



Sing in Harmony with AI:
Your Voice, Any Style

Unleash Your
Inner Singer
with AI

VOICECRAFT SVC VAE × 深度学习 歌声转换算法



基于深度学习和VAE结构的歌声转换算法研究



创新声音转换，为每位
歌者而生

VOICECRAFT



汇报人
高孙炜

汇报团队
高孙炜 (130%) 林
中天 (70%)



CONTENTS

目录



01

Docer

Part One

选题背景

为什么是SVC?



02

Docer

Part Two

项目概览

我们做了什么?



03

Docer

Part Three

项目评估

我们做的如何?



04

Docer

Part Four

项目总结

还可以改进的...





PART × 01

选题背景

为什么是SVC?



选题背景

SVC × 歌声转换

introduction

Singing Voice Conversion (SVC) 是音频处理领域的一个活跃研究方向，涉及到将一位歌手的歌声风格转换为另一位歌手的声音特征。

核心目的

实现声音特征的高度相似性

01

重大挑战

保留歌曲的原始情感和音质

02

研究地位

音频处理领域的重大研究方向

03

痛点分析

SVC × 歌声转换



痛点一 音频质量损失

声音的自然度和情感表达是评价转换质量的关键指标，任何在转换过程中产生的不自然现象都会直接影响用户体验

01



痛点二 多样性和可拓展性

当前许多系统在处理多样化的声音数据时效果有限，且对新声音样本的适应能力不强，对于不同语言和方言的适应性也是一大问题

02

技术目标

Our Goal



高度自然的
声音转换

宽泛的适应性



用户友好的
交互页面



PART × 02

项目概览

我们做了什么？





开发工具选择

- 基于Python使用vscode开发工具简化开发流程
- 使用anaconda进行软件包管理
- 使用pytorch的CUDA版本进行GPU加速

TensorBoard可视化



SVC

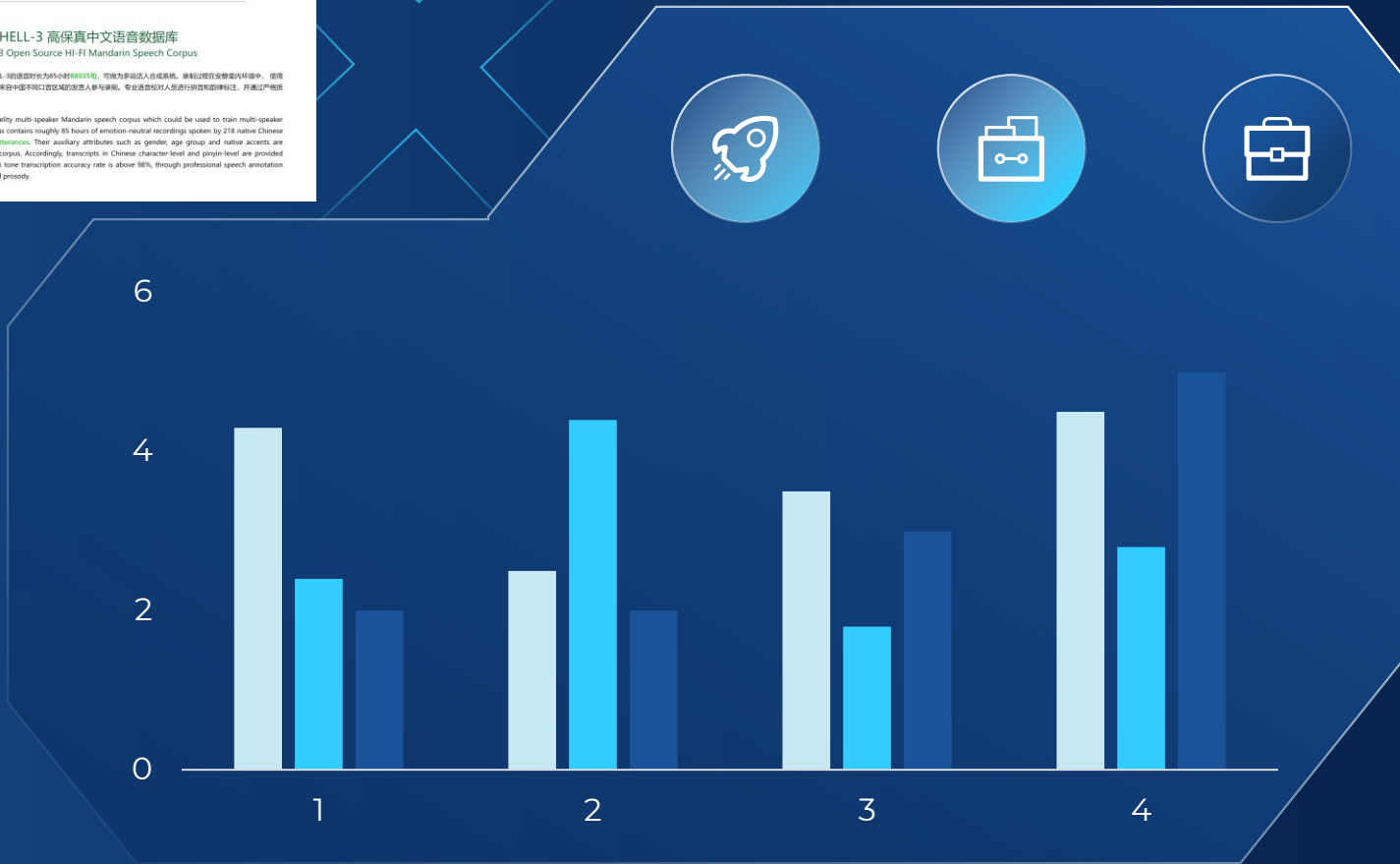
歌声转换

数据预处理



过程介绍

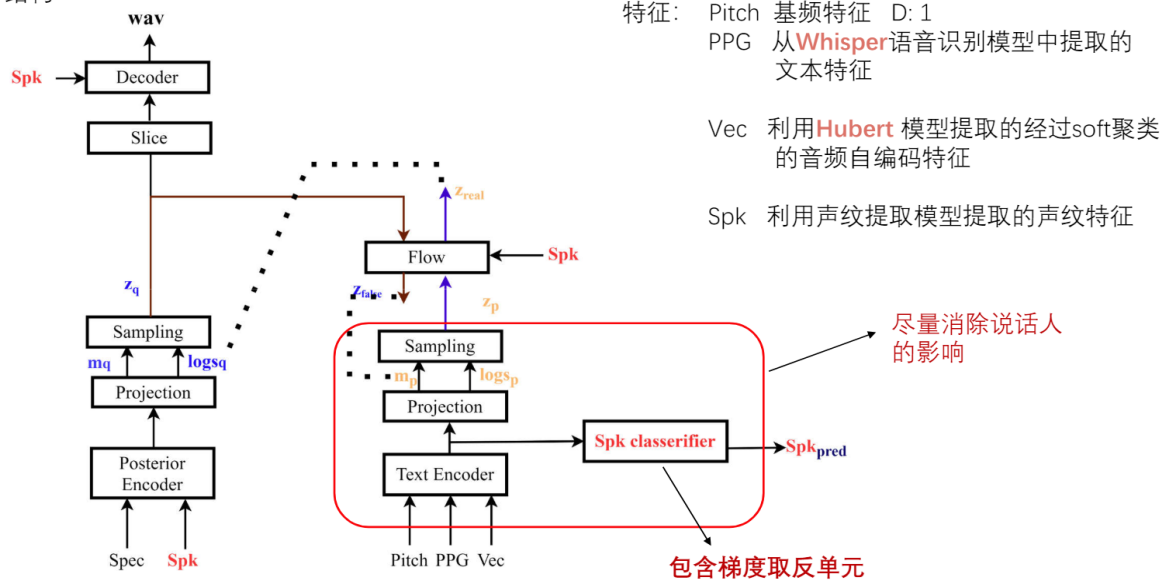
- ◆准备若干段原歌声音频文件和歌手声音的音频文件
- ◆对音频进行重采样为32K和16K
- ◆使用神经网络提取基频特征Pitch使用OpenAI的Whisper语言识别模型提取文本特征PPG
- ◆使用Hubert模型提取音频自编码特征Vec
- ◆使用声纹提取模型提取声纹特征Spk



03 模型搭建

- 搭建先验编码器（使用pytorch搭建卷积神经网络，对输入的预处理数据采样，消除说话人的影响）
- 搭建后验编码器
- 搭建解码器（将先后验编码器输出的采样值还原成音频文件）
- 搭建FLoW（将前后验编码器输出的采样值相互转换，以便比较）

模型整体结构



04 模型训练

■ 远程连接华为云ModelArts，在ModelArts上进行训练

ModelArts

总览

Workflow

自动学习

数据管理

开发环境

训练管理

部署上线

镜像管理

AI Gallery

专属资源池

全局配置

控制台

北京四

搜索云服务、

备案

资源

费用

企业

工具

工单

消息

帮助

简体

hid_gzswljfgcu9fwrs

VS Code扩展

PyCharm插件

ModelArts SDK

使用指南

创建

1 诚邀您参加notebook使用体验调研，您宝贵的意见和建议是我们持续提升产品体验的源动力，感谢您的参与!

1 状态为“运行中”的Notebook正在产生费用，不使用时，请及时停止。若您已经设置了自动停止，请注意Notebook的剩余运行时间。

查看所有

您最多可以创建10个Notebook，还可以创建8个Notebook

选择属性筛选，或输入关键字搜索

名称	状态	镜像	规格	描述	创建时间	创建者	操作
notebook-8be8 936bea3e-02fe-40a1-8adf-c9390a4...	运行中 (2)	pytorch_2_0.p...	GPU: 1*P100(...	--	2024/05/13 11...	hid_gzswljfgc...	打开 启动 停止 更多
notebook-1d80 936bea3e-e740-4f59-8039-8fa1f3e...	停止	pytorch1.8-cu...	GPU: 1*P100(...	--	2024/05/12 23...	hid_gzswljfgc...	打开 启动 停止 更多

总条数: 2 10 1

SSH: ModelArts-notebook-1d80

main*

行 1, 列 1

UTF-8

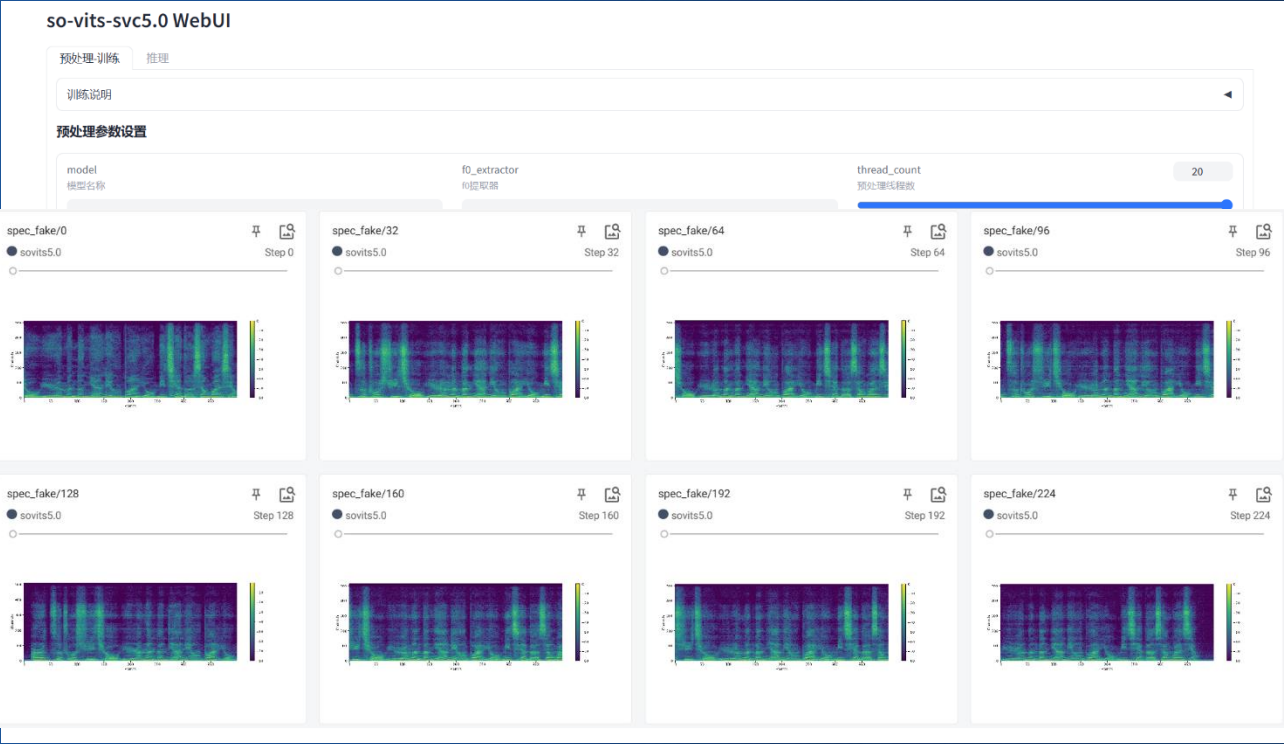
Python 3.7.10 (PyTorch-1.8) conda

剩余时间: 57分钟



05 UI绘制

- TensorBoard训练日志可视化
- Gradio训练、推理页面可视化





PART × 03

模型评估

我们做的咋样？



损失函数

Spk_loss:使预测的spk与真实spk不相似:

$$\text{loss}(x, y) = 1 - \cos(x_1, x_2), \text{ if } y = 1$$

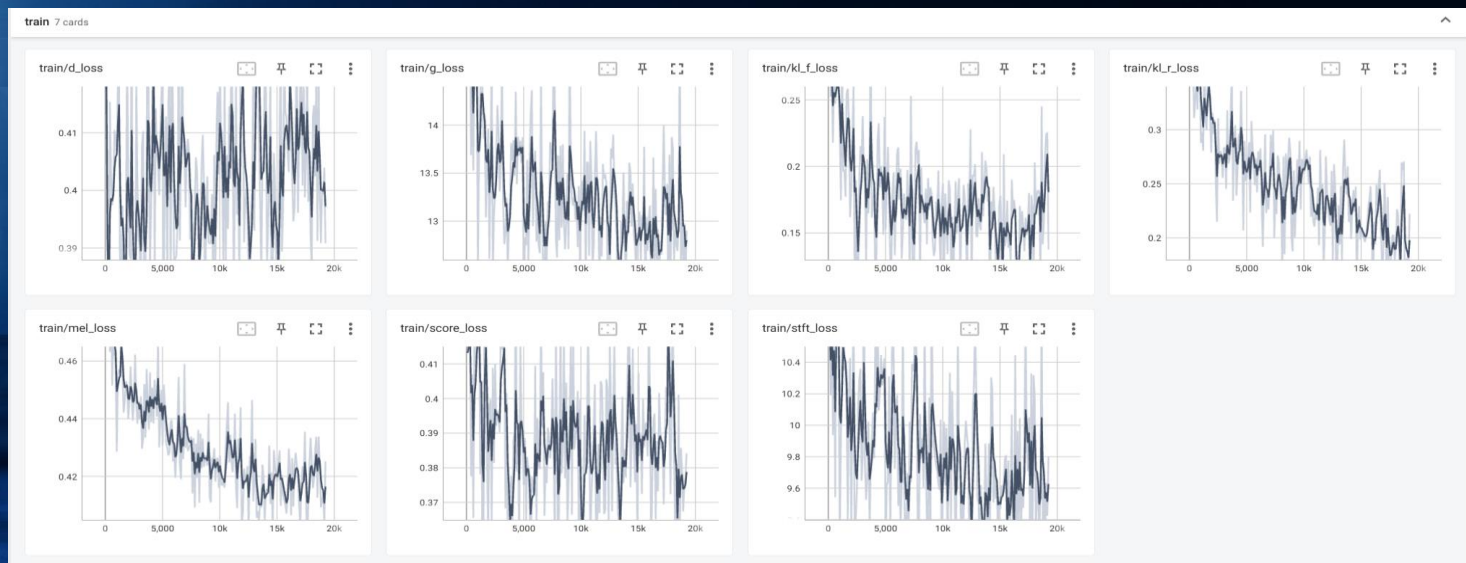
Mel_loss:生成音频和真实音频之间的L1 LOSS

Stft-loss: 生成音频和真实音频之间采用不同窗长、步进计算STFT

Score_loss(GAN)生成音频输入鉴别器, 输出为1

Feature_loss:生成音频和真实音频在鉴别器每层的输出都相似

KL_loss: 分布p和分布q接近, 分别计算 $KL(p||q)$ 和 $KL(q||p)$



产品主要优势



主要优势

声音的自然度

01

产品优势

高灵活性

02

主要优势

良好的用户体验

03

主要优势

应用的广泛性

04



产品主要优势



声音的自然度

通过结合损失函数Mel_loss和Stft_loss，模型能够在音频的波形和频谱特性上更好地逼近真实音频。Mel_loss通过比较输入和生成的音频的Mel频谱，有助于保持音频的音色和音质。而Stft_loss则进一步确保了在频域上的一致性，提高了音频的时间-频率分辨率。



高灵活性

svc技术使用简单的损失函数或生成对抗网络（GAN）来重建声学特征，捕捉源歌手和目标歌手的声​​音特征，可以通过微调来适应特定的应用场景。



良好的用户体验

歌声转换技术采用了端到端的非并行结构，使用生成对抗网络（GAN）技术和迁移学习。这意味着整个转换过程可以从输入直接得到输出。



应用的广泛性

SVC技术不仅可以应用于音乐制作，还可以用于SVC技术的应用范围可能会扩展到更多的媒体类型，如电影配音、虚拟角色配音、音乐制作等领域。它的灵活性和广泛的应用前景使其成为语音处理领域的一个热点。



Summarize

PART × 04

项目总结

还有可以改进的地方...



不足之处



¥ 456

算力平台

算力平台成本高昂，导致训练迭代轮数有限、同时数据集规模较小，导致模型损失图像显示上波动较大、尚未稳定



应用门槛

目前的使用需要用户提供转换目标音色文件或者编码捏造音色，用户门槛较高，最终目标应该让用户可以直接以录音方式实现声音转换





THANKS

THANK YOU

SVC × 声音转换

感谢观看



创新声音转换，为每位
歌者而生

VoiceCraft



汇报人
高孙炜

汇报团队
高孙炜 林中天