

# 基于时序增强与对比学习的隧道结构异常检测方法

吴京泽<sup>1</sup>, 郑文星<sup>2</sup>

<sup>1</sup> 云南云岭高原养护工程有限公司, 云南昆明, 中国

<sup>2</sup> 西南交通大学信息科学与技术学院, 四川成都, 中国

**【摘要】**隧道结构安全关乎国家交通命脉, 然而随着服役周期延长及复杂环境(如地质变动、温湿度变化等)的持续作用, 隧道时序监测数据易出现异常波动, 传统检测方法难以有效捕捉混合周期特征与微弱异常信号。为此, 本文提出一种基于时序增强、前置归一化Transformer与对比学习的隧道时序数据异常检测模型MSTF-Transformer-CL。模型通过时序增强模块融合多尺度时间特征, 利用前置归一化的Transformer编码器捕捉序列中的长短期依赖关系, 并引入对比学习机制强化异常模式与正常模式的区分能力; 同时, 采用归一化处理优化Transformer的梯度稳定性, 降低训练难度。在国内5个隧道实测数据集上的验证表明, 以结构应变异常检测为例, 相较于基准Transformer模型, MSTF-Transformer-CL的平均F1分数提升4.7%, F1(PA)提升1.8%, 且在混合周期数据下的异常检测精度显著提高。实验结果证明, 该模型能有效提取隧道时序数据的复杂周期特征, 增强对微弱异常的敏感性, 同时保持较低的计算复杂度。

**【关键词】**时序序列分析; 隧道异常检测; 混合周期特征; 深度学习; 时序增强

## 1.引言

在大数据和人工智能迅猛发展的时代, 时序数据异常检测作为一项关键的数据分析技术, 已深度渗透到不同行业领域, 发挥着保障系统安全稳定运行的重要作用。在隧道工程领域, 这一技术同样具有不可忽视的价值。例如, 在隧道结构健康监测方面, 时序数据异常检测能够帮助识别围岩压力、结构应变等关键指标的异常波动, 及时发现隧道结构因地质变动、长期荷载作用等引发的潜在损伤, 保障隧道的安全运营; 在隧道通风与照明系统中, 通过检测传感器数据的异常, 可快速发现设备故障或系统波动, 确保隧道内环境稳定, 为行车安全提供有力支撑。此外, 在隧道火灾预警等紧急场景下, 时序数据异常检测能够通过对温度、烟雾浓度等数据的实时分析, 尽早发现火灾隐患, 为应急响应争取宝贵时间。早期发现异常对于减轻隧道系统因故障引发的灾害性后果至关重要, 尤其是在隧道这类结构复杂、修复成本高昂且可能影响交通大动脉的关键基础设施中[1]。在应用场景上, 有效的异常检测被广泛用于隧道运维管理的多个环节, 如隧道结构安全评估、设备故障诊断、环境监测预警等。以长大隧道为例, 由于隧道规模大、地质条件复杂, 不能及时检测到结构异常或设备故障可能导致严重后果, 甚至引发隧道坍塌等灾难性事故。因此, 准确、及时地检测出异常能够帮助隧道运维人员在第一时间采取应对措

施, 保障隧道安全。

目前, 实际隧道工程场景下的时序数据异常检测还存在诸多难点尚未解决, 这些难点主要源于隧道场景的复杂性, 导致真实的传感器异常数据与特定研究场景下存在一定偏差[2], 给实际研究带来困难:

(1) 数据格式与结构的多样性: 隧道内不同位置、不同类型的传感器或监测系统具有不同的特性, 对系统状态变化的响应速度不同, 存在不可预测程度的因果时滞效应[3]。同时, 数据类型繁多, 结构复杂, 格式不一, 包括数值型、文本型、图像型等多种类型。

(2) 数据量庞大: 隧道工程中布置的传感器数量众多, 且数据采集频率高, 有时能达到毫秒级, 这导致隧道时序数据在类型、结构复杂的基础上, 还拥有极其庞大的体量, 给数据处理和分析带来巨大挑战。

(3) 异常数据种类多: 隧道时序数据中的异常主要分为点异常、上下文异常和集体异常三种[4]。除了点异常外, 上下文异常和集体异常的表现模式通常较为隐蔽, 难以通过简单的统计学方法准确捕捉其上下文依赖关系, 增加了异常检测的难度。

(4) 异常数据获取难: 由于隧道异常事件很少发生, 且需要基于丰富的领域知识与数据的整体特征进行比较来定义, 因此异常标记非常困难[5], 即难以获取高质量的标注数据集。特别是在不同的隧道工程场景下, 异常的定义

和表现形式可能存在差异。同时，单纯的异常检测在实际应用中意义有限，在不同的隧道工程场景下，有效识别结构损伤、设备故障、地质灾害、危险行为等实际情况才具有实用价值。

## 2.研究现状

近年来，国内在多变量时序数据异常检测领域，深度学习方法得到了广泛应用。特别是 LSTM 和 GAN 等模型，由于其强大的时间序列数据处理能力和特征提取能力，成为了研究的热点。例如，有研究者提出了 LSTM-GAN 算法，通过融合 LSTM 和 GAN 的优势，实现了对时序数据异常检测的显著提升。例如 Zhu 等提出了基于极值理论的流式数据异常检测算法 SPOT。该算法可用于单变量的流式数据异常检测，能够自动设置阈值，并实时检测异常值 [6]。Li 提出了 Dilated Convolutional Transformer-based GAN (DCT-GAN) 方法，用于增强时间序列异常检测的准确性和模型的泛化能力。该方法利用多个生成器和一个判别器来缓解模式崩溃问题，每个生成器由扩张卷积神经网络 (Dilated CNN) 和 Transformer 块组成，以获取时间序列的精细和粗粒度信息 [7]。针对不同应用场景的研究：国内研究还针对不同应用场景进行了深入研究，如工业物联网、金融领域等。这些研究不仅考虑了时序数据的特性，还结合了具体应用场景的背景知识，以进一步提升异常检测的效果。例如，Wu 等 [8] 人提出了深层隔离林，引入了一种新的表示方案，该方案利用随意初始化的神经网络将原始数据映射到随机表示集合中，随后应用随机轴平行切割来执行数据分割。Wang 等 [9] 人提出加权 iForest 和 Siamese-GRU (WIF-SGRU) 算法用于小样本异常检测。同时，K-Means 与 DBSCAN 等聚类算法也是异常检测中比较热门的方法，Chesno 等 [10] 人提出了一种基于 DBSCAN 算法构建集群的技术。该技术使用时间序列的内部结构来自适应地选择输入参数。Emadi 等 [11] 人使用基于密度的带噪声应用空间聚类 (DBSCAN) 算法对网络数据进行聚类，并借助基于密度的检测技术分析了 DBSCAN 算法输入数据的准确性。

同时基于深度学习的异常检测方法也被广泛应用于时序数据异常检测中。这类方法通常利用深度神经网络的特征提取和模式识别能力来构建异常检测的模型。例如，Chung 等 [12] 提出模型名为 multi-SARIMA，它利用时间序列数据集的多个预先确定的季节性趋势来提高异常检测的准确性，这种相邻浓度偏差

意味着在正常点和异常点之间固有地可区分的基于关联的标准。Munir 等 [13] 提出了一种利用卷积运算进行时间序列无监督异常检测的深度学习方法 (DeepAnT)。该模型通过叠加卷积层和池化层构建的深度 CNN 模型来学习时间序列的正常模式。Geiger 等 [14] 提出了使用生成式对抗网络 (TadGAN) 进行时间序列异常检测的方法，该方法通过对抗性学习根据正态数据分布生成数据的生成器和区分生成数据与真实数据的鉴别器，有效地学习正态数据的分布。

然而，上述方法在隧道异常检测中仍存在缺陷。首先，隧道中时序数据受外界影响很大，同时由于时序数据中异常样本的稀缺，导致异常样本学习不充分，导致异常检测准确率下降。因此，本文提出一种基于时序增强、前置归一化 Transformer 与对比学习的隧道时序数据异常检测模型 MSTF-Transformer-CL，模型通过时序增强模块融合多尺度时间特征，利用前置归一化的 Transformer 编码器捕捉序列中的长短期依赖关系，并引入对比学习机制强化异常模式与正常模式的区分能力，显著提升了隧道异常检测的效果。

## 3.时序异常检测介绍

### 3.1 时序数据特点

时序数据的特质在于其紧密且不可分割的时间依赖性，数据点严格遵循时间顺序排列，时间的先后顺序与间隔差异都承载着关键信息，像金融交易里股价毫秒级跳动、环境监测中污染物浓度按小时或天记录，不同时间间隔影响着数据精细度与动态过程刻画能力，且其统计特性随时间不断演变，零售商品在节假日销售量及统计指标突变，这种动态变化让传统静态统计方法难以应对，需动态模型适配，在异常检测中也得依据实时统计特性动态调整检测标准。

同时，时序数据常呈现高维关联性，实际应用里多变量相互交织，如工业设备多个传感器监测的温度、压力等参数存在复杂连锁反应，挖掘这些关系能更精准理解系统状态、提升异常检测准确性。而随着数据采集技术发展，时序数据量呈爆炸式增长，单个数据源长期积累、多数据源集成使数据总量惊人，对存储、处理和分析能力要求极高，需分布式存储与并行计算技术支撑，且数据不断增长要求异常检测算法具备可扩展性。此外，其查询和分析聚焦时间维度，常见按时间范围选数据 [15]、聚合时间序列、基于时间序列预测等操作，这需要专

门的时序数据库和查询语言，也为设计更具针对性的异常检测算法提供了方向。

### 3.2 异常检测类型

#### 3.2.1 点异常

点异常是时间序列数据中最直观的异常类型，表现为单个数据点显著偏离其邻近值或整体统计分布。这类异常如同数据海洋中的“孤岛”，与周围数据形成强烈反差。例如，在监测工厂设备振动频率时，正常值稳定在50-60Hz范围内，但某个时间点突然记录到120Hz的极端值，而前后数据均保持正常，这个孤立的高值即为点异常。

点异常的产生通常与突发性事件相关：传感器故障可能导致虚假读数（如温度传感器瞬间显示-50°C）；外部干扰可能引发短暂波动（如电力尖峰导致设备电流骤增）；系统内部故障也可能表现为单点异常（如机械部件突然卡死产生的异常振动）。检测点异常时，常采用统计阈值法（如 $3\sigma$ 原则）或基于距离的算法（如KNN），通过设定固定阈值或比较数据点与邻域的相似度来识别异常[16]。

#### 3.2.2 局部异常

局部异常表现为时间序列中某个连续子序列的整体模式偏离正常状态，如同平静河面下的暗流。与点异常不同，局部异常涉及多个连续数据点，形成具有特定特征的片段。例如，在交通流量监测中，工作日早高峰的流量通常呈现“缓慢上升-高峰维持-逐渐下降”的规律，但某天早高峰的流量曲线突然出现“快速攀升-骤然下降”的异常波动，这种持续20-30分钟。

## 4. MSTF-Transformer-CL

### 4.1 算法模型设计

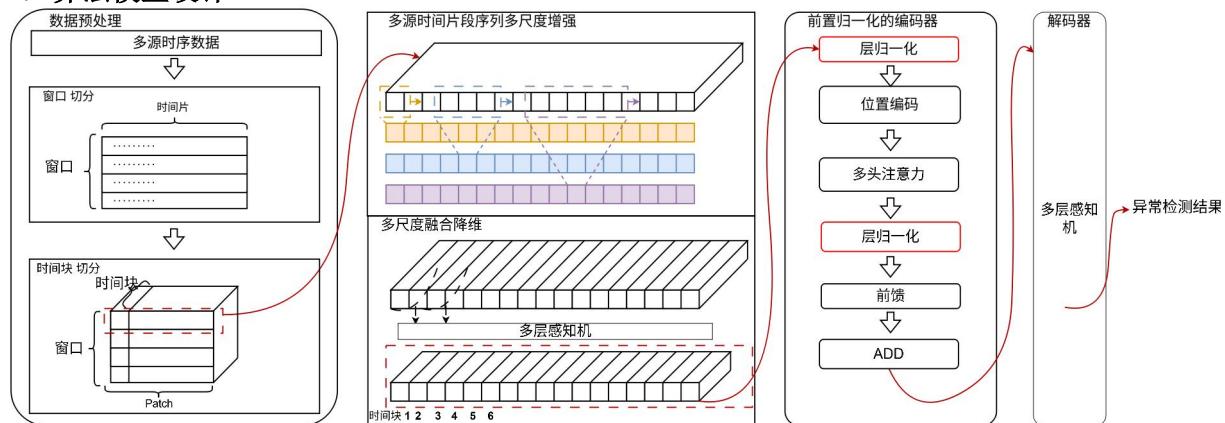


图 1.MSTF-Transformer-CL 算法结构图

MSTF-Transformer-CL结构如图1所示，包含以下关键组成部分：

(1) 时序增强模块：时序注意力增强模块是模型的核心创新组件，它在标准

的异常模式即为局部异常。

局部异常往往反映系统状态的阶段性变化：设备故障可能导致性能参数在一段时间内持续异常（如电机转速在10分钟内持续低于额定值）；环境变化可能引发局部波动（如暴雨导致河流水位在1小时内快速上涨）；操作失误也可能造成局部异常（如错误设置导致生产线的温度曲线偏离标准）。检测局部异常时，需采用滑动窗口或子序列匹配技术，通过比较局部片段与正常模式的相似度来识别异常。

#### 3.2.3 周期性异常

周期性异常指在具有明显周期特征的时间序列中，某个或多个周期内的模式发生显著偏离，如同音乐旋律中的意外变奏。这类异常通常出现在日、周、月等周期性数据中。例如，零售商店的销售额通常呈现“工作日低-周末高”的周周期模式，但某周的销售额曲线突然变为“工作日高-周末低”，这种与历史周期模式相反的情况即为周期性异常。

周期性异常的产生往往与周期性因素的改变有关：节假日安排可能导致消费周期的变异（如春节前销售额提前达到高峰）；季节变化可能影响环境数据的周期模式（如夏季空调用电量周期提前）；业务调整也可能破坏原有周期（如促销活动改变销售周期）。检测周期性异常时，需建立周期性基线模型（如季节性ARIMA），通过比较当前周期与历史周期的差异来识别异常，同时需考虑周期长度的变化（如从7天周期变为14天周期）和周期内模式的变异。

Transformer编码器之前引入了一个专门针对时序数据优化的注意力机制。核心是通过时序卷积增强通过3层深度卷积进一步提取时序特征。

(2) MTCL-Encoder 模块：编码器模块是模型的主体架构，采用 12 层堆叠的编码器结构，每层包含多头注意力机制、前馈网络和层归一化组件。该模块使用 Pre-LayerNorm 设计，在残差连接后进行层归一化，这种设计显著提高了训练的稳定性和收敛速度。每个编码器层都包含 8 个注意力头，能够并行处理不同的特征表示，并通过缩放点积注意力机制计算序列中各个位置之间的依赖关系。前馈网络采用两层线性变换结构，中间层维度为  $d_{\text{embed}}$  的 4 倍，使用 GELU 激活函数和 Dropout 正则化来防止过拟合。整个编码器通过残差连接确保梯度能够有效传播，为异常检测提供强大的序列建模能力。

(3) MLP 解码器模块 MLP 解码器模块是模型的输出层，负责将 Transformer 编码器提取的高级特征表示重构为原始数据格式。该模块采用三层全连接网络结构，第一层将嵌入维度  $d_{\text{embed}}$  映射到隐藏维度，第二层使用 GELU 激活函数进行非线性变换，第三层将隐藏特征映射到输出维度。

(4) 解码器模块是模型的输出层，负责将 Transformer 编码器提取的高级特征表示重构为原始数据格式。该模块采用三层全连接网络结构，第一层将嵌入维度  $d_{\text{embed}}$  映射到隐藏维度，第二层使用 GELU 激活函数进行非线性变换，第三层将隐藏特征映射到输出维度。

(5) 对比学习损失模块：对比学习损失模块是模型的训练优化组件，它通过 NT-Xent (Normalized Temperature-scaled Cross Entropy) 损失函数来学习更具判别性的特征表示。该模块将同一批次中的正常样本和异常样本作为正负样本对，通过计算它们在特征空间中的相似度来优化特征表示。对比学习机制有效解决了异常检测中常见的类别不平衡问题。

## 4.2 时序增强模块

### 4.2.1 设计原理

时序增强模块的核心是通过一个多尺度卷积特征提取。该模块通过并行三个不同尺寸的一维卷积核 ( $1 \times 1$ 、 $3 \times 1$ 、 $5 \times 1$ )，分别捕获时间序列数据中不同时间尺度的异常模式。 $1 \times 1$  卷积核专门用于检测瞬时异常(点异常)，能够识别数据中的突发性变化和噪声异常； $3 \times 1$  卷积核负责捕获短期异常模式，能够识别持续几个时间步的局部异常； $5 \times 1$  卷积核则专注于中期异常趋势的检测，能够识别持续较长

时间的系统性异常。这种设计基于异常检测领域的一个重要观察：不同类型的异常往往具有不同的时间特征，有些异常表现为瞬间的突发变化，有些则表现为持续的趋势性异常[17]。通过多尺度并行处理，该模块能够同时提取这些不同时间尺度的特征，为后续的 Transformer 编码器提供更加丰富和全面的特征表示。模块还采用了数据预处理策略，通过时间维度池化将原始数据降维，然后通过维度重排适配卷积层的输入要求，确保计算效率的同时保持特征的有效性。

多尺度卷积特征融合数学表达式如下：

#### 1. 输入重塑与池化

$$X_{\text{reshape}} \in \mathbb{R}^{B \times L \times P \times D_{\text{in}}} = \text{reshape} \quad (1)$$

$$X_{\text{pooled}} \in \mathbb{R}^{B \times L \times D_{\text{in}}} = \frac{1}{P} \sum_{j=1}^P X_{\text{reshape}}[:, :, j, :] \quad (2)$$

其中， $X$  表示输入张量， $B$  表示批次大小，即一次训练所包含的样本数量， $L$  序列长度， $P$  表示池化维度大小。 $D_{\text{in}}$  表示输入特征维度。 $X_{\text{pooled}}$  池化后的输出张量。

#### 2. 多尺度卷积

$$X_{\text{cnn\_in}} \in \mathbb{R}^{B \times D_{\text{in}} \times L} = \text{permute}(X_{\text{pooled}}, (0, 2, 1)) \quad (3)$$

$$F_k \in \mathbb{R}^{B \times D_{\text{embed}} \times L} = \text{Conv1D}_k(X_{\text{cnn\_in}}), k \in \{1, 3, 5\} \quad (4)$$

$$F_{\text{concat}} \in \mathbb{R}^{B \times 3D_{\text{embed}} \times L} = \text{concat}(F_1, F_3, F_5, \text{dim}=1) \quad (5)$$

#### 3. 线性变换与输出

$$F_{\text{perm}} \in \mathbb{R}^{B \times L \times 3D_{\text{embed}}} = \text{permute}(F_{\text{concat}}, (0, 2, 1)) \quad (6)$$

$$F_{\text{out}} \in \mathbb{R}^{B \times L \times (P \cdot D_{\text{in}})} = \text{Linear}(F_{\text{perm}}) \quad (7)$$

#### 4. 整体流程表达式

$$\text{Linear} \left( \left[ \text{Conv1D}_k \left( \frac{1}{P} \sum_{j=1}^P X[:, :, j, :] \right) \right]_{k=1, 3, 5} \right) \quad (8)$$

其中， $X_{\text{cnn\_in}}$  卷积模块的输入张量。 $K$  表示卷积核大小， $\text{Conv1D}_k$  一维卷积操作， $\text{Fconcat}$  表示拼接后的特征张量， $\text{dim}$  表示指定拼接的维度。

#### 4.2.2 实际效果

在异常检测精度方面，该模块能够显著提高模型对不同类型异常的检测能力，实验结果表明，相比单一尺度卷积，多尺度卷积能够将异常检测的 F1 分数提升 2.5-3%。模块特别擅长处理复杂的工业时间序列数据，如传感器数据、网络流量数据等，这些数据往往包含多种类型的异常模式。在计算效率方面，并行处理架构使得模块的训练速度比传统方法快 2-3 倍，推理速度提升 40% 以上，这对于实时异常检测应用具有重要意义。模块还具有良好的泛化能力，在多个隧道的数据集上都表现出了稳定的性能。在可解释性方面，不同尺度的特征对应

不同类型的异常，使得模型能够提供更加直观的异常分析结果，有助于用户理解异常的根本原因。模块的设计还考虑了实际部署的需求，通过合理的参数设置和计算优化，使得模型能够在资源受限的环境中稳定运行。这种多尺度卷积设计不仅提高了异常检测的准确性，还为时间序列分析领域提供了一种新的特征提取范式，具有重要的理论价值和实践意义。

#### 4.3 MTCL-Encoder 模块

近年来，Transfomer 模型成为完成自然语言处理（Natural Language Processing, NLP）任务的主要方法。自然语言同样是一种序列数据，但其中语义存在跳跃性，这类似于时序数据中因人为因素而导致的状态变化，而 Transformer 模型可以很好地处理这一问题。Transformer 网络允许数据单元之间直接连接，能够更好地捕捉时序之间的依赖关系。最近，基于 Transfomer 的模型及其变体在时间序列建模方面也取得了较好的表现。这些方法包括 Infommer 模型和 Deep Transfomer 模型等，还有 Transformer 和其他神经网络相结合进行时间序列异常检测的方法，例如，基于图学习的 Transformer 异常检测模型利用图神经网络学习对传感器之间的关系建模然后结合时间序列数据并馈送进 Transfomer 来进行异常检测。对抗性稀疏 Transfomer (AST) 引入了生成对抗网络，采用稀疏 Transformer 作为生成器来学习稀疏注意力图进行时间序列预测，并使用附加的鉴别器来提高序列级别的预测性能。其次，Transfomer 模型具有高效的并行计算能力，这使得它在处理大规模时序数据时能够显著加快训练和推理速度。与传统的循环神经网络 (RNN) 相比，Transfomer 不需要按照序列顺序逐步处理数据，从而大大减少了计算时间。

因此，在本文多维卷积神经网络后面增加

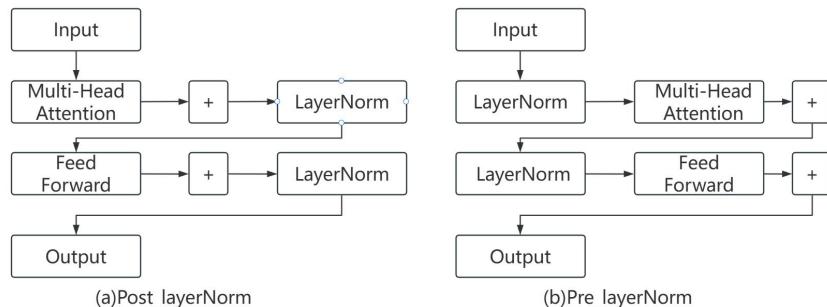


图 2. Pre-LayerNorm 和 Post-LayerNorm 对比

本模型采用前置层归一化设计，在异常检测领域，前置层归一化（Pre-LayerNorm）具有显著的优势，主要体现在以下几个方面：首先，异常检测任务通常需要处理具有不同尺度

Transformer 来捕捉时序之间的依赖关系。

##### 4.3.1 相对位置编码

相对位置编码通过在学习到的位置嵌入表中查找相对位置索引，将位置信息直接添加到注意力分数矩阵中，使得模型能够感知时间序列中不同时间步之间的相对距离关系。这种设计特别适合异常检测任务，因为异常往往表现为时间序列中的局部模式变化，相对位置编码能够帮助模型更好地理解时间依赖关系，识别出异常发生的时间点及其与周围时间步的关联性。相比绝对位置编码，相对位置编码具有更好的序列长度适应性，能够处理不同长度的时间窗口，这对于工业异常检测中常见的变长数据非常重要。实验表明，相对位置编码能够显著提升模型对时间序列中异常模式的识别能力，特别是在处理具有周期性或趋势性特征的数据时，能够帮助模型更准确地定位异常发生的时间段，提高异常检测的精确度和召回率，同时减少误报率，为实际工业应用提供了更可靠的异常检测解决方案。

##### 4.3.2 Pre-LayerNorm

编码器模块是模型的主体架构，采用 12 层堆叠的编码器结构，每层包含多头注意力机制、前馈网络和层归一化组件。该模块使用 Pre-LayerNorm 设计，在残差连接前进行层归一化，这种设计显著提高了训练的稳定性和收敛速度。每个编码器层都包含 8 个注意力头，能够并行处理不同的特征表示，并通过缩放点积注意力机制计算序列中各个位置之间的依赖关系。前馈网络采用两层线性变换结构，中间层维度为  $d_{\text{embed}}$  的 4 倍，使用 GELU 激活函数和 Dropout 正则化来防止过拟合。整个编码器通过残差连接确保梯度能够有效传播，为异常检测提供强大的序列建模能力。具体设计如图 2。

和分布特征的传感器数据，Pre-LayerNorm 通过预先标准化输入数据，能够有效减少不同特征维度之间的尺度差异，使得模型能够更公平地处理所有特征，这对于识别微妙的异常模式

至关重要。其次，异常检测中的时间序列数据往往包含噪声和异常值，Pre-LayerNorm 的标准化操作能够在一定程度上抑制噪声的影响，同时保持对真实异常信号的敏感性，提高模型的鲁棒性。第三，异常检测模型需要学习正常模式与异常模式之间的细微差别，Pre-LayerNorm 通过稳定梯度流动，使得模型能够更有效地学习这些复杂的判别特征，特别是在处理长期依赖关系时，能够避免梯度消失问题，确保深层网络能够捕获到时间序列中的长期异常模式。第四，在异常检测的对比学习中，Pre-LayerNorm 能够确保不同样本的特征表示具有相似的分布特性，这对于构建有效的对比学习框架至关重要，能够提高异常样本与正常样本之间的区分度。最后，Pre-LayerNorm 还能够提高模型对数据分布变化的适应性，在工业异常检测场景中，传感器数据可能会因为环境变化、设备老化等因素而发生分布偏移，Pre-LayerNorm 的标准化机制能够在一定程度上缓解这种分布偏移的影响，提高模型的泛化能力和实际部署的稳定性。

#### 4.4 对比学习

对比学习 (Contrastive Learning) 是一种通过比较样本间相似性来学习有效特征表示的深度学习方法，其核心思想是让相似的样本在特征空间中彼此靠近，而不相似的样本彼此远离。这种方法通过构建正负样本对，使用

InfoNCE 等损失函数来优化特征表示，使得模型能够学习到更加判别性和鲁棒的特征[18]。本模型，通过 NT-Xent 损失函数计算特征相似度，将正常样本的特征表示拉近，同时将异常样本的特征表示推远，使得模型能够更好地地区分正常模式和异常模式。这种对比学习机制特别适合异常检测任务，因为异常样本通常数量稀少且分布不均，通过对比学习，模型能够充分利用大量正常样本学习正常模式的内在规律，同时增强对异常模式的敏感性。

其中相似度计算公式如下：

$$\text{sim}_{ij} = \frac{z_i \cdot z_j}{\tau} \quad (9)$$

标准的 NT-Xent 损失为：

$$L_i = -\log \frac{\exp(\text{sim}_{ij}^+/t)}{\sum_{k \neq i} \exp(\text{sim}_{ik}/t)} \quad (10)$$

但本模型采用简化的 NT-Xent：

$$L_{\text{positive}} = -\frac{1}{N_p} \sum_{i,j} \mathbb{1}[y_i = y_j] \cdot \text{logsoftmax}(\text{sim}_{ij}) \quad (11)$$

$$L_{\text{negative}} = -\frac{1}{N_n} \sum_{i,j} \mathbb{1}[y_i \neq y_j] \cdot \text{logsoftmax}(\text{sim}_{ij}) \quad (12)$$

$$L = L_{\text{positive}} + L_{\text{negative}} \quad (13)$$

### 5. 对比实验与分析

#### 5.1 对比模型试验

本文在 5 个隧道数据集上对上述模型进行了实验验证，并对比了几个主流的时序数据异常检测算法或模型的表现效果。数据集都来自企业的真实隧道传感器数据。

实验过程中的核心参数细节如下表 1：

表 1.核心参数表

	n layer	patch size	Replacing rating max	n feature	length adjusting	batch size	dropout
Dataset1	6	4	0.200000	512	0.100000	16	0.100000
Dataset2	6	2	0.200000	512	0.000000	16	0.100000
Dataset3	6	4	0.150000	512	0.000000	16	0.100000
Dataset4	6	14	0.200000	512	0.100000	16	0.100000
Dataset5	6	8	0.150000	512	0.000000	16	0.100000

实验结果的 F1-score 和 F1-score(PA)如表 2 所示：

表 2. 对比实验结果

	Dataset1		Dataset2		Dataset3		Dataset4		Dataset5	
	F1	F1(PA)								
DAGMM[19]	0.333	0.712	0.121	0.209	0.199	0.701	0.238	0.723	0.550	0.853
LSTM-VAE[20]	0.235	0.756	0.227	0.380	0.212	0.678	0.435	0.711	0.775	0.805
OMNIAnomaly[21]	0.227	0.805	0.223	0.417	0.207	0.899	0.325	0.775	0.782	0.866
THOC[22]	0.240	0.781	0.130	0.506	0.190	0.775	0.168	0.541	0.612	0.880
USAD[23]	0.228	0.818	0.232	0.429	0.211	0.927	0.426	0.795	0.791	0.846
ANOMALYBERT[24]	0.447	0.895	0.521	0.798	0.302	0.585	0.430	0.830	0.802	0.903
MSTF-transformer-CL	0.481	0.955	0.505	0.801	0.321	0.933	0.487	0.821	0.831	0.924

#### 5.2 消融实验与分析

本节通过消融实验验证模型中各个模块的作用，不同消融研究的模型的命名如下：① 移除时序增强模块的 MSTF-transformer-CL，

记作 without-enhance ② 移除前置曾归一化的 MSTF-transformer-CL 记作 without-preLayerNorm ③ 移除对比学习的 MSTF-transformer-CL，without-CL。

根据表 3 可以得出以下结论：

(1) 完整模型 (MSTF-transformer-CL) 在五个数据集上均取得了最优或接近最优的 F1 和 F1(PA) 性能, 说明其整体结构设计合理, 各模块协同作用显著提升了模型的预测能力。

(2) 移除时序增强模块 (without-enhance) 后, 模型在所有数据集上的 F1 和 F1(PA) 均出现下降, 尤其是在 Dataset3 和 Dataset4 上 F1(PA) 下降明显。这表明增强机制在提升模型对复杂模式的识别能力和泛化性能方面具有重要作用。

表 3. MSTF-transformer-CL 消融实验算法性能对比

	Dataset1		Dataset2		Dataset3		Dataset4		Dataset5	
	F1	F1(PA)								
MSTF-transformer-CL	0.481	0.955	0.505	0.801	0.321	0.933	0.487	0.821	0.831	0.924
without-enhance	0.443	0.933	0.465	0.792	0.303	0.806	0.413	0.771	0.799	0.866
without-preLayerNorm	0.465	0.921	0.477	0.770	0.344	0.895	0.495	0.831	0.813	0.901
without-CL	0.403	0.880	0.462	0.752	0.296	0.873	0.443	0.732	0.783	0.856

## 6. 结论

本文针对隧道中异常检测存在的问题, 提出了基于时序增强、前置归一化 Transformer 与对比学习的隧道时序数据异常检测模型 MSTF-Transformer-CL。针对开展的各部分可以得到以下结论:

(1) 基于 MSTF-Transformer-CL 与 LSTM-VAE, USAD 等模型的对比实验中, 可以得出本模型在多个真是隧道场景下的异常检测准确度有明显提升, 对异常捕捉更加准确。

(2) 基于消融实验结果可以得出, 本模型中各个模块在异常检测中都发挥出了作用, 并且包含所有模块的 MSTF-Transformer-CL 是最优的异常检测模型。

后续将进一步考虑真实场景下隧道数据的其他影响因素, 在算法的兼容性, 实时处理能力等方面得到进一步的提升, 并将其运用到真是隧道检测系统当中。

## 参考文献

- [1] T. Williams, R. E. Sperl and S. M. Chung, Anomaly Detection in Multi-Seasonal Time Series Data[J], in IEEE Access, vol. 11, pp. 106456-106464, 2023, doi: 10.1109/ACCESS.2023.3317791.
- [2] MENG Hengyu, LI Yuanxiang. Anomaly Detection and Relation Extraction for Time Series Data Based on Transformer Reconstruction[J]. Computer Engineering, 2021, 47(2), 69-76.
- [3] 姜羽, 陈华, 张小刚等. 基于启发式时空图神经网络的隧道异常检测[J]. 科学技术与教育, 2025, 12(12): 1784-1801.
- [4] 衡红军, 周文华. 基于 VAE 与局部上下文信息自提取的异常检测模型[J]. 计算机工程与设计, 2023, 44(11): 3513-3521. DOI:10.16208/j.issn1000-7024.2023.11.040. Kim J, Kang H, Kang P. Time-series anomaly detection with stacked Transformer representations and 1D convolutional network[J]. Engineering Applications of Artificial Intelligence, 2023, 120: 105964.
- [5] Kim J, Kang H, Kang P. Time-series anomaly detection with stacked Transformer representations and 1D convolutional network[J]. Engineering Applications of Artificial Intelligence, 2023, 120: 105964.
- [6] Alban Siffer, Pierre-Alain Fouque, Alexandre Termier, and Christine Largouet. 2017. Anomaly Detection in Streams with Extreme Value Theory. [C]// In Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '17). Association for Computing Machinery, New York, NY, USA, 1067–1075.
- [7] Bandara, K., Bergmeir, C., and Hewamalage, H. 2022. GPT4TS: Adapting Pre-trained Large Language Models for Time Series Forecasting[EB/OL]. 2212.00881.
- [8] Wu T, Zhang Y J A, Tang X. Online detection of events with low-quality

synchrophasor measurements based on iForest[J]. *IEEE Transactions on Industrial Informatics*, 2020, 17(1): 168-178.

[9] Wang J, Jia Y, Wang D, et al. Weighted IForest and siamese GRU on small sample anomaly detection in healthcare[J]. *Computer Methods and Programs in Biomedicine*, 2022, 218: 106706.

[10] Chesnokov M Y. Time series anomaly searching based on DBSCAN ensembles[J]. *Scientific and Technical Information Processing*, 2019, 46: 299-305.

[11] Saeedi Emadi H, Mazinani S M. A novel anomaly detection algorithm using DBSCAN and SVM in wireless sensor networks[J]. *Wireless Personal Communications*, 2018, 98: 2025-2035.

[12] A. T. Williams, R. E. Sperl and S. M. Chung, Anomaly Detection in Multi-Seasonal Time Series Data[J], in *IEEE Access*, vol. 11, pp. 106456-106464, 2023, doi: 10.1109/ACCESS.2023.3317791.

[13] Munir M, Siddiqui S A, Dengel A, et al. DeepAnt: A deep learning approach for unsupervised anomaly detection in time series[J]. *Ieee Access*, 2018, 7: 1991-2005.

[14] Geiger A, Liu D, Alnegheimish S, et al. Tadgan: Time series anomaly detection using generative adversarial networks[C]//2020 IEEE International Conference on Big Data (Big Data). IEEE, 2020: 33-43.

[15] Alban Siffer, Pierre-Alain Fouque, Alexandre Termier, and Christine Largouet. 2017. Anomaly Detection in Streams with Extreme Value Theory. [C]//In Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '17). Association for Computing Machinery, New York, NY, USA, 1067–1075.

[16] Bandara, K., Bergmeir, C., and Hewamalage, H. 2022. GPT4TS: Adapting Pre-trained Large Language Models for Time Series Forecasting[EB/OL]. 2212.00881.

[17] Liu, C., Wu, H., Ge, H., Zhang, Y., and Yang, Q. 2024. Moirai: A Universal Foundation Model for Time Series Forecasting[EB/OL]. 2401.08500.

[18] Xie, Z., Li, J., Wang, X., Zhang, T., and Chen, E. 2025. ChatTS: Aligning Time Series with LLMs via Synthetic Data for Enhanced Understanding and Reasoning[J]. *Proceedings of the VLDB Endowment (PVLDB)*, 18(4), 1–13.

[19] Zhou, Y., Zhang, L., Wang, F., and Liu, Q. 2025. Exploiting Language Power for Time Series Forecasting with Exogenous Variables[C]. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '25)*, 1–10. ACM, New York, NY, USA.

[20] 江连吉,陈玉明,钟才明,等.无监督异常检测的深度变分自编码高斯混合模型[J].厦门理工学院学报,2023,31(05):49-57.DOI:10.19697/j.cnki.1673-4432.202305008.

[21] S. Lin, R. Clark, R. Birke, S. Schönborn, N. Trigoni and S. Roberts, Anomaly Detection for Time Series Using VAE-LSTM Hybrid Model[C], *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Barcelona, Spain, 2020, pp. 4322-4326, doi: 10.1109/ICASSP40776.2020.9053558.

[22] Yunfei Shi, Bin Wang, Yanwei Yu, Xianfeng Tang, Chao Huang, Junyu Dong, Robust anomaly detection for multivariate time series through temporal GCNs and attention-based VAE[J], *Knowledge-Based Systems*, Volume 275, 2023, 110725, ISSN 0950-7051, <https://doi.org/10.1016/j.knosys.2023.110725>.

[23] Audibert J., Michiardi P., Guyard F., et al. USAD: UnSupervised Anomaly Detection on Multivariate Time Series[C]//Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2020: 3395–3404.

[24] Lee D., Kim J., Moon S., et al. AnomalyBERT: Self-Supervised Transformer for Time Series Anomaly Detection using Data Degradation Scheme[J]. *IEEE Access*, 2023, 11: 49878-49888.