

于灵云

个人自述

我的研究兴趣包括 Talking face generation, Multi-modal learning, Articulatory movements-driven 3D Talking Head, Human-Computer Interaction 和 Video synthesis。

现主要的研究项目是给定任意一段音频或文本，生成同步的，逼真的人脸动画。

联系方式

电话: +86-13966763354

Email: yuly@mail.ustc.edu.cn

个人主页: <http://home.ustc.edu.cn/~yuly>

GitHub: <https://github.com/xiaoyun4>

教育背景

09/2015-现在	控制科学与工程（自动化系） 中国科学技术大学 (USTC)	硕博连读
09/2011-06/2015	电气工程与自动化 中国矿业大学 (CUMT)	本科(前 5%，获保研资格免试入学 USTC)

项目经历

10/2017-Present	基于语音/文本的高自然度的语音动画生成（虚拟主播） 独立完成人 <ul style="list-style-type: none"> 给定一段语音/文本，生成高自然度说话人的语音动画。我们将此任务分为两个部分:唇部关键点预测以及基于关键点，进行视频生成。 使用多模态输入，利用 time-delayed LSTM 进行唇部关键点预测。 对于视频生成，在预测唇部关键点基础上，提出了 Face2vid 网络结构。在 Face2vid 中，引入 optical flow 以及 self-attention 机制实现视频生成中的时间/空间依赖性。 以第一作者将项目成果发表于 ICDM2019（已接收，Regular paper, 9.08% acceptance rate）和 IEEE TCSVT (major revision)。
05/2016-09/2017	基于深度学习文本/语音的发音器官运动轨迹估计（3D 人脸动画） 独立完成人 <ul style="list-style-type: none"> 我们提出了一种新的网络架构 BLTRCNN, 用于通过文本和音频输入进行关节运动预测。 采用 bottleneck network 提取瓶颈特征来作为韵律特征的补充，以获得更好的性能。 将 CNN, LSTM 和 skip connection 融合不仅可以学到局部 higher-level features 还可以学习大量的上下文信息。 将声学特征和韵律特征结合作为输入，能够包含更多语言学信息以提高性能。 以第一作者将项目成果发表于 IEEE TMM 期刊和 MMM2019, VCIP2018 国际会议。
09/2017-11/2017	基于 3D 模型，Trump 协同发音规律统计 独立完成人 <ul style="list-style-type: none"> 此任务主要应用在 Trump 3D 人脸动画合成。 文本和语音信息作为输入，旨在利用 mouth shape 来研究相邻音素间的协同发音规则。 采用 time-delayed LSTM 实现韵律特征、语音特征与 mouth shape 的映射关系。 利用统计学的知识，统计出三音素间的协同发音规律。
02/2015-02/2016	基于隐马尔科夫（HMM）的发音器官运动轨迹估计 独立完成人 <ul style="list-style-type: none"> 此任务主要是实现给定一段文本提取上下文相关的韵律特征，利用传统方法 HMM 实现发音器官运动轨迹估计，并研究不同情感下运动规律。 以第一作者将项目成果发表于 Multimedia Tools and Applications 期刊和 CCPR2016 会议。
07/2019-04/2019	参加 ACM MM2019 举办的“AI Meets Beauty”图像识别挑战赛 合作完成人 <ul style="list-style-type: none"> 此比赛旨在为参与者提供 50 多万张美容及个人护理产品图像的大规模图像数据集，即

Perfect-500K 数据集的基础上, 实现细粒度下美容及个人护理产品的图像识别。为了解决任务中的背景, 位置, 光照以及包装的多样性的干扰, 我们提出了 **Generalized-attention Regional Maximal Activation of Convolutions (GR-MAC)** 有效地提取特征, 从而提高了检索性能。此外, 我们也使用 Densenet201 和 SEResnet152 进行特征提取并特征融合。

- 引入 **attention** 机制, 自适应的将目标区域赋予更大权重。
- 团队在此比赛中荣获**第一名**的成绩且项目成果发表于 ACM MM 2019。

实习经历

07/2017-01/2019 科大讯飞股份有限公司

核心技术研发员岗位

- 负责给定一段语音, 生成唇音同步的, 高真实的视频动画。项目已提交

论文发表

- **Lingyun Yu et al.** Mining Audio, Text and Visual information for Talking face generation **IEEE ICDM 2019 (Regular paper, 9.08% acceptance rate)**
- **Lingyun Yu, Jun Yu, and Qiang Ling.** 2018. BLTRCNN Based 3D Articulatory Movement Prediction: Learning Articulatory Synchronicity From Both Text and Audio Inputs. **IEEE Transactions on Multimedia (IF:5.452)** .
- **Lingyun Yu et al.** Multimodal Inputs Driven Talking Face Generation With Spatial-Temporal Dependency (**IEEE Transactions on Circuits and Systems for Video Technology**) (major revision **IF: 4.046**)
- **Lingyun Yu, Yu J, Wang Z.** A realistic 3D articulatory animation system for emotional visual pronunciation[J]. **Multimedia Tools and Applications**, 2017, 76(18): 19241-19262 (**IF:2.101**).
- **Lingyun Yu, Yu J, Ling Q.** Deep Neural Network Based 3D Articulatory Movement Prediction Using Both Text and Audio Inputs[C]//International Conference on Multimedia Modeling. Springer, Cham, 2019: 68-79.
- **Lingyun Yu, Yu J, Ling Q** Synthesizing 3D Acoustic-Articulatory Mapping Trajectories: Predicting Articulatory Movements by Long-Term Recurrent Convolutional Neural Network. **IEEE VICP (2018)**
- **Lingyun Yu, Changwei Luo, and Jun Yu.** "An Emotional Text-Driven 3D Visual Pronunciation System for Mandarin Chinese." Chinese Conference on Pattern Recognition. Springer, Singapore, 2016.
- **Yu J, Lingyun Yu.** Synthesizing Photo-Realistic 3D Talking Head: Learning Lip Synchronicity and Emotion from Audio and Video[C]//2018 25th IEEE International Conference on Image Processing. IEEE, 2018: 1448-1452.
- **Jun Yu et al., Lingyun Yu,** Beauty Pro duct Retrieval Based on Regional Maximum Activation of Convolutions with Generalized Attention. **ACM MM 2019 (Accepted CCF A)**
- **Yu J, Luo C, Lingyun Yu, et al.** Facial video coding/decoding at ultra-low bit-rate: a 2D/3D model-based approach[J]. **Multimedia Tools and Applications, 2016: 1-21. (IF: 2.101)**
- **Luo C, Li R, Lingyun Yu, et al.** Automatic Tongue Tracking in X-Ray Images[J]. Chinese Journal of Electronics, 2015, 24(4): 767-771. (**IF: 0.945**)
- **Chuanbin Liu, Hongtao Xie, Zhengjun Zha, Lingyun Yu, Zhineng Chen, and Yongdong Zhang.** Bidirectional Attention-Recognition Model for Fine-grained Object Classification. **IEEE Transactions on Multimedia. (Accepted IF: 5.452)**

项目参与情况

- 国家自然科学基金面上项目 (61572450): 能说会唱的三维虚拟人头: 语音、歌声和情感语义同步的人脸及发音器官可视化研究, 2016/01-2019/12, 80 万。
- 国家自然科学基金-通用技术基础研究联合基金项目 (U1736123): 海量数据小训练样本集环境下的 2D+3D 人脸表情融合识别研究, 2018/01-2020/12, 82 万。
- 国家自然科学基金青年科学基金项目 (61303150): 语音同步的高真实感三维人脸情感动画研究, 2014/01-2016/12, 25 万。
- 华为创新研究计划-旗舰计划项目 (ES2100100098): 基于语音驱动的人物头部和身体模型驱动研究项目,

2017/09-2018/12, 120 万。

- 科大讯飞股份有限公司技术研发项目 (ES2100100099): 唇形建模与视频口唇恢复系统一期, 2017/07-2018/01, 20 万。

获奖情况

2019/07	在 ACM Multimedia “AI Meets Beauty” 比赛中, 团队荣获第一名
2017/09	中国科学技术大学 “国家奖学金”
2017/11	Travel Award, 2018 IEEE VCIP
2015/05	中国矿业大学优秀毕业生
2014/10	获得保研资格免试入学中国科学技术大学
2015~2018	历年中国科学技术大学 “一等奖学金”
2011~2014	历年中国矿业大学“校级奖学金”或“院级奖学金”

知识技能

- 数理基础:** 了解常见机器学习算法, 如回归、聚类等, 掌握深度学习理论及常见模型, 如 CNN, RNN, GAN
- 英语水平:** CET-6/CET-4; 扎实的英语听读写能力, 英文文献阅读写作
- 计算机水平:** 掌握 Python、MATLAB 和 OpenCV, 能实现基本的机器学习算法
掌握 Caffe, Pytorch, Keras 以及 Tensorflow 深度学习框架