ZHILIN(HADLAY) ZHANG

✉ hadlay.zhang@gmail.com   in linkedin.com/in/Zhilin-Zhang
github.com/Hadlay-Zhang   🌐 hadlay-zhang.github.io

Research Interests

- **Multi-modal AI**, especially in the context of human-like perception and reasoning ability, and its application in the domains like (Medical) Visual Question Answering and Image Captioning.

- **LLM Agents**, specifically multi-agent systems with dynamic communication protocols and task decomposition mechanisms that can collaborate and coordinate to solve complex real-world problems.

Education

New York University (Tandon School of Engineering), New York, USA　　　　**Sep 2024 – Present**

Master of Science in Computer Science, Overall GPA: 3.9/4.0
Courses: *Machine Learning (A), Operating System (A), Programming Languages (A), Intro to Java (A), Algorithmic Machine Learning and Data Science (A-), Deep Learning (A-)*

Tongji University, Shanghai, China　　　　**Sep 2020 – Jul 2024**

Bachelor of Engineering in Computer Science, Overall GPA: 4.31/5.0; 88.11/100
Courses: *Data Structure (A), Software Engineering (A), Data Mining (A), Machine Learning (A), Computer Vision (A), Natural Language Processing (A)*
Honors and Awards: *Tongji University Scholarship for Undergraduate Freshmen (2020-2021); Tongji Undergraduate Third-Class Scholarship (2021-2022, 2023-2024)*

Publications and Preprints

[†] denotes equal contribution[1]. **C** = Conference, **J** = Journal, **P** = Preprint.

[P1] **Zhilin Zhang**[†], Xiang Zhang[†], Jiaqi Wei, Yiwei Xu, Chenyu You. *PosterGen: Aesthetic-Aware Paper-to-Poster Generation via Multi-Agent LLMs.* ArXiv 2025. [paper][code]

[P2] Jie Wang, Yichen Wang, **Zhilin Zhang**, Kaidi Wang, Zhiyang Chen. *SentiXRL: An advanced large language Model Framework for Multilingual Fine-Grained Emotion Classification in Complex Text Environment.* ArXiv 2024. [paper]

[C1] **Zhilin Zhang**, Jie Wang, Ruiqi Zhu, Xiaoliang Gong. *Efficient Bilinear Attention-based Fusion for Medical Visual Question Answering.* International Joint Conference on Neural Networks (IJCNN) 2025. [paper]

[C2] Zhanghao Qin, Hongming Zhang, **Zhilin Zhang**, Tianyu Wang, Hongtao Mao, Guangzhen Yao. *Enhancing Intra-Modality Compactness in Text-to-Image Person ReID.* International Joint Conference on Neural Networks (IJCNN) 2025.

[C3] **Zhilin Zhang**, Fangyu Wu. *Enhanced Textual Feature Extraction for Visual Question Answering: A Simple Convolutional Approach.* 6th International Conference on Computer Vision and Computational Intelligence 2025. [paper]

Research Experience

Research Intern, Stony Brook University　　　　**May 2025– Present**
mentored by Prof. Chenyu You　　　　*AI4Research, Multi-agents*

Graduate Research Assistant, Text2API　　　　**Oct 2024– May 2025**
mentored by Dr. Yujian Gan (University College London)　　　　*Python; LLM agents; Text2SQL*

- Developed multiple agents with a layered query parsing architecture and managed robust inter-agent communication to accurately interpret natural language queries and provide reliable feedbacks.
- Led the generation of intermediate training data for fine-tuning LLMs by implementing two strategies: rewriting basic APIs and generating APIs based on metadata.
- Fine-tuned the models (e.g., Llama-3.1) using LLaMA Factory and conducted comprehensive testing and evaluations.

Undergraduate Thesis, Medical Visual Question Answering　　　　**Sep 2023– Aug 2024**
mentored by Prof. Xiaoliang Gong (Tongji University)　　　　*Pytorch; Vision and Language*

---

[1]Google Scholar: https://scholar.google.com/citations?user=vWTBvyQAAAAJ

- Proposed a fusion framework **OMniBAN** by integrating Orthogonality Loss, Intra-modal Multi-head Self-Attention, and Cross-modal Bilinear Attention Networks to efficiently fuse visual (from **BiomedCLIP**) and textual (from **BioBERT**) representations for MedVQA.
- Conducted extensive comparative experiments on VQA-RAD and SLAKE datasets, and demonstrated that OMniBAN achieves competitive accuracy while reducing computational complexity (around $3.8\times$ **fewer FLOPs** and $1.5\times$ **fewer parameters**) compared to Transformer-based Co-Attention fusion models.

Undergraduate Research Assistant, Visual Question Answering (VQA) Mar 2023– May 2024
mentored by Dr. Jiaxi Yang (Columbia University) and Dr. Zhi Liu (Zhejiang Lab) *Vision and Language*

- Conducted a thorough investigation comparing different text models including TextCNN, GRU, LSTM, and Transformer Encoder on the VQA-v2 dataset, revealing that complex models like Transformer Encoder often underperform compared to simpler sequence models such as GRU and LSTM in VQA tasks.
- Proposed ConvGRU that integrates convolutional layers with GRU to improve local feature extraction from text, enhancing model performance on the VQA-v2 dataset without substantially increasing parameter complexity.
- Conducted extensive case studies demonstrating the practical benefits of the added convolutional layers in extracting local textual information and analyzed the text length distribution in the VQA-v2 dataset.

Undergraduate Research Assistant Jun 2022– Oct 2022
Shanghai Research Institute for Intelligent Autonomous Systems *Re-Identification*

- **Second inventor** for the patent: "*Intelligent Unmanned Parking Method based on Spatio-Temporal Vehicle Re-Identification*" Patent Link
- Put forward a solution for smart parking based on vehicle re-identification, spatial-temporal correlation, sensorless payment, and the architecture of the automated parking system.
- Drafted the patent claims and description, designed point-based path-finding algorithms and YOLOv5-based vehicle license plate recognition algorithms.

Internships

Research & Development Intern Sep 2023– Jan 2024
Momenta Automotive Technology Co., Ltd. *C++; Bash*

- Constructed thorough test scenarios for autonomous driving state transitions based on ROS, ensuring robust and accurate simulation environments.
- Developed and maintained test code using Shell scripts for various client releases of autonomous driving software, ensuring each release met stringent quality and functionality standards.
- Assisted in designing and developing clustering standards for autonomous driving fault diagnosis, collaborating across departments to align code with fault categorization standards, enhancing overall system reliability and safety.

Assistant Software Engineer Jun 2023– Sep 2023
Ronovo Surgical Technology Co., Ltd. *C++; Bash*

- Developed embedded audio communication software for the control center of the surgical robot and manual control panel, conducting thorough function testing to ensure reliability.
- Implemented a high-quality, low-latency two-way audio communication system using the GStreamer framework and audiornnoise plugin to optimize for performance and clarity.
- Examined and fine-tuned the audio pipeline, adjusting parameters to accommodate various audio equipment and sound cards procured by the hardware department, ensuring seamless integration and functionality.

Additional Skills

- **Standard Tests:** TOEFL (R28, L24, S25, W28), GRE (V154, Q170)
- **Languages:** C/C++, Python, PHP, Java, Shell
- **Tools & Libraries:** Git, NumPy, Pandas, Scikit-learn, Pytorch, Keras, LangChain

Academic Services

- **Conference Reviewing:** IJCNN 2025, ACML 2025
- **Journal Reviewing:** EAAI