

基于改进 Res2Net 与自适应多尺度窗口池化的 调制识别方法

王丹*, 李万杰, 江丰杨

(重庆邮电大学通信与信息工程学院, 重庆 400065)

摘要: 随着现代通信技术的快速发展,自动调制识别(Automatic Modulation Recognition, AMR)在频谱资源管理方面的重要性日益凸显,基于深度学习的AMR方法凭借其优异的性能成为当前研究热点。针对现有方法在复杂信道条件下多尺度特征融合能力不足、特征 token 化方式有效性与复杂度难以平衡的问题,提出一种基于改进 Res2Net 与自适应多尺度窗口池化的调制识别方法 Res2-AMWP。特征提取阶段利用改进的 Res2Net 对特征按通道分组并逐级融合,同时引入挤压与激励(Squeeze-and-Excitation, SE)注意力机制对通道进行自适应重标定。特征融合阶段提出自适应多尺度窗口池化(Adaptive Multi-scale Window Pooling, AMWP)模块将多尺度特征转化为更具判别性的 token 表征,并利用双向长短期记忆网络(Bidirectional Long Short-Term Memory, BiLSTM)捕获 token 间的上下文依赖。注意力分类头采用注意力池化机制进一步突出关键的 token 表征,由全连接层得到最终的识别结果。在公开数据集 RadioML2016.10a、RadioML2016.10b、RML22 上的实验结果表明,Res2-AMWP 的整体识别准确率分别达到 63.51%、65.36%、70.30%,相较于多种对比方法分别提高了 1.01%~7.33%、0.32%~6.5%、0.75%~8.40%,且模型的复杂度保持在较低水平,实现了精度与复杂度的较好平衡。

关键词: 自动调制识别;多尺度特征融合;特征 token 化;Res2Net;注意力机制;自适应多尺度窗口池化

基金项目: 重庆市自然科学基金创新发展联合基金(中国星网)资助项目(No.CSTB2023NSCQ-LZX0114)

中图分类号: TN911.3

文献标识码: A

文章编号: 0372-2112(2026)02-0562-16

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.12263/DZXB.20251229

Modulation Recognition Method Based on Improved Res2Net and Adaptive Multi-Scale Window Pooling

WANG Dan*, LI Wanjie, JIANG Fengyang

(School of Communications and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China)

Abstract: With the rapid development of modern communication technology, automatic modulation recognition (AMR) has become increasingly important in spectrum resource management, and deep learning-based AMR methods have become a current research hotspot due to their superior performance. To address the problems of insufficient multi-scale feature fusion capability and the difficulty in balancing the effectiveness and complexity of feature tokenization under complex channel conditions in existing methods, this thesis proposed a modulation recognition method termed Res2-AMWP based on an improved Res2Net and adaptive multi-scale window pooling. In the feature extraction stage, the improved Res2Net was adopted to group features by channel and fuse them progressively, while the squeeze-and-excitation (SE) attention mechanism was introduced to perform adaptive channel re-calibration. In the feature fusion stage, an adaptive multi-scale window pooling (AMWP) module was proposed to transform multi-scale features into more discriminative token representations, and a bidirectional long short-term memory network (BiLSTM) was employed to capture contextual dependencies among tokens. The attention-based classification head further highlighted key token representations through an attention pooling mechanism, and the final recognition results were obtained by fully connected layers. Experimental results on the public datasets RadioML2016.10a, RadioML2016.10b, and RML22 demonstrated that Res2-AMWP achieved overall recognition accuracies of 63.51%, 65.36%, and 70.30%, respectively, outperforming multiple baseline methods by 1.01%~7.33%, 0.32%~6.5%, and 0.75%~8.40% on the three datasets. Moreover, the model complexity remained at a relatively low level, achieving a good balance between accuracy and complexity.

Keywords: automatic modulation recognition; multi-scale feature fusion; feature tokenization; Res2Net; attention mechanism; adaptive multi-scale window pooling

Foundation Item(s): Innovation and Development Joint Fund of Chongqing Natural Science Foundation (China Satellite Network) (No.CSTB2023NSCQ-LZX0114)

0 引言

随着第五代移动通信(5th Generation mobile communication, 5G)、物联网(Internet of Things, IoT)、卫星互联网等现代通信技术的快速发展,电磁环境变得日益复杂,无线电信号与调制方式显著增加,当前有限的无线通信资源管理和利用面临巨大的挑战^[1-2]。在缺少先验知识的情况下,自动调制识别(Automatic Modulation Recognition, AMR)可根据信道状况和数据需求自适应选择最佳的调制方式,实现对非合作场景下信号调制类型的自动判别,为频谱资源的合理分配与高效使用提供了关键信息^[3]。目前,AMR已经在认知无线电、频谱检测、电子侦察等任务中得到了广泛应用^[4-5]。

传统的 AMR 方法包括基于似然比(Likelihood-Based, LB)的 AMR(LB-AMR)方法^[6]和基于信号特征提取(Feature-Based, FB)的 AMR(FB-AMR)方法^[7]。其中, LB-AMR 方法的基本思想是将调制识别建模为一个多重假设检验问题,在不同调制假设下建立接收信号的统计模型并构造似然函数,通过最大似然判决或似然比检验等方法实现识别。常见形式主要包括广义似然比检验(Generalized-Likelihood Ratio Test, GLRT)、混合似然比检验(Hybrid-Likelihood Ratio Test, HLRT)、平均似然比检验(Average-Likelihood Ratio Test, ALRT)^[8]。然而实际场景中的信道以及噪声参数往往未知且难以精确估计,导致似然函数的计算复杂度较高,工程部署难度较大。相比之下, FB-AMR 通常不依赖显式的信道与噪声信息,其计算开销更低,对于环境的适应性较强。该方法通常由特征提取与分类判决两部分组成。首先从接收信号中直接提取信号的差异性判别特征,如瞬时特征、频谱特征、变换域特征等;随后将提取的特征输入分类器完成判决,常见分类器包括支持向量机(Support Vector Machine, SVM)、 K 近邻算法(K -Nearest Neighbor, KNN)、随机森林(Random Forest, RF)等。尽管 FB-AMR 在运算复杂度与工程部署方面具有一定优势,但其分类器性能依赖特征设计,且提取的特征容易受到噪声影响,在复杂信道条件下的识别性能受限。

近年来,深度学习(Deep Learning, DL)凭借其强大的特征表示与端到端建模能力,在自然语言处理、计算机视觉、信号处理等多个领域取得了显著进展^[9]。受此影响,一系列学者将 DL 引入 AMR 领域,与传统方法相比,基于 DL 的 AMR(DL-AMR)方法对

于信号特征的学习能力更强,具有更好的识别性能^[10]。其中,卷积神经网络(Convolutional Neural Network, CNN)对于特征的局部特征提取能力较强, O'Shea 等人^[11]首次将 CNN 引入 AMR 中,直接利用时域信号的同相正交(In-phase/Quadrature, I/Q)分量对模型进行训练,实验结果证实,该方法的性能优于基于特征提取的传统方法。与 CNN 相比,循环神经网络(Recurrent Neural Network, RNN)引入了时序建模,能够捕获序列的局部相关性, Hong 等人^[12]构建了基于 RNN 的 AMR 方法,在公共数据集 RadioML2016.10a^[13]的识别结果显示其在 -4 dB 信噪比(Signal-to-Noise Ratio, SNR)下的识别性能优于 CNN。随后,研究者将 RNN 的改进形式——长短期记忆网络(Long Short-Term Memory, LSTM)^[14]与门控循环单元(Gated Recurrent Unit, GRU)^[15]引入 AMR,以建模特征的时序依赖,提升识别性能。

为进一步提升 DL-AMR 方法对特征的融合能力与上下文关联性,现有研究从特征提取方式与特征表示形式两方面进行改进。前者聚焦于多尺度表征学习,通过多分支卷积、跨尺度特征融合等方法同时捕获不同尺度的判别信息。Zhang 等人^[16]提出多尺度卷积网络 MSNet,通过在同一模块内并行设置不同大小的一维卷积核并进行特征拼接融合,以捕获不同尺度的判别信息,在 RadioML2016.10a 的 8 种数字调制类型的实验中,模型在 -6~14 dB SNR 下平均识别准确率达到 87.66%,优于 ResNet、LSTM 等对比模型。Tan 等人^[17]通过在多分支卷积结构中引入不同膨胀率的膨胀卷积,实现多尺度特征提取与融合,基于 RML2016.10a 消融实验结果表明,在 SNR>0 dB 条件下,仅引入多尺度特征提取模块即可使模型在三种小样本设置下相较基线提升 10%。另一类工作借鉴 Transformer 的 token 化表征思想,将原始信号或特征映射为多个 token,以便更有效地建模全局依赖关系,其实现方式主要分为卷积映射与线性映射两类:前者通过卷积层将特征直接映射为 token 序列,后者则将输入按固定窗口划分为多个 patch 并经线性层投影,对每个 Patch 进行展平和投影得到 token 序列。Dao 等人^[18]采用多尺度卷积核并行提取特征,并将各分支输出在通道维拼接融合后展平为 token 序列,作为 Transformer 的输入,其文献提出的 VT-MCNet 在 RadioML2018.01a^[19]下的整体识别准确率达到 64.8%。Cai 等人^[20]将 I/Q 序列按固定窗口划分为多个 patch,

再将每个 patch 展平后通过线性投影生成 token 序列, 最终输入 Transformer, 三个公开数据集的实验结果表明, 该模型在性能上优于 CNN、LSTM 与其他机器学习模型, 同时参数量更少。

尽管上述方法在公开数据集上已经取得了较好的性能, 但在复杂信道条件下仍存在局限性。一方面, 现有方法的多尺度信息交互多停留在通道拼接或线性融合等简单形式, 难以充分建模尺度间关联, 导致融合后的特征判别性提升受限。另一方面, 现有部分方法以固定尺度或步长生成 token 序列, 难以同时兼顾不同时间尺度下的关联信息。同时, 为进一步实现 token 的特征嵌入与多尺度建模, 往往需要额外的线性投影与卷积模块, 使得 token 化过程在表征有效性与复杂度之间难以取得平衡。

针对上述问题, 本文提出了一种基于改进 Res2Net 与自适应多尺度窗口池化的 AMR 方法 Res2-AMWP, 该方法在保证较低复杂度的同时, 显著提高了不同 SNR 下的识别准确率。本文的主要贡献可以总结如下:

(1) 构建了一种端到端的基于改进 Res2Net 与自适应多尺度窗口池化的网络模型 Res2-AMWP, 该模型主要包含改进的 Res2Net 多尺度特征提取模块、自适应多尺度窗口池化 (Adaptive Multi-Scale Window Pooling, AMWP) 模块、双向长短期记忆网络 (Bidirectional Long Short-Term Memory, BiLSTM) 以及注意力池化模块。

(2) 对 Res2Net^[21] 进行了改进, 改进 Res2Net 利用本文设计的卷积模块替换了标准卷积结构, 同时在

Res2Net 的基础上引入了挤压与激励 (Squeeze-and-Excitation, SE) 注意力机制^[22], Res2Net 利用通道分组级联卷积实现了多尺度特征的联合提取, 而 SE 注意力机制则对提取的特征进行通道重标定, 该设计在不显著提高参数量前提下提高了特征的表达能力。

(3) 提出 AMWP 模块, 该模块首先在特征图上采用不同尺寸的平均池化窗口生成多尺度 token, 随后为各尺度引入可学习的尺度权重, 最后对不同窗口尺度的 token 进行自适应拼接。AMWP 模块在较低复杂度的前提下增强了模型对跨尺度信息的利用能力。

(4) 在 AMR 领域的公开数据集 RadioML2016.10a、RadioML2016.10b、RML22 上进行了识别结果分析、对比实验、消融实验, 证明了本文提出的方法相比其他 AMR 方法具有更好的识别性能与泛化能力, 且各个模块的设计具有合理性。

1 系统模型

图 1 展示了单输入单输出的 DL-AMR 系统模型, 假设接收端以满足奈奎斯特速率的符号率对脉冲进行采样, 接收信号可以表示为

$$r(n) = h_n s_m(n) e^{j(\varphi_0 + 2\pi f_d n)} + w(n) \quad (1)$$

式中: $n = 0, 1, \dots, N-1$ 为接收端在不同采样时刻的采样点索引; $h_n \in (0, 1]$ 表示接收端在第 n 个采样时刻的等效信道增益, 用于描述由多径衰落、路径损耗等因素引起的时变振幅响应; $s_m(n)$ 表示发送信号在第 m 种调制方式下的第 n 个传输符号; j 为虚数单位, $j^2 = -1$; φ_0 为相位偏移; f_d 为归一化载波频率偏移; N 为符号总数; $w(n)$ 为加性高斯白噪声。

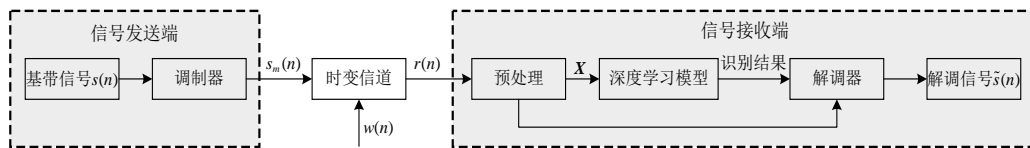


图 1 DL-AMR 系统模型

Figure 1 DL-AMR system model

为了便于深度学习模型的处理, 实际通信系统通常将输入信号表示为其同相分量 I_n 与正交分量 Q_n 相加的形式, 即 $r(n) = I_n + jQ_n$, 并将其实部与虚部分别作为两个独立特征通道, 按时间顺序构造成二维输入矩阵:

$$\mathbf{X} = \begin{bmatrix} I_0 & I_1 & \dots & I_{N-1} \\ Q_0 & Q_1 & \dots & Q_{N-1} \end{bmatrix} \quad (2)$$

DL-AMR 本质上可以看成是一个分类问题, 给定训练集 $\{(\mathbf{X}_i, Y_i)\}_{i=1}^{N_s}$, 其中 \mathbf{X}_i 表示第 i 个输入样本, Y_i 为其

真实类别标签, 对于每个输入样本, DL 模型通过多层卷积、时序建模等方法对其进行特征提取, 在此期间模型不断更新网络参数 ω , 使网络提取到的高层次判别特征能够最大化区分不同调制方式。设调制信号的类别总数为 L , 则模型对样本 \mathbf{X}_i 预测的概率向量可以表示为

$$\mathbf{p}_i = [p_{i,1}, p_{i,2}, \dots, p_{i,L}]^T \quad (3)$$

式中: $[\cdot]^T$ 为矩阵的转置操作; $p_{i,l} = P_\omega(Y=l|\mathbf{X}_i)$ 表示模型在参数 ω 下预测样本为第 l 种调制类型的概率。

DL-AMR 方法通过不断优化模型参数, 以最小化模

型的损失函数为目标,从而提升模型对信号调制类型的识别准确率。令 Y_i 的独热编码向量 $\mathbf{y}_i = [y_{i,1}, y_{i,2}, \dots, y_{i,L}]^T$, 其中,当 X_i 的真实类别为第 l 种调制类型即 $Y_i = l$ 时, $y_{i,l} = 1$, 否则 $y_{i,l} = 0$ 。当模型采用分类交叉熵作为损失函数时,其优化问题可以表示为

$$\omega^* = \arg \min_{\omega} \mathcal{L}_{CE}(\omega) = \arg \min_{\omega} \left(-\frac{1}{N_S} \sum_{i=1}^{N_S} \sum_{l=1}^L y_{i,l} \log p_{i,l} \right) \quad (4)$$

2 基于改进 Res2Net与自适应多尺度窗口池化的调制识别方法

本文提出的 Res2-AMWP 整体架构如图 2 所示,该网络以时域 I/Q 序列作为输入,结合多尺度表征学习与 token 序列建模的方式,实现端到端的调制识别。整体架构由多尺度特征提取、特征融合与注意力分类头三个部分组成。

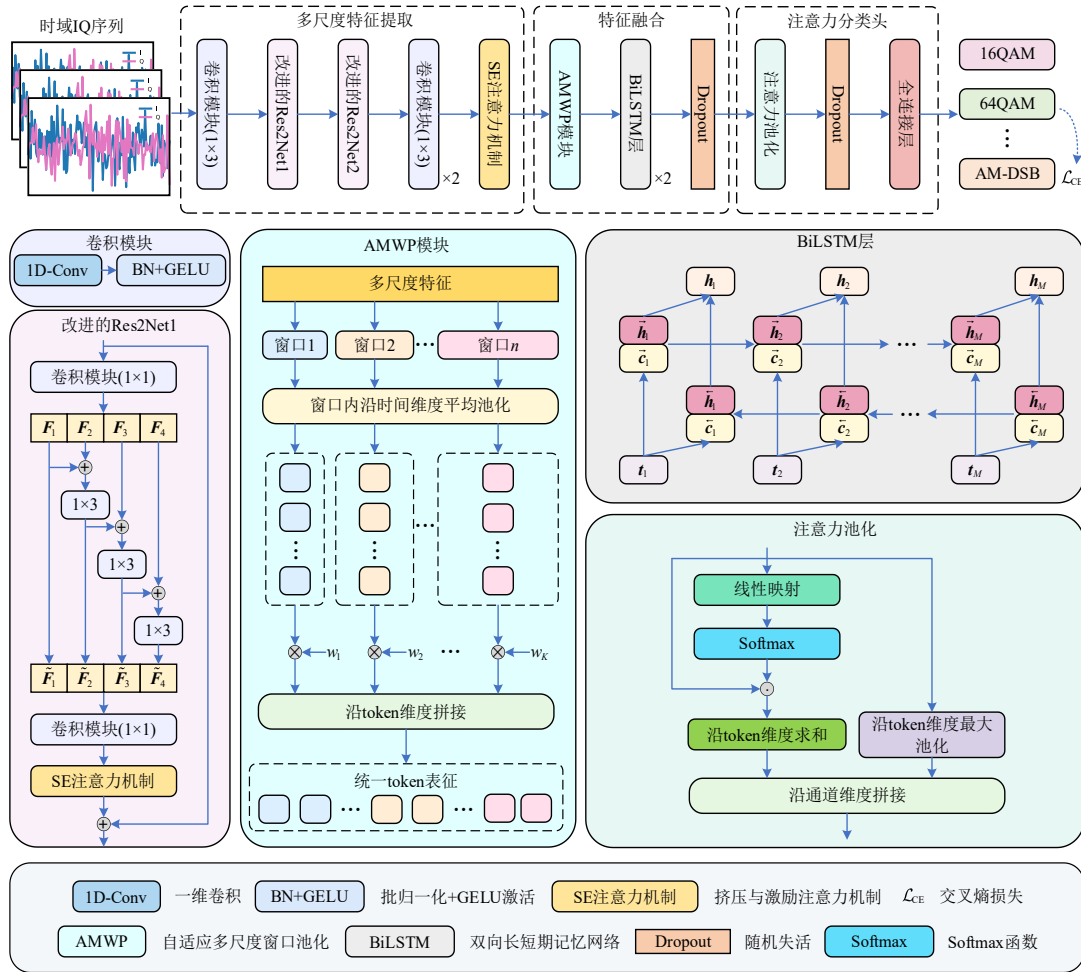


图2 Res2-AMWP整体架构

Figure 2 Overall architecture of Res2-AMWP

首先,在多尺度特征提取阶段,利用卷积模块初步提取特征,其中,卷积模块在一维卷积(1-Dimensional Convolution, 1D-Conv)后引入批归一化(Batch Normalization, BN)与高斯误差线性单元(Gaussian Error Linear Unit, GELU),以稳定特征分布并增强特征的非线性表达,图中的 1×3 、 1×1 表示卷积核所使用的尺寸。两层堆叠的改进 Res2Net 用于多尺度建模,提升网络在低 SNR 下的性能。其中,改进的 Res2Net1 把原始标准卷积替换为一维卷积模块,对输入的特征按照通道分组以进行特征提取,并对特征逐级融合,

随后利用 SE 注意力机制对特征进行通道重标定,提取具有判别性的多尺度特征。改进的 Res2Net2 去掉了从输入到 SE 输出的整体残差相加,其余结构与改进的 Res2Net1 一致,后续的两层卷积模块与 SE 注意力机制用于将所提取特征初步整合与增强。其次,在特征融合阶段,为了在保证复杂度的情况下有效进行进一步的特征融合,本文提出一种自适应多尺度窗口池化 AMWP 模块,利用不同池化窗口将多尺度特征映射为统一的 token 表征,同时定义尺度权重自适应学习不同尺度窗口的重要性,并引入 BiLSTM 建模 to-

ken间的上下文依赖关系。最后,在注意力分类头阶段,对融合后的token表征计算重要性权重,并进行加权汇聚得到全局特征表示,通过全连接层将高维特征映射到调制类别空间,输出对各调制类别的识别概率。

2.1 多尺度特征提取

在通信系统中,调制信号在不同SNR与信道扰动的环境下呈现出非平稳性。判别信息既存在于局部的短时变化中,也依赖于更长时间范围内的整体结构。单一尺度卷积难以充分捕获信号的上下文判别特征,而如果采用多分支卷积结构,往往会带来更多的参数量与计算开销。为此,在特征提取阶段对初始卷积模块提取的特征利用两层改进的Res2Net进行多尺度建模,在不显著增加参数量的前提下提高网络在低SNR下特征提取的有效性。实现方法为在每个Res2Net模块内对已有特征按通道进行分组并采用级联卷积逐级融合,从而在同一模块内形成具有不同感受野的多尺度特征表示。同时,引入SE注意力机制对输出通道进行自适应标定,突出更具判别性的通道信息。

首先,利用卷积模块对输入的I/Q序列构成的矩阵 \mathbf{X} 进行浅层特征提取,为统一表示,将该模块的运算定义为 $C_{d \times l}(\cdot)$,有

$$\mathbf{X}_C = C_{1 \times 3}(\mathbf{X}) = \sigma(\text{BN}(\text{1D-Conv}_{1 \times 3}(\mathbf{X}))) \quad (5)$$

式中: \mathbf{X}_C 为得到的特征张量; $\sigma(\cdot)$ 表示GELU激活函数; $\text{BN}(\cdot)$ 表示批归一化操作; $\text{1D-Conv}_{1 \times 3}(\cdot)$ 表示卷积核尺寸为 1×3 的一维卷积运算。

改进的Res2Net模块采用的卷积操作均为本文定义的卷积模块,以增强非线性表达与训练的稳定性。首先通过一个卷积模块将输入特征的通道维度映射到Res2Net模块的输出通道维度,并按照通道均分为四组子特征。

$$\mathbf{F} = [\mathbf{F}_1, \mathbf{F}_2, \mathbf{F}_3, \mathbf{F}_4] \quad (6)$$

式中: \mathbf{F} 表示经过通道映射之后的特征张量; $\mathbf{F}_1, \mathbf{F}_2, \mathbf{F}_3, \mathbf{F}_4$ 表示经过通道划分之后的特征张量。

随后,利用残差级联的方式逐级融合各组子特征,获得四组不同等效感受野的多尺度特征。

$$\begin{cases} \tilde{\mathbf{F}}_1 = \mathbf{F}_1 \\ \tilde{\mathbf{F}}_2 = C_{1 \times 3}(\tilde{\mathbf{F}}_1 + \mathbf{F}_2) \\ \tilde{\mathbf{F}}_3 = C_{1 \times 3}(\tilde{\mathbf{F}}_2 + \mathbf{F}_3) \\ \tilde{\mathbf{F}}_4 = C_{1 \times 3}(\tilde{\mathbf{F}}_3 + \mathbf{F}_4) \end{cases} \quad (7)$$

式中: $\tilde{\mathbf{F}}_1, \tilde{\mathbf{F}}_2, \tilde{\mathbf{F}}_3, \tilde{\mathbf{F}}_4$ 表示不同感受野的多尺度特征张量。

最后,将上述多尺度特征张量沿通道维度重新拼接,利用卷积模块对拼接后的特征进一步融合,并通

过SE注意力机制对通道特征重新标定,该过程可以表示为

$$\tilde{\mathbf{F}} = \text{SE}\left(C_{1 \times 1}\left(\text{Concat}(\tilde{\mathbf{F}}_1, \tilde{\mathbf{F}}_2, \tilde{\mathbf{F}}_3, \tilde{\mathbf{F}}_4)\right)\right) \quad (8)$$

式中: $\tilde{\mathbf{F}}$ 表示SE注意力机制作用得到的特征张量; $\text{SE}(\cdot)$ 表示SE注意力机制; $\text{Concat}(\cdot)$ 表示拼接操作。

SE注意力机制通过挤压 F_{eq} 、激励 F_{ex} 、缩放 F_{sc} 三个步骤对通道特征进行全局信息聚合并学习各通道的重要性权重,实现对关键通道的自适应增强与非关键通道的抑制,其流程如图3所示。

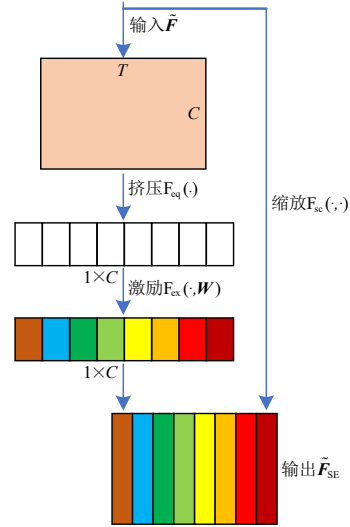


图3 SE注意力机制流程

Figure 3 Workflow of the SE attention mechanism

首先, F_{eq} 操作将输入特征张量 $\tilde{\mathbf{F}} \in \mathbf{R}^{C \times T}$ 沿时间维度将每个时间步对应的元素进行全局平均池化,得到通道描述向量 $\mathbf{v} \in \mathbf{R}^C$,将每个通道用一个具体数值表示:

$$\mathbf{v} = F_{\text{eq}}(\tilde{\mathbf{F}}) = \frac{1}{T} \sum_{t=1}^T \tilde{\mathbf{f}}_{:,t} \quad (9)$$

式中: $t \in \{1, 2, \dots, T\}$ 表示时间维度索引; $\tilde{\mathbf{f}}_{:,t}$ 表示特征张量在时间步索引 t 为时对应的列向量。

随后, F_{ex} 操作利用两层全连接层进行非线性映射,将 \mathbf{v} 进一步转化为通道权重向量 $\mathbf{s} \in \mathbf{R}^C$,其每个分量用于衡量不同通道的重要性。

$$\mathbf{s} = F_{\text{ex}}(\mathbf{v}, \mathbf{W}) = \eta(\mathbf{W}_2 \delta(\mathbf{W}_1 \mathbf{v})) \quad (10)$$

式中: $\delta(\cdot)$ 表示ReLU激活函数; $\eta(\cdot)$ 表示Sigmoid激活函数; $\mathbf{W}_1 \in \mathbf{R}^{C/r \times C}$ 与 $\mathbf{W}_2 \in \mathbf{R}^{C \times C/r}$ 为模型可学习的权重参数矩阵; r 为通道压缩比,本文设置为8。

最后, F_{sc} 操作将权重向量 \mathbf{s} 作用于原特征张量的对应通道,实现通道重标定,得到输出特征 $\tilde{\mathbf{F}}_{\text{SE}}$:

$$\tilde{\mathbf{F}}_{\text{SE}} = F_{\text{sc}}(\tilde{\mathbf{F}}, \mathbf{s}) = \tilde{\mathbf{F}} \odot \mathbf{s} \quad (11)$$

式中: \odot 表示逐通道乘法运算。

两层卷积模块与 SE 注意力机制用于将输出特征 $\tilde{\mathbf{F}}_{\text{SE}}$ 进行初步整合与增强, 得到最终的多尺度特征

$$\mathbf{F}_m = \text{SE}\left(C_{1 \times 3}\left(C_{1 \times 3}\left(\tilde{\mathbf{F}}_{\text{SE}}\right)\right)\right) \quad (12)$$

2.2 特征融合

特征融合阶段主要利用 AMWP 模块将多尺度特征提取阶段提取的多尺度特征张量转化为 token 表征, 接着利用 BiLSTM 对 token 建模双向时序依赖, 同时在 BiLSTM 后加入随机失活 (Dropout) 进行正则化, 以抑制过拟合并提升泛化性能。

具体而言, AMWP 的主要思想是对特征张量以多个尺度的窗口进行特征聚合, 将不同时间尺度下的统计信息直接映射为 token, 与现有卷积嵌入或者线性映射的方法相比, AMWP 采用平均池化实现 token 映射, 避免了为每个尺度额外构建卷积层或投影层, 因而计算开销更低。此外, 模型为不同时间尺度 token 引入可学习的尺度权重, 使模型能够根据特征分布自适应调整各尺度信息的贡献, 这在一定程度上缓解了固定尺度 token 难以同时覆盖不同时间尺度判别信息的问题。

设多尺度模块的输出特征张量为 $\mathbf{F}_m \in \mathbf{R}^{C \times T}$, 给定一个尺度集合 $S_K = \{2^k | k = 1, 2, \dots, K\}$, 其中, 本文设置尺度数 $K = 3$, 这一超参数设置将在 3.4.3 节的消融实验中进行验证。对于任一尺度 $s_k \in S_K$, 采用大小与步长均为 s_k 的一维平均池化窗口对原特征在时间维度进行池化, 这时每个通道维度上共有 $N = T/s_k$ 个窗口, 其中, 单个窗口的池化值可以表示为

$$x_{c,n}^{s_k} = \frac{1}{s_k} \sum_{t=(n-1)s_k+1}^{ns_k} x_{c,t} \quad (13)$$

式中: $n = 1, 2, \dots, N$ 表示窗口位置索引; $c = 1, 2, \dots, C$ 表示通道维度索引; $t = 1, 2, \dots, T$ 表示时间维度索引; $x_{c,t}$ 表示特征张量 \mathbf{F}_m 在第 c 个通道上的第 t 个时间索引处对应的元素; $x_{c,n}^{s_k}$ 表示在尺度 s_k 下, 第 c 个通道上的第 n 个窗口得到的平均池化值, 对应时间索引区间 $[(n-1)s_k + 1, ns_k]$ 。

首先, 将第 n 个窗口在所有通道上的池化值按通道顺序排列组成 token 向量 $\mathbf{x}_n^{s_k} = [x_{1,n}^{s_k}, x_{2,n}^{s_k}, \dots, x_{C,n}^{s_k}] \in \mathbf{R}^{1 \times C}$, 将各窗口在通道上组成的 token 向量按窗口索引 n 依次排列可得到 token 表征:

$$\mathbf{T}_{s_k} = [\mathbf{x}_1^{s_k}, \mathbf{x}_2^{s_k}, \dots, \mathbf{x}_N^{s_k}]^T \in \mathbf{R}^{N \times C} \quad (14)$$

为了使网络能够自适应调节每个尺度 token 的贡献, AMWP 模块为每个尺度 s_k 引入了一个可学习标量 α_k , 并通过 Softmax 函数得到归一化尺度权重:

$$w_k = \frac{\exp(\alpha_k)}{\sum_{i=1}^K \exp(\alpha_i)} \quad (15)$$

随后, 对尺度 s_k 对应的 token 表征进行整体缩放, 得到加权后的 token 表征:

$$\hat{\mathbf{T}}_{s_k} = w_k \mathbf{T}_{s_k} \quad (16)$$

最后, 将所有尺度对应的加权 token 表征沿着 token 维度进行拼接, 得到统一的多尺度 token 表征:

$$\mathbf{T}_M = \text{Concat}(\hat{\mathbf{T}}_{s_1}, \hat{\mathbf{T}}_{s_2}, \dots, \hat{\mathbf{T}}_{s_K}) \in \mathbf{R}^{M \times C} = [\mathbf{t}_1, \mathbf{t}_2, \dots, \mathbf{t}_M]^T \quad (17)$$

式中: M 为拼接后的 token 数量; $\mathbf{t}_1, \mathbf{t}_2, \dots, \mathbf{t}_M$ 表示拼接后得到的 token 向量; 通道数 C 表示每个 token 对应的特征嵌入维度。

获得统一多尺度表征 token 以后, 本文引入两层 BiLSTM 进一步对 token 进行上下文序列建模, 由于 AMWP 生成的 token 按窗口索引具有时间先后关系, 且不同尺度 token 在拼接后共同形成了由局部到长程的多尺度表示, BiLSTM 能够沿 token 顺序正向与反向同时传递信息, 从而有效捕获相邻窗口之间的相关性以及跨尺度的依赖关系。

2.3 注意力分类头

注意力分类头主要由注意力池化、Dropout 以及全连接层组成。与前文 AMWP 在窗口尺度上评估 token 重要性不同, 注意力池化通过注意力机制自适应评估单个 token 的重要性并进行加权汇聚, 同时结合最大池化补充显著特征, 随后在通道维度将两类汇聚特征拼接, 经过 Dropout 与全连接层输出分类预测。

设输入 token 表征为 $\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_M]^T \in \mathbf{R}^{M \times C}$, 其中, \mathbf{h}_m 表示第 m 个 token 的特征向量, 对于每个这样的特征向量, 首先利用线性映射对每个 token 进行重要性评分:

$$\mathbf{e} = \mathbf{H} \mathbf{w}_a, \quad \mathbf{w}_a \in \mathbf{R}^C \quad (18)$$

式中: $\mathbf{e} \in \mathbf{R}^M$ 表示每个 token 的重要性得分向量。

\mathbf{e} 经 Softmax 函数权重得到向量 $\boldsymbol{\beta} = [\beta_1, \beta_2, \dots, \beta_M]$, 其中, β_m 表示第 m 个 token 的重要性系数, 随后在 token 维度进行加权注意力汇聚, 得到全局注意力特征向量:

$$\mathbf{f}_{\text{att}} = \sum_{m=1}^M \beta_m \mathbf{h}_m \quad (19)$$

为了保留最显著的局部特征, 沿 token 维进行最大池化得到最大池化特征向量:

$$\begin{cases} f_{\max,c} = \max_{m \in \{1, 2, \dots, M\}} H_{m,c} \\ \mathbf{f}_{\max} = [f_{\max,1}, f_{\max,2}, \dots, f_{\max,C}]^T \end{cases} \quad (20)$$

式中: $f_{\max,c}$ 表示第 c 个通道上的最大池化值; \mathbf{f}_{\max} 为最大池化特征向量; $H_{m,c}$ 表示第 m 个 token 在第 c 个通道

维度上对应的元素。

为了形成更具判别能力的特征,将 f_{att} 与 f_{max} 在通道维度进行拼接:

$$f = \text{Concat}(f_{att}, f_{max}) \quad (21)$$

全连接层采用两层线性映射的分类头结构。第一层将融合后的全局特征投影到较低维的隐空间,并在其后依次引入 GELU 激活与 Dropout 正则化,以增强非线性表达能力并抑制过拟合;第二层线性映射输出各调制类别的得分,通过 Softmax 函数得到各调制类型的概率分布,最终选取最大概率对应的类别作为识别结果。

3 仿真实验与结果分析

3.1 实验数据集与配置

本文采用 AMR 领域的公开数据集 RadioML2016.10a、RadioML2016.10b、RML22^[23]进行了实验。其中,RadioML2016.10a 在模拟复杂信道条件下产生,该数据集在生成过程中加入了瑞利与莱斯衰落、载波频率偏移以及采样率偏移等典型失真因素,而 RadioML2016.10b 则是在 RadioML2016.10a 的基础上进一步扩展了样本规模。此外,RML22 在 RadioML2016 系列数据集生成的基础之上进一步增强了信道与硬件失真建模,采用更贴近移动通信场景的 3GPP 衰落模型模拟真实信道,并基于晶振误差对载波频率偏移以及采样率偏移进行了更合理地设置。相较于前两个数据集,RML22 的单帧持续时间更长,失真条件更接近真实接收环境,因此更能反映所提方法在模拟复

杂信道下的性能表现。各数据集的参数说明如表 1 所示。

本文所有实验全部基于 PyCharm Professional Edition 2023.1 集成开发环境、Python3.8.18 编程语言、Pytorch1.12.1 的深度学习框架,硬件平台采用 12th Gen Intel (R) Core (TM) i7-12700H CPU、NVIDIA GeForce RTX 3060 laptop GPU 并搭载 CUDA 11.3 并行计算平台对模型训练过程进行加速。数据集按照 6:2:2 的比例进行分层抽样划分,分别作为训练集、验证集、测试集。训练过程中,设置最大训练 epoch 为 100,初始学习率设置为 0.001,训练与测试阶段的批大小均设置为 64,损失函数采用多分类交叉熵(Cross Entropy, CE),并使用 Adam 对网络参数进行自适应梯度优化。为了有效提升训练效率并防止模型过拟合,训练过程中引入了学习率动态调整与早停策略,当验证集损失连续 3 个 epoch 没有下降则将当前学习率衰减为原来的 0.1,当验证集损失连续 10 个 epoch 未下降则终止训练。

评价指标方面,本文实验采用准确率(Accuracy, Acc)、整体准确率(Overall Accuracy, OA)、最大准确率(Maximum Accuracy, MA)、宏平均 F1 分数(Macro-F1)以及 Kappa 系数对模型的识别性能进行综合评估。其中,OA 为各 SNR 条件下 Acc 的平均值,MA 为各 SNR 条件下 Acc 的最大值。同时,本文考虑参数量(Parameters)、浮点运算数(Floating Point Operations, FLOPs)、训练时间(Train time)以及推理时间(Inference time)来综合衡量模型的复杂度与工程部署潜力。

表 1 数据集参数说明

Table 1 Description of dataset parameters

数据集	SNR 范围	样本格式	样本规模	调制方式
RadioML2016.10a	[-20:2:18] dB	(2,128)	220 000	数字调制(BPSK, QPSK, 8PSK, 16QAM, 64QAM, 4PAM, CPFASK, GFSK), 模拟调制(WBFM, AM-SSB, AM-DSB),共 11 种
RadioML2016.10b	[-20:2:18] dB	(2,128)	1 200 000	数字调制(BPSK, QPSK, 8PSK, 16QAM, 64QAM, 4PAM, CPFASK, GFSK), 模拟调制(WBFM, AM-DSB),共 10 种
RML22	[-20:2:20] dB	(2,128)	462 000	数字调制(BPSK, QPSK, 8PSK, 16QAM, 64QAM, 4PAM, CPFASK, GFSK), 模拟调制(WBFM, AM-SSB, AM-DSB),共 11 种

3.2 识别结果

图 4 展示了本文提出的 Res2-AMWP 模型对于 RadioML2016.10a、RadioML2016.10b、RML22 中所有调制信号类型的识别准确率随 SNR 变化的曲线。总体来看,Res2-AMWP 在以上三个数据集上的 OA 分别为 63.51%、65.36%、70.30%,这是由于 RML22 在信号生成过程中对信息源进行了更合理地选取,并对时钟偏移、多普勒等信道与硬件非理想因素进行了更贴合实际的参数化建模,从而提升了数据集的整体建模质

量,使模型能够学习到更稳定的判别特征,从而获得了更高的 OA。

从整体趋势来看,随着 SNR 从 -20 dB 增加到 0 dB,模型对于调制信号的识别准确率快速上升,其原因在于当 SNR 逐步改善后,噪声对信号结构的掩蔽效应减弱,I/Q 信号的特征分布更加清晰,使得模型能够提取到更具判别性的特征表示,从而显著降低误判的概率。从放大图可以看出,当准确率阈值设为 95% 时,模型在 1.5 dB 的 SNR 附近就能对 RadioML2016.10a 上

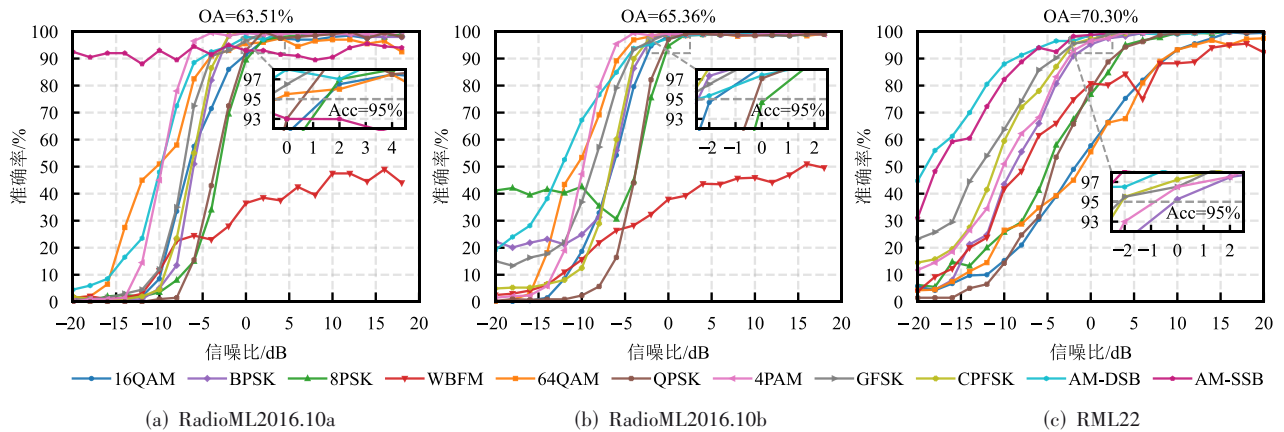


图4 Res2-AMWP模型对不同调制信号的识别准确率

Figure 4 Recognition accuracy of the Res2-AMWP model for different modulation signals

除 WBFM 外的调制类型的识别准确率超过阈值;在样本数量较多的 RadioML2016.10b 上,仅需要 SNR≈0 dB 便可以达到相同水平,而在更贴近真实信道情况的 RML22 数据集上,模型在 0 dB 的 SNR 附近也能对六种调制类型的识别准确率超过阈值,上述结果说明本文方法对于特征具有较强的判别能力与泛化性能。

为进一步分析在较强的噪声以及信道干扰等条件下 Res2-AMWP 模型对各调制类型的判别差异,图 5 给出了模型在 SNR=-4 dB 条件下的混淆矩阵。从结果可以看出,模型的主要误判集中在调制机理相近、特征分布相似的 16QAM 与 64QAM、QPSK 与 8PSK 以

及 WBFM 与 AM-DSB 上。究其原因,高阶 QAM 调制类型的星座图较为密集,最小符号间距更小,在受到噪声、多径等干扰时,星座点云更易发生扩散与旋转,导致分类边界模糊;而 QPSK 与 8PSK 在受到相关干扰时,二者的相位分布容易出现重叠,从而使模型难以分辨;此外,由于 WBFM 和 AM-DSB 是在具有静默周期的模拟语音信号采集得到的,而静默期主要包含载波信息,WBFM 的频率与相位特征更难稳定提取,模型往往将不确定样本判定为统计特征更为稳定的 AM-DSB。相比之下,BPSK、CPFSK、GFSK、4PAM、AM-SSB 未出现显著的混淆现象,说明其判别特征更具区分性,因此整体识别效果相对较好。

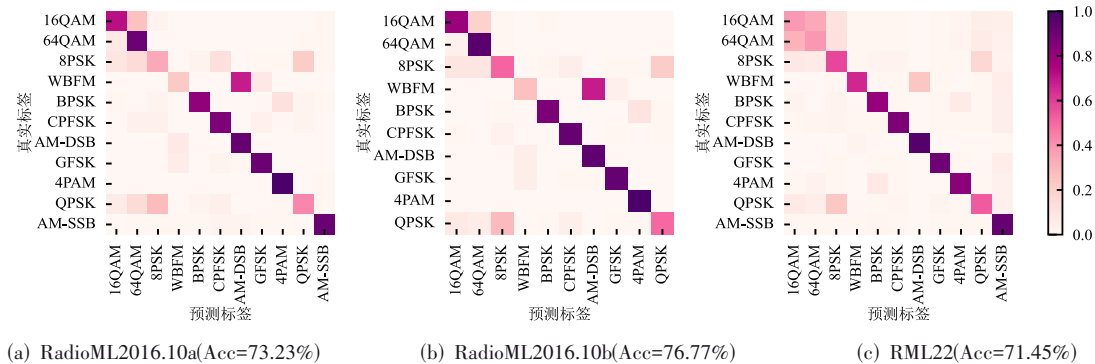


图5 Res2-AMWP模型在SNR=-4 dB条件下的混淆矩阵

Figure 5 Confusion matrix of the Res2-AMWP model at SNR=-4 dB

3.3 对比实验

为全面评估所提 Res2-AMWP 模型的有效性,本文选取了近年来 AMR 领域具有代表性的 DL-AMR 方法作为对比模型进行对比实验,包括 MCLDNN^[24]、LSTM2^[25]、AMC-Net^[26]、WACN-T^[27]、ResNet^[28]、MCNet^[29]、PET-CGDNN^[30],上述对比模型涵盖了 CNN、RNN、Transformer 等不同网络架构,表 2 汇总了各对比模型的主要特点。

3.3.1 性能对比

表 3 展示了不同模型的性能对比结果。从结果可以看出,在 RadioML2016.10a 上,本文提出的 Res2-AMWP 模型在低信噪比区间(SNR≤0 dB)相较于其他对比模型的 Acc、Macro-F1、Kappa 分别提高了 0.44%~4.17%、0.33%~4.96%、0.48%~4.59%,在中高信噪比区间(SNR≥0 dB),上述三项指标分别提高 1.69%~11.47%、1.69%~11.84%、1.86%~12.62%,同时,模型的

表 2 对比模型汇总

Table 2 Summary of compared models.

模型	主要特点
MCLDNN	采用多分支卷积分别提取 I、Q、I/Q 特征,并结合 LSTM 进行时序建模
LSTM2	采用幅度与相位分量作为输入,用两层 LSTM 进行时序建模
AMC-Net	设计频域矫正模块对 I/Q 信号去噪,并行多尺度卷积提取多尺度特征,最后采用 Transformer 编码器进行时序建模
WACN-T	将 I/Q 信号划分为局部窗口,以卷积构建的局部自注意力进行特征提取,实现局部细节与全局依赖的联合建模
ResNet	引入残差连接的 CNN,对输入的 I/Q 信号进行高效的特征提取
MCNet	采用并行的非对称卷积构建多流卷积块以学习 I/Q 信号的多尺度时空相关特征,并通过跳跃连接融合残差信息
PET-CGDNN	先通过参数估计与相位变换模块对接收 I/Q 信号进行相位偏移补偿,再采用轻量 CNN+GRU 混合网络提取空时特征

OA 与 MA 分别较对比模型提高了 1.01%~7.33%、1.55%~10.60%。在 RadioML2016.10b 上,模型在低信噪比区间的 Acc、Macro-F1、Kappa 三个指标分别取得了 0.32%~4.47%、0.28%~4.64%、0.36%~4.96% 的提升,在中高信噪比区间内,上述三个指标的提升分别为 0.26%~9.11%、0.26%~9.31%、0.29%~10.13%,此外,OA 与 MA 也分别提升了 0.32%~6.5%、0.30%~8.89%。在

RML22 上,模型在低信噪比区间内的 Acc、Macro-F1、Kappa 分别提升 0.94%~4.62%、0.68%~5.29%、1.04%~5.08%;在中高信噪比区间内,上述三项指标分别提升 0.51%~12.28%、0.50%~12.36%、0.56%~13.51%,OA 也相较对比模型提升 0.75%~8.40%;MA 除较 PET-CGDNN 略低 0.07% 外,相对其余模型提升 0.13%~13.13%。

表 3 不同模型的性能对比

Table 3 Performance comparison of different models

数据集	模型	SNR≤0 dB			SNR≥0 dB			OA/%	MA/%
		Acc/%	Macro-F1	Kappa	Acc/%	Macro-F1	Kappa		
RadioML2016.10a	MCLDNN	37.92	0.397 5	0.317 1	90.67	0.902 4	0.897 3	61.83	92.27
	LSTM2	37.00	0.382 2	0.307 0	91.07	0.906 6	0.901 8	61.58	92.50
	AMC-Net	38.85	0.406 6	0.327 4	91.13	0.907 1	0.902 4	62.50	92.00
	WACN-T	37.30	0.383 0	0.310 3	85.08	0.845 0	0.835 9	58.92	86.59
	ResNet	37.35	0.384 2	0.310 9	90.18	0.898 0	0.892 0	61.34	91.95
	MCNet	35.12	0.360 3	0.286 3	81.35	0.805 6	0.794 8	56.18	83.45
	PET-CGDNN	35.38	0.365 5	0.289 2	86.73	0.863 2	0.854 0	58.75	88.32
	Res2-AMWP	39.29	0.409 9	0.332 2	92.82	0.924 0	0.921 0	63.51	94.05
RadioML2016.10b	MCLDNN	40.76	0.414 0	0.341 8	93.14	0.926 5	0.923 8	64.44	93.73
	LSTM2	41.05	0.414 3	0.345 0	93.40	0.929 2	0.926 7	64.70	94.12
	AMC-Net	41.54	0.421 6	0.350 5	93.33	0.928 7	0.925 9	64.93	94.03
	WACN-T	41.74	0.422 6	0.352 6	93.37	0.928 9	0.926 4	65.04	93.93
	ResNet	40.69	0.410 6	0.341 0	93.36	0.929 7	0.926 2	64.48	94.09
	MCNet	37.59	0.379 0	0.306 6	84.55	0.839 2	0.828 3	58.86	85.53
	PET-CGDNN	39.94	0.401 5	0.332 7	92.71	0.922 5	0.919 0	63.83	93.68
	Res2-AMWP	42.06	0.425 4	0.356 2	93.66	0.932 3	0.929 6	65.36	94.42
RML22	MCLDNN	45.70	0.442 7	0.402 7	94.35	0.943 3	0.937 8	69.45	99.05
	LSTM2	43.90	0.439 8	0.382 9	92.75	0.927 4	0.920 2	67.72	97.75
	AMC-Net	45.80	0.448 5	0.403 7	94.43	0.943 9	0.938 7	69.50	98.80
	WACN-T	44.73	0.430 5	0.392 0	87.36	0.872 6	0.861 0	65.43	90.91
	ResNet	44.72	0.435 9	0.391 9	90.02	0.900 1	0.890 2	66.71	93.34
	MCNet	42.12	0.402 4	0.363 3	82.91	0.828 1	0.812 0	61.90	86.05
	PET-CGDNN	45.63	0.440 9	0.401 9	94.68	0.946 7	0.941 5	69.55	99.25
	Res2-AMWP	46.74	0.455 3	0.414 1	95.19	0.951 7	0.947 1	70.30	99.18

注:加粗数据为所在数据集上的最优结果。

上述性能对比实验的结果表明,本文所提出的 Res2-AMWP 通过改进的 Res2Net 多尺度残差特征提取、AMWP 自适应多尺度窗口池化、BiLSTM 时序建模、注意力池化的协同设计,有效增强了模型的抗噪能力,并在不同数据集上取得较为领先的性能,具有更强的跨数据集泛化能力。从不同信噪比区间的对比结果来看,Res2-AMWP 的优势更多体现在中高信噪比区间,这主要是由于该区间内噪声干扰减弱,模型能够充分利用多尺度特征提取及跨尺度融合带来的优势,从而得到更加稳定的判别特征。

图 6~图 8 分别展示了三个数据集上不同模型的识别准确率曲线。在 RadioML2016.10a 上,所提模型自 SNR 超过 -6 dB 后便开始展现出较为明显的性能领先优势。而在 RadioML2016.10b 上,该优势从 -4 dB 左右开始显现。此外,在更贴近实际通信场景的 RML22 上,

Res2-AMWP 在 -10 dB 的 SNR 附近便能保持最高识别准确率。Res2-AMWP 性能优势在中高信噪比区间内放大后的识别准确率曲线中可以更加直观地体现出来,以具有代表性的 0 dB 为例,所提模型在上述三个数据集上相较于 MCNet 分别提升了 10.23、14.18、13.98 个百分点。此外,本文利用可容忍识别精度曲线来刻画各模型达到相应识别 Acc 阈值所需要的最低 SNR,在 Acc=80% 的情况下,所提模型达到该阈值所需的 SNR 均为最小,在三个数据集上分别优于其他模型约 0.2~5.37 dB、0.15~2.79 dB、0.19~4.83 dB,这表明 Res2-AMWP 在较低信噪比条件下即可达到与其他模型在更高信噪比下相当的识别准确率,能够在较强干扰的情况下实现有效识别。上述结果表明 Res2-AMWP 对模拟复杂信道环境中的调制信号具有更强的判别能力,体现出了良好的跨数据集泛化能力。

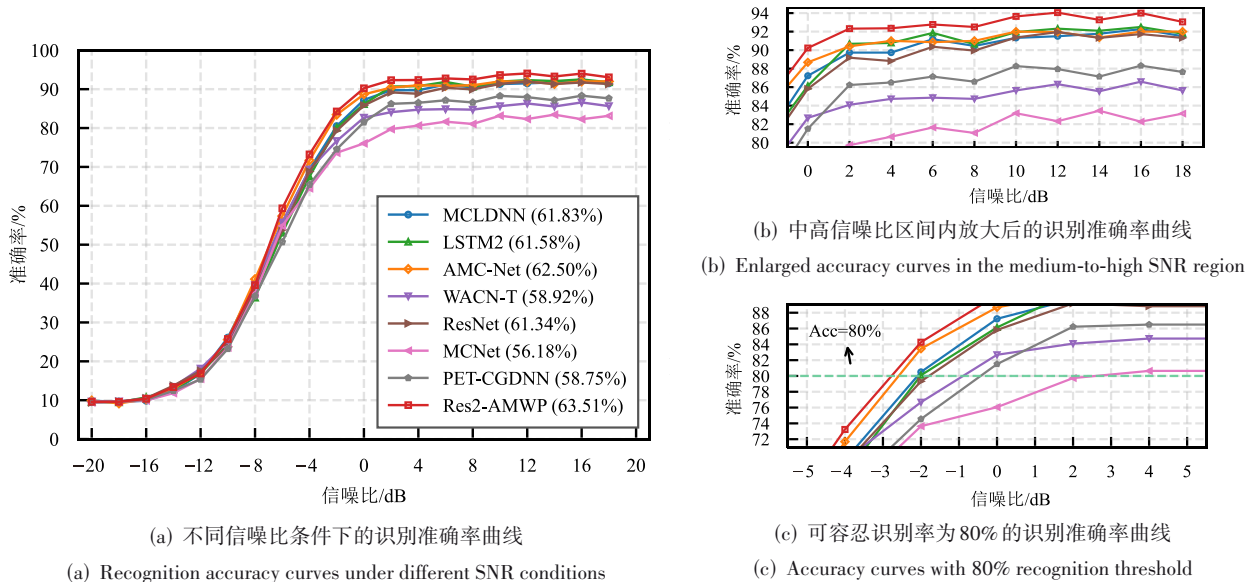


图 6 RadioML2016.10a 上不同模型的识别准确率曲线

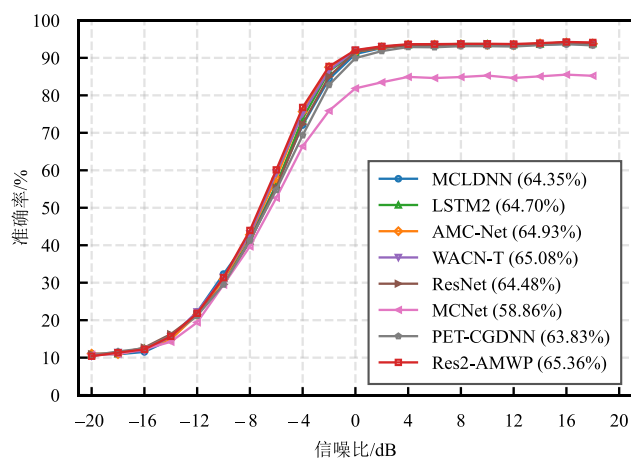
Figure 6 Recognition accuracy curves of different models on RadioML2016.10a

3.3.2 误差对比

t-SNE 是一种将高维数据映射到低维空间的非线性降维算法,可以在较大程度上保持样本的局部邻域结构、直观呈现特征的聚类与可分性,因此本文对各模型最后一层全连接分类器之前的特征张量进行 t-SNE 可视化,以对比分析不同模型在指定 SNR 条件下学习到的特征分布差异。图 9 展示了 RadioML2016.10a 数据集上 SNR=0 dB 情况下不同模型的 t-SNE 可视化对比。从可视化结果可以看出,所有模型在 16QAM 与 64QAM、QPSK 与 8PSK 以及 WBFM 与 AM-DSB 等调制类型之间更容易出现簇间混叠,这一现象的原因与 3.2 节识别结果中分析的一致。此外,相较于其他对

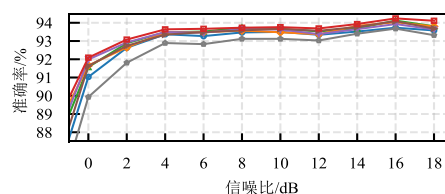
比模型,本文提出的 Res2-AMWP 模型能够形成整体上更为紧凑、类间边界更为清晰的聚类结构,表明其所学习到的特征在噪声干扰下仍具有更好的可分性。

以上 t-SNE 可视化从特征空间角度直观展示了 0 dB SNR 条件下不同模型所学习特征的类内紧凑性与类间可分性差异,但该结果主要体现了定性分析,难以精确量化各调制类别最终的误判数量。为进一步从识别结果角度对模型性能进行定量评估,并与 t-SNE 中呈现的潜在混叠现象相互印证,图 10 给出了各模型在 RadioML2016.10a 数据集上 SNR=0 dB 条件下不同模型的混淆矩阵对比,各图中对角线元素对应



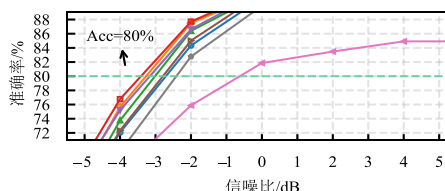
(a) 不同信噪比条件下的识别准确率曲线

(a) Recognition accuracy curves under different SNR conditions



(b) 中高信噪比区间内的识别准确率曲线

(b) Recognition accuracy curves in the medium-to-high SNR region

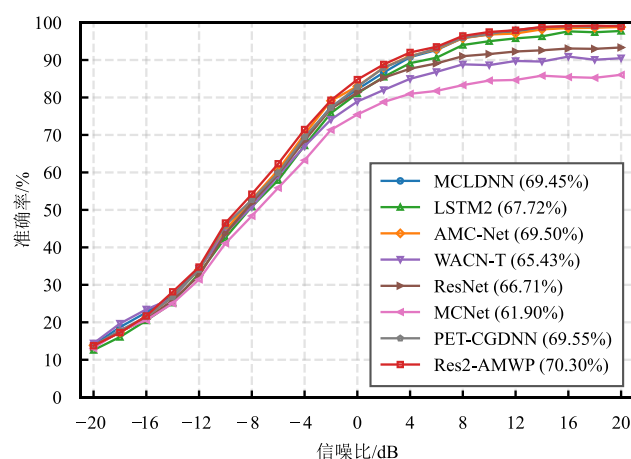


(c) 可容忍识别率为80%的识别准确率曲线

(c) Accuracy curves with 80% recognition threshold

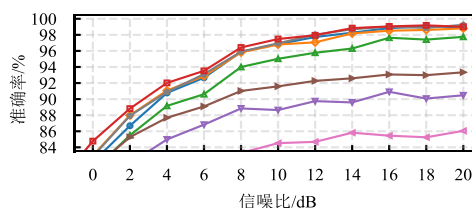
图7 RadioML2016.10b上不同模型的识别准确率曲线

Figure 7 Recognition accuracy curves of different models on RadioML2016.10b



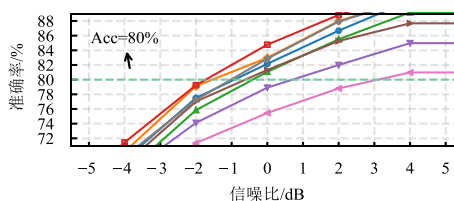
(a) 不同信噪比条件下的识别准确率曲线

(a) Recognition accuracy curves under different SNR conditions



(b) 中高信噪比区间内的识别准确率曲线

(b) Recognition accuracy curves in the medium-to-high SNR region



(c) 可容忍识别率为80%的识别准确率曲线

(c) Accuracy curves with 80% recognition threshold

图8 RML22上不同模型的识别准确率曲线

Figure 8 Recognition accuracy curves of different models on RML22

各调制类别的正确判别样本数,非对角线元素则反映不同类别之间的误判样本数。

从结果可以看出,Res2-AMWP在多数调制类型上的正确判别样本数更多。所有模型在WBFM上出现了较为明显的误判,其原因在3.2节已经讨论过,而Res2-AMWP模型虽在该类型的判别上并不占优,但整体表现与多数对比模型较为接近。相比较之下,本文提出的Res2-AMWP模型在易产生误判的16QAM、64QAM、8PSK、QPSK四类上的混淆样本数分别为17、9、21、14,相较其他对比模型明显减少。此外,该结果与前文t-SNE中呈现的现象相一致,表明所提出的

方法在噪声干扰下仍能对多数调制类型具有更强的区分能力。

3.3.3 复杂度对比

表4对比了多个指标,以全面评估模型在实际应用中的可行性。结果显示,PET-CGDNN在Parameters、Train time、Inference time上取得领先优势,整体复杂度最低,MCNet的FLOPs最低。本文提出的Res2-AMWP模型在Parameters与FLOPs处于适中水平,同时Train time与Inference time均保持在较低值,实现了精度与复杂度的较好平衡,在实际部署中具有一定的应用潜力。

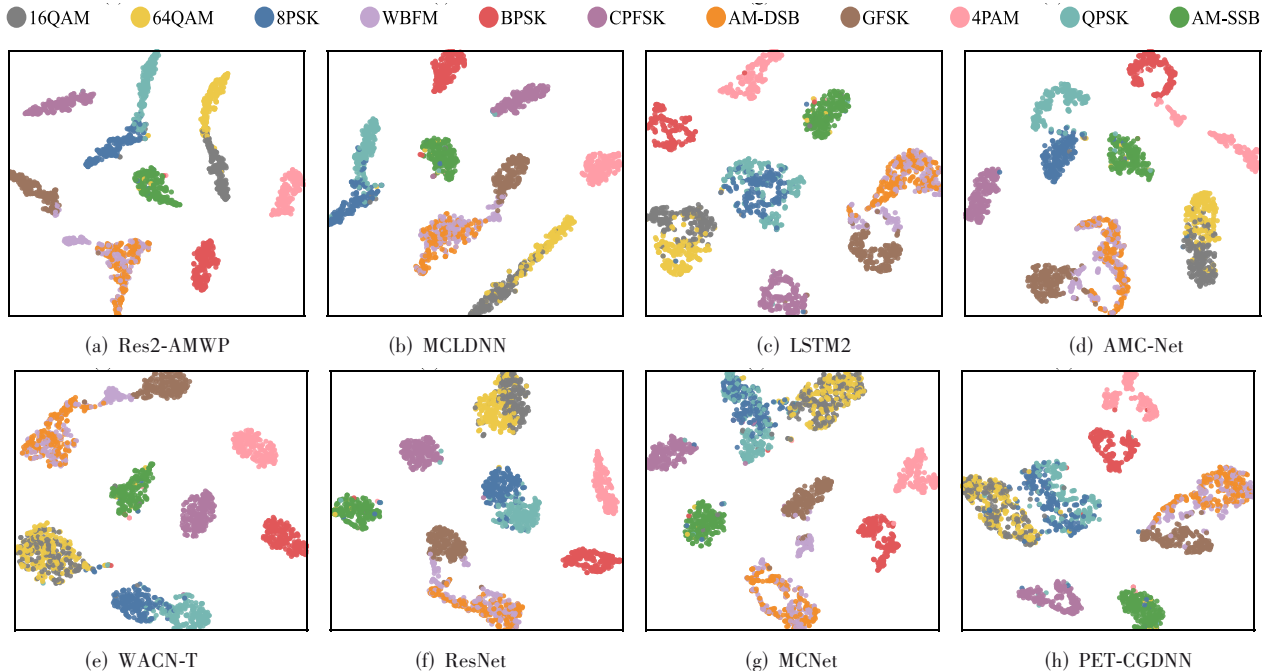


图9 RadioML2016.10a上SNR=0 dB条件下不同模型的t-SNE可视化对比

Figure 9 t-SNE visualization comparison of different models on RadioML2016.10a at SNR=0 dB

3.4 消融实验

3.4.1 模块消融实验

为了验证所提出网络中各关键模块对调制识别性能的具体贡献,本文在RadioML2016.10a上开展了消融实验。如前文所述,考虑到低信噪比与信道扰动会同时破坏信号的局部细节与长程结构,采用改进后的Res2Net进行多尺度残差特征提取并结合SE注意力机制以增强网络在低SNR条件下的判别能力。进一步地,AMWP模块旨在几乎不增加参数数量的前提下通过多尺度池化窗口将多尺度特征转化为更有效的token表征。注意力池化可对时序建模后的序列特征进行自适应汇聚,突出判别性更强的token表征。模块消融对比的结果如表5所示,其中,R1与R2分别表示特征提取中的改进后的Res2Net1与Res2Net2模块;MP表示AMWP模块;AP表示注意力池化模块;此外,符号 \times 与 \surd 分别表示该模块移除与保留的状态。需要说明的是,当同时去除两层改进Res2Net模块时,为保证特征的通道维度能够正常映射,本文使用一节2.1节所介绍的卷积核尺寸为 1×1 的卷积模块进行替代;当去除注意力池化模块时,则采用全局平均池化对序列特征进行聚合。

从模块消融对比的结果可以看出,完整模型在RadioML2016.10a数据集上取得了最优的性能,当分别移除Res2Net1与两层改进的Res2Net模块时,对应模型的OA分别下降了0.83%、1.85%,MA分别下降了0.64%、1.41%,且相比不使用改进Res2Net的模型,完

整模型在低信噪比区间的Acc提升了2.15%,高于中高信噪比区间的1.58%,说明改进的Res2Net通过逐级通道卷积与SE注意力机制的结合显式提取并增强了信号特征的局部细节与上下文信息,有效提高了模型在低信噪比下对调制类型的判别能力,且两层的叠加带来了更显著的增益。在去除了AMWP模块后,对应模型的OA与MA分别下降了1%、0.69%,同时注意到模型的参数量几乎没有发生变化,这表明AMWP模块在几乎不引入额外参数的情况下,通过时域多尺度窗口池化方式对不同时间尺度的信息进行融合,将多尺度特征转化为了更有效的token表征,有效提升了模型的性能。使用平均池化代替注意力池化后,模型的OA与MA下降分别为0.63%、0.50%,降幅相对较小,说明简单的平均池化无法有效汇聚token间的全局特征,而注意力池化为各个token自适应分配权重并加权聚合,有效突出了判别性更强的token表征。此外,其他对比指标Macro-F1与Kappa变化趋势与OA、MA一致,这说明各模块改善了模型对调制信号识别的一致性与均衡性。

3.4.2 token化方式消融实验

为验证所提AMWP模块token化方式的有效性,并分析不同token化方式对模型性能与复杂度的影响,表6展示了在数据集RadioML2016.10a上模型采用不同的token化方式(卷积映射、线性映射)的消融对比。其中,Train time与Inference time的单位与表4一致,线性映射与卷积映射分别采用线性投影矩阵以及多尺度

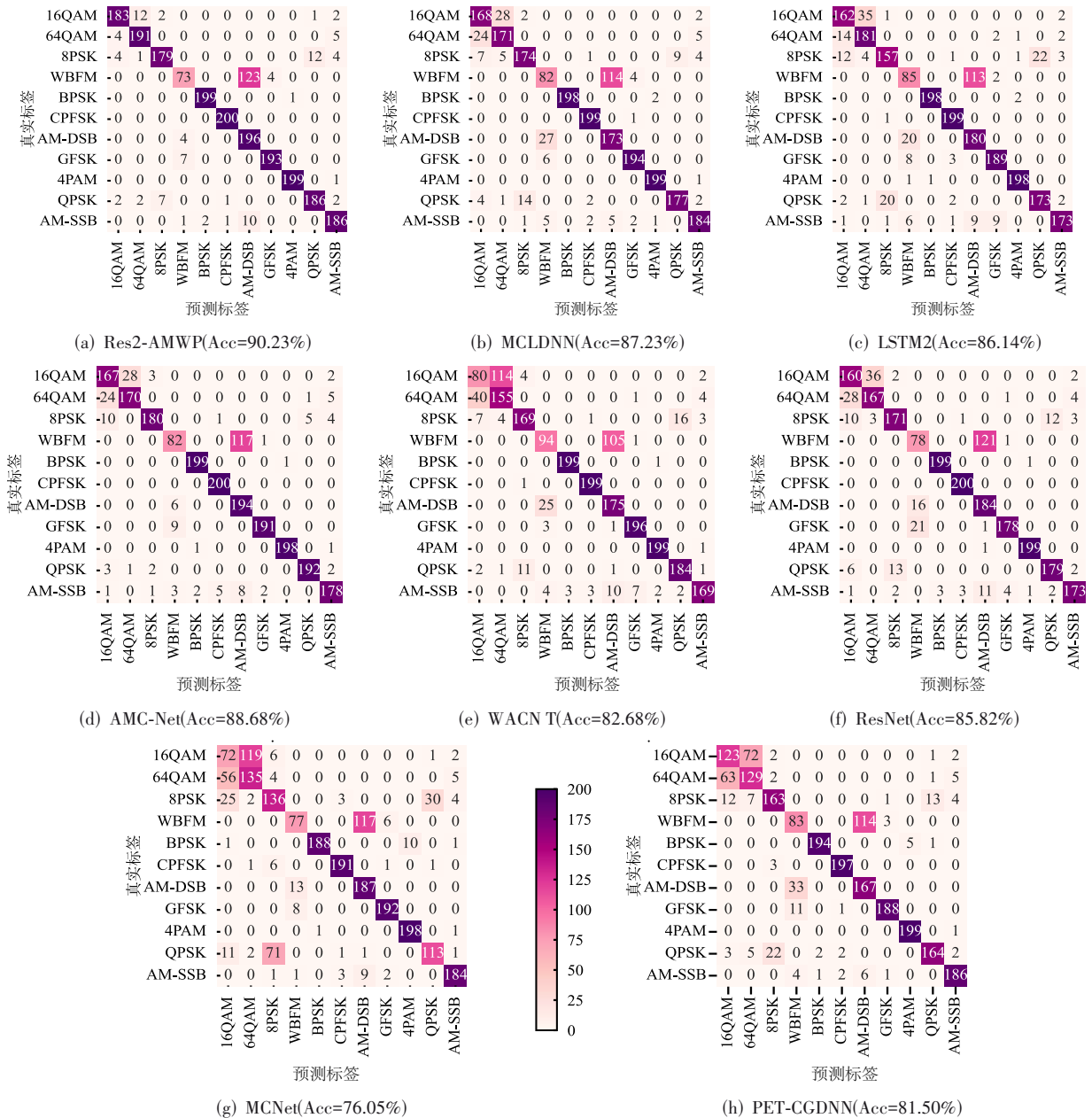


图 10 RadioML2016.10a 上 SNR=0 dB 条件下不同模型的混淆矩阵对比

Figure 10 Confusion matrix comparison of different models on RadioML2016.10a at SNR=0 dB

并行卷积(卷积核大小分别为 1×2 、 1×4 、 1×8)结构将特征映射为与 AMWP 模块输出维度一致的 token 表征。

从 token 化方式消融对比的结果可以看出,相较于无 token 化(无 AMWP 模块),线性映射、卷积映射、AMWP 的 OA 分别提高了 0.59%、0.26%、1.00%,而在 MA 指标上,线性映射与无 token 化持平,卷积映射出现了 0.13% 的小幅下降,AMWP 则提升了 0.69%,此外,在低信噪比与中高信噪比区间下,AMWP 均取得了所有识别性能指标的最优值,体现了其在不同信噪比下的稳定性能优势。在复杂度评估方面,线性映射与卷积映射均在不同程度上增加了模型的开销,其中,卷积映

射由于采用多尺度并行卷积结构,引入了更多的计算参数,使模型的整体复杂度达到最高,相比之下,AMWP 相较无 token 化的各项复杂度指标都最为接近。上述分析说明,AMWP 模块能够在较低复杂度与计算开销的情况下,将特征转化为更具判别性的 token 表征,且相较传统 token 化方式实现了更有效的特征融合。

3.4.3 AMWP 模块窗口尺度超参数消融实验

AMWP 中不同的窗口尺度组合能够覆盖不同时间尺度的统计特征,从而影响 token 表征对特征的跨尺度融合能力。为验证 AMWP 模块窗口尺度超参数对模型性能的影响,图 11 展示了 RadioML2016.10a 上

表 4 模型复杂度对比(A:RadioML2016.10a, B:RadioML2016.10b, C:RML22)

Table 4 Model complexity comparison(A:RadioML2016.10a, B:RadioML2016.10b, C:RML22)

模型	Parameters/M	FLOPs/M	Train time(A, B, C)(ms/样本)	Inference time(A, B, C)(ms/样本)
MCLDNN	0.41	97.39	(0.209, 0.183, 0.181)	(0.062, 0.060, 0.058)
LSTM2	0.20	50.92	(0.103, 0.099, 0.090)	(0.039, 0.038, 0.035)
AMC-Net	0.47	93.05	(0.273, 0.248, 0.243)	(0.061, 0.063, 0.065)
WACN-T	3.39	48.27	(0.973, 0.965, 1.019)	(0.149, 0.188, 0.167)
ResNet	3.85	87.40	(0.291, 0.278, 0.290)	(0.065, 0.063, 0.077)
MCNet	0.12	2.87	(0.284, 0.275, 0.291)	(0.062, 0.061, 0.072)
PET-CGDNN	0.07	16.36	(0.070, 0.076, 0.073)	(0.018, 0.020, 0.021)
Res2-AMWP	0.27	48.99	(0.258, 0.243, 0.266)	(0.054, 0.056, 0.056)

注:加粗数据为最优结果。

表 5 RadioML2016.10a 上模块消融对比

Table 5 Ablation comparison of modules on RadioML2016.10a

R1	R2	MP	AP	SNR≤0 dB			SNR≥0 dB			OA/%	MA/%	Parameters/M
				Acc/%	Macro-F1	Kappa	Acc/%	Macro-F1	Kappa			
×	√	√	√	38.35	0.398 7	0.321 9	92.05	0.916 4	0.912 5	62.68	93.41	0.25
×	×	√	√	37.14	0.384 1	0.308 6	91.24	0.908 0	0.903 6	61.66	92.64	0.23
√	√	×	√	38.24	0.395 4	0.320 6	91.76	0.913 3	0.909 3	62.51	93.36	0.27
√	√	√	×	38.65	0.402 8	0.325 1	92.17	0.917 6	0.913 9	62.88	93.50	0.25
√	√	√	√	39.29	0.409 9	0.332 2	92.82	0.924 0	0.921 0	63.51	94.05	0.27

注:加粗数据为最优结果。

表 6 RadioML2016.10a 上 token 化方式消融对比

Table 6 Ablation comparison of tokenization methods on RadioML2016.10a

token 化方式	SNR≤0 dB			SNR≥0 dB			OA/%	MA/%	Parameters/M	Train time	Inference time
	Acc/%	Macro-F1	Kappa	Acc/%	Macro-F1	Kappa					
无 token 化	38.24	0.395 4	0.320 6	91.76	0.913 3	0.909 3	62.51	93.36	0.27	0.246	0.052
线性映射	38.97	0.405 8	0.328 7	92.32	0.918 7	0.915 5	63.10	93.36	0.29	0.276	0.073
卷积映射	38.68	0.404 2	0.325 5	91.85	0.914 4	0.910 3	62.77	93.23	0.40	0.310	0.065
AMWP	39.29	0.409 9	0.332 2	92.82	0.924 0	0.921 0	63.51	94.05	0.27	0.258	0.054

注:加粗数据为最优结果。

不同窗口尺度数 K 的消融对比,其中,窗口尺度集合 $S_K = \{2^k | k = 1, 2, \dots, K\}$ 。

从窗口尺度数消融对比的结果可以看出,随着 K 从 1 增至 3,模型的 OA 由 62.48% 提升至 63.51%,说明多尺度窗口能够同时引入局部细节与更长时间依赖的统计特征,从而增强了 token 表征的判别性,有效提升了模型性能。而当 K 增加到 4 时,MA 进一步达到 94.18%,但 OA 降低了 0.22 个百分点,这是由于额外尺度在部分信噪比区间能够强化更大时间尺度的统计特征从而带来更高的峰值增益,但整体信息出现冗余,对有效特征的表达与融合造成了一定干扰;当 K 继续增至 5 后,两项性能指标均出现大幅下降,表明过大尺度可能带来了过度平滑与信息冗余,导致模型的整体性能受损。因此当 $K=3$,即窗口尺度集合为 $S_3 = \{2, 4, 8\}$ 时,模型能够达到最优的综合性能。

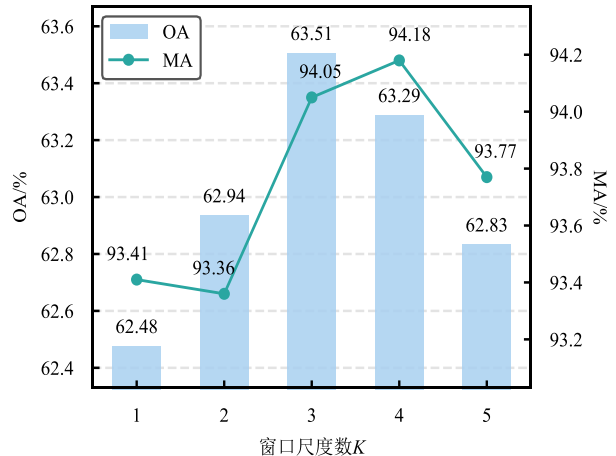


图 11 RadioML2016.10a 上窗口尺度数的消融对比

Figure 11 Ablation comparison of the number of window scales on RadioML2016.10a

4 结束语

针对现有方法在复杂信道条件下多尺度特征融合能力不足、特征 token 化方式有效性与复杂度难以平衡的问题,本文提出一种基于改进 Res2Net 与自适应多尺度窗口池化的调制识别方法 Res2-AMWP。其中,特征提取阶段将改进的 Res2Net 与 SE 注意力机制结合,有效提高了模型在低信噪比条件下的识别准确率。特征融合阶段提出自适应多尺度窗口池化 AMWP 模块,通过可学习的窗口池化机制将多尺度特征映射为更具判别性的 token 表征,较传统 token 化方式更好地兼顾了表征的有效性与复杂度。此外,利用 BiLSTM 建模 token 间的时序依赖关系,实现了有效的特征融合。后续注意力池化对 token 进行自适应汇聚,突出了关键 token 表征。在公开数据集 RadioML2016.10a、RadioML2016.10b、RML22 上的实验结果表明,Res2-AMWP 的整体识别准确率分别达到 63.51%、65.36%、70.30%,相较于多种对比方法分别提高了 1.01%~7.33%、0.32%~6.5%、0.75%~8.40%,且模型的复杂度保持在较低水平,实现了精度与复杂度的较好平衡。

后续研究工作将从以下三个方面展开:一是探索更强的特征学习策略,考虑引入多模态联合表示或对比学习等方法,以进一步提高模型在低 SNR 或复杂信道条件下的性能;二是深入研究 AMWP 的尺度选择与 token 构建机制,引入窗口尺度的动态设计,减少其对人工超参数的依赖;三是引入信号处理方面的相关方法,将信号的时频特征、高阶统计特征等信息与深度神经网络相结合,进一步探索噪声抑制与特征增强的策略。

参考文献

- [1] Kong Weisi, Jiao Xun, Xu Yuhua, et al. A transformer-based contrastive semi-supervised learning framework for automatic modulation recognition[J]. *IEEE Transactions on Cognitive Communications and Networking*, 2023, 9(4): 950-962.
- [2] Zhang Jiawei, Wang Tiantian, Feng Zhixi, et al. Toward the automatic modulation classification with adaptive wavelet network[J]. *IEEE Transactions on Cognitive Communications and Networking*, 2023, 9(3): 549-563.
- [3] 郑庆河, 刘方霖, 余礼苏, 等. 基于改进 Kolmogorov-Arnold 混合卷积神经网络的调制识别方法[J]. *电子与信息学报*, 2025, 47(8): 2584-2597.
Zheng Qinghe, Liu Fanglin, Yu Lisu, et al. An improved modulation recognition method based on hybrid Kolmogorov-Arnold convolutional neural network[J]. *Journal of Electronics & Information Technology*, 2025, 47(8): 2584-2597. (in Chinese)
- [4] Hao Xiaoyang, Feng Zhixi, Yang Shuyuan, et al. Automatic modulation classification via meta-learning[J]. *IEEE Internet of Things Journal*, 2023, 10(14): 12276-12292.
- [5] Liu Fugang, Pan Jingyi, Zhou Ruolin. Contrastive learning-based multimodal fusion model for automatic modulation recognition[J]. *IEEE Communications Letters*, 2024, 28(1): 78-82.
- [6] Zheng Jianping, Lyu Yanfang. Likelihood-based automatic modulation classification in OFDM with index modulation[J]. *IEEE Transactions on Vehicular Technology*, 2018, 67(9): 8192-8204.
- [7] Majhi S, Gupta R, Xiang Weidong, et al. Hierarchical hypothesis and feature-based blind modulation classification for linearly modulated signals[J]. *IEEE Transactions on Vehicular Technology*, 2017, 66(12): 11057-11069.
- [8] Chang Shuo, Yang Zheng, He Jiashuo, et al. A fast multi-loss learning deep neural network for automatic modulation classification[J]. *IEEE Transactions on Cognitive Communications and Networking*, 2023, 9(6): 1503-1518.
- [9] An T T, Lee B M. Robust automatic modulation classification in low signal to noise ratio[J]. *IEEE Access*, 2023, 11: 7860-7872.
- [10] Li Lin, Zhu Yaxin, Zhu Zhigang. Automatic modulation classification using ResNeXt-GRU with deep feature fusion[J]. *IEEE Transactions on Instrumentation and Measurement*, 2023, 72: 2519710.
- [11] O'Shea T J, Corgan J, Clancy T C. Convolutional radio modulation recognition networks[C]//*Engineering Applications of Neural Networks*. Cham: Springer, 2016: 213-226.
- [12] Hong Dehua, Zhang Zilong, Xu Xiaodong. Automatic modulation classification using recurrent neural networks[C]//2017 3rd IEEE International Conference on Computer and Communications (ICCC). Piscataway: IEEE, 2018: 695-700.
- [13] O'Shea T J, West N. Radio machine learning dataset generation with GNU Radio[C]//*Proceedings of the 6th GNU Radio Conference*. Boulder: The GNU Radio Foundation, 2016: 1-6.
- [14] Zhang Zufan, Luo Hao, Wang Chun, et al. Automatic modulation classification using CNN-LSTM based dual-stream structure[J]. *IEEE Transactions on Vehicular Technology*, 2020, 69(11): 13521-13531.
- [15] Huang Sai, Dai Rui, Huang Juanjuan, et al. Automatic modulation classification using gated recurrent residual network[J]. *IEEE Internet of Things Journal*, 2020, 7(8): 7795-7807.
- [16] Zhang Hao, Zhou Fuhui, Wu Qihui, et al. A novel automatic modulation classification scheme based on multi-scale networks[J]. *IEEE Transactions on Cognitive Com-*

munications and Networking, 2022, 8(1): 97-110.

- [17] Tan Haoyue, Zhang Zhenxi, Li Yu, et al. Multi-scale feature fusion and distribution similarity network for few-shot automatic modulation classification[J]. IEEE Signal Processing Letters, 2024, 31: 2890-2894.
- [18] Dao T T, Noh D I, Pham Q V, et al. VT-MCNet: High-accuracy automatic modulation classification model based on vision transformer[J]. IEEE Communications Letters, 2024, 28(1): 98-102.
- [19] O'Shea T J, Roy T, Clancy T C. Over-the-air deep learning based radio signal classification[J]. IEEE Journal of Selected Topics in Signal Processing, 2018, 12(1): 168-179.
- [20] Cai Jingjing, Gan Fengming, Cao Xianghai, et al. Signal modulation classification based on the transformer network[J]. IEEE Transactions on Cognitive Communications and Networking, 2022, 8(3): 1348-1357.
- [21] Gao Shanghua, Cheng Mingming, Zhao Kai, et al. Res2Net: A new multi-scale backbone architecture[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43(2): 652-662.
- [22] Hu Jie, Shen Li, Sun Gang. Squeeze-and-excitation networks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 7132-7141.
- [23] Sathyanarayanan V, Gerstoft P, El Gamal A. RML22: Realistic dataset generation for wireless modulation classification[J]. IEEE Transactions on Wireless Communications, 2023, 22(11): 7663-7675.
- [24] Xu Jialang, Luo Chunbo, Parr G, et al. A spatiotemporal multi-channel learning framework for automatic modulation recognition[J]. IEEE Wireless Communications Letters, 2020, 9(10): 1629-1632.
- [25] Rajendran S, Meert W, Giustiniano D, et al. Deep learning models for wireless signal classification with distributed low-cost spectrum sensors[J]. IEEE Transactions on Cognitive Communications and Networking, 2018, 4(3): 433-445.
- [26] Zhang Jiawei, Wang Tiantian, Feng Zhixi. AMC-Net: An effective network for automatic modulation classification[C]//ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing. Piscataway: IEEE, 2023: 10097070.
- [27] Feng Yuan, Peng Kexiao, Wei Jiaolong, et al. Window attention convolution network (WACN): A local self-attention automatic modulation recognition method[J]. IEEE Transactions on Cognitive Communications and Networking, 2025, 11(3): 1597-1608.
- [28] Liu Xiaoyu, Yang Diyu, El Gamal A. Deep neural network architectures for modulation classification[C]//2017 51st Asilomar Conference on Signals, Systems, and Computers. Piscataway: IEEE, 2017: 915-919.
- [29] Huynh-The T, Hua C H, Pham Q V, et al. MCNet: An efficient CNN architecture for robust automatic modulation classification[J]. IEEE Communications Letters, 2020, 24(4): 811-815.
- [30] Zhang Fuxin, Luo Chunbo, Xu Jialang, et al. An efficient deep learning model for automatic modulation recognition based on parameter estimation and transformation[J]. IEEE Communications Letters, 2021, 25(10): 3287-3290.

作者简介



王 丹 女,1981年9月出生于重庆市。博士,现为正高级工程师、重庆邮电大学硕士生导师。主要研究方向为移动通信物理层算法、物联网、信号处理。

E-mail: wangdan@cqupt.edu.cn



李万杰 男,2001年10月出生于重庆市。现为重庆邮电大学硕士研究生。主要研究方向为调制识别、信号处理。

E-mail: 1531603255@qq.com



江丰杨 男,2002年2月出生于浙江省杭州市。现为重庆邮电大学硕士研究生。主要研究方向为调制识别、信号处理。

E-mail: fengyang_jiang@163.com