

# SHAOBIN ZHUANG

✉ hahahahaha@sjtu.edu.cn · ☎ (+86) 15260330756 ·

## EDUCATION

---

**Shanghai Jiao Tong University (SJTU)**, Shanghai, China 2023 – 2028 (*Expected*)  
*Ph.D student* in Information and Communication Engineering, advised by **Prof. Yali Wang**.

**Xiamen University**, Xiamen, China 2019 – 2023  
*B.S. in Automation major.* GPA: 3.80/4.00 Rank: 2/91

## SELF DESCRIPTION

---

My research interests include efficient model fine-tuning (TimeStep Master), controllable generation (Get In Video and Vlogger), visual generation pretraining (BitDance and Video-GPT), and unified multimodal large language model (WeGen, WeTok and UniWeTok).

## PAPERS

---

\* refers to the co-first author. † refers to the corresponding authors.

- [1] **Shaobin Zhuang\***, Yuang Ai\*, Jiaming Han\*, Weijia Mao, Xiaohui Li, Fangyikang Wang, Xiao Wang, Yan Li, Shanchuan Lin, Kun Xu, Zhenheng Yang, Huaibo Huang, Xiangyu Yue, Hao Chen†, Yali Wang†, “UniWeTok: An Unified Binary Tokenizer with Codebook Size  $2^{128}$  for Unified Multimodal Large Language Model.” **Technical Report**.
- [2] Yuang Ai\*, Jiaming Han\*, **Shaobin Zhuang\***, Weijia Mao, Xuefeng Hu, Ziyang Yang, Zhenheng Yang, Huaibo Huang†, Xiangyu Yue†, Hao Chen†, “BitDance: Scaling Autoregressive Generative Models with Binary Tokens.” **Technical Report**.
- [3] **Shaobin Zhuang**, Zhipeng Huang, Ying Zhuang, Fangyikang Wang, Canmiao Fu, Binxin Yang, Chong Sun, Chen Li, Yali Wang†, “Video-GPT via Next Clip Diffusion.” **ICLR2026**.
- [4] **Shaobin Zhuang**, Yiwei Guo, Canmiao Fu, Zhipeng Huang, Zeyue Tian, Fangyikang Wang, Ying Zhang, Chen Li, Yali Wang†, “WeTok: Powerful Discrete Tokenization for High-Fidelity Visual Reconstruction.” **ICLR2026**.
- [5] **Shaobin Zhuang**, Kunchang Li, Xinyuan Chen†, Yaohui Wang†, Ziwei Liu, Yu Qiao, Yali Wang†, “Vlogger: Make Your Dream A Vlog.” **CVPR2024**.
- [6] **Shaobin Zhuang\***, Yiwei Guo\*, Yanbo Ding\*, Kunchang Li, Xinyuan Chen†, Yaohui Wang†, Fangyikang Wang, Ying Zhang, Chen Li, Yali Wang†, “TimeStep Master: Asymmetrical Mixture of Timestep LoRA Experts for Versatile and Efficient Diffusion Models in Vision.” **ICML2025**.
- [7] **Shaobin Zhuang\***, Zhipeng Huang\*, Binxin Yang, Ying Zhang, Fangyikang Wang, Canmiao Fu, Chong Sun, Zheng-Jun Zha, Chen Li, Yali Wang†, “Get In Video: Add Anything You Want to the Video.” **Arxiv**.
- [8] Zhipeng Huang\*, **Shaobin Zhuang\***, Canmiao Fu, Binxin Yang, Ying Zhang, Chong Sun, Zhizheng Zhang†, Yali Wang†, Chen Li, Zheng-Jun Zha, “WeGen: A Unified Model for Interactive Multimodal Generation as We Chat.” **CVPR2025**.
- [9] Yiwei Guo, **Shaobin Zhuang**, Kunchang Li, Yu Qiao, Yali Wang†, “TransAgent: Transfer Vision-Language Foundation Models with Heterogeneous Agent Collaboration.” **NeurIPS2024**.
- [10] Yanbo Ding, **Shaobin Zhuang**, Kunchang Li, Zhengrong Yue, Yu Qiao, Yali Wang†, “MUSES: 3D-Controllable Image Generation via Multi-Modal Agent Collaboration.” **AAAI2024**.
- [11] Zhengrong Yue, **Shaobin Zhuang**, Kunchang Li, Yanbo Ding, Yali Wang†, “V-Stylist: Video Stylization via Collaboration and Reflection of MLLM Agents.” **CVPR2025**.
- [12] Xiaohui Li, **Shaobin Zhuang**, Shuo Cao, Yang Yang, Yuandong Pu, Qi Qin, Siqi Luo, Bin Fu, Yihao Liu, “LinearSR: Unlocking Linear Attention for Stable and Efficient Image Super-Resolution.” **ICLR2026**.
- [13] Xinyuan Chen\*, Yaohui Wang\*, Lingjun Zhang, **Shaobin Zhuang**, Xin Ma, Jiashuo Yu, Yali Wang, Dahua Lin†, Yu Qiao†, Ziwei Liu†, “Seine: Short-to-Long Video Diffusion Model for Generative Transition and Prediction.” **ICLR2024**.
- [14] Xiaohui Li\*, Yihao Liu\*†, Shuo Cao, Ziyang Chen, **Shaobin Zhuang**, Xiangyu Chen, Yinan He, Yi Wang, Yu Qiao, “DiffVSR: Revealing an Effective Recipe for Taming Robust Video Super-Resolution Against Complex Degradations.” **ICCV2025**.

## INTERNSHIPS

---

**OpenGVLab, Shanghai AI Lab** Shanghai, China Apr. 2023 – Jul. 2024  
Long Video Generation; General Image/Video Generation Model; Image/Video Generation Model Fine-Tuning.

**WeChat, Tencent** Shenzhen, China

Jul. 2024 – Sep. 2025

Visual Generative Pre-Training; Generative Understanding Unified Model; Video Editing.

**TikTok, ByteDance** Shanghai, China

Sep. 2025 – Present

Unified Multimodal Large Language Model.

## SERVICE

---

**Conference Reviewer:** CVPR, ICCV, ECCV, AAAI, ICLR, ICML, NeurIPS, ECCV, BMVC