中北大学学报(自然科学版)
JOURNAL OF NORTH UNIVERSITY OF CHINA(NATURAL SCIENCE EDITION)
2026年 第47卷 第2期(总第226期)
Vol. 47 No. 2 2026(Sum No. 226)

文章编号：1673-3193(2026)02-0215-11 J. North Univ. China, Nat. Sci. Ed., 2026, 47(2): 215-225.

基于GADF与MobileViT-RMF的网络入侵检测模型

刘泽晋[1,3]，赵利辉[1*]，刘诗涵[1]，董和磊[2,3]

(1. 中北大学 软件学院，山西 太原 030051；2. 中北大学 半导体与物理学院，山西 太原 030051；
3. 中北大学 微纳器件与系统教育部重点实验室，山西 太原 030051)

摘　要：针对现有基于流量图像化的网络入侵检测方法存在时序关联表达不足、特征表征能力有限以及轻量化模型检测精度受限等问题，提出一种融合格拉姆角差场(Gramian Angular Difference Field，GADF)与残差多粒度聚焦模块(Residual Multi-grain Focus，RMF)的轻量化网络入侵检测模型MobileViT-RMF。该方法首先将网络流量的一维特征序列映射为具有时空关联性的二维GADF图像，以增强流量演化信息和类别差异特征的表达能力；随后在MobileViT轻量化骨干网络中嵌入RMF模块，通过多粒度特征聚焦机制强化局部关键信息提取与全局上下文建模，从而提升模型对复杂流量图像的判别能力。围绕CICIDS2017和UNSW-NB15两个公开数据集开展了实验验证。实验结果显示，所提模型在CICIDS2017数据集上的检测准确率达到99.75%，较基线模型MobileViT_V3_xxs提高2.74百分点；在UNSW-NB15数据集上的峰值准确率达到99.89%，体现出较好的跨数据集检测性能。研究结果表明，将GADF流量图像表示与RMF特征增强机制相结合，能够在保持模型轻量化特征的同时，提高网络流量图像的表征质量和分类性能，可为轻量化网络入侵检测模型的设计提供参考。

关键词：入侵检测；轻量化网络；图像分类；MobileViT

中图分类号：TP393.08　　**文献标识码：**A　　**doi:** 10.62756/jnuc.issn.1673-3193.2025.06.0007

引用格式：刘泽晋，赵利辉，刘诗涵，等. 基于GADF与MobileViT-RMF的网络入侵检测模型[J]. 中北大学学报(自然科学版)，2026，47(2)：215-225.

LIU Zejin，ZHAO Lihui，LIU Shihan，et al. Network intrusion detection model based on GADF and MobileViT-RMF[J]. Journal of North University of China(Natural Science Edition)，2026，47(2)：215-225.

Network Intrusion Detection Model Based on GADF and MobileViT-RMF

LIU Zejin[1,3]，ZHAO Lihui[1*]，LIU Shihan[1]，DONG Helei[2,3]

(1. School of Software，North University of China，Taiyuan 030051，China；
2. School of Semiconductor and Physics，North University of China，Taiyuan 030051，China；
3. Key Laboratory of Micro-Nano Devices and Systems，Ministry of Education，
North University of China，Taiyuan 030051，China)

Abstract: A lightweight network intrusion detection model，termed MobileViT-RMF，was proposed to address several limitations in existing traffic-visualization-based intrusion detection methods，including insufficient representation of temporal correlations in traffic data，limited feature characterization capability，and restricted

收稿日期：2025-06-12
基金项目：国家自然科学基金面上项目(52178825)
作者简介：刘泽晋(2000—)，男，硕士生，主要从事网络测试与网络安全方面的研究。
***通信作者：**赵利辉(1979—)，男，副教授，博士，主要从事网络测试与网络安全方向的研究。E-mail：leehwi@nuc.edu.cn。

detection accuracy under lightweight model constraints. The proposed method combined the Gramian Angular Difference Field (GADF) with a Residual Multi-grain Focus (RMF) module to improve both traffic representation and feature extraction. Firstly, one-dimensional network traffic feature sequences were converted into two-dimensional GADF images with spatiotemporal correlations, so that the evolutionary characteristics of traffic behavior and the discriminative differences among traffic categories could be more effectively preserved in the image space. On this basis, the RMF module was embedded into the lightweight MobileViT backbone to strengthen local key-feature extraction and global contextual modeling through a multi-granularity feature focusing mechanism, thereby the discriminative capability of the model for complex traffic images was enhanced. Experimental validation was conducted on the CICIDS2017 and UNSW-NB15 public datasets. The proposed model achieves a detection accuracy of 99.75% on CICIDS2017, which is 2.74 percentage points higher than that of the baseline MobileViT_V3_xxs model. On UNSW-NB15, the peak accuracy reaches 99.89%, indicating favorable cross-dataset detection performance. These findings suggest that the combination of GADF-based traffic image representation and RMF-based feature enhancement can improve the representation quality and classification performance of network traffic images while preserving the lightweight nature of the model. The proposed approach provides a feasible idea for the design of lightweight network intrusion detection models.

Key words: intrusion detection; lightweight network; image classification; MobileViT

0 引 言

随着5G与物联网技术的规模化部署，全球联网设备数量已突破289亿台[1]，根据Nexusguard[2]《2024年全球分布式拒绝服务攻击趋势白皮书》，2023年全球DDoS攻击峰值达到700 Gbit/s流量吞吐量，并发冲击突破8×10^7数据包/s(pps)，较2021年最高纪录(580 Gbit/s)激增20.7%。与此同时，加密协议的普及使网络流量加密率超过83%[3]，传统基于人工特征工程的入侵检测系统(Intrusion Detection System，IDS)在密文环境下逐渐失效。面对海量高维流量数据与未知攻击模式的挑战，研究者开始探索基于图像化与深度学习的检测范式，通过将网络流量转换为二维图像，并利用卷积神经网络(CNN)的视觉特征提取能力实现高效分类[4]。

在基于图像的入侵检测技术演进中，网络流量转换图像的方法与深度检测模型架构是两大核心研究领域[5]。在图像转换层面，现有方法可分为两类：1) 基于原始字节的视觉重构方法，Demmese等[6]直接将字节流重塑为二维图像以保留完整信息，但生成的噪声模式显著降低了模型收敛效率；2) 基于统计特征的映射技术，如Pham等[7]通过线性编码流量特征(持续时间、字节数等)生成灰度图像，虽简化了特征工程，但丢失了数据包级的细粒度信息；Attack等[8]进一步引入协议标志位的RGB通道映射，虽增强协议类型区分度，却因人工特征选择导致跨场景泛化能力受到限制。上述方法普遍面临时空信息割裂与视觉模式模糊的双重困境，制约了后续检测模型的性能上限。

在检测模型层面，研究者尝试通过不同架构平衡精度与效率：刘等[9]采用VGG16提取局部空间特征，但其庞大的参数量导致推理延迟激增。为融合时空信息，Wu等[10]提出了CNN-GRU混合模型，通过GRU层捕获流量时序特征，提升了DDoS检测的$F1$值，但模型结构更为复杂，参数量和计算开销比单独使用CNN更大，进一步加剧了部署和效率方面的挑战。

Wang等[11]和Avci等[12]开始将ViT(Vision Transformer)应用于该领域，其全局注意力机制虽能对长程依赖进行建模，但由于参数量过大而导致训练效率低下。认识到效率的重要性，姚等[13]使用MobileViT来处理流量图像，显著降低了模型的参数量和计算复杂度，提高了部署在资源受限设备上的可行性。但其没有针对网络流量图像的独特性质(特定纹理模式)进行优化，性能仍有提升空间。由此可见，现有模型尚未解决全局-局部特征协同与计算-精度均衡优化的关键问题。

针对上述问题，本研究提出融合格拉姆角差场(Gramian Angular Difference Field，GADF)编码与改进MobileViT的轻量化入侵检测框架。主要贡献包括：

1）设计基于GADF的流量图像转换方法，通过时序编码保留流量数据的时空关联性，解决传统线性映射导致的特征信息丢失的问题；

2）构建残差多粒度聚焦RMF模块（Residual Multi-granularity Fusion Block），集成自适应大核注意力ALKA、门控空间注意力SGA和尾部大核注意力SMA机制，实现跨通道特征依赖建模与关键区域定位优化。

1　问题的提出与解决

1.1　一维数据图像二维化

传统方法通过线性映射将流量特征映射为灰度像素值，虽可实现数据可视化，但存在特征空间坍缩问题。相邻时间窗的流量突变特征被离散化截断，特征向量间的非线性关系无法通过灰度梯度表达，流量抖动直接映射为像素值波动，形成了椒盐噪声等问题，使得原始流量特征的时序依赖性与空间拓扑关系在转换过程中被破坏。格拉姆角差场GADF[14]技术为时序数据特征提取提供了新思路。GADF通过极坐标编码重构时空关联性，其核心优势体现在以下三方面：

1）角度不变性降噪：GADF将一维流量序列映射至极坐标系，通过计算相邻时间窗的余弦相似度生成二维矩阵。该过程利用角度对幅值波动的低敏感性，有效抑制了流量抖动产生的脉冲噪声。

2）时空特征耦合：传统灰度图像仅保留特征幅值信息，而GADF则通过极径-角度双通道编码，同时捕获了流量特征的时间演化趋势与局部突变模式。这种耦合机制使CNN卷积核能同时提取时序动态与空间分布特征。

3）类间可分性增强：GADF变换后，不同类别流量样本在图像纹理分布和结构形态上表现出一定差异，有助于提升类别间的可区分性。对于DoS攻击流量，其短时突发变化、统计特征联动关系等信息能够在图像中以局部纹理或整体结构的形式体现出来，从而为后续模型识别异常模式提供了更直观的输入表达。

对网络流量数据的数值特征进行分段处理，通过极坐标转换构建角度矩阵，该过程为

$$\varphi_i = \arccos(x_i), \tag{1}$$

$$GADF = \left[\sin(\varphi_i - \varphi_j)\right], \tag{2}$$

式中：φ_i为第i个角度值；$\arccos(x_i)$为反余弦函数，输入范围[−1, 1]，输出对应角度(0～π)，x_i为第i个输入值。$\sin(\varphi_i - \varphi_j)$用于计算第$i$个与第$j$个角度的差值正弦，结果在[−1, 1]之间，[]表示矩阵，内部由正弦值填充。

然后计算不同时间窗口内特征向量的余弦相似度，最终生成具有时空关联性的二维图像数据。

1.2　模型改进

网络流量经GADF转换后形成具有时空关联性的二维图像：协议类型差异映射为不同方向的纹理模式；流量强度波动体现为像素值梯度分布；时间相关性转换为空间相邻性。这种转换使DoS攻击的放射状条纹与正常流量周期波动形成显著的视觉可分性，为模型分类提供了更好的输入条件。

图1中各模型气泡大小代表了其计算复杂度。

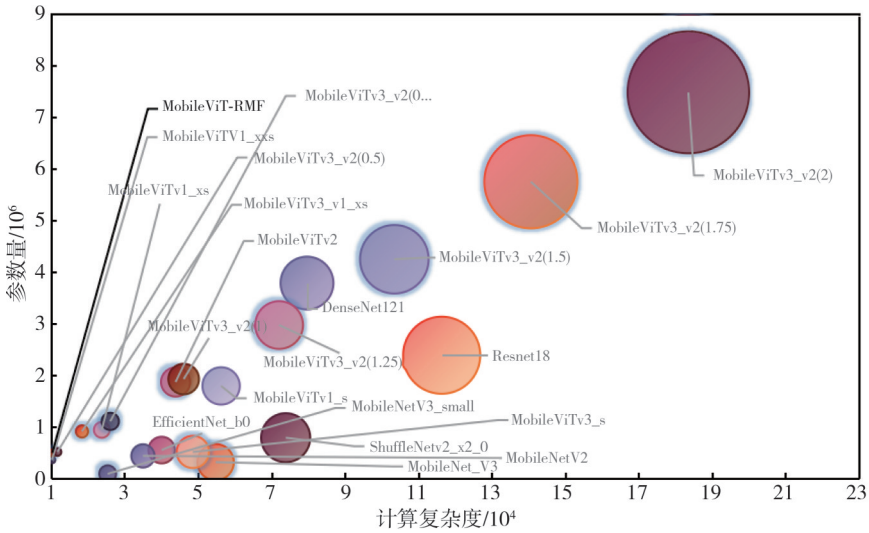

图1　MobileViT-RMF模型与现有轻量级模型的参数规模与计算复杂度比较

Fig. 1　Comparison of parameter size and computational complexity between MobileViT-RMF model and existing lightweight models

MobileViTv3_xxs模型通过深度可分离卷积与Transformer的混合架构，在参数量与计算量之间实现了最优平衡，成为后续改进MobileViT-RMF模型的理想基线，其核心优势在于：1）采用MV2模块捕获GADF图像中DoS攻击的微观纹理特征，通过逐点卷积实现通道维度特征重组；2）MobileViT模块建立了像素间的长程依赖关系，可以识别DoS攻击的宏观关联模式。然而，该模型对特征图的均匀处理方式难以区分突发流量与正常流量的周期性波动。本研究提出了残差多粒度聚焦RMF模块，结构如图2所示。

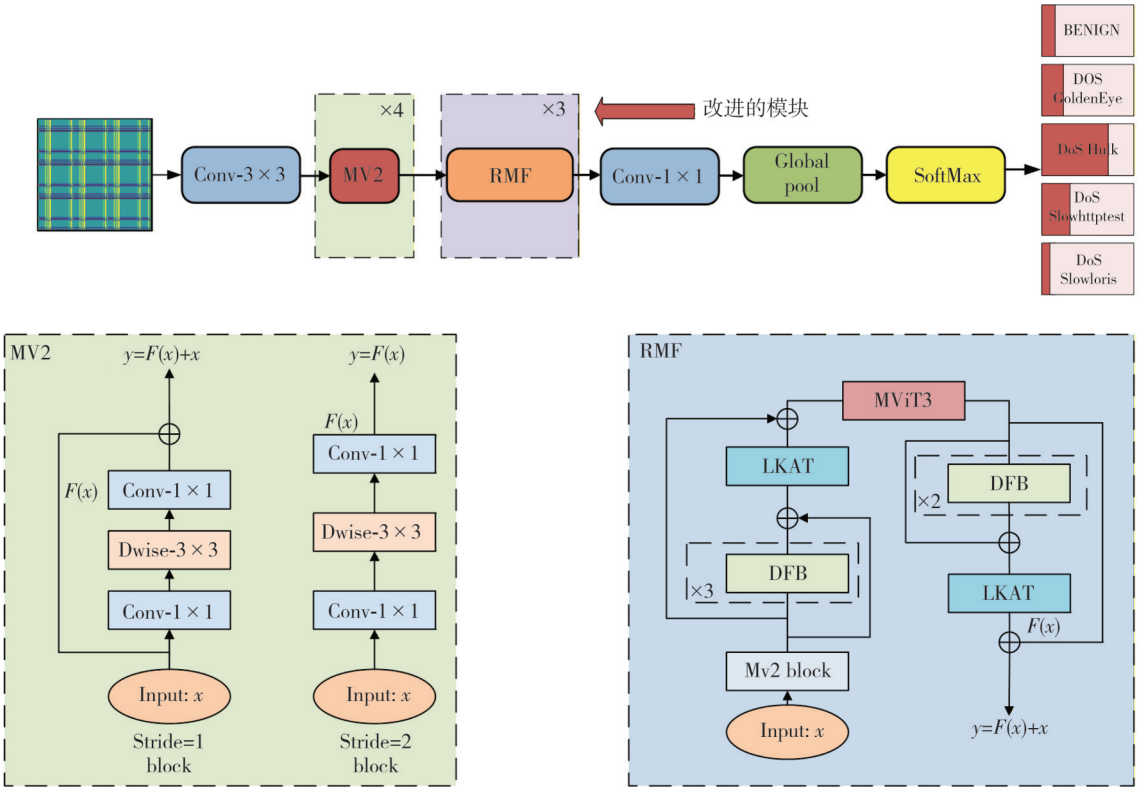

图2 改进MobileViT-RMF结构图

Fig. 2 Structure diagram of the improved MobileViT-RMF

1.2.1 RMF模块

本研究基于动态聚焦模块DFB(DynaFocus-Block)构建了RMF模块如图3所示，RMF模块通过分层堆叠的DFB单元协同工作，每个DFB包含自适应大核注意力ALKA(Adaptive Large Kernel Attention)、空间门控注意力SGA(Spatial Gate Attention)和自调制注意力SMA(Self-modulated Attention)三个核心组件。

将自适应大核注意力ALKA与空间门控注意力SGA堆叠来构建基础模块DFB，DFB通过多尺度和门控方案来改进大核注意力，生成覆盖不同粒度级别的注意力图。此方法有效促进了全局与局部信息的整合，并减少了由于大核尺寸可能引发的阻塞伪影问题。受最新视觉注意研究[15]的启发，本文使用ALKA来解决这些问题，方法是结合大核注意力和多尺度学习，ALKA由3个主要函数组成。

1）大核注意力：给定输入特征图$X\in\mathbb{R}^{C\times H\times W}$，LKA通过将一个$K\times K$卷积分解为3个分量来自适应地构建长程关系：一个$(2d-1)\times(2d-1)$深度可分离卷积$f_{DW}(\cdot)$，一个$\lceil k/d\rceil\times\lceil k/d\rceil$深度可分离空洞卷积$f_{DWD}(\cdot)$和一个逐点卷积$f_{PW}(\cdot)$，上述操作可表述为

$$LKA(X)=f_{PW}(f_{DWD}(f_{DW}(X)))。\tag{3}$$

2）多尺度机制：假设输入特征图$X\in\mathbb{R}^{C\times H\times W}$，该模块首先将其分割为$n$份$X_1,X_2,\cdots,X_n$，每份的大小为$\left\lfloor\frac{C}{n}\right\rfloor\times H\times W$。对于第$i$组特征$X_i$，一个由$\{k_i,d_i\}$分解的LKA被用来生成同质尺度注意力图$LKA_i$。

3）门控机制：对于第i组输入X_i，为了避免块效应，以及学习更多的局部信息，利用空间门控来动态地将$LKA_i(\cdot)$适应为$ALKA_i(\cdot)$，表示为

$$ALKA_i(X_i)=G_i(X_i)\otimes LKA_i(X_i),\tag{4}$$

式中：$G_i(\cdot)$是由$a_i \times a_i$深度可分卷积生成的第i个门控，$LKA_i(\cdot)$是由a_i-b_i-1分解的LKA。特别是，具有更大感受野的$MLKA_i$更倾向于反应长距离依赖，而较小的$MLKA_i$倾向于保留局部纹理。

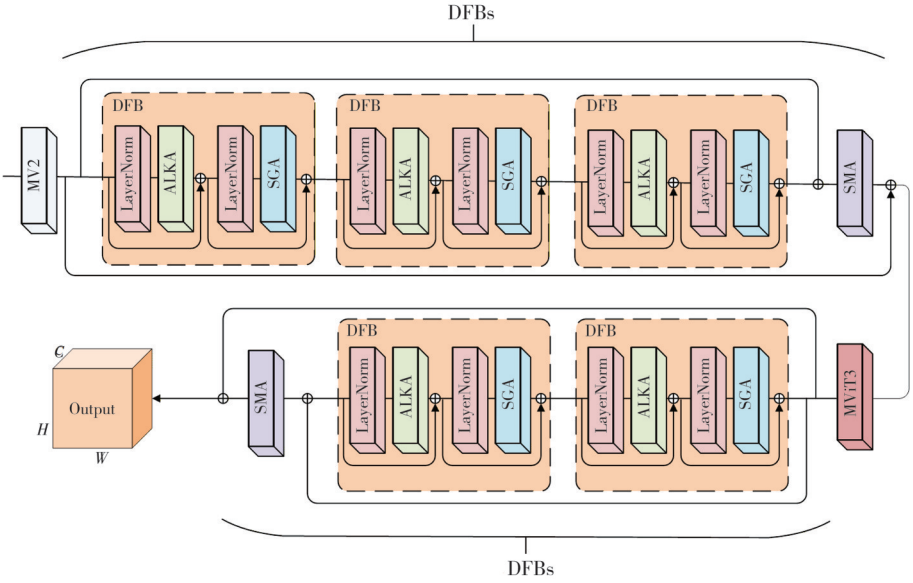

图 3　RMF模块结构图

Fig 3　RMF structure diagram

SGA结合了门控机制与空间注意力，可以剔除不必要的线性层，并聚集空间上下文信息。为了更有效地捕获空间信息，采用单层深度可分离卷积来对特征图进行加权。给定密集变换的X和Y，SGA的关键过程可以表示为

$$SGA(X, Y)=f_{DW}(X)\otimes Y, \tag{5}$$

式中：$f_{DW}(\cdot)$和$\otimes$分别为深度可分离卷积和逐元素乘法。

通过应用空间门控，SGA可以移除非线性层，并在考虑复杂性的情况下捕获局部连续性。ALKA、SGA与SMA的结构如图4所示。ALKA和SGA注意力模块能够分别捕捉长距离依赖且自适应调整特征表示，从而增强局部特征的表征能力。

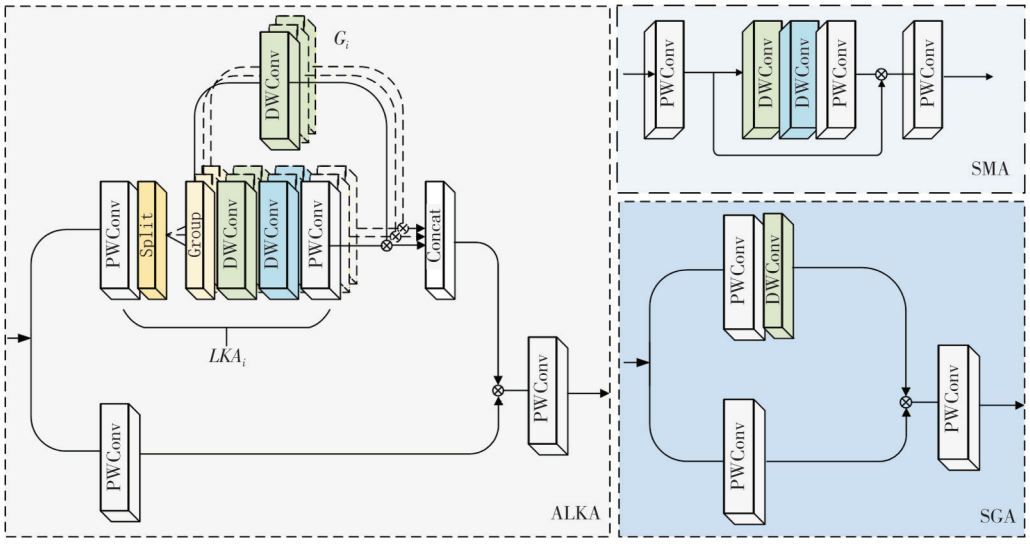

图 4　ALKA、SGA与SMA结构图

Fig. 4　Structural diagrams of ALKA, SGA, and SMA

DFB的具体计算过程为

$$F_{\mathrm{DFB}}=\begin{cases} X_1=LayerNorm(X_{\mathrm{in}}), \\ X_2=X_1+ALKA(X_1) \\ X_3=X_2+SGA(LayerNorm(X_2))。 \end{cases} \tag{6}$$

对于给定输入特征图$X_{\mathrm{in}}\in\mathbb{R}^{C\times H\times W}$，经过LayerNorm、ALKA、SGA等结构后得到残差主路径X_3。每个DFB堆栈包含N个单元，如式(7)所示。

$$X_{\text{out}}^{(k)}=X_{\text{in}}+\bigoplus_{i=1}^{k}F_{\text{DFB}}^{(i)}(X_{\text{in}}),\tag{7}$$

式中：k为DFB块堆叠的次数；$\bigoplus$为跨堆叠DFB单元的残差聚合。通过多次堆叠后在尾部使用SMA尾部注意力最终组成DFBs，具体计算过程为

$$F_{\text{DFBs}}^{(k)}=SMA(X_{\text{out}}^{(k)})+X_{\text{in}}。\tag{8}$$

最终实现两阶段特征调节，在MViT3模块前通过3层DFB堆栈提取多粒度空间模式，如式(9)所示。

$$X_{\text{pre}}=F_{\text{DFBs}}^{(k)}+X_{\text{in}},\ k=3。\tag{9}$$

在MViT3后使用两层DFB堆栈增强上下文感知，通过跳跃连接实现浅层纹理特征与深层语义特征的跨层融合，如式(10)所示。

$$X_{\text{post}}=MViT(X_{\text{pre}})+F_{\text{DFBs}}^{(k)}(MViT(X_{\text{pre}}),\ k=2。\tag{10}$$

1.2.2 融合RMF的MobileViT网络

整个训练过程采用焦点损失Focal Loss优化分类性能，以提高对入侵样本的检测能力。该模型结合局部与全局特征，有效提升了网络入侵检测的准确率。

MobileViT-RMF网络结构如表1所示。*Input*表示网络中每个模块的输入特征图尺寸，*Operator*表示每个特征层所经过的模块，*Cout*表示通过每个特征层后输出的通道数，L表示MViT3模块中Transformer模块的个数，s表示步长，即卷积核每次移动的距离[16]。

表1 MobileViT-RMF网络结构

Tab. 1 MobileViT-RMF network structure

Input	*Operator*	*Cout*	*L*	*s*
3×256×256	Conv2d	16	—	2
16×128×128	MV2	16	—	1
16×128×128	MV2	24	—	2
24×64×64	MV2	24	—	1
24×64×64	MV2	24	—	1
24×32×32	RMF	63	2	1
63×32×32	RMF	72	4	1
72×16×16	RMF	126	4	1
126×8×8	Conv2d	504	—	1
504×8×8	AvgPool2d	—	—	—
504	FC	—	—	—

2 实验与分析

2.1 数据准备

本研究采用加拿大网络安全研究所发布的CICIDS2017基准数据集，该数据集通过模拟真实企业网络环境而构建，包含连续5 d的完整网络流量记录，涵盖HTTP、FTP、SSH等常规协议流量及多种新型攻击模式[17]。实验选取BENIGN正常流量与DDoS、DoS GoldenEye、DoS Hulk、DoS Slowhttptest、DoS Slowloris五类典型DoS攻击数据，该分类体系可有效验证模型在泛洪攻击DDoS、应用层攻击Slowloris及混合攻击场景下的检测鲁棒性。

2.1.1 数据预处理

针对CICIDS2017数据集存在的非数值型特征、缺失值及异常值问题，本研究实施三级数据清洗策略：

1）删除包含无限大/小值的异常样本；

2）采用众数填充法补全离散型特征缺失值；

3）对连续型特征缺失值进行相邻时间窗线性插值。

为缓解类别不平衡问题，采用分层欠采样方法将多数类BENIGN样本量与少数类DoS Slowloris样本量的比例压缩至1∶1，相较Pradipta等[18]采用的SMOTE数据平衡的方法，该方法可有效避免过拟合风险。

在数据编码阶段构建多分类编码体系：采用二进制编码机制，将良性流量统一标记为0，各类攻击流量根据攻击类型分配1～5的离散编码值，实现分类变量的数值化转换。采用Min-Max归一化方法将原始序列$X=\{x_1,x_2,\cdots,x_n\}$归一化至区间$[-1,1]$。式中，X为特征向量，$X_{\min}$与$X_{\max}$分别为特征最小值与最大值。相较于文献[19]采用的Z-score标准化，该方法能更好地保留流量特征的相对比例关系。

$$X_{\text{norm}}=\frac{X-X_{\min}}{X_{\max}-X_{\min}}。\tag{4}$$

预处理后的CICIDS2017数据集包含80维有效特征，将时间窗同步设为80点可实现特征维度与转换矩阵的严格对应。不同时间窗对分类模型准确率的影响如表2所示。此设计确保了每个时间窗完整覆盖单次检测所需的全部特征维度，避免零填充或截断导致的信息损失。

表2 不同窗口大小对分类准确率的影响

Tab. 2 The impact of different window sizes on classification accuracy

窗口尺寸	图片大小/像素	分类准确率/%
40	256×256	96.53
50	256×256	96.99
60	256×256	98.41
70	256×256	98.44
80	256×256	99.75

最终构建的数据集包含34 292个有效样本，按7∶3比例采用分层随机抽样划分训练集与测试集，具体数据分布如表3所示。最终生成具有时空关联性的二维图像数据如图5所示。

表3　数据分布

Tab. 3　Data distribution

攻击名称	训练集	测试集	总计
BENIGN	5 036	1 039	6 075
DDoS	6 000	1 000	7 000
DoS_GoldenEye	6 000	1 000	7 000
DoS_Hulk	6 000	1 000	7 000
DoS_Slowhttptest	5 485	1 000	6 485
DoS_Sloworis	5 771	1 000	6 771

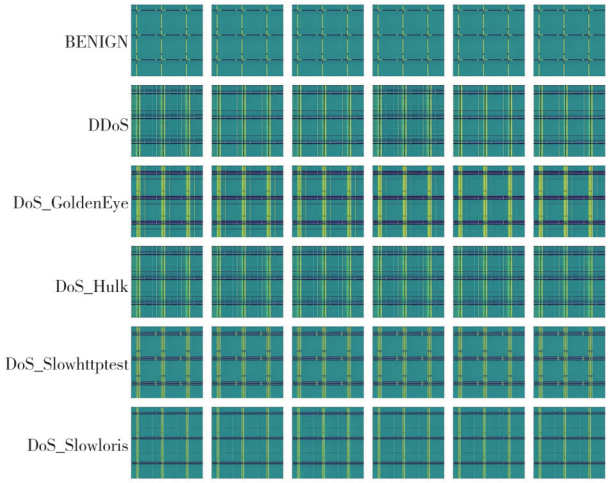

图5　CICIDS2017数据集部分样例展示

Fig. 5　A portion of sample displays from the CICIDS2017 dataset

2.2　实验环境

本实验选用Python 3.9.19作为编程语言，基于PyTorch 2.3.1深度学习框架构建网络模型，并使用CUDA 12.5版本进行加速计算。具体硬件配置如表4所示。

表4　实验环境硬件配置

Tab. 4　Hardware configuration of the experimental environment

CPU	GPU	内存	存储
AMD Ryzen9 5900HX	GeForce RTX 3080	SamsungDDR4 16×2GB	SAMSUNG MZVLB

实验的训练与测试环境为Windows 11操作系统。本次实验采用带权重衰减的自动梯度AdamW[20]为优化器，具体超参数如表5所示。

表5　实验环境超参数设置

Tab. 5　Hyperparameter settings for the experimental environment

迭代次数	批次大小	优化器	学习率	损失函数
100	32	AdamW	0.003	Focal Loss

2.3　评价指标

实验采用图像分类任务常用的评价指标：精确率P（Precision）、召回率R（Recall）、$F1$值（$F1$-score）和预测准确率Acc（Accuracy）。

$$P=\frac{N_{TP}}{N_{TP}+N_{FP}},\tag{5}$$

$$R=\frac{N_{TP}}{N_{TP}+N_{FN}},\tag{6}$$

$$F1=\frac{2P}{2N_{TP}+N_{FP}+N_{FN}},\tag{7}$$

$$Acc=\frac{N_{TP}+N_{TN}}{N_{TP}+N_{TN}+N_{FP}+N_{FN}},\tag{8}$$

式中：N_{TP}为正例预测正确的个数；N_{TN}为正例预测错误的个数；N_{FP}为负例预测错误的个数；N_{FN}为负例预测正确的个数。

2.4　对比实验

为了验证本文提出的MobileViT-RMF网络模型的性能，在数据集CICIDS2017上将其与EfficientNet、Resnet、Shufflenet、MobileNet、SqueezeNet、MobileViT等主流轻量级系列模型进行了对比实验。

2.4.1　与主流轻量级网络的对比

实验中，平均预测时间（Average Predicted Time，APT）是在稳定输入条件下对同一幅图片重复进行16次计算获得的平均推理时间，用以表征即时响应特性。所有评价指标结果如表6所示。

由表6可以看出，相较于基线模型MobileViT_V3_xxs，MobileViT-RMF在准确率上提升了2.74百分点，同时标准差由1.10降低至0.59，表明模型的输出结果更加稳定。在保持较低参数规模（1.03×10^6）与计算开销的前提下，所提模型在ACC、$F1$及召回率等指标上均达到当前对比模型中的最优水平。进一步对比可发现，MobileViT-RMF在各项评价指标上的标准差均控制在0.6以内，相较于MobileViT_V3_0.5及MobileViT_V2的结果均表现出更低波动性，说明引入RMF模块后模型在不同测试划分下具有更稳定的性能表现。

为了验证RMF模块的有效性，在相同训练条件下对比了MobileViT-RMF与基准MobileViT模型的分类准确率，结果如图6所示。可以看出，改进模型在90%的测试周期中保持了99%以上的准确率，而基准模型存在3次显著波动。面

对混合型DDoS攻击时，RMF模型在第16测试周期的准确率为99.32%，而基准模型下降至92.75%。实验结果验证了残差多粒度聚焦模块RMF对异常流量特征的稳定捕捉能力。

表6 MobileViT-RMF网络与轻量级网络各指标对比

Tab. 6 Comparison of MobileViT-RMF network and lightweight network indicators

网络	参数量/10^6	APT/ms	FLOPs/10^6	*Acc*/%	*F*1/%	*R*/%	*P*/%
ShuffleNet_V2[21]	7.39	20	780	96.61±2.32	96.60±2.39	96.61±2.31	96.73±1.68
EfficientNet_B0[22]	4.01	37	540	97.18±1.56	97.18±1.55	97.18±1.56	97.22±1.30
MobileNet_V1[23]	3.21	30	587	97.08±1.02	97.08±1.04	97.08±1.02	97.11±0.86
MobileNet_V2[24]	3.50	22	427	95.93±8.93	95.76±10.26	95.93±8.94	96.81±2.84
MobileNet_V3[25]	5.48	28	305	93.72±13.53	93.33±15.41	95.45±13.54	95.45±5.95
MobileViT_V1[26]	1.02	25	350	97.03±0.73	97.03±0.73	97.05±0.73	97.05±0.64
MobileViT_V2[27]	4.39	30	1451	97.08±3.87	97.00±4.49	97.08±3.90	97.39±1.67
MobileViT_V3_xxs[28]	0.73	26	278	97.01±1.10	97.02±1.10	97.03±1.10	97.26±0.93
MobileViT_V3_0.5[28]	1.17	28	381	97.68±0.76	97.68±0.77	97.70±0.75	97.70±0.61
SqueezeNet1_1[29]	0.72	21	263	85.11±11.14	81.16±14.62	85.11±11.14	90.66±5.85
MobileViT-RMF(ours)	1.03	23	279	99.75±0.59	99.75±0.59	99.75±0.59	99.76±0.53

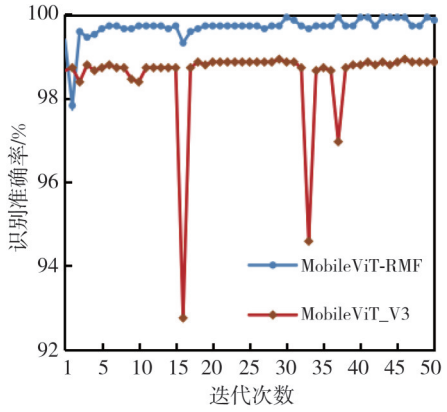

图6 模型准确率对比

Fig. 6 Comparison of model accuracy

图7中MobileViT-RMF与基准模型的交叉熵损失在50个测试周期内呈现显著差异。可以看出，改进模型展现出更优的损失收敛稳定性，其平均损失波动幅度较基准模型降低61.4%。具体而言，改进模型自第35测试周期起即进入稳定收敛状态，而基准模型在第47测试周期出现异常激增，突显了潜在梯度失稳的缺陷。RMF模型的最大损失增幅仅为基准模型的48.1%，进一步验证了该模块对异常流量特征的鲁棒性优化。

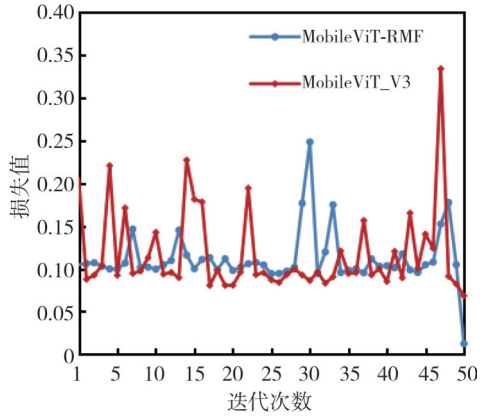

图7 模型交叉熵损失值对比

Fig. 7 Comparison of model accuracy

2.4.2 与CICIDS2017数据集上较优模型的对比

将本文模型与现有基于CICIDS2017数据集一维数据分类模型OC-SVM / RF[30]、Random Forest[31]、CNN-BiLSTM[32]、DBN[33]等进行比较，结果如图8所示。由图8可知，MobileViT-RMF在四项关键指标上均显著优于其他模型，展现出近乎完美的检测性能。OC-SVM虽在精确度和*F*1得分上表现较优，但其召回率与MobileViT-RMF存在1.39%的显著差距。CNN-BiLSTM出现了明显的性能失衡现象，其召回率与精确度差值达8.01百分点，暗示该模型存在严重的漏报风险。

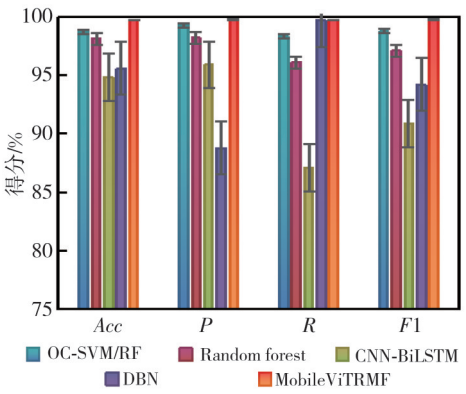

图8 模型性能对比柱状图

Fig. 8 Comparison of model performance bar chart

2.4.3 模型泛化性验证

为了评估MobileViT-RMF模型的跨数据集适应能力，在UNSW-NB15[34]数据集上对其进行了泛化验证实验。数据集同样经上述方法进行数据清洗、预处理及下采样至样本均衡，实验结果如图9所示。可以看出，MobileViT-RMF模型的峰值准确率达到99.89%。

输出混淆矩阵，其中矩阵的行表示真实类别，列表示预测结果，各主对角线元素反映类内

检测的准确率，非对角线元素揭示误判模式。

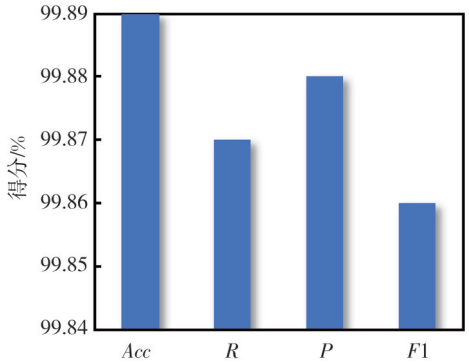

图 9　改进模型在UNSW-NB15数据集上的指标柱状图

Fig. 9　The improved model's indicator bar chart on the UNSW-NB15 dataset

图 10 为MobileViT-RMF模型在CICIDS2017数据集中性能检测的结果。可以看出，BENIGN样本被正确识别，DDoS、DoS_GoldenEye、DoS_Slowloris与Hulk等四类攻击均实现了100%的准确检测。GoldenEye攻击特有的不完整HTTP请求特征，经GADF转换后呈现放射状条纹。Hulk攻击的高并发GET请求模式映射为密集块状纹理。实验结果表明，本文提出的RMF模块成功捕获了两类攻击在空间频率域的差异化特征。DoS Slowhttptest检测的准确率为99.00%，1%的误判可能是由于Slowloris混合攻击策略导致了特征重叠。

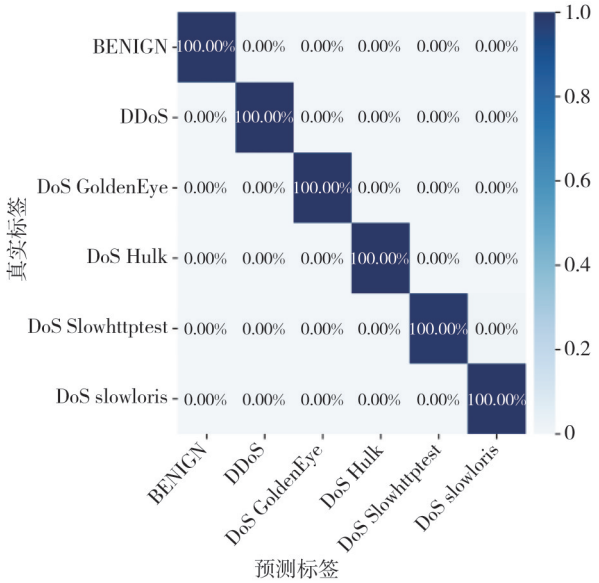

图 10　改进模型在CICIDS2017数据集上的混淆矩阵

Fig. 10　The confusion matrix of the improved model on the CICIDS2017 dataset

图 11 为本文模型在UNSW-NB15数据集上输出的混淆矩阵。可以看出，在对DoS攻击进行分类时准确率达到了99.33%，其中0.67%的样本被误判为Exploits攻击，产生误差的原因可能是部分Exploits样本使用了UDP Flood技术，与DoS攻击策略产生交叉导致生成图像特征区域重叠。而对BENIGN、Exploits、Fuzzers、Generic、Reconnaissance四类样本均实现了100%的精确识别，证明GADF转换通过时频域特征增强机制使正常流量与攻击流量具有显著的视觉可分性。

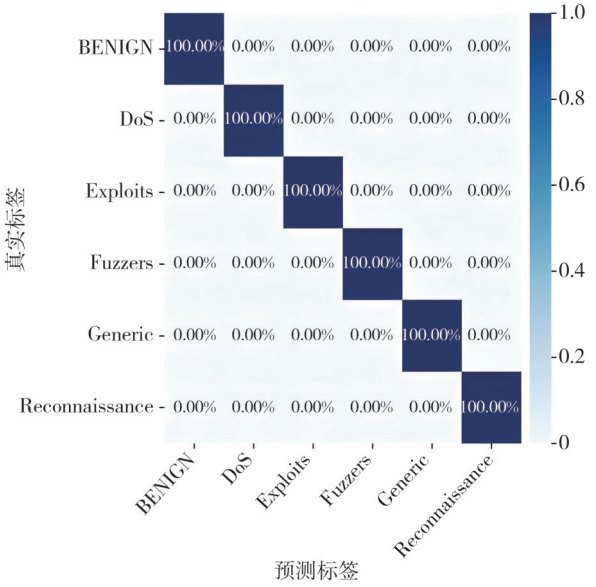

图 11　改进模型在UNSW-NB15数据集上的混淆矩阵

Fig. 11　The confusion matrix of the improved model on the UNSW-NB15 dataset

2.4.4　消融实验

为了验证本文改进模型中各个模块的有效性，在CICIDS2017数据集上进行了系统性消融实验，结果如表 7 所示。可以看出，基准模型的分类准确率为98.44%，单独引入SMA组件使准确率提升了0.44百分点，证明局部卷积与自注意力融合能有效捕获网络流量的空间相关性特征。单独应用SGA时准确率达98.60%，说明动态空间门控机制可强化关键区域的特征响应。单独部署ALKA导致性能下降至97.87%（下降0.57百分点），这表明多尺度特征融合需依赖其他组件进行通道维度校准。

ALKA与SMA协同，证实多尺度特征金字塔为局部增强操作提供了层次化语义支撑；ALKA+SGA组合的准确率达到99.33%，表明门控机制能自适应调节多尺度特征的权重分布；SGA+SMA组合出现性能抑制现象，推测全局注意力与局部增强可能存在特征响应域冲突，需通过ALKA的多尺度解耦实现协同优化。

表7 消融实验结果

Tab. 7 Ablation experiment results

RMF			CICIDS2017
ALKA	SGA	SMA	*Acc*/%
×	×	×	98.44
×	×	√	98.88
×	√	×	98.60
√	×	×	97.87
×	√	√	98.21
√	×	√	99.10
√	√	×	99.33
√	√	√	**99.99**

当同时集成ALKA、SGA与SMA时，模型的准确率达到99.99%，较基准提升1.55百分点。此现象表明：ALKA构建的多尺度特征空间为SGA提供了分粒度的注意力计算基础。SGA的动态门控引导SMA聚焦于局部模式的判别。

2.4.5 边缘计算设备部署试验分析

为了验证MobileViT-RMF模型在移动端的实际部署性能，将其与MobileViT模型共同部署于边缘计算设备Jetson Xavier NX平台。该平台硬件配置如表8所示。

平台软件环境采用JetPack 5.0.2系统，集成TensorRT 8.4.1.5推理引擎与CUDA 11.4并行计算平台，设置平台最高功耗为15W模式。

基于此平台将MobileViT、MobileViT-RMF模型设置为FP16量化。模型检测帧率结果如表9所示。

表8 平台硬件配置表

Tab. 8 Hardware configuration

平台	处理器	内存	算力
Jetson Xavier NX	6核Carmel CPU + 384核Volta GPU	8GB LPDDR4	21 TOPs

表9 模型检测帧率

Tab. 9 Detection frame rate of models

模型	参数量/10^6	FLOPs/10^6	帧率/(帧·s^{-1})
MobileViT	1.03	350	72
MobileViT-RMF	1.02	279	94

由表9可以看出，MobileViT-RMF实现了94帧·s^{-1}的实时推理性能，验证了其在边缘设备上的部署优势。与MobileViT相比，该模型在资源受限设备上表现出更优的工程适用性，为网络入侵检测等实时边缘计算场景提供了新方案。未来可通过模型剪枝和硬件指令集优化进一步提升性能。

3 结 论

本研究通过融合格拉姆角差场GADF编码与残差多粒度聚焦模块RMF，构建了轻量化入侵检测模型MobileViT-RMF。实验表明，该模型在CICIDS2017和UNSW-NB15数据集上的检测准确率分别达到99.75%与99.89%，验证了GADF在时空特征编码与RMF模块在多粒度特征融合上的有效性。

尽管本研究取得了上述成果，仍存在以下局限性：1）检测范畴方面，当前GADF-RMF框架主要针对已知攻击模式的特征优化，对零日攻击的检测能力尚未验证。这源于GADF编码依赖预定义攻击特征的时空模式提取，而未知攻击缺乏此类先验知识。2）流量适应方面，固定时间窗策略(80点)在处理非平稳流量时可能产生相位偏移。

针对这些不足，可从以下方面改进：1）开发动态时间窗机制，通过LSTM网络自适应调整分段策略；2）探索基于小样本学习的未知攻击检测范式，通过元学习(Meta-learning)构建可扩展的特征空间。

利益冲突声明/Conflict of Interests

所有作者声明不存在利益冲突。

All authors declare no relevant conflict of interests.

参考文献：

[1] KUMARI P, JAIN A K. A comprehensive study of DDoS attacks over IoT network and their countermeasures[J]. Computers & Security, 2023, 127: 103096.

[2] NEXUSGUARD. Distributed denial of service trend report 2024[EB/OL]. (2025-02-21)[2025-06-12]. https://www.nexusguard.com/file/nexusguard-ddos-trend-report-2024.

[3] ALI A, YOUSAF M M. Novel three-tier intrusion detection and prevention system in software defined network[J]. IEEE Access, 2020, 8: 109662-109676.

[4] LIAO H, MURAH M Z, HASAN M K, et al. A survey of deep learning technologies for intrusion detection in Internet of Things[J]. IEEE Access, 2024, 12: 4745-4761.

[5] MO Y M, LI H, WANG D S, et al. An intrusion detection system based on convolution neural network[J]. PeerJ Computer Science, 2024, 10: e2152.

[6] DEMMESE F A, NEUPANE A, KHORSANDROO S, et al. Machine learning based fileless malware traffic classification using image visualization[J]. Cybersecurity, 2023, 6(1): 32.

[7] PHAM V, SEO E, CHUNG T M. Lightweight convolutional neural network based intrusion detection system[J]. Journal of Communications, 2020, 15(11): 808-817.

[8] ATTACK W, ATTACK I, ATTACKB F. Ensemble of feature augmented convolutional neural network and deep autoencoder for efficient detection of network attacks [J]. Scientific Reports, 2025, 15: 4267.

[9] 刘文琪，胡涛，闫洁，等. 基于DeepInsight和迁移学习的入侵检测技术[J]. 工程科学学报，2024，46(12)：2238-2245.
LIU Wenqi, HU Tao, YAN Jie, et al. Network intrusion detection technology based on DeepInsight and transfer learning[J]. Chinese Journal of Engineering, 2024, 46(12): 2238-2245.(in Chinese)

[10] WU Z T, LONG Z H. Research on network traffic classification method based on CNN-RNN[M]. Singapore: Springer Nature, 2023.

[11] WANG L T, HU W, LIU J Y, et al. Encrypted traffic classification based on fusion of vision transformer and temporal features[J]. The Journal of China Universities of Posts and Telecommunications, 2023, 30(2): 73-82.

[12] AVCI C, TEKINERDOGAN B, CATAL C. Design tactics for tailoring transformer architectures to cybersecurity challenges [J]. Cluster Computing, 2024, 27(7): 9587-9613.

[13] 姚军，孙方超. 基于MobileViT的轻量型入侵检测模型研究[J]. 现代电子技术，2024，47(19)：33-39.
YAO Jun, SUN Fangchao. Research on lightweight intrusion detection model based on MobileViT [J]. Modern Electronics Technique, 2024,47(19): 33-39.(in Chinese)

[14] XIAO F Y, CHEN Y Y, ZHU Y H. GADF/GASF-HOG: Feature extraction methods for hand movement classification from surface electromyography[J]. Journal of Neural Engineering, 2020, 17(4): 046016.

[15] WANG Y, LI Y S, WANG G, et al. Multi-scale attention network for single image super-resolution [C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2024: 5950-5960.

[16] LUO S S, LI M J, CHEN M, Multi-scale integrated at-tention mechanism for facial expression recognition network[J]. Computer Engineering and Applications, 2023,59(1): 199-206.

[17] PANIGRANHI R, BORAH S. A detailed analysis of CICIDS2017 dataset for designing intrusion detection systems [J]. International Journal of Engineering & Technology, 2018, 7(3): 479-482.

[18] PRADIPTA G A, WARDOYO R, MUSDHOLIFAH A, et al. SMOTE for handling imbalanced data problem: A review[C]//2021 Sixth International Conference on Informatics and Computing (ICIC), 2021: 1-8.

[19] PEDREGOSA F, VAROQUAUX G, GRAMFORT A, et al. Scikit-learn: Machine learning in Python [J]. Journal of Machine Learning Research, 2011, 12(10): 2825-2830.

[20] ZHUANG Z X, LIU M R, CUTKOSKY A, et al. Understanding AdamW through proximal methods and scale-freeness [DB/OL]. (2022-01-31) [2025-06-16]. https://arxiv.org/abs/2202.00089.

[21] MA N N, ZHANG X Y, ZHENG H T, et al. Shufflenet V2: Practical guidelines for efficient CNN architecture design [DB/OL]. (2018-07-30) [2025-06-12]. https://arxiv.org/abs/1807.11164.

[22] BODAVARAPU P N R, SRINIVAS P V V S. Facial expression recognition for low resolution images using convolutional neural networks and denoising techniques[J]. Indian Journal of Science and Technology, 2021,14(12): 971-983.

[23] HOWARD A G, ZHU M, CHEN B, et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications [DB/OL]. (2017-04-17) [2025-06-12]. https://arxiv.org/abs/1704.04861.

[24] SANDLER M, HOWARD A, ZHU M, et al. MobileNetV2: Inverted residuals and linear bottlenecks [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018: 4510-4520.

[25] HOWARD A, SANDLER M, CHU B, et al. Searching for MobileNetV3 [C]//2019 IEEE/CVF International Conference on Computer Vision(ICCV), 2019: 1314-1324.

[26] MEHTA S, RASTEGARI M. Mobilevit: Lightweight, general-purpose, and mobile-friendly vision transformer [DB/OL]. (2021-10-05) [2025-06-12]. https://arxiv.org/pdf/2110.02178v1.

（下转第262页）