

基于知识蒸馏与注意力图的雷达信号识别方法

曲志昱 李根 邓志安*

(哈尔滨工程大学信息与通信工程学院 哈尔滨 150001)

(先进船舶通信与信息技术工业和信息化部重点实验室 哈尔滨 150001)

摘要: 针对传统雷达信号识别方法无法有效进行识别类型扩展问题, 该文提出一种基于知识蒸馏与注意力图的雷达信号识别方法。首先将雷达信号的平滑伪Wigner-Ville分布(SPWVD)作为输入; 然后设计了基于残差网络的增量学习网络结构, 利用基于知识蒸馏与注意力图的损失函数, 缓解类别增量过程中的灾难性遗忘; 最后采用基于样本特征均值距离的方法对数据集进行管理, 有效降低存储资源占用空间。实验表明, 该方法能在存储资源有限的情况下, 对扩展分类的信号快速完成训练, 且对原有分类和扩展分类信号均有良好的识别准确率。

关键词: 雷达信号识别; 增量学习; 知识蒸馏; 注意力图; 灾难性遗忘

中图分类号: TN957.51

文献标识码: A

文章编号: 1009-5896(2022)09-3170-08

DOI: 10.11999/JEIT210695

Radar Signal Recognition Method Based on Knowledge Distillation and Attention Map

QU Zhiyu LI Gen DENG Zhian

(College of Information and Communication Engineering, Harbin Engineering University,
Harbin 150001, China)

(Key Laboratory of Advanced Marine Communication and Information Technology,
Ministry of Industry and Information Technology, Harbin 150001, China)

Abstract: In order to solve the problem that traditional radar signal recognition methods can not effectively expand the recognition types, a radar signal recognition method based on knowledge distillation and attention map is proposed. Firstly, the Smooth Pseudo Wigner-Ville Distribution (SPWVD) of the radar signal is used as input; Then, the incremental learning network structure based on residual network is designed, and the loss function based on knowledge distillation and attention map is used to alleviate the catastrophic forgetting in the process of category increment; Finally, a method based on the mean distance of sample features is used to manage the data set, which reduces effectively the occupied storage resources. Experiments show that this method can quickly complete the training of the extended classification signal under the condition of limited storage resources, and has good recognition accuracy for the original classification and the extended classification signal.

Key words: Radar signal recognition; Incremental learning; Knowledge distillation; Attention map; Catastrophic forgetting

1 引言

雷达信号识别是电子对抗的重要环节^[1]。通过对截获的雷达信号进行识别, 可以判断敌方雷达的类型、威胁等级, 对于夺得战场的信息优势有重要意义^[2]。随着新体制雷达大规模投入使用^[3], 实际

战时侦察到的雷达信号可能为新调制类型信号, 因此雷达信号识别算法需要进行识别类型扩展。

近年来, 学者将机器学习与深度学习的算法应用在雷达信号识别领域。文献[4]将奇异值熵与分形维数作为特征向量, 利用支持向量机(Support Vector Machine, SVM)实现了对8种雷达信号的识别。文献[5,6]都是利用卷积神经网络去噪, 然后利用Inception网络进行分类识别, 在低信噪比下(-10 dB)仍有90%以上的识别正确率。文献[7]设计了栈式稀疏自动编码器, 将预处理过的8种雷达信号时频图像送入进行离线训练后, 整体识别正确率达到96.4%。

收稿日期: 2021-07-12; 改回日期: 2022-01-18; 网络出版: 2022-02-19

*通信作者: 邓志安 dengzhian@hrbeu.edu.cn

基金项目: 国家自然科学基金(61801143, 61971155)

Foundation Items: The National Natural Science Foundation of China (61801143, 61971155)

以上的识别算法虽然有较好识别正确率但是都是经过离线学习后识别固定类别的信号。当有识别类型扩展的需求时，需要已有类别与扩展类别的所有数据集混合后重新开始训练网络，这对于设备的存储资源和训练时间均有较高要求。文献[8]基于残差网络与3元损失函数学习得到特征映射，利用设计的基于样本库的识别方法实现识别类型扩展，但是随着类型的增加训练后的识别正确率有待提高。

本文采用增量学习的训练方法实现雷达信号识别类型扩展。针对增量学习过程中网络会出现灾难性遗忘^[9]问题，本文利用基于特征平均值距离的方法选择极少典型样本构成新训练集，设计了基于残差的增量学习网络结构，并利用知识蒸馏与注意力图^[10]缓解灾难性遗忘，实现了识别类型的有效扩展。仿真实验表明，在存储资源有限的条件下，本方法对原有分类与扩展分类的雷达信号均有良好的识别准确率。

2 知识蒸馏与注意力图

2.1 知识蒸馏

知识蒸馏^[11]是一种有效的模型压缩方法，它将复杂模型所学到的“知识”迁移到参数量少的简单模型中。在知识蒸馏过程中，训练简单模型时加入复杂模型的softmax层输出作为软标签辅助训练，用来提高简单模型的识别效果。在复杂模型产生软标签的过程中，加入温度变量 T 使各个类别产生较平滑的概率分布，如式(1)所示，这使得负标签所携带的信息相对放大，模型训练更加关注负标签

$$\text{softmax}(z_i) = \frac{\exp(z_i/T)}{\sum_{j=1}^N \exp(z_j/T)} \quad (1)$$

其中， z_i 为复杂模型所产生的软标签， T 为温度变量， N 为类别数量。由于增加了软标签，知识蒸馏的优化目标除了交叉熵损失外还多了一项蒸馏损失 L_{DL}

$$L_{DL} = - \sum_j^N y_j^T \ln \left(\hat{y}_j^T \right) \quad (2)$$

其中， y_j^T 为复杂模型所产生的软标签， \hat{y}_j^T 为简单模型加入温度变量 T 后的输出。

2.2 注意力图

随着深度卷积神经网络在解决复杂视觉问题方面取得了长足进步，学者对于网络做出决策时的“依据”也进行了研究。注意力图 (attention map) 是一种反映网络可解释性的重要方法，其能够表示网络做出输出决策时对输入图像最关注的区域。本文使用梯度加权类激活映射 (Gradient-weighted

Class activation mapping, Grad-Cam)^[12]表示注意力图。Grad-Cam将网络的特征图赋予不同的权重，然后通过生成热力图标记出输入图像中对预测影响较大的像素区域。卷积神经网络最后一层卷积含有最丰富的语义信息与空间信息^[13]，所以Grad-Cam将最后一层卷积作为特征图。当图像进入网络后，得到其全连接层输出，计算全连接层输出最大值对特征图的梯度，经过平均计算后得到特征图权重，如式(3)所示

$$\lambda_k^c = \frac{1}{N} \sum_i \sum_j \frac{\partial y^c}{\partial A_{i,j}^k} \quad (3)$$

其中， λ_k^c 为第 k 个特征图对应的权重， y^c 为最后一层全连接层输出的最大值， $A_{i,j}^k$ 为第 k 个特征图(i, j)位置的像素值， N 为每个特征图长和宽的乘积。由式(3)得到了每个特征图的权重，与特征图相乘加权后，经过ReLU函数便得到Grad-Cam，用其表示模型对于图像的注意力图，如式(4)所示

$$L_{G-C} = \text{ReLU} \left(\sum_k \lambda_k A_k \right) \quad (4)$$

本文利用Grad-Cam对线性调频信号 (Linear Frequency Modulation, LFM)、Frank编码信号、二相编码信号 (Binary Phase-coded Signal, BPSK)、V型调频信号 (DLFM)、偶2次调频信号 (Even Quadratic Frequency Modulation, EQFM) 与正弦调频信号 (Sinusoidal Frequency Modulation signal, SFM) 产生的热力图与叠加到原时频图像的效果图如图1所示。在热力图中，红色区域表示对网络输出影响较大的区域，由图1可以看出红色区域主要覆盖信号的时频结构，也就是说网络对信号调制类型识别时其决策依据主要为信号的时频结构。同时网络对于不同雷达信号的注意力图不同，将其作为额外的监督信息，在进行扩展类型识别时可以帮助缓解灾难性遗忘。

3 基于知识蒸馏与注意力图的雷达信号识别方法

本方法首先将雷达信号的平滑伪Wigner-Ville分布 (Smooth Pseudo Wigner-Ville Distribution, SPWVD) 作为网络的输入，将积累的扩展类型信号与存储的典型原有分类信号混合组成新的数据集。然后进行类别增量训练，训练过程中加入知识蒸馏损失与注意力图损失缓解训练过程中的灾难性遗忘。最后进行数据集管理，采用基于均值距离的样本选择方法，选择典型新类样本进行存储，同时删除部分原有分类样本维持存储样本总数不变。具体流程图如图2所示。

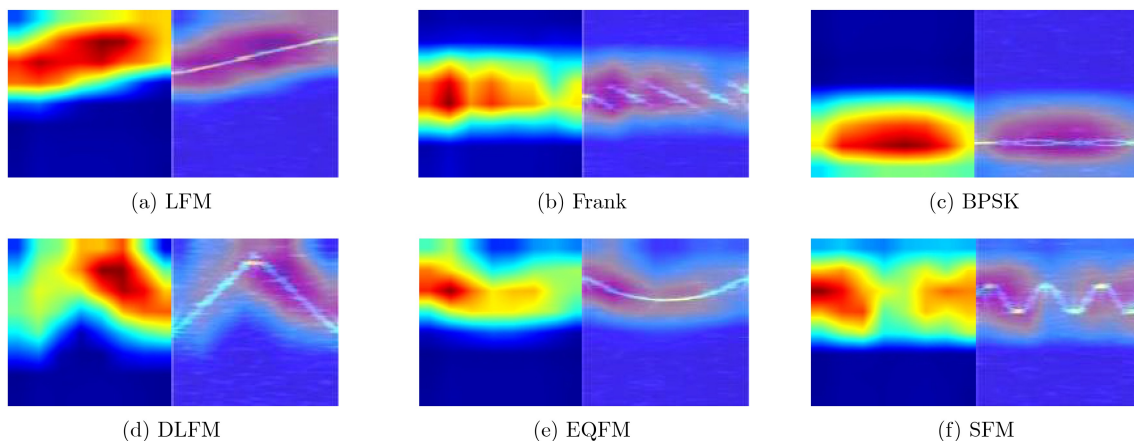


图1 雷达脉内调制信号时频图像Grad-Cam可视化效果

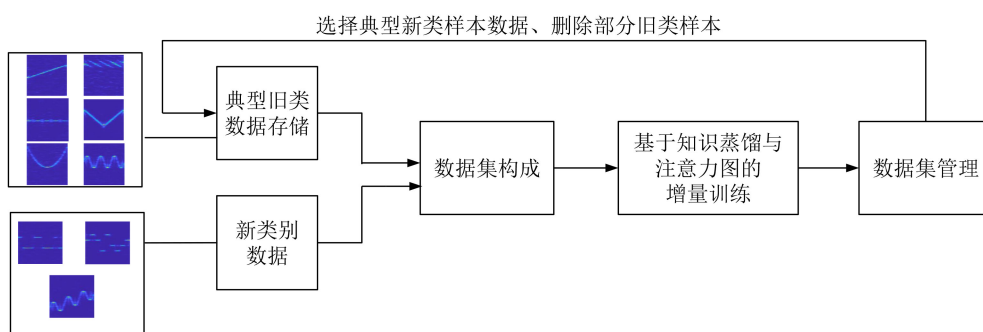


图2 基于知识蒸馏与注意力图的雷达信号识别方法流程

3.1 网络结构

本文采用基于增量学习的网络结构，信号经过SPWVD后送入网络进行训练，过程可以分为特征提取与分类识别，如图3所示。其中 θ_s 为用于特征提取的参数，一般为卷积神经网络中的卷积层。 θ_o 为识别已有类别所用到的参数，当需要进行类型扩展时增加全连接层的输出，所增加的参数用 θ_n 表示^[4]。

本文基于残差网络^[15]设计的网络结构如图4所示，共有4组残差单元，每组残差单元两层卷积，将雷达信号做SPWVD后，经过线性插值调整为 128×128 的图像作为输入。最初的卷积层加上4组残差单元组成特征提取参数 θ_s ，全连接层的参数作为 θ_o ，输出扩展后增加的参数为 θ_n 。最后一组卷积单元的输出含有对时频图像最丰富的语义信息且未丢失空间信息，所以将其作为特征图用于计算Grad-Cam，softmax层的输出含有对时频图像预测概率的信息，所以将其作为软标签。利用网络扩展前后对于扩展类型数据Grad-Cam与输出预测的不同作为额外的监督信息加入训练。

3.2 基于知识蒸馏与注意力图的增量训练

首先根据已有类别信号的数据集训练得到网络模型 M_{t-1} ，当需要进行识别类型扩展时，将新增类型与存储的典型原有样本混合作为新训练集，扩

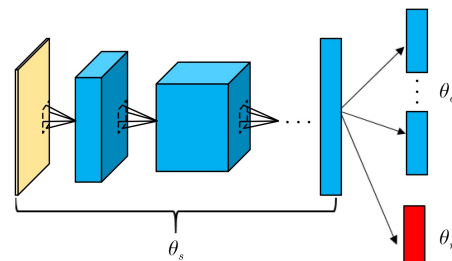


图3 增量识别卷积神经网络参数示意图

展 M_{t-1} 的全连接层输出得到模型 M_t 。其中 M_t 模型中的 θ_s 与 θ_o 用 M_{t-1} 的参数初始化，全连接层增加的 θ_n 随机初始化。最后在训练过程中加入知识蒸馏损失表示 M_{t-1} 与 M_t 对新数据集输出层面的差异，加入注意力图损失表示 M_{t-1} 与 M_t 对新数据集关注区域之间的差异。增量训练过程具体如图5所示。

如图5所示，知识蒸馏损失由 M_{t-1} 对新数据集信号经过温度变量 T 产生的输出与 M_t 对新数据集信号产生的输出计算得到。根据知识蒸馏的过程， M_{t-1} 对新数据集产生的输出作为软标签缓解 M_t 对已识别信号的“遗忘”，知识蒸馏损失 L_{DL} 的计算公式如式(5)所示

$$L_{DL} = - \sum_j^N y_{j, M_{t-1}}^T \ln(y_{j, M_t}^T) \quad (5)$$

其中， $y_{j, M_{t-1}}^T$ 为模型 M_{t-1} 加入温度变量 T 的输出，

y_{j, M_t}^T 为模型 M_t 加入温度变量 T 的输出, N 为已识别信号种类的数量。

本文使用Grad-Cam表示模型的注意力图, 利用 M_t 与 M_{t-1} 对新数据集Grad-Cam的差异表示注意力图损失。最小化注意力图损失使得 M_t 与 M_{t-1} 对信号时频图像的关注区域趋于相似。当用新数据集训练时, 由式(3)、式(4)得到模型 M_{t-1} , M_t 的Grad-Cam。然后如式(6)所示分别将其向量化

$$\left. \begin{aligned} Z_{t-1} &= \text{vector} \left(L_{G-C}^{M_{t-1}} \right) \\ Z_t &= \text{vector} \left(L_{G-C}^{M_t} \right) \end{aligned} \right\} \quad (6)$$

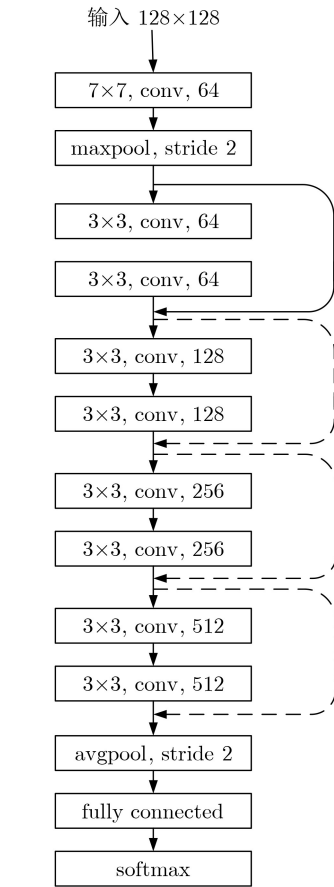


图4 网络结构

其中, $L_{G-C}^{M_{t-1}}$ 与 $L_{G-C}^{M_t}$ 分别为 M_{t-1} 与 M_t 所产生的Grad-Cam, $\text{vector}(\cdot)$ 表示将其向量化, Z_{t-1} 与 Z_t 为向量化的Grad-Cam。然后将 Z_{t-1} 与 Z_t 进行归一化后, 计算两个向量之间的欧氏距离作为 M_{t-1} 与 M_t 注意力的差异, 记为 L_{AML} 如式(7)所示

$$\begin{aligned} L_{AML} &= \left\| \frac{Z_t}{\max(Z_t)} - \frac{Z_{t-1}}{\max(Z_{t-1})} \right\|_2 \\ &= \left(\sum_{i=1}^n \left| \frac{z_t^i}{\max(Z_t)} - \frac{z_{t-1}^i}{\max(Z_{t-1})} \right|^2 \right)^{\frac{1}{2}} \end{aligned} \quad (7)$$

其中, $\max(\cdot)$ 为求向量中的最大值, n 为向量化后的长度。图6表示的是当模型对雷达信号时频图像关注区域不同时, 由式(7)所计算出的注意力图损失 L_{AML} 。在图6(b)模型关注区域相似所以 L_{AML} 较小, 图6(c)模型关注区域差别较大所以 L_{AML} 较大。

综上, 增量训练时将知识蒸馏损失、注意力图损失与交叉熵损失的线性加权作为最终的损失函数, 更新模型 M_t 中的所有参数。当训练完成后, 用 M_t 代替 M_{t-1} 作为下一次类别增量过程中识别原有类别的模型。

3.3 数据集管理

增量学习过程通常在存储资源有限的条件下进行^[6]。采用基于特征均值距离的方法选择每一类的典型样本进行存储可满足存储资源有限的条件。基于特征均值距离的选择方法如式(8)所示

$$\left. \begin{aligned} \mu &= \frac{1}{N} \sum_{i=1}^N \varphi(x_i) \\ p_k &= \min_{x \in X} \|\mu - \varphi(x)\|_2 \end{aligned} \right\} \quad (8)$$

其中, $\varphi(\cdot)$ 是网络参数 θ_s 所对应的特征提取函数, N 是样本数量, X 为训练集, x_i 是训练集中的样本, μ 为每一类经过特征提取后所得到的特征平均值, p_k 为在训练集中所选择的第 k 个典型样本。利

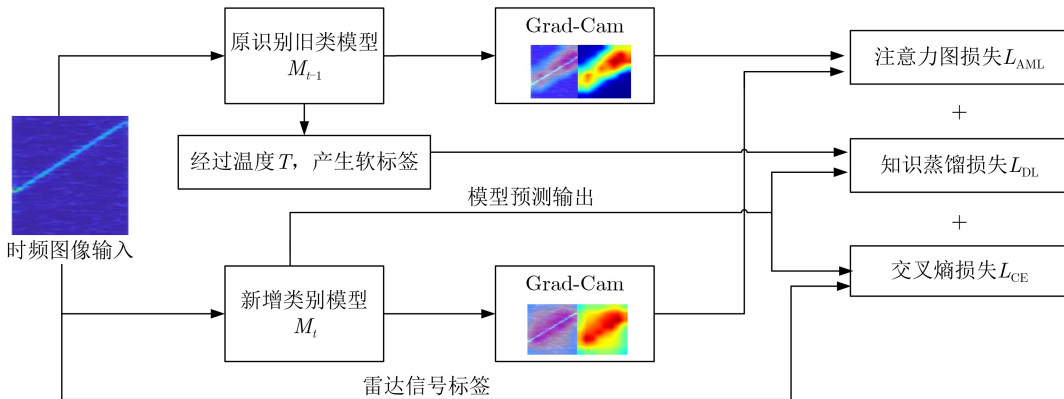


图5 基于知识蒸馏与注意力图的增量训练过程

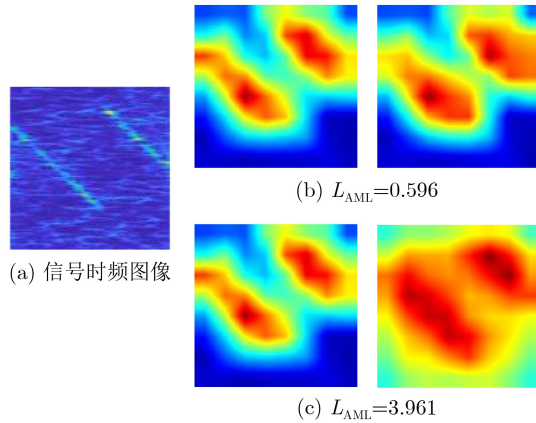


图6 模型关注区域差异示意图

用式(8)选择距离特征平均值最近的前 k 个样本进行存储。设所能存储的样本总数为 M ，在未进行识别类型扩展时，每类存储的典型样本数量为 k ， k 随着识别种类的增加而减少^[17]，具体为 $k = M/t$ ， t 为模型现在所能识别的种类。所以在进行数据集管理时包括两步，首先是选择典型的扩展分类信号进行存储，然后是删除部分原有类别样本，以保证在有限存储空间下快速完成训练。

4 实验结果与分析

4.1 实验设置

为评估本方法的有效性，使用MATLAB生成数据，每种雷达脉内调制信号的频率参数根据采样频率做归一化处理，每种信号的长度 $N=512\sim 1024$ 且每种信号的参数都有一个动态变化的范围。首先将参数如表1所示的6种雷达脉内调制信号作为已识别的信号类型，数据集样本信噪比范围在 $-10\sim 8$ dB，每隔2 dB每种信号产生500个样本，所以每种信号有5000个样本，其中4500个样本作为训练集，500个样本作为验证集。测试集数据信噪比在 $-12\sim 0$ dB，每隔1 dB产生500个样本数据。

基于以上训练集，训练得到模型 M_{t-1} 作为原有信号的识别模型。同时选取文献[18,19]所提出的模型与本文模型进行对比。文献[18]采用了基于AlexNet的模型进行识别，文献[19]采用了扩张残差网络进行识别。训练完成后用测试集测试，在不同信噪比环境下识别准确率如图7所示。由图7明显看出本文所用网络结构准确率较高，所用网络结构对信号有良好的特征提取能力与分类识别能力。

设存储容量为600个样本，训练完成的模型 M_{t-1} 利用基于样本特征均值的方法，每一类信号选取100个作为典型样本进行存储。仿真参数如表2所示的雷达脉内调制信号依次作为扩展类型。信噪比设置与已识别的6类信号相同，每类新增信号同

表1 已识别信号参数设置

信号类型	参数	变化范围
LFM	初始频率 f_c	0.01~0.45
	带宽 Δf	0.02~0.40
Frank	载频 f_0	0.10~0.45
	相位数	[4, 5, 6, 7, 8]
BPSK	Barker码	[5, 7, 11, 13]
	载频 f_0	0.10~0.45
DLFM	码长 T_b	(1/32~1/16)N
	初始频率 f_c	0.01~0.41
EQFM	带宽 Δf	0.05~0.45
	最小频率 f_{\min}	0.01~0.45
SFM	最大频率 f_{\max}	0.01~0.45
	带宽 Δf	0.05~0.40
	最小频率 f_{\min}	0.01~0.18
	最大频率 f_{\max}	0.20~0.45
	带宽 Δf	0.02~0.44

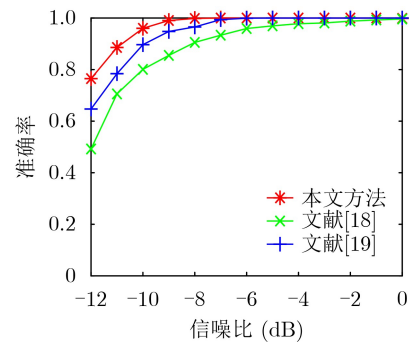


图7 准确率对比

表2 扩展类型信号参数设置

信号类型	参数	变化范围
2FSK	载频 f_1, f_2	0.10~0.45
	码宽 T_b	(1/32~1/8)N
4FSK	载频 $f_1\sim f_4$	0.10~0.45
	码宽 T_b	(1/32~1/8)N
LFM-SFM	初始频率 f_c	0.01~0.45
	最小频率 f_{\min}	0.08~0.18
	最大频率 f_{\max}	0.20~0.30
MLFM	载频 f_0	0.01~0.25
	带宽 Δf	0.10~0.45
P1, P2	载频 f_0	0.10~0.45
	步进频率 M	[4,5,6,7,8], P2取偶数
P3, P4	载频 f_0	0.10~0.45
	压缩比	16~64

样也为4500个样本。所有训练过程batchsize设为64，采用Adam优化器，初始学习率设为 1×10^{-3} ，随训练过程递减，知识蒸馏时设 $T=2$ ，知识蒸馏损失、注意力图损失与交叉熵损失的加权系数分别取0.8, 0.1, 0.1。

4.2 与联合训练效果对比

联合训练是将原有分类与扩展分类的所有数据混合重新开始训练，不存在灾难性遗忘问题，可以

看作增量学习识别效果的上限，但是对于存储及训练时间有较高要求。本次实验与联合训练方法进行对比，在识别6类调制信号的基础上依次增加新调制类型。

图8为当扩展类型数量为4类，采用基于联合训练的方法、未添加注意力图损失的方法与本文方法训练时，每隔10个batchsize所记录的损失值，由图8可以看出每种方法训练完成后模型都收敛。图9为扩展类型为4类时，以上3种方法所训练的模型在-5 dB下识别效果的混淆矩阵。由图9可以看出，在-5 dB时本文方法的识别效果好于未添加注意力图损失的方法，且只有几种信号的正确率略低于联合训练方法。图10为扩展类型数量为4时，联合训练与本方法训练时所存储的样本数量。因为本方法采用的损失函数缓解了灾难性遗忘，所以训练时只需600个典型原有类别的样本与45000个扩展类型样本。当新增识别4类训练完成后，网络总共

可以识别10类雷达脉内调制信号，根据样本选择方法每一类保存60个典型样本。同时，由于本方法训练集样本数量低于联合训练样本数量数倍，所以训练时间也会缩短。综上，本文方法进行雷达信号增量识别的效果与联合训练大致相当，但训练时间和存储要求远低于联合训练。

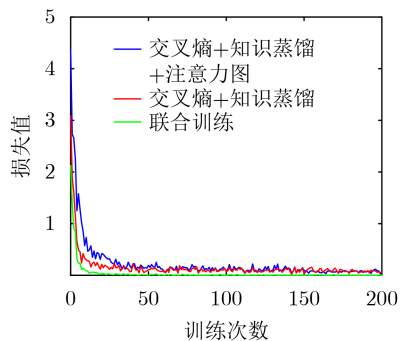
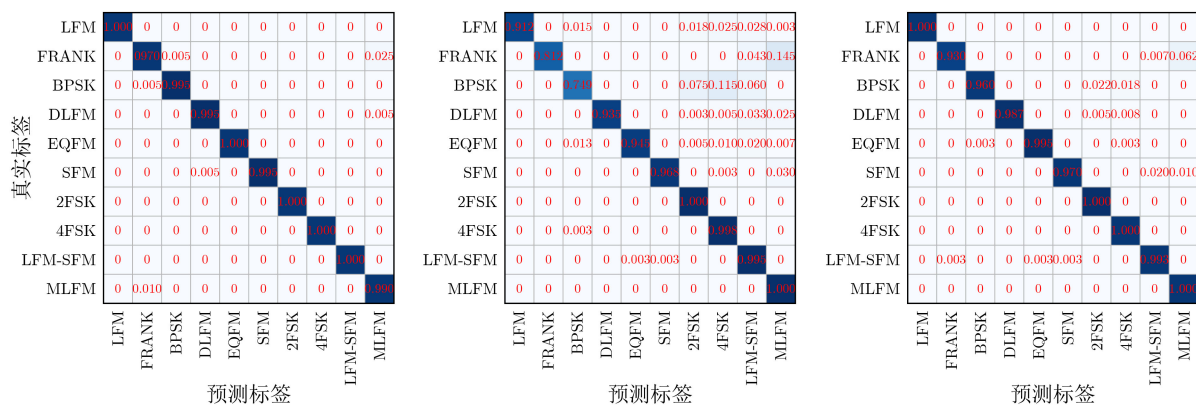


图8 不同模型损失



(a) -5 dB联合训练识别效果混淆矩阵 (b) -5 dB交叉熵+知识蒸馏识别效果混淆矩阵 (c) -5 dB本文方法识别效果混淆矩阵

图9 信噪比-5 dB识别效果对比

4.3 缓解灾难性遗忘有效性验证

为验证所采用的损失函数的有效性，在相同数据集情况下，与文献[20]基于连续微调的增量识别方法进行对比；为验证存储典型样本的有效性，与只采用扩展类型数据训练的方法进行对比。当依次扩展1~4类调制信号时，得到每种模型在-12~0 dB下对于所有调制信号的平均识别正确率如图11所示。

由图11对比可知，由于添加了知识蒸馏损失与注意力图损失，从而缓解了网络对已有信号类型的遗忘，本文方法识别正确率要明显高于文献[20]中只采用交叉熵损失的连续微调方法。同时，当只用扩展类型数据进行训练时，与本文采用典型原有样本存储的方法在低信噪下的识别正确率有很大差距。这是因为低信噪比情况下时频结构受损严重，识别正确率相对较低，当仅使用扩展类型数据训练

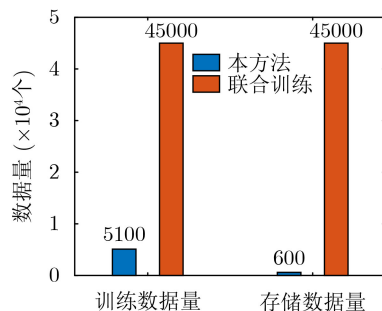


图10 联合训练与本文方法存储数量对比

没有原有数据参与时，模型对已识别信号更容易遗忘。综上，本文方法所采用的知识蒸馏、注意力图与典型样本存储能够缓解增量识别过程中的灾难性遗忘，提高模型的识别正确率。

4.4 增量识别扩展性验证

为测试本方法识别效果随扩展类型数量增加的

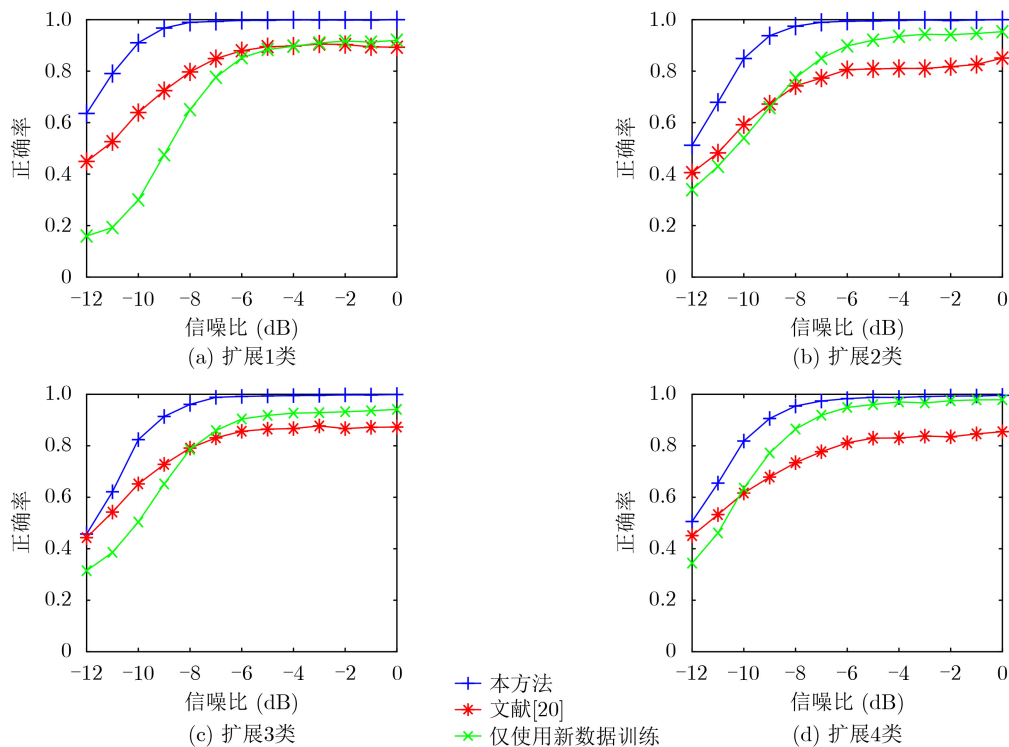


图 11 识别正确率与信噪比关系曲线

变化,设计依次增加8种调制信号。每一类进行增量训练后,用信噪比为 -7 dB的测试集进行测试。增量到8种信号时,模型总共可以识别14种信号,每一类信号存储43个典型样本。模型对调制信号的识别正确率与扩展数量的关系如图12所示。由图12可知,随着识别种类的增加模型识别效果逐渐下降。这主要是因为随着调制类型的增多,模型所学到的“知识”越多,再次新增类别时,“遗忘掉的知识”也会增多。同时由于存储限制,每一类的典型样本也会越来越少,也会造成识别效果下降。

5 结论

本文基于知识蒸馏、注意力图与典型样本选择等方法,提出了一种基于残差网络的雷达信号增量识别方法,有效地解决了雷达调制信号识别类型扩展问题。仿真结果表明,本方法对原有分类和扩展

分类的调制信号识别正确率逼近于联合训练模型得到的结果,但训练时间和存储要求远低于后者,同时本文所采用的损失函数能够有效缓解类别增量过程中的灾难性遗忘,使得模型对原有分类和扩展分类信号均有较好的识别正确率。

参考文献

- [1] 王星,周一鹏,周东青,等. 基于深度置信网络和双谱对角切片的低截获概率雷达信号识别[J]. 电子与信息学报, 2016, 38(11): 2972-2976. doi: 10.11999/JEIT160031. WANG Xing, ZHOU Yipeng, ZHOU Dongqing, et al. Research on low probability of intercept radar signal recognition using deep belief network and bispectra diagonal slice[J]. *Journal of Electronics & Information Technology*, 2016, 38(11): 2972-2976. doi: 10.11999/JEIT160031.
- [2] ZHANG Chunhong, HAN Yuetao, ZHANG Ping, et al. Research on modern radar emitter modelling technique under complex electromagnetic environment[J]. *The Journal of Engineering*, 2019(20): 7134-7138. doi: 10.1049/joe.2019.0579.
- [3] 徐玉芬. 现代雷达信号处理的数字脉冲压缩方法[J]. 现代雷达, 2007, 29(7): 61-64. doi: 10.3969/j.issn.1004-7859.2007.07.018. XU Yufen. Methods of digital pulse compression in modern radar signal processing[J]. *Modern Radar*, 2007, 29(7): 61-64. doi: 10.3969/j.issn.1004-7859.2007.07.018.
- [4] 曲志昱,毛校洁,侯长波. 基于奇异值熵和分形维数的雷达信

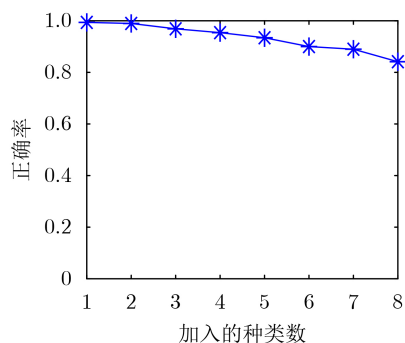


图 12 类别扩展数量与识别正确率关系曲线

- 号识别[J]. 系统工程与电子技术, 2018, 40(2): 303–307. doi: [10.3969/j.issn.1001-506X.2018.02.10](https://doi.org/10.3969/j.issn.1001-506X.2018.02.10).
- QU Zhiyu, MAO Xiaojie, and HOU Changbo. Radar signal recognition based on singular value entropy and fractal dimension[J]. *Systems Engineering and Electronics*, 2018, 40(2): 303–307. doi: [10.3969/j.issn.1001-506X.2018.02.10](https://doi.org/10.3969/j.issn.1001-506X.2018.02.10).
- [5] 肖易寒, 王亮, 郭玉霞. 基于去噪卷积神经网络的雷达信号调制类型识别[J]. 电子与信息学报, 2021, 43(8): 2300–2307. doi: [10.11999/JEIT200506](https://doi.org/10.11999/JEIT200506).
- XIAO Yihan, WANG Liang, and GUO Yuxia. Radar signal modulation type recognition based on denoising convolutional neural network[J]. *Journal of Electronics & Information Technology*, 2021, 43(8): 2300–2307. doi: [10.11999/JEIT200506](https://doi.org/10.11999/JEIT200506).
- [6] QU Zhiyu, WANG Wenyang, HOU Changbo, et al. Radar signal intra-pulse modulation recognition based on convolutional denoising autoencoder and deep convolutional neural network[J]. *IEEE Access*, 2019, 7: 112339–112347. doi: [10.1109/ACCESS.2019.2935247](https://doi.org/10.1109/ACCESS.2019.2935247).
- [7] 郭立民, 寇韵涵, 陈涛, 等. 基于栈式稀疏自编码器的低信噪比下低截获概率雷达信号调制类型识别[J]. 电子与信息学报, 2018, 40(4): 875–881. doi: [10.11999/JEIT170588](https://doi.org/10.11999/JEIT170588).
- GUO Limin, KOU Yunhan, CHEN Tao, et al. Low probability of intercept radar signal recognition based on stacked sparse auto-encoder[J]. *Journal of Electronics & Information Technology*, 2018, 40(4): 875–881. doi: [10.11999/JEIT170588](https://doi.org/10.11999/JEIT170588).
- [8] 石礼盟, 杨承志, 吴宏超. 基于深层残差网络和三元组损失的雷达信号识别方法[J]. 系统工程与电子技术, 2020, 42(11): 2506–2512. doi: [10.3969/j.issn.1001-506X.2020.11.12](https://doi.org/10.3969/j.issn.1001-506X.2020.11.12).
- SHI Limeng, YANG Chengzhi, and WU Hongchao. Radar signal recognition method based on deep residual network and triplet loss[J]. *Systems Engineering and Electronics*, 2020, 42(11): 2506–2512. doi: [10.3969/j.issn.1001-506X.2020.11.12](https://doi.org/10.3969/j.issn.1001-506X.2020.11.12).
- [9] PARISI G I, KEMKER R, PART J L, et al. Continual lifelong learning with neural networks: A review[J]. *Neural Networks*, 2019, 113: 54–71. doi: [10.1016/j.neunet.2019.01.012](https://doi.org/10.1016/j.neunet.2019.01.012).
- [10] DHAR P, SINGH R V, PENG K C, et al. Learning without memorizing[C]. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, USA, 2019: 5133–5141. doi: [10.1109/CVPR.2019.00528](https://doi.org/10.1109/CVPR.2019.00528).
- [11] HINTON G, VINYALS O, and DEAN J. Distilling the knowledge in a neural network[J]. *Computer Science*, 2015, 14(7): 38–39.
- [12] SELVARAJU R R, COGSWELL M, DAS A, et al. Grad-CAM: Visual explanations from deep networks via gradient-based localization[C]. 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 2017: 618–626. doi: [10.1109/ICCV.2017.74](https://doi.org/10.1109/ICCV.2017.74).
- [13] ZHOU Bolei, KHOSLA A, LAPEDRIZA A, et al. Learning deep features for discriminative localization[C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, USA, 2016: 2921–2929. doi: [10.1109/CVPR.2016.319](https://doi.org/10.1109/CVPR.2016.319).
- [14] LI Zhizhong and HOIEM D. Learning without forgetting[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018, 40(12): 2935–2947. doi: [10.1109/TPAMI.2017.2773081](https://doi.org/10.1109/TPAMI.2017.2773081).
- [15] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Deep residual learning for image recognition[C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, USA, 2016: 770–778. doi: [10.1109/CVPR.2016.90](https://doi.org/10.1109/CVPR.2016.90).
- [16] CASTRO F M, MARÍN-JIMÉNEZ M J, GUIL N, et al. End-to-end incremental learning[C]. The 15th European Conference on Computer Vision, Munich, Germany, 2018: 241–257. doi: [10.1007/978-3-030-01258-8_15](https://doi.org/10.1007/978-3-030-01258-8_15).
- [17] REBUFFI S A, KOLESNIKOV A, SPERL G, et al. iCaRL: Incremental classifier and representation learning[C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, USA, 2017: 5533–5542. doi: [10.1109/CVPR.2017.587](https://doi.org/10.1109/CVPR.2017.587).
- [18] 郭立民, 陈鑫, 陈涛. 基于AlexNet模型的雷达信号调制类型识别[J]. 吉林大学学报:工学版, 2019, 49(3): 1000–1008. doi: [10.13229/j.cnki.jdxbgxb20171056](https://doi.org/10.13229/j.cnki.jdxbgxb20171056).
- GUO Limin, CHEN Xin, and CHEN Tao. Radar signal modulation type recognition based on AlexNet model[J]. *Journal of Jilin University: Engineering and Technology Edition*, 2019, 49(3): 1000–1008. doi: [10.13229/j.cnki.jdxbgxb20171056](https://doi.org/10.13229/j.cnki.jdxbgxb20171056).
- [19] 秦鑫, 黄洁, 查雄, 等. 基于扩张残差网络的雷达辐射源信号识别[J]. 电子学报, 2020, 48(3): 456–462. doi: [10.3969/j.issn.0372-2112.2020.03.006](https://doi.org/10.3969/j.issn.0372-2112.2020.03.006).
- QIN Xin, HUANG Jie, ZHA Xiong, et al. Radar emitter signal recognition based on dilated residual network[J]. *Acta Electronica Sinica*, 2020, 48(3): 456–462. doi: [10.3969/j.issn.0372-2112.2020.03.006](https://doi.org/10.3969/j.issn.0372-2112.2020.03.006).
- [20] KÄDING C, RODNER E, FREYTAG A, et al. Fine-tuning deep neural networks in continuous learning scenarios[C]. The 13th Asian Conference on Computer Vision, Taipei, China, 2016: 588–605. doi: [10.1007/978-3-319-54526-4_43](https://doi.org/10.1007/978-3-319-54526-4_43).
- 曲志昱: 女, 副教授, 研究方向为电子侦察与对抗、阵列信号测向。
李根: 男, 硕士生, 研究方向为雷达信号识别。
邓志安: 男, 副教授, 研究方向为人工智能与宽带信号处理。