Yolo Yunlong Tang

E-mail: yunlong.tang@rochester.edu | Phone: (+1)585-616-0074

Homepage: yunlong10.github.io | Google Scholar | GitHub | LinkedIn

Research Area: *Multimodal Learning, Multimodal LLMs/Agents for Video Understanding*

Education

University of Rochester — *Aug. 2023 - Present*
Ph.D. Student in Computer Science, advised by Prof. Chenliang Xu — *Rochester, NY*

Southern University of Science and Technology (SUSTech) — *Aug. 2019 - Jun. 2023*
B.Eng. in Intelligence Science and Technology, advised by Prof. Feng Zheng — *Shenzhen, China*

Work & Internships

Amazon Ring AI — *May 2025 - Aug. 2025*
Applied Scientist Intern. Host: Wei Wang, Liqiang He — *Bellevue, WA*

ByteDance Multimedia Lab — *May 2024 - Aug. 2024*
Research Intern, mentored by Gen Zhan and Yiting Liao — *San Jose, CA*

SUSTech Visual Intelligence & Perception Lab — *Aug. 2022 - Jul. 2023*
Student Researcher, mentored by Teng Wang and Prof. Feng Zheng — *Shenzhen, China*

Tencent Data Platform — *Sept. 2021 - Aug. 2022*
Research Intern, mentored by Qin Lin and Wenhao Jiang — *Shenzhen, China*

Publications & Preprints

[1] **Yunlong Tang***, Junjia Guo*, Hang Hua, Susan Liang, Mingqian Feng, Xinyang Li, Rui Mao, Chao Huang, Jing Bi, Zeliang Zhang, Pooyan Fazli, and Chenliang Xu. *VidComposition: Can MLLMs Analyze Compositions in Compiled Videos?* In: *Proceedings of the Computer Vision and Pattern Recognition Conference* (***CVPR***). 2025.

[2] **Yunlong Tang***, Jing Bi*, Siting Xu*, Luchuan Song, Susan Liang, Teng Wang, Daoan Zhang, Jie An, Jingyang Lin, Rongyi Zhu, Ali Vosoughi, Chao Huang, Zeliang Zhang, Pinxin Liu, Mingqian Feng, Feng Zheng, Jianguo Zhang, Ping Luo, Jiebo Luo, and Chenliang Xu. *Video Understanding with Large Language Models: A Survey*. In: *IEEE Transactions on Circuits and Systems for Video Technology* (***TCSVT***). 2025.

[3] **Yunlong Tang**, Gen Zhan, Li Yang, Yiting Liao, and Chenliang Xu. *CaRDiff: Video Salient Object Ranking Chain of Thought Reasoning for Saliency Prediction with Diffusion*. In: *Proceedings of the AAAI Conference on Artificial Intelligence* (***AAAI***). 2025.

[4] **Yunlong Tang**, Daiki Shimada, Jing Bi, Mingqian Feng, Hang Hua, and Chenliang Xu. *Empowering LLMs with Pseudo-Untrimmed Videos for Audio-Visual Temporal Understanding*. In: *Proceedings of the AAAI Conference on Artificial Intelligence* (***AAAI***). 2025.

[5] **Yunlong Tang**, Junjia Guo, Pinxin Liu, Zhiyuan Wang, Hang Hua, Jia-Xing Zhong, Yunzhong Xiao, Chao Huang, Luchuan Song, Susan Liang, Yizhi Song, Liu He, Jing Bi, Mingqian Feng, Xinyang Li, Zeliang Zhang, and Chenliang Xu. *Generative AI for Cel-Animation: A Survey*. In: *Proceedings of the International Conference on Computer Vision* (***ICCV***) *Workshops*. 2025.

[6] Hang Hua*, **Yunlong Tang***, Chenliang Xu, and Jiebo Luo. *V2Xum-LLM: Cross-modal Video Summarization with Temporal Prompt Instruction Tuning*. In: *Proceedings of the AAAI Conference on Artificial Intelligence* (***AAAI***). 2025.

[7] **Yunlong Tang**, Siting Xu, Teng Wang, Qin Lin, Qinglin Lu, and Feng Zheng. *Multi-modal Segment Assemblage Network for Ad Video Editing with Importance-Coherence Reward*. In: *Proceedings of the Asian Conference on Computer Vision* (***ACCV***). 2022.

[8] **Yunlong Tang**, Jinrui Zhang, Xiangchen Wang, Teng Wang, and Feng Zheng. *LLMVA-GEBC: Large Language Model with Video Adapter for Generic Event Boundary Captioning*. In: *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops* (***CVPRW***). 2023.

[9] Jing Bi, **Yunlong Tang**, Luchuan Song, Ali Vosoughi, Nguyen Nguyen, and Chenliang Xu. *EAGLE: Egocentric AGgregated Language-video Engine*. In: *Proceedings of the 32nd ACM International Conference on Multimedia* (***ACM MM***). 2024.

[10] Jing Bi, Junjia Guo, **Yunlong Tang**, Lianggong Bruce Wen, Zhang Liu, and Chenliang Xu. *Unveiling Visual Perception in Language Models: An Attention Head Analysis Approach*. In: *Proceedings of the Computer Vision and Pattern Recognition Conference* (***CVPR***). 2025.

[11] Teng Wang*, Jinrui Zhang*, Junjie Fei*, Hao Zheng, **Yunlong Tang**, Zhe Li, Mingqi Gao, and Shanshan Zhao. *Caption Anything: Interactive Image Description with Diverse Multimodal Controls*. In: *arXiv*. 2023.

High-Impact Projects

1. **Awesome LLMs for Video Understanding** (`GitHub Stars 2.7k+`)
   Latest papers, codes, and datasets on Video-LLMs. Repository for the survey paper [2].
   `https://github.com/yunlong10/Awesome-LLMs-for-Video-Understanding`
2. **Caption-Anything** (`GitHub Stars 1.8k+`)
   Implementation of Caption-Anything [11], a versatile image processing tool that combines the capabilities of Segment Anything, Visual Captioning, and ChatGPT.
   `https://github.com/ttengwang/Caption-Anything`

Skills

Programming Languages:
Proficient: Python, C/C++, Linux Shell
Capable: JavaScript, Java, SQL, MATLAB

Natural Languages:
Mandarin Chinese (*native*), *English* (*fluent*), *Japanese* (*beginner*)

Tools & Frameworks:
PyTorch, Git, LaTeX, *OpenCV, FFmpeg, HuggingFace, vLLM*