Yunfei Xie

Email: xieyunfei01@gmail.com · Tel: +1-831-295-9285 · Homepage: https://yunfeixie233.github.io/

EDUCATION

Huazhong University of Science and Technology 09/2021-06/2025 (expected)
Candidate for **Bachelor of Engineering in Artificial Intelligence** GPA: 87.1/100
Scholarship & Honors: Science and Technology Scholarship (top 2% in school, ¥5000)

UC Santa Cruz 08/2024-present
Visiting Student at Baskin School of Engineering

- Advised by Prof. Yuyin Zhou and Prof. Cihang Xie
- Worked on multimodal models, datasets, benchmarks, and medical AI

Johns Hopkins University 08/2023-02/2024
Visiting Student at Whiting School of Engineering, Department of Computer Science

- Advised by Prof. Alan Yuille
- Worked on integrating efficient and explainable superpixel representations with transformers

PUBLICATIONS

Multimodality

[1] **Yunfei Xie,** Ce Zhou, Lang Gao, Juncheng Wu, Xianhang Li, Hong-Yu Zhou, Sheng Liu, Lei Xing, James Zou, Cihang Xie, Yuyin Zhou. **"MedTrinity-25M: A Large-scale Multimodal Dataset with Multigranular Annotations for Medicine"**
Submitted to The Thirteenth International Conference on Learning Representations (ICLR 2025, under review). [paper] / [project page] / [code] / [data]

[2] **Yunfei Xie,** Juncheng Wu, Haoqin Tu, Siwei Yang, Bingchen Zhao, Yongshuo Zong, Qiao Jin, Cihang Xie, Yuyin Zhou. **"A Preliminary Study of o1 in Medicine: Are We Closer to an AI Doctor?"**
Submitted to The Thirteenth International Conference on Learning Representations (ICLR 2025, under review). [paper] / [project page] / [code] / [data]

Segmentation

[3] **Yunfei Xie,** Cihang Xie, Alan Yuille, and Jieru Mei. **"From Pixels to Objects: A Hierarchical Approach for Part and Object Segmentation Using Local and Global Aggregation".**
Accepted by The 18th European Conference on Computer Vision (ECCV, 2024). [paper] / [code]

[4] **Yunfei Xie,** Alan Yuille, Cihang Xie, Yuyin Zhou, Jieru Mei. **"Few-Shot Medical Image Segmentation via Supervoxel Transformer".**
Submitted to the 39th Annual AAAI Conference on Artificial Intelligence (AAAI 2025, under review). [paper]

[5] **Yunfei Xie,** Ce Zhou, Jieru Mei, Xianhang Li, Cihang Xie, and Yuyin Zhou. **"Brain Tumor Segmentation Through Supervoxel Transformer"**
Accepted by IEEE International Symposium on Biomedical Imaging 2024 (IEEE ISBI 2024). [paper]

Generation

[6] Jiawei Mao, Xiaoke Huang, **Yunfei Xie,** Yuanqi Chang, Mude Hui, Yuyin Zhou. **"Story-Adaptor: A Training-free Iterative Framework for Long Story Visualization"**
Submitted to The Thirteenth International Conference on Learning Representations (ICLR 2025, under review). [paper] / [project page] / [code]

RESEARCH EXPERIENCE

Research Intern, VLAA lab at University of California, Santa Cruz 02/2024-Present

Summary:

- Worked on multimodal models, datasets, benchmarks, and medical AI
- Worked under the supervision of Prof. Yuyin Zhou and Prof. Cihang Xie, collaborating with Prof. James Zou (Stanford University) and Prof. Lei Xing (Stanford University)
- Led 3 projects and submitted 4 papers within 7 months

MedTrinity-25M: A Large-scale Multimodal Dataset with Multigranular Annotations for Medicine

- Proposed an automated pipeline for generating multi-granular visual and textual annotations for any given image, leveraging Retrieval-Augmented Generation (RAG) and medical-specific multimodal large language models (MLLMs)
- Developed the largest multimodal dataset for medicine based on the pipeline, covering over 25 million images across 10 modalities with multi-granular annotations for more than 65 diseases, including detailed ROIs
- Pretrained medical-specific MLLMs on MedTrinity-25M, achieving state-of-the-art performance in three medical VQA tasks, with average accuracy improvements of 10.27% over prior models
- The dataset is open-source and has received over 190 stars on GitHub within 2 months. One paper submitted to ICLR 2025 as the co-first author and project leader

A Preliminary Study of o1 in Medicine: Are We Closer to an AI Doctor

- Introduced a comprehensive benchmark on understanding, reasoning, multilinguality, and agent tasks encompassing 37 challenging medical datasets from high-impact journals, including The Lancet and NEJM.
- Explored OpenAI's o1-preview in medical benchmark; found o1-preview surpasses GPT-4 in overall accuracy by $\sim$6.4%, moving closer to an AI doctor
- Identified several weaknesses of o1 in clinical scenarios, including hallucination, inconsistent multilingual capabilities, and discrepancies in evaluation metrics, and explored potential solutions to address these issues
- One paper submitted to ICLR 2025 as the co-first author and project leader

Story-Adapter: A Training-free Iterative Framework for Long Story Visualization

- Proposed a training-free and efficient framework for generating high-quality, fine-grained long stories
- Designed an iterative paradigm that progressively optimizes image generation by repeatedly incorporating text and global constraints, resulting in more precise, fine-grained interactions and better semantic consistency across the entire story
- Achieved state-of-the-art performance of aFID and aCCS scores in generating long stories, demonstrating superior semantic coherence compared to previous methods
- One paper submitted to ICLR 2025

Research Intern, CCVL lab at Johns Hopkins University 08/2023-02/2024

Summary:

- Worked on integrating efficient and explainable superpixel representations with transformers
- Worked under the supervision of Prof. Alan Yuille
- Led 2 projects and published 2 papers within 4 months

From Pixels to Objects: A Hierarchical Approach for Part and Object Segmentation Using Local and Global Aggregation

- Developed a hierarchical superpixel-based model that simultaneously addresses two conflicting needs in segmentation: local detail for parts and global context for objects
- Designed dual aggregation framework: local aggregation for superpixels in part representation and global aggregation for group tokens in object representation
- Achieved state-of-the-art performance of mIoU on common datasets in part and object segmentation, surpassing previous methods by 2.8% and 2.0%
- One paper accepted by ECCV 2024 as the first author

Few-shot Medical Image Segmentation via Supervoxel Transformer

- Proposed differentiable supervoxels, a new representation that tends to follow the boundaries of the organ structures, which are more semantically meaningful
- Introduced SVFormer, the first 3D Transformer-based few-shot framework for medical segmentation, leveraging supervoxels to reduce redundancy while enhancing organ representation
- Validated SVFormer across three public datasets, consistently outperforming state-of-the-art methods in few-shot segmentation by 5.7%, 1.2% and 1.3% in accuracy
- One paper submitted to AAAI 2025 as the first author

ACADEMIC SERVICE

Reviewer for ICLR '25, ICLR '24, CVPR '24, and IEEE ISBI '24

SKILLS

English Proficiency: TOEFL iBT 101
Programming Languages: Python, C/C++, Shell, LaTeX, HTML