

# Yongliang Wu

yongliang0223@gmail.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.
- 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
- Participated in the training of Multi-Modal Language Model WeMM-2B. Took charge of organizing pretrain data and use recaption 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 the automatic short-form video generation project. Given a long-form video, the system can automatically select the highlight segments within it. Multiple short-form 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.
- Machine Learning Engineer**, Mettler-Toledo, Inc. – Changzhou, Jiangsu, China Jul 2022 – Dec 2022
- Trained a lightweight classifier network on supermarket product images using MobileNet v3. Initialized with contrastive learning for pre-training, and trained with triplet and center losses in metric learning. Achieved high zero-shot classification accuracy on the out-of-domain test set.

## Research Projects

- Enhance the Temporal Grounding Ability of Video-LLMs** Github Page.
- 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.
- Unlearning Concepts for Text-to-Image Diffusion Model** Paper Link. Github Page.
- 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.
- Multi-Modal In-Context Learning for Image Caption** Paper Link. Github Page.
- We first 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 compared with the baseline.

## Competitions

- Hour-long videoQA challenge of the Second Perception Test challenge @ECCV2024. Winner
- Long-Term videoQA challenge of the LOVEU Workshop @CVPR2024. Winner

## Publications

- A Glance at In-Context Learning** Apr 2024
- Yongliang Wu**, Xu Yang. In: Frontiers of Computer Science (FCS), 2024.
- Exploring Diverse In-Context Configurations for Image Captioning** Sept 2023
- Xu Yang, **Yongliang Wu**, Mingzhuo Yang, Haokun Chen, Xin Geng. In: Advances in Neural Information Processing Systems (NeurIPS), 2023.