Yongliang Wu

+86 15558561706 | yongliang0223@gamil.com | Homepage | Github | Google Scholar

Education

Southeast University, MS in Computer Science — Sept 2023 – Apr 2026

- GPA: 3.60/4.0. Average score: 87.2. Advisor: Xu Yang.

Southeast University, BS in Artificial Intelligence — Sept 2019 – Apr 2023

- GPA: 3.75/4.0. Average score: 89.5. Award: Southeast University President Scholarship.

Internship Experiences

Machine Learning Engineer, WeChat, Tencent Inc. – Beijing, China — June 2024 – Dec 2024

- Response for the training of a Multi-Modal Language Model WeMM-2B. Took charge of organizing pre-training data and use re-caption method to generate higher-quality image-text pairs. During the SFT stage, constructed a pipeline for automatic video annotation.

Machine Learning Engineer, OpusClip – San Francisco, California, United States — Nov 2023 – May 2024

- Response for developing an automatic short-form video generation system. Given a long-form video, the system can automatically select the highlight segments within it. Multiple shor-formt videos are then generated based on different topics at any ratio, and customized ones can also be produced according to the query provided by the user. The system is mainly LLM-driven and integrates several visual/audio foundation models as tools.

Research Projects

Enhance the Temporal Grounding Ability of Video-LLMs [Paper] [Code]

- We propose a simple Visual Prompt method, which involves annotating specific numbers on each video frame to represent the frame number. It has been found that this method can significantly enhance the model's temporal grounding ability in a training-free manner, achieving SOTA results. The paper is submitted to CVPR 2025.

Unlearning Concepts for Text-to-Image Diffusion Model [Paper] [Code]

- We propose a GAN-like adversarial training framework for unlearning the target concept in text-to-image diffusion model. Additionally, we introduce a gradient surgery approach, which serves to eliminate the conflict between the unlearn and retrain objectives. Through this, efficient unlearning is achieved while the utility of the model is maintained. The paper is accepted by AAAI 2025.

Multi-Modal In-Context Learning for Image Caption [Paper] [Code]

- We explore the problem of in-context examples selection in the Multi-Modal Large Language Model. For the visual and textual modalities respectively, we propose different selection strategies, achieving an average 20.9 CIDEr score improvement in the image caption task. The paper is accepted by NeurIPS 2023.

Competitions

Hour-long videoQA challenge of the Second Perception Test challenge @ECCV2024. *Winner*

Long-Term videoQA challenge of the LOVEU Workshop @CVPR2024. *Winner*

Publications

Unlearning Concepts in Diffusion Model via Concept Domain Correction and Concept Preserving Gradient
Yongliang Wu, Shiji Zhou, Mingzhuo Yang, Lianzhe Wang, Wenbo Zhu, Heng Chang, Xiao Zhou, Xu Yang. In: AAAI Conference on Artificial Intelligence (AAAI), 2025.

Video Repurposing from User Generated Content: A Large-scale Dataset and Benchmark
Yongliang Wu, Wenbo Zhu, Jiawang Cao, Yi Lu, Bozheng Li, Weiheng Chi, Zihan Qiu, Lirian Su, Haolin Zheng, Jay Wu, Xu Yang. In: AAAI Conference on Artificial Intelligence (AAAI), 2025.

Exploring Diverse In-Context Configurations for Image Captioning
Xu Yang, ***Yongliang Wu***, Mingzhuo Yang, Haokun Chen, Xin Geng. In: Advances in Neural Information Processing Systems (NeurIPS), 2023.