

罗明双

✉ 739314837@qq.com · ☎ (+86) 18813052708 · 🌐 github.com/luomingshuang

🎓 教育背景

中国科学院大学, 研究生

2018.09 – 2021.07

专业: 计算机技术

导师: 山世光, 研究员

湘潭大学, 本科

2014.09 – 2018.06

专业 电子信息科学与技术

🔬 科研成果

基于双向同步学习的多语言唇语识别工作发表到 2020 BMVC

2020

第一作者 提出多语言的唇语识别模型, 利用双向同步学习获得上下文信息, 得到更强的解码模型

- Mingshuang Luo, Shuang Yang, Shiguang Shan, Xilin Chen. Synchronous Bidirectional Learning for Multilingual Lip Reading. **BMVC (CCF-C)** 2020, <https://arxiv.org/abs/2005.03846>.

基于伪卷积策略梯度的唇语识别工作发表到 2020 IEEE FG

2019

第一作者 提出利用伪卷积策略梯度的方法解决曝光偏差的问题, 并获得更强的语言模型

- Mingshuang Luo, Shuang Yang, Shiguang Shan, Xilin Chen. Pseudo-Convolutional Policy Gradient for Sequence-to-Sequence Lip-Reading. **IEEE FG (CCF-C)** 2020, <https://arxiv.org/abs/2003.03983>.

👤 工作经历

小米集团, 小米集团技术委员会人工智能部 AI 实验室语音组

2021.07 – 目前

正式员工, 语音算法工程师

- 在小米集团语音首席科学家 **Daniel Povey** 的领导下, 进行新一代 kaldi (k2) 的开发
 - 负责为本项目 k2 增加更多数据集上的 recipes
 - 探索相关的流式识别算法
 - 截止目前已在 k2, icefall, Lhotse 等新一代 kaldi 开源库贡献 14,000+ 行代码
- 参与 Audio-Visual Speech Separation/Recognition 的预研工作

👤 实习/项目经历 (在校期间)

华为, 车载环境下唇语识别项目

2019.10 – 2021.04

作为项目主要成员负责唇语识别模型的设计和优化

- 探索了脸部不同区域及大小对模型识别性能和速度的影响
- 设计了基于唇部动作频率的 VAD 检测方法
- 分别设计并对比了基于 GRU 和 Transformer 的解码模型的识别性能, 速度和模型大小
- 模型最终能够获得 12% 的字错误率的识别性能, 每句的识别时间低于 300ms, 模型大小低于 200M

华为, 流式唇语识别项目

2020.11 – 2021.04

作为项目主要成员负责流式唇语识别模型的设计和优化

- 分别探索了双向、单向的 RNN-Transducer 和 Transformer-Transducer 的流式模型方案
- 设计了基于单向注意力和注意力掩码矩阵的 Transformer 流式唇语模型设计
- 对视觉前端分别进行 Conv2D 和 Conv3D 的方式进行特征提取
- 获得了 14.37 的字错误率、0.252 的实时率和 10ms 的单句识别时延的性能指标

⚙️ 个人技能

- Programming Languages: 熟悉 python, 了解 c/c++
- Framework: 熟悉 Pytorch 深度学习框架
- Research interest: 声纹识别, 唇语识别, 语音识别, 语音转换, 语音合成, 多模态融合, 声源定位, 自监督/无监督学习
- Professional Skills: Seq2Seq 模型设计 (GRU,Transformer), 多模态协同学习

🏆 获奖情况

- 获得 2018-2019 年度中国科学院大学”三好学生”称号
- 获得 2018-2019 年度中国科学院大学”优秀党员”称号
- 2017 年全国大学生电子设计竞赛湖南省二等奖
- 2016 年湘潭大学电子设计竞赛一等奖
- 2017 年国家励志奖学金
- 2014-2018 多次获得湘潭大学奖学金, 被评为湘潭市”优秀毕业生”