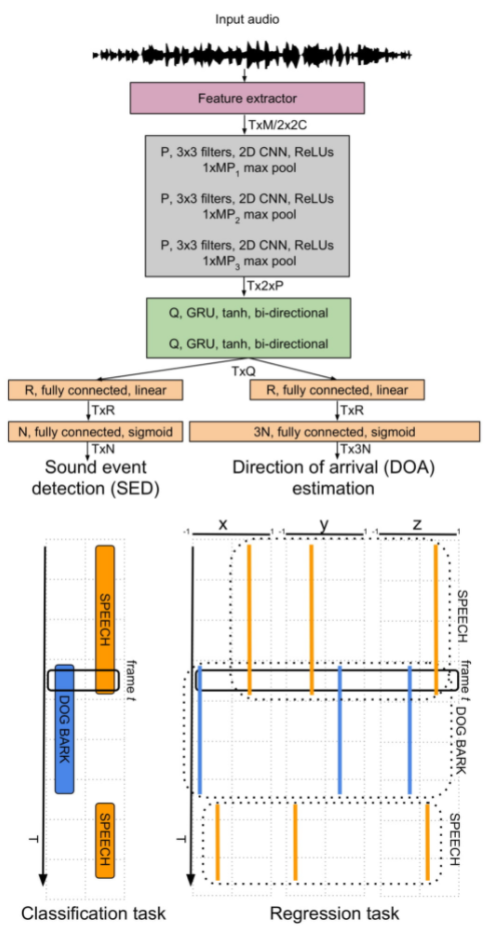
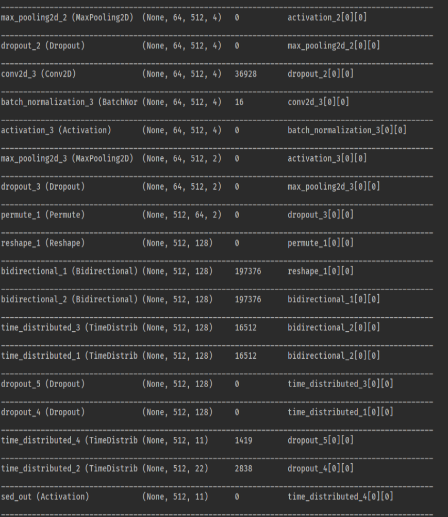
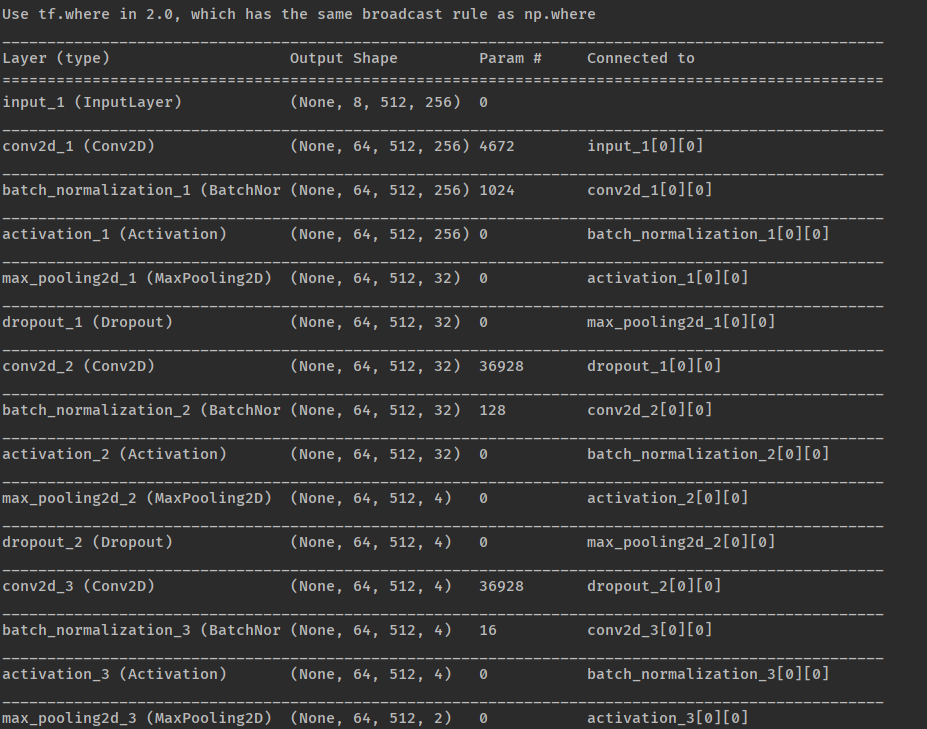
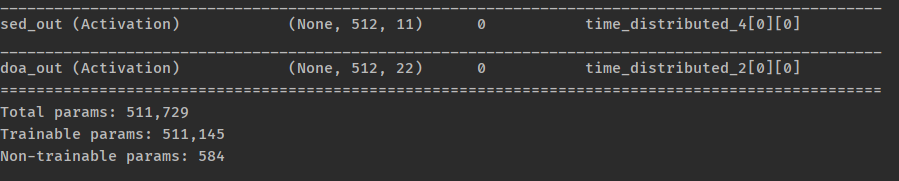
## 简介

本项目基于SELDnet架构对wav音频输入进行音频的分类与定位识别，其主要的网络架构如下所示：



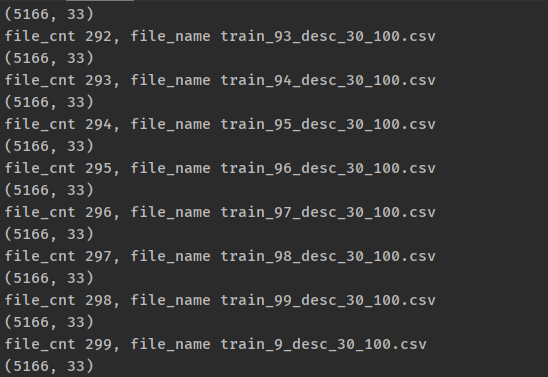
首先，这里采用的是多声道音频数据作为输入数据，经过特征提取，将音频信号提取成数字信号，转入到一系列的2D的CNN神经网络，激活函数均选用ReLu，每层CNN后均做添加最大池化层做下采样处理，然后经过两个GRU层后分别输入到两个并列的全连接神经网路层做分类，一个是用于声音的标签分类，另一个是用于定位的输出。代码中打印模型架构如下：

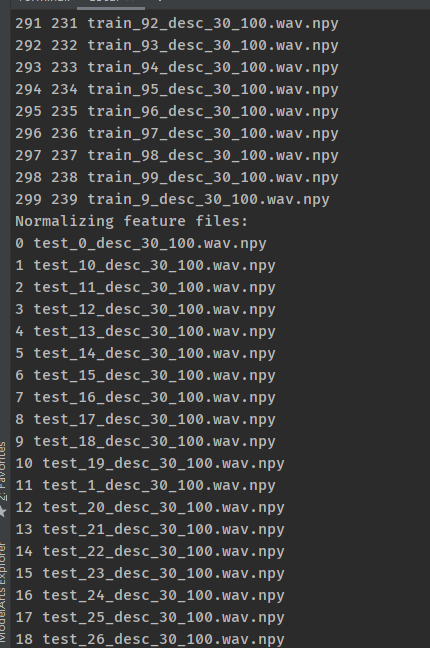




## 特征处理

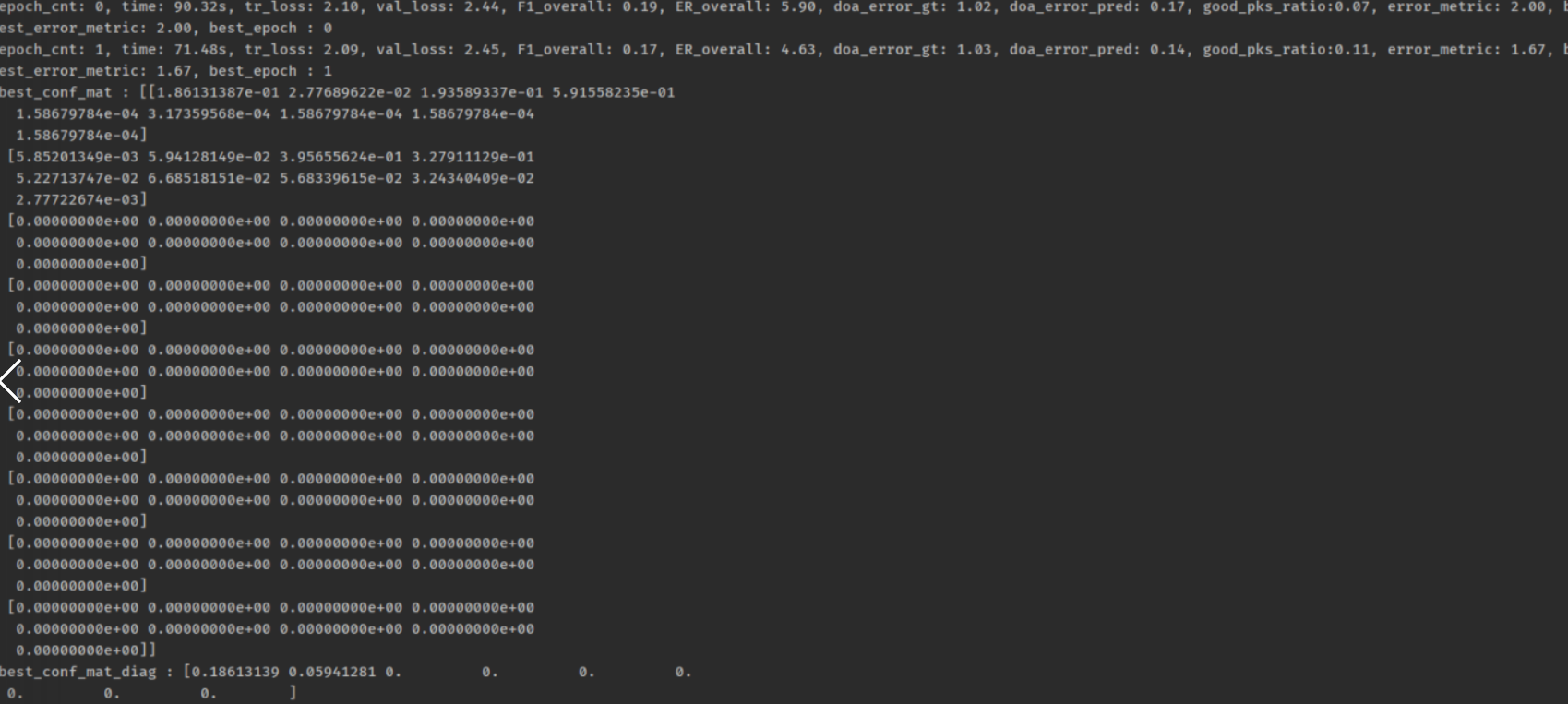
首先，输入音频为多通道的音频数据，需要对其提取相位数据和幅度数据作为后面模型输入训练的特征值，这里首先对数据集做特征提取，经过短时傅里叶变换，提取出其中频谱的相位和幅度，将其作为单独的输入特征并堆叠起来，作为模型的输入数据。特征提取部分打印如下：





## 模型搭建

模型搭建可以分为三大部分，第一部分是卷积神经网络，可以分为每层卷积+一层池化，这里的卷积层激活函数均选用ReLu函数，池化策略均选用最大池化，本层主要是对音频特征的时空信息进行低级计算；第二部分是循环神经网络，本层选用了GRU变体做卷积输出的特征信息做时间方向的提取和分类；第三部分由两列并行的全连接神经网络构成，主要作用是对以上两层提取的信息做分类，完成声音分类与定位方向检测。训练部分打印如下



## 结果输出

1. 输入为空时，做判断



1. 选择对应数据集的音频输入

