**电子科技大学成都学院**

**开源信息获取与应用**

项目设计报告

项目名称： 零基础入门语音克隆—声音克隆

项目负责人： 2022402060112 权健

项目成员： 2022402060211 刘倩

项目成员： 2022402060216 王梦媛

专　　业： 人工智能

班　　级： 20224020601 20224020602

任课教师姓名： 陈东祥

2023年06月制

摘 要

语音克隆（Voice cloning）是以语音为研究对象，通过语音信号处理和模式识别让机器自动识别和理解人类口述的语言。语音识别技术就是让机器通过识别和理解过程把语音信号转变为相应的文本或命令的高技术。它是一门涉及数字信号处理、人工智能、语言学、数理统计学、声学、情感学及心理学等多学科交叉的科学。这项技术可以提供比如自动客服、自动语音翻译、命令控制、语音验证码等多项应用。近年来，随着人工智能的兴起，语音识别技术在理论和应用方面都取得大突破，开始从实验室走向市场，已逐渐走进我们的日常生活。

本文基于Mockingbird项目，对该项目的实现背景和场景进行介绍，并写出了使用教程。

**关键词：**语音克隆，anaconda，NIPS

ABSTRACT

Voice cloning is the research object of speech, through speech signal processing and pattern recognition to automatically identify and understand human oral language. Speech recognition technology is a high technology that allows machines to transform speech signals into corresponding text or commands through the process of recognition and understanding. It is a multidisciplinary science involving digital signal processing, artificial intelligence, linguistics, mathematical statistics, acoustics, emotion science and psychology. The technology can provide applications such as automatic customer service, automatic voice translation, command control, voice captcha, and more. In recent years, with the rise of artificial intelligence, speech recognition technology has made great breakthroughs in both theory and application, and has begun to move from the laboratory to the market, and has gradually entered our daily life.

Based on the Mockingbird project, this article introduces the implementation background and scenarios of the project, and writes a tutorial for using it.

**Key Words:** Voice cloning, anaconda, NIPS

目 录

[第1章 引言 1](#_Toc15136)

[1.1 选题背景 1](#_Toc6252)

[1.2 研究目标和意义 1](#_Toc21599)

[1.3 研究思路 1](#_Toc7980)

[第2章 相关技术理论基础 2](#_Toc5188)

[2.1 语音识别的概念 2](#_Toc47)

[第3章 安装教程 5](#_Toc14279)

[第4章 使用教程 3](#_Toc549)1

[参考文献 3](#_Toc28317)8

[致谢 3](#_Toc16328)9

1. 引言
   1. 选题背景

该项目是“Real Time Voice Cloning”的中文版。该项目的实现是基于论文“Transfer Learning from Speaker Verification to Multispeaker Text-To-Speech Synthesis”，来自人工智能顶会NIPS，参考论文：Advances in Neural Information Processing Systems 31 (2018), 4485-4495。

* 1. 研究目标和意义

数据集中包含大约30万条人声音频，它们有着高度准确性和一致性，语速，音量基本相同，咬字清晰。音频时间短，基本在5s之内等特点。在这个项目中，我们可以学习到新的编程语言、框架和工具。这将有助于提高我们的技能水平。

* 1. 研究思路

在了解过部分项目原理后我们就可以开始项目部署和使用。

1. 相关技术理论基础
   1. Anaconda

Anaconda是一个开源的Python和R编程语言的发行版，用于科学计算、数据分析和机器学习任务。它包含了许多常用的科学计算和数据分析库，并提供了一个方便的环境管理工具，可以很方便地解决多版本python并存、切换以及各种第三方包安装问题。

以下是Anaconda的一些主要特点：

1. 方便的包管理；2.易于安装和管理多个Python环境；3.可以快速安装和更新大量科学包；4.提供了Jupyter Notebook等交互式开发工具；5.可以通过conda命令行工具快速创建、管理和删除各种类型的虚拟环境。

* 1. CUDA

CUDA 是 NVIDIA 发明的一种并行计算平台和编程模型。它通过利用图形处理器 (GPU) 的处理能力，可大幅提升计算性能CUDA（Compute Unified Device Architecture）。CUDA使开发人员能够利用GPU的能力来实现计算的可并行化部分，从而加快计算密集型应用程序的速度。CUDA™是一种由NVIDIA推出的通用[并行计算](https://baike.baidu.com/item/%E5%B9%B6%E8%A1%8C%E8%AE%A1%E7%AE%97/113443?fromModule=lemma_inlink" \t "https://baike.baidu.com/item/CUDA/_blank)架构，该架构使[GPU](https://baike.baidu.com/item/GPU?fromModule=lemma_inlink" \t "https://baike.baidu.com/item/CUDA/_blank)能够解决复杂的计算问题。它包含了CUDA[指令集架构](https://baike.baidu.com/item/%E6%8C%87%E4%BB%A4%E9%9B%86%E6%9E%B6%E6%9E%84/7029547?fromModule=lemma_inlink" \t "https://baike.baidu.com/item/CUDA/_blank)（ISA）以及GPU内部的并行计算引擎。开发人员可以使用[C语言](https://baike.baidu.com/item/C%E8%AF%AD%E8%A8%80?fromModule=lemma_inlink" \t "https://baike.baidu.com/item/CUDA/_blank)来为CUDA™架构编写程序，所编写出的程序可以在支持CUDA™的处理器上以超高性能运行。

* 1. CuDNN

 CuDNN（CUDA Deep Neural Network library）是NVIDIA开发的一个用于深度神经网络加速的GPU库。它提供了高性能的GPU加速的深度神经网络操作，包括卷积、池化、归一化、激活函数等。

CuDNN的主要特点和功能包括：

高性能加速：CuDNN使用GPU的并行计算能力，通过高度优化的算法和实现，提供了深度神经网络操作的高性能加速。这使得训练和推理过程更快速和高效。

简化开发：CuDNN提供了一系列的API和函数，使得开发者可以方便地在深度学习框架中集成和使用CuDNN。这简化了深度神经网络的开发过程，加速了模型的训练和推理。

硬件兼容性：CuDNN与NVIDIA的GPU硬件紧密集成，充分利用了GPU的计算能力。

* 1. PyTorch

PyTorch是一个开源的[Python](https://baike.baidu.com/item/Python/407313?fromModule=lemma_inlink" \t "https://baike.baidu.com/item/PyTorch/_blank)机器学习库，基于Torch，用于[自然语言处理](https://baike.baidu.com/item/%E8%87%AA%E7%84%B6%E8%AF%AD%E8%A8%80%E5%A4%84%E7%90%86/365730?fromModule=lemma_inlink" \t "https://baike.baidu.com/item/PyTorch/_blank)等[应用程序](https://baike.baidu.com/item/%E5%BA%94%E7%94%A8%E7%A8%8B%E5%BA%8F/5985445?fromModule=lemma_inlink" \t "https://baike.baidu.com/item/PyTorch/_blank)。PyTorch既可以看作加入了GPU支持的[numpy](https://baike.baidu.com/item/numpy/5678437?fromModule=lemma_inlink" \t "https://baike.baidu.com/item/PyTorch/_blank)，同时也可以看成一个拥有自动求导功能的强大的[深度神经网络](https://baike.baidu.com/item/%E6%B7%B1%E5%BA%A6%E7%A5%9E%E7%BB%8F%E7%BD%91%E7%BB%9C/6424200?fromModule=lemma_inlink" \t "https://baike.baidu.com/item/PyTorch/_blank)。PyTorch 是一种用于构建深度学习模型的功能完备框架，是一种通常用于图像识别和语言处理等应用程序的机器学习。PyTorch的一些重要特性有：

他是一个开源的Python机器学习库，基于Torch,用于自然语言处理等应用程序。

脚本语言（称为 TorchScript）在即时模式下易于使用且灵活。这是一种快速启动执行模式，从 Python 调用运算时立即执行运算，但也可以在 C++ 运行时环境中转换为图形模型，以提高速度和实现优化。

它支持 CPU、GPU、并行处理以及分布式训练。这意味着计算工作可以在多个 CPU 和 GPU 核心之间分配，并且可以在多台机器上的多个 GPU 上进行训练。

PyTorch 支持动态计算图形，能够在运行时更改网络行为。与大多数机器学习框架相比，提供了更大的灵活性优势。

PyTorch 拥有一组备受好评的 API，可用于扩展核心功能。

既支持用于实验的“即时模式”，也支持用于高性能执行的“图形模式”。

拥有从计算机视觉到增强学习等领域的大量工具和库。

* 1. FFmpeg

FFmpeg是一个开源的跨平台音视频处理工具集，它提供了一系列的音视频处理库和命令行工具，用于处理、转换和编辑音视频文件。以下是FFmpeg的一些主要特点：

功能完整：FFmpeg是领先的多媒体框架，能够解码(decode)、编码(encode)、转码(transcode)、复用(mux)、解复用(demux)、流(stream)、过滤(filter)和播放(play)人类和机器创建的几乎所有内容。

几乎支持所有格式：FFmpeg支持最模糊的古代格式直至最前沿。无论是由某些标准委员会、社区还是公司设计的。

跨平台高度可移植性：FFmpeg可以在各种构建环境：机器体系结构和配置下，跨Linux、Mac OS X、Microsoft Windows、BSD、Solaris等编译，运行并通过测试基础架构 FATE。

* 1. webrtcvad

webrtcvad是一个Python模块，它提供了一个简单的接口来处理音频数据并检测语音活动，同时用于在远场语音交互场景中检测语音，

它能够成功检测到最低能量的语音，同时在多噪环境下成功检测，漏检率和虚检率都较低。

1. 安装教程

一．因为该项目涉及虚拟环境和各种包，所以推荐annaconda，以下是安装教程 。

1.找到Anaconda官网

图形用户界面, 文本, 应用程序, 电子邮件

描述已自动生成

2.点击下载z

3.文件下载完成后，打开该文件

图形用户界面, 文本, 应用程序

描述已自动生成

4.根据以下图片勾选选项

图形用户界面, 文本, 应用程序, 聊天或短信

描述已自动生成

图形用户界面, 文本, 应用程序, 聊天或短信

描述已自动生成

（建议默认安装路径）

图形用户界面, 文本, 应用程序

描述已自动生成

（默认添加环境，必须勾选）

图形用户界面, 文本, 应用程序, 聊天或短信

描述已自动生成

（next）

图形用户界面, 文本, 应用程序, 聊天或短信

描述已自动生成

（finish）

图形用户界面, 文本, 应用程序

描述已自动生成

5.检查conda安装情况

（打开命令调试符，输入 conda，看格式是否相同，基本相同则安装成功）

图形用户界面, 文本

描述已自动生成

（输入conda -V，查看版本）

图形用户界面, 文本

描述已自动生成

（输入python，查看python版本）

文本

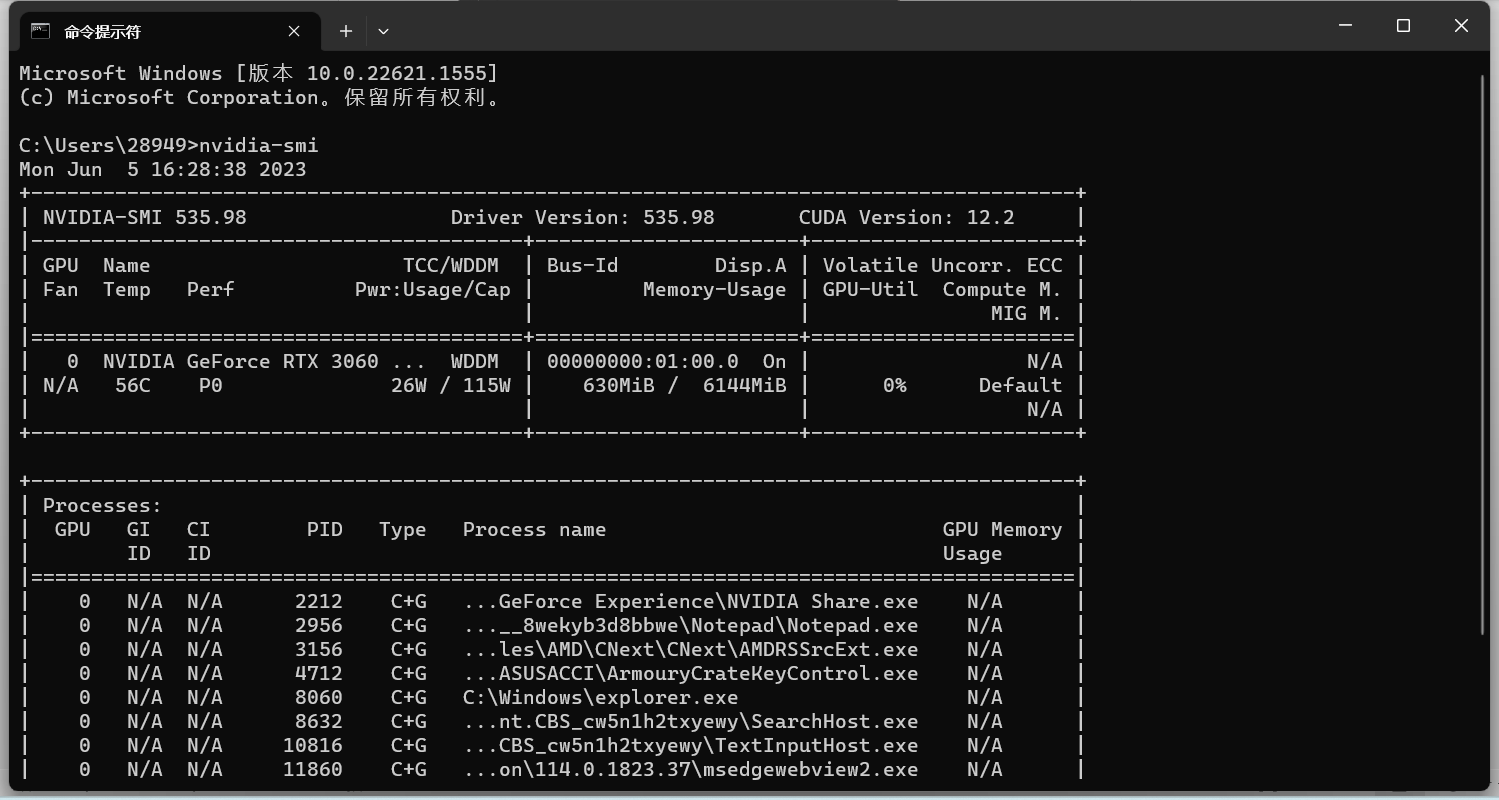
描述已自动生成

二．接下来开始cuda和cudnn的安装

在此之前，要查询你英伟达显卡驱动支持的cuda最高版本，再决定安装cuda的版本。

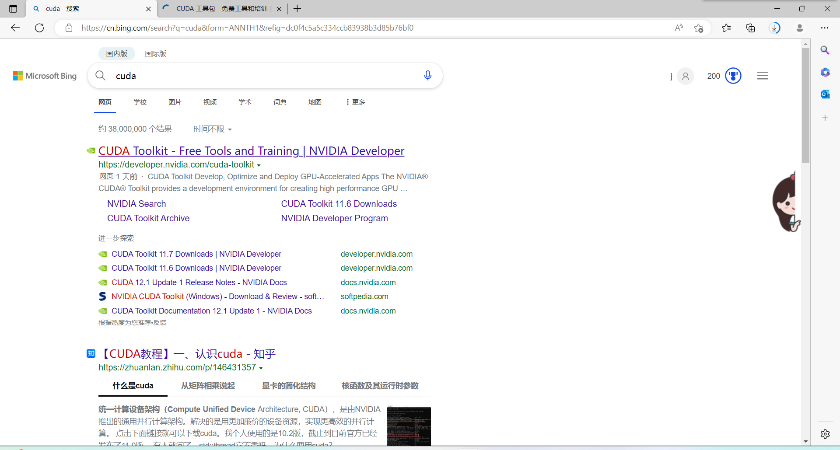
打开命令调试符，输入nvidia-smi,呈现如下。

右上角的（CUDA version：12.2）表示该显卡支持的最高CUDA版本为12.2。

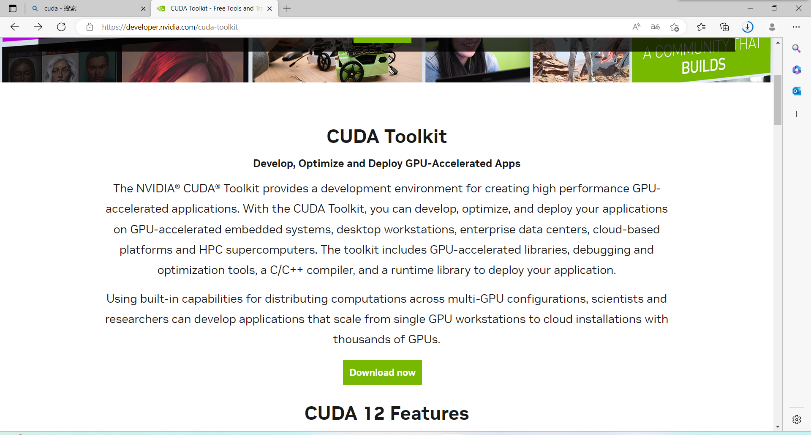


1.Cuda的安装

1.1找到下载CUDA的官网（https://developer.nvidia.com/cuda-toolkit）



（点击下方Download now）



（这里默认最新版本CUDA，如果不符合安装需求，点击下方Archive of Previous CUDA Releases，找到合适版本再进行操作）

图形用户界面, 文本

描述已自动生成

(选择操作系统，window版本后，选择exe（local），然后下载)

图形用户界面, 文本, 应用程序

描述已自动生成

1.2安装CUDA

（打开下载好的cuda文件，默认选项安装，直到安装结束）



电脑萤幕的截图

描述已自动生成

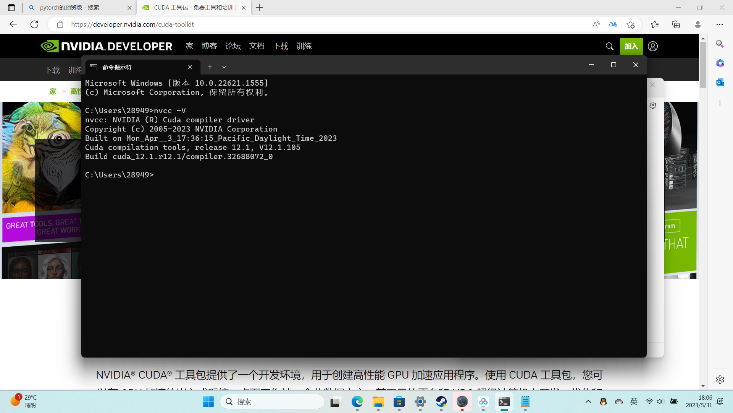


图形用户界面, 网站

描述已自动生成

1.3检查CUDA安装情况

（打开命令调试符，输入nvcc -V或nvcc –version,如下图所示则安装成功）



2.安装cudnn

2.1找到cudnn的官网

图形用户界面, 文本, 应用程序, 电子邮件

描述已自动生成

2.2点击下载（在此之前，先在网页登入英伟达账户）

图形用户界面

描述已自动生成

2.3勾选条例，点击Archived cuDNN Releases找到相应版本的cudnn

图形用户界面, 文本, 应用程序

描述已自动生成

图形用户界面

描述已自动生成

图形用户界面, 文本, 应用程序, 电子邮件

描述已自动生成



2.4解压下载好的压缩包，复制其中三个文件夹

图形用户界面, 应用程序

描述已自动生成

2.5找到cuda的安装目录，默认路径看下图

图形用户界面, 应用程序

描述已自动生成

图形用户界面, 应用程序

描述已自动生成

（选对版本号）

图形用户界面, 应用程序, Word

描述已自动生成

(将复制好的文件粘贴到下图的文件夹)

图形用户界面, 应用程序

描述已自动生成

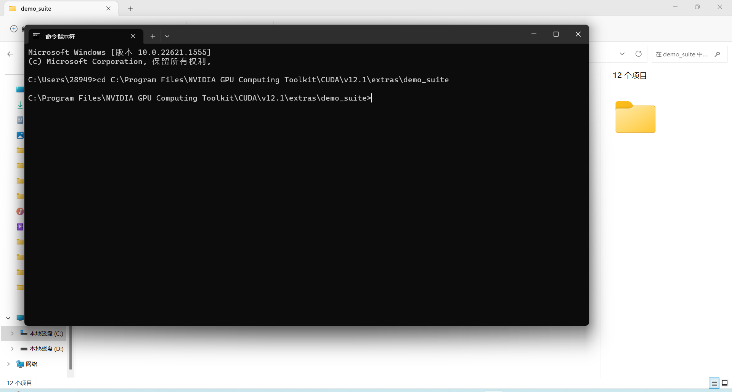
(同意一切操作)

图形用户界面, 应用程序

描述已自动生成

2.6检查cudnn安装情况

(用命令调试符打开extra文件夹里的demo\_suite文件夹)



(运行bandwidthTest.exe和deviceQuery.exe分别如下图则成功安装)

文本

描述已自动生成

文本

描述已自动生成

随后利用anaconda创建一个独立环境方便部署项目

（打开conda终端）

电脑截图

描述已自动生成

(创建虚拟环境,代码：conda create -n 环境名 python=X.X。输入y)

文本

描述已自动生成

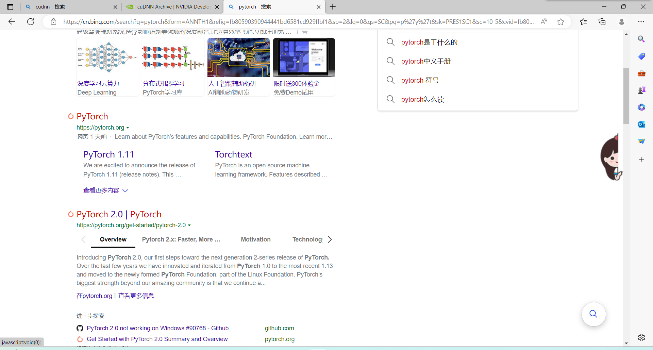
(激活环境，activate 环境名，再输入conda init bash初始化，避免安装出错)

文本

描述已自动生成

三．下面开始安装pytorch

1找到官网



（点击下载）

电脑萤幕的截图

描述已自动生成

（选择相对应的版本，如果没有，点击install previous versions of PyTorch查找）

图形用户界面, 应用程序

描述已自动生成

(复制下面的代码，打开conda终端，切换到虚拟环境安装)

文本

描述已自动生成

（安装完成后，从终端进入python，依次输入import torch，torch.cuda.is\_available() ）

（结果为true就是成功了）

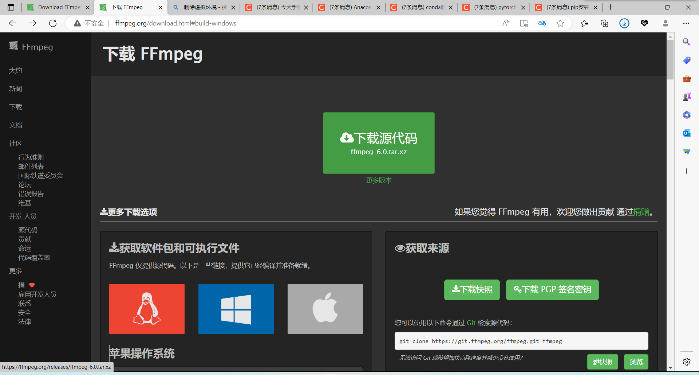
文本

描述已自动生成

四．接下来安装FFmpeg

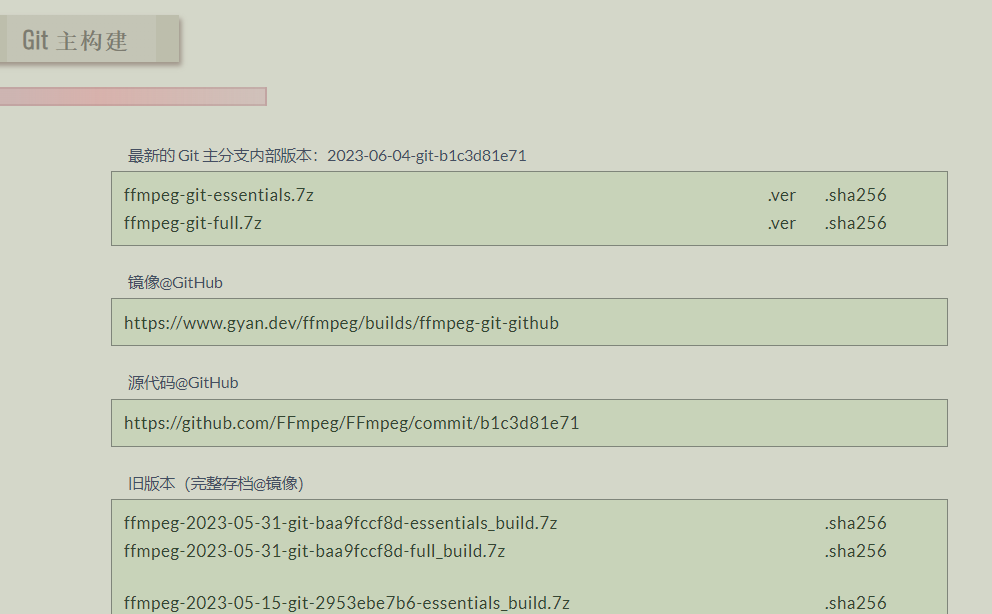
1.找到官网（http://ffmpeg.org/download.html#build-windows）下载压缩包



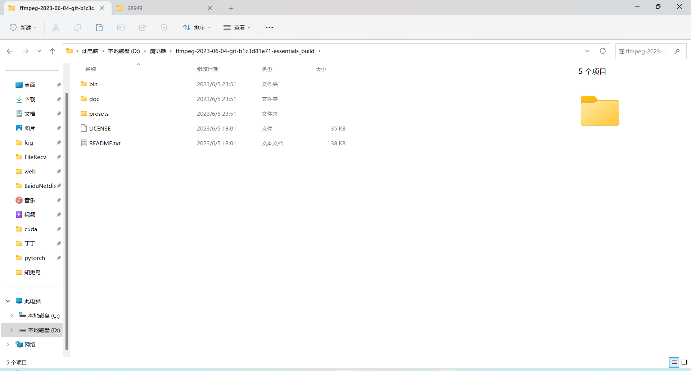


2点击下载压缩包，选择第一个

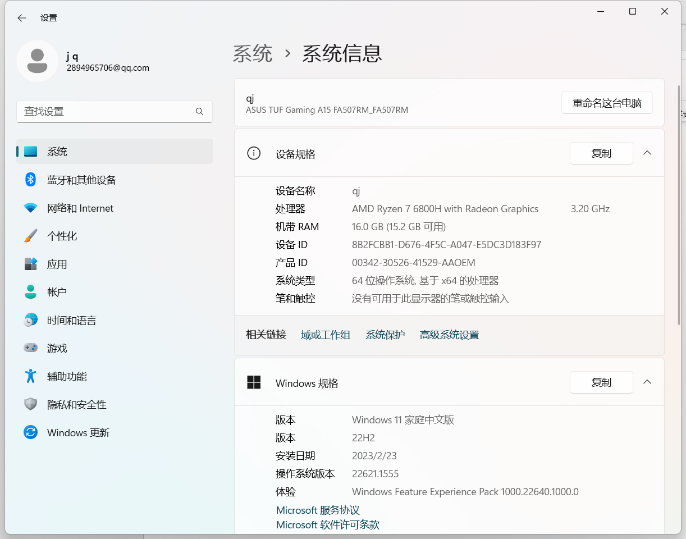




3.解压压缩包，复制其中bin文件的地址



4．配置环境变量，找到系统设置，点击环境变量，点开用户设置里的path，点新建，将文件地址复制进去，点击确定，再次点击确认。



图形用户界面, 文本, 应用程序

描述已自动生成

图形用户界面, 文本, 应用程序, 电子邮件

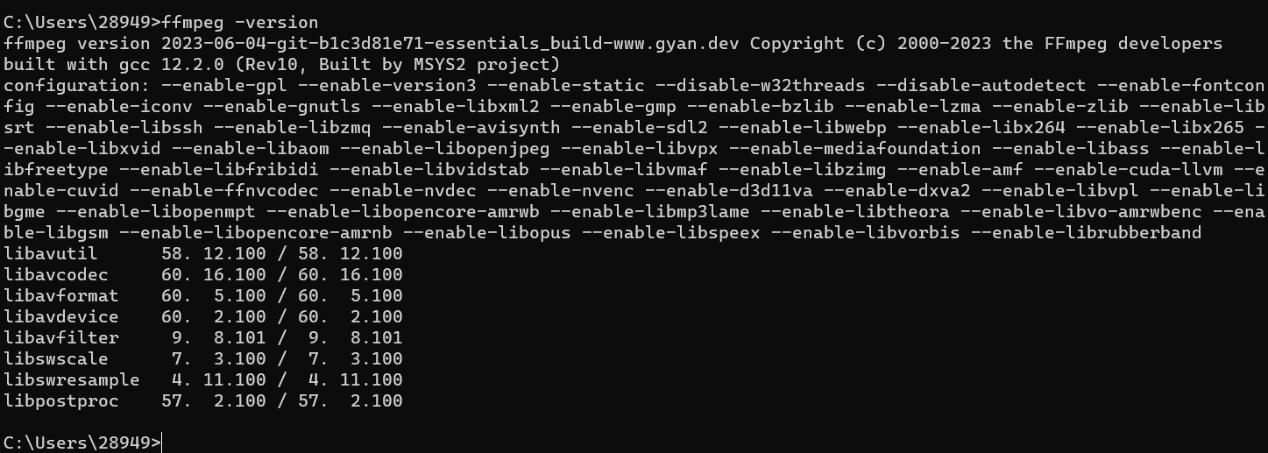
描述已自动生成

图形用户界面, 文本, 应用程序

描述已自动生成

5检查是否安装成功

（打开cmd，输入ffmpeg -version，如下既成功）

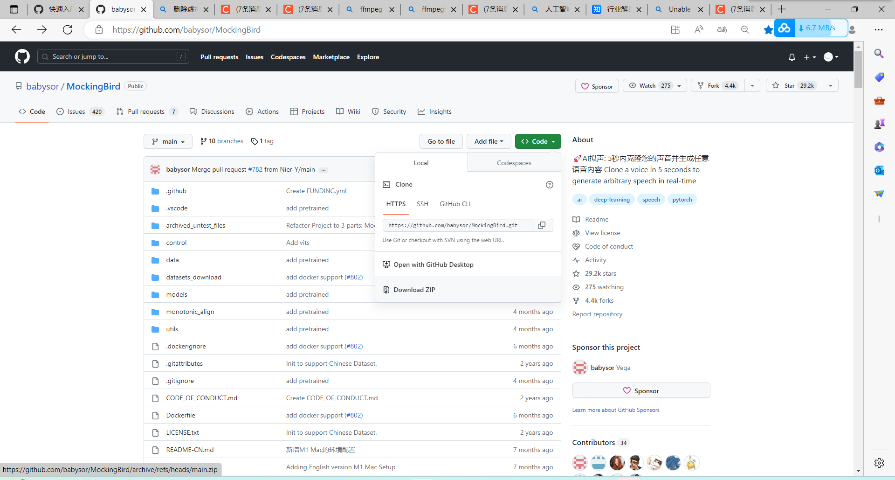


1. 使用教程

准备工作完成，接下来开始部署：

1. 下载code

（在github找到Mockingbird项目。地址：<https://github.com/babysor/MockingBird>）



1. 解压压缩包

图形用户界面, 应用程序

描述已自动生成

3用conda终端打开MockingBird-main文件夹，记得切换虚拟环境

文本

描述已自动生成

4运行pip install -r requirements.txt

文本

描述已自动生成

文本

描述已自动生成

5.输入pip install webrtcvad-wheels

文本

描述已自动生成

6.经过笔者试错，因为某种原因，以下库安装失败或是版本错误，为了保险起见，利用以下代码安装最新库

pip install PyQt5

pip install scipy

pip install scikit-learn

pip install librosa --ignore-installed llvmlite -i https://pypi.douban.com/simple --user

pip install sounddevice

pip install umap

pip install typer

pip install pydantic

pip install pandas

pip install streamlit

pip install fastapi

pip install loguru

pip install pyyaml

pip install unidecode

pip install inflect

pip install pypinyin。

7.使用模型途中可能遇到的问题及解决办法

Q1:resample() takes 1 positional argument but 3 were given

（这是librosa库的版本问题，要根据报错重写参数名，或者直接使用pip install librosa==0.9.2 降版本）

Q2:module ‘numpy‘ has no attribute ‘bool‘

（也是版本问题，新版numpy要使用“bool\_”，这里错误文件路径忘记保存了，根据报错找到错误文件重写代码即可）

Q3:module ‘numpy‘ has no attribute ‘float‘

（问题同上，在项目文件内根据路径MockingBird-main\control\toolbox\ui.py，找到“ui.py”文件修改第36行代码，将float改为cfloat）

电脑萤幕的截图

描述已自动生成

8.接下来利用别人的模型使用一下项目，我们这里使用第一个模型



9.在项目文件里data文件的子文件下创建一个名为synthesizer的文件夹，将模型放进去。

图形用户界面, 应用程序

描述已自动生成

图形用户界面, 应用程序

描述已自动生成

图形用户界面, 应用程序, Word

描述已自动生成

图形用户界面, 应用程序

描述已自动生成

10.接下来在conda终端里进入虚拟环境里的项目文件夹并输入python web.py

电脑截图

描述已自动生成

文本

描述已自动生成

11.导入5~15秒的音频就可以克隆出人声并将文本内容读出了（最终效果由模型决定）

图形用户界面, 应用程序

描述已自动生成

参考文献

Advances in Neural Information Processing Systems 31 (2018), 4485-4495。

致谢

首先我们要感谢陈东祥老师对我们此次的论文悉心指导，在格式和内容上都提出了许多宝贵的意见。在这个过程中，陈老师投入了很多心血和精力，其严谨的学术态度令我们受益匪浅，在此我们对陈老师致以深深的谢意！

其次在完成本次报告的过程中，我们小组的每一位成员分工合作，各司其职付出了很多努力，尽管完成的并不顺利，还遇到了大大小小的困难，但是在组长权健的悉心指导和带领下，最终得以完成，在这里我们对组长也表示由衷的感谢！

**电子科技大学成都学院**

**开源信息获取与应用** **课程项目报告成绩评定表**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 组长姓名 | 权健 | 组长学号 | | 2022402060112 | 专业 | 人工智能 |
| 课程名称 | 开源信息获取与应用 | | | | | |
| 设计题目 | 零基础入门语音克隆--声音克隆 | | | | | |
| 项目要求 | 给声音数据建模，准确克隆 | | | | | |
| 同组成员 | 刘倩、王梦媛 | | | | | |
| 项目设计的心得体会 | 通过本次项目，我们对语音识别有了更深刻的认识，也充分体会到团队合作的重要性。在完成项目的过程中，针对出现的问题，应该不断总结反思，从实际出发，考虑多方面的原因，及时想出解决办法。我们也认识到了自身还存在许多不足，应该在不断完善中突破自我。 | | | | | |
| 报告格式 | | | 报告内容 | | | 总分 |
|  | | |  | | |  |
| 教师评语 | 任课教师签字： | | | | | |
| 复核人签字：  时间： | | | | | | |