

黄茂森

邮箱: mhuangaz@connect.ust.hk

电话: (+852)93011609; (+86)19821232902

个人优势

- 1, 在学校里面系统学习了深度学习的理论知识, 包含机器学习基础, 计算机视觉和自然语言处理。自己在网上也寻找感兴趣的课程进行学习, 例如推荐系统 (基于小红书进行讲解)。同时参与了多段深度学习相关的课题, 在课题中锻炼了文献调研, 快速复现以及改进算法的能力。
- 2, 英语能力过关, 在学校里和外国人组队能流畅沟通, 并且能用全英文进行学术汇报;
- 3, 对大数据处理, 例如 pyspark 有一定的了解, 还在持续学习中; 擅长快速编写一些小程序提高效率, 例如利用语音识别(基于 whisper)和 chatGPT, 编写了一个小程序, 能总结上课的内容。
- 4, 本科毕业于软件工程专业, 编码有良好的规范; 在本科参加过 ACM 集训, 组队进行了几次比赛, 锻炼了思维能力。

教育经历

香港科技大学

Sep 2023 - present

硕士 大数据工程与技术

核心课程: 机器学习、Spark 技术、自然语言处理、推荐算法

同济大学

Sep 2019 – Jul 2023

本科 软件工程

核心课程: 软件工程、C++、数据结构与算法、机器学习、数据库

项目经历

数字人生成

Sep 2023 – Dec 2023

1. 现有的数字人技术, 例如 Wav2Lip, 以及 DiNet, 在口型的准确度上存在缺陷, 我们认为这是由于音频提取器不能很好地提取音频特征; 经过调研, 我们发现 HuBERT 和 ContentVec 或许能很好地提取音频特征;
2. 我们基于 Wav2Lip, 将音频提取器从 Spectrogram 更换成了 HuBERT 和 ContentVec, 将整个流程中涉及的多个网络均进行了重新训练。
3. 最后惊喜地发现, 在英语语料中训练出来的网络, 能够泛化到其他语言上 (测试了中, 日, 韩语), 同时不受背景杂音的干扰。这就意味着可以收集全网各种语言的数据, 模型能从中学习到人类发音的规律。
4. 项目链接: <https://github.com/Crestina2001/Talking-head>

领域泛化

Oct 2022 – Jul 2023

1. 主要研究在深度学习中, 如果测试集与训练集不符合独立同分布假设, 应该怎么办? 举个例子, 在图像识别任务中, 训练集是照片, 那训练出来的算法能否识别卡通图片呢?
2. 通过在多个数据集上测试, 证明平均模型权重能带来稳定的性能提升。
3. 使用 Stable Diffusion 进行数据增强, 带来了一定的性能提升;
4. 对比了 ViT 和 CNN, 发现 ViT 能更好地提取图像的全局信息, 而 CNN 更擅长提取局部特征, 举例来说, CNN 对于素描画的识别比 ViT 更好, 因为局部特征对于识别素描画很重要; 同时测试了 CLIP 的零样本泛化性能, 发现其在常见的数据分布(不涉及小物体, 多个物体)上有良好的表现。

车道线识别

Mar 2022 – Jul 2022

- 1, 基于 Ultra-Fast Lane Detection, 在 Tusimple 和 CULane 两个数据集上训练, 调优, 观察模型表现。
- 2, 最后模型能在网上下载的长达一小时的视频上(非测试集数据)有较好的表现。