



季圣鹏

☎ 18943038195 ✉ shengpengji@zju.edu.cn
💻 <https://novateurjzp.github.io/>

教育经历

吉林大学 软件学院软件工程 本科 （本科综合排名：前8%） 2019.09-2023.06
浙江大学 软件学院软件工程 硕士 （导师：赵洲） 2023.09-2026.03

荣誉奖项

2021：国家奖学金（GPA：5/392）
2022：美国大学生数学建模竞赛Finalist奖（7%，20000+队伍）
2023：吉林大学一等奖学金（GPA：1/392），吉林大学优秀毕业生，EMNLP 2023/ACM MM 2024 Reviewer

论文

- 《TextrolSpeech: A Text Style Control Speech Corpus With Codec Language Text-to-Speech》 ICASSP 一作
- 《Boosting Prompting Mechanisms for Zero-Shot Speech Synthesis》 ICLR co-worker
- 《Generating Neural Networks for Diverse Networking Classification Tasks via Hardware-Aware Neural Architecture Search》 CCFA IEEE Transactions on Computers co-worker
- 《A Fast and High-Fidelity Framework for Mobile Zero-Shot Text-to-Speech》 ACL 一作在投(3.5/3/3)
- 《Mega-TTS: Zero-Shot Text-to-Speech at Scale with Intrinsic Inductive Bias》 arxiv co-worker

实习经历

- 2023.03-2023.07 **微软亚洲研究院自然语言计算组研究型算法实习生** 导师：刘树杰，周龙
工作内容是关于大语言模型，Zero-shot text to speech。主要研究基于离散codec的级联模型VALL-E鲁棒性问题。由于VALL-E特殊的结构设计，即利用AR模型自回归式生成第一层codec，利用NAR模型同时并行生成剩余的二到八层codec。VALL-E常常会出现多词，少词，重复的现象。我通过在AR模型中引入自回归duration信息，并且通过force attention机制，让acoustic token强制关注到对应的semantic token。最终将VALL-E的WER降低了15%。在MSRA，我很长一段时间处于实验不work的状态。例如在引入duration信息实验中，我设计了并行引入，embedding相加，自回归引入等各种方式来提升效果，这段经历极大地锻炼了我科学分析能力。VALL-E代码是基于fairseq框架的集群多卡分布式部署，这也很好地锻炼了相关工程能力，使我能够较为熟练地基于开源代码进行优化，并且复现一些常见的基于shell相关的论文代码。
- 2022.11-2023.03 **平安科技集团机器学习语音团队研究型算法实习生** 导师：张旭龙
工作内容是关于语音合成，歌声美化场景。动手实现了端到端语音合成baseline，并完成了语音合成相关的很多专利提案。在歌声美化任务上，保留业余歌唱者的音色，对其音高进行修正，并进行技巧，情感上的提升。对于音高部分，主要是想通过频谱包络和MIDI标准作为神经网络的输入，预测得到音高信息。并且通过音色解耦，提取出了歌曲内容和表现力（情感韵律）相关特征。对于生成部分，先将Mel频谱通过扩散模型的前向传播过程进行加噪，然后将修改后的条件等信息约束加入到隐变量z中，并且使其满足高斯分布，通过逆采样生成美化后的Mel频谱。将Mel频谱喂入声码器中即可得到美化后的语音。为了提高模型的推理速度，降低采样步数，我们通过直接预测干净Mel频谱的方式降低采样步数t。
- 2021.10-2022.04 **清华大学深圳研究生院智能网络组研究型实习生** 导师：江勇，李清
主要研究在网络分类场景下，结合多种硬件指标，基于神经网络架构搜索得到自适应的模型。允许用户输入数据集和优化指标（延迟、FLOPs）后自适应构建模型（卷机、池化、激活等）。最终成果为一篇CCFA期刊IEEE Transactions on Computers论文。我在其中的工作主要包括：参与设计NAS模型；编写基于DARTS的部分代码；提出了迭代式的神经网络算子搜索策略；基于正则化的思想，构造硬件表，将latency和FLOPs加入到损失函数中；同时复现了大量网络领域中的深度学习模型；负责论文中的大部分实验。这段科研经历让我开始接触科研，并因此入门深度学习，基于Pytorch框架对CNN神经网络，RNN神经网络，含自注意力机制的神经网络均有较好的理论基础和代码基础。

- 荣耀研发管理部智慧语音平台与浙江大学百万经费校企合作项目
- 初步研究成果已经转化成论文MobleSpeech，并且模型已经部署到荣耀[Magic6全系列](#)中，在语音助手中进行商用.
- 高校方**独立算法负责人**，给予荣耀团队算法模型支持，荣耀团队负责转化部署
- 工作内容主要关于移动端的个性化语音合成。传统的个性化语音合成，例如YourTTS模型，往往在推理阶段需要一定量的数据进行微调，这对用户不太友好并且耗时耗力。最近2023年基于大数据大语言模型方案，例如VALL-E，SpearTTS，VoiceBox，SoundStorm，NaturalSpeech2，MegaTTS，往往只需要几秒的语音prompt就能完全克隆出未见说话人的声音。但是这些方案都在云测进行部署，在推理速度，显存占用，模型参数量上并不满足移动端的要求，我们首个在移动端实现了个性化语音合成系统MobileSpeech。由于保密属性，在我完成该项目的过程中，从数据获取，数据处理，到模型复现对比，算法框架设计，实验迭代，产生了很多的问题，也最终被一一解决。最终该项目已经通过荣耀团队一阶段验收。

研究方向

- 实习时间：六个月（导师同意研二可继续实习）
- 实习地点：杭州，上海
- 研究方向：语音合成，语音大模型相关。对AIGC生成式模型/多模态相关/有很炫酷Demo的方向也很感兴趣。渴望在新的实习阶段学到一些新的东西，持续追求能做出**高引用论文**的工作，在此基础之上期待能将论文部署成实际算法产品。