

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



- 📌 **ByteDance** May. 2024 – Aug. 2024
Research Intern, supervised by Gen Zhan and Dr. Yiting Liao San Jose, CA, US
- Conducting research on large language models for Video ROI/Saliency Detection.
 - Proposed CaRDiff [1], a framework that adopts MLLM with VSOR-CoT (Video Salient Object Ranking Chain of Thought) prompting to obtain saliency ranking map, which helps improve the performance of saliency diffusion model. Results are submitted to AAAI 2024.
- 📌 **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 the LLMVA-GEBC model [2] that won the championship.
 - Proposed LaunchpadGPT, utilizing a language model to generate Launchpad displaying video for music visualization. Results [3] 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 [4] writing.
- 📌 **Tencent** Sept. 2021 – Aug. 2022
Research Intern, supervised by 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] are 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**, G. Zhan, L. Yang, Y. Liao, and C. Xu, “CaRDiff: Video Salient Object Ranking Chain of Thought Reasoning for Saliency Prediction with Diffusion,” *arXiv preprint arXiv:2408.12009*, 2024.
- 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 S. Xu*, **Y. Tang***, and F. Zheng, “LaunchpadGPT: Language Model as Music Visualization Designer on Launchpad,” *arXiv preprint arXiv:2307.04827*, 2023.
- 4 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.

- 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)*, 2022, pp. 3519–3535.
- 6 **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.
- 7 J. Bi, **Y. Tang**, L. Song, A. Vosoughi, N. Nguyen, and C. Xu, "EAGLE: Egocentric AGgregated Language-video Engine," in *ACM Multimedia 2024*, 2024.
- 8 **Y. Tang**, D. Shimada, J. Bi, M. Feng, H. Hua, and C. Xu, "Empowering LLMs with Pseudo-Untrimmed Videos for Audio-Visual Temporal Understanding," *arXiv preprint arXiv:2403.16276*, 2024.
- 9 H. Hua*, **Y. Tang***, C. Xu, and J. Luo, "V2Xum-LLM: Cross-Modal Video Summarization with Temporal Prompt Instruction Tuning," *arXiv preprint arXiv:2404.12353*, 2024.
- 10 H. Hua*, **Y. Tang***, Z. Zeng*, L. Cao, Z. Yang, H. He, C. Xu, and J. Luo, "MMComposition: Benchmarking the Compositionality for Pre-trained Vision-Language Models," *yunlong10.github.io*, 2024.  URL: <https://yunlong10.github.io/projects/mmccomposition/>.
- 11 M. Feng, **Y. Tang**, Z. Zhang, and C. Xu, "Do More Details Always Introduce More Hallucinations in LVLm-based Image Captioning?" *arXiv preprint arXiv:2406.12663*, 2024.

Academic Service




Conference Reviewer  CVPR 2024, ACM MM 2024, ACL 2024, NeurIPS 2024
Journal 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

- 2024  **Fall CSC 245/445 Deep Learning**, University of Rochester.
Instructor: Prof. Chenliang Xu.
- 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

- 2024  **The First Place** in the **AIM Challenge on Video Saliency Prediction** at the ECCV 2024 Workshop.

Miscellaneous Experience (continued)

- 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.