

**毕业论文**

**（ 2021 届本科 ）**

**题 目**：  **基于深度学习的音频分析处理研究**

**学 院**：  **信息学院**

**专 业： 软件工程**

**班 级： 17软工2班**

**姓 名：**  **骆春江**

**学 号：**  **1759231**

**指导教师： 王甘霖、林海**

**目录**

[摘要： 1](#_Toc68726610)

[Abstract： 2](#_Toc68726611)

[第一章 绪论 3](#_Toc68726612)

[第二章 研究探索 3](#_Toc68726613)

[2.1研究目的 4](#_Toc68726614)

[2.2研究内容 4](#_Toc68726615)

[2.2.1深度学习 5](#_Toc68726616)

[2.2.2音频分析处理 6](#_Toc68726617)

[2.3研究意义 7](#_Toc68726618)

[2.4研究过程 7](#_Toc68726619)

[第三章 可行性分析 8](#_Toc68726620)

[3.1 技术可行性 8](#_Toc68726621)

[3.2 经济可行性 9](#_Toc68726622)

[3.3 操作可行性 9](#_Toc68726623)

[第四章 常见处理过程 10](#_Toc68726624)

[第五章 补充处理方式 11](#_Toc68726625)

[5.1 降噪处理 11](#_Toc68726626)

[5.2 情景还原 12](#_Toc68726627)

[第六章 面临的问题 12](#_Toc68726628)

[第七章 展望与规划 13](#_Toc68726629)

[结论： 14](#_Toc68726630)

[致谢: 15](#_Toc68726631)

[参考文献: 16](#_Toc68726632)

**基于深度学习的音频分析处理研究**

# 摘要：

本文主要围绕深度学习和音频分析与处理进行展开，以海洋生物作为课题的切入点，旨在通过计算机来实现对于陌生海洋音频的识别，分析音频信号中的主要发声源。同时围绕项目的科研价值进行联系与展开，具体化了对项目系统的可行性分析，得出在经济、技术、操作上是可行的。通过背景需要介绍了音频分析与处理常用处理方法及其优缺点以及项目常见处理过程的跟踪与还原，并从在理论上有建设性地提出了两种能应用到项目工程中的信号处理方法，对音频信号的放大再生与真实还原有着较好的辅助作用。针对研究课题所面临的问题，如数据集的缺失及时做了调整，并针对性地研究了可行的方法。对于未来工作的实施进行了规划，确保课题研究成果能应用到实际生活当中。

**关键词：**深度学习；音频分析；处理算法

**Research on Audio Analysis and Processing Based on Deep Learning**

# Abstract：

This article mainly focuses on deep learning and audio analysis and processing, taking marine life as the entry point of the subject, and aims to realize the recognition of unfamiliar ocean audio through the computer and analyze the main sound source in the audio signal. At the same time, the research value of the project was connected and expanded, and the feasibility analysis of the project system was embodied, and it was concluded that it is feasible in economy, technology and operation. Through the background needs, it introduces the common processing methods of audio analysis and processing and their advantages and disadvantages, as well as the tracking and restoration of common processing procedures of the project, and theoretically and constructively put forward two signal processing methods that can be applied to project engineering. It has a good auxiliary effect on the amplification and reproduction of the audio signal and the true restoration. In response to the problems faced by the research topics, such as the lack of data sets, timely adjustments were made, and feasible methods were studied in a targeted manner. The implementation of future work is planned to ensure that the research results of the subject can be applied to real life.

**Key Words：**Deep Learning; Audio Analysis; Processing Algorithm

# 第一章 绪论

选择本课题项目并非一时兴起，而是有较为细致的考虑，从以下三方面来进行阐述。

从深度学习的角度来说，现在的深度学习可以说是一门很成熟的技术，对于数据的处理较人工方式有着非常大的进步，大大节省了人力物力等相关支出品的消耗，这在各方面都是非常具有意义的。科学技术的进步很大程度上方便了人类的日常生活，而自从上世纪开始对人工智能极大精力的投入研究以来，本世纪，我们已经可以将较为成熟的技术应用到生活实际，并且这些内容的的确确提高了日常生活的效率，从中可见，这部分的投入是非常值得，并且可以继续深化下去。

而其次，从音频分析的角度来看，音频分析与深度学习的结合相比于深度学习在图像处理等方面的应用就占比而言仍旧是很小的一部分，有很大的发展上升空间，就目前来看，生活实际中已经出现非常多的应用，极大地便利了人类的生活，但音频分析与研究在其他方面的细化与深入还有很多路要走。而另外更为成熟的技术，如深度学习在图像分析处理中的应用，对其与音频文件的结合或多或少有着借鉴意义，从某种程度上来说，有可以少走弯路的可能。

最后，从分析目标即海洋生物的角度来看，课题项目与海洋生物的结合是考虑到学校的特色，正因为学校在这方面的专业程度以及较为深厚的科研基础，是的确能够在课题项目的某些方面提供更为专业的支持与帮助的，同时对于自身研究方向的修正有着较为良性的指引。借助于对于海洋生物的识别与分析，我们可以做到对相关环境有更为全面的了解，诸如气候、生态、物种多样性以及对未知的探索等方面都具有较为深远的意义。这多方面的结合大大拓宽了研究的道路方向、较为明显地提高了研究的层次。

# 第二章 研究探索

这一部分是对于研究目的、研究内容、研究建立基础、研究意义与研究过程做比较细致的阐述，围绕着目的开展的过程所做的一系列行动与举措对于后续的研究探索是具有非常重大的意义的。研究区别于其他事物的因素是因为它的方向性与不断实践而得出的理论与实际应当如何结合的巧妙性。对于个人来说是需要总结的。

## 2.1研究目的

本研究课题旨在：

(1)针对音频分析处理提出可行性算法，并尽可能地进行实现

(2)针对海洋音频处理系统的短缺，从数据的分析、处理上提供一种可行的操作方案

(3)实践深度学习在音频分析与处理方面的应用

## 2.2研究内容

本研究课题旨在将深度学习、音频分析处理与海洋研究相结合，希望通过对海洋音频的分析与处理，实现对海洋、海洋生物音频信号的研究，同时希望可以让处理系统拥有更大的通用性。课题项目主要是对目标音频文件进行研究，在获取到输入的前提下，再进行一系列的分析，得出最终的发声源，做到音频文件与声源的对应，对后续其他研究起到先行准备的作用。在这一过程中通过比对不同算法对课题项目的结果影响来得出较为合理的应用区间以此达到项目的运行。前期主要通过深度学习的发展史、音频分析的理论基础来对项目有一个最初的了解，经过与图形处理的对比，从项目开始就设定了最终的框架。

具体安排如下：

绪论部分主要就研究课题的背景、意义、目的与过程进行展开，通过背景知识来引出课题项目的需求，经过分析项目有其研究的可行性，从海洋生物音频入手来展开对海洋的研究诸如海洋生态、海洋资源、海洋气候、海洋文明等。

可行性分析主要围绕技术可行性、经济可行性、操作可行性三方面进行展开，对课题项目的落地提出理论基础。使用的技术是当今较为成熟的，具备扎实理论基础的，因此在技术方面是可行的；在经济方面只需提供常规深度学习任务的算力即可满足要求；操作上则更为方便，后期开发文档的补充大大降低了操作的难度。

常见处理过程则是就目前主流的处理技术来对课题项目进行步骤式分析，主要围绕成熟的音频分析处理技术展开，通过采样将音频信号数字化，之后将时域数字信号转为频域后进行傅里叶变换，提取声音特征后输入经过训练的声学模型，得到理想化的输出，

补充处理方式则是基于处理过程当中产生的疑难问题所提出的建设性意见，主要包括两方面。一是如何有效地降噪，二是如何做到原情景的还原。

面临的问题主要是就项目处理到后期依旧存在的问题进行剖析，通过比对同类课题所面临的一致性问题，尽量缩小化未被解决的难题的范围。像训练数据的缺失是否应采用数据增强或者数据融合来使数据多样化。对于不发声的海洋生物个体是否可以采取环境音频的方式来实现这一物种的研究。

展望与规划是对系统未来发展的思考与探索，考虑通过与图像处理技术相结合来提高系统的命中率与准确性，以及对于系统的迭代升级应考虑下一步的规划，再者希望通过日后的优化提升来提高特征匹配的速度与准确率。

### 2.2.1深度学习

深度学习是人工智能旗下机器学习的一个最重要的分支，是一种基于数学知识以及计算机算法外加尽可能多的训练数据、优秀的计算机运算能力来实现对内部参数的调节，以期达到问题设立目标的半理论、半经验相结合的建模方式。

与传统的机器学习在数据准备和预处理上具有很大的相似性。两者的核心区别在于对特征的提取一个主要采用人工方式，另一个则依靠机器自动处理。很明显深度学习是后者，这也或多或少地导致了其较差的可解释性，同时前期复杂的模型设计以及对于计算量的巨大需求势必要求较高的硬件配置，但与其出色的学习能力和优秀的覆盖范围、可移植性相比，以上缺点就显得不那么重要了。

深度学习常见的具有以下四种算法：卷积神经网络(CNN)、循环神经网络(RNN)、生成对抗网络(GANs)、深度强化学习(RL)。

CNN就像人类的视觉系统一样，能够在保证数据有效性的同时，将其缩小化，即降维[1]。基于以下三层：卷积层，池化层，全连接层。分别用作特征的保留，数据的缩小化，选择性地输出。此算法一般被用在图片的分类与检索，人脸识别即是其中的一种，还可用来做目标的定位检测，比如脸上的面部特征。

RNN主要用来处理序列数据，即按照一定顺序进行存放的数据文件，而顺序输入的其中一个较为显著的特征是前一步的操作会影响到后一步，这从短期来看具有较为明显的影响，从长期来看则影响不那么大。因此RNN一般被用作处理音频信号、文本字节流的识别应用，生活中所用的翻译也大多数基于这种方法，并且很明显的是翻译这一过程是比较成功的。

GANs采用互相对抗的方式使各自得到强化，并在不断的循环中得到满意的结果。这就导致了此算法是非常不稳定的，且不如前面几种容易训练，但倘若训练完成，则对于数据分布可以有更好的建模。RL是一种与奖励机制挂钩的算法，奖励一次后就进行强化，游戏AI通常采用此种方式。

### 2.2.2音频分析处理

声音是基于振动这个物理基础而言的，振动的产生伴随着能量的传递，因此可以作为一种信号来进行交流与沟通。而信号的频率是基于声波的，用波峰的数目赖进行表示，仅就人类而言，由声音携带的内容信号占总的外界信息来源的10%[2]。对于音频的处理可以理解为是对声音这种信号方式的处理，将声音转换为数字信号是输入进计算机进行处理的先决条件，这一过程通常是基于模拟信号的离散化展开的，即之后的采样、量化、编码等操作，同时这种保存在计算机中的信号具有还原性高、保存时间长的特点。

一般的，计算机通过波形图的形式来处理音频数据，即以脉冲形式对语音信号进行编码，以时间序列作为表现方式呈现，通常将此类数据作为模型输入数据。时间序列坐标轴的x轴代表时间；y轴代表幅值，也即平常所说的振幅。

为了将音频信号离散化，我们需要对信号进行采样，每隔一定时间间隔对模拟信号进行采集来获取离散后的信号，为了防止信息的大量丢失，采样频率是有所要求的，即基于奈奎斯特定理需要大于等于信号最高频率的2倍。

量化即将数据离散化，这一处理过程中势必会造成原数据信号的丢失，并且一旦完成就无法逆转，这就导致了在原信号的基础上可能会混入不具有使用价值的数据，这就是量化噪声，对于量化噪声的处理也是实践过程中需要考虑的因素[3]。量化也分为均匀量化和非均匀量化。均匀量化与非均匀量化的区别在于是否依据等量的信号区间来确定这个量化间隔，非均匀量化的操作其实就是均匀量化，只不过是在均匀量化之前先对它做了一个压扩操作，通俗地讲就是一种变换或者是一种映射，有A律与μ律两种，一般在国内采用A律算法比较多。

这样完成后就可以对处理好的数据进行编码，作为后期的输入。编码的过程中会加一些额外的数据，作为判断数据用来比对处理前的数据是否出错或是遗失。

对信号编码后就可以进行傅里叶变换来做进一步的处理。1980年提出的梅尔-频率倒谱系数(MFCCs)作为有效特征来对波形图进行处理，目前看来在对音频的处理上是最佳的方式[4]，将信号从时域到频域的转化则依靠傅里叶变换来完成，其作用就是统计各频率的波动值。

## 2.3研究意义

海洋与陆地在地球表面积上的占比大致为7:1，陆地生态的多样性不得不让我们对海洋生态产生巨大的想象空间。在如此巨大的海洋空间中，水下生物扮演着润滑剂的角色，对于人类个体、社会生态、生物多样性、地球发展均有着重要的意义。而从音频作为切入点则较以往的操作更具安全性与便捷性，同时这极大地拓宽了我们的研究领域。此前对于深海的探索需要依靠的不仅仅是极为先进的技术，同时对参与其中的技术人员也是一种考验，而有了本课题研究项目的帮助，可以较大地减轻一系列的压力负担。

通过从海中的声音信号作为对海洋进行分析的切入点，结合现代前沿的计算机技术，来实现对个体认知的延拓，实现对海洋的多方面探索与考察。比如拿到了一段陌生音频，经过系统检测，有很大可行度可以认为是某类海豚的，然后通过海豚这一海洋生物作为接下来的切入点，结合获取音频的位置，分析此处的海洋生态、海洋气候、海洋环境以及此类海豚的相关海洋信息，可以实现对海洋较为全面的考察，但前提是确保分析的结果没有问题，即系统的高准确率和低容错率，保证获得的结果是可信的，以此为前提，后续深入的研究才是有意义的。

传统的海洋生物识别研究一般依靠人工完成，且对于音频数据的重视程度远没有其他数据那么高。很明显这其中是有现实需求的。如果能处理好这部分需求，将会在一定程度上减轻研究海洋的科研压力，同时应用得当可为其他项目的研究提供先例。

## 2.4研究过程

前期从查找资料入手，包括但不限于技术博客、中外论文、视频课程等，对深度学习以及音频分析处理有一个大致的了解，方便之后程序的编写。开始着手编写程序之后，首先通过开源网站与技术博客的分享进行系统的试运行，以期对最终的结果有大致的了解。网络上的多数项目大多都是对人类语言的识别，如中英文；小部分有对音乐进行处理的，主要围绕音乐匹配、音乐识别。通过分析认为，处理的系统应该是一致的，区别在于训练的数据集与输入输出上，因此可通过此类系统进行项目的熟悉，用来了解大致的运作原理。

中期开始对系统进行适应性匹配，保证与项目的相关性；主要工作是进行了海洋生物数据的收集，但是情况不乐观，网络上分享的数据过于局限，仅海豚、企鹅等个别种类的数据，样本数量远不足以支持系统的数据训练，同时可作为后期输入的数据也十分有限，不一定能确保系统的检测率，项目进行到后期，此问题暂时仍旧找不到解决办法，本应采用的录音方式因周期过长、人力耗费过大而暂时不考虑实施，拟采用的方法是和专项的科研数据一同收集。后期进行了对系统的代码优化，减小算力的消耗，同时期望获得更高的性能。

在确定研究方向后，之后就需要进行研究课题的核心要素与基本需求，后期的工作都必须围绕着两者展开，在开发的过程中，关键理念是首先确保开发的项目具备基本的使用需求，即当一段音频文件具有检测需求的时候，系统能通过处理实现得出设定之处的输出结果，在本研究课题中即，通过对音频的检测来得出对应的海洋生物，以供他用。

# 第三章 可行性分析

## 3.1 技术可行性

以往对海洋音频的识别主要依靠人的辨识，存在着许多问题如对人力的依赖使得识别的周期比机器处理更长。现在得益于计算机硬件和技术的快速发展，以及深度学习领域的不断深入，大型、超大型的运算任务都可以尝试着解决，像此类问题的处理一旦训练完成便不需要开展二次工程，提高准确率的同时也降低了人力消耗。

计算机系统硬件及算力的发展提供了硬件基础。从早期的单任务、小存储、简单计算到现在的多任务、大存储、复杂计算，性能的一步步提升，也带动了算力的发展，这一改变使得运算任务的困难程度大幅度降低，对于单一训练任务具有较好的底层支持性。

深度学习相关算法的迭代升级则提供了理论基础[6]。从最初的神经机械学到以联结为主再到现如今大规模的模型、数据集与算法优化，使系统变得越来越智能，对于数据的处理越来越高效，理论模型日趋完美，符合时代的发展趋势。

本次采用的循环神经网络已经是较为成熟的技术，在人类语音的识别上具有较好的成功率，其较广的应用面以及后续一次次的优化与升级，在技术层面已经具备较完善的可实现性与成熟度，因此可以说是非常契合本项目课题的。

程序开发工具库的支持也大大地节省了很多不必要的麻烦，不必亲自从最底层开始实现，对于入门级别的项目建设具有非常良性的价值，同时程序代码的可读性也得到了里的增强，对于同一个方法的使用是唯一的，这大大方便了开发人员与后续使用的便捷性。

## 3.2 经济可行性

本课题的研究主要在于人力的投入，对于算力的需求，一般高校理论上均可满足。理论知识的储备主要来源于互联网的分享，课题实践也有基于开源项目的基础。针对程序实现而言，前期主要保证系统的可实现性，中期则进行系统的适应性匹配，后期则对系统进行代码的优化。这样的一个系统流程在成本方面是完全可控的，仅对单一任务的处理不需要大量的人力与算力，一般的研究环境足以满足项目需要。对于实验数据的获取是本研究课题需要下功夫的地方，但相较于以往的海洋探测分析，工作量明显减轻很多，且一次数据的获取可以达到后续多次的适配性与泛用性，从成本上说是大幅度降低了相关领域的经济投入。

项目使用软件是PyCharm社区版，深度学习框架是基于开源社区的，这些如果只是用来做研究是不收取费用的。需要额外进行的是初试训练数据的收集，经过信息的搜集，发现相关数据过于单一且数量稀少，网络上较为普遍的海洋生物音频文件主要以海豚为主，且海豚的声音信息较为单调，收集场景较为单一。目前看来这块就算资金投入也不一定可行，这方面需要获得专业技术人员的支持。但是与后期能收获的内容相比，一次收集，多次受益，具有非常高的科研价值。

假如不准备投入过多的经济、技术支持，可以采用以下两方面的备选措施：一是转为对陆生生物音频的研究，同样对于生态保护，动物研究具有非常高的科研价值，并此这方面能搜集到的资源信息也较海洋生物更为丰富；二是以实现系统的基本操作为主要目标，即识别处理功能，在保证核心功能没有问题的前提下，可于之后替换训练的数据集，然后再进行专门的适配与优化，这样也能达到同样的效果，如果采取本方式，那么项目可以采取人类的声学模型作为项目核心功能展示，这类型的数据及资源较方式一又更为丰富，且网络上的项目也具有一定的成熟度。

## 3.3 操作可行性

本课题的程序交互方式较为简单，模型完成训练后，处理过程相对来说不具备操作难度。只需拥有训练数据就能在已有框架下完成模型的训练，之后只需要输入读取检测数据就能实现对数据的识别，操作过程几乎不需要额外的培训，简单易用，无复杂设计。仅对数据进行检测与应用是不需要理解项目内的理论与执行的，而对于开发人员来说，项目程序内配备完整的代码注释，保证更换项目成员后也能进行二次开发。一般地，待到程序开发完毕，会专门配上相应的参考文档以供后续开发使用，以便提供更便捷的技术支持，缩小实际操作的门槛来提高操作的可行性。

基于深度学习的快速迭代升级，项目可能会存在较快的更新，同时维护工作也是不可缺少的一个环节，这是额外需要注意的点，因此项目中的注释是非常重要的，对于更新与维护

倘若在关键之处提供足够的信息，那么对于操作来上会轻松不少。另一个额外需要考虑的点是并不是所有类型的计算机都能运行程序，即部分计算机不具备系统的操作性，但这仅仅只是部分，大多数还是满足的。

# 第四章 常见处理过程

音频识别从广义的角度来说是一种对声音信号的处理技术，这是基于声音作为信号的一种表现形式来展开的，近似地可以将声音理解为随时间变化而变化的一个信号序列，或者说函数[7]。拿这个信号与已知的并且确认无误的信号源进行对比，得出想要的结果这就是简单的音频识别。

一般地，做音频识别需要如下步骤。

首先，需要对音频信号做预处理，即确保输入信号的真实性、确定性。现实生活中获取到的数据难免会存在噪声，这就无法保证输入信号就是我们所需要的最终信号，为了获取到单一的、真实的数据信号，通常我们需要进行静音切除、声音分帧和噪声的过滤，将这一影响因素降到最低来保证实验数据的准确性。静音消除就是为了消除静音片段对处理的干扰而将其进行消除[8]。

经过降噪和增强等操作后，从中提取有效的特征向量，即通过时域到频域的映射将一个波形转化为一个向量，一个包含声音信息的向量。常用的有线性预测倒谱系数(LPCC)和梅尔频率倒谱系数(MFCC)。前者是建立在声管模型上的，后者则建立在人的听觉上；前者针对于声道响应，后者针对人耳听觉[9]。一般采用后者作为特征表征。在理论上，MFCC需先做快速傅里叶变换，然后进行梅尔频率尺度转换，之后再通过滤波器组计算输出，最后对输出做对数运算，再进行离散余弦变换后得到MFCC。实际项目中，不必重新去实现MFCC方法，可以调用官方的处理库来实现对信号的处理。值得一提的是，现如今受限玻尔兹曼机(RBM)、卷积神经网络(CNN)等模型也取得了不错的效果[10]。

这部分特征向量又将作为声学模型的输入，来对模型进行训练，输出的一般为一段显示检测结果的文本信息，这一步主要是为了做音频匹配来得到符合的概率。高斯混合模型(GMM)和隐马尔可夫模型(HMM)作为经典的声学模型是比较成功的[11]。前者是利用混合的高斯随机变量分布来拟合音频特征形成的模型，这是因为输入的信号经过傅里叶变换后变成特征序列，现实中因声音并非是顺序的，且其不能处理非线性的数据特征，就受到了局限；而后者通过将状态的引入作为其核心，于是就诞生了能弥补高斯混合模型的隐马尔可夫模型，因状态是离散随机变量，有其不确定性，于是这一双随机过程就能很好地处理非线性音频数据。

现如今较为成熟的语音识别技术像语音助手、语音转录都是已经投入日常使用的，这些均基于语音交互的基础；像音乐识别等方向在目前现实生活中也具有较大的应用，如听歌识曲等操作，主要基于音频匹配的基础之上，通过对曲库中音频文件的标签进行匹配与检索来得到与输入相近似的识别结果[12]。处理流程都是一致的，区别在于流程中用到的算法有所差别，不同算法各有优劣，应考虑项目的具体情况进行适应匹配，考虑不同项目各自的核心功能与主要参照，同时项目的关键是处理好特征提取这一步骤[1]，在将模拟信号转为数字信号的过程中，直接转化为相应的特征序列，然后计算序列与数据库内已有文件序列的匹配度，一般称作距离函数，得到最优解即确定与哪个片段是最匹配的[13]。这个算法有其独特性，即在对于信号的处理上采用的是色度特征[14]，对于输入数据与比对数据两者的差异采取模糊处理，来提高两者的匹配度，此方法也有其局限性，不能够普遍适配与应用，故对于很多项目来说往往不采用。

# 第五章 补充处理方式

考虑到海洋环境的复杂性，需对处理算法进行深入化。比如海洋中除了海洋生物会发出音频信号外，水流、礁石可能都存在各自的声音，所以需要将此类噪音进行过滤，确保输入数据的准确性。此外对于输入数据的真实性可以提供二次检验来进行核对。目前想到的方式有以下两种：降噪处理与情景还原。

## 5.1 降噪处理

灵感来自于计算机网络工程中的一种器件——中继器。它在网络中的功能是将信号进行变形，待到放大后再将信号转发出去，实现传输距离的延伸，尽可能地减小传输过程中信号因噪声等原因而失真或者衰减的影响。应用到本课题中采取的处理方式有如下几个步骤：获取音频信号->信号滤波降噪->信号缩小->信号放大。即先将信号通过滤波操作过滤掉噪音，虽然在降噪过程中无可避免地会损失部分真实音频，但对这部分音频先执行缩小操作，再进行放大操作，即可将影响降到比较小的程度，且能保证音频的波动范围更窄，很明显，这是利于后期信号数据的处理的。

## 5.2 情景还原

因为海洋环境与陆地环境的不同，譬如音频传输介质、音源等因素，我们很难确定通过设备获取到的音频信号是否是立即可以投入使用的，或者说就是被我们所需要的。正如其名，所谓情景还原即在已经获取到数据的前提下，通过不同形式的模拟来达到还原真实场景的目的。一般而言，数据通过系统检测后再依靠已经有的数据做重叠、移位等操作，实现对初试数据的复原，来进行数据对比进行后期的分析。当然这一步操作必须基于数据的可靠性分析，否则音频信号的堆叠并不利于情景还原，反而容易造成音频信号的混乱，理论分析上是可行的，具体实践有待试验，目前仍旧停留在理论阶段。

# 第六章 面临的问题

尽管项目初期的实践与应用是较为轻松的，但仍旧在推进过程中遇到较大的阻力，部分得到了比较好的解决。另一部分则没有，但对于为什么没有得到解决也进行了细致的分析，并尽可能地提出了解决措施，相较于一帆风顺的实践研究，这一过程也有其独特的意义。

深度学习依靠大量的数据来进行计算，尽管已有许多开源数据集，但关于本研究课题的实验数据依旧十分稀少，部分来源于海豚和鲸鱼，而且对于相关数据集本地化工作仍有待继续深入，仅目前收集到的实验资料来看还远达不到训练的目的，同时因个人计算机仅为普通台式笔记本，不具备充分的算力，因此在项目中后期就将着眼点放于流程处理，弱化数据集在之中的作用，所以在项目设计中还需要考虑相关数据集成熟后，能立马投入使用，这在目前来看是可行的并且是已经实现了的。当然了，对于数据稀缺也是有其解决方法的：迁移学习、数据增强、数据生成[15]。数据迁移是指先进行数据的预处理，然后再做数据的调整；数据增强则是通过改变参数从时间或者频率上来形成新数据，不适用本研究课题；数据生成则是依靠已有模型来生成新的数据或者多种数据相融合的虚拟数据。

其次，对于部分海洋生物，一从生物学角度而言，它们也许是不发声的；二在早前记录的实验数据上并不能确定生物与音频数据的对应性。针对这部分问题一要考虑数据的重新收集，二则是依托此类生物的环境音频采取行之有效的措施。像部分生存于深海当中的龙虾，就单一个体而言，其有记录的发声资料几乎没有，可以近似地认为对于此类个体的处理是有困难的，好消息是当它们开始群体行动时产生的环境音是经常性被捕捉到的，且多次采集数据的同一性保证了试验数据的可用性与有效性。

# 第七章 展望与规划

首先是要确保系统能应用到实践当中，抛开样本数据的问题是完全可行的。对于训练算法依旧有必要进行优化，来确保一个稳定的学习速率，增大同等时间下的工作率。此外对于信号中的特征比配一直是一个避不开的考虑因素，至到目前为止，还没有一个称得上是完美的算法来做匹配。部分优化后的算法在提高速度的同时牺牲了准确率，从结果的角度来看是得不偿失的。

另外在基于音频分析的基础上，可以考虑同图像分析处理相结合，提高识别效率与准确性。研究过程后期进行的优化将是一直持续的，伴随着系统的发展与升级，理想的目标是更高的性能，更低的运算消耗，更强大的分析与处理能力。拟准备的处理方法是进行优化算法的适配，实践过程中只经过尝试，还未具体应用。

系统的最终目标是可以提供其自身的价值，或大或小，即希望投入是有产出的，理论上、实践上都应奔着此目标去实现。尽管项目本身就不是一成不变的，但总要有一个参考标准来对其进行约束，在系统中是系统得到的结果，在理论中是理论本身的可行性。当然，考虑到处理过程中面临的困难，我们可以选择性地进行实时改进，来对结果做更好的适配。

# 结论：

本论文通过深度学习的发展与理论基础以及音频信号处理的常见流程与方法等相关联知识的介绍，得出项目课题的实现意义，并接着做了对系统的可行性分析，指出系统的开发是有需求的且是合理的。围绕着研究目标展开的后期项目开发与实际处理过程中及时做出的方向调整，也取得了不错的效果，并且额外地提出了有建设性的处理过程的优化方法，希望能够得到继续的深化，也能在日常实践中投入使用，而不单单停留在理论方法层面。

当然，对于项目过程中论文遇到的主要问题也在论文快结尾处进行了分析与探讨，也尽可能地查找了相关的资料，提出了较为可行的方法。末尾对项目之后的发展做了大致的规划与安排，希望后续经过不断的迭代升级，系统可以拥有不俗的表现。

经实验，深度学习与音频识别的结合有效地完成了初期设定的基本任务，对于数据的处理有着较好的表现，同时拥有着较高的识别效率与准确率，对部分数据而言，甚至可以达到百分之百的识别率，这对于课题项目的推进起到了明显的良性的作用，对于之后的处理有着很好的借鉴作用。

随着深度学习的深入与优化，相信在未来依旧可以有顽强的生命力与表现力，并会被应用到生活的各个方面，正是有了现在的不断尝试，未来的发展才可能会有更多的尝试与创新。

# 致谢:

在本文即将完成的尾声当中，谨向这一路走来外界给与的帮助致以最真诚的感谢！

感谢王老师对此研究课题及本人能力的相信与肯定，以及对于研究进展与论文撰写的悉心指导和独特见解，在各个时间节点对于论文进度的关注，对学生论文的细致审核批改，细致入微的态度无不让学生们放心。

感谢开源社区的无私分享，让我一步步地成长，建立清晰的知识体系，提出自己的个人见解，修正过程中的各种错误。有了整合的资源与体系，大大降低了入门操作的门槛，同时开源社区良好的交流氛围对于项目问题的解决也提供了非常大的帮助。

感谢一路走来所翻阅文献的作者与参与者，正是你们用辛勤的付出搭建了一个个的知识框架，使得深层次的概念与研究有了实现的可能。从我自身来说，我靠着这些文献逐渐清晰地摸索出了一条深度学习与音频分析相结合的道路。

当然，还有许多人与事不能一一在此表达内心的感谢之情，还请见谅。

最后，再一次表达衷心的祝福与感谢！

# 参考文献:

1. 冯为江. 基于深度学习的两类典型语音处理问题研究[D].国防科学技术大学,2016.
2. Castellini Paolo et al. A neural network based microphone array approach to grid-less noise source localization[J]. Applied Acoustics, 2021, 177
3. 高雨轩. 基于深度学习的音乐音频分类研究[D].华南理工大学,2020.
4. 申学帅. 基于深度神经网络的回声消除算法的研究与实现[D].扬州大学,2020.
5. V. I. Solovyov et al. Analyzing the Models of Speech Recognition on the Basis of Neural Networks of Deep Learning for Examination of Digital Phonograms[J]. Cybernetics and Systems Analysis, 2021, 57(prepublish) : 133-138.
6. 唐健. 深度学习语音识别系统中的若干建模问题研究[D].中国科学技术大学,2020.
7. N. Alamdari and A. Azarang and N. Kehtarnavaz. Improving deep speech denoising by Noisy2Noisy signal mapping[J]. Applied Acoustics, 2021, 172
8. 刘明,黄继风.深度学习在音频信号处理中的研究探析[J].信息通信,2020(03):144-146.
9. Emotion recognition of audio/speech data using deep learning approaches[J]. Journal of Information and Optimization Sciences, 2020, 41(6) : 1309-1317.
10. 薛又岷. 基于深度学习的音频噪声处理系统[D].南京邮电大学,2020.
11. Ahsan Adeel and Mandar Gogate and Amir Hussain. Contextual deep learning-based audio-visual switching for speech enhancement in real-world environments[J]. Information Fusion, 2020, 59 : 163-170.
12. 陈艳. 基于深度学习的音频场景分类方法研究[D].山西大学,2019.
13. 魏金太,高穹.基于深度学习的对话重叠语音片段检测[J].中北大学学报(自然科学版),2021,42(01):34-39.
14. Victor I. Solovyov and Oleg V. Rybalskiy and Vadim V. Zhuravel. Method of Identification of Attributes of Phonogram Digital Editing Using Neural Networks of Deep Learning[J]. Journal of Automation and Information Sciences, 2020, 52(1) : 22-28.
15. 孙影影,贾振堂,朱昊宇.多模态深度学习综述[J].计算机工程与应用,2020,56(21):1-10.