

樊思明

个人主页 (siming.fun)

(+86)15818178828 ◇ gzfansiming@gmail.com

教育背景

电子科技大学 (UESTC)

2017 - 2021

信息与计算科学学士 (计算机科学方向)

四川, 中国

数学科学学院

加权平均分: 91.225(排名:5/37)

工作经历

商汤科技 移动智能事业部/研究院

2020-now

从单张图片的视角合成

2020.8-2021.2

- 视角合成/ 超分/ 去模糊/ 深度预测模型效果测试, GUI 编写
- 以 synsin 为 baseline, 增加 Partial Convolution, Skip-Connection, 修改 Rasterize 和 Render 方式来提升 ViewSynthesis 图像质量

从单张图片的流体动画、模拟 (SLR-SFS)

2021.3-2022.6

- 复现 Holynski 和 Mahapatra 的方法
- 基于 taichi 的流体模拟 2D/3D/Mesh 版本实现, 应用于单张图片输入上的流体流速编辑
- 流体和背景的分层表示

带次级运动的 Mesh 动画

2022.6-2022.11

- 用 Vega FEM(参考 zhengmianlun 的 DeepEmulator 数据生成方法) 在 SMPL 模型上生成带有次级运动动画的数据集
- 结合 vto/snug/snarf 的方法作为 baseline, 训练 vto dataset, 效果略好于复现的 vto 和 snug

人体人脸数据集处理和 Benchmark

2022.11-2023.7

- 在DNA-Rendering数据集中负责跑 Novel-ID, Cross-dataset Benchmark, 网页搭建
- 在RenderMe360数据集中负责部分数据处理工作, 网页搭建

视频精彩帧检测

2023-now

- BestMoment 智能抓拍项目 2023.5-2024.4
 - 预研阶段: 调研关键点匹配 (PoseMatching) 方案, baseline 搭建, 增加数据, Gradio Demo。
 - 交付阶段: 基于关键点规则初筛大规模公开数据集中符合定义的精彩帧, 以便低成本构建大规模训练数据 (十万级训练视频), 以及上线小模型的 MultiTask 训练 (增加关键点预测)。
- AAI2024: ShotVL(在投, 暂未开源) 2023.12-2024.8
 - BestShot Benchmark 构建: 对现有时序定位数据集 (TH14, FineAction) 筛选精彩帧, 重新标注用细粒度文本描述, 建立任意 Query 定位视频帧的 Benchmark。
 - GPT4o 训练数据构建: 利用 GPT4o 对现有图片生成细粒度描述和 Image QA, 对视频生成包含定位类问题的 Video QA, 总计生成 600K ImageQA 和 120K VideoQA。其中 VideoQA 的生成 pipeline 是首个用 GPT4o 生成视频内帧定位/区间定位类 QA 的标注生成策略。

- 合成姿态描述训练数据构建：用 PoseScript 跑通 13 个公开数据集的姿态描述，总计生成 18.6M 个，在 PoseScript 基础上增加人物朝向描述。
- baseline 训练：以 InternVL14B 和 LITA(Clip+Vicuna13B) 为 baseline 训练以上数据，在多个 benchmark(BestShot/Th14/AVA Benchmark) 的 zero-shot 结果上取得 SOTA，其中动作类检索 Query 比 InternVL 平均提高约 10%，姿态类 Query 检索比 InternVL 平均提高约 61%，并保持了与 InternVL 一致的通用检索能力。
- 后续拓展工作 (doing, 2024.09 now): 1. ShotVL2: Locate any frame in video by any fine-grained text , 2.AnyDiffEdit: Any image editing by any fine-grained text

猫狗 Re-ID/聚类项目

2024.08-2024.09

- 负责百万级 ID 爬取和数据处理、聚类算法 (Incremental DBScan 为 Baseline) 的策略优化工作。
- 部署模型 (mobilenet) 下已达到 ios 指标 (f1 score) 的 117%，远超交付标准。

论文

- ShotVL: Human-centric Highlight Frame Retrieval via Language Queries.
Wangyu Xue*, Chen Qian*, Jiayi Wu, Yang Zhou, Wentao Liu, Ju Ren, Siming Fan[†], Yaoxue Zhang[✉].
(AAAI2024 在投)
- DNA-Rendering: A Diverse Neural Actor Repository for High-Fidelity Human-centric Rendering.
Wei Cheng*, Ruixiang Chen*, Wanqi Yin*, Siming Fan*, Keyu Chen*, Honglin He*, Huiwen Luo, Zhongang Cai, Jingbo Wang, Yang Gao, Zhengming Yu, Zhengyu Lin, Daxuan Ren, Lei Yang, Ziwei Liu, Chen Change Loy, Chen Qian, Wayne Wu, Dahua Lin, Bo Dai, Kwan-Yee Lin. (ICCV2023)
- RenderMe-360: A Large Digital Asset Library and Benchmarks Towards High-fidelity Head Avatars.
Pan, Dongwei and Zhuo, Long and Piao, Jingtian and Luo, Huiwen and Cheng, Wei and Wang, Yuxin and Fan, Siming and Liu, Shengqi and Yang, Lei and Dai, Bo and Liu, Ziwei and Loy, Chen Change and Qian, Chen and Wu, Wayne and Lin, Dahua and Lin, Kwan-Yee. (NeurIPS 2023)
- Simulating Fluids in Real-World Still Images.
Siming Fan*, Jingtian Piao*, Chen Qian, Hongsheng Li, Kwan-Yee Lin[†](ICCV2023 Oral)

†:Project Lead *:Co-first ✉:corresponding

在校期间

LIDAR/RGB-LIDAR 3D 目标检测

2019

- pytorch 复现了 Pointnet, 包括未开源的预处理和可视化分析.(blog) 和 frustum-pointnets(不含 RGB 检测器)(github)。
- 基于 frustum-convnet 实现了 DenseFusion 的特征融合方法, 准确率在 kitti 验证集 (caronly,Moderate) 上从 85.52 提升到 86.44。(github)。
- 创建 kitti2nuScenes 无监督领域自适应评估数据集, 将 Domain Adaptive Faster RCNN 应用于 PointRCNN, 准确率 (caronly, Easy) 从 9.78 到 10.98.(github)

其他奖项

2018 电子科技大学程序设计挑战赛决赛校级三等奖

2018 电子科技大学第十一届 ACM 趣味程序设计竞赛三等奖

2019 电子科技大学第十一届 ACM 趣味程序设计竞赛三等奖 (rank 36/360)

2018 电子科技大学数学建模新生赛三等奖

2019 电子科技大学数学建模校赛大二及以上组校级三等奖 (rank 26-50/177)

奖学金

2018 优秀学生奖学金 (参与者的 10%)

2019 标兵学生奖学金 (参与者的 25%)

2020 标兵学生奖学金 (参与者的 25%)