香港科技大学(广州)
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY (GUANGZHOU)

数据科学与分析学域
DATA SCIENCE AND ANALYTICS THRUST
信息枢纽
INFORMATION HUB

DSA Research Experiences for Undergraduates

Research Project

Section1: Faculty Information

Full Name	CHU Xiaowen	Tel	
Thrust/Hub	DSA Thrust/ Information Hub	Office	E1 L6 603
Email	xwchu@hkust-gz.edu.cn		

Section2: Research Project Proposal

Project Title	**Ego-View Drone Video Understanding for Scene Understanding**
Project Description (max 800 words)	**1.Background** Ego-view video understanding has emerged as a critical research area driven by wearable devices (e.g., AR glasses) and autonomous driving. However, applying ego-view understanding to drone platforms introduces unique challenges and opportunities. Unlike ground-based scenarios (e.g., Ego4D datasets with head-mounted static cameras), drones operate in highly dynamic space environments, requiring real-time spatial reasoning and for the aerial tasks such as obstacle avoidance, terrain mapping, and target tracking. This project aims to bridge the gap between conventional ego-view research and the demands of aerial platforms by developing a dedicated dataset and evaluation framework for drone-centric video understanding. **2.Challenges** Drone ego-view video understanding faces the two key challenge. First, **high-speed maneuvers** (e.g., pitch, yaw, altitude changes) introduce severe motion blur and unstable viewpoints, degrading the performance of existing video stabilization and action recognition models. Second, Ground-centric datasets (e.g., Ego4D) lack annotations for **aerial-specific tasks**, such as obstacle localization, flight path planning, and terrain traversability analysis. **3. Objective** First, this project will collect the video data for drone, capturing diverse environments (urban, forest, aquatic) with visual and geospatial data from the open web. Second, we will utilize the existing data annotation method (such as the crowdsourcing) to annotate the language query-answer for the drone video data and extend it to the large-scale multimodal dataset for drone ego-view understanding. Third, we will develop a comprehensive benchmark to evaluate state-of-the-art multimodal large language models (MLLM) on aerial tasks, exposing limitations in spatial reasoning and temporal coherence. **4. Expected Outcomes**

	• **Ego-Drone Dataset**: A multimodal aerial ego-view dataset with rich annotations, enabling research in scene understanding. • **Benchmark Insights**: Quantitative analysis of MLLM models' limitations in handling aerial dynamics, providing guidelines for model adaptation (e.g., motion-robust architectures). • **Open-Source Tools**: Release data preprocessing code, evaluation scripts, and baseline models to accelerate community-driven research. **5. Required Competencies** The students are expected to have the following skills: • Fundamental understanding of Large Language Models (LLMs) and Multimodal Large Language Models (MLLMs). • Basic knowledge in Computer Vision (CV) and Natural Language Processing (NLP). • Basic knowledge in basic Python programming, Linux system operations and web crawler.
Proposed Research Duration	Start Date: 2025 / 03/ 10 End Date: 2025 / 09/ 10
Student/Researcher Duties	The students are expected to conduct a systematic review of the specific domain. Additionally, they are required to collect data from the internet using web crawlers and complete the benchmark evaluation using Python through MLLM.
Technical Skills Required	☑Python ☐ Machine Learning ☐ Big Data ☐ R ☑ Deep Learning ☐ SQL ☐ C/C++ ☐ Other: _____
Preferred Student/Researcher Background	LLM, MLLM, Python Programming, Web Crawler
Maximum Number of Students/Researchers	☑ 1 ☐ 2

Section3: Pre-Application Research Exposure Meeting

Faculty members are encouraged to schedule a Research Exposure Meeting to introduce students to their projects.

Preferred Date	**2025.03.07**
Preferred Time	14.00
Meeting Mode	☐ In-Person ☑ Online
Venue (if in-person)	
Meeting Link (if online)	https://meeting.tencent.com/dm/S2r9xLSqaFmz