第 36 卷第 4 期
2024 年 4 月

强 激 光 与 粒 子 束
HIGH POWER LASER AND PARTICLE BEAMS

Vol. 36, No. 4
Apr., 2024

·电磁频谱作战与电磁安全·

基于重要区域定位与掩码的射频指纹可视化分析*

刘文斌[1,2]， 范平志[1]， 杨佳煌[3]， 李雨锴[3]， 王钰浩[3]， 孟 华[3]

（1. 西南交通大学 信息科学与技术学院，成都 611756； 2. 中国电子科技集团公司第三十研究所，成都 610041；
3. 西南交通大学 数学学院，成都 611756）

摘 要： 针对时域脉冲信号样本的射频指纹提取与深度学习模型的可解释性，提出了一种基于 Grad-CAM 的重要区域可视化呈现方法，并通过重要区域的多次掩码测试，来分析重要区域对射频指纹识别结果的影响。基于 10 个辐射源的信号样本，对比了层数不同的两种 ResNet 模型的测试结果。测试发现该方法能够区分不同类型信号并呈现个体差异。分析表明，该方法能够发现不同辐射源发送相同信号时的重要区域定位差异，能可视化反映辐射源个体特征的空间距离，以及不同模型的特征表征与指纹定位准确度差异；同时发现对重要区域的掩码更容易产生误预测，证明特定信号存在与时频特征相关的射频指纹，并可辅助可视化定位影响射频指纹样本识别的关键点。

关 键 词： 可解释性；射频指纹；深度学习；可视化；信号特征

中图分类号： TN92 **文献标志码：** A **doi**: 10.11884/HPLPB202436.230380

Visual analysis method for RF fingerprint based on important region localization and masking

Liu Wenbin[1,2]， Fan Pingzhi[1]， Yang Jiahuang[3]， Li Yukai[3]， Wang Yuhao[3]， Meng Hua[3]

（1. *School of Information Science & Technology, Southwest Jiaotong University, Chengdu* 611756, *China*;
2. *The* 30*th Research Institute of China Electronics Technology Group Corporation, Chengdu* 610041, *China*;
3. *School of Mathematics, Southwest Jiaotong University, Chengdu* 611756, *China*）

Abstract: A Grad-CAM based visualizing method for important regions is proposed for the interpretability of RF fingerprint extraction and deep learning models of time-domain pulse signal samples. The impact of important regions on RF fingerprint recognition results is analyzed through multiple mask tests of important regions. Based on signal samples of 10 emitters, the test results of two ResNet models with different layers are compared. It is found that the proposed method can distinguish different types of signals and present individual differences. Analysis shows that this method can detect important regional localization differences when different emitters send the same signal, and can visually reflect the spatial distance of RF fingerprint characteristics, as well as the differences in feature representation and fingerprint localization accuracy of different models; At the same time, it is found that masks for important areas are more prone to false predictions, which proves the existence of RF fingerprints related to time-frequency characteristics in specific signals, and can assist in visualizing key points that affect the recognition of RF fingerprint samples.

Key words: interpretability, radio frequency fingerprint, deep learning, visualization, signal characteristics

射频指纹识别是近年来比较热门的一个研究方向，可通过空中接收的无线信号来分析辐射源射频电路中的微小失真或缺陷，从而对辐射源进行个体区分。该方法已逐步应用于通信、雷达等射频辐射源与办公设备等电磁泄漏辐射源检测中，如针对未授权对象的安全检测场景以及非合作条件下的目标检测场景[1-3]。传统的指纹识别方法，需要针对特定对象，针对性地设计特征提取方法，效率低且准确率不高，而通过诸如卷积神经网络的多维特征

* **收稿日期:** 2023-10-30； **修订日期:** 2023-12-19
基金项目: 西南交通大学交叉培育项目（2682023TPY027）
联系方式: 刘文斌, bingege389@sina.com.cn。
通信作者: 孟 华, menghua@swjtu.edu.cn。

提取来进行辐射源特征提取与识别，通用性强，拓展性好，准确率高，但会随着空间、时间、传感器的变化导致识别效果变化大，鲁棒性降低[4]。为此，需要提高深度学习的可解释性，分析出无意甚至是有意生成的对抗样本的影响，同时评估出深度学习网络模型的能力与安全边界。

可解释人工智能以及诸如类激活图（CAM）等可视化解释方法，已逐步研究并探讨应用于网络安全、信号检测等应用[5-7]。梯度加权类激活映射（Grad-CAM）利用卷积层上的空间信息，并使用梯度作为权重来获得类决策的属性，可定位重要区域，与被测对象相叠加进行图形化呈现时具有较好的解释性，已应用于光学图像以及基于时频图等图形化信号特征的可视化分析及模型解释[8-11]。

射频指纹识别相对于类型识别而言，样本差异更小。文献 [12] 针对电磁大数据非凡挑战赛数据集提出了一种测试架构，使用多种方法来分析样本的个体特征并评估模型的能力边界。本文在此基础上，针对该数据集的时域脉冲信号样本进行个体识别的同时，围绕样本与深度学习模型的可解释性，提出了一种基于 Grad-CAM 的信号重要特征区域的可视化分析方法，并通过重要区域的多次掩码测试，来分析重要区域对射频指纹识别结果的影响。

1 射频指纹识别对象与模型

1.1 对象

采用与文献 [12] 相同的电磁大数据非凡挑战赛目标个体识别开源数据集，共 10 个辐射源对象，前 5 个对象发送的是单频点脉冲，后 5 个对象发送的是线性调频脉冲。每个类别有 10000 条样本，其中信噪比覆盖 5、6、7……14 dB 共 10 种，即每个类别每种信噪比分别有 1000 条样本。训练集和测试集比例划分为 8:2，进行辐射源指纹识别训练与测试；从测试集中每个对象选择 30 个样本，用于评估深度学习的预测能力、Grad-Cam 呈现结果与掩码测试结果。

1.2 LittleResNet 与 MiddleResNet 模型

为了验证本文提出可视化分析方法的泛用性，在后续实验中使用应用广泛的残差卷积网络（ResNet）[13]。ResNet 是一种在计算机视觉和图像处理领域取得巨大成功的深度学习模型，并在射频指纹识别任务中也被广泛使用[14,15]。最早在传统的卷积神经网络中，网络的性能通常随着深度的增加而提高。然而，当网络达到一定的深度后，性能反而会下降，这被称为"退化问题"。这个问题并非由过拟合导致，因为即使在训练误差上也会观察到这种退化现象。为了缓解这种现象，微软研究院 K. He 等人提出了 ResNet 模型[13]。ResNet 使用了基于卷积神经网络的结构，同时引入了一种称为"残差连接"的结构，它允许网络的输入信息绕过一层或多层后直接流向后面的层。

本文使用的 LittleResNet 与 MiddleResNet 的模型结构如图 1 所示。其中 LittleResNet 模型的主体结构由 6 个重复的卷积模块构成，每个卷积模块由一个卷积核为 3，输出通道为 8 的一维卷积与对应的批归一化、ReLU 激活函数以及残差连接构成；而 MiddleResNet 的主体结构则是由 12 个同样的卷积模块构成。LittleResNet 和 MiddleResNet 都在 ResNet 的基础上进一步增加了中间层，即在每个残差块中增加了一个额外的卷积层，这样做的目的是增加模型的深度和复杂性，进一步提升模型的性能。与 MiddleResNet 相比，LittleResNet 通过减少模型的深度和参数量来实现模型的轻量化，它可以通过减少残差块的数量或减小每个残差块中的卷积层的通道数来减少模型的复杂性，能在保持一定性能的同时减少计算和存储资源的需求，适用于一些资源受限的场景。

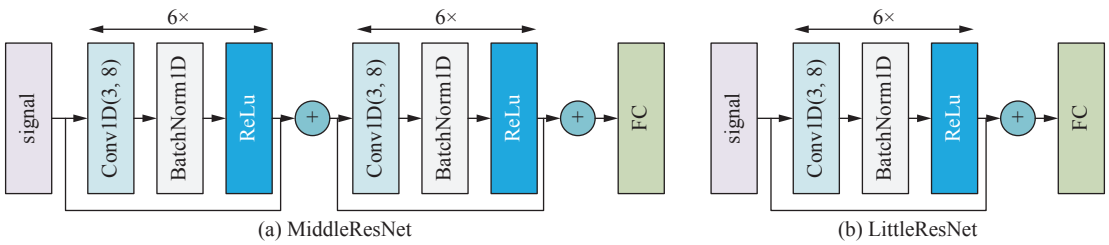


Fig. 1 Network architecture of MiddleResNet and LittleResNet

图 1 LittleResNet 与 MiddleResNet 网络结构

1.3 Grad-Cam 方法

梯度权重激活映射（Grad-CAM）是一种用于阐明卷积神经网络决策过程的可视化技术，它通过生成热力图来揭示输入图像的哪些区域对模型的输出预测产生了重大影响。相较于需要替换最后的分类器并重新训练模型的类激活映射（CAM），Grad-CAM 适用于任意结构的卷积神经网络。Grad-CAM 的主要思想是对分类得到的最大值

反向传播，得到选取特征层的梯度信息，并求得每个特征图对识别结果的贡献值。Grad-CAM 创建热力图的关键步骤如下。

（1）选择目标层。Grad-CAM 方法首先需要选择一个神经网络层来计算热力值。该层通常是网络中接近最后的特征提取层，因为此层具有丰富的空间信息和较高级别的语义信息。具体地，在 MiddleResNet 和 LittleResNet 结构中，本文选取最后一个卷积层作为映射层。而文献 [12] 使用的 Transformer 模型，不适用于此种方法。

（2）前向传播。将输入图像输入神经网络进行前向传播，直到所选择的目标层。

（3）计算梯度。对目标类别得分进行反向传播（即想要可视化的目标类别），计算目标层的特征图相对于目标类别的梯度$G_{i,j,k}^{c}$。这个梯度实质上描述了每个特征图像素值对目标类别影响的程度，计算公式如下

$$G_{i,j,k}^{c}=\frac{\partial y^{c}}{\partial A_{i,j}^{k}}$$

式中：y^{c}为模型的最大输出值，$A_{i,j}^{k}$为选定特征层第 k 个特征图坐标为(i,j)的元素。

（4）计算权重系数。对特征图上所有位置的梯度求平均值，得到每个特征图的权重系数。这个系数度量了每个特征图对目标类别的重要性，其计算公式如下

$$\alpha_{k}^{c}=\frac{1}{Z}\sum_{i}\sum_{j}\frac{\partial y^{c}}{\partial A_{i,j}^{k}}$$

式中：$Z=\sum_{i}\sum_{j}1$，即代表对所有梯度求平均值。

（5）生成 Grad-CAM。使用权重系数对目标层的特征图进行加权求和，得到 Grad-CAM 值。Grad-CAM 的每个像素值代表该位置对目标类别的重要性，其计算公式如下

$$L_{Grad-CAM}^{c}=ReLU\left(\sum_{k}\alpha_{k}^{c}A^{k}\right)$$

式中：$ReLU(\cdot)$代表 ReLU 激活函数，$ReLU(x)=\max(0,x)$。

（6）生成热力图。为了使得 CAM 的尺寸与原始图像一致，需要对 CAM 进行上采样插值到原始信号的长度。然后，将上采样的 CAM 叠加在原始信号上，生成最终的热力图。

这种方式生成的热力图能反映出模型在做出决策时，哪些区域的信息发挥了重要作用，即定位样本的重要区域，这有助于理解样本并解释模型的预测行为。

2 测试方法与结果

2.1 测试方法

比较 LittleResNet、MiddleResNet 模型以及文献 [12] 使用的 Transformer 模型，分析在相同测试样本条件下各模型的预测错误差异。针对测试集所有样本，开展第 1 项测试并进行准确率计算与对比评估；从测试集中各个辐射源对象 14 dB 信噪比时的 200 个样本中，随机选择 30 个样本（分为记为 sample1～sample30）开展第 2～第 4 项测试，即每个辐射源的 30 个测试样本分别依次输入到 3 种模型并输出预测结果，其中测试样本与预测结果记为 gt-*m*_pred-*n*_sample*p*，表示 *m*#辐射源对象的第 *p* 个样本，预测成了 *n*#辐射源对象，如果 *n* 与 *m* 相同，则预测正确；否则，预测错误。

测试 1：模型性能比对。确保模型在相同训练条件下，以测试集准确率为评估标准评价三种模型性能。

测试 2：预测结果比对。在无掩码时，对预测错误的样本进行分析，查看几种模型下预测错误的样本是否相同；

测试 3：重要区域可视化比对。分三种情形对重要区域可视化的差异进行比对，包括：①同一对象、同一模型的重要区域是否相同；②同一对象、不同模型时的重要区域是否相同；③不同对象、同一模型时的重要区域是否相同。

测试 4：掩码测试比对。在无掩码的原始信号基础上，对信号样本中热力值≥q（q 此处设置为 0.92）的样点进行掩码，即把相应样点的值赋为 0，进行第一次掩码后对掩码后的样本进行预测；然后，分别基于上一次掩码后的热力值进行第二次、第三次掩码及预测。

2.2 测试结果

（1）测试 1：模型性能对比

针对三种模型，本文使用训练过程的优化模型和学习率等参数如下：使用余弦衰减学习率调度器，并进行 5 轮

线性预热训练迭代，接着使用 AdamW 优化器进行 300 轮迭代的训练。批量大小为 128，初始学习率为 0.000 1，以及权值衰减为 0.005。在测试集上得到 LittleResNet、MiddleResNet、Transformer 三种模型参数量以及 5、9、14 dB 三种信噪比时和全部 10 种信噪比数据时的识别准确率实验结果如表 1 所示。通过表 1 可以发现，LittleResNet 与 MidlleResNet 的性能较为接近，其中 MiddleResNet 在整个测试集上的准确率仅比 LittleResNet 高 1.18%，但 LittleResNet 的参数量仅为 MiddleResNet 的 60%。此外，Transformer 模型的性能明显优于其他模型，但模型参数量也随之数倍增加。LittleResNet 模型结构更简洁，参数量与计算复杂度更低，而 MiddleResNet 参数量与计算复杂度较高，准确率有一定提升，可根据实际的算力能力与计算时延、准确率等需求情况，来选择不同的模型。

表 1 三种模型的性能对比结果

Table 1 Performance comparison results of three models

model	accuracy				parameter number/10^6
	5 dB	9 dB	14 dB	all data	
LittleResNet	0.760 5	0.877 5	0.960 5	0.881 2	0.153
MiddleResNet	0.766 5	0.899 0	0.968 5	0.893 0	0.254
Transformer	0.867 0	0.949 3	0.995	0.943 9	1.021

(2)测试 2: 预测结果对比

针对 MiddleResNet 和 LittleResNet 两种模型，并对比文献 [12] 采用的 Transformer 模型，得到各对象 30 个样本时预测错误的数量和序号，如表 2 所示。可以看出，Transformer 模型预测的准确率最高，MiddleResNet 和 LittleResNet 预测的准确率相近，符合测试(1)中的性能对比结果；LittleResNet 在某些类别上看似甚至略优于 MiddleResNet 模型，跟测试 1 对比可以分析得出，这种结果应是由于使用的测试样本量较少的实验偶然而导致。

表 2 预测错误的样本数量及其序号

Table 2 Number and sequence of mispredicted samples

emitter	Transformer	MiddleResNet	LittleResNet
1#	1 (sample23)	4 (sample3,5,12,17)	4 (sample5,10, 12, 23)
2#	0	7 (sample3,5,7,14,19,22,28)	6 (sample3,6,7,19,22,28)
3#	0	0	0
4#	1 (sample19)	2 (sample7,19)	2 (sample7,19)
5#	1 (sample9)	1 (sample14)	0
6#	0	0	0
7#	0	0	0
8#	0	0	0
9#	0	0	0
10#	0	0	0

为了解释样本与模型，可进一步分析具体哪些样本在使用何种模型时识别错误。通过表 2 可以发现，MiddleResNet 和 LittleResNet 预测结果中有较多重合的误预测样本，这说明两种模型具有相似性；同时发现，有少量样本预测结果不同，说明两种模型存在一定差异。此外还发现在预测 5#辐射源时，Transformer 和 MiddleResNet 都有一个样本预测出错，而 LittleResNet 都能准确预测，这说明整体识别性能最高的模型，针对具体某个对象的预测准确率不一定最高。同时还可以看出，3#以及 6#-10#辐射源样本都预测正确，说明这几个对象特征更明显。

(3)测试 3: 重要区域可视化对比

图 2 是 MiddleResNet 模型和 LittleResNet 模型对 6#辐射源对象的 sample1 和 sample5 两个样本的重要区域定位呈现，热力值分布在 [0,1] 区间，越红表示该区域重要性越高，越蓝表示该区域重要性越低。可以看出，同一对象、同一模型的不同样本，如图 2(a)、(c)所示，这两个基于 MiddleResNet 模型的样本虽然信号起止点位置不同，但计算出的重要区域相对位置十分相似；同样的，如图 2(b)、(d)所示，基于 LittleResNet 模型的的两个样本，重要区域也较为相似，这说明同一对象的不同样本在同一模型下的重要区域具有相似性。另外，对比图 2(a)、(b)为

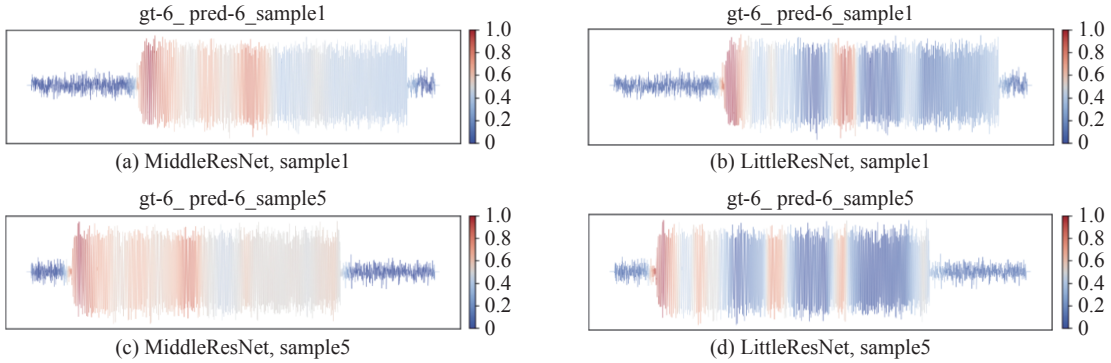


Fig. 2 Comparison of time-domain pulse signal heat maps between MiddleResNet and LittleResNet

图 2 两种模型的时域脉冲信号热力图对比

sample1 样本在两种模型下的重要区域呈现，发现其在整体区域虽有一定的相似性，但重点区域的位置和宽度有一定的差异，说明不同模型提取的特征区域会有差异。

采用与图 2 相同的热力值参数，对 MiddleResNet 模型和 LittleResNet 模型对 1#～10#各个辐射源对象的 sample1 的重要区域进行呈现对比，如图 3 所示。从图中可以看出，因为调制方式的差异，信号特征差异可区分为两类，其中辐射源 1#～5#可以明显归为一类，6#～10#归为另一类。6#～10#因为使用线性调频信号，频率随时间在线性变化，其特征区域也更多地呈现出了随频率变化的分段特征，而 1#～5#是单频率脉冲，主要呈现的是头部但并不稳定的瞬态特征。这也可以说明 6#～10#辐射源因为原始信号时频特征更明显，所以预测准确率更高。

此外从图 3 还可以看出，针对大部分对象，两种模型对同一对象的重要区域定位大部分相似，只是热力图颜色深浅有一定差异。而针对 3#辐射源对象，两种模型的热力图呈现的重要区域定位结果相反，其中 MiddleResNet 模型定位到了脉冲信号段，LittleResNet 定位到的是噪声信号段，虽然这可能并不会影响对测试集的预测准确率，但当信道条件发生变化时，噪声信号段随之发生改变，LittleResNet 模型预测结果很可能出错。这是因为 LittleResNet 模型关于 3#学习到的特征主要反映了该辐射源信号样本中噪声部分与其它辐射源信号的差异，而非实质的指纹特征；当信号的信噪比降低时，这种差异会因为更大的噪声干扰而减弱，进而影响模型的分类识别。与之相反，参数量更大的 MiddleResNet 模型能够更好地定位信号中的指纹特征部分。尽管两种模型在以准确率为评价指标的性能对比上表现接近，但通过本文提出的可视化分析方法可以评估得出 MiddleResNet 模型提取到了更接近指纹特征的辐射源特征。这验证了本文所提可视化方法可以从更深层的角度来分析模型是否提取到了本质的指纹特征，而非仅依靠准确率来衡量模型的特征表示能力。

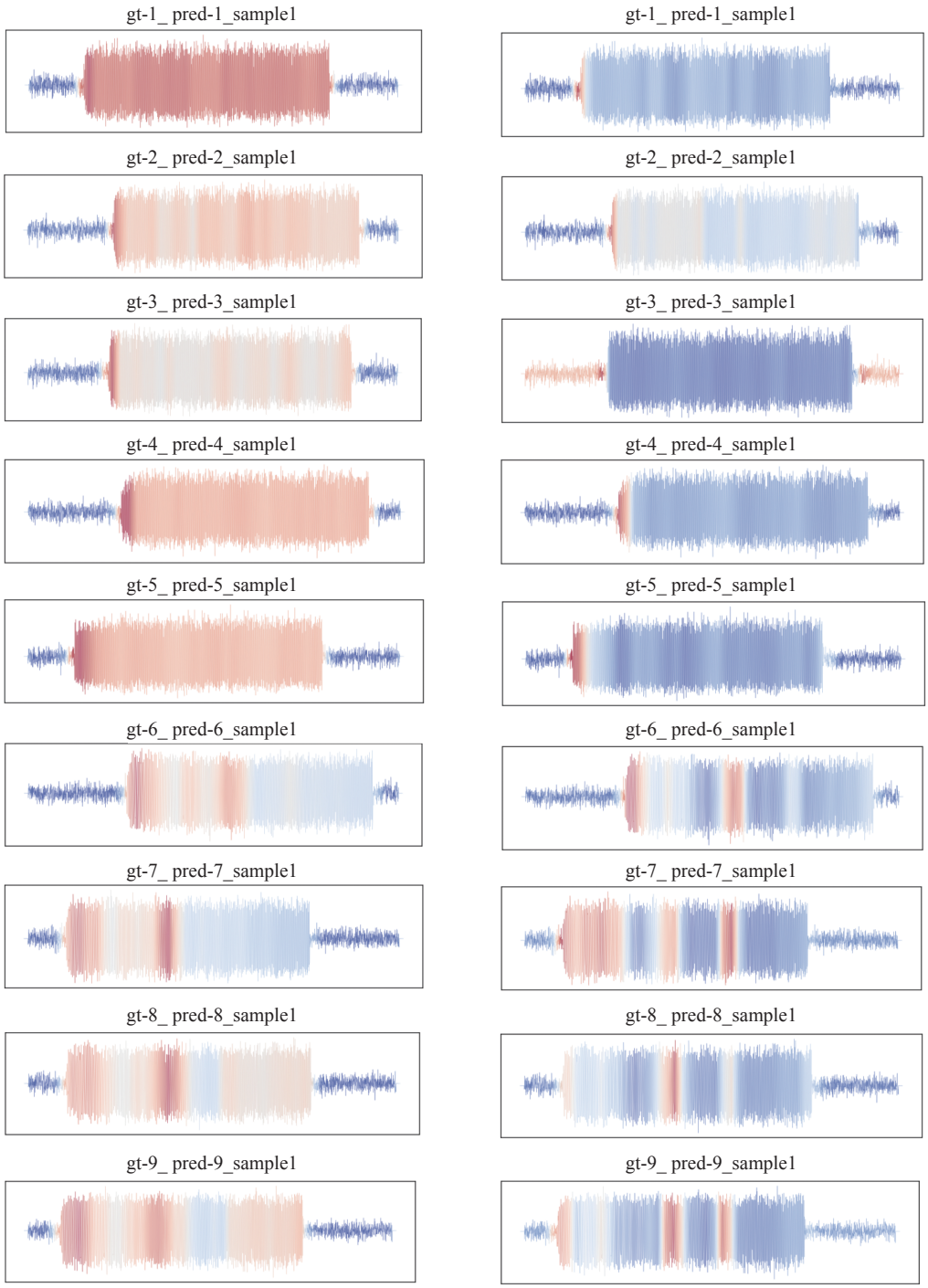
(4)掩码测试对比

基于 3 次掩码测试的发现，进行结果呈现分析。如对 3#辐射源进行掩码分析，在 MiddleResNet 模型下进行掩码时，结果如表 3 所示。该表共 30 列表示 30 个样本，第 1 行表示样本序号，第 2 行～第 5 行分别表示无掩码和第 1～3 次掩码。预测错误时，表格用红色填充并列出预测出的对象序号。可以看出，3#辐射源在掩码后主要被预测成了 2#和 1#辐射源。其中，3#辐射源的 sample3 样本在第 1 次和第 2 次掩码时预测成了 2#辐射源，但在第三次掩码后又预测成了 3#，如图 4(a)所示。

在实在过程中发现：(1)发现 3#辐射源所有样本在 LittleResNet 模型下进行掩码时，仍然全部预测成了 3#，如图 4(b)所示，这很可能是因为 3#样本在 LittleResNet 模型下重要区域定位在噪声段，即使遮掩了部分噪声段，其它噪声段依然存在，因此不会影响预测结果。而在 MiddleResNet 模型下，掩码遮挡了 3#的 sample3 样本部分头部位置时，会导致预测失败，这是因为其重要区域定位在头部，遮掩真正的重要区域会导致预测结果出错；(2)进行掩码测试时，统计从无掩码到 3 次掩码时各对象的 30 个样本中预测的错误样本数量和主要误预测成的目标对象，见表 4 中各列所示。可以看出，两种模型的错误预测样本的数量是相似的，易被误预测成的对象也是相似的，这说明两种模型针对相同原始样本和掩码后样本具有近似的预测能力。

3 结 论

本文通过基于 Grad-CAM 进行信号样本重要区域定位测试与掩码测试，得到的结果分析如下：(1)针对 LittleResNet、MiddleResNet 模型和典型辐射源信号样本，进行基于射频指纹的个体识别以及基于 Grad-CAM 的热



(a) MiddleResNet (b) LittleResNet

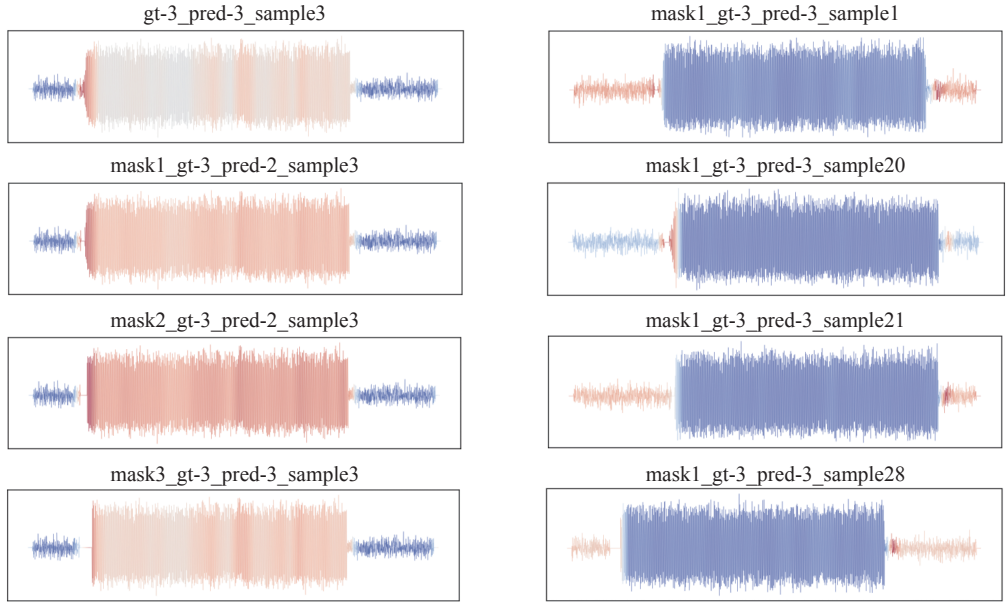
Fig. 3 Important areas of emitters when using MiddleResNe and LittleResNet

图 3 MiddleResNet 模型和 LittleResNet 模型的各对象的重要区域呈现

表 3 MiddleResNet 模型 3#辐射源对象掩码预测结果

Table 3 Prediction results when masking 3# and using the MiddleResNet model

sample	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
no masking	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
1st masking	3	3	2	3	3	3	3	3	1	3	3	1	3	3	3	3	3	3	3	3	3	3	3	3	2	3	2	3	3	3
2nd masking	3	3	2	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	2	3	3	3
3rd masking	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	2	3	3	3	2	3	3	3



(a) 0-3rd mask results using MiddleResNet for sample3　　(b) 1st mask results for samples using LittleResNet

Fig. 4　Test results by continuous mask and different samples comparison

图 4　样本掩码测试对比

表 4　无掩码到 3 次掩码时误预测样本数量与误预测成的对象

Table 4　Number of mispredicted samples and emitter(s) mainly mispredicted to when masking different times

emitter	MiddleResNet					LittleResNet				
	no masking	1st masking	2nd masking	3rd masking	emitter(s) mainly mispredicted to	no masking	1st masking	2nd masking	3rd masking	emitter(s) mainly mispredicted to
1#	4	27	29	29	2#	4	10	10	10	2#
2#	7	16	19	19	1#、3#	6	10	14	10	1#、3#
3#	0	5	2	2	2#、1#	0	0	0	0	——
4#	2	13	27	29	3#	2	12	16	10	3#、2#、5#、1#
5#	1	20	29	29	3#、4#	0	5	14	17	3#
6#	0	13	28	30	7#、10#	0	15	23	25	7#、10#
7#	0	7	9	16	10#、3#	0	1	1	7	10#、3#
8#	0	9	20	28	10#、7#、9#、3#	0	7	13	16	10#、7#、6#、9#
9#	0	8	14	17	10#、7#、3#	0	6	13	16	10#
10#	0	2	11	15	7#、3#	0	2	3	4	7#

力图呈现是可行的，对样本的重要特征区域定位是可行的；(2)发现同一辐射源在使用近似的深度学习模型时，其重要区域定位结果是相似的，而不同辐射源发送同类信号时其重要区域定位结果是有差异的，热力图的差异能反映辐射源个体特征的空间距离，以及不同模型的定位准确度差异；(3)掩码能够用于测试分析重要区域的作用，证明射频指纹存在重要区域位置，同时发现对重要区域的掩码更容易产生攻击效果，而多次掩码的预测结果变化类似于攻击测试，可以辅助定位影响射频指纹样本识别的扰动点。

后续可以改进和提升几个方面，如：在本文所提方法基础上，拓展更多的信号类型、特征呈现方式以及深度学习模型；分析引起重要区域定位到噪声段的原因；以更加智能的方式来进行干扰或掩码，更快更准确地发现重要特征位置；分析不同区域对信号射频指纹的贡献度，并对模型的鲁棒性以及数据集的构建提出优化方法。

参考文献：

[1] Al-Shawabka A, Restuccia F, D'Oro S, et al. Exposing the fingerprint: dissecting the impact of the wireless channel on radio fingerprinting[C]//Proceedings of the IEEE INFOCOM 2020 - IEEE Conference on Computer Communications. 2020: 646-655.

[2] Liu W B, Fan P Z, Wang M H, et al. Optical, acoustic and electromagnetic vulnerability detection for information security[J]. Journal of Physics:Conference Series, 2021, 1775: 012001.

[3] 刘文斌, 丁建锋, 寇云峰, 等. 物理隔离网络电磁漏洞研究[J]. 强激光与粒子束, 2019, 31: 103215.（Liu Wenbin, Ding Jianfeng, Kou Yunfeng, et al. Research on electromagnetic vulnerability of air-gapped network[J]. High Power Laser and Particle Beams, 2019, 31: 103215）

[4] José A. Gutiérrez del Arroyo Pérez. Learning robust radio frequency fingerprints using deep convolutional neural networks[D]. USA: Air Force Institute of Technology, 2022.

[5] Yang Zhou, Liu Ninghao, Hu Xiaben, et al. Tutorial on deep learning interpretation: a data perspective[C]//Proceedings of the 31st ACM International Conference on Information & Knowledge Management. 2022: 5156-5159.

[6] Srivastava G, Jhaveri R H, Bhattacharya S, et al. XAI for cybersecurity: state of the art, challenges, open issues and future directions[DB/OL]. arXiv preprint arXiv: 2206.03585, 2022.

[7] 李辉. 基于类激活图的卷积神经网络可视化解释方法研究[D]. 长春: 吉林大学, 2023.（Li Hui. Research on visual interpretable method of convolutional neural networks based on class activation mapping[D]. Changchun: Jilin University, 2023）

[8] Selvaraju R R, Cogswell M, Das A, et al. Grad-CAM: visual explanations from deep networks via gradient-based localization[C]//Proceedings of 2017 IEEE International Conference on Computer Vision. 2017: 618-626.

[9] Kim J, Oh J, Heo T Y. Acoustic scene classification and visualization of beehive sounds using machine learning algorithms and grad-CAM[J]. Mathematical Problems in Engineering, 2021, 2021: 5594498.

[10] 梁先明, 倪帆, 陈文洁, 等. 基于时频 Grad-CAM 的调制识别网络可解释研究[J/OL]. 西南交通大学学报, 2022. https://kns.cnki.net/kcms/detail/51.1277.u.20220608.1636.008.html.（Liang Xianming, Ni Fan, Chen Wenjie, et al. Interpretability of modulation recognition network based on time-frequency grad-CAM[J/OL]. Journal of Southwest Jiaotong University, 2022. https://kns.cnki.net/kcms/detail/51.1277.u.20220608.1636.008.html.）

[11] 倪帆. 基于可解释深度学习的通信信号调制识别算法研究[D]. 成都: 西南交通大学, 2022.（Ni Fan. Research on communication signal modulation recognition algorithm based on interpretable deep learning[D]. Chengdu: Southwest Jiaotong University, 2022）

[12] 刘文斌, 范平志, 李雨锴, 等. 辐射源个体识别的一种可解释性测试架构[J]. 太赫兹科学与电子信息学报, 2023, 21(6): 734-744.（Liu Wenbin, Fan Pingzhi, Li Yukai, et al. An interpretable testing architecture for specific emitter identification[J]. Journal of Terahertz Science and Electronic Information Technology, 2023, 21(6): 734-744）

[13] He Kaiming, Zhang Xiangyu, Ren Shaoqing, et al. Deep residual learning for image recognition[C]//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. 2016: 770-778.

[14] Xiao Yao, Wei Xizhang. Specific emitter identification of radar based on one dimensional convolution neural network[J]. Journal of Physics:Conference Series, 2020, 1550: 032114.

[15] Wu Bin, Yuan Shibo, Li Peng, et al. Radar emitter signal recognition based on one-dimensional convolutional neural network with attention mechanism[J]. Sensors, 2020, 20: 6350.