

# 面向复杂频谱环境的多模态融合与大模型范式协同的弱信号智能识别研究

BX2504513 张海仁

## 研究背景与问题定义

在复杂电磁环境中，多频段、多设备同时运行使得频谱空间呈现出高度非平稳与强噪声特性。微弱信号在实际采集过程中普遍存在信噪比较低、可观测性不足以及类别间特征高度重叠等问题，给自动化识别与监管带来了显著挑战。传统基于单一特征空间的建模方法难以全面刻画信号在时域、频域及时频耦合域中的综合特性，其性能在小样本和复杂背景条件下受到明显制约。

随着大模型技术的发展，多模态联合建模逐渐成为提升复杂任务理解能力的重要手段。大模型通过在不同模态之间建立统一的表示空间，实现跨信息源的协同推理，为复杂感知任务提供了新的研究范式。在频谱信号领域，如何将物理可解释的专家特征与深度学习自动提取的高维表征进行有效融合，是推动大模型思想落地的关键问题之一。基于上述背景，本研究围绕复杂频谱环境下的微弱信号识别任务，构建了一种以多模态融合为核心的智能识别框架，在方法论层面与大模型的多模态建模思想形成了紧密呼应。

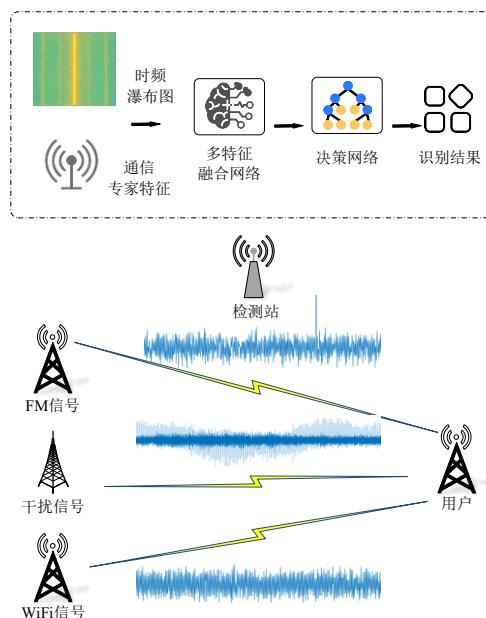


图 1 基于多模态融合的小样本微弱信号识别方案模型图

\*代码上传至 <https://github.com/lebusishu515/Spectrum-Classification-System-Based-on-Multi-modal-Fusion>

## 数据集构建

本研究所使用的原始数据集，源自校园内利用多接收天线节点对于多频段采集得到的实测操场频谱信号，频段主要集中于  $99\text{ MHz}$ 、 $125\text{ MHz}$  和  $2.45\text{ GHz}$ 。对于三种频段内的六类信号分别采集时长十分钟左右，且在采集时间上有所区分。六类信号主要参数如下：

### (1) FM 信号

FM 信号代表的典型信号类别是为黑广播信号，其为一种在窄频带出现的较规律的功率信号。因其非法性，黑广播信号往往功率较低，极容易淹没于强噪声中无法得到有效监管。采集时间为下午三点五十五分，选取的黑广播的中心频率约为  $99\text{ MHz}$ ，持续时间约为 10 分钟。由于采集设备本身的限制，每分钟可以采集 60 个独立且连续的数据片段，十分钟共形成了 600 条数据用于后续识别。

### (2) 噪声信号

噪声信号采集，来源于对非繁忙通信时段，WIFI 设备未开机状态下对无强信号源波段的监测，采集于下午四点二十五分。其来源是大气噪声、热噪声等对采集设备的影响。对应中心频率约为  $2.45\text{ GHz}$ ，采集区间为  $125\text{ MHz}$ ，持续时间约为 10 分钟，共生成数据 600 条。

### (3) $1\text{ MHz}$ 间隔的梳状波发生器信号

梳状波干扰，源于监测区域内放置的梳状波干扰源，采集于晚上八点十分左右。其中心频率为  $500\text{ MHz}$ ，采集区间为  $125\text{ MHz}$ ，其干扰样式为梳齿状，间隔为  $1\text{ MHz}$ 。实测时长约 10 分钟，共生成数据 600 条。

### (4) $5\text{ MHz}$ 间隔的梳状波发生器信号

采集于晚上八点二十分左右。与  $1\text{ MHz}$  间隔的梳状波干扰信号类似，中心频率和采集区间一致，间隔变为  $5\text{ MHz}$ 。实测时长约 10 分钟，共生成数据 600 条。

### (5) 通信繁忙时的 WIFI 信号

通信繁忙时段的 WIFI 信号，其采集时间选取在下午三点三十分左右，属于正常工作、学习的时间，因此通信较为频繁，用户请求较为拥挤。实测 WIFI 信号，中心频率

$2.45\text{ GHz}$ ,采集区间为  $125\text{ MHz}$ 。持续时间 10 分钟左右，生成数据 600 条。

#### (6) 通信空闲时的 WIFI 信号

较空闲时段 WIFI 信号，采集时间选取在凌晨六点零六分左右，属于清晨非工作时间，因此通信较为空闲，用户请求较少。实测此时的 WIFI 信号，中心频率  $2.45\text{ GHz}$ ，采集区间为  $125\text{ MHz}$ 。持续时间 10 分钟左右，生成数据 600 条。

对上述六种信号，分别将 I/Q 两路时域信号转化到频域的同时，选用了峭度、偏度、波形因子、峰值因子、脉冲因子、裕度因子、重心频率、频率均方根、频率标准差以及  $3\text{ dB}$  带宽等十个典型信号特征进行提取，形成六类信号对应的专家特征向量集。

### 多模态特征体系构建与融合动机

频谱信号本质上同时包含结构化物理信息与非结构化时频分布信息。单纯依赖时域或频域统计特征，虽具有明确物理意义，但难以表达信号随时间演化的动态模式；仅利用时频图像进行端到端学习，则可能忽略通信系统中已有的领域先验知识，降低模型在小样本条件下的稳定性。

本研究通过构建多模态特征体系，将频谱信号从两个互补视角进行表征。一方面，基于通信领域专家知识，从时域和频域中提取统计特征和物理特征，形成低维、结构化的专家特征向量。另一方面，通过短时傅里叶变换将原始信号映射为时频瀑布图，以图像形式刻画信号在时间与频率维度上的联合分布特性。进一步，引入数字孪生技术生成高质量补充样本增强训练。

#### 1、时频专家特征提取

时域分析是研究信号在时间维度演变过程的重要手段。通过对信号在不同时间点上的动态变化进行描绘，时域分析能够深入了解信号的行为特性，从而为后续的信号处理、识别和分类环节提供基础与数据保障。所得参数分为有量纲参数和无量纲参数两种。

有量纲的时域特征主要关注信号的整体性能。例如，最大值反映了信号的峰值，均值描述了信号的平均水平，方差和均方根提供了关于信号波动性的信息。这些参数

在描述信号的基本特性时非常有用，能够直观地揭示信号的幅度和波动情况。与此同时，无量纲的时域特征更注重信号的波形形状。峰值因子反映了信号的尖锐程度，脉冲因子描述了信号的脉冲特性。裕度因子和波形因子提供了关于信号形状的信息，偏度和峭度关注信号的分布特性。这些无量纲特征参数对于刻画信号的形态和特异性非常关键，有助于更全面地理解信号的结构和特征。

在实际应用中，这些时域特征参数不仅是对微弱信号进行定量分析的工具，更是信号处理和微弱信号识别领域的重要特征。它们为建立信号模型、进行信号识别分类提供了丰富的信息基础。因此，深入理解和运用时域分析中的有量纲和无量纲特征，对于提升信号处理技术和微弱信号识别算法的性能具有重要意义。下面对使用到的时域特征参数提取方法分别进行介绍：

### (1) 均值

均值  $\mu$  是信号的平均值，是信号的一阶矩。对信号绝对值求平均，得到的值  $MA$  称为信号的平均幅值。分别表示为：

$$\mu = \frac{1}{N} \sum_{n=1}^N x(n),$$

$$MA = \frac{1}{N} \sum_{n=1}^N |x(n)|,$$

其中， $N$  表示信号采集点数， $x(n)$  表示输入信号。

### (2) 方差

方差  $\sigma^2$  是描述信号波动程度的统计量，作为信号的二阶中心矩，其用公式表示为：

$$\sigma^2 = \frac{1}{N} \sum_{n=1}^N [x(n) - \mu]^2,$$

方差越大，表示信号的样本值越分散，波动越大。

### (3) 均方根

均方根  $RMS$  是信号的有效值，其表示为：

$$RMS = \sqrt{\left[ \sum_{n=1}^N x^2(n) \right] / N},$$

均方根是一种对信号振幅的有效度量，其常用于电信号、音频信号等领域，用于描述信号的实际振幅水平，相较于平均幅值，均方根更全面地反映了信号的强度。

#### (4) 峰值因子

峰值因子  $C_{pf}$  是信号峰值与有效值的比值，反映了信号波形峰值的极端程度。其表示为：

$$C_{pf} = |x(n)|_{\max} / RMS,$$

其中， $|x(n)|_{\max}$  表示信号的峰值。

#### (5) 脉冲因子

脉冲因子  $C_{if}$  是信号峰值与平均幅值的比值，其表达式为：

$$C_{if} = |x(n)|_{\max} / MA,$$

脉冲因子提供了有关信号波形的额外信息，特别关注了信号峰值与整体振幅的关系。较高的脉冲因子表明信号包含尖锐而瞬时的峰值，较低的脉冲因子则表示信号峰值相对平滑。在不同的工程和信号处理应用中，脉冲因子可以用于衡量信号的尖锐程度，在一些需要考虑信号波形特性的系统设计和分析工作中具有广泛的应用场景。

#### (6) 裕度因子

裕度因子  $C_{mf}$  是信号峰值与方根幅值的比值，同样用于反映信号的波形特性。其公式为：

$$C_{mf} = N \cdot |x(n)|_{\max} / \sqrt{\sum_{n=1}^N x(n)},$$

较高的裕度因子通常表明信号包含更尖锐的峰值，较低的裕度因子表明信号峰值相对平滑。

### (7) 偏度

偏度  $S_s$  作为信号时域形式的三阶标准矩，主要用于反映信号时域分布的不对称性，可以在一定程度上体现信号时域波形的特点，其计算公式如下：

$$S_s = \frac{E[(x - \mu)^3]}{\sigma^3},$$

其偏度衡量了信号分布的不对称性。当偏度为正时，表示信号分布偏向于右侧，尾部在右侧延伸；当偏度为负时，表示信号分布偏向于左侧，尾部在左侧延伸。偏度为零，表示信号分布相对对称。

### (8) 峭度

峭度  $K$  与偏度之间存在一定的相关性。同样用于描述信号时域波形的分布状况，反映了信号时域分布的陡峭程度。其计算公式如下：

$$K = \frac{E[(x - \mu)^4]}{\sigma^4},$$

峭度反映了信号分布的尾部形状，峭度较大表示信号分布的尾部更陡峭，有更多的极端值；峭度较小则表示分布的尾部相对较平缓。

频域是描述信号频率结构及不同频率信号幅度关系的领域，频域特征涵盖了重心频率、均方频率、频率方差以及  $3 dB$  带宽等参数。

以下是主要频域特征参数的含义及计算方法：

#### (1) 重心频率

重心频率  $FC$  基于数学中计算重心的方法，衡量了信号频域的分布特征，其公式如下：

$$FC = \frac{\sum_{k=1}^N f_k A_k}{\sum_{k=1}^N A_k},$$

其中， $A_k$  表示频率点  $f_k$  处对应的信号幅值。

#### (2) 均方频率

均方频率  $MSF$  与  $FC$  类似，用于展现信号频域的能量分布特征，其计算公式如下：

$$MSF = \frac{\sum_{k=1}^N f_k^2 A_k}{\sum_{k=1}^N A_k},$$

在均方频率中，权重是每个频率分量幅度的平方。均方频率提供了对频谱分布的另一种描述，强调了频率分布的能量加权平均。

### (3) 频率方差

频率方差  $VF$  表示了信号在频域分布上的离散性，其计算公式如下：

$$VF = \frac{\sum_{k=1}^N (f_k - MSF)^2 A_k}{\sum_{k=1}^N A_k},$$

### (4) 3 dB 带宽

3 dB 带宽  $W_{3dB}$  是指信号功率谱密度降低到最大值的一半时对应的频率范围。在频谱图中，3 dB 带宽通常是由功率谱密度曲线的最高点向下降低  $dB$  的两个频率点确定的。这个范围反映了信号的频域分布，尤其是关注信号的主要频率成分，其表示如下：

$$W_{3dB} = f_{upper} - f_{lower},$$

其中， $f_{upper}$  和  $f_{lower}$  分别是频谱中幅值降到峰值一半的上下边界频率。

### (5) 带宽

带宽  $W$  即信号传输过程中的占用频带宽度。在进行信号特征求解计算时，通常利用算法对信号的 90% 功率占用带宽进行计算，以体现特定信号在频域上的功率分布范围。这一计算通常涉及对功率谱密度曲线的积分或求和操作，以衡量信号在频域上的主要分布情况。

## 2、时频瀑布图

在时域信号向频域信号转换的过程中，常常需要借助傅里叶变换进行计算：

$$F(\omega) = \mathbb{F}[f(t)] = \int_{-\infty}^{+\infty} f(t)e^{-j\omega t} dt,$$

其中， $F(\omega)$  表示频域信号， $\mathbb{F}[\cdot]$  表示傅里叶变换， $\omega$  表示频率， $f(t)$  为时域信号。傅里叶变换常用于处理和分析统计学意义上的平稳信号，以用于生成频域信号。傅里叶正变换，表示从时空维度中提取的频谱信号，使用了从整个时间范围内获取到的信息，计算出某个特定频率值的频谱对应的函数值。这是由于积分范围来自选取的完整时间长度，其得到的信号信息对应所使用时间轴的全部范围，而不能轻易获得局部时间域内的特征。并且，傅里叶变换本身考虑的是无限长的时间维度，在实际编写代码中难以实现，所以往往使用短时傅里叶变换（Short Time Fourier Transform, STFT）进行代替。其思想核心是实现局部的平稳化，即将时间上持续较长的非平稳随机过程看作对许多短时段内的随机平稳过程进行叠加。其中，短时性是根据在时域信号上添加窗函数实现的。窗函数首先将原始数据的一部分进行了截断。对于截断之后的信号，进行范围内的变换。这种算法，有效规避了信号在很长时间内某一小部分受损，影响整个信号分析的问题，充分利用了局部信息，且易于实现。其计算过程如下：

$$S[f(\omega, u)] = \int f(t)g(t-u)e^{j\omega t} dt,$$

其中， $S[\cdot]$  表示 STFT 运算， $g(t-u)$  为窗函数。将窗函数与时域信号相乘，可以对信号进行片段的截取，截取后变换，即可得到相应频域信息。

为了更充分地表示频域信息和时域信息，使用时频瀑布图来表征二者之间的联系。所定义的瀑布图实际上反映的是信号功率在不同频点随着时间推移实现层层堆叠，体现在图片上即为：x 轴表示信号的不同频率，y 轴表示不同的时间节点，颜色深浅根据标度对应某个频率下信号功率谱密度的分贝值，即功率与该频点的频率之比。单位为  $dB / Hz$ ，对于六类不同频段信号，分别做出时频瀑布图如图 2 所示。

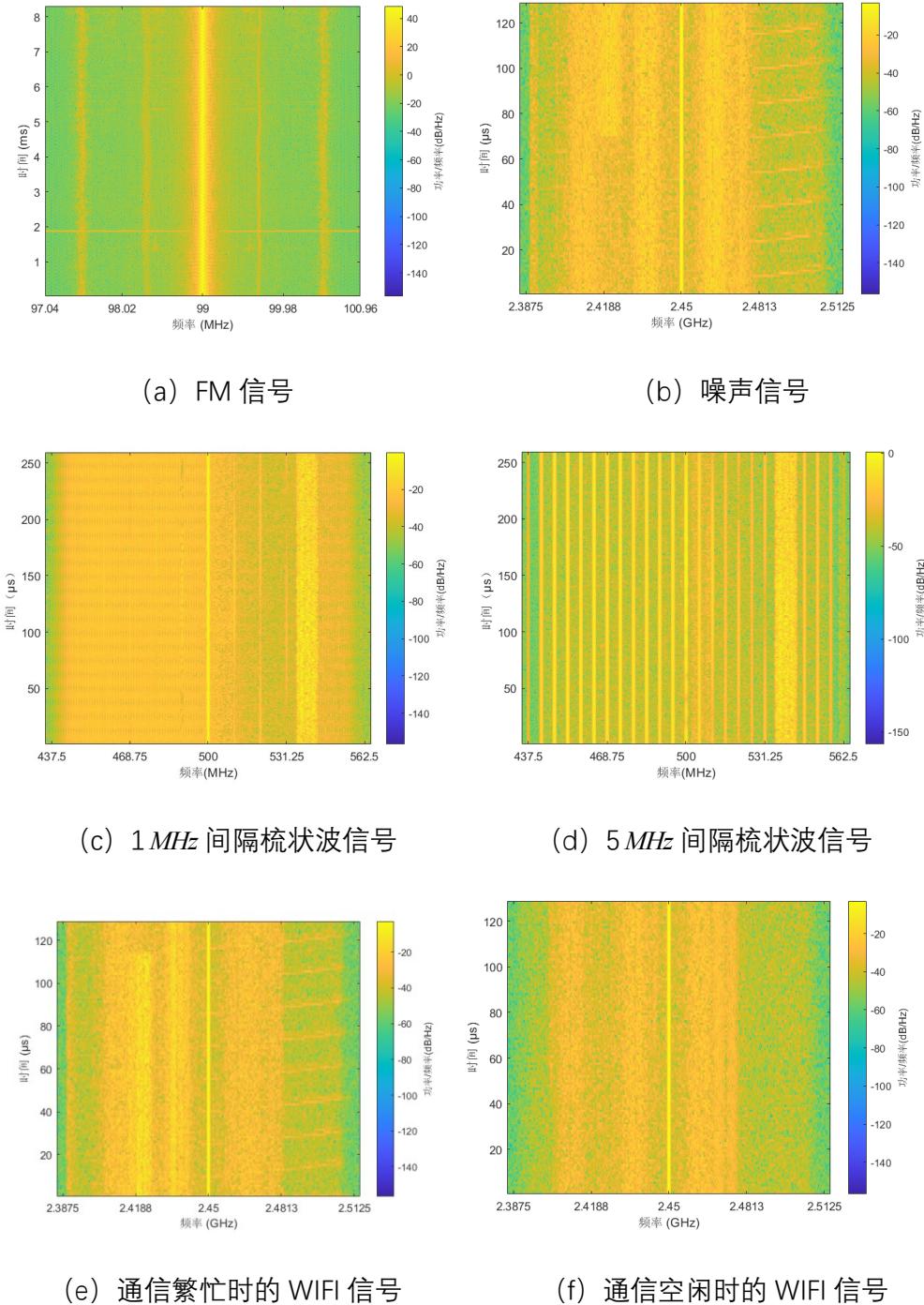


图 2 六类弱信号时频瀑布图

观察图 2 可知, FM 信号的频域特征, 随时间的变化, 其变化并不明显。只是峰值的位置出现略微移动, 但其移动的范围非常小。WIFI 信号由于自身自适应性强、按需分配的特点, 在比较宽的频率范围内都有功率分布, 且分布总体较为均匀。同时, 不同时段的 WIFI 通信信号和底噪很难通过肉眼区分。梳状波信号则有明显的间隔分布特性, 有规律地分布在等间隔的频点处。这些特征与以往对于这些信号的认知一致, 说

明频谱瀑布图是反映信号主要特征的有效方法。这种特征体系的构建为多模态融合奠定了基础，使得模型能够同时感知“信号是什么”以及“信号如何变化”，从而提升整体判别能力。

### 3、频谱数据孪生增强

频谱数据孪生指在通信系统中，通过模拟实际运行中的通信系统产生的数据，以创建与真实数据相似的虚拟数据集，作为实测数据的一种有效补充。数据孪生方法被广泛应用于通信系统的仿真、测试、和算法验证等方面。在多设备多频段信号的采集和处理过程中，由于部分频段获取到的数据质量和数量往往无法得到可靠保障，会出现样本量过少或样本质量不高的问题，对分类效果造成很大影响。

本研究使用频谱数据孪生技术生成了大量仿真条件下的新数据，作为真实采样得到数据集的有效补充，以此丰富样本量。由于干扰信号和 FM 广播信号具有高度实时性，相应频段的高质量信号采样数据往往不足，以具有代表性的 FM 信号和两类梳状波信号为例，生成孪生信号对应的基带频谱瀑布图如图 3 所示。

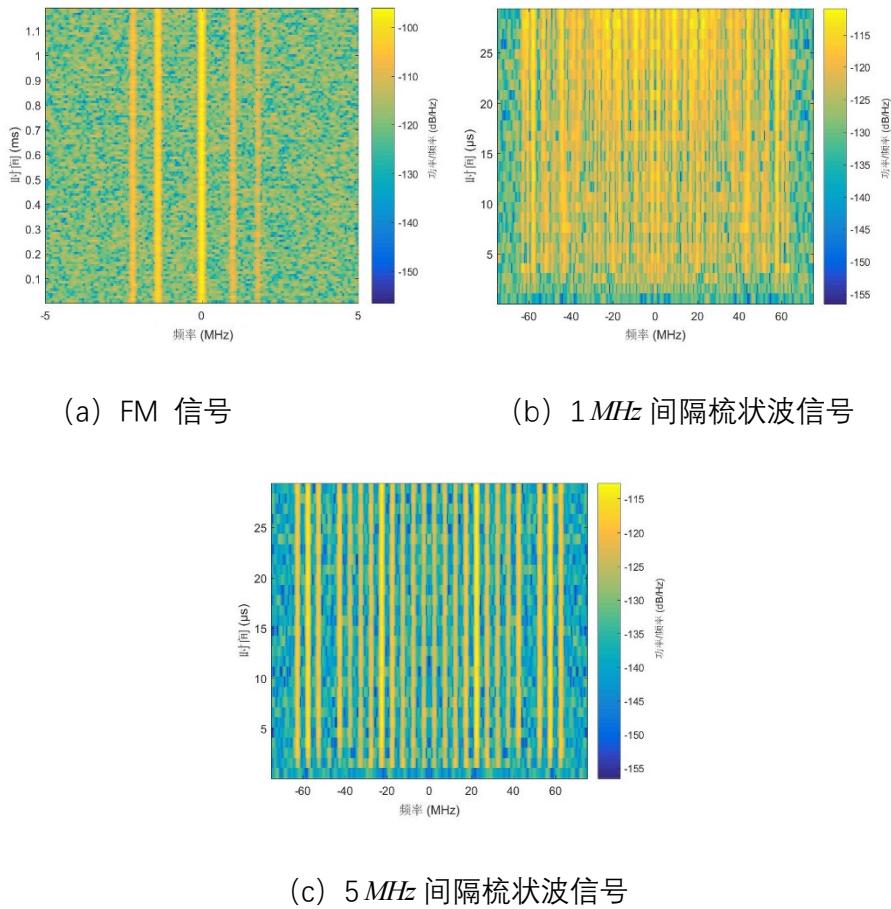


图 3 孪生微弱信号时频瀑布图

观察图 3 可知，孪生微弱信号时频瀑布图具有和实际数据集直接生成的时频瀑布图一致的时频特征。其可作为用于识别任务的多频段实测微弱信号时频瀑布图数据集的有效补充，用于解决网络训练时影响网络性能的部分频段采集到的高质量样本量过少的问题。

### 多模态融合网络设计与大模型关联性分析

在模型层面，本研究设计了一种多模态融合网络结构，用于对结构化专家特征与非结构化时频图像特征进行联合建模。时频瀑布图首先输入轻量化卷积神经网络进行特征提取，卷积网络在该过程中学习到具有判别性的时频模式，可视为一种视觉模态的特征编码器。专家特征则通过数值映射直接进入融合层，保留其物理可解释性。

在特征融合阶段，不同模态的表示被映射至统一的特征空间，并通过全连接网络进行联合学习，实现跨模态信息的深度协同。整体流程结构在图 4 中给出了系统性描述。以往基于神经网络的信号识别方法大多只运用了时域或者频域的单一域信息，对于数据信息的利用并不充分，对于低 SNR 的微弱信号，其没有挖掘出有助于识别精度进一步提升的更多特征。本研究使用了多个变换域的多域多特征融合，使用 CNN 和通信领域专家知识分别进行特征提取，再利用融合策略进行特征的充分表征，对于低 SNR 的微弱信号具有良好的识别能力。同时，使用了数字孪生技术仿真得到的数据进行补充，进一步提高了识别性能。

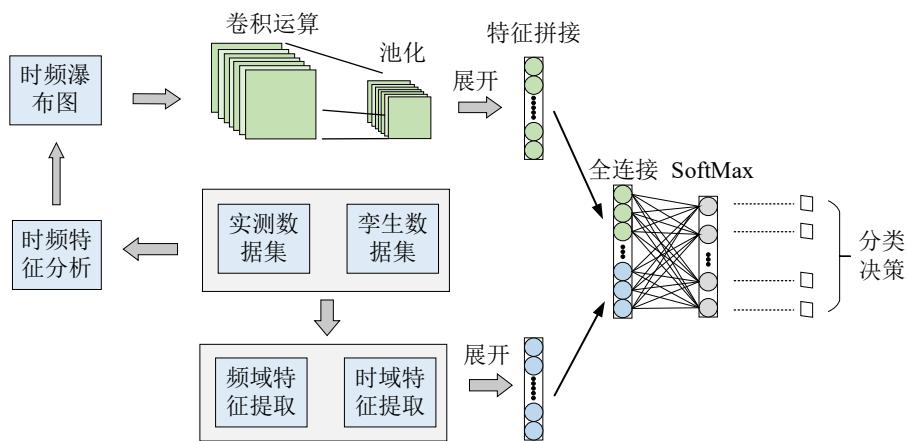


图 4 基于多特征融合的微弱信号识别算法流程图

不同的特征提取方法可以从不同的域或不同维度中提取出接收信号的更多细节，丰富深度神经网络的输入数据，有利于挖掘更多信号的深层特征，以提高微弱信号识别的精度。从大模型方法论视角看，该多模态融合过程体现了多模态大模型的核心思想。卷积特征提取器承担了类似预训练子模型的角色，通过学习通用时频模式为下游任务提供稳定表示；融合网络对应于多模态对齐与微调阶段，使不同信息源在统一语义空间中完成联合推理。相较于单模态模型，该结构在特征表达能力与泛化性能方面具有显著优势。

具体使用的网络模型结构如图 5 所示。

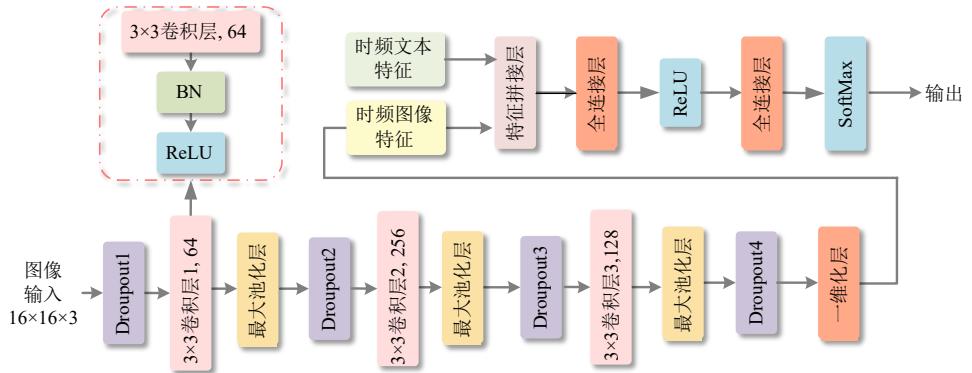


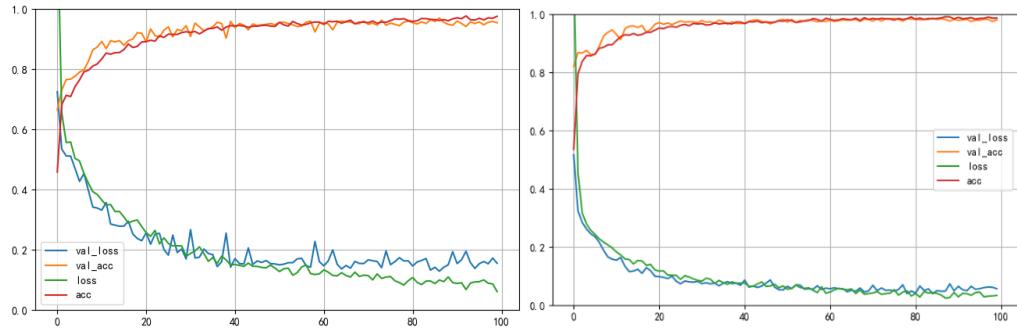
图 5 算法采用的神经网络模型结构图

如图 5 所示，本网络使用的所有卷积核大小都选取为  $3 \times 3$ ，池化核尺寸选取为  $2 \times 2$ ，这受到了 VGGNet 网络的启发。第一层使用了 64 个卷积核；第二层卷积层使用了 256 个卷积核；第三层卷积层使用了 128 个卷积核。全连接层为 2 层，节点数分别为 64 和 10。激活函数使用了 *Relu* 函数，由于网络较为轻量化，因此其收敛速度较快。同时，为了解决可能出现的过拟合问题，在每个卷积层之后、激活层之前还添加了批标准化层（Batch Normalization, BN）层，其核心是在每层卷积之后进行归一化，来解决数据分布发生改变的问题。在机器学习中，初始学习率和训练速度息息相关。往往更高的学习率伴随着更快的收敛速度，但过大的学习率使得模型受到异常数据的影响增大，反而很难收敛。因此，不断调整学习率对于模型训练非常重要。在采用了 BN 层之后，调整学习率的作用被大大弱化，这是由于 BN 层具有很好的快速收敛特性。同时，BN 层在解决过拟合问题中表现突出，这是因为其兼具很强的泛化能力。

## 多模态融合优势的实验验证与基准对比

为验证多模态融合策略的有效性，本研究在统一实验环境下，对基于专家特征的传统机器学习方法、基于时频图像的卷积神经网络方法以及多模态融合方法进行了系统对比。实验结果表明，单一模态方法在部分信号类别上虽能取得一定识别精度，但在信号特征相似或背景噪声复杂的场景中，误判率较高。

选取分类任务中最为常用的四个重要评价指标：训练精度、训练误差、验证精度、验证误差，对网络训练过程进行画图并分析。观察得到 60 次迭代后，模型基本收敛。精度指标以 6 类分类的平均准确率（Mean Average Precision, MAP）来衡量，这一指标来源于对查全率和查准率的综合考察。损失函数值（Loss）的意义则是表示预测值和目标值之间的差距，更小的损失函数标志着预测值更接近目标值。对于整个训练过程，绘制精度曲线、损失函数。如图 6 所示。



(a) 未添加孪生数据集的 CNN 模型      (b) 添加孪生数据集的 CNN 模型

图 6 基于时频瀑布图特征的微弱信号识别算法精度、损失函数图

观察图 6 可知，在进行了 60 次迭代之后，添加孪生数据的 CNN 识别算法，其精度和损失函数值基本都收敛到一个稳定值，验证集和测试集的精度和损失函数基本趋于重合。而未添加孪生数据的 CNN 识别算法，在 80 次迭代后才基本收敛，其收敛速度和分类精度显著低于利用孪生数据丰富了样本后的识别算法。在选取算法和网络模型完全一致的情况下，添加了孪生数据集的方法获得了更好的识别性能，MAP 达到了 97%，证明了利用孪生数据进行样本扩充的有效性。图 7 为本研究提出并重点介绍的基于多特征融合的识别方法，其在未添加孪生数据的情况下，于 50 次迭代后基本收敛，MAP 达到了 98% 以上，高于上述两种方法，显著高于仅基于未添加孪生数据的时频瀑布图或特征向量文本的方法。

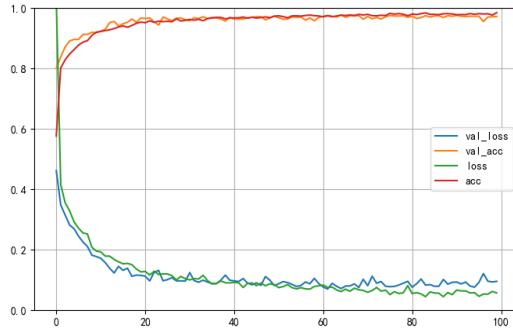
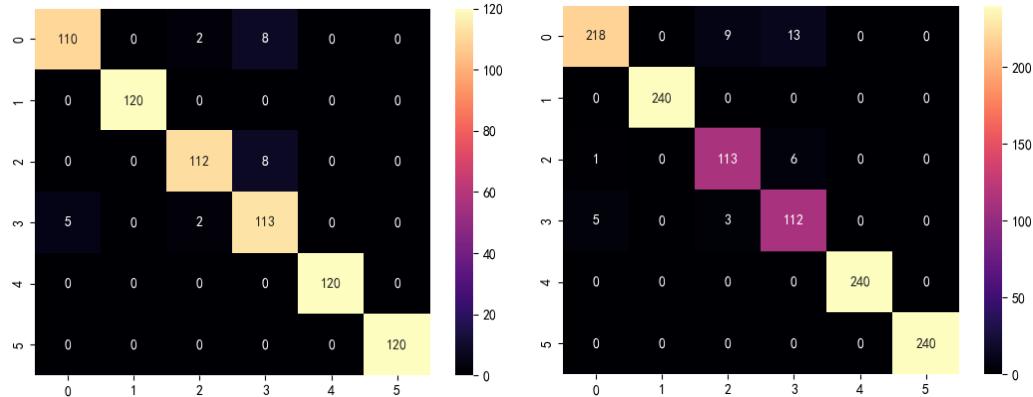
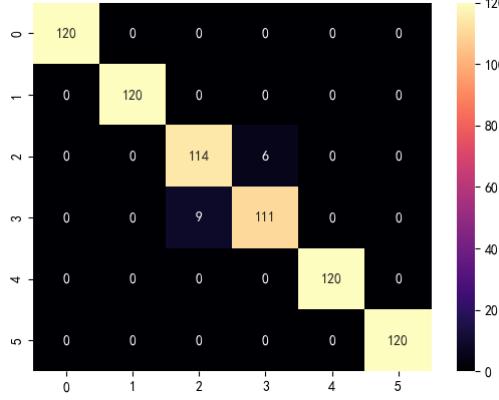


图 7 基于多特征融合网络的微弱信号识别算法精度、损失函数图

引入多模态融合后，模型在整体识别精度、召回率和 F1 指标上均表现出显著提升。尤其在不同通信状态下的 WiFi 信号与噪声信号区分任务中，多模态融合方法有效缓解了特征重叠带来的混淆问题，相关混淆矩阵结果在图 8 中呈现。观察混淆矩阵可知，相较于基础的未扩充数据集的基于 CNN 的识别算法，孪生数据集扩充后的 CNN 网络识别算法和基于多特征融合网络的识别算法均展现出了更好的分类效果，尤其是提出的基于多特征融合网络的微弱信号识别算法，其仅在对两类不同时段采集的 WiFi 信号存在极个别混淆，对于其他四类信号未发生误检和漏检情况，显著优于其他两种识别算法。



(a) 数据集未扩充的 CNN 网络识别算法 (b) 孪生数据集扩充后的 CNN 网络识别算法



(c) 基于多特征融合网络的识别算法

图 8 基于神经网络的微弱信号识别算法混淆矩阵

### 面向大模型嵌入的扩展设想

在现有多模态融合框架基础上，该研究具备良好的大模型扩展潜力。专家特征与时频图像特征可分别作为结构化模态与视觉模态输入，嵌入到更大规模的多模态预训练模型中，通过参数共享和任务微调实现能力迁移。进一步地，数据孪生机制可用于构建大规模合成训练集，为多模态大模型的预训练提供数据支撑。

在系统层面，可引入基于大模型的 Agent 机制，对频谱感知、特征选择与识别策略进行动态调度；同时结合检索增强生成方法，将历史频谱数据、通信协议知识或监管规则作为外部知识源，与多模态特征联合推理，从而实现从“信号识别”向“频谱理解与决策支持”的能力拓展。

### 总结

本研究围绕复杂频谱环境下的微弱信号识别问题，提出了一种基于多特征融合网络的微弱信号识别方法，其设计原则和实施策略旨在构建更全面、更富有信息量的特征表示，以更有效地捕捉和表示微弱信号的特征用于提高识别精度。提出的方法充分利用了专家特征和时频图像特征。该算法首先依靠专家知识对时频信息进行特征向量文本提取，随后生成时频瀑布图，再通过轻量化 CNN 进行图像特征提取。所得特征经过融合网络进行整合，最终经过分类器实现了微弱信号类型的准确识别。实验结果表明，该方法相较于传统方法以及其他基准算法，在微弱信号识别任务中优势显著，获得了 98% 以上的识别精度，展示了其在微弱信号识别领域的卓越性能。同时，为了应对实际频谱环境中多频段多设备采样时，部分频段频谱样本量稀疏的问题，引入了数

据孪生技术，通过生成仿真条件下的频谱数据样本，对实际采样信号数据集进行扩充。仿真结果表明，通过引入孪生数据，微弱信号识别任务的性能得到了显著提升，有效解决了频谱样本量稀疏的问题。

本研究通过构建一种以多模态融合为核心的智能识别方法，联合建模专家特征与时频图像特征，显著提升了识别性能与鲁棒性。从方法论角度看，该研究在频谱信号领域实现了多模态大模型思想的有效落地，为大模型技术在专业工程场景中的应用提供了可推广的研究范式与技术路径。