目前对比的方法有三篇

《ResNet-Conformer》、《CST-FORMER》、《SELD-SSAST》对原论文进行调研，两篇论文均使用了合成数据集，没有使用数据增强技术。CST-Former后处理使用了类的阈值判断，SELD-SSAST没有明确提到使用什么后处理技术。公平评判的话，我们也使用相同的合成数据集，不使用数据增强技术。ResNet-Conformer与SELD-SSAST方法配置保持一致。

**总框架**

进行Mixture of Adapter，Mixture中使用了路由机制，混合专家（MOE）机制，辅助损失机制。

图示

描述已自动生成

**挑战**

多任务耦合，同时进行声源检测和定位，建模目标不同。

环境多样性，房间结构、混响、麦克风设备排列导致严重的域偏移（domin- shift）。

多通道音频，空间结构性强，通道语义不一致。声音事件可能被遮蔽在背景噪声中，不同通道信息冗余或干扰显著。

泛化困难，在新场景，新设备下容易出现性能骤降，鲁棒性差。

**Motivation**

在声音事件定位与检测（SELD）是一项复杂的多任务建模问题，要求模型同时感知事件的语义内容和空间位置。在现实场景中，输入音频不仅具有明显的多通道结构、复杂的空间排列，而且在不同录音环境中存在明显的分布差异（如房间结构、混响、背景噪声等），这使得传统基于单一结构或统一参数的模型难以在不同任务目标和多源域条件下都取得良好性能。

尽管现有适配器（Adapter）方法能以较低成本对预训练模型进行微调，但现有方法大多采用单一形式的Adapter，往往难以同时应对上述多种建模需求。例如，卷积型 Adapter 适于局部空间建模，但忽略通道重要性；通道注意型 Adapter 可提升通道特异性建模，但在跨通道交互方面能力有限；而线性 Adapter 则在领域迁移中具备良好的适应性，但缺乏对细粒度结构的建模能力。声音的到达方向估计（DOA）需要捕捉微小空间结构差异（如左右声道时间差，相位差）。

为此，提出了一种基于Mixture-of-Adapters （MOA）的适配框架，融合了多种结构互补的轻量 Adapter，包括面向空间局部建模的 Conv-Adapter，强化局部空间感知、位置信息建模、面向通道选择的 SE-Adapter，自适应通道权重，突出有用信息、以及面向任务迁移的 Linear-Adapter，低成本适应新域，提升迁移能力。通过引入动态路由机制与专家融合策略，该方法能够在不同任务目标和复杂环境中动态选择最合适的适配路径，从而提升模型的鲁棒性与泛化能力，为 SELD 任务提供更具适应性与表达力的建模方式。

同时还对STARSS23数据集分析，发现在不同的房间录制的音频存在域偏移的问题，即数据分布偏移。通过引入一种专家混合的适配器架构，称为Mixture-of-Adapters（MoA）用来解决域泛化的新兴Adapter机制，将MoE思想与参数高效微调（PEFT）相结合，构建高效的Domain Generation训练框架。实现了在低计算成本下对多源域分布偏移的强适应性和泛化能力。

**方法**

**路由机制**

使用余弦相似度计算输入与专家之间的匹配度

支持Top-K机制，选择得到分高的K个专家进行融合

训练时添加噪声以增加探索性

**混合专家机制**

同时添加多个任务相关（Task-Aware）Adapter

**辅助损失机制**

收集每一个Adapter的损失，以平衡每个专家的利用率，实现负载平衡机制。

**方案一**

使用不同的Adapter，然后在进行Mixture of Adapter。

ConvAdapter + SEAdapter + LinearAdapter + …

**ConvAdapter**

**结构：**

深度可分离卷积，非线性层，逐点卷积三部分组成。

**Motivation：**

SELD任务中，局部空间结构（如声源位置信息）十分关键：

Depthwise卷积可有效提取局部模式，逐点卷积捕捉跨通道信息。

与传统卷积相比，计算量更低、更易训练，可插入在主干网络中提升局部建模能力。ConvAdapter通过点卷积+密集连接特征，可以在不增加大量参数的前提下，提升模型对细粒度特征的敏感性，尤其适合定位任务中的空间解耦。

**SEAdapter**

**Motivation**

在 SELD（Sound Event Localization and Detection）中，输入往往是多通道特征，通道之间的信息具有语义区分性。某些通道包含更多有用的事件信息或方向信息；其他通道可能是冗余或噪声信息。SEAdapter 可动态调整通道权重，突出关键信号，抑制无效信息，从而提升检测与定位性能。

**Linear Adapter**

**Motivation**

为了解决 SELD 任务中领域迁移和多任务建模的挑战，在主干网络中引入轻量级的 Linear Adapter。该模块在保持原始预训练参数不变的基础上，实现了对中间特征的任务相关调节，使模型能够在小样本或特定领域的 SELD 数据集上高效微调，同时大幅降低计算成本。快速适配新数据/任务，适合迁移学习。

**方案二**

**Motivation**

在声音事件定位与检测（SELD）任务中，环境因素如房间结构、混响强度、背景噪声和麦克风排列方式等会导致显著的声学特性变化。这种环境变化引发的域偏移是导致模型在非训练场景中性能下降的关键原因。传统方法通常采用统一模型参数处理所有环境，但忽略了各场景的差异性，导致泛化能力不足，特别是在跨环境部署时性能不稳定。

为了应对这一挑战，提出一种 环境感知（Environment-aware）适配方案，通过为每种环境单独训练一个专属 Adapter，捕捉其特有声学特征。同时，设计 Adapter Fusion 模块，使模型能够根据输入信号的环境特征，动态选择并融合多个环境特定的 Adapter 权重，实现个性建模与共享知识迁移的协同，从而显著提升在多场景、多域数据下的鲁棒性和泛化性能。

**方法**

用训练集的每个房间的环境数据训练一个相应的linear adapter，然后加载对应的预训练权重，在训练adapter fusion，让模型学会自动根据不同的环境（房间）调整adapter的权重。