Brandon Jinxiu Liu

Email : jinxiuliu0628@foxmail.com
Mobile : +86-13951891694

Homepage: https://brandon-liu-jx.github.io/

Education

- **South China University of Technology (SCUT), School of Future Technology** Guangdong, China
  B.Eng, Majoring in Artifical Intelligence, Senior Undergraduate *Aug 2021 - Jun 2025*
  *Main Courses: Deep Learning and Computer Vision **(4.0/4.0)**, Course Design of Deep Learning and Computer Vision **(4.0/4.0, Best project)**, C++ Programming Foundations **(4.0/4.0)**, Python Programming **(4.0/4.0)**, Data Structure **(4.0/4.0)**, Advanced Language Programming Training **(4.0/4.0)**, Artificial Intelligence and 3D Vision **(4.0/4.0)**, Calculus **(4.0/4.0)**, Optimization Method **(4.0/4.0)** .....*

Research Interests

Image/Video/3D Generation, Generative Multimodal Model
My long-term research goal is to develop an AI system capable of creating **multi-modal**, **controllable** (e.g., Maple, OpenStory++, OpenStory), **immersive** (e.g., DynamicScaler, Prompt-Image-to-Life), **structured** (e.g., R3CD), **interactive** (e.g., PiGIE), and **physically grounded** (e.g., ongoing projects) **2D, 3D, and 4D virtual worlds**. This vision harnesses the transformative power of **foundation models**—including LLMs, MLLMs, and diffusion models for images, videos, and 3D data—drawing inspiration from **human nature** and addressing the **real needs of designers and artists**.
I see my PhD journey as an **entrepreneurial venture** akin to founding a startup: guided by my **"North Star"**, focusing on long-term and important issues, fueled by **passion**, and dedicated to producing **impactful, practical research** that bridges **innovation from academia and utility from industry**.

Representative Projects

- DynamicScaler: Seamless and Scalable Video Generation for Panoramic Scenes
  Tech Report, The first and an unified framework for Arbitrary Size Text/Image to Rectangular/360 Degree Panoramic Video Generation
  Jinxiu Liu, Shaoheng Lin, Yinxiao Li, Ming-Hsuan Yang
- R3CD: Scene Graph to Image Generation with Relation-aware Compositional Contrastive Control Diffusion
  Accepted by AAAI 2024, Vancouver, Canada
  Jinxiu Liu, Qi Liu
- Openstory++: A Large-scale Dataset and Benchmark for Instance-aware Open-domain Visual Storytelling
  Zilyu Ye*, **Jinxiu Liu* ‡**, Ruotian Peng, Jinjin Cao, Zhiyang Chen, Ziwei Xuan, Mingyuan Zhou, Xiaoqian Shen , Mohamed Elhoseiny, Qi Liu, Guo-Jun Qi (*equal contribution, ‡Project Lead)
  Tech Report, Hugging Face Daily Papers
- OpenStory: A Large-Scale Open-Domain Dataset for Subject-Driven Visual Storytelling
  Accepted by CVPR 2024@VDU (Oral Presentation) Seattle, USA
  Zilyu Ye*, **Jinxiu Liu* ‡**, Zhiyang Chen, Ziwei Xuan, Mingyuan Zhou, Qi Liu, Guo-Jun Qi (*equal contribution, ‡Project Lead)
- PiGIE: Proximal Policy Optimization Guided Diffusion for Fine-Grained Image Editing (Extended Version)
  InstructRL4Pix: Training Diffusion for Image Editing by Reinforcement Learning
  Accepted by CVPR 2024@AI4CC, Seattle, USA
  Tiancheng Li*, **Jinxiu Liu* ‡**, Huajun Chen , Qi Liu (*equal contribution, ‡Project Lead)
- Maple: Multi-modal Pre-training for Contextual Instance-aware Visual Generation
  Tech Report, MLLM Pretraining for Interleaved Image-text Generation based on Openstory++
  Jinxiu Liu, Jinjin Cao, Zilyu Ye, Zhiyang Chen, Ziwei Xuan, Zemin Huang, Mingyuan Zhou, Qi Liu, Guo-Jun Qi
- Prompt Image to Life: Training-free Text-driven Image-to-video Generation
  Tech Report, Best Project in "MetaVerse and VR Course Project"
  Jinxiu Liu, Yuan Yao, Bingwen Zhu, Weijian Luo, Fanyi Wang, Yanhao Zhang, Yuxiao Wang, Qi Liu, Jiebo Luo, Guo-Jun Qi

- **Deep Neural Network Compression by Spatial-wise Low-rank Decomposition**

  Best Project (1/96) in "Optimization Method Course Project"

  Xiaoye Zhu*, **Jinxiu Liu***, Ye Liu, Michael Ngo, Zihan Ji (* equal contribution)

- PoseAnimate: Zero-shot High Fidelity Pose Controllable character animation

  Accepted by IJCAI 2024, Jeju, Korea

  Bingwen Zhu, Fanyi Wang, Peng Liu, Jingwen Su, **Jinxiu Liu**, Yanhao Zhang, Zuxuan Wu, Yu-Gang Jiang, Guo-Jun Qi

- FreeA: Human-object Interaction Detection using Free Annotation Labels

  Yuxiao Wang, Zhenao Wei, Xinyu Jiang, Yu Lei, Weiying Xue, **Jinxiu Liu**, Qi Liu

Experience

- ***Research Collaborator, Google DeepMind, supervised by Prof. Ming-Hsuan Yang and Dr. Yinxiao Li, Nov 2024 - Present***

  - I am exploring the integration of physics principles into video diffusion models, focusing on enhancing the ability of these models to simulate physical phenomena like gravity, motion, and collision dynamics. This project leverages in-context learning, enabling models to generate more realistic, physics-aware video content.
  - I propose a two-pronged approach to solving the challenge of physics simulation in generative models: (1) **Symbolic Control Powered by Physics Engines**—integrating a physics engine to precisely control physical parameters in the generative context, and (2) **In-Context Learning for Video Diffusion**—leveraging video examples to guide model generation and enhance realism through learned physical behaviors.
  - This research aims to bridge the gap between generative video diffusion models and physics simulations, enabling models to incorporate detailed physical parameters and produce dynamic, physically grounded content.

- ***Research Collaborator, Google DeepMind, supervised by Prof. Ming-Hsuan Yang and Dr. Yinxiao Li, Aug 2024 - Nov 2024***

  - I am developing DynamicScaler, a novel framework for generating scalable, 360° dynamic panoramas, addressing resolution and aspect ratio limitations in video generation. This is particularly impactful for AR/VR applications, where high-quality, dynamic scenes are essential.
  - DynamicScaler enables the creation of high-resolution dynamic panoramas with infinite scalability, transferring diffusion model knowledge across spatial dimensions to maintain spatial and temporal coherence.
  - The framework allows for the generation of large-scale, high-resolution dynamic content directly from text prompts, offering exceptional flexibility in motion, spatial scalability, and resolution for immersive AR/VR experiences.

- ***Research Intern, Stanford University, supervised by Prof. Jiajun Wu, June 2024 - Aug 2024***

  - I am exploring Novel View Synthesis using video diffusion models, focusing on generative spatial intelligence. My research aims to teach AI the ability to generate new perspectives of a scene, inspired by human spatial imagination—the ability to mentally reconstruct and visualize spaces not directly observed.
  - By modifying the denoising process in video diffusion models, I enable them to take a single 2D image as input and generate novel viewpoints of a scene, filling in occluded parts and creating new perspectives from depth map prediction. This process mimics how humans use prior knowledge to infer unseen visual details and anticipate how objects might appear from different angles.

- ***Research Intern, Westlake University & OPPO Research Institute, supervised by Prof. Guo-Jun Qi (IEEE Fellow), Sep 2023 - June 2024***

  - A framework is proposed for story video generation based on multimodal large language models, which can generate long stories and maintain the consistency of the main task across different scenes.
  - A Large-Scale Open-Domain Dataset for Subject-Driven Visual Storytelling is proposed and a paper is accepted by CVPR 2024@VDU as **Oral** presentation and the extended version Openstory++ has been featured as the **Huggingface Daily Paper**.

- ***Research Intern, SCUT, supervised by Prof. Qi Liu (IEEE Senior Member), Dec 2022 - Current***

- Developed a novel scene graph to image generation framework based on **diffusion models** and **contrastive control mechanisms.** And addressed the challenges of abstract relations in scene graphs and generated images that matched the scene graph specifications.
- A paper is accepted by AAAI 2024 (**First Author**)

- ***Course Design of Deep Learning , mentored by Prof. Mingkui Tan (IEEE Senior Member) and Prof. Huiping Zhuang***
  - Developed a text-based dialogue system called MiniHuggingGPT, a mini multi-modal application like HuggingGPT and MiniGPT-4, which can leverage three large-scale models for image captioning, image generation, and text conversation using natural language commands by instruction finetuning.
  - Provided a web-based interface based on gradio for easy interaction with the system and showcased various examples of its capabilities
  - Awarded as the ***Best Course Design (1/39).***

Honors and Awards

- The Taihu Oversea Scholarship (**ranked 1/500+ comprehensively**, 40000 RMB, Wuxi city governments)
- The Taihu Innovation Scholarship (**ranked 5/500+ comprehensively**, 10000 RMB, Wuxi city governments)
- TCL Corporate Scholarships (**ranked 1/25 comprehensively**, 20000 RMB, TCL Technology)
- SCUT Scholarship (**ranked 1/25 comprehensively**, 20000 RMB, SCUT)
- The Taihu Research Scholarship (**ranked 1/160 comprehensively**, 8000 RMB, Wuxi city governments)
- The Mathematical Contest in Modeling & The Interdisciplinary Contest in Modeling(MCM/ICM), **Meritorious Winner (Top 7% globally)**
- Asia and Pacific Mathematical Contest in Modeling (APMCM) – The Second Prize (Top 15% globally)