

# 哈尔滨工业大学深圳校区

## 毕业论文开题报告

题 目 基于深度学习的无线电信号

时频定位与制式识别方法研究

姓 名 黄文浩

学 号 220210809

学 院 信息科学与技术学院

专 业 通信工程

指 导 教 师 张行健副教授

日 期 2025 年 10 月 13 日

## 目 录

1 课题背景及研究的目的与意义 .....	1
1.1 课题背景 .....	1
1.2 研究的目的与意义 .....	1
2 研究现状及分析 .....	3
2.1 国内外研究现状 .....	3
2.2 国内外文献综述及简析 .....	3
2.2.1 传统信号定位与识别方法研究现状 .....	3
2.2.2 基于深度学习的目标检测方法研究现状 .....	5
2.2.3 基于深度学习的信号检测方法研究现状 .....	6
3 主要研究内容及研究方案 .....	8
3.1 研究内容 .....	8
3.1.1 针对不同尺度信号数据的处理方法研究 .....	8
3.1.2 基于 Transformer 检测头的信号检测模型研究 .....	8
3.1.3 针对窄带短时信号的模型优化策略研究 .....	9
3.2 研究方案 .....	9
4 进度安排及预期目标 .....	12
4.1 进度安排 .....	12
4.2 预期目标 .....	12
5 已具备和所需的条件和经费 .....	13
5.1 已具备的条件 .....	13
5.1.1 实验室已具备条件 .....	13
5.1.2 实验室经费保障 .....	13
5.2 所需的条件 .....	13
6 预计困难及解决方案说明 .....	14
6.1 技术难点与预计困难 .....	14
6.2 解决方案 .....	14
参考文献 .....	16

# 1 课题背景及研究的目的与意义

## 1.1 课题背景

随着物联网（Internet of Things, IoT）技术的迅猛发展，无线通信设备数量正呈爆炸式增长。从智能家居、工业自动化到智慧医疗与无人驾驶，海量的无线设备与通信技术在有限的频谱资源中交织共存，构成了日益拥挤且动态多变的复杂电磁环境。在此背景下，信号检测（Signal Detection）技术的重要性日益凸显，它不仅是提升频谱共享效率、降低信号间干扰的关键，也是保障无线通信安全、监控非法信号的必要手段。

与仅关注频段是否被占用的传统频谱感知（Spectrum Sensing, SS）不同，信号检测旨在同时完成两项核心子任务：（1）对环境中的信号进行精确的时频定位（即确定信号在何时占用哪些频谱资源）；（2）对定位到的信号进行准确的制式识别（即判断信号的调制类型或通信协议）。由此可知，信号检测是实现动态频谱接入（Dynamic Spectrum Access, DSA）和保障通信可靠性的关键前提。

然而，能量检测、匹配滤波的传统频段定位方式和基于机器学习的传统信号分类方式由于其对噪声功率敏感、依赖信号先验信息以及计算复杂度高等固有局限，难以完全适应当前非协作、高动态的复杂电磁环境。

近年来，深度学习（Deep Learning, DL）领域的突破性进展为应对上述挑战提供了全新思路。通过将一维时域信号经过短时傅里叶变换（Short-Time Fourier Transform, STFT）转化为二维时频图，信号感知任务得以巧妙地转化为计算机视觉领域中技术成熟的目标检测（Object Detection）问题。这种跨域迁移方法的优势在于，它能够借助目标检测模型实现端到端的处理，在无需精确信号或信道先验知识的条件下，自动学习信号在时频域的深层特征，从而在信号的精确定位与制式分类方面展现出巨大潜力。

## 1.2 研究的目的与意义

尽管计算机视觉领域的目标检测技术已日趋成熟，并涌现出众多先进方案，但其在信号检测领域的应用仍处于初步探索阶段，不仅相关研究成果有限，且现有的技术迁移模式也存在若干亟待解决的问题：

1. 数据集与真实场景存在偏差：目前，该领域的研究大多采用仿真生成或混合真实信号的数据集。这些数据集普遍存在对现实场景的过度简化，例如信号制

式与调制类型有限、信号时频密度较低，且通常忽略了信号间的重叠与干扰。此外，现有数据集普遍假设了固定的采样率与采样时长，这与实际监测中参数多变的情况存在显著差异。虽然这种简化有助于在研究初期快速验证算法的可行性，但却使其难以准确评估模型在真实复杂电磁环境下的性能与泛化能力。

2. 模型迁移缺乏针对性优化：现有的研究在模型迁移方面，大多直接套用计算机视觉领域的经典模型（如 Faster R-CNN），而未根据信号时频图的内在特征（相较于自然图像形态差异巨大）对模型结构进行针对性调整。虽然近期有研究开始尝试对模型进行改进，但仍主要局限于传统的卷积神经网络（Convolutional Neural Network, CNN）架构，未能引入近年来在计算机视觉领域取得突破性进展的 Transformer 架构（如 DETR、Deformable DETR 等），其潜力有待进一步发掘。

因此，本课题旨在针对上述问题展开深入研究。研究将在技术迁移的基础上，充分考虑更为复杂多变的信号检测场景，从数据处理、模型结构设计到训练策略等环节进行系统性的优化。其最终目标是实现目标检测技术与信号检测任务的深度融合，显著提升信号检测模型的准确率与鲁棒性。这一技术在具体实践中将直接赋能两大关键应用：（1）在 DSA 中提高信号对总体时频资源的利用率，同时降低信号之间的干扰；（2）在复杂电磁环境中识别已授权和未授权的信号，并精确定位非法信号，从而保障无线通信的电磁空间安全。

## 2 研究现状及分析

### 2.1 国内外研究现状

在无线通信领域，对宽带频谱中纷繁复杂的信号进行精确、快速的检测与识别，是现代频谱管理、认知无线电及电子对抗等应用的核心前提。针对这一目标，国内外的研究历程可清晰地划分为两个主要阶段：传统信号处理方法阶段与基于深度学习的智能感知阶段。

在传统方法阶段，信号的“定位”与“分类”遵循着两条相互独立的技术路径。一方面，信号定位技术从一维的能量检测、匹配滤波，发展到基于 STFT 的二维时频谱图分析，但这些方法普遍存在泛化能力差、抗噪声性能弱的瓶颈，且无法识别信号的具体制式。另一方面，信号分类技术无论是采用基于专家特征的传统机器学习，还是早期的卷积神经网络，虽然能够识别信号的调制类型，却无法提供其精确的时频位置信息。这种技术上的分离导致两条路径始终未能有效融合。

深度学习目标检测技术的崛起，为实现定位与分类的“一体化”提供了全新范式。该技术从经典的 Faster R-CNN、YOLO 等卷积网络模型，演进至以 DETR 及其改进型为代表的、更为先进的端到端 Transformer 架构，为信号检测任务提供了强大的理论与模型基础。受此启发，研究者们开始将目标检测框架应用于信号时频谱图，成功地将信号感知任务重构为视觉识别问题，并验证了此思路的可行性。

### 2.2 国内外文献综述及简析

#### 2.2.1 传统信号定位与识别方法研究现状

##### 2.2.1.1 传统信号定位方式

在信号处理的早期阶段，信号定位主要依赖于对一维时域数据的分析，其中能量检测（Energy Detection）与匹配滤波（Matched Filtering）是最具代表性的技术。

能量检测作为一种基础的非相干检测方法，通过在特定频段内对信号能量进行积分并与预设门限比较来判断信号是否存在。Quan 等人的研究<sup>[1]</sup> 为认知无线电引入了多频段联合能量检测技术，以识别频谱空洞。然而，这类将宽带频谱划分为多个独立窄带进行检测的方法会导致频谱分析的粒度较粗，频谱利用效率不高；Bkassiny 等人<sup>[2]</sup> 则结合了盲能量检测与循环平稳特征，并辅以无监督聚类进行分类，但这增加了系统的复杂性。

匹配滤波则是一种相干检测方法，其核心思想是通过将接收信号与已知信号模板进行互相关运算来实现信号检测。Bao 等人的工作<sup>[3]</sup> 利用信号与噪声在统计分布上存在差异这一先验信息，通过直方图分析自适应地寻找阈值，以定位占用频段。然而，然而，匹配滤波方法存在对先验知识强依赖的固有缺陷，即检测方必须预先精确掌握待测信号的完整信息（如波形、调制方式等），这使其在非合作通信、盲信号检测等场景下适用性严重受限。此外，当需要检测多种不同类型的信号时，必须设计相应的多个匹配滤波器，导致系统实现复杂度高、灵活性差。

为了克服一维分析的局限性，研究人员引入了时频分析工具，其中短时傅里叶变换是应用最广泛的一种。通过 STFT，一维的时序信号可以被转换为二维的时频谱图（Spectrogram），从而可以直观地展示信号频率随时间的变化情况。这一转变为研究者借鉴计算机视觉领域中成熟技术进行信号分析提供了可能性。

在对于二维时频谱图的分析中，基于数学形态学（Mathematical Morphology）的处理方法成为一个研究热点。该方法利用腐蚀、膨胀等一系列形态学操作，对时频谱图进行去噪、边缘增强和目标提取。例如，Mankun 等人<sup>[4]</sup> 利用图像增强和数学形态学技术对跳频信号的时频谱图进行分析，有效地提取了信号的轮廓和跳变模式；Phonsri 等人<sup>[5]</sup> 进一步结合了计算机视觉与双向神经网络，用于从含噪的时频谱图中提取通信信号。然而，这些方法的缺点也十分明显：首先，形态学操作对噪声和干扰非常敏感，且其参数（如结构元素的尺寸和形状）需要根据信号特征进行人工设计，缺乏自适应性，泛化能力较差；更重要的是，这些方法主要集中于“定位”，即找出信号在时频谱图中的位置和形状，但无法在此基础上进一步提供信号的调制类型、通信制式等深层语义信息。

随着深度学习技术的发展，一些研究开始尝试利用神经网络来处理频域数据。例如，Huang 等人<sup>[6]</sup> 采用全卷积网络（Fully Convolutional Network, FCN）对宽带功率谱进行分析，以实现载波信号的检测。这种方法虽然在信号的二维时频结构信息基础上利用了深度学习强大的特征提取能力，提升了频率定位的准确性，但仍未能解决时频联合定位的根本问题。

### 2.2.1.2 传统信号分类方式

在深度学习普及之前，基于传统机器学习的信号分类方法是主流。这类方法通常遵循“特征工程 + 分类器”的范式，即研究者需要手动设计并提取信号的高阶累积量、循环谱特征、小波变换特征等专家特征，再将这些特征输入到支持向量机（Support Vector Machine, SVM）、决策树等分类器中进行训练和分类。例如，在 2008 年<sup>[7]</sup> 和 2016 年<sup>[8]</sup> 的两项研究中，学者们分别利用小波特征和新颖的统计特征，结合 SVM 分类器，对数字信号进行了自动调制识别。这类方法在特定数据

集上取得了不错的分类效果，但其核心缺陷在于严重依赖专家知识进行特征设计，导致特征的泛化能力和鲁棒性有限。更关键的是，这些方法的设计目标是为信号片段赋予类别标签，完全无法提供信号在时频域中的精确位置，从而无法实现定位与分类的一体化。

深度学习，特别是卷积神经网络的引入，为信号分类带来了突破，它能够自动从原始数据中学习特征，避免了繁琐的手动特征工程。Bitar 等人<sup>[9]</sup>将信号功率谱密度图作为图像输入 CNN，成功对 WiFi、蓝牙等多种无线技术进行了分类，并初步探讨了重叠信号的分类问题，展示了深度学习在复杂电磁环境下的应用潜力。尽管如此，这类方法依旧沿袭了“先分割、后分类”的思路，未能解决信号在时间和频率上的精确定位问题。

综上所述，在传统信号处理框架下，信号的定位与分类始终是两条相互独立的技术路径。直至计算机视觉领域的目标检测技术取得长足发展，才为实现端到端的时频联合定位与分类提供了全新的可能。

## 2.2.2 基于深度学习的目标检测方法研究现状

深度学习的浪潮极大地推动了计算机视觉领域的发展，其中核心任务之一即为目标检测，旨在从图像中同时预测出目标的边界框（Bounding Box）并对目标进行分类。

在目标检测领域研究初期，以 Faster R-CNN<sup>[10]</sup> 为代表的两阶段（two-stage）检测器取得了巨大成功。该方法首先通过区域提议网络（Rigion Proposal Network, RPN）生成一系列可能包含目标的候选框（anchor），然后再对这些候选框进行二次分类和位置精修，精度较高但速度较慢。与之相对的是以 YOLO<sup>[11]</sup> 和 SSD<sup>[12]</sup> 为代表的单阶段（one-stage）检测器，它们取消了区域提议步骤，直接在特征图上预测目标的类别和位置，实现了更快的检测速度。其中，YOLO 系列是典型的无锚框（anchor-free）算法，而 SSD 则沿用了锚框（anchor-based）机制。

一个关键的里程碑是特征金字塔网络（Feature Pyramid Network, FPN）<sup>[13]</sup> 的提出。FPN 通过构建多尺度的特征图，并融合高层语义信息和底层位置信息，显著提升了模型对不同尺度目标（尤其是小目标）的检测能力。此后，引入 FPN 或其变体已成为现代检测器的标准配置。在 FPN 的基础上，RetinaNet<sup>[14]</sup> 通过引入 Focal Loss 解决了单阶段检测器中正负样本极度不平衡的问题，在保持高速的同时达到了媲美两阶段检测器的精度；FCOS<sup>[15]</sup> 则是一种更为彻底的 anchor-free 单阶段检测器，它将目标检测视为一种逐像素的预测任务，进一步简化了模型设计。

近年来，在自然语言处理领域取得成功的 Transformer<sup>[16]</sup> 架构也被逐渐引入到计算机视觉中。DETR<sup>[17]</sup> 是首个将 Transformer 完全应用于目标检测的工作，它将

目标检测视为一个集合预测问题，无需非极大值抑制（Non-Maximum Suppression, NMS）等后处理操作，实现了端到端的检测。但早期的 DETR 存在收敛速度慢、对小目标检测能力不足等问题，且未有效利用 FPN 的多尺度思想。为了解决这些问题，Deformable DETR<sup>[18]</sup> 应运而生，它通过引入可变形注意力机制，并结合多尺度特征，显著提升了模型的性能和收敛速度。

计算机视觉中目标检测技术的演进，为解决信号时频联合检测这一难题提供了直接的理论借鉴与丰富的模型储备。

### 2.2.3 基于深度学习的信号检测方法研究现状

借鉴于目标检测领域的成功，研究者们开始将相关框架应用于二维时频谱图的分析，以实现信号的时频定位与分类识别的一体化处理。

早期的探索性工作验证了这一思路的可行性。Ke 等人<sup>[19]</sup> 将 CNN 与长短期记忆网络（Long-Short Term Memory, LSTM）相结合，以分别提取信号的频率特征和时间（空间）特征，对非合作通信信号进行盲检测。然而，该研究并未直接采用成熟的目标检测框架。随后，主流的目标检测模型被陆续引入。Zha 等人<sup>[20]</sup> 采用了 SSD 框架进行信号检测与调制分类，Prasad 等人<sup>[21]</sup> 将 Faster R-CNN 框架用于宽带系统中的盲时频定位，并在同年<sup>[22]</sup> 对其进行了轻量化裁剪；Li 等人<sup>[23]</sup> 则采用了改进的 YOLOv3 模型进行信号的盲检测。这些工作证明了将经典目标检测模型迁移至信号领域的有效性，但也暴露出共同的局限：（1）所用数据集多为仿真生成，信号种类有限（例如仅包含 WiFi 信号或仅分为连续/突发两类），且未考虑信号重叠等复杂情况，与真实的复杂电磁环境存在显著脱节。（2）这些研究更侧重于方案的可行性验证，在模型结构创新和性能提升方面探索不足。

为提升模型对特定信号形态的检测能力，后续研究开始探索更具针对性的网络结构。Li 等人<sup>[24]</sup> 针对高频宽带频谱图，提出了一种无锚框的框架。该框架利用每个频率点下的所有时间点特征来联合预测信号的中心频率和形状属性。这项工作在数据集的复杂度上有所提升（单个样本中包含了更多的信号实例），但其设计的网络结构较为简单，完全基于卷积，缺乏对长距离依赖的建模能力，且其逐频点预测的机制使其天然无法处理频率上重叠的信号。

最新的研究则开始关注更具挑战性的场景和更先进的网络结构。Peng 等人<sup>[25]</sup> 提出了一种结合残差扩张网络和水平移位注意力的模型，其数据集包含了拥挤的信号环境，且在模型的性能上也有所提升。然而，该研究仍存在明显不足：（1）信号类别和调制方式依然有限（仅分为两类信号和三种调制方式）；（2）模型仍局限于 CNN 架构，未能引入表现出强大建模能力的 Transformer；（3）忽略了在真实测

量场景中，由于采样率和采样时间不同导致 STFT 时频谱图尺寸不统一的实际问题。

综上所述，将深度学习目标检测应用于信号时频联合检测已成为该领域的主流趋势，并展现出巨大潜力。然而，现有研究仍存在以下主要不足：(1) 大多数研究使用的数据集或为仿真生成，或在信号类型、密度、信噪比范围以及重叠复杂性上与真实电磁环境存在较大差距，限制了模型的泛化能力和实用价值；(2) 多数研究直接套用计算机视觉领域的成熟模型，缺乏对信号时频谱图特有物理属性（如信号的连续性、时变性、长宽比悬殊等）的针对性设计和优化，尤其在长距离依赖关系建模和重叠信号处理上能力欠缺；(3) 没有考虑在真实测量场景中，由于采样率和采样时间不同导致 STFT 时频谱图尺寸不统一的实际问题。因此，本课题拟围绕上述不足展开研究，旨在设计一种能够应对复杂真实电磁环境、具备更强泛化能力与工程实用性的新型深度学习模型，以解决信号时频联合检测与分类的挑战。

### 3 主要研究内容及研究方案

#### 3.1 研究内容

##### 3.1.1 针对不同尺度信号数据的处理方法研究

在真实的电磁监测场景中，采样率、采样时长等采集参数的多样性，对数据处理提出了关键挑战。为保证神经网络能够有效学习信号的内在模式，进行 STFT 时，必须设定固定的频率分辨率与时间分辨率。然而，这一约束将不可避免地导致采用不同采样参数的样本在 STFT 后生成尺寸大小不一的二维时频谱图。

传统的计算机视觉模型，特别是包含全连接层的网络，通常要求输入图像具有固定的尺寸。尽管可通过插值缩放或统一尺寸填充（Padding）来强制对齐，但这两种方法均不适用于本任务：前者会严重扭曲信号固有的时频结构，破坏其物理意义；后者在样本尺寸差异悬殊时，会引入大量无信息的零值填充，不仅造成计算资源的极大浪费，还可能干扰模型的有效特征学习。

针对此问题，本研究拟对目标检测领域的多尺度训练（Multiscale Training）策略进行创新性改造。不同于在视觉任务中将其作为一种可选的数据增强手段，本课题将多尺度处理升格为一种应对异构数据源的必要核心机制。具体而言，本研究将充分利用所选 DETR 模型（由 CNN 骨干网络、Transformer 及预测层构成）对输入尺寸无严格限制的结构优势，在训练流程中设计一种自适应批处理（Adaptive Batching）策略，即将尺寸相同的 STFT 数据动态地组织成批次（Batch）送入网络。该策略旨在使模型能够稳健、高效地处理来源多样化的真实信号数据，从根本上解决异构尺度输入的难题。

##### 3.1.2 基于 Transformer 检测头的信号检测模型研究

检测头（Detection Head），即负责从深度特征图中解码出目标位置和类别的网络末端部分，是决定检测器性能的关键组件。主流的检测头可依据其结构特性分为三种范式，且各有优劣：(1) 全连接检测头：其优点在于可以人为设定固定数量的预测输出，无需复杂的后处理。但其结构要求输入特征图必须为固定尺寸，这与前述的多尺度输入策略相悖，缺乏灵活性。(2) 全卷积检测头：该结构能够灵活地适应不同尺寸的输入，但其输出的预测框数量不固定，与输入尺寸相关，因此必须依赖 NMS 等复杂的后处理算法来滤除冗余的预测框。(3) Transformer 检测头：该结构结合了前两者的特点：既能处理可变尺寸的输入，又能通过其内部的对象

查询（Object Queries）机制输出固定数量的预测结果，从而省去了复杂的后处理环节，构建了真正意义上的端到端检测流程。

更重要的是，Transformer 架构的核心——自注意力机制（Self-Attention），赋予了模型强大的全局依赖建模能力，这对于捕捉信号间的谐波关系、跳频模式等全局性特征具有潜在优势。然而，从时频谱图的视觉先验来看，信号的形态多为局部性较强的矩形或条状。全局注意力机制是否会引入不必要的计算冗余，以及其对于信号局部特征的捕捉是否优于卷积网络，是一个需要通过实验深入探究的关键问题。因此，本研究将重点构建并评估基于 Transformer 的检测头，以验证其在信号检测任务中的实际效能与适用性。

### 3.1.3 针对窄带短时信号的模型优化策略研究

在时频谱图中，窄带短时信号（如猝发信号、短时通信等）表现为尺寸微小、持续时间短的“小目标”，这对检测模型提出了极高的要求。标准的检测模型由于特征图分辨率较低、感受野过大等问题，在处理这类小目标时常出现漏检的现象。为攻克这一难题，本研究拟引入 Deformable DETR 模型进行优化：(1) Deformable DETR 借鉴了 FPN 的思想，融合了骨干网络在不同阶段输出的多尺度特征图。高层特征图包含丰富的语义信息，有助于信号分类，底层特征图具有更高的空间分辨率，保留了精确的位置信息。通过将二者有效结合，模型能够同时兼顾对大、小不同尺度信号的检测能力；(2) 标准的 Transformer 注意力机制会对特征图上的所有像素点进行计算，对于仅占图像一小部分的稀疏信号而言，这会带来巨大的计算浪费。Deformable DETR 则是采用了可变形注意力机制，即不再对整个特征图进行全局计算，而是让网络根据输入动态地学习少数关键采样点的位置，并将注意力集中在这些与目标最相关的区域。这不仅大幅降低了模型的计算复杂度和内存消耗，还可以使得注意力能够更精准地聚焦于目标区域，从而显著提升对窄带短时信号这一类小目标的检测精度。

## 3.2 研究方案

为实现上述研究目标，本课题计划遵循以下技术路线，分阶段进行：

1. 数据集分析与预处理：(1) 对现有的真实电磁环境信号数据集进行探索性数据分析（Exploratory Data Analysis, EDA），明确其中信号制式类型的均衡性、不同信号的时宽与带宽分布、以及单个样本中的信号密度等关键特征。(2) 通过 STFT 将一维时域采样数据转换为二维时频谱图。在此过程中，需要仔细选择窗函数、窗长及重叠率等参数，以寻求时间和频率分辨率之间的最佳平衡。(3) 对部分样本

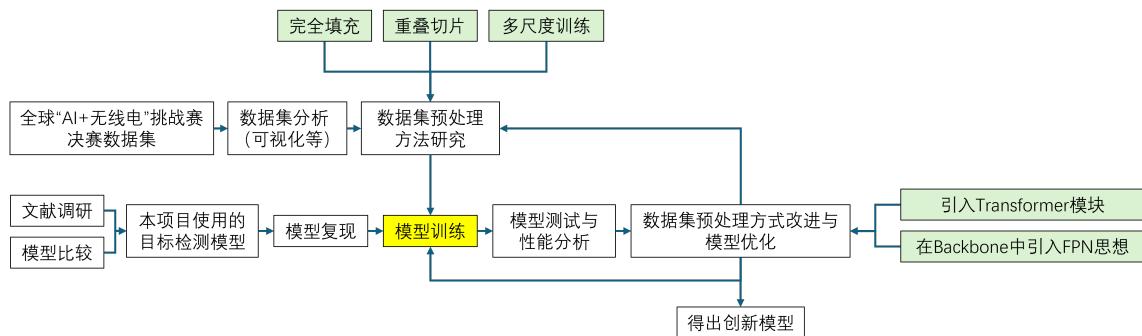


图 3-1 技术路线图

Fig.3-1 technology roadmap

进行可视化观察，以直观检验 STFT 参数的合理性，并确认是否需要对生成的时频谱图进行幅度归一化（如 Min-Max 归一化或 Z-Score 归一化）处理，以利于模型的稳定训练与快速收敛。

2. 基线模型搭建与实验：为了科学、定量地评估本课题所提方法的有效性，将首先搭建一个基线（Baseline）模型。考虑到 YOLO 系列在目标检测领域中速度与精度的良好平衡性，拟选择 YOLOv11 作为基线模型。将在预处理完成的数据集上对其进行训练与测试，并详细记录其平均精度均值（mAP）、精确率（Precision）、召回率（Recall）以及检测速度（FPS）等核心性能指标，为后续所有改进模型提供用于比较和评估的参照标准。

3. 核心模型的设计与实现：本阶段是研究的核心，将重点围绕 Deformable DETR 模型展开，将其迁移至信号时频检测这一特定场景并进行适配与优化。（1）搭建基于 Deformable DETR 的信号检测框架，该框架由 CNN 骨干网络、集成了多尺度特征融合与可变形注意力的 Transformer 编解码器，以及用于目标分类和边界框回归的线性预测层构成。（2）针对性地解决多尺度输入问题，在训练流程中设计数据加载策略，将尺寸相同的时频谱图组织在同一 Batch 中送入网络，充分利用模型对可变尺寸输入的兼容性。（3）将根据信号数据的特性对模型进行微调，例如调整 Transformer 中的对象查询数量以匹配数据集中信号的典型密度，或优化损失函数中分类损失与回归损失的权重，使模型更专注于本任务的特定挑战。

4. 综合实验与性能评估：在完成核心模型的设计与实现后，通过多维度实验，全面、深入地验证核心模型的性能与有效性。（1）整体性能对比：将所提出的 Deformable DETR 模型与基线模型（YOLOv11）在同一测试集上进行比较，分析两者在 mAP、精确率、召回率等关键指标上的差异，验证核心模型的总体优越性。（2）消融实验：为验证模型关键组件的有效性，将设计消融研究。例如，通过对比 Deformable DETR 与使用单尺度特征的基础版 DETR 的性能，来量化多尺度特征融合（FPN 思想）与可变形注意力机制对于检测窄带短时“小目标”信号

的具体提升效果。(3) 定性分析：除了定量的指标对比，还将对检测结果进行可视化分析。重点选取一些具有挑战性的样本，如包含低信噪比信号、密集重叠信号或极端长宽比信号的频谱图，直观对比不同模型在这些复杂场景下的检测效果，以展示本研究模型在鲁棒性与精确性上的优势。

5. 总结结论与论文撰写：系统性地整理和总结所有实验数据与分析结果，凝练出本研究的核心结论，明确所提出的信号检测模型相较于传统方法和基线模型的优势所在，并客观分析其可能存在的局限性。在此基础上，将对未来可能的研究方向进行展望，并依据学位论文的规范要求完成毕业设计的撰写工作。

## 4 进度安排及预期目标

### 4.1 进度安排

本课题的研究时间为 2025 年 9 月开始至 2026 年 5 月，根据之前的项目经验，结合自身的学习科研能力，针对本课题研究目标，特制定以下研究进度计划：

1. 2025 年 9 月至 2025 年 10 月：(1) 深入研读相关领域的前沿文献，完善并最终确定详细的技术实现方案；(2) 对信号数据集进行全面的统计性分析，完成 STFT 转换、数据清洗、归一化及标注格式统一等所有预处理工作；(3) 完成深度学习实验环境的搭建、配置与调试，并验证数据加载与基础训练流程的稳定性。
2. 2025 年 11 月至 2026 年 3 月：(1) 实现 YOLOv11 基线模型与基于 Deformable DETR 的核心检测模型；(2) 解决多尺度时频谱图的输入与批处理问题，(3) 在数据集上完成模型的完整训练、调优与多维度性能评估，获取全面的实验数据，并进行系统性的对比分析，得出研究结论。
3. 2026 年 4 月至 2026 年 5 月：(1) 根据已有的实验结果，撰写毕业论文，准备毕业设计答辩。

### 4.2 预期目标

本课题旨在通过上述研究，达成以下具体目标：

1. 深入分析真实电磁环境信号数据集的关键特性，完成科学的 STFT 数据预处理。最终构建一套能够支撑异构尺度输入训练的、规范化的实验与验证数据集，为后续研究提供坚实的数据基础。
2. 成功复现并将在信号时频谱图检测任务上适配、调优一个高性能的 YOLOv11 模型。获取其在各项关键指标（精确率、召回率、mAP 等）上的可靠性能数据，为衡量本课题核心模型的先进性提供一个强有力地参照基准。
3. 成功将先进的端到端目标检测框架 Deformable DETR 迁移并优化，使其深度适配信号时频联合检测任务。通过充分的实验对比，验证该模型相较于传统 CNN 基线模型，在复杂电磁环境下，特别是在处理密集、微小及重叠信号时的鲁棒性与性能优越性。
4. 针对真实场景中 STFT 时频谱图尺寸不一的核心工程挑战，成功设计并实现一种创新的多尺度自适应训练策略。该策略旨在从根本上解决异构数据输入的难题，避免因传统处理方式（插值、填充）导致的特征失真与资源浪费，提升模型的工程实用性。

## 5 已具备和所需的条件和经费

### 5.1 已具备的条件

#### 5.1.1 实验室已具备条件

本研究依托实验室现有的高性能计算服务器。

该服务器配备了 128GB 系统内存与 NVIDIA RTX 4090 高性能图形处理器 (GPU)，其强大的计算能力与充足的显存资源，能够为本课题中大规模数据集的处理及深度学习模型的训练提供充分的硬件支持。且服务器已部署了稳定、完善的科研环境，包括 Ubuntu 操作系统、Python 编程环境以及 PyTorch 等主流深度学习框架。相关开发库与依赖项均已配置完毕，可确保研究工作的顺利开展。

服务器已部署了稳定、完善的科研环境，包括 Ubuntu 操作系统、Python 编程环境以及 PyTorch 等主流深度学习框架。相关开发库与依赖项均已配置完毕，可确保研究工作的顺利开展。

鉴于本课题的核心工作为算法设计与软件实现，且所需硬件与软件条件均已具备，因此本研究无需申请额外经费支持，现有资源可完全保障课题的顺利完成。

#### 5.1.2 实验室经费保障

本课题主要通过编写代码完成，同时训练所需的硬件设备均以具备，故无需经费支持。

### 5.2 所需的条件

所需条件已经在 5.1 中详细列出并已具备，无需另外条件和经费。

## 6 预计困难及解决方案说明

### 6.1 技术难点与预计困难

1. 数据的异构性与复杂性：本研究所采用的数据集包含由不同采样参数导致的异构尺寸时频谱图，如何设计高效的数据加载与批处理机制以适应这种内在的异构性，是一个关键的工程挑战。此外，数据集中信号类型多样、时频密集且存在相互重叠，这对模型的特征表征与精细分辨能力提出了极高的要求。
2. 模型的高复杂性与超参数调优：本研究中超参数调优的复杂度较高，涉及模型结构参数（如模块层数）、优化器与学习率调度器参数（如学习率、Warmup 轮数）、以及损失函数各部分权重系数等众多变量。不恰当的超参数配置极易导致模型不收敛或陷入局部最优，因此，高效的超参数寻优策略至关重要。
3. 微小信号检测的瓶颈：在时频谱图中，大量关键信号（如猝发信号）呈现为时宽与带宽极窄的“小目标”。此类目标在经过骨干网络的多层次下采样后，其特征信息极易被稀释或丢失，是目标检测领域的公认技术瓶颈。如何有效克服这一瓶颈，确保模型对微小信号的检出率与定位精度，是衡量本研究算法先进性的核心指标。
4. 训练开销与显存瓶颈：引入 Transformer 架构将不可避免地带来计算量与显存占用的显著增长。在有限的硬件资源下，为防止显存溢出而过度减小批处理大小（Batch Size），不仅会严重降低训练效率，还可能损害模型的收敛性能与最终精度。

### 6.2 解决方案

针对上述可能遇到的困难，本课题预先制定了如下应对策略和解决方案：

1. 将设计并实现一个定制化的数据整理函数（Collate Function）。该函数能够在构建每个训练批次时，动态地将尺寸相同或相近的样本聚合，从而在不破坏原始数据结构的前提下，实现对异构数据的兼容与高效处理。
2. 将采用迁移学习范式，加载在大型公开数据集上预训练的模型权重作为优化起点，以加速收敛并提升性能。同时，将借鉴相关领域顶级会议论文中的成熟训练策略与超参数配置，构建一个可靠的初始设定。采用控制变量法进行系统性的超参数调优，并密切监控训练过程中的损失曲线与验证集性能。辅以早停（Early Stopping）策略避免不必要的计算开销，并通过对模型输出结果的可视化诊断，为针对性的结构与损失函数调整提供直观依据。

3. 充分利用 Deformable DETR 的多尺度特征融合能力，确保来自底层网络的高分辨率特征被有效传递至检测头。发挥可变形注意力机制的优势，引导模型将计算资源精准聚焦于微小信号所在的关键区域。
4. 将优先采用混合精度训练 (Automatic Mixed Precision, AMP)，在几乎不损失模型精度的前提下，大幅降低显存占用并提升训练速度。若批大小依然受限，将采用梯度累积技术，以时间换空间，在不增加显存消耗的情况下实现等效的大批次训练效果。在必要时，可利用实验室的多 GPU 资源，通过数据并行 (Data Parallelism) 等分布式训练框架，从根本上扩展可用计算与显存资源，突破单卡硬件瓶颈。

## 参考文献

- [1] QUAN Z, CUI S, SAYED A H, et al. Wideband Spectrum Sensing in Cognitive Radio Networks [C/OL] // 2008 IEEE International Conference on Communications, [S.I.], 2008 : 901-906. <http://dx.doi.org/10.1109/ICC.2008.177>.
- [2] BKASSINY M, JAYAWEERA S K, LI Y, et al. Wideband Spectrum Sensing and Non-Parametric Signal Classification for Autonomous Self-Learning Cognitive Radios [J/OL] . IEEE Transactions on Wireless Communications, 2012, 11 (7) : 2596-2605. <http://dx.doi.org/10.1109/TWC.2012.051512.111504>.
- [3] BAO D, DE VITO L, RAPUANO S. A Histogram-Based Segmentation Method for Wideband Spectrum Sensing in Cognitive Radios [J/OL] . IEEE Transactions on Instrumentation and Measurement, 2013, 62 (7) : 1900-1908. <http://dx.doi.org/10.1109/TIM.2013.2251821>.
- [4] MANKUN X, XIJIAN P, TIANYUN L, et al. A New Time-Frequency Spectrogram Analysis of FH Signals by Image Enhancement and Mathematical Morphology [C/OL] // Fourth International Conference on Image and Graphics (ICIG 2007), [S.I.], 2007 : 610-615. <http://dx.doi.org/10.1109/ICIG.2007.154>.
- [5] PHONSRI S, MUKHERJEE S S, SELLATHURAI M. Computer vision and bi-directional neural network for extraction of communications signal from noisy spectrogram [C/OL] // 2015 IEEE Conference on Antenna Measurements & Applications (CAMA), [S.I.], 2015 : 1-4. <http://dx.doi.org/10.1109/CAMA.2015.7428185>.
- [6] HUANG H, LI J-Q, WANG J, et al. FCN-Based Carrier Signal Detection in Broadband Power Spectrum [J/OL] . IEEE Access, 2020, 8 : 113042-113051. <http://dx.doi.org/10.1109/ACCESS.2020.3003683>.
- [7] PARK C-S, CHOI J-H, NAH S-P, et al. Automatic Modulation Recognition of Digital Signals using Wavelet Features and SVM [C/OL] // 2008 10th International Conference on Advanced Communication Technology : Vol 1, [S.I.], 2008 : 387-390. <http://dx.doi.org/10.1109/ICACT.2008.4493784>.
- [8] HASSANPOUR S, PEZESHK A M, BEHNIA F. Automatic Digital Modulation Recognition Based on Novel Features and Support Vector Machine [C/OL] // 2016 12th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS), [S.I.] , 2016 : 172-177. <http://dx.doi.org/10.1109/SITIS.2016.35>.

- [9] BITAR N, MUHAMMAD S, REFAI H H. Wireless technology identification using deep Convolutional Neural Networks [C/OL] //2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), [S.I.], 2017 : 1-6. <http://dx.doi.org/10.1109/PIMRC.2017.8292183>.
- [10] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks [J/OL] . IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39 (6) : 1137-1149. <http://dx.doi.org/10.1109/TPAMI.2016.2577031>.
- [11] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection [C] //Proceedings of the IEEE conference on computer vision and pattern recognition, [S.I.], 2016 : 779-788.
- [12] LIU W, ANGUELOV D, ERHAN D, et al. Ssd: Single shot multibox detector [C] //European conference on computer vision, [S.I.], 2016 : 21-37.
- [13] LIN T-Y, DOLLÁR P, GIRSHICK R, et al. Feature Pyramid Networks for Object Detection [C/OL] //2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), [S.I.], 2017 : 936-944. <http://dx.doi.org/10.1109/CVPR.2017.106>.
- [14] LIN T-Y, GOYAL P, GIRSHICK R, et al. Focal Loss for Dense Object Detection [C/OL] //2017 IEEE International Conference on Computer Vision (ICCV), [S.I.], 2017 : 2999-3007. <http://dx.doi.org/10.1109/ICCV.2017.324>.
- [15] TIAN Z, SHEN C, CHEN H, et al. FCOS: Fully Convolutional One-Stage Object Detection [C/OL] //2019 IEEE/CVF International Conference on Computer Vision (ICCV), [S.I.] , 2019 : 9626-9635. <http://dx.doi.org/10.1109/ICCV.2019.00972>.
- [16] VASWANI A, SHAZER N, PARMAR N, et al. Attention is all you need [J] . Advances in neural information processing systems, 2017, 30.
- [17] CARION N, MASSA F, SYNNAEVE G, et al. End-to-end object detection with transformers [C] //European conference on computer vision, [S.I.], 2020 : 213-229.
- [18] ZHU X, SU W, LU L, et al. Deformable detr: Deformable transformers for end-to-end object detection [J] . arXiv preprint arXiv:2010.04159, 2020.

- [19] KE D, HUANG Z, WANG X, et al. Blind Detection Techniques for Non-Cooperative Communication Signals Based on Deep Learning [J/OL] . IEEE Access, 2019, 7 : 89218-89225. <http://dx.doi.org/10.1109/ACCESS.2019.2926296>.
- [20] ZHA X, PENG H, QIN X, et al. A deep learning framework for signal detection and modulation classification [J] . Sensors, 2019, 19 (18) : 4042.
- [21] PRASAD K N R S V, DSOUZA K B, BHARGAVA V K, et al. A Deep Learning Framework for Blind Time-Frequency Localization in Wideband Systems [C/OL] //2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring), [S.I.], 2020 : 1-6. <http://dx.doi.org/10.1109/VTC2020-Spring48590>. 2020.9128779.
- [22] PRASAD K N R S V, D' SOUZA K B, BHARGAVA V K. A Downscaled Faster-RCNN Framework for Signal Detection and Time-Frequency Localization in Wideband RF Systems [J/OL] . IEEE Transactions on Wireless Communications, 2020, 19 (7) : 4847-4862. <http://dx.doi.org/10.1109/TWC.2020.2987990>.
- [23] LI R, HU J, LI S, et al. Blind Detection of Communication Signals Based on Improved YOLO3 [C/OL] //2021 6th International Conference on Intelligent Computing and Signal Processing (ICSP), [S.I.], 2021 : 424-429. <http://dx.doi.org/10.1109/ICSP51882.2021.9408998>.
- [24] LI W, WANG K, YOU L, et al. A New Deep Learning Framework for HF Signal Detection in Wideband Spectrogram [J/OL] . IEEE Signal Processing Letters, 2022, 29 : 1342-1346. <http://dx.doi.org/10.1109/LSP.2022.3179958>.
- [25] PENG T, YANG S, FENG Z, et al. Spectrum Sensing via Residual Dilated Network and Horizontal Shift Attention for Cognitive IoT [J/OL] . IEEE Internet of Things Journal, 2024, 11 (22) : 36817-36828. <http://dx.doi.org/10.1109/JIOT.2024.3429200>.