思考笔记

1. **自注意力机制**

注意力机制最早提出在2017年，来自Google Brain的实验团队

Attention is all you need

在了解自注意力机制前需要先了解注意力机制Attention Mechanism

注意力机制起源于**人类**的注意力分配过程，在处理信息时，大脑会根据信息的价值和重要性自动分配注意力。

**注意力机制**的实现通常包括**查询（Query）**、**键（Key）**和**值（Value）**三个部分。在给定输入序列的情况下，每个位置的输出都是基于输入序列中所有位置的加权和，权重由相应的Query和Key之间的关系决定。具体来说，注意力机制计算每个位置的权重，然后用这些权重对Value进行加权求和，以得到输出。

自注意力机制是注意力机制的一种特殊形式，特点是**Query、Key、Value**来自同一个输入序列。这样就可是使模型能够学习到输入序列中的内在关系和依赖性。在自注意力机制中，对于输入序列中的每个位置，模型都会计算它与所有其他位置之间的关系，并得到一个权重分布。然后，根据这个权重分布对输入序列进行加权求和，以得到每个位置的输出。这个过程也被称为“内部注意”或“自关注”。

1. **SE模块**

Squeeze-and-Excitation（SE）模块是一种提高卷积神经网络（CNN）性能的技术，最早由Jie Hu等人在2017年的论文《Squeeze-and-Excitation Networks》中提出。该模块的主要目的是通过显式建模特征通道之间的相关性来提高网络的表示能力。

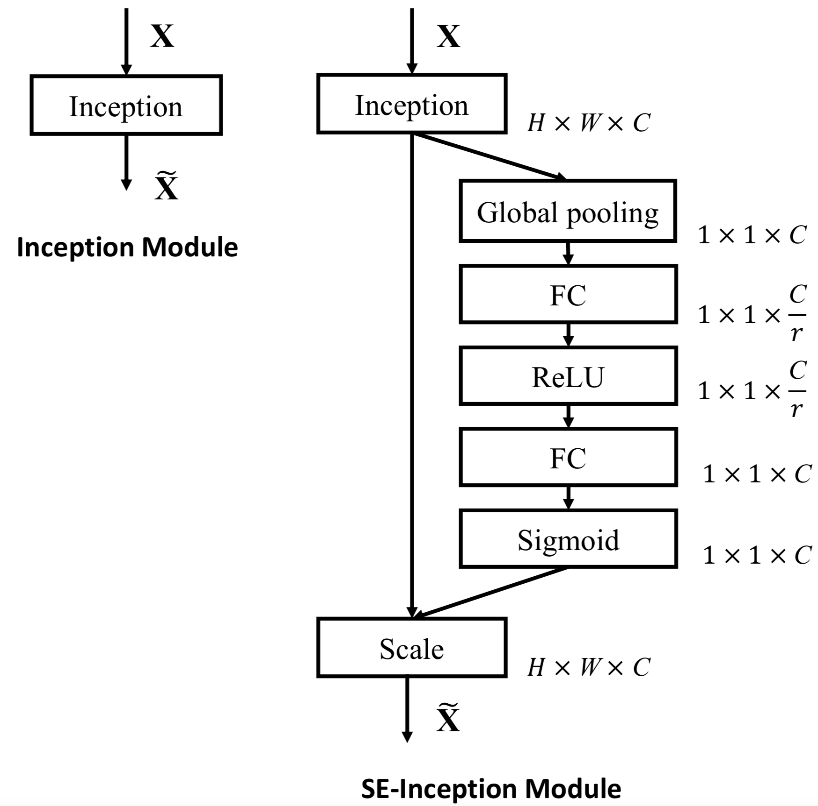
传统的卷积操作主要关注空间信息的提取，而忽略了**通道**之间的关系。SE模块通过引入显式建模通道间关系的机制来弥补这一不足。

结构：

**Squeeze：**通过全局平均池化将每个通道的空间信息压缩为一个全局描述。

**Excitation：**通过一个两层的全连接网络（包含ReLU和Sigmoid激活函数），为每个通道生成一个标量权重，这些权重反映了每个通道的重要性。

**融合：**将这些权重应用到原始特征图上，对每个通道进行重新校准。



1. **可行性分析**

对于声音事件检测定位这个项目，我们的改进目标是实现对特征的一个更好的捕获。那么只需要将工作重心放在原始代码的特征提取部分，原作者在进行特征处理时采用了CNN和TCN结合使用的方法，通过CNN提取输入数据的特征，然后通过TCN在时间维度上处理这些特征。这样的设计结合了CNN的特征提取能力和TCN在时间序列处理上的优势，能够有效的处理时频特征输入，对于本项目而言，也就是音频信号的短时傅里叶变换输出。

那么我们的改进目标是想要进一步提升模型对特征的捕获能力。主要集中在以下两个方面：

**引入SE模块增强通道关注力**

在当前的实现中，已经引入了SE模块，用于动态调整每个通道的权重，从而增强特征的表达能力。SE模块能够通过学习通道间的相关性，使得模型能够更加集中于重要的特征，对提升检测性能具有积极作用。

**整合自注意力机制提升全局特征关联性**

在当前结构的基础上，引入自注意力机制，增强模型对全局特征的关联性建模能力。有助于捕捉长距离依赖和全局上下文信息，特别适合处理复杂的音频信号格式。

1. **代码实施**

**SE模块的引入**

在代码中，将SE模块嵌入到每个卷积块中，可以确保每个阶段都能够动态地调整通道权重，优化特征的表达。

**AT模块的引入**

关于AT模块的引入问题，有两种方案：

**（1）AT模块放在CNN和TCN中间**

通过在CNN之后立即应用注意力机制，可以强化CNN提取的空间特征，使得后续的TCN能够处理更为精炼的特征。在TCN处理之前，通过注意力机制滤除冗余信息，使得时间卷积网络能够更有效地聚焦于重要特征。但是在处理完CNN特征后立即应用注意力机制，会增加计算复杂度，特别是在特征维度较高时。如果注意力机制的参数设置不当，可能会过度削弱某些特征，从而影响TCN的特征提取效果。

**（2）AT模块放在CNN和TCN之后**

通过在TCN处理完所有特征之后应用注意力机制，可以更好地捕捉全局的时空信息。在特征经过多层处理后应用注意力机制，可以进一步提升最终的特征表达，增强模型的检测和定位能力。但是如果在特征已经经过多层处理后再应用注意力机制，可能会有较多冗余信息，导致注意力机制的效果不如预期。在TCN之后应用注意力机制，计算复杂度和资源消耗同样会增加，特别是在处理长时间序列数据时。

最终采用在CNN和TCN之间引入AT模块，然后在TCN处理的内部同样引入AT模块。

1. **合理性**

整个模型结合了CNN和TCN的优势，CNN用于空间特征提取，TCN用于时间序列建模。这种结合能够有效处理音频信号的时频特征，适应了声音事件检测定位任务的需求。通过引入SE模块和注意力机制，可以进一步提升模型对复杂特征的抽象和学习能力，从而达到更好的特征提取效果。

理论上，这些改进是合理且具有潜力的，应该能够提升模型在声音事件检测定位任务中的性能。