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4期(总第228期)
Vol. 47 No. 4 2026(Sum No. 228)

文章编号：1673-3193(2026)04-0531-10 J. North Univ. China, Nat. Sci. Ed., 2026, 47(4): 531-540.

基于递归图和MobileViT-AESTF的网络入侵检测模型

刘诗涵，赵利辉*，刘泽晋，安 洋
（中北大学 软件学院， 山西 太原 030051）

摘 要：针对当前网络入侵检测中流量图像化导致的时序特征丢失以及轻量化模型在复杂攻击下识别精度不足的问题，提出一种基于递归图(Recurrence Plot，RP)与自适应高效时空融合模块(Adaptive Efficient Spatiotemporal Fusion，AESTF)的轻量化入侵检测模型MobileViT-AESTF。该方法利用相空间重构技术，通过递归图将一维流量时间序列映射为二维状态空间中的递归关系，有效保留了流量的非线性时序特征。针对MobileViT_V3在图像化流量处理中特征捕获能力不足的问题，设计了AESTF模块。该模块包含高效时空融合单元(Efficient Spatiotemporal Fusion，ESTF)和自适应特征激励器(Adaptive Feature Exciter，AFE)，其中ESTF通过双路径异构注意力机制并行提取空间与通道特征，AFE则基于神经元能量函数自适应增强关键判别特征。在CIC UNSW-NB15 Augmented数据集上的实验结果表明，MobileViT-AESTF在保持较低参数量(1.04×10^6)的同时，检测准确率、召回率及$F1$分数分别达到了99.84%、99.82%和99.85%。相比基准模型MobileViT_V3，其计算量降低，且在Fuzzers等复杂攻击上的识别性能更具鲁棒性。该模型在模型轻量化与检测精度之间取得了有效平衡，适用于资源受限环境下的网络入侵检测，具有较高的工程应用价值。

关键词：入侵检测；图像分类；轻量化网络；MobileViT
中图分类号：TP393.08 **文献标识码：**A **doi:** 10.62756/jnuc.issn.1673-3193.2025.09.0004
引用格式：刘诗涵，赵利辉，刘泽晋，等. 基于递归图和MobileViT-AESTF的网络入侵检测模型[J]. 中北大学学报(自然科学版)，2026，47(4)：531-540.
Liu Shihan, Zhao Lihui, Liu Zejin, et al. Network intrusion detection model based on recurrence plot and MobileViT-AESTF[J]. Journal of North University of China(Natural Science Edition), 2026, 47(4): 531-540.

Network Intrusion Detection Model Based on Recurrence Plot and MobileViT-AESTF

Liu Shihan, Zhao Lihui*, Liu Zejin, An Yang
(School of Software, North University of China, Taiyuan 030051, China)

Abstract: To address the issues of temporal feature loss caused by traffic visualization and the insufficient detection accuracy of lightweight models under complex attacks in network intrusion detection, a lightweight model named MobileViT-AESTF was proposed based on recurrence plot (RP) and an adaptive efficient spatiotemporal fusion (AESTF) module. Utilizing phase space reconstruction technology, the method mapped one-dimensional traffic time series into recursive relationships in a two-dimensional state

收稿日期：2025-09-07
基金项目：山西省基础研究计划青年科学研究项目(202203021212114)
作者简介：刘诗涵(2000—)，女，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。

space via RP, effectively retaining the non-linear temporal features of the traffic. To overcome the insufficient feature capture capability of MobileViT_V3 in processing visualized traffic, the designed AESTF module comprised an efficient spatiotemporal fusion (ESTF) unit and an adaptive feature exciter (AFE); specifically, the ESTF employed a dual-path heterogeneous attention mechanism to extract spatial and channel features in parallel, while the AFE adaptively enhanced key discriminative features based on a neuronal energy function. Experimental results on the CIC UNSW-NB15 Augmented dataset demonstrate that MobileViT-AESTF achieves an accuracy, recall, and $F1$-score of 99.84%, 99.82%, and 99.85% respectively, while maintaining a low parameter count of 1.04×10^6. Compared with the baseline MobileViT_V3, the proposed model reduces computational cost and exhibits greater robustness in identifying complex attacks such as Fuzzers. The model achieves an effective balance between lightweight design and detection accuracy, making it suitable for network intrusion detection in resource-constrained environments and offering significant engineering application value.

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

0 引 言

随着互联网技术的迅猛发展，网络安全问题日益严重，网络流量的复杂性不断增强，攻击手段也呈现出规模化、隐蔽化的趋势，据计算机犯罪研究中心(CCRC)发布的最新报告预测，到2025年，网络犯罪造成的损失将达到12万亿美元[1]。传统基于人工规则和统计特征的入侵检测系统(Intrusion Detection System，IDS)在面对网络攻击时，存在特征表达不足和实时响应瓶颈。近年来，随着人工智能技术的不断发展，基于机器学习和深度学习的入侵检测方法得到了广泛应用，其中图像化特征表达与轻量化模型设计成为两大核心方向[2-3]。

在图像化处理领域，研究者将网络流量转化为二维图像以提升网络攻击的识别能力：Demmese等[4]将恶意流量转换为灰度图像，结合CNN实现了文件类恶意软件的精准识别；刘文琪等[5]引入DeepInsight技术优化流量图像转换过程，通过迁移学习解决小样本泛化问题；Li等[6]则在工业场景中融合CNN与RNN，从图像化流量中捕捉长周期攻击模式。

轻量化技术则致力于突破设备部署的资源限制，提升入侵检测的识别速度：Pham等[7]通过剪枝压缩CNN参数量，为移动端部署奠定了基础；姚军等[8]提出基于MobileViT的轻量检测模型，在保持视觉特征提取能力的同时降低了计算复杂度；而Tiwari等[9]针对IIoT网关设计的专用优化框架将推理延迟控制在毫秒级，Alshehri等[10]则利用自注意力机制精简卷积运算，在工业设备上实现98%的检测精度。这些进展推动了轻量化模型在入侵检测领域的发展。

现有图像化与轻量化入侵检测研究虽取得一定进展，但二者的融合应用仍面临挑战：一是流量数据在图像化转换过程中，易出现时序特征丢失的问题[6]，影响对网络攻击的判定；二是受模型参数量限制，轻量化模型在处理图像数据时存在特征提取能力不足，以及检测精度难以突破瓶颈的问题。

针对上述问题，本研究提出融合递归图RP(Recurrence Plot)编码[11]与改进MobileViT_V3[12]的轻量化入侵检测框架。主要贡献包括：

1) 设计基于RP的流量图像转换方法，利用相空间重构技术将流量序列转化为递归图，通过刻画序列状态间的递归关系，克服传统线性映射导致的时序特征丢失问题；

2) 构建自适应高效时空融合模块(Adaptive Efficient Spatiotemporal Fusion，AESTF)，该模块核心由高效时空融合单元(Efficient Spatiotemporal Fusion，ESTF)和自适应特征激励器(Adaptive Feature Exciter，AFE)两大组件构成，通过增强特征通道间的相互作用与依赖建模，在保持模型轻量化的前提下，提升了检测精度和鲁棒性。

1 基于递归图和MobileViT-AESTF的模型

本研究提出基于递归图RP与轻量化模型MobileViT-AESTF的网络入侵检测框架。首先，引入递归图RP对网络流量数据进行二维转换，克

服传统映射引起的特征丢失问题；其次，选用MobileViT_V3模型作为基线模型，该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有效平衡，本研究在MobileViT_V3的基础上，构建了AESTF模块，在确保模型轻量化的基础上提升了检测性能。

该框架结构图如图1所示，首先对原始流量数据进行数据预处理操作，经数据清洗、Z-score标准化及重要性评分后，保留最具判别性的前N个特征；然后将处理后的一维数据输入递归图RP进行转换，通过计算欧几里得距离转换为递归矩阵，再映射为RGB图像；最后将生成的递归图输入到改进模型MobileViT-AESTF中，分别经过普通卷积层、MV2模块和AESTF模块进行特征提取，使用全局平均池化后放入全连接层并通过SoftMax层输出分类结果，以此来识别攻击行为。

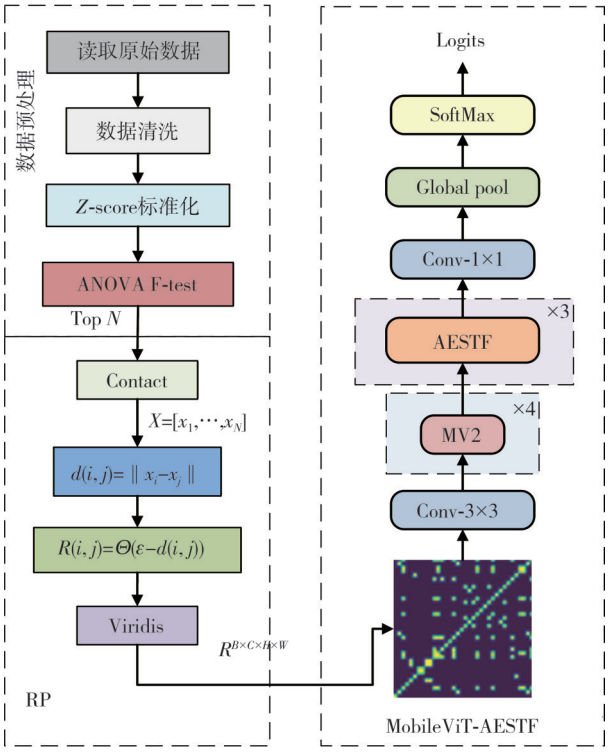

图1　入侵检测框架

Fig. 1　Intrusion detection framework

1.1　递归图RP

本研究引入递归图RP技术对流量特征进行二维转化，该方法通过对网络流量数据特征进行相空间重构，将其映射到二维空间，并且通过距离矩阵的构建，得到二维递归图像，具体流程如下：

1）对于长度为N的时间序列$X=[x_1,x_2,\cdots,x_N]$，计算相邻两个时间点i和j之间的欧几里得距离

$$d(i,j)=\|x_i-x_j\|=\sqrt{\left[(x_i-x_j)^2\right]}。\tag{1}$$

这个过程会生成一个$N\times N$的距离矩阵$\boldsymbol{D}$，其中每个元素$D[i,j]$表示时间点i和j之间的相似性度量。

2）选定距离阈值ε为0.1，不同距离阈值下模型的准确率如表1所示，将距离矩阵转换为二值递归矩阵$\boldsymbol{R}$，其中，$R(i,j)=1$表示时间点i和j的距离小于阈值ε（即相似），$R(i,j)=0$表示距离大于等于阈值ε（即不相似）。

$$R(i,j)=\Theta(\varepsilon-d(i,j)),\tag{2}$$

式中：Θ为Heaviside阶跃函数。

表1　不同距离阈值下模型的准确率

Tab. 1　The accuracy of the model under different distance thresholds

ε	$ACC/\%$
1.0	95.78
0.5	97.46
0.2	98.52
0.1	**99.84**
0.05	98.12
0.01	94.04

3）将二值递归矩阵$\boldsymbol{R}$作为二维图像进行处理，使用viridis颜色映射方案将二值矩阵转换为彩色图像，将输出的递归图作为模型的输入。

整个流程的核心思想是通过计算时间序列中相邻两个时间点之间的相似性来构建二维矩阵，从而将网络流量数据以二维化的方式呈现出来。为验证RP技术对检测性能的改善作用，本研究设计了对照实验：在相同模型架构（MobileViT-AESTF）与实验条件下，测试传统灰度图与递归图的模型性能，准确率分别为96.58%和99.84%。这一结果证明，RP技术通过相空间重构保留了流量数据的时空关联特征，克服了传统映射导致的时序特征丢失问题。

1.2　AESTF模块

AESTF模块的架构如图2所示，由时空融合单元ESTF和自适应特征激励器AFE构成，并嵌入残差连接结构中。其中，ESTF模块负责特征分离与融合，实现时空特征的融合，AFE模块负责特征精炼与增强，自适应地增强关键特征表示。时空特征中的时间特征指递归图中由时间序列相空间重构所隐含的跨时间状态演化关系；空间特

征对应递归图中不同像素位置所反映的状态相似性分布。AESTF模块分别以空间注意力建模递归结构分布，以通道注意力刻画不同特征维度对时间演化的贡献，实现时空特征的联合建模。

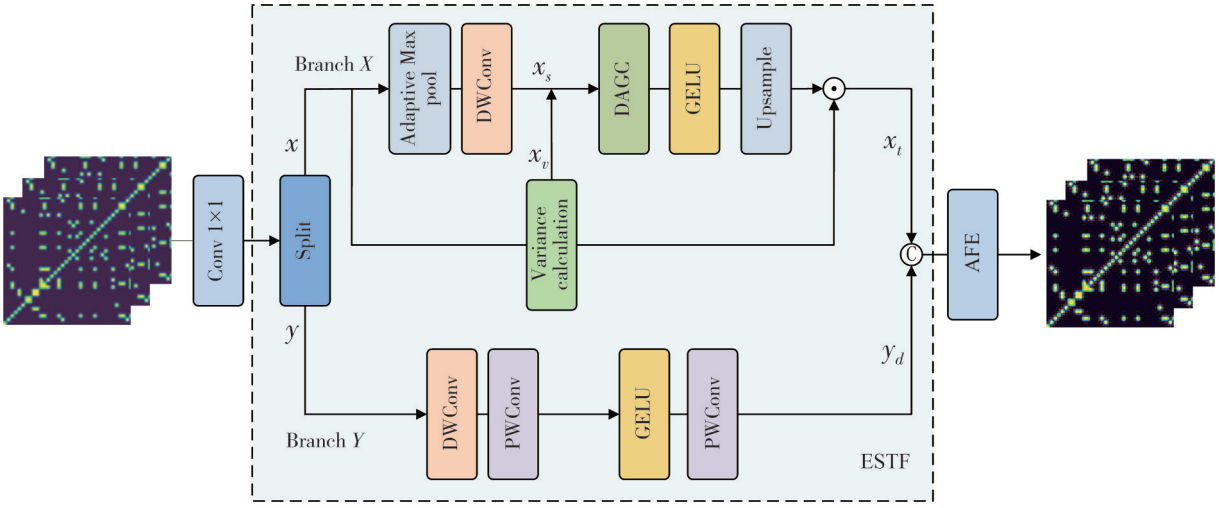

图2 AESTF模块结构

Fig. 2 AESTF structure

该模块的输入特征图为$F_{in}\in R^{B\times C\times H\times W}$，其中$B$为批量大小，$C$为通道数，$H$, W分别为特征图的高度和宽度。其输出F_{out}的计算过程如下：

1) 对输入的特征图进行一次前向增强计算，生成增强特征图，如式(3)所示。

$$F_{proc}=AFE(ESTF(F_{in})), \tag{3}$$

式中：$AFE(\cdot)$为自适应特征激励器的处理过程，其通过动态调整特征响应，增强模型对重要特征的关注度。

2) 将增强特征图与原始输入特征图进行逐元素累加，形成融合特征图。为强化模型对关键特征的捕获能力，将此融合特征图再次与原始输入特征图进行残差连接，得到最终输出特征图，如式(4)所示。

$$F_{out}=F_{in}+(F_{in}+F_{proc})。 \tag{4}$$

如表2所示，与标准残差连接相比，本文改进后的残差连接通过强化原始输入的权重占比，有效缓解了轻量化网络中的特征衰减问题，并在保留原始特征判别信息的同时，进一步提升了模型的检测性能。

表2 不同残差连接下模型的准确率

Tab. 2 The accuracy of the model under different residual connections

残差连接公式	ACC/%
$F_{out}=F_{in}+F_{proc}$	98.28
$F_{out}=F_{in}+(F_{in}+F_{proc})$	**99.84**

在网络入侵检测任务中，AESTF模块经过验证，在区分正常流量与异常行为方面表现出一定的有效性。特别是在处理经过递归图转化的网络流量数据时，该模块展现出较为稳定的检测性能。

1.2.1 ESTF模块

ESTF模块采用双路径异构注意力机制，ESTF为核心调度器，它将输入的特征图$F_{in}\in R^{B\times C\times H\times W}$分为两条并行的处理路径：空间注意力路径(Branch X)和通道注意力路径(Branch Y)，如图2所示。两条路径在特征建模维度和计算方式上具有异构性，通过并行建模提升特征表达能力。

对特征图进行特征分离，通过一个1×1卷积对输入特征进行线性变换，沿通道维度将其均匀地分割为两个独立的特征图x, y，如式(5)所示。

$$[x,y]=Split(Conv_{1\times 1}(F_{in})), \tag{5}$$

式中：$x\in R^{B\times \frac{C}{2}\times H\times W}$，$y\in R^{B\times \frac{C}{2}\times H\times W}$。这种分离策略使两条路径能够专注于不同的注意力维度，同时使计算量减半，在保证检测性能的同时提升了处理效率。

Branch X：该路径为空间注意力路径，旨在建模不同空间尺度下的特征分布差异，通过对特征图进行下采样获取多尺度特征；引入空间维度特征方差刻画通道内特征的空间离散性，以捕捉全局统计信息，再经可学习参数α和β自适应调节权重，强化递归图关键区域响应。具体实现如下：

1）多尺度特征提取：对特征图进行下采样，通过自适应最大池化（Adaptive Max Pooling，AMPool）保留显著的特征响应，利用一个轻量的深度可分离卷积（Depthwise Separable Convolution）来提取这些特征，得到x_s，如式（6）所示。

$$x_s = DWConv_{3\times3}(AMPool(x))。\tag{6}$$

2）全局方差信息：计算x在每个通道上的空间方差x_v，它反映了该通道特征的离散程度和全局统计信息，如式（7）所示。

$$x_v = Var(x)。\tag{7}$$

3）动态注意力门控（Dynamic Attention Gate Control，DAGC）：将x_s和x_v通过两个可学习的缩放参数α和β进行加权融合。其中，α初始值为1，确保初始时x_s无衰减，β初始值为0，保证方差信息无贡献，在训练过程中，通过反向传播机制动态优化更新。表3结果表明，自动学习模式下模型准确率达99.84%，优于固定参数配置。再经过一个1×1卷积和GELU激活函数，生成一个紧凑的门控信号G。这种设计使空间统计信息能够以加权融合的形式参与注意力生成，从而在保持轻量化的同时增强全局感知能力，如式（8）所示。

$$G = \phi(Conv_{1\times1}(x_s \odot \alpha + x_v \odot \beta)),\tag{8}$$

式中：$\odot$为逐元素相乘；$\phi(\cdot)$为GELU激活函数。

表3　不同参数配置下模型的准确率

Tab. 3　The accuracy of the model under different parameter configurations

参数配置	ACC/%
$\alpha=1.0$，$\beta=0.0$	98.32
$\alpha=0.5$，$\beta=0.5$	99.02
$\alpha=1.5$，$\beta=1.0$	98.86
自动学习	**99.84**

4）特征调制：将生成的门控信号G通过最近邻插值上采样至原始分辨率，并与原始特征图x进行逐元素相乘，得到经过空间注意力增强的特征x_t。具体计算过程如式（9）所示。

$$x_t = x \odot Up(G),\tag{9}$$

式中：$Up(\cdot)$为最近邻插值上采样。

Branch Y：作为通道注意力路径，用于建模不同特征通道之间的非线性依赖关系。该路径采用深度可分离卷积与逐点卷积相结合的结构，在扩大感受野、混合通道信息的同时，降低了参数量和计算成本。具体实现如下：

输入特征图经过3×3深度可分离卷积进行空间维度上的卷积运算，并通过1×1逐点卷积（Pointwise Conv，PW）完成通道扩展。随后引入GELU激活函数以实现非线性映射，通过第二个1×1逐点卷积将特征映射回原通道维度，其处理后的特征为y_d如式（10）所示。

$$y_d = PW\Big(\phi\big(PW(DWConv_{3\times3}(y))\big)\Big),\tag{10}$$

式中：$DWConv$用于空间特征提取；PW用于通道维度上的线性变换。

最后进行特征融合，将经过空间增强的特征x_t和经过通道增强的特征y_d进行逐元素相加，并通过一个1×1卷积将通道数恢复至C，从而将两种注意力信息有效地融合在一起，如式（11）所示。

$$X = Conv_{1\times1}(x_t + y_d)。\tag{11}$$

此X即为ESTF模块的最终输出。

1.2.2　AFE模块

在ESTF输出特征的基础上，引入AFE模块进行特征精炼。该模块基于空间抑制理论，通过神经元能量函数评估重要性。该能量函数设计借鉴了SimAM注意力机制的思想[13]，将其能量评估机制与递归图的时序演化特征相结合，实现关键特征响应增强。通过Sigmoid函数映射，AFE自适应为高判别神经元分配权重，在不增加模型复杂度的前提下提升特征的表达能力。

AFE模块为输入的每个神经元计算一个能量函数E。对于一个特征图X，其能量函数的计算公式为

$$E = \frac{(X-\mu)^2}{4(\sigma^2+\lambda)} + 0.5,\tag{12}$$

式中：$\mu = \dfrac{1}{HW}\sum_{i=1}^{HW} X_i$为通道内的空间均值；$\sigma^2 = \dfrac{1}{HW}\sum_{i=1}^{HW}(X_i-\mu)^2$为通道内的空间方差；常数4为归一化系数，用于将特征偏差项$(X-\mu)^2$缩放至合理区间，避免因数值过大导致能量函数饱和；0.5为偏移量，提升关键特征与非关键特征的区分度；λ是一个微小的正则化常数，防止分母为零。

最终，精炼后的特征F_{proc}是通过对能量函数进行$Sigmoid$缩放后与原特征相乘得到的，如式（13）所示。

$$F_{\text{proc}} = X \odot Sigmoid(E)。\tag{13}$$

AESTF模块整体过程如以下算法所示：

算法：AESTF模块
输入：特征图$F_{in}\in R^{B\times C\times H\times W}$
输出：增强特征图F_{out}
1. 特征分离：$[x,y]\leftarrow Split(Conv(F_{in}))$
2. Branch X(空间注意力)：
$x_s\leftarrow DWConv(AMPool(x))$；$x_v\leftarrow Var(x)$
$G\leftarrow GELU(Conv(x_s\odot\alpha+x_v\odot\beta))$
$x_t\leftarrow x\odot Up(G)$
3. Branch Y(通道注意力)：
$y_d\leftarrow PW\left(GELU\left(PW(DWConv(y))\right)\right)$
4. 特征融合：$X\leftarrow Conv(x_t+y_d)$
5. 特征精炼：$F_{proc}\leftarrow X\odot Sigmoid(Energy(X))$
6. 残差连接：$F_{out}=F_{in}+(F_{in}+F_{proc})$
7. **返回**：F_{out}

通过上述设计，AESTF模块能够以一种协同的方式，从递归图中提取与入侵行为相关的空间与通道特征，有效提升了网络入侵检测的准确率。

1.3 MobileViT-AESTF网络架构

MobileViT-AESTF由卷积层、MV2模块和AESTF模块进行特征提取，然后使用全局平均池化后放入全连接层并通过SoftMax层输出分类结果。网络结构如表4所示。在模型架构中，*Input*为模块输入特征图的维度尺寸，*Operator*为特征层的处理操作类型，C_{out}为输出通道数，S为卷积核在空间维度滑动的像素步距[14]。

表4 MobileViT-AESTF网络结构表

Tab. 4 MobileViT-AESTF network structure

Input	*Operator*	C_{out}	*s*
3×256×256	Conv2d	16	2
16×128×128	MV2	32	1
32×128×128	MV2	48	2
48×64×64	MV2	48	1
48×64×64	MV2	48	1
48×64×64	AESTF	96	1
96×32×32	AESTF	160	1
160×16×16	AESTF	160	1
160×8×8	Conv2d	640	1
640×1×1	AvgPool2d	—	—
64×1×1	FC	—	—

2 实验与分析

2.1 数据预处理

本研究采用的数据集为CIC UNSW-NB15 Augmented Dataset[15]。

2.1.1 数据清洗

针对原始数据集中存在的类别不平衡、数据类型不一致、异常值等问题，实施了一系列的数据预处理策略：1）删除FlowID、SrcIP、DstIP非数值列，针对列“Flow Bytes/s”和“Flow Packets/s”，检测其中的inf和nan，移除这些异常值行；2）通过分层下采样方法使样本量达到相对平衡；3）采用Z-score标准化方法将数值特征转换为标准正态分布；4）将良性流量编码为0，各类攻击流量分配1～6的离散编码值。经数据清洗后的数据集类别及数量如表5所示。

表5 数据清洗后的数据集

Tab. 5 The dataset after data cleaning

编码	标签类别	个数
0	Benign	5 000
1	DoS	4 467
2	Exploits	5 000
3	Fuzzers	5 000
4	Generic	4 632
5	Reconnaissance	5 000
6	Shellcode	2 102

2.1.2 特征提取

本研究对数据清洗后的79个特征采用方差分析方法计算特征和标签类型之间的分数，进行重要性排序，不同特征数量下模型的准确率如表6所示，因此筛选出最具判别性的前35个特征，如表7所示。

表6 不同特征数量下模型的准确率

Tab. 6 The accuracy of the model with different numbers of features

特征数量	*ACC*/%
25	97.96
35	**99.84**
45	99.25

表7 CIC UNSW-NB15数据集特征及其得分

Tab. 7 Characteristics and scores of the CIC UNSW-NB15 augmented dataset

序号	特征名称	得分
1	Dst Port	12.5
2	Fwd Seg Size Min	9.5
3	Packet Length Max	6.4
4	Fwd Packet Length Max	5.3
5	Total Length of Fwd Packet	5.0
6	Average Packet Size	5.0
⋮	⋮	⋮
33	FIN Flag Count	2.4
34	Bwd IAT Total	2.3
35	Flow IAT Max	2.3

将选择出来的特征通过相空间重构技术构建递归矩阵，关于递归图RP图像转换的方法已在本文第1章中详细阐述。如图3所示为生成的二维图像数据。最终构建的数据集包含31 201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。

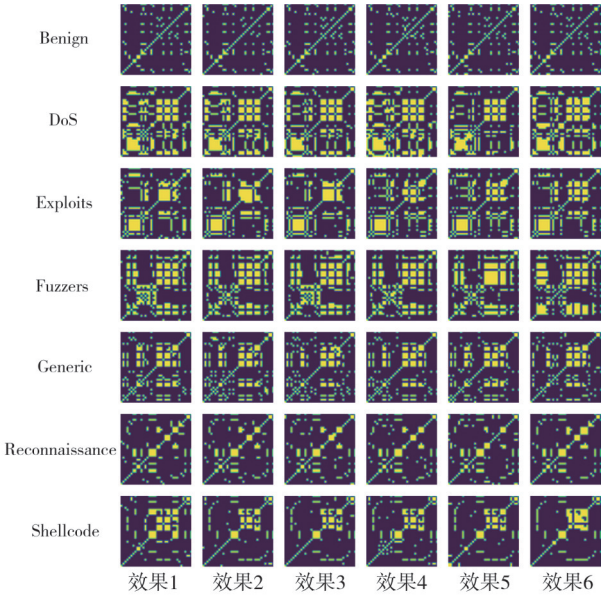

图 3　CIC UNSW-NB15数据集部分样例展示

Fig. 3　A portion of sample displays from the CIC UNSW-NB15 Augmented dataset

2.2　实验环境

本实验选用Python 3.9.13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。实验的训练与测试环境为Windows 11操作系统。

2.3　评价指标

本文通过图像分类任务常用评价指标[8]来测试和评估所提出模型的性能。

1）Accuracy（ACC）：准确预测量占总流量的百分比，准确率计算公式为

$$ACC=\frac{N_{TP}+N_{TN}}{N_{TP}+N_{TN}+N_{FP}+N_{FN}}\times 100\%。\quad(14)$$

2）Precision（P）：表示被分类为攻击流量数量占实际攻击流量的百分比，精确率的计算公式为

$$P=\frac{N_{TP}}{N_{TP}+N_{FP}}\times 100\%。\quad(15)$$

3）Recall（R）：表示估计预测的攻击数量占总攻击数量的百分比，召回率的计算公式为

$$R=\frac{N_{TP}}{N_{TP}+N_{FN}}\times 100\%。\quad(16)$$

4）$F1$-score（$F1$）对比：根据P和R返回对模型精度的更可靠度量，$F1$分数的计算公式为

$$F1=\frac{2N_{TP}}{2N_{TP}+N_{FP}+N_{FN}}\times 100\%,\quad(17)$$

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

2.4　对比实验

本节在数据集CIC UNSW-NB15上进行实验验证，系统评估了包括ShuffleNet、EfficientNet、MobileNet、MobileViT、SqueezeNet、FastViT、SBCFormer等主流轻量级模型以及本文提出的MobileViT-AESTF模型的性能。对CIC UNSW-NB15数据集上各轻量级模型的轻量化水平与分类性能进行系统分析，验证了改进模型在入侵检测方面的综合性能优势。

2.4.1　模型轻量级对比

基于CIC UNSW-NB15数据集的多分类入侵检测实验研究中，对上述一系列轻量化模型进行对比分析，如表 8 所示。

表 8　MobileViT-AESTF网络与轻量级网络各指标对比

Tab. 8　Comparison of MobileViT-AESTF network and lightweight network indicators

Network	参数量/10^6	FLOPs/10^9	ACC/%	$F1$/%	R/%	P/%
ShuffleNet_V2[16]	8.27	0.78	98.01	98.01	98.02	98.14
EfficientNet_B0[17]	5.12	0.54	98.12	98.21	98.16	98.28
MobileNet_V1[18]	3.21	0.59	98.25	98.29	98.30	98.29
MobileNet_V2[19]	3.50	0.43	97.31	97.12	97.34	98.15
MobileNet_V3[20]	5.58	0.31	95.88	95.95	95.20	96.25
MobileViT_V1[21]	1.02	0.35	98.28	98.45	98.34	98.35
MobileViT_V2[22]	4.39	1.45	98.17	98.21	98.31	98.45
MobileViT_V3_xxs[12]	0.73	0.28	98.34	98.54	98.46	98.44
MobileViT_V3_0.5[12]	1.17	0.38	98.56	98.58	98.58	98.60
FastViT[23]	3.14	0.27	98.70	98.43	98.64	98.24
SBCFormer[24]	5.36	0.92	99.32	99.36	99.33	99.42
SqueezeNet1_1[25]	**0.72**	**0.26**	86.59	83.65	86.59	92.15
MobileViT-AESTF(ours)	1.04	0.27	**99.84**	**99.82**	**99.85**	**99.82**

实验结果表明，MobileViT-AESTF通过自适应高效时空融合模块重构计算路径，达到了参数量和计算量的相对平衡。其参数量为1.04×10^6，虽略高于极简的SqueezeNet1_1（0.72×10^6）与Mobile-

ViT_V3_xxs(0.73×10⁶)，但通过动态特征复用机制，将计算量(FLOPs)压缩至ShuffleNet_V2的34.6%，较MobileNet_V1降低54.2%，甚至与参数量更低的SqueezeNet1_1近乎持平；相较于FastViT(3.14×10⁶，0.27GFLOPs)，参数量减少66.9%，计算量保持一致，而检测准确率提升1.14百分点；相较于SBCFormer(5.36×10⁶，0.92GFLOPs)，参数量仅为其19.4%，计算量压缩至29.3%，展现出更优的轻量化特性。

这种轻量化改进源于AESTF模块的异构注意力协同设计，即通过双路径分离与自适应特征精炼机制，在聚焦关键特征表达的同时减少冗余计算，实现了参数量与计算效率的优化平衡。

2.4.2 模型分类性能对比

表8表明，MobileViT-AESTF模型实现性能提升，准确率达99.84%，较EfficientNet_B0(98.12%)提升1.72百分点，较FastViT(98.70%)提升1.14百分点，较SBCFormer(99.32%)提升0.52百分点；精确率(99.82%)与召回率(99.85%)同步优化。$F1$值99.82%较MobileNet_V1(98.29%)提升1.53百分点，较FastViT(98.43%)提升1.39百分点，模型鲁棒性显著增强。

为了验证AESTF模块的有效性，本实验在相同训练条件下对比了改进模型MobileViT-AESTF与基准MobileViT_V3模型的分类准确率。如图4所示，改进模型在98%测试周期中保持99%以上准确率，而基准模型存在2次显著波动。

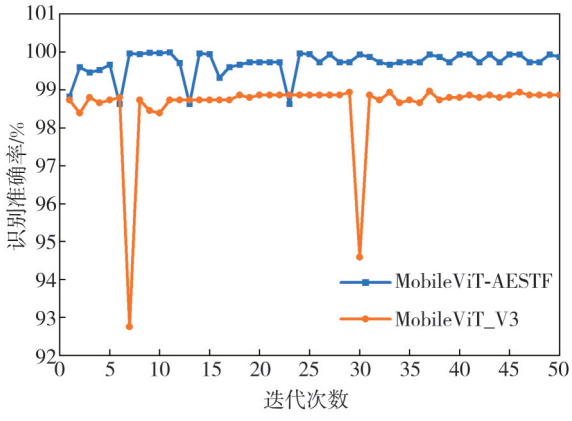

图4 模型准确率对比

Fig. 4 Comparison of model accuracy

实验结果验证了自适应高效时空融合模块AESTF对流量类别的稳定检测能力。

如图5所示，MobileViT-AESTF模型与基准模型的损失对比显示：改进模型的平均波动幅度(0.007 6)仅为基准(0.019 7)的38.6%；基准模型出现3次显著损失激增(第13、21、45周期，第47周期增幅达48.7%)，而改进模型仅在第34周期有一次短暂高损失，抑制了少数样本导致的梯度失衡；收敛稳定性上，改进模型自第36周期进入稳定收敛状态，体现了MobileViT-AESTF通过特征融合与残差连接优化带来的鲁棒性优势。

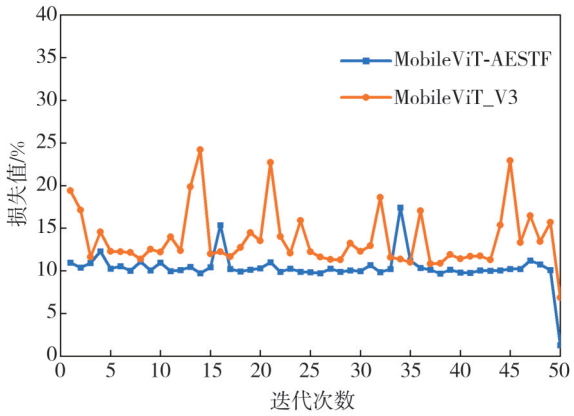

图5 模型损失值对比

Fig. 5 Comparison of model loss values

改进模型在CIC UNSW-NB15数据集上输出的混淆矩阵如图6所示。矩阵行表示真实类别，列对应预测结果。各主对角线元素反映类内检测准确率，非对角线元素表示误判情况。在对Fuzzers攻击进行分类时准确率为99.4%，其中0.6%的样本被误判为Dos攻击。其他样本均实现100%精确识别，验证了RP转换机制，将正常流量与攻击流量形成视觉可分性。

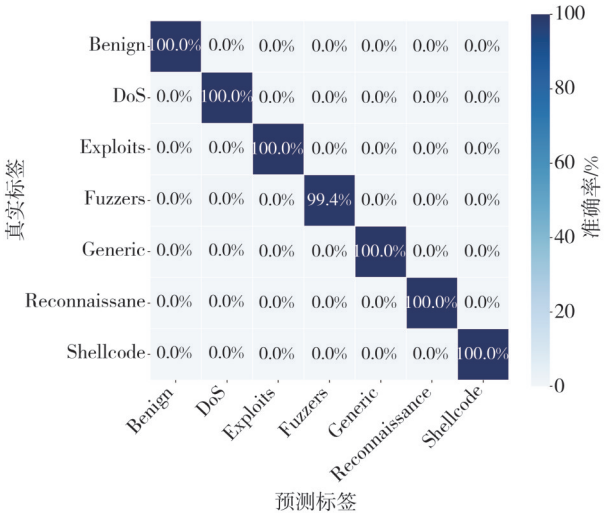

图6 混淆矩阵

Fig. 6 Confusion matrix

2.4.3 消融实验

为了深入验证所提出的AESTF模块中各个组件的有效性和协同作用，在CIC-UNSW-NB15数据

集上进行了系统性的消融实验。实验设计采用控制变量法，分别评估ESTF模块中的Branch X、Branch Y以及AFE模块对整体性能的贡献。

表9表明，当所有组件都被禁用时，模型取得98.34%的分类准确率，单独启用Branch X时，准确率略微下降至98.21%，表明Branch X在缺乏其他特征分支支持时存在一定的信息冗余。相比之下，单独启用Branch Y时，准确率提升至98.35%，显示出Branch Y具有更强的独立特征提取能力。当同时启用Branch X和Branch Y时，准确率提升至98.96%，相比基线配置提升了0.62百分点，验证了ESTF模块内部两个分支的互补性和协同效应。单独使用AFE模块时，模型准确率达到98.76%，相比基线提升了0.42百分点，证明了自适应特征增强机制的有效性。

表9　消融实验结果

Tab. 9　Ablation experiment results

ESTF		AFE	*ACC*/%
Branch X	Branch Y		
×	×	×	98.34
√	×	×	98.21
×	√	×	98.35
√	√	×	98.96
×	×	√	98.76
√	×	√	99.02
×	√	√	98.89
√	√	√	**99.84**

进一步验证组件协同效应发现：仅集成Branch X与AFE时，准确率为99.02%（较完整AESTF低0.82百分点），说明Branch Y的通道增强对特征交互至关重要；仅集成Branch Y与AFE时准确率为98.89%（低0.95百分点），进一步佐证Branch X在空间特征定位中的核心作用。当同时集成ESTF与AFE时，模型最终达到99.84%的准确率，较基准提升1.50百分点。这一结果表明：ESTF构建的空间、通道双注意力处理机制为AFE提供了有效的输入特征，而AFE通过动态能量缩放进一步强化了特征判别性，两者在特征处理过程中相互配合，提升了检测效率。

3　结　论

针对入侵检测中流量图像化存在的时空特征提取不足与模型资源消耗较大的问题，本文提出了一种基于递归图RP和轻量化模型MobileViT-AESTF的网络入侵检测方法。实验表明，该模型在入侵检测任务中达到99.84%的分类准确率，验证了ESTF（空间注意力＋通道注意力）对时空特征的有效融合，以及AFE对特征的自适应精炼能力，同时保持了模型轻量化。

当前研究仍存在两点局限性：1）AFE采用固定正则化常数λ，难以适应多样流量特征的分布差异；2）ESTF与AFE的连接方式未充分挖掘模块间特征交互的潜力。未来研究将聚焦以下3个方向：

1）引入自适应λ参数的能量函数，根据通道特征的均值与方差动态调整，提升特征精炼的针对性；

2）设计跨模块特征反馈机制，将AFE特征精炼后的结果融入ESTF的注意力计算，强化二者协同效应；

3）优化模型压缩与量化策略（如剪枝、权重量化），减少参数量与计算量，提升分类速度以适应边缘设备实时检测需求。

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] Black D. Cybercrime will cost $12TN next year, say experts[EB/OL]. (2024-01-24)[2025-08-04]. https://cybernews.com/news/cybercrime-will-cost-trillions-next-year/.

[2] 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.

[3] Han Y L, Han X W. ICPS multi-target constrained comprehensive security control based on DoS attacks energy grading detection and compensation[J]. Journal of Measurement Science and Instrumentation, 2024, 15(4): 518-531.

[4] Demmese F A, Neupane A, Khorsandroo S, et al. Machine learning based fileless malware traffic classification using image visualization[J]. Cybersecurity, 2023, 6(1): 1-18.

[5] 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)
[6] Li S M, Chai G Z, Wang Y H, et al. CRSF: an intrusion detection framework for industrial internet of things based on pretrained CNN2D-RNN and SVM [J]. IEEE Access, 2023, 11: 92041-92054.
[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] 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)
[9] Tiwari R S, Lakshmi D, Das T K, et al. A lightweight optimized intrusion detection system using machine learning for edge-based IIoT security [J]. Telecommunication Systems, 2024, 87(3): 605-624.
[10] Alshehri M S, Saidani O, Alrayes F S, et al. A self-attention-based deep convolutional neural networks for IIoT networks intrusion detection [J]. IEEE Access, 2024, 12: 45762-45772.
[11] Al-Jarrah O Y, El Haloui K, Dianati M, et al. A novel detection approach of unknown cyber-attacks for intra-vehicle networks using recurrence plots and neural networks[J]. IEEE Open Journal of Vehicular Technology, 2023, 4: 271-280.
[12] Wadekar S N, Chaurasia A, Mishra A. MobileViTv3: enhanced fusion for mobile vision transformers [C]//Proceedings of the European Conference on Computer Vision, 2023: 412-429.
[13] Yang L X, Zhang R Y, Li L D, et al. SimAM: A simple, parameter-free attention module for convolutional neural networks[C]//38th International Conference on Machine Learning, 2021: 11863-11874.
[14] Wang Y, Li Y S, Wang G, et al. Multi-scale attention network for single image super-resolution[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2024: 5950-5960.
[15] Canadian Institute for Cybersecurity (CIC), University of New Brunswick. CIC UNSW-NB15 Augmented Dataset [DB/OL]. [2025-09-07]. https://unb.ca/cic/datasets/cic-unsw-nb15.html.
[16] Ma N N, Zhang X Y, Zheng H T, et al. Shufflenet v2: Practical guidelines for efficient cnn architecture design[C]//European Conference on Computer Vision (ECCV), 2018: 116-131.
[17] Chen X, Liu Z, Wang Y, et al. Efficient mobile transformer for edge-AI security applications[C]//International Conference on Machine Learning, 2023: 2689-2699.
[18] Howard A G, Zhu M, Chen B, et al. Mobilenets: efficient convolutional neural networks for mobile vision applications[PP/OL]. Vl. arXiv(2017-04-17)[2025-09-07]. https://arxiv.org/abs/1704.04861.
[19] Sandler M, Howard A, Zhu M L, et al. MobileNetV2: Inverted residuals and linear bottlenecks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018: 4510-4520.
[20] Howard A, Sandler M, Chen B, et al. Searching for MobileNetV3 [C]//2019 IEEE/CVF International Conference on Computer Vision(ICCV), 2019: 1314-1324.
[21] Mehta S, Rastegari M. Mobilevit: light-weight, general-purpose, and mobile-friendly vision transformer [PP/OL]. Vl. arXiv(2021-10-05)[2025-09-07]. https://arxiv.org/abs/2110.02178.
[22] Mehta S, Rastegari M. Separable self-attention for mobile vision transformers[PP/OL]. Vl. arXiv(2022-06-06)[2025-09-07]. https://arxiv.org/abs/2206.02680.
[23] Vasu P K A, Gabriel J, Zhu J, et al. FastViT: a fast hybrid vision transformer using structural reparameterization [C]//IEEE/CVF International Conference on Computer Vision(ICCV), 2024: 5785-5794.
[24] Lu X Y, Suganuma M, Okatani T. SBCFormer: lightweight network capable of full-size imagenet classification at 1 fps on single board computers[C]//IEEE/CVF Winter Conference on Applications of Computer Vision, 2024: 1123-1133.
[25] Liu C, Zhang H, Wang Z, et al. Squeeze-attention: Lightweight network for real-time intrusion detection [J]. IEEE Transactions on Dependable and Secure Computing, 2024, 21(3): 1987-1999.