Huimin Zeng

✉ zenghuimin1998@gmail.com | ⌂ huimin-zeng.github.io | in LinkedIn | 🎓 Google Scholar

Research Interest

Multimodal LLMs for vision-language reasoning and 3D-grounded generation; interpretability and controllability in computational photography; prior work in low-level vision and 3DGS.

Selected Publications

- **Huimin Zeng**, Yue Bai and Yun Fu, "Arbitrary-Scale 3D Gaussian Super-Resolution with Diffusion Priors", Under Review. ArXiv:2508.16467
- **Huimin Zeng**, Jiacheng Li and Zhiwei Xiong, "Plug-and-Play Versatile Compressed Video Enhancement", **CVPR 2025**. Website
- **Huimin Zeng**, Jiacheng Li, Ziqiang Zheng and Zhiwei Xiong, "All-in-One Image Compression and Restoration", **WACV 2025** (**oral**). arXiv:2502.03649
- Ziqiang Zheng, Yiwei Chen, **Huimin Zeng**, Tuan-Anh Vu, Binh-Son Hua, Sai-Kit Yeung, "MarineInst: A Foundation Model for Marine Image Analysis with Instance Visual Description", **ECCV 2024** (**oral**). Website
- **Huimin Zeng**, Jie Huang, Jiacheng Li and Zhiwei Xiong, "Region-Aware Portrait Retouching with Sparse Interactive Guidance," IEEE Transaction on Multimedia (**TMM**), doi: 10.1109/TMM.2023.3262185
- **Huimin Zeng**, Weinong Wang, Xin Tao, Zhiwei Xiong, Yu-Wing Tai and Wenjie Pei, "Feature Decoupling-Recycling Network for Fast Interactive Segmentation," **ACM MM 2023**, doi/10.1145/3581783.3611837

Industry Experience

Sanofi

Full-time Research Intern *2025*

- *Mentor*: Dr. Wei Zhao & Dr. Yongjian Yang
- Evaluate the performance of MLLMs in pharmaceutical document understanding and reasoning.
- Discover the challenge of redundant tokens and complexity-resolution mismatch in long document VQA.
- Develop a hierarchical token fragmentation mechanism for efficient understanding and reliable reasoning.

Microsoft Research Asia (MSRA)

Full-time Research Intern *2023*

- *Mentor*: Dr. Bin Li & Dr. Jiahao Li
- Assess the performance of image codecs under challenging scenarios (*e.g.,* degraded inputs and extreme-low bitrates)
- Reveal long-termly overlooked drawbacks of clean-data-specific codecs in handling degraded inputs.
- Develop general neural image codec with the restoration ability for degradations of different types and degrees.
- Part of this internship is accepted to **WACV 2025**.

Kuaishou Technology

Full-time Research Intern   *2021*

- *Mentor*: Prof. Yu-Wing Tai & Weinong Wang
- Design the decoupling and recycling algorithm for efficient interactive segmentation.
- Deploy the efficient interactive segmentation algorithm on multiple lightweight backbones.
- Develop the interactive segmentation function of the Kuaiying APP.
- Part of this internship is accepted to **ACM MM 2023**.

Research Experience

MarineInst: A Foundation Model for Marine Image Analysis with Instance Visual Description

Boston, U.S.

Northeastern University   *12/2023 - 4/2024*

- Existing foundation models (*e.g.*, SAM and CLIP) confront challenges in terms of data distribution shift and intrinsic characteristics of marine visual data, leading to the failure of directly applying existing models to marine images, highlighting domain-specific designs.
- We propose MarineInst, a powerful and flexible marine VLM model, which could perform the instance visual description task in an automatic or interactive manner. Our instance visual description task includes instance segmentation and instance captioning.
- We propose MarineInst20M, the largest documented marine image dataset to date, with remarkable visual diversity and semantic instance mask annotations.
- MarineInst trained on MarineInst20M demonstrates strong performance on various marine analysis tasks (*e.g.*, object segmentation, semantic instance captioning and text-to-image synthesis).
- Accepted to **ECCV 2024 (oral)**.

Dynamic Token Strategy for Efficient Long Document Understanding

Boston, U.S.

Northeastern University   *6/2025 - 10/2025*

- Long document understanding with MLLMs remains challenging due to the attributes of high-resolution and fine-grained layouts.
- We analyze redundancies in long-document VQA and summarize factors that affect the efficiency as missing cross-page dependencies from independent page encoding, and computation waste from a mismatch between varying complexity and fixed-size patch size.
- We propose DTDoc, an MLLM for long-document VQA based on a dynamic token fracturing-recomposition strategy.
- Experimental results show the effectiveness of DTDoc in reducing tokens and serving as a plug-and-play improving strategy for long-document VQA.
- Ongoing project.

3D World Object-Interaction with MLLM Agent Collaboration

Boston, U.S.

Northeastern University   *2/2025 - 10/2025*

- Existing 3D large-scale scene ignores the inner relationship between objects, suffering from isolated editing for object updates.
- We formulate the 3D world generation into multi-agent collaborative tasks with prompt-guided layout generation, scene reconstruction, object recomposition/harmonization, thereby enabling object-interaction and recomposition.
- To address inconsistencies in appearance and scale, we introduce a harmonization module that seamlessly blends scene components and supports prompt-based artistic control.
- The integrated 3D world considers rationality between objects and supports high-quality results across zoom/resolution levels.
- Ongoing project.

Education

Northeastern University
Boston, U.S.
PhD. in Computer Engineering
09/2024 - Present
- **Advisor:** Prof. Raymond Fu
- **Research topic:** 3D Vision, MLLMs

University of Science and Technology of China
Hefei, China
M.S. in Information and Communication Engineering
09/2021 - 06/2024
- **Advisor:** Prof. Zhiwei Xiong
- **Research topic:** Low-level Vision, Interactive Tasks

Ocean University of China
Qingdao, China
B.S. in Electronic Information Engineering
09/2017 - 06/2021
- **Advisor:** Prof. Haiyong Zheng & Prof. Zhibin Yu
- **Research topic:** Image/Video Generation, Underwater Image Enhancement

Teaching & Service

Teaching Assistant	Undergraduate course "Object-Oriented Programming", "Data Structures".
Journal Reviewer	TPAMI, TKDD, TMM, NPJ Artificial Intelligence
Conference Reviewer	ACM MM 2023/2024, ECCV 2024, WACV 2025, CVPR 2025, AAAI 2026

Achievements & Awards

ChinaMM 2019 Underwater Image Enhancement Challenge (**Winner**)	*2019*
2019 National Artificial Intelligence Challenge on 4K UHD HDR (**Top 15%**)	*2020*
Outstanding Student Scholarship, USTC	*2023/2022*
Outstanding Freshman Scholarship, USTC	*2021*
The First Prize Scholarship, OUC	*2018/2019/2020*
The Research and Innovation Scholarship, OUC	*2019*

Programming

Languages	Python, C, C++, Matlab, LaTeX, Markdown
Frameworks	PyTorch, TensorFlow, Keras, OpenCV, PIL