

1、模型起始的输入参数结构：

(1, in\_chans, freq\_dim, time\_dim)

(1, 输入通道数, 频率维度,时间维度)

(beach\_size, in\_chans, freq\_dim, time\_dim)

也就是音频信号经过短时傅里叶变化的输出

2、接下来进入特征提取模块

第一步进行初始化卷积块，一共进行了**三次**，每个卷积块包括：

-- 一个卷积层

-- 一个批量归一化层

-- 一个激活函数

-- 一个最大池化层

-- 一个Dropout层

和上图相对应。