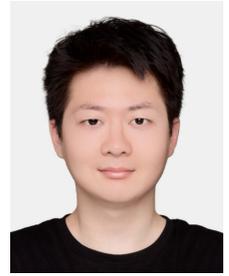


Zuhao Yang (杨祖豪)

📍 Singapore | 📞 (+65) 87906208 | ✉️ yang0756@e.ntu.edu.sg
🏠 Homepage | 🎓 Google Scholar | 🌐 github.com/mwxely



EDUCATIONAL BACKGROUND

- Nanyang Technological University** 2024.01 – 2028.01
Doctor of Philosophy in Computer Science and Engineering Supervisor: Prof. Shijian Lu
• Video-Centric Research: temporal grounding, complex reasoning, agentic tool use, and generation.
- Nanyang Technological University** 2022.08 – 2024.01
Master of Science in Artificial Intelligence cGPA: 4.8/5.0
- University of Alberta** 2017.09 – 2021.06
Bachelor of Science with Specialization in Computing Science cGPA: 3.7/4.0
• Outstanding Graduate with Distinction, Dean's Honor Roll, and International Student Scholarship.

SELECTED PUBLICATIONS

Large Multimodal Models: Video Understanding, Reasoning, Agents, and Efficiency

- *LongVT: Incentivizing "Thinking with Long Videos" via Native Tool Calling.*
Zuhao Yang*, Sudong Wang*, Kaichen Zhang*, ..., Xingxuan Li, Lidong Bing. **CVPR 2026**
- *OpenMMReasoner: Pushing the Frontiers for Multimodal Reasoning with an Open and General Recipe.*
Kaichen Zhang*, Keming Wu*, Zuhao Yang, ..., Xingxuan Li, Lidong Bing. **CVPR 2026**
- *SVAgent: Storyline-guided Long Video Understanding via Cross-modal Multi-agent Collaboration.*
Zhongyu Yang, Zuhao Yang, Shuo Zhan, Tan Yue, Wei Pang, Yingfang Yuan. **CVPR 2026**
- *A Comprehensive Study on Visual Token Redundancy for Discrete Diffusion-based MLLMs.*
Duo Li*, Zuhao Yang*, Xiaoqin Zhang, Ling Shao, Shijian Lu. **CVPR 2026**
- *ToDRE: Effective Visual Token Pruning via Token Diversity and Task Relevance.*
Duo Li*, Zuhao Yang*, Xiaoqin Zhang, Ling Shao, Shijian Lu. **arXiv 2025**
- *TimeExpert: An Expert-Guided Video LLM for Video Temporal Grounding.*
Zuhao Yang, Yingchen Yu, Yunqing Zhao, Shijian Lu, Song Bai. **ICCV 2025**

Deep Generative Models: Image, Video, and 3D Generation

- *Versatile Transition Generation with Image-to-Video Diffusion.*
Zuhao Yang, Jiahui Zhang, Yingchen Yu, Shijian Lu, Song Bai. **ICCV 2025**
- *ReChar: Revitalising Characters with Structure-Preserved Aesthetic Enhancements.*
Zhongyu Yang, Junhao Song, Zhang Luo, Zuhao Yang, et al. **SIGGRAPH Asia 2025**
- *AI-Generated Images as Data Sources: The Dawn of Synthetic Era.*
Zuhao Yang, ..., Xiaoqin Zhang, Ling Shao, Shijian Lu. **TPAMI (2/3 major revision)**
- *PaCaNet: A Study on CycleGAN with Transfer Learning for Diversifying Fused Chinese Painting.*
Zuhao Yang*, Huajun Bai*, Zhang Luo, Yang Xu, Wei Pang, ..., Yingfang Yuan. **arXiv 2023**

Natural Language Evaluation: Open-Ended Generation and Question-Answering Tasks

- *Evaluating Text Generation Quality Using Spectral Distances of Surprisal.*
Zhichen Liu, Yongyuan Li, Yang Xu, Yu Wang, Yingfang Yuan, Zuhao Yang. **EMNLP 2025**
- *QAEval: Mixture of Evaluators for Question-Answering Task Evaluation.*
Tan Yue, Rui Mao, Xuzhao Shi, Shuo Zhan, Zuhao Yang, Dongyan Zhao. **ACL 2025**
- *FACE: Evaluating Natural Language Generation with Fourier Analysis of Cross-Entropy.*
Zuhao Yang*, Yingfang Yuan*, Yang Xu*, Shuo Zhan, Huajun Bai, Kefan Chen. **NeurIPS 2023**

WORK EXPERIENCE

- Shanda AI Research Institute & MiroMind** 2025.04 – 2026.02
AI Scientist Intern
- Researched multimodal reasoning and agentic tool use for large multimodal models.
 - Collaborated with Dr. Lidong Bing and contributed to **MiroThinker** (RL infrastructure) development.

ByteDance Inc. & TikTok

2023.11 – 2025.03

Computer Vision Research Intern

- Researched video generation with diffusion models and video temporal grounding with MoEs.
- Collaborated with Dr. Song Bai and filed two patents (one US, one SG) upon project completion.

TMI Robotics Technology Co. Ltd

2021.05 – 2022.06

Natural Language Processing Research Engineer

- Developed the dialogue system of a medical robot (FAQ Service: Elastic Search + Sentence-BERT; Guide Service: TextRCNN + Knowledge Graph) and built department-specific knowledge graphs via named entity recognition, relation extraction and entity clustering. Filed one CN patent.
- Optimized model inference (OpenVINO, TensorRT) and dialogue pipeline (Rasa) with BERT compression and weight pruning.

ADDITIONAL INFORMATION

Open-Source Projects

- **Core Contributor**, [LMMs-Eval](#) (3.7k+ ★): a unified evaluation toolkit for multimodal models across text, image, video, and audio tasks.
- **Core Contributor**, [LMMs-Engine](#) (700+ ★): a lean and flexible training engine for large multi-modal models, built for rapid prototyping at scale.

Academic Services

- **Reviewer**: CVPR 24/25/26, ECCV 24/26, ACM MM 24/25/26, NeurIPS 24/25, ICLR 25, AISTATS 25/26, ICML 25, ICCV 25, BMVC 26; IEEE TPAMI, Pattern Recognition, SPIE JETI.
- **PC Member**: [SyntaGen](#), [Neural Rendering Intelligence](#), [Engineering Agentic Intelligence](#).

Filed Patents

- *Method, Device, and Medium for Video Temporal Grounding with Mixture of Experts*. US Patent.
- *Method, Device, and Medium for Generating Transition Videos with Diffusion Model*. SG Patent.
- *Method, Device, and Medium for Automatic Question-Answering*. CN Patent.