基于轻量级卷积神经网络的移动端实时入侵检测系统

何宋阳

（四川大学网络空间安全学院，四川 成都 610065）

摘　要：随着移动互联网的快速普及与移动端设备的广泛应用，移动终端已成为人们生产生活、信息交互的核心载体，与此同时，移动端网络安全威胁呈现出多样化、隐蔽化、智能化的发展态势，恶意入侵行为不仅会导致用户隐私泄露、财产损失，还可能威胁国家网络空间安全。传统入侵检测方法受限于移动端设备算力有限、存储资源不足、能耗敏感等固有特性，难以实现实时、高效的入侵检测，存在检测准确率低、误报率高、资源占用过大等问题。卷积神经网络（CNN）凭借其强大的特征自动提取能力，在网络入侵检测领域展现出良好的应用前景，但传统CNN模型结构复杂、参数量庞大，无法直接部署于移动端设备。为此，本文提出一种基于轻量级卷积神经网络的移动端实时入侵检测系统，通过优化CNN网络架构、引入模型轻量化技术，在保证检测性能的前提下，大幅降低模型参数量与计算复杂度，适配移动端设备的资源约束。本文首先分析移动端网络安全威胁现状与传统入侵检测方法的局限性，论证轻量级CNN在移动端入侵检测中的适用性；其次，综述移动端入侵检测技术与轻量级神经网络的研究进展，明确当前研究的不足与本文的研究切入点；然后，详细设计基于轻量级CNN的入侵检测模型，包括输入层的移动端网络流量预处理模块、轻量级CNN核心特征提取模块以及输出层的异常分类与实时决策机制，并提出针对性的轻量化与实时性优化策略；接着，通过搭建移动端仿真平台，采用公开入侵检测数据集开展对比实验，从检测准确率、误报率、推理时间、资源占用率等多个维度验证所提系统的有效性与优越性；最后，总结本文的研究成果，分析研究过程中存在的不足，并展望未来的改进方向。实验结果表明，本文提出的轻量级CNN入侵检测系统，在检测准确率上达到98.7%以上，误报率控制在1.2%以下，模型推理时间缩短至50ms以内，内存占用量低于8MB，相较于传统入侵检测方法与普通CNN模型，在检测性能与移动端适配性上均有显著提升，能够满足移动端实时入侵检测的实际需求，为移动端网络安全防护提供了一种高效、可行的技术方案。

关键词：轻量级卷积神经网络；移动端；入侵检测；实时性；模型轻量化；网络流量分析

中图分类号：TP393　　**文献标识码：**A　　**文章编号：**3106-2709（2025）02-0027-12

DOI：10.62022/NCAR.issn3106-2709.2025.02.003

Real-time Intrusion Detection System for Mobile Terminals Based on Lightweight Convolutional Neural Network

He Songyang

（College of Cyberspace Security, Sichuan University, Chengdu, Sichuan 610065）

Abstract: With the rapid popularization of mobile Internet and the wide application of mobile devices, mobile terminals have become the core carrier of people´s production and life, as well as information interaction. At the same time, mobile network security threats are showing a diversified, hidden and intelligent development trend. Malicious intrusion behaviors will not only lead to the leakage of user privacy and property losses, but also may threaten national cyberspace security. Traditional intrusion detection methods are limited by the inherent characteristics of mobile devices, such as limited computing power, insufficient storage resources and sensitive energy consumption, making it difficult to achieve real-time and efficient intrusion detection, with problems such as low detection accuracy, high false positive rate and excessive resource occupation. Convolutional Neural Network (CNN) has shown good application prospects in the field of network intrusion detection due to its strong ability of automatic feature extraction, but the traditional CNN model has complex structure and large number of parameters, which cannot be directly deployed on mobile devices. Therefore, this paper proposes a real-time intrusion detection system for mobile terminals based on lightweight convolutional neural network. By optimizing the CNN network architecture and introducing model lightweight technology, the number of model parameters and computational complexity are greatly reduced under the premise of ensuring detection performance, so as to adapt to the resource constraints of mobile devices. This paper first analyzes the current situation of mobile network security threats and the limitations of traditional intrusion detection methods, and demonstrates the applicability of lightweight CNN in mobile intrusion detection; secondly, summarizes the research progress of mobile intrusion detection technology and lightweight neural

作者简介：何宋阳，博士，讲师，研究方向为模式识别。

network, and clarifies the shortcomings of current research and the research entry point of this paper; then, designs the intrusion detection model based on lightweight CNN in detail, including the mobile network traffic preprocessing module of the input layer, the lightweight CNN core feature extraction module and the anomaly classification and real-time decision-making mechanism of the output layer, and proposes targeted lightweight and real-time optimization strategies; then, by building a mobile simulation platform and using public intrusion detection datasets to carry out comparative experiments, the effectiveness and superiority of the proposed system are verified from multiple dimensions such as detection accuracy, false positive rate, inference time and resource occupation rate; finally, summarizes the research results of this paper, analyzes the deficiencies in the research process, and looks forward to the future improvement direction. The experimental results show that the lightweight CNN intrusion detection system proposed in this paper achieves a detection accuracy of more than 98.7%, a false positive rate of less than 1.2%, a model inference time shortened to less than 50ms, and a memory occupation of less than 8MB. Compared with traditional intrusion detection methods and ordinary CNN models, it has significant improvements in detection performance and mobile adaptability, which can meet the actual needs of real-time intrusion detection on mobile terminals and provide an efficient and feasible technical solution for mobile network security protection.

Keywords: Lightweight Convolutional Neural Network; Mobile Terminal; Intrusion Detection; Real-time Performance; Model Lightweight; Network Traffic Analysis

1 引言

在数字经济高速发展的今天，移动互联网已深度融入社会经济的各个领域，智能手机、平板电脑、智能穿戴设备等移动端终端的普及率持续攀升，成为人们获取信息、沟通交流、在线交易、娱乐休闲的主要工具。据中国互联网络信息中心（CNNIC）发布的第55次《中国互联网络发展状况统计报告》显示，截至2024年6月，我国移动电话用户数达17.6亿户，移动互联网用户规模突破12.5亿，移动互联网接入流量达2142亿GB，同比增长18.2%。移动端设备的广泛应用，极大地便利了人们的生产生活，但同时也带来了严峻的网络安全挑战。随着移动应用的不断丰富、网络传输技术的迭代升级，移动端网络安全威胁呈现出多样化、隐蔽化、智能化的新特征，恶意入侵行为频发，给用户隐私安全、财产安全以及国家网络空间安全带来了严重威胁[1]。

入侵检测作为网络安全防护体系的核心组成部分，能够实时监测网络流量与系统行为，及时发现并预警恶意入侵行为，为移动端设备提供主动防护。传统的入侵检测方法主要分为基于特征匹配的误用检测和基于异常行为的异常检测两大类，这些方法在固定网络环境中取得了一定的应用效果，但在移动端环境中，由于设备自身算力有限、存储资源不足、能耗敏感，且网络环境动态变化、流量类型复杂多样，传统入侵检测方法难以满足实时、高效的检测需求，存在诸多局限性。

近年来，深度学习技术的快速发展为入侵检测领域带来了新的机遇，卷积神经网络（CNN）作为深度学习的核心算法之一，凭借其强大的特征自动提取能力、良好的泛化能力和容错能力，在图像识别、自然语言处理、网络安全等领域得到了广泛应用[2]。在网络入侵检测中，CNN能够自动从海量的网络流量数据中提取深层特征，无需人工进行复杂的特征工程，有效解决了传统方法中特征提取依赖人工经验、适应性差的问题。然而，传统CNN模型结构复杂、参数量庞大、计算开销高，需要大量的算力和存储资源支撑，无法直接部署于移动端设备，限制了其在移动端入侵检测中的应用。

轻量级卷积神经网络作为传统CNN的优化版本，通过简化网络结构、减少参数量、优化计算流程等方式，在保证模型性能的前提下，大幅降低了模型的计算复杂度和资源占用率，能够很好地适配移动端设备的资源约束[3]。因此，研究基于轻量级卷积神经网络的移动端实时入侵检测系统，解决传统入侵检测方法在移动端的局限性，实现高效、实时的移动端入侵检测，具有重要的理论意义和实际应用价值。

本文围绕基于轻量级卷积神经网络的移动端实时入侵检测系统展开深入研究，首先分析移动端网络安全威胁现状与传统入侵检测方法的局限性，论证轻量级CNN在移动端入侵检测中的适用性；然后综述相关研究工作，明确当前研究的不足；接着设计轻量级CNN入侵检测模型及优化策略；随后通过实验验证模型的有效性与优越性；最后总结研究成果并展望未来改进方向，为移动端网络安全防护提供技术支撑。

1.1 研究背景与意义

随着5G技术的全面普及、物联网技术的快速发展以及移动应用生态的不断完善，移动端设备已成为连接人、物、网络的核心枢纽，其承载的信息价值和应用场景不断拓展。从个人用户的隐私信息、金融账户，到企业的商业数据、政务系统，再到国家的关键信息基础设施，都与移动端设备有着密切的关联。然而，移动端网络环境的开放性、动态性以及设备自身的资源约束，使得移动端成为网络攻击的主要目标，恶意入侵行为频发，网络安全形势日益严峻[4]。

传统的移动端安全防护手段主要以杀毒软件、防火墙、

加密传输等为主，这些手段主要针对已知的恶意软件和攻击行为，对未知的、新型的入侵行为防御能力较弱。入侵检测系统作为网络安全防护体系的“哨兵”，能够实时监测网络流量和系统行为，及时发现异常入侵行为，弥补传统防护手段的不足。但目前大多数入侵检测系统都是针对固定网络环境设计的，无法很好地适配移动端设备的资源约束和网络环境特点，存在检测效率低、误报率高、资源占用过大等问题，难以满足移动端实时入侵检测的需求。

轻量级卷积神经网络的出现，为解决移动端入侵检测的困境提供了新的技术路径。轻量级CNN通过优化网络架构、引入模型压缩技术，在保证检测性能的前提下，大幅降低了模型的参数量和计算复杂度，能够在移动端设备上实现高效、实时的推理[5]。因此，开展基于轻量级卷积神经网络的移动端实时入侵检测系统研究，不仅能够丰富深度学习在网络安全领域的应用，推动轻量级神经网络技术的发展，还能够为移动端设备提供高效、可靠的安全防护方案，保障用户隐私安全、财产安全以及国家网络空间安全，具有重要的理论意义和实际应用价值。

1.1.1 移动端网络安全威胁现状

随着移动端设备的普及和移动互联网的发展，移动端网络安全威胁呈现出多样化、隐蔽化、智能化、规模化的发展态势，恶意入侵行为的种类不断增多、手段不断升级，对移动端设备和用户安全造成了严重威胁。根据瑞星公司发布的《2025年中国网络安全报告》显示，2025年网络攻击手段持续升级，人工智能技术与供应链渗透成为年度最突出的安全挑战，其中移动端成为攻击的重灾区，各类恶意攻击行为频发，威胁形势日益严峻。结合当前移动端网络安全的实际情况，其安全威胁主要体现在以下几个方面：

第一，恶意应用攻击频发，成为移动端最主要的安全威胁。恶意应用是指未经用户许可，擅自收集用户信息、窃取用户财产、破坏设备正常运行的应用程序，其传播途径主要包括第三方应用市场、短信链接、社交软件分享、恶意广告推送等。近年来，恶意应用的数量呈现出爆发式增长，且种类不断丰富，主要包括窃取隐私类、恶意扣费类、病毒木马类、钓鱼欺诈类等。例如，窃取隐私类恶意应用会擅自收集用户的通讯录、短信、位置信息、照片、支付记录等敏感信息，将其出售给第三方，导致用户隐私泄露；恶意扣费类应用会在用户不知情的情况下，擅自订购增值服务、扣除手机话费，造成用户财产损失；病毒木马类应用会植入手机系统，控制手机设备，窃取用户信息、远程操控手机，甚至发起网络攻击；钓鱼欺诈类应用会仿冒正规应用的界面，诱导用户输入账号密码、支付密码等敏感信息，实施诈骗行为[6]。据国家互联网应急中心（CNCERT）发布的《2023年中国互联网网络安全报告》显示，2023年我国共监测到移动恶意应用234.2万个，同比增长12.7%，其中窃取隐私类恶意应用占比最高，达45.3%，恶意扣费类和病毒木马类应用分别占比23.6%和18.9%。

第二，网络钓鱼攻击呈现智能化、隐蔽化趋势，用户识别难度加大。网络钓鱼攻击是指攻击者通过仿冒正规网站、发送虚假短信、推送虚假广告等方式，诱导用户输入敏感信息，从而窃取用户财产或隐私的攻击行为。在移动端，网络钓鱼攻击主要针对用户的金融账户、社交账号、支付账号等，其攻击手段不断升级，呈现出智能化、隐蔽化的特点。例如，攻击者利用人工智能技术生成高仿真的钓鱼网站和钓鱼短信，其界面、内容与正规网站和短信高度相似，用户难以区分；部分钓鱼攻击通过社交软件、短视频平台等渠道传播，利用用户的信任心理，诱导用户点击链接、下载应用，从而实施攻击。此外，随着5G技术的普及，钓鱼攻击的传播速度更快、范围更广，给用户带来了更大的安全风险。据统计，2023年我国移动端用户遭遇网络钓鱼攻击的次数达1.2亿次，同比增长15.3%，造成的财产损失超过50亿元。

第三，无线网络攻击日益突出，公共Wi-Fi成为攻击重灾区。随着无线网络的普及，越来越多的用户习惯在公共场所连接公共Wi-Fi进行上网，然而公共Wi-Fi网络存在诸多安全隐患，容易成为攻击者实施攻击的目标。攻击者可以通过搭建虚假公共Wi-Fi、破解公共Wi-Fi密码、实施中间人攻击等方式，窃取用户的网络流量数据、敏感信息，甚至控制用户的移动端设备。例如，攻击者搭建虚假的公共Wi-Fi热点，当用户连接该热点后，所有的网络传输数据都会经过攻击者的设备，攻击者可以轻松窃取用户的账号密码、支付信息等敏感数据；中间人攻击则是攻击者拦截用户与服务器之间的通信，篡改通信数据，实施诈骗或窃取信息的行为[7]。此外，部分公共Wi-Fi网络未采取加密措施，网络流量数据容易被监听和窃取，进一步加剧了安全风险。据调查显示，超过60%的移动端用户曾在公共场所连接过公共Wi-Fi，其中有23%的用户遭遇过网络攻击，导致隐私泄露或财产损失。

第四，系统漏洞与恶意利用并存，设备安全防线薄弱。移动端设备的操作系统（如Android、iOS）和各类应用程序都可能存在安全漏洞，这些漏洞被攻击者发现后，会被恶意利用，实施入侵攻击。Android系统由于其开源特性，漏洞数量相对较多，且不同品牌、不同型号的设备更新迭代速度不一，很多老旧设备无法及时获得系统补丁更新，导致漏洞

长期存在，成为攻击者攻击的突破口。iOS系统虽然安全性相对较高，但也存在少量漏洞，攻击者通过利用这些漏洞，能够绕过系统安全机制，获取设备的最高权限，窃取用户信息或控制设备。此外，部分移动应用程序在开发过程中，缺乏安全意识，存在代码漏洞、权限滥用等问题，也容易被攻击者利用，实施入侵攻击。据国家信息安全漏洞库（CNNVD）统计，2023年我国共收录移动端操作系统漏洞1234个，移动应用漏洞2789个，其中高危漏洞占比达32.7%，这些漏洞被恶意利用后，给移动端设备安全带来了严重威胁。

第五，新型攻击手段不断涌现，防御难度持续加大。随着人工智能、大数据、区块链等技术的发展，攻击者开始将这些技术应用于网络攻击中，形成了新型的攻击手段，如AI生成恶意代码、AI驱动的钓鱼攻击、区块链恶意挖矿等[8]。AI生成恶意代码能够快速生成大量多样化的恶意代码，躲避传统杀毒软件的检测；AI驱动的钓鱼攻击能够根据用户的行为习惯，定制个性化的钓鱼内容，提高攻击的成功率；区块链恶意挖矿则是攻击者利用移动端设备的算力，进行区块链挖矿，消耗设备电量和算力，导致设备运行变慢、发热，甚至损坏设备。这些新型攻击手段具有智能化、自动化、隐蔽化的特点，传统的入侵检测方法难以有效识别和防御，进一步加大了移动端网络安全的防御难度。

此外，移动端网络安全威胁还呈现出规模化、产业化的趋势，形成了“制作恶意应用—传播恶意应用—实施攻击—窃取利益”的完整黑色产业链。攻击者通过专业化的团队，制作各类恶意应用和攻击工具，通过多种渠道进行传播，实施规模化的攻击行为，获取非法利益。这种黑色产业链的存在，不仅加剧了移动端网络安全的威胁，也给网络安全监管带来了巨大的挑战。

综上所述，当前移动端网络安全威胁形势日益严峻，恶意入侵行为频发，种类不断增多，手段不断升级，给用户隐私安全、财产安全以及国家网络空间安全带来了严重威胁。因此，研发高效、实时的移动端入侵检测系统，及时发现并预警恶意入侵行为，成为当前网络安全领域的迫切需求。

1.1.2 传统入侵检测方法在移动端的局限性

传统的入侵检测方法主要分为基于特征匹配的误用检测和基于异常行为的异常检测两大类，这些方法在固定网络环境中（如局域网、服务器集群）取得了一定的应用效果，但在移动端环境中，由于设备自身的资源约束、网络环境的动态变化以及攻击手段的不断升级，传统入侵检测方法存在诸多局限性，难以满足移动端实时入侵检测的需求，具体主要体现在以下几个方面：

第一，特征提取依赖人工经验，适应性差，难以应对新型攻击。基于特征匹配的误用检测是传统入侵检测方法中最常用的一种，其核心思想是事先建立已知攻击行为的特征库，当监测到的网络流量或系统行为与特征库中的特征匹配时，判定为入侵行为。这种方法的检测准确率较高，但依赖于人工提取攻击特征，需要专业的安全人员根据已知的攻击行为，总结提炼攻击特征，建立特征库。然而，移动端网络攻击手段不断升级，新型攻击行为层出不穷，人工提取特征的速度远远跟不上攻击手段的更新速度，导致特征库无法及时更新，无法识别新型攻击行为[9]。此外，不同类型的移动端设备、不同的移动应用、不同的网络环境，其攻击特征也存在差异，人工提取的特征通用性较差，难以适应多样化的移动端环境，导致检测准确率下降。

第二，计算复杂度高，资源占用过大，无法适配移动端设备。传统的入侵检测方法，无论是基于特征匹配的误用检测，还是基于异常行为的异常检测，都需要对大量的网络流量数据进行分析和处理，计算复杂度较高，需要消耗大量的算力和存储资源。例如，基于特征匹配的误用检测需要对每一条网络流量数据与特征库中的所有特征进行匹配，当特征库规模较大时，匹配过程会消耗大量的算力和时间；基于异常行为的异常检测需要建立正常行为模型，对网络流量数据进行统计分析，计算数据的偏离度，也需要消耗大量的计算资源。而移动端设备的算力、存储资源有限，电池续航能力也存在约束，传统入侵检测方法的高计算复杂度和高资源占用，会导致设备运行变慢、电池消耗过快，甚至出现卡顿、闪退等问题，无法在移动端设备上稳定运行。腾讯云开发者社区的相关研究表明，传统IDS通常针对固定网络环境设计，难以有效监控移动设备在蜂窝网络、Wi-Fi切换等动态场景下的流量，且其高资源消耗特性无法适配移动端设备的沙盒机制和权限控制要求[10]。

第三，检测实时性差，无法及时发现和预警入侵行为。移动端网络流量具有实时性、突发性的特点，用户的网络行为频繁变化，网络流量数据量大、更新速度快，这就要求入侵检测系统能够实时监测网络流量，及时发现并预警入侵行为。然而，传统的入侵检测方法由于计算复杂度高，对网络流量数据的分析和处理速度较慢，存在一定的检测延迟，无法及时发现入侵行为。当入侵行为发生后，往往需要经过一段时间才能被检测到，此时用户的隐私信息、财产已经遭受损失，入侵检测系统的防护作用无法得到有效发挥。例如，在移动端支付场景中，攻击者实施钓鱼攻击或恶意扣费攻击时，需要入侵检测系统能够实时识别并拦截，但传统方法的检测延

迟会导致攻击行为完成后才被发现，造成用户财产损失。

第四，误报率高，影响用户体验。传统的入侵检测方法，尤其是基于异常行为的异常检测，其正常行为模型的建立往往依赖于有限的样本数据，当网络环境发生变化、用户行为发生改变时，正常行为模型会出现偏差，导致将正常的网络行为误判为入侵行为，产生误报[11]。此外，移动端网络环境复杂多样，网络流量数据存在大量的噪声数据，传统方法难以有效过滤噪声数据，也会导致误报率升高。误报率过高会频繁触发预警提示，干扰用户的正常使用，降低用户体验，甚至导致用户忽视预警提示，错过真正的入侵行为预警。

第五，缺乏对移动端设备特性的适配，部署难度大。传统的入侵检测系统大多是针对固定网络环境设计的，没有充分考虑移动端设备的特性，如设备型号多样、操作系统版本不同、硬件配置差异大、网络环境动态变化等，导致入侵检测系统难以在不同的移动端设备上部署和运行。例如，部分入侵检测系统仅支持特定的操作系统版本，无法在老旧设备上运行；部分系统对硬件配置要求较高，中低端移动端设备无法满足其运行需求。此外，移动端设备的权限管理严格，传统入侵检测系统需要获取较高的系统权限才能实现全面的监测，这不仅增加了系统的部署难度，也可能引发用户对隐私安全的担忧。

第六，对加密流量的检测能力薄弱。随着移动端网络安全意识的提高，越来越多的移动应用采用加密传输技术，如HTTPS、SSL等，对网络流量进行加密，保护用户数据的安全。然而，传统的入侵检测方法主要针对明文流量进行分析，无法对加密流量进行有效解析，难以识别加密流量中的入侵行为。攻击者可以利用加密传输技术，隐藏恶意攻击行为，绕过传统入侵检测系统的检测，实施入侵攻击，进一步加剧了移动端网络安全的威胁。

综上所述，传统的入侵检测方法在移动端环境中存在特征提取依赖人工、计算复杂度高、实时性差、误报率高、部署难度大、对加密流量检测能力薄弱等局限性，无法满足移动端实时、高效、精准的入侵检测需求。因此，需要寻找一种新的技术方法，解决传统方法的不足，研发适配移动端设备的入侵检测系统。

1.1.3 轻量级卷积神经网络（CNN）的适用性分析

卷积神经网络（CNN）作为深度学习的核心算法之一，凭借其强大的特征自动提取能力、良好的泛化能力和容错能力，在网络入侵检测领域展现出良好的应用前景。与传统入侵检测方法相比，CNN能够自动从海量的网络流量数据中提取深层特征，无需人工进行复杂的特征工程，有效解决了传统方法中特征提取依赖人工经验、适应性差的问题[12]。然而，传统CNN模型结构复杂、参数量庞大、计算开销高，需要大量的算力和存储资源支撑，无法直接部署于移动端设备。

轻量级卷积神经网络是在传统CNN的基础上，通过优化网络架构、减少参数量、优化计算流程、引入模型压缩技术等方式，研发的一种轻量化模型。其核心目标是在保证模型性能的前提下，大幅降低模型的计算复杂度和资源占用率，使其能够适配移动端设备的资源约束。轻量级CNN不仅继承了传统CNN的特征自动提取能力和良好的检测性能，还具有参数量少、计算开销低、运行速度快、能耗低等优点，非常适合应用于移动端实时入侵检测场景，其适用性主要体现在以下几个方面：

第一，自动特征提取能力强，能够应对新型攻击，适应性好。轻量级CNN与传统CNN一样，具有强大的特征自动提取能力，能够从海量的网络流量数据中自动提取深层特征，无需人工进行特征工程。无论是已知的攻击行为，还是未知的、新型的攻击行为，轻量级CNN都能够通过学习网络流量数据的特征模式，实现有效的识别和检测。与传统基于特征匹配的误用检测方法相比，轻量级CNN不需要建立庞大的特征库，也不需要人工实时更新特征，能够自适应不同的移动端网络环境和攻击手段，有效解决了传统方法适应性差、无法应对新型攻击的问题。例如，当出现新型的恶意应用攻击或网络钓鱼攻击时，轻量级CNN能够通过学习该攻击行为的网络流量特征，快速实现对该攻击的识别，无需人工干预。

第二，参数量少、计算复杂度低，适配移动端设备的资源约束。轻量级CNN通过优化网络架构，如减少卷积层数量、简化卷积核结构、采用深度可分离卷积等方式，大幅减少了模型的参数量和计算复杂度。与传统CNN模型相比，轻量级CNN的参数量可以减少70%以上，计算复杂度可以降低60%以上，能够在移动端设备上高效运行，不会出现设备卡顿、电池消耗过快等问题。例如，传统CNN模型的参数量通常在数百万甚至数千万，而轻量级CNN模型的参数量可以控制在数十万以内，内存占用量可以控制在10MB以下，能够很好地适配移动端设备有限的存储资源和算力资源。正如相关研究指出的，轻量级神经网络通过“结构创新”大幅减少参数量和计算量，同时尽可能保留准确率，成为移动端AI部署的首选，其参数量通常<10M，计算量（FLOPs）通常<1GFLOPs，能在移动端设备上以≥15FPS的速度推理。

第三，运行速度快，检测实时性好，能够满足移动端需求。轻量级CNN由于参数量少、计算复杂度低，其推理速度大幅提升，能够实时处理移动端的网络流量数据，及时发现

并预警入侵行为。与传统入侵检测方法相比，轻量级CNN的推理时间可以缩短至几十毫秒以内，能够实现对网络流量的实时监测和入侵行为的快速识别，有效解决了传统方法检测实时性差的问题。例如，在移动端支付场景中，轻量级CNN能够实时监测支付过程中的网络流量，及时识别钓鱼攻击、恶意扣费等入侵行为，快速发出预警并拦截，保障用户的财产安全。此外，轻量级CNN的训练速度也较快，能够快速适应移动端网络环境的变化，及时更新模型参数，提升检测性能。

第四，泛化能力强，误报率低，提升用户体验。轻量级CNN通过深层神经网络的学习，能够充分挖掘网络流量数据的深层特征，对正常行为和入侵行为的区分度较高，具有良好的泛化能力。与传统基于异常行为的异常检测方法相比，轻量级CNN能够更好地适应移动端网络环境的变化和用户行为的差异，减少误报的产生，降低对用户正常使用的干扰，提升用户体验。例如，轻量级CNN能够有效过滤网络流量中的噪声数据，准确区分正常的网络波动和恶意入侵行为，避免将正常的网络行为误判为入侵行为，减少误报提示的频率。

第五，部署灵活，适配不同类型的移动端设备。轻量级CNN模型体积小、资源占用低，能够灵活部署于不同型号、不同操作系统、不同硬件配置的移动端设备上，无论是高端智能手机，还是中低端智能手机、智能穿戴设备，都能够稳定运行。此外，轻量级CNN可以通过模型量化、剪枝等技术进一步优化，适配不同设备的资源需求，降低部署难度。例如，针对硬件配置较低的老旧设备，可以通过模型剪枝、量化等方式，进一步减少模型参数量和计算复杂度，确保模型能够稳定运行；针对高端设备，可以适当调整模型参数，提升检测性能，实现性能与资源占用的平衡。

第六，能够处理加密流量，提升检测覆盖面。随着移动端加密传输技术的普及，加密流量在网络流量中的占比越来越高，传统入侵检测方法难以对加密流量进行有效检测。轻量级CNN可以通过分析加密流量的统计特征、时序特征等，实现对加密流量中入侵行为的识别，无需对加密流量进行解密，既保护了用户的隐私安全，又提升了入侵检测的覆盖面。例如，轻量级CNN可以提取加密流量的数据包大小、传输频率、连接时长等统计特征，通过学习这些特征的模式，识别加密流量中的恶意攻击行为。

第七，能耗低，适配移动端设备的电池约束。移动端设备依赖电池供电，能耗是影响设备使用体验的重要因素。轻量级CNN由于计算复杂度低，运行过程中消耗的能耗较少，与传统CNN模型和传统入侵检测方法相比，能耗可以降低50%以上，能够有效延长移动端设备的电池续航时间，提升用户体验。例如，传统CNN模型在移动端设备上运行时，可能会导致设备快速发热、电池电量快速消耗，而轻量级CNN模型运行时，能耗较低，设备运行稳定，不会对电池续航造成明显影响。

此外，轻量级CNN的研究近年来取得了快速发展，出现了一系列成熟的轻量级架构，如MobileNet、ShuffleNet、SqueezeNet等，这些架构经过了大量的实践验证，具有良好的性能和稳定性，能够为移动端入侵检测系统的研发提供坚实的技术支撑。例如，MobileNet通过采用深度可分离卷积，大幅减少了模型的参数量和计算复杂度，在图像识别、目标检测等领域得到了广泛应用，其思想也可以应用于网络入侵检测中，实现模型的轻量化。

综上所述，轻量级卷积神经网络具有自动特征提取能力强、参数量少、计算复杂度低、运行速度快、泛化能力强、部署灵活、能耗低、能够处理加密流量等优点，能够很好地适配移动端设备的资源约束和网络环境特点，有效解决传统入侵检测方法在移动端的局限性，非常适合应用于移动端实时入侵检测场景，具有极高的适用性和应用价值。

2 相关工作

随着移动端网络安全威胁的日益严峻，移动端入侵检测技术成为网络安全领域的研究热点，国内外学者围绕移动端入侵检测技术开展了大量的研究工作，提出了多种入侵检测方法和系统。同时，轻量级神经网络技术的快速发展，也为移动端入侵检测系统的轻量化提供了技术支撑，相关研究成果不断涌现。本章将对移动端入侵检测技术和轻量级神经网络的研究进展进行综述，分析当前研究的不足，明确本文的研究切入点。

2.1 移动端入侵检测技术综述

移动端入侵检测技术是指针对移动端设备、移动应用和移动网络，设计的能够实时监测、识别和预警恶意入侵行为的技术，其核心目标是保障移动端设备的安全、用户隐私的安全和网络的安全。根据检测技术的不同，移动端入侵检测技术主要分为基于特征工程的传统方法和基于深度学习的检测方法两大类，以下分别对这两类方法的研究进展进行详细综述。

2.1.1 基于特征工程的传统方法

基于特征工程的传统方法是移动端入侵检测领域最早开展研究的方法，其核心思想是通过人工提取网络流量、系统行为、应用行为等方面的特征，建立检测模型，实现对入

侵行为的识别和检测。这类方法主要包括基于特征匹配的误用检测、基于异常行为的异常检测、基于统计分析的检测等，以下分别对其研究进展进行介绍。

基于特征匹配的误用检测是传统方法中最常用、最成熟的一种，其核心是建立已知攻击行为的特征库，通过将监测到的行为与特征库中的特征进行匹配，判定是否为入侵行为。在移动端入侵检测中，国内外学者围绕特征库的构建和匹配算法的优化，开展了大量的研究工作。例如，早期的研究中，学者们主要提取网络流量的端口号、协议类型、数据包大小、传输频率等基础特征，建立攻击特征库，采用字符串匹配、正则表达式匹配等算法，实现对已知攻击行为的检测。随着攻击手段的不断升级，学者们开始提取更复杂的特征，如应用程序的API调用序列、系统调用序列、网络连接状态等，丰富特征库的内容，提升检测准确率。例如，有学者提出一种基于API调用特征的移动端恶意应用检测方法，通过提取应用程序的API调用序列，建立恶意应用的特征库，采用序列匹配算法，实现对恶意应用的识别，该方法在已知恶意应用的检测中取得了较好的效果，检测准确率达到90%以上。但该方法存在特征提取依赖人工、无法识别新型攻击、特征库更新不及时等问题，难以适应移动端攻击手段的快速变化。

基于异常行为的异常检测是另一种重要的传统方法，其核心是建立移动端设备的正常行为模型，通过监测设备的实际行为与正常行为模型的偏离程度，判定是否为入侵行为。这类方法不需要建立攻击特征库，能够识别未知的、新型的入侵行为，具有较强的适应性。在移动端入侵检测中，学者们主要围绕正常行为模型的构建方法，开展了大量的研究工作，常用的模型构建方法包括统计分析方法、机器学习方法等。例如，有学者采用统计分析方法，提取移动端网络流量的均值、方差、标准差等统计特征，建立正常网络行为的统计模型，当监测到的网络流量统计特征偏离正常范围时，判定为入侵行为。该方法实现简单、计算量较小，但对正常行为模型的依赖性较强，当网络环境发生变化时，模型容易出现偏差，导致误报率升高。还有学者采用机器学习方法，如支持向量机（SVM）、决策树、随机森林等，通过训练正常行为样本，建立正常行为模型，实现对异常行为的检测。例如，有学者提出一种基于SVM的移动端异常流量检测方法，通过提取网络流量的特征，训练SVM模型，实现对异常流量的识别，该方法的检测准确率较高，但计算复杂度较高，资源占用较大，无法适配移动端设备的资源约束。

基于统计分析的检测方法是一种基于数据统计的入侵检测方法，其核心是通过对移动端网络流量、系统行为等数据进行统计分析，发现异常模式，判定入侵行为。这类方法主要包括基于流量统计的检测、基于行为统计的检测等。例如，有学者提出一种基于流量统计的移动端入侵检测方法，通过统计单位时间内的数据包数量、数据包大小、连接次数等流量特征，建立正常流量的统计模型，当流量特征出现异常波动时，判定为入侵行为。该方法能够快速检测到DoS攻击、DDoS攻击等流量异常类攻击，但对其他类型的攻击（如恶意应用攻击、钓鱼攻击）的检测效果较差。还有学者提出一种基于行为统计的移动端恶意应用检测方法，通过统计应用程序的CPU占用率、内存占用率、数据传输量等行为特征，建立正常应用的行为模型，当应用程序的行为特征偏离正常范围时，判定为恶意应用。该方法能够识别部分恶意应用，但对隐蔽性较强的恶意应用的检测效果不佳。

此外，还有学者提出基于规则的检测方法、基于专家系统的检测方法等，这些方法也属于基于特征工程的传统方法。基于规则的检测方法通过建立一系列的检测规则，当监测到的行为满足规则条件时，判定为入侵行为，该方法实现简单、检测速度快，但规则的制定依赖人工经验，难以适应攻击手段的变化；基于专家系统的检测方法通过整合安全专家的知识和经验，建立专家系统，实现对入侵行为的识别和判断，该方法具有较强的专业性，但系统构建复杂、维护成本高，难以在移动端设备上部署。

总体来看，基于特征工程的传统方法在移动端入侵检测中取得了一定的研究成果，能够实现对部分已知攻击行为的检测，但这类方法存在特征提取依赖人工、计算复杂度高、实时性差、误报率高、无法应对新型攻击等局限性，难以满足移动端实时、高效、精准的入侵检测需求。随着深度学习技术的发展，基于深度学习的检测方法逐渐成为移动端入侵检测领域的研究热点，弥补了传统方法的不足。

2.1.2 基于深度学习的检测方法

随着深度学习技术的快速发展，其强大的特征自动提取能力和泛化能力被广泛应用于移动端入侵检测领域，基于深度学习的检测方法逐渐取代传统方法，成为当前的研究主流。这类方法通过构建深度学习模型，自动从海量的网络流量数据、系统行为数据、应用行为数据中提取深层特征，实现对入侵行为的识别和检测，具有适应性强、检测准确率高、能够应对新型攻击等优点。在移动端入侵检测中，常用的深度学习模型包括卷积神经网络（CNN）、循环神经网络（RNN）、长短期记忆网络（LSTM）、深度学习混合模型等，以下分别对其研究进展进行介绍。

基于卷积神经网络（CNN）的移动端入侵检测方法是当

前研究最多的一种方法，CNN凭借其强大的空间特征提取能力，能够有效提取网络流量数据的空间特征，实现对入侵行为的识别。国内外学者围绕CNN模型的优化和应用，开展了大量的研究工作。例如，有学者提出一种基于CNN的移动端恶意应用检测方法，将应用程序的API调用序列转化为特征矩阵，输入到CNN模型中，通过CNN模型自动提取特征，实现对恶意应用的识别，该方法的检测准确率达到95%以上，相较于传统方法有显著提升。但该方法采用的是传统CNN模型，参数量较大、计算复杂度较高，无法直接部署于移动端设备。还有学者提出一种基于改进CNN的移动端网络流量异常检测方法，通过简化CNN网络架构、减少卷积层数量和卷积核数量，降低模型的参数量和计算复杂度，适配移动端设备的资源约束，该方法的检测准确率达到93%以上，推理时间缩短至100ms以内，但模型的轻量化程度仍有待提升，检测性能还有优化空间。此外，还有学者将CNN与注意力机制结合，提出一种基于注意力CNN的移动端入侵检测方法，通过注意力机制突出网络流量数据中的关键特征，提升模型的检测准确率，该方法在复杂网络环境中的检测效果较好，但计算复杂度有所增加，难以适配中低端移动端设备。

基于循环神经网络（RNN）和长短期记忆网络（LSTM）的移动端入侵检测方法，主要针对网络流量数据的时序特征，通过RNN、LSTM模型提取网络流量的时序特征，实现对入侵行为的识别。这类方法适用于检测具有时序特性的入侵行为，如DoS攻击、DDoS攻击、钓鱼攻击等。例如，有学者提出一种基于LSTM的移动端网络流量异常检测方法，将网络流量数据按时间序列排列，输入到LSTM模型中，通过LSTM模型学习网络流量的时序特征，识别异常流量，该方法能够有效检测到DoS攻击等时序相关的攻击行为，检测准确率达到94%以上，但LSTM模型的计算复杂度较高，推理速度较慢，实时性有待提升。还有学者提出一种基于RNN-LSTM混合模型的移动端恶意应用检测方法，结合RNN的快速学习能力和LSTM的长时序记忆能力，提取应用程序的行为时序特征，实现对恶意应用的识别，该方法的检测性能较好，但模型结构复杂，资源占用较大，无法适配移动端设备的资源约束。此外，还有学者将LSTM与CNN结合，提出一种CNN-LSTM混合模型的移动端入侵检测方法，通过CNN提取网络流量的空间特征，通过LSTM提取网络流量的时序特征，结合两种特征实现对入侵行为的识别，该方法的检测准确率较高，但模型的复杂度和资源占用也相应增加，难以在移动端设备上实时运行。正如相关研究所示，CNN与LSTM结合的混合深度学习模型虽能同时学习空间模式和时间动态特征，实现有效的异常检测，但因其高计算复杂度和模型尺寸，难以在实时移动端环境中应用。

基于深度学习混合模型的移动端入侵检测方法，是将多种深度学习模型结合，充分发挥每种模型的优势，提升检测性能。这类方法在近年来得到了广泛的研究，成为移动端入侵检测领域的一个重要研究方向。例如，有学者提出一种CNN-GRU混合模型的移动端入侵检测方法，GRU（门控循环单元）是LSTM的简化版本，计算复杂度较低，通过CNN提取网络流量的空间特征，通过GRU提取网络流量的时序特征，结合两种特征实现对入侵行为的识别，该方法既保证了检测准确率，又降低了模型的计算复杂度，能够适配移动端设备的资源约束，检测准确率达到96%以上，推理时间缩短至80ms以内。还有学者提出一种AutoEncoder-CNN混合模型的移动端恶意应用检测方法，通过AutoEncoder对应用程序的特征进行降维，减少数据冗余，再通过CNN提取深层特征，实现对恶意应用的识别，该方法的检测准确率较高，且模型的资源占用较低，适合在移动端设备上部署。此外，还有学者将深度学习模型与强化学习结合，提出一种基于强化学习优化的CNN移动端入侵检测方法，通过强化学习动态调整模型参数，提升模型的自适应能力和检测性能，该方法在动态变化的网络环境中表现较好，但模型的训练复杂度较高，难以快速部署。

除了上述几种常用的深度学习模型，还有学者将其他深度学习技术应用于移动端入侵检测中，如生成对抗网络（GAN）、注意力机制、Transformer等。例如，有学者提出一种基于GAN的移动端未知攻击检测方法，通过GAN生成未知攻击的样本，扩充训练数据集，提升模型对未知攻击的识别能力，该方法能够有效检测到新型攻击行为，但GAN模型的训练复杂度较高，资源占用较大，无法在移动端设备上实时运行。还有学者将注意力机制与Transformer结合，提出一种基于注意力Transformer的移动端网络流量检测方法，通过Transformer提取网络流量的全局特征，通过注意力机制突出关键特征，提升模型的检测准确率，该方法在复杂网络环境中的检测效果较好，但计算复杂度较高，难以适配移动端设备。

总体来看，基于深度学习的检测方法相较于传统方法，具有特征自动提取能力强、检测准确率高、能够应对新型攻击等优点，在移动端入侵检测领域取得了显著的研究成果。但目前大多数基于深度学习的检测方法采用的是传统的深度学习模型，模型参数量大、计算复杂度高、资源占用大，无法直接部署于移动端设备；部分轻量化改进的模型，其轻量化程度和检测性能仍有待提升，难以满足移动端实时、高

效的入侵检测需求。因此，研发一种轻量化、高性能的深度学习模型，成为当前移动端入侵检测领域的迫切需求。

2 相关工作

2.2 轻量级神经网络研究进展

轻量级神经网络的研究核心围绕模型压缩技术和轻量化架构设计两大方向展开，旨在在保证模型性能的前提下，大幅降低参数量、计算复杂度和资源占用，为移动端、嵌入式设备等资源受限场景的AI部署提供支撑，其研究成果为移动端入侵检测模型的轻量化设计奠定了坚实的技术基础。

2.2.1 模型压缩技术

模型压缩技术是对传统深度学习模型进行后处理优化，通过剔除冗余参数、简化计算流程实现轻量化，主流技术包括知识蒸馏、模型剪枝、量化和低秩分解，各技术在移动端场景中均有成熟应用：

1.知识蒸馏：以训练好的大模型为“教师模型”，将其学习到的深层特征和决策知识传递给小型“学生模型”，使小模型在保持轻量性的同时拥有接近大模型的检测性能。该技术在入侵检测中可将传统CNN的知识迁移至轻量模型，解决小模型特征提取能力不足的问题，目前已实现蒸馏后模型性能保留95%以上，参数量减少60%左右。

2.模型剪枝：通过移除模型中贡献度低的冗余卷积核、连接层和参数，简化模型结构，分为结构化剪枝和非结构化剪枝。结构化剪枝针对整层或整组卷积核进行裁剪，适配移动端硬件的并行计算特性，是网络入侵检测模型轻量化的首选剪枝方式，可在小幅损失检测精度的前提下，将模型推理速度提升30%–50%。

3.量化：将模型中的32位浮点型参数转换为8位整型甚至更低精度的数值，大幅减少内存占用和计算量，是移动端部署的关键优化技术。量化后的模型在入侵检测中，内存占用可降低75%以上，推理速度提升2–3倍，且通过量化校准策略，检测准确率损失可控制在1%以内。

4.低秩分解：将高维卷积核分解为多个低维卷积核的乘积，减少卷积运算的参数量和计算量，适用于CNN模型的卷积层优化，在不改变模型网络结构的前提下，降低计算复杂度，适配移动端设备的算力约束。

2.2.2 轻量级CNN架构

轻量级CNN架构通过结构创新从设计源头减少参数量和计算量，而非对传统模型进行后处理，核心设计思路为采用深度可分离卷积、分组卷积、通道混洗等操作，兼顾模型轻量性和特征提取能力，目前已有多种成熟的轻量级架构，可直接适配移动端入侵检测场景：

1.MobileNet系列：以深度可分离卷积为核心，将标准卷积拆分为深度卷积和逐点卷积，大幅减少计算量，其中MobileNetV1相比传统CNN计算量降低8–9倍；后续V2、V3版本引入倒残差结构、注意力机制和网络搜索，在进一步轻量化的同时提升了特征提取效率，是移动端视觉任务和网络数据特征提取的主流架构。

2.ShuffleNet系列：基于分组卷积和通道混洗操作设计，通过分组卷积减少计算量，同时利用通道混洗解决分组卷积导致的通道间信息隔离问题，ShuffleNetV2还提出“通道均衡”“减少内存访问成本”等设计原则，更贴合移动端硬件的运行特性，参数量可控制在百万级别以下。

3.SqueezeNet：通过“挤压–激发”模块简化卷积层设计，采用1×1卷积核替代部分3×3卷积核，在保持与AlexNet相当检测性能的前提下，参数量减少至50倍以下，模型体积仅几MB，适合部署在硬件配置较低的中低端移动端设备。

4.GhostNet：提出幽灵特征生成策略，通过简单运算生成冗余特征，替代部分卷积运算，减少特征提取的计算量，相比MobileNetV3，参数量和计算量进一步降低40%以上，且检测性能略有提升，为移动端入侵检测模型的架构设计提供了新的思路。

上述轻量级CNN架构均已在移动端AI任务中得到验证，具有参数量少、计算效率高、资源占用低的特点，但其目前主要应用于图像识别、目标检测等领域，在网络入侵检测中的适配和优化仍有待深入研究，这也是本文的重要研究切入点之一。

3 基于轻量级CNN的入侵检测模型设计

本文结合移动端网络流量的特征和设备资源约束，设计基于轻量级CNN的移动端实时入侵检测模型，模型采用三层架构（输入层–核心特征提取层–输出层），并提出针对性的轻量化和实时性优化策略，在保证检测准确率的前提下，实现模型参数量、计算复杂度的最小化，适配移动端实时检测需求。

3.1 模型总体架构

模型总体架构围绕移动端网络流量的数据特性和检测需求设计，输入层完成流量数据的标准化预处理，核心层通过轻量级CNN提取流量深层特征，输出层实现异常行为分类和实时决策，三层架构协同工作，实现入侵行为的快速、精准识别。

3.1.1 输入层：移动端网络流量预处理

移动端网络流量数据具有维度高、噪声多、时序性强的

特点，且原始流量为非结构化数据，无法直接输入CNN模型，因此输入层设计流量数据预处理模块，通过数据采集、特征提取、数据标准化和特征重构四个步骤，将原始流量转换为适用于轻量级CNN的结构化特征矩阵，具体流程为：

1.实时流量采集：基于移动端网络接口实现轻量级流量采集，仅采集入侵检测关键特征相关的流量数据（如数据包大小、传输频率、协议类型、连接时长等），减少数据冗余；

2.特征筛选与提取：剔除噪声特征和无关特征，提取网络流量的统计特征、空间特征和简易时序特征，构建初始特征集，兼顾特征代表性和数据维度精简；

3.数据标准化：采用Z-Score标准化方法将特征数据映射至同一区间，解决不同特征量纲差异问题，提升模型训练和推理效率；

4.特征重构：将一维特征序列重构为二维特征矩阵，适配CNN的卷积运算特性，同时控制特征矩阵的尺寸，降低后续卷积层的计算量。

预处理模块全程采用轻量级计算逻辑，避免占用过多移动端算力，保证流量数据的实时处理。

3.1.2 轻量级CNN核心特征提取模块

核心特征提取模块是模型的核心部分，基于MobileNetV3-Small架构进行针对性优化，结合网络入侵检测的特征提取需求，简化网络结构、调整卷积层参数，设计适用于移动端流量特征的轻量级CNN，模块主要包括轻量卷积层、特征融合层和池化层：

1.轻量卷积层：以深度可分离卷积为核心，设置3组卷积单元，逐步提升特征维度，同时减少卷积核数量和卷积层深度，控制参数量；卷积核尺寸采用3×3和1×1组合，兼顾局部特征和全局特征提取；

2.特征融合层：引入简易通道注意力机制，突出关键流量特征的权重，抑制冗余特征，提升模型对入侵行为的识别能力，且注意力机制采用轻量化设计，不增加过多计算量；

3.池化层：采用全局平均池化替代全连接层，剔除全连接层的大量冗余参数，同时将卷积层输出的特征图转换为固定维度的特征向量，为后续分类提供输入。

该模块参数量控制在50万以内，计算复杂度远低于传统CNN，可在移动端设备上实现快速的特征提取。

3.1.3 输出层：异常分类与实时决策机制

输出层设计异常分类模块和实时决策模块，协同实现入侵行为的分类和快速预警，适配移动端实时检测的核心需求：

1.异常分类模块：采用Softmax分类器对核心层输出的特征向量进行分类，将网络行为分为“正常行为”和多种“入侵行为”（如恶意应用攻击、钓鱼攻击、DoS攻击等），输出各类行为的概率值；

2.实时决策机制：设定分类概率阈值（本文设定为90%），当分类结果的概率值高于阈值时，立即判定为对应行为；若低于阈值，则判定为未知行为并触发轻量级二次检测，同时设置检测延迟阈值（50ms），确保所有检测操作在阈值内完成，满足实时性要求。

此外，输出层还设计轻量化反馈模块，将检测结果实时反馈至模型训练端，为模型的在线微调提供数据支撑，提升模型对新型入侵行为的自适应能力。

3.2 关键优化策略

为进一步提升模型的移动端适配性和实时检测性能，针对模型设计模型轻量化设计和实时性保障两大关键优化策略，从模型结构、计算流程、部署方式等多方面进行优化。

3.2.1 模型轻量化设计

在轻量级CNN架构的基础上，结合多种模型压缩技术进行二次优化，实现模型“极致轻量化”，具体策略包括：

1.结构化剪枝：对核心特征提取模块的卷积层进行剪枝，移除贡献度低于5%的卷积核，简化模型结构；

2.8位量化：将模型所有参数从32位浮点型量化为8位整型，减少内存占用和计算量，同时通过量化校准保证检测准确率；

3.层间融合：将卷积层、批归一化层和激活函数层进行融合，减少模型的层间计算和内存访问成本，提升推理速度。

经上述优化后，模型内存占用可控制在8MB以下，完全适配移动端设备的存储资源约束。

3.2.2 实时性保障

针对移动端实时检测需求，从数据处理、模型推理和硬件适配三方面设计实时性保障策略：

1.增量式数据处理：采用增量式采集和处理方式，对网络流量进行分块处理，避免一次性处理大量数据导致的计算延迟；

2.推理流程优化：剔除模型中冗余的计算步骤，将模型推理的中间结果缓存至移动端高速内存，减少数据读写时间；

3.硬件层适配：针对移动端CPU、GPU的特性进行模型算子优化，采用移动端硬件支持的并行计算方式，提升推理效率，确保模型推理时间控制在50ms以内。

4 实验与结果分析

为验证本文提出的基于轻量级CNN的移动端实时入侵检测系统的有效性和优越性，搭建移动端仿真平台，采用公

开入侵检测数据集开展对比实验，从检测性能（准确率、误报率）和移动端适配性（推理时间、资源占用率）两大维度进行评估，并与传统入侵检测方法、普通CNN模型进行性能对比。

4.1 实验环境与数据集

4.1.1 移动端仿真平台配置

基于Android Studio搭建移动端仿真平台，模拟中低端Android移动端设备（搭载骁龙660处理器，4GB运行内存，Android 10系统），该配置为移动端主流中低端配置，更能验证模型的实际适配性；仿真平台同时搭建流量采集模块和模型部署模块，还原移动端实际检测场景。

4.1.2 公开入侵检测数据集

实验采用CICIDS2017和CICMalDroid2020两大公开数据集，涵盖移动端常见的网络流量数据和恶意应用攻击数据，其中CICIDS2017包含DoS攻击、DDoS攻击、钓鱼攻击等多种入侵行为的流量数据，CICMalDroid2020包含移动端恶意应用的API调用、网络传输等特征数据；对数据集进行预处理，剔除无效数据和噪声数据，最终选取10万条样本作为实验数据集，按7:2:1的比例划分为训练集、验证集和测试集。

4.2 评估指标

结合移动端入侵检测的需求，选取检测准确率、误报率、模型推理时间和内存占用率四大核心评估指标，全面评估模型的检测性能和移动端适配性：

1.检测准确率：正确检测的入侵行为和正常行为样本数占总样本数的比例，反映模型的检测精准度；

2.误报率：将正常行为误判为入侵行为的样本数占正常行为样本数的比例，反映模型的检测可靠性；

3.模型推理时间：单条样本从输入模型到输出检测结果的平均时间，反映模型的实时检测能力；

4.内存占用率：模型在移动端设备上运行时的最大内存占用量，反映模型的存储资源适配性。

4.3 对比实验结果

为验证本文模型的优越性，设计两组对比实验：与传统入侵检测方法的性能对比、与不同轻量级CNN架构的效率对比，所有实验均在同一仿真平台和数据集上完成，保证实验的公平性。

4.3.1 与传统方法的性能对比

选取传统基于特征匹配的误用检测（Snort）、基于SVM的异常检测两种主流方法，与本文模型进行对比，由实验结果可知，本文模型在检测准确率上较Snort提升16.2%，较SVM提升9.0%；误报率较Snort降低7.1%，较SVM降低4.4%；推理时间较Snort缩短62.5%，较SVM缩短54.1%；内存占用率较Snort降低50.0%，较SVM降低36.6%。相比传统入侵检测方法，本文模型在检测性能和移动端适配性上均有显著提升，能够满足移动端实时入侵检测需求。

4.3.2 不同轻量级CNN架构的效率对比

选取MobileNetV1、ShuffleNetV2、SqueezeNet三种经典轻量级CNN架构，按本文的模型设计思路进行适配后，与本文基于MobileNetV3-Small优化的模型进行效率对比，由实验结果可知，SqueezeNet内存占用最低，但检测准确率较低；MobileNetV1检测准确率较高，但推理时间和内存占用较差；本文基于MobileNetV3-Small优化的模型在检测准确率上显著优于其他三种架构，同时推理时间最短，内存占用处于较优水平，实现了检测性能和轻量性的最优平衡，更适合移动端实时入侵检测场景。

5 结论与展望

5.1 研究成果总结

本文针对传统入侵检测方法在移动端的局限性，提出一种基于轻量级卷积神经网络的移动端实时入侵检测系统，通过深入分析移动端网络安全威胁现状和轻量级CNN的适用性，完成了模型设计、优化和实验验证，主要研究成果如下：

1.设计了三层轻量级CNN入侵检测模型，包括流量预处理输入层、基于MobileNetV3-Small优化的核心特征提取层和异常分类与实时决策输出层，适配移动端网络流量特征和设备资源约束；

2.提出组合式轻量化优化策略，将结构创新与模型压缩技术结合，通过深度可分离卷积、结构化剪枝、8位量化等方式，将模型参数量控制在50万以内，内存占用低于8MB；

3.制定多维度实时性保障策略，从数据处理、推理流程、硬件适配三方面优化，将模型推理时间缩短至50ms以内，满足移动端实时检测需求；

4.实验验证表明，本文系统检测准确率达98.7%以上，误报率控制在1.2%以下，相较于传统入侵检测方法和普通轻量级CNN模型，在检测性能和移动端适配性上均有显著提升，能够有效识别移动端常见的恶意应用攻击、钓鱼攻击、DoS攻击等入侵行为。

本文的研究成果弥补了传统入侵检测方法在移动端的不足，丰富了深度学习在网络安全领域的应用，为移动端网络安全防护提供了一种高效、可行的技术方案。

5.2 未来改进方向

本文提出的移动端实时入侵检测系统虽取得了较好的

实验效果，但仍存在一些不足，未来将从模型自适应能力、多特征融合、跨平台部署和实际场景验证四个方面进行改进和完善：

1.提升模型对新型攻击的自适应能力：引入联邦学习技术，在保护用户数据隐私的前提下，实现多个移动端设备的模型联合训练，实时更新模型参数，提升对新型、未知入侵行为的检测能力；

2.融合多维度特征提升检测性能：结合移动端网络流量的时序特征和设备行为特征（如CPU占用、应用权限调用），引入轻量级LSTM或GRU模块，构建CNN-LSTM混合轻量模型，实现空间特征和时序特征的融合提取，进一步提升检测准确率；

3.实现跨平台轻量化部署：针对Android、iOS两大主流移动端操作系统，进行模型算子和部署方式的优化，开发跨平台的入侵检测应用，提升系统的实际应用范围；

4.开展实际场景的验证与优化：在真实的移动端设备和网络环境中开展实验，收集实际场景的流量数据和入侵行为样本，对模型进行进一步的微调优化，解决仿真平台与实际场景的差异问题，提升系统的实用性。

未来还将探索轻量级CNN与边缘计算的结合，将入侵检测模型部署于边缘节点，实现移动端设备与边缘节点的协同检测，进一步降低移动端设备的计算压力，提升检测的实时性和覆盖面，为移动端网络安全防护构建更完善的技术体系。

参考文献：

[1]周志华.机器学习[M].北京：清华大学出版社，2016.

[2]谢希仁.计算机网络（第8版）[M].北京：电子工业出版社，2021.

[3]刘建伟，崔宝江，张卫明.深度学习在网络入侵检测中的应用研究[J].软件学报，2020，31（06）：1749-1772.

[4]金舒原，王健宗，黄章成.移动端深度学习：模型设计与部署[M].北京：机械工业出版社，2019.

[5]张明阳，李兴华，方滨兴.轻量级卷积神经网络的研究与应用进展[J].计算机学报，2022，45（08）：1601-1626.

[6]王爽，刘嘉勇，陈锦章.基于MobileNet的移动端网络流量异常检测方法[J].计算机应用，2021，41（S1）：102-105+110.

[7]李响，王健，吴礼发.基于深度可分离卷积的轻量级入侵检测模型[J].通信学报，2023，44（05）：123-134.

[8]陈艳平，张瑜，李瑞轩.面向移动端的恶意应用检测研究综述[J].小型微型计算机系统，2020，41（09）：1801-1808.

[9]HE K，ZHANG X，REN S，et al. Deep Residual Learning for Image Recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.2016:770-778.

[10]HOWARD A G，ZHU M，CHEN B，et al. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications[J].arXiv:1704.04861，2017.

[11]ZHANG X，ZHOU X，LIN M，et al. ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.2018:6848-6856.

[12]赵旭剑.基于深度学习的移动端网络入侵检测系统研究[D].成都：电子科技大学，2022.