基于元学习和特征增强的网络入侵检测模型

蒋章涛[1], 李 欣[1,2], 薛 迪[1], 彭奕杰[1]

[1]中国人民公安大学 信息网络安全学院 北京 中国 100045
[2]安全防范技术与风险评估公安部重点实验室 北京 中国 100026

摘要 目前, 为有效检测网络上各种恶意攻击, 研究人员开发了大量的入侵检测系统。但面对零日攻击等新型攻击, 现有的检测系统难以在短时间内获得足够的攻击样本并进行充分训练, 导致识别准确率和网络系统安全性大幅下降。为解决以上问题, 本文提出一种基于元学习和特征增强的网络入侵检测模型 MCIDS(Meta-CLIP based Intrusion Detection System), 旨在模仿安全专家利用元知识进行学习、检测未知入侵流量的能力, 仅使用少量样本解决新的入侵检测任务。模型首先将网络流量数据转换为灰度图传入 CLIP 图像编码器进行特征增强, 以提升网络流量特征在高维向量空间中的表示能力。随后将其生成的高维向量转换二维数据传入任务池。在此基础上, 模型基学习器在任务支持集上更新局部参数, 元学习器在任务查询集上更新全局参数。在每个训练周期中, MCIDS 通过检验全局参数的线性组合与其在子空间投影的偏离程度, 决定是否扩大参数数量, 以保证最优初始化参数集合可覆盖所有检测任务。最后, 模型通过从少量新型攻击数据中获得的梯度步骤进行元更新, 从而在该任务上产生良好的检测效果。所构建的 MCIDS 在 UNSW-NB15 和 CSE-CIC-IDS2018 数据集上均取得了良好的检测效果, 对比 MAML+CNN、Meta-SGD 和无监督 Kitsune 模型, 检测准确率均值分别提高了 2.93%、4.74%和 16.07%。

关键词 网络入侵检测; 元学习; 特征增强
中图法分类号 TN915.08 DOI 号 10.19363/J.cnki.cn10-1380/tn.2026.01.17

Network Intrusion Detection Model based on Meta-Learning and Feature Enhancement

JIANG Zhangtao[1], LI Xin [1,2], XUE Di[1], PENG Yijie[1]

[1] Information Network Security Academy, People's Public Security University of China, Beijing 100045, China
[2] China Security Prevention Technology and Risk Assessment Key Laboratory of Ministry of Public Security, Beijing 100026, China

Abstract In response to the pressing need for effective detection of various malicious attacks on networks, a plethora of intrusion detection systems have been developed by researchers. However, in the face of new types of attacks such as zero-day attacks, it is difficult for existing detection systems to obtain sufficient attack samples and train them adequately in a short period of time, which leads to a significant decrease in recognition accuracy and network system security. To solve the above problems, this paper proposes a meta-learning and feature-enhanced network intrusion detection model MCIDS (Meta-CLIP based Intrusion Detection System), which is designed to mimic the ability of security experts to use meta-knowledge for learning and detecting unknown intrusion traffic, and to use only a small number of samples to solve the new intrusion detection task. The model first converts network traffic data into grey-scale maps to be passed into a CLIP image encoder for feature enhancement to improve the representation of network traffic features in a high-dimensional vector space. Subsequently, its generated high-dimensional vectors are converted into 2D data to be passed into the task pool. On this basis, the model base learner updates local parameters on the task support set and the meta-learner updates global parameters on the task query set. In each training cycle, MCIDS decides whether to expand the number of parameters by examining how much the linear combination of global parameters deviates from its projection in the subspace to ensure that the optimal initialised parameter set can cover all detection tasks. Finally, the model is meta-updated with gradient steps obtained from a small amount of novel attack data, which produces good detection results on this task. The constructed MCIDS achieves good detection results on both UNSW-NB15 and CSE-CIC-IDS2018 datasets, and the mean detection accuracy is improved by 2.93%, 4.74%, and 16.07% compared to MAML+CNN, Meta-SGD, and Unsupervised Kitsune models, respectively.

Key words network intrusion detection; meta-learning; feature enhancement

通讯作者: 李欣, 教授, Email: lixin@ppsuc.edu.cn。
本课题得到国家重点研发计划课题跨域多源视频监控网络安全体系研究(No. 2022YFC3301101)资助。
收稿日期: 2024-07-08; 修改日期: 2024-09-20; 定稿日期: 2026-01-07

1 引言

随着互联网技术在当代社会中的广泛应用, 大量信息网络系统通过网络进行互联互通。但与此同时, 针对各种网络的攻击技术也在不断迭代更新, 任何网络中的攻击或入侵都可能会造成严重影响, 网络空间的安全问题受到了学界的高度关注。

目前, 为提高网络系统安全性、有效检测网络上各种攻击者发起的恶意攻击, 大量的入侵检测系统(Intrusion Detection System, IDS)被研究开发, 成为一种关键的网络安全设备。如图 1 所示, IDS 的出现使得恶意攻击者通过已知的攻击类型发起入侵行为的攻击手段失去了作用。从传统机器学习的角度看, IDS 可以定义为一个监督学习系统, 旨在通过对网络数据的分析来实现准确的行为分类。但传统机器学习模型对于数据量具有极大的依赖性, 存在较高的训练功耗和训练成本[1]。此外, 基于网络入侵数据进行监督学习的 IDS 系统在面对新型网络攻击, 如零日攻击, 检测系统难以在短时间内获得足够的攻击样本并进行充分训练, 导致识别准确率大幅下降。因此, 恶意攻击者针对传统 IDS 的缺陷, 通过零日攻击等未知攻击手段可以成功实现对网络系统的入侵, 其威胁性大大增加。

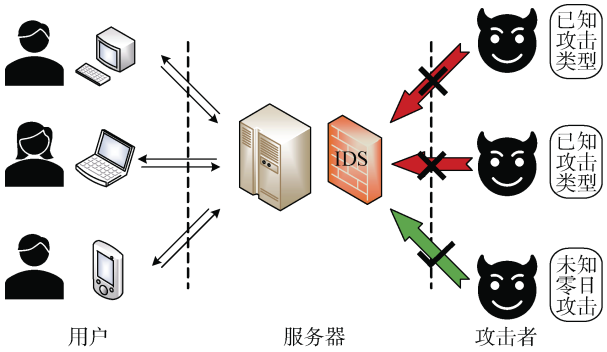

图 1 实施入侵行为的恶意攻击者

Figure 1 Malicious attacker committing an intrusion

因此, 基于小样本[2]和非监督学习的网络入侵检测成为网络空间安全领域的重要研究方向之一。近年来, 小样本学习领域发展迅速, 一些算法利用元学习的思想来解决小样本学习问题。元学习旨在模仿人类在没有先验知识情况下利用元知识进行学习、推理未知任务的能力。通过在多种学习任务上训练, 使元学习模型能够仅使用少量的训练样本就可解决新的学习任务, 即模型学会学习。此类学习方法在各个网络空间安全领域已取得了众多研究进展。在窃取隐私领域, Ni 等[3]报道了一种新的非接触式无线充电侧信道, 利用这种信道可以从正在充电的智能手机中推断出用户的隐私, 其提出采用元学习方法, 以快速调整预先训练好的模型, 使其能够部署到新的攻击场景和环境中, 取得了不错的攻击效果。在无线感知领域, Gong 等[4]提出一种自适应移动传感系统 MetaSense, 其采用元学习技术离线训练多个生成的小样本任务, 以适应目标用户状况。利用真实世界中运动和音频传感器数据进行的评估表明, MetaSense 不仅在准确率方面比最先进的迁移学习方法高出 18%, 而且目标用户所需的适应时间也大大减少。在网络入侵领域, Sirinam 等[5]提出了一种新的网络指纹攻击, 其使用三重网络进行 N 次小样本学习。此外, 该攻击只需要 5 个示例就能识别一个网站, 因此在收集和训练完整数据集不切实际的各种情况下, 都能发挥巨大作用, 可以使得计算资源相对较少的攻击者也能以相当有效的性能执行网络指纹攻击。Xu 等[6]首次提出应用于网络入侵检测领域的 FC-Net 元学习模型, FC-Net 直接从原始流量数据中学习用于网络流量分类的先验知识。在获得足够的先验知识后, 通过少量的样本检测出新类型的流量。但 FC-Net 仅可实现对流量的二分类任务, 难以有效应对存在多种流量种类的多分类任务。Finn 等[7]提出的元学习算法 MAML 可以在小样本的情况下解决分类或回归等任务。Lu 等[8]提出了一种基于 MAML 和卷积神经网络的物联网入侵检测模型, 模型直接将流量特征转换为灰度图后进行训练和推理, 但仍然存在特征利用不充分的问题。为解决以上问题, 本文提出了基于元学习和特征增强的网络入侵检测模型。首先, 模型通过 CLIP 对流量特征灰度图进行特征再提取以增强流量数据特征; 通过改进 MAML 的模型对流量数据特征进行元学习训练, 最终构建出网络入侵检测模型 MCIDS。本文工作的主要贡献如下。

(1) 构建了基于预训练 CLIP 的流量特征增强模块, 提高了对转换为灰度图的网络流量样本特征利用效率, 扩大了不同种类流量数据在高维度特征向量空间中的差异, 提高了模型分类效果。

(2) 构建了基于 XB-MAML 算法的入侵检测元学习模块, 通过自适应构建多个初始化参数作为参数空间中的基向量, 每个初始化基向量根据任务分布进行线性组合, 实现覆盖更多的入侵检测任务分布, 增强了模型的泛化能力。

(3) 融合了特征增强和元学习模块, 构建出小样本网络入侵检测模型 MCIDS, 解决了检测系统难以根据少量攻击样本进行充分训练以面对新型检测任

务的问题, 提高了网络入侵流量检测准确率。

2 相关工作

2.1 网络入侵检测技术

随着互联网的快速发展, 网络攻击行为的数量与种类日益增长, 网络空间安全变得愈发关键, 网络入侵检测技术已经成为网络攻击识别和防御的重要手段。根据检测技术不同, 入侵检测系统划分为基于统计的入侵检测、基于机器学习的入侵检测和基于深度学习的入侵检测等。

2.1.1 基于统计的入侵检测

基于统计的入侵检测系统通过利用统计方法来监测网络中的潜在异常行为。这类方法通过分析网络流量和用户行为的历史数据, 假设数据服从某种概率分布, 建立正常活动的统计模型, 任何显著偏离这些模型的行为都可能被识别为入侵。Eskin[9]基于异常流量数据远少于正常流量数据的假设, 提出了一种基于混合模型的检测方法, 能够从噪声数据中有效识别出异常流量。Thottan 和 Ji[10]提出一种使用统计信号处理技术的网络异常检测模型。连晓伟等[11]通过分析应用层的流量特征, 提取协议功能码序列作为载荷特征, 并结合传统的流量统计特征对流量进行识别。然而, 基于统计的检测方法达到较高的准确性需要对人工对数据特征进行统计分析和描述, 并且可能存在较高的误报率。

2.1.2 基于机器学习的入侵检测

基于机器学习的入侵检测系统旨在将机器学习算法应用于识别网络中的非典型行为和潜在威胁。系统在已标注的网络流量数据上进行训练, 得到一个用于网络流量分类的机器学习模型, 例如贝叶斯模型、决策树模型、逻辑回归模型等。徐鹏和林森[12]为解决朴素贝叶斯方法过分依赖于样本在样本空间的分布、具有潜在的不稳定性的问题, 提出使用 C4.5 决策树的方法来处理流量分类问题。Wang 等[13]提出了一种整合特征选择与分类组合的稀疏逻辑回归入侵检测系统, 其在 KDD-cup 1999 数据集[14]上的准确率达到了 97.86%。但基于机器学习的入侵检测系统存在训练数据的选取评估和类别标注成本高和模型迁移学习能力差等缺点。无监督算法应用于网络流量检测可解决数据标注问题。Mirsky 等[15]使用自动编码器神经网络来识别流量模式, 可实现无监督在线检测网络攻击。Fu 等[16]通过构建小型内存图来分析图的连通性、稀疏性和统计特征, 检测异常交互模式, 构建基于无监督机器学习的恶意流量检测系统。但数据特征选择对无监督模型的分类效果影响较大。

2.1.3 基于深度学习的入侵检测

基于深度学习的入侵检测是通过深层神经网络实现网络流量样本的多层特征学习技术。基于深度学习的入侵检测无需人工设计流量样本特征, 采用端到端的形式直接处理原始数据并输出高层特征表示。大大节省了安全专家人工进行特征提取的成本, 提高了检测效率和效果。Gao 等[17]提出一种基于深度信念网络的入侵检测模型, 使用无监督贪婪算法进行预训练以帮助网络捕捉样本的高阶特征。Xiao 等[18]提出一种基于卷积神经网络(Convolutional Neural Network, CNN)的网络入侵检测模型 CNN-IDS, 通过特征降维移除冗余特征, 应用 CNN 自动提取有效信息进行入侵检测。Ma 等[19]提出一个基于生成对抗网络和双向长短期记忆网络的网络入侵检测方法, 模型可为少数类攻击生成新样本, 在 NSL-KDD 数据集[20]和 UNSW-NB15 数据集[21]上有效提高了检测性能和准确率。以上基于深度学习的入侵检测模型虽然取得了不错的效果, 但很大程度上依赖于数据规模, 在实践中具有明显的局限性。

2.2 元学习算法

元学习可主要分为基于度量和基于优化的两种方法。基于度量的方法通过学习数据的特征向量表示, 将新类别的种类归属问题转换为向量间度量问题。Vinyals 等[22]提出了基于记忆和注意力机制的匹配网络模型对样本嵌入向量的距离进行度量和比较。 Oreshkin 等[23]提出任务依赖的距离度量方法, 根据不同任务进行数据特征提取, 自适应学习距离缩放系数, 从而提高模型泛化能力。基于优化的方法源自 Finn 等[7]提出的模型无关的元学习算法(model-agnostic meta-learning, MAML)。其训练目标不是为数据样本拟合一个最优模型参数, 而是通过不同任务的数据来获得元知识, 找到最优初始化参数 θ, 从而使得模型在遇到新任务时, 能够借助已学习的元知识进行参数的迅速收敛。由于 MAML 出色的效果, 目前现有工作大多采用基于 MAML 的元学习算法。Antoniou 等[24]针对 MAML 稳定性不足、计算量大和网络结构敏感的问题, 提出了在训练过程中更稳定、收敛速度更快且准确率更高的模型 MAML++。Meta-SGD[25]是一种基于随机梯度下降的元学习算法, 在一个元学习过程中, 模型不仅能学习学习器的初始化, 还能学习其更新方向和学习率。为实现利用多个初始化来扩大任务覆盖范围的目的, Jiang 等[26]提出的 MUSML 引入了子空间学习, 并行地将初始化微调到给定任务, 并用加权损失进行元

更新。Lee 和 Yoon[27]针对 MUSML 需要预定义初始化数量的问题, 提出 XB-MAML 算法。算法在多个训练任务上得到若干初始化参数, 每个初始化参数作为参数空间中的基向量。在流量检测任务阶段, 每个初始化参数根据任务分布进行线性组合, 若当前基向量无法覆盖任务分布, 模型将自适应增加基向量, 扩充参数空间的秩, 形成新的初始化参数, 其在多个小样本数据集上取得了显著的效果。

2.3 网络流量特征增强技术

目前基于网络流量的入侵检测系统取得了良好的检测效果, 但仍然存在分类方法高度依赖有效特征组合的问题。增强网络流量特征在高维向量空间的表示是提高入侵检测系统效果的重要方法。Wei 等[28]针对小规模不平衡数据集特征差异小的问题, 提出了基于特征增强的 FE-MTDM 入侵检测模型。其根据高斯分布和 K-means 算法对原始流量特征进行聚类, 得到聚类特征, 随后对原始特征和聚类特征进行共同训练, 从而检测网络流量。当前大型视觉语言模型已经展现出卓越性能, 已有大量研究基于预训练的跨模态大模型增强数据特征来提升模型整体效果。CLIP[29]是 2021 年由 Open AI 提出的在大规模自然语言监督下训练的图像分类器。其核心理念是图文对比学习, 通过 4 亿条图像文本对数据的预训练, 实现在高维特征空间中有效捕捉图像与文本之间的关联、具有高效编码图像和文本的能力。CLIP 针对图片的特征提取能力日益提高, 现已有研究将其作为视觉编码器来提高模型高维度特征表示的能力。Shen 等[30]通过实验证明使用 CLIP 作为图像编码器可以带来更好的泛化性能。Li 等[31]为解决良性和攻击特征之间的纠缠问题, 提出了基于 CLIP 特征增强的入侵检测系统, 通过跨模态特征理解能力优化和增强恶意后门样本特征表达的方法, 分别对良性样本和攻击样本特征进行处理, 整体提升了模型效果, 提高了系统检测准确率。众多实验证明, 基于预训练的 CLIP 图像编码器对于提取图像高维特征的显著优势, 可针对网络流量等数据进行特征增强, 但以上基于大型预训练视觉语言模型的方法忽略了在少样本情况下的入侵检测应用效果。

3 基于元学习和特征增强的入侵检测

网络环境的复杂多变, 传统入侵检测系统面临着快速适应新型威胁的挑战。元学习, 作为一种可帮助模型在小样本情况下快速适应新任务的学习范式, 为入侵检测提供了新的解决方案。基于元学习的入侵检测方法通过训练一个元模型来指导快速学习任务, 使系统在面对新的或少见的攻击类型时, 能够迅速调整并有效识别。目前, 元学习在网络安全流量检测领域已经取得了一些研究成果[6,8], 本文在现有研究的基础上, 针对模型泛化能力弱和特征利用不充分的问题做出了相应改进。

3.1 利用元学习方法实现入侵检测

基于元学习的入侵检测技术旨在通过训练得到一个良好的模型初始化参数, 使模型在面临新的入侵流量种类时, 可仅用少量训练数据得到的梯度来更新模型参数从而达到更好的泛化性能。

现有的网络流量入侵检测系统通常采用将流量数据转换为灰度图或 RGB 三通道图像的方法, 并利用卷积神经网络(Convolutional Neural Networks, CNN)来提取图像特征, 以实现流量的分类。鉴于 CNN 在图像分类领域展现出的卓越性能, 本文将网络流量样本数据转化为二维灰度图像数据并选择 CNN 作为模型的视觉提取模块, 对网络流量图像样本进行处理。通过网络流量特征视觉化的转换, CNN 能够有效地捕捉到这些图像中潜在的模式和异常行为。以包含 15 种流量类型的 CSE-CIC-IDS2018 数据集为例, 图 2 展示了视觉化处理后的该网络流量数据集部分样本。

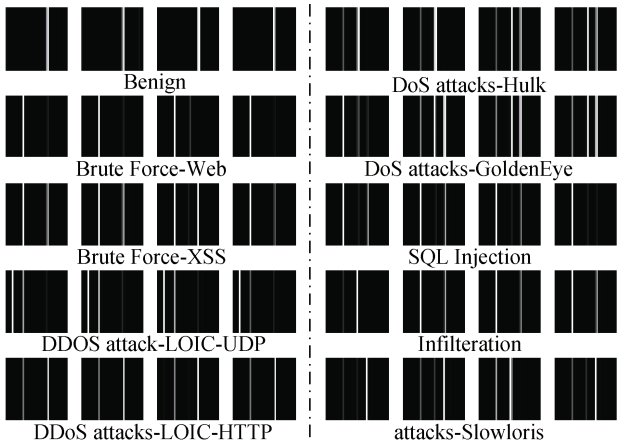

图 2 CSE-CIC-IDS2018 网络流量部分样本

Figure 2 Sample of CSE-CIC-IDS2018 network traffic

本文基于 MAML 算法构造入侵检测模型, 训练数据以任务为单位, 具体分为训练任务、验证任务和测试任务。其中, 每个任务包含支持集和查询集。在基于元学习的入侵检测模型训练过程中, 网络流量数据可以按照不同流量种类划分为不同任务, 并在每个任务中根据实践要求设置不同的支持集和查询集比例。模型的训练框架分为内外两层循环, 内层循环又称为基学习器, 外层循环又称为元学习器。基学习器用于学习每个训练任务中支持集上的数据特征,

其通过计算在每个任务 $\mathcal{T}_i$ 上的训练误差 $\mathcal{L}_{\mathcal{T}_i}(f_\theta)$，按照式(1)进行梯度下降优化更新局部参数 θ_i' 。

$$\theta_i' = \theta - \alpha \nabla_\theta \mathcal{L}_{\mathcal{T}_i}(f_\theta) \tag{1}$$

其中，θ 为全局参数，α 为基学习器的学习率，$\nabla_\theta \mathcal{L}_{\mathcal{T}_i}(f_\theta)$ 为根据训练误差 $\mathcal{L}_{\mathcal{T}_i}(f_\theta)$ 计算得到的梯度项。

元学习器利用基学习器的局部参数 θ_i' 在不同任务 $\mathcal{T}_i$ 查询集上得到的损失函数进行梯度下降优化全局参数 θ，以寻找全局最优初始化模型参数，实现更好的泛化能力和适应性，具体计算如式(2)所示。

$$\theta = \theta - \beta \nabla_\theta \sum_{\mathcal{T}_i \sim p(\mathcal{T})} \mathcal{L}_{\mathcal{T}_i}(f_{\theta_i'}) \tag{2}$$

其中，β 为元学习器的学习率，$p(\mathcal{T})$ 为整体任务分布。

在不同网络流量数据集中，数据的分布差异较大，仅依靠单个模型参数难以实现大范围任务的覆盖。为实现元学习更好的泛化能力，一些工作[26-27,32]开始尝试使用多个初始化参数来覆盖更多的任务分布。为提升网络流量入侵检测系统识别未知流量的能力，本文在元学习的基础上引入了 XB-MAML 算法，将初始化参数转换为自适应可扩展的基参数的线性组合，以实现模型泛化能力的增强。初始化参数可定义为集合 $\Theta = \{\theta_1, \theta_2, ..., \theta_N\}$，其中 N 表示初始化参数集合中基参数的数量。模型在处理特定网络流量分类任务 $\mathcal{T}_i$ 时，使用每个基参数对样本数据进行处理，将计算得到损失值的相反数输入到 Softmax 函数中作为线性组合系数来动态调整和生成新的初始化参数 θ_i^*，即如式(3)所示。

$$\theta_i^* = \sum_{j=1}^{n} \lambda_{ij} \cdot \theta_j \tag{3}$$

其中，λ_{ij} 表示在任务 $\mathcal{T}_i$ 中第 j 个基参数进行线性组合的系数。随后 θ_i^* 按照基学习器的优化方法在训练集上进行优化更新。进行元更新时，模型通过点积的形式对基参数向量施加正则化损失以保证基向量的正交性。最终通过元学习器，即式(2)得到最优全局初始化参数 θ^*。初始化参数 θ^* 在子空间 V 内的投影表示为 θ_{proj}^*。当现有基参数无法实现对任务分布的覆盖时，模型将自适应地扩大基的数量。XB-MAML 提出通过量化参数 θ^* 相对于 θ_{proj}^* 的偏离程度作为是否扩大基参数数量的依据，具体如式(4)所示。

$$\varepsilon = \frac{\left\| \theta^* - \theta_{proj}^* \right\|_2^2}{\left\| \theta^* \right\|_2^2} \tag{4}$$

当 ε 较大时，表明原始参数与其在特定子空间中的投影差异较大，即模型的当前基参数可能不足以覆盖全部任务特性。基于此，当 ε 在多个训练周期中持续增大时，即代表当前参数空间中基的数量不足，需要增大数量。在具体训练过程，如果在第 H 个周期中，ε 的平均值 $E(H)$ 大于前一个周期 ε 的平均值 $E(H-1)$ 时即在计数器中增加一个计数，当该数量超过阈值 c 时，模型将增加一个初始化基参数。

上述基于元学习的入侵检测方法具体流程如算法 1 所示:

算法 1 基于元学习的入侵检测方法

输入: 整体入侵流量数据任务分布 $p(\mathcal{T})$，每个任务 i 包括支持集 $\mathcal{S}_i$ 和查询集 $\mathcal{Q}_i$，基学习器学习率 α，元学习器学习率 β，随机初始化参数集合 Θ，扩展基向量阈值 c，训练周期 H。

输出: 最优全局初始化参数 θ^*

RANDOMLY INITIALIZE θ^*
WHILE NOT DONE DO
  SAMPLE BATCH OF TASKS $\mathcal{T}_i \sim p(\mathcal{T})$
  FOR ALL $\mathcal{T}_i$ DO
    CONVERT $\mathcal{S}_i$, $\mathcal{Q}_i$ INTO IMAGE
    $\left\{\mathcal{L}_{\mathcal{T}_i}(f_{\theta_j})\right\}_{j=1}^{N} = \left\{\mathcal{L}(\mathcal{S}_i, \theta_j)\right\}_{j=1}^{N}$
    λ_{ij} : Softmax $\left\{-\mathcal{L}_{\mathcal{T}_i}(f_{\theta_j})\right\}_{j=1}^{N}$
    $\theta_i^* = \sum_{j=1}^{n} \lambda_{ij} \cdot \theta_j$
    $\theta_i^* = \theta^* - \alpha \nabla_{\theta^*} \mathcal{L}_{\mathcal{T}_i}(f_{\theta^*})$
  END FOR
  FOR $j = 1 : N$ DO
    EVALUATE $\mathcal{L}_{reg,j}$
    INITIALIZE $\mathcal{L}_{total,j} = 0$
    FOR ALL $\mathcal{T}_i$ DO
      $\mathcal{L}_{total,j} = \mathcal{L}(\mathcal{Q}_i, \theta_i^*) + \mathcal{L}_{reg,j}$
    $\theta_j = \theta_j - \beta \nabla_{\theta_j} \sum_{\mathcal{T}_i \sim p(\mathcal{T})} \mathcal{L}_{total,j}$
  END FOR
  INITIALIZE $count = 0, E(H) = 0$
  CONSTRUCT SUBSPACE
  V=span $\{\theta_1, \theta_2, ..., \theta_N\}$
  FOR ALL $\mathcal{T}_i$ DO
    $\theta_{i,proj}^*$:Projection θ_i^* onto subspace V

$$\varepsilon = \frac{\left\| \theta_i^* - \theta_{i,proj}^* \right\|_2^2}{\left\| \theta_i^* \right\|_2^2}$$

$E(H)$: Average ε
END FOR
IF $E(H) > E(H-1)$ THEN
$count = count + 1$
ELSE
$count = 0$
END IF
IF $count > c$ then
ADD INITIAL MODEL
END IF

3.2 网络流量特征增强

网络流量的数据特征是影响入侵检测系统训练效果和检测准确率指标的重要因素。传统的特征提取方法依赖于手工设计规则或统计特征，虽然有一定效果，但在处理复杂和新型网络流量数据时，存在成本高、效果差等局限性。因此，增强网络流量特征在高维向量空间中的表示是提高检测效果的关键。当前大型视觉语言模型已经展现出卓越性能，已有大量研究基于预训练的跨模态大模型增强数据特征来提升模型整体效果。因此，本文使用基于 CLIP 的图像编码器对网络流量图像数据进行进一步特征提取和增强，捕捉关键区域特征，扩大不同种类流量数据在高维度特征向量空间中的距离，以提高模型分类效果。CLIP 基于残差网络(Residual Network, ResNet)[33]和视觉转换器(Vision Transformer, ViT)[34]预训练多个图像编码器，按照模型和参数量可以分为 ResNet-50、ResNet-101、ResNet-50×4、ResNet-50×16、ResNet-50×64 和 ViT - B/32、ViT-B/16、ViT-L/14、ViT-L/14@336px。本研究选取 CLIP 图像编码效果最好的基座模型 ViT-L/14@336px 作为网络流量数据的特征增强器。

具体特征增强流程如图 3 所示，模型将每一个原始网络流量灰度图像传入 CLIP 编码器，得到维度为 768 维的向量，随后将其转换成 24×24 的二维向量传入后续模型中进行元学习。

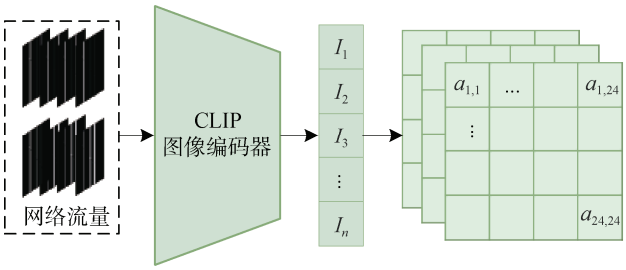

图 3 特征增强流程图

Figure 3 Flow chart of feature enhancement

3.3 基于元学习和特征增强的入侵检测模型设计

基于元学习和特征增强的入侵检测模型具体流程如图 4 所示。

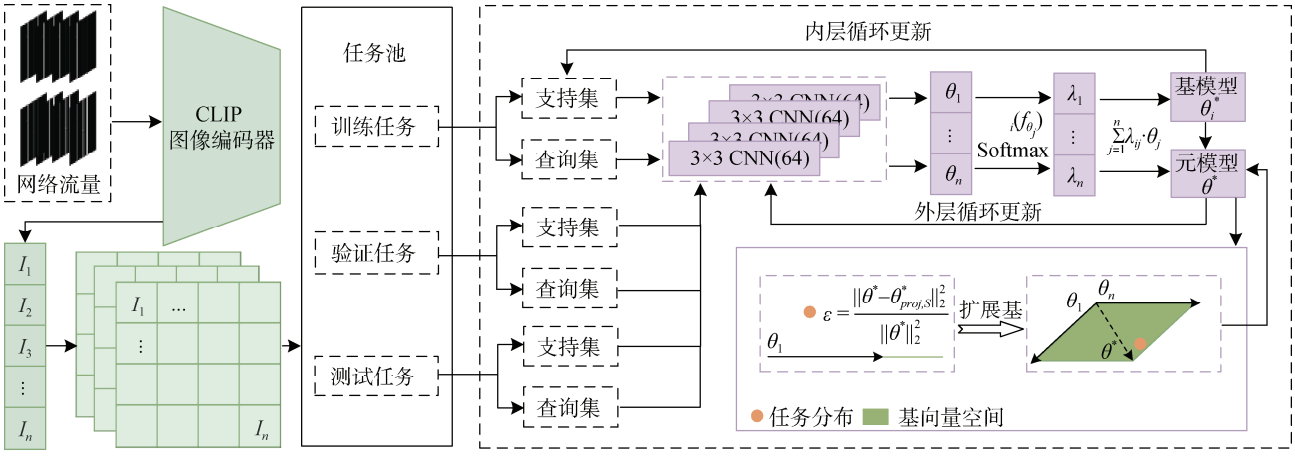

图 4 基于元学习和特征增强的入侵检测模型流程图

Figure 4 Flowchart of intrusion detection model based on meta-learning and feature enhancement

本文提出使用元学习框架实现小样本网络入侵检测，并引入特征增强技术提高方法准确率。构建了融合 XB-MAML 算法和预训练大型视觉语言模型 CLIP 的网络入侵检测模型 MCIDS，具体流程如图 4 所示。首先，MCIDS 将网络流量数据转换为灰度图数据后传入 CLIP 图像编码器。随后将图像编码器输出的 768 维特征向量转换为 28×28 的二维数据。经过数据预处理后，模型将所有数据传入任务池，按数

据集具体情况划分为训练任务、验证任务和测试任务，每个任务包含一定比例的支持集和查询集。在此基础上，基学习器在训练任务支持集上进行学习，更新局部参数。元学习器在训练任务查询集的基础上，更新全局参数。每个训练周期结束后，模型在验证任务上进行验证，监控模型的泛化能力，防止过拟合。在每个训练周期中，MCIDS 按照算法 1 中 20~32 行流程检验当前基参数与其在子空间中的投影的偏离程度决定是否扩大基参数数量，以保证模型可覆盖全部任务分布，提高整体的泛化能力。完成设定全部训练周期后，得到最优全局初始化参数 θ^*，结束模型训练。最终将测试任务输入 MCIDS，检验模型效果。

4 实验设置

本章对实验环境、数据集、模型参数设置和实验结果进行了全面的概述。

4.1 实验环境

本文所有实验均在 Ubuntu 20.04 操作系统上进行，使用 NVIDIA A100-PCIE-40GB,Intel(R) Xeon(R) Gold 6326 CPU @ 2.90GHz,377G 内存。实验环境为 Python 3.9, CUDA 11.1, PyTorch 1.9.0。

4.2 数据集

实验采用网络入侵检测领域常使用的入侵检测流量数据集，包括 UNSW-NB15[21]和 CSE-CIC-IDS 2018[35]。UNSW-NB15 是由新南威尔士大学网络靶场实验室捕获的 100 GB 原始网络流量数据包。该数据集包含 9 种攻击流量类型，共有 254 万条样本，但整体数据存在样本不平衡问题，Begin 类样本高达 221 万条，而 Worms 类样本仅有 174 条。CSE-CIC-IDS2018 是由加拿大通信安全机构和加拿大网络安全研究所联合设计的网络流量数据集，旨在为分析、测试和评估入侵检测系统提供支持。该数据集包括 7 种不同的攻击场景，15 种攻击流量类型，共约 1600 万条样本数据。但 CSE-CIC-IDS2018 也存在类不平衡性，只有 17%左右的样本包含入侵流量。针对类别不平衡问题，元学习可以有效提高模型在面对少量样本情况下的入侵检测准确率。

为便于比较基于相同入侵检测模型在不同数据集上的性能，Sarhan 等[36]针对多个网络流量数据集提出标准特征集 NetFlow V2。经实验验证，采用该标准特征集训练的模型具有与采用原数据特征集模型一致的分类性能。故本实验采用特征标准化处理后的网络流量入侵检测数据集。同时，为符合小样本训练环境，实验对以上数据集进行采样处理，每种流量类型随机选取 600 条样本，不足 600 条的种类进行随机过采样处理。实验遵循 Vinyals 等[22]提出 MAML 时采用的实验方案，即 *N*-way *K*-shot 小样本分类。N-way 代表选择 *N* 类未知流量，*K*-shot 代表每个种类为模型提供 *K* 条不同样本，以评估模型对流量数据中未知样本的分类能力。

4.3 模型参数设置

模型在 1Shot 和 5Shot 数据条件下训练过程如图 5 所示，纵坐标分别为训练任务检测准确率和损失函数值。由图 5 可知，同一数据集下，*K* 值越大，模型学习越充分，准确率越高，损失值越小。此外，模型在前 5000 次迭代周期效果提升显著，在 10000 个迭代周期以后趋于稳定，20000 次作业已经达到较好效果，因此，本文实验将迭代周期设置为 20000 次。

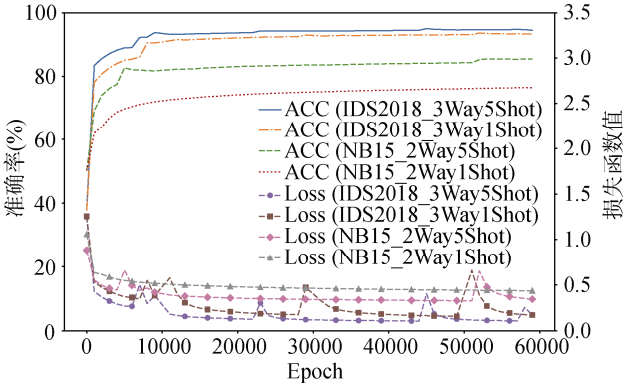

图 5 数据设置和迭代周期对检测性能的影响

Figure 5 Influence of data Settings and iteration cycles on detection performance

对于元学习模型训练过程来说，不同超参数的选取对模型检测效果影响显著。在本文实验过程中，结合文献[22]工作、经验准则和实验微调优化，选择以下模型超参数，具体如表 1 所示。

表 1 实验超参数

Table 1 Experimental hyperparameters

参数名称	参数值
周期	20000
批大小	3
内层更新步骤	3
内层优化学习率 α	0.03
外层优化学习率 β	0.001
扩展基向量阈值 c	500

4.4 对比实验

为验证 MCIDS 模型效果，本节对不同元学习模型和本文所提模型在不同数据集上的表现进行对比

分析。实验数据按照训练任务、测试任务和验证任务 3∶1∶1 的比例对数据集中的流量类型进行随机拆分, 具体拆分情况如表 2 所示。为更准确评估模型检测效果, 对流量类别进行了 3 次随机划分, 随后分别进行训练、验证和测试实验, 实验结果取 3 次测试结果的平均值。

表 2　实验数据拆分情况

Table 2　Splitting of experimental data

数据集	任务类别	流量种类
UNSW-NB15	训练任务	6
	验证任务	2
	测试任务	2
CSE-CIC-IDS2018	训练任务	9
	验证任务	3
	测试任务	3

按以上网络流量数据设置对各基线模型进行训练和测试。其中 MAML+CNN[8]是 Lu 等提出的基于 MAML 框架和 CNN 的网络入侵数据分析模型。该模型在多个方面的表现均优于传统机器学习模型和基于深度学习的模型。Meta-SGD[25]是一种基于随机梯度下降(Stochastic Gradient Descent, SGD)的元学习算法。Meta-SGD 将内循环学习率设为向量形式, 拥有高灵活性, 不仅可学习最优初始化参数, 还可学习最优更新方向和更新速率。Kitsune 是 Mirsky 等[15]基于无监督算法(Unsupervised Learning, UL)进行零日攻击检测的模型, 其使用自动编码器神经网络来区分正常和异常流量模式。Kitsune 可以在无监督情况下, 以高效的在线方式学习检测本地网络上的攻击。

不同数据拆分情况下, 实验结果取多次测试均值, 具体结果如表 3 所示, 图 6 直观地展示了不同模型在不同实验数据设置上的差距。

实验结果证明, 本文提出的 MCIDS 模型在不同小样本学习实验环境下, 针对入侵网络流量检测准确率较以往模型均具有一定提升, 整体效果优异, 在 CSE-CIC-IDS2018 数据集 3Way 20Shot 情形下, 模型最高准确率可达 94.46%, 相较 MAML+CNN、Meta-SGD 检测准确率分别提高了 3.1%、5.06%, 证明了本模型在相同算法框架下的效果。此外, 与无监督检测模型 Kitsune 相比, 本模型准确率平均提高 16.07%, 证明了本文所提模型与无监督算法相比具有明显的优越性。

4.5　消融实验

与其他元学习模型的对比实验结果证明了 MCIDS 模型融合特征增强和自适应可扩展基参数元

表 3　使用不同模型在不同实验设置下的性能

Table 3　Performance of different models in different experimental Settings

数据集	实验设置	模型	准确率
UNSW-NB15	2Way 1Shot	MAML+CNN	53.94 %
		Meta-SGD	57.10%
		MCIDS(ours)	**57.51 %**
	2Way 5Shot	MAML+CNN	62.99%
		Meta-SGD	58.51%
		MCIDS(ours)	**66.05%**
	2Way 10Shot	MAML+CNN	66.68%
		Meta-SGD	64.75%
		MCIDS(ours)	**69.10%**
	2Way 20Shot	MAML+CNN	69.51 %
		Meta-SGD	67.50%
		MCIDS(ours)	**70.53%**
	UL	Kitsune	61.56%
CSE-CIC-IDS2018	3Way 1Shot	MAML+CNN	81.58%
		Meta-SGD	82.59%
		MCIDS(ours)	**88.22%**
	3Way 5Shot	MAML+CNN	89.69%
		Meta-SGD	88.55%
		MCIDS(ours)	**91.43%**
	3Way 10Shot	MAML+CNN	91.33%
		Meta-SGD	88.77%
		MCIDS(ours)	**93.20%**
	3Way 20Shot	MAML+CNN	91.36%
		Meta-SGD	89.40%
		MCIDS(ours)	**94.46%**
	UL	Kitsune	66.68%

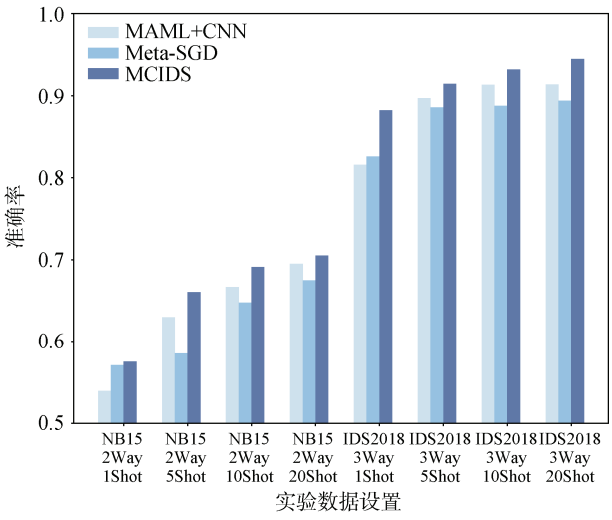

图 6　不同模型在不同实验设置下性能对比柱状图

Figure 6　Histogram of performance comparison of different models under different experimental settings

学习的有效性。为进一步验证模型各模块效果, 本节对 MCIDS 模型进行消融实验。消融实验考虑 MCIDS 模型的 4 种变体。

1) MAML: 仅使用 MAML 算法进行检测。

2) MAML＋CLIP: 使用特征增强和 MAML 算法进行异常流量检测。

3) XB-MAML: 仅使用 XB-MAML 算法进行检测。

4) MCIDS(XB-MAML＋CLIP): 融合特征增强和 XB-MAML 算法进行检测。

消融实验结果如表 4 所示。

表 4 消融实验结果

Table 4 Ablation experimental results

数据集	实验设置	模块	准确率
UNSW-NB15	2Way 1Shot	MAML	53.94%
		CLIP+MAML	56.17%
		XB-MAML	55.16%
		CLIP+XB-MAML(MCIDS)	**57.51%**
	2Way 5Shot	MAML	62.99%
		CLIP+MAML	64.51%
		XB-MAML	63.47%
		CLIP+XB-MAML(MCIDS)	**66.05%**
	2Way 10Shot	MAML	66.68%
		CLIP+MAML	67.55%
		XB-MAML	66.66%
		CLIP+XB-MAML(MCIDS)	**69.10%**
	2Way 20Shot	MAML	69.51%
		CLIP+MAML	70.30%
		XB-MAML	69.19%
		CLIP+XB-MAML(MCIDS)	**70.53%**
CSE-CIC-IDS2018	3Way 1Shot	MAML	81.58%
		CLIP+MAML	85.53%
		XB-MAML	84.89%
		CLIP+XB-MAML(MCIDS)	**88.22%**
	3Way 5Shot	MAML	89.69%
		CLIP+MAML	91.10%
		XB-MAML	90.96%
		CLIP+XB-MAML(MCIDS)	**91.43%**
	3Way 10Shot	MAML	91.33%
		CLIP+MAML	92.53%
		XB-MAML	92.29%
		CLIP+XB-MAML(MCIDS)	**93.20%**
	3Way 20Shot	MAML	91.36%
		CLIP+MAML	93.29%
		XB-MAML	93.83%
		CLIP+XB-MAML(MCIDS)	**94.46%**

由实验结果可知, 删除部分模块的模型检测准确率与 MCIDS 相比, 均有不同程度的下降。以数据集 CSE-CIC-IDS2018 数据集 3Way 1Shot 为例, 第一, 基于 XB-MAML 算法的模型较基线模型提高了 3.31%, 证明了将初始化参数转换为自适应可扩展的基参数的线性组合这一方法的有效性和在网络流量检测领域的适配性。第二, 基于 CLIP 特征增强和 MAML 算法的入侵检测模型较基线模型提高了 3.95%, 证明了基于 CLIP 的特征增强方法对提高小样本网络流量检测效果的有效性。第三, 融合 CLIP 特征增强和 XB-MAML 算法的 MCIDS 模型较基于单个改进方法的模型效果分别提升 3.63%和 2.69%, 相较基线模型提高 6.64%, 证明了模型两种改进方法的互补性。因此, 本文基于元学习和特征增强的小样本网络入侵检测模型 MCIDS 具有合理性和有效性, 达到了较好的检测效果。

5 总结与未来工作

本文提出了一种融合元学习和特征增强的网络入侵检测模型 MCIDS。模型通过利用 CLIP 图像编码器进行流量数据特征增强, 随后通过元学习算法自适应组合多个初始化参数, 实现仅使用少量训练样本进行网络入侵检测应用的效果。与其他基于元学习和无监督算法进行入侵检测的模型相比, 本文所构建的 MCIDS 入侵检测模型具有更优秀的检测效果。并且消融实验的结果验证了本文采用方法的合理性和有效性。

在未来工作中, 可以尝试更多网络入侵流量数据类型, 扩展样本多样性, 验证模型在更大任务分布数据集上的效果。此外, 尝试推进模型在实际场景中的应用示范, 进一步评估在实战场景中的应用效果。

参考文献

[1] Mehedi S T, Anwar A, Rahman Z, et al. Dependable intrusion detection system for IoT: A deep transfer learning based approach[J]. *IEEE Transactions on Industrial Informatics*, 2023, 19(1): 1006-1017.

[2] Wang Y Q, Yao Q M, Kwok J T, et al. Generalizing from a few examples: A survey on few-shot learning[J]. *ACM Computing Surveys*, 2020, 53(3): 1-34.

[3] Ni T, Li J F, Zhang X K, et al. Exploiting Contactless Side Channels in Wireless Charging Power Banks for User Privacy Inference *via* Few-Shot Learning[C]. *The 29th Annual International Conference on Mobile Computing and Networking*, 2023: 1-15.

[4] Gong T, Kim Y, Shin J, et al. MetaSense: Few-Shot Adaptation to Untrained Conditions in Deep Mobile Sensing[C]. *The 17th Conference on Embedded Networked Sensor Systems*, 2019: 110-123.

[5] Sirinam P, Mathews N, Rahman M S, et al. Triplet Fingerprinting: More Practical and Portable Website Fingerprinting with N-Shot Learning[C]. *The 2019 ACM SIGSAC Conference on Computer*

and Communications Security, 2019: 1131-1148.

[6] Xu C Y, Shen J Z, 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.

[7] Finn C, Abbeel P, Levine S. Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks[C]. *The 34th International Conference on Machine Learning - Volume 70*, 2017: 1126-1135.

[8] Lu C M, Wang X F, Yang A M, et al. A few-shot-based model-agnostic meta-learning for intrusion detection in security of internet of things[J]. *IEEE Internet of Things Journal*, 2023, 10(24): 21309-21321.

[9] Eskin E. Anomaly Detection over Noisy Data Using Learned Probability Distributions[C]. *The Seventeenth International Conference on Machine Learning*, 2000: 255-262.

[10] Thottan M, Ji C Y. Anomaly detection in IP networks[J]. *IEEE Transactions on Signal Processing*, 2003, 51(8): 2191-2204.

[11] Lian X W, Ma Y, Chen Y (L /Y), et al. Shodan traffic identification based on load characteristics and statistical characteristics[J]. *Computer Engineering*, 2021, 47(1): 117-122.
(连晓伟, 马垚, 陈永乐, 等. 基于载荷特征与统计特征的 Shodan 流量识别[J]. *计算机工程*, 2021, 47(1): 117-122.)

[12] Xu P, Lin S. Internet traffic classification using C4.5 decision tree[J]. *Journal of Software*, 2009, 20(10): 2692-2704.
(徐鹏, 林森. 基于 C4.5 决策树的流量分类方法[J]. *软件学报*, 2009, 20(10): 2692-2704.)

[13] Wang Y. A Multinomial Logistic regression modeling approach for anomaly intrusion detection[J]. *Computers & Security*, 2005, 24(8): 662-674.

[14] Stolfo S J, Lee W, Fan W, et al. Cost-based modeling and evaluation for data mining with application to fraud and intrusion detection: Results from the JAM project[C]. *DARPA Information Survivability Conference and Exposition*, 2000, 2: 130-144.

[15] Mirsky Y, Doitshman T, Elovici Y, et al. Kitsune: An Ensemble of Autoencoders for Online Network Intrusion Detection[EB/OL]. 2018: arXiv: 1802.09089. https://arxiv.org/abs/1802.09089.

[16] Fu C P, Li Q, Xu K. Detecting Unknown Encrypted Malicious Traffic in Real Time *via* Flow Interaction Graph Analysis[EB/OL]. 2023: arXiv: 2301.13686. https://arxiv.org/abs/2301.13686.

[17] Gao N, Gao L, Gao Q L, et al. An Intrusion Detection Model Based on Deep Belief Networks[C]. *2014 Second International Conference on Advanced Cloud and Big Data*, 2015: 247-252.

[18] Xiao Y H, Xing C, Zhang T N, et al. An intrusion detection model based on feature reduction and convolutional neural networks[J]. *IEEE Access*, 2019, 7: 42210-42219.

[19] Ma Z X, Li J, Song Y F, et al. Network intrusion detection method based on FCWGAN and BiLSTM[J]. *Computational Intelligence and Neuroscience*, 2022, 2022(1): 6591140.

[20] Tavallaee M, Bagheri E, Lu W, et al. A Detailed Analysis of the KDD CUP 99 Data Set[C]. *2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications*, 2009: 1-6.

[21] Moustafa N, Slay J. UNSW-NB15: A Comprehensive Data Set for Network Intrusion Detection Systems (UNSW-NB15 Network Data Set)[C]. *2015 Military Communications and Information Systems Conference*, 2015: 1-6.

[22] Vinyals O, Blundell C, Lillicrap T P, et al. Matching Networks for One Shot Learning[C]. *Neural Information Processing Systems*, 2016.

[23] Oreshkin B N, López P R, Lacoste A. TADAM: Task Dependent Adaptive Metric for Improved Few-Shot Learning[C]. *Neural Information Processing Systems*, 2018.

[24] Antoniou A, Edwards H, Storkey A. How to train your MAML[C]. *International conference on learning representations*, 2018.

[25] Li Z G, Zhou F W, Chen F, et al. Meta-SGD: Learning to Learn Quickly for Few-Shot Learning[EB/OL]. 2017: arXiv: 1707.09835. https://arxiv.org/abs/1707.09835.

[26] Jiang W, Kwok J, Zhang Y. Subspace learning for effective meta-learning[C]. *International Conference on Machine Learning*, 2022: 10177-10194.

[27] Lee J J, Yoon S W. XB-MAML: Learning Expandable Basis Parameters for Effective Meta-Learning with Wide Task Coverage[EB/OL]. 2024: arXiv: 2403.06768. https://arxiv.org/abs/2403.06768.

[28] Wei N, Yin L H, Zhou X M, et al. A feature enhancement-based model for the malicious traffic detection with small-scale imbalanced dataset[J]. *Information Sciences*, 2023, 647: 119512.

[29] Radford A, Kim J W, Hallacy C, et al. Learning Transferable Visual Models from Natural Language Supervision[EB/OL]. 2021: arXiv: 2103.00020. https://arxiv.org/abs/2103.00020.

[30] Shen S, Li L H, Tan H, et al. How much Can CLIP Benefit Vision-and-Language Tasks? [EB/OL]. 2021: arXiv: 2107.06383. https://arxiv.org/abs/2107.06383.

[31] Li Z Q, Sun H, Xia P F, et al. Efficient Backdoor Attacks for Deep Neural Networks in Real-World Scenarios[EB/OL]. 2023: arXiv: 2306.08386. https://arxiv.org/abs/2306.08386.

[32] Zhou P, Zou Y T, Yuan X T, et al. Task Similarity Aware Meta Learning: Theory-Inspired Improvement on MAML[C]. *Conference on Uncertainty in Artificial Intelligence*, 2021.

[33] He K M, Zhang X Y, Ren S Q, et al. Deep Residual Learning for Image Recognition[C]. *2016 IEEE Conference on Computer Vision and Pattern Recognition*, 2016: 770-778.

[34] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An Image Is Worth 16x16 Words: Transformers for Image Recognition at Scale[EB/OL]. 2020: arXiv: 2010.11929. https://arxiv.org/abs/2010.11929.

[35] Sharafaldin I, Habibi Lashkari A, Ghorbani A A. Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization[C]. *The 4th International Conference on Information Systems Security and Privacy*, 2018: 108-116.

[36] Sarhan M, Layeghy S, Portmann M. Towards a standard feature set for network intrusion detection system datasets[J]. *Mobile Networks and Applications*, 2022, 27(1): 357-370.

蒋章涛 于 2023 年在中国人民公安大学网络安全与执法专业获得学士学位。现在中国人民公安大学网络空间安全执法技术专业攻读硕士学位, CCF 学生会员。研究领域为网络安全、深度学习。Email: 2023211505@stu.ppsuc.edu.cn

李欣 于 2007 年在浙江大学获得博士学位。现任中国人民公安大学信息网络安全学院院长, 博士生导师, 教授, CCF 会员(51691M)。研究领域为网络安全、人工智能。Email: lixin@ppsuc.edu.cn

薛迪 于 2023 年在中国人民公安大学网络安全与执法专业获得学士学位。现在中国人民公安大学网络空间安全执法技术专业攻读硕士学位。研究领域为视觉问答、大语言模型。Email: 2023211517@stu.ppsuc.edu.cn

彭奕杰 于 2024 年在中国人民公安大学网络安全与执法专业获得学士学位。现在中国人民公安大学网络空间安全执法技术专业攻读硕士学位。研究领域为大模型、深度学习。Email: 2516789559@qq.com