徐睿卓

ruizhuoxu@163.com · (+86) 13250921863 · 2000.08.24 · 浙江湖州

教育背景

北京邮电大学, 人工智能学院，模式识别实验室 2022.09 – 至今

在读硕士研究生, 2025.06 毕业, 导师：邓伟洪教授(引用量 1.2 万 +)

一等学业奖学金, 院级优秀研究生, *1 篇 SCI 一区论文第一作者*

浙江工业大学, 信息工程学院，通信工程 2018.09 – 2022.06

学士, 绩点排名:*1 / 115*, 获得推荐免试攻读研究生资格, 省级优秀毕业生

省政府奖学金, 校级优秀学生, 全国大学生智能汽车竞赛-双车接力组 (全国一等奖)

实习/项目经历

字节跳动 - 生活服务算法 - GenAI - 算法实习生 2024 年 07 月 – 至今

业务方向: 生活服务领域多模态内容理解

- 负责多个多模态分类模型迭代，熟悉从数据构建、清洗到模型训练、调优整个完整的工程链路。针对所使用的多模态基座模型，采用一系列训练技巧，提升了任务指标；
- 在不同任务上，对 LLM 进行 SFT，探索了不同 Prompt 对训练的影响；针对不同任务，使用不同的小模型和大模型结合方案，在保证推理 QPS 的情况下，有效改善了任务指标；
- 针对使用的多模态模型基座（小模型），使用生服领域数据构建正负样本进行对比学习，生产高质量生服视频多模态表征。利用 Mask 机制进行训练效率优化，在相同的计算资源下，将训练速度提升大约 4 倍，并改善了表征质量；

字节跳动 - TikTok 短视频 - Data-Rec Core(基础推荐) - 算法实习生 2024 年 02 月 – 7 月

业务方向: 基于多模态技术的 TikTok 短视频内容理解 (三层级多元化标签分类)

- 利用多模态技术 (Video Frames, ASR, OCR, Title et al.) 为每日新投稿短视频打上多元化标签, 供下游业务调用 (推荐、搜索 et al.), 节省了大量标注人力;
- 将多模态框架从 CNN 转移到 Transformer, 使用学习率逐层衰减稳定训练, 实现了模态结构的统一, 便于后续模型扩展和数据扩增;
- 针对噪声标签问题, 利用教师模型的输出计算样本不确定性分数, 对蒸馏损失进行加权, 有效提升了模型平均召回率 (AR65%@P80->AR67%@P80)
- 利用训练好的基线模型根据阈值划分从无标签数据生产高质量伪标签, 将训练数据量从 5000 万扩充至 1 亿 2 千万, 大大提高了基线指标 (AR67%@P80->AR70%@P80);
- 针对所使用的多模态分类模型结构, 设计了一套多模态自监督预训练方法 (MIM + MLM + CLIP Alignment), 在大规模有标签数据量的场景下, 进一步提高了模型指标 (AR70%P80->AR71%@P80);

奇虎 360 - 北京 - 深度学习应用部 - 算法实习生 2021 年 12 月 – 2022 年 05 月

研究方向: 基于对比学习的语音模态和文本模态对齐

- 提出了一个新颖的端到端模型，将语音模态和文本模态相结合用于口语语言理解 (SLU) 任务；
- 利用 CIF 机制，实现了语音特征序列和文本特征序列长度的对齐；利用 InfoNCE 损失，通过将语音特征序列和文本特征序列对应位置的 token 作为正样本对，其它位置作为负样本对，进行对比学习，实现了两者特征空间的对齐，在保留语音特有信息的同时充分发挥了语言模型的强大能力；
- 相较于其他两阶段的方法或只使用语音模型的方法，本方法在口语语言理解数据集的情感分析子集上取得了 1.15% 的召回率和 0.82% 的 F1 分数提升；

中国移动研究院 - 北京 - 合作项目 2022 年 09 月 – 2023 年 12 月

研究方向: 基于深度相机的人脸和人体分析

- 针对消费级深度相机采集得到的低质量人脸深度图像，实现了一个完整的数据预处理和数据增强管道，有效地降低了数据噪声的干扰并扩增了训练数据量，改善了基于深度图像的人脸识别性能;
- 利用隐式神经表示技术，提出了一个新颖的深度人脸图像去噪网络，将空间坐标信息作为深度人脸去噪和细化的先验，有效地改善了深度人脸图像的质量并提高了人脸识别准确率;
- 提出了一个轻量级的分组卷积融合模块，实现了深度图模态和法线图模态在特征层面的有效融合，有利于模型对人脸形状和姿态的感知，进一步提高了人脸识别准确率。
- 相较于之前最先进的工作，本项目所提方法在深度人脸图像的去噪和细化指标 PSNR、SSIM 以及 RMSE 上分别提升了 0.56db，0.96 和 0.117，总体人脸识别率提升了 4.27%。

论文

- **Depth Map Denoising Network and Lightweight Fusion Network for Enhanced 3D Face Recognition.** (Pattern Recognition 2024 - 第 1 作者 -> 模式识别顶级期刊)

  ***Ruizhuo Xu**, Ke Wang, Chao Deng, Mei Wang Junlan Feng, Weihong Deng et al.*

  三维人脸识别、深度图去噪、隐式神经表示

  - 首次将隐式神经表示技术引入深度人脸图像去噪领域，利用空间坐标信息指导深度人脸去噪;引入位置编码，并提出了一个多尺度解码融合策略，有效地提升了深度人脸去噪的性能;
  - 提出了一个轻量级三维人脸识别网络，通过一个多分支卷积融合模块实现了深度图模态和法线图模态的深度融合，在提高三维人脸表征质量的同时，平衡了计算开销;
  - 利用提出的深度人脸去噪网络和人脸识别网络，在多个三维人脸数据集上均取得了 SOTA 的去噪和识别性能;

- **Skeleton2vec: A Self-supervised Learning Framework with Contextualized Target Representations for Skeleton Sequence** (准备投稿 - 第 1 作者)

  ***Ruizhuo Xu**, Linzhi Huang, Mei Wang, Jiani Hu, Weihong Deng*

  自监督预训练、掩码预测、基于骨架的行为识别

  - 之前基于掩码预测的骨架序列预训练工作往往采用局部的、低层次的预测目标 (如：原始关节点)，这是次优的;为此，我们提出了 Skeleton2vec 框架，利用 EMA 更新的教师编码器生成全局上下文化的高层次特征表示作为预测目标，迫使编码器学习到的表征具有更强的时空联系。
  - 针对骨架序列具有较高时空关联性可能会产生的信息泄露问题，我们提出了基于运动感知的管道遮蔽策略 (Motion-aware Tube Masking)，迫使模型建模更好的长程时空联系，并持续关注运动语义丰富的区域;
  - 在三个大规模 3D 骨架行为识别数据集上的多个测试协议下，均取得了 SOTA 的性能;

- **WaBERT: A Low-resource End-to-end Model for Spoken Language Understanding and Speech-to-BERT Alignment** (共同 1 作)

  *Lin Yao, Jianfei Song, **Ruizhuo Xu**, Yingfang Yang, Zijian Chen, Yafeng Deng*

  模态对齐、对比学习、口语语言理解

  - 口语语言理解任务主要有两种主流方法：(1) 两阶段法：首先将语音通过语音识别模型转成文本，作为语言模型的输入，微调语言模型做下游任务;(2) 端到端法：直接微调预训练好的语音模型做下游任务;前者会丢失语音特有的信息并易受语音识别错误的影响;后者缺乏语言模型强大的语言理解能力;为此，我们提出一个新的端到端方案，结合语音模型和语言模型用于口语语言理解任务;在不丢失语音特有信息的同时，充分发挥语言模型的能力;
  - 利用 CIF 机制实现语音模态和文本模态特征序列长度的对齐，利用 InfoNCE Loss 对齐语音模态和文本模态的特征空间;推理时，语音输入经过语音编码器提取声学特征，提取得到的特征再作为语言编码器的输入，输出下游任务的结果;

⚙ 专业技能

- 编程能力: 了解 Python、C/C++、Linux、Latex、Git、Vim;
- 深度学习: 熟悉使用 Pytorch 开发深度学习模型，在算法改进和工程开发方面有相关实操经验;
- 个人证书: 英语（CET-4、CET-6）;

♡ 主要荣誉

- 第十六届全国大学生智能汽车竞赛，全国一等奖 *(4 / 225)* 2021.08
- 第十五届全国大学生智能汽车竞赛，浙江省三等奖 2020.08
- 省级优秀毕业生 2022.06
- 院级优秀研究生 2023.09
- 省政府奖学金 & 校级优秀学生 2019.09 / 2020.09 / 2021.09