

基于条件归一化流的并行时空注意力的无监督异常检测模型

李新宇*, 于晋伟, 杨卫华

太原理工大学数学学院, 山西 太原

收稿日期: 2026年3月2日; 录用日期: 2026年3月26日; 发布日期: 2026年4月7日

摘要

预测性维护作为提升工业生产效率的重要环节, 其核心挑战在于准确识别故障并发出早期预警。在实践中, 工业生产中记录的数据具有两大特征: 数据规模庞大与标签缺失。因此需建立针对多元时间序列的无监督异常检测模型。本文提出一种新型无监督异常检测模型——基于条件归一化流的并行时空注意力无监督异常检测(PAF)模型。该模型基于“异常数据远少于正常数据”的假设, 结合高斯分布理论, 将异常数据定位于低概率密度区域。基于此假设, 采用条件归一化流变换数据分布并进行密度估计, 进而设定异常评分实现有效判别。该模型的核心特征在于运用并行注意力结构捕捉原始数据的时空信息, 作为条件归一化流的前置信息条件。相较于现有相关研究, PAF展现出更优异的检测性能。最终, 我们通过两个公开数据集验证了PAF模型的先进性和有效性。

关键词

时空注意力, 条件归一化流, 异常检测, 无监督学习

Conditional Normalization Flow-Based Unsupervised Anomaly Detection Model with Parallel Spatio-Temporal Attention

Xinyu Li*, Jinwei Yu, Weihua Yang

College of Mathematics, Taiyuan University of Technology, Taiyuan Shanxi

Received: March 2, 2026; accepted: March 26, 2026; published: April 7, 2026

Abstract

Predictive maintenance, as a crucial component for improving industrial production efficiency, faces

*通讯作者。

文章引用: 李新宇, 于晋伟, 杨卫华. 基于条件归一化流的并行时空注意力的无监督异常检测模型[J]. 应用数学进展, 2026, 15(4): 170-181. DOI: 10.12677/aam.2026.154147

a core challenge in accurately identifying faults and issuing early warnings. In practice, the data recorded in industrial production have two main characteristics: multivariate data and missing labels. Therefore, there is an urgent need to establish unsupervised anomaly detection models for multivariate time series. This paper proposes a novel unsupervised anomaly detection model—the Parallel Spatiotemporal Attention Flow-based (PAF) model. This model is based on the assumption that “anomalous data are far less frequent than normal data” and, combined with Gaussian distribution theory, positions anomalous data in low-probability density regions. Based on this assumption, conditional normalizing flows are used to transform the data distribution and perform density estimation, thereby setting anomaly scores to achieve effective detection. The core feature of this model is the use of a parallel attention structure to capture the spatiotemporal conditions of the original data as prior information for the normalizing flow. Compared with existing related studies, PAF demonstrates superior detection performance. Finally, the effectiveness and advancement of the PAF model are validated using two publicly available datasets.

Keywords

Spatio-Temporal Attention, Conditional Normalization Flow, Anomaly Detection, Unsupervised Learning

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

在实际工业生产过程中，要使设备长期高效运行，需要良好的保养和维护。

预测性维护(Predictive Maintenance)是一种基于状态的维护方式，在机器运行过程中，通过对其主要部件进行定期状态监测和故障诊断，确定设备状态，预测设备状态的未来发展趋势，并根据设备状态发展趋势和可能的故障模式，预先制定预测性维护计划，确定何时、何地、如何以及需要哪些技术和维护措施进行维修。基于设备状态发展趋势及可能的故障模式，预先制定预测性维护方案，确定设备维修的时间节点、内容范围、实施方式及所需技术与物资保障。预测性维护融合了设备状态监测、故障诊断、故障预测、维护决策支持与维护活动，是新兴的维护方法。

预测性维护的主要难点在于故障状态的精准诊断。对此，众多学者已作出卓有成效的贡献。

在异常检测领域，多位学者取得了显著成果。Feng [1]等人展示了元学习在异常检测领域的应用。Zhang [2]等人针对样本稀疏且运行条件多变的工业故障诊断场景，提出基于领域差异引导的比较特征学习方法，有效提升了诊断性能。Zhang [3]等人探索了基于先验知识增强的自监督特征学习技术，用于稀疏标注数据下的机器故障诊断，实现了高效故障识别。Li [4]等人提出可解释的图小波去噪网络用于智能故障诊断，有效消除噪声并提供可解释的诊断结果。Xu [5]等人研究了物理约束变分神经网络，通过整合物理约束评估外啮合齿轮泵磨损状态，融合物理模型与数据驱动模型以提升磨损状态评估的精度与鲁棒性。该方法展现出卓越的准确性与鲁棒性，尤其在面对缺失数据、噪声及不同强度异常值时表现突出。

流模型作为生成模型的子类，以其通过学习数据底层概率分布来生成新样本的能力而闻名。近年来，这类模型取得了显著进展，并在多个领域得到广泛应用。Durkan [6]等人提出基于样条函数的流式生成模型“神经样条流式生成”，在密度估计任务中展现出显著效能。Ho [7]等人提出的 Flow++模型通过融合变分去量化和架构设计，实现了流生成模型的增强，在图像生成任务中展现出良好前景。Grathwohl [8]等人开发的 FFIORD 模型基于自由形式连续动力学，采用常微分方程实现高效密度估计与生成，构建了可

扩展的可逆生成模型。Dinh [9]等人提出实数非体积保持变换(RealNVP), 作为概率模型无监督学习的可逆可学习变换, 可精确计算对数似然、采样及潜在变量推断, 并成功应用于自然图像建模。流模型在数据结构转换与密度估计任务中持续展现卓越性能, 使其成为生成式建模领域的重要模型。

从多元时间序列中获取时空信息是关键的研究领域, 其核心在于有效捕捉并利用数据在时间与空间维度上的特性。近年来, 随着数据量的激增和计算能力的提升, 学者们提出了多种创新方法来应对这一挑战。Kim [10]等人提出了一种用于时间序列分类的概率插值技术, 该技术能有效处理缺失数据并提升分类模型的准确性。Zhou [11]等人探讨了深度潜在状态空间模型在时间序列生成中的应用, 突显其捕捉复杂时间序列数据动态特性的优势。Raghu [12]等人提出基于多维度自监督学习的临床时间序列数据序列分析方法, 实现了对无标签数据集的高效特征学习。Li [13]等人提出面向原型的多变量时间序列无监督异常检测方法, 在无监督环境下有效实现了异常识别。Park [14]等人探索了神经随机微分博弈在时间序列分析中的应用, 并提出了一种创新的时间序列建模方法。Lai [15]等人提出了一种基于多维自注意力机制的剩余寿命预测方法, 通过多维自注意力机制自动捕捉时间序列中不同维度之间的复杂依赖关系。该方法在多个工业数据集上表现出优异的预测性能与鲁棒性, 为智能运维与预测性维护提供了新的技术路径。这些成果为解决时空数据固有的复杂性做出了重要贡献。

在现有研究中, Dai [16]等人和 Zhou [17]等人在将流模型用于多元时间序列异常检测方面取得了丰硕成果。Dai [16]等人提出了一种图增强的归一化流方法 GANF [16], 并在密度估计、异常检测和时间序列分布漂移识别中展示了其有效性。Zhou [17]在 GANF [16]基础上引入实体感知归一化流, 生成实体特异性密度估计, 相较先前工作实现了更优的区分能力。本研究思路部分借鉴 GANF [16]与 MTGFlow [17], 并对两者进行了针对性改进。

本文的主要贡献总结如下:

(1) 提出一种获取时空信息的新方法: 建立并行型结构以获取时空信息作为条件归一化流模型的条件信息。该并行结构相较于已有研究中的串联结构更能有效保留时空信息, 多头注意力机制在不同维度获取时空信息的同时有效缩小了模型规模。该并行结构支持跨维度数据的同步处理, 增强了模型捕捉数据中复杂模式与相互关联关系的能力。通过并行处理, 模型能更有效地保留原始数据的特征信息, 减少干扰, 从而提升异常检测与预测任务的性能表现。

(2) 条件归一化流与时空特征学习: 该模型采用条件归一化流与时空特征学习方法, 通过整合数据分布特征与多头注意力机制进行预测, 从而提升异常检测的有效性。该模型包含两个核心模块: 可完成高斯分布和实际分布转化的分布转换模块, 以及基于多头注意力机制的并行式时空特征学习模块。此设计能更有效地捕捉和利用数据中的时空信息, 从而提升异常检测的准确性。

本文剩余部分结构如下: PAF 模型与方法章节详细阐述了提出的本研究所提出的 PAF 模型架构及其基本原理。数值实验章节通过对比其他学者的研究, 验证了 PAF 并行式结构的有效性。结论章节总结了本文完成的工作。

2. PAF 模型与方法

在实际应用中, 现有数据往往呈现高斯分布: 正常数据占大多数而异常数据稀少或极难获取。准确标注数据标签的过程既耗时又耗费资源。因此, 开发高效的无监督学习方法对于应对现实世界挑战至关重要。基于高斯分布的理论基础, 我们提出如下定义:

定义 1 (高斯分布假设):

在设备正常运行条件下, 绝大多数采集到的检测数据代表正常状态, 仅有极小比例可能表明潜在故障。通过利用高斯分布特性, 当数据段 x^c 表现出异常行为时, 将其识别为低概率区域; 当其正常运行时,

则归入高概率区域。该区分机制构成了对区段 x^c 评分的基准——此评分越高表明该区段处于异常区域的可能性越大。预设阈值 θ' 作为判定区段 x^c 属于正常或异常状态的标准。

根据定义 1, 我们提出一种新型异常检测框架, 命名为 PAF。该框架利用高斯分布特性, 实现高效的无监督异常检测。与 GANF [16]和 MTGFlow [17]采用的串联式结构进行时空特征提取不同, PAF 引入并行结构且结合注意力机制以增强时空信息的获取能力。这种并行方法不仅更有效地保持了时空信息的完整性, 还减少了不同维度信息之间的干扰。对比实验结果详见数值实验。整体框架如图 1 所示。

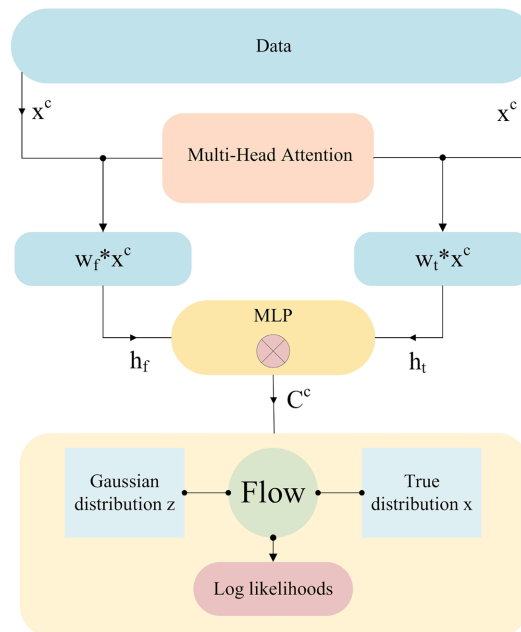


Figure 1. PAF flowchart
图 1. PAF 流程图

2.1. PAF 模型简介

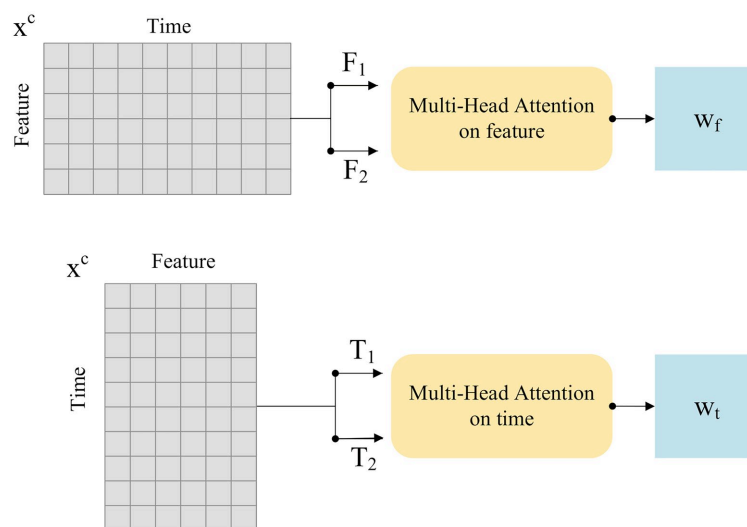


Figure 2. Multi-head attention mechanism acquires weights across different dimensions
图 2. 多头注意力机制获取不同维度的权重

在大小为 T 的滑动窗口内，多元时间序列 x^c 分两路，结合多头注意力机制。分别获取特征间空间结构信息权重 w_f 及时间序列权重 w_t 。具体结构如图 2 所示。分别捕捉特征间动态关联的隐藏状态 h_f 与时序信息的隐藏状态 h_t 。时空条件 C^c 通过 h_f 与 h_t 的点积运算生成，随后经多层感知器(MLP)处理。最终，用于辅助实体感知归一化流模型，通过生成时间序列分布的实体特异性密度估计值实现优化。

2.2. 多头注意力机制

缩放点积注意力机制的核心概念[18]，包含以下步骤：给定一个查询向量(Q)和一组键值对(K - V pairs)，该机制计算查询向量与每个键值对的点积。随后通过 softmax 函数将这些点积转换为概率分布，从而评估每个值的相对重要性。输出值由这些概率值的加权求和得出，权重 softmax 函数赋予的概率值。该过程的数学表达式如下：

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V, \quad (1)$$

其中 Q ， K 和 V 分别对应查询矩阵、键矩阵和值矩阵。键矩阵的维度记为 d_k 。当 d_k 较大时，点积的绝对值也趋于增大，这会将 softmax 函数推入梯度最小的区域，从而导致梯度消失问题。为解决此问题，引入缩放因子 $1/\sqrt{d_k}$ ，该因子有助于更有效地分配注意力权重，从而提升模型泛化能力。

在多头注意力架构[18]中，通过将多个缩放后的点积注意力结果进行拼接，再利用权重矩阵将输出投影回单头维度。该过程的公式定义如下：

$$\begin{aligned} \text{MultiHead}(Q, K, V) &= \text{Concat}(\text{head}_1, \dots, \text{head}_h)W^O \\ \text{where } \text{head}_i &= \text{Attention}(QW_i^Q, KW_i^K, VW_i^V), \end{aligned} \quad (2)$$

其中 h 表示头部数量， W^Q 、 W^K 、 W^V 和 W^O 均为可训练参数矩阵。多头注意力框架中的每个头部均实现缩放点积注意力机制。本研究采用多头注意力机制，从具有时间演变特征的多变量时间序列输入数据中进行学习。该机制能够同步捕捉数据中的动态相互依赖关系，使模型能够更精细地理解数据中的时空关系，从而增强其处理和分析多变量时间序列的能力。

2.3. 并行时空结构

多元时间序列数据蕴含着丰富的复杂信息。首先，作为时间序列数据，其本质上具有时间特征。其次，数据中由不同监测站点代表的独特元素之间存在相互依赖关系，而非独立运作。揭示这些不同元素之间的相互关系，准确提取时空信息对于开发高效实用的模型至关重要。

在处理多元时空数据时，信息融合的拓扑结构对原始特征的保真度具有决定性影响。并行时空结构相较于串联式结构，在保留原始数据特征方面展现出显著优势。这主要源于并行架构的设计哲学：它允许不同模态的数据流在独立的通道中进行处理，避免了因强制排序而引入的跨模态干扰[19]。这种隔离机制确保了各原始数据流的内生统计特性与独立变化模式在融合前期得以完整保存。更重要的是，并行结构能够通过设计独立的子空间或状态模型，分别捕捉时空数据中并发的因果与非因果依赖关系，从而避免了串联结构中因前级处理误差向后级累积放大所导致的特征畸变[20]。

利用多头注意力架构在不同的维度捕捉多元时间序列数据中的空间特征和时序特征，初始阶段，原始数据分两路，在空间(不同特征间)维度应用多头注意力机制处理，获取空间结构信息权重 w_f 。在时间维度应用多头注意力机制处理，获取时间序列权重 w_t 。通过该方法提取的隐藏状态 h_f 和 h_t 分别封装了时间信息与空间信息：

$$h_f = x^c * w_f, \quad (3)$$

$$h_t = x^c * w_t, \quad (4)$$

获取的信息被整合以获得最终的时空信息 C^c :

$$C^c = \text{ReLU}(h_f * h_t) W_1, \quad (5)$$

其中 W_1 用于增强 C^c 的表达能力。在多元时空信息融合框架中, 采用点积操作进行特征交互具有显著的优越性。首先, 点积融合能够高效地实现高维特征的对齐与相似性度量。与拼接或加法等简单融合策略不同, 点积操作在高维特征空间中充当了一种可微分的注意力机制, 能够自适应地计算时间信息与空间信息在对应通道上的响应相关性, 从而精准地捕捉时空维度上的深层依赖关系[21]。

2.4. 条件归一化流

根据定义, 需建立处理后的数据与高斯分布假设之间的关联。为此, 我们选用了称为条件归一化流的生成模型, 以实现将数据的潜在分布映射至已知分布(即高斯分布)的过程。

归一化流作为一种无监督密度估计技术, 通过一系列可逆仿射变换将原始分布重塑为任意目标分布。原始数据分布 \mathcal{X} 的密度常难以直接确定, 而将其转化为已知分布可使问题更易处理。例如, 当已知原始分布样本 $x \in \mathcal{R}^D \sim \mathcal{X}$ 与目标分布样本 $z \in \mathcal{R}^D \sim \mathcal{Z}$ 时, 目标在于寻找一种双射且可逆的变换, 实现 \mathcal{X} 与 \mathcal{Z} 之间的互转关系[16] [17]。该方法可利用高斯分布的理论, 对原始数据分布进行分析与解释。

基于变量替换原理, 可得:

$$P_x(x) = P_z(z) \left| \det \frac{\partial f_\theta}{\partial x^T} \right|. \quad (6)$$

利用映射函数的可逆性及雅可比行列式 $\left| \det \partial f_\theta / \partial x^T \right|$ 的可计算性, 流模型的目标是实现 $\hat{z} = z$, 其中 $\hat{z} = f_\theta(x)$ 。这意味着模型旨在将输入数据 x 转换为符合高斯分布的目标分布 z 。

当引入时空信息等补充条件 C 作为输入时, 流模型可提升其密度估计性能[16] [17]。此类整合附加条件输入的流模型称为条件归一化流, 对应映射表达为 $z = f_\theta(x|C)$ 。通过最大似然估计(MLE)优化 f_θ 以更新参数 θ :

$$\theta^* = \arg \max_{\theta} \left(\log(P_z(f_\theta(x|C))) + \log \left(\left| \det \frac{\partial f_\theta}{\partial x^T} \right| \right) \right). \quad (7)$$

该优化过程使模型预测与观测数据趋于一致, 从而提升模型对潜在数据分布的精确估计能力。

2.5. PAF 的联合优化

该模型融合多元时间序列中的时空依赖关系。通过利用推导出的时空条件, 模型能精确预测时间序列的密度分布, 从而实现实体感知型归一化流。为降低各模块收敛至局部最优解的风险, 我们采用联合优化方法处理所有模块[17], 显著提升了模型整体性能。完整参数集 W^* 通过最大似然估计法进行求解, 确保模型参数优化后能最佳拟合观测数据。

$$\begin{aligned} W^* &= \arg \max_W \log(P_x(x)) \\ &\approx \arg \max_W \frac{1}{NK} \sum_{c=1}^N \sum_{k=1}^N \log \left(P_{z_k} \left(f_\theta(x_k^c | C_k^c) \right) \left| \det \frac{\partial f_\theta}{\partial x_k^{cT}} \right| \right) \\ &\approx \arg \max_W \frac{1}{NK} \sum_{c=1}^N \sum_{k=1}^N -\frac{1}{2} \|\hat{z}_k^c - \mu_k\|_2^2 + \log \left| \det \frac{\partial f_\theta}{\partial x_k^{cT}} \right|, \end{aligned} \quad (8)$$

其中 N 为窗口总数。

2.6. 异常评分

通过将窗口序列 x_k^c 作为输入，可估算序列中所有实体的密度。异常得分 S_c 源自这些实体负对数似然值的均值[17]。更高异常得分表明 x_k^c 位于低密度区域，数据异常的概率更高。该得分计算方式如下：

$$S_c = -\frac{1}{K} \sum_{k=1}^K \log(P_{x_k}(x_k^c)). \quad (9)$$

3. 数值实验

数值实验的主要目标是验证我们提出的并行时空注意力机制能否有效增强条件归一化流模型的异常检测能力。本实验旨在评估引入该并行时空注意力机制后，数据集内的异常检测性能是否得到显著提升。实验结果将揭示本方法在增强模型识别异常模式能力方面的有效性。

3.1. 实验数据

为验证我们提出的并行时空注意力机制能否有效增强条件归一化流模型的异常检测能力，实验所使用的数据在同类研究中也使用。在使用相同数据的情况下，所作出的对比结果更具有参考价值 and 实际意义。

本实验所使用的公开数据集如下：

PSM [22] (聚合服务器指标)：该数据集汇总了 eBay 管理的多个服务器节点的性能指标。通过该数据集，可识别出表明服务器性能潜在问题或安全漏洞的异常值。

MSL [23] (火星科学实验室漫游车)：该数据集源自火星科学实验室漫游车，具体为好奇号漫游车。其中包含遥测数据，用于检测漫游车在火星任务期间运行参数中的异常情况。

3.1.1. 数据集划分

在实验设置中，遵循 GANF [16]的数据集划分方案，将原始测试数据按 60%训练集、20%验证集、20%测试集进行划分。训练数据用于模型训练，验证数据用于超参数调优，测试数据用于评估模型性能。其余数据集则采用 60%训练集 40%测试集的划分比例。

3.1.2. 预处理

去除缺失值，并将数据归一化至[0, 1]区间。随后将数据划分为固定长度序列，采用 60 的窗口大小与 10 的步长。窗口大小决定每个序列的时间步数，步长则决定序列间的间隔步长。最终数据被输入模型进行训练与评估。

3.1.3. 实现细节

所有数据集均采用窗口大小 60、步长 10 的配置。其他具体参数与 MTGFlow [17]中的参数保持一致。所有实验均在 NVIDIA GeForce RTX 4060 Laptop GPU 8GB 上使用 PyTorch 2.3.0 + cu121 运行，执行 300 个 epoch。

3.2. 评价指标

为准确评估模型性能并便于比较，我们采用了同类研究中(GANF [16], MTGFlow [17])均使用到的评价指标：

AUROC (接收者操作特征曲线下面积)：该指标适用于评估模型在不同阈值设置下区分正常与异常窗口的整体效能。

AUPRC (精确率召回率曲线下面积): 该指标在数据失衡的异常检测场景中尤为有效, 可衡量模型在各决策阈值下捕捉真实异常的同时最小化误报的能力。

这些指标能全面评估模型在不同运行条件和阈值设置下的检测能力。

3.3. 对比实验

我们将本方法与已发表的半监督和无监督方法进行比较: 半监督方法包括: DeepSAD [24]。无监督方法包括: DROCC [25]、USAD [26]、GANF [16]、MTGFlow [17]。

我们通过与五种已发表方法进行比较, 验证了所提方法的性能优势。采用 AUROC 和 AUPRC 指标量化评估效果, 具体详见表 1 和表 2, 图 3 和图 5。括号内数值表示不同初始化种子间的方差。最优结果以粗体标注。

Table 1. AUROC

表 1. AUROC

模型	年份	AUROC	
		PSM	MSL
DROCC [25]	2020	74.3 (± 2.0)	53.4 (± 1.6)
DeepSAD [24]	2019	73.2 (± 3.3)	61.6 (± 0.6)
USAD [26]	2020	78.0 (± 0.2)	57.0 (± 0.1)
GANF [16]	2022	81.8 (± 1.5)	64.5 (± 1.9)
MTGFlow [17]	2024	85.7 (± 1.5)	67.2 (± 1.7)
PAF	Ours	87.2 (± 1.3)	69.7 (± 0.9)

Table 2. AUPRC

表 2. AUPRC

模型	年份	AUPRC	
		PSM	MSL
DROCC [25]	2020	60.7 (± 11.4)	13.2 (± 0.9)
DeepSAD [24]	2019	66.7 (± 10.8)	26.3 (± 1.7)
USAD [26]	2020	57.9 (± 3.6)	31.3 (± 0.0)
GANF [16]	2022	73.8 (± 4.7)	31.1 (± 0.2)
MTGFlow [17]	2024	76.2 (± 4.8)	31.1 (± 2.6)
PAF	Ours	80.3 (± 1.7)	31.8 (± 0.6)

在 PSM 数据集上, PAF 取得 AUROC 分数 87.2 (较现有最佳结果高出 1.3 个百分点)及 AUPRC 分数 80.3 (较现有最优结果高出 4.1 个百分点)。在 MSL 数据集上, PAUD 获得 69.7 的 AUROC 分数(较现有最优解高出 2.5 个百分点)及 31.8 的 AUPRC 分数(较现有最优解高出 0.7 个百分点)。对比分析表明, 本方法在异常检测任务中始终优于已有方法, 验证了其稳定性和有效性。

在模型结构上, 本研究所提出的 PAF 在不同评价指标中均优于采用了串联式结构的同类研究 MTGFlow [17], 详见表 1 和表 2, 图 3 和图 4。说明并行时空结构在保留原始数据特征方面优于已有的串联式结构。证明了改进的并行时空结构的有效性。

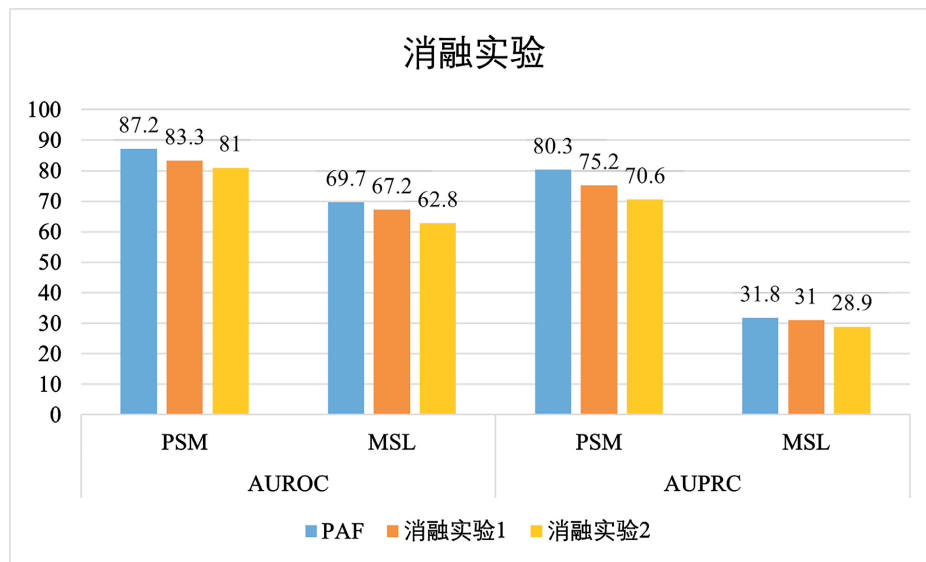


Figure 3. AUROC

图 3. AUROC

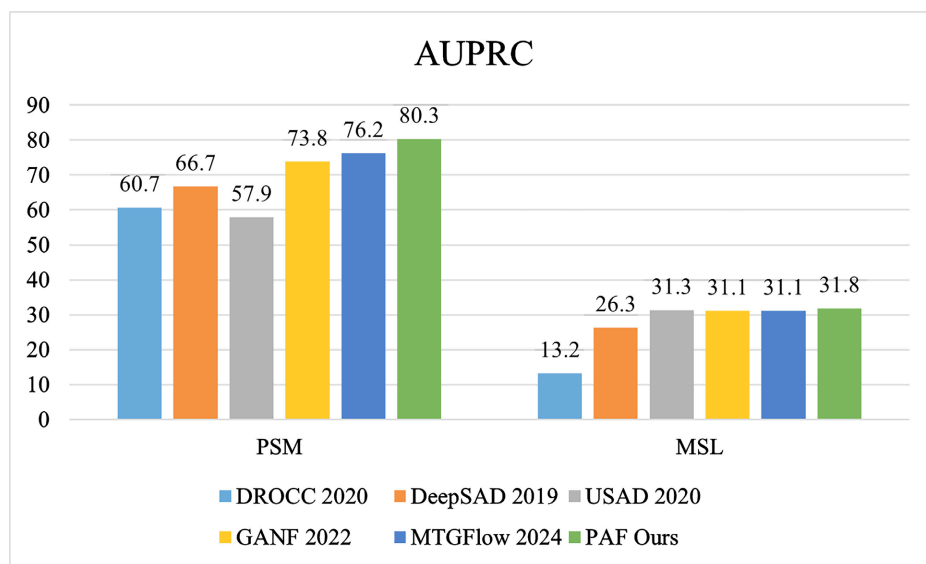


Figure 4. AUPRC

图 4. AUPRC

在模型运行时间上，同类研究 MTGFlow [17] 的训练时间约每 55 个 epoch 花费 1 小时，PAF 在信息获取结构作出针对性改进且保持其余参数不变的情况下，PAF 的整体训练时间缩减到了约每 94 个 epoch 花费 1 小时，模型判别精度有一定的提升，在训练时间上大幅下降。证明了改进的并行时空结构在模型规模上更为精简。

3.4. 消融实验

为验证时空并行模型的各部分有效性，我们设置了消融实验进行验证。消融实验分为两组：消融实验 1 (去除空间注意力模块)，消融实验 2 (去除时序注意力模块)。对比结果详见表 3 和图 5，最佳结果以粗体标注。

Table 3. Ablation experiment
表 3. 消融实验

模型	AUROC		AUPRC	
	PSM	MSL	PSM	MSL
PAF	87.2 (± 1.3)	69.7 (± 0.9)	80.3 (± 1.7)	31.8 (± 0.6)
消融实验 1	83.3 (± 1.8)	67.2 (± 1.1)	75.2 (± 2.8)	31.0 (± 0.5)
消融实验 2	81.0 (± 1.4)	62.8 (± 0.7)	70.6 (± 1.9)	28.9 (± 1.1)

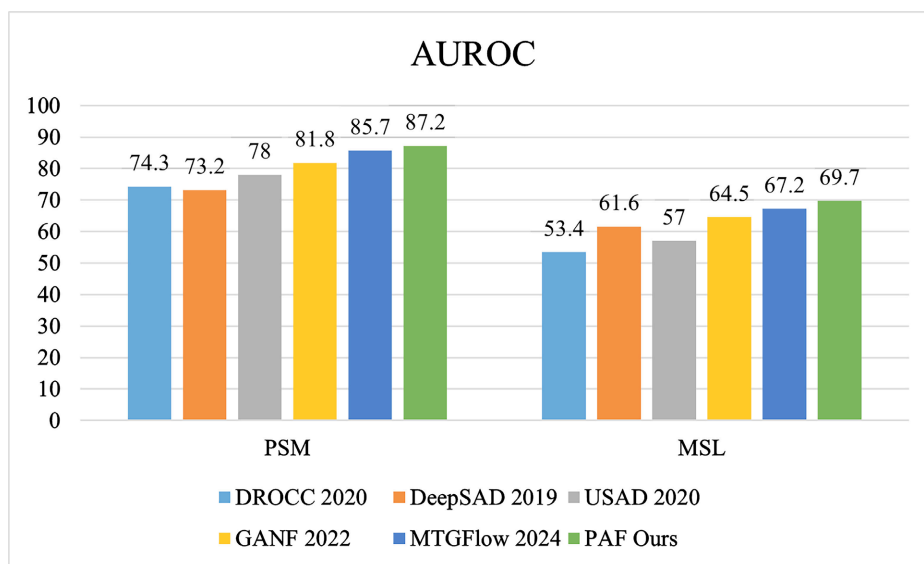


Figure 5. Ablation experiment
图 5. 消融实验

无论缺失空间注意力模块还是时序注意力模块，均导致模型判别精度下降，但程度各异。具体而言：去除空间注意力模块使 PSM 和 MSL 数据集的 AUROC 值分别降低 3.9 和 2.5 个百分点，AUPRC 值分别降低 5.1 和 0.8 个百分点。去除时序注意力模块使 PSM 和 MSL 数据集的 AUROC 值分别降低 6.2 和 6.9 个百分点，AUPRC 值分别降低 9.7 和 2.9 个百分点。消融实验结果表明，综合多维信息的时空结构能够捕捉更丰富的信息，从而有助于构建更高效精准的模型。

3.5. 实验结论

对比实验和消融实验表明，本研究提出的并行式时空结构在保留时空信息方面优于已有研究的串联式时空结构。该并行结构最大限度地减少了不同信息组件之间的干扰，从而提升了模型性能，同时使模型相较于已有研究更加精简。跨数据集的比较结果进一步验证了 PAF 模型的稳定性和有效性。对于多元时间序列而言，时空信息均不可或缺。

4. 结论

在本研究中，我们提出了一种新型异常检测方法 PAF，该方法在无标签条件下仅利用正常运行状态的数据即可工作。我们采用条件归一化流实现对多元时间序列数据的异常检测与隐式故障预警。PAF 的卓越性能主要归功于多头注意力机制的融合，在保留获取时空信息功能的同时减小了模型规模，以及并行结构从多元时间序列数据中提取的全面时空信息，这些要素共同为条件归一化流提供了高效的前置条

件信息。未来我们将致力于优化模型，以提升其在更广泛实际场景中的适用性与实用价值。

基金项目

本文所述工作得到山西省重大研究项目(编号：202202020101006)的支持。

参考文献

- [1] Feng, Y., Chen, J., Xie, J., Zhang, T., Lv, H. and Pan, T. (2022) Meta-Learning as a Promising Approach for Few-Shot Cross-Domain Fault Diagnosis: Algorithms, Applications, and Prospects. *Knowledge-Based Systems*, **235**, Article ID: 107646. <https://doi.org/10.1016/j.knosys.2021.107646>
- [2] Zhang, T., Chen, J., Liu, S. and Liu, Z. (2023) Domain Discrepancy-Guided Contrastive Feature Learning for Few-Shot Industrial Fault Diagnosis under Variable Working Conditions. *IEEE Transactions on Industrial Informatics*, **19**, 10277-10287. <https://doi.org/10.1109/tii.2023.3240921>
- [3] Zhang, T., Chen, J., He, S. and Zhou, Z. (2022) Prior Knowledge-Augmented Self-Supervised Feature Learning for Few-Shot Intelligent Fault Diagnosis of Machines. *IEEE Transactions on Industrial Electronics*, **69**, 10573-10584. <https://doi.org/10.1109/tie.2022.3140403>
- [4] Li, T., Sun, C., Li, S., Wang, Z., Chen, X. and Yan, R. (2022) Explainable Graph Wavelet Denoising Network for Intelligent Fault Diagnosis. *IEEE Transactions on Neural Networks and Learning Systems*, **35**, 8535-8548. <https://doi.org/10.1109/tnnls.2022.3230458>
- [5] Xu, W., Zhou, Z., Li, T., Sun, C., Chen, X. and Yan, R. (2022) Physics-Constraint Variational Neural Network for Wear State Assessment of External Gear Pump. *IEEE Transactions on Neural Networks and Learning Systems*, **35**, 5996-6006. <https://doi.org/10.1109/tnnls.2022.3213009>
- [6] Durkan, C., Bekasov, A., Murray, I. and Papamakarios, G. (2019) Neural Spline Flows. *Proceedings of the 33rd International Conference on Neural Information Processing Systems*, Vancouver, 8-14 December 2019, 7511-7522.
- [7] Ho, J., Chen, X., Srinivas, A., Duan, Y. and Abbeel, P. (2019) Flow++: Improving Flow-Based Generative Models with Variational Dequantization and Architecture Design. *Proceedings International Conference on Machine Learning (ICML)*, Long Beach, 9-15 June 2019, 2722-2730.
- [8] Grathwohl, W., Chen, R.T.Q., Bettencourt, J., Sutskever, I. and Duvenaud, D. (2018) FFJORD: Free-Form Continuous Dynamics for Scalable Reversible Generative Models. arXiv:1810.01367.
- [9] Dinh, L., Sohl-Dickstein, J. and Bengio, S. (2016) Density Estimation Using Real NVP. arXiv:1605.08803.
- [10] Kim, S., Kim, H., Yun, E., Lee, H., Lee, J. and Lee, J. (2023) Probabilistic Imputation for Time-Series Classification with Missing Data. *Proceedings International Conference on Machine Learning (ICML)*, Honolulu, 23-29 July 2023, 16654-16667.
- [11] Zhou, L., Poli, M., Xu, W., Massaroli, S. and Ermon, S. (2023) Deep Latent State Space Models for Time-Series Generation. *Proceedings International Conference on Machine Learning (ICML)*, Honolulu, 23-29 July 2023, 42625-42643.
- [12] Raghu, A., Chandak, P., Alam, R., Gutttag, J. and Stultz, C. (2023) Sequential Multi-Dimensional Self-Supervised Learning for Clinical Time Series. *Proceedings International Conference on Machine Learning (ICML)*, Honolulu, 23-29 July 2023, 28531-28548.
- [13] Li, Y., Chen, W., Chen, B., Wang, D., Tian, L. and Zhou, M. (2023) Prototype-Oriented Unsupervised Anomaly Detection for Multivariate Time Series. *Proceedings International Conference on Machine Learning (ICML)*, Honolulu, 23-29 July 2023, 19407-19424.
- [14] Park, S., Park, B., Lee, M. and Lee, C. (2023) Neural Stochastic Differential Games for Time-Series Analysis. *Proceedings of the 40th International Conference on Machine Learning*, Honolulu, 23-29 July 2023, 27269-27293.
- [15] Lai, Z., Liu, M., Pan, Y. and Chen, D. (2022) Multi-Dimensional Self Attention Based Approach for Remaining Useful Life Estimation. arXiv:2212.05772.
- [16] Dai, E. and Chen, J. (2022) Graph-Augmented Normalizing Flows for Anomaly Detection of Multiple Time Series. arXiv:2202.07857.
- [17] Zhou, Q., He, S., Liu, H., Chen, J. and Meng, W. (2024) Label-Free Multivariate Time Series Anomaly Detection. *IEEE Transactions on Knowledge and Data Engineering*, **36**, 3166-3179. <https://doi.org/10.1109/tkde.2024.3349613>
- [18] Vaswani, A., et al. (2017) Attention Is All You Need. *Proceedings of the 31st International Conference on Neural Information Processing Systems*, Long Beach, 4-9 December 2017, 6000-6010.
- [19] Bhatti, A.A. (2009) Reduced Order Multiport Parallel and Multidirectional Neural Associative Memories. *Biological Cybernetics*, **100**, 395-407. <https://doi.org/10.1007/s00422-009-0310-0>

-
- [20] Virbitskaite, I.B., Bozhenkova, E.N. and Erofeev, E. (2015) Space-Time Viewpoints for Concurrent Processes Represented by Relational Structures. *Proceedings of the 24th International Workshop on Concurrency, Specification and Programming*, Rzeszow, 28-30 September 2015, 222-233.
- [21] Pan, J., Lin, C., Nie, L., Liu, M. and Zhao, Y. (2024) Multimodal Spatiotemporal Aggregation for Point Cloud Accumulation. *Journal of Visual Communication and Image Representation*, **103**, Article ID: 104243. <https://doi.org/10.1016/j.jvcir.2024.104243>
- [22] Abdulaal, A., Liu, Z. and Lancewicki, T. (2021) Practical Approach to Asynchronous Multivariate Time Series Anomaly Detection and Localization. *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, 14-18 August 2021, 2485-2494. <https://doi.org/10.1145/3447548.3467174>
- [23] Hundman, K., Constantinou, V., Laporte, C., Colwell, I. and Soderstrom, T. (2018) Detecting Spacecraft Anomalies Using LSTMs and Nonparametric Dynamic Thresholding. *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, London, 19-23 August 2018, 387-395. <https://doi.org/10.1145/3219819.3219845>
- [24] Ruff, L., Vandermeulen, R.A., Görnitz, N., Binder, A., Müller, E., Müller, K.-R. and Kloft, M. (2019) Deep Semi-Supervised Anomaly Detection. arXiv:1906.02694.
- [25] Goyal, S., Raghunathan, A., Jain, M., Simhadri, H.V. and Jain, P. (2020) DROCC: Deep Robust One-Class Classification. *Proceedings International Conference on Machine Learning (ICML)*, 13-18 July 2020, 3711-3721.
- [26] Audibert, J., Michiardi, P., Guyard, F., Marti, S. and Zuluaga, M.A. (2020) USAID: Unsupervised Anomaly Detection on Multivariate Time Series. *Proceedings 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 6-10 July 2020, 3395-3404.