Yunlong Tang

Ph.D. Student, University of Rochester

250 Hutchison Rd, Rochester, NY 14620

Phone: (+1)585-616-0074

E-mail: yunlong.tang@rochester.edu

Homepage: https://yunlong10.github.io/

Last updated: Friday 22nd November, 2024

Education

University of Rochester, Rochester, NY, USA — 2023 - now
Ph.D. in Computer Science
Advisor: Chenliang Xu

Southern University of Science and Technology (SUSTech), Shenzhen, China — 2019 - 2023
B.Eng. in Intelligence Science and Technology
Advisor: Feng Zheng

Research Internships

ByteDance, San Jose, CA, USA — May 2024 - Aug. 2024
Mentors: Yiting Liao and Gen Zhan
Deliverables: [1]

SUSTech VIP Lab, Shenzhen, China — Aug. 2022 - Jul. 2023
Mentors: Feng Zheng and Teng Wang
Deliverables: [2, 3, 4]

Tencent, Shenzhen, China — Sep. 2021 - Aug. 2022
Mentors: Wenhao Jiang and Qin Lin
Deliverables: [5]

Teaching

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

Teaching Assistant at SUSTech
CS308 Computer Vision Instructor: Feng Zheng — Spring 2023
CS308 Computer Vision Instructor: Feng Zheng — Fall 2022

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: *Review*. 2024.

[2] 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.

[3] **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: *CVPR Workshops*. 2023.

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

[5] **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.

[6] **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.

[7] **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: *arXiv*. 2023.

[8] Jing Bi, **Yunlong Tang**, Luchuan Song, Ali Vosoughi, Nguyen Nguyen, and Chenliang Xu. *EAGLE: Egocentric AGgregated Language-video Engine*. In: *ACM Multimedia (ACM MM)*. 2024.

[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] **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: *Review*. 2024.

[11] Hang Hua*, **Yunlong Tang***, Chenliang Xu, and Jiebo Luo. *V2Xum-LLM: Cross-modal Video Summarization with Temporal Prompt Instruction Tuning*. In: *Review*. 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] 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: *ECCV Workshops*. 2024.

Open-Sourced Project & Repository Contributions

1. **VidComposition**
   Implementation of the benchmark [6] for evaluating MLLMs' capability of understanding video compositions.
   `https://yunlong10.github.io/VidComposition/`

2. **Awesome LLMs for Video Understanding**
   Latest papers, codes, and datasets on Video-LLMs. Repository for the survey paper [7].
   `https://github.com/yunlong10/Awesome-LLMs-for-Video-Understanding/`

3. **Caption-Anything**
   Implementation of Caption-Anything [2], a versatile image processing tool that combines the capabilities of Segment Anything, Visual Captioning, and ChatGPT.
   `https://github.com/ttengwang/Caption-Anything/`

4. **LLMVA-GEBC**
   Implementation of the winner solution [3] to Generic Event Boundary Captioning (GEBC) track in Long-form Video Understanding Challenge (LOVEU) at CVPR 2023 Workshop.
   `https://github.com/zjr2000/LLMVA-GEBC`

Selected Honors and Awards

The First Place [14] in the AIM Challenge on Video Saliency Prediction at ECCV 2024 Workshop 2024
The First Place [3] 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

Skills

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

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

Tools & Frameworks:
PyTorch, Git, LaTeX, OpenCV, FFmpeg, HuggingFace, LangChain, ComfyUI

Service

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

Journal Reviewer
TPAMI, TMM.