

基于 Foster 的调制识别增量学习方法

刘浩楠¹, 谢军^{1,2}, 薛飞^{1,2}

(1. 中国电子科技集团公司第五十四研究所, 石家庄 050081;

2. 河北省电磁频谱认知与管控重点实验室, 石家庄 050081)

摘要: 针对新的调制识别信号在通信场景中动态出现的问题, 提出了一种基于 Foster 的增量学习调制识别算法, 该算法采用基于通道共享阈值机制的深度残差收缩网络, 通过自适应软阈值化机制实现噪声抑制并提取有效特征; 结合一种两阶段学习范式动态扩展新模块来适应新类别, 通过残差拟合模块动态扩展网络容量以适应新调制类型, 为解决训练过程中参数不断增多导致参数爆炸的问题, 通过一种知识蒸馏策略减少特征维度和冗余参数, 保持模型的主干部分; 实验结果表明该方法性能明显优于其他增量学习算法, 能够有效解决动态环境下的自动调制识别, 展现出良好的应用价值。

关键词: 类别增量学习; 深度学习模型; 特征增强; 知识蒸馏; 调制识别

Method for Incremental Learning Based on Foster Modulation Recognition

LIU Haonan¹, XIE Jun^{1,2}, XUE Fei^{1,2}

(1. The 54th Research Institute of CETC, Shijiazhuang 050081, China;

2. Key Laboratory of Electromagnetic Spectrum Cognition and Control in Hebei Province,
Shijiazhuang 050081, China)

Abstract: A modulation recognition algorithm based on the Foster incremental learning is proposed to address the challenge of dynamically emerging novel modulation signals in communication scenarios. The algorithm employs a deep residual shrinkage network incorporating a channel-shared threshold mechanism, which suppresses noise while extracting discriminative features by an adaptive soft-thresholding mechanism. Furthermore, the framework integrates a dual-phase learning paradigm to dynamically incorporate new modules, thereby accommodating novel modulation categories. The residual fitting module dynamically expands the network capacity to adapt to new modulation types. To solve the issue of parameter explosion caused by the increase of parameters during the training process, a knowledge distillation strategy is used to reduce feature dimensionality and redundant parameters while maintaining the model backbone. Experimental results demonstrate that this method is superior in performance to other incremental learning algorithms. It can effectively achieve automatic modulation recognition in dynamic environments, with a significant application value.

Keywords: class incremental learning; deep learning; feature enhancement; knowledge distillation; modulation recognition

0 引言

在信息化时代迅猛发展的背景下, 通信信号侦查的重要性日益凸显, 信号调制与解调作为通信领域核心技术之一, 其正确识别直接关系到通信质量和效率, 在民用与军用场景中均发挥着关键作用。因此在复杂环境中实现高效调制识别已成为推动通信技术发展与电子战装备升级亟待解决的核心问题。

调制识别方法主要分为三类: 基于似然函数的方

法^[1]、基于特征提取的方法^[2-3]和基于深度学习的方法^[4]。基于似然函数的方法需要先验知识构造似然函数, 通过对概率密度函数进行积分才能计算似然值进行分类。随着信号调制方式日益复杂, 基于似然函数的方法需要进行大量计算与存储, 计算复杂度高。基于特征提取的方法分为3个阶段, 参数初始化、数据清洗和归一化处理, 要求人工从信号中提取特征, 如瞬时特征^[5], 高阶统计量^[6-7]、星座图特征^[8]等。特征的提取需要依赖领域专家的经验, 不能适应动态变化的信道条

收稿日期:2025-05-09; 修回日期:2025-06-18。

基金项目:国家自然科学基金(U22B2002)

作者简介:刘浩楠(2001-),男,硕士研究生。

引用格式:刘浩楠,谢军,薛飞.基于 Foster 的调制识别增量学习方法[J].计算机测量与控制,2025,33(11):316-323.

件。而且传统特征的鲁棒性不足, 在有限样本条件下, 人工特征的可区分性会受限, 导致模型泛化能力下降。总而言之, 基于似然函数的方法和基于特征提取的方法依赖人工经验进行调制方式识别。这不仅要求操作人员具有一定专业素养, 也难以保证识别结果的客观性与稳定性, 导致存在较大的误差。

文献 [9] 在 2016 年将卷积神经网络引入到调制识别领域, 实现了 0 dB 以上时能够以大于 70% 的准确率识别 11 种调制信号, 深度学习技术得以被广泛认可。在文献 [10] 中, 研究者采用了卷积神经网络和短时傅里叶变换结合的方法, 使分类信号的准确率进一步提高。文献 [11] 中, 研究者运用长短期记忆网络 (LSTM) 处理 I/Q 数据, 并将其转换为幅度和相位信息。与直接输入 I/Q 数据相比, 这种预处理方法取得了更好的效果, 强调了数据预处理在自动调制识别中的关键作用。文献 [12] 则采用了卷积长短时记忆网络 (ConvLSTM) 进行分类识别, 该模型融合了 CNN 和 LSTM 的优势, 能够处理原始时域波形, 实现了更高效的调制识别。文献 [13] 对文献 [9] 提出的 CNN 模型进行了改进, 设计了一个包含 4 个卷积层和 2 个全连接层的结构。这个改进的模型不仅降低了算法的复杂性, 还缩短了训练时间, 表现出了优异的性能。

然而上述方法不适用于实际通信环境下的信号调制识别任务, 由于数据隐私和高昂的训练成本, 每次出现新类别时重新训练模型是不现实的, 导致模型缺乏对新型调制方式的适应能力。因此, 有必要使模型能够持续学习新类别, 即增量学习。传统的增量学习算法主要有基于回放的 Incremental Classifier and Representation Learning (iCaRL) 算法^[14] 和基于正则化的 Elastic Weight Consolidation (EWC) 算法^[15]。iCaRL 算法存

储一部分旧任务的数据样本, 当学习新任务时, 将这些旧样本与新样本混合训练, 使模型在学习新知识的同时复习旧知识, 从而缓解灾难性遗忘, 但需要的存储空间要求高。EWC 算法约束重要参数变化, 防止旧任务的知识被覆盖, 但当新旧任务差异较大时, EWC 的约束可能导致模型无法充分学习新任务。

针对以上问题以及在动态实际环境中学习新类型调制识别信号的需求, 本文提出一种基于 Feature Boosting and Compression for Class Incremental Learning (Foster)^[16] 的增量学习方法。首先采用基于通道共享阈值机制的深度残差收缩网络获得有效特征, 然后通过特征增强和特征压缩, 也就是一种两阶段的学习机制, 模型能够持续学习新型调制信号, 更好地适应复杂多变的动态环境。

1 系统框架及原理

1.1 基于增量学习的信号调制识别框架

增量学习是一种机器学习范式, 旨在模拟人类持续学习新知识并保留已有知识的能力。与传统的一次性学习不同, 增量学习模型在训练过程中不是一次性训练所有数据, 而是在模型已经具备一定知识的基础上, 逐步接收新的数据, 并在此基础上更新模型。这种方法使得模型能够适应不断变化的环境和数据分布, 并且可以应对数据无法一次性获取或者存储的场景。

如图 1 所示, 采用通道共享阈值的深度残差收缩网络 (DRSN-CS)^[17] 作为特征提取器对信号进行特征提取。在网络训练的过程中, 新增的调制识别信号被视为独立的学习任务。当任务 t 到来时, 特征提取器 F_t 通过一种两阶段学习范式的增量学习方法实现特征空间的渐进式扩展, 其中特征提取器是由新的特征提取器扩展而得。同时采用特征向量融合策略将新旧类别特征进行组合, 并输入至分类器 W_t 完成预测识别。

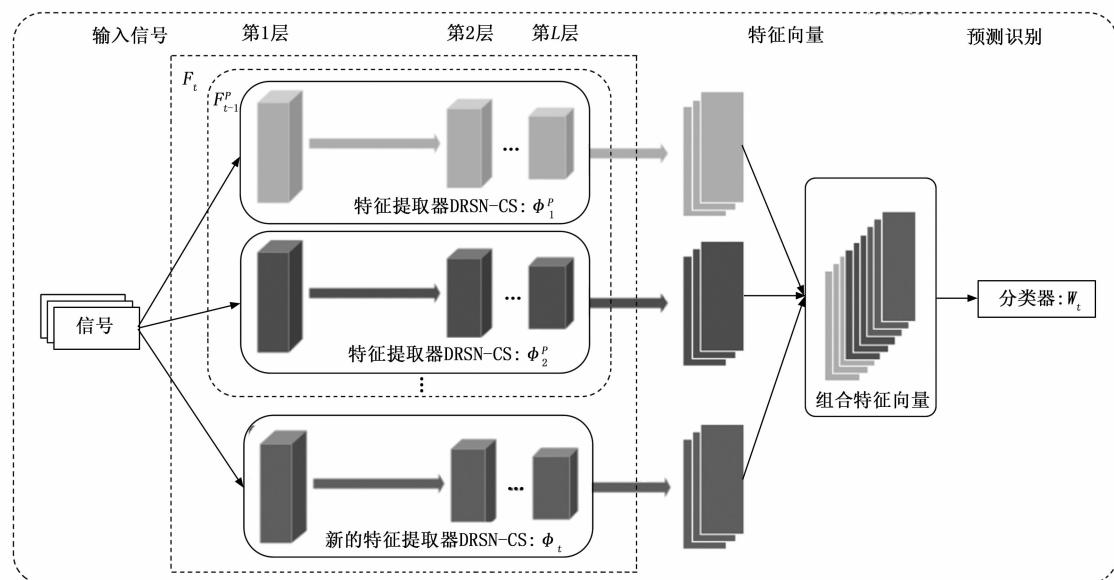


图 1 基于增量学习的调制识别框架

1.2 软阈值化

软阈值化^[18]作为信号去噪领域的核心处理技术，在诸多领域得到广泛应用。其基本流程是将原始信号中的噪声及无关紧要的部分，也就是接近零的特征置零。以经典的小波阈值去噪为例该方法的性能高度依赖于滤波器设计。但是，滤波器的设计始终是一个较大的技术挑战，要求相关人员具备深厚的信号处理专业知识。随着深度学习技术的发展，通过其自动的特征学习能力，为解决这一问题提供了新的思路。基于数据驱动的深度学习框架能够通过梯度下降算法自动优化滤波器参数，避免了传统人工设计的局限性。软阈值化的函数形式可表示为：

$$y = \begin{cases} x - \tau & x > \tau \\ 0 & -\tau < x < \tau \\ x + \tau & x < -\tau \end{cases} \quad (1)$$

其中： x 是输入特征， y 是输出特征， τ 是阈值，即一个正参数。与在 ReLU 激活函数中将负特征设为零不同，软阈值化将接近零的特征设为零，从而保留有用的负特征。因此，将软阈值化与深度学习相结合，有望构建更具判别性的特征提取方法，实现更高效的噪声抑制。

软阈值化的过程如图 2 (a), (b) 所示。软阈值的导数只有 1 和 0 两种情况，能够有效地避免参数爆炸和梯度消失的问题。

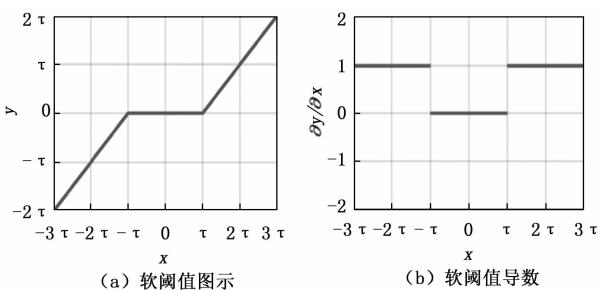


图 2 软阈值及其导数

针对传统信号去噪算法中阈值参数需要人工经验设定的局限性，DRSN-CS 提出了一种自适应阈值学习机制。具体而言，该模型的阈值参数能够根据输入信号的统计特性进行动态调整，从而避免了传统方法中因固定阈值设置而导致的性能下降问题。不仅解决了最优阈值因信号特征不同而变化的难题，还显著降低了算法对人工调参的依赖性。

1.3 特征提取器

卷积层的本质是微分思想的离散化实现。不同于全连接层的全局密集计算，卷积层用滑动窗口（卷积核）在输入数据上逐区域扫描，通过加权求和生成特征图。例如， 3×3 的卷积核仅需 9 个参数，而全连接层处理相同尺寸的输入时参数量呈指数级增长。参数量的减少

不仅提升了计算效率，更有效缓解了过拟合问题。卷积操作支持多通道输入，并且通过多层堆叠卷积操作，网络能够逐级提取从低级特征到高级语义特征，形成层次化表征，最终生成具有强表达能力的特征图。

如图 3 所示，在卷积神经网络中，特征图和卷积核本质上是三维张量结构。如图 4 所示，卷积核通过参数化的局部滤波操作在输入特征张量上进行离散卷积运算，从而生成输出特征图的一个特征通道。而每个卷积层通常由多个卷积核构成，这使得输出特征图能够形成多个通道。每个卷积核独立地在输入数据上进行操作，提取不同的特征模式，从而在输出特征图的不同通道中捕捉到输入数据的多样化特征信息。

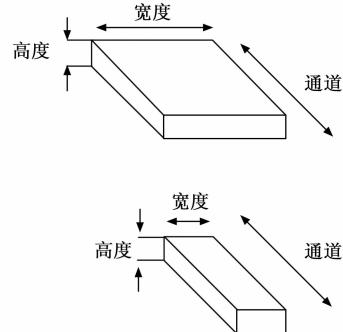


图 3 特征图和卷积核

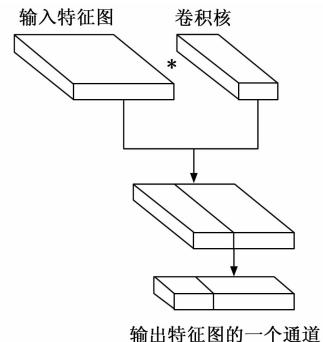


图 4 卷积过程

深度残差网络（ResNets）^[19]是近年来兴起并备受关注的深度学习方法。残差构建单元（RBUs, residual building units）是其基本组成部分。一个 RBU 由两个 BN 层、两个 ReLU 激活函数、两个卷积层和一个恒等快捷连接（identity shortcut）组成。恒等快捷连接是使 ResNet 优于一般卷积网络的部分。通过使用恒等快捷连接，梯度可以有效地流向靠近输入层的早期层，从而更有效地更新参数。

实际通信场景中，信号信噪比往往呈现动态波动特性，这种不稳定性导致传统调制识别方法在低信噪比条件下准确率显著下降。为提高模型在噪声环境下的鲁棒性，本研究引入 DRSN-CS^[20]模块，该模块以 ResNet

的整体架构为基础, 它由一个输入层、一个卷积层、多个 RBSU、一个 BN 层、一个 ReLU 层、一个 GAP 层和一个输出全连接 (FC) 层组成, 并通过自适应软阈值化机制实现噪声抑制并提取有效特征。其中, 软阈值化作为一个非线性变换层被嵌入到构建单元中。如图 5 (a) 所示, 通道共享阈值的残差收缩模块 (RSBU-CS) 包含一个专门的阈值估计模块。该模块首先对特征图 x 的绝对值执行全局平均池化操作, 生成一维特征向量。随后, 通过一个两层全连接网络, 得到一个缩放参数, 其结构设计与文献 [21] 相似。在网络末端采用 Sigmoid 激活函数将输出缩放参数归一化至 (0, 1) 区间, 数学表达式为:

$$\alpha = \frac{1}{1 + e^{-z}} \quad (2)$$

其中: z 表示 RSBU-CS 中两层全连接网络的输出, α 为对应的归一化缩放参数。缩放参数 α 与 $|x|$ 的均值相乘后得到最终阈值。这种设计基于两个关键考量: 首先, 软阈值化要求阈值必须为正数; 其次, 阈值应合理约束, 若超过特征图最大绝对值将导致输出全零。因此, RSBU-CS 中的阈值计算可表示为:

$$\tau = \alpha \cdot \text{average}_{i,j,c}(|x_{i,j,c}|) \quad (3)$$

其中: τ 是阈值, i 、 j 和 c 分别是特征图 x 的宽度、高度和通道索引。通过这种方式, 阈值可以保持在合理范围内, 从而避免软阈值化的输出全部为零。

DRSN-CS 架构如图 5 (b) 所示, 其核心构建模块为 RSBU-CS。该网络通过堆叠多个 RSBU-CS 模块实现噪声特征的逐层抑制。相较于传统方法, DRSN-CS 的优势在于, 阈值参数通过深度网络自动学习获得, 避免了人工经验设定的主观性, 无需依赖信号处理领域的先验知识。

如表 1 所示 DRSN-CS 架构的超参数, 包括 1 层卷积层和 8 层残差收缩模块。每层超参数为卷积核数量、

卷积核大小、卷积步长。

表 1 DRSN-CS 架构超参数

输出大小	DRSN-CS
$2 \times 1024 \times 1$	输入
$64 \times 1024 \times 1$	Conv(64, 3, 1)
$64 \times 1024 \times 1$	RSBU-CS(64, 3, 1)
$64 \times 1024 \times 1$	RSBU-CS(64, 3, 1)
$128 \times 512 \times 1$	RSBU-CS(128, 3, 2)
$128 \times 512 \times 1$	RSBU-CS(128, 3, 1)
$256 \times 256 \times 1$	RSBU-CS(256, 3, 2)
$256 \times 256 \times 1$	RSBU-CS(256, 3, 1)
$512 \times 128 \times 1$	RSBU-CS(512, 3, 2)
$512 \times 128 \times 1$	RSBU-CS(512, 3, 1)

如图 5 (b) 所示, 本文网络的输入为 2×1024 的调制信号, 经过 1 层卷积层包括 64 个卷积核, 卷积步长为 1 和 8 层残差收缩模块的逐层特征提取, 最终输出 512×128 的深层特征表示。在残差收缩模块中, 输入特征维度为 $C \times W \times 1$ (通道 \times 宽度 \times 高度); S 为卷积步长, K 为卷积核数量, M 为全连接层输出维度。该模块首先对输入进行两次连续卷积运算, 随后对输出特征取绝对值并进行全局平均池化操作, 将特征宽度 W 降维至 1。全局平均池化 (GAP) 是一种从特征图的每个通道计算平均值的操作。通常, 它用于最后的输出层之前。GAP 可以减少最终全连接输出层中使用的权重数量, 从而降低深度学习网络过拟合的可能性。接着通过两个全连接层和处理得到中间变量 z , 经 Sigmoid 后获得归一化参数 α 。归一化 (BN) 的目的是减少内部协变量偏移, 即特征的分布在训练迭代过程中会不断变化。最终, 将 α 与全局平均池化输出的均值相乘得到自适应软阈值, 并与原始输入特征进行阈值化处理。

2 Foster 增量学习算法

Foster 增量学习算法从梯度提升视角提出了一种分

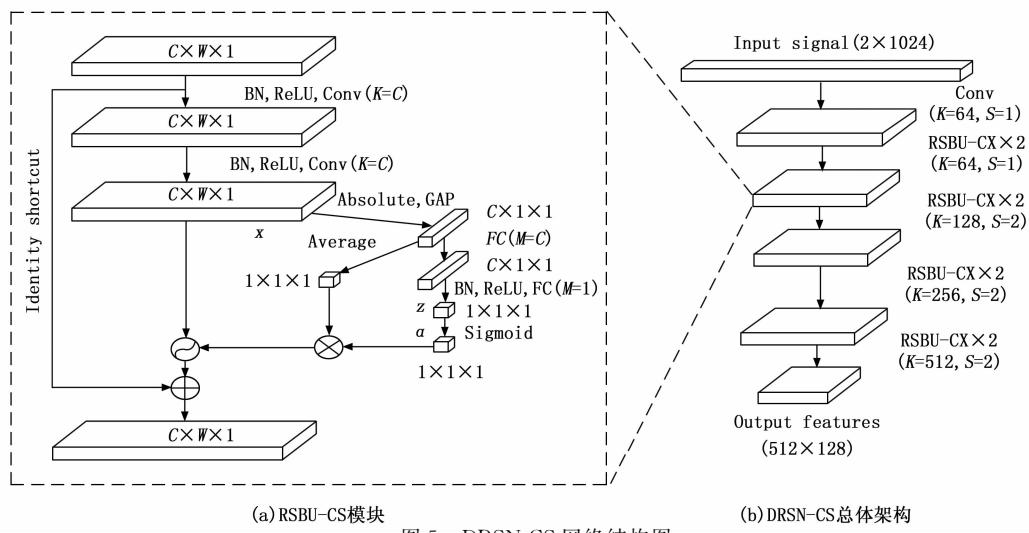


图 5 DRSN-CS 网络结构图

析和实现增量学习目标的新视角。梯度提升借助加法模型逐步逼近真实目标模型，后续模型专注于拟合先前模型输出与目标之间的残差。在增量学习中，新类别的持续引入引发分布漂移，导致预测输出与真实标签间产生残差。因此，该算法采用一个类似的提升框架来解决增量学习问题，利用加法模型逐步拟合残差，不同模型分别专注于特定任务（对应不重叠的类别集）。该方法在两方面具有显著优势：一方面，新模型提升了可塑性，提升对新类别的判别能力；另一方面，通过训练新模型对所有类别分类，揭示原始模型可能忽略的关键要素。不过，这种动态扩展机制不仅会造成参数量攀升，并可能引发特征层面新旧模型的不一致。为此，我们对提升模型进行压缩，剔除冗余参数及不一致特征，既能规避基于动态结构方法的弊端，又能留存关键信息，进而增强模型鲁棒性。

总之，该范式解耦为提升（Boosting）和压缩（Compression）两步。第一步为提升，旨在缓解新类别引入导致的性能下降。具体操作为保留旧模型及其全部冻结参数，拓展新的可训练特征提取器并与旧模型的特征提取器拼接，同时初始化一个受限全连接层，将超级特征转化为逻辑值（logits）。第二步聚焦于消除特征增强带来的冗余参数和无意义维度：采用一种高效的蒸馏策略，即使在有限的新任务数据下，也能以可忽略的性能损失将提升模型的知识迁移到单一模型中。

2.1 特征增强

在第一阶段，FOSTER 动态扩展新模块以适应目标模型与原始模型输出之间的残差。通过这种方式，模型能够增强对新类别的区分能力，并发现原始模型可能忽略的重要特征。

假定在阶段 t 已经得到了前一个阶段的模型 F_{t-1} ：

$$F_{t-1}(x) = (W_{t-1})^T \Phi_{t-1}(x) \quad (4)$$

其中： $\Phi_{t-1}(\cdot) : R^D \rightarrow R^d$ ， $W_{t-1} \in R^{d \times |\hat{y}_{t-1}|}$ 。根据 gradient boosting（梯度提升）受到启发，训练一种新模型以适应 y 和 F_{t-1} 的残差。新模型 F_t 同样包含 $\Phi_{t-1}(\cdot) : R^D \rightarrow R^d$ ， $W_{t-1} \in R^{d \times |\hat{y}_t|}$ 。因此，训练过程可以被表示为：

$$F_t(x) = F_{t-1}(x) + E_{(x,y)} \in D_t[l(y, F_t(x) + F_{t-1}(x))] \quad (5)$$

和 gradient boosting 一样，令 $l(\cdot, \cdot)$ 为均方误差损失。理想情况下：

$$y = F_{t-1}(x) + F_t(x) = S\left(\begin{bmatrix} W_{t-1}^T \\ 0 \end{bmatrix} \Phi_{t-1}(x)\right) + S\left(\begin{bmatrix} (W_t^{(o)})^T \\ (W_t^{(n)})^T \end{bmatrix}\right) \quad (6)$$

S 是 softmax 操作， $O \in R^{d \times |\hat{y}_t|}$ 是零矩阵或者是 Φ_{t-1} 在 D_t 上微调优化得到的。该过程可以表示为如下优化问题：

$$\theta_t^* = \arg_{\theta} \min_{\text{Dis}} \left[y, S\left(\begin{bmatrix} (W_{t-1}^T) \\ 0 \end{bmatrix} \Phi_{t-1}(x)\right) + S\left(\begin{bmatrix} (W_t^{(o)})^T \\ (W_t^{(n)})^T \end{bmatrix} \Phi_t(x)\right) \right] \quad (7)$$

Dis 为距离指标，如欧式距离等， θ_t 为 F_t 的参数。之后，将 softmax 后求和换为求和后 softmax，同时使用 KL 散度作为衡量指标，得到目标函数如下：

$$\theta_t^* = \arg_{\theta} \min_{\text{KL}} \left(y \| S\left(\begin{bmatrix} (W_{t-1}^T) & (W_t^{(o)})^T \\ 0 & (W_t^{(n)})^T \end{bmatrix}\right) \right) \quad (8)$$

因此， F_t 可以被看作扩大的线性分类器 W_t 以及拼接后的超特征提取器 $\Phi_t(x)$ ：

$$W_t^T = [(W_{t-1})^T \quad (W_t^{(o)})^T \quad (W_t^{(n)})^T] \quad (9)$$

$$\Phi_t(x) = \begin{bmatrix} \Phi_{t-1}(x) \\ \Phi_t(x) \end{bmatrix}$$

输出：

$$W_t^T \Phi_t(x) = \begin{bmatrix} W_{t-1}^T \Phi_{t-1}(x) + (W_t^{(o)})^T \Phi_t(x) \\ (W_t^{(n)})^T \Phi_t(x) \end{bmatrix} \quad (10)$$

在面对新任务训练时，类别不平衡的训练集是一个常见的问题，它会导致模型的预测出现偏差。具体来说，当训练集中某些类别的样本数量远多于其他类别时，模型往往会在预测时更倾向于数量较多的类别。这种现象被称为类别不平衡（class imbalance），它会导致模型的性能在少数类别上显著下降，因为模型没有足够的样本来学习这些类别的特征。

同时，特征增强策略在处理新任务时也面临挑战。当训练集类别不平衡时，特征增强倾向于忽略样本数量较少的类别的残差。这是因为模型更容易受到多数类的影响，而少数类的残差在学习过程中可能被忽视。这会导致模型在新任务中对少数类别的识别能力下降，从而影响整体性能。FOSTER 算法采用了 Logits Alignment 技术：

$$\gamma W_t^T \Phi_t(x) = \begin{bmatrix} \gamma_1 (W_{t-1})^T \Phi_{t-1}(x) + (W_t^{(o)})^T \Phi_t(x) \\ \gamma_2 (W_t^{(n)})^T \Phi_t(x) \end{bmatrix} \quad (11)$$

其中： $0 < \gamma_1 < 1$ ， $\gamma_2 > 1$ ，且 γ 为由 γ_1 和 γ_2 构成的对角矩阵。通过该缩放策略，旧类别的 logits 绝对值被压缩，而新类别的 logits 绝对值被放大，从而迫使模型 F 为旧类别生成更大的 logits 值，为新类别生成更小的 logits 值。

通过各类别的归一化有效数量 E 来确定缩放因子 γ_1 和 γ_2 。该有效数量可视为等比数列的求和结果，其中 n 表示实例数量， β 为可调节的超参数：

$$E_n = \begin{cases} \frac{1-\beta_n}{1-\beta}, & \beta \in [0, 1) \\ n, & \beta = 1 \end{cases} \quad (12)$$

$$\gamma_1, \gamma_2 = \left(\frac{E_{n_{old}}}{E_{n_{old}} + E_{n_{new}}}, \frac{E_{n_{new}}}{E_{n_{old}} + E_{n_{new}}} \right) \quad (13)$$

因此目标函数可表述为:

$$L_{LA} = KL(y \parallel S(\gamma W_t^T \Phi_t(x))) \quad (14)$$

仅仅让 $F_t(x)$ 去适应残差通常是不够的。考虑极限情况: 残差为 0, 则 $F_t(x)$ 不能学到有关旧类别的信息, 因此需要对新模型进行提示, 整个过程分为两步。首先初始化一个新的线性分类器 $W_t^{(a)} \in R^{d \times |\hat{y}_t|}$, 将新的特征 $\Phi_t(x)$ 变换为所有见过的类别的概率, 使新特征能够用于分类:

$$L_{FE} = KL(y \parallel S(W_t^{(a)})^T \Phi_t(x)) \quad (15)$$

同时, 使用 one-hot 编码向量训练新的分类器可能会导致模型对新类别过拟合, 为此使用知识蒸馏:

$$L_{KD} = KL(S(F_{t-1}(x)) \parallel S(F_{t-1}^{(x)} + W_t^{(a)})^T \Phi_t(x))) \quad (16)$$

总的损失为:

$$L_{\text{Boosting}} = L_{LA} + L_{FE} + L_{KD} \quad (17)$$

2.2 特征压缩

第二阶段旨在通过有效的知识蒸馏策略消除冗余参数和特征维度, 以维持单一的主干模型。这一步骤有助于避免动态结构方法中参数数量和特征维度的爆炸性增长, 同时保持模型的关键信息和鲁棒性。每一个任务都增加一个网络会带来极大的空间开销, 因此, 作者考虑对模型进行蒸馏, 将 F_t 的特征空间进行压缩, 同时不会带来明显的性能下降。损失函数:

$$L_{BKD} = KL(\omega \otimes S(F_t(x)) \parallel S(F_t^{(s)}(x))) \quad (18)$$

其中: \otimes 为 tensor 的点乘 (自动广播), ω 是权重向量。

3 实验与结果分析

3.1 实验平台与数据集

实验使用开源自动调制识别数据集 Augmod^[22]。该数据集仅使用正 SNR 数据 (已验证全范围结果一致), 并在此数据集基础上添低信噪比数据。通过基带等效模型模拟接收信号, 可调节现有数据集中固定或未知的参数, 从而研究分类器对参数的鲁棒性。数据集包含 7 种线性调制, SNR 在 $-2 \sim 40$ dB 之间, 共生成 210 000 个样本 (每个 SNR-调制组合 5 000 个)。

每个样本为 1 024 点 I/Q 信号。在基线数据集基础上添加了传播信道和解调器引入的多类损伤: 随机相位偏移、时延、滚降、采样率变化和高斯噪声。具体参数范围如表 2 所示。

本次实验是在 windows11 环境下使用 python3.8 编程语言和 pytorch2.1.0 的深度学习框架搭建的神经网络模型, 开发环境为 pycharm。硬件环境 CPU 使用 Intel Core i7-14700K, GPU 使用 NVIDIA V100, 内存大小为 16 G。

表 2 Augmod 数据集参数

参数	描述
调制样式	BPSK、8 PSK、16 QAM、32 QAM、64 QAM、8 QAM、QPSK
样本数	210000
相位	$[0, 2\pi]$
时延	$[0, 1]$
信噪比	$\{-2, 0, 10, 20, 30, 40\}$
相对频率偏移	$\pm [10^{-6}, 5 \times 10^{-1}]$

3.2 评价指标

本文实验根据网络模型在测试集上每个任务的分类效果作为评价标准:

$$Acc = \frac{\sum_{i=0}^{(N-1)} [f(x_i) == y_i]}{N} \quad (19)$$

式中, Acc 为分类的精确度, N 为总数据集, y_i 为样本 x_i 的种类, $f(x_i)$ 为样本 x_i 的预测种类。在增量学习算法评估中, 所有任务分类准确率的平均值是衡量算法性能的关键指标:

$$Acc_{avg} = \frac{\sum_{i=0}^{M-1} Acc_i}{M} \quad (20)$$

式中, Acc_{avg} 为所有任务分类准确率的均值, Acc_i 为第 i 个任务的分类准确率, M 为任务的总数。

3.3 模型训练和评估

在进行增量学习之前, 对数据集中的所有调制识别信号进行训练和识别, 对不同特征提取器的调制识别性能进行对比。基于训练结果对超参数进行设定, 将学习率确定为 0.01, 批量大小设为 128, 训练周期设为 100 个, 遵循训练集: 验证集: 测试集 = 6 : 2 : 2 的比例划分, 对各信号进行均方根归一化处理, 使其功率归一化至 1。采用 CNN、LSTM、Transformer、CLDNN 和 DRSN-CS 模型对数据集进行训练, 并在测试集上进行检验, 如图 6 所示为识别结果。

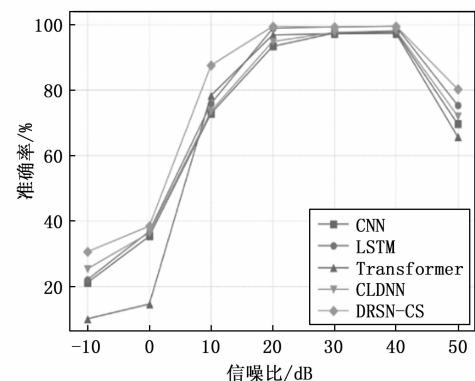


图 6 不同特征提取器的调制识别准确率

由图 6 可知, 通过对不同的特征提取器对相同的数据集进行训练结果, DRSN-CS 模型整体识别率达

到 80.02%，相较于 CNN、CLDNN、Transformer、LSTM 各提升了 9.46%、7.52%、14.15%、4.96%。Transformer 的整体准确率最差，原因主要是其在 0 dB 时识别效果差；LSTM 由于引入了门控机制，整体准确率较好。当信噪比在 10 dB 以下时，DRSN-CS 模型凭借其独特的架构优势和深层特征提取能力较各种模型分别提升了 10.64%、8.23%、17.84%、7.57%。因此 DRSN-CS 模型作为增量学习的特征提取器，以实现对新类别信号的学习。

为进行增量学习，将调制识别信号按照类型划分为 3 个渐进式学习的任务。任务 1 包含 32 QAM、16 QAM、8 QAM；任务 2 包含 BPSK、8 PSK；任务 3 包含 QPSK、64 QAM。特征增强阶段，批量大小设为 128，采用余弦退火学习率调度，初始学习率确定为 0.01，逐步衰减以稳定训练过程。特征压缩阶段，使用较小的批量 64，采用较低的学习率进行知识蒸馏，避免过拟合并保留旧知识。本文采用了 6 种方法对划分后的任务序列进行增量学习，包括微调法、LwF、iCaRL、EW_ACE、联合学习以及本文所提方法。实验结果表明如图 7 所示。

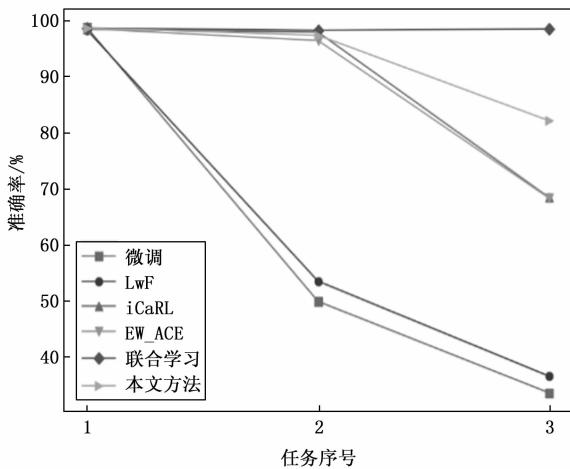


图 7 不同增量算法下调制识别准确率

各方法在连续任务学习中的准确率呈现显著差异，具体性能指标对比如表 3 所示。

根据表 3 和图 7 所示的实验数据，联合学习方法在算法性能上表现最为突出。该方法使用完整数据集进行训练，为调制识别模型提供了充分的样本支撑，能够提取数据的深层特征，从而实现最高的调制识别准确率。但是，正因为使用的是完整数据集导致需要过多的计算资源，在计算复杂度和时间消耗方面显著高于其他方法。其余 5 种方法的具体性能差异如下：

1) 微调与 LwF 方法存在显著的任务间性能失衡，尽管这两种方法在新任务上的识别准确率可达到 98%，但其对旧任务的识别准确率普遍低于 10%，致使整体

表 3 增量学习 6 种方法的识别准确率

方法	任务序号	调制样式	训练任务 1	训练任务 2	训练任务 3	平均正确识别率/%	单论训练时间 /s
			准确率/%	准确率/%	准确率/%		
微调	1	8 QAM、16 QAM、32 QAM	98.76	0.74	0	98.76	7.23
	2	BPSK、8 PSK	—	98.67	1.21	49.71	6.97
	3	QPSK、64 QAM	—	—	98.76	33.32	7.18
LwF	1	8 QAM、16 QAM、32 QAM	98.36	10.29	1.24	98.36	7.11
	2	BPSK、8 PSK	—	96.39	9.88	53.34	9.12
	3	QPSK、64 QAM	—	—	97.81	36.31	9.07
iCaRL	1	8 QAM、16 QAM、32 QAM	98.53	98.26	60.64	98.53	7.21
	2	BPSK、8 PSK	—	97.75	94.16	98.01	17.74
	3	QPSK、64 QAM	—	—	50.31	68.37	17.62
ER_ACE	1	8 QAM、16 QAM、32 QAM	98.65	98.41	26.57	98.65	7.22
	2	BPSK、8 PSK	—	94.64	94.52	96.51	11.87
	3	QPSK、64 QAM	—	—	83.67	68.25	11.74
联合学习	1	8 QAM、16 QAM、32 QAM	98.76	98.29	98.47	98.76	7.12
	2	BPSK、8 PSK	—	98.42	98.82	98.35	13.76
	3	QPSK、64 QAM	—	—	98.52	98.60	20.62
本文方法	1	8 QAM、16 QAM、32 QAM	98.77	98.47	53.86	98.77	7.11
	2	BPSK、8 PSK	—	96.34	94.93	97.45	9.91
	3	QPSK、64 QAM	—	—	97.64	82.14	9.89

平均准确率降至 33% 以下。相较而言，iCaRL、ER_ACE 及本文方法在训练完新任务后仍保持较高的旧任务识别准确率。导致该现象的原因是微调与 LwF 缺乏样本回放机制，仅依赖模型结构调整与正则化约束来维持了跨任务的知识迁移能力。而 iCaRL、ER_ACE 及本文方法会存储旧任务样本并联合新数据进行持续训练，缓解了灾难性遗忘，有效提升了模型的多任务泛化能力。

2) 本文提出的识别方法在增量学习中表现出显著优势，较性能最优的 iCaRL 方法识别准确率提升 13.77%。在任务 1，任务 2 训练完成后，iCaRL、ER_ACE 及本文方法对新旧任务的识别性能差异较小，但在任务 3 训练完成后，iCaRL、ER_ACE 的训练序列 1 和训练序列 3 的识别准确率出现明显下降。这种现象的原因可能在于训练序列 2 中的 PSK 信号与训练序列 1、3 中的 QAM 信号存在显著特征差异。在任务 3 的训练过程中，由于训练数据间分布极不平衡，导致出现灾难性遗忘，模型对任务 1 和任务 3 的数据产生了混淆。

3) 本文方法相较于联合学习以及 iCaRL 算法，单轮次训练所需时间最少，充分说明本文算法具有更高的

训练效率。

因此, 联合学习虽然能提供增量学习的理论上限, 但其需要对全部数据进行重新训练, 在实际应用中面临较高的计算和时间成本。本文提出的方法在准确率上虽略低于联合学习, 但训练时间显著缩短, 更具实用价值。

4 结束语

本文针对动态环境下调制识别信号可持续学习的问题, 采用了 DRSN-CS 作为特征提取器, 通过软阈值化操作去除调制识别任务中输入信号噪声和干扰。软阈值操作前后的特征变化, 可以理解网络是如何识别和过滤掉噪声信息。考虑到灾难性遗忘的问题, 利用两阶段学习范式动态扩展新模块来适应新类别。首先利用一个模块拟合目标模型和原始模型输出之间的残差以实现特征增强; 其次, 通过知识蒸馏去除冗余参数和特征维度。为验证本文方法的有效性, 基于开源数据集开展了对比实验。实验结果表明, 相较于多种主流信号特征提取方法和增量学习算法, 本文方法在识别准确率和鲁棒性方面均展现出更优越的效果, 展现出良好的应用价值。

参考文献:

- [1] HAMEED F, DOBRE O A, POPESCU D C. On the likelihood-based approach to modulation classification [J]. IEEE Transactions on Wireless Communications, 2009, 8 (12): 5884–5892.
- [2] 覃国津, 迟楠. 基于特征提取的人工智能调制格式识别算法研究现状 [J]. 激光与光电子学进展, 2023, 60 (9): 228–237.
- [3] HUANG S, LIN C S, XU W J, et al. Identification of active attacks in internet of things: joint model and datadriven automatic modulation classification approach [J]. IEEE Internet of Things Journal, 2021, 8 (3): 2051–2065.
- [4] 杨宵, 姚爱琴, 孙运强, 等. 基于 CBAM-GRU 的通信信号自动调制识别 [J]. 遥测遥控, 2024, 45 (5): 73–81.
- [5] MULLER F C B F, CARDOSO C, KLAUTAU A. A front end for discriminative learning in automatic modulation classification [J]. IEEE Commun. Lett, 2011, 15 (4): 443–445.
- [6] 谭晓衡, 褚国星, 张雪静, 等. 基于高阶累积量和小波变换的调制识别算法 [J]. 系统工程与电子技术, 2018, 40 (1): 171–177.
- [7] CORUK R B, GOKDOGAN B Y, BENZAGHTA M, et al. On the classification of modulation schemes using higher order statistics and support vector machines [J]. Wireless Personal Communications, 2022, 126 (2): 1363–1381.
- [8] WANG L, LI Y, Constellation based signal modulation recognition for MQAM [J]. in 2017 IEEE 9th International Conference on Communication Software and Networks (ICCSN), Guangzhou, 2017: 826–829.
- [9] O'SHEA T J, CORGAN J, CLANCYTC. Convolutional radio modulation recognition networks [C] // Proceedings IEEE Int. Conf. Eng. Appl. Neural Netw, 2016: 213–226.
- [10] EJAZ W, HASAN N U, KIM H S. Idetection: intelligent primary user detection for cognitive radio networks [J]. Services and Technologies, 2012: 153–157.
- [11] RAJENDRAN S, MEERTW, GIUSTINIANOD, et al. Deep learning models for wireless signal classification with distributed lowcost spectrum sensors [C] // IEEE Transactions on Cognitive Communications and Networking, 2018, (4) 3: 433–445.
- [12] WEST, NATHAN E, O'SHEA T. Deep architectures for modulation recognition [J]. IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN), IEEE, 2017: 2964–2975.
- [13] TEKBYK, EKTIAR, GÖRÇİN A, et al. Robust and fast automatic modulation classification with CNN under multipath fading channels [C] // 2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring), 2020: 1–6.
- [14] REBUFFI, SYLVESTRE-ALVISE, et al. Icarl: incremental classifier and representation learning [J]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 2001–2010.
- [15] KIRKPATRICK, JAMES, et al. Overcoming catastrophic forgetting in neural networks [C] // Proceedings of the National Academy of Sciences, 114, 2017 (13): 3521–3526.
- [16] WANG F Y, et al. Foster: feature boosting and compression for classincremental learning [J]. European Conference on Computer Vision, 2022: 398–414.
- [17] 崔良中, 孙佳杰, 牛雅萌. 基于类别增量学习的自动调制识别方法 [J]. 海军工程大学学报, 2025: 1–9.
- [18] 曹岩, 朱真峰. DRSTN: 深度残差软阈值化网络 [J]. 计算机科学, 2024, 51 (s1): 93–99.
- [19] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition in proc [J]. IEEE Conf. Comput. Vision Pattern Recognit, Seattle, 2016, 770–778.
- [20] ZHAO M H, et al. Deep residual shrinkage networks for fault diagnosis [J]. IEEE Transactions on Industrial Informatics, 2019: 4681–4690.
- [21] MA S, CHU F, HAN Q. Deep residual learning with de-modulated time frequency features for fault diagnosis of planetary gearbox under nonstationary running conditions [J]. Mech. Syst. Signal Process, 2019, 127: 190–201.
- [22] HU J, SHEN L, SUN G. Squeeze-andexcitation networks [J]. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 7132–7141.