

基于多尺度特征融合的轻量级自动调制识别网络

李世豪, 钱 博

沈阳理工大学信息科学与工程学院, 辽宁 沈阳

收稿日期: 2025年11月2日; 录用日期: 2025年11月25日; 发布日期: 2025年12月1日

摘 要

针对现有自动调制识别模型在资源受限场景下部署困难的问题, 本文建立了基于多尺度特征融合的轻量级自动调制识别网络, 结合多尺度深度可分离卷积、通道注意力机制与循环神经网络的高效识别架构, 采用端到端的单输入设计, 通过多分支深度可分离卷积提取信号的时频特征, 并利用LSTM单元捕捉时序依赖关系, 实现了轻量准确的信号调制方式识别功能。实验表明, 所提模型在保持高识别精度的同时, 显著降低了计算复杂度与内存占用, 适用于边缘计算设备部署。

关键词

自动调制识别, 轻量级网络, LSTM

A Lightweight Automatic Modulation Recognition Network Based on Multi-Scale Feature Fusion

Shihao Li, Bo Qian

School of Information Science and Engineering, Shenyang Ligong University, Shenyang Liaoning

Received: November 2, 2025; accepted: November 25, 2025; published: December 1, 2025

Abstract

To address the deployment challenges of existing automatic modulation recognition (AMR) models in resource-limited scenarios, this paper introduces a lightweight network based on multi-scale feature fusion. The proposed architecture efficiently integrates multi-scale depthwise separable convolution, a channel attention mechanism, and a recurrent neural network. It employs an end-to-end,

single-input design. This design uses multi-branch depthwise separable convolutions to extract time-frequency features and LSTM units to capture temporal dependencies. This enables accurate and lightweight modulation recognition. Experimental results show that our model maintains high recognition accuracy while significantly reducing computational complexity and memory usage. It is, therefore, suitable for deployment on edge computing devices.

Keywords

Automatic Modulation Recognition, Lightweight Network, LSTM

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随着无线通信技术的飞速演进,自动调制识别作为非合作通信中的关键技术,在频谱监测、认知无线电以及军事侦听等领域扮演着日益重要的角色。传统方法依赖人工设计的特征与专家经验,在复杂多变的信道环境中泛化能力有限,难以适应实际应用的需求[1]。深度学习技术的引入改变了这一局面,其端到端的学习机制能够直接从原始信号中提取判别性特征,显著提升了识别系统的鲁棒性与准确率。

O'Shea 等人于 2016 年开创性地将卷积神经网络应用于调制识别任务,直接将基带 I/Q 信号作为输入,对 11 类调制方式进行分类,在 0 dB 信噪比下识别准确率达到 80%,为该领域的研究奠定了基础[2]。随着深度学习技术的深入应用,研究者逐渐将重心转向提升模型的识别性能,提出了多种结构复杂的深度网络。例如, Xu 等人设计的多通道卷积长短时深度神经网络(MCLDNN)通过多路径输入分别处理 I/Q 两路信号,并融合 CNN 与 LSTM 以联合提取信号的时域与空域特征[3]。在 RML2016.10a 数据集上的实验表明,该模型在 0 dB 至 18 dB 信噪比范围内的平均识别率达到 92%,在 12 dB 时最高识别精度达 92.95%。然而,该模型的参数量高达 40.5 万,计算复杂度限制了其在资源受限场景中的应用。Liu 等人进一步提出基于时间注意力机制与 LSTM 的编码器-解码器结构(LSTM-AL),通过注意力机制强化对关键时间步的建模能力,在 AM-SSB、BPSK、CPFSK 和 16-QAM 等多种调制类型上均取得了优于传统 LSTM 模型的识别效果,展现出复杂网络在特征提取方面的潜力[4]。

随着研究的深入,为了满足实际部署中对计算效率和资源消耗的要求,轻量级网络架构逐渐成为研究热点。Zhang 等人提出的 PET-CGDNN 模型通过引入相位估计与补偿模块提升了对相位偏移的鲁棒性,并结合 CNN 与 GRU 实现轻量化设计。在 RML2016.10a 数据集上,该模型在未剪枝情况下的平均识别准确率为 60.44%,最高识别精度达 91.36%,而参数量仅为 7.1 万。通过剪枝技术,该模型能在保持 90% 以上识别精度的同时,将参数量进一步压缩至 1.4 万[5]。Zhu 等人设计的 MSGNet 采用深度可分离卷积与门控循环单元(GRU)构建轻量级多特征融合网络,从幅度/相位、I/Q 样本以及频域实部/虚部三个通道中提取空间、时间和频率特征。在 RML2016.10a 数据集上,MSGNet 在 0 dB 至 18 dB SNR 范围内的平均识别准确率达到 92.42%,最高识别精度在 14 dB 时为 93.4%,而模型参数量仅为 14.9 万,在精度和效率间取得了良好平衡[6]。Zeng 等人探索了基于脉冲神经网络(SNN)与 Transformer 的混合模型,利用 SNN 的稀疏激活和二值化特性,在 RadioML2018.01a 数据集上实现了与全精度模型相当的识别精度,同时将模型尺寸压缩至全精度模型的 1/20,功耗降低超过 95%,为在非地面网络(NTN)等资源受限环境中的部署提供了新的解决方案[7]。

尽管上述轻量级模型在降低复杂度方面取得了显著进展,但它们大多侧重于单一优化策略: PET-CGDNN 依赖于剪枝技术, MSGNet 采用多特征融合但未充分利用时序建模潜力, SNN-Transformer 混合模型则面临训练复杂性和时序建模精度的挑战。鉴于此, 本文提出一种轻量级多尺度特征融合网络。该网络采用多分支深度可分离卷积结构提取不同尺度下的信号特征, 引入压缩激励模块增强判别性特征的权重, 并利用 LSTM 单元建模长时依赖关系。与现有方法相比, 所提模型在保持较高识别精度的同时, 显著降低了计算复杂度与内存占用, 为资源受限场景下的实时调制识别提供了可行的技术路径。

2. 信号模型与提出的网络模型

2.1. 信号模型

信号通过信道并被采样后, 等效基带信号可以表示为:

$$y[l] = A[l]e^{j(\omega l + \varphi)}x[l] + n[l], l = 1, \dots, L \quad (1)$$

其中 $x[l]$ 是发射机以某种调制方案调制的信号, $n[l]$ 表示复加性高斯白噪声(AWGN), $A[l]$ 代表信道增益, ω 是频偏, φ 是相偏, $y[l]$ 表示接收机观察到的第 l 个值, L 是一个信号样本中的符号数。为了便于数据处理和调制识别, 接收到的信号可以以同相/正交(I/Q)形式存储为:

$$y = [\text{Re}\{y[1]\}, \dots, \text{Re}\{y[L]\}; \text{Im}\{y[1]\}, \dots, \text{Im}\{y[L]\}] \quad (2)$$

2.2. 多尺度深度可分离注意力网络

本文提出的多尺度深度可分离注意力网络(Multi-Scale Depthwise-Separable Attention Network, MDSA-Net)模型网络结构如图 1 所示。

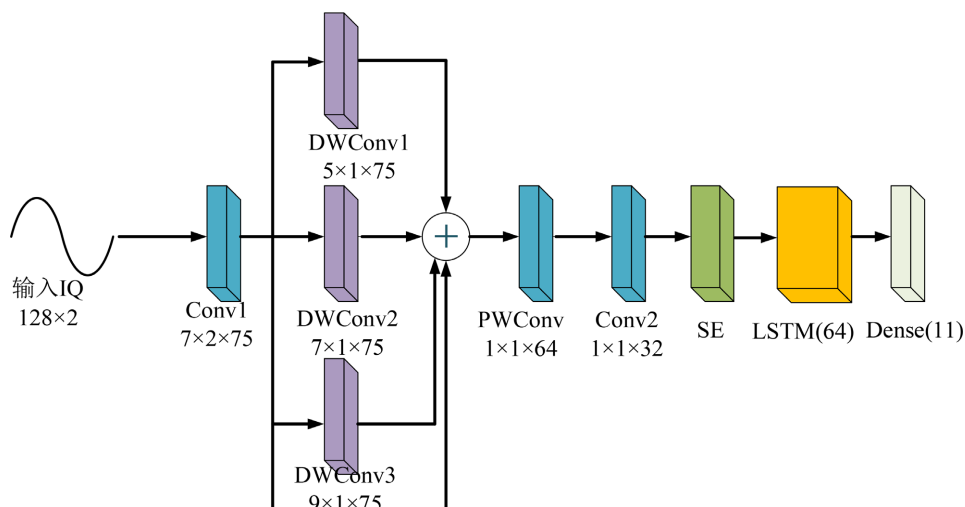


Figure 1. The structure of the proposed MDSA-Net

图 1. 提出的 MDSA-Net 结构

网络接受一个维度为(B, 128, 2)的输入张量, 其中 B 为批处理大小。首先通过一个重塑层将其转换为 (B, 128, 2, 1) 的四维张量, 以适应后续的二维卷积操作。前端特征提取部分采用一个卷积核为 7×2 的卷积层, 配备 75 个滤波器, 用于捕获信号在时间和 I/Q 维度上的局部相关性。后接批量归一化层和 ReLU 激活函数, 以加速训练收敛并引入非线性。

随后, 网络采用三个并行的深度卷积分支(DWConv1、DWConv2、DWConv3), 卷积核尺寸分别为 5

$\times 1$ 、 7×1 和 9×1 。这种多尺度设计能够从不同感受野提取信号特征，在扩大特征覆盖范围的同时，通过深度卷积的操作显著减少参数量。多尺度特征通过残差连接与原始输入相加进行融合后，经由一个 1×1 的逐点卷积(PWConv)将通道数整合为 64，自此完成多尺度深度可分离卷积。接着使用另一个 1×1 卷积进一步将通道数压缩至 32，实现特征的降维和深化。在特征精炼阶段，网络引入压缩激励(SE)注意力模块，通过全局平均池化获取通道级统计信息，再经由两个全连接层生成通道权重，对特征图进行自适应重校准，增强判别性特征的响应。SE 模块结构图如图 2 所示。

时序建模部分将空间特征重塑为序列格式，输入至包含 64 个隐藏单元的 LSTM 层中，捕捉调制信号中的长程时序依赖关系。最后通过全连接层和 Softmax 激活函数输出 11 类调制方式的概率分布。网络各层的参数配置均经过精心设计，在保证性能的同时最大限度地降低了计算复杂度。

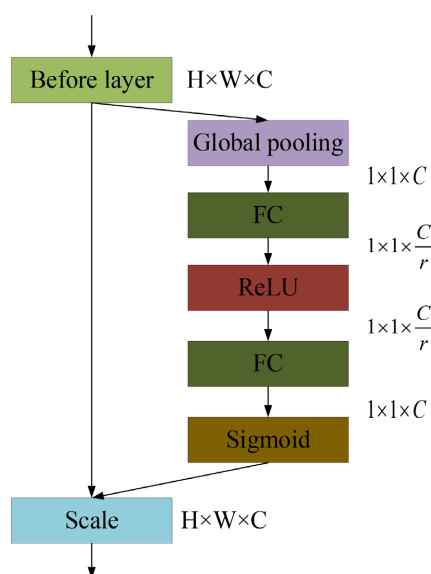


Figure 2. Structural architecture of the SE module network
图 2. SE 模块网络结构

3. 数据集与实现细节

本实验在 Windows 11 操作系统环境下开展，选用 Python 作为主要开发语言，并基于 TensorFlow 2.5.0 框架构建深度学习模型。为提升训练效率，系统配置了 CUDA 8.1 以启用 GPU 加速。实验所用硬件平台搭载 Intel(R) Core(TM) i7-14700HX 处理器及 NVIDIA GeForce RTX 4070 Laptop GPU，为模型训练提供充足算力支持。

本实验数据采用 RadioML2016.10a 公开调制信号数据集，该数据集涵盖多种常用调制类型的 IQ 信号，同时引入多径衰落、高斯白噪声及频率偏移等信道损伤因素，能够有效模拟真实通信环境。该数据集共包含 220,000 组样本，涵盖 8 种数字调制与 3 种模拟调制共 11 类信号。每个样本以 2×128 的向量形式输入网络，其中 2 表示同相与正交分量，128 代表信号序列长度。数据集覆盖从 -20 dB 至 18 dB 的信噪比范围，以 2dB 为间隔共包含 20 个不同信噪比等级。

在数据划分方面，按 6:2:2 比例将数据集随机划分为训练集、验证集与测试集。模型训练采用分类交叉熵作为损失函数，使用 Adam 优化器进行参数更新，批次大小设置为 128。训练过程中引入动态学习率调整策略：若验证集损失连续 5 个训练周期未下降，学习率衰减为原值的 0.5；若连续 50 个周期未下降，则提前终止训练并保存验证损失最小的模型参数。

4. 仿真分析

为验证本文提出的 MDSA-Net 信号调制方式识别模型在不同信噪比条件下的调制信号的识别性能, 选用 MCLDNN、CNN [8]、LSTM [9]、PET-CGDNN 和 IC-AMCNet [10] 作为对比模型。所有模型在相同设置下进行训练与测试, 各模型在 RadioML2016.10a 数据集上的不同信噪比条件下的识别准确度对比结果如图 3 所示。

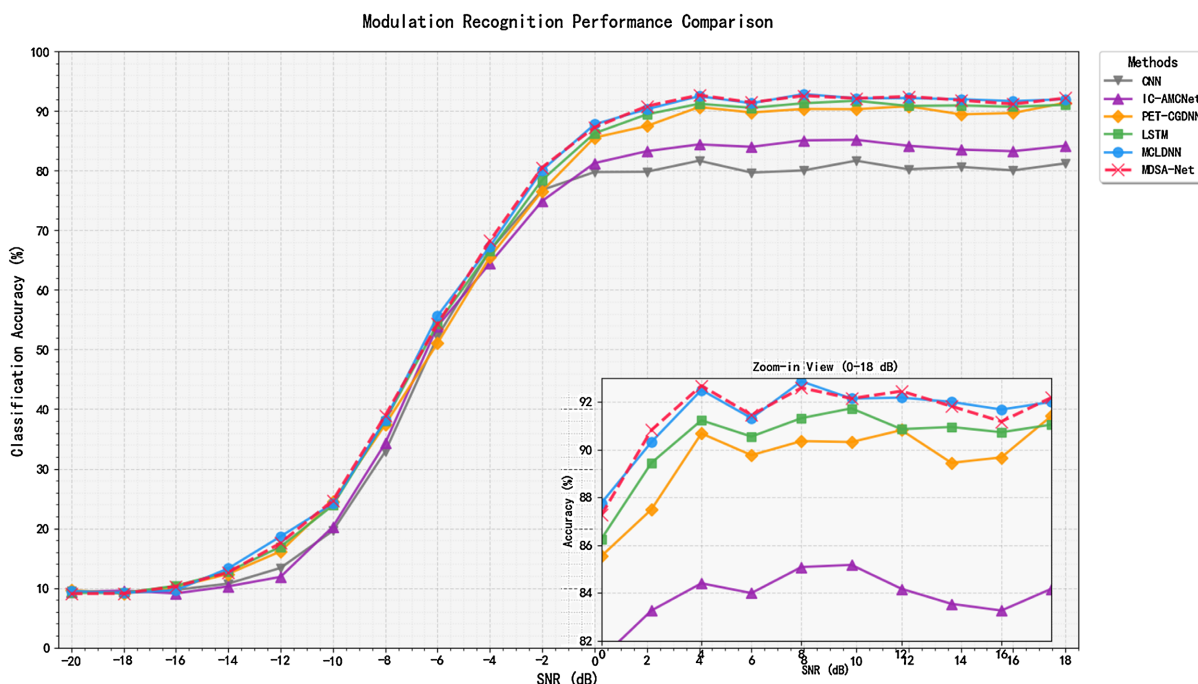


Figure 3. Comparison chart of accuracy rates of different networks

图 3. 不同网络准确率对比图

从图 3 能够看出本文提出的 MDSA-Net 调制方式识别模型和另外 5 种调制方式识别模型对信号调制方式进行分类所得的准确率随着信噪比的上升而上升, 其中, 在低信噪比区间(低于 -8 dB), 所有模型的分类型准确率都显著受到了噪声的影响, 表现普遍偏低且增长趋势一致。在此区间, CNN 与 IC-AMCNET 模型的分类型准确率低于其他模型。在 -8 dB 至 0 dB 信噪区间, 所有模型的分类型准确率都保持较高的增长速度。信噪比超过 4 dB 后, 各网络分类型准确率趋于稳定且性能差异开始显著, 本文提出的网络模型、MCLDNN、PET-CGDNN 和 LSTM 保持的较高准确率, 本文所提的网络模型与 MCLDNN 分类型准确率相近, 分类型准确率在 92% 左右, LSTM 的分类型准确率在 91% 左右, PET-CGDNN 的分类型准确率在 90% 左右, 而 CNN 与 IC-AMCNET 的分类型准确率均低于 90%, 分别为 80% 左右和 86% 左右。

在表 1 比较了本文提出的轻量级神经网络 MDSA-Net 与另外五个网络的参数量、计算量、单条样本测试时间和迭代次数。参数量反映模型的存储开销和表示复杂度, 计算量反映每次前向传播的运算负担, 单条样本测试时间体现推理延迟, 迭代次数反映训练过程中需要的更新轮次或收敛迭代水平。这四项指标合起来能够比较直观地评估模型的复杂度、训练开销和推理延迟。从表 1 所示的模型复杂度分析结果可以看出, 本文提出的 MDSA-Net 在模型效率方面展现出多方面的优势。在模型复杂度方面, MDSA-Net 的参数量仅为 36,896, 显著低于其他对比模型, 比参数量次优的 PET-CGDNN 减少了约 48.7%, 体现了极致的参数效率。在计算复杂度方面, MDSA-Net 的 8.37 M 计算量同样表现优异, 远低于 MCLDNN 的

97.35 M 和 CNN 的 157.73 M，显示出其在计算资源使用方面的高效性。在训练效率方面，MDSA-Net 仅需 93 次迭代即可收敛，这一指标优于 CNN 的 106 次和 IC-AMCNET 的 140 次，体现了其良好的训练特性。虽然在测试时间方面，MDSA-Net 的 6.07 毫秒每样本并非最优，但仍优于 MCLDNN 的 8.02 毫秒，且在可接受范围内。这些综合指标表明，MDSA-Net 在模型复杂度、计算效率和训练速度等方面取得了良好的平衡，特别适合在资源受限的环境中部署应用。

Table 1. Comparison of complexity among different networks
表 1. 不同网络的复杂度对比

网络模型	参数量	计算量	测试时间(ms/样本)	迭代次数
MDSA-Net	36,896	8.37 M	6.07	93
MCLDNN	405,175	97.35 M	8.02	91
CNN	858,123	157.73 M	1.61	106
LSTM	200,075	50.83 M	5.64	84
PET-CGDNN	71,871	16.39 M	4.95	86
IC-AMCNET	1,264,01	29.69 M	1.55	140

综上所述，虽然 MDSA-Net 的准确率与最优的 MCLDNN 相当，但其更低的参数量、计算量和测试时间使其在实践中更具优势，尤其适合需要高效计算和快速响应的应用场景。作为一个轻量级的模型，MDSA-Net 尤其适合部署在资源受限的设备或边缘计算设备上，能够在这些平台上提供高效的计算能力。MDSA-Net 在准确率和计算效率之间达到了良好的平衡，是一种在实际应用中具有高性价比的轻量级深度学习模型。

为验证本文提出的多尺度深度可分离卷积分支设计的合理性，本小节通过系统地移除或修改网络中模块，定量分析各组件对最终性能的贡献。本文选择 5×1 、 7×1 、 9×1 三种卷积核尺寸构建三支结构，其设计基于如下考虑：在调制识别任务中，不同尺寸的卷积核对应于不同的感受野，能够捕获信号在不同时间尺度上的特征。较小的 5×1 卷积核适于捕捉信号的局部细节和快速瞬变特征(如相位突变);中等的 7×1 卷积核能够覆盖更典型的符号周期，提取更具代表性的时域波形结构；而较大的 9×1 卷积核则有助于建模更长程的依赖关系，识别如连续相位调制中的渐变趋势。选择三个分支旨在模型复杂度和特征丰富度之间取得平衡：分支过少可能无法充分覆盖特征尺度，而分支过多则会引入不必要的参数冗余，导致精度提升的边际效益递减。为评估多尺度设计的合理性，设计以下消融变体：**Model-A (Baseline)**：完整的三支架构，包含所有三个不同核尺寸的 DWConv 分支。**Model-B (双分支)**：移除 9×1 的 DWConv 分支，仅保留 5×1 和 7×1 两个分支，以验证大感受野分支的贡献。**Model-C (单分支)**：仅保留中间尺寸的 7×1 DWConv 分支，移除其他两个分支，评估多尺度设计相对于单尺度的优势。

各模型在 RadioML2016.10a 数据集上的性能对比如图 4 所示，复杂度与准确率指标详见表 2。从图 4 中可以观察到，缺乏多尺度深度可分离卷积分支的 Model-C 在高信噪比(SNR 大于 0 dB)范围内的表现明显逊色于具有双分支结构的 Model-B 以及完整的 MDSA-Net 模型。相比之下，完整的 MDSA-Net 相较于双分支架构略有提升，表明大感受野分支对提高网络准确率具有一定的积极作用。

表 2 进一步说明了不同分支结构对网络性能的影响。从表中可以看出，虽然三种模型在参数量和计算量上差异不大，但 Model-A (MDSA-Net)在准确率表现上明显优于其他两个模型。Model-A 的最高准确率为 92.68%，略高于 Model-B 的 92.36%和 Model-C 的 91.68%。在 0~18 dB 范围内，Model-A 的平均准

确率为 91.46%，也优于 Model-B 的 91.33%和 Model-C 的 90.40%，显示出它在高信噪比条件下的优势。整体而言，三种模型的参数量和计算复杂度差异较小，但 Model-A 凭借其在准确率上的优势，尤其在高信噪比环境下，显示出最佳的整体性能。这也说明了本文所提出的多尺度深度可分离卷积分支设计能够在保持低复杂度的同时，使模型获得较高的准确率性能。

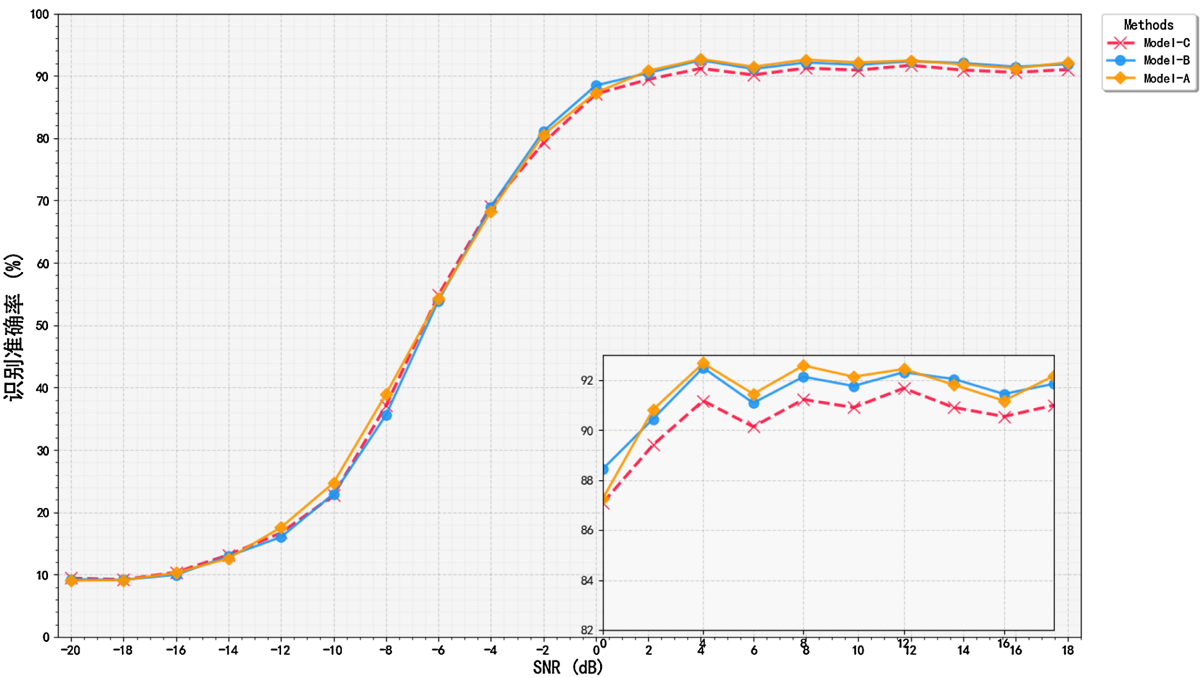


Figure 4. Comparison chart of ablation experiments
图 4. 消融实验对比图

Table 2. Comparison of complexity among different networks
表 2. 不同网络的复杂度对比

网络模型	参数量	计算量(FLOPS)	最高准确率	0~18 dB 平均准确率	平均准确率
Model-A (MDSA-Net)	36,896	8.37 M	92.68%	91.46%	61.99%
Model-B (双分支)	35,921	8.21 M	92.36%	91.33%	61.67%
Model-C (单分支)	35,246	8.11 M	91.68%	90.40%	61.29%

5. 总结

本文针对对现有自动调制识别模型在资源受限场景下部署困难的问题，提出了一种轻量级多尺度特征融合网络(MDSA-Net)。该网络融合了多尺度深度可分离卷积、通道注意力机制与 LSTM 时序建模模块，在保证识别精度的同时，显著降低了模型的参数量与计算复杂度。通过在公开数据集 RadioML2016.10a 上的实验验证，所提模型在不同信噪比条件下均表现出优异的识别性能，尤其在模型效率方面具有明显优势：参数量仅为 36,896，计算量为 8.37 M，远低于传统复杂网络，同时具备较快的收敛速度与可接受的推理延迟。MDSA-Net 在准确率与计算效率之间取得了良好平衡，尤其适用于边缘计算设备与物联网终端等资源受限环境中的实时调制识别任务。未来的研究工作将进一步探索模型的量化与剪枝优化，以进一步提升其在实际系统中的部署能力。

参考文献

- [1] 濮希同. 面向调制识别的深度学习模型轻量化关键技术研究[D]: [硕士学位论文]. 成都: 电子科技大学, 2025.
- [2] O'Shea, T.J., Corgan, J. and Clancy, T.C. (2016) Convolutional Radio Modulation Recognition Networks. In: Jayne, C. and Iliadis, L., Eds., *Engineering Applications of Neural Networks*, Springer International Publishing, 213-226. https://doi.org/10.1007/978-3-319-44188-7_16
- [3] Xu, J., Luo, C., Parr, G. and Luo, Y. (2020) A Spatiotemporal Multi-Channel Learning Framework for Automatic Modulation Recognition. *IEEE Wireless Communications Letters*, **9**, 1629-1632. <https://doi.org/10.1109/lwc.2020.2999453>
- [4] Liu, K.T. and Wu, S. (2024) Automatic Modulation Classification Based on Time-Attention Mechanism and LSTM Neural Networks. 2024 7th International Conference on Information Communication and Signal Processing (ICICSP), Zhoushan, 21-23 September 2024, 112-117. <https://doi.org/10.1109/icicsp62589.2024.10809243>
- [5] 张富鑫. 基于深度学习的自动调制识别技术研究[D]: [硕士学位论文]. 成都: 电子科技大学, 2023.
- [6] Zhu, Z., Zhou, N., Wang, Z., Liang, J. and Wang, Z. (2024) MSGNet: A Multi-Feature Lightweight Learning Network for Automatic Modulation Recognition. *IEEE Communications Letters*, **28**, 2553-2557. <https://doi.org/10.1109/lcomm.2024.3414617>
- [7] Zeng, D., Xiao, Y., Liu, W., Du, H., Zhang, E., Zhang, D., et al. (2025) Efficient Automatic Modulation Classification in Nonterrestrial Networks with Snn-Based Transformer. *IEEE Internet of Things Journal*, **12**, 2562-2573. <https://doi.org/10.1109/jiot.2024.3467284>
- [8] Tekbiyik, K., Ekti, A.R., Gorcin, A., Kurt, G.K. and Kececi, C. (2020) Robust and Fast Automatic Modulation Classification with CNN under Multipath Fading Channels. 2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring), Antwerp, 25-28 May 2020, 1-6. <https://doi.org/10.1109/vtc2020-spring48590.2020.9128408>
- [9] Rajendran, S., Meert, W., Giustiniano, D., Lenders, V. and Pollin, S. (2018) Deep Learning Models for Wireless Signal Classification with Distributed Low-Cost Spectrum Sensors. *IEEE Transactions on Cognitive Communications and Networking*, **4**, 433-445. <https://doi.org/10.1109/tccn.2018.2835460>
- [10] Hermawan, A.P., Ginanjar, R.R., Kim, D. and Lee, J. (2020) CNN-Based Automatic Modulation Classification for beyond 5G Communications. *IEEE Communications Letters*, **24**, 1038-1041. <https://doi.org/10.1109/lcomm.2020.2970922>