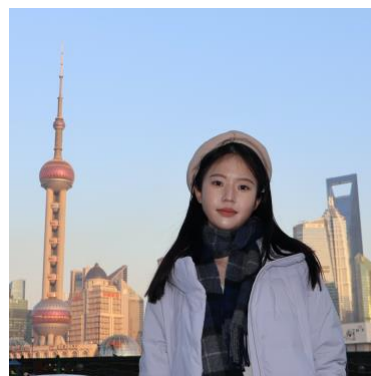


黄艾琳 Ailin Huang

武汉大学 计算机科学与技术专业 本科 2019 – 2023

北京大学 国家发展研究院 经济学辅修学位 2024 – 2026

p2oileen@whu.edu.cn | huangailin@stu.pku.edu.cn | [谷歌学术](#) | [GitHub](#) | [知乎](#)

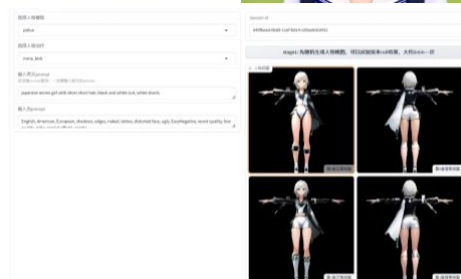
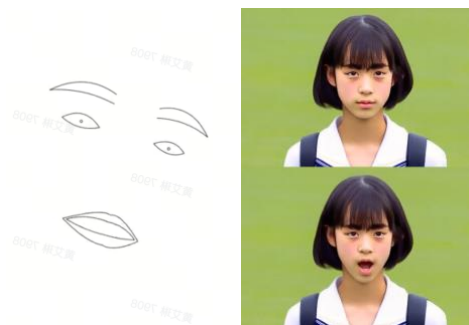


工作经历

- 2025.01 ~ 2025.05 ✨ 阶跃星辰 AIGC 组 | 算法研究员
负责 CRef 相关能力迭代，开发 Story Visualization 评测 benchmark，文章于 NeurIPS 在投（共一，Project Lead）。
- 2023.11 ~ 2025.1 ✨ 阶跃星辰 Agent 组 | 算法研究员
多模模型探索，LLM 数据处理、SFT/RL 训练、语音大模型训练等工作。
开源 **130B Step-Audio** 语音大模型和**技术报告**，负责 Rap/歌声功能迭代。
- 2023.07 ~ 2023.11 🧠 旷视研究院 AIC 组 | 助理研究员
转正后负责大语言模型的有监督微调工作。图像处理方面发表一篇 **WACV24**（二作）。
- 2021.01 ~ 2023.06 🧠 旷视研究院 AIC 组 | 算法研究员（实习） | Mentor: [周舒畅](#)
全职北京线下实习，研究领域：视频插帧，神经渲染，CLIP，3D 人体，Talking Head 等。发表一篇 **MM22**（一作），一篇 **IJCAI23**（共一），一篇 **CVPR23 Highlight**（三作）。详见[谷歌学术](#)或下方研究经历。
- 2021.03 ~ 2021.06 🏠 北航×旷视 深度学习课程 | 开发助教 | 教师: [王静远](#)
编写基于 MegEngine 框架的[代码和 tutorial](#)，包括蒸馏，对抗样本生成，攻击，ISP，RL 等。[线上课程](#)

项目经历

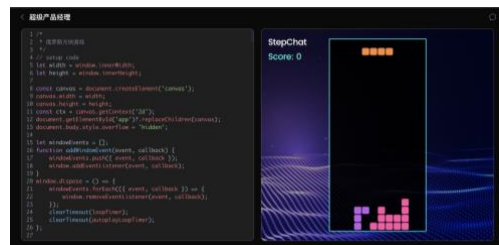
- 2025.03 基于 ComfyUI 的 VR 场景 workflow 实现
通过带掩码的图像进行带 CRef 能力的真实场景中的二次元人物生成。
- 2025.01 基于 Flux 模型的文生图大模型 CRef 能力迭代
在 2B Flux 模型 Megatron pipeline 上复现 PuLID 算法，用于人脸 ID 输入以生成符合人物 ID 的图像。
- 2024.10 音乐生成/干声哼唱/RAP 任务：
1. 音乐 midi 生成模型初步实现，并探索了音乐转类乐谱的降维表示方法，如 chroma、midi 获取等。
2. 集中解决 text2rap/vocal 存在的问题，如无限重复、丢字、难听和风格缺失等。
3. 优化 text2rap/vocal 效果并上线，达到业内领先水平
- 2024.06 人脸表情识别、视频预测与 Talking Head 合成
1. 探索多种 token（如 4m21、AU、blendshape、landmark、SVG、live portrait embedding）作为人脸表示以降维，利用大模型做表情识别。相同数据用于预测类任务训练。
2. 数据收集，使用 SVG 表示方法降低人脸表示的维度，提高表情识别的效率。
3. 训练 LLM 学习从声音 VQtoken 到 live portrait 表示，实现从人声到脸部动作的驱动。
- 2024.04 人脸/human 3D 美术资产生成。实现以下结果：
1. 通过 Prompt 生成人物语言 Text
2. 人物语言 TTS
3. Speech to blendshapes，使用 blendshape 驱动人脸表情
4. 从 Prompt 和 3D 模型深度图获取高质量人物贴图。
5. 通过生成高质量贴图修改 3D 模型 UV，以达到模型染色目的
6. 以 Neural 方式达成更连续的染色效果。
- 2024.02 百万 RL 数据



清洗大量互联网爬取的数据，由标注员进行标注，制作为 DPO 训练可使用的数据。包括回流数据，GPT4 真实对话数据等。该 RL 数据被用于对于基础大模型的微调，以及 Reward Model 训练。

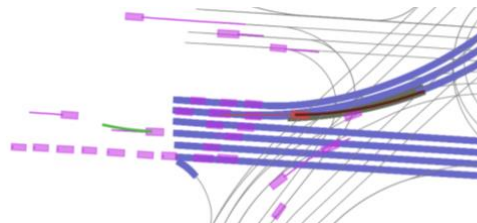
2024.01 参与公司内部大模型多轮对话测试 benchmark 的搭建。
手工设置了一系列多轮对话问题，并使用 GPT4 作为评判者进行评分和文字内容评价，自动生成离散评分和连续评分的报表。该多轮 bmk 在公司内部广泛使用，用于判断一些常见问题（跨轮重复，单条回复重复，坏分支，指令遵循能力）等的表现。

2024.01 上海人工智能展会 demo 项目：超级产品经理
用语音或文字交流，来修改游戏代码，并即时渲染出新的俄罗斯方块游戏。
支持模糊的 prompt 要求来改变游戏代码。
例如：我希望所有方块都变成红的，将该 prompt 映射到代码的修改，并渲染得到新游戏内容。

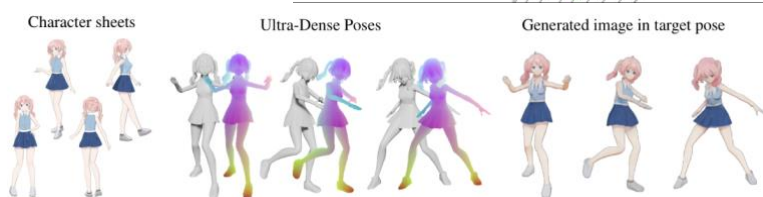


2023.11 参与冒泡鸭应用项目中人物 AI 的 DPO 训练流程。
负责从游戏人物中获取 DPO 数据和模型训练，使得模型表现和人物设定尽可能贴近。
使用几个基础模型的生成结果作为负例，以贴近真实数据的说话风格。在使用 DPO 数据训练后，人物的发言逐渐趋同于训练数据，具有生动性，且口癖和发言习惯类似。

2022.11 - 参加 2023 Nuplan 自动驾驶比赛，负责使用强化学习 PPO 等算法生成决策信息。
其中包括换道决策，最远距离决策等，通过设计并编写各类奖励函数，实现车辆前进，避障，稳定驶向终点等功能。[总榜第五名](#)
右图展示了一个 scenario 的行驶实例。



2022.06 旷视 2022 MegTech Open Day 展览：展出手绘图片转动画的 demo，该 demo 接收多张单人物的手绘图像，和一个动作序列，生成该人物的动画视频。
负责 demo 的制作。[GitHub Repo](#)



2021.12 使用 MegEngine 复现插帧 RIFE 算法，并优化 StyleCLIP / HyperStyle 等文字驱动的人脸编辑算法。
制作文字驱动的人脸编辑 pipeline，其中，RIFE 算法被用于将编辑前后的人脸进行融合，以增加编辑前后人脸的相似程度。
[RIFE 的复现 Repo](#) [人脸编辑的 Repo: Oh-My-Face](#)



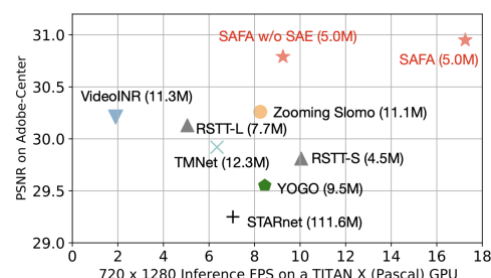
研究经历

- 2025 ViStoryBench: Comprehensive Benchmark Suite for Story Visualization [GitHub](#)
Cailin Zhuang[†], **Ailin Huang**[†], Wei Cheng, Jingwei Wu, Yaoqi Hu, Jiaqi Liao, Hongyuan Wang, Xinyao Liao, Weiwei Cai, Hengyuan Xu, Xuanyang Zhang, Xianfang Zeng, Zhewei Huang, Gang Yu, Chi Zhang Submitting to NeurIPS 2025 Datasets and Benchmarks Track
对于 Story Visualization 任务收集了类别广泛的剧本-图像对应数据集，并针对该领域独有的一些问题开发对应 benchmark，例如 cross/self 多人 ID 一致性、分镜风格一致性、Prompt 遵循程度等，并测试了 27 个开源/非开源工作给出结论，提供定性/定量分析结果证实 benchmark 的有效性。
作为共一/Project Lead，负责完成论文前期调研，提出论文指标主要方法，全部数据集收集和整理工作，部分论文写作工作，组织多人并参与 benchmark coding 开发工作。
- 2025 Step-Audio: Unified Understanding and Generation in Intelligent Speech Interaction [Paper](#) | [GitHub](#)
StepAudio Team
阶跃星辰 Agent 组工作，开源 130B 语音文本多模态大模型，支持精细控制，可跨方言、情绪进行动态调整，实现对话、歌唱、RAP 等功能。线上应用见跃问 APP。负责 RAP 和歌唱功能的实现。
- 2022 Scale-Adaptive Feature Aggregation for Efficient Space-Time Video Super-Resolution [Paper](#) | [GitHub](#)
Zhewei Huang, **Ailin Huang**, Xiaotao Hu, Shuchang Zhou WACV2024

提出一种针对尺度自适应的特征聚合方法，并生成光流对视频的时间和空间进行扩展：即同时进行视频插帧和超分辨率。

我们专注于构建基于光流的高级特征传播的关键组件，SAFA 可以自适应地调整处理后的特征尺寸以适应不同的场景，从而产生一个简单而高效的 STVSR 结构。经过实验测试，SAFA 在几个公共指标中实现了 SOTA。

负责部分实验工作和论文撰写。



2022 A Dynamic Multi-scale Voxel Flow Network for Video Prediction

Xiaotao Hu, Zhewei Huang, **Ailin Huang**, Shuchang Zhou

使用光流进行视频预测。我们提出了一个多尺度 Voxel-Flow 块，使我们的模型能够处理不同尺度的运动，以提高模型泛化性。实验证明，在推理时，DMVFN 可以根据输入自适应地选择合适的子网络，加快推理时间并保持模型性能。

负责部分实验工作和论文撰写。

[Paper](#) | [GitHub](#)
CVPR2023 Highlight



2022 Perceptual Conversational Head Generation with Regularized Driver and Enhanced Renderer

Ailin Huang[†], Zhewei Huang[†], Shuchang Zhou

ACM Multimedia* 2022 Conversational Head Generation Challenge 冠军

该比赛方案训练一个通用的音频-行为驱动，模拟一个高质量的渲染器，将行为表示转化为视频。

我们对 baseline 进行了精细的调整，并使用一个前后景融合模块将渲染结果进行后处理，探索多种集成学习方法。

负责绝大部分实验工作和论文撰写。

[Paper](#) | [GitHub](#)
ACM-MM2022 Main Track



荣誉奖励

- 2024 湖北省三湖连江青年学者称号 (Not real, just for fun)
- 2022 ACM MM2022 Conversational Head Generation Challenge | [Track1 银牌](#) [Track2 金牌](#)
- 2020 – 2021 ICPC2020 南京赛区银牌, ICPC2021 银川赛区银牌, CCPC2020 威海赛区银牌
- 2020 旷视人工智能开源大赛 初赛 [第一名](#) (未入职时参加)
- 2018 高中信息学奥林匹克竞赛 (NOIP) 省级一等奖

其他

- Part-time job ACM-ICPC 2021 昆明站出题人, 算法竞赛教学 100+ 课时
- 兴趣/生活 单板滑雪, 游泳, 桌游, 养一只布偶猫 🐱