光电技术及应用

DOI: 10.16818/j.issn1001－5868.20250226003

基于自适应多通道注意力和知识蒸馏电力通信业务识别方法

夏炳森*,唐元春,周钊正,陈霖锋
(国网福建省电力有限公司经济技术研究院,福州 350012)

摘 要: 随着电力通信网络的迅速发展,高效准确的业务识别对于保障电力系统的稳定运行至关重要。因此,文章提出一种基于自适应多通道注意力和知识蒸馏的电力通信业务识别方法,旨在提高识别精度并优化计算资源的使用。首先,提出一种自适应多通道注意力网络,该网络融合了自适应多通道机制、注意力机制与交互注意力机制,通过多个并行特征通道有效捕捉输入数据中的多样化特征信息,并引入自适应权重机制,实现对不同通道贡献的动态加权调整。其次,设计了一种自适应知识传递蒸馏方法,该方法通过两个阶段逐步传递教师模型的知识,并随着迭代次数动态调整传递过程,使轻量级学生模型在显著降低计算资源消耗的同时,依然能够保持较高的识别精度。最后,在Moore数据集上的实验结果表明,所提方法在精确率、准确率、召回率和*F*1分数方面均优于对比方法,具有广泛的实际应用前景。

关键词: 电力通信网络; 业务识别; 注意力; 知识蒸馏
中图分类号: TN915.853 **文章编号:** 1001－5868(2026)01－0210－09

Power Communication Service Recognition Method Based on Adaptive Multichannel Attention and Knowledge Distillation

XIA Bingsen, TANG Yuanchun, ZHOU Zhaozheng, CHEN Linfeng
(**Econormic and Technological Research Institute, State Grid Fujian Electric Power Co., Ltd., Fuzhou 350012, CHN**)

Abstract: With the rapid development of power communication networks, efficient and accurate service recognition are crucial to ensure the stable operation of power systems. Therefore, this paper presents a power communication service recognition method based on adaptive multichannel attention and knowledge distillation to improve recognition accuracy and optimize the use of computational resources. First, an adaptive multichannel attention network (AMCAN) is introduced, which integrates adaptive multichannel mechanisms, attention mechanisms, and interactive attention mechanisms. This network effectively captures diverse feature information in the input data, through multiple parallel feature channels, and introduces an adaptive weight mechanism to dynamically adjust the contributions of different channels. Second, an adaptive knowledge transfer distillation method is designed that gradually transfers knowledge from the teacher model in two stages, dynamically adjusting the transfer process with each iteration. Using this approach, the lightweight student model can maintain a high recognition accuracy while significantly reducing computational resource consumption. Finally, experiments on the Moore dataset demonstrate that the

收稿日期:2025－02－26.
基金项目:2024年国网福建省电力有限公司科技项目(52130N240008).
* **通信作者:**夏炳森

proposed method outperforms other comparison methods in terms of precision, accuracy, recall, and $F1$ score, showing broad practical application prospects.

Key words: power communication network; service recognition; attention; knowledge distillation

0　引言

电力通信网络是专门为电力系统的运行、监控与管理而设计的通信基础设施，旨在实现电力设备、调度中心与用户之间的高效数据传输与信息交互。随着信息技术的迅猛发展，电力通信网络逐渐向基于IP架构的全数字化网络转型，承载了更多样化的业务应用，如负荷监控、设备管理、故障诊断等。这不仅提升了电力系统的智能化水平，也为实现智能电网和可再生能源的集成提供了坚实的通信支持[1]。根据《DL/T5391-2007》标准，电力通信业务按照业务流类型的不同可以划分为四种主要类型：数据业务、语音业务、视频业务和多媒体业务。

作为连接电力设备、调度中心与用户的重要枢纽，电力通信网络需要承载大量实时数据和业务应用，这些业务的稳定运行直接关系到电力系统的安全与效率。不同的电力业务流对网络资源有着不同的需求，例如电力调度需要极高的实时性，而用户用电数据则更关注带宽。如果这些业务流无法得到准确识别和分类，可能会导致资源分配不当，进而影响电网的稳定性和效率。因此，准确识别和分类网络中的不同业务流对于优化网络资源配置、提升服务质量和保障系统安全具有重要意义。

随着电力系统规模的不断扩大和复杂性的日益增加，电力通信网络在保障电网的安全、稳定和高效运行中扮演着至关重要的角色。然而，现有的业务识别与分类技术仍面临诸多挑战。一方面，电力通信网络中的业务流种类繁多，数据维度多样且复杂，传统的基于端口和DPI的方法难以有效处理高维、非线性数据的特征，导致识别精度不高[2]。另一方面，在实际应用场景中，电力通信网络中的流量呈现出动态变化的特征，传统方法无法自适应调整对数据特性的处理，导致资源分配不均和网络拥塞问题日益严重。此外，随着大数据、物联网等新技术的引入，电力通信网络需要处理的业务流量和数据量急剧增加，现有方法在计算资源消耗和处理效率上表现出明显的不足，难以满足现代电力系统对业务识别的高效性和鲁棒性要求[3]。传统方法在应对复杂数据特征、动态网络环境和高计算资源需求时表现出较大的局限性，尤其在面对电力通信网络中多样化的业务流和不断变化的网络状态时，现有技术无法提供足够的精细化和高效性支持。因此，开发一种能够自适应调整、资源高效利用且具备高识别精度的创新性识别方法，成为提升电力通信网络整体性能的关键。

针对上述问题，本文提出一种基于自适应多通道注意力和知识蒸馏的电力通信业务识别方法。本文的主要工作和创新点总结如下：

首先，提出了一种自适应多通道注意力网络AMCAN，该网络融合了自适应多通道机制、注意力机制与交互注意力机制，通过多个并行特征通道有效捕捉输入数据中的多样化特征信息，并引入自适应权重机制，实现对不同通道贡献的动态加权调整，从而提升电力通信业务识别的精度。

其次，设计了一种自适应知识传递蒸馏方法，该方法通过两个阶段逐步传递教师模型的知识，并随着迭代次数动态调整传递过程，使轻量级学生模型在显著降低计算资源消耗的同时，依然能够保持较高的识别精度。

最后，在Moore数据集上的实验结果表明，所提方法在精确率、准确率、召回率和$F1$分数方面均优于对比方法。

1　基于自适应多通道注意力的电力通信业务识别方法

本章将详细介绍所提出的自适应多通道注意力网络AMCAN，该网络融合了自适应多通道机制、注意力机制与交互注意力机制，用于电力通信业务的精准识别。AMCAN在多个并行特征通道中利用注意力机制与交互注意力机制来捕捉输入数据中不同的特征信息，并结合自适应权重机制，实现对各通道贡献的动态加权调整。通过对输入数据的特性进行灵活适应，AMCAN能够提升不同业务类型的识别准确率与鲁棒性。

1.1 构建单通道AMCAN电力通信业务识别模型

单通道的AMCAN架构如图1所示。在AMCAN的单个通道内，采用基于注意力机制和交互注意力机制的单通道网络，用于特征的提取和建模。每个单通道的网络结构包括嵌入层、注意力层和特征交互层和输出层。

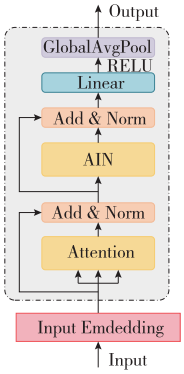


图1 单通道AMCAN电力通信业务识别模型

用于电力通信业务识别的输入特征$X\in R^{N\times d}$，其中N是样本数量，d是每个样本的特征维度。为了有效捕捉输入样本中不同特征之间的局部依赖关系，采用一维卷对输入样本进行特征编码。卷积核的大小为k，步长为s，通道数为C，对应的卷积层可以表示为：

$$X'=\mathrm{Conv1D}(X,W,b) \tag{1}$$

式中，$W\in R^{C\times k}$是卷积核的权重，$b\in R^{C}$是偏置，$X'\in R^{N\times d'\times C}$表示卷积后的输出特征。为了增加非线性，在卷积操作后应用激活函数ReLU，以进一步提取重要特征，减少特征维度，在卷积层后添加最大池化。具体表示为：

$$E=\mathrm{MaxPooling1D}\left[\mathrm{RELU}(X')\right] \tag{2}$$

式中，池化后的特征维度为$E\in R^{N\times d'\times C}$，其中$d'$是经过池化操作后的特征长度。

基于输入的嵌入特征E需经过线性映射后得到查询，键和值分量，如下所示：

$$\begin{aligned} Q&=EW_q+b_q\\ K&=EW_k+b_k\\ V&=EW_v+b_v \end{aligned} \tag{3}$$

式中，$Q=\{q_1,q_2,\cdots,q_m\}\in R^{N\times d'\times C}$表示查询分量，$K=\{k_1,k_2,\cdots,k_m\}\in R^{N\times d'\times C}$表示键分量，$V=\{v_1,v_2,\cdots,v_m\}\in R^{N\times d'\times C}$表示值分量，$q_m,k_m,v_m$分别指示三个矩阵的列向量。并且$W_q\in R^{C\times d'}$，$W_k\in R^{C\times d'}$，$W_v\in R^{C\times d'}$表示可学习的权重向量，$b_q$，$b_k$，$b_v$表示偏置项。对$Q$，$K$，$V$进行缩放点积注意力来计算这些局部信息，其公式如下：

$$O=\mathrm{Softmax}\left[\frac{Q(K)^{\mathrm{T}}}{\sqrt{d'}}\right]V \tag{4}$$

式中，$O\in R^{N\times d'\times C}$。然后将注意力输出$O$与原始输入$E$进行残差连接，并通过层归一化：

$$O'=\mathrm{LayerNorm}(O+E) \tag{5}$$

注意力层的输出O'通过AIN层（Attention Interaction Network）捕捉高阶特征交互，特征交互公式如下：

$$z_{ij}=O'[i]\odot O'[j],\forall i,j(i<j) \tag{6}$$

式中，$z_{ij}\in R^{d'}$是特征交互表示。交互特征的注意力权重α_{ij}表示为：

$$\alpha_{ij}=\mathrm{Softmax}\left[f(z_{ij})\right] \tag{7}$$

AIN层的输出表示为：

$$H=\sum_{i<j}\alpha_{ij}z_{ij} \tag{8}$$

式中，$H\in R^{N\times d'\times C}$。将输入特征$O'$与与AIN层输出$H$进行残差连接，并通过层归一化：

$$Output_{\mathrm{norm}}=\mathrm{LayerNorm}(H+O') \tag{9}$$

在AIN层输出$Output_{\mathrm{norm}}$后，通过线性层和RELU激活函数进行处理：

$$L=\mathrm{RELU}(W_{\mathrm{lin}}Output_{\mathrm{norm}}+b_{\mathrm{lin}}) \tag{10}$$

式中，$L\in R^{N\times d'\times C}$。使用全局池化来得到每个通道的全局特征表示：

$$R=\mathrm{GlobalAvgPool}(L) \tag{11}$$

式中，$R\in R^{N\times d'}$。

1.2 构建自适应多通道的AMCAN电力通信业务识别模型

自适应多通道的AMCAN如图2所示，将1.1节的单通道AMCAN扩展为M个通道。该模型通过构建自适应权重模块对M个通道的输出进行加权融合。在自适应多通道的AMCAN模型中，使用多层感知器（MLP）来计算自适应权重的过程可以通过以下步骤实现：首先从每个通道的输出中提取全局特征，接着MLP通过多层全连接网络对输入特征进行非线性映射，最终生成每个通道的自适应权重。

假设每个通道的输出为$R\in R^{N\times d'}$，其中N是样本数量，d'是全连接层的输出维度。针对每个通道的输出R，将所有通道的全局特征拼接成一个特征矩阵G：

$$G=[R_1;R_2;\cdots;R_m] \tag{12}$$

式中，$G\in R^{N\times Md'}$。然后通过一个具有两个隐藏层的 MLP 网络生成每个通道的权重。具体过程如下：

$$h=\mathrm{MLP}(G) \tag{13}$$

式中，$h\in R^{M}$。最终通过 Softmax 得到 M 个通道的自适应权重：

$$\alpha=\mathrm{Softmax}(h) \tag{14}$$

式中，$\alpha=\{\alpha_1,\alpha_2,\cdots,\alpha_m\}$ 表示每个通道的自适应权重。

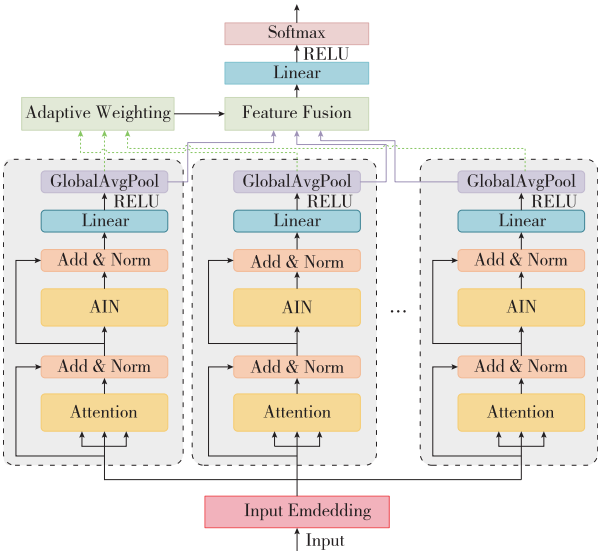


图 2　自适应多通道 AMCAN 电力通信业务识别模型

接下来，通过加权融合操作将所有通道的特征结合起来，并输入全连接层和 Softmax 层，最终生成分类结果。加权融合的目标是将各通道输出的特征按照自适应权重进行加权求和，生成一个统一的特征表示。具体计算过程如下：

$$F_{\text{fused}}=\sum_{m=1}^{M}\alpha_m\cdot R_m \tag{15}$$

式中，$F_{\text{fused}}\in R^{N\times d'}$ 表示融合后的特征矩阵，R_m 是第 m 个通道的特征表示，α_m 为该通道对应的自适应权重。加权融合后的特征 F_{fused} 经过全连接层进行进一步的特征提取。全连接层的计算公式如下：

$$F=\mathrm{RELU}\left(F_{\text{fused}}\cdot W_f+b_f\right) \tag{16}$$

式中，$W_f\in R^{d'\times d_{\text{out}}}$ 是全连接层的权重矩阵，$b_f\in R^{d_{\text{out}}}$ 是偏置项，d_{out} 表示输出特征的维度。最后，将经过全连接层的输出特征 F 输入到 Softmax 层，生成分类结果。Softmax 层用于多类分类任务，其公式如下：

$$\boldsymbol{Output}=\mathrm{Softmax}(F) \tag{17}$$

式中，$\boldsymbol{Output}\in R^{d_{\text{out}}}$ 表示各个类别的概率。

为了训练模型，采用基于交叉熵的损失函数来衡量模型预测结果与真实标签之间的差异。同时，模型中的所有可学习参数通过反向传播算法进行更新。损失函数采用交叉熵损失（Cross-Entropy Loss），其定义如下：

$$L=-\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C}y_{i,j}\log\left(\hat{y}_{i,j}\right) \tag{18}$$

式中，$y_{i,j}$ 是用于电力通信业务识别的第 i 个样本在第 j 类的真实标签，$\hat{y}_{i,j}$ 是模型预测第 i 个样本为第 j 类的概率，C 表示样本的维度。

2　自适应知识传递蒸馏框架

知识蒸馏是一种压缩深度学习模型的技术，它可以抽取“教师”模型的知识，然后将这些知识迁移到更小、更轻的学生模型中[22]。在传统的知识蒸馏方法中，教师模型的预测输出被作为软标签（soft labels），学生模型通过最小化与这些软标签之间的损失来学习教师模型的知识。电力通信业务识别需要极高的实时性和准确性，现有的模型由于网络结构复杂，导致训练和推理速度较慢，无法满足快速业务识别的要求。利用知识蒸馏可以压缩模型，提高业务识别的速度和精度。然而，随着任务复杂性的提升和深度学习网络结构的扩展，简单的知识迁移方法无法充分利用教师模型中的丰富信息。为了解决充分利用教师模型中的丰富信息，本文提出了一种自适应知识传递蒸馏框架，用于将电力通信业务识别模型压缩为轻量级模型以提高业务识别的速度和准确率。如图 3 所示，该框架包含两个阶段分别是预训练阶段和自适应知识传递阶段。

AMCAN 教师模型指导训练学生模型分为两个步骤，首先计算蒸馏损失，分别将 AMCAN 教师网络的输出 $Y^{(t)}$ 和 AMCAN 学生网络的输出 $Y^{(s)}$ 经过一个蒸馏温度为 T 的蒸馏器，具体表示为：

$$Z^{(t)}=\mathrm{Softmax}\left(\frac{Y^{(t)}}{T}\right) \tag{19}$$

$$Z^{(s)}=\mathrm{Softmax}\left(\frac{Y^{(s)}}{T}\right) \tag{20}$$

蒸馏损失定义为 $Z^{(t)}$ 和 $Z^{(s)}$ 的 KL 散度，具体表示为：

$$L_{\text{KD}}=\frac{T^2}{N}\sum_{i=1}^{N}\sum_{j=1}^{C}Z_{i,j}^{(s)}\log\frac{Z_{i,j}^{(s)}}{Z_{i,j}^{(t)}} \tag{21}$$

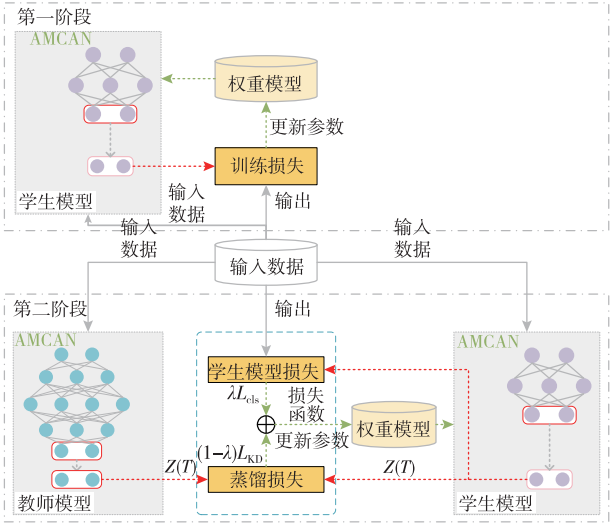


图3 自适应知识传递蒸馏框架

其次是计算AMCAN学生模型相对于原始标签的损失，采用交叉熵损失函数来表示AMCAN学生模型的损失，具体表示为：

$$L_{cls}=-\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C}Z_{i,j}^{(s)}\log\left(\hat{Z}_{i,j}^{(s)}\right) \tag{22}$$

式中，$Z_{i,j}^{(s)}$表示样本的真实标签，$\hat{Z}_{i,j}^{(s)}$表示样本的预测标签。

自适应知识传递算法通过权重参数λ来调节对两个损失函数的不同关注。具体来说，自适应算法对模型输出层的损失进行操作，通过损失层的权重来控制和切换阶段，λ代表当前对相应任务的关注程度，损失项的权重对学习过程有很大的影响，损失值大的任务比损失值小的任务更新得更快。在第一阶段，λ被设置为1，只进行对原始数据的学习任务，因为其他阶段任务的权重为0，停止了其损失的反向传播。其损失函数为公式(22)。第一阶段的学生模型参数更新表示为：

$$\Delta\boldsymbol{w}_1=\frac{\partial L_{cls}}{\partial w_1} \tag{23}$$

在第二阶段，根据训练伦次的个数，λ逐渐减小，直到0，目的是同时学习原始数据和预测数据，逐步将教师知识转移到学生模型。λ表示为：

$$\lambda=1-\left(\frac{B-B_1}{B_{\max}-B_1-B_2}\right) \tag{24}$$

式中，B_1和B_2分别是阶段一和阶段二中的训练轮次的数目，B是当前训练轮次。而$B_{\max}$是最大纪元的数目。一旦前一阶段的任务训练好，对后面任务的关注就会迅速增加。第二阶段的AMCAN学生模型的损失函数表示为：

$$L=\lambda\cdot L_{cls}+(1-\lambda)\cdot L_{KD} \tag{25}$$

第二阶段的AMCAN学生模型参数更新表示为：

$$\Delta\boldsymbol{w}_2=\frac{\partial L}{\partial w_2}=\lambda\cdot\frac{\partial L_{cls}}{\partial w_2}+(1-\lambda)\cdot\frac{\partial L_{KD}}{\partial w_2} \tag{26}$$

3 仿真与性能分析

实验使用Python以及pytorch2.0.1在配备了8块RTX3090GPU的Linux服务器上仿真，最小批量大小为32，使用Pytorch中内置的Adam作为优化器。选用公开的Moore数据集对所提算法进行验证，该数据集使用249个特征来表示一个流，包括流的持续时间，数据包的数量、大小、相邻数据包到达的平均间隔时间等参数。该数据集较为详尽地描述了一个流的绝大部分特性，被研究界所广泛使用。本文选用Moore数据集的前三个数据包的数据进行网络训练和测试，同时考虑到前三个数据包中所包含的ATTACK类型数量过少无法满足实验需求，因此还合并了剩余7个数据包中的ATTACK。表1是所用样本数据的具体类型以及数据的分布情况。

表1 实验所用数据样本的数量

类型	数目
WWW	54 835
MAIL	8 320
FTP-CONTROL	2 110
FTP-PASV	512
P2P	533
DATABASE	773
FTP-DATA	3 326
MULTIMEDIA	373
SERVICES	626
ATTACK	1 793
总计	73 201

3.1 对比模型性能分析

实验选取CNN，2D-CNN[23]，CNN-LSTM[24]，SGC-LSTM[25]作为对比模型。为了评估模型的分类性能，使用最为常见的准确率、精确率、召回率和$F1$分数作为模型评价指标，准确率的定义如下：

$$Accuracy=\frac{\sum_{i=1}^{N}TP_i}{\sum_{i=1}^{N}TP_i+FP_i+FN_i} \tag{27}$$

针对每个类别i，精确率、召回率和$F1$分数的计算公式为：

$$Precision_i=\frac{TP_i}{TP_i+FP_i} \tag{28}$$

$$Recall_i=\frac{TP_i}{TP_i+FN_i} \tag{29}$$

$$F1_i=2\times\frac{Precision_i\times Recall_i}{Precision_i+Recall_i} \tag{30}$$

式中，TP_i是分类为类别i且实际属于类别i的样本数，FP_i是分类为类别i且实际不属于类别i的样本数，FN_i是实际属于类别i但未被正确分类的样本数。多分类的整体精确率、召回率和$F1$分数是各类别的加权平均，具体计算如下：

$$Precision=\sum_{i=1}^{N}\boldsymbol{w}_i\times Precision_i \tag{31}$$

$$Recall=\sum_{i=1}^{N}\boldsymbol{w}_i\times Recall_i \tag{32}$$

$$F1=\sum_{i=1}^{N}\boldsymbol{w}_i\times F1_i \tag{33}$$

式中，$\boldsymbol{w}_i$是类别i的样本占比。

各个模型的分类性能指标如表2所示。其中CNN模型的性能相对较低，其精确率为0.869，召回率为0.721，$F1$分数为0.788，该模型在分类任务中的表现存在一定的局限性，尤其在召回率方面，容易漏检部分正类样本。相比之下，2D-CNN模型在所有指标上均有显著提升，精确率为0.906，召回率为0.776，$F1$分数为0.836，这表明2D-CNN在捕捉空间特征方面更加有效。进一步结合时序特征的CNN-LSTM模型表现出更好的综合性能，其精确率为0.928，召回率为0.863，$F1$分数为0.894。SGC-LSTM模型在此基础上进一步提升，精确率达0.962，召回率为0.927，$F1$分数为0.944，表明其在复杂任务中具有更高的识别能力。而所提出的AMCAN模型以最高的精确率0.981、召回率0.961和$F1$分数0.971取得了最佳表现，表明其在处理多维特征和时序数据方面具有极高的准确性和鲁棒性。

表2　在Moore数据集上对比模型的性能指标

模型	*Accuracy*	*Precision*	*Recall*	*F1*
CNN	0.720	0.869	0.721	0.788
2D-CNN	0.787	0.906	0.776	0.836
CNN-LSTM	0.854	0.928	0.863	0.894
SGC-LSTM	0.937	0.962	0.927	0.944
AMCAN	**0.973**	**0.981**	**0.961**	**0.971**

表3展示了在Moore数据集上不同算法对各类的精确率表现。可以看出，传统的CNN模型在大多数类别上的表现较为一般，精确率在0.367～0.768之间。2D-CNN和CNN-LSTM相较于CNN有一定的提升，特别是CNN-LSTM在各类中均表现稳定，精确率在0.831～0.874之间。SGC-LSTM进一步提高了精确率，几乎所有类别的精确率均超过0.900。然而，本文提出的AMACN模型在所有类别上的表现均超过了现有的对比模型，精确率在0.944～0.990之间，其中在P2P类上达到了最高的0.990，整体表现优于SGC-LSTM和其他对比模型。这表明AMACN模型能够更好地捕捉输入特征中的关键信息，显著提升了电力通信业务的分类精度。

表3　在Moore数据集上对比模型中各类的精确率

模型	WWW	MAIL	F-C	F-P	P2P	DB	F-D	MULT	SERV	ATT
CNN	0.767	0.768	0.400	0.509	0.457	0.509	0.412	0.506	0.528	0.367
2D-CNN	0.788	0.784	0.791	0.774	0.747	0.741	0.798	0.800	0.800	0.802
CNN-LSTM	0.853	0.856	0.860	0.852	0.831	0.838	0.867	0.853	0.864	0.874
SGC-LSTM	0.936	0.945	0.938	0.950	0.953	0.961	0.939	0.906	0.920	0.947
AMCAN	0.971	0.978	0.973	0.970	0.990	0.987	0.980	0.960	0.944	0.972

3.2　混淆矩阵分析

AMACN网络在Moore数据集上的混淆矩阵如图4所示，从图中可以观察到所提出的AMACN模型在Moore数据集上的分类表现显著优于其他方法。

首先，在WWW类中，模型展现了极高的精确率，绝大多数样本（10 659个）被准确分类，误分类数量极少，显示了模型在处理这一类通信业务时的优越性。MAIL类的分类效果同样也较为出色，共有1 628个样本被正确分类，但部分样本被误分类为FTP-CONTROL和P2P类，说明在这几类较为相似的数据特征上，模型存在一定的混淆现象。对于P2P和SERVICES类，模型在大多数样本中都能够做出正确分类，其中P2P类106个样本中绝大部分分类正确，SERVICES类118个样本中也表现良好。然而，ATTACK类的分类表现虽然总体较好，但少量样本被误分类为DATABASE和WWW类，这可能是由于这几类特征的某些重叠。此外，

DATABASE, FTP-CONTROL, FTP-DATA 和 FTP-PASV 类的误分类较为明显,表明在这些类别之间,特征区分度较低,模型难以完全分辨。MULTIMEDIA 类则表现相对较差,部分样本被误分类为 WWW 和 MAIL,反映了模型在处理多媒体类业务时的分类难度。

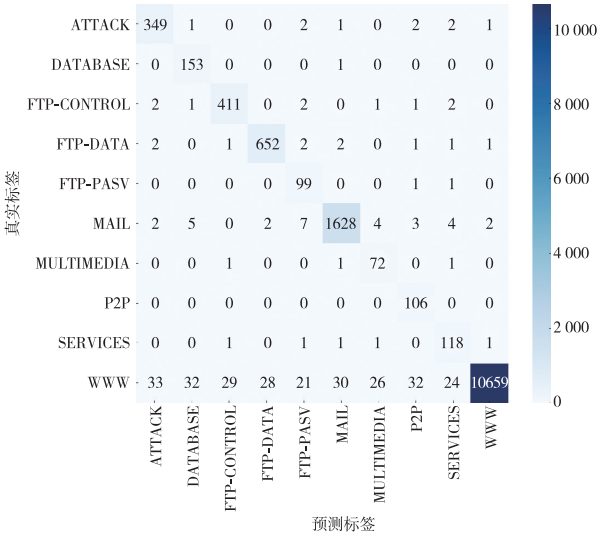
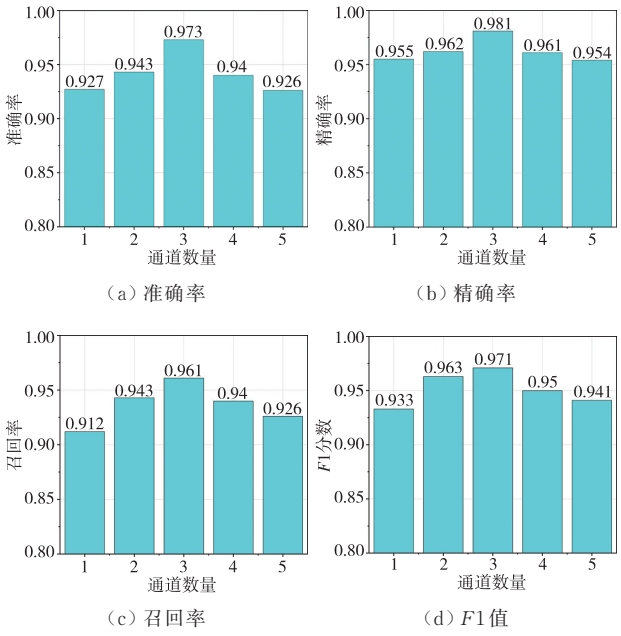


图 4 混淆矩阵分析

3.3 不同通道数量对模型性能的影响

图 5 中展示了不同通道数量对 AMCAN 网络在准确率、精确率、召回率和 $F1$ 值方面的影响。



(a) 准确率 (b) 精确率

(c) 召回率 (d) $F1$ 值

图 5 不同通道数量对 AMCAN 模型性能的影响

可以看出,当通道数量为 3 时,模型的准确率为 0.973,$F1$ 分数为 0.971,模型的表现相对较优,四项指标均达到了较高的值。随着通道数量的增加,性能趋于稳定,但在某些情况下,通道数量过多可能不会显著提升模型的表现。然而,当通道数量过多时,模型可能会捕捉到数据中的噪声,导致过拟合,从而降低泛化能力。因此,当模型在通道数量较为适中时,例如 3 个通道,可以在特征提取和模型复杂性之间实现最佳平衡,从而达到较好的性能,而通道数量过多反而可能导致性能趋于饱和甚至下降。

3.4 自适应知识蒸馏对 AMCAN 的影响

对 AMCAN 教师模型和 AMCAN 学生模型在 Moore 数据集上的检测性能进行了对比,包括分类性能和推理性能。图 6 展示了学生模型和教师模型的分类性能,评估指标包括准确率、精确率、召回率和 $F1$ 分数。结果显示,两者都表现出较高的性能,但学生模型要略优于教师模型。这表明,在知识蒸馏的过程中,学生模型不仅有效继承了教师模型的优点,还在某些方面实现了性能的提升,展示了自适应知识蒸馏对模型性能优化的有效性。

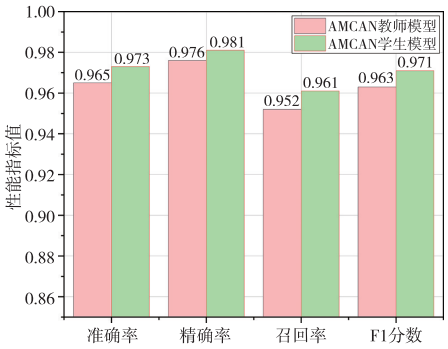
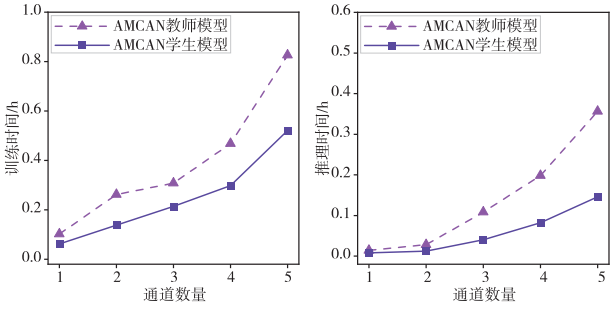


图 6 AMCAN 教师模型与 AMCAN 学生模型的检测性能比较

图 7 展示了学生模型和教师模型在不同通道数量下的训练时间和推理时间对比。



(a) 训练时间 (b) 推理时间

图 7 AMCAN 教师模型与 AMCAN 学生模型的性能比较

可以观察到,随着通道数量的增加,教师模型和学生模型的训练和推理时间都呈现一定的增长,

但是学生模型的训练时间和推理时间显著低于教师模型。无论是训练还是推理，学生模型在所有通道数量下均表现出更快的速度，尤其是在通道数量较多的情况下，学生模型的优势更为明显，这表明自适应知识蒸馏对模型压缩的有效性。结合前面的检测性能分析，说明学生模型不仅在模型压缩和加速方面取得了成功，而且在性能上也能与教师模型相媲美。整体而言，学生模型通过自适应知识蒸馏有效继承了教师模型的知识，在保证检测性能的同时显著提高了效率，非常适用于时效性要求较高的应用场景。

4　结论

本文提出了一种基于自适应多通道注意力和知识蒸馏的电力通信业务识别方法，该方法通过自适应多通道注意力网络AMCAN，有效捕捉多维特征信息，并动态调整不同通道的权重，从而提升了模型在处理复杂时序和空间数据时的表现。所提出的自适应知识传递蒸馏方法通过将教师模型的知识逐步传递给学生模型，实现了在减少计算资源消耗的同时保持较高的识别精度。在Moore数据集上的实验结果表明，学生模型在准确率、精确率、召回率和$F1$分数等指标上均优于教师模型，展现了其在知识蒸馏过程中实现性能优化的潜力。同时，学生模型在训练和推理时间上显著优于教师模型，进一步证明了该方法在模型压缩与加速方面的有效性。因此，所提方法不仅能在复杂场景中保证高精度的业务识别，还能显著提升计算效率，非常适合应用于时效性要求较高的电力通信网络业务识别任务。

参考文献：

[1] Gungor V C, Sahin D, Kocak T, et al. Smart grid technologies: communication technologies and standards[J]. IEEE Transactions on Industrial Informatics, 2011, 7(4): 529-539.

[2] Azab A, Khasawneh M, Alrabaee S, et al. Network traffic classification: techniques, datasets, and challenges[J]. Digital Communications and Networks, 2024, 10(3): 676-692.

[3] Shi J, Liu M, Hou C, et al. Online encrypted mobile application traffic classification at the early stage: challenges, evaluation criteria, comparison methods[C]//2021 IEEE 6th International Conference on Computer and Communication Systems (ICCCS), 2021: 1128-1135.

[4] Garcia J, Korhonen T, Andersson R, et al. Towards video flow classification at a million encrypted flows per second[C]//2018 IEEE 32nd International Conference on Advanced Information Networking and Applications (AINA), 2018: 358-365.

[5] Moore A W, Zuev D. Internet traffic classification using Bayesian analysis techniques[C]//Proceedings of the 2005 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems, 2005: 50-60.

[6] Saber A, Fergani B, Abbas M. Encrypted traffic classification: combining over-and under-sampling through a PCA-SVM[C]//2018 3rd International Conference on Pattern Analysis and Intelligent Systems (PAIS), 2019: 1-5.

[7] Li R, Xiao X, Ni S, et al. Byte segment neural network for network traffic classification[C]//2018 IEEE/ACM 26th International Symposium on Quality of Service (IWQoS), 2019: 1-10.

[8] Zhang Y, Chen X, Jin L, et al. Network intrusion detection: based on deep hierarchical network and original flow data[J]. IEEE Access, 2019, 7: 37004-37016.

[9] Oeung P, Shen F. Imbalanced Internet traffic classification using ensemble framework[C]//2019 International Conference on Information Networking (ICOIN), 2019: 37-42.

[10] Ke W, Wang Y, Lei X, et al. Spark-based feature selection algorithm of network traffic classification[C]//2017 13th International Conference on Computational Intelligence and Security (CIS), 2018: 140-144.

[11] Zhang F, Wang Y, Ye M. Network traffic classification method based on improved capsule neural network[C]//2018 14th International Conference on Computational Intelligence and Security (CIS), 2018: 174-178.

[12] Wang Y, An J, Huang W. Using CNN-based representation learning method for malicious traffic identification[C]//2018 IEEE/ACIS 17th International Conference on Computer and Information Science (ICIS), 2018: 400-404.

[13] 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), 2017: 712-717.

[14] Zhang Y, Chen X, Guo D, et al. PCCN: parallel cross convolutional neural network for abnormal network traffic flows detection in multi-class imbalanced network traffic flows[J]. IEEE Access, 2019, 7: 119904-119916.

[15] Li D, Zhu Y, Lin W. Traffic identification of mobile apps based on variational autoencoder network[C]//2017 13th International Conference on Computational Intelligence and Security (CIS), 2018: 287-291.

[16] Yim J, Joo D, Bae J, et al. A gift from knowledge distillation: fast optimization, network minimization and transfer learning[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017: 7130-7138.

[17] Zhang L, Shi Y, Shi Z, et al. Task-Oriented Feature Distillation [C]//NIPS'20: Proceedings of the 34th International Conference on Neural Information Processing Systems, 2020: 14759 - 14771.

[18] Ji M, Heo B, Park S. Show, attend and distill: knowledge distillation *via* attention-based feature matching [J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2021, 35(9): 7945-7952.

[19] Chen P, Liu S, Zhao H, et al. Distilling knowledge *via* knowledge review [C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021: 5006-5015.

[20] Li T, Li J, Liu Z, et al. Few sample knowledge distillation for efficient network compression [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020: 14627-14635.

[21] Mitsuno K, Nomura Y, Kurita T. Channel planting for deep neural networks using knowledge distillation [C]//2020 25th International Conference on Pattern Recognition (ICPR), 2021: 7573-7579.

[22] Hong H, Kim H. Feature distribution-based knowledge distillation for deep neural networks [C]//2022 19th International SoC Design Conference (ISOCC), 2023: 75-76.

[23] Zhou Y, Shi H, Zhao Y, et al. Encrypted network traffic identification based on 2D-CNN model[C]//2021 22nd Asia-Pacific Network Operations and Management Symposium (APNOMS), 2021: 238-241.

[24] Sun P, Liu P, Li Q, et al. DL-IDS: extracting features using CNN-LSTM hybrid network for intrusion detection system [J]. Security and Communication Networks, 2020, 2020: 8890306.

[25] Pan Y, Zhang X, Jiang H, et al. A network traffic classification method based on graph convolution and LSTM [J]. IEEE Access, 2021, 9: 158261-158272.

作者简介：

夏炳森(1981—),男,电力工程高级工程师,硕士研究生,主要从事电力通信规划设计及电力物联网建设研究。