

GPT-SoVITS

[下载地址](#) 解压后双击 go-webui.bat 即可启动Web-UI界面

启动后不要关闭命令行界面

网页预览:

0-前置数据获取工具

1-GPT-SoVITS-TTS

2-GPT-SoVITS-变声

0a-UVR5人声伴奏分离&去混响去延迟工具

☐ 是否开启UVR5-WebUI

UVR5进程输出信息

0b-语音切分工具

音频自动识别输入路径, 可文件可文件夹

切分后的子音频的输出根目录

threshold:音量小于这个值视作静音的备选切割点

min_length:每段最小多长, 如果隔一段大短一段和前面连起来直到超过这个值

max_interval:最短切割间隔

hop_size:怎么算音量曲线, 越小精度越大计算量越高 (不是精度越大效果越好)

max_sil_kept:切完后静音最多留多长

max归一化后最大值多少

alpha_min:混多少比例归一化后降噪效果

切割使用的进程数

语音切割进程输出信息

打开语音切割

0c-中文批量离线ASR工具

批量ASR(中文only)输入文件夹路径

ASR进程输出信息

0d-语音文本校对标注工具

☐ 是否开启打标WebUI

.txt标注文件的路径

打标工具进程输出信息

使用:

1. 去除伴奏,分离人声

0-前置数据获取工具
1-GPT-SoVITS-TTS
2-GPT-SoVITS-变声

0a-UVR5人声伴奏分离&去混响去延迟工具

☐ 是否开启UVR5-WebUI

UVR5进程输出信息

0b-语音切分工具

音频自动切分输入路径, 可文件可文件夹

切分后的子音频的输出根目录
output/slicer_opt

threshold:音量小于这个值视为静音的最低切分点
-34

min_length:每段最小多长, 如果每一段太短一直和后面段落合并直到超过这个值
4000

min_interval:最短切分间隔
300

hop_size:怎么算音量曲线, 越小精度越大计算量越高 (不推荐精度大效果越差)
10

max_sl_kept:切完后静音最多留多长
500

开启语音切分

max归一化后最大值多少
0.9

alpha_min:需多少比例归一化后做降噪
0.25

切分使用的进程数
4

设置切分进程输出信息

0c-中文批量离线ASR工具

开启离线批量ASR

批量ASR(中文only)输入文件夹路径
D:\VC1000\GPT-SoVITS\raw\xxx

ASR进程输出信息

0d-语音文本校对标注工具

☐ 是否开启打标WebUI

标注文件的地址
D:\VC1000\GPT-SoVITS\raw\xxx.lst

打标工具进程输出信息

点击选择框,等待新页面跳出(可能会有些延迟,请耐心等待)

伴奏人声分离&去混响&去回声

人声伴奏分离批量处理, 使用UVR5模型。
合格的文件夹格式举例: E:\codes\py39\uits_vc_gpu\白登霍华测试样例\去文件管理器地址栏替换就行了。
模型分为三类:
1、保留人声: 不带和声的音频选这个, 对主人声保留比HP5更好。内置HP2和HP3两个模型, HP3可能轻微漏伴奏但对主人声保留比HP2稍微好一点;
2、仅保留主人声: 带和声的音频选这个, 对主人声可能有保留。内置HP5一个模型;
3、去混响、去延迟模型 (by FoxJoy):
[1]MDX-Net(onnx_dereverb):对于双通道混响是最好的选择, 不能去除单通道混响;
[234]DeEcho:去除延迟效果, Aggressive比Normal去除得更彻底, DeReverb额外去除混响, 可去除单声道混响, 但是对高频重的板式混响去不干净。
去混响/去延迟, 附:
1、DeEcho-DeReverb模型的耗时是另外2个DeEcho模型的接近2倍;
2、MDX-Net-Dereverb模型挺慢的;
3、个人推荐的最干净的配置是先MDX-Net再DeEcho-Aggressive。

输入待处理音频文件夹路径
C:\Users\Desktop\to-do-songs

☒ 也可批量输入音频文件, 二选一, 优先文件夹

Drop File Here
- OR -
Click to Upload

模型
HP5_only_main_vocal

指定输出主人声文件夹
output\ivr5_opt

指定输出主人声文件夹
output\ivr5_opt

导出文件格式
☒ wav ☒ flac ☒ mp3 ☐ m4a

7

转换

8

输出信息

①需要处理的音频目录

②可以直接选择上传音频文件

③去除杂音模型

推荐使用HP5_only_main_vocal

其他模型请根据实际情况选择

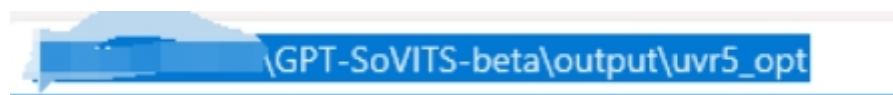
④处理完输出人声文件夹(一般默认不改)

⑤处理完输出非人声(杂音)的文件夹(一般默认不改)

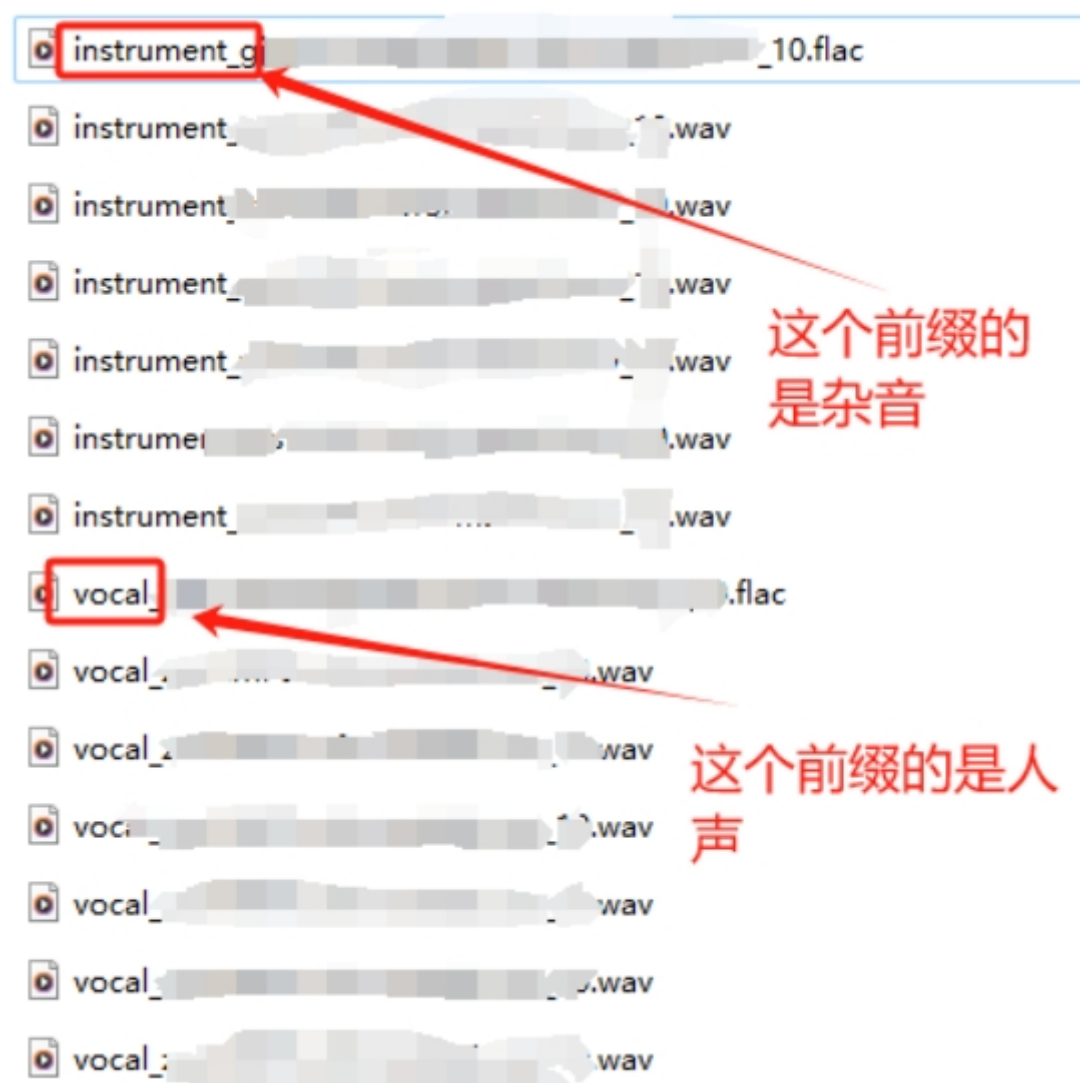
⑥音频输出格式(都能识别,没特殊要求随便选)

⑦输出信息为Success则为转换成功

进入项目输出文件夹,默认情况下进入项目根目录下output\uvr5_opt文件夹



把杂音以及前一个模型的音频删除



2.语音切分

0-前置数据集获取工具

1-GPT-SoVITS-TTS

2-GPT-SoVITS-交互

0a-UVR5人声伴奏分离&去混响去延迟工具

☒ 是否开启UVR5-WebUI

UVR5进程输出信息

UVR5已开启

0b-语音切分工具

音频自动切分输入路径，可文件可文件夹

①

切分后的子音频的输出根目录

②

output\slicer_opt

threshold:音量小于这个值视作静音的静音切割点

-34

min_length:每段最小多长，如果是一段大长一直和后面连起来直到超过这个值

4000

max归一化后最大值多少

0.9

alpha_mix:混多少比例归一化后降噪

0.25

min_interval:最短切割间隔

300

hop_size:怎么算音量曲线，越小精度越大计算量越高（不想精度放大效果越好）

10

max_sil_kept:切割后静音最多留多长

500

③

开启语音切割

切速使用的进程数

4

语音切割进程输出信息

0c-中文批量离线ASR工具

开启离线批量ASR

批量ASR(中文only)输入文件夹路径

D:\RVC1006\GPT-SoVITS\raw\xxx

ASR进程输出信息

0d-语音文本校对标注工具

☐ 是否开启打标webui

.txt标注文件的路径

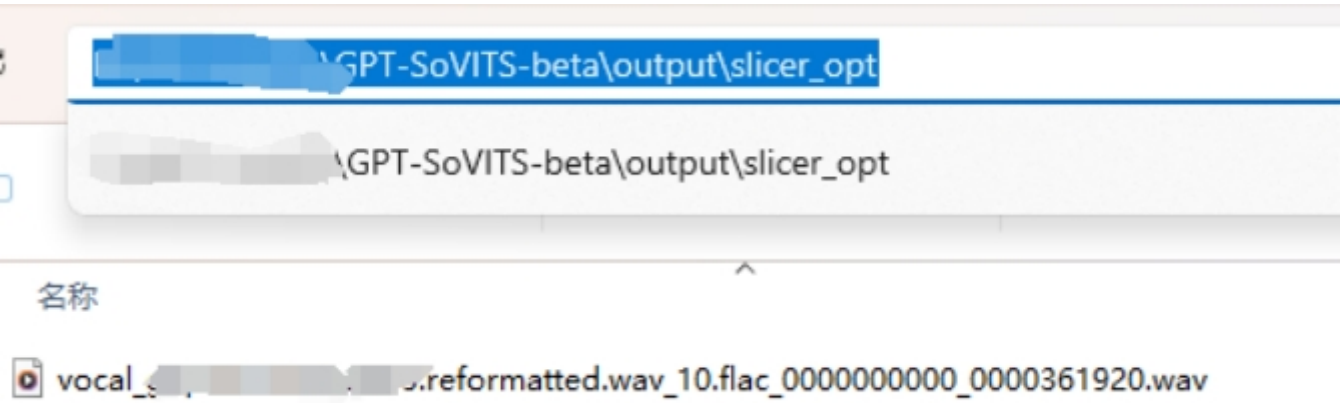
D:\RVC1006\GPT-SoVITS\raw\xxx.txt

打标工具进程输出信息

- ①将上一步音频处理输出目录路径填入,默认为根目录下output\uvr5_opt
- ②选择切割后语音输出文件夹,一般无需更改
- ③开始语音切割,输出信息为切割结束且命令行界面无报错时则为切割成功

注:参数按项目需求自行调整,一般默认参数即可

输出文件在根目录的output\slicer_opt



3.批量处理切割文件

0-前置数据集获取工具

1-GPT-SoVITS-TTS

2-GPT-SoVITS-变声

0a-UVR5人声伴奏分离&去混响去延迟工具

☐ 是否开启UVR5-WebUI

UVR5进程输出信息

0b-语音切分工具

请从自动切分输入路径，
可文件可文件夹

D:\GPT-SoVITS\GPT-
SoVITS-
beta\output\sav15_opt

切分后的子音频的输出模
目录

output\slicer_opt

threshold:音量小于这个值
视为静音的静音切分点

-34

min_length:每段最小多
长，如果第一段太长一直
和后面段连起来直到超过
这个值

4000

max_interval:最短切分间隔

300

hop_size:怎么算音量曲线，越小精度
越高计算量越高（不是精度越高效果
越好）

10

max_sil_kept:切分后静音最多留多长

500

max归一化后
最大值
多少

0.9

alpha_max:
需多少比
例归一化
后音频进
来

0.25

切割使用
的进程数

4

语音切割进程输出信息

切割结束

开启语音切割

0c-中文批量离线ASR工具

开启离线批量ASR

②

批量ASR(中文only)输入文件夹路径

D:\RVC1006\GPT-SoVITS\raw\xxx

①

ASR进程输出信息

0d-语音文本标注工具

☐ 是否开启打标WebUI

Jar标注文件的地址

D:\RVC1006\GPT-SoVITS\raw\xxx.list

打标工具进程输出信息

①输入上面处理完的文件夹,默认为根目录下output\slicer_opt文件夹

②点击处理

输出文件默认在根目录下output\asr_opt下,为.list文件

4. 文本标注

0-前置数据集获取工具
1-GPT-SoVITS-TTS
2-GPT-SoVITS-变声

0a-UVR5人声伴奏分离&去混响去延迟工具

☐ 是否开启UVR5-WebUI

UVR5进程输出信息

0b-语音切分工具

音频自动切分输入路径, 可文件可文件夹
D:\GPT-SoVITS\GPT-SoVITS-beta\output\uvr5_opt

切分后的子音频的输出根目录
output\slicer_opt

threshold:音量小于这个值视作静音的静音切割点
-34

min_length:每段最小多长, 如果某一段太短一直和相邻段连起来直到超过这个值
4000

min_interval:最短切分间隔
300

hop_size:怎么算音量曲线, 越小精度越大计算量越高 (不想精度太大效果越好)
10

max_sil_kept:切分后静音最多留多长
500

开启语音切分

max归一化后最大值多少
0.9

alpha_mix:需多少比例归一化后音量进来
0.25

切分使用的进程数
4

语音切分进程输出信息
切分结束

0c-中文批量离线ASR工具

开启离线批量ASR

批量ASR(中文only)输入文件夹路径
D:\RVC1000\GPT-SoVITS\raw\xxx

ASR进程输出信息

0d-语音文本校对标注工具

☐ 是否开启打标WebUI

标注文件的根路径
D:\RVC1000\GPT-SoVITS\raw\xxx.list

打标工具进程输出信息

①将上一步输出目录路径下的.list文件填入,默认为根目录下output\asr_opt\slicer_opt.list

②勾选,等待打标页面启动

5. 文本核对

Change Index		Submit Text		Merge Audio		Delete Audio		Previous Index		Next Index	
<div>Index 0</div>		<div>Audio 1</div>		<div>Split Audio</div>		<div>Save File</div>		<div>Invert Selection</div>			
选择下标		语音切割秒数		切割选中语音		保存文件					
Text 0		<div></div>		Output Audio		<div>0:04 / 0:04</div>		<div><input type="checkbox"/> Yes</div>			
Text 1		<div></div>		Output Audio		<div>0:02 / 0:02</div>		<div><input type="checkbox"/> Yes</div>			
Text 2		<div></div>		Output Audio		<div>0:00 / 0:04</div>		<div><input type="checkbox"/> Yes</div>			
Text		<div></div>		Output Audio		<div></div>		<div><input type="checkbox"/> Yes</div>			
Text		<div></div>		Output Audio		<div></div>		<div><input type="checkbox"/> Yes</div>			
Text		<div></div>		Output Audio		<div></div>		<div><input type="checkbox"/> Yes</div>			
Text		<div></div>		Output Audio		<div></div>		<div><input type="checkbox"/> Yes</div>			
Text		<div></div>		Output Audio		<div></div>		<div><input type="checkbox"/> Yes</div>			

这里需要核对声音和文本是否对应,多听几遍

切割功能一般不在这里使用,建议在第一步后使用剪映等软件进行处理(比网页端更准确)

核对完后需要点击保存文件

6. 训练格式化

0-前置数据集获取工具 1-GPT-SoVITS-TTS 2-GPT-SoVITS-变声

实验模型名: xxx ① 显卡信息: [显示卡信息]

预训练的SoVITS-G模型路径: GPT_SoVITS/pretrained_models/s2G488k.pth 预训练的SoVITS-O模型路径: GPT_SoVITS/pretrained_models/s2O488k.pth 预训练的GPT模型路径: GPT_SoVITS/pretrained_models/s1bert256.pth

1A-训练集格式化工具 1B-微调训练 1C-推理

输出logs/实验名目录下应有23456开头的文件和文件夹

*文本标注文件: D:\RVC1806\GPT-SoVITS\raw\xxx.list ② *训练集音频文件目录: 填写音频所在目录！读取的音频文件完整路径-该目录 创建 txt文件里要对应文件名（不是全路径）。 ③

1Aa-文本内容

GPU卡号以分割，每个卡号一个进程: 0-0 预训练的中文BERT模型路径: GPT_SoVITS/pretrained_model\chinese_bert4keras\chinese_bert4keras.pth 开启文本获取 文本进程输出信息: [输出框]

1Ab-SSL白监督特征提取

GPU卡号以分割，每个卡号一个进程: 0-0 预训练的SSL模型路径: GPT_SoVITS/pretrained_model\ssl\ssl_model.pth 开启SSL提取 SSL进程输出信息: [输出框]

1Ac-语义token提取

GPU卡号以分割，每个卡号一个进程: 0-0 开启语义token提取 语义token提取进程输出信息: [输出框]

1Aabc-训练集格式化一键三连

开启一键三连 ④ 一键三连进程输出信息: [输出框]

这些一般不需要去修改

①项目名称(自定义)

②.list文件路径,默认为根目录下output\asr_opt\licer_opt.list

③第二步中的输出文件夹,默认为根目录下output\licer_opt

④开始批处理

7. 微调训练

0-前置数据集获取工具 1-GPT-SoVITS-TTS 2-GPT-SoVITS-变声

*实验模型名:

显卡信息: 0 NVIDIA GeForce RTX 4070 Ti SUPER

预训练的SoVITS-G模型路径: GPT_SoVITS/pretrained_models/s2G488k_

预训练的SoVITS-D模型路径: GPT_SoVITS/pretrained_models/s2D488k_

预训练的GPT模型路径: GPT_SoVITS/pretrained_models/s1bert25

1A-训练集格式化工具 1B-微调训练 1C-推理

1Ba-SoVITS训练。用于分享的模型文件输出在SoVITS_weights下。

每张显卡的batch_size: 8 总训练轮数total_epoch, 不建议太高: 8 文本模块学习率权重: 0.4 保存频率save_every_epoch: 4 ☒ 是否仅保存最新的ckpt文件以节省硬盘空间 ☒ 是否在每次保存时间也将最终小模型保存至weights文件夹 GPU卡号以分割, 每个卡号一个进程: 0

开启SoVITS训练 ①

SoVITS训练输出信息

SoVITS训练完成

1Bb-GPT训练。用于分享的模型文件输出在GPT_weights下。

每张显卡的batch_size: 8 总训练轮数total_epoch: 15 ☒ 是否仅保存最新的ckpt文件以节省硬盘空间 ☒ 是否在每次保存时间也将最终小模型保存至weights文件夹 保存频率save_every_epoch: 5 GPU卡号以分割, 每个卡号一个进程: 0

开启GPT训练 ②

GPT训练输出信息

GPT训练完成

这里需要结合实际生成的声音进行参数的微调整

8. 推理

0-前置数据集获取工具 1-GPT-SoVITS-TTS 2-GPT-SoVITS-变声

*实验模型名: 后Q2

显卡信息: 0 NVIDIA GeForce RTX 4070 Ti SUPER

预训练的SoVITS-G模型路径: GPT_SoVITS/pretrained_models/s2G488k_

预训练的SoVITS-D模型路径: GPT_SoVITS/pretrained_models/s2D488k_

预训练的GPT模型路径: GPT_SoVITS/pretrained_models/s1bert25

1A-训练集格式化工具 1B-微调训练 1C-推理

选择训练文件存放在SoVITS_weights和GPT_weights下的模型。默认的一个是底模, 体验5秒Zero Shot TTS用。

*GPT模型列表: GPT_SoVITS/pretrained_m *SoVITS模型列表: GPT_SoVITS/pretrained_m GPU卡号, 只能填1个整数: 0

刷新模型路径 ①

☒ 是否开启TTS推理WebUI ②

TTS推理WebUI进程输出信息

TTS推理进程已开启

①处理完后点击刷新模型,前面模型列表中会出现刚才训练好的模型

②点击勾选框,打开新页面(页面加载较慢,请耐心等待)



①点击刷新模型

②③选择前面训练好的模型

④选择参考语音(非常重要)

选择一条该音色的参考语音,最好从前面分割的语音中选择,咬字清晰

⑤参考语音中说的内容

⑥需要推理的内容

⑦点击推理

到这里模型训练基本已经结束了

注:语音切割按实际情况自行判断使用哪种,一般短句选择不切

GPT模型在根目录\GPT_weights下

SoVITS模型在根目录\SoVITS_weights下

一个GPT模型+SoVITS模型+参考语音为一个组合(对应的参考语音需要一起保存,一般保存一条即可)

进行新模型训练时请把\output下三个子文件夹内的内容清空,否则会影响下个模型