# Xiangyu Zeng

XiangyuZeng2001@Outlook.com · +86 15264367990

https://lanxingxuan.github.io/ · Google Scholar Link

## Education Background

**Shandong University** Sep. 2020 – Jun. 2024

*B.Eng.* *Artificial Intelligence*

*Course GPA: 92.1/100* *Class Rank: 3/52* *Comprehensive GPA: 102.1/110* *Comprehensive Rank: 1/52*

- *Recipient of National Scholarship, Shandong University Dean's Scholarship, First-Class Academic Scholarship (×2), and Outstanding Graduate of Shandong University*

**Nanjing University** Sep. 2024 – Present

*Ph.D. Candidate* *Computer Science and Technology*

- *Recipient of Nanjing University Presidential Special Scholarship*

## Competition Experience

- 23rd CCF CSP Certification: 370 points (Top 1%) Sep. 19, 2021
- ACM ICPC Asia Regional Contest: Silver Medal Nov. 14, 2021
- CCSP (East China): Bronze Medal Dec. 13, 2021
- ACM ICPC Asia Regional Contest: Bronze Medal Apr. 3, 2022
- CUMCM (Shandong): First Prize Sep. 18, 2022

## Internship Experience

**Shanghai AI Laboratory** Jan. 2024 – Present

*Research Intern* *OpenGVLab*

- *Focused on pretraining and fine-tuning of multimodal video foundation models, achieved progress in long video temporal modeling and contributed to a paper accepted at ICLR 2025; participated in the InternVideo2.5 project.*

**Huawei** Jan. 2025 – Jun. 2025

*Collaborative Project* *Noah's Ark Lab*

- *Worked on online video understanding, with emphasis on memory mechanisms and real-time perception for applications such as autonomous driving; contributed to a paper submitted to NeurIPS 2025.*

## Research Experience

- **TimeSuite: Improving MLLMs for Long Video Understanding via Grounded Tuning**
  - Accepted to **ICLR 2025**, first author
  - *Without relying on any external expert decoder, TimeSuite can achieve expert-level performance in grounding tasks while maintaining considerable generalization QA capability and strong zero-shot capabilities.*
  - *The introduction of grounding tasks enhances the comprehensive understanding of long videos. We validated the feasibility of enhancing MLLM's comprehensive capabilities by integrating expert tasks.*

- **Adaptive Edge-Aware Semantic Interaction Network for Salient Object Detection in Optical Remote Sensing Images**
  - Accepted to **TGRS 2023**, first author
  - *Developed AESINet, a novel edge-aware semantic interaction network for salient object detection in optical remote sensing images, incorporating LDAM for unsupervised edge enhancement, MFEM for handling*

*varying object scales, and DSIM for robust detection in cluttered scenes.*
*- Achieved superior performance over 14 state-of-the-art methods across three benchmark datasets.*

- **StreamForest: Efficient Online Video Understanding with Persistent Event Memory**
  - Submitted to **NIPS 2025**, co-first author (rank1)
  *- Developed StreamForest, a novel architecture for real-time streaming video understanding, featuring a Persistent Event Memory Forest for efficient long-term memory and a Fine-grained Spatiotemporal Window for enhanced short-term perception.*
  *- Introduced OnlineIT, an instruction-tuning dataset, and ODV-Bench, a benchmark for autonomous driving scenarios.*
  *- Achieved state-of-the-art performance across multiple benchmarks while maintaining high accuracy under extreme visual token compression.*

***Participating Works***

- **Online Video Understanding: OVBench and VideoChat-Online**
  - Accepted to **CVPR 2025**

- **Task Preference Optimization: Improving Multimodal Large Language Models with Vision Task Alignment**
  - Accepted to **CVPR 2025**

- **Make Your Training Flexible: Towards Deployment-Efficient Video Models**
  - Accepted to **ICCV 2025**

- **VTTS: Visual Test-Time Scaling to Reinforce Multimodal Reasoning by Iterative Perception**
  - Submitted to **NIPS 2025**

- **Learning Goal-Oriented Language-Guided Navigation with Self-Improving Demonstrations at Scale**
  - Submitted to **NIPS 2025**

- **VideoChat-Flash: Hierarchical Compression for Long-Context Video Modeling**
  - Github **400+ Star**

- **InternVideo2.5: Empowering Video MLLMs with Long and Rich Context Modeling**
  - Technical Report

- **VideoChat-R1: Enhancing Spatio-Temporal Perception via Reinforcement Fine-Tuning**
  - Technical Report

*Last updated:* ***June 26, 2025***