HAOJIAN HUANG

TEL.: +86-135 0246 0463 · E-MAIL: haojianhuang927@gmail.com Github: JethroJames | Personal Links: wonghougin.me

EDUCATION

| Department of Computer Science, University of Hong Kong |
|---------------------------------------------------------------------------|
| MSc in Computer Science |
| College of Computer Science and Technology, Harbin Engineering University |

- **Bachelor of Engineering in Information Security**
- GPA: 86.94/100 (Ranking: 5/25)

PROFESSIONAL EXPERIENCES

HUAWEI | NOAH'S ARK LAB | Decision Making & Reasoning -- Intern

Researching high-level planning and reasoning for robots, leveraging diffusion-based slow thinking combined with reinforcement fine-tuning to enhance their ability to handle complex downstream tasks (e.g., stacking cups, folding clothes).

TeleAI -- Intern (Supervisor: Associate Professor Chen Mulin & Professor Xuelong Li)

Led the development of innovative technologies in Zero-Shot Learning, Multimodal NER, and Video Temporal Grounding, significantly enhancing performance and interpretability, with research findings published in top conferences like ACM MM 2024 and AAAI 2024.

Hangzhou Innovation Institute of Beihang University -- Intern

Researched multimodal traffic prediction on spatio-temporal graphs, built a robust codebase, and led data integration, aiding project success and patent efforts.

Big Data Development Center, Ministry of Agriculture and Rural Affairs of the People's Republic of China -- Researcher at the **Comprehensive Data Analysis Department** 03/2023-06/2023

Deployed literature crawlers and temporal analysis algorithms for data analysis, authored a paper on ultra-short-term wind speed prediction for wind farms submitted to top-tier journal EAAI, and aided in report development.

Key Laboratory of Multi-disciplinary Collaborative Cognitive AI Technology and Application in Heilongjiang Province --03/2022-03/2023 Intern

Led the National Undergraduate Innovation Program, organizing field research, literature reviews, software development, and thesis writing, focusing on real-time people flow systems using deep learning and a wind speed forecasting model with attentiongated RNN, submitted to Q1 top journal Energy.

Huazhong University of Science and Technology - RA (Supervisor: Professor Yang Ruotian)

Researched Trustworthy AI in multi-view clustering, built an efficient and scalable code framework from scratch, and wrote a high-quality paper detailing motivation, methodology, experiments, and results.

RESEARCH PROJECTS

Temporal Regularization Makes Your Video Generator Stronger

- **Role:** Led the research on temporal quality in video generation, proposing and implementing the FluxFlow strategy while overseeing experiments and result analysis.
- Achievements: improving coherence and diversity across models on UCF-101 and VBench, while maintaining spatial fidelity.

VistaDPO: Video Hierarchical Spatial-Temporal Direct Preference Optimization for Large Video Models 11/2024-01/2025

- Role: Led the design and implementation of VistaDPO, focusing on hierarchical text-video alignment and the creation of the VistaDPO-7k dataset for fine-grained preference optimization.
- Achievements: Developed a novel framework improving video-language alignment across three levels, significantly enhancing Large Video Models (LVMs) performance on Video QA, Captioning, and Hallucination benchmarks.

BoViLA: Bootstrapping Video-Language Alignment via LLM-Based Self-Questioning and Answering 09/2024-12/2024

- **Role:** Managed project, pioneering self-questioning paradigm to enhance video-language alignment.
- Achievements: Activated LLM-based multimodal alignment, boosting precision and contextual awareness in video-language tasks via high-quality question generation.

VideoGen-of-Thought: A Collaborative Framework for Multi-Shot Video Generation

- Role: Spearheaded the design and development of VGoT, a training-free architecture for multi-shot video generation, focusing on narrative logic and cross-shot consistency.
- Achievements: Proposed a modular framework for multi-shot video generation, achieving logical storytelling and visual consistency, with innovations in script generation, identity preservation, and cross-shot smoothing.

Trusted Unified Feature-Neighborhood Dynamics for Multi-View Classification

- Role: Lead the entire project process, developing core methods for unified feature neighborhood representation and integrated dynamic Markov random field fusion.
- Achievements: Achieved superior accuracy, reliability, and stability in complex, high-conflict scenarios, outperforming traditional methods for enhanced decision precision.

Hong Kong, China 09/2024-Present

> Harbin, China 09/2020-07/2024

10/2024-Present

02/2024-08/2024

10/2023-02/2024

07/2021-05/2023

01/2025-03/2025

08/2024-11/2024

06/2024-08/2024

Beyond Uncertainty: Evidential Deep Learning for Robust Video Temporal Grounding

Role: Led the project, focusing on systematically addressing video temporal uncertainty via novel evidence-based deep learning.
Achievements: Significantly enhanced model robustness and interpretability for open-ended user queries, setting a new standard for uncertainty management in video grounding tasks.

CREST: Cross-modal Resonance through Evidential Deep Learning for Enhanced Zero-Shot Learning 02/2024-04/2024

- **Role:** Led the entire project, focusing on developing the core methodology for robust multimodal resonance in zero-shot learning.
- Achievements: Enhanced the reliability of modal fusion and decision-making, achieving remarkable performance in open-domain zero-shot scenarios, addressing challenges related to data scarcity and uncertainty.

FineCLIPER: Multi-modal Fine-grained CLIP for Dynamic Facial Expression Recognition with AdaptERs 01/2024-04/2024

- Role: Managed the project, pioneering fine-grained semantic extraction technology and adapter-based fine-tuning methods.
- Achievements: Notably enhanced CLIP's comprehension of motion features and emotional semantics, boosting its capability in dynamic facial expression recognition.

PATENT CERTIFICATE

An Autonomous Snow-Sweeping Robot, with Publication Number: CN202121512613 (Sole First Inventor)

Selected Publications

Trusted AI: Focused on exploring the potential of Evidence Theory from a set and distribution perspective for robust multimodal alignment, and tailoring credible AI solutions for downstream tasks such as multi-view and multi-modal applications.

- Adaptive Weighted Multi-View Evidential Clustering With Feature Preference (Student First Author, KBS, Q1 Top)
- L2-Regularization Based Two-Way Weighted Neutrosophic Clustering With Manhattan and Euclidean Distances (FSS, Q1 Top)
- CREST: Cross-modal Resonance through Evidential Deep Learning for Enhanced Zero-Shot Learning (First Author, ACM MM 2024)
- Trusted Unified Feature-Neighborhood Dynamics for Multi-View Classification (First Author, AAAI 2025)
- BoViLA: Bootstrapping Video-Language Alignment via LLM-Based Self-Questioning and Answering (arXiv, IJCAI 2025 round2 review)
- DependEval: Benchmarking LLMs for Repository Dependency Understanding (arXiv, ACL 2025, under review)

Personalized AGI: Introduced GaussianVTON, a 3D virtual try-on framework (The first 3D virtual try-on framework based on image prompts) leveraging Gaussian Splatting for image-cue-driven 3D editing; Researched and designed a three-stage refinement strategy, extensively validated through experiments for effectiveness in 3D virtual try-on and 3D editing.

• GaussianVTON: 3D Human Virtual Try-ON via Multi-Stage Gaussian Splatting Editing with Image Prompting (Third Author, arXiv)

Video Understanding and Generation: Enhancing pretrained LVLM through SFT/RL-Tuning for video understanding and exploring self-supervised learning and applications for Video Generation Models.

- FineCLIPER: Multi-modal Fine-grained CLIP for Dynamic Facial Expression Recognition with AdaptERs (ACM MM 2024)
- Beyond Uncertainty: Evidential Deep Learning for Robust Video Temporal Grounding (Co-First Author, arXiv)
- VistaDPO: Video Hierarchical Spatial-Temporal Direct Preference Optimization for Large Video Models (First Author, under review at ICML 2025)
- VideoGen-of-Thought: A Collaborative Framework for Multi-Shot Video Generation (*arXiv*, under review at ICCV 2025)
- Temporal Regularization Makes Your Video Generator Stronger(*arXiv*, under review at ICCV 2025)

AWARDS / HONORS

Competition Wins:

- Meritorious Award at the Mathematical Contest in Modeling (MCM) for Undergraduate Students, USA
- Provincial First Prize at the National College Students Mathematical Modeling Contest
- Provincial Second Prize at the Northeast China Mathematical Modeling League

Kaggle Expert:

- Silver Medals (Top 2%)*3 in CV-related Competitions
- Silver Medals (Top 3%)*2 in NLP-related Competitions
- Bronze Medal (Top 7%)*1 in an NLP-related Competition

University-level Honors:

- University-level First-Class Scholarship (once)
- University-level Second-Class Scholarship (twice)
- University-level Third-Class Scholarship (twice)
- Third Place in the "Electromechanical Cup" Pull-up Team Competition

SERVICES

Reviewer

ICML/ICLR/NeurIPS/ACM MM | TIP/TAFFC

Clubs and Organizations

CareerSynapse (Leader)

SKILLS

- IT skills: Python (PyTorch&TensorFlow), Java, C/C++, JavaScript, CSS, HTML, MATLAB, Linux, MySQL, OpenCV, React-Native, Gurobi, Git, Shell, LaTeX, Markdown, Web Establishment, Qt, SPSS, Arduino
- Languages: English (fluent, CET-6, IELTS 6.5); Mandarin (native)