Qiao SUN

qiaosun22.github.io   qiaosun22@m.fudan.edu.cn   github.com/qiaosun22
Scholar   220 Handan Rd., Yangpu, Shanghai, China   (1)213-583-9491   WeChat

Education

Massachusetts Institute of Technology Cambridge, MA, U.S.
Visiting Student Researcher *Jul. 2024 - Dec. 2024*
Research Focus: 4D World Model, Embodied AI
Advised by Prof. Chuang Gan & Dr. Yilun Du

Fudan University Shanghai, China
M.Eng. in Electrical and Computer Engineering | GPA: 3.48/4.0 *Sep. 2022 - Jun. 2025*
Research Experiences: Natural Language Processing, Computer Vision, Multimodal Learning, Nursing Robots, Memristor-based Analog Computing
Highlighted Courses: Computer Vision (A, 4.0), Robotics (A, 4.0), Frontiers in Intelligent Robotics (A, 4.0), Applied Mathematical Methods (B+, 3.3), Data Science (B+, 3.3)
Advised by Prof. Shijie Guo, Prof. Nanyang Ye & Dr. Qinying Gu

Tianjin University Tianjin, China
B.Eng. (Major) in Civil Engineering | GPA: 85.10/100 *Sep. 2015 - Jun. 2019*
Highlighted Courses: Advanced Mathematics (93), Selected Explanation of Mathematical Methods (96), Introduction to College Physics (93), Probability Theory and Mathematic Statistics (92), Basic Techniques of Electrical Engineering (91), Geographic Information System and Engineering (91)
B.Mgmt. (Minor) in Financial Management *Sep. 2016 - Jun. 2019*

Research Intrests

[1] Embodied AI, 3D (4D) Computer Vision & World Models
[2] Vision-Language Learning & Multimodal AI Methods
[3] Natural Language Processing & Novel Applications of Foundation Models

Publications & Preprints (by year)

2024

[1] **Learning 4D Embodied World Models**
Under review at *CVPR 2025* [Preprint] [Website]
Haoyu Zhen[*], **Qiao Sun**[*], Pengxiao Han, Yilun Du, Chuang Gan

[2] **VL-Rotate: Vision Model Modulated by Language Model for Few-Shot OoD Rotated Object Detection**
Under review at *CVPR 2025* [Preprint]
Weihan Yin, **Qiao Sun**, Lin Zhu, Liujia Yang, Nanyang Ye.

[3] **MiniConGTS: A Near Ultimate Minimalist Contrastive Grid Tagging Scheme for Aspect Sentiment Triplet Extraction**
Accepted at Main Conference of EMNLP 2024 [Paper] [Code] [PaperWithCode]
Qiao Sun, Liujia Yang, Minghao Ma, Nanyang Ye, Qinying Gu

[4] **Enhancing Nursing and Elderly Care with Large Language Models: A Framework for AI-Driven Patient Monitoring and Interaction**

[*] Equal Contributions

Accepted at *COLING 2025* [Preprint]
Qiao Sun, Nanyang Ye, Qinying Gu, Jiexin Xie, Shijie Guo

2023
[5] **DV2DM: A Learning-based Visible Difference Predictor for Videos**
Submitted to *TPAMI* [Preprint]
Qi Fan*, **Qiao Sun***, Nanyang Ye, Qinying Gu

[6] **Synergistic Development of Perovskite Memristors and Algorithms for Robust Analog Computing**
Submitted to *Nature Communications* [Preprint]
Qiao Sun*, Qinying Gu*, Yifei Wang, Liujia Yang, Nanyang Ye, Huaqiang Wu

Research Experience

Learning 4D Embodied World Models [1]
Conducted During Visiting at MIT-IBM Watson AI Lab. *June. 2024 - Present*

- *Proposed a 4D Embodied World Model*: Developed an extendable and scalable framework to efficiently predict spatiotemporally consistent high-quality dynamic evolutions of 3D scenes in response to manipulator actions. Expanded pretrained video diffusion models with additional depth and normal output channels, enabling spatially and temporally awared RGB-DN video generation for dynamic scene understanding.
- *Created the First High-Quality RGB-DN Dataset*: Produced the first-ever RGB-DN dataset with temporally consistent annotations, establishing a foundation for robust training of 4D embodied AI models. Employed novel guidance techniques and harnessed 500+ GPUs for efficient and precise annotation of depth and normal maps.
- *Bilateral Reconstruction with RGB-DN Integration*: Achieved high-quality, temporally coherent 4D mesh reconstruction by bilaterally integrating depth and normal information, capturing fine-grained spatial and temporal dynamics. Enabled predicting efficient, precise, and explainable 4D representations of the embodied world, tailored for robotic manipulation tasks.
- *Policy Learning with 4D-Aware Representations*: Developed policy learning frameworks leveraging 4D-aware representations, significantly enhancing embodied performances in inverse dynamics modeling, and adaptive planning, outperforming prior 2D and 3D models.

VL-Rotate: Vision-Language Learning for Few-Shot OoD Rotated Object Detection [2]
Conducted During Visiting at JHCCS, SJTU. *Jan. 2024 - Jun. 2024*

- *Addressed OoD Rotated Object Detection Challenges*: Tackled the limitations of existing vision-language models in few-shot out-of-distribution (OoD) scenarios, focusing on enhancing rotated object detection in dense scenes.
- *Integrated CLIP-based Text Priors*: Designed VL-Rotate, leveraging CLIP's text encoder to align semantic representations across modalities, improving object representations in embedding space.
- *Developed Gradient-Guided Regularization*: Introduced Masked Feature Heuristics Dropout (MFHD), selectively deactivating classification features, and CLIP-guided Fine-Tuning (CFT) to guide the model's training phase effectively.
- *Achieved State-of-the-Art Results*: Demonstrated improvements of up to 45.09% in domain adaptation and 5.24% in domain generalization for few-shot OoD scenarios, outperforming prior approaches.

MiniConGTS: A Near Ultimate Minimalist Contrastive Grid Tagging Scheme for Aspect Sentiment Triplet Extraction [3]
Conducted During Internship at Shanghai AI Lab. *Oct. 2023 - Feb. 2024*

- *Rigorously Analyzed Inefficiencies in Existing Methods*: Critically analyzed the over-reliance on complex tagging schemes, external semantic augmentations, and intricate classifiers. With rigorous proof, proposed an internally optimized approach combining a minimalist tagging scheme with token-level contrastive learning.

*Equal Contributions

- *Introduced a Minimalist Tagging Scheme*: Designed a tagging mechanism with the fewest label categories to date, facilitating the learning process and reducing computational complexity without compromising expressiveness.
- *Optimized Contextual Representations*: Tailored a novel token-level contrastive learning strategy to enhance pretrained contextual embeddings from within the model, free from additional computational overhead or reliance on external semantic augmentation.
- *Achieved SotA Efficiency and Overwhelmed GPT-4*: Improved computational efficiency by up to 90% compared to prior methods, while achieving State-of-the-Art performance in ASTE tasks. Conducted the first benchmark evaluation of GPT-4o in few-shot learning and Chain-of-Thought scenarios, highlighting the continued relevance of the pretraining-finetuning paradigm.

Enhancing Nursing and Elderly Care with Large Language Models: A Framework for AI-Driven Patient Monitoring and Interaction [4]

Conducted as Part of Master's Thesis Research, Fudan University. *Sep. 2023 - Aug. 2024*

- *Proposed a Novel Task for AI in Healthcare*: Identified a new scenario and task for elderly and disabled care, distinct from traditional clinical settings, addressing the exacerbating societal pressure from aging populations and declining birth rates. Designed a comprehensive LLM-driven framework for real-world solution.
- *Specialized Multimodal Dataset and Data Mixing Strategy*: Developed a high-fidelity nursing dataset, including text, annotated images, and multi-turn dialogues. Employed a fine-grained data mixing strategy to balance domain-specific specialization with general capabilities.
- *Innovated a Multi-Stage Fine-Tuning Pipeline*: Conducted Incremental Pre-Training (IPT), Supervised Fine-Tuning (SFT), and Chain-of-Thought (CoT) reasoning to enhance multiple state-of-the-art foundational language models for healthcare-specific tasks.
- *Designed the First Nursing Competency Benchmark*: Created a novel evaluation metric for foundational nursing capabilities. Experimental results demonstrated significant improvements in baseline models, validating the effectiveness of the proposed pipeline.

DV2DM: A Learning-based Visible Difference Predictor for Videos [5]

Conducted During Visiting at JHCCS, SJTU. *Sep. 2023 - Dec. 2023*

- *Proposed a New Task for Video VDP*: Addressed the challenges of evaluating visual differences in video content, distinct from traditional image quality metrics, by focusing on pixel-wise visible differences and subjective human visual perception in dynamic video contexts.
- *Developed the ViLocVis Dataset*: Curated the first-ever large video VDP dataset, consisting of over 1,000 video pairs with annotated by 10 volunteers under 12 diverse viewing conditions.
- *Designed a State-of-the-Art Deep Learning Method*: Proposed DV2DM, a novel method based on a customized U-ViT architecture with a Siamese U-shaped network, featuring dual encoding branches and a unified decoding branch for effective spatio-temporal feature extraction. Incorporated environmental factors into the model architecture, improving its adaption to visual differences.
- *Demonstrated Versatility Through Applications*: Extended DV2DM to real-world scenarios such as content-adaptive watermarking, visually lossless video compression, invisible adversarial attacks, and video super-resolution quality metrics.

Synergistic Development of Perovskite Memristors and Algorithms for Robust Analog Computing Leveraging Bayesian Optimization [6]

Conducted During Internship at Shanghai AI Lab. *Apr. 2023 - Dec. 2023*

- *Proposed a Synergistic Framework*: Unified perovskite memristor fabrication optimization and robust analog DNN development to address non-idealities in memristor systems. Applied BO to identify ideal materials and fabrication conditions, enhancing energy efficiency and adaptability. Introduced BO-guided noise injection to improve analog DNNs' resilience against memristor imperfections.
- *Validated on Diverse Tasks*: Achieved up to 100-fold performance gains across image classification, autonomous driving, and large vision-language models.
- *Experimental Demonstration*: Verified the approach on a 10×10 optimized memristor crossbar with high classification accuracy and low energy consumption.

Internships

MIT-IBM Watson AI Lab *Jul. 2024 - Present*
Supervisor: Prof. Chuang Gan & Dr. Yilun Du
4D World Model, Embodied AI.

JHCCS[*], Shanghai Jiao Tong University *Jul. 2023 - Jul. 2024*
Supervisor: Prof. Nanyang Ye
Natural Language Processing, Computer Vision, Multimodal Learning.

Shanghai AI Lab *Mar. 2023 - Dec. 2023*
Supervisor: Dr. Qinying Gu & Prof. Tianfan Xue
Memristor-Based Analog Computing and Artificial Intelligence.

Western Securities, R&D Center *Sep. 2019 - Jul. 2020*
Supervisor: Yumeng Zhang
Data-Driven Financial Engineering, Fund-of-Fund (FoF) Investment Strategies.

Services

Volunteer at COLING 2025, Abu Dhabi (upcoming)	*Jan. 2025*
Reviewer for NAACL 2025 and ACL ARR October	*Nov. 2024*
Volunteer at EMNLP 2024 and its NLP4Science Workshop, Miami	*Nov. 2024*

Honors & Awards

First-class Academic Scholarship at Fudan University	*2024*
Outstanding Internship Award at Shanghai AI Lab	*2023*
First-class Academic Scholarship at Fudan University	*2023*
Asia and Pacific Mathematical Contest in Modeling, Second Prize	*2018*
Tianjin College Students Innovation and Entrepreneurship Competition, First prize	*2017*
Henan High School Students' Chemistry Competition, Second Prize	*2015*
Robot Competition in the National Computer Production Activity, Third Prize	*2014*

Skills

Programmng:
(Proficient) Python (PyTorch, Pillow, OpenCV-python, SciKit-Learn, Open3D, Transformers, etc.), Shell, SLURM, Git, Blender, MuJoCo, LaTeX
(Familiar) C/C++, CUDA, MATLAB, HTML/CSS
Math: Matrix Theory, Kolmogorov Probability Theory, Advanced Statistics, Complex Analysis, Differential Geometry
Computer Science: Parallel Computing, Network, Compiler Theory, Computer Organization and Architecture

[*] John Hopcroft Center for Computer Science