



嘉興大學
JIAXING UNIVERSITY
方正为人 勤慎治学

(2024 届)

学士学位论文

基于多源异构特征融合的小样本网络入侵检测系统

学 院:	信息科学与工程学院 (机械工程学院)
专 业:	软件工程
班 级:	软件 201
学 号:	202059065119
姓 名:	詹勇
指导教师:	许聪源

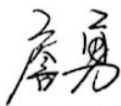
教 务 处 制

二〇二四年五月

诚信声明

我声明，所呈交的论文是本人在老师指导下进行的研究工作及取得的研究成果。据我查证，除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究成果，也不包含为获得嘉兴大学或其他教育机构的学位或证书而使用过的材料。我承诺，论文中的所有内容均真实、可信。

论文作者签名：



签名日期：2024 年 5 月 13 日

授 权 声 明

学校有权保留送交论文（设计）的原件，允许论文（设计）被查阅和借阅，学校可以公布论文（设计）的全部或部分内容，可以影印、缩印或其他复制手段保存论文（设计），学校必须严格按照授权对论文（设计）进行处理，不得超越授权对论文（设计）进行任意处置。

论文（设计）作者签名：詹勇 签名日期：2024 年 5 月 13 日

指导教师签名： 签名日期：2024 年 5 月 13 日

基于多源异构特征融合的小样本网络入侵检测系统

摘 要：现有的网络入侵检测系统主要依赖深度学习技术，并且需要大量样本进行模型训练，这些系统虽在数据丰富的环境中表现出色，但在样本受限的情况下性能显著下降。尽管已有的小样本网络入侵检测系统减少了对大规模数据的依赖，但在模型微调阶段对可迁移性和适应性的要求较高。本研究提出了一种基于多源异构特征融合的小样本网络入侵检测系统，采用数据异构与特征融合技术，无需额外预训练数据即可获得较高的检测性能。此外，该系统通过利用小样本数据进行模型微调，有效增强了在特定场景下的检测能力。系统还整合了异构数据输入与两种特异化特征提取技术，并提出了六种特征融合策略，增强了模型的泛化能力和检测精度。实验结果显示，该系统在两个网络入侵检测基准数据集上的多分类准确率最高可达到 93.40%和 98.50%，优于已有系统的检测效果。

关键词：数据异构；特征融合；小样本；入侵检测；深度学习

A Few-Shot Network Intrusion Detection System Based on Multi-Source Heterogeneous Feature Fusion

Abstract: Existing network intrusion detection systems primarily rely on deep learning techniques and require substantial data samples for model training. While these systems perform well in data-rich environments, their efficacy significantly diminishes under limited sample conditions. Current few-shot network intrusion detection systems reduce the dependency on large-scale data but demand high transferability and adaptability during the model fine-tuning phase. This study introduces a few-shot network intrusion detection system based on the fusion of multi-source heterogeneous features, which achieves high detection performance without the need for additional pre-trained data. The system enhances detection capabilities in specific scenarios by fine-tuning models using few-shot data. It also integrates heterogeneous data inputs with two specialized feature extraction techniques and introduces six feature fusion strategies, thereby enhancing the model's generalization ability and detection accuracy. Experimental results show that our system achieves multi-class accuracy rates of up to 93.40% and 98.50% on two network intrusion detection benchmark datasets, surpassing existing systems.

Keywords: data heterogeneity ; feature fusion ; few-shot ; intrusion detection ; deep learning

目 录

1 引言.....	1
1.1 研究背景和意义.....	1
1.2 论文结构安排.....	2
1.3 本文的主要贡献.....	3
2 相关理论与技术.....	4
2.1 网络入侵检测.....	4
2.2 数据不平衡与小样本问题.....	5
2.3 自注意力机制.....	6
2.4 多模态和特征融合.....	7
2.5 迁移学习.....	8
3 多源异构特征融合检测系统.....	10
3.1 数据预处理.....	10
3.1.1 网络流量来源.....	10
3.1.2 网络流量预处理和标注.....	10
3.1.3 异构数据特征提取策略.....	11
3.2 异构特征融合系统构建.....	12
3.2.1 数据异构.....	13
3.2.2 网络特征图特征提取器.....	13
3.2.3 流量特征集特征提取器.....	15
3.2.4 六种特征融合方法.....	15
3.2.4 迁移模型.....	18

4 系统设计与验证.....	20
4.1 网络入侵检测数据集.....	20
4.2 检测系统中的评价指标.....	21
4.3 基线模型的参数设置.....	21
4.4 实验设置.....	22
4.4.1 样本敏感度实验.....	23
4.4.2 特征融合实验.....	23
4.4.3 迁移实验中的模型调优.....	23
4.5 检测结果.....	24
4.5.1 样本敏感度实验.....	24
4.5.2 特征融合实验.....	29
4.5.3 迁移实验.....	32
4.5.4 结果总结.....	34
5 比较与讨论.....	36
5.1 与同类工作的对比.....	36
5.2 极小样本情况.....	36
5.3 检测系统的消融实验.....	37
6 总结与展望.....	39
6.1 总结.....	39
6.2 展望.....	40
致谢.....	41
参考文献.....	42

1 引言

1.1 研究背景和意义

随着互联网的快速发展，网络安全面临愈加复杂的挑战，尤其是在数字化转型加速推进的背景下，网络已成为支撑各行各业的基础设施。这不仅为网络攻击者提供了丰富的目标，也使得攻击手段不断演化，并且涉及从经济利益到国家安全的多样化攻击。尽管我国关键信息基础设施发展迅猛，但我们仍面临全球范围内最严峻的网络安全挑战之一。例如，在 2022 年，西北工业大学遭遇了一次严重的网络攻击，事件曝光后，不仅没有遏制黑客团体的活动，反而触发了更加激进的网络行动。2023 年，武汉市的地震监测系统也成为攻击目标。7 月 26 日，武汉市应急管理部门公开了这一事件的详情，国家计算机病毒处理中心与国内网络安全公司 360 联合调查发现，一个境外黑客组织利用木马程序非法访问了地震速报系统，窃取了关键的地震烈度数据。这种数据因能揭示地下结构和地质特征，具有高度敏感性，甚至可能用于推断某地区是否存在军事活动，例如工业爆破或地下军事设施，这种敏感信息的泄露直接威胁到国家安全。事实上，近年来我国关键基础设施频繁遭受多个敌对势力的网络攻击，这些网络攻击不仅隐蔽而且迅速在关键领域扩散，造成了直接的数据和服务中断以及关键数据的泄露，也加深了对国家安全的潜在威胁。预计未来，这种高风险的网络安全态势仍将持续。

习近平总书记提出全面国家安全观念已有十年，随着科技的快速进步，网络安全已成为新时代国家安全的关键领域和主战场。传统的网络入侵检测系统面临日新月异的攻击模式，难以有效应对。网络入侵检测系统（Network Intrusion Detection System, NIDS）面临着前所未有的挑战。这些挑战不仅来源于数据的高维度复杂性和攻击方式的多样化，还包括难以在小样本情况下保持高精度的检测率。随着攻击手段的演变，网络入侵的手段正变得越来越复杂，从传统的病毒、蠕虫攻击演变到高度复杂的高级持续威胁和零日攻击等。这种多样化的攻击模式对现有的网络入侵检测技术提出了严峻的挑战。特别是在安全要求较高的环境中，任何攻击都可能导致严重后果，因此只能在攻击尚未广泛发生之前获得少量样本，这使得传统系统的检测效率和准确性大幅降低。面对这种对安全防护要求极为严

格的网络安全形势，开发能够适应新型威胁并快速响应的高效网络入侵检测技术变得尤为迫切。针对这些挑战，本研究提出了一种基于小样本的网络入侵检测系统，该系统通过整合多源异构数据和特征融合方法，并利用深度学习与迁移学习来提升小样本环境下的检测率和准确性。这种方法的核心创新在于其能够有效处理数据稀缺的情况，通过这种高效的数据融合和特征处理机制，为网络安全防护体系提供技术支撑，从而更好地应对当前网络安全面临的复杂多变的威胁，填补现有网络安全防御手段在小样本环境下的不足。

1.2 论文结构安排

本文所研究的多源异构特征融合方法组织结构图如图 1.1 所示。其中，本文的第一章用于介绍研究背景和意义，最后概述其结构安排和主要贡献。第二章则用于介绍本文所使用的相关技术与理论，以及近期的相关工作。在第三章中重点阐述了异构特征融合系统的构建，介绍了数据异构、特征提取、特征融合、迁移学习等内容。第四章则是对第三章所提的检测系统进行具体实验，包括样本敏感度、特征融合、迁移实验等多组实验，通过这些实验在不同维度上揭示了模型性能。第五章则是对第四章的补充，侧重于与同类工作进行对比，探讨相较于同类检测系统的优劣情况，并给出两个值得探讨的实验：极小样本情况和检测系统的消融实验，最后通过第六章进行总结与展望。

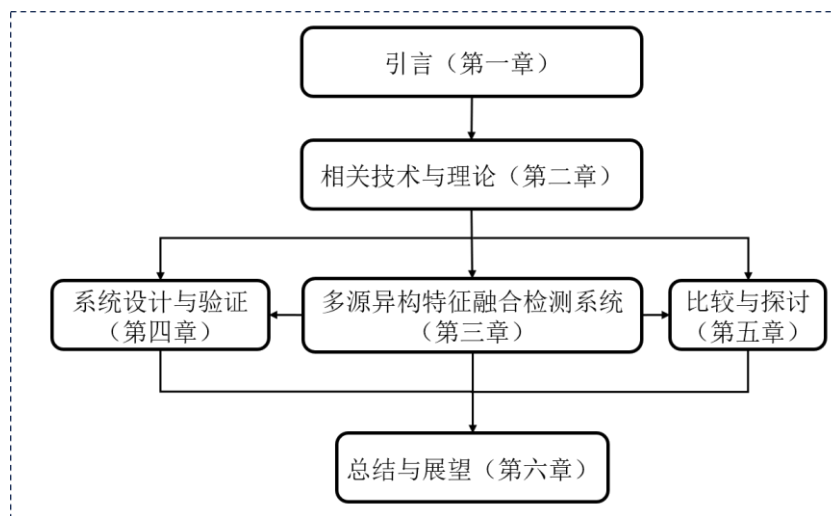


图 1.1 全文组织结构图

1.3 本文的主要贡献

在传统的网络入侵检测模型中，对大规模数据的依赖和在小样本情况下的性能下降是两个主要问题，针对这些问题，本文提出了一种基于多源异构特征融合的小样本网络入侵检测系统，集成了数据异构处理、特征提取和特征融合模块的高效入侵检测系统，具体贡献如下。

（1）多源数据异构技术的应用：该系统能够处理和融合来自不同数据源的信息，通过从多个维度描述和处理数据，生成两种高效的异构数据形式：流量特征图和网络特征集。这种技术的应用减少了对大规模标准数据集的依赖，优化了数据的利用率，为模型提供了更丰富的信息背景。

（2）并行的特征提取技术：通过结合卷积神经网络和 Transformer 的多头自注意力机制，我们的系统能够对异构后的数据进行深入的特征提取。这两种特征提取模型有效地处理和分析了网络流量的原始数据及其结构化特征，显著提高了特征的表达能力和检测模型的准确性。

（3）特征融合与迁移学习策略的整合：系统不仅采用了多样化的特征融合技术，还引入了迁移学习模型，使该系统能在小样本环境下进行更深层次的信息提取。这种结合方法有效地提高了模型的适应性和准确率，尤其是在新型或少见的攻击类型检测上显示出显著的效果。

通过这些技术的集成应用，本系统不仅提高了网络入侵检测的准确率，还增强了对小样本数据处理的能力，有效地应对了网络安全领域中的复杂挑战。

2 相关理论与技术

2.1 网络入侵检测

网络入侵检测系统旨在监测网络流量和系统活动,以发现和响应异常行为或可能的安全威胁。网络入侵检测的概念在 1987 年由 James Anderson 提出,他在一份报告中详细描述了利用审计追踪来检测不当的计算机使用。随着互联网的普及和网络安全威胁的增加,入侵检测系统^[1](Intrusion Detection System, IDS)开始获得广泛地应用,此时期诞生了许多商业和开源的入侵检测产品,例如 Snort^[2]。随着攻击手法的进化,仅依赖签名的检测方法开始显得力不从心,因此研究者和开发者开始探索基于行为的检测方法^[3],通过学习网络的正常行为模式,然后检测与此模式显著不同的行为来识别潜在的威胁。随着机器学习和人工智能技术的发展,入侵检测系统开始整合这些技术来提高检测的准确性和适应性。现代 IDS^[4]可以分析大规模数据,从中提取特征,并能够检测到先前未知的攻击和复杂的威胁。Souradip 等人^[5]通过一系列优化,包括去除多重共线性、采样和降维,从而有效检测网络攻击和异常,并且所提出的模型是轻量级的,可以部署在功率和存储能力有限的网络节点上。罗等人^[6]则提出一种基于贝叶斯攻击图的网络入侵意图分析方法,引入攻击代价和入侵意图对网络安全影响的问题,结合贝叶斯信念网络量化攻击图,建立静态风险评估模型,并利用入侵意图动态更新模型,在预测攻击路径方面也具有可行性。Ankit 等人^[7]提出了一种自动编码和主成分分析(Principal Component Analysis, PCA)技术,该方法能够捕获特征之间的线性和非线性关系,同时降低输入数据的维数,为解决网络流量数据的高维性和复杂性提供了新的方向。Jafar 等人^[8]提出了一个混合深度学习框架,结合了卷积神经网络(Convolutional Neural Network, CNN)和长短期记忆网络(Long Short-Term Memory Network, LSTM)的优势,以提高网络入侵检测系统的检测率。Maya 等人^[9]将自举聚合和梯度提升决策树两种集成技术相结合,提出了一种双集成模型。Zakieh 等人^[10]提出了一种基于非洲秃鹫优化算法的改进版本,在其中加入了正弦余弦算法,避免陷入局部最优并加速全局搜索能力。Liu 等人^[11]认为目前流行的深度学习算法存在精度相对较低、严重依赖于人工选择特征等局限性,对此提

出了一种增强的基于经验的成分分析方法来选择相关特征,综合了经验模态分解和 PCA 的优点,保留了大部分相关特征,再利用 LSTM 对选取的攻击节点进行分类。Miel 等人^[12]提出了一种新的多阶段分层入侵检测方法,这是一种可扩展的 IDS,能够检测未知的零日攻击,并且易于部署。Vladimir 等人^[13]认为目前的研究大多数都是基于至少五年或更长时间的数据集上进行训练和测试的,这使得人们对其安全性能的了解很模糊,此外,它们往往被设计为孤立的、自我关注的组件。因此提出了一个模块化的网络入侵检测体系结构,能够基于真实场景模拟网络攻击,同时评估其防御能力。Ratul 等人^[14]也注重于实时流量检测,提出了一种结合粒子群优化和遗传算法的相关特征选择,建立双阶段的 NIDS。Amir 等人^[15]提出了一种基于并行深度自动编码器的新型轻量级结构,该结构利用特征向量中各个值的局部和周围信息。这种类型的特征分离使我们能够提高模型的准确性,同时大大减少了参数的数量、内存占用和对处理能力的需求。

2.2 数据不平衡与小样本问题

在 NIDS 的数据集中,正常流量的样本通常远多于异常流量样本,造成了严重的数据不平衡问题。这种不平衡会影响检测模型的学习效果,使模型偏向于多数类,从而降低对少数类(如攻击行为)的检测能力。最早期的入侵检测系统主要关注特征工程和规则设计,较少考虑数据分布的问题。随着数据科学的发展,研究者开始注意到数据不平衡对模型性能的影响。此后开始应用传统的数据重采样技术,如过采样少数类^[16]和欠采样多数类,以改善模型在少数类上的识别能力。随着集成学习和成本敏感学习的引入^[17],随机森林、Boosting 被用于提高对少数类的检测精度。同时成本敏感学习开始得到应用,强调在学习过程中对不同类别分配不同的错误成本。现在,广泛利用深度学习来处理数据不平衡的问题,例如使用自动编码器来学习正常和异常模式,以及应用对抗性训练技术模拟少数类样本。Marija 等人^[18]的实验表明对较少类型的特征粗粒度处理非常重要,所提方法仅需要 3 个实例即可准确地检测攻击。Radhika 等人^[19]认为不同攻击类别之间的不平衡,导致降低了机器学习模型检测此类恶意流量的学习性能,因而提出了正则化 Wasserstein 生成对抗网络来增加少数攻击样本以获得平衡数据集。所提出的 WGAN-IDR 的增强性能优于其他增强方法。Danish 等人^[20]注重于优先考虑数

据集中的关键元素,将更多的计算资源分配给可能包含指示安全威胁的模式或异常的数据段。该机制与 Bi-LSTM 相结合,增强了检测系统从有限数据集中有效学习的能力。通过整合 SHAP 机制来提高检测系统的透明度,可信度和可解释性。Xiao 等人^[21]认为现有的网络入侵检测方法主要是利用传统的机器学习或深度学习技术,根据网络流的统计特征对入侵进行分类。其中的特征提取依赖于经验,直到网络流结束后才能进行,从而延迟了入侵检测。为解决这一问题,提出了一种基于图嵌入技术的检测方法,并使用随机森林对图向量进行分类,使用子图结构自动提取流图特征,并且只依赖于每个双向网络流的少量初始交互数据包。Mohamed 等人^[22]提出了一种利用机器学习和人工智能的智能混合模型,再加上特征约简技术,包括奇异值分解和主成分分析,以及 k-means 聚类模型增强信息增益,保证提取的特征取得较高的准确性和可靠性。Nan 等人^[23]认为现有的检测方法对小规模不平衡数据集的检测率较低,故提出了一种基于特征增强的恶意流量检测模型,特征增强方法根据高斯特征值对原始流量特征进行分组,并使用 k-means 算法生成聚类特征。将原始特征和生成特征输入到基于浅层神经网络和随机森林算法构建的双重分类模型中,用于网络流量检测。刘等人^[24]认为机器学习模型在处理网络入侵问题时,施加细微扰动会导致模型得出错误的结果,为应对此类威胁,从攻击、防御 2 个角度系统分析,并根据对抗攻击阶段提出了一个多维分类法。Ankit 等人^[25]使用 Bagging 分类器来解决类不平衡问题,该方法使用深度神经网络(Deep Neural Network, DNN)作为基本估计器,该方法在解决入侵检测数据集的类不平衡问题的同时实现了泛化,具有双重的可取性和优点。Rajkumar 等人^[26]注重于数据生成,首先进行数据预处理以提高训练数据的质量,然后使用自适应合成过采样技术生成少数类样本以克服类不平衡,其次,通过在五个基分类器的递归特征消除中嵌入 SHAP 特征重要性来进行特征选择,最后,将这些特征输入到动态集成选择技术中,通过改变 k 值进行分类。

2.3 自注意力机制

自注意力机制是一种计算模型内各部分之间相互关系的技术,它通过对输入数据的不同部分赋予不同的权重,从而捕捉它们之间的依赖关系。这种机制最初是在自然语言处理领域获得广泛应用,并逐渐扩展到其他领域如图像处理和网络

安全。自注意力机制最初由深度学习研究者探索,用于解决自然语言处理中的各种问题,如文本分类和机器翻译^[27]。这一阶段,研究者开始理解到在模型中显式建模输入之间的依赖关系的重要性。在2017年Vaswani等人^[28]首次提出了Transformer架构,这是一种完全基于自注意力机制的模型,极大地推动了自注意力技术的发展。该模型显示出在多个任务上的卓越性能,尤其是在机器翻译和文本理解方面。随着Transformer的成功,自注意力机制开始被应用到更多领域,包括图像识别^[29]、语音处理和生物信息学。在网络安全领域,研究者探索使用自注意力来捕捉网络流量数据中的复杂模式,以提升对复杂网络攻击的识别能力。近年来自注意力机制应用于网络入侵检测,特别是在检测网络流量中的异常活动方面,展示了自注意力在提高复杂攻击检测能力方面的有效性。Xue等人^[30]提出了一种新的基于n-gram频率和时间感知Transformer的入侵检测模型,该模型可以从包级和会话级分层学习流量特征,所提出的时间感知转换器用来学习IDS的会话级特征。时间感知转换器考虑数据包之间的时间间隔,并学习会话的时间特征进行分类。Qing等人^[31]提出了一种高效的异常检测方法AnoGLA,该方法考虑了网络结构和节点属性之间复杂的通信模式。在网络流量中构建了图结构数据,并利用图卷积网络进行建模。并将LSTM与注意机制相结合,提取图在不同时刻的变化信息。Fernando等人^[32]提出了一种具有自注意力机制的多层感知器,研究了这类算法中特征选择的相关性,并分析了注意机制的重要性,以改进同一模型内的特征评估。R等人^[33]提出了一种基于多头自注意力的门控图卷积网络架构,能够识别包括DoS和零日攻击在内的多种威胁,并通过优化聚类算法中簇头的选择,提高了检测的准确性和效率。

2.4 多模态和特征融合

多模态和特征融合技术旨在结合来自不同数据源的数据或特征,以提高系统的决策精度。在网络入侵检测中,通过结合多种数据源的信息,可以更有效地识别和响应潜在的网络威胁。在初步探索阶段研究者开始尝试将来自不同网络层^[34](如物理层、网络层和应用层)的信息进行整合,以增强入侵检测系统的性能。在此之后开始出现更系统的特征融合策略,如早期融合、晚期融合和混合融合,这些方法通过在不同的阶段整合数据来提高检测的精度和效率。随着机器学习技

术的发展,特征融合技术开始与各种学习算法结合^[35],如支持向量机、随机森林和深度学习,以处理更复杂的数据集和提高对复杂攻击行为的识别能力。直到深度学习技术的引入使得特征融合达到了一个新的高度, Ren-Hung 等人^[36]评估了三个基于主机的数据源——网络流量、系统日志和主机统计数据。它评估并比较了它们在不同攻击阶段和类型中的组合检测能力,网络流量数据由 CNN 处理,以改进自动特征选择。系统日志数据采用 LSTM 和注意力模型进行处理,以增强对时间关系的探索。主机统计数据通过 DNN 进行处理,通过对多种数据类型的具体化处理提高模型的表现情况。Ankit 等人^[37]设计了基于 DNN 的 IDS,该技术通过使用标准差和均值与中位数之差的统计重要性融合来选择特征,在提出的方法中,特征根据其基于统计重要性融合得到的排名进行剪裁,旨在筛选出具有高可辨别性和偏差的相关特征,从而更有效地学习数据。刘等人^[38]提出了一种基于 VAE-CWGAN 和特征统计重要性融合的检测方法,先使用 VAE-CWGAN 模型生成新样本以解决数据集类不平衡问题,再融合其统计重要性来进行特征选择。Juan 等人^[39]提出了一种基于梯度重要度增强的特征融合技术,将特征融合和特征增强相结合,使模型能够更加关注与分类相关的样本特征。Hong 等人^[40]首先设计了一种对抗样本生成算法来生成对抗样本并评估物联网网络入侵检测器的性能,提出了一种新的框架,可以通过特征分组和多模型融合来防御对抗性攻击,为物联网网络入侵检测做出了宝贵的贡献。张等人^[41]提出了一种可变融合的随机注意力胶囊网络的入侵检测模型,通过特征动态融合,使得模型能够更好地捕捉数据特征,并使用随机注意力机制,减少了对训练数据的依赖,使模型更具有泛化能力。Xiao 等人^[42]注重于融合来自安全信息和事件管理系统的异构威胁情报,重构多步骤攻击场景,发现关键攻击路径。将结构化威胁信息表达对异构威胁情报进行格式化,并将它们拼凑在一起,并使用语义关联权值和社区检测算法来挖掘攻击场景。

2.5 迁移学习

迁移学习是一种机器学习方法,它允许从一个任务学到的知识被应用到另一个相似的任务上,以此提高新任务的学习效率和性能。在网络入侵检测领域,迁移学习比较有效,因为它允许模型利用在其他领域或数据集上学到的知识来加快

并增强对新威胁的检测能力。最初,迁移学习主要用于图像和语音识别领域^[43],但很快研究者开始探索其在网络安全领域的潜在应用,特别是在处理稀有攻击类型或小样本数据问题时。随着深度学习技术的成熟,迁移学习在网络入侵检测中的应用变得更加广泛和有效。通过预训练的 DNN 模型,研究者能够有效地迁移在其他领域学到的复杂特征提取能力到网络入侵检测任务上。Farhan 等人^[44]提出的 IDS 是基于 Transformer 的迁移学习来学习网络特征表示和不平衡数据中的特征交互。David 等人^[45]证明了迁移学习在计算有限的环境中仅使用原始网络流量进行入侵检测的可行性,将一维卷积神经网络模型与再训练随机森林模型相结合,在边缘设备上仅使用 5000 个训练样本即可到达 96% 的检测结果。Shahid 等人^[46]提出了基于深度迁移学习的入侵检测系统,通过协同集成 CNN、遗传算法和 Bagging,在复杂的工业物联网网络环境中有很好的表现效果。Jia 等人^[47]提出了一种基于异常的入侵检测系统,该系统具有联邦学习功能,并且采用了基于实例的迁移学习,在本地采用基于实例的迁移学习,能够解决用非独立同分布数据训练本地模型的紧迫问题。

3 多源异构特征融合检测系统

本章将详细介绍我们所开发的多源异构特征融合的检测系统，该系统的核心是一系列精细的方法，我们将这些方法框架化，视为一个综合的系统，用以提升网络入侵检测的准确性和效率。这套检测系统是建立在多个模块的基础上，包括数据异构与输入模块、特征提取模块以及特征融合器模块，这些模块协同工作，提取异构数据并进行特征融合，以形成一个较高性能的网络入侵检测系统。

3.1 数据预处理

3.1.1 网络流量来源

在进行网络入侵检测研究时，获取高质量且具有代表性的网络流量对于模型训练和系统的检测性能至关重要。本研究所选用的网络流量来源是加拿大信息安全研究中心（Canadian Institute for Cybersecurity, CIC）提供的 CICIDS2017^[48]和 CICIDS2018^[48]数据集，选择这两个数据集的原因包含多个方面。首先，这些数据集涵盖了广泛的攻击类型，如 DDoS、DoS、渗透和僵尸网络等，其多样性有助于提升模型对复杂网络环境的泛化能力和鲁棒性。其次，这两个数据集在收集和处理过程中注重数据的真实性和完整性，提供了详尽的标签信息和攻击行为的上下文信息，为深入分析和模式识别提供了依据。此外，CICIDS2017 和 CICIDS2018 由于其高质量和广泛地应用，在网络安全研究中得到了广泛认可，成为评估不同检测方法的共同基准。最后，这些数据集的结构设计兼顾了现在的需求和未来可能的发展。因此，选择 CICIDS2017 和 CICIDS2018 作为本研究的数据集，为探索小样本网络入侵检测方法提供了坚实的基础。

3.1.2 网络流量预处理和标注

为了充分利用这些数据集，本研究采取了精细的预处理和标注策略，以确保数据的准确性和分析的有效性，针对两种数据集进行标注的流程如图 3.1 所示。

（1）CICIDS2017 数据集的预处理和标注：

CICIDS2017 数据集中的流量通过五元组来区分，五元组具体包括源 IP 地

址、源端口、目的 IP 地址、目的端口和传输层协议。在预处理阶段，我们首先使用五元组信息从 pcap 包中拆分出单条流量，并与 csv 文件中的相应五元组信息进行匹配，这里的 csv 文件使用的是 CICFlowMeter 进行提取的。一旦匹配成功，我们将相关的标签（例如，正常或各类攻击）应用于每条流量数据。这种方法不仅有助于精确地分类和标注数据，还确保了数据集的完整性和可用性。

（2）CICIDS2018 数据集的预处理和标注：

相比之下，CICIDS2018 的处理更为复杂。由于官方提供的 csv 文件已进行匿名化处理，仅保留了目的端口和协议信息，这不足以精确定位单条流量。为解决这一问题，我们首先将 pcap 文件输入到 CICFlowMeter-V3 工具中，生成对应的 csv 数据，在这一步生成的数据是无标签的。然后，我们使用生成的 csv 中的时间戳与官方数据集进行匹配，由于文件中每一行的时间戳信息是唯一的，因此可以用于标记流量，这一过程虽复杂，但确保了数据的精确标注和后续分析的有效性。

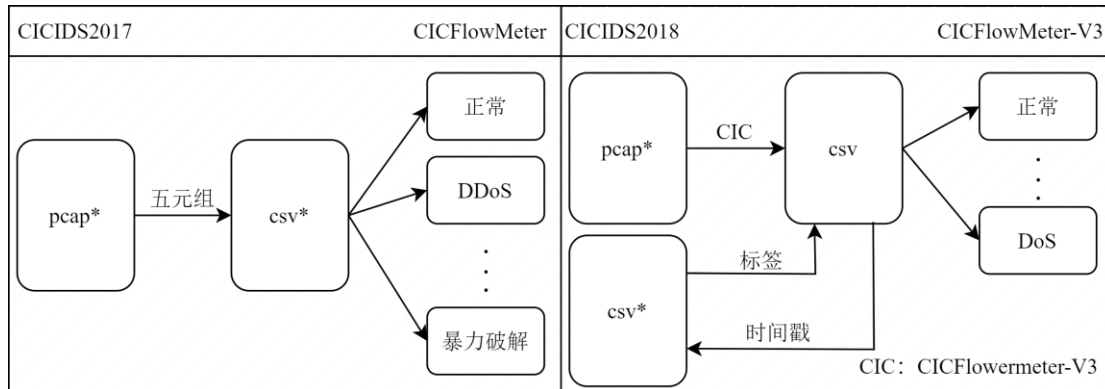


图 3.1 对 CICIDS2017 和 CICIDS2018 进行标注的流程

注：右上*表示使用的是官方的原始文件，CIC 表示的是 CICFlowMeter-V3。

3.1.3 异构数据特征提取策略

为了处理标记后的数据，我们采用了两种异构数据处理方法。首先，我们从每条流量中选取前 16 个包的前 256 个字节作为流量特征图。这种方法虽然可能导致原始数据的部分信息丢失，但可以提高检测的及时性，并减少样本的大小。其次，我们从完整流量中提取出详细的流量特征集，以便于使用基于流的入侵检测^[49]，包括包的数量、总流量、持续时间、平均包大小、最大包大小、最小包大小、包大小分布、到达时间的均值和方差、TCP 标志统计等，这些特征将用于深

入分析和模型训练^[50]。通过这种方法，我们可以最大限度地利用数据集，确保模型训练的高效和准确，网络特征集的组成成分如表 3.1 所示。此外需要注意的是，这里面的 Packet Size Distribution_1 表示数据大小在 0~200 这个范围，Packet Size Distribution_2 表示数据大小在 201~400，其余内容同理。TCP Flags 用于表示 TCP 头的多个标志位，这些标志标识传输状态，如同步（SYN）、确认（ACK）、结束（FIN）等。TCP Flags Count_A 表示 TCP 中 ACK 标识的计数，类似的 FA、SA、S、R、PA 分别代表同时包含 FIN-ACK 的计数、同时包含 SYN-ACK 的计数、包含 SYN、RST 以及同时包含 SYN-ACK 的计数。这里面的 Packet Size Distribution Combination 和 TCP Flags Combination 用于探究特定分布组合的情况，这归属于离散特征，其余特征均归属与连续特征，此外，这里的 id 和 Label 不参与其中，仅用于匹配和标记并不作为训练的特征。

表 3.1 网络特征集的组成成分

网络特征集的组成成分		
id	Packet Size Distribution_2	TCP Flags Count_FA
Number of Packets	Packet Size Distribution_3	TCP Flags Count_SA
Total Traffic	Packet Size Distribution_4	TCP Flags Count_S
Duration	Packet Size Distribution_5	TCP Flags Count_R
Average Packet Size	Packet Size Distribution_6	TCP Flags Count_PA
Max Packet Size	Inter Arrival Time Mean	Packet Size Distribution Combination
Min Packet Size	Inter Arrival Time Variance	TCP Flags Combination
Packet Size Distribution_1	TCP Flags Count_A	Label

3.2 异构特征融合系统构建

本文所提的多源异构特征融合的小样本网络入侵检测系统，是将原始流量数据通过异构处理成两种类型的数据，并两种类型进行单独处理，分别经过流量特征图特征提取模型（Traffic Feature Graph Feature Extraction Model, G-Model）和网络特征集特征提取模型（Network Feature Set Feature Extraction Model, S-Model）处理，最后在特征融合器中使用特征融合的方法，处理输入的内容，此外还引入迁移学习，使其适应小样本的环境，流程图如图 3.2 所示。

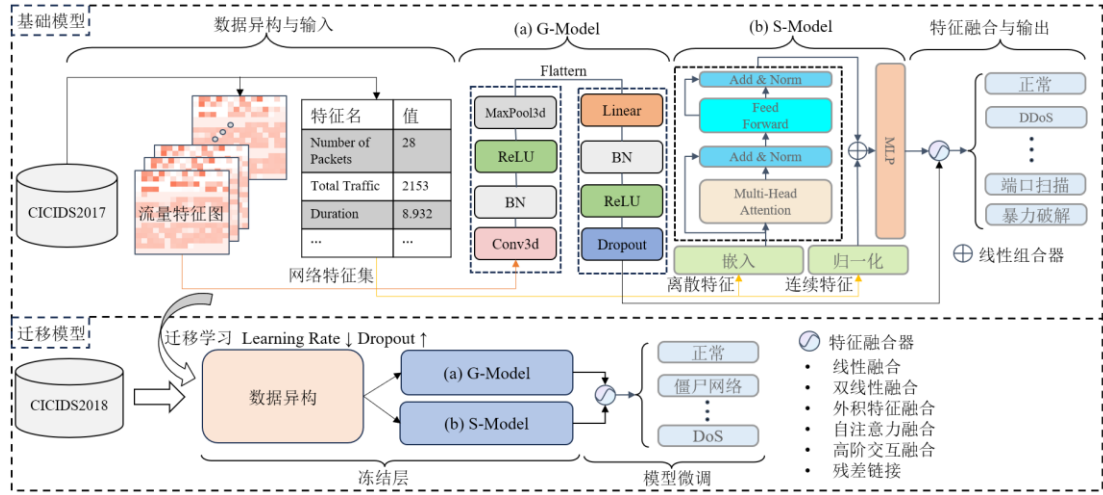


图 3.2 多源异构特征融合流程图

注：G-Model 和 S-Model 中虚线框所包含的内容表示需要多次堆叠。

3.2.1 数据异构

我们针对原始的网络流量提出了两种不同的表示方法：流量特征图和网络特征集。流量特征图描绘了网络数据包的内在联系，而网络特征集则综合了诸如数据包数量、总流量和会话持续时间等关键特征。从微观上，选择一条流量的前 16 个数据包的内容，将其按照顺序依次层叠，作为 G-Model 的输入；从宏观上、针对上述的流量，提取出诸如数据包数量、总流量和会话持续时间等关键连续特征和数据包分布组合、TCP 标志分布组合关键离散特征，作为 S-Model 的输入。

3.2.2 网络特征图特征提取器

G-Model 用于处理第一种异构数据，其中第 1~2 层的参数如表 3.2 所示，其中 Conv3d 表示的是使用 3D 卷积层，参数分别表示（输入通道，输出通道，卷积核数，步长，填充）；BatchNorm3d 表示的是 3D 卷积神经网络中的批归一化层，用于规范化各层输出，减少内部协变量偏移，在批归一化层中，每个特征通道的输出都会被规范化，使得输出的均值接近 0，标准差接近 1。这通常有助于网络的训练过程中更快收敛，同时提高模型在训练数据之外的泛化能力，其中的参数表示为（特征数量），这个参数指定要规范化的特征通道数，与其前一层卷积层的输出通道数相匹配，该层将学习相应的规范化参数；ReLU 是一个修正线性单元激活函数。它的作用是为模型引入非线性，使得模型能够学习更复杂的函

数关系,ReLU 激活函数没有在训练过程中需要调整的参数,所以将其置为 Nan;MaxPool3d 执行输入数据在空间维度下采样操作,通过在由 kernel_size 定义的窗口中取最大值来减小数据的尺寸,参数可表示为(卷积核数,步长),其中的(2,2)则表示池化窗口大小为 $2 \times 2 \times 2$,池化操作的步长为 2。这样就有效地将输入数据在空间维度上缩小了一半;Flatten 是扁平化层没有可学习的参数,所以也置为 Nan。它只是将输入维度展平成一个 1D 向量。

将处理成 1D 向量的内容输入第 3~7 层,这 5 层的参数如表 3.3 所示,其中 Linear 表示的是全连接层,也称为密集层,参数如(4096,2048)表示每层的输入和输出大小;BatchNorm1d 是 1D 批归一化层,它们在每个批次中规范化前一层的激活值,保持激活值的平均值接近 0,标准差接近 1。例如,参数(2048)指的是该层中的特征数量;Dropout 这一层在训练时随机将一部分输入单元设置为 0,有助于防止过拟合。参数(0.1)指定了 dropout 比率;即每个单元有 10%的概率被舍弃。

表 3.2 G-Model 中第 1~2 层参数及其输出维度

层	参数	输出维度
Conv3d-1 Conv3d-2	(1,32,3,1,1) (32,64,3,1,1)	[-1,32,16,16,16] [-1,64,8,8,8]
BatchNorm3d-1 BatchNorm3d-2	(32) (64)	
ReLU	Nan	
MaxPool3d	(2,2)	[-1,32,8,8,8] [-1,64,4,4,4]
Flatten	Nan	[-1,4096]

表 3.3 G-Model 中第 3~7 层参数及其输出维度

层	参数	输出维度
Linear-3 Linear-4 Linear-5 Linear-6 Linear-7	(4096,2048) (2048,1024) (1024,512) (512,256) (256,128)	[-1,2048] [-1,1024] [-1,512] [-1,256] [-1,128]
BatchNorm1d-3 BatchNorm1d-4 BatchNorm1d-5 BatchNorm1d-6 BatchNorm1d-7	(2048) (1024) (512) (256) (128)	
ReLU	Nan	
Dropout	(0.1)	

3.2.3 流量特征集特征提取器

S-Model 用于处理第二种异构数据，并针对连续的内容和离散的内容进行单独处理如图 3.3 所示，对于离散输入首先通过嵌入层处理，在嵌入层之前需要指定输出维度和离散特征的个数，该层会将离散的值映射到一个连续的高维空间，每个类别特征都有一个单独的嵌入矩阵，其大小为类别数量与输出维度的乘积。然后，将经过嵌入层后的张量输入 Transformer^[28]中，这些块中包含自注意力、残差连接和归一化以及前馈全连接层等内容，并且根据任务不同，可以进行多次堆叠。这种模型可以关注输入特征中不同部分的相互关系。每个编码器块都有多个自注意力头，使得模型可以在不同的子空间中并行学习特征间的关系。对于连续输入，它们首先会被标准化，这个过程涉及从每个特征中减去均值并除以标准差，如公式（3-1）所示，其中 X 是原始数据， μ 是原始数据的均值， σ 是原始数据的标准差， X' 是标准化后的数据。

$$X' = \frac{X - \mu}{\sigma} \quad (3-1)$$

这一步有助于加速模型的训练过程并改善性能。标准化后的连续特征然后会与经过 Transformer 处理后的离散特征进行拼接，组成两种模态的混合输入。

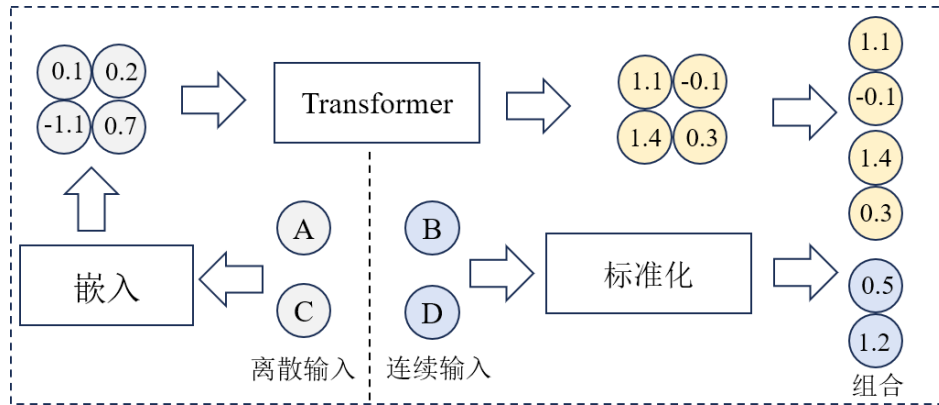


图 3.3 S-Model 中对离散特征和连续特征处理流程图

3.2.4 六种特征融合方法

本文提出六种融合方法，分别是线性融合、双线性融合、外积特征融合、自注意力融合、高阶交互融合、残差融合。将通过 G-Model 处理后的向量定义为 $x_1 \in \mathbb{R}^{d_1}$ ，通过 S-Model 处理后的向量定义为 $x_2 \in \mathbb{R}^{d_2}$ ，其中 d_1 和 d_2 分别表示输

出的维度，并设置默认输出维度相同（ $d_1 = d_2 = d$ ）。

（一）线性融合（Linear Fusion, LF）

这部分的线性融合具体指的是单线性融合，只进行简单的拼接操作，具体如公式（3-2）所示，其中 $\mathbf{x}_{\text{concat}} \in \mathbb{R}^{d_1+d_2}$ ，表示拼接后的向量具有 $d_1 + d_2$ 的维度。

$$\mathbf{x}_{\text{concat}} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} \quad (3-2)$$

将拼接后的结果再进行后续的线性变换，这部分公式如（3-3）所示，其中

$$\text{output} = \mathbf{W}_{\text{concat}} \mathbf{x}_{\text{concat}} + \mathbf{b} \quad (3-3)$$

$\mathbf{W}_{\text{concat}} \in \mathbb{R}^{o \times (d_1+d_2)}$ ，将 $(d_1 + d_2)$ 维空间映射到 o 维的输出空间， $\mathbf{b} \in \mathbb{R}^o$ 是偏置项，与输出的维度相对应。

（二）双线性融合（Bilinear Fusion, BF）

双线性融合不仅仅是简单地合并或连接特征，而是允许每对特征之间都有独立的权重。这在处理需要高度非线性特征组合的任务比较有效，具体的数学公式如（3-4）所示，其中 $\mathbf{W}^{(i)} \in \mathbb{R}^{d_1 \times d_2}$ 是第 i 个输出维度对应的权重矩阵，这是一个

$$\text{output}_i = \mathbf{x}_1^T \mathbf{W}^{(i)} \mathbf{x}_2 + \mathbf{b}_i \quad (3-4)$$

三维张量 $\mathbf{W} \in \mathbb{R}^{o \times d_1 \times d_2}$ 包含了每个输出维度特定的权重矩阵， \mathbf{b}_i 是偏置项，对应输出的第 i 维， o 则表示输出向量的维度。

（三）外积特征融合（Outer Product feature Fusion, OPF）

在这部分的外积是克罗内克积（Kronecker Product），能够组合每个特征维度，捕捉不同输入特征之间所有可能的相互作用，具体的数学计算公式如（3-5）所示，其中 \mathbf{P} 是一个 $d_1 \times d_2$ 的列向量，

$$\mathbf{P} = \mathbf{x}_1 \otimes \mathbf{x}_2 \quad (3-5)$$

每个元素定义为 $P_{ij} = x_{1i} x_{2j}$ 其中 i 是 \mathbf{x}_1 的索引， j 是 \mathbf{x}_2 的索引，再将 \mathbf{P} 展开成 \mathbf{P}_{flat} ， $\mathbf{P}_{\text{flat}} = \text{flatten}(\mathbf{P})$ ，最后的全连接层如公式（3-6）所示，其中 $\mathbf{W} \in \mathbb{R}^{o \times (d_1 \times d_2)}$ ， $\mathbf{b} \in \mathbb{R}^o$ ，

$output \in \mathbb{R}^o$ 。

$$output = W \cdot P_{\text{flat}} + b \quad (3-6)$$

（四）自注意力融合（Self-Attention Fusion, SAF）

首先，线性组合两个向量的维度，这部分的公式与公式（3-2）保持一致，其中 $x_{\text{combined}} \in \mathbb{R}^{d_1+d_2}$ ，表示拼接后的向量具有 d_1+d_2 的维度，其次设置查询(Query)、键(Key)、值(Value)。其中 $W^Q, W^K, W^V \in \mathbb{R}^{(d_1+d_2) \times (d_1+d_2)}$ 是可学习的参数矩阵，

$$\begin{cases} Q = W^Q \cdot x_{\text{combined}} \\ K = W^K \cdot x_{\text{combined}} \\ V = W^V \cdot x_{\text{combined}} \end{cases} \quad (3-7)$$

利用查询和键计算注意力分数，并使用这些分数对值进行加权，如公式（3-8）所示，在这其中 d_k 是键向量的维度， softmax 用来保证每一行的总和为 1。

$$output = \text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (3-8)$$

（五）高阶交互融合（Higher-order Interaction Fusion, HIF）

高阶交互通过引入特征的平方和特征间的交叉乘积，可以增强模型处理非线性关系的能力。主要考虑了特征的二次形式以及特征对的乘积，这种方法不仅捕获了特征的线性依赖（通过原始特征），还捕获了特征的单个非线性（通过平方项）和特征对之间的非线性相互作用（通过交叉乘积项）。这可增强模型对复杂数据结构的理解，尤其是在处理具有内在非线性关系的数据时。首先合并两个向量，如公式（3-2）所示，再对其进行二项式展开，展开的内容包括原始特征 x_1 和 x_2 它们的平方 x_1^2 和 x_2^2 ，以及交叉项乘积 x_1x_2 ，其中的 x_1x_2 公式如（3-9）所示，

$$x_1x_2 = x_1 \odot x_2 \quad (3-9)$$

在这一部分使用的是哈达玛积（Hadamard Product），其中 x_1x_2 的第 i 个元素是 x_1 中第 i 个元素与 x_2 中第 i 个元素相乘所得。最终高阶交互的公式如（3-10）所示。

$$\mathbf{x}_{expanded} = [\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_1^2, \mathbf{x}_2^2, \mathbf{x}_1\mathbf{x}_2] \quad (3-10)$$

(六) 残差融合 (Residual Fusion, RF)

对于残差^[51]融合来说, 分为直接路径和残差路径, 首先需要组合两个向量如公式 (3-11) 所示, $\mathbf{x}_{combined}$ 为组合后的向量。直接路径如公式 (3-9) 所示, 这个过

$$\mathbf{y}_{direct,combined} = \mathbf{W}_{direct}\mathbf{x}_{combined} + \mathbf{b}_{direct} \quad (3-11)$$

程映射了输入特征到新的空间, 为后续的非线性变换提供了基础。其中 $\mathbf{W}_{direct} \in \mathbb{R}^{o \times (d+d)}$ 是权重矩阵, 残差路径的数学公式表示如 (3-12) 所示, 首先进

$$\mathbf{z}_{combined} = \mathbf{W}_{res1}\mathbf{x}_{combined} + \mathbf{b}_{res1} \quad (3-12)$$

行线性变换如公式 (3-12) 所示, 再经过 ReLU 激活函数, ReLU 函数将所有负值映射为零, 而保留正值不变, 这样的非线性变换是神经网络能够捕捉复杂函数的关键。在此基础上再进行线性变

$$\mathbf{z}_{activated,combined} = \text{ReLU}(\mathbf{z}_i) \quad (3-13)$$

换得到公式 (3-14) 中的输出, 进一步加强了模型对于输入数据的理解。每个残差块都试图捕捉特征的增量, 从而允许模型层层递进地学习更复杂的特征组合, 其中 $\mathbf{W}_{res1}, \mathbf{W}_{res2} \in \mathbb{R}^{o \times (d+d)}$ 和 $\mathbf{b}_{res1}, \mathbf{b}_{res2} \in \mathbb{R}^o$ 是残差路径中各层的权重矩阵和偏置项。

$$\mathbf{y}_{res,combined} = \mathbf{W}_{res2}\mathbf{z}_{activated,combined} + \mathbf{b}_{res2} \quad (3-14)$$

最后将直接路径和残差路径合并再通过激活函数输出, 如公式 (3-15) 所示。

$$\mathbf{y} = \text{ReLU}(\mathbf{y}_{direct,combined} + \mathbf{y}_{res,combined}) \quad (3-15)$$

3.2.5 迁移模型

迁移学习^[52]是一种有效的深度学习方法, 它允许我们将从一个领域 (源域) 学习到的知识迁移到另一个相关的领域 (目标域), 如图 3.4 所示。这种技术特别适用于目标域的标注数据稀缺的情况, 因为它可以利用源域的大量数据来提高模型在目标域上的泛化能力。具体而言, 这一部分充分利用多个数据源的数据,

使其适应小样本情况，特别是当目标域中的标注样本相对较少时，它可以帮助改善模型在新数据上的性能。先在一个大的、丰富的源数据集上训练一个模型，让模型学习通用的特征表示。再冻结模型的前半部分，只允许后续特征融合的参数可以被修改，在新的目标域上对模型进行微调，与此同时，为了防止在小样本环境下出现过拟合的情况，我们增加了 Dropout 超参数的大小，同时降低了学习率，防止破坏原始的参数。

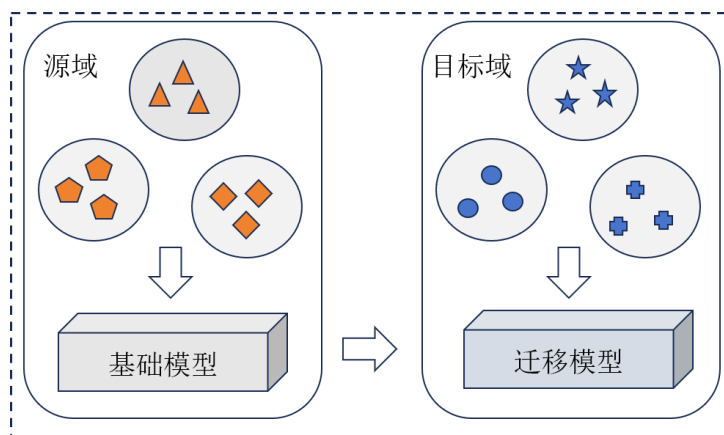


图 3.4 迁移学习示意图

4 系统设计与验证

4.1 网络入侵检测数据集

本文使用两个广泛认可的网络流量数据集，CICIDS2017 和 CICIDS2018，这些数据集由 CIC 提供。包含了正常和恶意流量的详细样本，用以模拟现实世界中的网络环境，从而有效地评估网络入侵检测系统（Network Intrusion Detection System, NIDS）的性能。

在 CICIDS2017 数据集上包括最常见的网络攻击类型，如 DDoS、DoS、心脏出血（Heartbleed）、端口扫描、Web 攻击（例如 SQL 注入和命令注入）等。数据集按时间顺序分布在五天中，每天收集的数据都代表不同类型的攻击活动或正常流量。CICIDS2018 数据集在 2017 年的基础上进一步扩展，增加了更多种类的网络攻击和更高的数据多样性，在选择 CICIDS2017 和 CICIDS2018 数据集中的样本时，我们均匀地收集了每种类型 1000 个样本，以保持数据平衡。这种方法旨在避免模型在训练过程中出现偏差，并确保对所有类别的泛化性能进行公正评估。此外选择每种类型的样本数量到 1000 个是为了保持对计算资源的合理需求，同时确保数据集的大小足以训练出具有鲁棒性的模型，具体所选的类型及其描述表 4.1 和表 4.2 所示。

表 4.1 所选 CICIDS2017 数据集的类型及其描述

标签	类型	描述
Brute Force	暴力破解	通过尝试多种用户名和密码组合，直到成功破解。
Port Scan	端口扫描	通过扫描服务器各个端口，试图发现开放的端口及其服务。
Web Attacks	Web 攻击	针对 Web 应用程序的攻击，包括 SQL 注入、命令注入等。
DDoS	DDoS	分布式拒绝服务攻击，造成目标服务器或网络资源不可达。
Normal	正常	代表正常的、没有恶意活动的网络流量。

表 4.2 所选 CICIDS2018 数据集的类型及其描述

标签	类型	描述
Botnet	僵尸网络	通过抓取浏览器的用户记录和表单来窃取个人信息。
DDoS	DDoS	分布式拒绝服务攻击，造成目标服务器或网络资源不可达。
DoS	DoS	针对 Web 应用程序的攻击，包括 SQL 注入、命令注入等。
Brute Force	暴力破解	通过尝试多种用户名和密码组合，直到成功破解。
Normal	正常	代表正常的、没有恶意活动的网络流量。

4.2 检测系统中的评价指标

在本研究中，我们采用以下评价指标来全面评估模型的性能，包括准确率（Accuracy Rate, ACC）、精确率（Precision Rate, PR）、检测率（Detection Rate, DR）、F1 分数（F1-Score），为了对多分类问题进行更加精确的评估，我们引入宏平均的概念，宏平均在多分类问题中非常关键，尤其是当数据集不平衡时，即各类别的样本数量不相同。这种情况下，少数类的性能可能会在模型的总体评估中被忽略。宏平均通过为每个类别赋予相同的权重来解决这一问题，确保模型对所有类别的泛化能力。具体地，选择了宏平均精确率（Macro Precision Rate, Macro-PR）和宏平均检测率（Macro Detection Rate, Macro-DR）以及宏平均 F1 分数（Macro F1-Score, Macro-F1），需要注意的是，整个模型的 ACC 不区分具体不同的类别，仅仅关注所有预测正确的情况。具体的表达式如公式（4-1）所示。

$$\left\{ \begin{array}{l} ACC = \frac{\sum_{i=1}^N C_{ii}}{\sum_{i=1}^N \sum_{j=1}^N C_{ij}} \\ Precision_i = \frac{TP_i}{TP_i + FP_i} \\ Detection_i = \frac{TP_i}{TP_i + FN_i} \\ Macro-PR = \frac{1}{N} \sum_{i=1}^N Precision_i \\ Macro-DR = \frac{1}{N} \sum_{i=1}^N Detection_i \\ Macro-F1 = 2 \cdot \frac{Macro-PR \times Macro-DR}{Macro-PR + Macro-DR} \end{array} \right. \quad (4-1)$$

4.3 基线模型的参数设置

本文实验中用到的软/硬件环境如下：CPU 为 Intel (R) Xeon (R) Platinum 8352V CPU @ 2.10GHz，内存为 128 GB，操作系统为 Ubuntu 20.04，GPU 为 RTX 3090 (24GB)。采用 CUDA 11.3 作为 GPU 加速库，使用了 Python 3.8 和深度学习框架 PyTorch 1.10.0。

本方法中需要修改的模型结构和参数设置较多，为了便于处理和探究方法的

优劣，我们按照如表 4.3 所示的方式组成基线模型，后续实验中若不加以额外说明，则均表示所选值或参数与基线模型中的参数设置相同。其中的 heads 表示为 Transformer 模型中的多头注意力机制中头的数量，dim 表示的是模型中嵌入维度，depth 表示的模型的深度，也表示了模型重复堆叠的次数。attn_dropout 表示在注意力权重中使用的 dropout 比率，用于防止注意力层的过拟合。ff_dropout 表示在前馈网络中使用的 dropout 比率，也是用于防止过拟合。mlp_hidden_mults 表示在多层感知机（Multilayer Perceptron，MLP）中用于确定隐藏层大小的乘数因子，这里为（4，2），意味着每个连续的隐藏层大小是前一个层的 4 倍和 2 倍。lr（learning rate）则表示为学习率的大小，决定了在反向传播时参数更新的步长大小，需要注意的是在迁移模型中的 lr 和 dropout 相较于基础模型分别下降和上升了一定程度以防止破坏原始模型。

表 4.3 基线模型的架构及其超参数配置

模型架构（堆叠次数/融合方法）	超参数	数值
上层 G-Model（2）	padding	1
下层 G-Model（5）	dropout	0.1
S-Model-Transformer（6）	heads	8
	dim	32
	depth	6
	attn_dropout	0.1
	ff_dropout	0.1
S-Model-MLP	mlp_hidden_mults	（4，2）
基础模型（LF）	lr	0.001
	dropout	0.1
迁移模型（LF）	lr	0.0001
	dropout	0.3

4.4 实验设置

为了更深入地评估和优化我们的模型，我们在 4.4 节中设计了 3 个阶段性实验。每个阶段都从不同角度探讨检测系统中不同模型的性能，并找出提升效果的关键策略，最后得到最佳的模型组合。

4.4.1 样本敏感度实验

在本实验中，先根据表 4.3 中的参数配置，构建基线模型，并对模型的样本数量进行限制。这一步骤有助于初步评估模型在数据受限条件下的表现，并为训练轮次等实验参数设定基准。随后，我们使用不同数量的样本在基础模型上进行实验，以探索其对样本数量的敏感度。同时，我们引入额外的微调数据集进行迁移学习，并对迁移后的模型实施相同的样本数量限制。此过程旨在评估迁移学习在提升小样本环境下模型表现的效果、基础模型对样本数量的依赖性，以及迁移模型对样本数量的敏感性。

此外，我们进行了双向交叉迁移实验，我们进一步探讨了迁移模型的适应性和灵活性。在该实验中，迁移模型的预训练模型使用的是基础模型。为了对比基础模型和迁移模型在不同数据集上的表现，我们交替使用 CICIDS2017 和 CICIDS2018 作为预训练即基础模型的数据集，以探索和评估模型在不同时间点、不同网络环境下收集的数据上的泛化能力和适应性。这一系列实验设计旨在全面了解模型在各种条件下的表现，从而为进一步的优化和应用提供依据。

4.4.2 特征融合实验

在本节中，我们对六种不同的特征融合策略进行了探讨，并比较了这些策略对模型性能指标的具体影响。并在 CICIDS2017 和 CICIDS2018 这两个基准数据集上进行全面评估。通过这一系列的实验，得到更加适合本模型的融合策略，并深入探究各种融合策略针对不同类型的表现情况，最终得到最佳的融合方法。

4.4.3 迁移实验中的模型调优

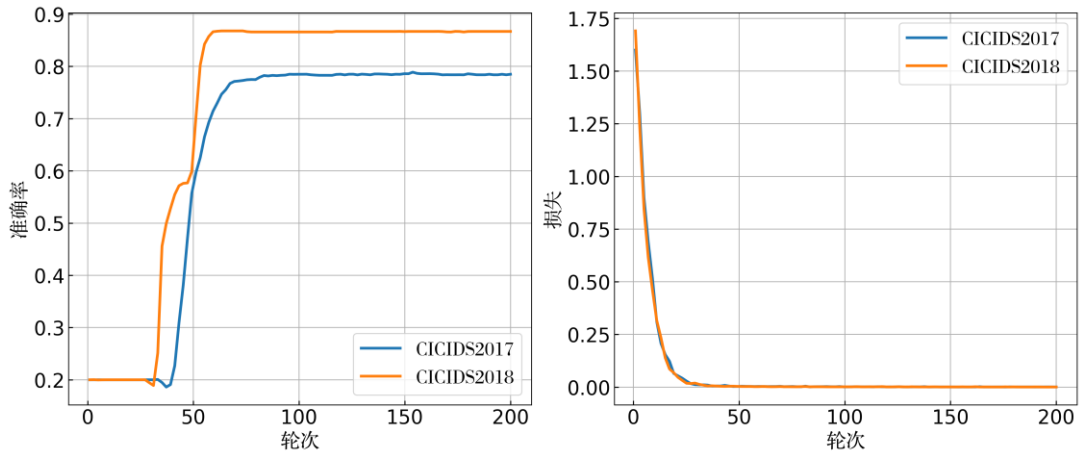
在完成上述的实验后，对模型的整个变化趋势和最佳的模型组合有较为明确的认识，在这一步主要是完善迁移实验在模型上的调优，主要是处理现有的可训练的原始类样本较少，并且可训练的新类样本也较少，以此来模拟在小样本环境下，原始类和新类的可用样本都较少的情况。伴随网络入侵技术的发展和更新，这一场景也变得较为常见。在这一部分中，我们交替将 CICIDS2017 数据定义成源域，将 CICIDS2018 数据定义成目标域，在源域和目标域中都加以小样本的限制，并且使用合适的特征融合策略，最后判断经过迁移后的模型是否可以在样本

量受限的情况下表现良好，并且判断特征融合策略是否有较高的泛化性能，在目标域的表现结果是否有所提升。

4.5 检测结果

4.5.1 样本敏感度实验

在本节中，我们对基础模型加以小样本限制，并分别在 CICIDS2017 和 CICIDS2018 数据集上进行实验，同时不进行迁移学习，检测结果如图 4.1 所示，其中 (a) 表示的是在 CICIDS2017 和 CICIDS2018 的准确率表现情况 (b) 表示的是损失的表现情况。

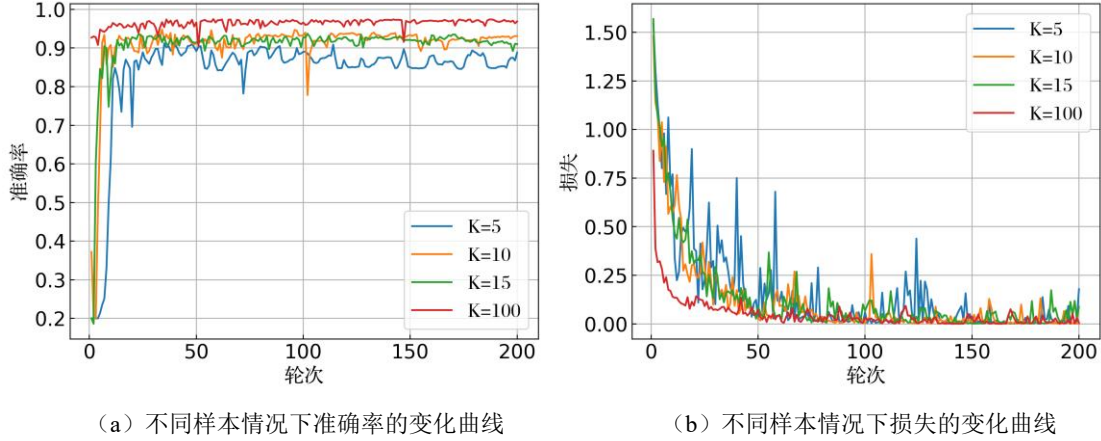


(a) 准确率根据训练轮次的变化曲线

(b) 损失值根据训练轮次的变化曲线

图 4.1 基础模型在 CICIDS2017 和 CICIDS2018 数据集上的表现情况

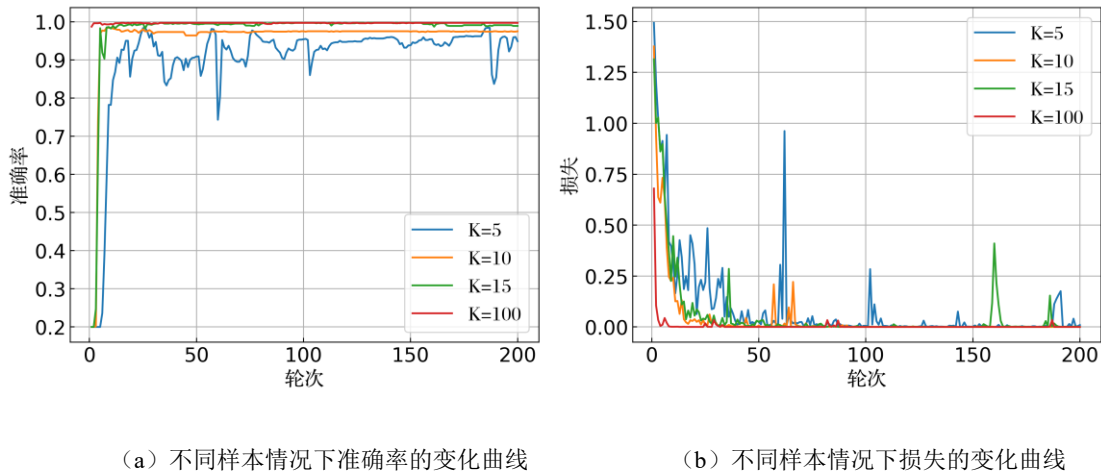
可以发现，在基础模型中，模型的损失值在 0~100 轮次迅速下降，在 100~200 轮次基本稳定，波动较小，因此，我们选择轮次为 200 轮作为后续实验的初值，并且为了进一步探究模型的波动与样本数量的关系，我们进一步设置不同的样本数量，用于反映不同样本 (K) 值下，模型的表现情况，用于探究模型对 K 值的敏感程度以及模型的波动情况，实验的结果如 4.2 和 4.3 所示。



(a) 不同样本情况下准确率的变化曲线

(b) 不同样本情况下损失的变化曲线

图 4.2 基础模型在 CICIDS2017 数据集的不同样本上的表现情况



(a) 不同样本情况下准确率的变化曲线

(b) 不同样本情况下损失的变化曲线

图 4.3 基础模型在 CICIDS2018 数据集的不同样本上的表现情况

从图 4.3 可以发现，在基础模型仅需 25 个轮次，准确率就能迅速上升，若不限制样本数量 ($K=100$)，模型迅速收敛且准确率和损失的波动较小，但限制样本数量后 ($K=5$ 、 10 、 15)，模型的准确率均有不同程度的降低，且随着样本数量的降低，模型的波动程度越来越大。

为进一步提升模型的性能，并降低模型的波动程度，让模型更加稳定，在基础模型的架构不变的基础上引入迁移模型，先保存基础模型的 K 值为 100 的模型，并进行迁移学习，注意这里我们将基础模型记为 **B (Base-Model)**，所需样本数量通过“-”进行连接，此外如果是迁移学习那么将第一个字母记为 **T (Transfer-Model)**，所需的基础模型（源域）的样本和迁移学习（目标域）的样本也通过“-”链接（例如仅使用基础模型并且样本数量是 10，则记为 **B-10**，若使用源域 100 的样本数量进行训练，后续迁移到 10 样本数量下的目标域则会被记为 **T-100-10**）。为探究迁移学习在是否能提升小样本环境下模型的表现情况，

我们分别在两个数据集上设计了不同样本（5、10、15）组成（T-100-5、T-100-10、T-100-15）的变化曲线，如图 4.4 和 4.5 所示，需要说明的是在迁移模型中我们降低了学习率等相关参数，此外由于初始的轮次在 200 时模型还未完全收敛，故将训练的轮次进行适当的增加。

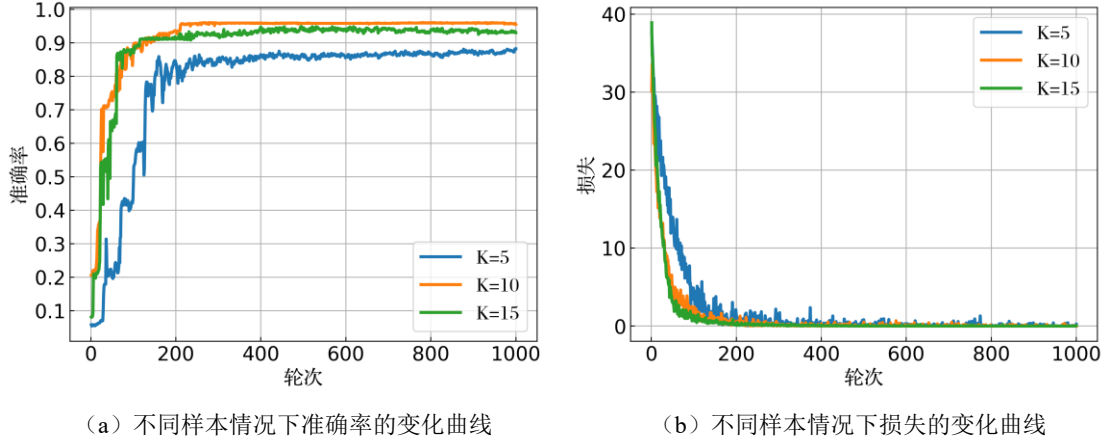


图 4.4 迁移模型在 CICIDS2017 数据集的不同样本上的表现情况

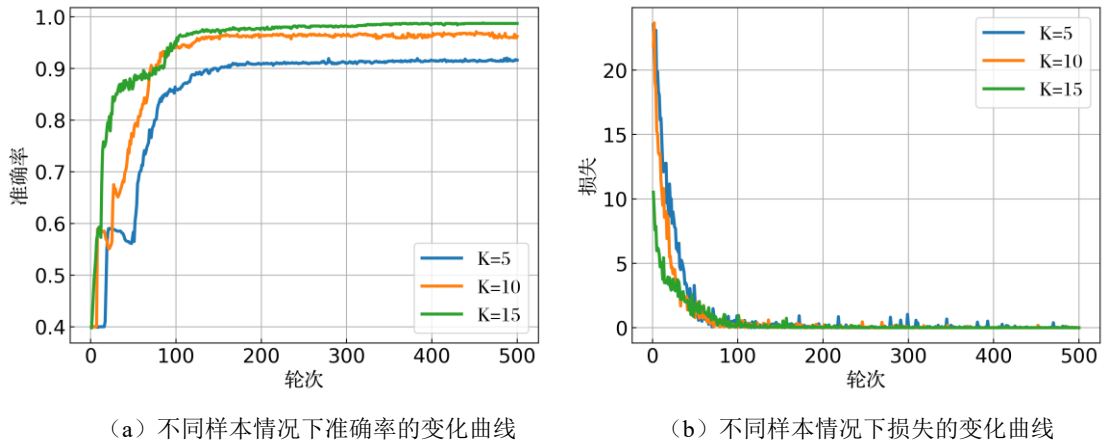
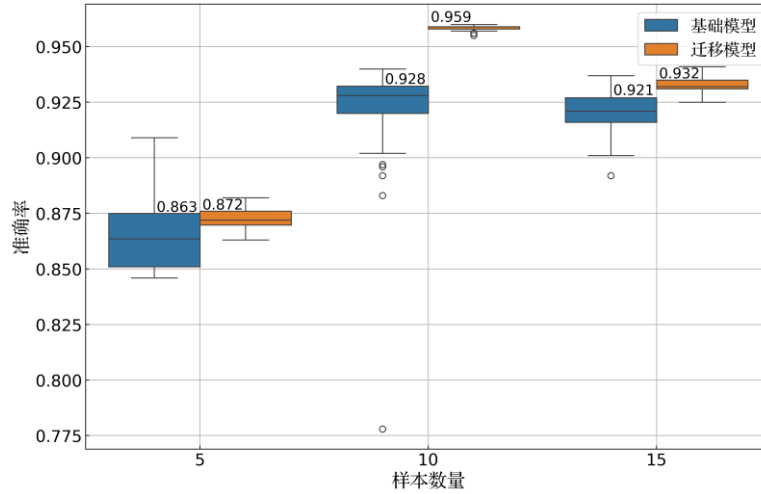
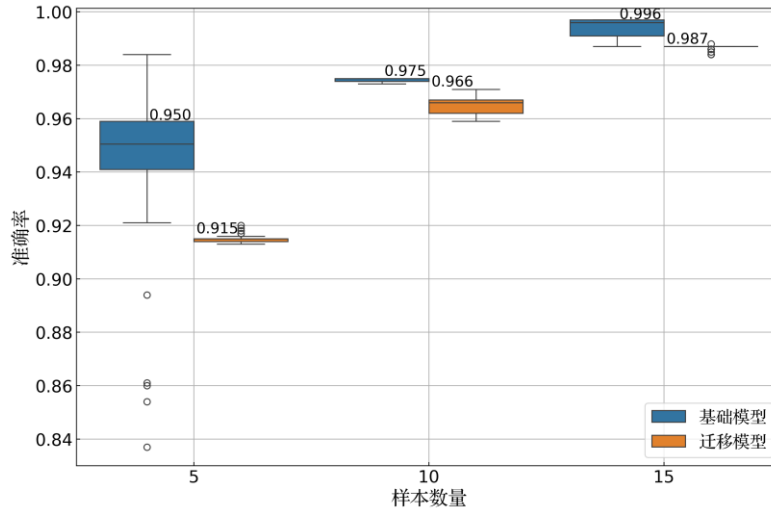


图 4.5 迁移模型在 CICIDS2018 数据集的不同样本上的表现情况

从图 4.5 的准确率曲线也可以发现，随着样本数量的增加，模型总体上的准确率都有所增加，并且涨幅较为明显，此外通过对比图 4.1~4.5 可以初步判断使用迁移模型会变得更加稳定。为了更加详细地反映基础模型和迁移模型的波动情况，我们绘制了如图 4.6 所示的箱线图，其中的数值展示的是中位数。



(a) 在 CICIDS2017 数据集上的表现情况



(b) 在 CICIDS2018 数据集上的表现情况

图 4.6 两种模型在两种数据集上准确率的波动情况

从波动情况可以发现，基础模型对数据集有较高的敏感程度，对于 CICIDS2017 数据集来说，表现欠缺，但对于 CICIDS2018 来说表现良好，使用 CICIDS2018 作为迁移模型的训练数据，也可以对 CICIDS2017 的检测起到促进作用，并且一定程度上减少了模型的波动，提高了稳定性。但是反过来，当使用 CICIDS2017 作为迁移模型的训练数据，无论是从准确率还是损失来说，都不如原始的基础模型，这也进一步印证了本模型对数据集有一定的敏感度。

仅仅使用这些指标还不足以全面评估模型，于是我们更进一步，搜集了多种模型在两种数据集上的表现结果，具体结果如表 4.4 和 4.5 所示，需要注意的是，这里的 K 值既反映的样本也反映了模型（例如，B-5 表示的是使用基础模型，

并且训练的样本数量是 5)，从图中可以发现，如果目标域的数据过少，那么使用迁移学习的效果竟会降低，如表 4.4 的 B-5 和 T-100-5 所示，降低了 2.8%，这可能是在目标域数据非常有限的情况下，即使是经过微调的模型也可能过度学习这些少量数据上的特定特征，而无法泛化到未见过的数据上。因此我们增加在迁移模型上的样本数量，结果均保持上升，分别如表 4.4 的 B-10、B-15 和 T-100-10、T-100-15 所示，分别上涨 1.2%、2.3%。

此外我们还从表 4.5 中发现无论是在哪一种情况下的迁移模型均不如基础模型，这也和我们之前的猜测类似，模型对于数据集的敏感度较大，在这里的迁移模型使用的是双向交叉迁移实验，这表明用于预测 CICIDS2018 的基础模型来自于 CICIDS2017，而从表 4.4、4.5 中在 CICIDS2017 数据集 B-100 的表现要明显低于 CICIDS2018 数据集 B-100 数据集的表现情况，这导致在迁移学习即使有较多的样本，也不能很好的调整模型的参数，反而会降低的模型的性能。因此，我们发现了迁移模型对初始化模型的数据集较为敏感。

表 4.4 两种模型在 CICIDS2017 数据集上的表现情况

模型	K	ACC	Macro-DR	Macro-F1	Macro-PR
基础模型	B-5	0.910	0.9147	0.9135	0.9123
	B-10	0.948	0.9497	0.9493	0.9489
	B-15	0.937	0.9422	0.9408	0.9395
	B-100	0.975	0.9755	0.9754	0.9752
迁移模型	T-100-5	0.882	0.8903	0.8882	0.8861
	T-100-10	0.960	0.9627	0.9620	0.9614
	T-100-15	0.949	0.9503	0.9500	0.9497

表 4.5 两种模型在 CICIDS2018 数据集上的表现情况

模型	K	ACC	Macro-DR	Macro-F1	Macro-PR
基础模型	B-5	0.984	0.9844	0.9843	0.9842
	B-10	0.984	0.9845	0.9844	0.9843
	B-15	0.997	0.9970	0.9970	0.9970
	B-100	0.997	0.9970	0.9970	0.9970
迁移模型	T-100-5	0.920	0.9242	0.9231	0.9221
	T-100-10	0.971	0.9712	0.9712	0.9711
	T-100-15	0.988	0.9881	0.9881	0.9880

4.5.2 特征融合实验

在这一节中，我们尝试使用更加复杂的特征融合进行实验，引入更加复杂的层模型，来保证模型能够有足够多的层结构权重可以在迁移学习中被修改，为此我们引入更加细化的特征融合的方式，使其更加适应小样本环境。

首先，我们细致探讨了六种特征融合策略，对比了它们对模型性能指标的影响，从而揭示了模型对不同特征融合技术的敏感度，通过调整特征融合的参数，我们可以评估哪些融合策略对模型的准确性、鲁棒性和泛化能力影响最大。这些实验共同构成了对所提出模型全面性能评估的实验框架。通过这一系列实验旨在通过多角度的评估和优化，全面提升模型的性能，并深入了解其在各种条件下的表现。通过这些细致的实验设计，我们期望最终能够提出一个健壮、高效且具有较强泛化能力的模型

需要注意的是，由于在特征融合中的学习率会被降低，为了保证模型能足够收敛故，在本实验中将训练轮次范围被划定到 1000 轮，并收集最后的 100 轮的数值，用于反映六种特征融合策略的表现情况，并通过准确率和损失这两个指标进行判断，如图 4.7 所示，从准确率和损失值随着训练轮次的波动曲线可以发现，对默认的 LF 来说，虽然损失比较稳定，但是准确率也较低，在 0.900 附近波动，此外 BF 和 RF 的损失也有较大的波动，而使用 SAF 中的准确率和损失情况均表现较好。

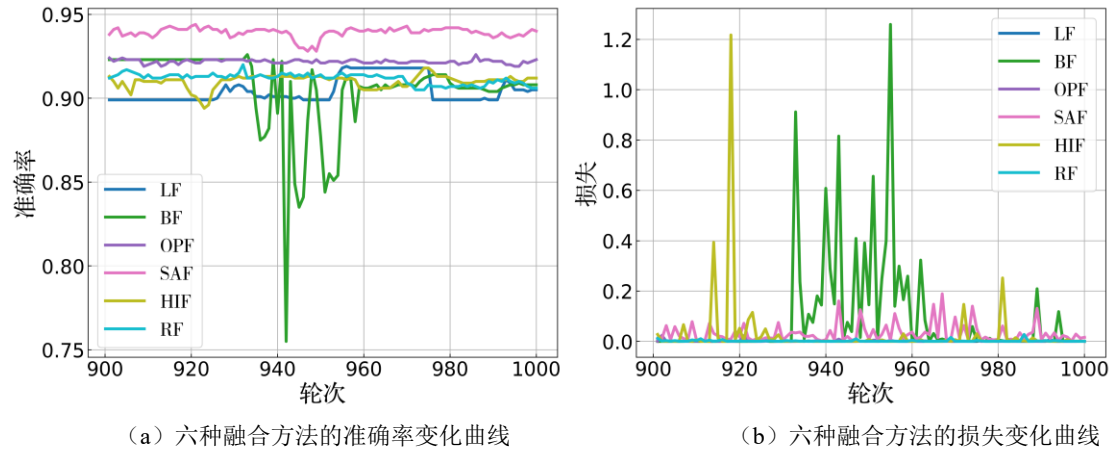


图 4.7 在 CICIDS2017 数据集上的六种特征融合方法表现情况

此外我们获取这六种方法在 100 轮测试集上的表现结果，重点在于探究多种特征融合方法在测试集上的表现情况，实验结果如图 4.8 所示，可以发现模型在

测试集上的表现情况与图 4.7 的实验结果保持一致。

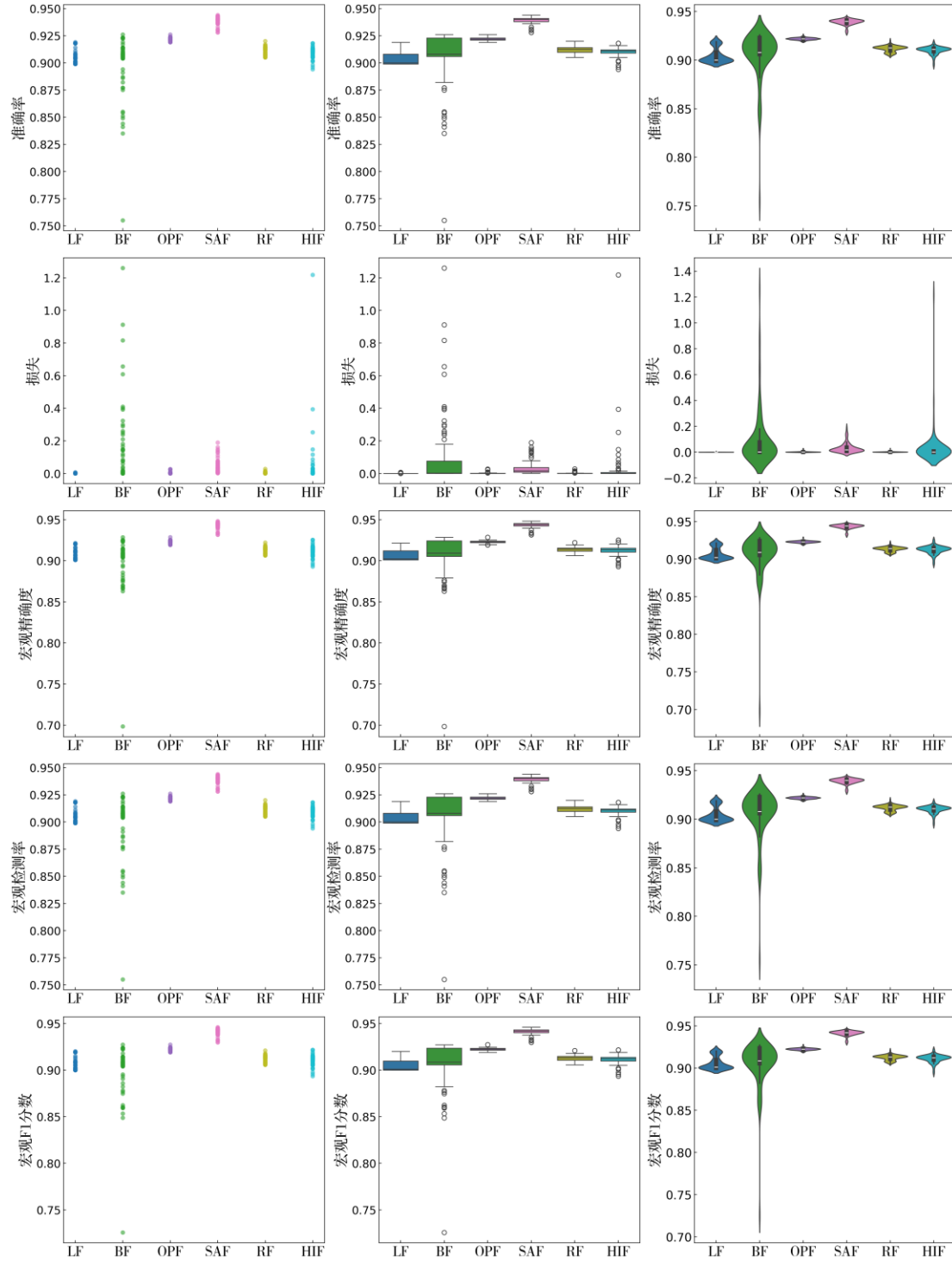


图 4.8 六种特征融合方法在多种评价指标下的表现情况

当然，图 4.8 仅仅反映模型的波动情况，不能更加细致地反映模型对每一种类型的检测情况，故为了进一步反映模型在各个类型上的表现情况，我们在上述 6 种方法中，重复测试 100 轮实验，并累加对应的混淆矩阵，如图 4.9，从图中

可以发现,对于默认的基础模型(LF)来说,在处理DDoS和端口扫描问题上表现欠佳,这一问题也延续到OPF中,但是无论是哪一种方法的平均准确率相较于基准模型都有一定幅度的上升,分别上升0.55%、2.86%、1.72%、0.11%、0.71%,其中SAF的上升幅度最大。

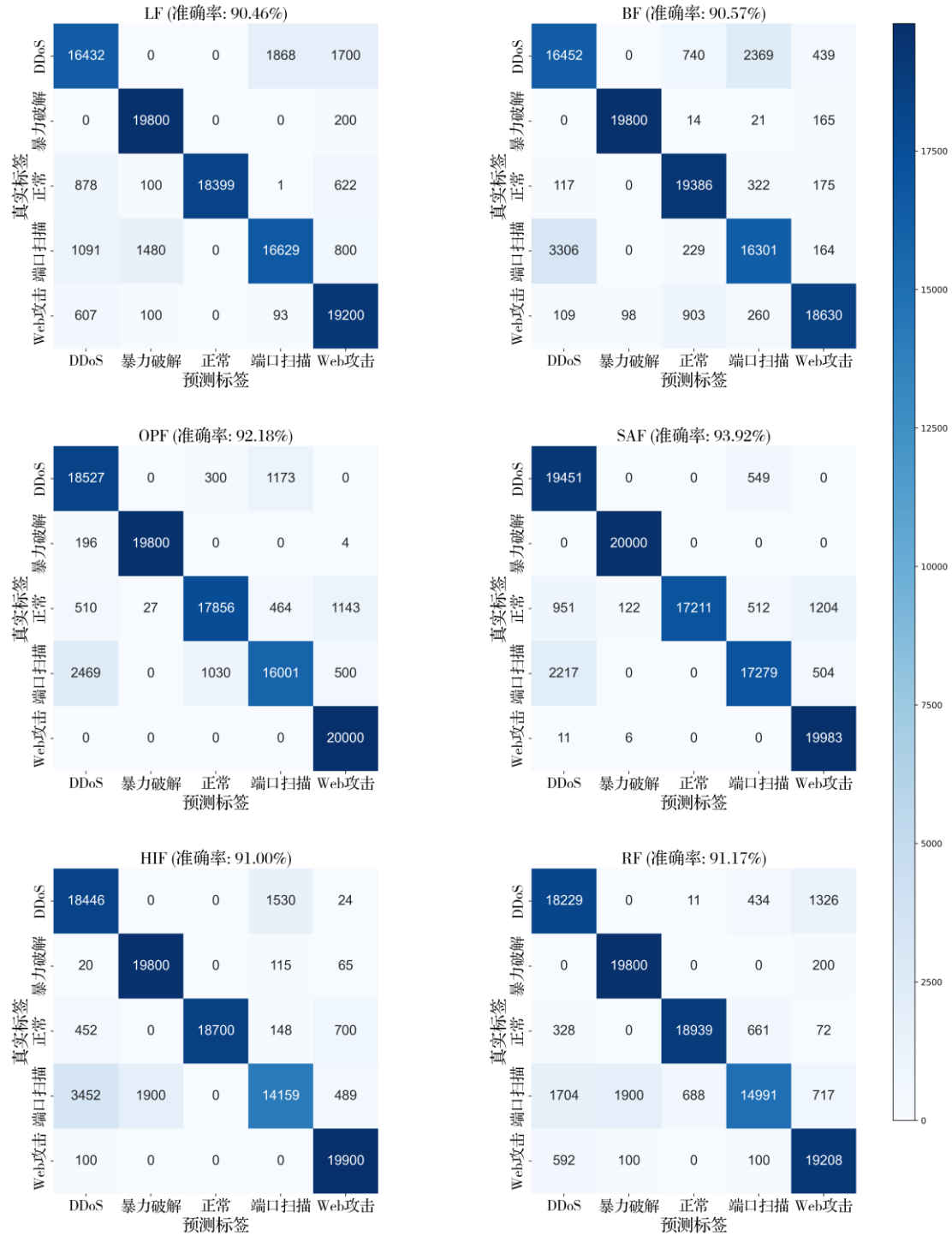


图 4.9 六种特征融合方法的混淆矩阵

综上所述，从数据波动情况来说，LF 和 BF 波动情况较大，并且 BF 中还存在较多的异常值，其余的方法的波动较小；从多种评估指标来看 SAF 在 ACC、Macro-PR、Marco-DR、Marco-F1 均表现最佳；从各种类型下的检测结果来看，SAF 可以有效提高，在基线模型上表现不好的 DDoS 和端口扫描等类型的检测结果。综合来看，我们选择 SAF 替换默认的 LF，原因如下：从图 4.7 表明 SAF 下的准确率相较于基线模型有所提升，表明此时的模型可提取和学习更加复杂的内容，这为适应小样本环境提供有力支持。从图 4.8 和 4.9 中表明，该方法的实验结果较为稳定，且能提高针对不同类型的检测效果，这也为后续能较好地处理迁移学习从源域到目标域类型的转变提供了有力保障。

4.5.3 迁移实验

图 4.10 中的横坐标代表不同模型和不同样本之间的组合情况，其中 B-5 表示使用基础模型，并且用于训练的样本数量是 5，B-100 以此类推表示使用基础模型并且用于训练的样本数量是 100，此外，T-5-5 表示使用的迁移模型，并且训练的模型来自 B-5，在此基础上额外使用 5 个训练样本在模型上进行微调，对于具体的数据集来说，如 CICIDS2017 的 T-5-5 表明先在 CICIDS2018 数据集上使用 5 个样本得到预训练模型，再在 CICIDS2017 数据集上使用 5 个样本进行模型微调，CICIDS2018 的 T-5-5 同理。

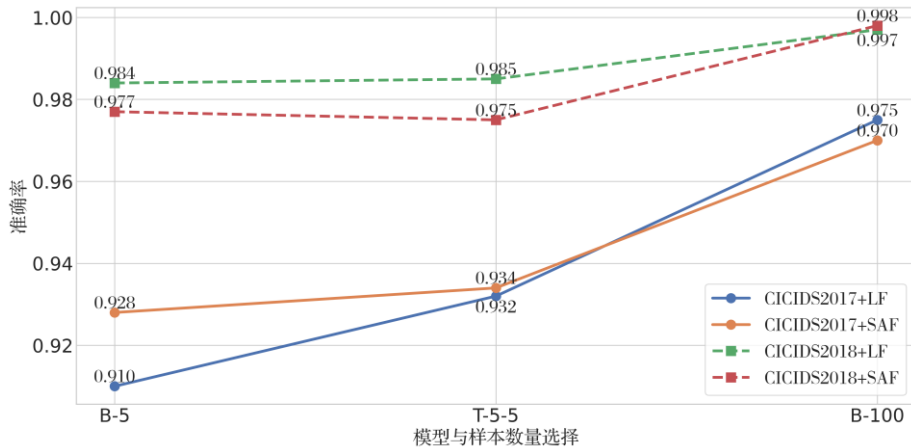


图 4.10 不同模型与不同样本组合下准确率的表现情况

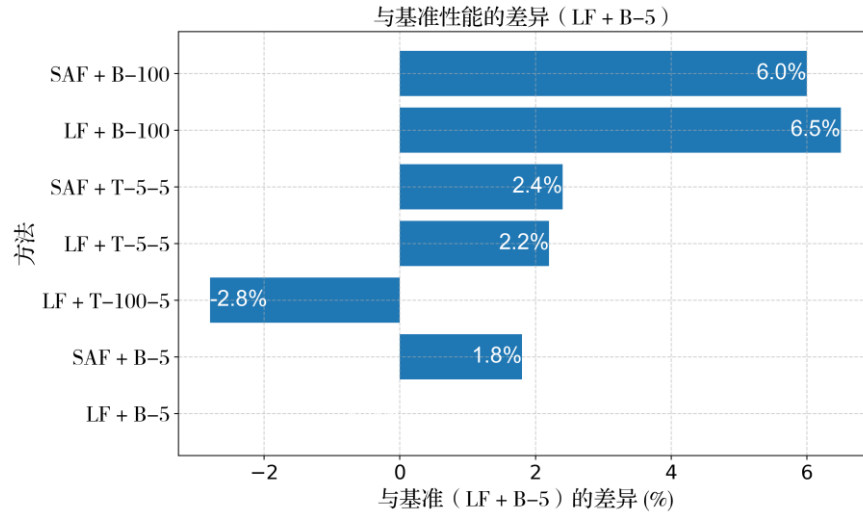
注：B-5 表示的使用基础模型并且训练的样本数量是 5，B-100 表示的使用基础模型并且训练的样本数量是 100，T-5-5 表示的是使用迁移模型并且在源域的训练样本数量是 5，在目标域的训练样本也是 5。

从图 4.10 中可以发现，使用迁移模型会在一定程度上提高基础模型在小样本情况下的准确率，尤其是针对在基础模型上表现略差的提升较为明显，如在 CICIDS2017 数据集上的表现情况。但对于基础模型初始表现就较好的情况来说，提升就较为微弱，仅仅上升 0.1%，这同样反映了模型对不同的数据集有的一定敏感程度。此外，即使迁移模型能够提高模型的表现情况，但相较于足够充足的训练样本来说，即使不进行迁移学习（B-100）的表现情况仍与使用后的结果（T-5-5）有较大的提升空间。但是，相较于 4.5.1 节的检测结果来说，选择合适的特征融合的方法，能有效解决或缓解如下 2 个问题：迁移模型需要在目标域中使用较多的数据集才能超过在基础模型上的表现结果、迁移模型对初始化模型的数据集较为敏感。

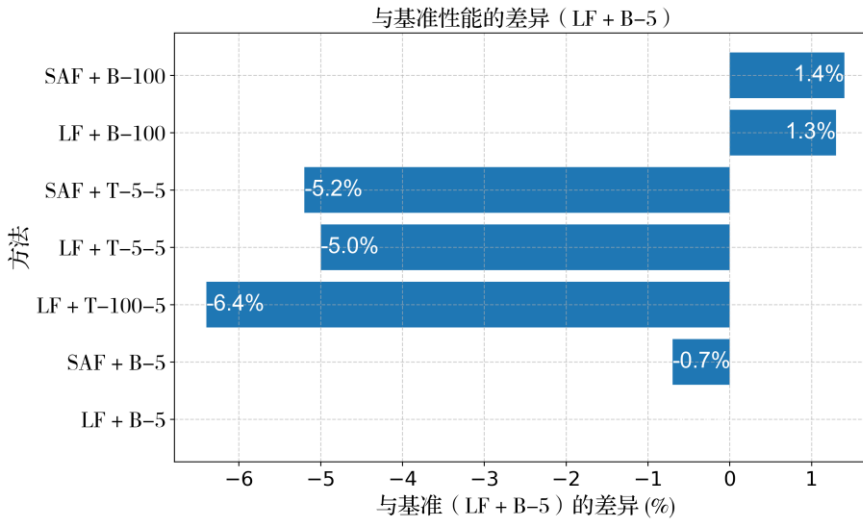
经过 SAF 的调整后，相较于基准（B-5）在准确率上的差异情况如图 4.11 所示，其中差异百分比的计算公式如（4-2）所示，其中的是相较于基准的性能差异百分比， A_i 表示的是当前方法的准确率， $A_{baseline}$ 表示的是基准方法的准确率。

$$\Delta_i = \left(\frac{A_i - A_{baseline}}{A_{baseline}} \right) \times 100\% \quad (4-2)$$

在 CICIDS2017 数据集上，使用 SAF 替换 BF 在 B-5 上涨 2%，此外仅需要 5 个源域和目标域的样本（T-5-5）即可相较于基准上涨 2.6%，同时我们也发现了（T-100-5）的表现情况竟相较于基准降低了 3.1%，这是由于在源域样本数量较少时，模型可能，更依赖于目标域的数据来调整其参数，从而减少了过拟合源域特征的风险，并可能在目标域上获得更好的泛化能力。相反，当源域样本过多时，模型可能会在源域数据上过度拟合，导致准确率下降。这一点从图 4.11（a）和（b）中均能体现。



(a) 在 CICIDS2017 数据集上的表现情况



(b) 在 CICIDS2018 数据集上的表现情况

图 4.11 不同模型和不同样本组合下相较于基准的准确率差异情况

注：图（a）中的基准（LF + B-5）的准确率为 91.0%，在图（b）中的基准（LF + B-5）的准确率为 98.4%

4.5.4 结果总结

本文提出的基础模型能较好地处理网络入侵检测的小样本多分类问题，并且在样本数量有所限制（ $k=5、10$ ）的情况仅仅使用基础模型的准确率分别可达到 91.0% 和 94.8%。此外，还探究了多种特征融合实验对模型性能的影响，通过图 4.9 可以发现，本文所提六种特征融合方法的平均准确率相较于基准模型都有一定幅度的上升，通过图 4.8 可以发现使用当将特征融合的方法改成自注意力融合（SAF）后，会显著提高模型的多种指标的表现情况，同时有效降低波动程度。

最后，通过比较不同数据集、不同融合方法、不同样本数量、不同模型下的指标分析如图 4.10 和 4.11 所示，可以发现，在基础模型中选择合适的特征融合的方法，会提高模型的准确率，能够将准确率从 91.0%提升到 92.8%，此外，若有可以利用的先验数据，则可通过迁移学习进一步提升模型的准确率，分别为 93.4%和 98.5%，涨幅为 0.6%和 0.1%，但是影响最大的还是数据集本身和可供训练的样本数量，在样本数量足够充足（B-100）的情况下，即使不使用迁移模型和特征融合方法，仍然在两个数据集上到达 97.5%和 99.8%，但在源域和目标域可供训练的样本数量都较少的情况下，本研究的意义就显得较为明显。综上所述，可得到以下两个结论：

（1）本文所提的基础模型能较好地处理小样本环境下的多分类问题，仅在样本数量为 5 时，通过复杂的特征融合的方法，可以在两个数据集下的准确率最高可达 98.40%和 92.80%。

（2）本文所提的模型同样适用于样本量足够充足的情况，在样本量足够的情况下分别可到 99.8%和 97.5%，此外，若有可以利用的先验数据，则可通过迁移学习后的模型进一步提升准确率，分别从 92.8%提升到 93.4%、从 98.4%提升到 98.5%，提升了 0.6%和 0.1%。

5 比较与讨论

5.1 与同类工作的对比

为了探究本文中模型的表现情况，我们选择了近期的同类工作进行对比，在数据集选择中，本文所使用的数据包括 csv+pcap，但现有的小样本网络入侵检测的方法使用混合数据较少，大多使用的是 csv 或者 pcap，使用异构数据集的方法还较少，同时这也是本文的创新点，为了更加全面地与同类工作进行比较，本文所选择的同类工作如表 5.1 所示，其中的 FE-MTDM 模型所使用的样本数量是 CICIDS2017 官方数据集的 1%，GDE 模型所使用的样本总数是 140 样本，其余模型的样本数量均是针对每种类型的样本数量而言，并不是针对所选的总样本数。

表 5.1 与同类工作在样本数量、类型、数据集、准确率的对比

模型/方法	样本数	类型	数据集	准确率
continual meta-learning (2022) [53]	10	多分类	CICIDS2017 (pcap)	97.56%
FS-IDS (2022) [54]	5	二分类	CICIDS2017 (pcap)	97.51%
GDE (2023) [55]	140	多分类	CICIDS2018 (pcap)	99.13%
FML (2023) [56]	10	多分类	CICIDS2017 (pcap)	87.27%
Res-Natural GAN (2024) [57]	15	二分类	CICIDS2018 (pcap)	95.75%
MetaMRE (2023) [58]	10	二分类	CICIDS2017 (pcap)	93.30%
MetaMRE (2023) [58]	10	多分类	CICIDS2017 (pcap)	91.80%
FE-MTDM (2023) [23]	1%	多分类	CICIDS2017 (pcap)	99.70%
基础模型	5	多分类	CICIDS2017 (csv+pcap)	92.80%
迁移模型	5	多分类	CICIDS2017 (csv+pcap)	93.40%
基础模型	5	多分类	CICIDS2018 (csv+pcap)	98.40%
迁移模型	5	多分类	CICIDS2018 (csv+pcap)	98.50%

5.2 极小样本情况

在探究样本数量为 5 的实验后，我们尝试探究样本数量为 1 的情况，我们探索了迁移学习和特征融合是否会在极小样本 (K=1) 的情况下的表现情况，实验结果如图 5.1 和表 5.2 所示，可以发现在基础模型 (LF + CICIDS2017 + B-1、LF + CICID2018 + B-1) 的表现结果仅仅能到达 74.1%和 71.3%

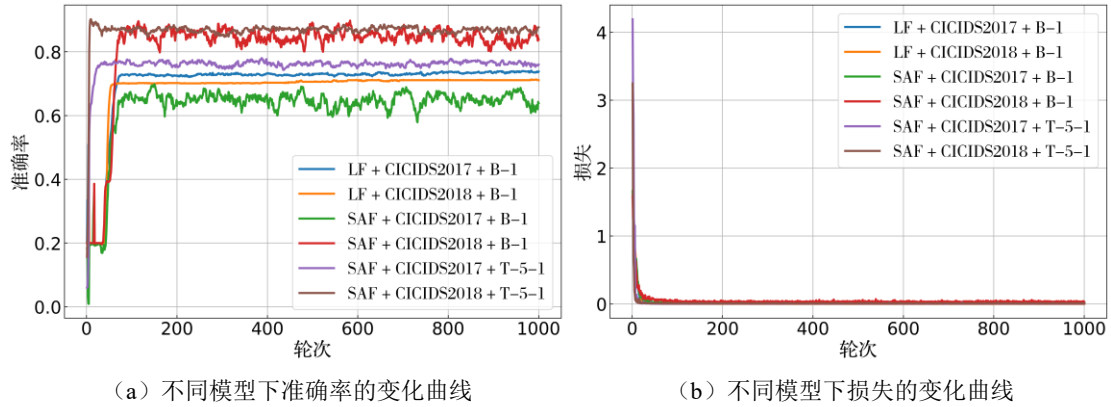


图 5.1 在不同模型、样本、数据集组合下的表现情况

但在尝试引入不同的融合模型后的表现效果出现截然不同的情况，针对 CICIDS2017 数据集来说，下降了 4.3%、但在 CICIDS2018 数据集上却上升了 18.5%，这也证明了本模型对数据集的敏感度较大，这一现象在极小样本情况下被放大，此外可以发现在 (T-5-1)，也就是经过迁移学习后均相较于原始的 (B-1) 有所涨幅，分别上升 8.2%、0.4%。

表 5.2 不同模型、样本、数据集在测试上准确率的表现情况

模型及样本	CICIDS2017	CICIDS2018
SAF + B-1	69.8%	89.8%
SAF + T-5-1	78.0%	90.2%
LF + B-1 (基准)	74.1%	71.3%

5.3 检测系统的消融实验

最后，在构建本模型时，我们利用了异构数据，并专门设计了不同的特征提取模块 (G-Model、S-Model)。为了探究这些内容对模型性能的具体影响，我们依次单独移除模型中不同的特征提取模块，这一过程能帮助我们理解不同模块对实验的贡献度和作用，而且还探讨了这些模块的相互作用如何影响整体实验结果。此外，我们引入了迁移实验的概念，将消融实验中的见解用于优化迁移模型的设计和应用，具体来说我们让经过 G-Model 和 S-Model 的输出后的维度依次变成 0，消去其中一个模型的影响，实验结果如下图所示，其中 S-Model+G-Model 表示既使用了 G-Model 又使用 S-Model，实验结果如图 5.2 所示。

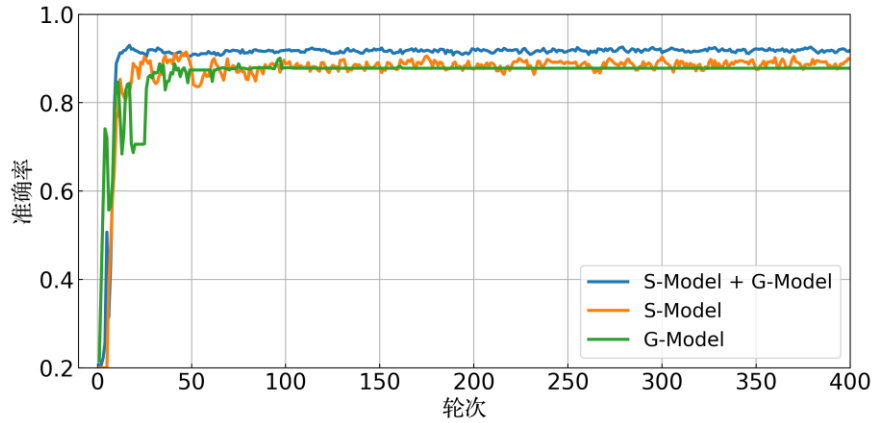


图 5.2 不同特征提取器下的准确率随着训练轮次的表现情况

可以发现这三种模型的总体趋势比较相似，在前 100 个轮次中准确率迅速上升，后续波动均不大，此外为了反映模型的更多细节，我们在测试集上使用更多的评价指标进行分析，实验结果如图 5.3 所示。

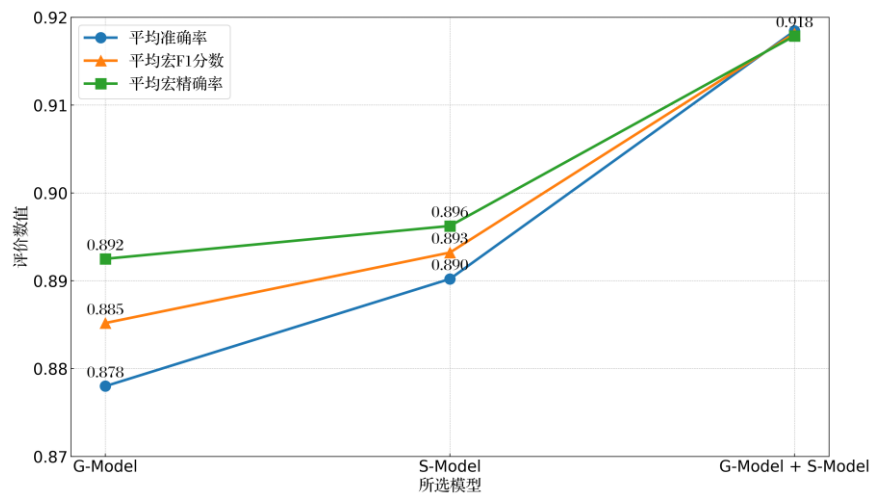


图 5.3 不同模型在多种评价指标下的表现情况

可以发现，在只使用 G-Model 或者 S-Model，模型的准确率、平均宏 F1 分数和平均宏精确率都有一定幅度的下降，对于只使用 S-Model 来说分别下降 2.2%、2.5%、2.8%，对于只使用 G-Model 来说分别下降 2.6%、3.3%、4.0%。

6 总结与展望

6.1 总结

现有的小样本网络入侵检测系统,大多只依赖于单一的数据类型或数据源进行检测,未能充分利用数据,并且针对的是二分类的问题,在这一情况下,只能判断是否是攻击而不能区分具体是哪一种攻击,这也为防御和排查带来了一定的难度,因此本文提出了多源异构小样本检测系统,该系统能够充分利用数据集,并处理多分类问题。

(1) 提出了有针对性的两种特征提取模型, **G-Model** 和 **S-Model** 分别用于处理网络特征图和流量特征集。通过深入分析和处理网络流量的原始数据与结构化特征,使用卷积神经网络和 **Transformer** 的多头自注意力机制设计的特征提取模块,并且注重对流量特征集和网络特征图的个性化处理,对于流量特征图,我们将其按照时间顺序重构堆叠,保证卷积核能够提取到有效的相邻特征。对于网络特征图,我们将其特征拆分,划分为离散内容和连续内容,并分别进行处理。对于上述的模块,都进行多次的层叠以便于模型能够提取更加深层次的特征。此外,我们在 **CICIDS2017** 数据集上进行消融实验,实验结果表明,针对单一的特征提取器,本文所提的方法的准确率分别提升了 2.8%和 4.0%。

(2) 提出了多种特征融合方法,有效提高在小样本情况下的准确率,并提高模型的稳定性。尤其是针对 **SAF** 来说,特征融合技术为网络入侵检测提供了更丰富和有效的特征表示,增强了模型的检测能力。在这一部分将 **G-Model** 的输出向量和 **S-Model** 的输出向量通过特征融合器,再通过 **Transformer** 的注意力机制,融合多种异构信息。最后在两个网络入侵的基准数据集上涨幅 2.86%,这一点在 $K=1$ 的情况下更为突出,涨幅可达 8.2%,与传统的基于单一数据源或未采用特征融合技术的方法相比,本研究提出的方法在多项性能指标上均表现出显著优势,证明了其在小样本网络入侵检测领域的应用潜力。

(3) 提出了基于迁移学习的迁移模型,充分利用多源数据,仅需要较少的源域样本即可进一步提高准确率。在这一部分,我们先在已有的数据集上进行模型的预训练,再使用另一数据集进行迁移学习,分别提升 0.6%和 0.1%,在两种

基准数据集上分别可达 93.40%和 98.50%。

6.2 展望

本文所设计的检测系统,通过各种实验展示出在小样本情况下不错的分类效果,未来的工作将集中在以下几个方面:

(1) 高效的数据预处理技术: 将研究并采用更先进的数据预处理和特征提取技术。通过自动化和优化数据处理流程,可以显著减少数据准备阶段的工作量 and 时间, 提高处理异构数据集时的效率。

(2) 改善数据集敏感度: 为了解决方法对数据集敏感度高的问题, 我们计划采用数据增强技术, 如过采样、欠采样和生成对抗网络, 这些技术可以生成更多样化的训练样本, 从而提高模型在不同数据集上的鲁棒性和准确性。

致谢

在本科学习生涯的尾声，撰写这篇论文的过程中获得了诸多宝贵的支持和帮助。在此，我诚挚地向所有在这一过程中给予我无限支持和帮助的人表达我的深切感激。

首先，我衷心感谢我的导师许聪源老师。他在学术上对我进行了精心指导，在整个实验流程和论文结构上提供了巨大的帮助，使我受益匪浅。许老师的严谨的学术态度和博学深厚的专业知识极大地激发了我对科研的热情。

虽然疫情几乎贯穿了我整个大学生涯，带来了无数挑战，但这也是一段充满真挚情感的时光。感激信息科学与工程学院的所有教师和同学们，他们在我学习与生活的每一步都给予了我巨大的支持。此外，我也感谢学校为我们提供的卓越学习环境，包括丰富的科研资源和图书馆设施，这些都极大地丰富和深化了我的学术追求。

我还要特别感谢我的家人，尤其是我的父母和姐姐。他们始终如一地给予我肯定和支持，坚定不移地支持我的每一个决定，他们的无条件爱是我前进的动力。

最后，再次感谢所有在论文研究和写作过程中给予我灵感和帮助的朋友们，感谢在这个充满挑战的时代里，我有幸学习和成长，感谢命运让我与这么多杰出的人相遇、前行。

知不足而后进，望山远而力行。

参考文献

- [1] Denning D E. An intrusion-detection model[J]. IEEE Transactions on software engineering, 1987, 13(2): 222-232.
- [2] Roesch M. Snort: Lightweight intrusion detection for networks[C]//Proc of the 13th USENIX Conference on System Administration (LISA). 1999, 99(1): 229-238.
- [3] Modi C, Patel D, Borisaniya B, et al. A survey of intrusion detection techniques in Cloud[J]. Journal of Network and Computer Applications, 2013, 36(1): 42-57.
- [4] Buczak A L, Guven E. A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection[J]. IEEE Communications Surveys & Tutorials, 2016, 18(2): 1153-1176.
- [5] Roy S, Li J, Choi B J, et al. A lightweight supervised intrusion detection mechanism for IoT networks[J]. Future Generation Computer Systems, 2022, 127: 276-285.
- [6] 罗智勇, 杨旭, 刘嘉辉, 等. 基于贝叶斯攻击图的网络入侵意图分析模型[J]. 通信学报, 2020, 41(09): 160-169.
- [7] Thakkar A, Kikani N, Geddam R. Fusion of linear and non-linear dimensionality reduction techniques for feature reduction in LSTM-based Intrusion Detection System[J]. Applied Soft Computing, 2024, 154: 111378.
- [8] Alzubi J A, Alzubi O A, Qiqieh I, et al. A Blended Deep Learning Intrusion Detection Framework for Consumable Edge-Centric IoMT Industry[J]. IEEE Transactions on Consumer Electronics, 2024, 70(1): 2049-2057.
- [9] Louk M H L, Tama B A. Dual-IDS: A bagging-based gradient boosting decision tree model for network anomaly intrusion detection system[J]. Expert Systems with Applications, 2023, 213: 119030.
- [10] Sharifian Z, Barekatin B, Quintana A A, et al. Sin-Cos-bIAVOA: A new feature selection method based on improved African vulture optimization algorithm and a novel transfer function to DDoS attack detection[J]. Expert Systems with Applications, 2023, 228: 120404.
- [11] Zhiqiang L, Mohiuddin G, Jiangbin Z, et al. Intrusion detection in wireless sensor network using enhanced empirical based component analysis[J]. Future Generation Computer Systems,

- 2022, 135: 181-193.
- [12] Verkerken M, D’hooge L, Sudyana D, et al. A Novel Multi-Stage Approach for Hierarchical Intrusion Detection[J]. IEEE Transactions on Network and Service Management, 2023, 20(3): 3915–3929
- [13] Ciric V, Milosevic M, Sokolovic D, et al. Modular deep learning-based network intrusion detection architecture for real-world cyber-attack simulation[J]. Simulation Modelling Practice and Theory, 2024, 133: 102916.
- [14] Chowdhury R, Sen S, Goswami A, et al. An implementation of bi-phase network intrusion detection system by using real-time traffic analysis[J]. Expert Systems with Applications, 2023, 224: 119831.
- [15] Basati A, Faghih M M. PDAE: Efficient network intrusion detection in IoT using parallel deep auto-encoders[J]. Information Sciences, 2022, 598: 57-74.
- [16] Chawla N V, Bowyer K W, Hall L O, et al. SMOTE: Synthetic Minority Over-sampling Technique[J]. Journal of Artificial Intelligence Research, 2002, 16: 321-357.
- [17] Haibo He, Garcia E A. Learning from Imbalanced Data[J]. IEEE Transactions on Knowledge and Data Engineering, 2009, 21(9): 1263-1284.
- [18] Milosevic M S, Ciric V M. Extreme minority class detection in imbalanced data for network intrusion[J]. Computers & Security, 2022, 123: 102940.
- [19] Chapaneri R, Shah S. Enhanced detection of imbalanced malicious network traffic with regularized Generative Adversarial Networks[J]. Journal of Network and Computer Applications, 2022, 202: 103368.
- [20] Attique D, Hao W, Ping W, et al. Explainable and Data-Efficient Deep Learning for Enhanced Attack Detection in IIoT Ecosystem[J]. IEEE Internet of Things Journal, Early Access, DOI: 10.1109/JIOT.2024.3384374.
- [21] Hu X, Gao W, Cheng G, et al. Toward Early and Accurate Network Intrusion Detection Using Graph Embedding[J]. IEEE Transactions on Information Forensics and Security, 2023, 18: 5817-5831.
- [22] Behiry M H, Aly M. Cyberattack detection in wireless sensor networks using a hybrid feature reduction technique with AI and machine learning methods[J]. Journal of Big Data, 2024, 11(1): 16.

- [23] Wei N, Yin L, Zhou X, et al. A feature enhancement-based model for the malicious traffic detection with small-scale imbalanced dataset[J]. *Information Sciences*, 2023, 647: 119512.
- [24] 刘奇旭, 王君楠, 尹捷, 等. 对抗机器学习在网络入侵检测领域的应用[J]. *通信学报*, 2021, 42(11): 1-12.
- [25] Thakkar A, Lohiya R. Attack Classification of Imbalanced Intrusion Data for IoT Network Using Ensemble-Learning-Based Deep Neural Network[J]. *IEEE Internet of Things Journal*, 2023, 10(13): 11888-11895.
- [26] Batchu R K, Seetha H. An integrated approach explaining the detection of distributed denial of service attacks[J]. *Computer Networks*, 2022, 216: 109269.
- [27] Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate[C]//*Proc of the 3rd Int Conf on Learning Representations*, 2015.
- [28] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C]//*Proceedings of the 31st International Conference on Neural Information Processing Systems*. New York: ACM Press, 2017: 6000-6010.
- [29] Parmar N, Vaswani A, Uszkoreit J, et al. Image transformer[C]//*Proceedings of the 35th International conference on machine learning*. PMLR, 2018, 80: 4055-4064.
- [30] Han X, Cui S, Liu S, et al. Network intrusion detection based on n-gram frequency and time-aware transformer[J]. *Computers & Security*, 2023, 128: 103171.
- [31] Ding Q, Li J. AnoGLA: An efficient scheme to improve network anomaly detection[J]. *Journal of Information Security and Applications*, 2022, 66: 103149.
- [32] Rendón-Segador F J, Álvarez-García J A, Varela-Vaca A J. Paying attention to cyber-attacks: A multi-layer perceptron with self-attention mechanism[J]. *Computers & Security*, 2023, 132: 103318.
- [33] Reka R, Karthick R, Saravana Ram R, et al. Multi head self-attention gated graph convolutional network based multi-attack intrusion detection in MANET[J]. *Computers & Security*, 2024, 136: 103526.
- [34] Atighetchi M, Pal P, Webber F, et al. Adaptive use of network-centric mechanisms in cyber-defense[C]//*Sixth IEEE International Symposium on Object-Oriented Real-Time Distributed Computing*, 2003. IEEE, 2003: 183-192.
- [35] Zhang Y, Lee W. Intrusion detection in wireless ad-hoc networks[C]//*Proceedings of the 6th*

- annual international conference on Mobile computing and networking. 2000: 275-283.
- [36] Hwang R H, Lee C L, Lin Y D, et al. Host-based intrusion detection with multi-datasource and deep learning[J]. Journal of Information Security and Applications, 2023, 78: 103625.
- [37] Thakkar A, Lohiya R. Fusion of statistical importance for feature selection in Deep Neural Network-based Intrusion Detection System[J]. Information Fusion, 2023, 90: 353-363.
- [38] 刘涛涛, 付钰, 王坤, 等. 基于 VAE-CWGAN 和特征统计重要性融合的网络入侵检测方法[J]. 通信学报, 2024, 45(02): 54-67.
- [39] Fu J Juan, Zhang X lan. Gradient importance enhancement based feature fusion intrusion detection technique[J]. Computer Networks, 2022, 214: 109180.
- [40] Jiang H, Lin J, Kang H. FGMD: A robust detector against adversarial attacks in the IoT network[J]. Future Generation Computer Systems, 2022, 132: 194-210.
- [41] 张兴兰, 尹晟霖. 可变融合的随机注意力胶囊网络入侵检测模型[J]. 通信学报, 2020, 41(11): 160-168.
- [42] Zang X, Gong J, Zhang X, et al. Attack scenario reconstruction via fusing heterogeneous threat intelligence[J]. Computers & Security, 2023, 133: 103420.
- [43] Pan S J, Yang Q. A Survey on Transfer Learning[J]. IEEE Transactions on Knowledge and Data Engineering, 2010, 22(10): 1345-1359.
- [44] Ullah F, Ullah S, Srivastava G, et al. IDS-INT: Intrusion detection system using transformer-based transfer learning for imbalanced network traffic[J]. Digital Communications and Networks, 2024, 10(1): 190-204.
- [45] Bierbrauer D A, De Lucia M J, Reddy K, et al. Transfer learning for raw network traffic detection[J]. Expert Systems with Applications, 2023, 211: 118641.
- [46] Latif S, Boulila W, Koubaa A, et al. DTL-IDS: An optimized Intrusion Detection Framework using Deep Transfer Learning and Genetic Algorithm[J]. Journal of Network and Computer Applications, 2024, 221: 103784.
- [47] Zhang J, Luo C, Carpenter M, et al. Federated Learning for Distributed IIoT Intrusion Detection Using Transfer Approaches[J]. IEEE Transactions on Industrial Informatics, 2023, 19(7): 8159-8169.
- [48] Sharafaldin I, Habibi Lashkari A, 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. Funchal, Madeira, Portugal: SCITEPRESS - Science and Technology Publications, 2018: 108-116.
- [49] Sperotto A, Schaffrath G, Sadre R, et al. An Overview of IP Flow-Based Intrusion Detection[J]. IEEE Communications Surveys & Tutorials, 2010, 12(3): 343-356.
- [50] Leevy J L, Khoshgoftaar T M. A survey and analysis of intrusion detection models based on CSE-CIC-IDS2018 Big Data[J]. Journal of Big Data, 2020, 7(1): 104.
- [51] He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA: IEEE, 2016: 770-778.
- [52] Zhuang F, Qi Z, Duan K, et al. A Comprehensive Survey on Transfer Learning[J]. Proceedings of the IEEE, 2021, 109(1): 43-76.
- [53] Xu H, Wang Y. A Continual Few-shot Learning Method via Meta-learning for Intrusion Detection[C]//2022 IEEE 4th International Conference on Civil Aviation Safety and Information Technology (ICCASIT). Dali, China: IEEE, 2022: 1188-1194.
- [54] Xu C, Shen J, Du X. A Method of Few-Shot Network Intrusion Detection Based on Meta-Learning Framework[J]. IEEE Transactions on Information Forensics and Security, 2020, 15: 3540-3552.
- [55] Yan Y, Yang Y, Shen F, et al. GDE model: A variable intrusion detection model for few-shot attack[J]. Journal of King Saud University - Computer and Information Sciences, 2023, 35(10): 101796.
- [56] Hu Y, Wu J, Li G, et al. Privacy-Preserving Few-Shot Traffic Detection Against Advanced Persistent Threats via Federated Meta Learning[J]. IEEE Transactions on Network Science and Engineering, 2024, 11(3): 2549-2560.
- [57] Yan Y, Yang Y, Shen F, et al. Meta learning-based few-shot intrusion detection for 5G-enabled industrial internet[J]. Complex & Intelligent Systems, Early Access, DOI: 10.1007/s40747-024-01388-1
- [58] Yang C, Xiong G, Zhang Q, et al. Few-shot encrypted traffic classification via multi-task representation enhanced meta-learning[J]. Computer Networks, 2023, 228: 109731.