计 算 机 研 究 与 发 展
Journal of Computer Research and Development
https://crad.ict.ac.cn
63(10): 2632−2649, 2026

基于自适应多分辨率特征解耦的动态异构网络时间序列分类

刘 瑶[1] 张栏戈[1] 鲁梦瑶[1] 侯 睿[2] 姜睿临[1] 文宇博[1] 丁 熠[1]
[1](电子科技大学信息与软件工程学院 成都 610054)
[2](电子科技大学计算机科学与工程学院 成都 611731)
(liuyao@uestc.edu.cn)

Adaptive Multi-Resolution Feature Decoupling for Time Series Classification in Dynamic Heterogeneous Networks

Liu Yao[1], Zhang Lanyi[1], Lu Mengyao[1], Hou Rui[2], Jiang Ruilin[1], Wen Yubo[1], and Ding Yi[1]
[1](*School of Information and Software Engineering*, *University of Electronic Science and Technology of China*, *Chengdu* 610054)
[2](*School of Computer Science and Engineering*, *University of Electronic Science and Technology of China*, *Chengdu* 611731)

Abstract With the deep integration of cloud-edge-end collaborative computing and IoT, the massive heterogeneous terminals in dynamic heterogeneous networks have become the core protection objects of cybersecurity systems. The multivariate time-series data generated by these intelligent terminals during continuous operation, including industrial control sensor logs and network traffic features, carry complete semantic information about terminal behavior. Accurate classification of such time-series data is a critical foundation for terminal threat detection. However, the data distribution imbalance caused by terminal heterogeneity poses three key challenges to existing time-series classification models. First, constrained by fixed-resolution processing paradigms, models struggle to capture multi-granularity threaten features across different time scales. Second, existing methods forcibly map all resolution features into a unified high-dimensional space, ignoring intrinsic differences in feature dimension distribution across time resolutions, introducing noise, feature redundancy, and limiting deployment feasibility on resource-constrained edge terminals. Third, performing joint classification with multiple resolutions at inference incurs significant time and memory overhead from the feature extraction and fusion of each resolution branch, further exacerbating the inherent tension between lightweight deployment and high-accuracy detection. To address the alignment between time resolutions and feature dimensions, we propose an adaptive multi-resolution representation learning (AdaMRL) model for dynamic heterogeneous networks, achieving bidirectional alignment and decoupling of time resolutions and feature dimensions. The core contributions of this work are as follows. First, a multi-resolution Matryoshka feature architecture is constructed, equipping each time resolution with an independent Matryoshka representation learning layer to adaptively select an appropriate dimensional subspace for representation learning. Second, a multi-resolution joint loss supervision mechanism is proposed, which applies synchronized constraints to representations at each resolution during training to enhance cross-resolution feature robustness and accelerate model convergence. Third, an adaptive dimension selection and ensemble strategy is designed, dynamically matching the optimal feature dimensionality to each resolution at inference, with collaborative decision-making via soft-voting. Extensive

收稿日期: 2026-06-06; **修回日期**: 2026-07-13
基金项目: 国家自然科学基金项目(U22B2061); 国家重点研发计划项目(2022YFB4300603)
This work was supported by the National Natural Science Foundation of China (U22B2061) and the National Key Research and Development Program of China (2022YFB4300603).
通信作者: 刘瑶(liuyao@uestc.edu.cn)

experiments on 10 UEA multivariate time-series benchmark datasets and the industrial IoT edge security dataset Edge-IIoTset demonstrate that AdaMRL achieves state-of-the-art classification performance. Compared with baseline models, the proposed method improves average classification accuracy by 0.78% and reduces redundant feature parameter computation by 27.3% under optimal adaptive configuration, striking an effective balance between detection accuracy and computational efficiency, and providing robust support for real-time security protection of edge heterogeneous terminals.

Key words time series classification; multi-resolution representation learning; Matryoshka representation learning; dynamic heterogeneous networks; network intrusion detection

摘 要 随着云边端协同计算与物联网技术的深度融合，动态异构网络中的海量终端已成为网络安全体系中的核心防护对象。这些智能终端在持续运行过程中产生网络流量、工控传感日志等多元时间序列数据，承载着终端行为的完整语义信息。因此，对此类时间序列数据进行精准分类是实现终端威胁检测的关键技术。然而，终端异质性引发的数据分布不均衡问题，使现有时序分类模型面临三重挑战：1）受限于固定分辨率的处理范式，模型难以捕捉不同时间尺度下的多粒度威胁特征；2）现有方法通常将所有分辨率特征映射至统一高维空间，忽视了不同时间分辨率在特征维度分布上的内在差异，导致噪声引入与特征冗余，制约了模型在计算资源受限的边缘终端上的部署适配能力；3）使用多分辨率进行联合推理，各分辨率分支的特征提取与融合过程带来显著的计算与存储开销，加剧了轻量化部署与高精度检测之间的固有矛盾。为了解决时间分辨率与特征维度适配的问题，提出了面向动态异构网络的自适应多分辨率表示学习（adaptive multi-resolution representation learning, AdaMRL）模型，实现了时间分辨率与特征维度的双向对齐解耦。首先，构建了多分辨率嵌套表征架构，提出多分辨率与嵌套维度自适应匹配机制，为每个时间分辨率独立配备嵌套表示学习层，自动选择最优的维度进行嵌套表示学习；其次，提出了多分辨率联合损失监督机制，在训练阶段对各分辨率表征进行同步约束，增强跨分辨率特征的鲁棒性并加快模型收敛；最后，设计了自适应维度选择与集成策略，在推理阶段为不同分辨率动态匹配最优特征维度，并通过软投票机制实现协同决策。在 10 个 UEA 多变量时间序列基准数据集与工业物联网边缘安全基准数据集 Edge-IIoTset 上的广泛实验表明，AdaMRL 达到了当前先进的分类性能。与基准模型相比，所提方法平均分类准确率提升 0.78 个百分点，并在最佳自适应配置下将冗余特征参数计算量降低 27.3%，在检测精度与计算效率之间实现了有效平衡，为边缘异质终端的实时安全防护提供了有力支撑。

关键词 时间序列分类；多分辨率表示学习；嵌套表示学习；动态异构网络；网络入侵检测

中图法分类号 TP393

DOI: 10.7544/issn1000-1239.202660465 **CSTR:** 32373.14.issn1000-1239.202660465

随着物联网、云边端协同计算以及下一代无线通信技术的蓬勃发展，动态异构网络正逐步占据主导地位[1]。在此类异构网络架构中，智能工控设备、智联汽车等数量庞大的异质终端相互交织、深度互联，共同构成了错综复杂的网络生态。在实际运行中，异质终端及边缘采集器会持续不间断地生成海量多元时间序列数据，涵盖设备运行日志、感知传感器数据、网络流量序列等多种形态。因此，对这些时序数据进行精准模式识别与分类，是实现终端威胁检测（如异常识别、入侵检测、故障预警）的核心技术[2]。

近年来，基于深度学习的多分辨率时序模型（如TimeMixer[3], TimesNet[4]）通过频域周期发现、二维时序重构、多尺度分解与跨尺度混合等方式，在捕捉高频局部细节和低频周期趋势方面展现出较强的表征能力，有效拓展了模型对复杂时间模式的感知范围，极大地推动了时序分析领域的研究进展[5]。然而，现有方法的核心关注点集中于如何提取和融合多尺度时间模式，却未讨论不同时间分辨率特征所需的表征空间维度这一关键问题。在动态异构网络复杂多变的环境下，将上述前沿模型部署于时序分类任务时却面临着检测精度不足与系统部署失配两大核心难题。经过深入研究分析，本文将其归纳为时间分辨率与特征维度不适配问题，并从以下 2 个层面展开阐述。首先，异质终端由于采集环境和硬件差异，产生的时序数据往往伴随极高的数据异构性与信噪比差异[6]。现有模型通常采用固定的特征提取范式，

强制所有时间分辨率特征映射到具有固定特征维度的统一表征空间。这种设计导致细粒度分辨率(短期趋势特征包含大量高频随机噪声)获得过多的特征维度,极易发生过拟合进而引发威胁检测的高误报率;相反,经过下采样浓缩了全局威胁行为模式的粗粒度分辨率(长期季节性特征代表低频稳定模式),若特征维度受限则无法被充分表达。其次,异质终端在计算、存储等资源方面存在显著差异。传统多分辨率分类模型在推理阶段无法根据不同时间分辨率自适应选择匹配的特征维度,由此带来庞大的计算开销,严重制约了模型在资源受限边缘终端上的轻量化与云端部署能力。

为了解决异质终端时序数据分类任务面临的精度与效率的挑战,本文提出一种面向动态异构网络的自适应多分辨率表示学习框架(adaptive multi-resolution representation learning, AdaMRL)。该框架的核心目标是实现时间分辨率与特征表征空间维度的双向对齐解耦,使得模型能够根据时序数据中存在的多分辨率时间模式自适应匹配最佳的特征空间维度。具体而言,本文在基准多分辨率架构的基础上将嵌套表示学习 MRL(Matryoshka representation learning)[7]由单一表征的多维度截断机制扩展为多分辨率场景下的分辨率与表征维度自适应匹配机制。在表征学习阶段,AdaMRL 为每个时间分辨率配备独立的 MRL 分类头,通过多分辨率联合损失函数,促使不同分辨率的时间特征在多维度特征空间的对齐解耦。在推理部署阶段,AdaMRL 设计了自适应维度匹配与集成策略,模型根据验证集性能为每个分辨率分支独立选择提升分类性能最大的嵌套维度。例如,当细粒度分支主要包含非判别性高频噪声时,模型自动选择较低维度以抑制特征冗余;当细粒度突变本身具有类别判别意义时,模型则保留较高维度以增强特征表达能力。因此,这种自适应的嵌套式特征表示机制具备灵活适配异质终端多分辨率时间模式表达的潜力,有望在更低计算开销下提升模型的协同决策精度。综上所述,本文的主要贡献如下:

1)提出了一种新颖的 AdaMRL 框架用于异构网络的时序数据分类任务。针对异质终端中时序数据特征复杂、异构性强的特点,将多分辨率与特征维度空间进行对齐解耦的嵌套表示学习,有效缓解了多分辨率特征冗余与噪声敏感问题。

2)提出了多分辨率联合损失函数与自适应维度选择机制。在训练阶段,通过联合损失能够同时约束不同分辨率和嵌套维度下的分类预测,使低维前缀也具备判别能力;在推理阶段,通过多时间分辨率分支和多嵌套维度之间的自适应匹配,不仅有效提升了分类准确率,还为资源受限的边缘终端提供了参数可收缩的轻量化部署方案。

3)在 10 个涵盖多元复杂场景的 UEA 基准数据集和 1 个工业物联网边缘安全基准数据集上的实验表明,本文提出的 AdaMRL 模型达到了先进的分类性能。相比于现有基准方法,该模型不仅提升了分类准确性,还可以大幅降低模型参数量与计算开销。

1 相关工作

1.1 时间序列分类

时间序列分类(time series classification, TSC)作为时间序列分析和数据挖掘领域的重要子任务,被称为 21 世纪数据挖掘领域的十大挑战性研究问题之一[8-9]。TSC 旨在将连续观察的时序信号映射到具有明确语义的离散标签类别集合,将连续信号转化为语义标签对各领域的决策支持具有重要意义[10]。因此,TSC 任务被广泛应用于医学信号检测[11]、声纹信号识别[12]、网络流量分类[13]等领域。传统的 TSC 经典方法包括最近邻(nearest neighbor, NN)分类器[14]、支持向量机(support vector machine, SVM)[15]和动态时间规整(dynamic time warping, DTW)[16]等。然而传统模型受限于固有的非线性表达能力不足,难以适应复杂多变量的时间序列分类任务的需求。

近年来,随着深度学习技术的不断发展,一系列公开高质量的数据集(如 UCR[17]和 UEA[18]数据集)被提出,促进了 TSC 研究的不断进步。例如卷积神经网络[19](convolutional neural network, CNN)、循环神经网络[20](recurrent neural networks, RNN)、LSTM[21](long short term memory)等模型。此外,为了进一步缓解这些模型在长期依赖特征、多变量复杂关系方面的局限性,基于 Transformer[22]的模型可以利用全局注意力机制对长期特征进行有效捕获;基于图卷积网络模型[23](graph neural network, GNN)可以有效建模不同变量间的非欧依赖关系。由于时间序列存在固有的季节性和趋势性,卷积网络如 Inception-Time[24]和 PatchTST[25]在捕捉长程依赖上取得了突破,但难以处理多时间分辨率特征。因此,有必要对时间序列中的多周期、多分辨率进行系统性的建模。例如 TimesNet[4]、TimeMixer[3]利用快速傅里叶变换(fast Fourier Transform, FFT)算法[26]对序列进行多分辨率周期模式的提取,进一步提升了模型预测性能

和稳定性。然而，这些方法将多分辨率周期模式视为同等重要，并且将不同分辨率特征映射于固定维度的特征空间[27]，忽略不同模式特征分布的差异性。与这些方法不同的是，本文构建了一种多分辨率与特征维度自适应对齐表示框架，旨在缓解了多分辨率特征冗余与噪声敏感问题，提升分类准确性。

1.2 异质终端流量检测

随着 5G 通信技术的发展，用户可以便捷地利用不同设备便捷访问互联网提供的各种服务[28]。根据国际电信联盟（International Telecommunication Union, ITU）的报告[29]，在 2019 年全球互联网用户就已经达到 41 亿。在云边协同与物联网高度发达的动态异构网络（如 FlexiViT[30]）中，海量异质终端用户给互联网服务提供商（ISP）带来了严峻的安全挑战。因此，迫切需要找到一种高效的网络流量分类的解决方案，以满足服务质量保障、内容过滤、合法监管等需求[31]。早期的网络流量分类方法主要利用数据包的端口号将网络流量分类到其对应的协议[32-33]。这类基于端口的分类方法优势在于简单高效、计算资源要求低。但随着各类应用程序使用伪装和随机部署策略[34]，直接导致了这类简单算法的失效。基于 DPI[35-36] 的方法通过检查数据包载荷内容克服了端口依赖的局限，但面临计算资源需求高和无法处理加密流量的双重挑战，限制了其在大规模网络中的应用。

近年来，研究人员开始关注将深度学习技术用于网络流量分类中。将原始网络流量作为输入，深度学习网络可以自动捕获网络流量的空间性或时间性，避免了手工特征工程，使得这类方法成为一种更强大、更通用的解决方案[37]。Wang[38] 最早使用人工神经网络（ANN）和堆叠式自动编码器（SAE）解决网络流量分类问题。Lotfollahi 等人[39] 则建立了一个包含卷积神经网络和 SAE 结合的框架，用于网络流量分类。Bu 等人[40] 提出了基于具有多个 MLP 卷积模块的深度学习网络对网络流量进行分类。为了进一步提升分类准确率，研究者相继提出了 YaTC[41]、EAPT[42]、Netmamba[43] 和 SoK[44] 等方法进一步提升了分类性能。然而，现有深度学习模型依赖于复杂的预训练流程或大规格参数，导致模型泛化性受限，并且难以直接部署到资源受限的边缘异构终端上。此外这些方法往往忽视了对不同分辨率特征模式的充分提取和建模[45]，限制了模型对复杂流量模式的表征能力。

1.3 MRL

MRL 最早是由 Kusupati 等人[46] 提出的一种表示学习方法，旨在解决固定容量表征对特定任务适应性不足或过度拟合的问题。MRL 通过在训练阶段联合优化多个嵌套维度（nested dimensions）的分类权重，使得单个模型能够在推理时动态截断特征维度，从而在不同维度下均保持优异的表征能力。MRL 已经被广泛用于视觉与文本处理[47]、多模态检索[48] 等任务以提升特征表示的效率[49]，并且在大模型 token 特征的压缩[50] 方面具有较好的表现。此外，Yi 等人[51] 将 MRL 引入异构联邦学习，通过融合通用表示和个性化表示来构建 MRL 特征表示，提升了模型的学习能力。

受上述工作的启发，考虑到异构终端同样具备独特的特征变化模式，本文创新性地将 MRL 引入到动态异构网络的时序分类任务中。基于 MRL 能够在单一表征空间内学习一组嵌套前缀，使同一特征向量在多个截断维度下保持有效表征的核心能力。本文进一步提出了自适应嵌套式特征维度选择机制，实现了多时间分辨率分支与多特征表示维度的对齐与解耦。与现有方法不同，本文关注的是多分辨率时序特征中不同分辨率分支的维度需求差异性：不同分辨率分支分别捕获局部突变、随机扰动、周期趋势或全局行为模式等不同尺度的时序特性，其最优嵌套维度并不遵循固定的单调规律（即高分辨率对应高维、低分辨率对应低维的简单映射）。因此，AdaMRL 采用数据驱动的策略，为每个分辨率分支自适应选择最优嵌套维度，避免了人为预设维度分配规则带来的次优性，实现了分辨率与表征维度的精细化匹配。

2 自适应多分辨率表示学习框架

2.1 问题定义

给定多元时间序列输入 $\boldsymbol{X}\in\mathbb{R}^{T\times C}$，其中 T 为序列长度，C 为变量数量，分类任务目标是预测其类别标签 $y\in\{1,2,\cdots,K\}$，K 表示划分类别的总数量。在多分辨率时序建模中，输入序列经过下采样或分辨率划分后形成 M 个时间分辨率分支。记第 m 个分辨率分支经过特征提取和全局平均池化后的特征矩阵为 $\boldsymbol{H}^{(m)}\in\mathbb{R}^{N\times d}$，其中 N 为样本数量，d 为统一特征维度。传统多分辨率模型通常默认所有分辨率分支共享相同的表示维度，即 $d_1=d_2=\cdots=d_M=d$。

然而，不同分辨率分支的信息密度、噪声水平和语义抽象程度并不一致，其所需的最小有效维度可能存在显著差异。若采用统一固定维度，则可能

造成高信息密度分支维度不足,或低信息密度分支维度冗余,从而出现分辨率分支到表示维度错配的问题。

2.2 总体框架

本节详细介绍 AdaMRL 的技术细节。如图 1 所示,AdaMRL 旨在解决 TSC 任务中多分辨率时间模式与特征维度匹配的难题。整体架构由 4 个核心模块构成:1)多分辨率混合的时序特征提取;2)多分辨率嵌套表示;3)多分辨率联合监督训练;4)自适应维度选择与集成推理。

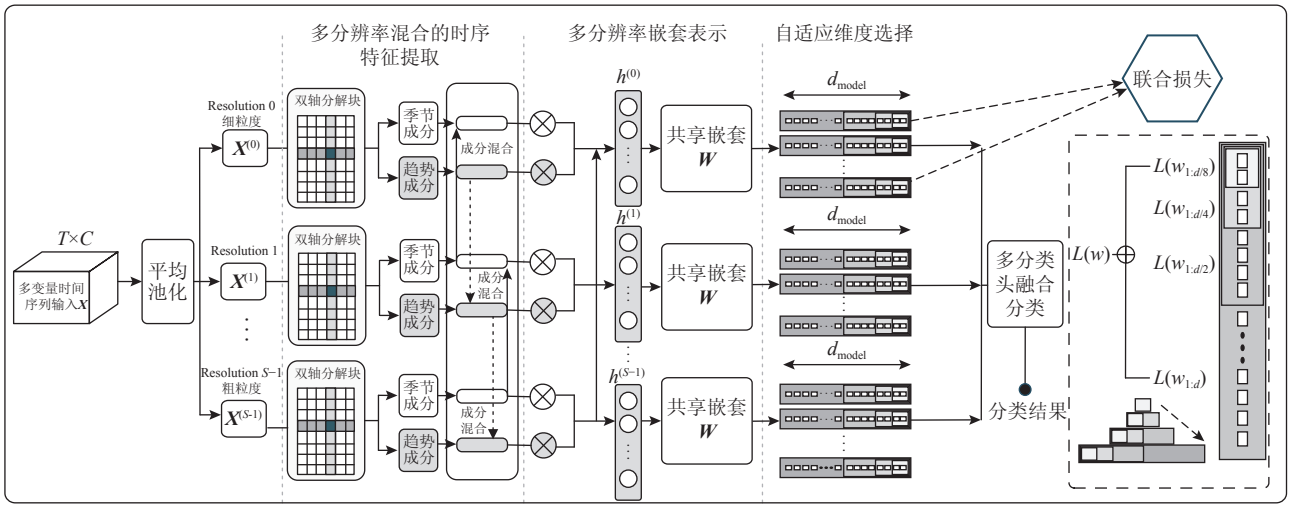


Fig. 1 AdaMRL framework

图 1 AdaMRL 框架

2.3 多分辨率混合的时序特征提取

为了捕捉时间序列中从高频局部波动到低频全局趋势的多层次信息,首先对输入序列$\boldsymbol{X}$进行多分辨率划分。通过使用卷积核大小为 2 的卷积操作进行S次下采样,生成了M个不同时间分辨率的子序列集合$\{\boldsymbol{X}^{(0)},\boldsymbol{X}^{(1)},\cdots,\boldsymbol{X}^{(m)},\cdots,\boldsymbol{X}^{(M-1)}\}$,此时$M=S+1$,其中$\boldsymbol{X}^{(0)}$为原始细粒度分辨率(fine-grained resolution)分支,$\boldsymbol{X}^{(M-1)}$为高度压缩的粗粒度分辨率(coarse-grained resolution)分支。每个分辨率分支的序列经过线性嵌入(linear embedding)后,得到中间特征图为$\boldsymbol{X}_{2\mathrm{D}}\in\mathbb{R}^{H\times W\times d_{\mathrm{model}}}$,$H$和$W$分别对应重塑后的周期长度与时序步数,$d_{\mathrm{model}}$为特征嵌入空间的表示维度。为了进一步挖掘异构网络流量中交织的多周期特性并消除单一分辨率的局限性,将$\boldsymbol{X}_{2\mathrm{D}}$作为多分辨率混合的时序特征提取模块的输入。由于时域分析难以同时捕获多个尺度的周期模式,本文引入快速傅里叶变换(FFT)对其进行频域分析,通过频谱能量分布直接识别数据中的显著周期成分。FFT 将时域信号分解为不同频率的正弦和余弦分量,使各周期模式在频域中以离散峰值形式显现,如式(1)所示。

$$P_K=TopK(|\boldsymbol{F}|),\ \boldsymbol{F}=FFT(\boldsymbol{X}_{2\mathrm{D}}),\tag{1}$$

其中$\boldsymbol{F}$表示复数频谱,其对应的振幅谱$|\boldsymbol{F}|$反映了不同频率成分的能量分布。通过能量排序算法提取前K个最高能量的主导频率,得到显著周期模式$\{P_1,\cdots,P_k,\cdots,P_K\}$。接下来,针对分辨率第$m$分支,定义$f_{m,k}$为该分支对应的第$k$个主导频率,将时序特征向量重塑为二维矩阵,并进一步补零得到完整的时间特征图像$\boldsymbol{Z}_{m,k}\in\mathbb{R}^{P_k\times f_{m,k}\times d_{\mathrm{model}}}$。

在得到频域增强的多分辨率图像$\boldsymbol{Z}_{m,k}$后,为了对时序序列中相互正交的季节和趋势模式进行显式建模,本文设计了双轴注意力(dual-axis attention)机制。鉴于图像的行对应周期内的时序片段(捕获趋势性),而列代表跨周期的同一时间点(反映季节性),该机制分别沿行轴和列轴进行注意力计算,从而解耦并提取 2 种正交的时间模式。

1)多分辨率季节注意力($Attention_{\mathrm{season}}$)。季节注意力旨在捕捉序列内部重复出现的规律性、周期性波动特征。季节注意力通过计算跨周期相同位置或同一周期内不同时间步之间的关联度,使模型能够精确感知周期性模式的轻微漂移。将多分辨率时间特征$\boldsymbol{Z}_{m,k}$作为自注意力机制$Attention_{\mathrm{season}}$的输入$\boldsymbol{Q}_{\mathrm{S}},\boldsymbol{K}_{\mathrm{S}},\boldsymbol{V}_{\mathrm{S}}$,将时间轴重塑为周期性矩阵,通过捕获相同相对位置点的相关性来提取局部模式,得到多分辨率季节特征表示$\boldsymbol{X}_{\mathrm{S}}$。

$$\boldsymbol{X}_{\mathrm{S}}=Attention_{\mathrm{season}}(\boldsymbol{Q}_{\mathrm{S}},\boldsymbol{K}_{\mathrm{S}},\boldsymbol{V}_{\mathrm{S}})=Softmax\left(\frac{\boldsymbol{Q}_{\mathrm{S}}\boldsymbol{K}_{\mathrm{S}}^{\mathrm{T}}}{\sqrt{d_k}}\right)\boldsymbol{V}_{\mathrm{S}}。\tag{2}$$

2)多分辨率趋势注意力($Attention_{\text{trend}}$)。趋势注意力侧重于建模长期的的方向性演化特征。趋势注意力通过跨越长时周期的依赖建模,可以赋予模型学习长期演化的能力。将多分辨率时间特征$\boldsymbol{Z}_{m,k}$作为$Attention_{\text{trend}}$自注意力机制的输入$\boldsymbol{Q}_{\text{t}},\boldsymbol{K}_{\text{t}},\boldsymbol{V}_{\text{t}}$,在压缩的时间轴上利用自注意力机制捕获趋势特征,得到多分辨率趋势特征表示$\boldsymbol{X}_{\text{t}}$。

$$\boldsymbol{X}_{\text{t}} = Attention_{\text{trend}}(\boldsymbol{Q}_{\text{t}},\boldsymbol{K}_{\text{t}},\boldsymbol{V}_{\text{t}}) = Softmax\left(\frac{\boldsymbol{Q}_{\text{t}}K_{\text{t}}^{\text{T}}}{\sqrt{d_k}}\right)\boldsymbol{V}_{\text{t}}。 \tag{3}$$

因此,通过双轴注意力解耦的趋势和季节性特征输出可进一步表示为:

$$\boldsymbol{X}_{\text{s}},\boldsymbol{X}_{\text{t}} = Dual\text{-}axis\ Attention(\boldsymbol{Z}_{m,k}), \tag{4}$$

其中,季节项$\boldsymbol{X}_{\text{s}}$侧重于局部特征,趋势项$\boldsymbol{X}_{\text{t}}$侧重于全局特征。在经过双轴分解块后,不同时间分辨率的特征通过自底向上(bottom-up)和自顶向下(top-down)的路径进行跨分辨率信息交互。经过h层混合后,得到融合的输出表征$\boldsymbol{Z} = Reshape(\boldsymbol{X}_{\text{s}} + \boldsymbol{X}_{\text{t}})$。对于第$m$个分辨率,特征被重构为$\boldsymbol{Z}^{(m)} \in \mathbb{R}^{T_m \times d_{\text{model}}}$。为了适配分类任务,沿时间维度应用全局平均池化(global average pooling, GAP),将时序特征聚合为定长的分辨率感知特征向量$\boldsymbol{h}^{(m)}$:

$$\boldsymbol{h}^{(m)} = \frac{1}{T_m}\sum_{t=1}^{T_m}\boldsymbol{Z}_{t,:}^{(m)} \in \mathbb{R}^{d_{\text{model}}}, \tag{5}$$

其中,T_m表示第m个分辨率的时间维度长度,$\boldsymbol{Z}_{t,:}^{(m)}$为第$m$个分辨率下重构后的特征图$\boldsymbol{Z}^{(m)}$在特定时间步$t$上的所有特征表示。此外,考虑到传统的多分辨率分类模型通常仅选取$\boldsymbol{h}^{(0)}$进行后续分类,而本文提出的AdaMRL则保留所有分辨率的特征向量的集合$\mathcal{H} = \{\boldsymbol{h}^{(0)},\boldsymbol{h}^{(1)},\cdots,\boldsymbol{h}^{(M-1)}\}$,并在后续阶段进行多分辨率嵌套表示学习,进而将不同分辨率特征自适应映射于不同的表征维度空间。

2.4 多分辨率嵌套表征

考虑到不同分辨率包含的噪声水平与信息密度存在显著差异,强制所有分辨率特征映射到固定维度的高维空间会导致特征冗余与过拟合。因此,本文在特征维度上引入了MRL,使得各分辨率自适应选择合适的维度空间进行MRL。首先,定义一组预设的嵌套特征表示学习维度集合$D = \{d_1,d_2,\cdots,d_n\}$,满足$d_1 < d_2 < \cdots < d_n \leqslant d_{\text{model}}$。对于标准线性分类器$\boldsymbol{W} \in \mathbb{R}^{d_{\text{model}} \times K}$,MRL通过截断权重矩阵的前$d_i$维,即$\boldsymbol{W}_{1:d_i} \in \mathbb{R}^{d_i \times K}$,来生成对应维度的逻辑值(Logits)。在AdaMRL中,所有分辨率共享同一个MRL分类头。对于第m个分辨率的特征向量$\boldsymbol{h}^{(m)}$,模型并行生成n个不同维度的分类预测输出:

$$\hat{\boldsymbol{y}}_{d_i}^{(m)} = \boldsymbol{W}_{1:d_i}\cdot\boldsymbol{h}_{1:d_i}^{(m)} + b_{1:d_i},\quad \forall d_i \in D。 \tag{6}$$

通过这种方式,模型为每一个时间分辨率生成了一个嵌套的分类预测元组$\hat{\boldsymbol{Y}}^{(m)} = (\hat{\boldsymbol{y}}_{d_1}^{(m)},\cdots,\hat{\boldsymbol{y}}_{d_n}^{(m)})$。该设计不仅实现了多分辨率特征与表征空间维度的对齐解耦,且由于权重共享,并未引入额外的模型参数。

2.5 多分辨率联合监督训练

为了促使多分辨率混合的时序特征提取模块能够学习到对各时间分辨率以及各特征维度更鲁棒的通用表示,本文重构了训练阶段的目标函数,设计了一种全分辨率嵌套联合监督机制。对于特定的时间分辨率$m \in \{0,1,\cdots,M-1\}$,其生成的分辨率感知特征向量为$\boldsymbol{h}^{(m)} \in \mathbb{R}^{d_{\text{model}}}$。基于MRL的定义,模型在特征维度上会受到多个嵌套前缀的监督。定义单个分辨率的嵌套损失$\mathcal{L}_{\text{MRL}}^{(m)}$为:

$$\mathcal{L}_{\text{MRL}}^{(m)}(\boldsymbol{h}^{(m)},y;\boldsymbol{W},\{c_i\}) = \sum_{i=1}^{D} c_i\cdot\mathcal{L}_{\text{CE}}(\boldsymbol{W}_{1:d_i}\cdot\boldsymbol{h}_{1:d_i}^{(m)},y), \tag{7}$$

其中,D表示预设的嵌套维度层级总数,$d_i \in D$为第i个嵌套维度的具体大小,满足$d_1 < d_2 < \cdots < d_D$,$\boldsymbol{W}_{1:d_i}$表示共享分类权重矩阵的前d_i维截断,c_i为不同嵌套层级的相对重要性权重系数,用于表示高维空间与低维空间表征的学习强度。

在全局训练过程中,为了实现跨分辨率的特征对齐,本文将所有M个时间分辨率的嵌套损失进行聚合,得到最终的联合训练总损失$\mathcal{L}_{\text{total}}$。

$$\mathcal{L}_{\text{total}} = \frac{1}{M}\sum_{m=0}^{M-1}\mathcal{L}_{\text{MRL}}^{(m)}, \tag{8}$$

其中$\mathcal{L}_{\text{MRL}}^{(m)}$表示第$m$个时间分辨率对应的单个嵌套损失。该损失函数通过内层的$\mathcal{L}_{\text{MRL}}$强制模型将最具辨别性的特征压缩在低维空间中,从而实现推理时的维度自适应截断;通过外层的全分辨率求和,确保模型能够同时从多分辨率的趋势和季节性特征中提取一致的行为语义。这种双重监督机制有效地缓解了深层网络在处理异构、非平稳流量序列时的梯度消失问题,可以提升模型在极端资源受限环境下的分类性能。

2.6 自适应维度选择与集成推理

在推理分类阶段,本文提出了一种基于验证集的自适应维度搜索算法,旨在为不同分辨率量身定制最佳的特征截断维度。

1)配置搜索(验证阶段)。在每一轮训练结束时,模型在验证集$\mathcal{D}_{\text{val}}$上评估所有分辨率、所有维度的组合表现。对于给定的分辨率m,本文搜索能够最大化

验证准确率的最优嵌套维度索引d_m^*。

$$d_m^*=\arg\max_{j\in\{1,\cdots,n\}} Accuracy(\hat{\boldsymbol{y}}_{d_j}^{(m)},y_{\text{val}}),\tag{9}$$

其中，$\hat{\boldsymbol{y}}_{d_j}^{(m)}$表示第$m$个分辨率在第$j$个嵌套维度$d_j$下生成的分类预测输出逻辑值，$y_{\text{val}}$表示验证集上的真实标签，$Accuracy(\cdot)$是在验证集上用于评估组合表现的分类准确率函数。记录最佳配置的列表$C^*=\{d_0^*,d_1^*,\cdots,d_{D-1}^*\}$，并随最佳模型权重一同保存。该机制通常会自动为包含高频噪声的细粒度分辨率选择低维表征空间(正则化降噪)，为粗粒度分辨率(捕捉全局模式)保留高维表征空间。

2)在自适应软投票集成阶段。在对未知测试集进行推理时，模型直接加载最优配置C^*。本文提取每个分辨率在对应最优维度下的逻辑值$\hat{\boldsymbol{y}}_{m_{d_s^*}^*}^{(s)}$，使用Softmax函数将其归一化为概率分布，并进行算术平均(Soft-Voting)，得到最终的集成预测概率$\hat{P}$。

$$\hat{P}=\frac{1}{M}\sum_{s=0}^{M-1}Softmax\left(\hat{\boldsymbol{y}}_{m_{d_s^*}^*}^{(s)}\right),\tag{10}$$

其中$m_{d_s^*}^*$指代第s个分辨率专属的最优特征截断维度。因此，该自适应集成机制有效避免了无效高维特征在集成时引入的分类判断干扰，最终的分类结果$\hat{C}$取决于概率最大的类别：$\hat{C}=\arg\max(\hat{P})$。

3 实　　验

为了全面评估本文提出的AdaMRL框架的性能，本文在10个时间序列分类基准数据集和1个物联网边缘安全基准数据集上设计了5组实验。

1)在10个UEA时序分类基准数据集上的对比实验表明，AdaMRL在整体性能上超越了现有最先进(state-of-the-art, SOTA)方法，验证了模型的有效性和泛化能力。

2)在工业物联网边缘安全基准数据集Edge-IIoTset上的对比实验表明，AdaMRL能够有效检测海量异质终端的网络威胁，验证了模型在真实动态异构网络环境下的准确性和鲁棒性。

3)在4个UEA代表性数据集上进行了消融实验，分别验证了AdaMRL核心模块对模型性能的贡献。

4)在2个UEA代表性数据集上进行了案例分析实验，通过对比共享嵌套维度与独立维度配置的实验结果，验证了自适应特征维度选择机制的有效性，证明AdaMRL针对不同分辨率任务自适应选择了不同的特征维度。

5)从参数量(Params)、浮点运算次数(FLOPs)以及真实边缘设备推理延迟(Inference Latency)三个维度评估了AdaMRL的计算效率。实验结果表明，尽管引入了嵌套表示与多头分类机制，模型仍保持较低的计算开销和参数量，验证了AdaMRL在资源受限的异构网络终端上的部署可行性。

3.1 实验设置

本文主要使用了以下2个公开的基准数据集，如表1所示。

Table 1　Datasets Introduction

表1　数据集介绍

数据集	领域	类别数	变量维度	序列长度	简要介绍
EthanolConcentration	化学/光谱	4	3	1 751	预测水和蜂蜜溶液中的乙醇浓度
FaceDetection	医疗/脑磁图	2	144	62	基于脑磁图的人脸与打乱图像识别
Handwriting	运动/传感器	26	3	152	基于智能手表运动数据的手写字母识别
Heartbeat	医疗/音频	2	61	405	正常与异常心音的听诊音频分类
JapaneseVowels	语音识别	9	12	29	根据语音特征识别日语元音说话人
PEMS-SF	交通/传感器	7	963	144	旧金山湾区高速公路交通占用率分类
SelfRegulationSCP1	医疗/脑电图	2	6	896	脑电信号中皮层慢电位的自我调节分类
SelfRegulationSCP2	医疗/脑电图	2	7	1 152	脑电信号皮层慢电位自我调节 (变体)
SpokenArabicDigits	语音识别	10	13	93	识别十个阿拉伯语口语数字的发音特征
UWaveGestureLibrary	运动/传感器	8	3	315	基于三轴加速度计数据的简单手势识别
Edge-IIoTset	网络流量	15	61	20	边缘计算环境中的网络入侵

1)UEA多变量时间序列分类基准数据集。该数据集是当前时间序列分类领域最权威的评估基准之一，包含30个子数据集，涵盖人类活动识别、心电图分类等多个应用领域。该数据集呈现显著的异构性：样本规模跨度为27~50 000，序列长度范围为8~17 984，特征维度分布在2~1 345之间，能够全面评估分类算

法在不同场景下的泛化能力。本文选择了其中 10 个当前 SOTA 模型常用的子数据集进行实验验证。这些数据集涵盖声音识别(SpokenArabicDigits)、人体动作捕捉(Handwriting、FaceDetection)、医疗生理信号(Heartbeat)等多种真实场景。数据集在序列长度、变量维度和类别分布上呈现显著异构性，能够充分评估模型的泛化能力。

2)Edge-IIoTset 工业物联网边缘安全基准数据集[52]。该数据集通过构建真实异构网络测试床捕获而成，测试床包含物联网传感器、边缘计算节点和核心路由设备。数据集不仅包含正常网络流量，还涵盖 14 种主流网络攻击，包括分布式拒绝服务攻击(DDoS/DoS)、SQL 注入、恶意软件渗透等。

3.2 评价指标

时序多分类任务和网络威胁检测任务常面临严重的数据不平衡问题，例如正常流量远多于攻击流量。因此，本文采用基于混淆矩阵的多种评估指标，而非单一准确率。设 *TP*(true positive)为真正例，*TN*(true negative)为真负例，*FP*(false positive)为假正例，*FN*(false negative)为假负例，核心指标定义如下：

1)准确率(*Accuracy*)。表示模型预测正确的样本占所有测试样本的比例，用于衡量分类的正确度：

$$Accuracy=\frac{TP+TN}{TP+TN+FP+FN}。\tag{11}$$

2)精确率(*Precision*)。指模型预测为某类样本中，真正属于该类的比例：

$$Precision=\frac{TP}{TP+FP}。\tag{12}$$

3)召回率(*Recall*)。指在所有正例样本中，被模型成功分类的比例，直接反映了模型的检出能力：

$$Recall=\frac{TP}{TP+FN}。\tag{13}$$

4)*F*1-Score。由于 *Precision* 和 *Recall* 在实际中往往存在制约关系，*F*1-Score 作为两者的调和平均值，能够对模型在类别不平衡数据集上的综合性能做出最客观的判定：

$$F1\text{-Score}=2\times\frac{Precision\times Recall}{Precision+Recall}。\tag{14}$$

5)误报率(false positive rate, *FPR*)。表示本属于某一类的正例样本却被模型错误地标记为其他类别比例。极低的误报率是系统能够落地部署的前提。

$$FPR=\frac{FP}{FP+TN}。\tag{15}$$

鉴于 AdaMRL 在 Edge-IIoTset 数据集上是对正常流量和 14 种网络攻击流量进行多分类识别，为了全面反映整体性能，本文在计算总体的 *Accuracy*、*Precision*、*Recall* 和 *F*1-Score 时采用加权平均(weighted average)，即将每个类别的指标乘以该类别在测试集中的样本权重后求和。

3.3 模型泛化性实验

为了全面且严谨地评估 AdaMRL 的分类性能，本文选取了 9 个在时序分类领域极具竞争力的基准模型进行对比。这些模型涵盖了经典深度网络模型(LSSL, M.TCN)、线性模型(DLinear)、基于 Transformer 模型(FEDformer, PatchTST, iTransformer)与基于多分辨率模型(TimesNet, MPTSNet, TimeMixer++)。

1)LSSL[53] 状态空间深度模型通过连续时间系统的离散化，能够高效地捕捉时序信号中的长程依赖；

2)FEDformer[54] 提取基准模型，通过频域分析或全卷积网络构建坚实的底层时序表征底座；

3)PatchTST[55] 是基于 Transformer 架构的前沿模型，创新性地采用通道独立与切块(Patching)机制，大幅提升了对局部时序语义的捕获能力；

4)DLinear[56] 是极简且高效的线性架构代表，利用移动平均方法将时序强行解耦为趋势与季节成分，在多项任务中展现出超越复杂模型的鲁棒性；

5)TimesNet[4] 是多周期架构的代表作，通过快速傅里叶变换将一维时序重塑为二维张量，并利用 2D-CNN 提取复杂的跨周期时空依赖；

6)iTransformer[57] 是采用“变量倒置”策略的 Transformer 模型，将每个时间变量的完整序列视为独立的 Token 进行注意力计算，有效缓解了多变量之间的特征干扰；

7)M.TCN[58] 是时间卷积网络，通过多分辨率的大核卷积机制，在保持计算线性复杂度的同时显著扩大了特征感受野；

8)MPTSNet[59] 是多周期时序网络架构，通过强化多分辨率下的周期感知与全局演化模式提取，在复杂分类任务中表现卓越；

9)TimeMixer++[6] 是当前时序学术界顶尖的多分辨率架构，通过跨分辨率的“过去可分解混合机制”，实现了宏观全局趋势与微观局部波动的深度融合。

表 2 给出了 AdaMRL 与 9 种基准模型在 10 个 UEA 多变量时间序列数据集上的分类准确率对比。整体来看，AdaMRL 的平均准确率达到 76.13%，高于 MPTSNet 的 75.35% 和 TimeMixer++的 75.29%，说明所提方法在不同类型的多元变量时间序列分类任务中具有较好的整体竞争力。进一步分析各个子数据集的结果，AdaMRL 的性能表现与数据集的内存特征

Table 2 Accuracy of Multivariate Time Series Classification

表 2 多元时间序列分类准确率 %

数据集	FEDformer（ICLR'22）	LSSL（ICLR'22）	PatchTST（ICLR'23）	DLinear（AAAI'23）	TimesNet（ICLR'23）	iTransformer（ICLR'24）	M.TCN（ICLR'24）	MPTSNet（AAAI'25）	TimeMixer++（ICLR'25）	AdaMRL（本文）
EthanolConcentration	31.2	31.1	32.8	36.2	35.7	28.1	36.3	43.3	39.9	**43.8**
FaceDetection	66.0	66.7	68.3	68.0	68.6	66.3	70.8	69.8	**71.8**	69.4
Handwriting	28.0	24.6	29.6	27.0	32.1	24.2	30.6	34.4	26.5	**36.6**
Heartbeat	73.7	72.7	74.9	75.1	78.0	75.6	77.2	75.6	**79.1**	79.0
JapaneseVowels	98.4	98.4	97.5	96.2	98.4	96.6	**98.8**	98.6	97.9	97.8
PEMS-SF	80.9	86.1	89.3	75.1	89.6	87.9	89.1	**94.2**	91.0	92.5
SelfRegulationSCP1	88.7	90.8	90.7	87.3	91.8	90.2	93.4	92.8	93.1	**94.2**
SelfRegulationSCP2	54.4	52.2	57.8	50.5	57.2	54.4	60.3	57.2	**65.6**	60.1
SpokenArabicDigits	**100.0**	**100.0**	98.3	81.4	99.0	96.0	98.7	99.5	99.8	99.5
UWaveGestureLibrary	85.3	85.9	85.8	82.1	85.3	85.9	86.7	88.1	88.2	**88.4**
平均准确率/%	70.66	70.85	72.50	67.89	73.57	70.52	74.19	75.35	75.29	**76.13**
本模型占优的数据集数量	8	8	10	10	9	10	7	7	6	

注：最优结果用黑体数值表示，次优结果用下划线数值表示。

密切相关，包括时序结构复杂度、噪声水平、多尺度特征分布，以及类别判别模式等。

首先，AdaMRL 更适用于局部动态波动明显、细粒度扰动与全局趋势并存的数据集。例如，在 Handwriting 和 UWaveGestureLibrary 等运动传感器数据集中，样本往往同时包含短时局部抖动、动作轨迹变化和较长时间范围内的整体运动趋势。固定维度的多分辨率模型容易将不同分辨率分支映射到相同表征维度，导致局部噪声过表达或全局趋势表达不足。相比之下，AdaMRL 通过多分辨率分支和自适应嵌套维度选择，为不同时间分辨率匹配不同的维度表示，从而更好地平衡局部细节与全局模式。在 Handwriting 数据集上，AdaMRL 取得 36.6% 的准确率，高于次优 MPTSNet 的 34.4%；在 UWaveGestureLibrary 数据集上，AdaMRL 取得 88.4% 的准确率，也高于 TimeMixer++的 88.2%。这表明，当数据中存在显著的多尺度动态变化且不同分辨率分支具有不同特征维度需求时，AdaMRL 的分辨率和维度自适应匹配机制更容易发挥作用。

然而，在周期结构较强、领域先验较明确的数据集上，AdaMRL 的优势并不一定明显。例如，在 JapaneseVowels 和 SpokenArabicDigits 等语音识别数据集上，FEDformer、LSSL、M.TCN、MPTSNet 或 TimeMixer++等方法取得了更高或相近的结果。这是因为 JapaneseVowels 序列长度较短、语音模式相对规整，已有方法对说话人声学特征和局部时序模式的建模已经较为充分；SpokenArabicDigits 中部分基准模型已达到接近饱和的准确率，AdaMRL 的自适应维度选择难以进一步带来显著增益。因此，对于类别边界清晰、周期模式稳定或已有模型已接近性能上限的数据集，AdaMRL 的提升幅度相对有限。

更为典型的是，在 Heartbeat 等蕴含严格且复杂周期性演化规律的数据集上，TimeMixer++和 MPTSNet 表现更为出色。这是由于心跳本质上是多频段叠加的生理声学信号，其判别性信息主要分布在频域的周期性模式中。TimeMixer++通过快速傅里叶变换将一维序列重塑为二维时频张量，利用多分辨率时间图像显式提取跨周期的时空依赖，这种基于频域的二维空间周期捕获能力在处理纯声学和生理周期信号时，比 AdaMRL 侧重时域分辨率混合的架构更具针对性和有效性。

上述分析说明，AdaMRL 的贡献不应仅理解为平均准确率的小幅提升，而在于提供了一种面向多尺度异构时序数据的分辨率与维度自适应匹配机制。AdaMRL 的优势主要体现在多尺度时序特征复杂、判别性互补且时域空域依赖显著的场景中，并在保证性能提升的同时显著提高了模型效率。

3.4 动态异质网络场景实验

为了进一步验证 AdaMRL 框架在真实动态异构网络环境下针对海量异质终端威胁检测的有效性，本文在前沿的工业物联网边缘安全基准数据集 Edge-IIoTset 上进行了对比实验。与早期的 UNSW-NB15[60] 和 Bot-IoT[61] 相比，该数据集包含了更多样化的异质工业互联网终端流量模式。在基准模型上，本实验

不仅对比了基于时间序列分类的模型，还对比了网络威胁检测的先进模型 YaTC。以这类在网络安全领域处于 SOTA 级别的专用深度模型作为基线，能够直接验证 AdaMRL 是否具备在特定工业物联网威胁检测场景下落地部署的实战能力。

与传统的同构网络流量数据集不同，Edge-IIoTset 包含了高度动态与异构的终端遥测数据。该数据集不仅涵盖了传统的 IT 网络协议（如 TCP, UDP, HTTP, DNS），还深度融合了物联网轻量级通信协议（MQTT）以及工业控制底层协议（Modbus/TCP）。这种异构的底层特征（如 tcp.window_size, mqtt.topic_len, mbtcp.len 等）映射了真实云边端网络中智能工控设备、传感器与边缘网关的复杂通信场景。由于 Edge-IIoTset 原始数据以离散网络流（Flow）记录形式存在，不能直接作为多元时间序列输入。本文首先按照 7∶3 比例将流统计特征数据划分为训练集和测试集。随后，分别在训练集和测试集内部进行滑动窗口重构，将该窗口内的流统计特征转换为固定长度的多元时间序列样本。具体而言，设窗口长度 $T=20$，滑动步长 $s=10$，每个窗口内包含连续的网络流量特征序列；窗口标签采用多数投票策略确定，即选取窗口内出现次数最多的流级标签作为该窗口的样本标签。最终，重构后的样本以.ts 格式存储，并作为 Edge-IIoTset 时序分类的输入。实验任务是多分类（共 15 个类别），将终端流量划分为正常流量（Normal）以及包含 DDoS、SQLInjection、Ransomware 等在内的 14 种攻击类型。

考虑到 Edge-IIoTset 不平衡的数据分布特性，本实验重点对比了 AdaMRL 与 6 种基准模型的加权准确率、精确率、召回率、*F*1-Score、宏平均准确率，以及安全领域最为关注的误报率，另外还针对异常流量中样本占比较高和较低的 7 种网络攻击做了详细的指标对比。

如表 3 所示，AdaMRL 在该高度异构的物联网数据集上展现出最优性能：整体加权准确率达到 99.41%，*F*1-Score 达到 99.40%，更关键的是误报率仅 0.12%。由于 Edge-IIoTset 存在显著类别不均衡，准确率和 *F*1-Score 可能受到 Normal 多数类样本影响。为进一步验证 AdaMRL 对少数攻击类别的检测能力，本文计算了宏平均准确率并统计各类别的检测结果。由表 3 可见，AdaMRL 在保持 99.41% 的准确率和 99.40% 的 *F*1-Score 的同时，宏平均准确率达到 92.04%，说明模型性能并非仅由 Normal 多数类样本贡献。

此外，Edge-IIoTset 的逐类检测结果表明，AdaMRL

Table 3 Performance of Models and Attacks on Edge-IIoTset Dataset

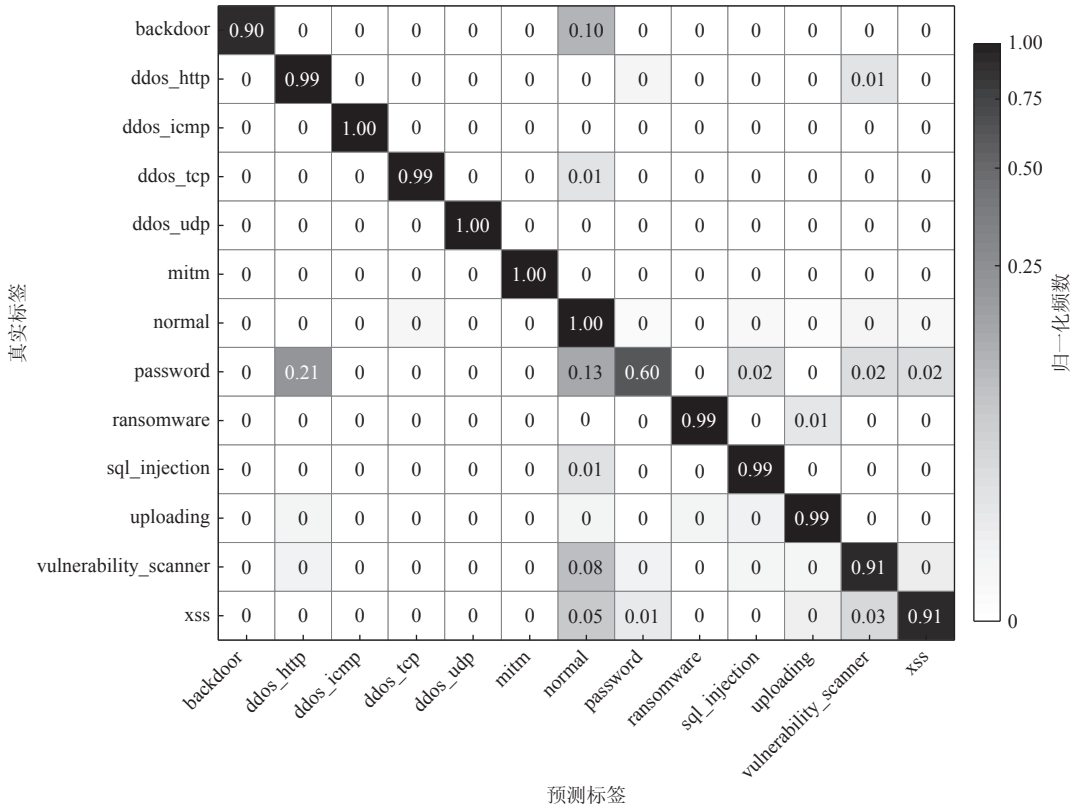
表 3 模型和攻击在数据集 Edge-IIoTset 上的表现

模型名称	准确率/%	精确率/%	召回率/%	*F*1-Score/%	误报率/%	宏平均准确率/%
AdaMRL（本文）	99.41	99.39	99.41	99.40	0.12	92.04
YaTC	98.55	98.72	98.68	98.70	1.61	91.48
1D-CNN	96.18	98.01	96.75	97.08	5.38	91.18
CNN-GRU	96.38	97.32	96.84	96.68	3.21	89.62
CNN-LSTM	96.35	97.29	96.52	96.43	3.27	89.50
TCN	96.72	98.28	97.16	97.21	3.82	91.21
攻击	**准确率/%**	**精确率/%**	**召回率/%**	***F*1-Score/%**	**误报率/%**	**样本数量**
DDoS_UDP	100.00	100.00	100.00	100.00	0.00	30 392
DDoS_ICMP	100.00	100.00	100.00	100.00	0.00	16 985
SQLinject	99.93	98.48	98.57	98.53	0.37	12 706
DDoS_TCP	99.92	97.70	98.88	98.29	0.53	12 515
Backdoor	90.00	90.00	90.00	90.00	10.00	33
MITM	100.00	100.00	100.00	100.00	0.00	90
Password	78.05	78.05	60.38	68.09	39.62	183

对 DDoS_UDP、DDoS_ICMP、DDoS_TCP、MITM、SQLinject 等攻击类别仍保持较高 *F*1-Score 值，说明自适应嵌套维度选择并未削弱具有判别意义的短时攻击突变特征的表征能力。对于 Password 类别，由于 Password 攻击的载荷通常隐藏在合法的 TCP 连接中，缺乏 DDoS 攻击那样显著的流量突发特征。这种低频高隐的特性使得短时窗口内的流统计特征（如流速率）不足以触发明显的判别信号，导致模型对其敏感度下降，这也说明 AdaMRL 对极少数攻击类别仍有进一步优化空间。

图 2 所示的行归一化混淆矩阵进一步表明，在以真实类别为基准进行归一化后，矩阵中每一行的数值总和为 1，各攻击类别在不同预测类别上的分布比例能够直观反映模型对该类别样本的识别准确性。通过观察可以发现，少数类攻击在对角线位置（即正确分类）上仍保持较高的占比，并未被系统性地误判为 Normal 类，说明 AdaMRL 在面对类别分布不均衡时，仍能较好地保持对少数类攻击特征的敏感性与区分能力。

这一结果表明，当面对多源异构、特征稀疏、信噪比差异巨大的混合时序数据时，传统固定维度的深度模型容易因过度拟合某些高频微小扰动而产生误报。而 AdaMRL 得益于其新颖的弹性嵌套表示与自适应维度匹配机制，能够有效地在细粒度时间分辨率上有效捕捉具有判别意义的短时突变（Burst），



注：*dm* 为模型特征维度，*df* 为前馈网络隐藏层维度。

Fig. 2 Edge-IIoTset normalized confusion matrix (*dm*=32, *df*=128)

图 2 Edge-IIoTset 归一化混淆矩阵（*dm*=32，*df*=128）

并在粗粒度分辨率上刻画攻击行为的全局时序特征。这种机制在保障对复杂注入攻击和泛洪攻击的高检出率的同时，最大程度地降低了误报率。

真实物联网环境中的网络流量数据往往充斥着强异质性与不规则性。例如，底层流量日志解析过程中，时间戳字段常混杂特定年份后缀、不规则空格以及时区偏移。这种非标准格式不仅增加了特征提取难度，也使解析后的时序数据上存在微观不规则抖动。本实验在流量数据上的优异结果表明，面对包含微观时间错位和异步采样的异构序列，传统注意力切块机制或线性分解极易发生特征对齐失败；而 AdaMRL 通过多分辨率嵌套架构，在粗粒度层级能够吸收底层解析带来的时序抖动，展现出较好的预处理鲁棒性。

AdaMRL 性能优异的另一个关键点在于其多分辨率联合损失监督实现了时间分辨率与特征维度的双向对齐解耦。无论底层攻击流量来自高频触发的网络探针，还是低频交互的物联网网关，AdaMRL 都能在宏观维度上捕获协同攻击时的演化共性，从而大幅提升模型在动态异构网络环境下的多维协同推理与分类能力。

3.5 消融实验

为了验证 AdaMRL 中各个核心组件的贡献，本文在 4 个代表性数据集 FaceDetection, Handwriting, Heartbeat, JapaneseVowels 上进行了消融实验。本文定义了 4 种模型消融变体。1）w/o Multi-Resolution: 不使用多分辨率，仅使用单一的精细分辨率；2）w/o Adaptive: 保留多分辨率，但强制所有分辨率使用最大特征维度进行集成；3）w/o JointLoss: 不使用联合监督损失函数，仅对主分辨率进行单头损失函数监督。4）w/o Soft-Voting: 不使用等权软投票机制，直接采用最佳的单个分辨率分支的分类结果。

消融实验结果如表 4 和图 3 所示，去除任何一个核心组件都会导致模型性能下跌。

1）多分辨率交互机制的有效性。在“w/o Multi-Resolution”变体中，移除了平均池化下采样模块与多分辨率机制，模型仅在单一精细时间分辨率上利用双轴注意力进行特征提取。实验结果表明，该变体的分类平均准确率从 70.6% 下降至 65.1%（下降 5.5 个百分点）。这一性能退化源于异构物联网流量的时间非平稳特性。单一高分辨率视图易陷入高频瞬态噪声，导致模型误报率上升。引入多分辨率混合

Table 4 Accuracy of AdaMRL in Ablation Study

表 4 AdaMRL 在消融实验中的准确率 %

数据集	原模型	w/o Multi-Resolution	w/o Adaptive	w/o Joint Loss	w/o soft-Voting
FaceDetection	69.4	65.6	66.2	67.5	69.1
Handwriting	36.6	24.4	25.4	30.1	32.1
Heartbeat	78.5	76.6	76.6	78.1	76.6
JapaneseVowels	97.8	93.8	92.7	96.5	96.3
平均准确率	70.6	65.1	65.2	68.1	68.5
Δ		5.5	5.4	2.5	2.1

注: “Δ”表示变体与原模型的平均准确率下降百分点。

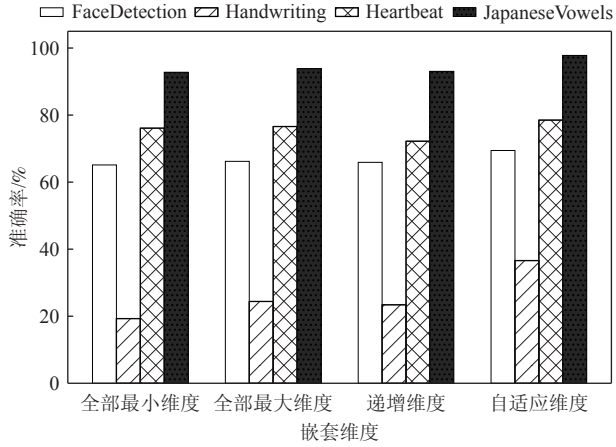


Fig. 3 Results of ablation study on nested dimension configuration

图 3 嵌套维度配置消融实验结果

机制后，自顶向下（top-down）与自底向上（bottom-up）的跨分辨率信息交互使模型能够同时捕获微观的包级别突变与宏观的演变趋势，从而显著增强对网络威胁行为的辨别能力。

2）自适应维度选择的决定性作用。在 w/o Adaptive 中，4 个数据集的平均准确率从 70.6% 下降至 65.2%（下降 5.4 个百分点）。这验证了本文提出的分辨率和维度双向解耦的必要性。此外，为了进一步验证嵌套维度配置的必要性，本文比较了 4 种维度配置方式：全部最小维度、全部最大维度、递增维度和自适应维度。AdaMRL 模型使用的是自适应维度配置，由验证集为每个分辨率分支选择最优嵌套维度。如图 3 所示，自适应维度配置在多个代表性数据集上取得最佳结果。最小维度配置由于表征容量不足，在复杂数据集上性能受限；最大维度配置虽然提供了更高容量，但可能引入冗余特征，性能并不稳定；递增维度能够部分缓解固定维度问题，但仍依赖人工规则，难以适配不同数据集的特征结构。相比之下，自适应维度配置能够根据不同分辨率分支的特征需求动态选择维度，因此表现更优。这说明 AdaMRL 中的自适应嵌套维度选择并非简单增加模型复杂度，而是缓解固定维度映射和人工维度规则局限性的关键设计。

3）联合损失的正则化效应。在 w/o Joint Loss 中，平均准确率下降至 68.1%（下降 2.5 个百分点）。这表明，在训练阶段对每个分辨率独立施加嵌套分类头的损失约束，能够有效防止深层网络在不同粒度特征上的表征退化。

4）多分辨率集成策略有效性。为进一步验证多分辨率分支融合方式的合理性，本文比较了硬投票、最优单分支和 soft-Voting 三种集成策略。硬投票仅基于各分支最终类别标签进行多数决策，忽略了类别概率分布中的置信度信息；最优单分支仅使用验证集表现最好的单一分辨率分支，无法利用其他分辨率分支中互补的局部或全局特征；soft-Voting 则对各分辨率分支的类别概率进行平均融合，能够保留更丰富的预测置信度信息。

如图 4 所示，soft-Voting 在 4 个代表性数据集上均取得最高的分类准确率，整体优于硬投票和最优单分支。该结果说明，多分辨率分支之间存在互补关系，仅依赖单一最佳分支会损失部分尺度信息；同时，概率级融合比类别级硬投票更能利用各分支的判别置信度。由于 soft-Voting 无需引入额外可学习参数或复杂权重估计过程，因此在提升分类性能的同时保持了推理阶段的简洁性，符合边缘部署场景对模型复杂度的要求。

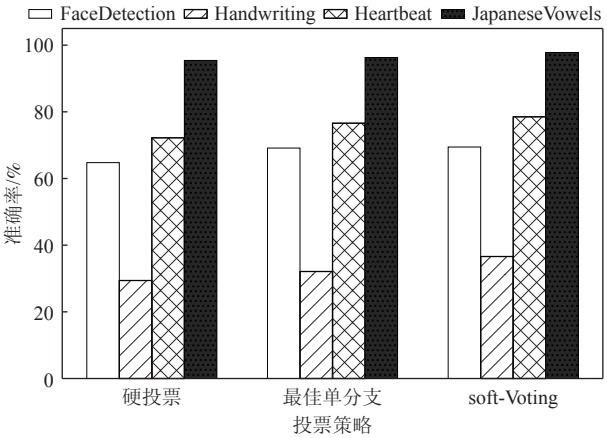


Fig. 4 Results of ablation study on Ensemble strategy

图 4 投票策略消融实验结果

3.6 自适应嵌套维度选择分析实验

为进一步分析 AdaMRL 的自适应嵌套维度选择机制，本文统计了不同分辨率分支与不同嵌套维度组合下的分类准确率，结果如图 5 所示。其中，Res0

表示原始或最高时间分辨率分支, Res3 表示经过多次下采样后的最低时间分辨率分支。需要说明的是, 嵌套维度选择作用于分辨率分支输出后的特征向量, 通过选择不同的嵌入维度来调节特征表达能力, 而不是直接对原始时间序列进行滤波, 因此选择不同的嵌套维度并不等价于时间域低通滤波。

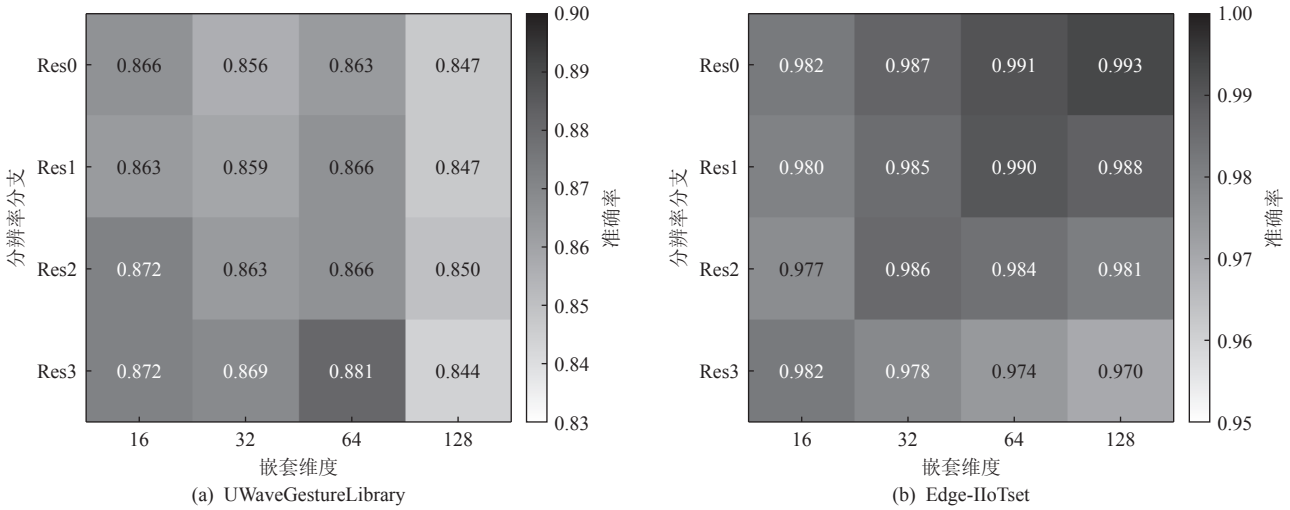


Fig. 5 Classification accuracy heatmaps with combinations of multi-resolution branch and nested dimension

图 5 多分辨率分支与嵌套维度组合下的分类准确率热力图

由图 5 可以看出, 2 个数据集呈现出不同的最优组合模式。在 UWaveGestureLibrary 中, 最优结果出现在 Res3 分支和 64 维嵌套维度组合下, 准确率为 0.881, 说明对于存在局部抖动的动作传感器数据, 低分辨率分支有助于弱化非判别性局部扰动, 并突出整体动作趋势。而在 Edge-IIoTset 中, 最优结果出现在 Res0 分支和 128 维嵌套维度组合下, 准确率为 0.993, 说明在网络攻击检测场景中, 短时突变和局部异常本身具有重要判别意义, 高分辨率分支和较高嵌套维度更有利于保留细粒度攻击特征。模型在验证集上的最优配置索引揭示了这一现象的内在机制: 不同数据集、不同分辨率分支对应的最优嵌套维度并不完全一致。这种分辨率与嵌套维度自适应匹配机制是 AdaMRL 性能提升的关键。

综合分析 2 个数据集结果可以发现, 最优分辨率与嵌套维度组合并不存在固定模式。对于局部扰动较多、整体轨迹趋势更具判别性的运动传感器数据, 低分辨率分支和中等嵌套维度更有利; 而对于高频突变本身具有判别意义的网络攻击数据, 高分辨率分支和较高嵌套维度更有利。该结果说明, AdaMRL 的自适应机制并不是简单地通过低维截断削弱高频信息, 而是根据验证集分类结果为不同数据集和不同分辨率分支选择合适的嵌套维度, 从而实现分辨率与特征维度之间的自适应匹配。

为进一步验证不同时间分辨率分支与特征维度之间存在匹配关系, 本文对各分辨率分支在全局平均池化之后、MRL 分类头之前的特征矩阵进行奇异值分解(singular value decomposition, SVD), 并计算累积解释方差(cumulative explained variance, CEV)。对于第m个分辨率分支, 记其特征矩阵为$\boldsymbol{H}^{(m)}\in\mathbb{R}^{N\times d}$。其中, N为样本数量, d为基础特征维度。首先对$\boldsymbol{H}^{(m)}$进行中心化处理:

$$\tilde{\boldsymbol{H}}^{(m)}=\boldsymbol{H}^{(m)}-\mathbf{1}\boldsymbol{\mu}^{(m)}, \tag{16}$$

其中, $\boldsymbol{\mu}^{(m)}$ 表示第m个分辨率分支特征矩阵在样本维度上的均值向量, $\mathbf{1}$ 为全 1 列向量。随后, 对中心化后的特征矩阵进行 SVD 分解:

$$\tilde{\boldsymbol{H}}^{(m)}=\boldsymbol{U}^{(m)}\boldsymbol{\Sigma}^{(m)}\boldsymbol{V}^{(m)\mathrm{T}}, \tag{17}$$

其中, $\boldsymbol{U}^{(m)}$和$\boldsymbol{V}^{(m)}$分别为左奇异向量矩阵和右奇异向量矩阵, $\boldsymbol{\Sigma}^{(m)}$为对角奇异值矩阵, 其对角元素满足$\sigma_1^{(m)}\geqslant\sigma_2^{(m)}\geqslant\cdots\geqslant\sigma_r^{(m)}\geqslant 0$,其中$r=\mathrm{rank}(\tilde{\boldsymbol{H}}^{(m)})$。从数学意义上看, SVD 将特征矩阵分解为一组相互正交的主方向, 奇异值$\sigma_i^{(m)}$反映了第i个主方向所包含的特征变化强度。由于奇异值平方与该方向上的方差贡献成正比, 因此可以利用$\sigma_i^{(m)^2}$衡量该维度方向对整体特征分布的解释能力。第i个主方向的解释方差比定义为:

$$EVR_i^{(m)}=\frac{\sigma_i^{(m)^2}}{\sum_{j=1}^{r}\sigma_j^{(m)^2}}。 \tag{18}$$

进一步地, 前K个主方向的累积解释方差定义为:

$$CEV_K^{(m)} = \sum_{i=1}^{K} EVR_i^{(m)} = \frac{\sum_{i=1}^{K} \sigma_i^{(m)^2}}{\sum_{j=1}^{r} \sigma_j^{(m)^2}} 。 \quad (19)$$

$CEV_K^{(m)}$表示仅保留前 K个主要特征方向时，能够解释第m个分辨率分支整体特征变化的比例。本文将达到 95% *CEV* 所需的最小维度定义为该分辨率分支的最小充分表示维度：

$$d_{95}^{(m)} = \min\left\{K \mid CEV_K^{(m)} \geqslant 0.95\right\} 。 \quad (20)$$

若不同分辨率分支的CEV曲线上升速度和$d_{95}^{(m)}$存在差异，则说明不同时间分辨率分支的特征信息在维度方向上的分布并不一致。换言之，某些分辨率分支的主要判别信息可能集中在较少维度中，而另一些分辨率分支则需要更多维度才能充分表达其特征变化。因此，SVD-CEV 分析可以从特征分布角度刻画时间分辨率与嵌套维度之间的耦合关系。时间分辨率的变化不仅改变了时间模式的尺度，也会影响该分支特征在维度空间中的集中程度和充分表示所需维度。若对所有分辨率分支强制采用统一固定维度，则可能导致部分分支维度不足，或部分分支维度冗余，从而造成分辨率和维度的错配。

以 Handwriting 数据集为例，如图 6 所示，不同分辨率分支的 CEV 曲线上升速度存在明显差异。最高分辨率分支 Res0 达到 0.95 的 *CEV*，需要 10 个维度，而 Res 1、Res 2 和 Res 3 分别需要 9、8 和 7 个维度。该结果表明，不同分辨率分支的充分表示维度并不一致，时间分辨率与特征维度之间存在明显耦合关系。也就是说，不同分辨率分支并不适合简单映射到相同固定维度空间，而需要根据其特征分布差异自适应选择合适的嵌套维度。该分析从特征矩阵的主方向分布角度进一步验证了 AdaMRL 中分辨率和维度自适应匹配机制的必要性。

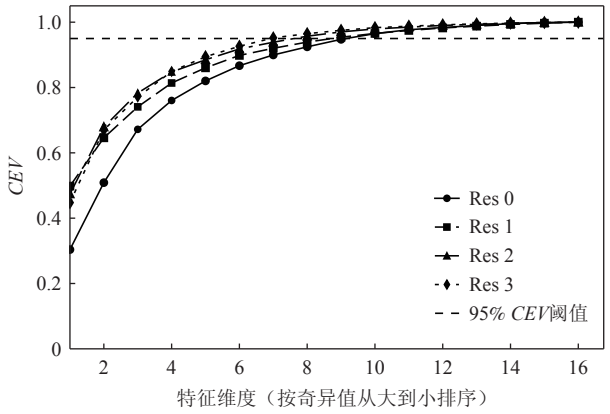


Fig. 6 Comparison of using shared nested dimensions and using different dimensions alone

图 6 采用共享嵌套维度和单独采用不同维度的对比

3.7 模型效率实验

在动态异构网络中，时序分类模型需部署至算力受限、功耗敏感的边缘异构终端，以执行实时威胁检测与异常防护。然而，传统深度多分辨率时序模型通常将所有时间分辨率的序列映射到统一的高维特征空间，这不仅引入大量冗余计算，还导致内存占用过高，制约了分布式边缘安全体系的部署。因此，本节从参数量（Params）、浮点运算次数（FLOPs）以及推理延迟（Inference Latency）三个维度，对 AdaMRL 的模型效率进行分析。

尽管 AdaMRL 引入了多分辨率多头分类机制，但得益于 MRL 的共享权重特性，模型在保持高精度的同时展现出优异的参数量和计算效率。首先，所有时间分辨率在特征映射阶段共享同一分类矩阵 $\boldsymbol{W}$，使得参数量相较于单头网络无显著增加。其次，自适应机制在验证阶段已筛选出最优截断维度，因此在推理时模型仅需执行部分矩阵乘法运算。

如表 5 所示，与使用全维度特征的 TimesNet、PatchTST 和 TimeMixer++相比，AdaMRL 在准确率更高的同时，浮点运算次数降低了 27.3%，推理吞吐量提升了 31.5%。这表明 AdaMRL 不仅性能优异，在实际部署中也具备显著的轻量化优势。

Table 5 Comparison of Model Efficiency and Computational Cost in UWaveGestureLibrary Dataset

表 5 在 UWaveGestureLibrary 数据集中对模型效率和计算成本的比较

模型	参数量	FLOPs	吞吐量/(seq·s^{-1})	准确率/%
DLinear	0.12×10^6	0.25×10^9	9 200	82.1
TimesNet	2.25×10^6	1.15×10^9	4 782	85.3
PatchTST	3.85×10^6	0.97×10^9	3 925	85.8
TimeMixer+	3.70×10^6	1.22×10^9	3 646	88.2
AdaMRL（本文）	1.97×10^6	0.81×10^9	5 412	88.4

为进一步验证 AdaMRL 在边缘资源受限环境下的推理可行性，本文在 UWaveGestureLibrary 数据集上构建 CPU-only 推理评测，并与原始 TimeMixer++在相同主干超参数下进行对比。实验仅包含评估推理阶段，不包含训练和反向传播；推理批大小设为 1，以模拟边缘节点在线接收单个时序窗口并实时分类的场景。资源约束设置为 R1_light、R2_medium 和 R3_strict 三档，分别对应 4/2/1 个 CPU 线程和 2 048/1 024/

512 MB 内存预算，每个档位重复 3 次，表中报告均值，实验结果如表 6 所示。资源受限 CPU-only 推理实验进一步表明，AdaMRL 在 1 线程、512 MB 内存预算下仍能完成单样本推理，并相较 TimeMixer 同时降低延迟、峰值内存和单位样本能耗，验证了其在边缘异构终端上的部署潜力。

Table 6 Comparison of Reasoning Performance in Resource Constrained Environments

表 6 资源受限环境下的推理性能对比

资源档位	模型	准确率/%	每样本平均延迟/ms	P50/ms	P95/ms	吞吐量（samples/s）	峰值 RSS/MB	内存预算通过	每样本能耗/J
R1_light	AdaMRL（本文）	88.44	25.5	25.3	26.4	38.0	275.468	是	0.395
	TimeMixer++	78.44	27.7	27.8	29.3	34.8	557.456	是	0.431
R2_medium	AdaMRL（本文）	88.44	34.3	34.6	36.0	28.5	275.324	是	0.352
	TimeMixer++	78.44	35.9	35.6	38.5	27.1	561.870	是	0.369
R3_strict	AdaMRL（本文）	88.44	46.6	46.6	47.7	21.0	273.781	是	0.237
	TimeMixer++	78.44	48.6	48.2	50.5	20.1	558.530	否	0.249

注：P50 表示中位数延迟，P95 表示 95% 分位延迟。

3.8 参数敏感性实验

为了深入评估 AdaMRL 框架在不同超参数配置下的鲁棒性与性能变化规律，本文在 Heartbeat 代表性数据集上进行了参数敏感性分析，如表 7 所示。本文重点考察了 3 个核心超参数对模型分类性能的影响：基础特征维度 d_{model}、多分辨率下采样层数 L 以及嵌套层级数 M。

Table 7 Parameter Sensitivity Analysis of AdaMRL on Heartbeat Dataset

表 7 Heartbeat 数据集上对 AdaMRL 的参数敏感性分析

超参数	取值	准确率/%	Δ
d_{model}	2	76.6	−2.4
	4	77.6	−1.4
	8	79.0	0.0
	16	76.1	−2.9
L	1	75.6	−3.4
	2	76.1	−2.9
	3	79.0	0.0
	4	76.6	−3.4
M	1	73.9	−5.1
	2	77.4	−1.6
	4	79.0	0.0
	8	79.1	+0.1

注：“Δ”表示与最优准确率之间的差值。

1）基础特征维度 d_{model} 的影响

d_{model} 决定了模型基础特征空间的容量。传统的深度时序模型在维度大幅缩减时往往会面临严重的表征能力坍缩。然而，从表 7 可以观察到，当 d_{model} 从 128 降至 32 时，AdaMRL 的准确率下降幅度极为平缓（例如在 Heartbeat 上仅下降了约 2.9 个百分点）。这一卓越的维度鲁棒性主要归功于本文引入的 MRL。MRL 强制模型在低维前缀空间中同样学到极具判别性的特征，使得 AdaMRL 即使在极度受限的计算资源（低维度）下，依然能够保持高效的威胁检测与分类能力。

2）多分辨率下采样层数 L 的影响

L 决定了模型在时间轴上提取粗粒度宏观特征的深度。随着 L 的增加，模型逐渐捕捉到更长周期的全局趋势。如表 7 所示，模型的分类准确率呈现出先升后降的趋势。当 L=2（包含 3 个时间分辨率）时，模型达到了性能峰值。这表明，适度的下采样能够形成更长时间尺度的分辨率分支，有助于模型补充全局趋势信息；但下采样层数过多时，局部突变和短时异常模式可能被过度压缩，导致分类性能下降。得益于自适应维度选择机制，AdaMRL 能够通过为过粗分辨率分配极低权重或截断至极低维度，有效抑制了深层下采样带来的负面影响。

3）嵌套层级数 M 的影响

M 控制着特征维度空间的搜索粒度。M=1 相当于退化为不具备自适应截断能力的传统固定维度模型。实验结果表明，当 M 从 1 提升至 4 时，模型的准确率获得了显著的跃升（提升约 5.1 个百分点）。这是因为更多的嵌套层级为自适应搜索算法提供了更丰富的维度截断选项，使模型能够更精准地为不同分辨率的信噪比匹配最优维度。当 M 继续增加至 8 时，性能增益逐渐趋于饱和。综合考虑计算开销与分类精度，本文在主实验中将 M=4 作为最佳的超参数平衡点。

4 结　论

本文针对动态异构网络环境下多元时间序列分类任务中时间多分辨率特征与嵌入维度对齐解耦的问题，提出了自适应多分辨率表示学习框架（AdaMRL）。该框架实现了时间维度和特征维度的联合解耦。具体而言：1）引入多分辨率嵌套表示学习架构，为每个时间分辨率独立配备嵌套表示学习层，使模型能够在不同分辨率下自适应匹配特征维度；2）设计多分辨率联合损失监督机制，在训练阶段对全分辨率下的嵌套特征进行同步约束，提升了跨分辨率特征的鲁棒性与收敛效率；3）提出自适应维度选择与集成策略，在推理阶段为不同分辨率动态匹配最优特征维度，并通过软投票机制实现协同决策。

在 10 个 UEA 多变量时间序列基准数据集上的实验表明，AdaMRL 在分类准确率上达到了先进水平，平均准确率提升 0.78 个百分点。在 Edge-IIoTset 数据集上，模型实现了 99.41% 的准确率与 0.12% 的误报率，验证了其在动态异构网络威胁检测中的有效性。消融实验、自适应维度分析及效率对比进一步证明了各核心模块的有效性以及模型的计算效率和资源消耗优势。

综上所述，AdaMRL 在准确率与效率上达到了较优平衡，为动态异构网络的安全智能系统提供了可扩展的技术方案。然而，本文研究仍存在以下局限：1）模型在极端资源受限的边缘设备（如 MCU）上的部署能力尚未充分验证；2）自适应维度选择机制依赖验证集，在小样本场景下可能不稳定；3）当前框架主要面向分类任务，对时序预测等其他任务的适用性有待进一步探索。未来工作将重点关注模型轻量化与跨任务泛化能力的提升。

作者贡献声明：刘瑶提出方法思路和整体设计；张栏弋实现方法与论文撰写；鲁梦瑶、侯睿、姜睿临、文宇博负责部分算法和实验；丁熠提出指导意见。

参　考　文　献

[1] Al-Garadi M A, Mohamed A, Al-Ali A K, et al. A survey of machine and deep learning methods for Internet of things (IoT) security[J]. IEEE Communications Surveys & Tutorials, 2020, 22(3): 1646−1685

[2] Landauer M, Skopik F, Stojanović B, et al. A review of time-series analysis for cyber security analytics: From intrusion detection to attack prediction[J]. International Journal of Information Security, 2024, 23(6): 3895−3923

[3] Wang Shiyu, Wu Haixu, Shi Xiaoming, et al. TimeMixer: Decomposable multiscale mixing for time series forecasting[C]//Proc of the Int Conf on Learning Representations (ICLR). Appleton, WI: ICLR, 2024: 4166−4192

[4] Wu Haixu, Hu Tengge, Liu Yong, et al. TimesNet: Temporal 2D-variation modeling for general time series analysis[C]//Proc of the Int Conf on Learning Representations (ICLR). Appleton, WI: ICLR, 2023: 6423−6445

[5] Ren Liqiang, Jia Shuyi, Wang Haipeng, et al. A survey of time series classification based on deep learning[J]. Journal of Electronics & Information Technology, 2024, 46(8): 3094−3116 (in Chinese)
(任利强, 贾舒宜, 王海鹏, 等. 基于深度学习的时间序列分类研究综述[J]. 电子与信息学报, 2024, 46(8): 3094−3116)

[6] Wang Shiyu, Li Jiawei, Shi Xiaoming, et al. Timemixer++: A general time series pattern machine for universal predictive analysis[C]//Proc of the Int Conf on Learning Representations (ICLR). Appleton, WI: ICLR, 2025: 1662−1698

[7] Kusupati A, Bhatt G, Rege A, et al. Matryoshka representation learning[C]//Proc of the 36th Annual Conf on Neural Information Processing Systems (NeurIPS). Red Hook, NY: Curran Associates, 2022: 30233−30249

[8] Yang Qiang, Wu Xindong. 10 challenging problems in data mining research[J]. International Journal of Information Technology & Decision Making, 2006, 5(4): 597−604

[9] Liu Ying, Li Xu, Lü Zheng, et al. Time series classification method based on deep residual probability random forest[J]. Control and Decision, 2024, 39(7): 2315−2324 (in Chinese)
(刘颖, 李旭, 吕政, 等. 基于深度残差概率随机森林的时间序列分类方法[J]. 控制与决策, 2024, 39(7): 2315−2324)

[10] Fawaz H I, Forestier G, Weber J, et al. Deep learning for time series classification: A review[J]. Data Mining and Knowledge Discovery, 2019, 33(4): 917−963

[11] Rajkomar A, Oren E, Chen K, et al. Scalable and accurate deep learning with electronic health records[J]. NPJ Digital Medicine, 2018, 1(1): 18−27

[12] Ding Biyun, Zhang Tao, Wang Chao, et al. Acoustic scene classification: A comprehensive survey[J]. Expert Systems with Applications, 2024, 238: 121902

[13] Azab A, Khasawneh M, Alrabaee S, et al. Network traffic classification: Techniques, datasets, and challenges[J]. Digital Communications and Networks, 2024, 10(3): 676−692

[14] Lines J, Bagnall A. Time series classification with ensembles of elastic distance measures[J]. Data Mining and Knowledge Discovery, 2015, 29(3): 565−592

[15] Kampouraki A, Manis G, Nikou C. Heartbeat time series classification with support vector machines[J]. IEEE Transactions on Information Technology in Biomedicine, 2008, 13(4): 512−518

[16] Jeong Y S, Jeong M K, Omitaomu O A. Weighted dynamic time warping for time series classification[J]. Pattern Recognition, 2011, 44(9): 2231−2240

[17] Dau H A, Bagnall A, Kamgar K, et al. The UCR time series

archive[J]. IEEE/CAA Journal of Automatica Sinica, 2019, 6(6): 1293−1305

[18] Bagnall A, Dau H A, Lines J, et al. The UEA multivariate time series classification archive, 2018[J]. arXiv preprint arXiv: 1811.00075, 2018

[19] Zhao Bendong, Lu Huanzhang, Chen Shangfeng, et al. Convolutional neural networks for time series classification[J]. Journal of Systems Engineering and Electronics, 2017, 28(1): 162−169

[20] Hüsken M, Stagge P. Recurrent neural networks for time series classification[J]. Neurocomputing, 2003, 50: 223−235

[21] Karim F, Majumdar S, Darabi H, et al. LSTM fully convolutional networks for time series classification[J]. IEEE Access, 2017, 6: 1662−1669

[22] Foumani N M, Tan C W, Webb G I, et al. Improving position encoding of transformers for multivariate time series classification[J]. Data Mining & Knowledge Discovery, 2024, 38(1): 1−22

[23] Liu Huaiyuan, Yang Donghua, Liu Xianzhang, et al. Todynet: temporal dynamic graph neural network for multivariate time series classification[J]. Information Sciences, 2024, 677: 120914

[24] Fawaz H I, Lucas B, Forestier G, et al. InceptionTime: Finding AlexNet for time series classification[J]. Data Mining and Knowledge Discovery, 2020, 34(6): 1936−1962

[25] Nie Yuqi, Nguyen N H, Sinthong P, et al. A time series is worth 64 words: Long-term forecasting with Transformers[C]//Proc of the Int Conf on Learning Representations (ICLR). Appleton, WI: ICLR, 2023: 33132−33155

[26] Nussbaumer H J. The Fast Fourier Transform[M]//Fast Fourier Transform and Convolution Algorithms. Berlin: Springer, 1981: 80−111

[27] Zeng Ailing, Chen Muxi, Zhang Lei, et al. Are Transformers effective for time series forecasting?[C]//Proc of the 37th AAAI Conf on Artificial Intelligence (AAAI). Palo Alto, CA: AAAI, 2023: 11121−11128

[28] Neyadi E A, Shehhi S A, Shehhi A A, et al. Discovering public Wi-Fi vulnerabilities using raspberry pi and kali Linux[C]//Proc of the 12th Annual Undergraduate Research Conf on Applied Computing (URC). Piscataway, NJ: IEEE, 2020: 1−4

[29] Li Junfeng, Li Dan, Huang Yukai, et al. High-performance and scalable network address translation system on general-purpose platforms[J]. Journal of Beijing University of Posts and Telecommunications, 2021, 44(2): 14−19 (in Chinese)
(李峻峰, 李丹, 黄昱恺, 等. 通用平台高性能可扩展网络地址转换系统[J]. 北京邮电大学学报, 2021, 44(2): 14−19)

[30] Nascita A, Aceto G, Ciuonzo D, et al. A survey on explainable artificial intelligence for internet traffic classification and prediction, and intrusion detection[J]. IEEE Communications Surveys & Tutorials, 2024, 27(5): 3165-3198.

[31] Nascita A, Aceto G, Ciuonzo D, et al. A survey on explainable artificial intelligence for internet traffic classification and prediction, and intrusion detection[J]. IEEE Communications Surveys & Tutorials, 2024, 27(5): 3165-3198

[32] Moore A W, Papagiannaki K. Toward the accurate identification of network applications[C]//Proc of Int Workshop on Passive and Active Network Measurement. Berlin: Springer, 2005: 41−54

[33] Madhukar A, Williamson C. A longitudinal study of P2P traffic classification[C]//Proc of the 14th IEEE Int Symp on Modeling, Analysis, and Simulation. Piscataway, NJ: IEEE, 2006: 179−188

[34] Sarker I H, Kayes A S M, Watters P. Effectiveness analysis of machine learning classification models for predicting personalized context-aware smartphone usage[J]. Journal of Big Data, 2019, 6(1): 1−28

[35] Fernandes S, Antonello R, Lacerda T, et al. Slimming down deep packet inspection systems[C]//Proc of /IEEE INFOCOM Workshops 2009. Piscataway, NJ: IEEE, 2009: 1−6

[36] Wang Xiaofei, Jiang Junchen, Tang Yi, et al. StriD^{2}FA: Scalable regular expression matching for deep packet inspection[C]//Proc of 2011 IEEE Int Conf on Communications (ICC). Piscataway, NJ: IEEE, 2011: 1−5

[37] Qi Jinshan, Liang Xun, Li Zhiyu, et al. Representation learning of large-scale complex information network: Concepts, method and challenges[J]. Chinese Journal of Computers, 2018, 41(10): 2394−2420 (in Chinese)
(齐金山, 梁循, 李志宇, 等. 大规模复杂信息网络表示学习: 概念、方法与挑战[J]. 计算机学报, 2018, 41(10): 2394−2420)

[38] Wang Zhanyi. The applications of deep learning on traffic identification[J]. BlackHat USA, 2015, 24(11): 1−10

[39] Lotfollahi M, Jafari Siavoshani M, Shirali Hossein Zade R, et al. Deep packet: A novel approach for encrypted traffic classification using deep learning[J]. Soft Computing, 2020, 24(3): 1999−2012

[40] Bu Zhiyong, Zhou Bin, Cheng Pengyu, et al. Encrypted network traffic classification using deep and parallel network-in-network models[J]. IEEE Access, 2020, 8: 132950−132959

[41] Zhao Ruijie, Zhan Mingwei, Deng Xianwei, et al. Yet another traffic classifier: A masked autoencoder based traffic transformer with multi-level flow representation[C]//Proc of the AAAI Conf on Artificial Intelligence. Palo Alto, CA: AAAI, 2023, 37(4): 5420−5427

[42] Zhan Mingming, Yang Jin, Jia Dongqing, et al. EAPT: An encrypted traffic classification model via adversarial pre-trained transformers[J]. Computer Networks, 2025, 257: 110973

[43] Wang Tongze, Xie Xiaohui, Wang Wenduo, et al. Netmamba: Efficient network traffic classification via pre-training unidirectional mamba[C]//Proc of 2024 IEEE 32nd Int Conf on Network Protocols (ICNP). Piscataway, NJ: IEEE, 2024: 1−11

[44] Wickramasinghe N, Shaghaghi A, Tsudik G, et al. SoK: Decoding the enigma of encrypted network traffic classifiers[C]//Proc of 2025 IEEE Symp on Security and Privacy (SP). Piscataway, NJ: IEEE, 2025: 1825−1843

[45] Huang Xunhua, Zhang Fengbin, Fan Haoyi, et al. Multimodal adversarial learning based unsupervised time series anomaly detection[J]. Journal of Computer Research and Development, 2021, 58(8): 1655−1667 (in Chinese)
(黄训华, 张凤斌, 樊好义, 等. 基于多模态对抗学习的无监督时间序列异常检测[J]. 计算机研究与发展, 2021, 58(8): 1655−1667)

[46] Kusupati A, Bhatt G, Rege A, et al. Matryoshka representation learning[J]. Advances in Neural Information Processing Systems, 2022, 35: 30233−30249

[47] Hu W, Dou Z Y, Li L H, et al. Matryoshka query Transformer for large vision-language models[J]. Advances in Neural Information

Processing Systems, 2024, 37: 50168−50188

[48] Cai Mu, Yang Jianwei, Gao Jianfeng, et al. Matryoshka multimodal models[C]//Proc of Int Conf on Learning Representations. WI: ICLR, 2025: 46254−46272

[49] Wan Chen, Li Wenzhong, Ding Wangxiang, et al. A multivariate time series forecasting algorithm based on self-evolution and pretraining[J]. Chinese Journal of Computers, 2022, 45(3): 513−525 (in Chinese)
(万晨, 李文中, 丁望祥, 等. 一种基于自演化预训练的多变量时间序列预测算法[J]. 计算机学报, 2022, 45(3): 513−525)

[50] Bachmann R, Allardice J, Mizrahi D, et al. Flextok: Resampling images into 1D token sequences of flexible length[C]//Proc of the 42nd Int Conf on Machine Learning. PMLR, 2025, 267: 2241−2295

[51] Yi Liping, Yu Han, Ren Chao, et al. Federated model heterogeneous Matryoshka representation learning[J]. Advances in Neural Information Processing Systems, 2024, 37: 66431−66454

[52] Gharib A, Bakhit A, Bostani H, et al. Edge-IIoTset: A new comprehensive realistic cyber security dataset of IoT and IIoT applications[J]. IEEE Access, 2022, 10: 3101−3123

[53] Gu A, Goel K, Re C. Efficiently modeling long sequences with structured state spaces[C]//Proc of the 10th Int Conf on Learning Representations (ICLR). Appleton, WI: ICLR, 2022, 14323−14349

[54] Zhou Tian, Ma Ziqing, Wen Qingsong, et al. FEDformer: Frequency enhanced decomposition Transformer for long-term series forecasting[C]//Proc of the 39th Int Conf on Machine Learning (ICML). Brookline, MA: PMLR, 2022: 27268−27286

[55] Huo Menghao, Lu Kuan, Li Yuxiao, et al. CT-PatchTST: Channel-time patch time-series Transformer for long-term renewable energy forecasting[C]//Proc of the 2025 10th Int Conf on Computer and Information Processing Technology (ISCIPT). Piscataway, NJ: IEEE, 2025: 86–95

[56] Liu Yong, Hu Tengge, Zhang Haoran, et al. iTransformer: Inverted transformers are effective for time series forecasting[C]//Proc of the Int Conf on Learning Representations (ICLR). Appleton, WI: ICLR, 2024: 4004−4028

[57] Luo Donghao, Zhang Xinfeng, Wu Chuanyan, et al. ModernTCN: A modern pure convolution structure for general time series analysis[C]//Proc of the 12th Int Conf on Learning Representations (ICLR). Appleton, WI: ICLR, 2024: 34209−34251

[58] Yang M, Shahzad M, Zhu X X. MptsNet: Integrating multiscale periodic local patterns and global dependencies for multivariate time series classification[J]. arXiv preprint, arXiv: 2503.05582, 2025

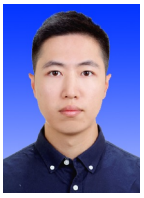
[59] Moustafa N, Slay J. UNSW-NB15: A comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set)[C]//Proc of the 2015 Military Communications and Information Systems Conf (MilCIS). Piscataway, NJ: IEEE, 2015: 1−6

[60] Koroniotis N, Moustafa N, Sitnikova E, et al. Towards the development of realistic BOTNET dataset in the Internet of things for network forensic analytics: Bot-IoT dataset[J]. Future Generation Computer Systems, 2019, 100: 779−796

[61] Wang Di, Qiao Ying. A review of key technologies for time series databases[J]. Journal of Computer Research and Development, 2024, 61(3): 603−625 (in Chinese)
(王迪, 乔颖. 时序数据库关键技术综述[J]. 计算机研究与发展, 2024, 61(3): 603−625)



Liu Yao, born in 1978. PhD, associate professor. Member of CCF. Her main research interests include multimodal knowledge representation and reasoning, and network security.
刘　瑶, 1978 年生。博士, 副教授。CCF 会员。主要研究方向为多模态知识表征与推理、网络安全。



Zhang Lanyi, born in 2002. Master candidate. His main research interests include time series classification and network security.
张栏弋, 2002 年生。硕士研究生。主要研究方向为时序分类、网络安全。



Lu Mengyao, born in 2003. Master candidate. Her main research interests include time series prediction and network security.
鲁梦瑶, 2003 年生。硕士研究生。主要研究方向为时序预测、网络安全。



Hou Rui, born in 1995. PhD candidate. His main research interests include time series forecasting and spatio-temporal data mining.
侯　睿, 1995 年生。博士研究生。主要研究方向为时序预测、时空数据挖掘。



Jiang Ruilin, born in 2003. Master candidate. His main research interest includes time series reasoning.
姜睿临, 2003 年生。硕士研究生。主要研究方向为时间序列推理。



Wen Yubo, born in 2003. Master candidate. His main research interests include knowledge representation and event reasoning.
文宇博, 2003 年生。硕士研究生。主要研究方向为知识表征、事件推理。



Ding Yi, born in 1985. PhD, professor. Member of CCF. His main research interests include multimodal learning, computer vision, and knowledge reasoning.
丁　熠, 1985 年生。博士, 教授。CCF 会员。主要研究方向为多模态学习、计算机视觉、知识推理。