# Ziyang Wang

ziyangw@cs.unc.edu | ziyangw2000.github.io/ | Ziyang412 | ziyang-wang-882188203

## Education

**University of North Carolina at Chapel Hill** — Chapel Hill, NC
Ph.D. Candidate in Computer Science, advised by Prof. Mohit Bansal. — Aug. 2022 - Exp. May. 2027

**University of North Carolina at Chapel Hill** — Chapel Hill, NC
M.S. in Computer Science, during Ph.D. study. — Aug. 2022 - Exp. Dec. 2024

**University of Electronic Science and Technology of China** — Chengdu, China
Bachelor of Engineer, Major: Software Engineering. Overall GPA: 3.92. — Sept. 2018 - Jun. 2022

## Research Interest

My main research interest is video-language understanding and multimodal AI. Particularly, I am interested in the challenge of reasoning over long and complex videos.

## Work Experience

**UNC-NLP** — Chapel Hill, NC
Graduate Research Assistant — Aug. 2022 - PRESENT

- Advised by Prof. Mohit Bansal, also work closely with Prof. Gedas Bertasius
- Built models to tackle challenging video understanding tasks like retrieval and long video understanding.
- Published several academic papers in CV/NLP conferences including ICCV, EACL, EMNLP.

**Salesforce AI Research** — Palo Alto, CA
Research Intern — May. 2025 - Feb. 2026

- Advised by Dr. Juan Carlos Niebles (manager), Dr. Michael S. Ryoo, Dr. Junnan Li and Dr. Honglu Zhou.
- Work on the agentic long video understanding.
- Finish the project “Active Video Perception: Iterative Evidence Seeking for Agentic Long Video Understanding”.

**Meta FAIR Perception** — Menlo Park, CA
Research Intern — May. 2024 - Dec. 2024

- Advised by Dr. Ronghang Hu (manager), Dr. Christoph Feichtenhofer, Dr. Po-Yao (Bernie) Huang and Dr. Daniel Bolya.
- Work on the next generation video foundation encoders.
- Work on long video encoder training and zero-shot application for short video encoder on long video tasks.

**Amazon Alexa AI** — Seattle, WA
Applied Scientist Intern — May. 2023 - Oct. 2023

- Advised by Dr. Heba Elfardy (manager), Dr. Kevin Small, Dr. Markus Dreyer
- Improved the multimodal large language model’s ability to retrieve both visual and textual outputs.
- Published the paper named ”Unified Embeddings for Multimodal Retrieval via Frozen LLMs” in EACL2024 (findings).

**Tsinghua University** — Beijing, China
Research Intern, advised by Prof. Jingjing Liu — Oct. 2021 - May. 2022

- Worked on a project to improve the visual question answering ability of the MLLMs.
- Also work on the project that focuses on Vehicle-Infrastructure Cooperative 3D Object Detection.

## Papers

2026 **EgoMemReason: A Memory-Driven Reasoning Benchmark for Long-Horizon Egocentric Video Understanding**
**Ziyang Wang***, Yue Zhang*, Shoubin Yu, Ce Zhang, Zengqi Zhao, Jaehong Yoon, Hyunji Lee, Gedas Bertasius, Mohit Bansal
**Preprint**

2026 **Multimodal Fact-Level Attribution for Verifiable Reasoning** — Seoul, South Korea
David Wan, Han Wang, **Ziyang Wang**, Elias Stengel-Eskin, Hyunji Lee, Mohit Bansal

**ICML 2026**

2026 **Active Video Perception: Iterative Evidence Seeking for Agentic Long Video Understanding** — Denver, USA
**Ziyang Wang**, Honglu Zhou, Shijie Wang, Junnan Li, Caiming Xiong, Silvio Savarese, Mohit Bansal, Michael S. Ryoo, Juan Carlos Niebles
**CVPR 2026 Findings**

2026 **SiLVR: A Simple Language-based Video Reasoning Framework**
Ce Zhang*, Yan-Bo Lin*, **Ziyang Wang**, Mohit Bansal, Gedas Bertasius
**TMLR 2026**

2026 **TimeRefine: Temporal Grounding with Time Refining Video LLM** — Tucson, USA
Xizi Wang, Feng Cheng, **Ziyang Wang**, Huiyu Wang, Md Mohaiminul Islam, Lorenzo Torresani, Mohit Bansal, Gedas Bertasius, David Crandall
**WACV 2026**

2025 **Video-RTS: Rethinking Reinforcement Learning and Test-Time Scaling for Efficient and Enhanced Video Reasoning** — Suzhou, China
**Ziyang Wang***, Jaehong Yoon*, Shoubin Yu, Md Mohaiminul Islam, Gedas Bertasius, Mohit Bansal
**EMNLP 2025 (Main)**

2025 **MEXA: Towards General Multimodal Reasoning with Dynamic Multi-Expert Aggregation** — Suzhou, China
Shoubin Yu*, Yue Zhang*, **Ziyang Wang**, Jaehong Yoon, Mohit Bansal
**EMNLP 2025 (Findings)**

2025 **VideoTree: Adaptive Tree-based Video Representation for LLM Reasoning on Long Videos** — Nashville, USA
**Ziyang Wang***, Shoubin Yu*, Elias Stengel-Eskin*, Jaehong Yoon, Feng Cheng, Gedas Bertasius, Mohit Bansal
**CVPR 2025**

2025 **DAM: Dynamic Adapter Merging for Continual Video QA Learning** — Tucson, USA
Feng Cheng*, **Ziyang Wang***, Yi-Lin Sung, Yan-Bo Lin, Mohit Bansal, Gedas Bertasius
**WACV 2025**

2024 **Unified Embeddings for Multimodal Retrieval via Frozen LLMs** — Malta
**Ziyang Wang**, Heba Elfardy, Markus Dreyer, Kevin Small, Mohit Bansal
**EACL 2024 (Findings)**

2024 **A Simple LLM Framework for Long-Range Video Question-Answering** — Miami, USA
Ce Zhang, Taixi Lu, Md Mohaiminul Islam, **Ziyang Wang**, Shoubin Yu, Mohit Bansal, Gedas Bertasius
**EMNLP 2024 (Main)**

2023 **Unified Coarse-to-Fine Alignment for Video-Text Retrieval** — Paris, France
**Ziyang Wang**, Yi-Lin Sung, Feng Cheng, Gedas Bertasius, Mohit Bansal
**ICCV 2023**

2022 **Language-Augmented Pixel Embedding for Generalized Zero-Shot Learning** — Journal
**Ziyang Wang**, Yunhao Gou, Jingjing Li, Lei Zhu, Heng Tao Shen
**TCSVT 2022**

2021 **Region Semantically Aligned Network for Zero-Shot Learning** — Virtual
**Ziyang Wang***, Yunhao Gou*, Jingjing Li, Yu Zhang, Yang Yang
**CIKM 2021 (long oral)**

## Service

### Top CV/NLP conferences

Reviewer

- Engaged in the peer-review process in ECCV, ICLR, CVPR, COLM, WACV, ACL Rolling Reviews, IJCV.

**T4V workshop @ CVPR 2023 & 2024 & 2025 & 2026**

Organizer and Program Commitee

- Co-organize the Transformers for Vision (T4V) workshop @ CVPR 2025 and 2026.
- Worked as a program committee member to help the reviewing process and other aspect of the workshop (2023 and 2024).

## Honor

**Bloomberg Data Science Ph.D. Fellowship Award** New York City, NY

Bloomberg L.P. March. 2026

- Recognized for my research in long context multimodal understanding and multimodal agents.