Yunlong TANG

Email: yunlong.tang@rochester.edu | Mobile: (+1) 585-616-0074 | Website: 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, CN*

PROFESSIONAL EXPERIENCES

SUSTech VIP Lab Aug. 2022 - Jul. 2023

Undergraduate Student Researcher, supervised by Prof. Feng Zheng *Shenzhen, CN*

- 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 language model to generate music visualization in the form of Launchpad displaying video. Results [3] accepted to International Computer Music Conference (ICMC), 2023.
- Collaborated on Caption-Anything project, contributed to the segmentation module for supporting interactive visual prompts, and involved in the technical report [4] writing.

Tencent Sept. 2021 - Aug. 2022

Research Intern, supervised by Ms. Qin Lin and Dr. Wenhao Jiang *Shenzhen, CN*

- Proposed and developed multi-modal segment assemblage network (M-SAN) and importance-coherence reward for training. The method improves efficiency and accuracy when 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".

PUBLICATIONS

(* equal contribution)

[1] **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, Feng Zheng, Jianguo Zhang, Ping Luo, Jiebo Luo, Chenliang Xu, "Video Understanding with Large Language Models: A Survey", *in arXiv:2312.17432,* 2023.

[2] **Yunlong Tang**, Jinrui Zhang, Xiangchen Wang, Teng Wang, Feng Zheng, "LLMVA-GEBC: Large Language Model with Video Adapter for Generic Event Boundary Captioning", *in arXiv:2306.10354*, 2023.

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

[4] Teng Wang*, Jinrui Zhang*, Junjie Fei*, Hao Zheng, **Yunlong Tang**, Zhe Li, Mingqi Gao, Shanshan Zhao, "Caption Anything: Interactive Image Description with Multimodal Controls", *in arXiv:2305.02677*, 2023.

[5] **Yunlong Tang**, Siting Xu, Teng Wang, Qin Lin, Qinglin Lu, Feng Zheng, "Multi-modal Segment Assemblage Network for Ad Video Editing with Importance-Coherence Reward", *in Proceedings of 16th Asian Conference on Computer Vision (ACCV)*, 2022.

TEACHING EXPERIENCE

SUSTech Sept. 2022 - Jun. 2023

Teaching Assistant for SUSTech CS308 Computer Vision *Shenzhen, CN*

ACADEMIC SERVICE

- **Journal Reviewer**: IEEE Transactions on Multimedia (TMM)

HONORS & AWARDS

- The First Place in Generic Event Boundary Captioning Track of LOVEU Challenge at CVPR 2023.
- Excellent Graduate for Exceptional Performance, SUSTech, 2023.
- Excellent Undergraduate Thesis, the Department of Computer Science and Engineering, SUSTech, 2023.
- The First Class of Merit Student Scholarship for Exceptional Performance, SUSTech, 2021-2022.
- Research Innovation Award, Shude College, SUSTech, 2020-2021.

SKILLS LIST

- Programming Languages: Python, C++, Java, JavaScript, MATLAB
- Libraries/Tools: PyTorch, HuggingFace, OpenCV, FFmpeg, LangChain
- Language: Chinese (native), English (fluent)

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.