DOI:10.19651/j.cnki.emt.2314716

基于时空特征自适应融合网络的流量分类方法*

杨 宇 唐东明 李驹光 肖宇峰
(西南科技大学信息工程学院 绵阳 621010)

摘 要:针对当前网络流量瞬时涌现导致网络安全事故骤增、网络管理负担加重等问题,基于深度学习技术提出了 ResNet 和一维 Vision Transformer 并行的网络结构对网络流量进行识别并分类。其中 ResNet 可以提取到流量数据在空间上深层次的特征,能够保证流量识别的准确率;一维 Vision Transformer 可以提取到更具代表性的时序特征。利用注意力机制将两种特征进行自适应融合得到更全面的特征表示,以提高网络识别流量的能力。在 ISCX VPN-nonVPN 数据集上进行实验表明:所提方法在流量的应用程序分类实验中的准确率达到了 99.5%,相较于单独的 ResNet 和一维 Vision Transformer 以及经典的一维 CNN 和 CNN+长短时记忆网络分别提高了 0.9%、3.6%、6.6% 和 3.3%。在 USTC-TFC 2016 数据集上,所提方法在能够轻松识别流量是否为恶意流量的基础上,实现了对 13 种应用程序的分类,且平均分类准确率达到了 98.92%,证明了其具有识别恶意流量并完成细粒度分类任务的能力。

关键词:流量分类;ResNet;vision Transformer;多头注意力机制;特征融合
中图分类号:TP393 **文献标识码**:A **国家标准学科分类代码**:520.3040

Traffic classification based on spatiotemporal feature adaptive fusion network

Yang Yu Tang Dongming Li Juguang Xiao Yufeng
(School of Information Engineering, Southwest University of Science and Technology, Mianyang 621010, China)

Abstract: In response to the current surge in network traffic leading to a sudden increase in network security incidents and an added burden on network management, a network architecture based on deep learning techniques has been proposed. This architecture involves the parallel use of ResNet and one-dimensional Vision Transformer for the identification and classification of network traffic. ResNet is capable of extracting deep spatial features from flow data, ensuring high accuracy in traffic recognition. Meanwhile, the one-dimensional Vision Transformer excels at capturing more representative temporal features. By employing an attention mechanism to adaptively merge these two types of features, a more comprehensive feature representation is obtained to enhance the network′s capability in traffic identification. Experiments conducted on the ISCX VPN-nonVPN dataset demonstrate that the proposed method achieves an accuracy of 99.5% in application-based traffic classification experiments. Compared to standalone ResNet and one-dimensional Vision Transformer, as well as classical one-dimensional Convolutional Neural Networks (1D-CNN) and CNN combined with Long Short-Term Memory (CNN + LSTM), the proposed method shows improvements of 0.9%, 3.6%, 6.6%, and 3.3%, respectively. On the USTC-TFC 2016 dataset, the proposed method not only easily identifies malicious traffic but also accomplishes the classification of 13 different applications, with an average classification accuracy of 98.92%. This proves its ability to recognize malicious traffic and perform fine-grained classification tasks.

Keywords: traffic classification; ResNet; vision Transformer; multi-head attention mechanism; feature fusion

0 引 言

近年来,随着互联网应用的快速发展以及互联网用户数量的不断增加,截止到 2023 年 8 月,我国互联网普及率已达到 76.4%。在用户量如此之大的情况下,极易出现网络流量的瞬时涌现,严重增加了网络管理负担。除此之外,

收稿日期:2023-10-09
* 基金项目:国家自然科学基金(12175187)项目资助

广泛应用的流量加密技术也使得恶意流量泛滥,严重影响了网络安全。针对这些问题,利用深度学习技术对网络流量进行分类可以帮助网络管理员及时发现和应对潜在的安全威胁,提高网络的安全性、稳定性以及管理效率,因此网络流量分类技术对于网络安全和管理具有重要意义。

早期的流量分类主要运用的是基于端口号和深度包检测(deep packet inspection, DPI)[1]的方法。基于端口号对流量进行分类是利用已知的应用端口号与标准端口号进行匹配对比实现的。如 FTP 网络应用端口号为 21、SSH(安全登录)默认端口号为 22。然而,随着网络技术的发展,动态端口技术和端口伪装技术被大量运用,使得基于端口号的方法可用性大大降低[2]。基于 DPI 的方法需要对流量数据进行解析,根据已知的协议签名来检测数据包的有效负载,进而对流量进行分类。当流量为加密流量时,该方法将会失效。

随着机器学习技术的兴起,研究人员意识到可以将机器学习应用于加密流量识别。如 Shafiq 等[3]使用了 C4.5 决策树、朴素贝叶斯、支持向量机(support vector machine, SVM)等多种算法对实时互联网数据进行网络流量分类。庞兴龙等[4]对半监督学习在流量分析问题上的研究进展进行了总结,阐述了半监督分类、聚类以及降维 3 个方面在网络流量分析中的实际应用,并指出了未来研究所面临的挑战以及新的方向。虽然机器学习在流量分类的应用取得了一定成果,但此类方法需要人工提取特征,依赖于专家经验,具有主观因素,且费时费力,而深度学习可以很好地解决这一问题。

深度学习是一个端到端自动化的过程,采用卷积神经网络(convolutional neural network,CNN)、循环神经网络(recurrent neural network,RNN)等神经网络对数据进行特征提取并作为 Softmax 层的输入,最后可以直接输出预测的标签。这种端到端结构能得到输入和输出的直接映射,天然具有协同效应,更有可能获得全局最优解[5]。因此当前许多流量分类都是基于深度学习技术实现的。

Wang 等[6]提出了一种使用一维 CNN(one-dimensional CNN,1D-CNN)对加密流量进行分类的方法,他们将流量中的每个字节作为一个像素点,截取每个数据包的前 784 个字节生成对应的灰度图,并将其作为神经网络的输入。实验使用的是“ISCX VPN-non VPN”数据集,在 12 类应用程序分类实验中得到了 92%的准确率。Zou 等[7]提出了一种 CNN 和长短时记忆网络(long short-term memory,LSTM)相结合的深度神经网络结构,通过实验得到了较高的分类准确率。Rezaei 等[8]总结出了一个基于深度学习对网络流量进行分类的通用框架,包括数据收集、预处理、特征选择和模型选择,为流量分类工作提供指南。Xie 等[9]从可解释性角度,选择自注意力(Self-Attention)对网络流量进行在线分类,准确率达到了 92%±0.86%,在时间上,其能在 2 ms 左右对数据包进行识别。Shapira 等[10]在对加密网络流量分类和应用程序识别实验中,将基本流量数据转换为直观的图片“FlowPic”,然后应用深度学习在图像领域的分类技术来对流量类型进行识别,如浏览、聊天、视频以及应用程序等。实验结果表明,该方法能够完成高精度的网络流量类型识别。张彦晖等[11]提出了一种基于卷积注意力门控循环网络的加密流量分类方法,该方法相对于 CNN+LSTM 和 CNN+LSTM+Attention 等方法提高了分类准确率、实时性和训练效率。HaoLi[12]将 CNN 和多头注意力(multi-head attention,MA)进行结合,并使用其提出的一个丢弃阈值来提高通过特征工程获得的数据集的质量,与 C4.5 和 1D-CNN 相比,作者所用方法的准确率最高,达到了 93.8%。王帅等[13]将特征生成和 LSTM 相结合的模型应用于网络流量分类。通过实验比较了原数据和特征数据在分类问题上的准确性,并将其与 CNN 进行对比。所提方法的细分类准确度为 93.9%,粗分类准确率为 99.2%。Lin 等[14]提出基于 Transformer 的加密流量双向编码器表示,作者先将该模型在大规模未标记的数据集中进行预训练,以得到数据包级的较深程度的上下文表示,然后把预训练得到的模型在下游任务上进行微调,最后在 5 个加密流量分类任务中实现了非常先进的性能。郭祥等[15]提出了通过对 Inception 模块进行改进,并将其作为残差块嵌入卷积神经网络的方式来构建模型,并对损失函数进行了改进,该方法比 C4.5 和 1D-CNN 在精确率上分别高出 13.91%和 9.50%。

综上所述,多数研究采用单一的网络模型提取流量特征,缺乏从多角度进行特征提取的机制;在应用卷积神经网络和循环神经网络实现流量数据特征提取时,研究人员更倾向于先用卷积神经网络提取空间特征,再用循环神经网络从空间特征中提取时序特征,这类单一网络或多种网络串行策略无法形成对网络流量数据的更多元化、更全面的特征表示。其次,在提取流量的时序特征时,注意力机制的并行计算使得其相对于循环神经网络有更高效的计算和训练速度。

为了进一步提高对网络流量的识别能力,本文针对以上不足,从多角度出发,在现有的研究基础上提出采取并行策略设计时空特征自适应融合网络(spatiotemporal feature adaptive fusion network, SFAFNet),分别应用残差神经网络(residual neural network, ResNet)和一维 Vision Transformer(one-dimensional vision transformer,1D-VIT)提取流量数据的空间特征和时序特征,再运用注意力机制将提取到的两种特征进行深度融合,生成新的更全面的特征表示,以提高流量分类的准确率。其中,1D-VIT 是 Vision Transformer(VIT)模型的变体,改进后的模型可以提取到更具代表性的时序特征。实验结果表明,本文所提模型相比于经典算法在准确率等方面均具有一定的优势。方法设计的具体细节在第 1 节中进行介绍。第 2 节则是对本文所提方法进行实验验证的结果与分析。最后对本文进

行了总结并构思下一步的研究工作。

1　方法设计

本文所设计的 SFAFNet 网络总体结构如图 1 所示。

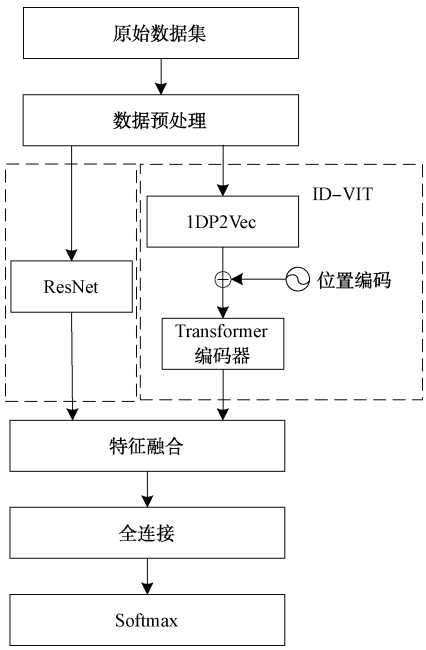

图 1　网络结构图

1.1　数据预处理

数据预处理阶段的主要目标是优化原始数据集，以提高其可用性和适用性，使其可以直接应用于本文所设计的深度学习模型中，从而提高实验效果和结果的可靠性。

对于本文的任务要求，首先需要将多个 Pcap 包按照需求进行合并，得到每种流量类型对应的 Pcap 包。Pcap 格式的流量数据有 24 字节的 Pcap 全局报头，在全局包头后面有多条数据包数据记录，其中每一个数据包都包含了 16 字节的数据包包头和数据包数据。Pcap 全局报头和数据包包头在实际中可能携带噪音，且对于流量的应用程序分类意义不大，如果在数据预处理阶段移除或忽略该部分数据，可以简化模型复杂性和数据的处理过程，同时降低过拟合风险，提高网络泛化能力。

考虑到若进行在线流量分类，当网络流量中有对时间延迟敏感的流量数据时，从流层面分析会影响时效性。因此，本文将区别于 Wang[6] 和 Shapira[10] 等人的处理方法，选择从更细粒度的流量包层面对流量数据进行分析，将流量数据中每一个数据包转换成灰度图像，而非将网络流或会话进行转换。

网络流量转灰度图像的过程是将原始数据进行升维，转换为神经网络更加易于处理的形式。转换出来的图片的大小会直接影响神经网络模型的分类效果。ISCXVPN2016 数据集中的数据包长度均小于 1 500 字节[15]，因此本文选择数据包有效载荷的前 1 024 字节，超过 1 024 字节的数据做截断处理，不满 1 024 字节的数据则以 0 填充，其中的截断和填充操作在一定意义上保护了用户的隐私。为加快训练速度、避免梯度消失和梯度爆炸等问题，把每个字节转化为 0～255 之间的数值后，需要进行归一化处理，得到 1 024 个大小在 0～1 之间的值，将每一个值都作为一个像素点，通过 Python 代码将其转换为 32×32 大小的灰度图像作为神经网络的输入。为防止出现样本不平衡现象，本文将每类应用程序流量都对应生成同样张数的灰度图样本。

最后将生成的图片转换为 idx 文件，在 idx 文件中，每张处理好的流量图片都会有与其应用程序相对应的标签。整个数据预处理流程如图 2 所示。

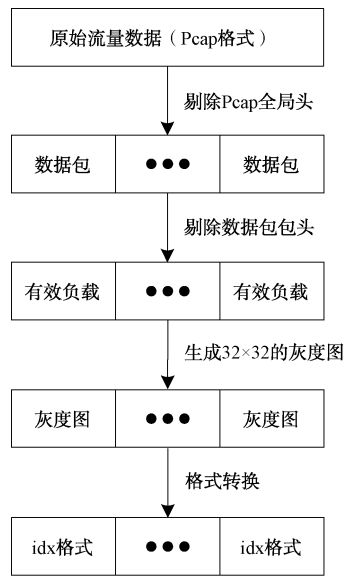

图 2　数据预处理流程图

1.2　空间特征提取分支

ResNet 在图像识别、目标检测等领域都有很广泛的运用，它是由 He 等[16]在 2015 年提出的一种深度卷积神经网络。其在传统卷积神经网络中引入了残差学习，即跳跃连接结构，它可以从某一层网络层获取激活，然后迅速反馈给另外一层，甚至是神经网络的更深层。

在单纯的 CNN 中，每一层都会进行 $y=F(x)$ 计算。其中 $\boldsymbol{x}$ 是输入，$\boldsymbol{y}$ 是输出，$\boldsymbol{F}$ 是激活函数，本文选择 ReLU 函数作为激活函数。

当存在误差时，随着 CNN 的层数不断增加，多次进行的 $y=F(x)$ 操作会使得误差越来越大，梯度在反向传播的过程中也会越来越发散，进而产生网络退化的问题。引入短连接结构，会将层与层之间的映射关系变为 $y=F(x)+x$。此时存在一个恒等学习的短连接：若这些残差块不能学习到新的东西，那么它可以等于恒等学习，让网络不至于产生退化；若这些残差块能够学到一些有用信息，那么它比学习恒等函数表现得更好。该结构很好的解决了深层网络中梯度弥散和爆炸以及精度下降的问题，使得网络在保证

精度的同时能做到越来越深。卷积网络越深,其学习到的面越广,学习到的特征也由浅至深,能有效提高分类任务的准确率,这也是本文在空间特征提取分支上采用 ResNet 网络结构的原因。

本文在空间特征提取分支所用残差学习块如图 3 所示。每个残差块包含两个卷积层,每个卷积层使用 3×3 的卷积核和 1 个填充(padding)来进行计算。输出通道数(out_channel)和第一层的卷积步长(stride)根据需求进行设置。图中的组归一化(Group Norm,GN)是一种强大的深度学习优化技术,它在通道方向计算每个分组的均值和方差,将输入数据规范化到均值为 0、方差为 1 的标准正态分布,从而减少不同分组的数据分布差异,可以起到提高模型泛化性能、改善梯度传播以及使模型适应不同尺度的特征的作用,且相对于 Batch Norm,GN 不受批处理大小的约束。

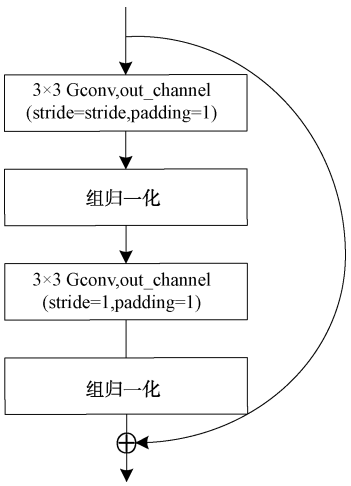

图 3　残差学习块

深层网络拥有强大的表示能力和学习能力,但在处理小型图片时,有可能会过度记忆训练集中的每个样本,导致模型失去对新数据的泛化能力,从而出现过拟合现象,因此本文的空间特征提取分支网络结构并不会设计得很深,其结构如表 1 所示。

表 1　空间特征提取分支网络结构

网络层	输出大小
输入	1×32×32
卷积层	64×32×32
最大池化	64×16×16
残差块	64×16×16
残差块	128×8×8
残差块	256×4×4
残差块	512×2×2
平均池化	512×1×1
全连接	480×1×1

在数据预处理阶段,流量数据被处理为对应的 32×32 大小的灰度图像,因此空间特征提取分支的输入是 1×32×32,在图像输入网络后,首先对图像使用 3×3 的卷积核进行一次恒等卷积,并扩大输出通道至 64,在之后使用最大池化可以在保留主要特征不变的同时,减少计算量,然后通过四个连续的残差块进一步提取更抽象、更高级别的特征,之后使用平均池化得到 512×1×1 的特征向量,将其作为图像的空间特征。为方便特征融合,最后需要用全连接层将特征向量压缩至 480×1×1。

1.3　时序特征提取分支

Vaswani 等[17]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。

Transformer 模型的编码器部分的核心部件是 MA,主要目的是在输入序列中寻找相关信息并进行加权汇总以产生输出序列。多头注意力机制是自注意力机制的一种扩展形式,具体来说,自注意力机制将一个输入序列中的各个元素与自身进行加权叠加,得到该序列的表示。而多头注意力机制则通过并行地使用多组不同的注意力机制来得到更丰富的表示。每组注意力机制被称为"头",每个头都学习一种不同的表示,并将这些表示进行拼接、投影得到最终的表示。

自注意力机制会使用不同的权值矩阵 $\boldsymbol{w}_q$、$\boldsymbol{w}_k$ 和 $\boldsymbol{w}_v$ 经过线性变换,将输入序列 $\boldsymbol{M}$ 映射到新的表示空间,得到对应的查询矩阵 $\boldsymbol{Q}$、键矩阵 $\boldsymbol{K}$ 和值矩阵 $\boldsymbol{V}$,然后计算 $\boldsymbol{Q}$ 和 $\boldsymbol{K}$ 的相似度得分,最后通过降维和归一化计算注意力权重。如式(1)所示。

$$Attention(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V}) = Softmax\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}}{\sqrt{d_k}}\right)\boldsymbol{V} \tag{1}$$

其中,d_k 表示键矩阵的维数。

多头注意力机制则根据设置的头的个数(本文设置头的个数为 12)先将 $\boldsymbol{Q}$、$\boldsymbol{K}$、$\boldsymbol{V}$ 进行对应次数的线性映射,再并行执行注意力函数得到对应的 $head_i$,最后会将所有的 $head_i$ 拼接起来并再次投影,得到最终值,如式(2)、(3)所示。

$$head_i = Attention(\boldsymbol{Q}\boldsymbol{W}_i^{\boldsymbol{Q}},\boldsymbol{K}\boldsymbol{W}_i^{\boldsymbol{K}},\boldsymbol{V}\boldsymbol{W}_i^{\boldsymbol{V}}) \tag{2}$$

$$MA(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V}) = Concat(head_1,\cdots,head_n)\boldsymbol{W}^o \tag{3}$$

从公式角度理解,多头注意力机制可以通过多个头同时进行注意力计算,从而使模型能够关注不同方面的信息,提取不同层次的特征,可帮助模型充分利用输入序列中的上下文信息,提高模型性能,尤其对于具有时序特征的输入数据。

本文在数据预处理阶段截取了每个数据包除开包头的前 1 024 个字节作为有效数据,如果将每个字节类比为 NLP 领域生成对应的词向量,这将导致网络的计算量变得非常复杂。为减少计算量,同时使两个分支网络拥有相同

的输入，需要将 Transformer 模型应用于图像领域。Dosovitskiy 和 Liu 等[18-19]分别提出了 VIT 模型和 Swin Transformer 模型，其中 VIT 模型更适合处理较小的图片，而本文生成的流量图片大小为 32×32。因此，时序特征提取模块参考了 VIT 的网络结构和思路，如图 4 所示。

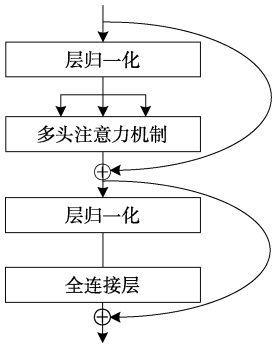

图 4　时序特征提取模块

对于第一个时序特征提取模块，首先需要将流量图片输入到图 1 中的 1DP2Vec（one-dimensional patch to vector）进行处理，将其转换为时序特征提取模块可以直接应用的形式。VIT 利用二维 CNN 将图片划分为多个 patch，并为每个 patch 生成相应的词向量。然而，网络流量本质上是一维序列数据，相比于二维 CNN，1D-CNN 可以更好地学习到更具代表性的流量时序特征[5]，因此本文中的 1DP2Vec 是基于 1D-CNN 的。

采用 1D-CNN 将输入图像在时间方向上依次进行“分割”，且每“分割”出来一块图片就类比 NLP 领域生成一个维度为 480 的词向量。1DP2Vec 操作可以类比于图 5。

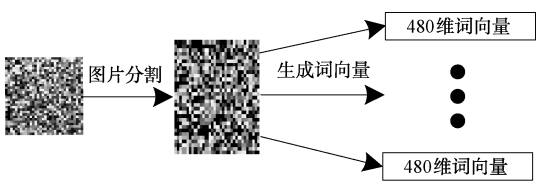

图 5　1DP2Vec 操作示意图

位置编码（positional encoding）是一种旨在为每个图像元素分配一个位置向量，以便将位置信息纳入模型中的技术。为保留流形图片的时序特点，需要在 1DP2Vec 的输出添加位置编码，再将整体作为时序特征提取分支的输入。LayerNorm 的主要作用是将神经元的输出进行归一化，消除输入数据在不同位置、网络深度或批次间带来的分布不一致性，提高模型的泛化能力和训练速度。时序特征提取模块中的短连接是参考 ResNet 设计理念，为防止有多个时序特征提取模块时出现梯度消失和网络精度下降。模块中全连接的作用是将多头注意力机制提取出来的时序特征再抽象出更高级别的特征，使得后续的模型训练和预测更加准确和有效。同时，为降低计算复杂度，保持与空间特征提取分支网络结构提取到的特征维度一致，全连接层会在提取到更多维度的特征后进行信息压缩，保持输出为 480 维。

1.4　特征融合

特征融合[20-21]指的是将多个来源的特征信息结合起来，形成一个更多元化、更全面的特征表示。

由于不同来源的特征通常包含着互补的信息，比如本文从空间和时序两个方面提取对应的特征，将它们合并起来可以提供更全面、更准确的信息，从而提高模型的性能。特征融合可以减少单一特征的失效对整体效果的影响，增强模型的鲁棒性。通过对多个从不同角度获取的特征进行结合，可以减少模型对于单一特征的过分依赖，降低过拟合的风险。对于复杂多样的任务，不同来源的特征可以提供不同维度的信息，通过融合可以实现更好的适应性和灵活性。

综上，特征融合拥有诸多优点。本文旨在利用并行的 ResNet 和 1D-VIT 分别提取流量数据的空间特征和时序特征并将两者进行融合，以达到更好的效果。在将两种特征进行融合时如何合理有效地设计融合方式及权重十分重要。常见的特征融合方法有相加（Add）和拼接（ConCat），其各特征权重的选择通常设定为一固定参数，这往往不能合理地适应不同特征的融合需求。为解决这一问题，本文采用注意力机制设计自适应动态分配权重方法将空间特征和时序特征进行深度融合。注意力机制与自注意力机制的区别在于前者的 $\boldsymbol{Q}$、$\boldsymbol{K}$、$\boldsymbol{V}$ 不完全相同，本文按照如式（4）所示的公式将两种特征进行融合。

$$\boldsymbol{f}=\begin{bmatrix} Attention(\boldsymbol{f}_1,\boldsymbol{f}_2,\boldsymbol{f}_1)+Attention(\boldsymbol{f}_1,\boldsymbol{f}_2,\boldsymbol{f}_2) \\ \boldsymbol{f}_1 \\ \boldsymbol{f}_2 \end{bmatrix} \tag{4}$$

其中，$\boldsymbol{f}_1$ 和 $\boldsymbol{f}_2$ 分别表示空间特征和时序特征，$\boldsymbol{f}$ 表示融合后得到的整体特征。

1.5　分类器及模型训练

将经过融合的特征输入分类器进行分类。本文的分类器由全连接层（fully connected，FC）和 Softmax 层构成。特征向量经过全连接层后可以得到样本的特征向量和样本分类得分的映射，其意义在于将较高层次的抽象特征转换为人类或机器能够更好理解的形式，可以将 Softmax 操作看作是一个归一化过程，经过 Softmax 层后，全连接层输出的样本分类得分将转换为概率。Softmax 运算公式如（5）所示。

$$Softmax(z_i)=\frac{e^{z_i}}{\sum_{j=1}^{n}e^{z_i}} \tag{5}$$

损失函数是用于评价模型预测结果与真实值之间差距的一种函数。一般来说，当模型的损失函数值越小，则模型的性能越好。因此，定义一个好的损失函数能够帮助我们评估模型的优劣，并促使模型学习到更好的参数以提高准确性。本文选择交叉熵损失函数。其公式如式（6）所示。

$$H(p,q) = -\sum_{i=1}^{n} p(x_i)\log(q(x_i)) \tag{6}$$

由式(6)可知,交叉熵损失函数是凸函数,因此可以在每次迭代中更快地找到全局最优解,从而加速模型的训练。

2　实验与分析

本文实验是基于硬件为一台内存 16 GB,搭载 NVIDIA GeForce RTX 2060 显卡的 64 位 Windows11 操作系统 PC;软件环境是以 Python3.9.16 为基础语言环境,采用 2.0.0 版本的 Pytorch 框架,对应的 CUDA 版本为 11.7。

在该软硬件环境下搭建本文第 1 章中提出的神经网络结构,设置网络训练的批处理大小为 64,学习率为 0.001,训练次数为 30 次,并使用 Diederik 等[22] 提出的 Adam (adaptive moment estimation)优化器对参数进行学习。

2.1　实验数据集

本文选择公开数据集 ISCX VPN-nonVPN (ISCXVPN2016)和 USTC-TFC 2016 对所提方法进行验证。ISCXVPN2016 数据集是由 Wireshark 抓包工具抓取获得的具有代表性的真实流量数据集,其数据主要为 Pcap 格式。数据集中有通过 VPN 加密的会话和常规会话(即 non VPN 会话),所以流量类型总共有 14 类,分别为 Chat、VPN-Chat、P2P 和 VPN-P2P 等,整个数据集大小有 27.5 GB。数据集具体内容如表 2 所示。

表 2　ISCXVPN2016 数据集内容表

流量类型 (VPN-nonVPN)	应用程序
Web Browsing	Firefox、Chrome
Email	SMPTS、POP3S、IMAPS
Streaming	Vimeo、Youtube
Chat	ICQ、AIM、Skype、Facebook、Hangouts
VoIP	Facebook、Skype、Hangouts (voice calls (1h duration))
P2P	uTorrent、Transmission、Bittorrent

表 2 列出了 ISCXVPN2016 数据集中所有类型的流量,以及与每种流量类型对应的应用程序。常规流量可以用常规方法进行识别,而加密流量的识别则需要用深度学习才能达到令人满意的效果。因此,本文在选择十种使用了 VPN 加密的应用程序流量,应用程序类型分别为:Facebook、Hangouts、ICQ、netflix、Email、Skype、Spotify、Vimeo、Voipbuster 和 Youtube。

USTC-TFC 2016 数据集是由王伟等[23] 创建的,其中包含了 10 类正常加密流量(Benign)和 10 类恶意加密流量(Malware),如表 3 所示。

在 USTC-TFC 2016 数据集中,本文将先将每种应用程序的 Pcap 文件按照流量类型进行合并,以进行恶意流量识别实验,后选择了 13 种应用程序进行细粒度的分类实验。

表 3　USTC-TFC 2016 数据集内容表

流量类型	应用程序
正常加密流量	BitTorrent,Facetime,FTP,Gmail,MySQL,Outlook,Skype,SMB,Weibo,WorldOfWarcraft
恶意加密流量	Cridex,Geodo,Htbot,Miuref,Neris,Nsis-ay,Shifu,Tinba,Virut,Zeus

2.2　模型评价

本文选择准确率(Accuracy)、假阳率(False Positive Rate,FPR)、F1 得分以及可视化混淆矩阵来评估本文网络结构的分类效果。准确率代表流量应用程序类型预测正确的概率;FPR 表示被错误地判断成正例的负例样本所占所有负例样本的比例;F1 得分是模型性能的综合指标,它是由精确率 Precision 和召回率 Recall 计算得出;混淆矩阵是深度学习中总结分类模型预测结果的情形分析表,它以矩阵形式将测试集的预测结果按照真实的类别与模型预测的类别两个判断标准进行汇总,可以直观地展现出模型的预测效果。准确率、FPR 以及 F1 计算公式如式(7)~(9)所示。

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{7}$$

$$FPR = \frac{FP}{FP + TN} \tag{8}$$

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \tag{9}$$

其中,Precision 和 Recall 计算公式如下。

$$Precision = \frac{TP}{TP + FP} \tag{10}$$

$$Recall = \frac{TP}{TP + FN} \tag{11}$$

其中,TP 表示正确预测为正例的样本数,TN 表示正确预测为负例的样本数,FP 表示将负例误判为正例的样本数,FN 表示将正例误判为负例的样本数。根据 TP、TN、FP 和 FN 的定义,可通过混淆矩阵中的值对其进行计算。

$$TP_i = (i,i) \tag{12}$$

$$TN_i = total - (FP_i + FN_i + TP_i) \tag{13}$$

$$FP_i = \sum_{j=1,j\neq i}^{n} (j,i) \tag{14}$$

$$FN_i = \sum_{j=1,j\neq i}^{n} (i,j) \tag{15}$$

其中,i 表示处于第 i 行第 i 列的元素,$total$ 表示为样本的总数,求和符号上的"n"对应于 n 分类任务。

2.3　模型有效性实验

为了验证所提方法的有效性,本文在 ISCXVPN2016 数据集上进行了模型有效性分析实验,实验采用了 5 种模型进行对比实验和消融实验。这 5 种模型分别为本文设计

的特征融合网络 SFAFNet(时序特征提取模块数为 1)、经典的 1D-CNN 和 CNN+LSTM 以及本文所提网络结构的两个分支 ResNet 和 1D-VIT。

图 6 展示了整个训练过程中五种模型在测试集上的准确率变化曲线。图中横轴表示训练次数,纵轴表示准确率,每种曲线代表一种模型。从曲线的变化趋势来看,SFAFNet 相较于其他四种方法的训练和收敛都更快。

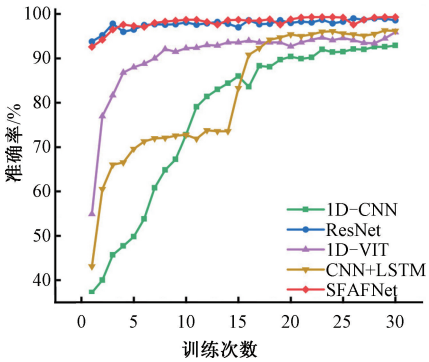

图 6　准确率曲线

图 7 为 5 种网络模型在同样的超参数设置下进行流量应用程序分类的平均准确率、F1 得分和 FPR 值的柱状图。

从图中可以直观地看到本文所提网络 SFAFNet 在网络流量应用程序分类任务中的平均准确率最高,达到了 99.3%,且 FPR 最低,为 0.000 7,相对于其他模型具有显著优势。本文的 F1 得分是将每类的 F1 得分进行求均值得到,所以在数值上与准确率非常接近。

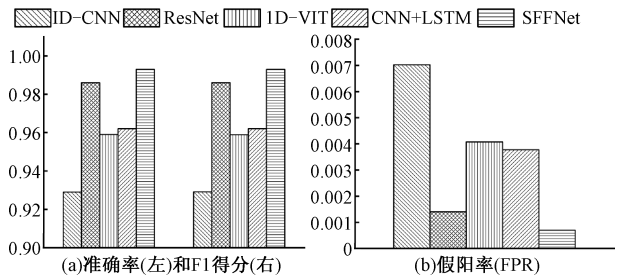

图 7　指标柱状图

为进一步观察不同模型对每类样本的分类效果,在图 8 中列出了 5 种方法的混淆矩阵分布图。其中,混淆矩阵的"行"为预测标签,"列"为真实标签。从图中可以观察到,SFAFNet(图 8(e))有 5 类应用程序的识别准确率达到 100%,而其他模型准确率达到 100% 的不超过 4 个;SFAFNet 对于 Youtube 和 Email 识别的准确率最低,为 98%,而 1D-CNN(图 8(a))对 Netflix 的识别准确率仅为 79%,ResNet(图 8(b))对 Netflix 的识别准确率为 95%,1D-VIT(图 8(c))对 Hangouts 的识别准确率为 87%,CNN+LSTM(图 8(d))对 Spotify 的识别准确率为 90%。通过对实验结果的分析,可以看出 SFAFNet 相较于其他四种模型具有显著优势,充分证明了其有效性。

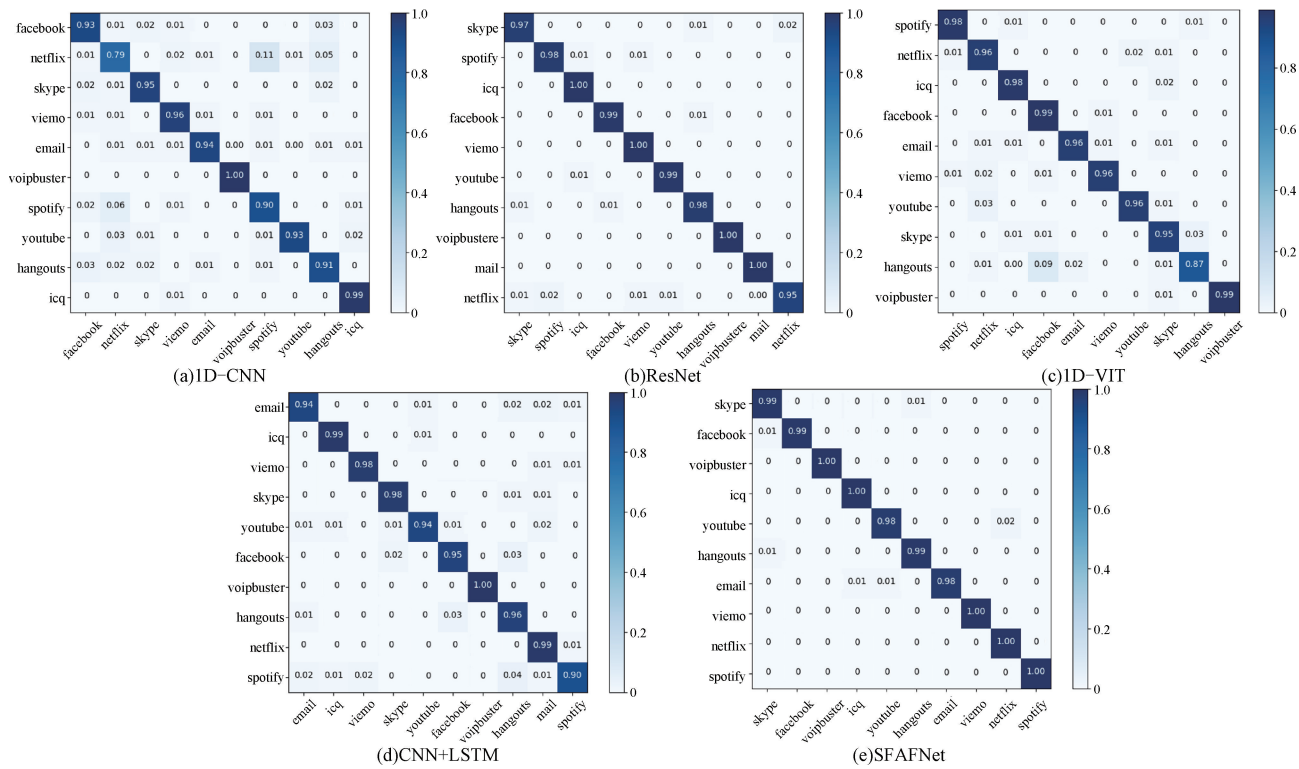

图 8　5 种模型的混淆矩阵

2.4　VIT 改进有效性实验

为验证本文对 VIT 的改进是否有性能的提升,将改进前后的 VIT 在 ISCXVPN2016 数据集上进行了对比实验。VIT 模型改进前后对比实验结果如表 4 所示。从表中数据可以反应出改进后的 VIT 在性能上有一定的提高,验证了对于本文任务来说,1D-CNN 比 2D-CNN 可以更好地学习到更具代表性的流量时序特征。

表 4　VIT 改进前后实验结果表

模型	模型分类准确率/%	F1 得分	假阳率
VIT	99.1	0.990 98	0.000 9
1D-VIT	99.3	0.993	0.000 7

2.5　时序特征提取模块数取值实验

为探索时序特征提取模块的个数对 SFAFNet 模型的影响,本文在 ISCXVPN2016 数据集上进行了模块数取值实验,采用十种不同取值的模块数进行实验,并将实验结果进行比较分析,以得出最佳方案。时序特征提取模块的数量与模型分类准确率对应关系如表 5 所示。

表 5　时序特征提取模块数实验结果表

时序特征提取模块个数	模型分类准确率/%
1	99.3
2	99.2
3	99.1
4	99.5
5	99.2
6	99.1
7	99.1
8	99.0
9	98.7
10	98.2

根据实验结果显示,当时序特征提取模块的个数为 4 时,此时准确率最高,达到了 99.5%;当模块个数超过 8 个时,准确率明显下降,出现了过拟合问题。综合考虑,最终模型的时序特征提取模块个数被确定为 4。

2.6　恶意流量识别并分类实验

为验证 SFAFNet 模型具有识别恶意流量并完成细粒度分类的能力,本文在 USTC-TFC 2016 数据集上进行了对应的实验。实验结果表明 SFAFNet 模型能完成对恶意流量的精确识别。

5 种模型对 13 种应用程序进行分类的平均准确率如表 6 所示。从表中数据可以看到 SFAFNet 模型的分类准确率最高,达到了 98.92%,较第 2 名高出了 0.77%,充分证明了该模型可以完成识别恶意流量并进行细粒度分类任务。

表 6　五种模型的应用程序分类准确率表

模型	平均分类准确率/%	假阳率
1D-CNN	86.69	0.010 1
ResNet	98.15	0.001 4
1D-VIT	92.31	0.005 8
CNN+LSTM	95.85	0.003 2
SFAFNet	98.92	0.000 8

图 9 为 SFAFNet 模型对 13 种应用程序分类的混淆矩阵分布图,其中 8 种应用程序的准确率可以达到 100%。

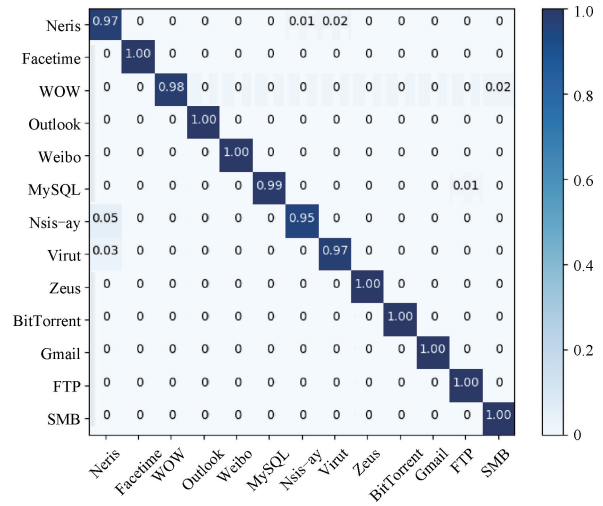

图 9　应用程序分类的混淆矩阵

3　结　　论

随着网络技术的更新迭代,互联网环境日益复杂,针对如何提高网络的安全性以及网络资源分配的效率等问题,本文提出了一种基于 ResNet 和 1D-VIT 并行的网络模型 SFAFNet,并通过公开数据集对流量进行了应用程序分类实验。SFAFNet 具有以下特征:将细粒度的包层面数据转换为灰度图像,并将其作为 SFAFNet 的输入;使用注意力机制将分支网络分别提取到的空间特征和时序特征进行深度融合,利用多种数据来源提供的信息优势,得到更为全面的特征表示,提升了模型性能及流量分类的准确率;为提取到更具代表性的时序特征对 VIT 进行了改进。实验结果表明,相对于现有的部分方法 SFAFNet 具有显著优势,其有更高的分类准确率、良好的鲁棒性和泛化能力。本文模型对于恶意应用程序流量的分类能力还可以进一步提高,后续可以在本文的基础上针对模型的超参数开展进一步的算法优化工作,使模型拥有更好的性能;同时针对真实网络环境下复杂流量数据的分类问题,考虑采用增量学习等方法对分类模型实现进一步的改进。

参考文献

[1]　KHALIFE J, HAJJAR A, DÍAZ-VERDEJO J.

Performance of OpenDPI in identifying sampled network traffic[J]. Journal of Networks, 2013, 8(1): 71.

[2] 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.

[3] SHAFIQ M, YU X, LAGHARI A A, et al. Network traffic classification techniques and comparative analysis using machine learning algorithms[C]. 2016 2nd IEEE International Conference on Computer and Communications(ICCC), IEEE, 2016: 2451-2455.

[4] 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.

[5] DAHL G E, SAINATH T N, HINTON G E. Improving deep neural networks for LVCSR using rectified linear units and dropout[C]. 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, IEEE, 2013: 8609-8613.

[6] WANG W, ZHU M, WANG J, et al. End-to-end encrypted traffic classification with one-dimensional convolution neural networks [C]. 2017 IEEE International Conference on Intelligence and Security Informatics(ISI), IEEE, 2017: 43-48.

[7] ZOU Z, GE J, ZHENG H, et al. Encrypted traffic classification with a convolutional long short-term memory neural network [C]. 2018 IEEE 20th International Conference on High Performance Computing and Communications; IEEE 16th International Conference on Smart City; IEEE 4th International Conference on Data Science and System, 2018: 329-334.

[8] REZAEI S, LIU X. Deep learning for encrypted traffic classification: An overview [J]. IEEE Communications Magazine, 2019, 57(5): 76-81.

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

[10] 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.

[11] 张彦晖,吕娜,刘鹏飞,等.基于卷积注意力门控循环网络的加密流量分类方法[J].信号处理,2021,37(7):1180-1188.

[12] HAOLI. Traffic classification algorithm using CNN and multi-head attention mechanism for representation learning[C]. Journal of Physics: Conference Series. IOP Publishing, 2022, 2258(1): 012001.

[13] 王帅,董育宁,李涛.基于 LSTM 和特征生成的网络流量分类[J].应用科学学报,2022,40(5):758-769.

[14] LIN X, XIONG G, GOU G, et al. Et-bert: A contextualized datagram representation with pre-training transformers for encrypted traffic classification [C]. Proceedings of the ACM Web Conference 2022, 2022: 633-642.

[15] 郭祥,姜文刚,王宇航.基于改进 Inception-ResNet 的加密流量分类方法[J].计算机应用,2023,43(8):2471-2476.

[16] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016:770-778.

[17] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]. Advances in Neural Information Processing Systems, 2017, 30.

[18] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16×16 words: Transformers for image recognition at scale[J]. ArXiv Preprint, 2020, ArXiv:2010.11929.

[19] LIU Z, LIN Y, CAO Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 10012-10022.

[20] DAI Y, GIESEKE F, OEHMCKE S, et al. Attentional feature fusion[C]. Proceedings of the IEEE/CVF Winter cConference on Applications of Computer Vision, 2021: 3560-3569.

[21] 陈瑞东,秦会斌.多特征融合与卡尔曼预测的车辆跟踪算法[J].电子测量技术,2023,46(7):32-38.

[22] KINGMA D P, BA J. Adam: A method for stochastic optimization [J]. ArXiv Preprint, 2014, ArXiv: 1412.6980.

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

作者简介

杨宇,硕士研究生,主要研究方向为网络流量分类。

E-mail:yy2285482325@163.com

唐东明(通信作者),博士,副教授,主要研究方向为网络通信系统,云计算及边缘计算。

E-mail:tangdongming@swust.edu.cnm

李驹光,博士,正高级工程师,主要研究方向为控制理论,计算机控制系统。

E-mail:lijuguang@swust.edu.cn

肖宇峰,博士,教授,主要研究方向为网络通信系统、智能机器人系统。

E-mail:xiaoyf_swit1@163.com