

解决方案实践

# 基于 GPT-SoVITS 快速实现语音克隆

文档版本 1.0.0  
发布日期 2024-03-30



版权所有 © 华为技术有限公司 2024。保留一切权利。

非经本公司书面许可，任何单位和个人不得擅自摘抄、复制本文档内容的部分或全部，并不得以任何形式传播。

## 商标声明



HUAWEI和其他华为商标均为华为技术有限公司的商标。

本文档提及的其他所有商标或注册商标，由各自的所有人拥有。

## 注意

您购买的产品、服务或特性等应受华为公司商业合同和条款的约束，本文档中描述的全部或部分产品、服务或特性可能不在您的购买或使用范围之内。除非合同另有约定，华为公司对本文档内容不做任何明示或暗示的声明或保证。

由于产品版本升级或其他原因，本文档内容会不定期进行更新。除非另有约定，本文档仅作为使用指导，本文档中的所有陈述、信息和建议不构成任何明示或暗示的担保。

# 安全声明

## 漏洞处理流程

华为公司对产品漏洞管理的规定以“漏洞处理流程”为准，该流程的详细内容请参见如下网址：

<https://www.huawei.com/cn/psirt/vul-response-process>

如企业客户须获取漏洞信息，请参见如下网址：

<https://securitybulletin.huawei.com/enterprise/cn/security-advisory>

---

# 目录

---

<b>1 方案概述</b> .....	<b>1</b>
<b>2 资源和成本规划</b> .....	<b>3</b>
<b>3 实施步骤</b> .....	<b>5</b>
3.1 准备工作.....	5
3.2 快速部署.....	8
3.3 开始使用.....	14
3.4 快速卸载.....	28
<b>4 附录</b> .....	<b>30</b>
<b>5 修订记录</b> .....	<b>31</b>

# 1 方案概述

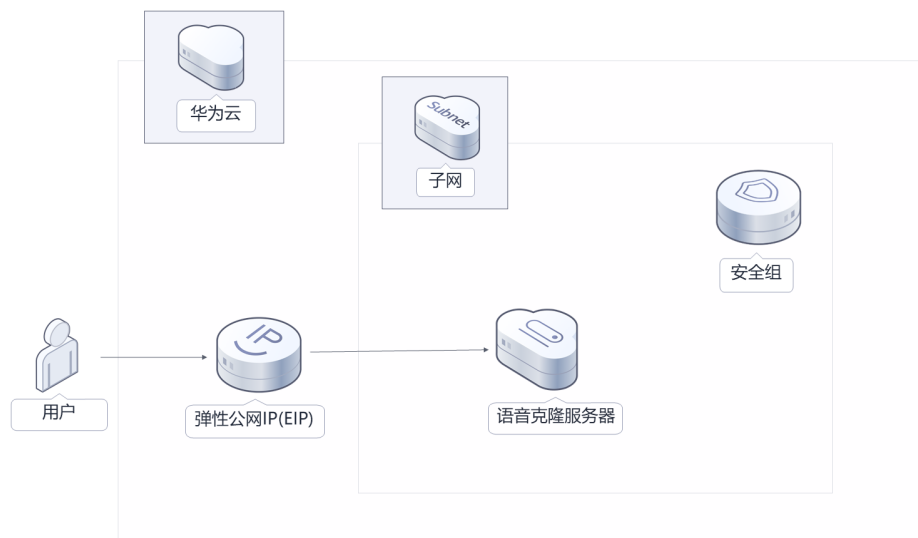
## 应用场景

该解决方案基于GPT-SoVITS，帮助您在华为云弹性云服务器 ECS上构建语音克隆 WebUI应用。**GPT-SoVITS**是一个开源语音克隆项目，支持中文、英文、日文，并提供语音分离、语音分段和文本转语音 (TTS) 等功能。适用于语音克隆、语音合成、跨语言语音处理等多种场景。

## 方案架构

该解决方案部署架构如下图所示：

图 1-1 方案架构图



该解决方案会部署如下资源：

- 创建一台Linux GPU加速型**弹性云服务器 ECS**，用于搭建语音克隆WebUI应用系统。

- 创建一个**弹性公网IP EIP**，绑定到云服务器，用于提供访问公网和被公网访问能力。
- 创建安全组，通过配置安全组规则，为云服务器提供安全防护。

## 方案优势

- 高效性  
仅需5秒的声音样本，即刻体验文本到语音的转换；仅需1分钟的训练素材，即可微调训练出声音相似度较高的语音克隆模型。
- 个性化  
该项目可以根据用户的语音样本，生成个性化的语音克隆，并支持跨语言推理。
- 一键部署  
一键轻松部署，即可完成弹性云服务器及弹性公网IP等资源的快速发放，以及语音克隆应用部署。

## 约束与限制

- 该解决方案部署前，需注册华为账号并开通华为云，完成实名认证，且账号不能处于欠费或冻结状态。
- 服务器中存放文件的所有路径不要出现中文及引号。

# 2 资源和成本规划

该解决方案主要部署如下资源，不同产品的花费仅供参考，具体请参考华为云[官网价格](#)，实际以收费账单为准：

表 2-1 资源和成本规划（按需计费）

华为云服务	配置示例	每月预估花费
弹性云服务器ECS	<ul style="list-style-type: none"><li>● 按需计费：16.49元/小时</li><li>● 区域：华北-北京四</li><li>● 计费模式：按需计费</li><li>● 规格：GPU加速型 P2v   8核   64GB   加速卡：1 * NVIDIA V100-SMX2 / 1 * 16G</li><li>● 镜像：Ubuntu 18.04 server 64bit with Tesla Driver 460.73.01 and CUDA 11.2</li><li>● 系统盘：高IO   100GB</li><li>● 购买量：1</li></ul>	11,872.08 元
弹性公网IP EIP	<ul style="list-style-type: none"><li>● 按需计费：0.80元/GB</li><li>● 区域：华北-北京四</li><li>● 计费模式：按需计费</li><li>● 线路：动态BGP</li><li>● 公网带宽：按流量计费</li><li>● 购买数量：1</li></ul>	0.80 元/GB
合计	-	11,872.08 元 + 流量费用

表 2-2 资源和成本规划（包年包月）

华为云服务	配置示例	每月预估花费
弹性云服务器ECS	<ul style="list-style-type: none"><li>● 区域：华北-北京四</li><li>● 计费模式：包月</li><li>● 规格：GPU加速型 P2v   8核   64GB   加速卡：1 * NVIDIA V100-SMX2 / 1 * 16G</li><li>● 镜像：Ubuntu 18.04 server 64bit with Tesla Driver 460.73.01 and CUDA 11.2</li><li>● 系统盘：高IO   100GB</li><li>● 购买量：1</li></ul>	7,924.80元
弹性公网IP EIP	<ul style="list-style-type: none"><li>● 按需计费：0.80元/GB</li><li>● 区域：华北-北京四</li><li>● 计费模式：按需计费</li><li>● 线路：动态BGP</li><li>● 公网带宽：按流量计费</li><li>● 购买数量：1</li></ul>	0.80 元/GB
合计	-	7924.80 元 + 流量费用



# 3 实施步骤

- 3.1 准备工作
- 3.2 快速部署
- 3.3 开始使用
- 3.4 快速卸载

## 3.1 准备工作

### 创建 rf\_admin\_trust 委托（可选）

**步骤1** 进入华为云官网，打开[控制台管理](#)界面，鼠标移动至个人账号处，打开“统一身份认证”菜单。

图 3-1 控制台管理界面



图 3-2 统一身份认证菜单



步骤2 进入“委托”菜单，搜索“rf\_admin\_trust”委托。

图 3-3 委托列表



- 如果委托存在，则不用执行接下来的创建委托的步骤
- 如果委托不存在时执行接下来的步骤创建委托

步骤3 单击步骤2界面中的“创建委托”按钮，在委托名称中输入“rf\_admin\_trust”，委托类型选择“云服务”，选择“RFS”，单击“下一步”。

图 3-4 创建委托

委托 / 创建委托

\* 委托名称

\* 委托类型  普通帐号  
将帐号内资源的操作权限委托给其他华为云帐号。  
 云服务  
将帐号内资源的操作权限委托给华为云服务。

\* 云服务

\* 持续时间

描述

0/255

步骤4 在搜索框中输入“Tenant Administrator”权限，并勾选搜索结果。

图 3-5 选择策略

委托“rf\_admin\_trust”将拥有所选策略

策略列表

名称	类型
<input checked="" type="checkbox"/> Tenant Administrator	系统角色
<input type="checkbox"/> 全部云服务管理员 (IAM管理权限)	

步骤5 选择“所有资源”，并单击下一步完成配置。

图 3-6 设置授权范围

根据当前选择的策略，系统会显示以下授权范围方案，建议您选择最小授权，可进行选择。了解如何根据应用场景选择合适的授权范围方案

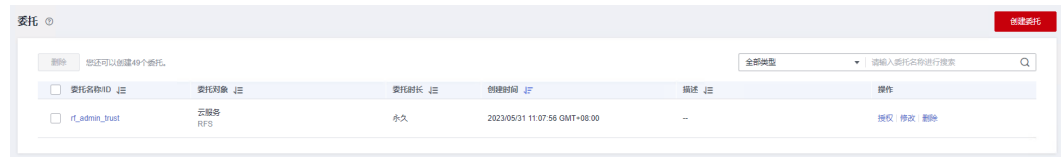
选择授权范围方案

所有资源  
授权后，IAM用户可以按照权限使用帐号中所有资源，包括企业项目、区域项目和全局服务资源。

[展开其他方案](#)

步骤6 “委托”列表中出现“rf\_admin\_trust”委托则创建成功。

图 3-7 委托列表



----结束

## 3.2 快速部署

本章节主要帮助用户快速部署该解决方案。

表 3-1 参数填写说明

参数名称	类型	是否必填	参数解释	默认值
vpc_name	String	必填	虚拟私有云名称，该模板使用新建VPC，不允许重名。取值范围：1-64个字符，支持中文、英文字母、数字、_（下划线）、-（中划线）、.（点）。	GPT-SoVITS-demo
security_group_name	String	必填	安全组名称，该模板新建安全组，安全组规则请参考 <a href="#">安全组规则修改（可选）</a> 进行修改。取值范围：1-64个字符，支持数字、字母、中文、_（下划线）、-（中划线）、.（点）。	GPT-SoVITS-demo
eip_bandwidth_size	Number	必填	弹性公网IP EIP带宽大小，该模板计费方式为按流量计费。取值范围：1-300Mbit/s。	300
ecs_name	String	必填	弹性云服务器 ECS名称，不支持重名。取值范围：1-64个字符，支持中文、英文字母、数字、_（下划线）、-（中划线）、.（点）。	GPT-SoVITS-demo

参数名称	类型	是否必填	参数解释	默认值
ecs_flavor	String	必填	弹性云服务器规格，需选取GPU加速型，请参考 <a href="#">弹性云服务器规格清单</a> 。（使用前请到 <a href="#">华为云服务器控制台</a> 查询，需选择Ubuntu 18.04 server 64bit with Tesla Driver 460.73.01 and CUDA 11.2镜像支持的规格，不然会导致方案创建失败。）	p2v.2xlarge.8(8vCPUs 64GiB T4-16GiB)
ecs_password	String	必填	弹性云服务器初始化密码。取值范围：长度为8-26个字符，密码至少包含大写字母、小写字母、数字和特殊字符（!@\$%^_-=+[{()}].,/?~#*）中的三种，管理员账户默认root。	空
charging_mode	String	必填	计费模式，默认自动扣费。可选值为：postPaid（按需计费）、prePaid（包年包月）。	postPaid
charging_unit	String	必填	订购周期类型。仅当charging_mode为prePaid（包年/包月）生效，此时该参数为必填参数。可选值为：month（月），year（年）。	month
charging_period	Number	必填	订购周期，仅当charging_mode为prePaid（包年/包月）生效，此时该参数为必填参数。当charging_unit=month（周期类型为月）时，取值范围：1-9；当charging_unit=year（周期类型为年）时，取值范围：1-3。	1

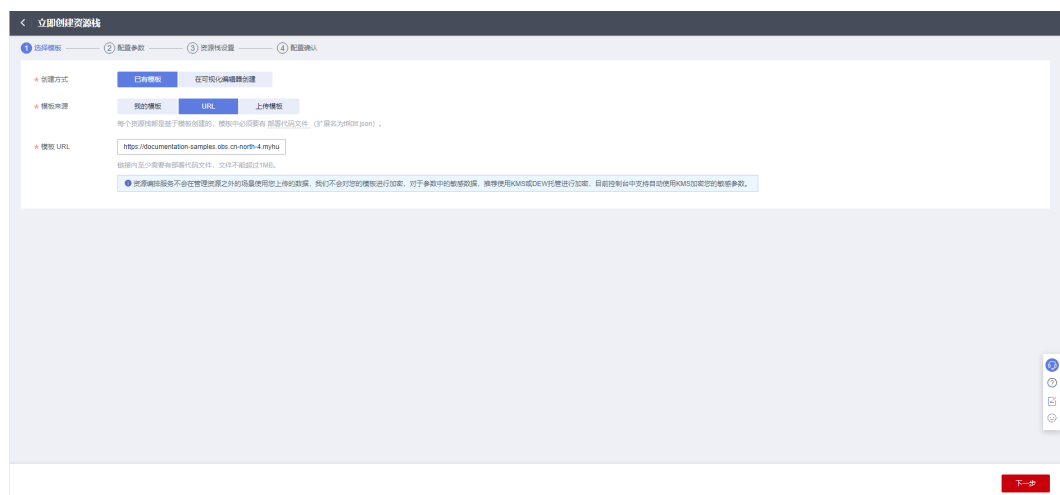
**步骤1** 登录[华为云解决方案实践](#)，选择“基于GPT-SoVITS快速实现语音克隆”并单击，跳转至该解决方案一键部署界面。

图 3-8 解决方案实施库



步骤2 单击“一键部署”，跳转至该解决方案创建资源栈部署界面。

图 3-9 创建资源栈



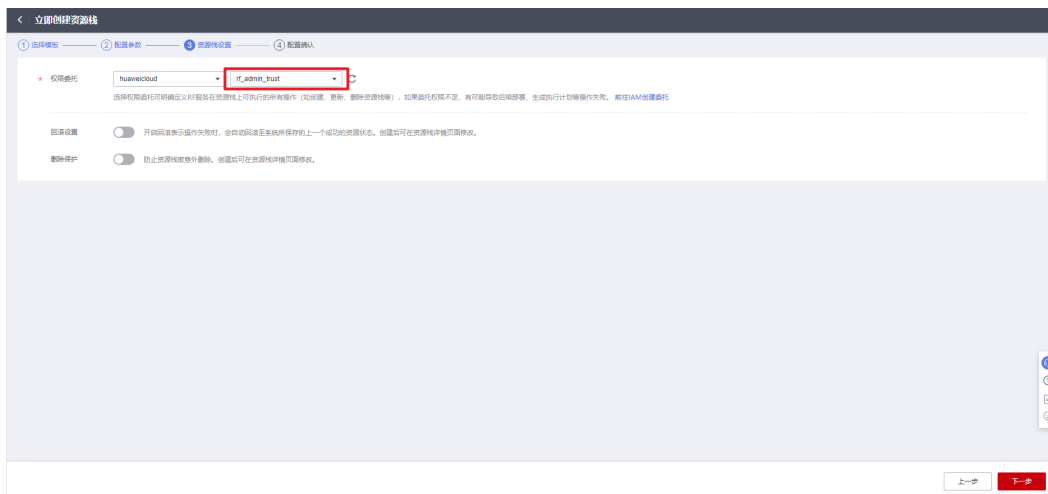
步骤3 单击“下一步”，参考表3-1完成自定义参数填写。

图 3-10 参数配置



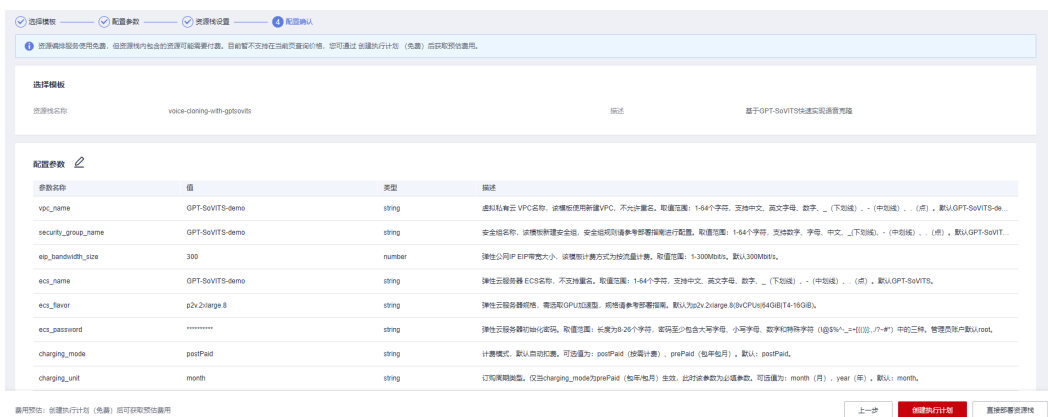
**步骤4** (可选) 在资源设置界面中，“权限委托”下拉框中选择“rf\_admin\_trust”委托，单击“下一步”。

**图 3-11 资源栈设置**



**步骤5** 在配置确认界面中，单击“创建执行计划”。

**图 3-12 创建执行计划**



**步骤6** 在弹出的创建执行计划框中，自定义填写执行计划名称，单击“确定”。

图 3-13 创建执行计划



步骤7 单击“部署”，弹出执行计划提示信息，单击“执行”确认执行。

图 3-14 执行计划确认

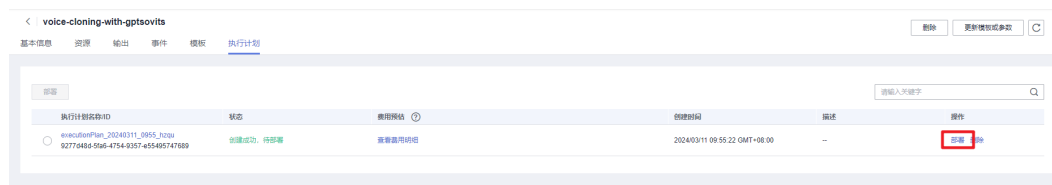


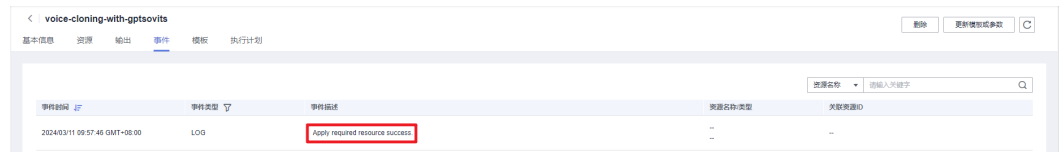


图 3-15 确认执行



步骤8 等待解决方案自动部署。部署成功后，单击“事件”，回显结果如下：

图 3-16 资源创建成功



步骤9 在“输出”中查看WebUI访问说明。堆栈部署成功后，GPT-SoVITS环境搭建脚本开始执行，耐心等待20分钟左右（受网络波动影响）。

图 3-17 输出



----结束

## 3.3 开始使用

### 说明

- 有关GPT-SoVITS详细使用指导请参考[官方视频](#)及[GPT-SoVITS指南](#)
- 涉及到GPT-SoVITS的业务端口如下：
  - 9874: GPT-SoVITS WebUI主页面访问端口
  - 9873: UVR5 音频处理WebUI页面访问端口
  - 9872: TTS推理WebUI页面访问端口
  - 9871: 语音文本校对页面访问端口
  - 22: 用于登录Linux弹性云服务器，默认只对VPC子网网段开放，用户需要修改该白名单用户IP
- 后续文档将展开GPT-SoVITS的基础使用说明

### 安全组规则修改（可选）

安全组实际是网络流量访问策略，包括网络流量入方向规则和出方向规则，通过这些规则为安全组内具有相同保护需求并且相互信任的云服务器、云容器、云数据库等实例提供安全保护。

如果您的实例关联的安全组策略无法满足使用需求，比如需要添加、修改、删除某个TCP端口，请参考以下内容进行修改。

- **添加安全组规则**：根据业务使用需求需要开放某个TCP端口，请参考[添加安全组规则](#)添加入方向规则，打开指定的TCP端口。
- **修改安全组规则**：安全组规则设置不当会造成严重的安全隐患。您可以参考[修改安全组规则](#)，来修改安全组中不合理的规则，保证云服务器等实例的网络安全。
- **删除安全组规则**：当安全组规则入方向、出方向源地址/目的地址有变化时，或者不需要开放某个端口时，您可以参考[删除安全组规则](#)进行安全组规则删除。

## GPT-SoVITS WebUI 应用使用

### 📖 说明

该程序的使用主要分为以下几个步骤，用户可以根据实际需要选择执行。

- **UVR5人声伴奏分离&去混响（可选）**

该步骤主要是提取音频文件中的纯人声，以便于后续使用。如果音频文件中没有背景音乐或者其他伴奏，则可以跳过该步骤。分离好的语音文件默认在/home/project/GPT-SoVITS/output/uvr5\_opt这个文件夹下，处理完的音频文件带vocal字样的是人声。

- **语音切分（必选）**

对纯人声音频文件进行语音切割。切分完后文件默认在/home/project/GPT-SoVITS/output/slicer\_opt这个文件夹下。

- **中文批量离线ASR工具（必选）**

该步骤用于语音识别。达摩ASR只能用于识别中文，fast whisper可以标注99种语言。输出文件默认在/home/project/GPT-SoVITS/output/asr\_opt这个路径下，会有一个.list文件，需要对该文件进行改名。

- **语音文本校对标注（可选）**

对list文件进行文字校对，修改每一句话的停顿，让AI学习到每个字该怎么读。对于后续的模式训练会很有帮助。

- **训练集格式化（必选）**

设置一个模型名，对该模型进行训练。

- **微调训练（必选）**

对模型进行微调，得到SoVITS和GPT两个模型，模型分别在/home/project/GPT-SoVITS/SoVITS\_weights和/home/project/GPT-SoVITS/GPT\_weights目录下。

- **推理（必选）**

即文本转语音。输入文本，用上述步骤训练出来的模型，生成新的语音。

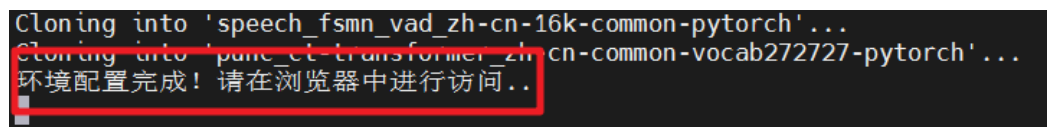
**步骤1** 进入华为云找到**弹性云服务器 ECS**控制台，找到本方案创建的云服务器。

图 3-18 GPT-SoVITS WebUI 服务器



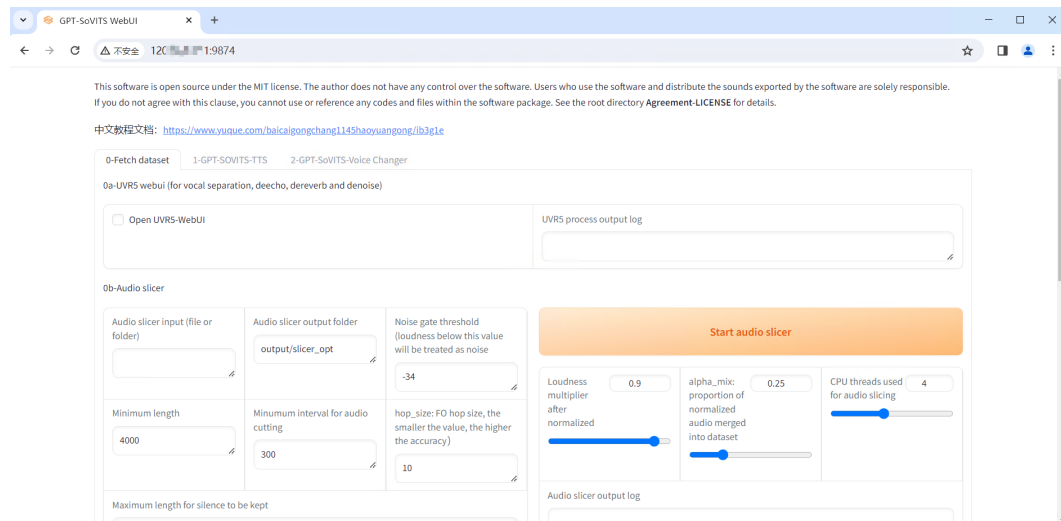
**步骤2** （可选）用户可以通过查看环境部署脚本日志，判断当前环境是否准备好。登录服务器，输入tail -f /tmp/install\_gptsovits.log，查看回显如下图所示，则代表服务器环境搭建成功。退出查看日志，在键盘同时按下“Ctrl+C”。

图 3-19 环境部署脚本日志



**步骤3** 查看**快速部署 步骤3.2-9**访问说明，即在浏览器中输入http://EIP:9874，其中EIP为服务器公网IP，访问GPT-SoVITS WebUI页面。

图 3-20 GPT-SoVITS WebUI 应用界面



**步骤4** GPT-SoVITS应用运行日志路径为：/home/project/GPT-SoVITS/webui.log。登录服务器，输入tail -f /home/project/GPT-SoVITS/webui.log，可以查看该服务区运行过程中的动态日志。

图 3-21 应用系统日志

```
(base) root@gpt-sovits-demo:~# tail -f /home/project/GPT-SoVITS/webui.log
Traceback (most recent call last):
  File "/home/project/GPT-SoVITS/tools/uvr5/webui.py", line 101, in uvr
    del pre_fun.model
UnboundLocalError: local variable 'pre_fun' referenced before assignment
Traceback (most recent call last):
  File "/home/project/GPT-SoVITS/tools/uvr5/webui.py", line 101, in uvr
    del pre_fun.model
UnboundLocalError: local variable 'pre_fun' referenced before assignment
Running on local URL: http://0.0.0.0:9873
clean_empty_cache
```

**步骤5** 查看当前服务运行的进程ID。在服务器输入ps -ef | grep webui.py命令，查看当前服务运行的进程ID。

图 3-22 进程 ID

```
(base) root@gpt-sovits-demo:~# ps -ef | grep webui.py
root      4203      1   0 14:36 ?        00:00:06 python3 webui.py
root      4440    4403   0 14:50 pts/0    00:00:00 grep --color=auto webui.py
(base) root@gpt-sovits-demo:~#
```

**步骤6** 关闭服务。在服务器输入kill -9 +进程ID，即可关闭当前运行的服务。

**步骤7** 重新启动服务。在服务器中输入以下命令，重新在后台运行服务。

```
conda activate GPTSoVits
cd /home/project/GPT-SoVITS/
python3 webui.py > /home/project/GPT-SoVITS/webui.log 2>&1 &
```

图 3-23 启动服务

```
(base) root@gpt-sovits-demo:~# conda activate GPTSoVits
(GPTSoVits) root@gpt-sovits-demo:~# cd /home/project/GPT-SoVITS/
(GPTSoVits) root@gpt-sovits-demo:/home/project/GPT-SoVITS# python3 webui.py > /home/project/GPT-SoVITS/webui.log 2>&1 &
[1] 2111
(GPTSoVits) root@gpt-sovits-demo:/home/project/GPT-SoVITS# █
```

----结束

## UVR5 人声伴奏分离&去混响（可选）

步骤1 勾选Open UVR5-WebUI（处理完成后可去勾选节省内存），待右侧框中出现“UVR5 opened”后，打开新的窗口输入http://EIP:9873。

图 3-24 打开 UVR5-WebUI

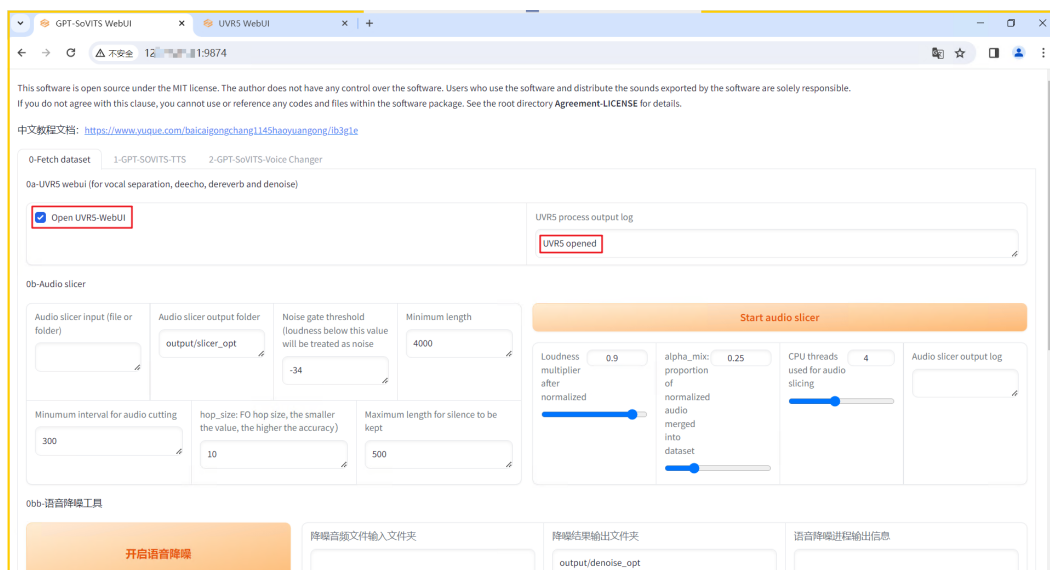
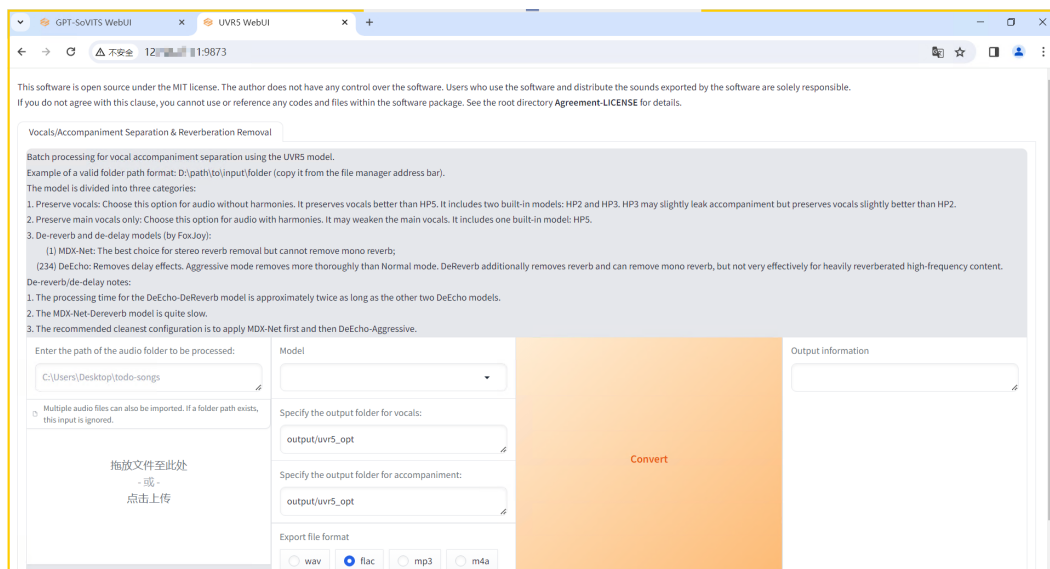
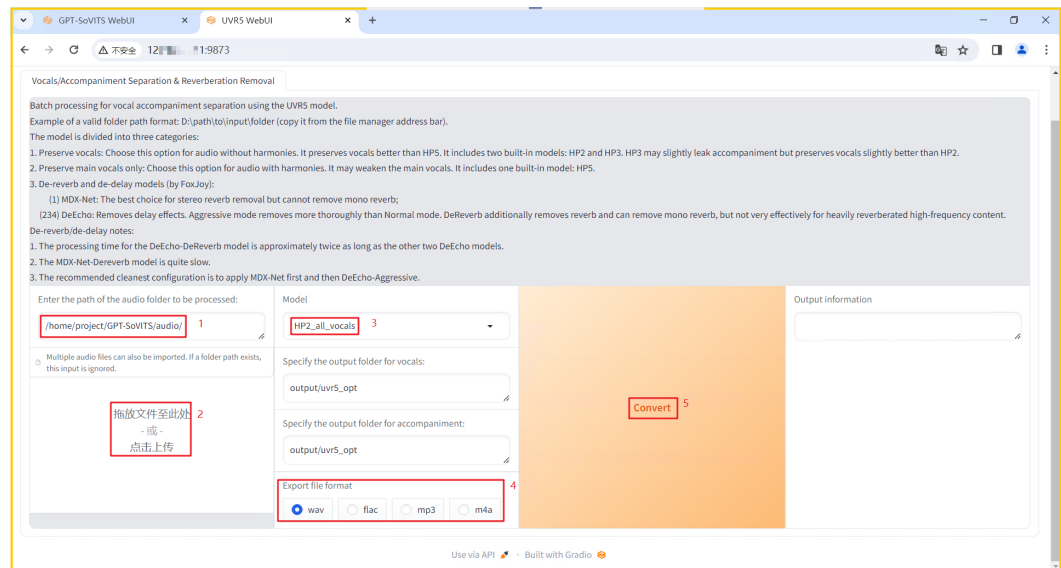


图 3-25 UVR5-WebUI



步骤2 上传需要处理的音频文件（时长1分钟到30分钟）到服务器并记录文件夹路径，或者直接拖拽文件到浏览器页面指定位置。根据页面提示说明选择合适的模型处理音频，选择需要的文件格式，单击“Covert”进行音频处理。

图 3-26 音频处理



**步骤3** 待页面出现“Success”后，代表音频文件处理成功。登录服务器，进入“/home/project/GPT-SoVITS/output/uvr5\_opt/”目录，即可找到伴奏分离的音频文件。删掉不需要的伴奏文件，保留带vocal字样的人声文件。

图 3-27 音频处理完成

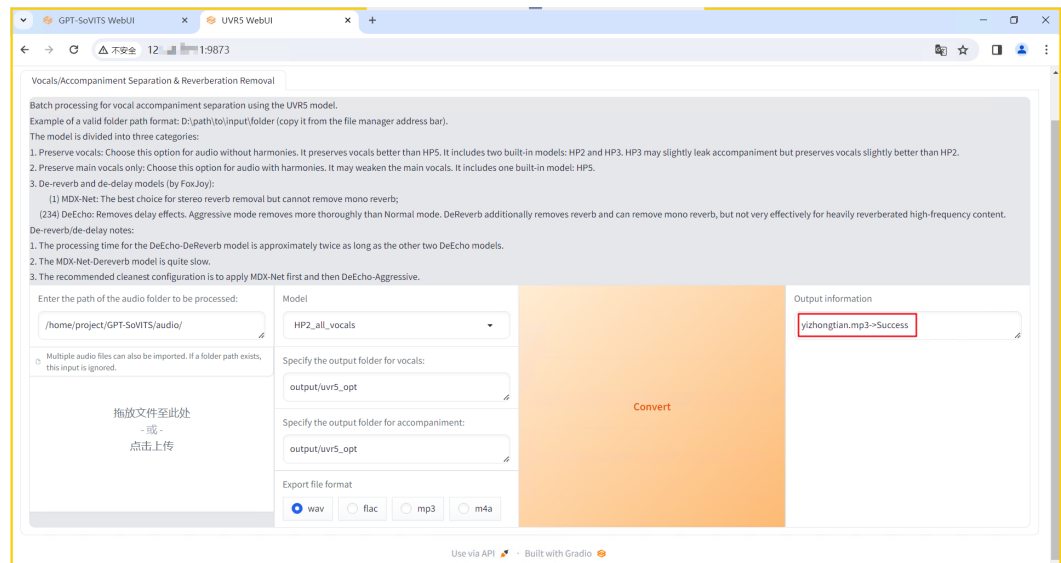
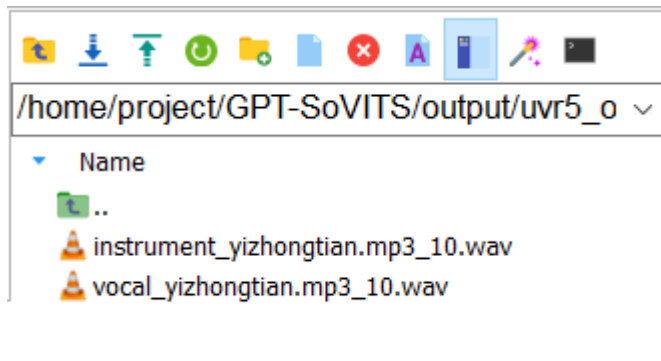


图 3-28 分离后的音频文件

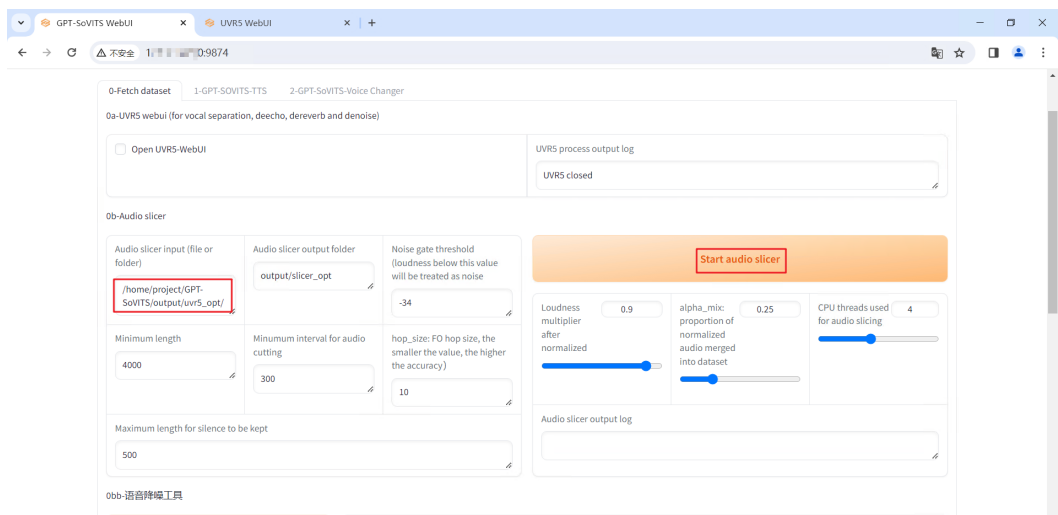


----结束

## 语音切分（必选）

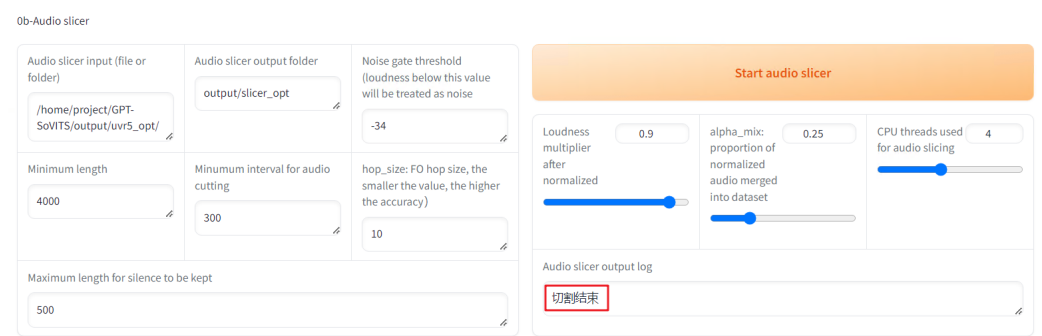
**步骤1** 上传纯人声音频文件到服务器，在GPT-SoVITS WebUI页面，输入纯人声音频文件或处理后的纯人声音频文件所在目录，单击“Start audio slicer”，进行音频切割。

图 3-29 语音切割



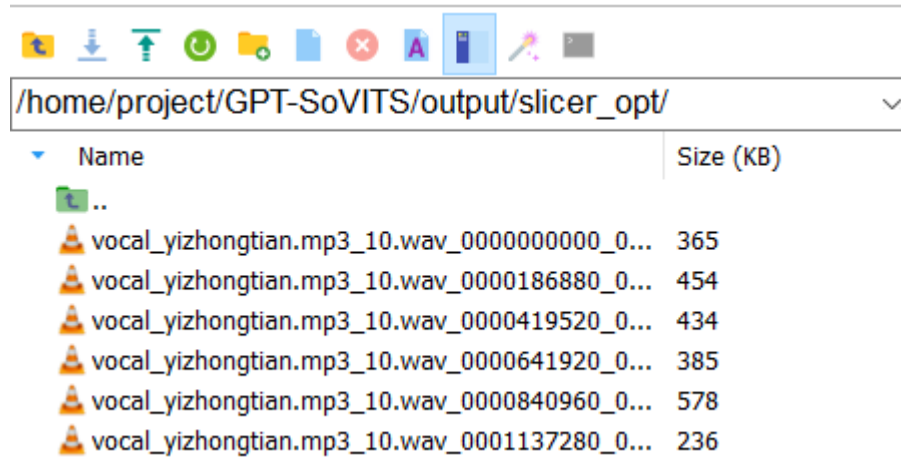
**步骤2** 页面出现“切割结束”，即代表语音文件切割完成。默认文件路径/home/project/GPT-SoVITS/output/slicer\_opt/。

图 3-30 语音切割完成



**步骤3** 登录服务器，查看切分后的音频文件。

图 3-31 语音片段文件

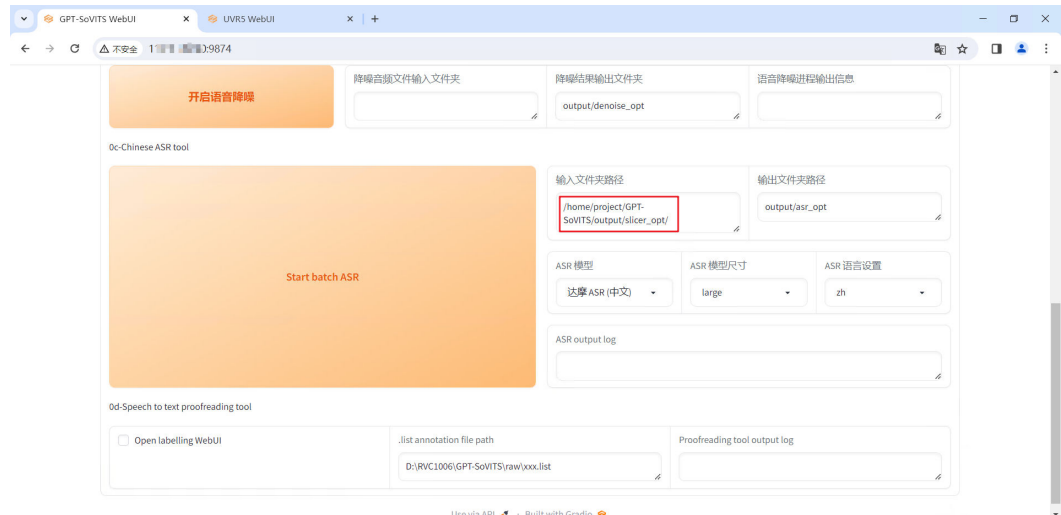


----结束

## 中文批量离线 ASR 工具（必选）

**步骤1** 以中文音频为例，在GPT-SoVITS WebUI页面，输入切割好的音频文件所在目录，单击“Start batch ASR”，进行语音转文字。

图 3-32 中文音频 ASR 处理



**步骤2** 页面出现“ASR任务完成，查看终端进行下一步”，即代表语音转文字文本文件生成成功。默认文件路径/home/project/GPT-SoVITS/output/asr\_opt。

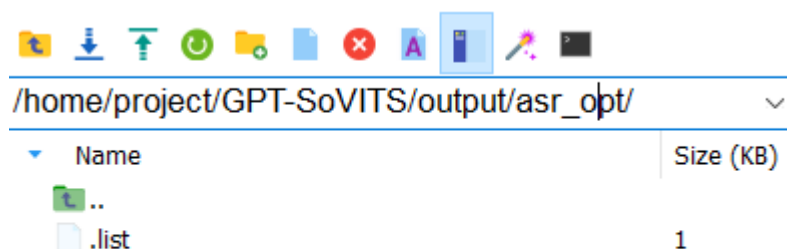


图 3-33 ASR 处理完成



**步骤3** 登录服务器，查看文本文件.list，给文件重新修改一个名字（支持字母、数字、下划线），并保存。

图 3-34 list 文件

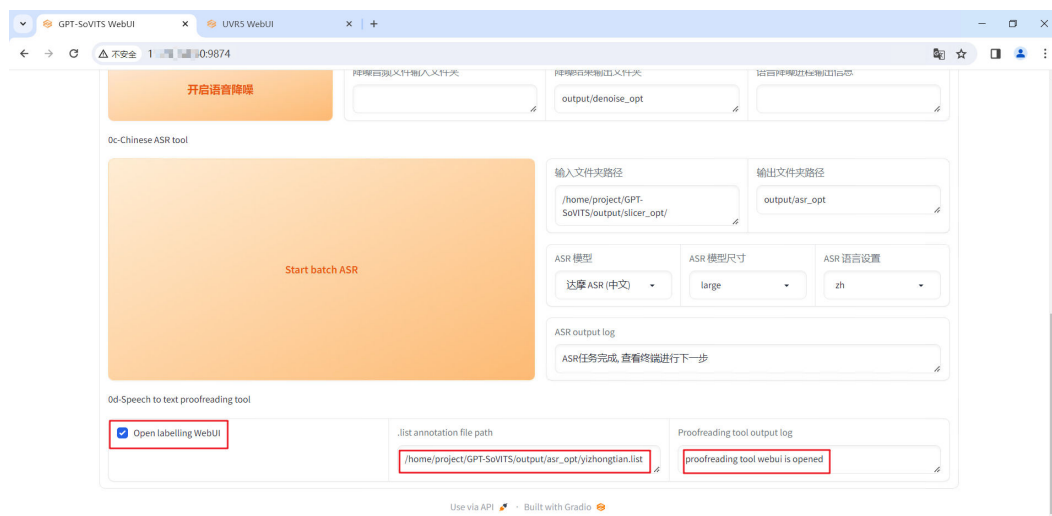


----结束

## 语音文本校对标注（可选）

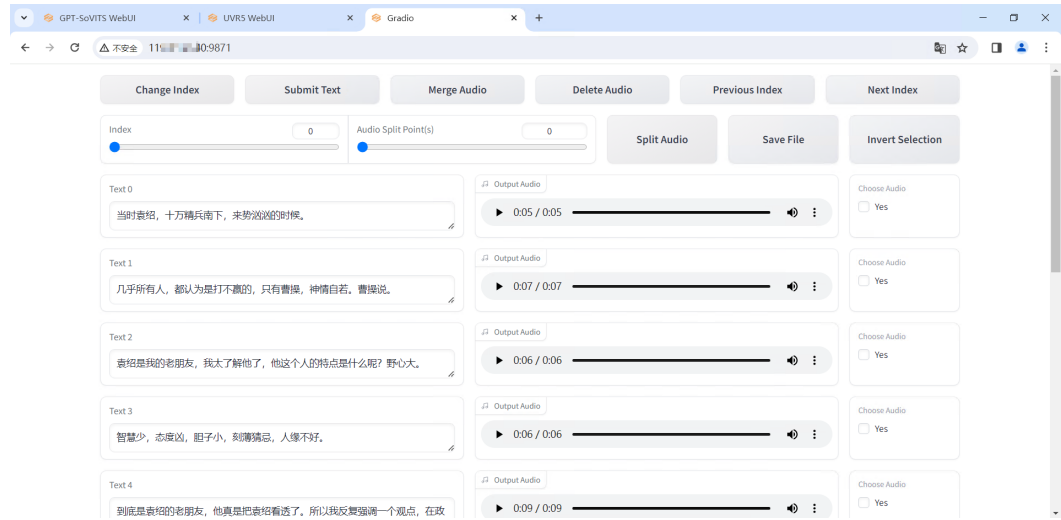
**步骤1** 在GPT-SoVITS WebUI页面，输入ASR工具处理好的list文件的路径（注意不是文件夹路径，需要带上文件名），勾选“Open labelling WebUI”，待页面出现“proofreading tool webui is opened”，打开新的窗口输入http://EIP:9871。

图 3-35 开启 Gradio WebUI



**步骤2** 在新打开的页面，对照语音和文本。根据需要，调整文本，例如标点符号和停顿位置一致和错别字。修改完成后单击“Submit Text”提交修改。

图 3-36 文本校对

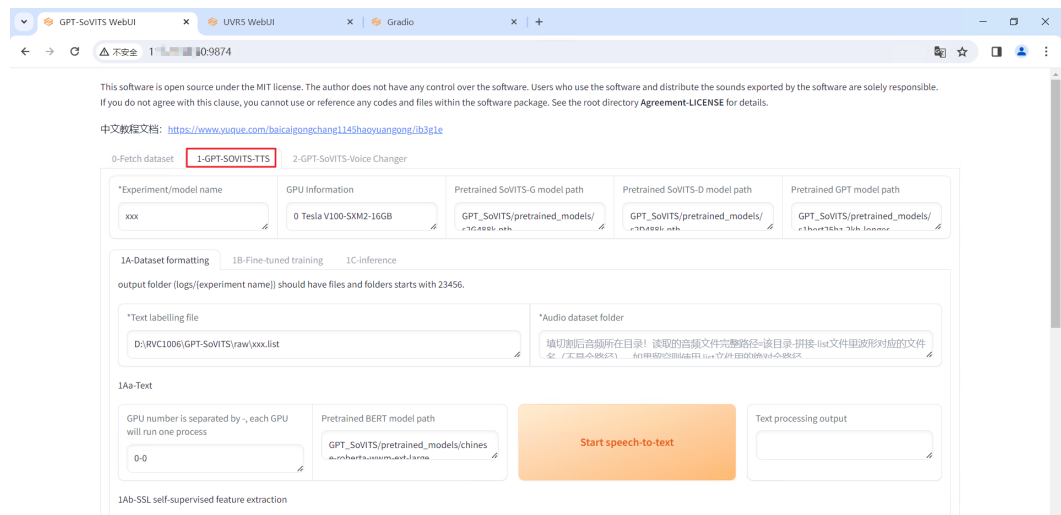


---结束

## 训练集格式化（必选）

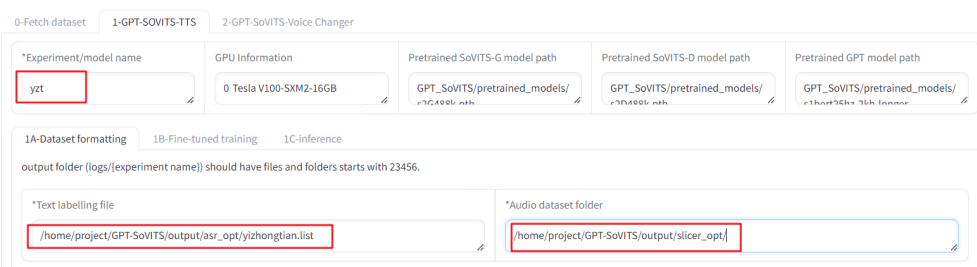
**步骤1** 在GPT-SoVITS WebUI页面，单击“1-GPT-SOVITS-TTS”，对模型进行训练。

图 3-37 GPT-SOVITS-TTS 界面



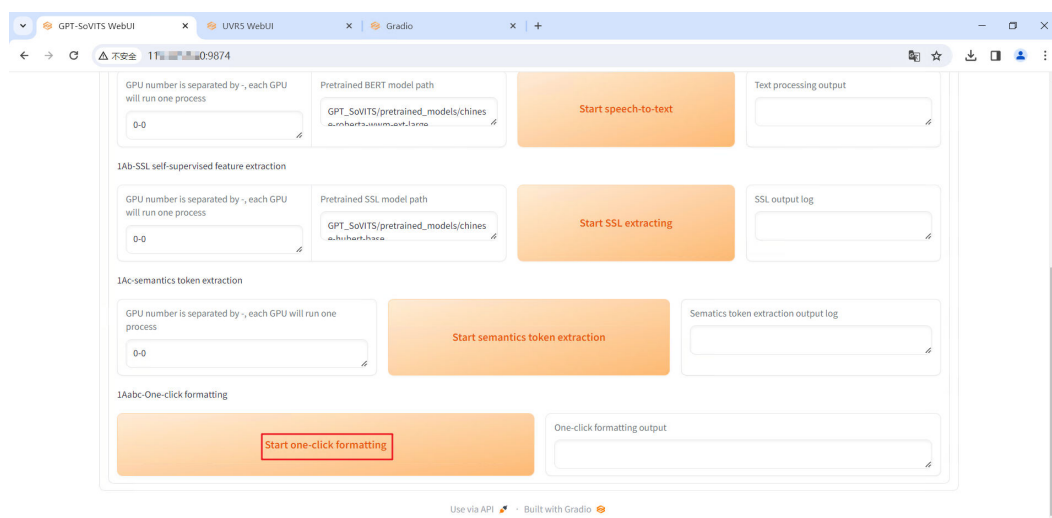
**步骤2** 输入模型名字（model name）、list文件所在路径（Text labelling file）以及切割后的音频文件所在目录（Audio dataset folder）。

图 3-38 模型参数填写



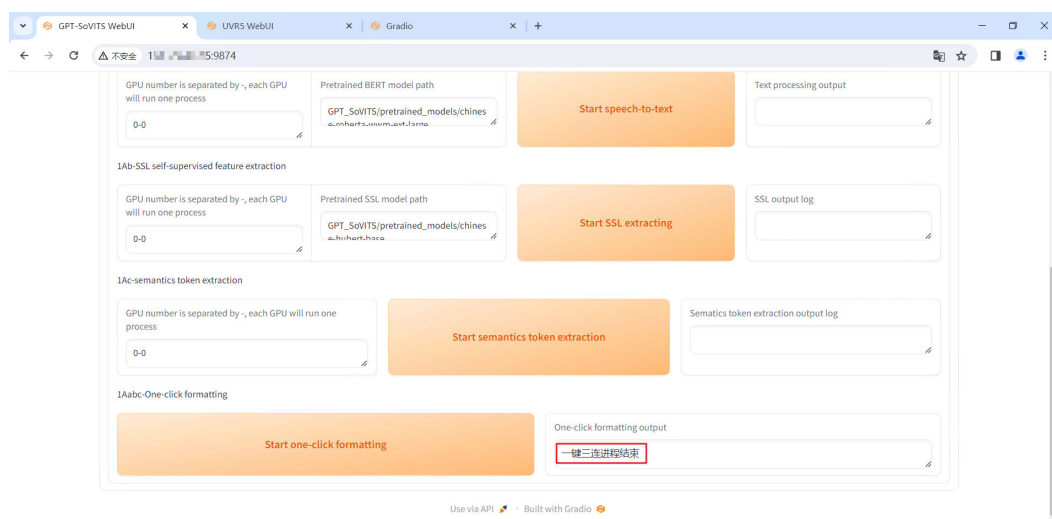
步骤3 向下滚动页面，单击一键三连“Start one-click formatting”，开始对模型进行初步训练。

图 3-39 一键三连



步骤4 训练完成后，页面显示如下。

图 3-40 一键三连完成

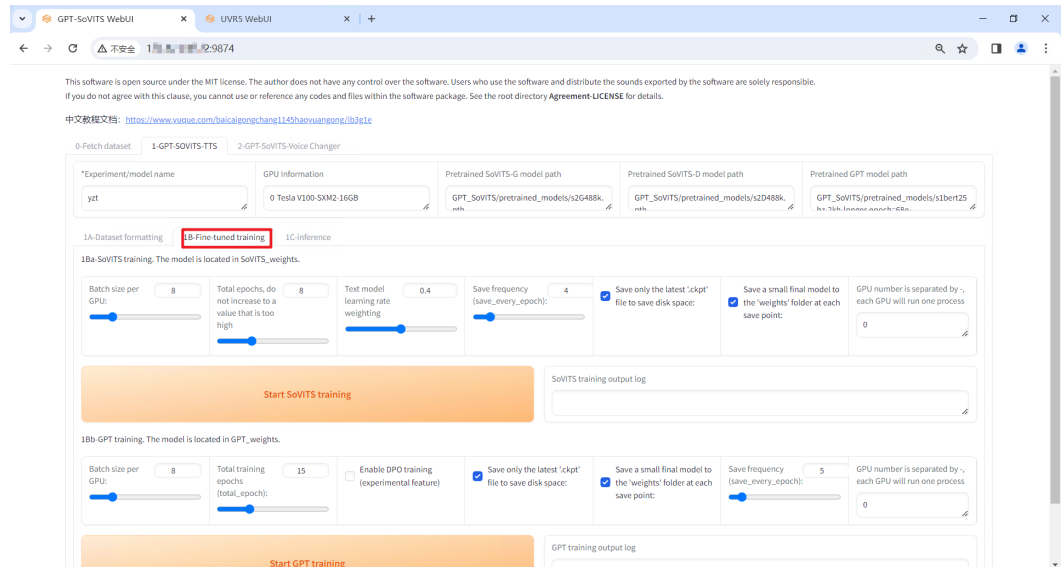


----结束

## 微调训练（必选）

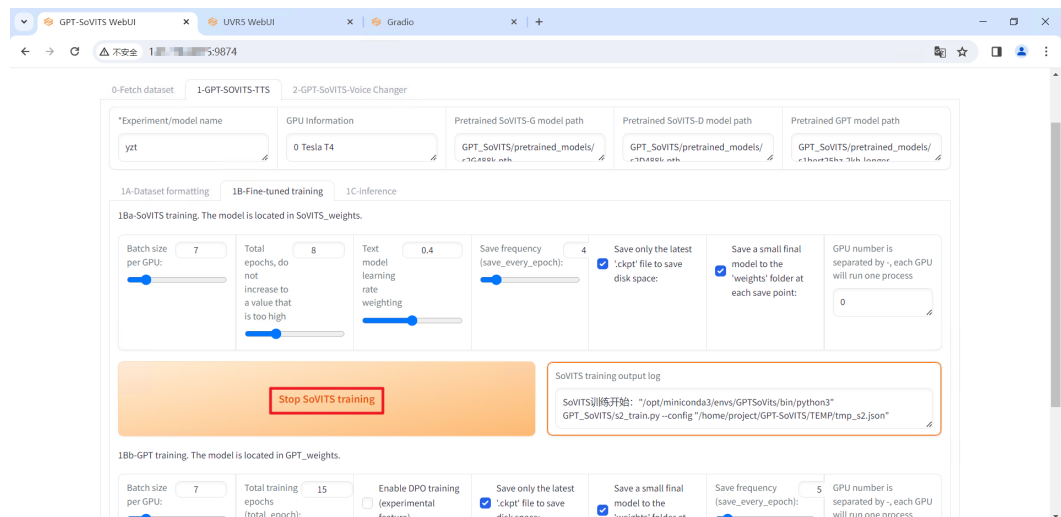
**步骤1** 在“1-GPT-SOVITS-TTS”页面，单击“1B-Fine-tuned training”，对模型进行微调。

图 3-41 微调训练



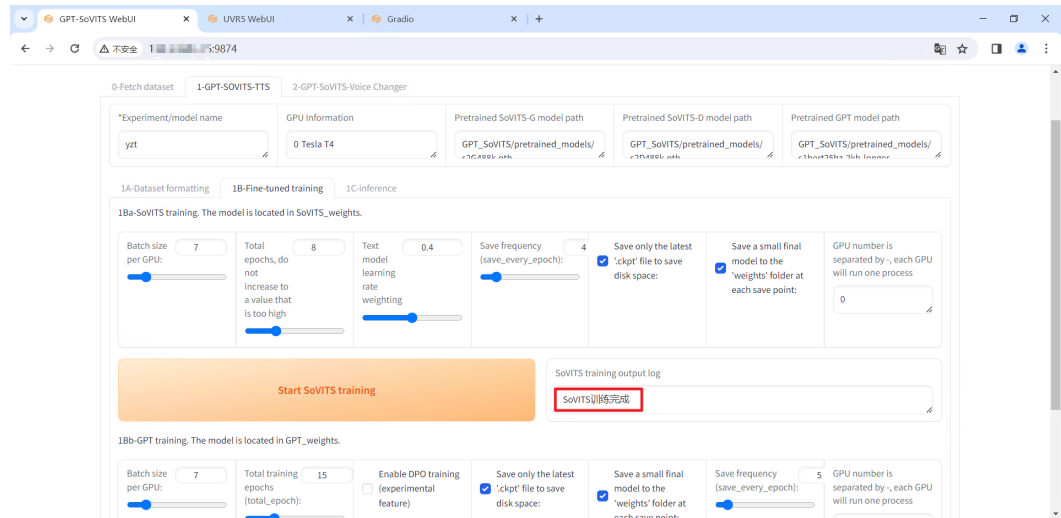
**步骤2** 保持默认参数，单击“Start SoVITS training”，进行SoVITS模型训练（如果有报错，减小Batch size per GPU数值）。

图 3-42 SoVITS 模型训练



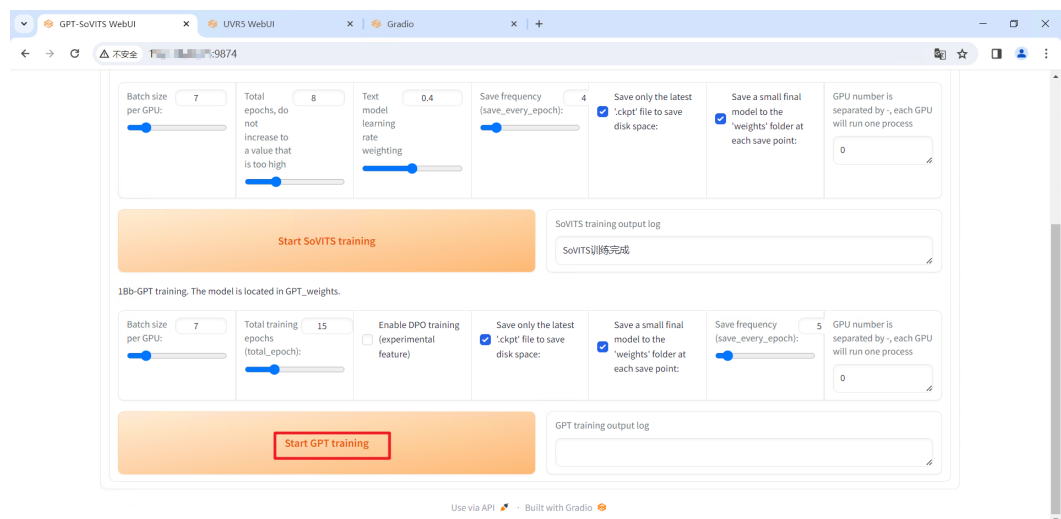
**步骤3** SoVITS模型训练完成后，页面显示如下。

图 3-43 SoVITS 模型训练完成



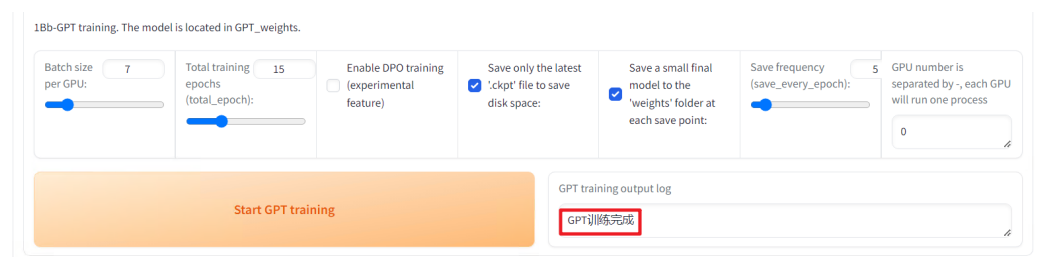
步骤4 向下滚动页面，保持默认参数，单击“Start GPT training”，进行GPT模型训练（如果有报错，减小Batch size per GPU数值）。

图 3-44 GPT 模型训练



步骤5 GPT模型训练完成后，页面显示如下。

图 3-45 GPT 模型训练完成

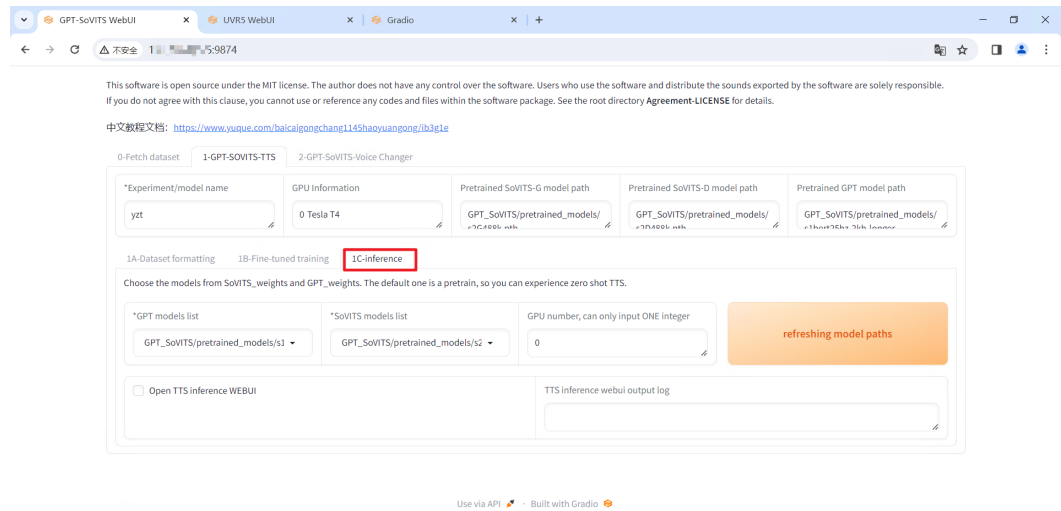


---结束

## 推理（必选）

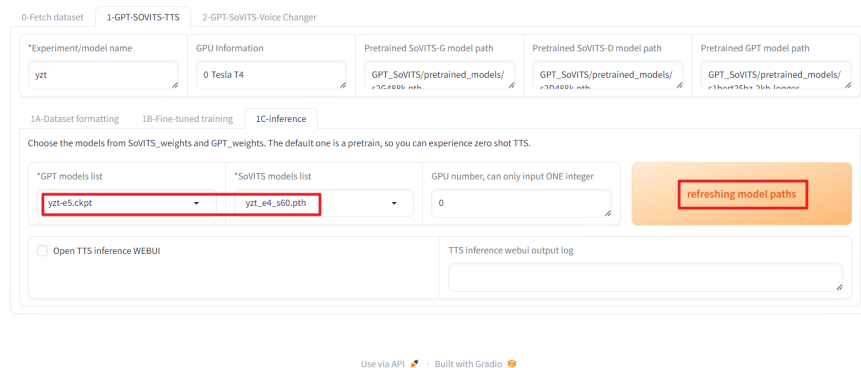
步骤1 在“1-GPT-SOVITS-TTS”页面，单击“1C-inference”，准备开始推理。

图 3-46 推理



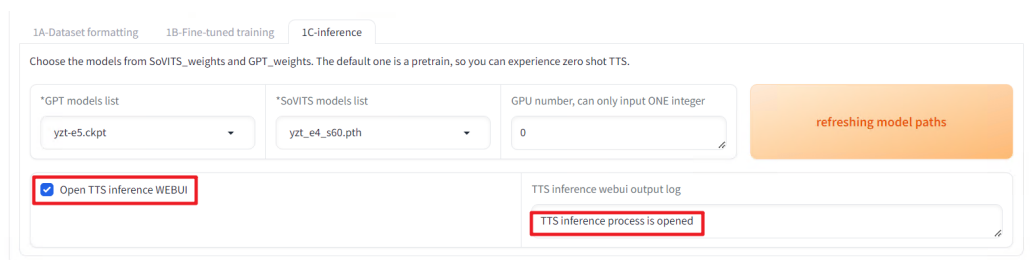
步骤2 单击“refreshing model paths”，刷新模型路径，即可选择上面训练好的模型。

图 3-47 刷新模型路径



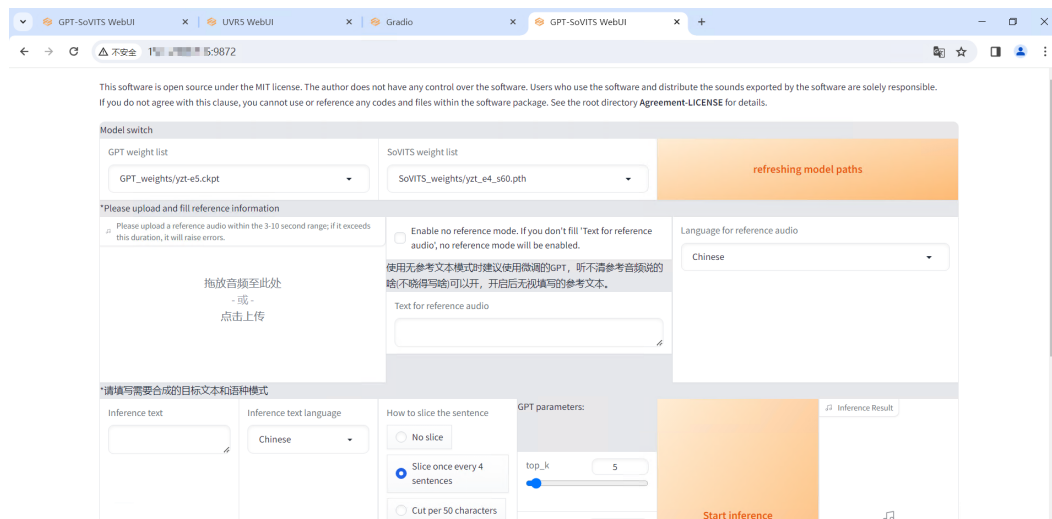
步骤3 选择任一模型，勾选“Open TTS inference WEBUI”，待页面出现“TTS inference process is opened”则开启成功。

图 3-48 开启 TTS 推理页面 WebUI



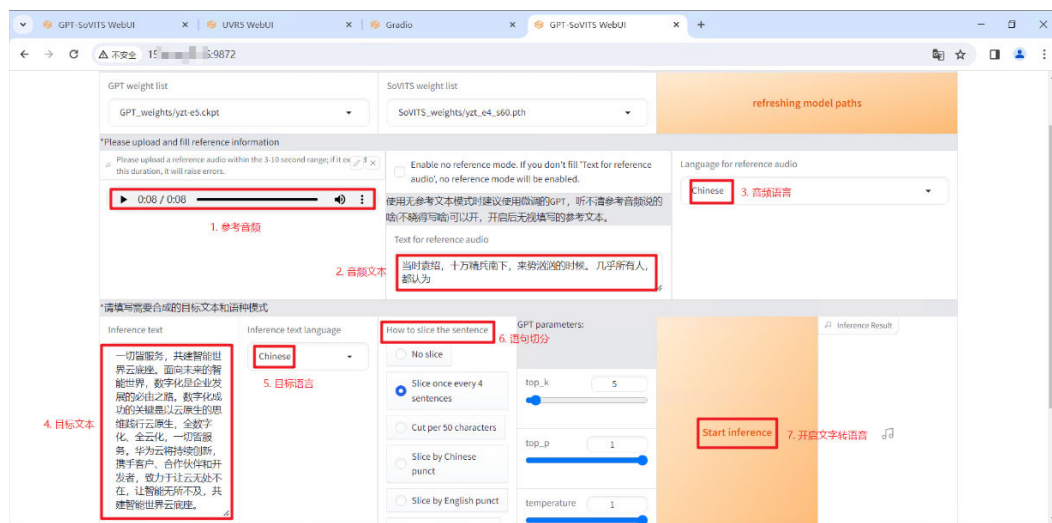
步骤4 打开新的窗口输入http://EIP:9872，其中IP为服务器公网IP，进入TTS推理页面WebUI。

图 3-49 推理页面



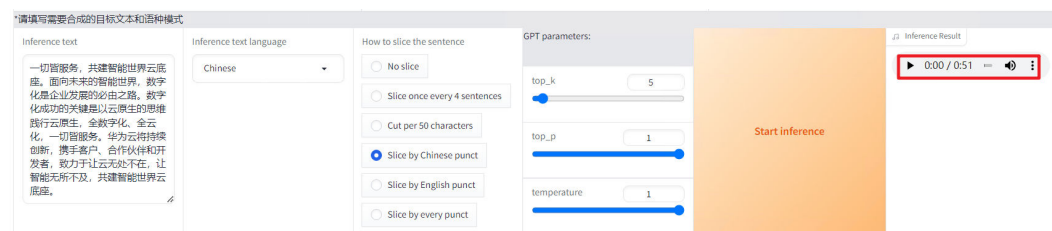
**步骤5** 上传一段3-10秒长的同一目人物参考音频文件。建议是数据集中的音频，推荐5秒时长，填写参考音频的文本。输入需要转语音的文本，单击“Start inference”开始进行语音合成。

图 3-50 语音合成参数配置



**步骤6** 语音生成完成后，如下所示。

图 3-51 文本生成语音



----结束

## 3.4 快速卸载

**步骤1** 登录[资源编排服务 RFS](#)资源栈，找到该解决方案创建的资源栈，单击资源栈名称右侧“删除”按钮，在弹出的“删除资源栈”提示框输入Delete，单击“确定”进行解决方案卸载。

图 3-52 一键卸载





图 3-53 删除资源



----结束

# 4 附录

## 名词解释

基本概念、云服务简介、专有名词解释

- **弹性云服务器 ECS**：是一种可随时自助获取、可弹性伸缩的云服务器，可帮助您打造可靠、安全、灵活、高效的应用环境，确保服务持久稳定运行，提升运维效率。
- **弹性公网IP EIP**：提供独立的公网IP资源，包括公网IP地址与公网出口带宽服务。可以与弹性云服务器、裸金属服务器、虚拟VIP、弹性负载均衡、NAT网关等资源灵活地绑定及解绑。
- **虚拟私有云 VPC**：是用户在云上申请的隔离的、私密的虚拟网络环境。用户可以自由配置VPC内的IP地址段、子网、安全组等子服务，也可以申请弹性带宽和弹性IP搭建业务系统。

# 5 修订记录

发布日期	修订记录
2024-3-30	第一次正式发布。