基于并联分支联合编码的网络恶意流量分类研究

马自强[1,2,3]，崔梦真[1]，杨天宇[1]，张宁宁[4]
（1.宁夏大学信息工程学院，宁夏 银川 750021；2.宁夏大数据与人工智能省部共建协同创新中心，宁夏 银川 750021；
3.宁夏“东数西算”人工智能与信息安全重点实验室，宁夏 银川 750021；
4.国家互联网应急中心宁夏分中心，宁夏 银川 750021）

摘　要：针对单分支模型无法完整表述流量信息且并联分支模型无法很好地融合各个支路提取特征的问题，提出了一种基于并联分支联合编码的网络恶意流量分类模型。在特征提取方面，采用分裂注意力残差网络和双向长短期记忆网络（Bi-LSTM）对原始流量分别提取空间特征和时序特征。在分支融合方面，采用交叉多头自注意力机制来利用分支间的关联性，并获得有效的特征融合。最终输入全连接层中进行网络恶意流量分类。在公开数据集USTC-TFC2016上的广泛实验表明，所提模型在准确率、精确度、召回率和F1值等关键性能指标上均表现出显著优势；在针对新型恶意流量的在线学习能力评估中，所提模型同样表现出优越的性能。

关键词：流量分类；并联分支联合；双向长短期记忆网络；交叉多头自注意力机制
中图分类号：TN919
文献标志码：A
DOI：10.11959/j.issn.1000–436x.2025154

Research on network malicious traffic classification based on parallel branch joint coding

MA Ziqiang[1,2,3], CUI Mengzhen[1], YANG Tianyu[1], ZHANG Ningning[4]
1. College of Information Engineering, Ningxia University, Yinchuan 750021, China
2. Collaborative Innovation Center for Ningxia Big Data and Artificial Intelligence Co-founded by Ningxia Municipality and Ministry of Education, Yinchuan 750021, China
3. Ningxia Key Laboratory of Artificial Intelligence and Information Security for Channeling Computing Resources from the East to the West, Yinchuan 750021, China
4. Ningxia Internet Emergency Center Ningxia Branch, Yinchuan 750021, China

Abstract: To address the limitations of single-branch models in fully representing traffic information and parallel branch models in effectively fusing features extracted from multiple pathways, a malicious traffic classification model based on parallel branch joint coding was proposed. For feature extraction, the split attention residual network and the bidirectional long short-term memory network (Bi-LSTM) were employed to extract spatial features and temporal features from the original traffic, respectively. For branch fusion, the cross multi-head self-attention mechanism was adopted to exploit inter-branch associations and obtain an effective fusion of features. Finally, the fused features were input into the fully connected layer for malicious network traffic classification. Extensive experiments on the public dataset USTC-TFC2016 demonstrate that the proposed model exhibits significant advantages across key performance metrics, including accuracy, precision, recall, and F1-score. Moreover, in the evaluation of the online learning ability for new types of malicious traffic, the model also demonstrates superior performance.

Keywords: traffic classification, parallel branch joint, bidirectional long short-term memory network, cross multi-head self-attention mechanism

收稿日期：2025–04–21；修回日期：2025–08–19
通信作者：马自强，ziqiang@nxu.edu.cn
基金项目：宁夏自然科学基金资助项目（No.2020AAC03035）
Foundation Item：Ningxia Natural Science Foundation (No.2020AAC03035)

0 引言

随着互联网的迅速发展，网络流量的增长速度也在不断加快。近年来，网络安全问题也越来越受到重视。网络中的任何入侵都会影响许多领域[1]。因此，保证网络的安全变得尤为关键。网络恶意流量分类的目的是将不同的网络流量组织成不同的类别[2]。精确的网络流量分类扮演着重要角色[3]，它是网络管理和安全的基础和关键。

目前，按照流量分类方法的思路不同大致可将其分为3类：基于端口的方法、基于深度包检测（DPI, deep packet inspection）的方法和基于深度学习的方法。基于端口的方法通过假设大多数应用程序使用的默认端口号来推断服务或应用程序的类型。然而，端口伪装、端口随机和隧道技术等方法的出现使该方法很快失效。基于深度包检测的方法在侵犯用户隐私的同时无法检测未知的恶意流量。因此，近年来，研究人员选择使用基于深度学习的方法分类网络恶意流量[4-5]。该方法先从流量中提取各种特征，再使用深度学习模型来检测恶意行为。周成胜等[6]先通过提取多会话特征数据并转换为图像，再利用深度学习在图像识别领域的优势，将加密流量识别问题转换为图像识别问题。王梦寒等[7]使用CICFlowmeter工具提取数据包的80多维特征信息，然后提出了一种基于随机森林–双向长短期记忆网络（Bi-LSTM, bidirectional long short-term memory）的网络异常流量检测方法。Jablaoui等[8]将卷积神经网络（CNN, convolutional neural network）和长短期记忆网络用于处理网络流量数据的空间特性和时序特性，通过结合不同模型的优点，提高网络异常流量检测的准确性和效率。然而，这些方法有2个缺点。首先，现有的方法不能从不同的方面提取流量特征，提取的特征较片面，导致模型难以全面捕捉流量的复杂结构和动态变化，进而影响分类的准确性和泛化能力。其次，现有的并联分支模型将各个分支结果进行拼接，没有考虑到分支间复杂的关联关系，进而影响模型的性能。

针对以上问题，本文提出了并联分支联合编码的网络恶意流量分类模型。该模型首先将数据包转化为灰度图像，不需要人工提取和选择网络流量的特征，从而适用于所有类型的流量；然后分别采用分裂注意力残差网络和Bi-LSTM并行地对原始流量图像提取空间特征和时序特征，充分利用了流量的多维信息。特征融合模块引入交叉多头自注意力机制，动态调整2个分支的特征表示，交叉多头自注意力机制通过多头的设计，能够从多个不同的子空间中捕捉分支间的关系。每个头可以学习到空间特征与时序特征的关联信息，从而全面地挖掘分支之间的复杂关系，实现更有效的特征融合。单头自注意力机制上下文信息局限于当前的序列，其捕捉的依赖关系相对单一。相比单头自注意力机制，交叉多头自注意力机制通过在不同序列间交互，可以引入额外的上下文信息。最后将融合的特征输入全连接层进行恶意流量分类。实验结果表明，该方法可以有效地提取和融合流量图像的空间特征和时序特征，在网络恶意流量分类的关键性能指标上表现出优越的性能，并具有较好的泛化能力和在线学习能力。

综上所述，本文的主要贡献如下。

1)针对提取的特征较片面的问题，本文设计了分裂注意力残差网络和Bi-LSTM对原始流量分别提取空间特征和时序特征。

2)针对并联分支模型未能充分考虑分支间复杂关联关系的问题，本文提出了基于交叉多头自注意力机制的特征融合模块。该模块通过动态调整空间特征和时序特征的表示，挖掘2种特征之间的关联互补信息，显著提升了模型的分类性能和泛化能力。

3)针对并联分支联合编码模型是否能达到好的效果的问题，在公开数据集USTC-TFC2016上开展了广泛实验。实验结果表明，该模型在准确率、精确度、召回率和F1值等关键性能指标上具有显著优势，这意味着在实际网络环境中，所提模型能够更精准地识别不同类别的流量，减少漏报和误报情况，有效降低网络遭受攻击的风险。

1 相关工作

1.1 深度学习在流量分类中的应用

由于端口随机、端口分配、流量加密等，传统的恶意流量分类失效。近年来，研究人员将研究方向转向了基于统计特征和基于深度学习的恶意流量分类方法。基于统计特征的分析方法不依赖于对网络协议内容的详细解析，即使在流量加密的情况下，仍然可以通过分析流量的统计特征检测恶意行为且能够适应不同类型的网络和应用环境。但是选

择合适的统计特征对结果影响较大，如果特征选择不当，可能会导致模型性能下降。深度学习降低特征选择的影响，它的一个重要优势是可以直接从原始数据进行表征学习[9]，而不是人工选择特征。Wang 等[10]首次将 USTC-TFC2016 数据集应用于异常流量分类的模型训练中，并使用 CNN 构建模型。Ferrag 等[11]创建了一个名为 Edge-IIoTset 的网络安全数据集，并验证了在该数据集上使用机器学习和深度学习方法的性能。Ma 等[12]提出了多尺度稀疏时频交叉注意力网络（MSTFCAN），并提出了一种增强和融合多尺度趋势和季节性特征的机制，同时利用稀疏时频交叉注意机制来提取和融合每个尺度的时域和频域特征。Ullah 等[13]提出了基于 Transformer 的不平衡网络流量迁移学习入侵检测系统（IDS-INT），借助 Transformer 迁移学习技术，利用其语义锚点对详细的特征表示进行学习，结合合成少数类过采样技术（SMOTE, synthetic minority over-sampling technique）平衡异常流量，利用 CNN 提取深度特征，再以 CNN-LSTM 混合模型实现多类型攻击检测。从之前的研究可以发现，基于深度学习的恶意流量分类研究已有了长足的发展，但是目前研究大多基于规则即依赖特征集，且大多是单一的模型，很少研究并联分支模型在恶意流量分类方面的应用。因此，本文提出了基于并联分支编码的网络恶意流量分类模型，该模型不但可以并行地提取流量图像的空间特征和时序特征，还挖掘了 2 种特征的关联关系，从而有效地融合了 2 种特征。

1.2　并联分支联合网络的应用

大部分卷积网络设计的研究都集中在深度、滤波器大小和特征通道数量等方面的实证调查，但最近的研究指出，采用分支结构，即将计算沿着并行但不同的线程进行拆分，然后聚合不同线程的输出，可以显著提高性能[14-17]。Ahmed 等[18]提出了一种用于学习网络中分支之间连接的算法。该算法通过优化一个针对最终任务定义的单个损失函数来同时学习网络的权值。作者使用了 3 个不同的数据集展示算法的图像分类的效果，结果显示，该算法性能优于深度神经网络的聚合残差变换网络（ResNeXt, aggregated residual transformation for deep neural network）。Fan 等[19]提出了一种多分支注意 Transformer，其中注意力层是多个分支的平均值，每个分支是一个独立的多头注意力层。在机器翻译、代码生成和自然语言理解方面，这种 Transformer 变体带来了显著的改进。Hao 等[20]提出了一种用于 12-导联心电图图像心肌梗死自动筛查的多分支融合框架，该框架由多分支网络、特征融合和分类网络组成。实验结果表明，该方法对基于心电图像的心肌梗死筛查是有效的。Chen 等[21]提出了一种基于多分支局部注意网络（MBLANet）的遥感图像场景分类方法，该方法将卷积局部注意模块（CLAM, convolutional local attention module）嵌入残差网络（ResNet, residual network）主干的所有下采样块和残差块中。在 3 个数据集上的实验表明，该方法优于最先进的方法。宋玉琴等[22]提出了一种基于注意力机制的多分支特征级联图像去雨网络模型。该模型结合多种注意力机制，形成不同类型的多分支网络，同时在网络分支间构建了阶段注意融合机制。实验结果表明，该算法主观视觉效果有效提升，去雨能力更强。以往的工作表明，多分支联合模型在各个领域中表现出色。因此，本文提出了并联分支联合编码模型，将其应用于网络恶意流量分类领域，验证该模型在网络恶意流量分类的性能。

2　基于并联分支联合编码的恶意流量分类

在已有的关于流量分类的工作中，研究人员从不同的角度提取了流量的特征。周子云等[23]提出一种改进 EfficientNet 来增强流量图像空间特征的有效提取。赵忠斌等[24]通过融合多头注意力机制提取流量的时序特征。以往的工作单方面地提取流量图像的空间特征或时序特征，会造成特征提取较片面的问题。魏德宾等[25]采用混合注意力的 Transformer 和 Bi-LSTM 的双模态网络提取全局特征和时空特征，再加权融合 2 种特征用于流量分类，这种方式没有考虑到 2 种特征的关联。基于以上问题，本文提出了基于并联分支联合编码的网络恶意流量分类模型，该模型由 4 个模块组成，分别为数据预处理、特征提取、特征融合和流量分类。在数据预处理模块中，将流量数据转化为灰度图。在特征提取模块中，分别采用分裂注意力残差网络和 Bi-LSTM 提取流量图像的空间特征和时序特征。在特征融合模块中，采用了交叉多头自注意力机制，使用空间特征来调整时序特征的特征表示和使用时序特征来调整空间特征的特征表示，并对 2 种特征的

相互作用建模，提取了2种特征间的关联互补信息，融合后的特征不仅包含了空间和时序的独立信息，还包含了它们之间的关联信息。在流量分类模块中，将融合的特征输入全连接层进行流量分类。模型结构如图1所示。

2.1 特征提取

特征提取模块分为空间特征提取模块和时序特征提取模块，流量图像中可能会呈现出特定的形状和纹理。空间特征描述流量数据在二维平面上的分布模式或结构，反映不同位置（如图像像素）之间的关联性或差异性。时序特征关注的是流量图像随时间的变化情况，描述流量数据在时间维度上的变化规律。特征提取模块采用并行处理的方式，允许模型同时从不同的角度提取流量数据的特征。这种设计确保空间特征和时序特征在提取过程中互不干扰，从而保留更多的原始信息。具体而言，空间特征提取模块通过分裂注意力残差网络捕捉流量图像的局部结构和全局模式，时序特征提取模块则通过Bi-LSTM捕捉流量数据在时间维度上的动态变化。

2.1.1 空间特征提取

空间特征提取采用分裂注意力残差网络来提取流量图像的空间特征，分裂注意力残差网络使用不同的卷积以提取不同尺度的特征信息。对于图像而言，多尺度的信息有助于网络更好地对图像信息进行选择，但多尺度带来的问题是计算量的增加，所以采用1×1的卷积用于减少特征的维度，再采用多尺度的结构提取特征信息。对于分裂注意力残差卷积网络，分为以下几个步骤。

1)特征图分组。将特征图分为几个组，特征图组的数量由超参数n给出，将得到的特征图组称为分组1,…,分组n，并引入了新的超参数k，k表示分组内的分裂数，将这些分裂的组称为子组1,…,子组k。因此，特征组的总数$G=n\times k$，每个分组的组合表示由分组内的子组经过卷积、注意力机制，之后沿着通道维度拼接后得到。其中，注意力机制用来确定每个子组的特征表示的权重组合，以便更好地捕获重要的特征并抑制不重要的信息，并采用残差网络减轻深度神经网络中的梯度消失问题，使网络学习更深层次的特征表示。

2)分组的组合表示。每个子组首先分别经过3×3、1×1、3×3卷积，这样做的目的是在提取不同尺度的特征信息的同时减少计算量，可以提高模型对复杂数据的表达能力和学习能力。其次经过注意力机制，用来计算每个通道的注意力权值，根据计算

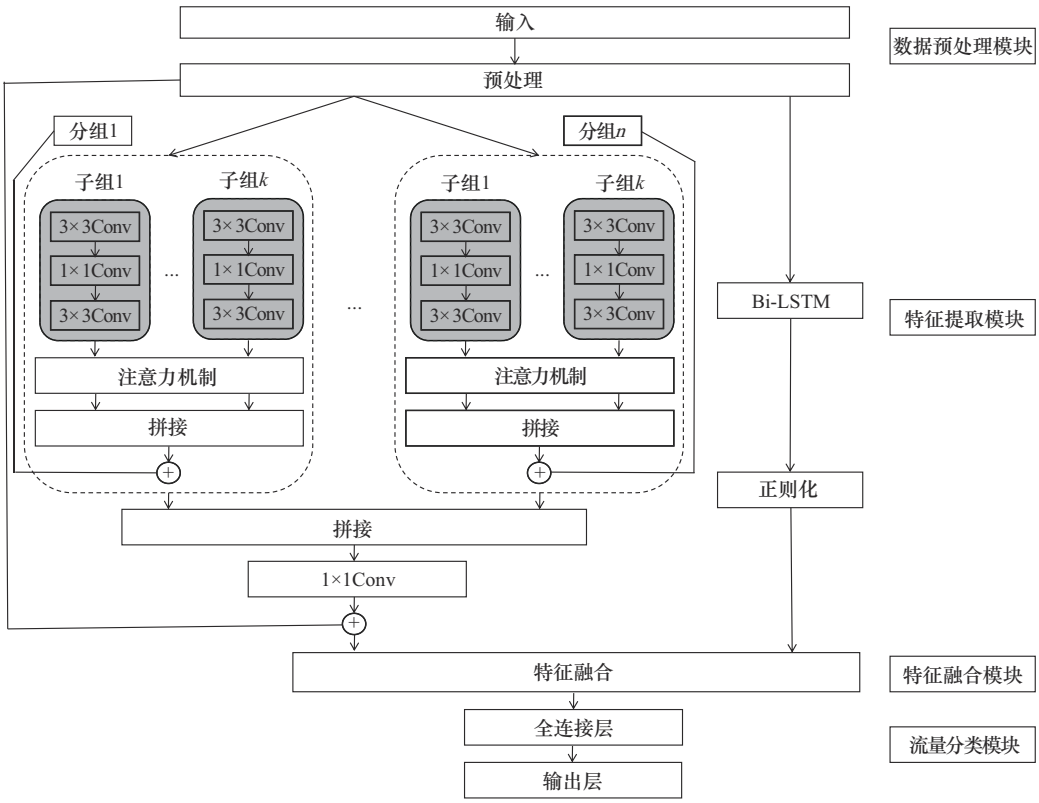


图1 基于并联分支联合编码的网络恶意流量分类模型结构

得到的注意力权值，对之前3×3、1×1、3×3卷积层处理的特征进行加权融合。然后将每个分组的特征沿着通道维度拼接，借助跳跃连接机制（跳跃连接是深度学习中一种在网络层间建立直接连接的技术，使输入信息能跨越部分层与经过这些层处理后的输出相加），将拼接后的特征与尚未进行分裂处理的分组特征执行相加操作，有助于更快地传递重要信息，减少信息的丢失。经过这些操作之后，最终得到分组的组合表示。

3)分裂注意力残差卷积网络的输出。将每个分组的组合表示拼接起来，经过1×1的卷积层后和原始的流量图像相加实现跳跃连接，得到分裂注意力残差卷积网络的输出，即流量图像的空间特征。

2.1.2　时序特征提取

采用Bi-LSTM来提取流量的时序特征，LSTM能在一定程度上缓解传统循环神经网络（RNN, recurrent neural network）存在的梯度爆炸和梯度消失问题，Bi-LSTM是LSTM的一种变体形式，可以通过正序和逆序2种方式学习数据的时序关系。将流量图像通过Bi-LSTM网络进行处理，再经过Dropout层减少过拟合，得到Bi-LSTM网络的输出，即流量图像的时序特征。

2.2　特征融合

为了挖掘2个分支间更有效的关联，通过构建特征融合模块来获得2个分支的联合特征。该模块主要由交叉多头自注意力机制、正则化层、残差连接和归一化层组成，结构如图2所示。

交叉多头自注意力机制通过在2个分支之间交叉计算注意力分数，来增强2个分支的表达能力。首先，通过空间特征来调整时序特征。假设空间特征表示为X_a，时序特征的表示为X_b，对于空间特征X_a，使用多头自注意力机制得到X_s，其中$\boldsymbol{K}$、$\boldsymbol{Q}$、$\boldsymbol{V}$来源于X_a，对于每个头$i(i=1,\cdots,h)$，对输入的键矩阵$\boldsymbol{K}$、查询矩阵$\boldsymbol{Q}$和值矩阵$\boldsymbol{V}$分别进行线性变换，得到每个头对应的查询、键和值矩阵，其计算过程如式(1)~式(3)所示。其次，计算每个头的注意力输出，过程如式(4)和式(5)所示。再次，将h个头的输出沿着特征维度拼接起来得到X_s，计算过程如式(6)所示。对于时序特征X_b，将时序特征X_b的$\boldsymbol{K}$和$\boldsymbol{V}$切换为空间特征X_a的$\boldsymbol{K}$和$\boldsymbol{V}$，$\boldsymbol{Q}$来源于X_b，时

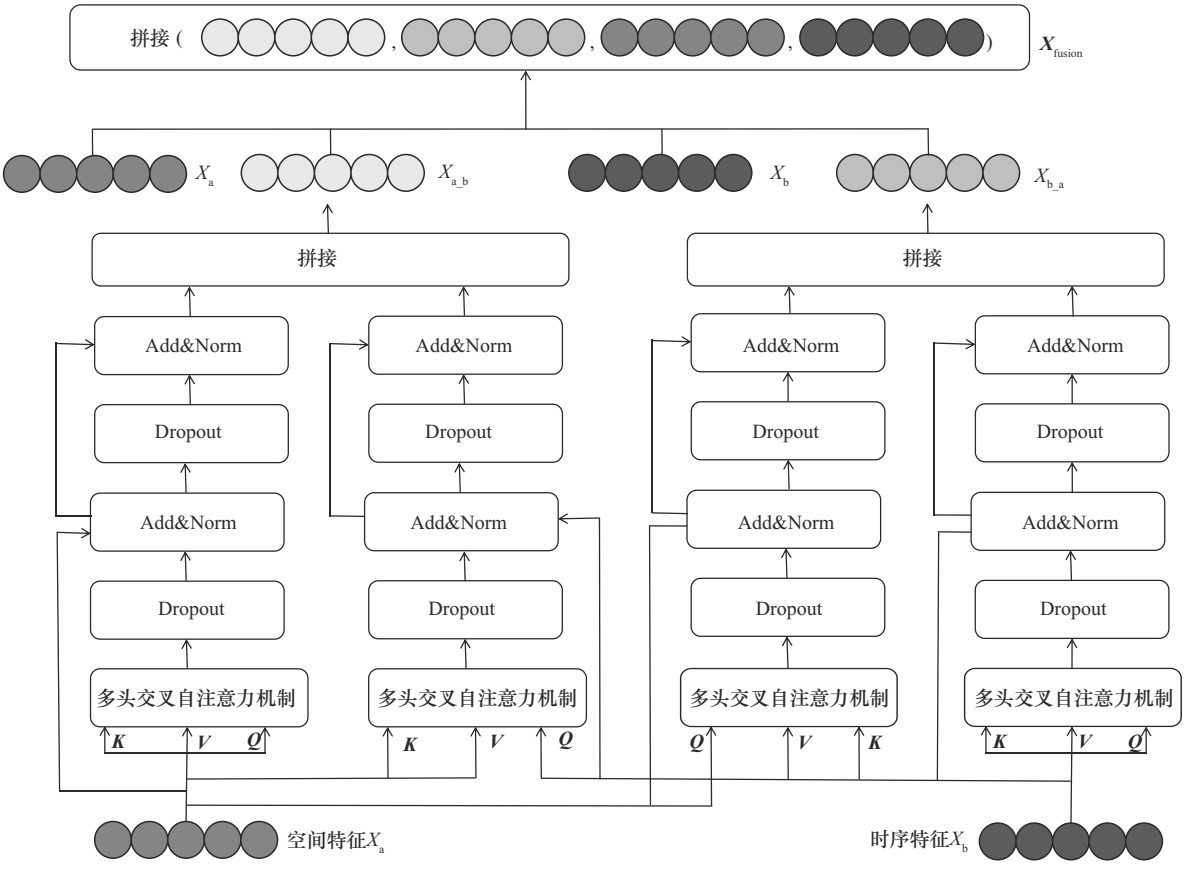


图2　特征融合模块结构

序特征的查询矩阵计算过程如式(7)所示。通过计算$\boldsymbol{Q}$和$\boldsymbol{K}$的点积得到注意力分数，其作用相当于衡量时序特征的每个位置和空间特征的每个位置的匹配程度，较高的注意力分数表示2个位置之间的关联较强，计算过程如式(8)和式(9)所示。最后，将h个头的输出沿着特征维度拼接起来得到X_{t}，计算过程如式(10)所示。将得到的特征X_{s}和X_{t}分别输入正则化层、残差连接和归一化层得到X_{a}'和X_{b}'，并将X_{a}'和X_{b}'进行拼接，获得空间特征调整时序特征的联合特征$X_{\mathrm{a_b}}$，$X_{\mathrm{a_b}}$的计算过程如式(11)所示。

$$\boldsymbol{Q}_{\mathrm{a}} = \boldsymbol{Q}\boldsymbol{W}_{\mathrm{a}}^{\boldsymbol{Q}} \tag{1}$$

$$\boldsymbol{K}_{\mathrm{a}} = \boldsymbol{K}\boldsymbol{W}_{\mathrm{a}}^{\boldsymbol{K}} \tag{2}$$

$$\boldsymbol{V}_{\mathrm{a}} = \boldsymbol{V}\boldsymbol{W}_{\mathrm{a}}^{\boldsymbol{V}} \tag{3}$$

$$\mathrm{Attention}\left(\boldsymbol{Q}_{\mathrm{a}}^{i},\boldsymbol{K}_{\mathrm{a}}^{i},\boldsymbol{V}_{\mathrm{a}}^{i}\right) = \mathrm{Softmax}\left(\frac{\boldsymbol{Q}_{\mathrm{a}}^{i}\left(\boldsymbol{K}_{\mathrm{a}}^{i}\right)^{\mathrm{T}}}{\sqrt{d_k}}\right)\boldsymbol{V}_{\mathrm{a}}^{i} \tag{4}$$

$$\mathbf{head}_{\mathrm{a}}^{i} = \mathrm{Attention}\left(\boldsymbol{Q}_{\mathrm{a}}^{i},\boldsymbol{K}_{\mathrm{a}}^{i},\boldsymbol{V}_{\mathrm{a}}^{i}\right), i = 1,\cdots,h \tag{5}$$

$$X_{\mathrm{s}} = \mathrm{Multihead}\left(\boldsymbol{Q}_{\mathrm{a}},\boldsymbol{K}_{\mathrm{a}},\boldsymbol{V}_{\mathrm{a}}\right) = \mathrm{Concat}\left(\mathbf{head}_{\mathrm{a}}^{1},\cdots,\mathbf{head}_{\mathrm{a}}^{h}\right) \tag{6}$$

其中，$\boldsymbol{W}_{\mathrm{a}}^{\boldsymbol{Q}}$、$\boldsymbol{W}_{\mathrm{a}}^{\boldsymbol{K}}$、$\boldsymbol{W}_{\mathrm{a}}^{\boldsymbol{V}}$是可学习的权重矩阵，$d_k$是矩阵$\boldsymbol{Q}_{\mathrm{a}}^{i}$、$\boldsymbol{K}_{\mathrm{a}}^{i}$的列数，$\mathbf{head}_{\mathrm{a}}^{i}$代表第$i$个注意力头的注意力矩阵。

$$\boldsymbol{Q}_{\mathrm{b}} = \boldsymbol{Q}\boldsymbol{W}_{\mathrm{b}}^{\boldsymbol{Q}} \tag{7}$$

$$\mathrm{Attention}\left(\boldsymbol{Q}_{\mathrm{b}}^{i},\boldsymbol{K}_{\mathrm{a}}^{i},\boldsymbol{V}_{\mathrm{a}}^{i}\right) = \mathrm{Softmax}\left(\frac{\boldsymbol{Q}_{\mathrm{b}}^{i}\left(\boldsymbol{K}_{\mathrm{a}}^{i}\right)^{\mathrm{T}}}{\sqrt{d_k}}\right)\boldsymbol{V}_{\mathrm{a}}^{i} \tag{8}$$

$$\mathrm{head}_{\mathrm{b}}^{i} = \mathrm{Attention}\left(\boldsymbol{Q}_{\mathrm{b}}^{i},\boldsymbol{K}_{\mathrm{a}}^{i},\boldsymbol{V}_{\mathrm{a}}^{i}\right), i = 1,\cdots,h \tag{9}$$

$$X_{\mathrm{t}} = \mathrm{Multihead}\left(\boldsymbol{Q}_{\mathrm{b}},\boldsymbol{K}_{\mathrm{a}},\boldsymbol{V}_{\mathrm{a}}\right) = \mathrm{Concat}\left(\mathbf{head}_{\mathrm{b}}^{1},\cdots,\mathbf{head}_{\mathrm{b}}^{h}\right) \tag{10}$$

其中，$\boldsymbol{W}_{\mathrm{b}}^{\boldsymbol{Q}}$是可学习的权重矩阵，$d_k$是矩阵$\boldsymbol{Q}_{\mathrm{b}}^{i}$、$\boldsymbol{K}_{\mathrm{a}}^{i}$的列数，$\mathbf{head}_{\mathrm{b}}^{i}$代表第$i$个注意力头的注意力矩阵。

$$X_{\mathrm{a\_b}} = \mathrm{Concat}\left(X_{\mathrm{a}}',X_{\mathrm{b}}'\right) \tag{11}$$

其中，X_{a}'和X_{b}'分别代表X_{s}和X_{t}输入正则化层、经过残差连接并通过归一化层后所得到的结果。

通过时序特征来调整空间特征的自注意力，首先，将空间特征X_{a}的$\boldsymbol{K}$和$\boldsymbol{V}$切换为时序特征X_{b}的$\boldsymbol{K}$和$\boldsymbol{V}$，通过计算$\boldsymbol{Q}$和$\boldsymbol{K}$的点积得到注意力分数，来获取空间特征的每个位置和时序特征的每个位置的匹配程度。然后，将得到的特征向量分别输入正则化层、残差连接和归一化层得到X_{a}''和X_{b}''，并将X_{a}''和X_{b}''进行拼接，获得时序特征调整空间特征的联合特征$X_{\mathrm{b_a}}$。最后，将联合特征$X_{\mathrm{a_b}}$、$X_{\mathrm{b_a}}$和没有经过融合的空间特征X_{a}和时序特征X_{b}拼接起来，得到最终的特征向量$\boldsymbol{X}_{\mathrm{fusion}}$，如式(12)所示。

$$\boldsymbol{X}_{\mathrm{fusion}} = \mathrm{Concat}\left(X_{\mathrm{a\_b}},X_{\mathrm{b\_a}},X_{\mathrm{a}},X_{\mathrm{b}}\right) \tag{12}$$

其中，$X_{\mathrm{a_b}}$和$X_{\mathrm{b_a}}$分别是通过空间特征调整时序特征和时序特征调整空间特征的结果，X_{a}和X_{b}分别是输入特征融合模块的时间特征和时序特征。

特征融合模块通过动态调整各特征的重要性，有助于模型自适应地加权特征，从而增强特征选择和融合的灵活性。通过2种特征相互调整，可以学习到特征之间的关联关系，进而形成新的组合特征，更好地拟合训练数据来提升模型的性能。

3 数据预处理

3.1 实验数据集介绍

本次实验的数据集来源是公开数据集USTC-TFC2016。数据集主要包括2个部分，恶意流量和正常流量。第一部分是2011—2015年CTU研究人员从真实的网络环境中收集到的10种恶意软件流量[26]。恶意流量信息如表1所示。

表1 恶意流量信息

名称	处理方式	数量/个
Cridex	原始流量	8 563
Geodo	截取流量	7 295
Htbot	原始流量	5 326
Miuref	原始流量	7 750
Neris	合并流量	6 532
Nsis-ay	原始流量	7 420
Shifu	截取流量	5 236
Tinba	截取流量	5 960
Virut	原始流量	3 256
Zeus	原始流量	3 426

第二部分包含了使用IXIA-BPS[27]采集的10种正常流量，这是一种专业的网络流量仿真设备。关于模拟方法的信息，可以在相应的产品网站上找到。为了尽可能地反映更多的流量，10种流量包含8类的通用应用程序。正常流量信息如表2所示。

表2　正常流量信息

名称	种类	数量/个
BitTorrent	P2P	5 855
Facetime	Voice/Video	3 254
FTP	Data Transfer	5 125
Gmail	Email/WebMail	6 987
MySQL	Database	5 214
Outlook	Email/WebMail	3 620
Skype	Chat/IM	9 253
SMB	Data Transfer	5 731
Weibo	Social Network	5 632
WorldofWarcraft	Game	4 852

将数据集中的90%作为训练集，10%作为测试集，然后使用训练好的模型对测试集的流量进行分类。本文分类任务是20分类，即将数据集中的20个类别（Cridex、Geodo、Htbot、Miuref、Neris、Nsis-ay、Shifu、Tinba、Virut、Zeus、BitTorrent、Facetime、FTP、Gmail、MySQL、Outlook、Skype、SMB、Weibo、WorldofWarcraft）进行分类。

3.2　预处理流程

原始流量预处理流程如图3所示。流量切分是一种处理网络流量的技术，通过截取长度为N B的流量数据，可以获取主机建立连接的关键特征信息。在流量切分过程中，对于长度不足N B的数据包，用0x00填充至N B；对于长度超过N B的数据包，直接进行流量切分，使其长度统一为N B。接下来，将统一长度的数据包转化为二维矩阵形式，再将二维矩阵转化为灰度图。具体来说，每个数据包的二进制形式对应灰度图中的一个像素值。为了更好地处理数据包中的不同类别，采用独热编码技术。独热编码[28]是一种将不同类别映射为二元特征向量的技术，其中只有一个元素为1，表示该数据包属于该类别，其余元素都为0。通过这种方式，可以将类别变量转化为可计算的形式，消除类别之间的顺序关系和数值偏好。标签平滑可以减轻模型对训练数据中标签噪声的过度拟合，具体是通过将真实的离散标签分布与一个平滑的连续分布进行插值。这样做的目的是对分类错误的标签有一定的容错率，即对错误的标签并不完全舍弃，而是以较小的概率选择错误的标签[29]，使模型对训练数据中的噪声更加鲁棒，从而提高泛化能力。

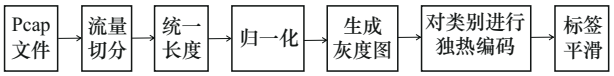


图3　原始流量预处理流程

3.3　流量分类可视化分析

对经过数据预处理后生成的图像进行分析，每个灰度图像的大小为1 024（即32×32）B。

USTC-TFC2016所有类别流量的可视化如图4所示。从图4中可以明显看出，除了少数图像在外观上非常相似，如FTP、SMB、Outlook，其他的流量类别之间很容易区分。

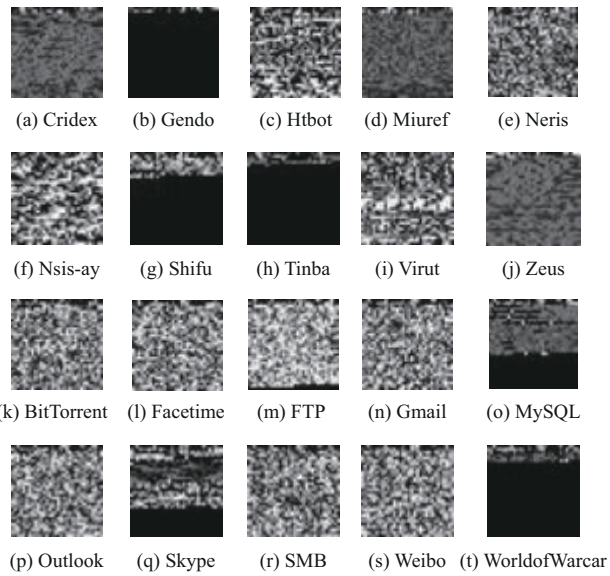


图4　USTC-TFC2016所有类别流量的可视化

同一类流量的特征一致性如图5所示，本节随机选择了来自同一流量类别的9个图像进行展示。从图5中可以看出，同一类流量图像具有相似性。例如，Cridex流量是由Cridex恶意软件在网络中进行数据传输所产生的流量，它会每隔一段时间就向特定的IP地址发送请求，以接受新的指令或上传窃取到的数据。另外，Neris、BitTorrent、Cridex、Zeus这4类流量图有较大的差别，这是因为不同类别的网络流量，其背后的生成机制和数据特征差异显著。

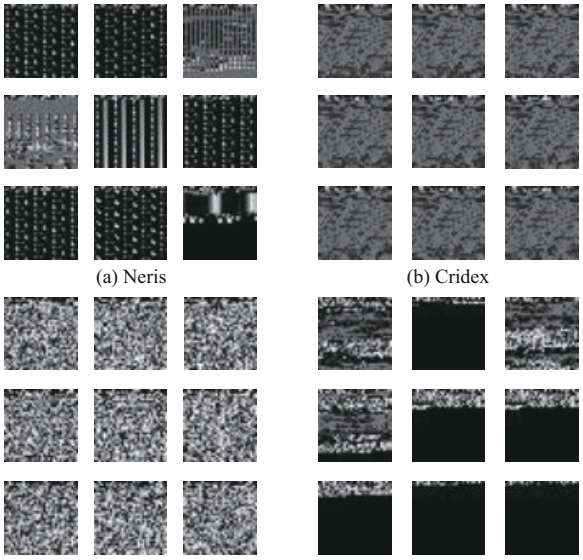


图5　同一类流量的特征一致性

4　实验结果与分析

4.1　实验环境与设置

本次实验使用的处理器是英特尔i7-9700，计算机的操作系统是Windows11，内存为8 GB，实验模型采用Python3.9编程语言、Tensorflow深度学习框架，其中batch_size选择32，epoch设置为30轮，优化器使用RMSprop，学习率为0.001，标签平滑中将标签软化程度epsilon设置为0.1。关于分裂注意力残差网络，将标签软化程序超参数n和k分别设为4；关于Bi-LSTM网络，将隐藏单元数设置为64；关于多头交叉注意力机制，将输出维度和注意力头数分别设置为128和8。

4.2　评价标准

本文采用准确率（Acc）、精确度（Pre）、召回率（Rec）、F1值（F1）为评价指标。准确率、精确度、召回率、F1值都是值越大模型效果越好，其计算式分别为

$$\mathrm{Acc}=\frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{TP}+\mathrm{TN}+\mathrm{FP}+\mathrm{FN}} \tag{13}$$

$$\mathrm{Pre}=\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}} \tag{14}$$

$$\mathrm{Rec}=\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}} \tag{15}$$

$$\mathrm{F1}=2\times\frac{\mathrm{Pre}\times\mathrm{Rec}}{\mathrm{Pre}+\mathrm{Rec}} \tag{16}$$

其中，TP（true positive）是模型将正例正确地分类为正例的数量，FP（false positive）是模型将负例错误地分类为正例的数量，TN（true negative）是模型将负例正确地分类为负例的数量，FN（false negative）是模型将正例错误地分类为负例的数量。

4.3　消融实验

为了论证本文所提出的分裂注意力残差卷积网络、Bi-LSTM网络以及特征融合模块的有效性，本节进行了多组消融实验，以比较不同结构的模型性能。这些消融实验包括以下8种模型。

1)本文模型。

2)仅提取流量图像的空间特征的模型。

3)仅提取流量图像的时序特征的模型。

4)在完整模型基础上，删除特征融合模块，仅采用拼接的方式融合空间特征和时序特征的模型。

5)在特征融合模块中，仅使用空间特征调整时序特征，删除时序特征调整空间特征部分的模型。

6)在特征融合模块中，仅使用时序特征调整空间特征，删除空间特征调整时序特征部分的模型。

7)在特征提取模块中，将分裂注意力残差网络替换为ResNet的模型。

8)在特征提取模块中，去掉分裂注意力残差网络中的残差连接的模型。

消融实验实现对不同结构的模型性能比较如表3所示。由表3看出，与本文模型相比，在仅使用空间特征或时序特征的情况下，模型的准确率分别下降了1.6%和2.8%，这验证了同时考虑空间特征和时序特征的模型比只考虑单一特征的模型表现更好。空间特征和时序特征通常具有互补性，空间特征通常捕捉流量数据包中的静态信息或局部模式，时序特征通常捕捉流量数据包中的动态变化和趋势，两者结合可以提供更全面的数据表示。在删除特征融合模块，仅采用拼接方式融合空间特征和时序特征的情况下，模型的准确率下降了1.5%，与仅提取空间特征的模型相比，只提升了0.01%，这说明了简单的拼接空间特征和时序特征并不能有效地利用2种特征，也验证了本文模型的特征融合模块的有效性。简单的拼接方式假设2种特征是独立的，忽略了它们之间的潜在关联和交互作用，从而无法捕捉空间特征和时序特征之间的复杂交互关系，也无法识别2种特征中存在的冗余部分以及干扰特征，导致无法充分利用2种特征的互补性。在特征融合模块中，仅使用空间特征调整时序特征或者仅使用时序特征调整空间特征，均造成模型性能

表3　消融实验实现对不同结构的模型性能比较

模型	准确率	精确度	召回率	F1 值
本文模型	0.983	0.983	0.983	0.983
仅提取流量图像的空间特征的模型	0.967	0.968	0.967	0.967
仅提取流量图像的时序特征的模型	0.955	0.958	0.955	0.955
删除特征融合模块，仅采用拼接方式融合空间特征和时序特征的模型	0.968	0.969	0.968	0.968
在特征融合模块中，仅使用空间特征调整时序特征的模型	0.981	0.982	0.981	0.981
在特征融合模块中，仅使用时序特征调整空间特征的模型	0.974	0.976	0.974	0.974
在特征提取模块中，将分裂注意力残差网络替换为ResNet的模型	0.971	0.972	0.971	0.971
在特征提取模块中，去掉分裂注意力残差网络中的残差连接的模型	0.978	0.979	0.978	0.978

的下降，说明了空间特征和时序特征相互调整的必要性。仅使用一种特征调整另一种特征，无法充分捕捉空间特征和时序特征之间的双向交互关系。相较于仅使用单一特征、简单拼接特征和单向调整特征，本文模型在分类性能上取得了显著提升。这一结果表明，本文模型有效解决了现有模型中特征提取片面以及无法充分考虑空间特征和时序特征之间关联关系的问题，在特征提取模块中，将分裂注意力残差网络替换为ResNet的模型准确率下降了1.2%，验证了本文提出的分裂注意力残差网络的有效性；在特征提取模块中，去掉分裂注意力残差网络中的残差连接的模型性能下降，验证了分裂注意力残差网络中的残差连接设计的合理性。综上所述，本文模型在准确率、精确度、召回率和F1值都表现得最好，说明本文模型结构和模块是合理的。

4.4　流量图像大小*N*的选择

本文选择了*N*的取值为196、784、1 024、1 600、2 304、4 096，即生成的图像形状分别为14×14、28×28、32×32、40×40、48×48、64×64。不同的*N*值会影响本文模型的分类效果，图6为不同*N*值下模型准确率变化趋势。当*N*=1 024时，本文模型的性能最好；当*N*值太小时，图像特征包含的信息量太少；当*N*值太大时，图像特征包含的噪声太多。

4.5　对比实验分析

本节将以下几种模型与本文模型在准确率、精确度、召回率和F1值上进行比较。

1)ResNeXt。通过引入新的块结构将网络分成多个路径或“基数”，可以并行训练。通过增加基数，ResNeXt可以捕获更丰富的特征集，而不会显著增加参数数量或计算成本。

2)CBAM-ResNet50[30]。这是一种基于卷积块注意力模块（CBAM）的改进型ResNet50模型。

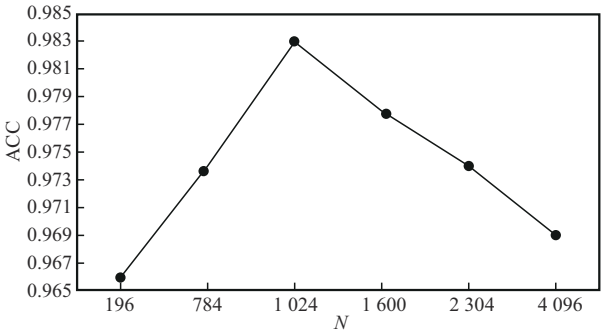


图6　不同*N*值下模型准确率变化趋势

通过在ResNet50中引入CBAM，提高了模型对输入数据的注意力，从而增强了模型的性能。

3)MSCANet[31]。这是一种基于深度分组卷积、多尺度卷积和自注意力机制的模型。该模型利用自注意力机制来调整和优化不同特征图的权值，从而提高模型的分类性能。

4)CNNBiGRU[32]。这是一种结合卷积神经网络和双向门控循环单元的模型。将卷积神经网络输出作为双向门控循环单元提取序列特征的输入，提取流的结构特征和序列特征。

5)CNN-BiLSTM[33]。这是一种多模型并行融合网络模型，融合了1D-CNN和Bi-LSTM，采用共同的全连接层进行特征融合。

6)DeepPacker[34]。这是一种采用2种深度神经网络结构，即堆叠自动编码器（SAE, stacked auto-encoder）和CNN。

7种模型的性能比较如表4所示。通过比较可以清晰地看出，本文模型具有较高的分类性能，且在准确率、精确度、召回率和F1值上均表现出色。CBAM-ResNet50和MSCANet模型都采用了残差网络和注意力机制，但只提取了流量图像的空间特征而没有考虑时序特征，单一的特征无法更全面地描

述流量数据的复杂行为。CNNBiGRU模型通过串联的方式提取了空间特征和时序特征，但缺乏对2种特征之间关联关系的充分考虑，可能导致部分有用信息在特征传递过程中被忽略或丢失，从而造成特征损失，影响模型的分类性能。CNN-BiLSTM通过拼接的方式融合空间特征和时序特征，简单的拼接虽然实现了特征的组合，但会忽略特征之间的关系和交互信息，组合的特征没有包含空间特征和时序特征之间的关联互补信息，导致模型无法充分利用2种特征的协同作用，从而限制了分类性能的提升。DeepPacker模型的输入是固定长度的字节向量而非流量图像，仅专注于提取字节流量的序列特征，会导致所提取的特征较为单一。本文模型同时考虑了空间特征、时序特征以及2种特征的关联关系，从而有更好的性能。

表4 7种模型的性能比较

模型设置	准确率	精确度	召回率	F1值
ResNeXt	0.976	0.971	0.970	0.970
CBAM-ResNet50	0.936	0.938	0.936	0.936
MSCANet	0.970	0.970	0.970	0.970
CNNBiGRU	0.874	0.88	0.874	0.873
CNN-BiLSTM	0.976	0.976	0.976	0.976
DeepPacker	0.964	0.965	0.963	0.964
本文模型	0.983	0.983	0.983	0.983

4.6 泛化能力分析

ISCXTor2016数据集是一个研究网络安全的数据集，主要用于分析和检测网络中的Tor流量。数据集中有nonTor和Tor两部分，分别从这两部分选取10个类别用于验证模型的泛化能力并与其他深度学习模型比较。5种模型在20分类上的性能比较如表5所示。本文模型与CNN-BiLSTM模型相比性能相近，但与其他模型相比有较大的优势。实验结果证明，本文模型在ISCXTor2016数据集上表现优秀，具有较好的泛化能力。

4.7 在线学习能力

为了评估本文模型在应对新型恶意流量时的在线学习能力，本节设计了如下实验。首先，自行采集了与USTC-TFC2016数据集中的恶意流量类型不同的8种新型流量数据（Trogan、Tor、DDoS、Brute force attack、Common Alerts、Malicious Scanning、Trojan horse、webshell、normal），并将这些新型恶意流量数据添加至原始数据集中，形成了增强数据集。接下来，利用增强数据集对本文模型和其他深度学习模型进行了对比实验。最后，通过比较各模型在面对新型恶意流量时的准确率、精确率、召回率和F1值来评估本文模型的在线学习能力。不同模型在增强数据集上的性能比较如表6所示。本文模型在面对新型流量时，并联分支提取特征和特征融合模块提高了模型学习新型流量特征的能力。融合后的特征包括了空间特征、时序特征以及2种特征的关联关系信息，更全面的特征表达使本文模型的在线学习能力强于其他模型。

表5 5种模型在20分类上的性能比较

模型设置	准确率	精确度	召回率	F1值
CBAM-ResNet50	0.927	0.925	0.919	0.921
MSCANet	0.965	0.960	0.961	0.960
CNNBiGRU	0.925	0.926	0.922	0.923
CNN-BiLSTM	0.988	0.987	0.982	0.984
本文模型	0.989	0.985	0.985	0.985

表6 不同模型在增强数据集上的性能比较

模型设置	准确率	精确度	召回率	F1值
CBAM-ResNet50	0.931	0.937	0.931	0.931
MSCANet	0.979	0.979	0.979	0.979
CNNBiGRU	0.903	0.904	0.903	0.903
CNN-BiLSTM	0.982	0.982	0.982	0.982
本文模型	0.988	0.989	0.988	0.988

5 结束语

本文提出的并联分支联合编码的网络恶意流量分类在网络安全、网络性能优化、网络管理等方面具有重要意义。该模型将网络流量转换为灰度图，这种处理方式适用于所有类型的流量，然后通过2个分支分别使用分裂注意力残差卷积网络和Bi-LSTM网络提取流量图像的空间特征和时序特征，通过特征融合模块挖掘2种特征的关联关系，以更好地融合两分支的特征。实验证明，并联分支联合编码模型在准确度、泛化能力和在线学习能力等方面均表现出优越性能。这一研究成果为网络流量分类领域发展提供了重要的参考和启示。并联分支联合编码模型也为提高模型的鲁棒性创造了新的思路，未来将考虑并联分支模型面对对抗性流量时如

何保持较高的性能。与此同时，已有不少将大语言模型应用于网络流量分析的成功案例。大语言模型能够深度剖析流量相关的语义信息，未来考虑将并联分支联合编码模型与大语言模型相结合，理解更多维度的流量信息。

参考文献：

[1] 张文哲, 杨栋, 魏松杰. 交互博弈引导的网络流量异常检测建模方法研究[J]. 信息安全学报, 2024, 9(2): 36-46
ZHANG W Z, YANG D, WEI S J. Interactive-gaming guided modeling and detection for network traffic anomaly detection[J]. Journal of Cyber Security, 2024, 9(2): 36-46.

[2] XIE R J, WANG Y X, CAO J H, et al. Rosetta: enabling robust TLS encrypted traffic classification in diverse network environments with TCP-aware traffic augmentation[C]//Proceedings of the ACM Turing Award Celebration Conference - China 2023. New York: ACM Press, 2023: 131-132.

[3] 马继烨, 朱国胜, 卫操, 等. 一种噪声容忍的网络流量分类方法[J]. 计算机科学, 2023, 50(S2): 775-781.
MA J Y, ZHU G S, WEI C, et al. Noise tolerant algorithm for network traffic classification method[J]. Computer Science, 2023, 50(S2): 775-781.

[4] LU J W, TANG C W, CHEN Z C, et al. Intraflow temporal correlation-based network traffic prediction[J]. Computer Networks, 2025, 256: 110913.

[5] PENG Y F, GUO Y Y, HAO R, et al. Network traffic prediction with attention-based spatial–temporal graph network[J]. Computer Networks, 2024, 243: 110296.

[6] 周成胜, 孟楠, 赵勋, 等. 基于深度学习的多会话协同攻击加密流量检测技术研究[J]. 信息安全研究, 2025, 11(1): 66-73.
ZHOU C S, MENG N, ZHAO X, et al. Encrypted traffic detection technology for multi-session coordinated attack based on deep learning[J]. Journal of Information Security Research, 2025, 11(1): 66-73.

[7] 王梦寒, 邓永晖, 魏波. 基于RF-BiLSTM的网络异常流量检测方法[J]. 通信技术, 2024, 57(12): 1297-1304.
WANG M H, DENG Y H, WEI B. Network abnormal traffic detection method based on RF-BilSTm[J]. Communications Technology, 2024, 57 (12): 1297-1304.

[8] JABLAOUI R, LIOUANE N. An effective deep CNN-LSTM based intrusion detection system for network security[C]//Proceedings of the 2024 International Conference on Control, Automation and Diagnosis (ICCAD). Piscataway: IEEE Press, 2024: 1-6.

[9] YANG H, ZHANG J F, SUN J, et al. MDAA: an unsupervised anomaly detection method for terminal traffic in new power system based on MDAA[J]. International Journal of Network Security, 2024, 26(3): 375-385.

[10] WANG W, ZHU M, ZENG X W, et al. Malware traffic classification using convolutional neural network for representation learning[C]//Proceedings of the 2017 International Conference on Information Networking (ICOIN). Piscataway: IEEE Press, 2017: 712-717.

[11] FERRAG M A, FRIHA O, HAMOUDA D, et al. Edge-IIoTset: a new comprehensive realistic cyber security dataset of IoT and IIoT applications for centralized and federated learning[J]. IEEE Access, 2022, 10: 40281-40306.

[12] MA H P, HUANG X Y, RUAN K, et al. MSTFCAN: multiscale sparse temporal-frequency cross attention network for traffic prediction[J]. Computer Networks, 2025, 258: 111035.

[13] ULLAH F, ULLAH S, SRIVASTAVA G, et al. IDS-INT: intrusion detection system using transformer-based transfer learning for imbalanced network traffic[J]. Digital Communications and Networks, 2024, 10(1): 190-204.

[14] 孙爽, 何立风, 朱纷, 等. 基于多分支特征融合的密集人群计数网络[J]. 计算机工程与设计, 2024, 45(3): 814-821.
SUN S, HE L F, ZHU F, et al. Multi-branch feature fusion network for crowd counting[J]. Computer Engineering and Design, 2024, 45(3): 814-821.

[15] ZAMPOKAS G, BOUGANIS C S, TZOVARAS D. Latency driven spatially sparse optimization for multi-branch CNNs for semantic segmentation[C]//Proceedings of the 2024 IEEE/CVF Winter Conference on Applications of Computer Vision Workshops (WACVW). Piscataway: IEEE Press, 2024: 949-957.

[16] DEVARAJ G P, RAVI R. Advancing skin cancer diagnosis with a multi-branch ShuffleNet architecture[J]. International Journal of Imaging Systems and Technology, 2024, 34(2): e23051.

[17] RASTOGI D, JOHRI P, TIWARI V, et al. Multi-class classification of brain tumour magnetic resonance images using multi-branch network with inception block and five-fold cross validation deep learning framework[J]. Biomedical Signal Processing and Control, 2024, 88: 105602.

[18] AHMED K, TORRESANI L, AHMED K, et al. Connectivity learning in multi-branch networks[J]. arXiv Preprint, arXiv: 1709.09582, 2017.

[19] FAN Y, XIE S F, XIA Y C, et al. Multi-branch attentive transformer[J]. arXiv Preprint, arXiv: 2006.10270, 2020.

[20] HAO P Y, GAO X, LI Z H, et al. Multi-branch fusion network for Myocardial infarction screening from 12-lead ECG images[J]. Computer Methods and Programs in Biomedicine, 2020, 184: 105286.

[21] CHEN S B, WEI Q S, WANG W Z, et al. Remote sensing scene classification via multi-branch local attention network[J]. IEEE Transactions on Image Processing, 2021, 31: 99-109.

[22] 宋玉琴, 赵继涛, 商纯良. 基于注意力机制的多分支特征级联图像去雨网络[J]. 光电子•激光, 2024, 35(4): 379-387.
SONG Y Q, ZHAO J T, SHANG C L. A multi-branch feature cascade image deraining network based on the attention mechanism[J]. Journal of Optoelectronics·Laser, 2024, 35(4): 379-387.

[23] 周子云, 黄洪. 改进EfficientNet图像分类的恶意流量检测模型[J]. 四川轻化工大学学报(自然科学版), 2023, 36(6): 49-56.
ZHOU Z Y, HUANG H. Improved malicious traffic detection model for EfficientNet image classification[J]. Journal of Sichuan University of Science & Engineering (Natural Science Edition), 2023, 36(6): 49-56.

[24] 赵忠斌, 蔡满春, 芦天亮. 融合多头注意力机制的网络恶意流量检测[J]. 数据与计算发展前沿, 2022, 4(5): 60-67.
ZHAO Z B, CAI M C, LU T L. Network malicious traffic detection incorporating multi-head attention mechanism[J]. Frontiers of Data & Computing, 2022, 4(5): 60-67.

[25] 魏德宾, 江亲龙, 温京龙, 等. GMTBLC: 基于深度学习的双模态网络流量分类[J]. 电信科学, 2024, 40(12): 93-106.
WEI D B, JIANG Q L, WEN J L, et al. GMTBLC: a deep learning-based bi-modal network traffic classification method[J]. Telecommuni-

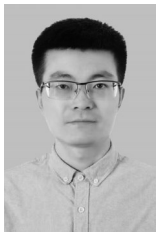
cations Science, 2024, 40(12): 93-106.
[26] CTU University. The stratosphere IPS project dataset[R]. 2016.
[27] Ixia Corporation. Ixia breakpoint overview and specifications[R]. 2016.
[28] BAGUI S, NANDI D, BAGUI S, et al. Machine learning and deep learning for phishing email classification using one-hot encoding[J]. Journal of Computer Science, 2021, 17(7): 610-623.
[29] 张永, 刘纪奎, 柯文龙. 基于并行可分离卷积和标签平滑正则化的脑电情感识别[J]. 电信科学, 2023, 39(5): 116-128.
ZHANG Y, LIU J K, KE W L. EEG emotion recognition based on parallel separable convolution and label smoothing regularization[J]. Telecommunications Science, 2023, 39(5): 116-128.
[30] 游小荣, 李淑芳, 邵红燕. 融合注意力机制与改进ResNet50的服装图像属性预测方法[J]. 现代纺织技术, 2025, 33(1): 58-64.
YOU X R, LI S F, SHAO H Y. A clothing image attribute prediction method integrating attention mechanism and improved ResNet50[J]. Advanced Textile Technology, 2025, 33(1): 58-64.
[31] 熊敬伟, 潘继飞, 毕大平, 等. 面向雷达行为识别的多尺度卷积注意力网络[J]. 西安电子科技大学学报, 2023, 50(6): 62-74.
XIONG J W, PAN J F, BI D P, et al. Multi-scale convolutional attention network for radar behavior recognition[J]. Journal of Xidian University, 2023, 50(6): 62-74.
[32] 杨永平, 王思婷. 基于CNN结合BiGRU的恶意流量分类算法研究[J]. 计算机科学, 2024, 51(S2): 867-875.
YANG Y P, WANG S T. Study on malicious traffic classification algorithm based on CNN combined with BiGRU[J]. Computer Science, 2024, 51(S2): 867-875.
[33] 李向军, 王俊洪, 王诗璐, 等. 基于多模型并行融合网络的恶意流量检测方法[J]. 计算机应用, 2023, 43(S2): 122-129.
LI X J, WANG J H, WANG S L, et al. Malicious traffic detection method based on multi-model parallel fusion network[J]. Journal of Computer Applications, 2023, 43(S2): 122-129.
[34] LOTFOLLAHI M, JAFARI SIAVOSHANI M, SHIRALI HOSSEIN ZADE R, et al. Deep packet: a novel approach for encrypted traffic classification using deep learning[J]. Soft Computing, 2020, 24(3): 1999-2012.

[作者简介]

马自强（1990– ），男，回族，新疆乌鲁木齐人，博士，宁夏大学副教授，主要研究方向为密码应用、系统安全、区块链、恶意流量识别、网络舆情分析。

崔梦真（2001– ），女，河南郑州人，宁夏大学硕士生，主要研究方向为流量分析。

杨天宇（2000– ），男，山东烟台人，宁夏大学硕士生，主要研究方向为恶意流量识别。



张宁宁（1992– ），男，宁夏隆德人，国家互联网应急中心宁夏分中心工程师，主要研究方向为网络加密流量、复杂网络攻击检测、流量溯源等。