Yunlong Tang

yunlong.tang@rochester.edu (+1) 585-616-0074
https://yunlong10.github.io/

Education

University of Rochester Aug. 2023 – Jun. 2028 (Expected)
Ph.D. Student in Computer Science, advised by Prof. Chenliang Xu *Rochester, NY, US*

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*

Professional Experience

SUSTech VIP Lab Aug. 2022 – Jul. 2023
Undergraduate Student Researcher, supervised by Prof. Feng Zheng *Shenzhen, China*

- Participated in the Generic Event Boundary Captioning competition at CVPR 2023 Long-form Video Understanding Workshop, proposed and developed the LLMVA-GEBC model [2] that won the championship.
- Proposed LaunchpadGPT, which aims to utilize a language model to generate music visualization in the form of Launchpad displaying video. Results [4] accepted to International Computer Music Conference (ICMC), 2023.
- Collaborated on the Caption-Anything project, contributed to the segmentation module for supporting interactive visual prompts, and involved in the technical report [3] writing.

Tencent Sept. 2021 - Aug. 2022
Research Intern, supervised by Ms. Qin Lin and Dr. Wenhao Jiang *Shenzhen, China*

- Proposed and developed multi-modal segment assemblage network (M-SAN) and importance-coherence reward for training. The method improves efficiency and accuracy compared to current automatic advertisement video editing techniques. Results [5] accepted to ACCV 2022.
- Deployed the model in Tencent servers online to perform efficient and accurate ad video editing, and filed the patent *An Approach for Automatic Ad Video Editing*.

Research Publications

1. **Y. Tang**, J. Bi, S. Xu, L. Song, S. Liang, T. Wang, D. Zhang, J. An, J. Lin, R. Zhu, *et al.*, "Video Understanding with Large Language Models: A Survey," *arXiv preprint arXiv:2312.17432*, 2023.
2. **Y. Tang**, J. Zhang, X. Wang, T. Wang, and F. Zheng, "LLMVA-GEBC: Large Language Model with Video Adapter for Generic Event Boundary Captioning," *arXiv preprint arXiv:2306.10354*, 2023.
3. T. Wang, J. Zhang, J. Fei, H. Zheng, **Y. Tang**, Z. Li, M. Gao, and S. Zhao, "Caption anything: Interactive Image Description with Diverse Multimodal Controls," *arXiv preprint arXiv:2305.02677*, 2023.
4. S. Xu, **Y. Tang**, and F. Zheng, "LaunchpadGPT: Language Model as Music Visualization Designer on Launchpad," *arXiv preprint arXiv:2307.04827*, 2023.
5. **Y. Tang**, S. Xu, T. Wang, Q. Lin, Q. Lu, and F. Zheng, "Multi-modal Segment Assemblage Network for Ad Video Editing with Importance-Coherence Reward," in *Proceedings of the Asian Conference on Computer Vision (ACCV)*, Dec. 2022, pp. 3519–3535.

Academic Service

Joural Reviewer — IEEE Transactions on Multimedia (TMM)

Skills

Languages — English (fluent), Mandarin Chinese (native).
Coding — Python, C++, Java, MATLAB, LaTeX.
Web Dev — Html, css, JavaScript.
Misc. — PyTorch, Hugging Face, OpenCV, FFmpeg, LangChain.

Miscellaneous Experience

Teaching Assistant

2023 **Spring CS308 Computer Vision**, SUSTech.
Instructor: Prof. Feng Zheng.

2022 **Fall CS308 Computer Vision**, SUSTech.
Instructor: Prof. Feng Zheng.

Certification

2021 **Certified in Machine Learning, Modeling, and Simulation Principles** from Massachusetts Institute of Technology (MIT). Credential ID: 5ed6ad60-3f98-4009-b342-95bdae56fef5.

Awards and Achievements

2023 **The First Place** in Generic Event Boundary Captioning Track of LOVEU (Long-form Video Understanding) Challenge at CVPR 2023 Workshop.

Excellent Graduate for Exceptional Performance, SUSTech.

Excellent Undergraduate Thesis, Department of Computer Science and Engineering, SUSTech.

2022 **The First Class of Merit Student Scholarship for Exceptional Performance**, SUSTech.

2021 **Research Innovation Award**, Shude College, SUSTech.

On-going Projects

- **Audio-Visual LLM for Fine-grained Video Understanding**: aiming to enhance the fine-grained audio-visual video understanding capabilities of audio-visual LLMs through pseudo temporal boundary alignment.
- **Blind Assistant Agent for Online Video Accessibility**: aiming to generate multimodal and comprehensive video descriptions to improve online video accessibility for individuals who are blind or have low vision.
- **Instruction-tuning for Cross-modal Video Summarization**: focusing on fine-tuning Vid-LLM with instructions and interleaved video-text prompts to adeptly handle both video-to-video and video-to-text summarization tasks.