About Me

Kejun Gao

Academic Background

I am currently an undergraduate student in the **Department of Computer Science and Technology** at **Tsinghua University**, a member of the Class of 2023.

As of the end of my sophomore year, my academic standing is **GPA 3.90** (ranked **23/171** in the cohort). Notably, **all core Computer Science courses and fundamental Mathematics courses were completed with a 4.0**.

Prior to this, I completed my high school education in Jinan, Shandong Province, where I achieved a score of **704** on the National College Entrance Examination, ranking among the **Top 10** students in the province.

Research Experience

During my undergraduate studies, I have been actively involved in research internships under the guidance of **Prof. Song-Hai Zhang**, **Prof. Xiao-Lin Hu**, and **Prof. Hao Zhao**.

Key Achievements

- **SRT Project (Prof. Zhang):** My individual work on the Student Research Training (SRT) project under Prof. Zhang received the **A+ grade (Highest)**, demonstrating exceptional research output.
- **Research Grant (Prof. Hu):** During my internship with Prof. Hu, I successfully initiated and secured a **20,000 RMB** grant through the competitive university-level "Xuetui Plan" (Student Research Promotion Program).

Research Interests

Broadly speaking, my research centers on the **perception, understanding, and generation of visual and auditory modalities**. My ultimate goal is to construct a **highly realistic and interactive audiovisual world**, serving as a foundation to achieve **Spatial Intelligence** capable of robust perception, understanding, and reasoning.

Specifically, I focus (or plan to focus) on the following areas:

1. 3D Vision

- **3D Representation Learning:** Considering the current fragmented landscape of 3D representations (e.g., Voxels, NeRF, Gaussian Splatting), I aim to explore and define a **unified 3D representation paradigm**.
- **3D Scene Generation & Reconstruction:** Generating interactive and Physically Realistic 3D scenes.

2. Spatial Audio

(For a comprehensive overview of this field, refer to the survey: [ASAudio: A Survey of Advanced Spatial Audio Research](#))

- **Acoustic Field Reconstruction & Generation:** Inspired by works such as [NeRAF](#) and [AV-DAR](#), I plan to develop a **feedforward Audio-Visual Gaussian Splatting (AV-GS) framework**, which supports multiple sound sources and generalizes across scenes without requiring per-scene optimization.

- **Unified Spatial Audio Generation:** Developing a unified framework capable of generating spatial audio from diverse inputs, including text, egocentric video, 360° video, and audio.

- **Spatial Audio Perception & Reasoning:** Investigating how **Multimodal LLMs (MLLMs)** perform reasoning using spatial audio, and how Embodied AI frameworks (e.g., **Vision-Language-Action (VLA)** models) utilize spatial audio for decision-making.

3. Multimodal Large Language Models (MLLM)

- **Robustness in Audiovisual Contexts:** Benchmarking and enhancing MLLM robustness under challenging conditions, such as noisy environments, multi-speaker audio, or extremely low-quality visual inputs.

- **Enhancing Spatial Intelligence:** Improving performance on downstream spatial tasks, including but not limited to:

  - **Vision-and-Language Navigation (VLN)**

  - **3D Object Detection & Grounding**

  - **Spatial Question Answering (Spatial QA)**

- **MLLM Paradigm Research:** Critically examining current multimodal alignment paradigms (e.g., designing specific discrete tokenizers or continuous encoders for each modality)，I aim to investigate whether this is the path to true multimodal intelligence or merely an expedient, short-sighted solution for simplicity.

Publications / Works

- 《叙事工坊：交互式叙事场景构建》 **(Narrative Workshop: Interactive Narrative Scene Construction)** [Project Page](#)

  - **Authors:** Hanxi Zhu, **Kejun Gao**, et al.

  - **Venue:** Chinagraph 2024 **Best Paper Award**; Submitted to Chinese Journal of Computers (CCF-A), 2025.

  - **Highlights:** Proposed an LLM-based optimization strategy for narrative scene layout. Won the **Best Paper Award** at Chinagraph 2024.

- **ViewSeeker: Locating Camera via Monocular RGB Image with MaskXY Derivatives**

  - **Authors:** Hanxi Zhu, **Kejun Gao**, et al.

  - **Venue:** Under Review at IEEE Transactions on Visualization and Computer Graphics (TVCG, CCF-A).

  - **Highlights:** Designed MaskXY derivatives based on detection/segmentation outputs and proposed a differentiable rendering framework for target view navigation.

- **Efficient Audio-Visual Speech Separation with Discrete Lip Semantics and Multi-Scale Global-Local Attention** [Project Page](#)

- **Authors:** Kai Li*, **Kejun Gao***, et al.

- **Venue:** Under Review at ICLR 2026 (Scores Before Rebuttal: 6/6/6/4).

- **Highlights:** Proposed a lightweight multi-modal speech separation model using a "Semantic-Reconstruction" dual-path visual encoding framework. Surpasses SOTA on all metrics, achieving **>50%** parameter reduction, **>2.4×** lower MACs, and **>6×** faster inference (GitHub 159 stars).

- **LottieGPT: Tokenizing Vector Animation for Autoregressive Generation**

  - **Authors:** Junhao Chen*, **Kejun Gao***, et al.

  - **Venue:** Under Review at CVPR 2026.

  - **Highlights:** The first generation framework based on the Lottie vector animation format. Outperforms SOTA methods (e.g., OmniSVG) in single SVG generation and surpasses closed-source models (e.g., Sora2, Kling) in vector video generation.

Awards

- **2024-2025 Academic Year:** Awarded the **"KuanDe" Comprehensive Excellence Scholarship** (宽德综合优秀奖学金).

- **2023-2024 Academic Year:** Awarded the **"Tsinghua Friend-Huawei" Comprehensive Excellence Scholarship** (清华之友-华为综合优秀奖学金).