Hao Zhou

(+86)199-5898-5988 auzhouhao@mail.scut.edu.cn

Education

- **South China University of Technology** *Sep 2022 – Jul 2026 (expected)*
  B.Eng, Majoring in Automation, Senior Undergraduate
  GPA: 3.7/4.0
  Main Course: Signal Analysis and Processing (4.0/4.0), Calculus (4.0, 4.0)/4.0, Linear Algebra (4.0/4.0) .....

Publication

1. **Retrieval is Not Enough: Enhancing RAG through Test-Time Critique and Optimization**
   Jiaqi Wei*, Hao Zhou*, Xiang Zhang, Di Zhang, Zijie Qiu, Wei Wei, Jinzhe Li, Wanli Ouyang, Siqi Sun
   Advances on Neural Information Processing Systems (NeurIPS), 2025.

Research Experience

- **MLL Lab, Northwestern University** *Jun 2025 – Sep 2025*
  Remote Research Intern, Mentor: Canyu Chen
  - Investigated the robustness of LLM-based search agents under imperfect retrieval environments characterized by six noise types: vague, misleading, false, irrelevant, incomplete, and redundant information.
  - Constructed a noisy retrieval environment given the correct information and trained search agents under this environment leveraging RL frameworks to evaluate their resilience to low-quality content.
  - Introduced a score-based confidence mechanism that enables the LLM to self-evaluate its trust in retrieved passages, using the confidence scores to guide the reasoning process.
- **Shanghai Artificial Intelligence Laboratory** *Sep 2024 – Mar 2025*
  Research Intern, Advisor: Prof. Yu Cheng
  - Developed a **medical assistant agent** for end-to-end healthcare workflows, integrating multiple LLMs to handle different subtasks (e.g., appointment scheduling, medical QA), enabling end-to-end automation of real-world clinical workflows.
  - Constructed a reasoning-augmented dataset by sampling from open-source medical QA corpora and **distilling** reasoning models to generate high-quality reasoning traces with an optimized inference backend (e.g., vLLM), then fine-tuned the LLMs using **LoRA** to enhance reasoning capability and medical QA accuracy.
  - Co-led a **joint first-author** study focused on retrieval-augmented generation (RAG); introduced contrastive critique fine-tuning, which trains critique models on prompts paired with both high- and low-quality responses to generate critiques iteratively for test-time refinement in RAG.
- **One-Shot Industrial Defect Segmentation Challenge (ECCV 2024)** *Jul 2024 – Aug 2024*
  Collaborator, Mentor: Xiaoyang Wang
  - Background: we need to tackle one-shot defect segmentation under severe data imbalance, requiring generalization to unseen defect types and adaptation to novel product categories without retraining.
  - Proposed a three-part solution: (1) high-resolution patch slicing (448×448) to preserve small-scale defect details; (2) an enhanced FPTrans-based dual-stream ViT with residual connections for improved feature retention; (3) **visual prompting** via red foreground masks to guide the support encoder toward defect regions.
  - Received **Third Prize** globally, demonstrating strong generalization in few-shot industrial defect segmentation.

- **Biometrics and Intelligence Perception Lab, SCUT** *Sep 2023 – May 2024*
  Research Intern, Advisor: Prof. Wenxiong Kang (IEEE Fellow)
  - Gained foundational experience in machine learning and deep learning, and surveyed published works in knowledge distillation and gait recognition.
  - Applied knowledge distillation (KD) techniques to improve gait recognition performance on outdoor datasets, enhancing model generalization under diverse viewing conditions.

Project Experience

- **China Undergraduate Engineering Practice and Innovation Ability Competition** *Jun 2023 – Oct 2023*
  - Built an end-to-end robotic system for real-time classification of four waste categories using YOLOv5 algorithm.
  - Collected and annotated custom dataset; trained models and deployed them on **Nvidia Jetson** devices with accelerated inference using *TensorRT*; implemented hardware-software communication with the STM32 microcontroller; received **Second Prize** in the competition.

- **Summer School, National University of Singapore (NUS)** *Jul 2023*
  - Applied classical machine learning algorithms (e.g., Decision Tree, Random Forest) to complete a traffic sign classification task involving seven classes.
  - Achieved a **Distinction** grade based on model performance and timely project completion.

- **Intramural Robot Competition** *Mar 2023 – May 2023*
  - Developed real-time visual algorithms in C++ using the OpenCV library on Linux to enable block detection and grasping in multi-terrain environments.
  - Co-designed and built a terrain-adaptive robotic system with teammates; received the **Open Source Award** (ranked 1st/21) in the same track.

Honors and Awards

- **Bronze Prize**, One Shot Industrial Defect Segmentation Challenge (ECCV2024)
- The Second Prize **(Top 6%)**, China Undergraduate Engineering Practice And Innovation Ability Competition
- Distinction Grade **(highest honor)**, Summer School of National University of Singapore
- Open Source Award **(ranked 1st/21 comprehensively)**, Intramural Robot Competition