

现在你可以通过深度学习用别人的声音来说话了

灰灰  
欢迎大家关注公众号“磐创AI”，学习更多AI知识。

+ 关注他

34 人赞同了该文章

语音合成（Text-to-speech，TTS）是指文本到音频的人工转换，也可以说给定一段文字去生成对应的人类读音。人类通过阅读来完成这项任务，而一个好的TTS系统是让计算机自动完成这项任务。

在打造这样一个系统时，一个非常有趣的地方是为生成的音频选择哪个声音，是男人还是女人的声音？声音是大还是小？

在使用深度学习创建TTS时，这有一个限制，你必须收集文本-音频的数据集，而录制演讲的演讲者可能是固定的——因为你不可能有无限多个演讲者！

所以，如果你想为自己或他人的声音创建音频，唯一的方法就是收集一个全新的数据集。

来自谷歌名为 Voice Cloning 的人工智能研究使计算机可以用任何类型声音发出声音。

**Voice Cloning是怎样工作的**

很明显，为了让电脑能够使用任何类型声音发出声音，它需要理解两件事:它读的是什么以及它是如何读的。

因此，谷歌研究人员设计的语音克隆系统有两个输入:我们想要读取的文本和我们想要用来阅读文本的语音样本。

例如，如果我们想让蝙蝠侠读“我爱披萨”这句话，那么我们会给系统两样东西：“我爱披萨”的文本和一小段蝙蝠侠的声音样本，这样它就知道蝙蝠侠的声音应该是什么样的。而电脑的输出应该是蝙蝠侠说“我爱披萨”的声音音频！

从技术角度来看，系统分为三个部分：

▲ 赞同 34 ▼

💬 1 条评论

➦ 分享

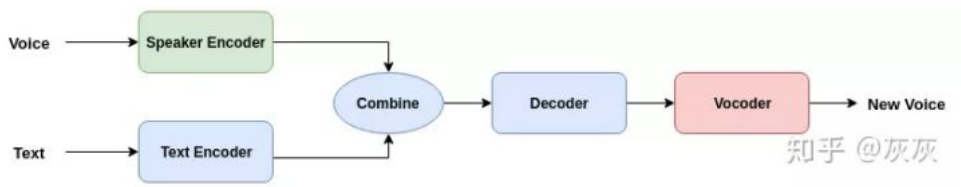
♥ 喜欢

★ 收藏

⋮

图

3. 使用声码器将声谱图转换成我们可以听到的音频波形。



在文章中，我们对这三个部分分别进行了训练。

在过去几年中，语音合成系统在深度学习社区中得到了很多研究关注。事实上，有很多基于深度学习针对语音合成的解决方案都非常有效。

这里的关键是，系统能够将编码器从语音中学到的“知识”应用到文本中。

在分别编码后，将语音和文本组合在一个公共的嵌入空间中，然后进行解码，生成最终的输出波形。

克隆声音的相关实现代码

多亏了人工智能社区中开放源码思想的美妙之处，在这里有一个公开可用的语音克隆实现！你可以按照下面的介绍这样来使用它。

1. 首先克隆仓库：

```
git clone https://github.com/CorentinJ/Real-Time-Voice-Cloning.git
```

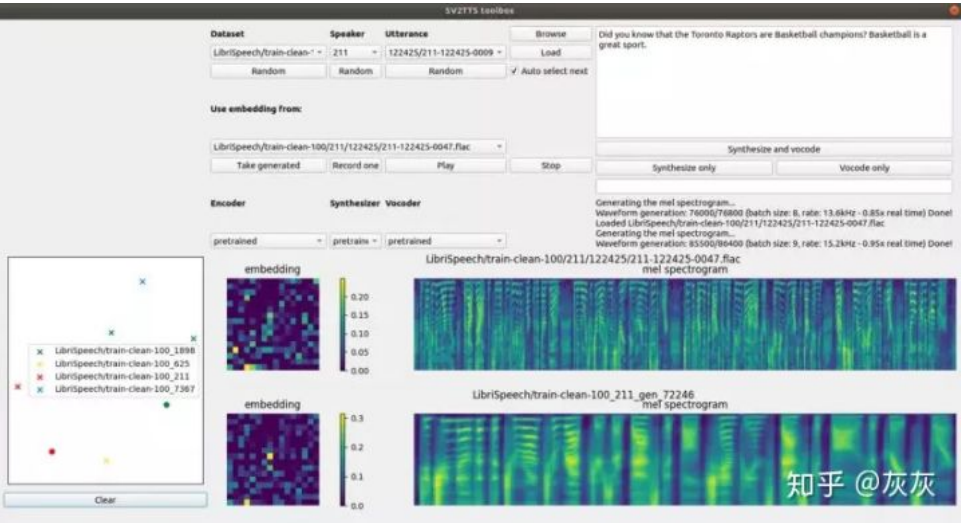
1. 安装必要的库并确保使用的是Python 3的版本：

```
pip3 install -r requirements.txt
```

- 1. 在README文件中，你还可以找到下载预训练模型和数据集的连接，并尝试一些示例。
- 2. 最后，你可以运行以下命令打开GUI:

```
python demo_toolbox.py -d <datasets_root>
```

下面是我这边运行操作的一个截图：



正如你所看到的，右边的输入框中是我刚才朗读过的内容：“Did you know that the Toronto Raptors are Basketball champions? Basketball is a great sport.”

Z'í'ǎǎ 从你选择的语音样本中选择数据集, Dǎ'ǎǎ 选择说话的人, l'í'ǎǎ 选择输入语音所说的短语。如果想听输入的语音是怎样的, 只需点击 ħĕĕĕ 按钮。

一旦你点击 **Run** 按钮，算法就会运行。一旦运行结束，你将听到扬声器以某个声音朗读你的输入文本。

你甚至可以录制你自己的声音作为输入，只需要点击 **⌘E** 或 **⌘Q** 按钮。相信我，这会非常有趣！

更进一步

如果你想了解更多关于这个算法的工作原理，你可以阅读谷歌的官方论文。这里还有一些音频样本结果。

论文链接: <https://doi.org/10.1016/j.jmb.2019.04.008>

你也许还想看：

知识图谱里的知识表示:  $E \times R \times E$

## 一文综述Hübn̄' 读写\kt̚ Ǟ̄ Ę̄' 文件各种骚操作

## 从零开始实现穿衣图像分割完整教程（附Hunb̄e代码演练）

欢迎扫码关注:



发布于 6 0 5 n ǎ ( ǎ (

深度学习 (k XXH±X Hēānj)

## 语音识别

文章被以下专栏收录



磐创2AE

欢迎大家关注公众号“磐创24h” 学习更多24h知识。

关注专栏

## 推荐阅读

▲ 贊同 78

5 条评论

分享

♥ 喜欢

★ 收藏

...



Supervised Text to Speech and Automatic Speech Recognition

《2 篇 "krm əkóh:ŋh ǎxI KkqΣ2gÉ" 阅读笔记

星辰漫游者

更新：此前信陵指出X为Y 的函数，所以合并有些问题。经过考虑维持原推导不变，此处推导过程可以认为X和Y 无关说明之前的推导仍存在问题，感谢？信陵指出，具体请查看评论区。我再好好想H

水木八刀

语音识别的主要任务在于从语音到文本 其处理流程主要分为三个部分：特征提取：将语音信号转化为特征向量，代表有h NLL和Xl rh声学模型：衡量语音特征和文本之间的距离，判断该语音听起H

÷ h Xkmm 发表于光通信和深mm

题目这是该是第一个赛，比赛地EXl" nǐǎǎ个È6XĒ n除此之外，砍手豪

5 条评论

切换为时间排序

写下你的评论mm



乐笑怀

8 个月前

记得李宏毅深度学习课提过这个。对抗训练的变分自编码器，将语音的口音信息对抗训练到隐变量一些纬度，然后再语音转换的时候通过目标语音的编码隐变量纬度代入目标隐变量然后解码输出。



5