研究与开发

GMTBLC：基于深度学习的双模态网络流量分类

魏德宾，江亲龙，温京龙，王欣睿
（大连大学信息工程学院，辽宁 大连 116622）

摘　要：网络流量分类对于网络安全维护和网络管理至关重要，在服务质量（quality of service，QoS）保证、入侵检测等任务中得到了广泛的应用。针对传统流量分类模型对特征提取不足，导致分类准确率较低等问题，提出了基于混合注意力（group mix attention, GMA）的Transformer和双向长短期记忆（bi-directional long short term memory，Bi-LSTM）网络的双模态网络流量分类（group mix transformer and Bi-LSTM for traffic classification，GMTBLC）方法。在数据预处理阶段，通过数据包的有效载荷生成会话内的包级别图像，以减少信息干扰。在分类阶段，图像首先由包混合Transformer（packet group mix transformer，PCMT）模块处理，该模块使用Transformer和GMA捕获全局特征。同时，会话图像由时空特征提取（spatio-temporal feature extraction，SFE）模块处理，其中数据包的空间特征由带有残差连接的卷积神经网络提取，数据包的时间特征由双向LSTM提取。在融合分类层中，通过动态加权机制融合上述全局特征和时空特征，最终完成网络流量分类。在公共数据集ISCX和USTC-TFC2016上进行的实验表明，该模型的分类准确率达99.31%，精确率、召回率和F1值均达到98%以上，相比其他模型分类效果更优。

关键词：流量分类；深度学习；注意力机制；Transformer；长短期记忆网络
中图分类号：TP393
文献标志码：A
doi: 10.11959/j.issn.1000−0801.2024251

GMTBLC: a deep learning-based bi-modal network traffic classification method

WEI Debin, JIANG Qinlong, WEN Jinglong, WANG Xinrui
School of Information Engineering, Dalian University, Dalian 116622, China

Abstract: Network traffic classification is crucial for network security maintenance and management, and it has been widely applied in tasks, such as quality of service (QoS) assurance and intrusion detection. To address the issues of traditional traffic classification models, such as insufficient feature extraction and low classification accuracy, a dual-modal network traffic classification method based on group mix attention (GMA) with a transformer and a bi-directional long short-term memory (Bi-LSTM) network, named group mix transformer and Bi-LSTM for traffic classification (GMTBLC), was proposed. In the data preprocessing phase, packet-level images within sessions were gen-

收稿日期：2024−10−18；**修回日期：**2024−11−02
通信作者：魏德宾，weidebin@163.com
基金项目：国家自然科学基金资助项目（No. 61931004）
Foundation Item: The National Natural Science Foundation of China (No. 61931004)

erated from the payloads of data packets to reduce information interference. In the classification phase, the images were firstly processed by the packet group mix transformer (PCMT) module, which utilized the transformer and GMA to capture global features. Simultaneously, session images were processed by the spatio-temporal feature extraction (SFE) module, of which the spatial features of packets were extracted by a convolutional neural network with residual connections, and temporal features of packets were extracted by a bi-directional long short-term memory network. In the fusion classification layer, the above global and spatiotemporal features were integrated using a dynamic weighting mechanism to complete network traffic classification. Experimental results on ISCX and USTC-TFC2016 datasets demonstrate that the proposed model achieves a classification accuracy of 99.31%, with precision, recall, and F1-score all above 98%, and outperforms the other models in classification effectiveness.

Key words: traffic classification, deep learning, attention mechanism, transformer, LSTM

0 引言

随着互联网的快速发展，网络业务种类繁多，网络流量呈爆发式增长，这不仅给网络服务质量（quality of service，QoS）保障带来了很大困难，也使得网络安全管理变得越来越困难[1-3]。为了解决这些问题，网络管理员可以通过对网络流量进行分类，提高安全性、稳定性和管理效率，从而及时发现并应对潜在的安全威胁。另一方面，通过对应用程序进行分类，监控和管理网络流量可以为高优先级的数据提供更快的通道。因此，网络流量分类技术对于提升网络的QoS具有重要意义，并在网络安全和管理中占据重要地位。

早期的流量分类方法主要基于端口和深度包检测（deep packet inspection，DPI）[4]技术。通过已知的注册端口号，应用程序的流量可以根据匹配的标准端口号进行分类。Moore和Papagiannaki[5]采用基于传统端口匹配的技术对网络流量进行分类。然而随着网络技术的发展，越来越多的应用程序使用动态分配端口，或者使用常用的通信协议端口进行伪装，使得基于端口号分类的方法可用性降低。基于DPI的方法将数据包载荷与一组存储的签名进行匹配，进而对网络流量进行分类。Sen等[6]使用应用级签名对P2P应用流量进行分类；Roughan等[7]使用统计应用签名进行分类；Khalife等[8]提出了OpenDPI工具，使用行为和统计分析对流量进行分类；Deri等[9]提出nDPI对会话证书进行分析，从而对流量进行分类；Zhao等[10]使用有效负载的特定签名字符串进行流量分类和识别。然而，基于DPI的技术，只适用于未加密的流量，并且需要大量的计算。

机器学习技术的发展，可以将机器学习应用于加密流量识别[11]，传统的基于机器学习的方法通过统计特征来区分不同的流量，例如每个流的持续时间或平均数据包大小。Auld等[12]提出了一种使用数据包长度、到达时间间隔和数据包的传输方向等一系列数据包特征进行P2P流量分类的贝叶斯神经网络。庞兴龙等[13]对半监督学习在流量分类上的优势进行总结，并阐述了半监督分类、半监督聚类和半监督降维在流量分类中的实际应用。Sheikh等[14]综合统计分析机器学习技术，结合多种统计特征来提高分类的准确性。Dong等[15]提出CMSVM算法，用于解决网络流量识别中类别不平衡的问题。大多数机器学习方法依赖于手动设计特征，这在网络流量迅速增长的情况下尤为耗时且容易出错，不能进行自动化特征提取，难以对复杂多样的网络流量进行准确分类。综上，与深度学习（deep learning，DL）相结合的方法已成为主流。

DL方法与机器学习方法不同，不需要分析数据结构，定义数据特征，能够自主进行数据特

征学习，学习样本数据的内在规律和表示层次[16]。目前，卷积神经网络（convolutional neural network，CNN）和递归神经网络（recurrent neural network，RNN）是主要的端到端流量分类方法。通过神经网络提取网络流量的特征，再经过Softmax输出所属的类别概率，可以对流量进行分类。Wang等[17]提出将数据包转换为图片，然后使用1D-CNN进行处理。Dong等[18]提出了MPNN方法，基于多个神经网络结构，使用单独的神经网络模块处理每个应用，从而提高准确率。Xie等[19]提出SAM的方法，将网络数据包的字节视为一种语言，用于流量分类。Shapira等[20]利用流中与时间相关和大小相关的特征，将流量转换为直观的“FlowPic”图片，利用CNN进行流量分类。Yang等[21]提出了实时端到端的网络流量分类框架ConViTML，融合了CNN和Vison Transformer，可以直接提取包含会话的基本特征和结构特征，然后通过数据包关系网络进行流量分类。Dong等[22]提出了基于半监督的双深度Q网络SSDDQN，利用自动编码器重构流量特征，然后使用深度神经网络进行分类。为了解决类别不平衡的问题，Dong等[23]提出了GADCN模型，拟合和扩展了流量图像，保持了类之间的平衡。然而，上述研究存在一定的局限性，即仅从一个模态对网络流量进行了分类，没有选择从不同模态进行流量特征提取。

最近的一些研究工作已经开始使用多模态深度学习方法对网络流量进行分类。多模态可以从不同角度提取流量特征，从而提高分类器性能，改善单一模态特征提取不充分的问题。Aceto等[24]提出了MIMETIC多模态深度学习框架，将数据包有效载荷的前576 byte和4个协议特征作为输入，并行使用1D-CNN和门控循环单元（gate recurrent unit，GRU）对有效载荷和协议特征进行特征提取。Wang等[25]提出的多模态加密流量分类框架AppNet，截取数据包的前1 014 byte，使用1D-CNN进行特征提取，同时并行使用LSTM学习数据包的长度序列特征，最后将从两个视角学到的特征连接起来进行分类。Lin等[26]提出PEAN多模态深度学习加密流量分类框架，该框架以数据包字节和长度序列作为输入，利用自注意力机制学习双向流中网络数据包之间的深层关系。Lin等[27]结合CNN和LSTM，提出了TSCRNN，Zhu等[28]提出的CMTSNN也使用了CNN和LSTM的级联结构。然而，上述研究亦存在一定的局限性。首先，CNN的局部卷积操作使得对于长距离的感知能力较弱，在处理全局关系上表现不如Transformer。其次，Transformer的自注意力机制虽然能够捕捉输入序列中各个位置之间的依赖关系，但是在注意力计算过程中，通过要查询的信息（Q）和被查询的向量（K）进行点积运算，再与查询得到的值（V）生成的注意力图只能捕获单一粒度上token到token之间的相关性。最后，TSCRNN和CMTSNN只从网络流量中提取时空特性，而没能完全把握流量中的数据包之间的关系。

综上，为了提高模型对网络流量分类的准确率，本文针对现有问题，从多模态角度出发，基于已有研究提出了基于混合注意力的Transformer和双向长短期记忆网络的双模态，网络流量分类（group mix transformer and bidirectional long short term memory for traffic classification，GMTBLC）方法。该方法使用混合Transformer提取会话的全局特征，用带有残差连接的卷积神经网络提取数据包的空间信息，同时使用LSTM提取数据包之间的时间特征，通过动态加权机制融合全局特征和时空特征，最终基于融合特征对网络流量进行分类，以提高流量分类的准确率。实验结果表明，本文所提方法的分类准确率相对于现有的一些方法有一定的提高，模型性能更好。

1 方法设计

为了从多角度充分提取网络流量的特征，本文提出了基于深度学习的多模态网络流量分类模型GMTBLC，GMTBLC如图1所示。模型整体共分为4部分：数据预处理、特征提取、特征融合和分类。数据预处理模块实现将捕获到的原始流量转换为特定的格式，以便输入模型中。特征提取模块由两条并行的分支组成，分别为左路的全局特征提取分支和右路的时空特征提取分支。特征融合模块通过动态加权机制融合两个分支提取的特征，生成用于分类的融合特征。最后通过全连接和Softmax输出分类结果。

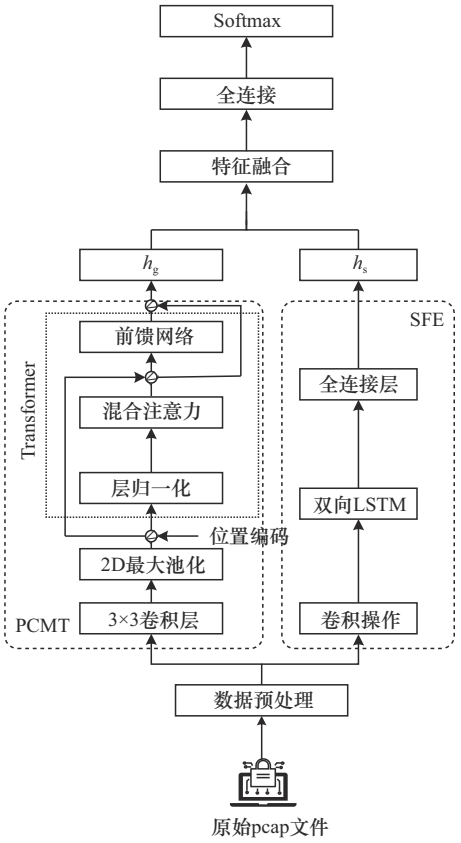


图1 GMTBLC的结构

1.1 数据预处理

数据预处理的主要目的是优化原始数据集，以提高实用性。公开的网络流量数据集可分为原始数据集和经过处理的流量特征数据集。在实际使用中，网络流量数据大多保存为pcap或pcapng格式的文件，为每一种流量生成一个pcap文件，pcap文件数据格式如图2所示，其中包含24 byte的全局报头，之后是多个数据包记录，每个记录有16 byte的包头。

因此，在数据预处理中应移除或忽略这些部分，简化模型处理。为了避免影响时延敏感数据的时效性，选择从数据包层面进行分析，将每个数据包转换为灰度图像，而不是对流会话进行转换。转换过程中，提取数据包有效载荷的前784 byte，超出部分截断，不足部分用0填充。由于TCP和UDP报头的长度不同（TCP为20 byte，UDP为8 byte），本文在UDP报头的末尾填充0。将数据包转换为28×28的图片格式，数据预处理过程如图3所示。数据预处理过程包括以下4个步骤。

（1）流量分割：根据会话标准将pcap文件分割为更小的文件。

（2）流量清除：删除重复数据包和不携带有效载荷的数据包。清除包头中的IP地址和MAC地址。

（3）统一数据长度：将UDP报头填充为20 byte，只保留前784 byte。

（4）将数据包转换为图像：将每个会话分割为数据包，进一步将每个数据包转换为灰度图像。

1.2 全局特征提取分支

Vaswani等[29]在2017年提出了Transformer神经网络模型，该模型目前广泛应用于自然语言处理（natural language processing，NLP）和其他领域，并取得了优异的结果。

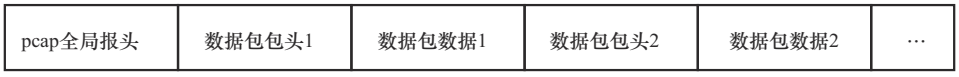


图2 pcap文件数据格式

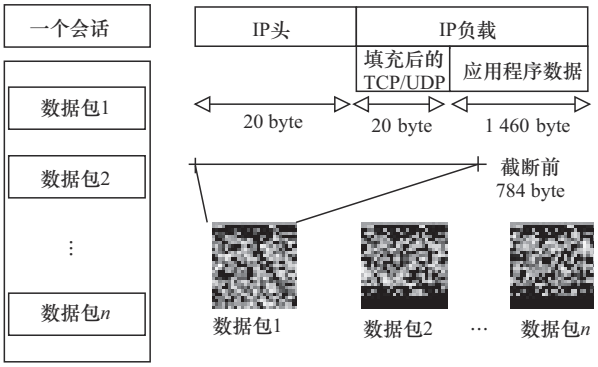


图3　数据预处理过程

Transformer 成功的关键得益于多头注意力（multi-head attention，MHA）模块，其主要作用是在输入序列中查找相关信息并加权汇总生成输出序列[12]。MHA 使得网络设计具备长距离依赖建模、全局感受野、灵活性和鲁棒性等优势[30]。通常的注意力计算，将 V 与 Q 和 K 之间的相关性重新线性组合，这些相关性通常在单个 token 之间计算，没有考虑不同维度上不同 token 组之间的相关性，自注意力示例如图 4 所示，图 4 展示了 7 个四维 token 之间的相关性计算，在这里相关性计算只关注单个 token。

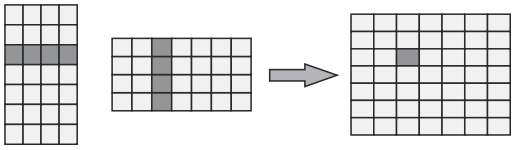


图4　自注意力示例

可以观察到，生成的注意力图仅捕捉了单一粒度上 token 对之间的相关性，并将注意力图与 V 相乘，从而忽略了不同 token 组之间的关联。为了解决这一局限性，本文引入了混合注意力（group mix attention，GMA）机制。GMA 将 token 分割为片段，并通过组聚合器生成组代理替代个别 token，混合注意力示例如图 5 所示。为了计算两个高亮组之间的相关性，将它们聚合成两个组代理，以便进一步相乘。GMA 旨在高效地计算 token 到 token、token 到 token 组、token 组到 token 组的关联，提供更全面的建模方法。

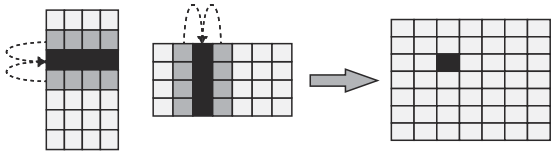


图5　混合注意力示例

全局特征提取过程如算法 1 所示。双模态网络流量分类（packet group mix transformer，PCMT）由多个 Transformer 编码器子层组成，每个编码器都由一个 GMA 机制和一个带有 ReLU 激活函数的全连接层组成。每个子层都使用了层归一化和残差连接。

算法 1　全局特征提取过程

输入　归一化网络会话图像 S，全局特征维度 d_{model}，Transformer 编码器层数 n_{layers}

输出　全局特征 h_{g}

CV = Conv2d(1,16,3,1)；/*卷积操作*/

MP = MaxPool2d(CV,kernel_size = 2)；/*　最大池化*/

PE = PositionEmbedding(MP)；/*位置嵌入*/

for $i \in n_{\text{layers}}$

LN_i = LayerNorm(PE)；/*归一化*/

W_i = GroupMixAttention(LN_i,d_{model})；/*混合注意力计算*/

RC = Addition(W_i, PE)；/*残差连接*/

RC = FeedForward(RC, d_{model})；/*前馈层*/

end for

h = Linear(RC)；

return h

原始字节的预处理结果 $x_t \in \text{R}^{m \times d_1 \times d_1}$，输入神经网络中，每个数据包 P_i 被视为一个补丁，使用由顺序模块组成的卷积层提取浅层低维特征。首先使用单通道输入和单通道输出的二维卷积层，卷积核大小设为 3，跨度为 1，无填充。然后，执行池化窗口为 2 的 2D 最大池化操作，以步幅为 2 滑动输入。从 CNN 提取的每个会话的数据

包基本向量特征$w_t \in \mathrm{R}^{m\times d_2\times d_2}$可表示为：

$$w_t = f(x_t, \theta) \tag{1}$$

其中，θ表示CNN的训练参数，x_t是输入样本。由于Transformer将序列数据作为输入，因此将降维后的每个数据包扁平化大小为d_2^2的序列格式。随后，利用Transformer对数据包的序列信息进行建模，扁平化操作为：

$$z_t = W \times w_t + \mathrm{PE} \tag{2}$$

其中，W为线性扁平化操作，是一个线性变换矩阵。$\times$表示矩阵相乘，$\mathrm{PE} \in \mathrm{R}$是位置嵌入，$z_t \in \mathrm{R}^{m\times d_2^2}$是最终会话嵌入。

目前，会话嵌入z_t包含数据包的基本特征。为了进一步提取会话的结构信息，采用了视觉Transformer（vision transformer，ViT）网络的经典Transformer编码模块，它由GMA和前馈网络（feed forward network，FFN）层组成，GMA模块结构设计如图6所示。

首先，将$Q/K/V$条目均匀地分为5个部分，并

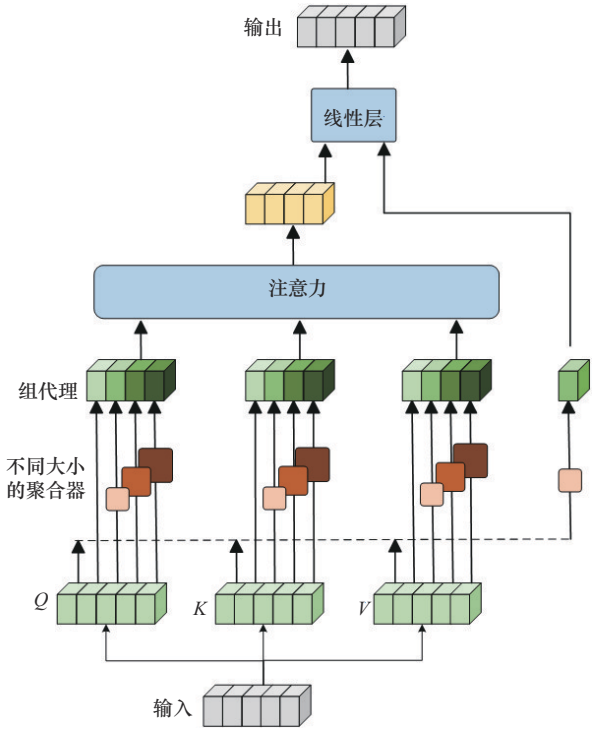


图6　GMA模块结构设计

对其中4个部分进行分组聚合。分别用X_i^Q、X_i^K、X_i^V（$i \in [1,2,3,4]$）表示从$Q/K/V$条目中分割出的部分。为了生成组代理Q'、K'和V'，本文先对这些部分执行聚合操作$\mathrm{Agg}^i(X_i^Q)$、$\mathrm{Agg}^i(X_i^K)$和$\mathrm{Agg}^i(X_i^V)$。然后，将所有4个$i \in \{1,2,3,4\}$聚合特征连接起来，输出组代理Q'、K'和V'。最后，在组代理上执行注意力计算，最终输出Attention。

$$\mathrm{Attention} = \frac{Q'}{\sqrt{d}}\mathrm{Softmax}(K'^{\mathrm{T}}V') \tag{3}$$

FFN层由两个线性变换组成，通过激活函数进行非线性转换。首先，输入特征经过第一个线性层，映射到高维空间。随后，映射后的特征通过激活函数（通常是GeLU）进行非线性变化。最后，变换后的特征通过第二个线性层映射回原始维度。FFN层结构如图7所示。

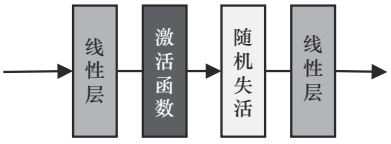


图7　FFN层结构

前馈（feed forward，FF）模块旨在利用多层感知机（multi-layer perceptron，MLP）结构，对每个数据包的特征进行灵活且自适应的非线性变换，从而捕捉更丰富的特征表示，增强模型的表达能力。

最终，用$s_t \in \mathrm{R}^{m\times d_2^2}$表示每个会话的最终表示，其中，$m$是会话中数据包的数量，$d_2$是嵌入维度，通过线性展开将特征矩阵映射为一维特征向量h_g，视为会话的全局特征。

$$h_g = \mathrm{Linear}(s_t) \tag{4}$$

1.3　时空特征提取分支

时空特征提取（spatio-temporal feature extraction，SFE）模块由两个核心组件组成，带有残差连接的卷积层（ResConv）和双向LSTM层（Bi-LSTM）。ResConv提取每个数据包的高级空间特征，并在对整个会话图像进行卷积时解决时间信

息混乱的问题。然后，这些带有时间特征的空间特征被输入双向LSTM中，以进一步捕捉会话的时空特征，时空特征提取过程如算法2所示。

算法2　时空特征提取过程

输入　归一化网络会话图像S_n，全局特征维度d_t

输出　全局特征h_{st}

$P_n=S_n$；/*将会话图像分割为数据包图像*/

$L=\text{Length}(P_n)$；/*获取数据包的长度*/

$\text{PE}=\text{PositionEmbedding}(\text{MP})$；/*位置嵌入*/

$t=1$；

for　$P \in P_n$

$F_m=\text{BN}(\text{Conv2d}(P))$；/*卷积操作和归一化*/

$W_p=\text{FC}(\text{GAP}(F_m)),\text{FC}(\text{GMP}(F_m))$；/*全局最大池化和全局平均池化*/

$F_{\text{spatial}}=\text{FC}(W_p)$；

$F'_m=\text{Addition}(F_{\text{spatial}},F_m)$；/*残差连接*/

$F_s(t)=\text{Flatten}(F'_m)$；/*扁平化*/

$t=t+1$；

end for

$h_{\text{f}},h_{\text{b}}=\text{Bi-LSTM}(F_s(1),\cdots,F_s(t))$；/*双向LSTM*/

$h_{st}=\text{Concat}(h_{\text{f},n},h_{\text{b},n})$；

$h_{st}=\text{Linear}(h_{st})$；/*线性变换*/

return h_{st}

1.3.1　残差卷积层

卷积层负责对数据包图像的特征提取，对每个数据包进行卷积操作。在数据处理阶段，每个数据包被处理为28×28的灰度图像，生成一组特征图F_m。由于每个通道的特征图在整个图像中的贡献不同，通过在F_m上应用全局最大池化和全局平均池化，每个特征图被压缩成一个标量，仍保留F_m的重要信息，F_m会被压缩成两个向量，这些向量被输入带有激活函数的全连接层，生成每个通道的权重向量，代表了每个通道的相对重要性。接下来通过将生成的权重与原始特征图F_m进行加权，放大重要特征，同时抑制不重要的特征。最后将原始特征图F_m和加权后的特征图相加，形成一个残差连接。这个过程可被形式化为：

$$F_m=\text{BN}(\text{Conv}(X)) \tag{5}$$

$$F'_m=\text{FC}(\text{FC}(\text{GAP}(F_m))+\text{FC}(\text{GMP}(F_m)))+F_m \tag{6}$$

其中，BN(·)代表批量标准化。最后将输出的F'_m扁平化得到F_s。

$$F_s=\text{Flatten}(F'_m) \tag{7}$$

1.3.2　双向LSTM

通过ResConv获取每个数据包的空间特征后，这些特征按顺序输入双向LSTM中。双向LSTM的性质使其能同时捕捉数据包序列的前向和后向信息，这可以增强GMTBLC捕捉数据包之间相关性的建模能力。每个数据包的空间特征$F_s(t)$都会以时间顺序和反向时间顺序输入LSTM单元中，以捕捉空间特征$F_s(t)$的时间相关性。

$$h_{\text{f},t}=\text{LSTM}(F_s(t),h_{\text{f},t-1}) \tag{8}$$

$$h_{\text{b},t}=\text{LSTM}(F_s(t),h_{\text{b},t+1}) \tag{9}$$

其中，$h_{\text{f},t}$表示在时间步长t的LSTM前向隐藏状态，$h_{\text{b},t}$表示在时间步长t的LSTM后向隐藏状态。将来自前向和后向隐藏状态序列的最后隐藏状态连接起来，获得会话的时空特征，序列到向量双向LSTM网络如图8所示。最后，将连接的隐藏状态通过具有激活函数的全连接层进行非线性变化。

$$h_{st}=\text{FC}([h_{\text{f},n},h_{\text{b},n}]) \tag{10}$$

其中，n代表时间步长的最大值，即数据包序列的长度。

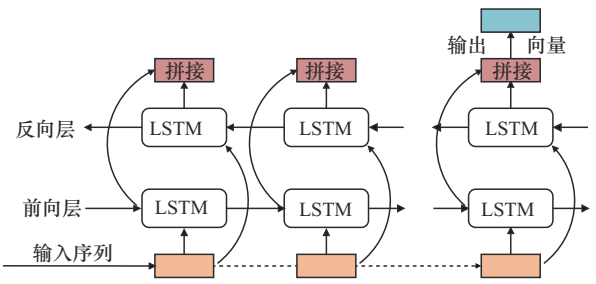


图8　序列到向量双向LSTM网络

1.4 特征融合

特征融合[31]通过将来自不同来源的多个特征组合，以构建一个更具代表性的特征集合。

从多个不同的角度学习到的特征具有互补的信息，对本文而言，从数据包中学习到的全局特征和时空特征，将其进行特征融合，模型能更好地理解数据，提升预测精度。并且当某一特征表现较弱时，其他特征可以提供补充信息，从而提高模型的稳定性和鲁棒性。

在传统方法中，如简单的拼接或固定权重融合，无法充分利用这两种特征之间的相关性，也不能适应性地调整权重。在GMTBLC方法中，提出了一种在特种融合层内部的动态权重机制，用于计算全局特征和时空特征的权重系数。

动态权重机制为不同会话的全局特征和时空特征分配不同的权重。此外，引入了温度参数τ，用于控制不同特征提取模块之间的权重平衡，防止特征提取模块在训练过程中失效，可以表示为：

$$z=\tanh(W_{g,s}[h_g;h_{st}]+b_{g,s}) \tag{11}$$

$$\xi=W_z z+b_z \tag{12}$$

$$\alpha_g=\frac{\exp(\xi_1/\tau)}{\exp(\xi_1/\tau)+\exp(\xi_2/\tau)} \tag{13}$$

$$\alpha_s=\frac{\exp(\xi_2/\tau)}{\exp(\xi_1/\tau)+\exp(\xi_2/\tau)} \tag{14}$$

$$h=\alpha_g h_g+\alpha_s h_{st} \tag{15}$$

其中，$W_{g,s}$、$b_{g,s}$、W_z和b_z是可学习的参数，$\tanh(\cdot)$表示双曲正切函数，$\exp(\cdot)$表示指数函数，$\xi\in R^2$是特征的未归一化权重，ξ_1、ξ_2是全局特征和时空特征的权重。温度参数τ是一个超参数，它平衡了对不同特征的关注并在归一化权重时缓解了极端情况。

1.5 分类器和模型训练

将经过融合的特征输入全连接层。融合特征经过全连接层可以映射为流量属于各类的概率，通过比较概率的大小，得出对应的类别索引，进行流量分类。

损失函数用于量化模型的预测结果与实际结果之间的差异，通过最小化损失函数，可以找到模型参数的最优值，从而提高模型的预测准确度。在训练过程中，通过不断调整模型参数，以减小损失函数的值，模型逐渐学习到数据的内在规律和特征，从而能够更准确地进行预测。为了使模型可以学习到分别训练的两部分的最佳性能，需要对损失函数进行改进。新的损失函数不仅衡量了最终的分类结果，还考虑了每部分学习的特征的分类性能。因此，在获得h_g和h_{st}特征向量后，分别对它们进行全连接和Softmax运算，并得到γ_1和γ_2。然后，将h_g和h_{st}连接起来，再次通过全连接和Softmax运算进行分类，得到最终分类结果γ_3。相应地，在每部分使用交叉熵作为它们的损失函数，得到loss_1、loss_2和loss_3。整个模型的损失loss_{total}是上述3个损失之和，具体为：

$$\gamma_1=\text{Softmax}(W_{h_1}\cdot h_g+b_{h_1}) \tag{16}$$

$$\gamma_2=\text{Softmax}(W_{h_2}\cdot h_{st}+b_{h_2}) \tag{17}$$

$$\gamma_3=\text{Softmax}(W_{h_3}\cdot h+b_{h_3}) \tag{18}$$

$$\text{loss}_\theta=\frac{1}{T}\sum_{i=1}^{T}\ln(\gamma_\theta)_i,\quad \theta=\{1,2,3\} \tag{19}$$

$$\text{loss}_{total}=\sum_{\theta=1}^{3}\text{loss}_\theta \tag{20}$$

其中，$\{W,b\}$是对应的全连接层的参数，T是样本训练数，Y_i为第i个网络流的类。

2 实验与分析

本文实验的显卡为NVIDIA Geforce RTX 3090，内存为24 GB，操作系统为Ubuntu20.06，Python环境版本为3.8，Pytorch采用2.0.0版本，对应的CUDA版本为11.6。

2.1 数据集

在评估所提的端到端网络流量分类框架的有

效性时，首先，在数据集选择方面，必须使用原始网络流量数据集，而不是经过处理或提取的文件。其次，为了获得精确的评估结果值，必须为每个原始流量样本贴上适当的标签。

本文选择数据集USTC-TFC2016和ISCX VPN-nonVPN对GMTBLC方法进行实验验证。USTC-TFC2016数据集由Wang等[32]创建，其中包含了10类正常加密流量（Benign）和10类恶意流量（Malware），USTC-TFC2016数据集统计见表1，其中，N为正常流量，M为恶意流量。

表1　USTC-TFC2016数据集统计

类名	类型	样本数量
BitTorrent	N	7 502
Facetime	N	6 000
FTP	N	6 319
Gmail	N	5 111
MySQL	N	7 026
Outlook	N	7 475
Skype	N	6 089
SMB	N	5 473
Weibo	N	4 569
Cridex	M	8 197
Geodo	M	6 690
Htbot	M	5 952
Miuref	M	4 952
Neris	M	8 425
Nsis-ay	M	6 033
Shifu	M	9 576
Tinba	M	8 504
Virut	M	6 138

ISCX VPN-nonVPN数据集是加拿大Dalhousie大学提供的一个专门用于虚拟专用网络（virtual private network，VPN）流量和非VPN流量分析的数据集。它包含来自VPN和非VPN连接的网络流量。其中包含17种应用程序的流量，如Facebook、YouTube等，ISCX VPN-nonVPN数据集统计见表2。

表2　ISCX VPN-nonVPN数据集统计

类别	数据包数量
AIM chat	1 785
Email	17 578
Facebook	11 233
FTPs	3 784 620
Gmail	11 014
Hangouts	984 241
ICQ	1 106
Netflix	299 057
SCP	447 792
SFTP	416 813
Skype	23 710
Spotify	40 592
Tor	326 251
Torrent	108 227
Vimeo	145 947
Voipbuster	2 480
YouTube	209 785

2.2　模型评价

为了评估所提方法的性能，本文采用3个评估指标：准确率（accuracy）、F1分数和可视化混淆矩阵。准确率是指预测正确的结果占总样本的百分比，F1分数是衡量模型性能的综合指标，通过精确率和召回率计算得出，混淆矩阵用于对比分类结果与实际值，可以把分类结果的精度显示在一个矩阵里面。本文根据以下公式计算4个指标的精确值，用于评估所提方法的整体性能。

$$\text{Accuracy}=\frac{\text{TP}+\text{TN}}{\text{TP}+\text{FP}+\text{TN}+\text{FN}} \tag{21}$$

$$\text{Precision}=\frac{\text{TP}}{\text{TP}+\text{FP}} \tag{22}$$

$$\text{Recall}=\frac{\text{TP}}{\text{TP}+\text{FN}} \tag{23}$$

$$\text{F1}=\frac{2\cdot\text{Precision}\cdot\text{Recall}}{\text{Precision}+\text{Recall}} \tag{24}$$

其中，TP是正确分类为正例的样本数，TN是正确分类为负例的样本数，FP是错误分类为正例的样本数，FN是错误分类为负例的样本数。

2.3 超参数设置

本文进行了特征融合层中温度参数τ的选择研究，参数τ依赖于一个动态权重机制，该机制调整原始字节特征和长度序列特征之间的权重分布，以实现不同特征的自适应融合。从两个数据集中每个类别选取200个样本进行训练。在训练过程中，τ的值由1增加至600，步长为10。经过30轮训练，不同τ值模型准确率变化曲线如图9所示。τ值越大，准确率明显提高。相反，τ值较小，准确率下降，这说明全局特征和时空特征的权重分配不合理，特别是，当τ值大于300时，分类准确率趋于稳定。因此，选择合适的τ值，可以有效地提高分类性能，从图9可以看出，当τ值为400时，所得的准确率最大，所以，本文取τ值为400。

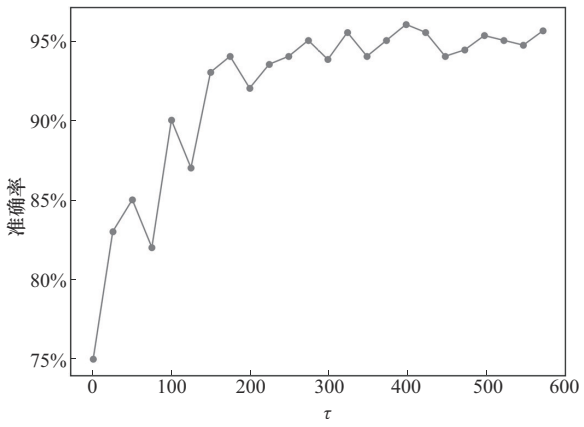


图9 不同τ值模型准确率变化曲线

2.4 与主流分类模型对比

为了验证GMTBLC方法的有效性，本文在两个数据集上进行了模型分析实验。将所提模型与经典算法和目前的先进模型进行对比分析，所选的模型为1D-CNN[17]、TSCRNN[27]、CMTSNN[28]、Flow-GNN[33]，1D-CNN采用一维卷积网络，TSCRNN和CMTSNN结合了递归和卷积神经网络架构，Flow-GNN采用集合深度学习。

指标柱状图如图10所示，展示了5种网络模型在数据集USTC-TFC2016上进行流量分类的平均准确率和F1分数的柱状图。对图10进行分析可知，本文方法的分类结果明显优于其他4种方法，各指标均取得了最好的结果。图10中的结果显示，GMTBLC在数据集USTC-TFC2016上，流量分类的准确率高达99.31%，较1D-CNN、TSCRNN、CMTSNN和Flow-GNN分别提高了5.02%、0.63%、0.76%和0.31%。相对于其他模型的准确性有所提高。同时，在流量分类任务中，F1分数也最高，达到了99.43%。

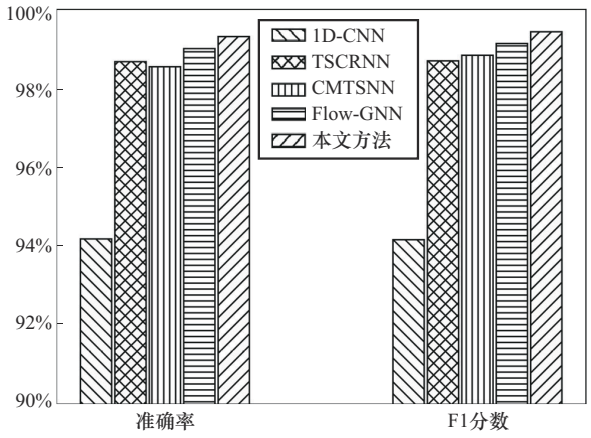


图10 指标柱状图

进一步观察每种模型对数据集中每种类别的分类效果，5种模型的混淆矩阵分布如图11所示。

在混淆矩阵中，“行”表示真实类别，“列”表示预测类别。从图11可以看出，本文方法在USTC-TFC2016数据集上对7种应用程序的分类准确率达到了100%，而其他模型分类准确率达到100%的类别数均少于本文方法。这一优势归因于本文方法在特征提取方面的独特设计，使其能够更全面地捕获加密流量中的特征信息。相比之下，1D-CNN的准确率最低，仅有4种应用程序的分类准确率为100%。特别是对Gmail的分类准确率仅为77%，远低于本文方法，这是因为1D-CNN对复杂的加密流量模式不敏感，在捕捉流量特征的广度和深度上有所不足。

Flow-GNN的分类准确率虽然最接近本文方

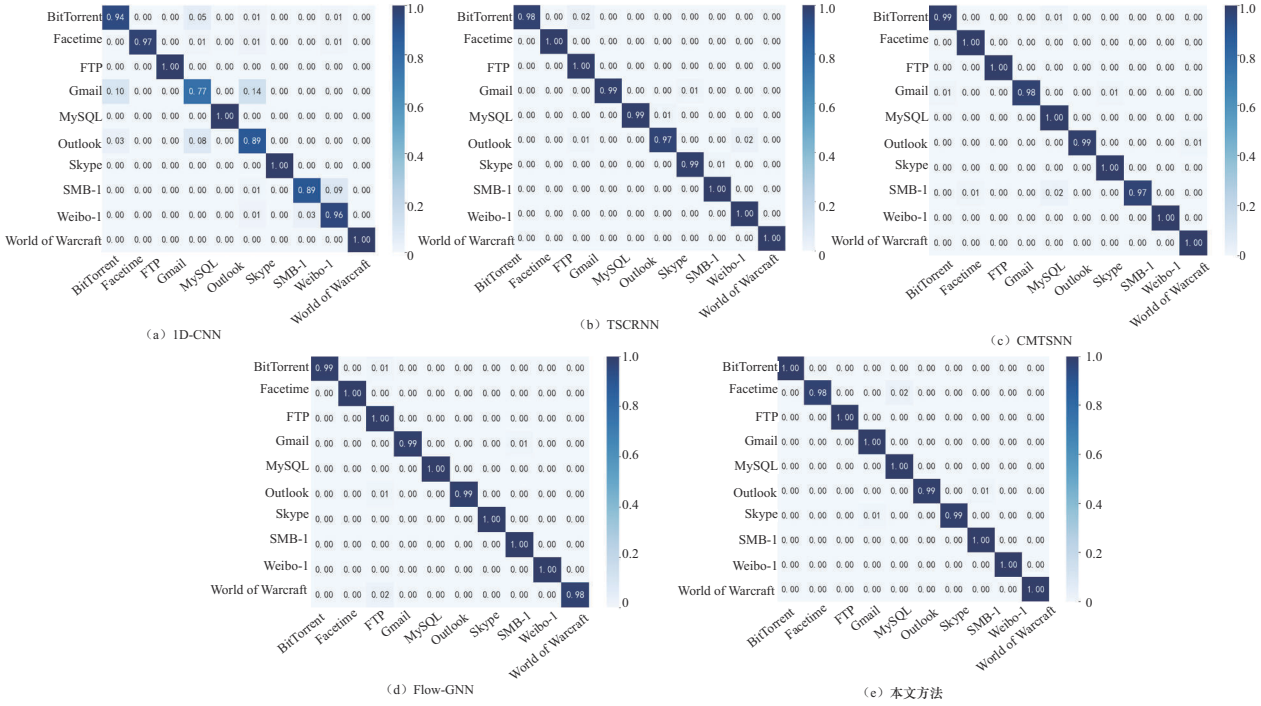


图 11　5 种模型的混淆矩阵分布

法，且对 6 种应用程序的分类准确率达到了 100%，但对某些无法完全正确预测的类别，本文方法的平均准确率仍然更高。这是因为本文方法采用的混合 Transformer 与卷积神经和 LSTM 相结合的特征融合策略，在全局特征和时空特征的捕获上更加精准，从而在分类不确定性较高的类别上也能保证较好的性能。模型 TSCRNN 和 CMTSNN，分别对 5 种和 6 种应用程序进行了准确识别，但总体效果依然不如本文方法。这表明，本文方法在特征提取的有效性和分类稳定性方面优于其他模型，进一步验证了其在加密流量分类任务中的有效性和泛化能力。

2.5　复杂性分析

1D-CNN 的时间复杂度为 $O(NkC)$，取决于输入序列长度 N，输出通道 C 和卷积核大小 k。TSCRNN 的时间复杂度由两部分组成，卷积层部分与 1D-CNN 类似，RNN 部分的时间复杂度为 $O(Nd^2)$，总体时间复杂度为 $O(NkC+Nd^2)$。CMTSNN 的复杂度主要来源于多尺度卷积层，假设有 M 个尺度，每个尺度的卷积核大小为 k_m，输入长度为 N，输出通道数为 C，则多尺度卷积的时间复杂度为 $O(MNk_mC)$。本文方法的时间复杂度为 $O(N^2d+Td^2)$，其中，T 为数据包长度序列。

分析时间复杂度可知，在模型的复杂度方面，本文方法略高于其他 4 种方法，在相同超参数设置的情况下，1D-CNN 的复杂最低，因为模型较为简单，这也限制了其特征提取的深度，在处理流量分类时表现不佳。TSCRNN 结合了卷积层和 RNN，相较于 1D-CNN 有更强的特征学习能力，复杂度也更高。CMTSNN 结合多尺度卷积和时序网络，参数量较少，复杂度低于 TSCRNN。Flow-GNN 的复杂度较高，因为通过图神经网络捕获流量间的关联特征，当数据量较大时，会显著增加计算量。

5 种模型的参数和分类时间见表 3，列出了 5 种模型的参数量、模型大小以及批量大小为 64 时的分类时间。从表 3 中可以看出，因为 1D-CNN 的模型简单，所以参数量和模型大小最小，分类

时间也最短。TSCRNN结合卷积和RNN结构，参数量明显增加，因此模型更大，分类时间也最长。这一结构能够提取时序特征，适合需要捕捉序列信息的任务，但由于其复杂的RNN计算，处理速度受到了显著影响，不适合实时应用。CMTSNN采用多尺度卷积网络，参数量较少，但在捕捉特征上有较好表现，分类时间次于1D-CNN。该模型在保证较小参数量的同时，能提取多尺度特征。Flow-GNN使用图神经网络提取流量特征，能有效处理复杂数据，但由于图卷积的高计算需求，分类时间较长，该模型在复杂任务中具有较强表现，但对资源的需求较高。本文方法的参数量和模型大小最高，复杂度最高，其结构结合了Transformer、卷积和LSTM，在时空特征提取上表现出色，因此在分类精度上更具优势。虽然分类时间不如1D-CNN或CMTSNN快，但相较于TSCRNN和Flow-GNN表现更佳，适合对精度要求高的加密流量分类任务。

表3　5种模型的参数和分类时间

模型	参数量	模型大小/MB	分类时间/ms
1D-CNN	308 160	1.18	90.35
TSCRNN	2 897 104	11.05	1191.92
CMTSNN	939 946	3.50	112.36
Flow-GNN	4 379 232	16.70	501.44
GMTBLC	8 216 282	31.34	362.54

2.6　消融实验

（1）Transformer验证

为了验证本文分类模型对Transformer注意力机制改进的有效性，设置对比实验，使用原始的MHA和改进的GMA进行对比。同时，考虑到不同数量的多头注意力机制对模型性能的影响，使用不同头数的Transformer同使用GAM的Transformer在ISCX VPN-nonVPN数据集上进行实验对比。Transformer模型验证对比结果见表4，表4中2-Transformer、4-Transformer和6-Transformer分别表示编码器中的自注意力层的个数为2、4和6。GMA-Transformer表示使用混合注意力机制的Transformer编码器。

表4　Transformer模型验证对比结果

模型	准确率	精确率	召回率	F1分数
2-Transformer	95.45%	95.20%	95.32%	95.26%
4-Transformer	97.30%	97.02%	97.15%	97.08%
6-Transformer	97.66%	97.32%	97.50%	97.40%
GMA-Transformer	98.70%	98.52%	98.63%	98.41%

由表4可知，使用2头注意力机制的编码器分类模型各项性能指标最低。当头数为4和6时，性能评估指标相差不大。同时，使用混合注意力机制的编码器分类模型各项性能指标均优于使用多头注意力机制的编码器分类模型。由此可知，使用GMA机制能有效地解决传统多头注意力机制的局限性，提高模型的分类性能。

（2）多模态验证

为了验证多模态的有效性，本文针对多模态进行对比实验，分别包括完整模型GMFLNet、仅使用PCMT模块和仅使用SFE模块3种对比模型，在数据集ISCX VPN-nonVPN上进行实验验证，多模态验证对比结果见表5。

表5　多模态验证对比结果

模型	准确率	精确率	召回率	F1分数
PCMT	96.67%	96.30%	96.45%	96.37%
SFE	95.33%	95.20%	95.21%	95.18%
PCMT+SFE	98.70%	98.52%	98.63%	98.41%

由表5可知，仅使用单一模态进行网络流量分类时，其性能指标较多模态均有所下降。其中，仅使用PCMT模块，准确率为96.67%，而仅使用SFE模块进行流量分类时，准确率最低，仅有95.33%。由此可见，使用多模态机制能有效地解决单一模态特征提取不足的缺点，从而提高分类器的整体性能。

3　结束语

本文提出了一种新的网络流量分类算法，通

过使用混合注意力代替Transformer的多头注意力来学习数据包字节之间的关系，获取数据包的全局特征，同时使用带有残差连接的卷积神经网络和双向LSTM提取数据包的时空特征，最后通过动态加权机制，将学习到的全局特征和时空特征进行融合，得到用于分类的融合特征。实验结果显示，与其他方法相比，GMTBLC在流量分类方面，准确率和鲁棒性等方面具有一定的优势。尽管如此，本文方法在流量分类方面仍有提升空间，未来可在本文基础上进一步优化模型参数，提升性能，同时降低模型的复杂度。另外，用于训练的网络流量数据集可能包含用户不愿上传的敏感信息，在未来的工作中，可以引入联合学习，探索本文方法在分布式架构中的可扩展性。

参考文献：

[1] SADEGHZADEH A M, SHIRAVI S, JALILI R. Adversarial network traffic: towards evaluating the robustness of deep-learning-based network traffic classification[J]. IEEE Transactions on Network and Service Management, 2021, 18(2): 1962-1976.

[2] DORIGUZZI-CORIN R, MILLAR S, SCOTT-HAYWARD S, et al. Lucid: a practical, lightweight deep learning solution for DDoS attack detection[J]. IEEE Transactions on Network and Service Management, 2020, 17(2): 876-889.

[3] MOLINA-CORONADO B, MORI U, MENDIBURU A, et al. Survey of network intrusion detection methods from the perspective of the knowledge discovery in databases process[J]. IEEE Transactions on Network and Service Management, 2020, 17(4): 2451-2479.

[4] 郭丽, 刘磊. 基于多层感知器的流量分类方法研究[J]. 电子测量与仪器学报, 2019, 33(7): 56-64.
GUO L, LIU L. Research on refined classification method based on multilayer perceptron[J]. Journal of Electronic Measurement and Instrumentation, 2019, 33(7): 56-64.

[5] MOORE A W, PAPAGIANNAKI K. Toward the accurate identification of network applications[M]//Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005: 41-54.

[6] SEN S, SPATSCHECK O, WANG D M. Accurate, scalable in-network identification of p2p traffic using application signatures[C]//Proceedings of the 13th International Conference on World Wide Web. New York: ACM, 2004: 512-521.

[7] ROUGHAN M, SEN S, SPATSCHECK O, et al. Class-of-service mapping for QoS: a statistical signature-based approach to IP traffic classification[C]//Proceedings of the 4th ACM SIGCOMM Conference on Internet Measurement. New York: ACM, 2004: 135-148.

[8] KHALIFE J M, HAJJAR A, DÍAZ-VERDEJO J. Performance of OpenDPI in identifying sampled network traffic[J]. Journal of Networks, 2013, 8(1): 71-81.

[9] DERI L, MARTINELLI M, BUJLOW T, et al. nDPI: open-source high-speed deep packet inspection[C]//Proceedings of the 2014 International Wireless Communications and Mobile Computing Conference (IWCMC). Piscataway: IEEE Press, 2014: 617-622.

[10] ZHAO Y, YANG Y R, TIAN B, et al. Edge intelligence based identification and classification of encrypted traffic of Internet of things[J]. IEEE Access, 1895, 9: 21895-21903.

[11] 杨宇, 唐东明, 李驹光, 等. 基于时空特征自适应融合网络的流量分类方法[J]. 电子测量技术, 2024, 47(3): 166-174.
YANG Y, TANG D M, LI J G, et al. Traffic classification based on spatiotemporal feature adaptive fusion network[J]. Electronic Measurement Technology, 2024, 47(3): 166-174.

[12] AULD T, MOORE A W, GULL S F. Bayesian neural networks for Internet traffic classification[J]. IEEE Transactions on Neural Networks, 2007, 18(1): 223-239.

[13] 庞兴龙, 朱国胜. 基于半监督学习的网络流量分析研究[J]. 计算机科学, 2022, 49(S1): 544-554, 611.
PANG X L, ZHU G S. Survey of network traffic analysis based on semi supervised learning[J]. Computer Science, 2022, 49 (S1): 544-554, 611.

[14] SHEIKH M S, PENG Y Q. Procedures, criteria, and machine learning techniques for network traffic classification: a survey [J]. IEEE Access, 2022, 10: 61135-61158.

[15] DONG S. Multi class SVM algorithm with active learning for network traffic classification[J]. Expert Systems with Applications, 2021, 176: 114885.

[16] 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(7): 1-9.
YANG Y P, WANG S T. Research on malicious traffic classification algorithm based on CNN combined with BiGRU[J]. Computer Science, 2024, 51(7): 1-9.

[17] WANG W, ZHU M, WANG J L, et al. End-to-end encrypted traffic classification with one-dimensional convolution neural networks[C]//Proceedings of the 2017 IEEE International Conference on Intelligence and Security Informatics (ISI). Piscat-

away: IEEE Press, 2017: 43-48.

[18] DONG S, LI R X. Traffic identification method based on multiple probabilistic neural network model[J]. Neural Computing and Applications, 2019, 31(2): 473-487.

[19] XIE G R, LI Q, JIANG Y. Self-attentive deep learning method for online traffic classification and its interpretability[J]. Computer Networks, 2021, 196: 108267.

[20] SHAPIRA T, SHAVITT Y. FlowPic: a generic representation for encrypted traffic classification and applications identification[J]. IEEE Transactions on Network and Service Management, 2021, 18(2): 1218-1232.

[21] YANG L, GUO S T, LIU D F, et al. ConViTML: a convolutional vision transformer-based meta-learning framework for real-time edge network traffic classification[J]. IEEE Transactions on Network and Service Management, 2024, 21(3): 3344-3357.

[22] DONG S, XIA Y J, PENG T. Network abnormal traffic detection model based on semi-supervised deep reinforcement learning[J]. IEEE Transactions on Network and Service Management, 2021, 18(4): 4197-4212.

[23] DONG S, XIA Y J, PENG T. Traffic identification model based on generative adversarial deep convolutional network[J]. Annals of Telecommunications, 2022, 77(9): 573-587.

[24] ACETO G, CIUONZO D, MONTIERI A, et al. MIMETIC: mobile encrypted traffic classification using multimodal deep learning[J]. Computer Networks, 2019, 165: 106944.

[25] WANG X, CHEN S H, SU J S. App-net: a hybrid neural network for encrypted mobile traffic classification[C]//Proceedings of the IEEE INFOCOM 2020 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS). Piscataway: IEEE Press, 2020: 424-429.

[26] LIN P, YE K J, HU Y S, et al. A novel multimodal deep learning framework for encrypted traffic classification[J]. IEEE/ACM Transactions on Networking, 2023, 31(3): 1369-1384.

[27] LIN K D, XU X L, GAO H H. TSCRNN: a novel classification scheme of encrypted traffic based on flow spatiotemporal features for efficient management of IIoT[J]. Computer Networks, 2021, 190: 107974.

[28] ZHU S Z, XU X L, GAO H H, et al. CMTSNN: a deep learning model for multiclassification of abnormal and encrypted traffic of Internet of things[J]. IEEE Internet of Things Journal, 2023, 10(13): 11773-11791.

[29] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[J]. Advances in Neural Information Processing Systems, 2017, 30: 6000 – 6010.

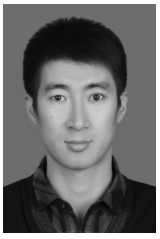
[30] GE C J, DING X H, TONG Z, et al. Advancing vision transformers with group-mix attention[EB]. 2023.

[31] DAI Y M, GIESEKE F, OEHMCKE S, et al. Attentional feature fusion[C]//Proceedings of the 2021 IEEE Winter Conference on Applications of Computer Vision (WACV). Piscataway: IEEE Press, 2021: 3559-3568.

[32] 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.

[33] HUOH T L, LUO Y, LI P L, et al. Flow-based encrypted network traffic classification with graph neural networks[J]. IEEE Transactions on Network and Service Management, 2023, 20(2): 1224-1237.

[作者简介]



魏德宾（1978–），男，博士，大连大学信息工程学院副教授，主要研究方向为天地一体化网络传输技术、流量工程等。

江亲龙（1999–），男，大连大学信息工程学院硕士生，主要研究方向为网络流量分类、网络切片等。

温京龙（2000–），男，大连大学信息工程学院硕士生，主要研究方向为边缘计算、深度强化学习等。

王欣睿（2001–），男，大连大学信息工程学院硕士生，主要研究方向为卫星网络的主动队列管理、路由策略等。