arXiv:2304.00709v3 [cs.LG] 9 Jul 2024

MSS-PAE：拯救基于自编码器的离群点检测，避免意外重构

Xu Tan[a]，Jiawei Yang[b]，Junqi Chen[a]，Sylwan Rahardja[c]，Susanto Rahardja[a,d,*]

[a]*School of Marine Science and Technology, Northwestern Polytechnical University, Xi'an, P.R.China*[b]*Faculty of Technology, University of Turku, Turku, Finland*[c]*School of Computing, University of Eastern Finland, Joensuu, Finland*[d]*Engineering Cluster, Singapore Institute of Technology, Singapore*

摘要

自编码器（AEs）凭借其内在的学习能力，常用于机器学习任务。这一独特特性可用于离群点检测（OD）。然而，传统的基于自编码器的方法面临决策过度自信以及离群点出现意外重构结果的问题，从而限制了其在异常检测中的性能。为缓解这些问题，我们首先分析了均方误差（MSE）和负对数似然（NLL），并阐明了在基于自编码器的异常检测中纳入偶然不确定性的重要性。随后，我们提出加权负对数似然（WNLL），以调整不确定性在不同异常检测场景中的影响。此外，我们提出均值漂移评分（MSS）方法，利用数据的局部关系来减少自编码器导致的误判为内点的数据。在32个真实世界异常检测数据集上的实验验证了所提方法的有效性。WNLL与MSS结合后，相较于最佳基线方法，性能相对提升了41%。此外，MSS使多个基于自编码器的离群点检测器的检测性能平均提升了20%。所提方法有望推动自编码器在异常检测领域的发展。

关键词：
离群点检测、异常检测、自编码器、不确定性估计、均值漂移

1. 引言

离群点检测（OD），也称为异常检测，是数据挖掘领域的一项基础技术。离群值可定义为“

*通讯作者*Email addresses:* xutan@ieee.org (Xu Tan), jiaweiyang@ieee.org (Jiawei Yang),jqchen@ieee.org (JunqiChen), sylwanrahardja@ieee.org (Sylwan Rahardja),susantorahardja@ieee.org (Susanto Rahardja)

在其所属样本中明显偏离其他成员的对象” [1]。离群点具有两个主要特征：（1）在特征方面不同于内点；（2）与内点相比，在数据集中较为罕见 [2]。离群点检测的目标是从正常样本（内点）中识别离群点，以保持数据的干净和安全，或发现异常情况与行为。离群点检测有许多应用，例如欺诈检测[3], 网络入侵检测 [4], 疾病诊断 [5], 视频监控 [6]和轨迹分析 [7]。通常，离群点检测被视为无监督或半监督任务，因为在现实应用中很难获得真实标签 [8]——这意味着大多数数据由尚未识别的内点和离群点组成。

多年来，人们提出了许多异常检测方法，可将其分为不同类别。基于邻近度的方法 [9, 10, 11, 12] 衡量原始数据空间中内点与离群值之间的相似度。基于统计的方法 [13, 14, 15] 对内点的统计分布进行建模。基于分类的方法 [16, 17, 18]将数据转换到替代空间，再通过划分替代空间来识别离群值。基于集成的方法 [19, 20, 21, 22] 利用数据的随机性和多样性，将多个较弱的离群点检测器组合起来，生成一个较强的检测器。基于概率的方法 [23, 24]分析数据属性的概率分布。

传统异常检测方法由于特征提取能力较弱，在面对高数据维度和复杂数据流形时通常会遇到问题。因此，近年来，具有强大学习能力的深度学习技术引起了异常检测领域专家的关注。基于学习的方法，例如基于自编码器（AE）的方法 [25, 26, 27, 28, 29, 30] 和基于生成对抗网络（GAN）的方法 [31, 32, 33, 34], ，利用神经网络强大的学习能力来捕捉内点的内在特征。基于自编码器的方法假设，与离群点相比，内点能够得到更好的重构。基于生成对抗网络的方法通过生成器生成离群点，或通过判别器区分内点与离群点。目前，基于生成对抗网络的方法通常难以训练，这是由于其对抗性所致，因此基于自编码器的方法在所有基于学习的方法中最受欢迎 [35]。

大多数基于AE的方法利用重构误差（例如均方误差（MSE））检测离群点。优化均方误差可以解释为在底层采用高斯误差模型的情况下最大化似然的一种特殊情形[36]。这相当于假设对数似然函数的方差项为 1，且不随数据样本或属性而变化。尽管这一假设简化了自编码器的优化过程和架构，却限制了自编码器的判别能力。更具体地说，采用均方误差损失的自编码器可能会做出过度自信却错误的判断，这不利于异常检测任务。

一些研究人员曾考虑上述方差项（也称为偶然不确定性），以提升异常检测性能 [27, 37,38, 39, 40]。然而，他们缺乏对该方差项如何影响异常检测任务的分析，尤其是对表格数据进行异常检测的情况。因此，本研究阐明了

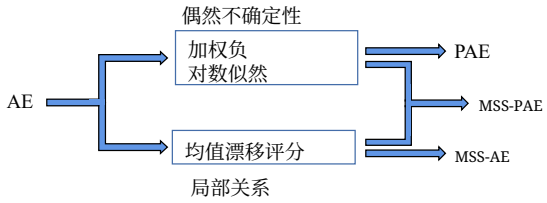


图 1：本文如何改进基于自编码器的异常检测方法示意图。WNLL 纳入偶然不确定性，以缓解自编码器的过度自信问题；MSS 引入局部关系，以检测重构良好的离群点。

方差项影响自编码器重构结果的机制，并解释了它为何能提升异常检测性能。进一步分析负对数似然（NLL）函数中的权衡后，我们发现，恰当的加权值能够显著提升异常检测性能，而这一点在文献中被忽视了。因此，我们提出了加权负对数似然（*WNLL*），这是一种适用于各种异常检测任务的有效且灵活的评分函数。

其次，传统的基于自编码器的异常检测方法存在异常重构问题：一些离群点意外地被良好重构，因而难以与内点区分 [41]。这一现象可归因于传统自编码器未考虑局部关系，而只捕捉数据的全局特征。本文通过调整自编码器的评分策略来缓解这一问题，提出了 均值漂移异常评分（*MSS*）方法。该方法使用均值偏移后的数据点，而非原始数据点来计算异常值分数。因此，当某个离群点以相对较高的似然被重构时，重构结果将与输入的均值偏移结果不相似，从而产生较高的离群分数。该方法简便有效，且易于应用于大多数基于自编码器的方法，相关内容将在第 4 节中展示。

图 1 展示了本研究将如何改进基于自编码器的异常检测方法。加权负对数似然考虑了偶然不确定性，以缓解自编码器的过度自信问题；MSS 则引入局部关系来检测重构良好的离群点。

总之，本研究的主要贡献如下：

- 从理论上分析了传统自编码器中的过度自信问题，并阐释了偶然不确定性对重构任务和基于自编码器的异常检测的影响。

- 提出了加权负对数似然，以缓解基于自编码器的异常检测中的过度自信问题，并通过平衡偶然不确定性的影响，显著提升了检测性能。

- 提出 MSS，通过引入局部关系来减少 AE 中意外出现的重构良好的离群点所造成的危害，从而提升鲁

棒性和准确性。除了效果显著外，它还可以轻松应用于其他基于自编码器的异常检测方法。

- 在 32 个真实世界异常检测数据集上开展实验，将所提出的方法与 5 种典型的基于自编码器的异常检测方法以及 8 种非基于自编码器的最先进离群检测方法进行比较。实验结果证明了我们的方法有效且优越。

本文其余部分组织如下：第 2 节回顾基于自编码器的异常检测原理、近年来提出的典型基于自编码器的异常检测方法，以及神经网络不确定性量化相关研究。第 3 节详细介绍所提出的负对数似然、加权负对数似然和 MSS 方法，并分析其理论优势。随后，第 4 节介绍实验设置并报告实验结果。最后，第 5 节总结全文。

2. 预备知识与相关工作

2.1. 用于离群点检测的自编码器

自编码器是一种经过训练、能够将输入复制到输出的神经网络 [42]。它由两个主要组件构成：编码器和解码器。以传统的全连接自编码器为例，给定一个多变量数据向量$\mathbf{x} \in \mathbb{R}^D$ 作为输入，编码器会将其转换为低维潜在表示，同时保留输入数据中最关键的特征。编码器输出的潜在表示z </style}表示为

$$\mathbf{z} = \mathrm{Encoder}(\mathbf{x}, \theta), \tag{1}$$

其中 θ 表示编码器的网络参数。然后，解码器使用潜在表示重构输入，其输出 $\mathbf{x}$ 表示为

$$\hat{\mathbf{x}} = \mathrm{Decoder}(\mathbf{z}, \phi), \tag{2}$$

其中 ϕ 表示解码器的网络参数。

由于全连接层可以视为线性降维变换，因此编码器的作用类似于主成分分析，其目标是找到反映原始数据特征的主要成分 [43]。然而，复杂的真实世界数据难以进行线性分析，因此在每个全连接层之后都应用非线性激活函数，以增强网络的表达能力。常见的典型激活函数包括 Sigmoid函数、Tanh函数、修正线性单元（ReLU）及其变体。

自编码器的优化和离群分数计算都依赖于其损失函数。由于自编码器的目标是将输入数据重构为输出，因此输入与输出之间的偏差可以用作损失函数。这也称为重构误差。最常用的损失函数

是传统自编码器的均方误差函数，其定义如下：

$$\mathrm{MSE}(\mathbf{x}, \hat{\mathbf{x}}) = ||\mathbf{x} - \hat{\mathbf{x}}||_2^2, \tag{3}$$

其中 $||\cdot||_2$ 表示 L2范数。

自编码器首先将输入$\mathbf{x}$ 转换为潜变量$\mathbf{z}$，然后将$\mathbf{z}$ 转换为输出$\hat{}\mathbf{x}$。潜变量表示$\mathbf{z}$ 的维度设计得小于输入层的维度。它充当瓶颈，仅允许关键信息通过 [44]。因此，自编码器必须捕捉数据的关键特征，才能更好地重构输入。

考虑一个同时包含内点和离群点的数据集。内点数量众多，且通常呈现相似的模式，因此自编码器很容易学到它们的特征。离群点在数据集中数量较少且各不相同，因此自编码器难以学到它们的特征。因此，内点的重构误差往往较低，而离群点的重构误差往往较高。

2.2. 基于自编码器的离群点检测方法

自编码器在基于无监督学习的异常检测方法中发挥着关键作用。自编码器既可以单独使用，也可以与其他方法结合使用 [45, 30]。除了将自编码器应用于各种异常检测应用外，一些研究人员还致力于提升自编码器本身的性能，使其更好地满足异常检测的需求。Zhou 等人提出了鲁棒深度自编码器（RDA） [26], ，其灵感来自鲁棒主成分分析。RDA提出了一种新的损失函数，可迭代地从训练集中移除异常成分，从而减少自编码器受到的污染。Chen等人提出了用于离群点检测的随机神经网络（RandNet） [46] ，通过在不同训练数据子集上使用随机架构训练多个自编码器，提升自编码器的鲁棒性。An等人提出了一种基于变分自编码器（VAE）的异常检测方法 [27]。该方法使用变分推断和重构概率，以获得更具理论依据且更客观的离群分数。Ishii 等人提出了低成本自编码器（LCAE） [47]。该方法在每个训练轮次中仅使用重构误差较低的数据，从而减少训练过程中离群点造成的污染。Gong 等人提出了记忆增强自编码器（MemAE）[48], ，通过从数量有限的已记录代表性正常模式中重构样本，提升了自编码器的鲁棒性。Lai 等人提出了一种基于鲁棒子空间恢复层（ RSRAE） [28]的自编码器模型。该模型利用鲁棒子空间恢复层增加离群点的重构难度。类似地，Yu 等人提出了一种基于正交投影约束（OPCAE）[49], 的自编码器模型，该模型使用 Kautlr-Thomas 变换，在编码后仅保留内点信息。Guo等人提出了一种名为特征分解自编码器（FDAE） [29],的自编码器架构，该架构融合了 RDA [26] 和 RSRAE [28]的优点。

总而言之，近期大多数基于自编码器的方法都致力于保护自编码器，使其免受离群点在训练过程中造成的有害影响。然而，这些方法

通常具有复杂的网络架构和训练流程，因此难以使用，在实际应用中的表现也不够理想。此外，他们没有意识到自编码器有时会在非预期区域做出过度自信却错误的判断，这将在下一小节中详细讨论。

2.3. 量化神经网络中的不确定性

尽管神经网络在各种机器学习任务中表现出令人瞩目的能力，但在某些情况下，它们往往会做出过度自信的决策 [50, 51, 52, 40,53]。例如，回归模型可能会对没有任何训练数据的区域给出置信度很高的预测 [52]。这些过度自信却错误的决策可能在实际应用中造成严重后果，尤其是在自动驾驶和医学诊断等高风险应用中 [54]。虽然更优秀的模型和大量高质量训练数据可以减少错误决策，但要考虑到所有可能的情况仍不切实际 [55]。因此，神经网络必须能够知道自己知道什么 [50], ，也就是说，进行不确定性量化。

通常，神经网络的不确定性可分为两种主要类型：偶然不确定性和认知不确定性 [56, 57, 53]。偶然不确定性也称为数据不确定性，表示观测数据中固有的噪声或随机性。根据噪声是否依赖于具体的数据实例，还可进一步分为异方差不确定性和同方差不确定性 [57, 55]。认知不确定性也称为模型不确定性，表示网络无法确定模型参数是否捕捉到了数据的内在规律。这可能由训练错误、模型能力不足，或因遇到未知样本或训练数据不足而导致的知识缺乏所引起 [58]。由于偶然不确定性和认知不确定性的性质不同，研究人员采用了不同的方法对它们进行量化。对于异方差偶然不确定性，常见做法是让网络估计观测噪声的方差 [57, 55]。要量化认知不确定性，通常会使用贝叶斯神经网络 [59, 60, 61] 或集成方法 [50, 62]。

由于不确定性与数据分布密切相关，研究人员认识到，不确定性量化有助于基于自编码器的离群点检测。一些研究量化了偶然不确定性。An 等人提出使用重构概率代替均方误差作为变分自编码器的损失函数，并指出异常数据会具有更高的偶然不确定性 [27]。Pol 等人使用负对数似然训练变分自编码器，并将经偶然不确定性归一化的重构误差作为离群分数 [37]。Mao 等人利用经负对数似然训练的自编码器检测异常像素，这些像素的重构误差与偶然不确定性的比值较高。另一些研究人员则量化了认知不确定性。Legrand 等人和 Baur 等人使用蒙特卡洛 Dropout 训练贝叶斯自编码器，以计算认知不确定性 [63, 64]。Legrand 等人将认知不确定性与重构误差结合，作为最终的离群分数；而 Baur 等人指出，脑部磁共振图像中的病变组织表现出较低的认知不确定性。Daxberger 等人训练了一个贝叶斯变分自编码器，使用

马尔可夫链蒙特卡罗计算认知不确定性，并将认知不确定性较大的数据视为离群点 [65]。Park 等人仅训练贝叶斯变分自编码器的编码器，并利用蒙特卡洛 Dropout 计算综合离群分数函数，其中包含库尔贝克-莱布勒散度和认知不确定性 [66]。一些研究人员还探索了偶然不确定性与认知不确定性的组合 [39, 67]

尽管不确定性量化已应用于以往的研究，但其对 OD 任务的影响机制尚未得到阐明，尤其是在一般表格数据上的 OD 中。值得注意的是，不同机器学习或 OD 场景中不确定性的含义可能有所不同。例如，在回归任务中，偶然不确定性反映了输入和目标测量值中的噪声，而网络无法学会对其进行校正。但在分类任务中，它反映的是信息不足，无法识别某一类数据 [58]。传统的不确定性研究主要关注回归、分类或分割。然而，对于采用自编码器等重构模型的异常检测任务，情况有所不同。本文详细分析了自编码器中的偶然不确定性，并阐释了其在表格异常检测中的优势。

3. 拯救自编码器免于意外重建

本节针对传统自编码器遭遇意外重建问题的两个原因进行了分析；该问题会导致其在某些特定情况下离群点检测性能不理想。此外，我们提出了两种有效解决该问题的方法。

3.1. 自编码器对自身的重构结果过度自信

自编码器的优化目标可以解释为：根据数据噪声或随机性，在某种潜在概率分布下最大化重构似然。假设该分布为对角多元高斯分布，则可得式 4：

$$\begin{cases} \mathbf{x}^{\mathrm{rec}} \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \\ \boldsymbol{\mu} = [\mu_1, \mu_2, \cdots, \mu_D]^T, \\ \boldsymbol{\sigma}^2 = [\sigma_1^2, \sigma_2^2, \cdots, \sigma_D^2]^T, \\ \boldsymbol{\Sigma} = \mathrm{diag}(\boldsymbol{\sigma}^2), \end{cases} \tag{4}$$

其中，$\mathbf{x}^{\mathrm{rec}}$ 表示输入$\mathbf{x}$的重建随机变量；$\mathrm{diag}(\cdot)$表示对角矩阵；$\boldsymbol{\mu}$ 表示自编码器的重建均值；$\boldsymbol{\sigma}^2$ 表示重建方差。由此可得式 5：

$$P(\mathbf{x}^{\mathrm{rec}}) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\boldsymbol{\Sigma}|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}^{\mathrm{rec}}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x}^{\mathrm{rec}}-\boldsymbol{\mu})}, \tag{5}$$

其中 $P(\cdot)$表示概率密度函数，$|\boldsymbol{\Sigma}|$ 表示矩阵 $\boldsymbol{\Sigma}$的行列式，且 $\boldsymbol{\Sigma}^{-1}$ 表示 $\boldsymbol{\Sigma}$的逆矩阵。对数

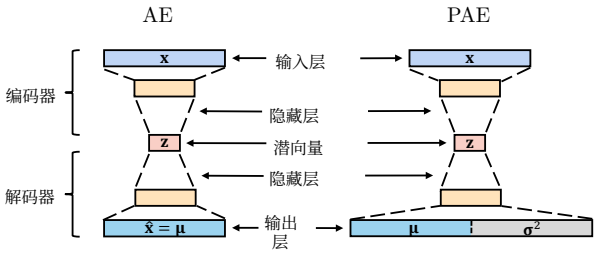


图 2：自编码器和 PAE 的典型结构。自编码器具有对称结构，且潜向量的大小远小于输入层；而 PAE 的输出层大小是输入层的两倍。

概率密度函数可表示为式 6：

$$\ln P(\mathbf{x}^{\mathrm{rec}}) = -\frac{D}{2}\ln 2\pi - \frac{1}{2}\sum_{d=1}^{D}\ln\sigma_d^2 - \frac{1}{2}\sum_{d=1}^{D}\frac{(x_d^{\mathrm{rec}}-\mu_d)^2}{\sigma_d^2}. \tag{6}$$

假设所有方差值均为 1，则

$$\begin{aligned}\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_D^2 = 1,\\ \boldsymbol{\Sigma} = \mathbf{I},\end{aligned} \tag{7}$$

其中$\mathbf{I}$表示单位矩阵。因此，对数概率密度函数（也称为$\mathbf{x}^{\mathrm{rec}}$ 的对数似然）可由公式 8 表示：

$$\begin{aligned}\ln P(\mathbf{x}^{\mathrm{rec}}) &= -\frac{D}{2}\ln 2\pi - \frac{1}{2}\sum_{d=1}^{D}(x_d^{\mathrm{rec}}-\mu_d)^2\\ &= C - \frac{1}{2}||\mathbf{x}^{\mathrm{rec}} - \boldsymbol{\mu}||_2^2,\end{aligned} \tag{8}$$

其中，C 表示一个不会影响优化过程的常数。

传统自编码器旨在重建输入$\mathbf{x}$，这意味着在$\mathbf{x}^{\mathrm{rec}} = \mathbf{x}$时最大化对数似然 $\ln P(\mathbf{x}^{\mathrm{rec}})$，并将均值 $\boldsymbol{\mu}$ 视为自编码器的输出 $\hat{\mathbf{x}}$。因此，优化目标是最大化$C - \frac{1}{2}||\mathbf{x} - \hat{\mathbf{x}}||_2^2$。显然，这等价于最小化 $||\mathbf{x} - \hat{\mathbf{x}}||_2^2$，这正是一般均方误差的形式。因此，给定训练集$\mathbf{X} \in \mathbb{R}^{N\times D}$，传统自编码器的损失函数可由公式 9 表示：

$$\mathrm{MSE}(\mathbf{X}, \hat{\mathbf{X}}) = \frac{1}{N}\sum_{i=1}^{N}||\mathbf{x}_i - \hat{\mathbf{x}}_i||_2^2, \tag{9}$$

其中，N 表示训练集中的样本数量。

自编码器领域的研究者主要使用公式 8 的形式或其变体（例如 $||\cdot||_2$（均方根误差）或 $||\cdot||_1$（平均绝对误差））作为损失函数。然而，使用均方误差损失的前提是

所有重构概率分布的方差值 $\boldsymbol{\sigma}^2$ 均为 1。在此假设下，自编码器仅估计分布的均值 $\boldsymbol{\mu}$ 。这一假设简化了自编码器的优化和架构，但它是次优解，因为无法反映实际情况。该假设认为每个数据实例和属性的随机性都是恒定且相互独立的（即同方差），而这种情况在现实中极为罕见。这一缺陷使采用均方误差的自编码器无法实现更高的重构精度，并会导致异常检测任务出现过度自信问题。因此，有必要将方差 $\boldsymbol{\sigma}^2$纳入考虑，从而得到负对数似然损失：

$$\mathrm{NLL}(\mathbf{X}, \boldsymbol{\mu}, \boldsymbol{\sigma}) = \frac{1}{N} \sum_{i=1}^{N} \sum_{d=1}^{D} \frac{(x_{i,d} - \mu_{i,d})^2}{\sigma_{i,d}^2} + \ln \sigma_{i,d}^2. \tag{10}$$

其中，$x_{i,\ d}$、 $\mu_{i,\ d}$ 和 $\sigma_{i,\ d}^2$ 分别表示样本$\mathbf{x}_i$的第 d 个属性、输出均值 $\boldsymbol{\mu}_i$和输出方差 $\boldsymbol{\sigma}_i^2$。这是对式 6 的简单变形，省去了常数项和系数，并取了相反数。由于计算负对数似然（NLL）需要两个变量，即 $\boldsymbol{\mu}_i$ 和 $\boldsymbol{\sigma}_i^2$，因此相应地需要一个具有两个输出的网络。所以，传统自编码器被改造为概率自编码器（*PAE*），如图 2 所示。其区别在于，网络输出层的维度是输入层大小的两倍，由 $\boldsymbol{\mu}_i$ 和 $\boldsymbol{\sigma}_i^2$组成。输出方差后接 Softplus激活函数，以确保数值为正。$\boldsymbol{\mu}_i$ 表示重构结果。 $\boldsymbol{\sigma}_i^2$ 可称为异方差偶然不确定性，这一概念已在文献中确立，详见第 2.3 节的介绍。然而，偶然不确定性与基于自编码器的异常检测之间的内在关系尚未得到解释，而异常检测中不确定性量化的潜力也尚未得到充分探索。

PAE 由编码器和解码器两个组件构成。编码器将高维输入数据转换为低维潜变量，而解码器则执行相反的操作。因此，PAE 的过程可视为自监督两步回归。在此过程中，多个输入实例可能被编码为同一个潜变量，随后又被解码为相同的输出。这种现象有两个可能的原因：一是内点固有的随机性或噪声在降维过程中被消除；二是离群点的模式与模型学到的知识相矛盾。

训练完成后，重构结果与训练数据的流形相吻合，而估计得到的偶然不确定性则反映了训练集的质量。由于输出与潜变量一一对应，我们可以根据潜变量的取值概括偶然不确定性的特性：当数据落在潜变量的最小—最大范围内（该范围由训练集决定）时，偶然不确定性反映了多种因素，包括数据的随机性或噪声、缺失数据的区域，以及数据稀疏的区域（很可能是离群点）；当数据落在潜变量的最小—最大范围之外时，偶然不确定性就会变得毫无意义且难以捉摸。更重要的是，将偶然不确定性与重构误差结合起来，会产生更积极的影响。这表明

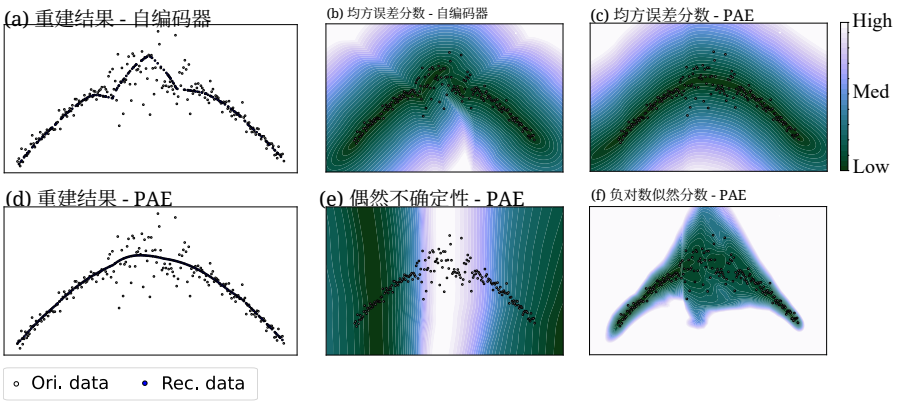


图3：第一个案例包含一组正弦形数据，不同区域具有不同程度的高斯噪声。(a) 显示，自编码器的重建结果受到噪声数据的严重影响。(b) 显示，根据自编码器计算出的均方误差分数在重构后呈现不规则性。(c) 和 (d) 显示，PAE 的重建结果平滑，且基本贴合原始曲线。(e) 显示，PAE 在水平轴上 (-0.5,0.5) 区间内输出较高的偶然不确定性。(f) 显示，PAE 对存在数据的区域赋予较低的负对数似然分数，而对噪声区域赋予相对较高的分数。图中的“Ori.”和“Rec.”分别表示“原始数据”和“重构”。

我们应关注类似负对数似然的形式，从而得到加权负对数似然：

$$\mathrm{WNLL}(\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\sigma}) = \sum_{d=1}^{D} \alpha \frac{(x_d - \mu_d)^2}{\sigma_d^2} + (1-\alpha) \ln \sigma_d^2, \tag{11}$$

其中，x_d、μ_d 和 σ_d 分别表示样本的第 d 个属性、$\mathbf{x}$、输出均值和输出方差值 $\boldsymbol{\sigma}^2$。$\alpha \in [0,1]$ 是控制 WNLL 两个组成部分之间权衡的超参数。显然，第一个组成部分同时受重构误差和偶然不确定性的影响，而第二个组成部分仅受偶然不确定性的影响。此外，偶然不确定性对这两个组成部分的影响不同。α的值越小，偶然不确定性对 WNLL 的正向影响越强。在本研究中，我们在测试阶段使用 WNLL 作为评分函数。通过调整 α，PAE 能够适应数据集的不同特征。这使其无需额外修改训练过程，便能为各种异常检测应用带来显著益处。

3.2. 使用负对数似然和加权负对数似然解决过度自信问题

为验证上一节提到的观点，并揭示负对数似然和加权负对数似然如何有益于异常检测，我们通过可视化研究了四种情况。针对每种情况，均采用相同的设置训练自编码器和 PAE。由于这些情况是二维且高度非线性的，我们将所有模型中每个网络

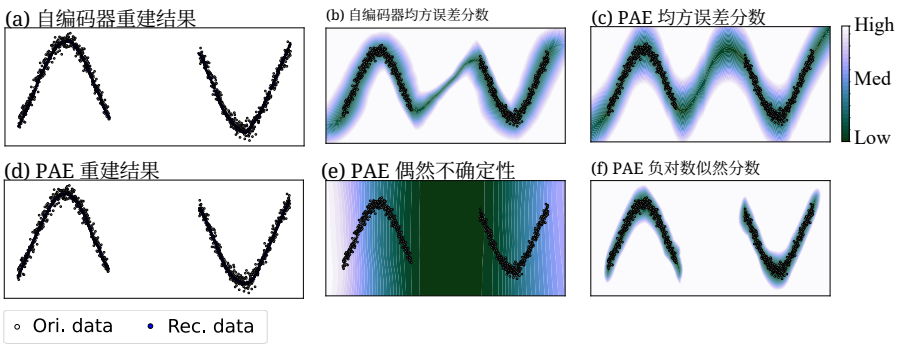


图 4：第二种情况包含两组密度相同的正弦形数据。(a) 和 (d) 显示，自编码器和 PAE 都很好地重构了原始数据。(b) 和 (c) 显示，两组数据之间的一些区域也得到了良好重构。(e) 显示，PAE 在两组数据之间的空白区域输出了低偶然不确定性。(f) 显示，PAE 仅对靠近数据的区域赋予较低的负对数似然分数。

层的单元数设置为 $[2,32,32,1,32,32,2]$，，并使用缩放指数线性单元作为激活函数，最后一层除外。

第一种情况中的数据取自正弦曲线的一部分，不同区域中的高斯噪声各不相同。如图 3.(a) 所示，原始数据以白色点标出，水平轴上 (-0.5,0.5) 区间内的噪声明显大于其他区域。我们可以观察到，自编码器的重构结果（蓝色点）受噪声数据的严重影响。根据自编码器计算出的均方误差分数也显示，重构后数据变得不规则。然而，PAE 显著缓解了这一问题，其重构结果得到了明显改善，并且与原始曲线相似，如图 3.(d) 所示。此外，图 3.(e) 显示，PAE 在水平轴上 (-0.5,0.5) 区间内输出了高偶然不确定性。这表明偶然不确定性捕捉到了原始数据中的噪声。同时考虑重构误差和偶然不确定性后，我们发现 PAE 对存在数据的区域赋予了较低的负对数似然分数。这表明 PAE 试图捕捉训练集的分布。另一方面，噪声区域的分数相对较高，这是由较高的不确定性造成的。因此，这一特性有助于将噪声数据视为离群值的应用。

第二种情况包含两组等密度的正弦形数据，如图4所示。在这种情况下，自编码器和 PAE 都很好地重构了原始数据。然而，两组数据之间的一些区域（即空白区域）也被重构为误差较低。这出乎意料，因为模型只需学习训练集中的数据模式。这种现象称为模型过度自信；如果我们只关注重构误差，它会对异常检测造成不利影响。值得庆幸的是，PAE 估计的偶然不确定性可以解决这个问题。图4(e)显示，两组数据之间空白区域的不确定性较低。尽管有些

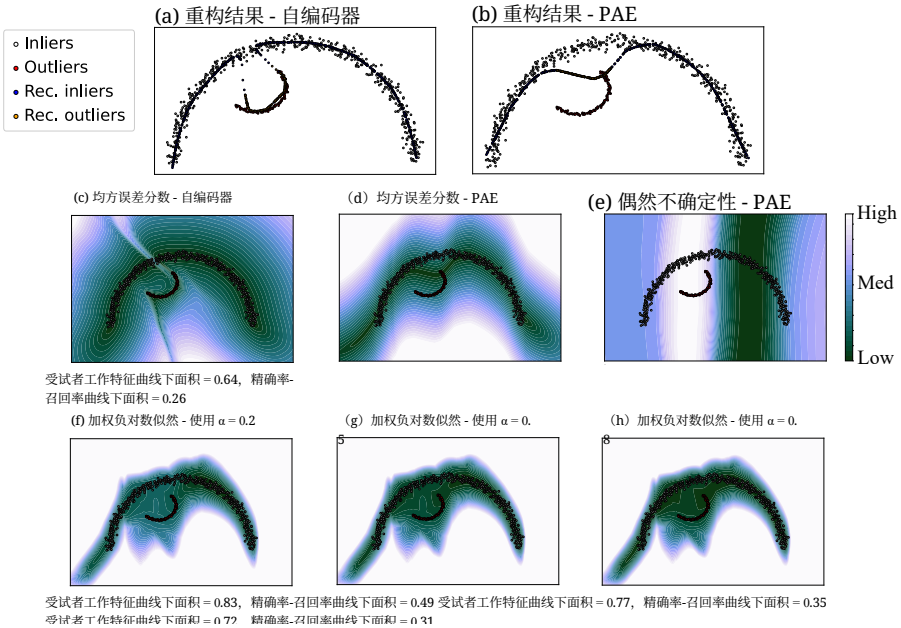


图 5：第三种情况包含一大组半圆形内点和一小组弧形离群点，它们在水平轴上的范围有所重叠。（a）和（c）显示，自编码器错误地拟合了离群点，导致异常检测性能不佳。（b）和（d）显示，PAE 的重构结果与内点流形相差不大。（e）显示，PAE 在离群点周围输出较高的偶然不确定性，使离群点更容易被区分。（f）至（h）显示，使用加权负对数似然时，PAE 以较小的 α 实现了更好的异常检测性能。

两组数据之间的区域出现了不理想的低重构误差，但根据公式10的第一项，较低的偶然不确定性会放大最终误差。因此，PAE 得到了理想的负对数似然分数，如图4(f)所示。这样，空白区域中潜在的离群点就不会被误判。

第三种和第四种情形下的数据中，内点占比较大，而集体异常值占比较小。除了测试自编码器和 PAE 的重构性能外，还测试了采用不同评分函数时的异常检测性能。采用受试者工作特征曲线下面积（AUROC）和精确率-召回率曲线下面积（AUPRC）作为异常检测性能指标。这两个指标都适用于标签不平衡分类任务，取值范围为 0 到 1，其中 1 表示性能最佳。

第三种情况包含一大组半圆形内点和一小组弧形离群点，它们在水平轴上的范围有所重叠。如图5所示，我们可以观察到，自编码器错误地拟合了离群点，导致异常检测性能不佳。相比之下，PAE 的结果有所不同。其重建结果与内点没有显著偏离。然而，潜在空间部分区域中内点与离群点的重叠

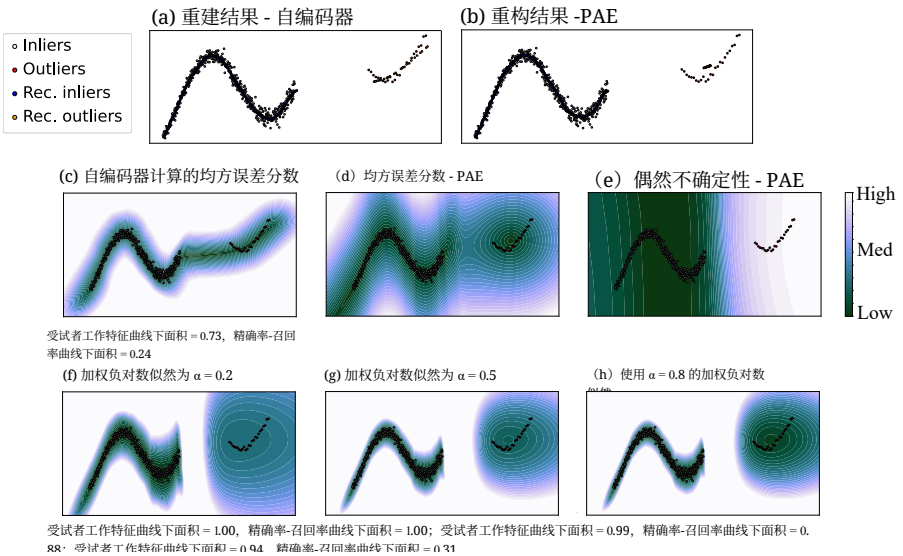


图 6：第四种情况包含一大群正弦形状的内点和一小群弧形离群点，且它们在水平轴上没有重叠。（a）和（c）显示，自编码器能够很好地重构内点，但它不仅在离群点附近输出较低的均方误差分数，也在内点与离群点之间输出较低分数。（b）和（d）显示，PAE 未能很好地重构离群点，但仍为一些离群点赋予了相对较高的均方误差分数。（e）显示，PAE 在上述两个区域都输出了较高的偶然不确定性。（f）至（h）显示，PAE 通过降低加权负对数似然中的 α，取得了更好的异常检测性能。

使 PAE 感到困惑，导致离群点所在区域具有较高的偶然不确定性。因此，尽管 PAE 希望为所有数据分配较低的负对数似然分数，但对于离群点高度集中的区域，它只能分配相对较高的分数。鉴于此，尽管部分内点受到怀疑，但大多数离群点仍能被识别出来。为更好地避免离群点被错误地识别为正常数据，我们可以降低加权负对数似然中的 α，因为这能增强偶然不确定性的正向作用。因此，当 $\alpha = 0.2$ 时，PAE 的异常检测性能优于 $\alpha = 0.5$（等同于负对数似然）。

最后一种情况包含一大群正弦形状的内点和一小群弧形离群点，且它们在水平轴上没有重叠。如图6所示，自编码器在离群点周围以及内点与离群点之间都错误地生成了均方误差分数较低的区域，如图6.(c)所示。这一现象损害了异常检测性能。相比之下，对于 PAE 而言，离群点的稀疏性和少数性使其在训练期间对离群点周围区域的信息感到不确定，因此在该区域输出较高的偶然不确定性。在这种情况下，偶然不确定性的正面作用对于区分至关重要，因此，较小 α 的加权负对数似然带来了更好的异常检测性能。

最后，我们将 PAE 中偶然不确定性的表现总结如下：

1）在数据具有较大随机性或噪声的区域，不确定性会增加；2）在潜在空间中内点与离群点重叠的区域，不确定性会增加；3）在数据稀疏且数量较少的区域（很可能是离群值），不确定性会增加；4）在数据分布的空洞区域，不确定性会降低。

此外，我们还针对异常检测任务中 NLL 和 WNLL 的使用提出了一些建议：

1) 当训练数据包含多个内点流形（模式）时，NLL 和 WNLL 能有效缓解过度自信问题；此时建议将 WNLL 中的 α 设为较高值；2) 当训练数据包含离群点时，NLL 和 WNLL 同样有效；此时建议将 WNLL 中的 α设为较低值。3) 对于同时存在多个内点模式和离群点污染的应用，实践者应根据具体分析，谨慎调整 WNLL 中的 α 。

3.3. 自编码器忽略了局部关系

自编码器经过训练以学习数据的正常模式，而离群点直观上应具有较大的重构误差。虽然可以直接将均方误差等重构误差用作离群分数，但在某些情况下，该分数的可靠性可能较低。首先，如果训练数据集受到离群点污染，自编码器就会被误导，学到数据的异常模式，导致离群点的重构误差降低，从而更难将离群点与内点区分开来。其次，有些离群点的潜在特征符合内点的模式，但其特征值偏离正常范围。这些离群点可能会被自编码器意外地很好地重构，因此难以与内点区分。尽管 NLL 和 WNLL 可以缓解离群点污染问题，但仍有必要为任何基于自编码器的异常检测方法寻找互补的解决方案。

上述问题可以通过自编码器捕获数据的全局特征来解释，因为它们没有明确考虑局部关系。因此，为弥补这一不足，我们将数据的局部关系作为自编码器的辅助信息。均值漂移方法可以作为一种可行的解决方案。均值漂移的概念最早由 Fukunaga等人于 1975年[68], 提出，用于估计密度函数的梯度。此后，它被应用于多种领域，包括聚类 [69], 图像分割[70],和目标跟踪[71]。

均值漂移的过程可描述如下。给定数据集 $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}2, \cdots, \mathbf{x}_N\}$，均值漂移首先计算 X 中每个样本的 k 近邻。例如，对于样本 $\mathbf{x}_i$，其 k 近邻集合为 $\mathbf{NL}_i = \{\mathbf{x}_i^1, \mathbf{x}_i^2, \cdots, \mathbf{x}_i^k\}$。随后，均值漂移将样本自身加入邻居集合 $\mathbf{NL}_i = \{\mathbf{x}_i, \mathbf{x}_i^1, \mathbf{x}_i^2, \cdots, \mathbf{x}_i^k\}$，并计算该列表的均值，如下所示：

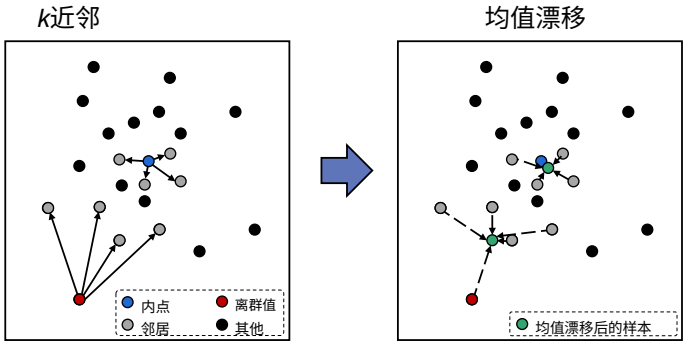


图 7：对内点（上）和离群值（下）进行均值漂移的二维示意图，其中使用了 $m = 1$ 和 $k = 4$ 。

如式 12 所示。

$$\begin{cases} \mathbf{x}_i^{\mathrm{MS}(1,k)} = \dfrac{1}{|\mathrm{NL}_i|} \displaystyle\sum_{\mathbf{x} \in \mathrm{NL}_i} \mathbf{x}, \\ \mathbf{X}^{\mathrm{MS}(1,k)} = \{\mathbf{x}_1^{\mathrm{MS}(1,k)}, \mathbf{x}_2^{\mathrm{MS}(1,k)}, \cdots, \mathbf{x}_N^{\mathrm{MS}(1,k)}\}. \end{cases} \tag{12}$$

此过程可以重复多次，以获得更加紧凑的数据分布。具体而言，可以对$\mathbf{X}^{\mathrm{MS}(1,k)}$ 应用相同的过程，得到$\mathbf{X}^{\mathrm{MS}(2,k)}$，其中 MS$(m, k)$表示使用 k 个邻居并进行 m 次漂移的均值漂移过程。图 7 展示了均值漂移的二维示意图。

为了将均值漂移应用于基于自编码器的异常检测方法，提出了以下过程：在训练阶段，可以使用公式 9 或公式 10 的损失函数训练自编码器。在评分阶段，可以使用公式 13 和公式 14 所示的评分函数计算测试样本$\mathbf{x}$ 的离群分数。

$$\mathrm{MSS\text{-}MSE}(\mathbf{x}) = ||\mathbf{x}^{\mathrm{MS}(m,k)} - \hat{\mathbf{x}}||_2^2, \tag{13}$$

or,

$$\mathrm{MSS\text{-}WNLL}(\mathbf{x}) = \sum_{d=1}^{D} \alpha \frac{(x_d^{\mathrm{MS}(m,k)} - \mu_d)^2}{\sigma_d^2} + (1-\alpha) \ln \sigma_d^2, \tag{14}$$

其中，$\mathbf{x}^{\mathrm{MS}(m,k)}$ 表示测试样本$\mathbf{x}$的均值偏移结果。

3.4 使用 *MSS* 排除重构良好的离群点

MSS 的概念可以从两个角度进行分析。从距离度量的角度来看，重构误差可以视为对象在原始数据空间中的位置偏差。如果自编码器训练充分，内点的重构误差相对较小，这意味着重构内点与原始位置非常接近。同时，内点的均值偏移结果也接近其原始位置，

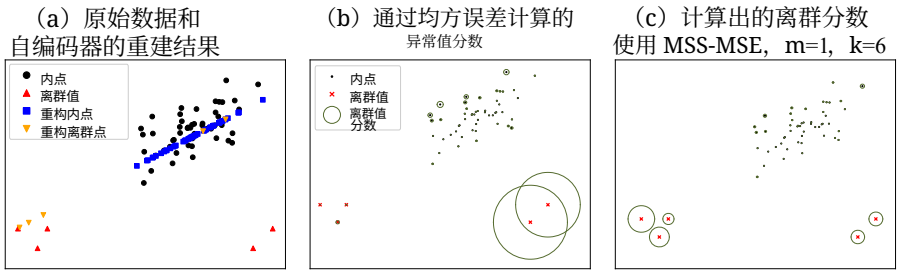


图 8：MSS 影响的二维示例。使用均方误差（MSE）损失函数，基于这些数据训练了一个自编码器（AE）。（b）和（c）分别展示了由 MSE 和 MSS-MSE 与 $m = 1$，$k = 6$ 生成的离群分数。绿色圆圈的半径越大，离群分数越高。

因此，重构内点与均值偏移内点之间的距离会很小。对于离群点，均值偏移结果通常更接近内点，而远离其原始局部区域。由于自编码器学习不充分，离群点的重构结果在距离上偏差较大，且方向随机。因此，与内点相比，离群点的均值偏移结果与重构结果之间的距离会显著更大，使其易于区分。从重建似然的角度来看，尽管离群点的重构似然相对较高，其均值偏移结果却与重构结果并不相似。由于均值偏移结果落在似然分布中似然较低的位置，因此其离群分数较高。

从理论上讲，使用 k-近邻均值与重构数据之间的距离，而不是原始数据与重构数据之间的距离，具有两项优势。首先，一些数据共享相同的均值，因此特征空间中的相似数据可能具有相似的离群分数 [72]。这样可以避免离群分数空间中的局部方差。其次，均值被视为 k-近邻的代表，因此在进行 [73]评分时，同时考虑了对象级因素和群组级因素。因此，这为检测点异常值和集体异常值提供了一种稳健的方法。

图 8 展示了 MSS 效果的一个示例。位于左下方的离群点恰好沿着内点的特征方向分布，因此其重构误差出乎意料地小，使得难以将这些离群点与内点区分开来。然而，由于它们距离内点较远，应用 MSS 并使用 $m = 1$ 和 $k = 6$ 后，便可以轻松识别它们。

图 9 展示了 MSS 方法在*Cardiotocography*数据集上的一个示例。在此案例中，MSS 应用于传统自编码器，称为 MSS-AE。为了更好地展示，使用 t分布随机邻居嵌入（t-SNE）将*Cardiotocography* 中的数据降至二维 [74]。图 9.(a) 展示了数据的真实标签图，其中黑点表示内点，红点表示离群点。可以观察到，离群点分布在图的左上区域和右下区域。

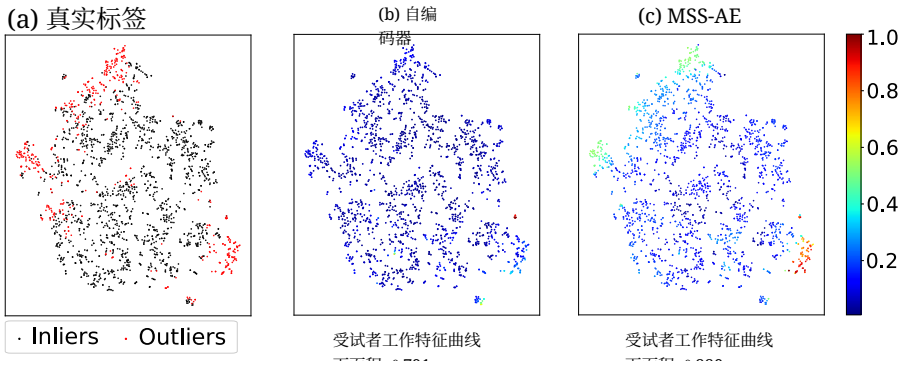


图 9：在 *Cardiotocography* 数据集上应用 MSS方法的示意图。 *Cardiotocography* 中的数据通过 t-SNE 降维至二维。(a)展示了数据的真实标签图，其中黑点表示内点，红点表示离群点。(b)和(c)分别展示了由自编码器和 MSS-AE 生成的离群分数图。

该图中。图 9.(b) 和 (c) 分别展示了由 AE 和 MSS-AE 生成的离群分数图。色标对应每个对象的离群分数。性能通过受试者工作特征曲线下面积进行评估。AE 能够区分位于右下方的部分离群点，受试者工作特征曲线下面积达到 0.791，但也无法区分左上方的离群点。此外，异常值分数的区分度也不够高。相比之下，MSS-AE 能够区分位于右下方和左上方的离群点。MSS-AE 提供了区分度更高的离群分数，并将受试者工作特征曲线下面积显著提升至 0.880。

MSS 也可应用于 PAE，从而得到 MSS-PAE。图 10 展示了 MSS-PAE 的训练和测试流程。第 4 节评估了 MSS-PAE 的性能。

4. 实验

本研究在 32 个常用的真实世界离群点检测数据集上进行了实证实验。研究将所提出的方法与 5 种典型的基于自编码器的异常检测方法（包括自编码器本身）进行了比较，证明了应用 MSS 的有效性。此外，还与 8 种经典的非基于自编码器的最先进异常检测方法进行了比较。

实验回答了以下问题：1) WNLL 在真实场景中有多有效、适应性有多强？（第 4.3.1 节）2) 超参数 k 和 m 如何影响 MSS？（第 4.3.2 节）3) MSS 是否有益于其他基于AE的方法？（第 4.3.3 节）4) 将所提出的方法与 5 种基于AE的方法和 8 种非基于自编码器的异常检测方法进行比较时，哪种方法表现最佳？（第 4.3.3 节）

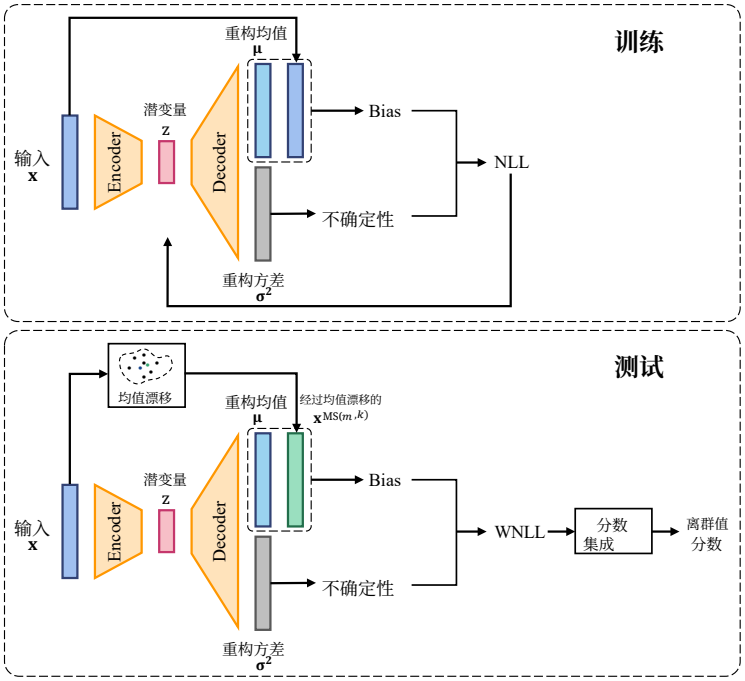


图 10：MSS-PAE 训练和测试流程示意图。图中的“Rec.”表示“重构”。

4.1. 数据集和基线方法

从 DAMI[1] 和 ODDS[2] 数据集存储库中获取了 32 个真实世界的离群点检测数据集。这些数据集的概况见表 1。我们按 3:1 的比例将每个数据集划分为训练集和测试集。随后，从训练集中选取三分之一的样本组成验证集。三个集合中的异常值比例保持不变；验证集和测试集包含数据标签，而训练集不包含。此外，使用均值-标准差归一化 [75]将数据的每个属性归一化为零均值和单位方差。均值和方差取自训练集。对于基于自编码器的方法而言，这一步至关重要，因为它平衡了数据各属性的尺度，避免数值尺度相对较大的属性对模型优化产生意外的更大影响。

为展示所提出方法的优越性，我们选取了 5 种典型的基于自编码器的方法和 8 种非基于自编码器的先进异常检测方法作为基线方法。基于自编码器的方法包括 AE [25], RDA [26], LCAE [47], RSRAE [49] 以及

[1]DAMI：www.dbs.ifi.lmu.de/research/outlier-evaluation/DAMI [2]
ODDS：odds.cs.stonybrook.edu

表 1：32 个真实世界离群点检测数据集的信息。

Name	维度	样本	离群点	比例
ALOI	27	49,534	1,508	3.0%
Annthyroid	21	7,129	534	7.5%
Arrhythmia	259	450	206	45.8%
Breastw	9	683	239	35.0%
Cardiotocography	21	2,114	466	22.0%
玻璃	7	214	9	4.2%
心脏病	13	270	120	44.4%
InternetAds	1,555	1,966	368	18.7%
电离层	32	351	126	35.9%
字母	32	1,600	100	6.3%
淋巴造影	3	148	6	4.1%
乳腺摄影	6	11,183	260	2.3%
MNIST	100	7,603	700	9.2%
Musk	166	3,062	97	3.2%
光学数字	64	5,216	150	2.9%
PageBlocks	10	5,393	510	9.5%
帕金森病	22	195	147	75.4%
PenDigits	16	9,868	20	0.2%
Pima	8	768	268	34.9%
Satellite	36	6,435	2,036	31.6%
Satimage-2	36	5,803	71	1.2%
Shuttle	9	1,013	13	1.3%
SpamBase	57	4,207	1,679	39.9%
语音	400	3,686	61	1.7%
Stamps	9	340	31	9.1%
甲状腺	6	3,772	93	2.5%
Vertebral	6	240	30	12.5%
元音	12	1,456	50	3.4%
波形	21	3,443	100	2.9%
Wilt	5	4,819	257	5.3%
Wine	13	129	10	7.8%
WPBC	33	198	47	23.7%

FDAE [29]。这些方法的详细介绍见第 2.2 节。非基于自编码器的异常检测方法包括：基于邻近度的方法——局部离群因子（LOF） [10], 均值漂移离群检测器（MOD） [12], 和密度递增路径（DIP） [76], ； 基于分类的方法——单类支持向量机（OCSVM）[16], ； 基于集成的方法——孤立森林（iForest） [19] 和深度孤立森林（DIF） [77], ； 基于概率的方法——基于直方图的离群分数（HBOS） [23] 和基于经验累积分布的离群检测（ECOD） [24]。我们复现了基于自编码器的方法，并将结果与其原始结果进行比较；此外，还应用所提出的 MSS 方法来提升性能。所有方法均使用 Python 实现；除 DIP 外，非基于自编码器的方法均采用 PyOD [78]中的实现，DIP 则使用其原始代码实现。评估指标采用受试者工作特征曲线下面积（AUROC）。

表 2：基于AE的方法的架构设置。

维度 (D)	层数	单元数
<20	3	[D, D // 2, D]
≥20, <100	5	[D, D // 2, D // 4, D // 2, D]
≥100, <200	7	[D, D // 2, D // 4, D // 8, D // 4, D // 2, D]
≥200	7	[D, D // 2, D // 4, D // 16, D // 4, D // 2, D]

4.2. 实验设置

4.2.1. 网络架构

PAE 的架构已在第 3.1 节介绍。该模型使用负对数似然（NLL）作为训练损失函数，并使用加权负对数似然（WNLL）作为评分函数。MSS 是一种通用评分方法，可应用于大多数基于自编码器的方法，包括本文提出的 PAE。例如，应用于 AE 得到 MSS-AE，应用于 RSRAE 得到 MSS-RSRAE，依此类推。

由于所有数据均为表格数据，因此使用全连接层构建自编码器。层数取决于输入数据的维度。网络架构的具体设置见表 2。随后，在自编码器的每一层之后（最后一层除外）添加修正线性单元（ReLU）激活函数，以增强网络的非线性。ReLU 被广泛用作深度神经网络中的激活函数 [79]。它模拟人脑中的神经元机制，并在许多任务中展现出良好的性能，同时只需极少的计算资源。本实验中的所有基于自编码器的方法均采用这些设置。

4.2.2. 超参数

所提出的方法需要三个超参数，分别为 α、 m和 k。对于 WNLL，测试了五种不同的 α： $\{0.20$、0.33、0.50、0.66、$0.8\}$。结果见第 4.3.1 节。对于 MSS，均值漂移时间 m 在 1 到 3 之间变化。最近邻的数量 k 从候选列表$\mathbf{k}$中选取。在我们的实验中，$\mathbf{k} = [1, 2, \cdots, \min(99, N-1)]$,，其中 N表示训练集中样本的数量。性能结果见第 4.3.2 节。

4.2.3. 训练策略

所有基线方法和所提出的方法均在训练集上训练，并借助验证集进行调优。训练完成后，使用最佳模型在测试集上进行测试，以获得最终结果。此外，为了提高基于AE的方法的稳定性，我们采用了集成技术。我们为自编码器生成了20个初始

表 3：不同 α 下自编码器和 PAE 的受试者工作特征曲线下面积结果。

数据集	AE	PAE* 在不同 α 下				
		0.20	0.33	0.50	0.66	0.80
ALOI	0.563	0.545	0.559	0.562	0.556	**0.564**
Annthyroid	0.608	0.799	0.793	0.790	**0.850**	0.835
Arrhythmia	0.714	0.736	**0.740**	0.624	0.706	0.649
Breastw	0.972	0.994	**0.995**	0.994	0.971	0.971
Cardiotocography	0.811	0.818	0.818	0.819	0.858	**0.870**
玻璃	0.804	0.961	**0.980**	0.873	0.912	0.882
心脏病	0.813	0.865	**0.871**	0.859	0.831	0.768
InternetAds	**0.706**	0.685	0.681	0.671	0.674	0.697
电离层	0.971	**0.985**	0.971	0.914	0.957	0.956
字母	**0.857**	0.855	0.768	0.808	0.764	0.807
淋巴造影	**0.971**	0.914	0.914	0.914	0.914	0.914
乳腺摄影	**0.893**	0.884	0.882	0.891	0.877	0.874
MNIST	0.882	**0.948**	0.945	0.931	0.906	0.891
Musk	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
光学数字	0.760	**0.936**	0.912	0.865	0.850	0.874
PageBlocks	0.952	**0.962**	0.946	0.949	0.942	0.953
帕金森病	0.627	**0.815**	0.662	0.711	0.611	0.576
PenDigits	0.879	**0.951**	0.929	0.913	0.897	0.907
Pima	0.651	**0.756**	0.736	0.698	0.696	0.683
Satellite	0.750	**0.855**	0.845	0.817	0.777	0.740
Satimage-2	**0.993**	0.983	0.983	0.988	0.988	0.988
Shuttle	0.995	**0.996**	0.988	0.992	0.992	0.992
SpamBase	0.565	**0.847**	0.793	0.712	0.646	0.650
语音	0.551	0.541	0.532	0.561	**0.597**	0.495
Stamps	0.931	0.844	**0.952**	0.950	0.948	0.948
甲状腺	0.975	0.973	**0.991**	0.977	0.979	0.982
Vertebral	0.607	0.676	**0.772**	0.349	0.747	0.758
元音	0.937	0.968	**0.976**	0.972	0.955	0.931
波形	0.781	**0.881**	0.790	0.776	0.771	0.771
Wilt	0.383	0.822	0.823	0.818	0.861	**0.879**
Wine	**0.914**	0.879	0.897	0.897	**0.914**	**0.914**
WPBC	0.499	**0.548**	0.516	0.501	0.484	0.479
AVG.	0.791	**0.851**	0.843	0.815	0.826	0.819
BEST	7	**13**	9	1	3	5

* 表示所提出的方法。

状态，并分别进行训练。训练完成后，使用验证集评估这些模型，选出最佳的5个模型。最后，针对一个测试样本，利用这5个模型计算出5个离群分数，然后对这些分数进行归一化并取平均，作为最终分数。为便于复现，所提出方法的源代码已发布在GitHub上 [3]。

4.3. 实验结果

4.3.1. 概率自编码器的评估

本节评估了 PAE 的性能。如表 3 所示，在 α 的全部五个取值下，PAE 在 32 个真实世界数据集上的平均性能均优于通用 AE。相较于 AE,

[3] github.com/Ra1demmmm/基于自编码器的-离群值-检测

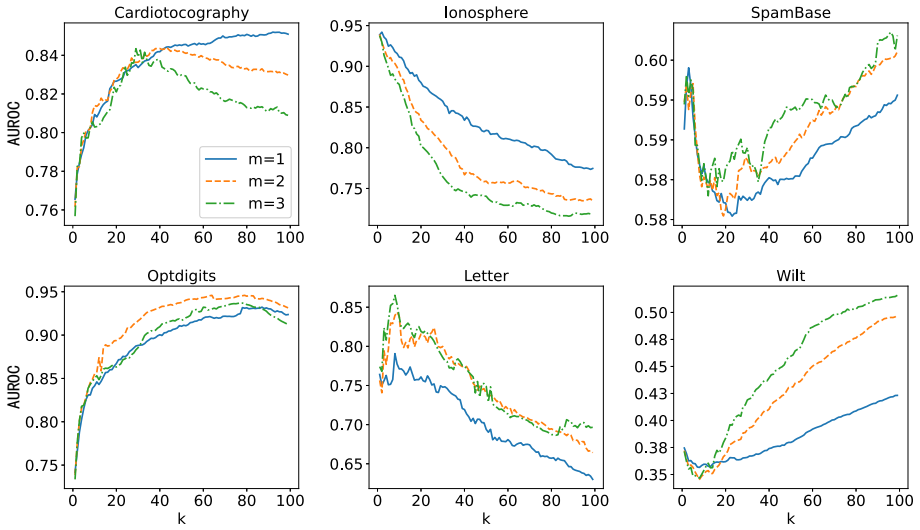


图 11 k 1 99 6 *Car-* ：*diotocography* *Ionosphere* *SpameBase* *Optdigits* *Letter* *Wilt* 当 从 增加至 时，数据集（ 、 、 、 、 和 ）上的性能曲线。每张图包含三条不同的曲线， m 表示均值平移的时间。对于每个数据集，均在相应的训练集上计算 -NN 关系（完整数据集的大小为 $\frac{3}{4}$ ），并在相应的验证集上评估 AUROC 结果。

PAE（$\alpha = 0.20$）的平均AUROC 相对提高幅度最大，达到 29%（0.851 对比 0.791）。此外，在某些数据集上，PAE 的表现显著优于 AE，例如甲状腺数据集 （0.850 对比 0.608）、玻璃 （0.980 对比 0.804）、光学数字 （0.936 对比 0.760）、*SpamBase* （0.847 对比 0.565）和 *Wilt* （0.879 对比 0.383），这表明 PAE 在特定应用中的优势。

α 的最佳值取决于具体应用。例如，在数据集*ALOI*、*Cardiotocography*、*Wilt* 和葡萄酒上，随着 α 增加，性能有所提升，这表明在这些应用中，重构误差或偶然不确定性的负面影响更为关键。相反，在数据集光学数字、*PenDigits*、*Satellite* 和波形上，随着 α 降低，性能有所提升，这表明在这些应用中，偶然不确定性的正面影响更为关键。由此可推断，应根据具体应用选择合适的 α 值，以获得最佳异常检测性能。

实验表明，其他所有组的平均表现都优于 $\alpha = 0.50$ 的值（该值等同于 NLL），这证明了WNLL中可调参数的重要性。此外， $\alpha = 0.20$ 的值在13个数据集（41%）上取得了最佳表现，平均AUROC表现也最佳（0.851）。这清楚地表明，自编码器的不确定性对异常检测至关重要，在许多应用中甚至比常用的偏差度量更重要。过去对这一变量的重视不足，而本实验中的发现可能带来重大变化。这一发现有望促使所有基于AE的方法在异常检测应用中得到进一

表4：自编码器与MSS-AE、PAE与MSS-PAE的受试者工作特征曲线下面积结果。

数据集	AE	MSS-AE			PAE	MSS-PAE		
		$m=1$	$m=2$	$m=3$		$m=1$	$m=2$	$m=3$
ALOI	**0.563**	0.558	0.563	0.561	0.545	0.630	0.658	**0.664**
Annthyroid	0.608	0.603	0.630	**0.641**	0.799	0.856	0.886	**0.893**
Arrhythmia	0.714	0.720	0.724	**0.730**	0.736	0.716	0.761	**0.762**
Breastw	0.972	**0.994**	0.987	0.987	0.994	0.994	**0.996**	0.992
Cardiotocography	0.811	0.875	**0.888**	0.877	0.818	0.841	**0.844**	**0.844**
玻璃	0.804	0.922	0.863	**0.951**	**0.961**	0.804	0.824	0.843
心脏病	0.813	0.859	**0.869**	0.799	0.865	**0.892**	0.883	0.886
InternetAds	0.706	0.724	**0.801**	0.783	0.685	**0.797**	**0.797**	**0.797**
电离层	0.971	**0.978**	0.975	0.973	**0.985**	0.962	0.964	0.962
字母	0.857	0.844	**0.860**	0.853	**0.855**	0.748	0.742	0.798
淋巴造影	0.971	**1.000**	**1.000**	**1.000**	0.914	0.971	**1.000**	**1.000**
乳腺摄影	**0.893**	0.876	0.885	0.885	0.884	0.882	**0.885**	0.882
MNIST	**0.882**	0.876	0.881	**0.882**	0.948	**0.949**	**0.949**	0.948
Musk	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
光学数字	0.760	0.957	0.970	**0.981**	0.936	**0.994**	0.993	0.992
PageBlocks	0.952	**0.955**	0.948	0.953	**0.962**	0.958	0.957	0.954
帕金森病	0.627	0.639	**0.769**	0.736	0.815	0.766	0.796	**0.819**
PenDigits	0.879	0.887	**0.974**	0.973	0.951	0.963	**0.977**	0.971
Pima	0.651	0.689	0.691	**0.699**	0.756	**0.757**	0.740	0.722
Satellite	0.750	0.788	**0.789**	0.763	0.855	0.856	**0.857**	**0.857**
Satimage-2	**0.993**	**0.993**	0.991	0.989	**0.983**	0.981	0.975	0.967
Shuttle	0.995	0.997	**1.000**	0.997	0.996	**1.000**	0.997	0.997
SpamBase	0.565	0.580	0.622	**0.623**	0.847	**0.886**	0.879	0.878
语音	0.551	0.567	0.586	**0.651**	0.541	**0.658**	0.477	0.527
Stamps	0.931	**0.961**	0.915	0.950	0.844	**0.954**	0.942	0.837
甲状腺	**0.975**	0.972	0.963	0.960	0.973	**0.994**	0.991	0.990
Vertebral	0.607	**0.755**	0.750	0.745	0.676	0.687	**0.712**	0.703
元音	0.937	0.906	0.925	**0.958**	**0.968**	0.945	0.940	0.943
波形	0.781	0.792	0.826	**0.872**	0.881	0.895	0.897	**0.901**
Wilt	0.383	0.427	0.478	**0.511**	0.822	**0.884**	0.873	0.676
Wine	0.914	0.948	**1.000**	**1.000**	0.879	0.948	**0.966**	**0.966**
WPBC	0.499	**0.506**	0.413	0.477	**0.548**	0.415	0.428	0.450
AVG.	0.791	0.817	0.829	**0.836**	0.851	**0.862**	**0.862**	0.857
BEST	6	9	11	**14**	8	**12**	11	11

状态所有基于自编码器的方法在异常检测应用中。

4.3.2. 均值漂移异常评分的评估

最近邻的数量 k 和均值平移所需的时间 m 是MSS方法的两个重要参数。实际上，这些参数都取决于具体应用。图11展示了在多个数据集上 k 对MSS-AE的影响。该图呈现了不同数据集的性能随 k 增加而变化的情况，以及每个数据集各自不同的最优 k 。与所有基于 k-NN的方法类似，这一因素取决于具体应用的数据结构。

MSS-AE 和 MSS-PAE（$\alpha = 0.20$）的 AUROC 结果列于表4中。对三种方法均测试了均值平移所用时间的变化 $m = 1$，2，3，并使用验证集为每种情况找到了最优 k 。在32个数据集上的平均结果表明，MSS-AE 和 MSS-PAE 分别优于对应的 AE 和 PAE。 $m = 3$时，MSS-AE 的平均 AUROC 表现最佳，为0.836； $m = 1$时，MSS-PAE 的表现最佳，为0.858，在所有竞争方法中排名最高。总计

表5：6种基于自编码器的方法及其 MSS 版本的受试者工作特征曲线下面积结果。

数据集	AE [25]	MSS-自编码器 *	PAE *	MSS-PAE *	RDA [26]	MSS-RDA *	LCAE [47]	MSS-LCAE *	RSRAE [49]	MSS-RSRAE *	FDAE [29]	MSS-FDAE *
ALOI	0.563	0.563	0.564	**0.706**	0.562	0.557	0.553	0.584	0.549	0.571	0.547	0.540
Annthyroid	0.608	0.641	0.850	**0.893**	0.608	0.617	0.749	0.689	0.611	0.606	0.623	0.617
Arrhythmia	0.714	0.730	0.740	**0.762**	0.714	0.729	0.724	0.713	0.711	0.724	0.709	0.639
Breastw	0.972	0.994	0.995	**0.996**	0.972	0.995	0.988	0.993	0.995	**0.996**	0.993	0.989
Cardiotocography	0.811	0.888	0.870	0.899	0.811	0.888	0.812	0.883	0.812	**0.901**	0.812	**0.901**
玻璃	0.804	0.951	**0.980**	0.853	0.863	**0.980**	0.922	0.863	0.745	0.941	0.755	0.951
心脏病	0.813	0.869	0.871	0.905	0.813	0.885	0.770	**0.909**	0.799	0.845	0.750	0.848
InternetAds	0.706	0.801	0.697	**0.839**	0.707	0.801	0.702	0.765	0.653	0.754	0.665	0.711
电离层	0.971	0.978	**0.985**	0.978	0.968	0.968	0.980	0.977	0.965	0.962	0.949	0.952
字母	0.857	0.860	0.855	0.798	0.820	0.833	0.855	0.832	**0.867**	0.794	0.829	0.793
淋巴造影	0.971	**1.000**	0.914	**1.000**	0.971	**1.000**	0.971	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
乳腺摄影	0.893	0.885	0.891	0.901	0.892	0.890	0.896	0.895	**0.907**	**0.907**	0.890	0.890
MNIST	0.882	0.882	0.948	**0.949**	0.882	0.889	0.925	0.919	0.891	0.886	0.892	0.882
Musk	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
光学数字	0.760	0.981	0.936	0.995	0.760	0.981	0.759	0.950	0.795	**0.998**	0.818	0.990
PageBlocks	0.952	0.955	0.962	0.958	0.952	0.956	0.944	0.948	0.959	**0.961**	0.958	**0.961**
帕金森病	0.627	0.769	0.815	**0.861**	0.620	0.769	0.755	0.725	0.690	0.778	0.653	0.764
PenDigits	0.879	0.974	0.951	**0.993**	0.841	0.970	0.925	0.976	0.764	0.946	0.831	0.960
Pima	0.651	0.699	0.756	**0.757**	0.651	0.693	0.677	0.677	0.681	0.723	0.661	0.667
Satellite	0.750	0.789	0.855	0.857	0.745	0.791	0.830	**0.858**	0.817	0.829	0.840	0.847
Satimage-2	0.993	0.993	0.988	0.987	0.993	0.993	**0.994**	0.993	0.992	0.990	0.993	0.978
Shuttle	0.995	**1.000**	0.996	**1.000**	0.995	0.997	0.960	**1.000**	0.989	**1.000**	0.977	0.987
SpamBase	0.565	0.623	0.847	**0.886**	0.554	0.623	0.548	0.734	0.594	0.771	0.570	0.716
语音	0.551	0.651	0.597	**0.673**	0.525	0.651	0.529	0.634	0.507	0.528	0.563	0.488
Stamps	0.931	0.961	0.952	0.954	0.957	0.961	0.941	**0.974**	0.941	0.957	0.941	0.918
甲状腺	0.975	0.972	0.991	**0.994**	0.976	0.972	0.980	0.980	0.982	0.982	0.971	0.970
Vertebral	0.607	0.755	0.772	**0.841**	0.593	0.755	0.676	0.808	0.602	0.745	0.703	0.695
元音	0.937	0.958	**0.976**	0.945	0.925	0.939	0.919	0.929	0.901	0.886	0.727	0.707
波形	0.781	0.872	0.881	**0.901**	0.781	0.817	0.834	0.758	0.783	0.859	0.802	0.812
Wilt	0.383	0.511	0.879	**0.891**	0.388	0.503	0.483	0.559	0.529	0.529	0.529	0.536
Wine	0.914	**1.000**	0.914	**1.000**	0.914	**1.000**	0.897	0.983	0.914	**1.000**	0.966	**1.000**
WPBC	0.499	0.506	0.548	0.484	0.499	0.506	0.526	0.575	0.612	0.521	0.585	**0.764**
AVG.	0.791	0.844	0.868	**0.889**	0.789	0.841	0.813	0.846	0.799	0.840	0.797	0.827
BEST	1	4	4	**19**	1	4	2	6	4	9	2	6

在32个数据集上，MSS-AE 在27个数据集（84%）上的表现优于 AE，MSS-PAE 在24个数据集（75%）上的表现优于 PAE，这证明了 MSS 的有效性。表4显示，MSS 显著提升了 AE 或 PAE 表现较差的数据集上的性能，例如 $InternetAds$、$Optidigits$、语音 和 $Vertebral$。这证明了 AE 类方法与 MSS 相结合是有效的。

4.3.3. 与基线方法的比较

为展示所提出的WNLL函数的有效性以及MSS方法的通用性，我们在表5中比较了PAE、5种基于自编码器的异常检测方法及其均值漂移版本的性能。本节中所有方法的参数均经过调优，以达到最优。所有MSS-版本的结果均使用 m 的最优值获得，而PAE和MSS-PAE的结果则使用 α的最佳值获得。表中结果明确显示，所有MSS-版本均优于其对应的原始版本，在平均AUROC上平均相对性能提升了20%，证明了MSS方法的通用性。在所有标准版本中，PAE的平均AUROC表现最佳，相较于排名第二的LCAE（0.868对0.813），相对提升了29%。在所有MSS-版本中，MSS-PAE的平均AUROC表现同样最佳，相较于排名第二的MSS-LCAE（0.889对0.846），相对提升了28%。这一结果表明，所提出的WNLL是有效的，尤其是在与MSS方法联合使用时。

为进一步证明所提出方法的优越性，表 6 汇总了所提出方法与 8 种非基于 AE 的异常检测方法的性能比较

表 6：所提方法与非自编码器基线的受试者工作特征曲线下面积结果 .

数据集	ECOD [24]	iForest [19]	LOF [10]	OCSVM [16]	HBOS [23]	MOD [12]	DIF [77]	DIP [76]	AE [25]	PAE *	MSS-AE *	MSS-PAE *
ALOI	0.540	0.544	0.785	0.543	0.494	0.769	0.546	**0.788**	0.563	0.564	0.563	0.706
Annthyroid	0.739	0.652	0.696	0.601	0.670	0.716	0.504	0.728	0.608	0.850	0.641	**0.893**
Arrhythmia	0.732	0.740	0.705	0.711	0.755	0.718	0.530	0.731	0.714	0.740	0.730	**0.762**
Breastw	0.992	0.989	0.696	0.995	0.983	0.987	0.975	0.993	0.972	0.995	0.994	**0.996**
Cardiotocography	0.802	0.706	0.628	0.767	0.619	0.598	0.689	0.556	0.811	0.870	0.888	**0.899**
玻璃	0.853	0.922	0.706	0.863	0.848	0.941	0.931	0.814	0.804	**0.980**	0.951	0.853
心脏病	0.687	0.721	0.777	0.730	0.774	0.758	0.723	0.845	0.813	0.871	0.869	**0.905**
InternetAds	0.722	0.764	0.632	0.653	0.736	0.660	0.750	0.696	0.706	0.697	0.801	**0.839**
电离层	0.789	0.895	0.931	0.885	0.892	0.932	0.903	0.906	0.971	**0.985**	0.978	0.978
字母	0.549	0.575	0.878	0.571	0.555	**0.905**	0.554	0.872	0.857	0.855	0.860	0.798
淋巴造影	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	0.914	**1.000**	0.971	0.914	**1.000**	**1.000**
乳腺摄影	**0.914**	0.864	0.805	0.863	0.811	0.844	0.776	0.843	0.893	0.891	0.885	0.901
MNIST	0.751	0.819	0.846	0.857	0.534	0.860	0.584	0.851	0.882	0.948	0.882	**0.949**
Musk	0.967	1.000	0.997	**1.000**	**1.000**	0.997	0.966	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
光学数字	0.615	0.799	0.542	0.557	0.915	0.494	0.649	0.835	0.760	0.936	0.981	**0.995**
PageBlocks	0.914	0.905	0.933	0.930	0.766	0.921	0.929	0.915	0.952	**0.962**	0.955	0.958
帕金森病	0.350	0.414	0.789	0.384	0.676	0.669	0.484	0.609	0.627	0.815	0.769	**0.861**
PenDigits	0.392	0.758	0.936	0.539	0.761	0.976	0.764	0.985	0.879	0.951	0.974	**0.993**
Pima	0.537	0.592	0.641	0.631	0.656	0.634	0.637	0.690	0.651	0.756	0.699	**0.757**
Satellite	0.608	0.724	0.566	0.694	0.806	0.695	0.748	0.696	0.750	0.855	0.789	**0.857**
Satimage-2	0.969	0.992	0.976	0.996	0.980	0.998	0.996	**0.998**	0.993	0.988	0.993	0.987
Shuttle	0.723	0.820	0.993	0.989	0.803	0.989	0.900	0.988	0.995	0.996	**1.000**	**1.000**
SpamBase	0.679	0.622	0.474	0.537	0.690	0.538	0.552	0.662	0.565	0.847	0.623	**0.886**
语音	0.507	0.527	0.698	0.506	0.518	**0.734**	0.582	0.627	0.551	0.597	0.651	0.673
Stamps	0.870	0.931	0.944	0.941	0.786	0.937	0.917	**0.961**	0.931	0.952	**0.961**	0.954
甲状腺	0.977	0.977	0.964	0.953	0.982	0.963	0.941	0.973	0.975	0.991	0.972	**0.994**
Vertebral	0.596	0.552	0.646	0.629	0.549	0.536	0.371	0.591	0.607	0.772	0.755	**0.841**
元音	0.632	0.767	0.939	0.792	0.682	**0.984**	0.760	0.979	0.937	0.976	0.958	0.945
波形	0.674	0.759	0.746	0.776	0.753	0.734	0.648	0.800	0.781	0.881	0.872	**0.901**
Wilt	0.350	0.437	0.690	0.295	0.323	0.667	0.478	0.640	0.383	0.879	0.511	**0.891**
Wine	0.897	0.862	0.931	0.879	0.966	0.931	0.897	0.879	0.914	0.914	**1.000**	**1.000**
WPBC	0.482	0.479	0.526	0.499	0.536	0.509	0.494	0.528	0.499	**0.548**	0.506	0.484
AVG.	0.713	0.753	0.782	0.736	0.744	0.800	0.722	0.812	0.791	0.868	0.844	**0.889**
BEST	2	1	1	2	2	4	0	5	1	5	5	**21**

。结果表明，所提出的 PAE、MSS-AE 和 MSS-PAE 在平均 AUROC 上优于所有基线方法。其中，MSS-PAE 在全部 32 个数据集中的 21 个（66%）上取得最佳表现，平均 AUROC 也最高，达到 0.889（相较于 8 种基线方法中表现最佳的 DIP，相对提升了 41%）。这些结果有力证明了结合特征提取（神经网络）、不确定性估计（概率分析）和局部结构信息（近邻图）的有效性。这表明，异常检测研究者应考虑采用多种工具来设计异常检测器，以适应各种复杂的现实场景。

5. 结论

本文通过两种新策略缓解了意外重建问题。首先，分析了基于自编码器的异常检测中的过度自信问题，以及偶然不确定性的作用。分析结果支持使用负对数似然训练自编码器，并使用加权负对数似然进行异常评分。针对各种离群点检测场景，本文提出了若干关于应用负对数似然和加权负对数似然的建议。其次，本文提出了MSS方法，通过引入局部关系信息，减少基于自编码器的异常检测方法误判为内点的数据。

在32个真实世界异常检测数据集上的实验结果表明，所提出的方法显著提升了自编码器在异常检测中的性能。具体而言，

加权负对数似然在适应不同异常检测应用时非常有效且灵活，而所提出的MSS方法也可有效应用于其他基于自编码器的异常检测方法。此外，结合了所提出两种方法的MSS-PAE在所有基线方法中表现最佳。因此，实验支持在实践中同时使用所提出的加权负对数似然和MSS，以获得最佳性能。我们认为，MSS方法可作为大多数基于自编码器的异常检测方法的通用插件。此外，负对数似然和加权负对数似然也可用于提升其他基于自编码器方法的性能，并可与其他损失函数结合使用。本文未对此进行评估，但未来研究可以加以验证。

参考文献

[1] F. E. Grubbs，样本中离群观测的检测程序，Technometrics 11 (1) (1969) 1–21。[2] M. Goldstein、S. Uchida，多变量数据无监督异常检测算法的比较评估，PLoS ONE 11 (4) (2016) e0152173。[3] J. Yang、S. Rahardja、S. Rahardja，点击欺诈检测：用于从变长用户行为时间序列中提取特征的 Hk-index，载于：IEEE 信号处理机器学习国际研讨会（MLSP）论文集，2022年，第 1–6 页。[4] P. Ghosh、A. Losalka、M. J. Black，利用高斯混合变分自编码器抵御对抗攻击，载于：AAAI 人工智能会议论文集，第 33 卷，2019年，第 541–548 页。[5] J. Yang、G. I. Choudhary、S. Rahardja、P. Franti，利用注意力熵对心搏间期时间序列进行分类，IEEE Transactions on Affective Computing 14 (1) (2023) 321–330。[6] N. Li、F. Chang，基于多变量高斯全卷积对抗自编码器的视频异常检测与定位，Neurocomputing 369 (2019) 92–105。[7] J. Yang、X. Tan、S. Rahardja，Mipo：如何利用表格数据离群点检测器检测轨迹离群点，Remote Sensing 14 (21) (2022) 5394。[8] R. Domingues、M. Filippone、P. Michiardi、J. Zouaoui，离群点检测算法的比较评估：实验与分析，Pattern Recognition 74 (2018) 406–421。[9] S. Ramaswamy、R. Rastogi、K. Shim，从大型数据集中挖掘离群点的高效算法，载于：ACM SIGMOD 数据管理国际会议论文集，2000年，第 427–438 页。[10] M. M. Breunig、H.-P. Kriegel、R. T. Ng、J. Sander，LOF：识别基于密度的局部离群点，载于：ACM SIGMOD 数据管理国际会议论文集，2000年，第 93–104 页。

[11] H. Du、S. Zhao、D. Zhang、J. Wu，基于聚类的局部离群点检测新方法，载于：IEEE 计算机通信会议研讨会（INFOCOM WKSHPS）论文集，2016年，第 802–811 页。[12] J. Yang、S. Rahardja、P. Fr¨anti，均值漂移离群点检测与过滤，Pattern Recognition 115 (2021) 107874。[13] X.-m. Tang、R.-x. Yuan、J. Chen，利用子空间学习和高斯混合模型检测能源分解中的离群点，International Journal of Control and Automation 8 (8) (2015) 161–170。[14] P. I. Dalatu、A. Fitrianto、A. Mustapha，线性与非线性回归模型在离群点检测中的比较研究，载于：软件计算与数据挖掘国际会议（SCDM）论文集，2017年，第 316–326 页。[15] L. J. Latecki、A. Lazarevic、D. Pokrajac，利用核密度函数进行离群点检测，载于：模式识别中的机器学习与数据挖掘国际研讨会（MLDM）论文集，2007年，第 61–75 页。[16] B. Sch¨olkopf、J. C. Platt、J. Shawe-Taylor、A. J. Smola、R. C. Williamson，估计高维分布的支撑集，Neural Computation 13 (7) (2001) 1443–1471。[17] D. M. Tax、R. P. Duin，支持向量数据描述，Machine Learning 54 (1) (2004) 45–66。[18] L. Ruff、R. A. Vandermeulen、N. G¨ornitz、A. Binder、E. M¨uller、K.-R. M¨uller、M. Kloft，深度半监督异常检测，arXiv 预印本 arXiv:1906.02694 (2019)。[19] F. T. Liu、K. M. Ting、Z.-H. Zhou，孤立森林，载于：IEEE 数据挖掘国际会议（ICDM）论文集，2008年，第 413–422 页。[20] X. Tan、J. Yang、S. Rahardja，用于离群点检测的稀疏随机投影孤立森林，Pattern Recognition Letters 163 (2022) 65–73。[21] T. Pevn`y，LODA：轻量级在线异常检测器，Machine Learning 102 (2) (2016) 275–304。[22] Y. Zhao、X. Hu、C. Cheng、C. Wang、C. Wan、W. Wang、J. Yang、H. Bai、Z. Li、C. Xiao 等，SUOD：加速大规模无监督异构离群点检测，机器学习系统论文集（MLSys）3 (2021) 463–478。[23] M. Goldstein、A. Dengel，基于直方图的离群分数（HBOS）：一种快速的无监督异常检测算法，KI-2012：海报与演示环节 9 (2012)。[24] Z. Li、Y. Zhao、X. Hu、N. Botta、C. Ionescu、G. Chen，ECOD：利用经验累积分布函数进行无监督离群点检测，IEEE Transactions on Knowledge and Data Engineering (2022)。[25] S. Hawkins、H. He、G. Williams、R. Baxter，利用复制器神经网络进行离群点检测，载于：数据仓储与知识发现国际会议（DaWaK）论文集，2002年，第 170–180 页。

[26] C. Zhou、R. C. Paffenroth，《利用鲁棒深度自编码器进行异常检测》，载于：ACM SIGKDD知识发现与数据挖掘国际会议（KDD）论文集，2017年，第665–674页。[27] J. An、S. Cho，《基于变分自编码器的异常检测：利用重构概率》，IE专题讲座 2 (1) (2015年) 1–18。[28] C.-H. Lai、D. Zou、G. Lerman，《用于无监督异常检测的鲁棒子空间恢复层》，载于：第八届国际学习表征会议（ICLR）论文集，2020年，第1–28页。[29] Y. Guo、X. Zhu、Z. Hu、Z. Zhan，《利用特征分解的自编码器进行无监督异常检测》，载于：机器学习与计算国际会议（ICMLC）论文集，2022年，第258–265页。[30] F. Angiulli、F. Fassetti、L. Ferragina，《Latent out：利用潜在空间分布的无监督深度异常检测方法》，Mach. Learn. (2022年) 1–27。[31] T. Schlegl、P. Seeb¨ock、S. M. Waldstein、U. Schmidt-Erfurth、G. Langs，《利用生成对抗网络引导标志物发现的无监督异常检测》，载于：医学图像处理国际会议（IPMI）论文集，2017年，第146–157页。[32] S. Akcay、A. Atapour-Abarghouei、T. P. Breckon，《Ganomaly：通过对抗训练进行半监督异常检测》，载于：亚洲计算机视觉会议（ACCV）论文集，2018年，第622–637页。[33] B. I. Ibrahim、D. C. Nicolae、A. Khan、S. I. Ali、A. Khattak，《基于VAE-GAN的零样本离群点检测》，载于：计算机科学与智能控制国际研讨会（ISCSIC）论文集，2020年，第1–5页。[34] Z. Yang、T. Zhang、I. S. Bozchalooi、E. Darve，《用于异常检测的记忆增强生成对抗网络》，IEEE Trans. Neural Netw. Learn. Syst. 33 (6) (2022年) 2324–2334。[35] R. Chalapathy、S. Chawla，《异常检测的深度学习：综述》，arXiv预印本 arXiv:1901.03407 (2019年)。[36] D. A. Nix、A. S. Weigend，《估计目标概率分布的均值和方差》，载于：1994年IEEE神经网络国际会议论文集（ICNN’94），第1卷，IEEE，1994年，第55–60页。[37] A. A. Pol、V. Berger、C. Germain、G. Cerminara、M. Pierini，《利用条件变分自编码器进行异常检测》，载于：2019年第18届IEEE机器学习与应用国际会议（ICMLA），IEEE，2019年，第1651–1657页。[38] Y. Mao、F.-F. Xue、R. Wang、J. Zhang、W.-S. Zheng、H. Liu，《利用不确定性预测自编码器检测胸部X射线图像中的异常》，载于：《医学图像计算与计算机辅助干预——MICCAI 2020：第23届国际会议，秘鲁利马，2020年10月4–8日，论文集，第VI部分23》，施普林格，2020年，第529–538页。

[39] H. Choi、E. Jang、A. A. Alemi，《WAIC，但为什么？用于鲁棒异常检测的生成式集成》，arXiv预印本 arXiv:1810.01392 (2018年)。[40] B. X. Yong、A. Brintrup，《具有不确定性量化功能的贝叶斯自编码器：迈向可信异常检测》，Expert Systems with Applications 209 (2022年) 118196。[41] N. Merrill、A. Eskandarian，《用于深度学习中鲁棒无监督异常检测的改进自编码器训练与评分》，IEEE Access 8 (2020年) 101824–101833。[42] I. Goodfellow 、 Y. Bengio 、 A. Courville，《深度学习》，MIT出版社，2016年。[43] S. Wold、K. Esbensen、P. Geladi，《主成分分析》，Chemom. Intell. Lab. Syst. 2 (1-3) (1987年) 37–52。[44] B. B. Thompson、R. J. Marks、J. J. Choi、M. A. El-Sharkawi、M.-Y. Huang、C. Bunje，《自编码器新颖性评估中的隐式学习》，载于：神经网络国际联合会议（IJCNN）论文集，第3卷，2002年，第2878–2883页。[45] B. Zong、Q. Song、M. R. Min、W. Cheng、C. Lumezanu、D. Cho、H. Chen，《用于无监督异常检测的深度自编码高斯混合模型》，载于：国际学习表征会议，2018年。[46] J. Chen、S. Sathe、C. Aggarwal、D. Turaga，《利用自编码器集成进行离群点检测》，载于：SIAM数据挖掘国际会议（SDM）论文集，2017年，第90–98页。[47] Y. Ishii、M. Takanashi，《利用稳健估计和自编码器进行低成本无监督离群点检测》，J. Inf. Process. 27 (2019年) 335–339。[48] D. Gong、L. Liu、V. Le、B. Saha、M. R. Mansour、S. Venkatesh、A. v. d. Hengel，《记忆正常性以检测异常：用于无监督异常检测的记忆增强深度自编码器》，载于：IEEE/CVF计算机视觉国际会议（ICCV）论文集，2019年，第1705–1714页。[49] Q. Yu、M. Kavitha、T. Kurita，《基于正交投影约束的自编码器框架可提升异常检测效果》，Neurocomputing 450 (2021年) 372–388。[50] B. Lakshminarayanan、A. Pritzel、C. Blundell，《利用深度集成进行简单且可扩展的预测不确定性估计》，神经信息处理系统进展 30 (2017年)。[51] E. Nalisnick、A. Matsukawa、Y. W. Teh、D. Gorur、B. Lakshminarayanan，《深度生成模型知道自己不知道什么吗？》，arXiv预印本 arXiv:1810.09136 (2018年)。[52] D. Hafner、D. Tran、T. Lillicrap、A. Irpan、J. Davidson，《用于函数不确定性的噪声对比先验》，载于：《人工智能中的不确定性》，机器学习研究论文集，2020年，第905–914页。[53] M. Sensoy、L. Kaplan、F. Cerutti、M. Saleki，《利用生成模型的不确定性感知深度分类器》，载于：AAAI人工智能大会论文集，第34卷，2020年，第5620–5627页。

[54] X.-Y. Zhang、G.-S. Xie、X. Li、T. Mei、C.-L. Liu，《拒绝学习综述》，IEEE汇刊111 (2) (2023) 185–215。[55] N. St˚ahl、G. Falkman、A. Karlsson、G. Mathiason，《深度学习中不确定性量化的评估》，载于：知识库系统中的不确定性信息处理与管理国际会议，施普林格，2020年，第556–568页。[56] A. Der Kiureghian、O. Ditlevsen，《偶然不确定性还是认知不确定性？这重要吗？》，《结构安全》31 (2) (2009) 105–112。[57] A. Kendall、Y. Gal，《计算机视觉的贝叶斯深度学习需要哪些不确定性？》，《神经信息处理系统进展》30 (2017)。[58] J. Gawlikowski、C. R. N. Tassi、M. Ali、J. Lee、M. Humt、J. Feng、A. Kruspe、R. Triebel、P. Jung、R. Roscher 等，《深度神经网络中不确定性的综述》，《人工智能评论》56 (增刊1) (2023) 1513–1589。[59] C. Blundell、J. Cornebise、K. Kavukcuoglu、D. Wierstra，《神经网络中的权重不确定性》，载于：国际机器学习会议，机器学习研究论文集，2015年，第1613–1622页。[60] Y. Gal、Z. Ghahramani，《将丢弃法作为贝叶斯近似：表示深度学习中的模型不确定性》，载于：国际机器学习会议，机器学习研究论文集，2016年，第1050–1059页。[61] C. Nemeth、P. Fearnhead，《随机梯度马尔可夫链蒙特卡罗》，《美国统计学会杂志》116 (533) (2021) 433–450。[62] T. Pearce、F. Leibfried、A. Brintrup，《神经网络中的不确定性：近似贝叶斯集成》，载于：国际人工智能与统计会议，机器学习研究论文集，2020年，第234–244页。[63] A. Legrand、H. Trannois、A. Cournier，《利用自编码器神经网络中的不确定性进行异常检测》，载于：2019年IEEE第二届人工智能与知识工程国际会议（AIKE），IEEE，2019年，第32–35页。[64] C. Baur、B. Wiestler、S. Albarqouni、N. Navab，《用于高分辨率脑部磁共振成像无监督高信号异常检测的贝叶斯跳跃自编码器》，载于：2020年IEEE第17届生物医学成像国际研讨会（ISBI），IEEE，2020年，第1905–1909页。[65] E. Daxberger、J. M. Hern´andez-Lobato，《用于无监督分布外检测的贝叶斯变分自编码器》，arXiv预印本 arXiv:1912.05651 (2019)。[66] S. Park、G. Adosoglou、P. M. Pardalos，《解读变分自编码器的率失真并利用模型不确定性进行异常检测》，《数学与人工智能年刊》90 (7) (2022) 735–752。[67] B. X. Yong、T. Pearce、A. Brintrup，《贝叶斯自编码器：分析并修正用于分布外检测的伯努利似然》，arXiv预印本 arXiv:2107.13304 (2021)。

[68] K. Fukunaga、L. Hostetler, 《密度函数梯度的估计及其在模式识别中的应用》, IEEE信息理论汇刊 21 (1) (1975年) 32–40。[69] Y. Cheng, 《均值漂移、模态搜索与聚类》, IEEE模式分析与机器智能汇刊 17 (8) (1995) 790–799。[70] D. Comaniciu、P. Meer, 《均值漂移：一种稳健的特征空间分析方法》, IEEE模式分析与机器智能汇刊 24 (5) (2002年) 603–619。[71] R. T. Collins, 《通过尺度空间进行均值漂移斑块跟踪》, 载于：IEEE计算机视觉与模式识别国际会议论文集（CVPR）, 第2卷, 2003年, 第II–234页。[72] J. Yang, 《离群点检测技术》, 博士论文, 东芬兰大学（2020年）。[73] J. Yang、Y. Chen、S. Rahardja, 《用于改进异常检测器的邻域代表》, 《信息科学》（2022年）。[74] L. Van der Maaten、G. Hinton, 《使用t-SNE可视化数据》, 《机器学习研究杂志》9 (11) (2008年)。[75] S. Kandanaarachchi、M. A. Mu˜noz、R. J. Hyndman、K. Smith-Miles, 《无监督离群点检测中的归一化与算法选择》, 《数据挖掘与知识发现》34 (2) (2020年) 309–354。[76] J. Zhao、F. Deng、J. Zhu、J. Chen, 《搜索密度递增路径以寻找局部密度峰值，实现无监督异常检测》, 《IEEE大数据汇刊》（2023年）。[77] H. Xu、G. Pang、Y. Wang、Y. Wang, 《用于异常检测的深度孤立森林》, IEEE知识与数据工程汇刊（2023年）。[78] Y. Zhao、Z. Nasrullah、Z. Li, 《PyOD：用于可扩展离群点检测的Python工具箱》, 《机器学习研究杂志》20 (2019年) 1–7。[79] P. Ramachandran、B. Zoph、Q. V. Le, 《搜索激活函数》, arXiv预印本 arXiv:1710.05941 (2017年)。