)\delta(z_j)) \quad (8)$$

其中, μ_j 、 σ_j^2 、 γ_j 是输入数据为 $\mathbf{x}$ 、权重参数为 ϕ 的神经网络输出, 下标 j 代表向量的第 j 个元素。近似后验分布的形式与式 (7) 的钉板先验分布类似, 这样选择的近似后验分布允许后验分布与先验分布相匹配, 可以帮助模型更好地编码隐变量所在空间的信息。

与标准 VAE (variational auto-encoder) 相同, 通过最大化变分下界来最大化样本的边缘概率分布, 经过推导得到稀疏变分自编码器的变分下界为:



$$L(\theta, \phi; \mathbf{x}) \approx \sum_{j=1}^J \left[\frac{\gamma_j}{2} (1 + \log(\delta_j^2) - u_j^2 - \delta_j^2) + (1 - \gamma_j) \log \left(\frac{1 - \alpha}{1 - \gamma_j} \right) + \gamma_j \log \left(\frac{\alpha}{\gamma_j} \right) \right] + \frac{1}{L} \sum_{l=1}^L \log p_{\theta}(\mathbf{x} | \mathbf{z}_l) \quad (9)$$

1.3 深度支持向量网络

支持向量数据描述 (support vector data description, SVDD) 是一种单值分类方法^[20]。异常检测问题本质也是一种单值分类问题, 即将正常数据看作目标类数据, 各种异常数据均看作离群类数据, 因此可以使用单值分类的思想解决异常检测问题。SVDD 方法旨在找到一个包含大多数目标类数据、中心点为 $\mathbf{c} \in \mathcal{F}_k$ 、半径为 $R > 0$ 的最小超球体, 其中 $\mathcal{F}_k$ 表示特征空间, R 为超球体半径。

设输入数据集为 $\{\mathbf{x}^{(i)}\}, i=1, 2, \dots, N$, SVDD 在特征空间上 $\mathcal{F}_k$ 的最小超球体本质是求解目标函数:

$$\begin{aligned} \min_{R, \xi, \mathbf{c}} \quad & R^2 + \frac{1}{\nu N} \sum_i \xi^{(i)} \\ \text{s.t.} \quad & \|\phi(\mathbf{x}^{(i)}) - \mathbf{c}\|^2 \leq R^2 + \xi^{(i)}, \quad \xi^{(i)} > 0, i=1, 2, \dots, N \end{aligned} \quad (10)$$

其中, $\xi^{(i)}$ 是松弛变量, ν 是惩罚参数, $\nu \in (0, 1)$, $\phi(\mathbf{x}^{(i)})$ 是输入 $\mathbf{x}^{(i)}$ 在特征空间上的映射, $\|\phi(\mathbf{x}^{(i)}) - \mathbf{c}\|^2$ 是输入 $\mathbf{x}^{(i)}$ 的映射值与超球体中心 $\mathbf{c}$ 之间的欧氏距离。

受支持向量数据描述的启发, 本文通过训练一个神经网络把网络输出拟合到最优超球体中, 由于网络把输出紧密地映射到球体中, 因此它能提取数据特征分布变化的共同因素。本文采用的深度支持向量网络目标函数为:

$$\min_{\phi} \frac{1}{N} \sum_i \|\phi(\mathbf{x}^{(i)}; \phi) - \mathbf{c}\|^2 + \lambda \sum_{l=1}^L \|\phi^{(l)}\|^2 \quad (11)$$

其中, ϕ 表示层数为 L 的神经网络的参数, $\phi(\mathbf{x}^{(i)}; \phi)$ 表示输入为 $\mathbf{x}^{(i)}$ 时的神经网络输出, $\phi^{(l)}$ 表示第 l 层神经网络的参数, λ 是正则化系数, 第二项用以防止神经网络出现过拟合从而使网络具

有较好的泛化能力。

当网络训练完成以后, 设测试数据 $\mathbf{x}'$, 获得其对应的神经网络输出 $\phi(\mathbf{x}'; \phi)$ 。这里定义测试数据 $\mathbf{x}'$ 的异常分数 $s(\mathbf{x}')$ 为 $\phi(\mathbf{x}'; \phi)$ 到超球体中心 $\mathbf{c}$ 的距离, 如式 (12) 所示。

$$s(\mathbf{x}') = \|\phi(\mathbf{x}'; \phi) - \mathbf{c}\|^2 \quad (12)$$

1.4 网络模型

本文提出的基于钉板分布稀疏变分自编码器的异常检测模型网络结构如图 1 所示。

模型主要由两部分组成: 稀疏变分自编码网络和深度支持向量数据描述网络。稀疏变分自编码器使用深度神经网络作为特征提取网络得到隐变量的均值 μ 、方差 σ^2 以及隐变量非零的概率 γ , 通过重采样得到隐变量 $\mathbf{z}$, 解码得到重建图像。稀疏变分自编码器通过采用式 (9) 为代价函数训练获得。将自编码器得到的隐变量均值 $\mathbf{c}$ 作为超球体中心, 构建深度支持向量网络并进行异常检测。深度支持向量网络通过采用式 (11) 为代价函数训练获得。在模型训练完后, 测试数据经稀疏变分自编码器的编码器输入深度支持向量网络, 结合均值 $\mathbf{c}$ 对输入数据进行异常检测。

1.5 异常检测

在训练阶段, 以钉板分布为先验构建并训练稀疏变分自编码器, 由获得的自编码器提取正常类别数据的特征并求取特征的均值 $\mathbf{c}$ 。以均值 $\mathbf{c}$ 为超球体中心, 结合稀疏变分自编码器构建和训练深度支持向量数据描述网络。在测试阶段, 测试数据 $\mathbf{x}'$ 先进入深度支持向量数据描述网络, 然后计算式 (12) 所示异常分数。根据式 (13) 检测数据是否为异常, 当数据的异常得分小于阈值时, 则该数据位于超球体之内, 判断为正常数据; 当数据的异常得分大于阈值时, 则该数据位于超球体之外, 判断为异常数据。

$$S(\mathbf{x}') = \begin{cases} \text{正常} & , s(\mathbf{x}') \leq \text{阈值} \\ \text{异常} & , s(\mathbf{x}') > \text{阈值} \end{cases} \quad (13)$$

2 实验结果与分析

2.1 实验配置及数据集

为了验证方法的有效性,对方法进行了验证。采用的计算机配置和编程环境如下: Intel Core i5-7300HQ@2.50GHz CPU, 内存为 8 GB, Windows10 系统下的 python3.6 编译器, 编程环境为 TensorFlow 和 Keras, 编程语言为 Python。

本文使用深度神经网络的常见参数^[21]搭建稀疏变分自编码器, 由 5 个全连接层组成, 网络结构如图 1 所示。其中编码网络的网络维度是 D-500-500-2000-H, 解码网络是编码网络的镜像, 网络维度是 H-2000-500-500-D, D、H、500、2 000 均是全连接层的节点数, 其中 D 是输入维数, 本文中 $D=784$ 。

本文中取 $H=32$ 、 $\alpha=0.2$ (该取值的获取过程见第 2.3 节实验部分)。深度支持向量数据描

述网络正则化系数 $\lambda=0.1$ 。模型参数采用正态分布随机值初始化, batch size 设为 100。模型参数更新方法采用 Adam 优化器, 学习率设置为 0.000 1, 采用指数衰减, 每 20 000 步衰减为原来的 0.9。

本文使用 MNIST (modified national institute of standards and technology database) 数据集^[22]和 Fashion-MNIST 数据集^[23]进行实验。MNIST 数据集包含 10 种不同数字(0~9)图像, 如果指定其中 1 种为正常数据, 则余下 9 种为异常数据。Fashion-MNIST 数据集包含 10 种不同时尚商品图像, 如果指定其中 1 种为正常数据, 则余下 9 种为异常数据。如果两个数据集均采用正常类别的数据作为训练数据, 各有约 6 000 个样本, 它们的测试集各有 10 000 个样本, 包括正常数据和异常数据。实验中图像像素值均归一化到 [0,1] 内。

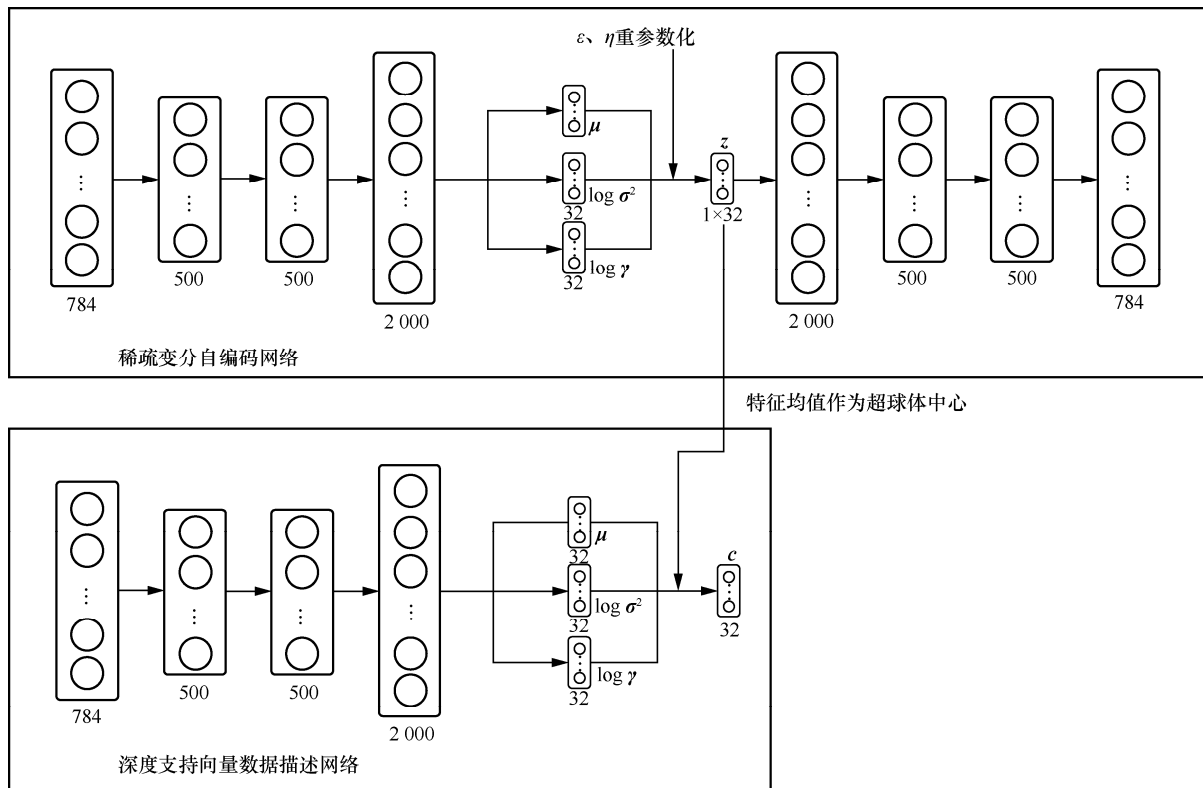


图1 异常检测模型网络结构



2.2 评价标准

本文使用曲线下面积 (area under curve, AUC) 评估异常检测模型性能。AUC 是受试者工作特征 (receiver operating characteristic, ROC) 曲线下的面积。ROC 由混淆矩阵得到, 二分类的混淆矩阵见表 1。其中, 真阳性 (true positive, TP) 表示真值和预测值均为正常的样例; 假阴性 (false negative, FN) 表示预测值为异常的正常样例; 假阳性 (false positive, FP) 表示预测值为正常的异常样例; 真阴性 (true negative, TN) 表示真值与预测值均为异常的样例。

表 1 二分类的混淆矩阵

真值	预测值	
	True	False
True	TP	FN
False	FP	TN

根据混淆矩阵以及式 (14)、式 (15) 计算出伪阳率 (false positive rate, FPR) 和真阳率 (true positive rate, TPR)。改变式 (13) 中的阈值, 可以得到模型在不同阈值下的 FPR 和 TPR 值, 将伪阳率作为横坐标, 真阳率作为纵坐标, 即得 ROC 曲线。考虑到正常、异常样本数量存在分布不均的情况, 不适宜以准确率作为指标评估模型, 而 AUC 兼顾了模型对正常样本和异常样本的分类能力, 因此本文采用 AUC 评估异常检测模型。

$$TPR = \frac{TP}{TP + FN} \quad (14)$$

$$FPR = \frac{FP}{FP + TN} \quad (15)$$

2.3 消融实验

(1) 编码器编码和解码器解码实验

为了验证本文算法所得数据特征的稀疏性, 本文搭建了稀疏变分自编码器, 并使用 MNIST 数

据集和 Fashion-MNIST 数据集进行训练。MNIST 数据集部分结果如图 2 所示, Fashion-MNIST 数据集部分结果如图 3 所示, 可以看出, 本文提出的稀疏变分自编码器能够对数据稀疏编码, 实现了从图像数据到稀疏隐变量 $\mathbf{Z}$ 的映射; 解码器对数据进行较好地重建, 实现从稀疏隐变量 $\mathbf{Z}$ 到图像的重建。

(2) 隐变量 $\mathbf{Z}$ 的维度对重构图像的影响

稀疏自编码器不仅将复杂的高维数据映射到低维, 同时隐变量 $\mathbf{Z}$ 的稀疏性对维度是否携带信息进行了区分。不为 0 的维度携带了图像中的特征信息, 为 0 的维度则不携带特征信息。改变隐变量 $\mathbf{Z}$ 中的非零元素, 重建图像表现出不同的特性。例如, 隐变量 $\mathbf{Z}$ 中非零维度值的变化对重建图像的影响如图 4 所示, 图 4 (a) 重建图像数字 3 的字体粗细发生改变, 说明箭头标记的维度携带了字体粗细这一特性信息; 图 4 (b) 重建图像从平底鞋逐渐变为高跟鞋, 说明箭头标记的维度携带了鞋底的形状信息; 图 4 (c) 重建图像从长袖逐渐变为短袖, 说明箭头标记的维度携带了袖子的长度信息。

(3) 隐藏层节点数 H 和稀疏度 α 对检测性能的影响

隐藏层节点数 H 和稀疏度 α 是网络的两个主要参数, 其中 H 也是编码模块中稀疏编码 $\mathbf{z}$ 的维数。隐藏层节点数 H 和稀疏度 α 对网络的检测性能影响比较大。为了选取最优的隐藏层节点数 H 以及钉板分布中参数 α 的组合, 本文分别对隐藏层节点数 H 以及参数 α 取值的不同组合在 MNIST 数据集上进行了实验。实验以 AUC 值作为评价指标, 不同实验参数组合 (H, α) 下的实验结果如图 5 所示, 表示了在不同的稀疏度和不同的隐层节点数 H 下的 AUC 值变化情况。可以看出, 当隐藏层节点数 $H=32$, 参数 $\alpha=0.2$ 时, AUC 值最大, 异常检测效果较好, 因此本文取 $H=32$ 、 $\alpha=0.2$ 。

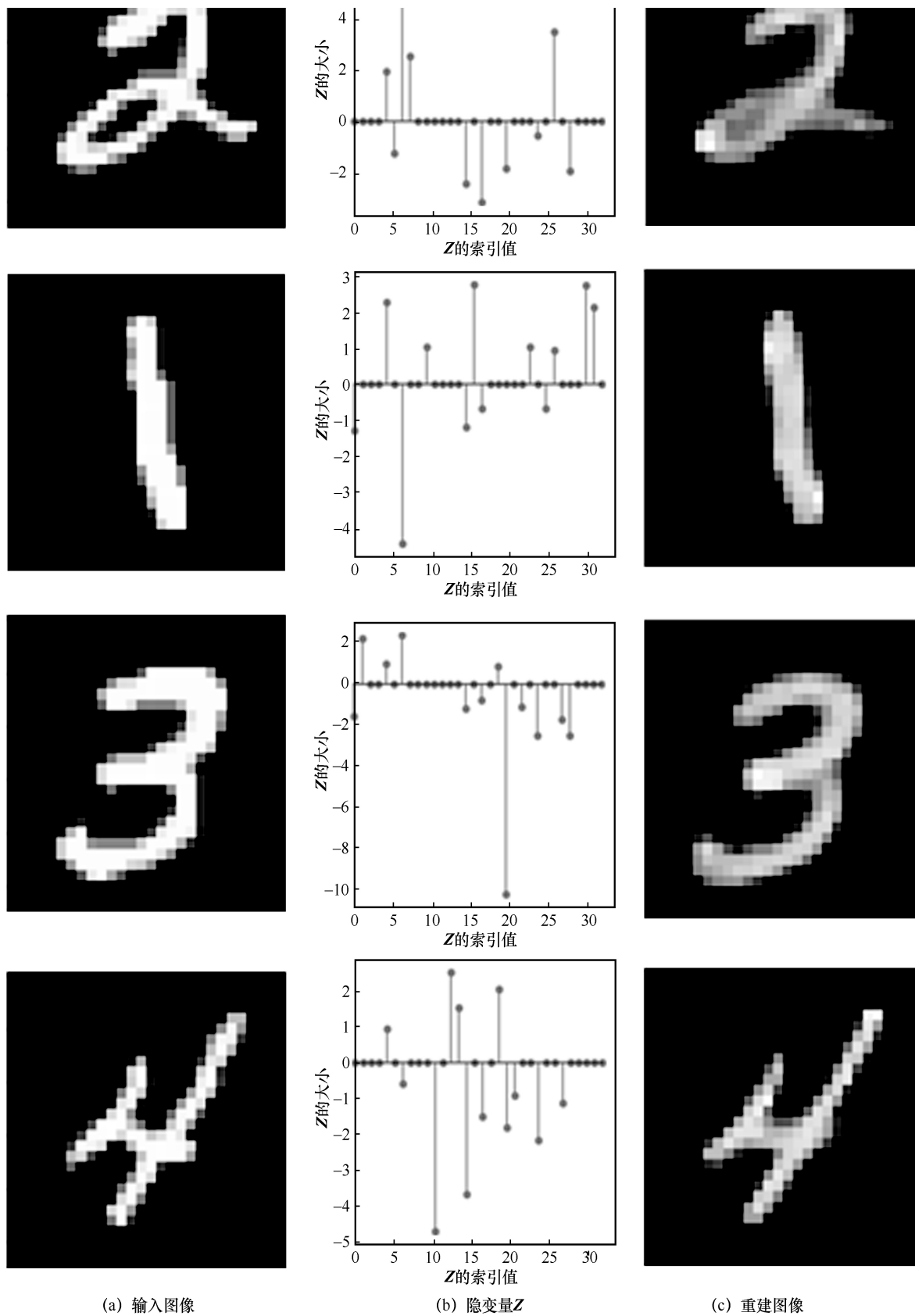


图2 MNIST 数据集部分结果

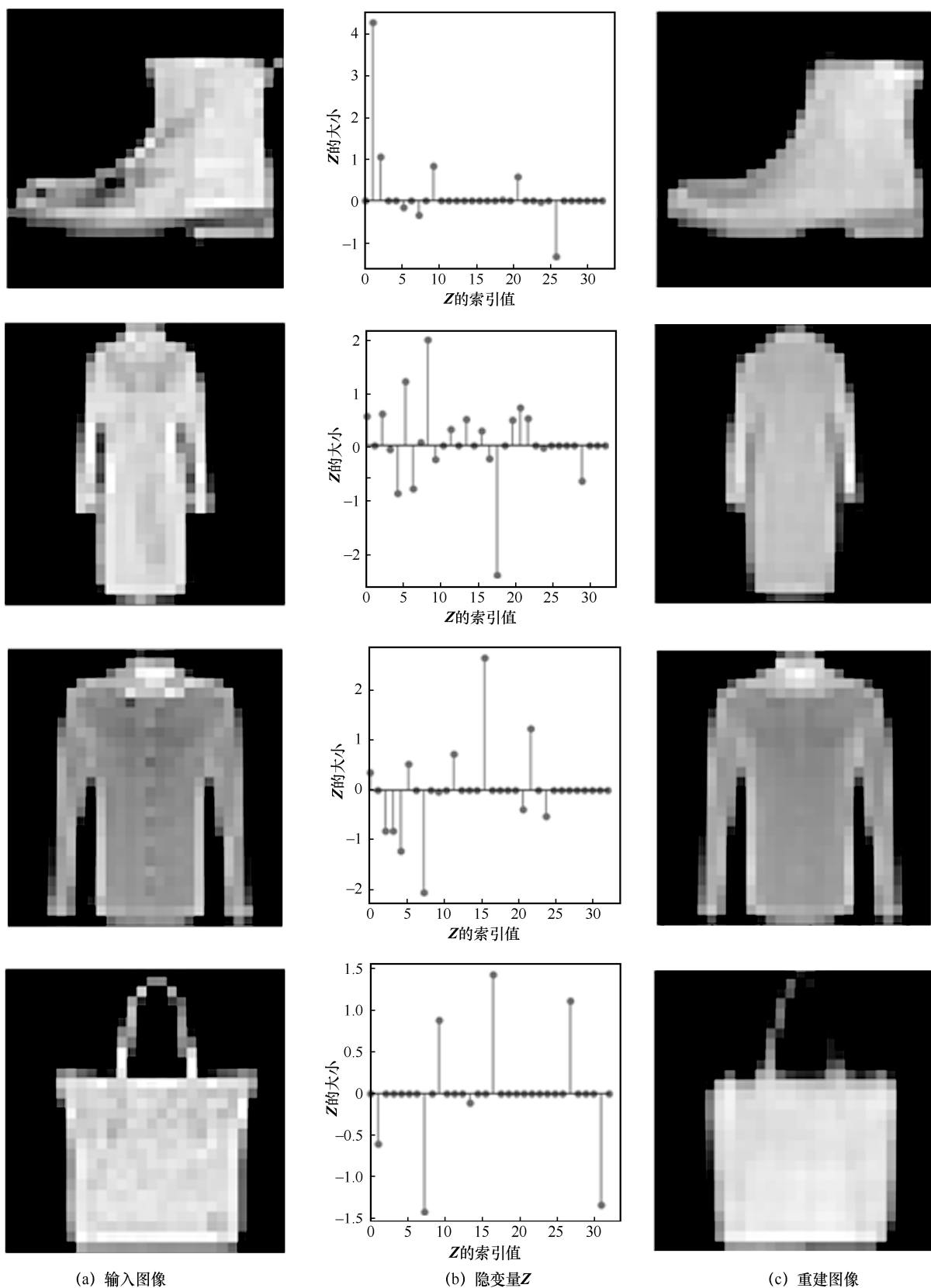
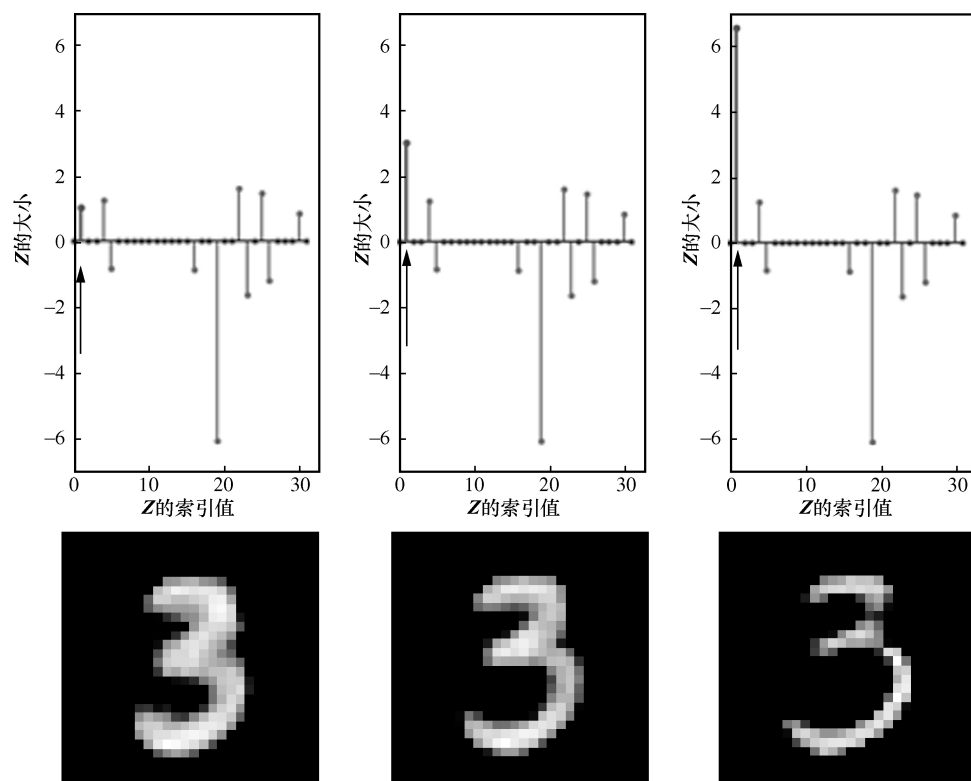
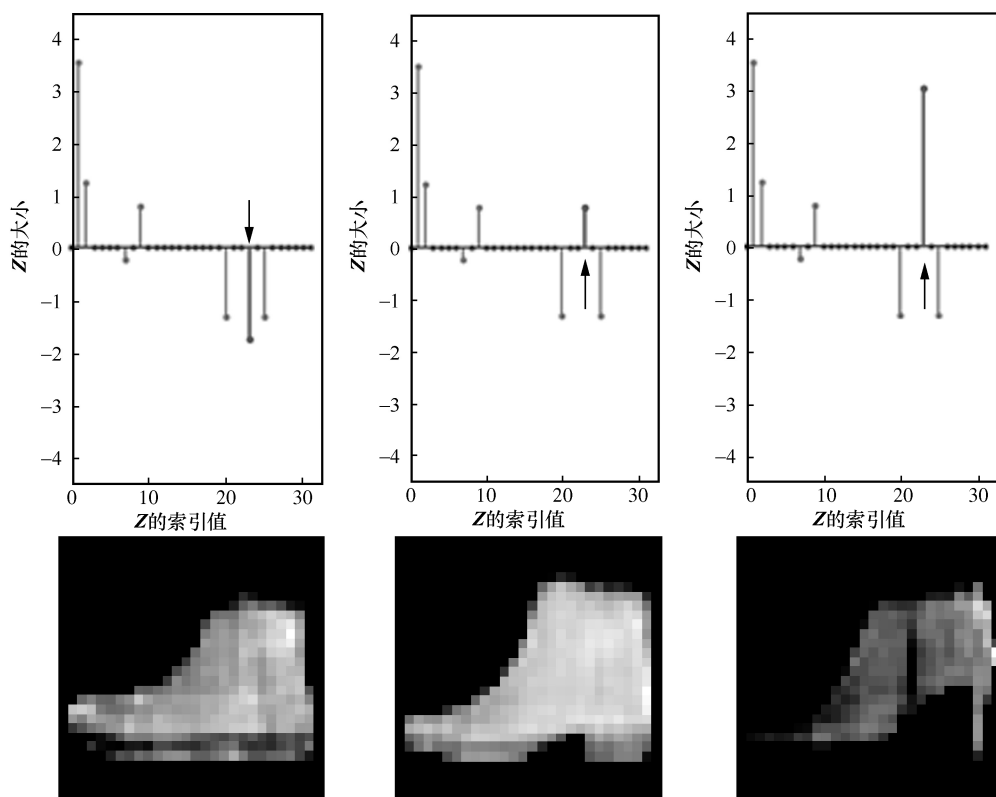


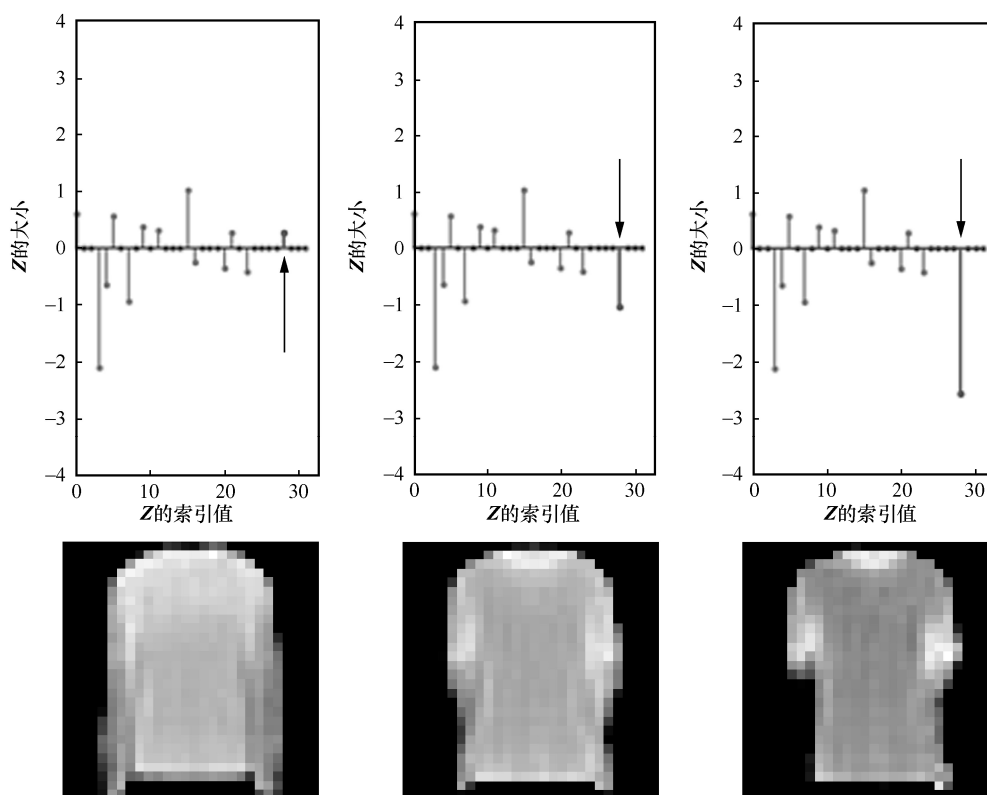
图3 Fashion-MNIST 数据集部分结果



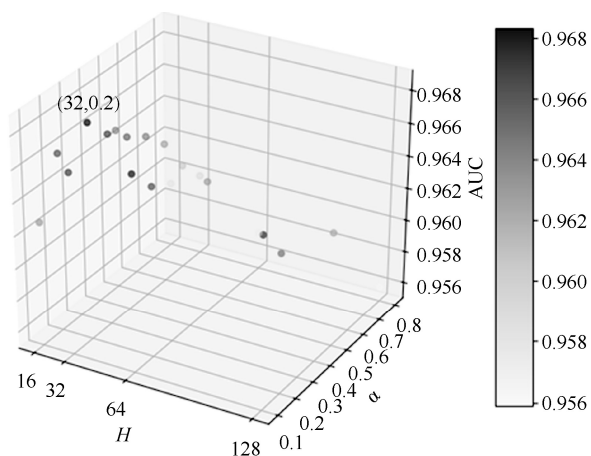
(a) 数字3



(b) 鞋底



(c) 衣服袖子

图4 隐变量 Z 中非零维度值的变化对重建图像的影响图5 不同实验参数组合(H, α)下的实验结果

2.4 异常检测实验

MNIST 数据集在各种异常检测方法下的实验结果对比见表 2, 展示了本文算法与 OC-SVM^[24]、IF^[9]、DCAE (deep convolutional auto-encoder)^[25]、ANOGAN (anomaly detection with generative adversarial network)^[15]等算法在 MNIST 数据集上的

实验结果, 加粗显示代表效果优于其他算法。可以看出, 本文算法的平均 AUC 值较其他优秀的异常检测算法有显著的提高, 表明本文算法在 MNIST 数据集中实现了较好的检测效果, 能够进行有效的异常检测。

Fashion-MNIST 数据集在各种异常检测方法下的实验结果对比见表 3。可以看出, 本文算法的平均 AUC 值较其他优秀的异常检测算法有显著的提高, 这说明本文算法在 Fashion-MNIST 数据集中实现了较好的检测效果, 能够进行有效的异常检测。

表 2 和表 3 的结果是模型由所有训练集样本参与训练、所有测试样本参与测试的结果。本文参照文献[26]的做法, 由训练集中 80% 的正常样本作为训练样本, 剩余的 20% 正常样本参与测试, 并且从测试集中随机抽取异类测试样本, 使其构成新的测试集的一半。在 MNIST 和

表 2 MNIST 数据集在各种异常检测方法下的实验结果对比

NORMAL CLASS	OC-SVM ^[24]	IF ^[9]	DCAE ^[25]	ANOGAN ^[15]	本文算法
0	98.6%±0.0%	98.0%±0.0%	97.6%±0.7%	96.6%±1.3%	98.1%±0.7%
1	99.5%±0.0%	97.3%±0.0%	98.3%±0.6%	99.2%±0.6%	99.6%±0.1%
2	82.5%±0.1%	88.6%±0.0%	85.4%±2.4%	85.0%±2.9%	96.5%±1.1%
3	88.1%±0.0%	89.9%±0.0%	86.7%±0.9%	88.7%±2.1%	95.4%±0.8%
4	94.9%±0.0%	92.7%±0.0%	86.5%±2.0%	89.4%±1.3%	97.6%±0.7%
5	77.1%±0.0%	85.5%±0.0%	78.2%±2.7%	88.3%±2.9%	94.6%±0.9%
6	96.5%±0.0%	95.6%±0.0%	94.6%±0.5%	94.7%±2.7%	97.7%±0.5%
7	93.7%±0.0%	92.0%±0.0%	92.3%±1.0%	94.5%±1.8%	97.0%±0.6%
8	88.9%±0.0%	89.9%±0.0%	86.5%±1.6%	84.9%±2.1%	95.3%±0.8%
9	94.1%±0.0%	94.5%±0.0%	90.4%±1.8%	92.4%±1.1%	96.3%±0.6%
平均值	91.3%±0.0%	92.3%±0.0%	89.7%±1.4%	91.3%±1.9%	96.8%±0.7%

表 3 Fashion-MNIST 数据集在各种异常检测方法下的实验结果对比

NORMAL CLASS	OC-SVM ^[24]	IF ^[9]	DCAE ^[25]	ANOGAN ^[15]	本文算法
0 (T-shirt)	91.4%±0.0%	91.3%±0.0%	95.2%±0.4%	95.8%±0.5%	95.7%±0.4%
1 (Trousers)	98.2%±0.0%	97.5%±0.0%	98.3%±0.3%	98.5%±0.6%	99.0%±0.1%
2 (Pullover)	88.2%±0.1%	87.4%±0.0%	84.3%±0.7%	88.8%±0.9%	91.2%±0.5%
3 (Dress)	91.6%±0.0%	92.2%±0.0%	89.7%±0.5%	88.7%±2.1%	96.2%±0.7%
4 (Coat)	89.3%±0.0%	90.8%±0.0%	86.5%±1.3%	90.4%±0.8%	92.7%±0.3%
5 (Sandal)	89.1%±0.0%	92.6%±0.0%	82.6%±0.7%	90.7%±0.6%	93.1%±1.6%
6 (Shirt)	84.5%±0.1%	80.0%±0.0%	79.3%±0.4%	84.3%±1.4%	84.8%±0.9%
7 (Sneaker)	98.5%±0.0%	98.3%±0.0%	92.3%±0.6%	94.5%±0.8%	98.6%±0.5%
8 (Bag)	89.5%±0.0%	88.4%±0.0%	91.7%±0.9%	94.1%±1.2%	97.9%±0.7%
9 (Ankle boot)	97.7%±0.0%	97.6%±0.0%	91.2%±0.5%	92.6%±0.7%	98.8%±0.4%
平均值	91.8%±0.0%	91.6%±0.0%	89.1%±0.6%	91.6%±1.1%	94.8%±0.6%

Fashion-MNIST 数据集上和 GPND (generative probabilistic novelty detection)^[26]、ALOCC DR (adversarially learned one-class classifier for discrimination and reconstruction)^[27]、ALOCC D (adversarially learned one-class classifier for discrimination)^[27]、AEC (auto-encoder)^[28]和 OCGAN (one-class generative adversarial network)^[29]方法进行了比较, 80%正常样本作为训练样本的模型实验结果对比见表 4。可以看出, 本文算法具有更佳检测效果。

本文还参照文献[30]的做法, 在 Fashion-MNIST 中, 改变异常样本在测试集中的不同占比, 与 GPND 和 GradCon (gradient constraint)^[30]进行了性能比较, Fashion-MNIST 集中异常样本在测

试集不同占比下的实验结果对比见表 5。可以看出, 测试集在不同异常样本的占比下, 本文方法具有更好的检测效果。

表 4 80%正常样本作为训练的模型实验结果对比

对比项	MNIST	Fashion-MNIST
ALOCC DR ^[27]	88.0%	75.3%
ALOCC D ^[27]	82.0%	60.1%
AEC ^[28]	89.9%	90.8%
GPND ^[26]	93.2%	90.1%
OCGAN ^[29]	97.7%	92.4%
本文算法	98.3%	95.2%

由表 2~表 5 知, 本文算法具有较好的异常检测效果, 并优于当前的一些异常检测算法。



表5 Fashion-MNIST 集中异常样本在测试集不同占比下的实验结果对比

测试集	10%	20%	30%	40%	50%
GPND ^[26]	92.8%	93.2%	93.3%	93.3%	93.3%
GradCon ^[30]	93.8%	93.3%	93.5%	93.6%	93.4%
本文算法	95.1%	95.3%	95.3%	95.3%	95.2%

3 结束语

本文提出了一种基于稀疏变分自编码器和深度支持向量网络的异常检测算法,使用钉板分布作为变分自编码器的先验,模拟隐变量所在空间的稀疏性,得到数据特征的稀疏表示,并利用稀疏变分自编码器构建深度支持向量网络对特征空间进行压缩,以最优超球体区分正常数据与异常数据。实验结果表明本文算法优于其他优秀方法,实现了较好的异常检测效果。未来的工作可以设计更加高效的网络结构,进一步地提升特征提取能力以提高异常检测的效果,从而适用于更多复杂场景下的异常检测。

参考文献:

- [1] HAWKINS D M. Identification of outliers[M]. London: Chapman and Hall, 1980.
- [2] SHENH D, KURSUN E. Label augmentation via time-based knowledge distillation for financial anomaly detection[EB]. 2021.
- [3] TUFAN E, TEZCAN C, ACARTÜRK C. Anomaly-based intrusion detection by machine learning: a case study on probing attacks to an institutional network[J]. IEEE Access, 9: 50078-50092.
- [4] TSCHUCHNIGM E, GADERMAYR M. Anomaly detection in medical imaging: a mini review[EB]. 2021.
- [5] CUI Y J, LIU Z X, LIAN S G. A survey on unsupervised industrial anomaly detection algorithms[EB]. 2022.
- [6] PATRIKARD R, PARATE M. Anomaly detection using edge computing in video surveillance system: review[J]. International Journal of Multimedia Information Retrieval, 2021, 11: 85-110.
- [7] BOGDOLL D, NITSCHKE M, ZÖLLNER J M. Anomaly detection in autonomous driving: a survey[EB]. 2022.
- [8] JÖRG SANDER. LOF: identifying density-based local outliers[J]. Acm Sigmod Record, 2000, 29(2):93-104.
- [9] LIU F T, TING K M, ZHOU Z H. Isolation forest[C]//Proceedings of 2008 Eighth IEEE International Conference on Data Mining. Piscataway: IEEE Press, 2008: 413-422.
- [10] ERFANI S M, RAJASEGARAR S, KARUNASEKERA S, et al. High-dimensional and large-scale anomaly detection using a linear one-class SVM with deep learning[J]. Pattern Recognition, 2016, 58: 121-134.
- [11] LI X Y, KIRINGA I, YEAP T, et al. Exploring deep anomaly detection methods based on capsule net[C]//Advances in Artificial Intelligence. Switzerland: Springer, 2020: 375-387.
- [12] ZIMMERER D, KOHL S A A, PETERSEN J, et al. Context-encoding variational autoencoder for unsupervised anomaly detection[EB]. 2018.
- [13] SINGH I, HEMACHANDRA N. Anomaly detection using capsule networks for high-dimensional datasets[EB]. 2021.
- [14] TONOLINI F, JENSEN B S, MURRAY S R. Variational sparse coding[EB]. 2019.
- [15] SCHLEGL T, SEEBÖCK P, WALDSTEIN S M, et al. Unsupervised anomaly detection with generative adversarial networks to guide marker discovery[C]//Information Processing in Medical Imaging. Switzerland: Springer, 2017: 146-157.
- [16] ŠKVÁRA V, PEVNÝ T, ŠMÍDL V. Are generative deep models for novelty detection truly better? [EB]. 2018.
- [17] SHEIKH A S, SHELTON J A, LÜCKE J. A truncated EM approach for spike-and-slab sparse coding[J]. Journal of Machine Learning Research, 2012, 15: 2653-2687.
- [18] COURVILLE A, BERGSTRA J, BENGIO Y. Unsupervised models of images by spike-and-slab RBMs[EB]. 2011.
- [19] KINGMA D P, WELING M. Auto-encoding variational Bayes[EB]. 2013.
- [20] TAX D M J, DUIN R P W. Support vector data description[J]. Machine Learning, 2004, 54(1): 45-66.
- [21] VAN DER MAATEN L. Learning a parametric embedding by preserving local structure[J]. Journal of Machine Learning Research, 2009, 5: 384-391.
- [22] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.
- [23] XIAO H, RASUL K, VOLLGRAF R. Fashion-MNIST: a novel image dataset for benchmarking machine learning algorithms[EB]. 2017.
- [24] MANEVITZ L M, YOUSEF M. One-class SVMs for document classification[J]. Journal of Machine Learning Research, 2001, 2(12): 139-154.
- [25] MAKHZANI A, FREY B J. Winner-take-all autoencoders[C]//Proceedings of Advances in Neural Information Processing Systems(NIPS). Montreal: Neural Information Processing Sys-

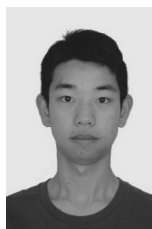
- tems Foundation, Inc. 2015: 2791-2799.
- [26] PIDHORSKYI S, ALMOHSEN R, ADJEROH D A, et al. Generative probabilistic novelty detection with adversarial autoencoders[C]//Proceedings of Advances in Neural Information Processing Systems(NIPS). Montreal: Neural Information Processing Systems Foundation, Inc. 2018: 6822-6833.
- [27] SABOKROU M, KHALOOEI M, FATHY M, et al. Adversarially learned one-class classifier for novelty detection[C]//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 3379-3388.
- [28] SAKURADA M, YAIRI T. Anomaly detection using autoencoders with nonlinear dimensionality reduction[C]//MLSDA'14: Proceedings of the MLSDA 2014 2nd Workshop on Machine Learning for Sensory Data Analysis. 2014: 4-11.
- [29] PERERA P, NALLAPATI R, XIANG B. OCGAN: one-class novelty detection using GANs with constrained latent representations[C]//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2019: 2893-2901.

- [30] KWON G, PRABHUSHANKAR M, TEMEL D, et al. Back-propagated gradient representations for anomaly detection[C]//Proceedings of the European Conference on Computer Vision (ECCV). Heidelberg: Springer, 2020: 206-226.

[作者简介]



陈华华（1975—），男，博士，杭州电子科技大学副教授，主要研究方向为数字图像处理、计算机视觉和模式识别。



陈哲（1995—），男，杭州电子科技大学硕士生，主要研究方向为异常检测。