

企业网络故障预测与威胁识别的多模态自监督蒸馏方法

曾 浩

(中国融通集团信息技术有限公司, 北京 100089)

摘要: 为解决现代企业网络中故障预测与安全威胁识别面临的数据稀缺、模型泛化能力弱及部署效率低等问题, 提出一种基于多模态自监督表征与跨域知识蒸馏的企业网络故障预测与安全威胁识别统一框架; 具体地, 提出包含跨模态一致性约束的增强型多模态掩码自编码, 并引入随机模态丢弃以提升缺模态鲁棒性; 同时提出类条件相关性对齐, 结合置信筛选与类别重加权实现类别粒度的二阶统计对齐; 在目标域上以少量标注进行多任务联合学习, 并配合温度缩放完成概率校准与阈值决策; 实验显示, 相对最佳公开基线在数据集上 $RMSE$ 下降约 6.9%、 F_1 提升约 2.1%, 低标注比例更具优势, 并在缺模态与跨域场景中保持稳定; 同时, 蒸馏得到的轻量学生模型便于部署到边缘与资源受限环境, 在保持主要精度指标的条件下, 对缺模态与跨域漂移表现出较好的稳健性, 并通过可解释性分析与运维指标验证了该方法在实际应用中的高效性与可行性。

关键词: 多模态表征; 自监督学习; 知识蒸馏; 故障预测; 威胁识别

A Multimodal Self-Supervised Distillation Method for Enterprise Network Fault Prediction and Threat Identification

ZENG Hao

(China Rongtong Group Information Technology Co., Ltd., Beijing 100089, China)

Abstract: To address data scarcity, weak model generalization, and low deployment efficiency of fault prediction and security threat detection methods in modern enterprise networks, this paper proposes a unified framework based on the multimodal self-supervised representation learning and cross-domain knowledge distillation. Specifically, an enhanced multimodal masked autoencoder with cross-modal consistency constraints is presented, which introduces a random modality dropout to improve their robustness under missing modalities. Meanwhile, a class-conditioned correlation alignment is proposed, which achieves second-order statistical alignment at the category granularity by combining confidence-based filtering with class reweighting. A multi-task joint learning method is conducted on the target domain with limited labeled data, which completes probability calibration and threshold selection by temperature scaling. Experiments show that, compared with the best public baselines, this method decreases the $RMSE$ by about 6.9% and increases the F_1 by about 2.1% on the dataset, with more pronounced advantages under low-label settings, maintaining stability in missing-modality and cross-domain scenarios. Moreover, it is convenient for the lightweight student model obtained by the distillation to deploy in edge and resource-constrained environments while maintaining the key accuracy metrics, which shows a good robustness for the missing modalities and cross-domain shifts, further verifying its effectiveness and feasibility in real-world applications through interpretability analysis and O&M metrics.

Keywords: multimodal representation; self-supervised learning; knowledge distillation; fault prediction; threat identification

收稿日期:2025-09-22; 修回日期:2025-11-07。

作者简介:曾 浩(1985-),男,博士,高级工程师。

引用格式:曾 浩. 企业网络故障预测与威胁识别的多模态自监督蒸馏方法[J]. 计算机测量与控制, 2026, 34(5): 146-153, 162.

0 引言

在企业数字化转型的浪潮下，现代企业网络已演变为一个复杂的信息物理系统，承载着海量的业务数据流，并持续产生包括网络流量、系统日志、设备性能指标在内的多模态数据^[1]。这些数据构成了企业网络运行状态的全景视图，为实现智能化的网络运维提供了前所未有的机遇。从技术视角看，对企业网络的有效管理依赖于两个核心环节：一是对多模态数据的持续、精准测量，以全面感知网络状态；二是在精确测量的基础上，进行主动、智能的控制，以保障网络服务的连续性和安全性^[2]。而实现主动控制的关键，在于对潜在的网络故障进行提前预测，并对各类安全威胁进行准确识别。然而，在实践中，网络故障管理与安全威胁应对往往被视为两个独立的领域，忽视了二者之间内在的深刻联系，例如分布式拒绝服务（DDoS, distributed denial of service）攻击在网络指标上的表现与网络拥塞等性能故障高度相似^[3]。因此，传统的单任务、单模态分析方法难以捕捉故障与威胁之间的复杂关联，一个能够融合多模态数据、协同分析性能与安全事件的统一框架，对于提升网络运维的整体效能至关重要^[4-5]。

在国内外相关研究中，传统的网络故障预测方法多依赖于时间序列分析模型，如自回归积分滑动平均模型（ARIMA, auto regressive integrated moving average），而威胁识别则依赖于基于规则或签名的模式匹配，这些方法在处理现代网络数据的高维度、高动态性和非线性特征时显得力不从心^[6]。近年来，以循环神经网络（RNN, recurrent neural network）及其变体和 Transformer 模型为代表的深度学习技术，凭借其强大的非线性建模能力，在网络数据分析领域取得了显著进展。尽管如此，深度学习的应用仍面临三大挑战。首先是数据稀缺性问题，高性能的深度学习模型通常需要大量高质量的标注数据，而在企业网络环境中，获取带有精确故障或攻击标签的数据集成本高昂且耗时^[7]。其次是模型泛化性问题，由于不同企业网络的拓扑结构、业务模式存在显著差异，在一个网络环境（源域）中训练的模型，直接应用于另一个网络环境（目标域）时，往往会因领域漂移问题而导致性能急剧下降^[8]。最后是模型效率问题，为了追求高精度，深度学习模型通常具有庞大的参数量和高昂的计算复杂度，意味着难以在资源受限或对实时性要求极高的监控场景中进行部署^[9]。与此同时，现有多模态自监督/预训练方法大体可以分为两类：一类以多模态掩码自编码方法为代表，采用“逐模态掩码一同模态重构”的范式，跨模态语义往往依赖共享编码器被动获得；另一类则通过模态 dropout 或随机遮蔽提升对缺模态输入的鲁棒性，但并未针对企业网络这类

时间同步的多源运维数据做显式的跨模态对齐建模。

为应对上述挑战，本文提出一种基于多模态自监督表征与跨域知识蒸馏（MM-SSKD, multi-modal self-supervised knowledge distillation）的统一框架。方法以增强型多模态掩码自编码作为教师侧自监督核心，在大规模无标签数据上学习稳健的共享表征；在迁移阶段结合条件相关性对齐（C-CORAL, class-conditional correlation alignment）与知识蒸馏（KD, knowledge distillation），将教师的跨域判别能力压缩至轻量学生，最终通过统一的多任务目标同时建模故障回归与威胁分类。

1 基于多模态自监督表征与跨域知识蒸馏的方法

本文提出的 MM-SSKD 框架旨在构建兼具准确性与高效性的智能网络分析模型，其核心思想是将复杂的、计算密集型的知识学习过程与轻量化的、快速响应的实际部署过程相解耦。该框架包含两个阶段，如图 1 所示。

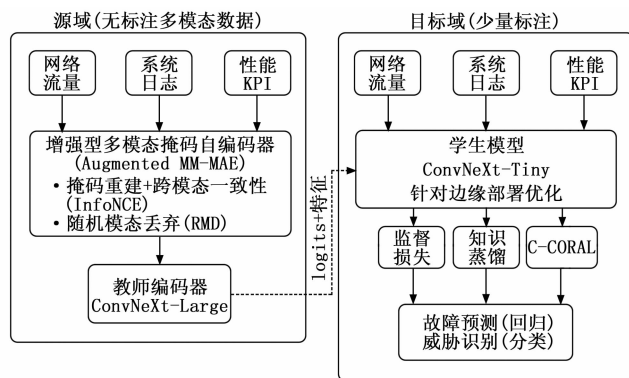


图1 本文 MM-SSKD 框架

阶段一：教师模型自监督预训练。选用基于 ConvNeXt^[10] 的大容量模型作为教师，在多源海量无标签的通用网络数据（源域）上训练。训练不直接面向下游任务，而是通过增强型多模态掩码自编码（Augmented MM-MAE）这一自监督代理任务，学习网络数据的内在结构与跨模态关联，获得具有泛化能力的特征表示。

阶段二：学生模型跨域知识蒸馏与微调。选用与教师结构相似但更小的轻量模型作为学生，面向特定企业网络环境（目标域）部署。在目标域少量有标签数据上，以复合损失进行训练。学生在继承教师表征能力的同时，快速适应目标域分布，以较低计算成本获得高性能。

1.1 多模态网络数据预处理与表征

为了让模型能够有效处理来源多样、格式各异的网络数据，首先需要对数据进行标准化的预处理与表征。本文主要关注以下 3 种核心模态的数据。

1) 网络流量数据：主要来源于网络流量包（PCAP）或流记录（NetFlow）。通过解析得到协议类型、源/目

的 IP 与端口、包数量、字节数、连接持续时间、包长序列等结构化特征。

2) 系统日志数据: 来源于防火墙、入侵检测系统、服务器及各类应用, 通常为非结构化或半结构化文本。采用词嵌入 (Word Embedding) 将每条日志消息映射为一个固定维度的向量。

3) 性能指标数据: 主要指通过 SNMP 或监控代理采集的设备时间序列, 如 CPU 利用率、内存占用率、磁盘 I/O 速率、网络接口带宽等, 这些数据直接反映设备与服务器的健康状况。

完成各模态的特征提取与向量化后, 按实体标识与时间戳进行对齐, 并采用固定长度的滑动时间窗口聚合; 对采样率不同的性能指标使用线性插值补齐。最终在特征维上拼接得到统一多模态张量 $X \in \mathbb{R}^{N \times T \times M \times D}$, 其中 N 为批次大小, T 为时间序列长度, M 为模态数量, D 为每个模态的特征维度。

为提升在不完整采集场景下的鲁棒性, 训练阶段引入随机模态丢弃, 即在每个批次中以概率 $p \in [0.1, 0.2]$ 随机将某一或若干模态在整个窗口置零, 促使模型学习利用其余模态进行推理。

1.2 增强型多模态掩码自编码自监督学习

在统一的窗口级多模态张量上, 本文提出包含跨模态一致性约束的增强型多模态掩码自编码代理任务 (Augmented MM-MAE), 用于对大容量教师网络进行自监督预训练, 使其在无标签条件下学习跨模态的共享语义与可迁移表征。与标准 MM-MAE^[11] 相比, 本文在“逐模态掩码—重构”的基础流程上引入对比式跨模态一致性学习, 并在训练时叠加随机模态丢弃 (RMD, random modal dropout)^[12], 以显式暴露“缺模态/弱模态”情形, 增强在真实运维中常见的数据不完整场景下的鲁棒性。教师网络采用 ConvNeXt^[10] 的高容量骨干, 以提升表征上限并稳定跨模态特征, 为后续蒸馏提供高质量源表示。

具体地, 对第 i 个窗口, 模态 m 的时序特征 $X_i^{(m)}$ 以掩码率 ρ_m 采样二值掩码 $M_i^{(m)}$, 得到遮蔽输入 $\tilde{X}_i^{(m)} = [1 - M_i^{(m)}] \odot X_i^{(m)}$, 经共享编码器—轻量解码器还原出 $\hat{X}_i^{(m)}$ 。本文仅在遮蔽位置计算重构误差并在模态上累加:

$$\mathcal{L}_{\text{MM-MAE}} = \sum_m \frac{\|M_i^{(m)} \odot [X_i^{(m)} - \hat{X}_i^{(m)}]\|_2^2}{\max(1, \|M_i^{(m)}\|_0)} \quad (1)$$

单纯的重构会促使模型在各模态内学习到“可还原”的低级统计, 但未必能捕获“跨源一致”的高级语义。为此, 将各模态的编码特征经投影头 $g^{(m)}$ 凝练为窗口级嵌入 $h_i^{(m)} = g^{(m)}\{E[\tilde{X}_i^{(m)}]\}$, 并以“同窗不同模态”为正对、“批内异窗”为负对构造跨模态对比任务, 采用 InfoNCE^[13] 作为一致性项:

$$L_{\text{CM}} = -\frac{1}{|P|} \sum_{(i,a,b)} \log \frac{\exp[\text{sim}(h_i^{(a)}, h_i^{(b)})/\tau]}{\sum_j \exp\{\text{sim}[h_i^{(a)}, h_j^{(b)}]/\tau\}} \quad (2)$$

$$\text{sim}(u, v) = \frac{u^T v}{\|u\| \|v\|} \quad (3)$$

其中温度 τ 控制分布平滑度与负样本难度。该对比目标可视作最大化不同模态间的互信息下界, 从优化层面把“同一事件的多源观测”在表示空间内拉近, 并自然压制“跨实例的偶然同形”带来的虚假关联。最终的自监督目标为:

$$\mathcal{L}_{\text{SSL}} = \mathcal{L}_{\text{MM-MAE}} + \gamma \mathcal{L}_{\text{CM}} \quad (4)$$

其中: $\gamma > 0$ 平衡“可还原性”与“跨模态一致性”的强弱。通过式 (4) 的联合训练, 教师在不依赖标签的情况下同时学习到“模态内可逆结构”和“模态间语义配准”。

1.3 跨域知识蒸馏与类条件域对齐

在教师模型通过自监督学习掌握了通用网络知识后, 下一步是通过知识蒸馏 (KD) 与类条件相关性对齐 (C-CORAL) 策略, 将知识高效地迁移给轻量化的学生模型。学生模型采用与教师模型同构但层数更少、通道数更窄的 ConvNeXt-T^[10] 架构, 以保证特征层对齐的结构兼容和部署效率。

训练目标由监督项贴合任务指标 $\mathcal{L}_{\text{task}}$, 蒸馏项承接教师软知识 $\mathcal{L}_{\text{KD}}$, 对齐项在类别粒度上修复域偏移 $\mathcal{L}_{\text{align}}$ 三部分组成。首先给出总目标函数, 随后分别解释各分量的意义和实现:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{task}} + \mathcal{L}_{\text{KD}} + \mathcal{L}_{\text{align}} \quad (5)$$

1) 监督任务 (Task)。目标域同时包含故障预测 (连续量) 与威胁识别 (离散类) 两类下游需求, 采用多任务学习框架, 用可调权重平衡两者, 使训练过程在全局上既关注连续指标的数值误差, 又关注分类边界的可分性与对长尾类别的敏感度。具体定义如下:

$$\mathcal{L}_{\text{task}} = \alpha \mathcal{L}_{\text{pred}} + \beta \mathcal{L}_{\text{cls}} \quad (6)$$

$$\mathcal{L}_{\text{pred}} = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2} \quad (7)$$

$$\mathcal{L}_{\text{cls}} = -\frac{1}{N} \sum_{i=1}^N a_{t,i} (1 - p_{t,i})^{\gamma_f} \log p_{t,i} \quad (8)$$

直观地, $\mathcal{L}_{\text{pred}}$ (RMSE) 对大偏差更为敏感, 能抑制预测的系统性漂移; $\mathcal{L}_{\text{cls}}$ 采用 Focal Loss, 通过焦点参数 γ_f 抑制易分类样本的主导效应, 并用类权重 $a_{t,i}$ 处理类别不均衡与长尾问题。

2) 知识蒸馏 (KD)。仅靠目标域少量标注, 很难把学生模型直接推到“通用且鲁棒”的解。为此, 采用双路径蒸馏: 一条在响应层 (logits) 上对齐, 传递教师对类间相似性的“软判断”; 另一条在特征层对齐, 传递教师在中间层形成的空间结构与判别线索。训练时

教师端停止梯度，学生端以蒸馏与监督联合反向传播：

$$\mathcal{L}_{KD} = \lambda_{\text{logits}} \mathcal{L}_{\text{logits_KD}} + \lambda_{\text{feat}} \mathcal{L}_{\text{feat_KD}} \quad (9)$$

$$\mathcal{L}_{\text{logits_KD}} = T^{\gamma} \text{KL} \left[\text{softmax} \left(\frac{\mathbf{z}_t}{T} \right) \middle| \text{softmax} \left(\frac{\mathbf{z}_s}{T} \right) \right] \quad (10)$$

$$\mathcal{L}_{\text{feat_KD}} = \sum_l \frac{1}{d_l} \|\mathbf{W}_l \Phi_t^{(l)} - \Phi_s^{(l)}\|_2^2 \quad (11)$$

其中： $\mathbf{z}_t$ 和 $\mathbf{z}_s$ 分别为教师和学生的 logits 向量， $T > 0$ 为蒸馏温度， $\Phi_t^{(l)}$ 和 $\Phi_s^{(l)}$ 分别为第 l 层的教师和学生特征， $\mathbf{W}_l$ 是将教师特征映射到学生维度的线性变换， λ_{logits} 和 λ_{feat} 为两条蒸馏路径的权重。

3) 类条件域对齐 (C-CORAL)。源域与目标域的数据统计常存在系统差异。如果只做整体分布对齐 (如传统 CORAL^[14])，往往会过度平滑类间结构，在类分布不均或多峰时尤其明显。为此，本文提出在类别粒度上对齐协方差，即对每个类别分别估计源/目标域的二阶统计并最小化其差异；目标域缺标签时，用教师预测生成伪标签，只选择置信度高于阈值 τ 的样本参与对齐，并通过类别权重抑制不平衡带来的估计方差。除特别说明外，统一将伪标签置信度阈值设为 $\tau = 0.8$ ，类别权重按通过阈值后的目标域样本数做逆比例缩放并归一化：

$$\mathcal{L}_{\text{align}} = \frac{1}{4d^2} \sum_{c=1}^C w_c \|\mathbf{C}_S^{(c)} - \mathbf{C}_T^{(c)}\|_F^2 \quad (12)$$

$$\mathbf{C}_S^{(c)} = \frac{1}{n_S^{(c)} - 1} [\mathbf{H}_S^{(c)} - 1\boldsymbol{\mu}_S^{(c)\top}]^T [\mathbf{H}_S^{(c)} - 1\boldsymbol{\mu}_S^{(c)\top}] \quad (13)$$

$$\mathbf{C}_T^{(c)} = \frac{1}{n_T^{(c)} - 1} [\mathbf{H}_T^{(c)} - 1\boldsymbol{\mu}_T^{(c)\top}]^T [\mathbf{H}_T^{(c)} - 1\boldsymbol{\mu}_T^{(c)\top}] \quad (14)$$

其中： C 为类别数， c 为类别索引； $\mathbf{H}_S^{(c)} \mathbf{H}_T^{(c)}$ 和 分别是源/目标域中属于类别 c 的样本特征矩阵； $\boldsymbol{\mu}_S^{(c)}$ 和 $\boldsymbol{\mu}_T^{(c)}$ 是对应的类均值向量； 1 表示与样本数匹配的全 1 列向量，用于去均值； $\mathbf{C}_S^{(c)}$ 和 $\mathbf{C}_T^{(c)}$ 为类内协方差矩阵； $w_c > 0$ 为类别权重（逆比例缩放高置信样本数以抑制不平衡）； $\tau \in (0, 1]$ 为伪标签置信阈值。

C-CORAL 的优势在于：对齐类内的形状与相关结构，而不是把所有样本“搅匀”，因此在类别重叠、长尾或多模态分布的场景下能更好地保留决策边界。同时，置信度筛选与类别重加权减少了伪标签噪声与样本不平衡带来的负面影响。

通过最小化复合损失函数 $\mathcal{L}_{\text{total}}$ ，学生模型不仅学习了如何完成目标域的特定任务，还继承了教师模型强大的通用表征能力，并通过更精细的类条件域对齐进一步适应了目标域的数据分布，从而在数据有限的情况下也能达到很高的性能。

1.4 概念漂移适应与“测量—控制”闭环

为应对网络环境的概念漂移，采用滑动时间窗口进行在线更新：定期用最新且高置信度的预测作为伪标

签，对学生模型做小步在线蒸馏/自训练，使其随网络状态演化自适应；同时使用温度缩放 (Temperature Scaling) 校准输出概率，并以预期校准误差 (Expected Calibration Error) 持续监控置信度，稳定告警阈值。

框架形成测量—预测—控制闭环：以多模态数据测量网络状态，利用 MM-SSKD 预测潜在故障与威胁；预测结果（如高风险告警）可自动触发控制策略，如联动防火墙限速或 IP 隔离，或生成运维工单交由人工处理，实现智能化闭环运维。

2 实验结果与分析

2.1 实验设置

为全面评估 MM-SSKD 框架的有效性，实验采用严格隔离的源域与目标域，并设置防数据泄漏流程。

源域选择 UNSW-NB15 (2015)^[15] 与 CIC-IDS2017 (2017)^[16] 的周二至周五数据作为无标注语料，用于教师模型的多模态自监督预训练。所有样本均按事件时间戳排序，仅参与自监督目标，不进入任何监督训练或评测。为增强多模态性，按会话/日志键对齐并统一到固定长度时间窗。

目标域用于学生模型的蒸馏与微调，包含两部分：其一采用 2018 AIOps KPI 数据集（公开的企业运维多 KPI 时序及异常标注），用于回归子任务；其二为威胁识别数据，使用 CIC-IDS2017^[16] 的周一正常流量与 CIC-IDS2018^[17] 的攻击子集混合而成，以模拟同一采集环境下的时间漂移。目标域按时间顺序划分为训练/验证/测试集 (70%/15%/15%)，并通过哈希比对与 IP/子网过滤确保与源域零重叠；整体数据流严格单向。

评价指标包括回归任务的均方根误差 (RMSE, root mean squared error) 和平均绝对误差 (MAE, mean absolute error)，以及分类任务的 F1 分数 (F1-Score) 和曲线下面积 (AUC, area under the receiver operating characteristic curve)。对比方法涵盖传统与深度模型 LSTM^[18]、GRU^[19]、TCN^[20]、XGBoost-TS^[21]、Transformer^[22]；无监督域适配 DANN^[23]、CORAL^[14]、CDAN^[24]、SHOT^[25]；面向类级域对齐的 ClassDA^[26] 和 CDAN+E^[24]；少样本/半监督方法 Prototypical Networks^[27] 和 Mean Teacher^[28]；以及消融变体 Supervised-Student。

实现基于 PyTorch，在 NVIDIA A100 GPU 上运行。学习率、批次大小、优化器、损失权重与蒸馏温度等超参数通过贝叶斯优化在预定义搜索空间内选择。其中，教师端采用 AdamW，学习率 1×10^{-4} ，批次大小 256，InfoNCE 温度 0.07， $\gamma = 0.2$ ；学生端同样采用 AdamW，学习率 1×10^{-3} ，批次大小 64。监督多任务损失中回归项与分类项的权重均取 $\alpha = \beta = 0.5$ ，以平衡

连续量预测和不均衡分类的优化目标。知识蒸馏部分采用温度 $T=2.0$ ，并将两条蒸馏路径的权重分别设为 $\lambda_{\text{logits}}=1.0$ 和 $\lambda_{\text{feat}}=0.5$ 。为便于复现实验中使用的 3 种模态数据，预处理阶段的关键配置如表 1 所示。

表 1 数据预处理细则

模态	预处理内容	关键参数	维度
流量 (Flow)	从流记录中选取协议、IP/端口、包数、字节数、持续时间、方向统计等共 22 个结构化字段，并按 1 s 聚合到对应时间窗	窗口 = 300 s； 步长 = 60 s；时间对齐精度 = 1 s；时间容差 = ± 0.5 s	22 维
日志 (Log)	对 message 字段做正则归一化(IP/URL/数字 $\rightarrow$ 占位符)，中文用 jieba，英文用 SpaCy 分词，去停用词后映射为 300 维词向量，在窗口内做平均	每条日志 ≤ 64 token；嵌入 = 300 维；空秒补 0	300 维
性能 指标 (KPI)	将 12 路监控/性能指标先升采样到 1 s，数值型用线性插值，离散/告警型前向填充，在窗口内计算 {mean, max, min, std}	指标数 = 12； 统计 = 4；窗口 = 300 s；步长 = 60 s	48 维 (12 $\times$ 4)
拼接	按时间戳和实体 ID 对齐后拼接	/	370 维

2.2 性能对比与分析

为了评估 MM-SSKD 框架的整体性能，本文在构建的目标域测试集上对各模型进行了综合评测。

表 2 各模型在目标域测试集上的综合性能对比
(均值 $\pm$ 标准差, $N=5$)

模型	RMSE $\downarrow$	MAE $\downarrow$	F1-Score $\uparrow$	AUC $\uparrow$
LSTM	0.452 $\pm$ 0.011	0.288 $\pm$ 0.008	0.842 $\pm$ 0.007	0.913 $\pm$ 0.005
GRU	0.445 $\pm$ 0.009	0.281 $\pm$ 0.007	0.849 $\pm$ 0.006	0.919 $\pm$ 0.004
TCN	0.421 $\pm$ 0.012	0.269 $\pm$ 0.009	0.865 $\pm$ 0.008	0.931 $\pm$ 0.006
XGBoost-TS	0.489 $\pm$ 0.015	0.312 $\pm$ 0.010	0.821 $\pm$ 0.011	0.895 $\pm$ 0.008
Transformer	0.391 $\pm$ 0.008	0.245 $\pm$ 0.006	0.888 $\pm$ 0.005	0.946 $\pm$ 0.003
DANN	0.399 $\pm$ 0.010	0.253 $\pm$ 0.007	0.880 $\pm$ 0.007	0.941 $\pm$ 0.005
CORAL	0.405 $\pm$ 0.011	0.258 $\pm$ 0.008	0.876 $\pm$ 0.008	0.938 $\pm$ 0.006
CDAN	0.385 $\pm$ 0.009	0.241 $\pm$ 0.006	0.895 $\pm$ 0.006	0.951 $\pm$ 0.004
ClassDA	0.353 $\pm$ 0.009	0.231 $\pm$ 0.006	0.908 $\pm$ 0.005	0.958 $\pm$ 0.003
CDAN+E	0.348 $\pm$ 0.009	0.226 $\pm$ 0.006	0.912 $\pm$ 0.005	0.961 $\pm$ 0.003
SHOT	0.402 $\pm$ 0.011	0.255 $\pm$ 0.008	0.878 $\pm$ 0.007	0.939 $\pm$ 0.005
Prototypical Networks	0.412 $\pm$ 0.012	0.260 $\pm$ 0.008	0.881 $\pm$ 0.008	0.940 $\pm$ 0.006
Mean Teacher	0.406 $\pm$ 0.011	0.255 $\pm$ 0.008	0.886 $\pm$ 0.008	0.944 $\pm$ 0.005
Supervised-Student	0.415 $\pm$ 0.009	0.263 $\pm$ 0.007	0.875 $\pm$ 0.006	0.939 $\pm$ 0.004
MM-SSKD (C-CORAL)	0.324 $\pm$ 0.007	0.197 $\pm$ 0.005	0.931 $\pm$ 0.004	0.972 $\pm$ 0.002

从表 2 的结果可以看出，本文提出的 MM-SSKD 学生模型在四项指标上均优于所有基准模型，包括传统时间序列模型、近期的深度结构，以及一系列主流无监督/半监督域适配方法。值得注意的是，在同时考虑类级域对齐的代表性方法 ClassDA 和 CDAN+E 的情况下，本文方法仍然保持最好的结果：相较于性能最优的基线 CDAN+E，MM-SSKD 将 RMSE 从 0.348 降至 0.324 (相对下降约 6.9%)，将 F1-Score 从 0.912 提升到 0.931 (约 2.1% 相对提升)，AUC 也由 0.961 提升到 0.972。这说明，在已有的对抗式/类级域对齐策略基础上，再结合“自监督预训练+跨域知识蒸馏+类条件相关性对齐”，可以在处理不均衡、甚至同类多峰的目标域时更好地保持类间决策边界，从而获得更高的识别性能。

2.3 少标注学习能力分析

为验证模型在目标域标注数据稀缺场景下的性能，本文在不同比例 (0.5%，1%，5%，10%) 的标注数据上训练模型，并与强基线模型进行对比。

如图 2 所示，MM-SSKD 方法在极少量标注数据 (如 0.5%) 的情况下，性能远超包括 Transformer、CDAN 以及少样本基线 Prototypical Networks 在内的所有对比模型，证明了自监督预训练与知识蒸馏在提升样本效率方面的巨大价值。

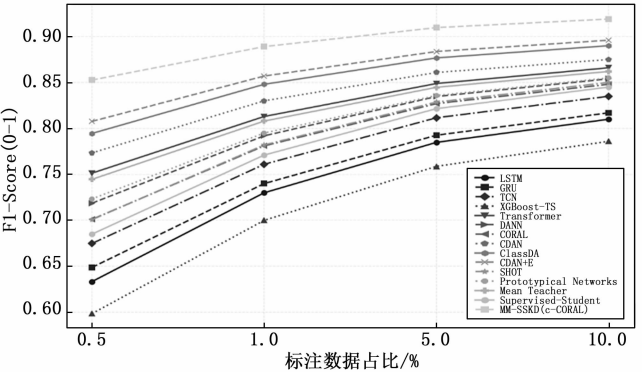


图 2 少标注比例下各方法的 F1-Score 对比

2.4 消融与敏感性分析

为了进一步剖析 MM-SSKD 框架中各个关键组件的贡献，消融实验的结果如表 3 所示。所有变体与完整模型之间的性能差异均通过配对 t 检验，结果显示具有统计学显著性 ($p<0.05$)，证明了 MM-SSKD 框架中每个组成部分的不可或缺性。

本文提出的增强型 SSL 代理任务和 C-CORAL 对齐策略，相较于它们的标准版本，均带来了显著的性能提升。值得注意的是，在增强型 SSL 内部进一步拆分后可以看到：仅去掉跨模态一致性或仅去掉随机模式丢弃都会带来性能下降，但两者的单独版本仍优于使用标准

MM-MAE 的设置。此外，进一步加入“无特征蒸馏”和“无 logits 蒸馏”两个单路径变体，结果表明二者都能带来比“仅 SSL 微调”更高的性能，但仍低于同时开启两条路径的完整模型，这说明判别知识蒸馏与特征对齐蒸馏是互补的，双路径设计确实能够带来额外增益。

表 3 消融实验结果 (均值±标准差, N=5)

模型变体	RMSE↓	F1-Score↑
w/o 多模态(仅使用流量数据)	0.438±0.011	0.867±0.009
w/o 自监督学习(监督蒸馏)	0.372±0.009	0.903±0.007
w/o 知识蒸馏(SSL 微调)	0.359±0.008	0.915±0.006
w/o C-CORAL(使用标准 CORAL)	0.332±0.008	0.926±0.004
w/o 增强型 SSL(使用标准 MM-MAE)	0.335±0.007	0.924±0.005
w/o 跨模态一致性 (只做缺模态鲁棒性)	0.331±0.006	0.927±0.004
w/o 随机模态丢弃(只做跨模态对齐)	0.329±0.006	0.928±0.004
w/ logits 蒸馏(无特征蒸馏)	0.330±0.007	0.923±0.005
w/ 特征蒸馏(无 logits 蒸馏)	0.329±0.007	0.924±0.005
MM-SSKD (完整模型)	0.324±0.007	0.931±0.004

为探究关键超参数对模型性能的影响，对 3 类参数做了敏感性分析：(1) 增强型 MM-MAE 中的掩码率；(2) 学生模型深度；(3) 双路径蒸馏中的权重与温度。结果如图 3 所示。可以看到，模型在较广的参数范围内(掩码率 60%~80%，学生深度为教师的 40%~60%)均能保持稳健性能；在蒸馏部分，当 logits 路径权重取 1.0、特征路径权重取 0.5 时，F1 与 RMSE 同时达到较优水平，而继续增大特征路径权重会略微降低稳定性；温度方面， $T=2$ 处于最优或近最优区间，温度升高到 5、10、15 后，F1 逐步下降、RMSE 逐步上升，说明过大的温度会使教师分布过度平滑、削弱可迁移的

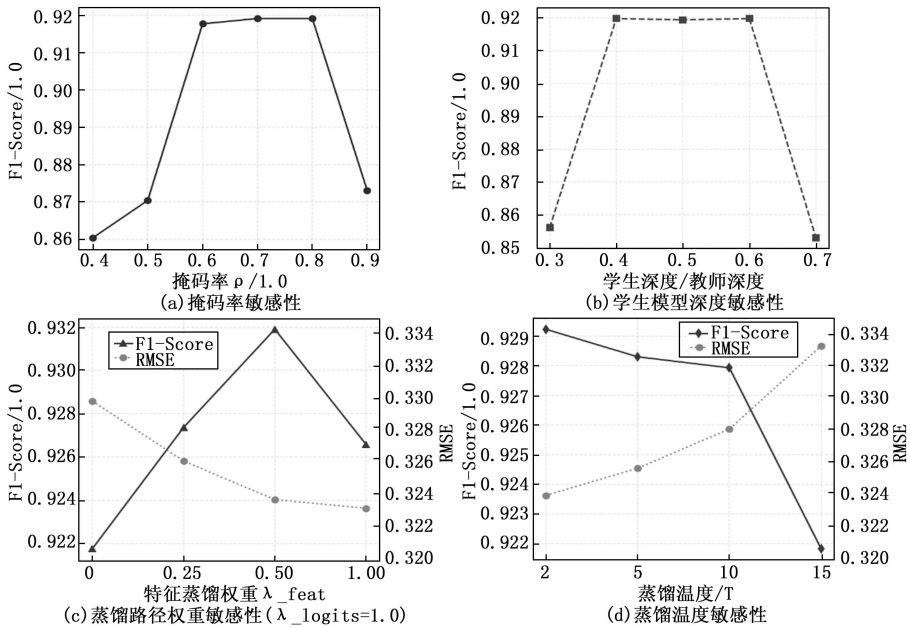


图 3 关键超参数敏感性分析

最后，为考察 C-CORAL 中两个实现相关超参数的影响，结果见表 4。可以看到，阈值从 0.7 提高到 0.8 时，AUC 和 F1 均有小幅提升；继续提高到 0.9 时，由于参与对齐的目标域样本减少，性能略有回落。去掉类权重后，总体 AUC 变化不大，但 F1 (主要是长尾类) 有约 0.3 个百分点下降。

表 4 C-CORAL 参数敏感性分析

设置	AUC↑	F1-Score↑
$\tau=0.7$,带权重	0.969	0.929
$\tau=0.8$,带权重	0.972	0.931
$\tau=0.9$,带权重	0.970	0.930
$\tau=0.8$,无类权重	0.971	0.928

2.5 鲁棒性、校准与在线适应性分析

为测试模型在数据采集不完整情况下的表现，在测试阶段随机丢弃不同比例的某一模态数据。同时，增加了一个未在训练中使用随机模态丢弃 (RMD) 策略的对照组，结果如图 4 所示。结果显示，使用了 RMD 策略的模型展现出强大的鲁棒性，而未使用该策略的模型性能下降非常迅速。这一现象与表 2 中的结果保持一致：

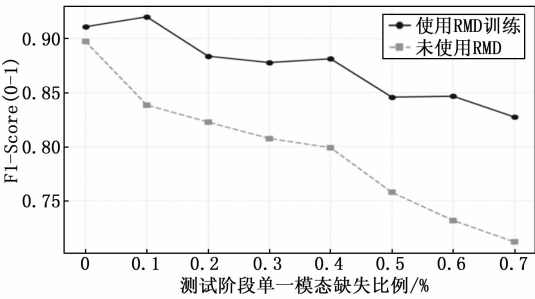


图 4 随机模态丢弃带来的鲁棒性提升

当在自监督预训练阶段就显式暴露随机模态丢弃时，教师所产生的蒸馏目标在目标域同样具备更强的缺模态鲁棒性。

为验证温度缩放对模型置信度的校准效果，计算了校准前后的预期校准误差 (ECE)。如表 5 所示，校准后模型的 ECE 显著降低。更进一步，构建了一个简化的运营成本模型 (总成本 = 误报成本 × 误报数 + 漏报成本 × 漏报数)，并绘制了“阈值 - 代价曲线” (图 5)。结果表明，校准后的模型能够在更低的阈值下达到最低运营成本，证明了模型校准在实际业务中的价值。

表 5 模型校准前后性能对比

模型	ECE ↓	Brier Score ↓
MM-SSKD(校准前)	0.048	0.081
MM-SSKD(校准后)	0.015	0.063

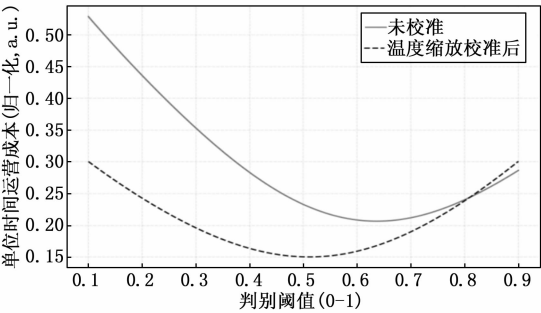


图 5 阈值-代价曲线

2.6 外部可迁移性与在线适应性

为检验模型在截然不同的网络环境下的泛化能力，选取未参与任何训练与模型选择的 Server Machine Dataset (SMD)^[29] 进行零样本迁移测试。如表 6 所示，所有模型性能均有下降，但 MM-SSKD 的性能下降幅度最小。此外，经过 24 小时的在线自适应微调后（每日额外消耗约 0.5 GPU-hour），MM-SSKD 的性能得到显著恢复，证明了其强大的外部可迁移性和在线适应能力。

表 6 SMD 上的零样本迁移与在线适应结果

模型	初始 F1-Score ↑	24h 在线适应后 F1-Score ↑
Transformer	0.75	0.78
CDAN	0.79	0.81
MM-SSKD	0.85	0.90

2.7 模拟运营影响与闭环验证

为量化本方法在真实“测量—预测—控制”闭环中的价值，进行了一项模拟实验，对比了部署 MM-SSKD 前后，一个典型网络安全运营中心（SOC）的关键绩效指标（KPIs）。结果如表 7 所示。部署 MM-SSKD 后，通过自动化预测与告警联动，各项运营指标均得到大幅改善，证明了本框架在提升实际运维效率、降低人力成本方面的巨大潜力。

表 7 部署 MM-SSKD 前后的模拟运营 KPI 对比

指标	部署前(人工+规则)	部署后(MM-SSKD 联动)	提升率/%
平均检测时间 (MTTD)	2 小时	5 分钟	95.8
误报工单率/%	15	3	80.0
平均响应时间 (MTTR)	4 小时	30 分钟	87.5
分析师重复工作量/%	60	10	83.3

2.8 可解释性与机制分析

为深入理解知识蒸馏与域对齐的作用机制，利用 t-SNE 对不同阶段的特征分布进行了可视化（图 6），并使用中心核对齐（CKA）计算了师生网络间的层级特征相似度（图 7）。图 6 清晰地显示，经过 C-CORAL 和知识蒸馏后，目标域的特征分布与源域高度对齐，且类别边界更清晰。CKA 热力图则表明，经过蒸馏的学生网络与教师网络在各层级的特征表示上具有高度的相似性，直观地证明了知识迁移的有效性。

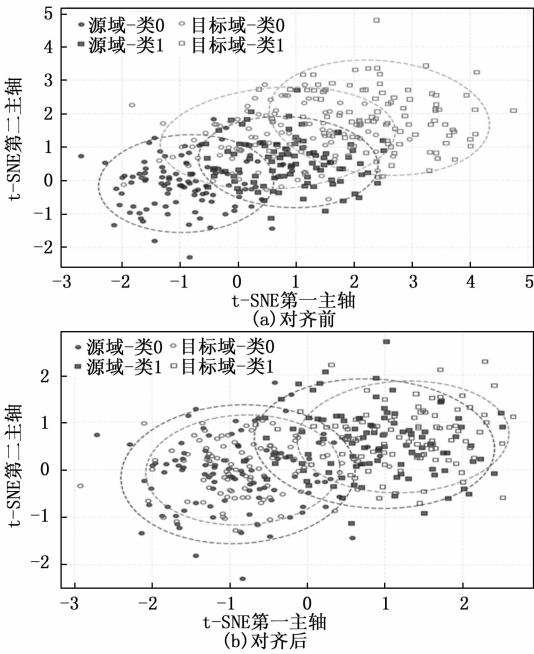


图 6 t-SNE 特征分布可视化

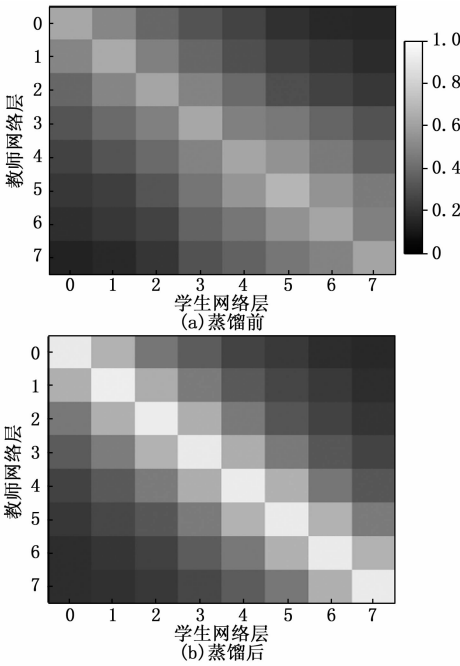


图 7 师生网络层级特征相似度 (CKA) 热力图

3 结束语

围绕企业网络的预测与防护一体化需求, 本文构建了以增强型多模态自监督为底座、双路径跨域蒸馏与类条件相关性对齐为核心的 MM-SSKD 框架, 统一解决企业网络中的故障预测与威胁识别两类任务。实验与可解释性分析表明, 该框架在少量目标域标注、跨域迁移以及单一或部分模态缺失的场景下均能保持稳定优势; 蒸馏后的学生模型以更低的时延与资源占用满足在线部署需求, 概率校准与阈值选择使实际告警决策更稳健。消融结果验证了多模态自监督、类条件对齐与随机模态丢弃等关键组件的有效性。

未来研究可以聚焦于更复杂的生产网络与多任务协同场景, 完善跨域对齐策略与持续学习机制, 并探索面向运维闭环的在线自适应与安全策略联动。

参考文献:

- [1] 张 恒, 蔡志平, 李 阳. SDN 网络测量技术综述 [J]. 中国科学: 信息科学, 2018, 48 (3): 293-314.
- [2] LANDAU-FEIBISH S, LIU Z, REXFORD J. Compact data structures for network telemetry [J]. ACM Computing Surveys, 2025, 57 (8): 1-31.
- [3] LI K, ZHOU W, LI P, et al. Distinguishing DDoS attacks from flash crowds using probability metrics [C] // Proceedings of the 2009 International Conference on Network and System Security (NSS 2009), Gold Coast, Australia; IEEE, 2009: 9-17.
- [4] DIAZ-DE-ARCAYA J, TORRE-BASTIDA A I, ZARATE G, et al. A joint study of the challenges, opportunities, and roadmap of MLOps and AIOps: a systematic survey [J]. ACM Computing Surveys, 2023, 56 (4): 1-84.
- [5] ZHANG L, JIA T, JIA M, et al. A Survey of AIOps in the Era of Large Language Models [J]. ACM Computing Surveys, 2025, 58 (2): 44: 1-44: 35.
- [6] 高雪梅, 董 晔, 许文强, 等. 基于 ARIMA 模型的中亚粮食生产量时空变化分析与预测 [J]. 中国科学院大学学报 (中英文), 2025, 42 (4): 472-486.
- [7] KHRAISAT A, GONDAL I, VAMPLEW P, et al. Survey of intrusion detection systems: techniques, datasets and challenges [J]. Cybersecurity, 2019, 2 (1): 1-22.
- [8] LAYEGHY S, BAKTASHMOTLAGH M, PORTMANN M. DI-NIDS: domain invariant network intrusion detection system [J]. Knowledge-Based Systems, 2023, 273: 110626.
- [9] RIZVI S, SCANLON M, MCGIBNEY J, et al. Deep learning based network intrusion detection system for resource-constrained environments [C] // International Con-

- ference on Digital Forensics and Cyber Crime. Cham; Springer Nature Switzerland, 2022: 355-367.
- [10] LIU Z, MAO H, WU C Y, et al. A ConvNet for the 2020s [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2022). New Orleans, LA, USA; IEEE Computer Society, 2022: 11976-11986.
- [11] BACHMANN R, MIZRAHI D, ATANOV A, et al. MultiMAE: multi-modal multi-task masked autoencoders [C] // Computer Vision-ECCV 2022: Proceedings of the 17th European Conference. Tel Aviv, Israel; Springer, 2022: 348-367.
- [12] NEVEROVA N, WOLF C, TAYLOR G, et al. ModDropp: adaptive multi-modal gesture recognition [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 38 (8): 1692-1706.
- [13] HENAFF O. Data-Efficient Image Recognition with Contrastive Predictive Coding [C] // Proceedings of the 37th International Conference on Machine Learning (ICML 2020). PMLR, 2020: 4182-4192. (PMLR, 119)
- [14] SUN B, SAENKO K. Deep CORAL: correlation alignment for deep domain adaptation [C] // European Conference on Computer Vision (ECCV 2016). Amsterdam, The Netherlands; Springer, 2016: 443-450.
- [15] MOUSTAFA N, SLAY J. UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set) [C] // Military Communications and Information Systems Conference (MilCIS 2015). Canberra, Australia; IEEE, 2015: 1-6.
- [16] SHARAFALDIN I, LASHKARI A H, GHORBANI A A. Toward generating a new intrusion detection dataset and intrusion traffic characterization [C] // Proceedings of the 4th International Conference on Information Systems Security and Privacy (ICISSP 2018). Funchal, Madeira, Portugal; SciTePress, 2018: 108-116.
- [17] SHARAFALDIN I, GHARIB A, LASHKARI A H, et al. Towards a reliable intrusion detection benchmark dataset [J]. Software Networking, 2018 (1): 177-200.
- [18] HOCHREITER S, SCHMIDHUBER J. Long short-term memory [J]. Neural computation, 1997, 9 (8): 1735-1780.
- [19] CHO K, VAN MERRIENBOER B, GULCEHRE C, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation [C] // Conference on Empirical Methods in Natural Language Processing (EMNLP 2014). 2014: 1724-1734.

(下转第 162 页)