Yunlong Tang

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

Homepage: https://yunlong10.github.io/ | Google Scholar | LinkedIn

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

Education

University of Rochester — *Aug. 2023 - Present*
Ph.D. 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

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

- Proposed Video Salient Object Ranking Chain of Thought [1] (**AAAI'25**), utilizing multimodal LLM (MLLM) to obtain salient ranking and improve the diffusion-based video saliency prediction.
- Participated in AIM Challenge on Video Saliency Prediction [15] (**ECCVW'24**).

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

- Participated in Long-form Video Understanding Challenge (**CVPRW'23**), proposed LLMVA-GEBC [7], and won the championship in Generic Event Boundary Captioning (GEBC) track.
- Proposed LaunchpadGPT [8] (**ICMC'23**) to visualize music with autoregressive language model.
- Collaborated on Caption-Anything [14] project, contributed to the interactive visual prompt module.

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

- Proposed multi-modal segment assemblage network and importance-coherence reward [6] (**ACCV'22**), achieving efficiency and accuracy in automatic advertisement video editing, contributing to patent [11].

Publications & Preprints

[1] **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: *AAAI Conference on Artificial Intelligence* (***AAAI***). 2025.

[2] **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: *AAAI Conference on Artificial Intelligence* (***AAAI***). 2025.

[3] **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: *Review*. 2024.

[4] **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: *Review*. 2024.

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

[6] **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.

[7] **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.

[8] Siting Xu*, **Yunlong Tang***, and Feng Zheng. *LaunchpadGPT: Language Model as Music Visualization Designer on Launchpad*. In: *Proceedings of the International Computer Music Conference* (***ICMC***). 2023.

[9] Hang Hua*, **Yunlong Tang***, Ziyun Zeng*, Liangliang Cao, Zhengyuan Yang, Hangfeng He, Chenliang Xu, and Jiebo Luo. *MMCOMPOSITION: Revisiting the Compositionality of Pre-trained Vision-Language Models*. In: *Review*. 2024.

[10] 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.

[11] Qin Lin, **Yunlong Tang**, Qinglin Lu, Nuo Pang, Wenhao Jiang, and Feng Zheng. *Video Editing Method and Device, Electronic Equipment and Storage Medium*. CN Patent 115,883,878. 2024.

[12] Mingqian Feng, **Yunlong Tang**, Zeliang Zhang, and Chenliang Xu. *Do More Details Always Introduce More Hallucinations in LVLM-based Image Captioning?* In: *Review*. 2024.

[13] Chao Huang, Susan Liang, **Yunlong Tang**, Yapeng Tian, Anurag Kumar, and Chenliang Xu. *Scaling Concept with Text-Guided Diffusion Models*. In: *Review*. 2024.

[14] 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.

[15] Andrey Moskalenko, Alexey Bryncev, Dmitry Vatolin, Radu Timofte, Gen Zhan, Li Yang, **Yunlong Tang**, Yiting Liao, Jiongzhi Lin, Baitao Huang, Morteza Moradi, Mohammad Moradi, Francesco Rundo, Concetto Spampinato, Ali Borji, Simone Palazzo, Yuxin Zhu, Yinan Sun, Huiyu Duan, Yuqin Cao, Ziheng Jia, Qiang Hu, Xiongkuo Min, Guangtao Zhai, Hao Fang, Runmin Cong, Xiankai Lu, Xiaofei Zhou, Wei Zhang, Chunyu Zhao, Wentao Mu, Tao Deng, and Hamed R Tavakoli. *AIM 2024 Challenge on Video Saliency Prediction: Methods and Results*. In: *Proceedings of the European Conference on Computer Vision* (***ECCV***) *Workshops*. 2024.

Open-Sourced Project Contributions

1. **Awesome LLMs for Video Understanding** (`GitHub Stars 1.6k+`)
   Latest papers, codes, and datasets on Video-LLMs. Repository for the survey paper [4].
   `https://github.com/yunlong10/Awesome-LLMs-for-Video-Understanding/`

2. **Caption-Anything** (`GitHub Stars 1.7k+`)
   Implementation of Caption-Anything [14], a versatile image processing tool that combines the capabilities of Segment Anything, Visual Captioning, and ChatGPT.
   `https://github.com/ttengwang/Caption-Anything/`

3. **VidComposition**
   High-quality benchmark [3] for evaluating MLLMs' capability of understanding video compositions.
   `https://yunlong10.github.io/VidComposition/`

4. **LLMVA-GEBC**
   Implementation of the winner solution [7] to GEBC track in Long-form Video Understanding Challenge (LOVEU) at CVPR 2023 Workshop.
   `https://github.com/zjr2000/LLMVA-GEBC/`

Honors and Awards

The First Place in the AIM Challenge on Video Saliency Prediction at ECCV 2024 Workshop *2024*
The First Place in the GEBC Track of LOVEU Challenge at CVPR 2023 Workshop *2023*
Excellent Graduate for Exceptional Performance, SUSTech *2023*
Excellent Undergraduate Thesis, Department of Computer Science and Engineering, SUSTech *2023*
The First Class of Merit Student Scholarship for Exceptional Performance, SUSTech *2022*
Research Innovation Award, Shude College, SUSTech *2021*

Teaching

Teaching Assistant at University of Rochester
CSC 245/445 Deep Learning *Instructor: Prof. Chenliang Xu* *Fall 2024*

Teaching Assistant at SUSTech
CS308 Computer Vision *Instructor: Prof. Feng Zheng* *Spring 2023*
CS308 Computer Vision *Instructor: Prof. Feng Zheng* *Fall 2022*

Service

Conference Reviewer
CVPR 2024, ACM MM 2024, ACL 2024, NeurIPS 2024, ICLR 2025

Journal Reviewer
TPAMI, TMM

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, LangChain, ComfyUI*