YUHAO SU

su.yuh@northeastern.edu 617-676-7360 yuhaosu.github.io GitHub

Seeking 2027 full-time applied scientist roles in computer vision, multimodal learning, and generative AI.

EDUCATION

Northeastern University, Boston, USA *Sep. 2020 - May 2027 (expected)*
Ph.D. Candidate in Computer Science, Khoury College of Computer Sciences
Thesis: Towards Practical AI Task Assistants: Video Understanding at Lower Cost

The University of Minnesota, Minneapolis, USA *Sep. 2018 - Jun. 2020*
B.A. in Mathematics and Computer Science. Transferred from Shandong University in China.

EXPERIENCE

Research Intern, Samsung Research America *May 2026 - Aug. 2026*
Advisor: Dr. Deen Dayal Mohan *Mountain View, CA, USA*

- Research on **multimodal AI** with large vision-language models.

Research Intern, Vision and Robotics Team, UII America *May 2025 - Aug. 2025*
Advisors: Dr. Zhongpai Gao, Dr. Anwesa Choudhuri and Dr. Ziyan Wu *Boston, MA, USA*

- Introduced **MedVidBench, a large-scale multi-task medical video understanding dataset** with **531K** video-instruction pairs and **8 tasks** covering video, segment, and frame-level understanding.
- Developed **MedGRPO, a multi-task RL framework** with cross-dataset reward normalization for balanced training and medical LLM judge for caption evaluation. Improves over SFT baseline across tasks (e.g., +**0.074** mIoU@0.3 on temporal action grounding, +**0.588** LLM score on video summary). [CVPR 2026]

Graduate Research Assistant, NOVA Lab, Northeastern University *Sep. 2021 - Present*
Advisor: Prof. Ehsan Elhamifar *Boston, MA, USA*

- Developing **egocentric spatial binaural audio generation**: generating binaural spatial audio given egocentric video and text descriptions. [In progress]
- Developed an **interactive temporal action segmentation framework** that learns from user feedback to improve predictions without retraining. Improves edit distance +**10.9** to +**14.5**. [ECCV 2026 submission]
- Developed an **object correspondence framework** matching objects between egocentric (first-person) and exocentric (third-person) views. Improves IoU by **10.16**% (ego-to-exo) and **6.04**% (exo-to-ego). [WACV 2026]
- Developed a **two-stage active learning framework for temporal action segmentation** using minimal annotations. First selects diverse videos, then identifies the most representative frames for labeling. Achieving **95**% of full-supervision performance with only **0.35**% labeled frames.[ECCV 2024]

SELECTED PUBLICATIONS

Yuhao Su, Ehsan Elhamifar. "Interactive Online Temporal Action Segmentation: Enabling Test-Time Learning from User Corrections," ECCV 2026 submission.

Yuhao Su, Anwesa Choudhuri, Zhongpai Gao, Benjamin Planche, Van Nguyen Nguyen, Meng Zheng, Yuhan Shen, Arun Innanje, Terrence Chen, Ehsan Elhamifar, Ziyan Wu. "MedGRPO: Multi-Task Reinforcement Learning for Heterogeneous Medical Video Understanding," CVPR 2026.

Yuhao Su, Ehsan Elhamifar. "RegionAligner: Bridging Ego-Exo Views for Object Correspondence via Unified Text-Visual Learning," WACV, 2026.

Yuhao Su, Ehsan Elhamifar. "Two-Stage Active Learning for Efficient Temporal Action Segmentation," ECCV, 2024.

HONORS & AWARDS

Maximillian Lando Scholarship, the University of Minnesota *May 2019*
Awarded for strong math performance by the School of Mathematics.

SERVICE

Teaching Assistant CS6140 Machine Learning, Northeastern University
Reviewer/ Committee CVPR, ICCV, TPAMI, IJCNN, ICASSP, APAH@ICCV2025

RESEARCH INTERESTS

Multimodal Learning: Vision+Language, Vision+Audio, Video LLMs, Audio Generation
Video Understanding: Temporal Action segmentation, Object correspondence, Video Captioning
Learning & Optimization: Reinforcement Learning, Active Learning, Human-in-the-loop, Feedback Learning