Ce Zhang

cezhang@cs.unc.edu | ceezh.github.io | twitter.com/cezhhh | Google Scholar

Education

UNC-Chapel Hill — Chapel Hill, NC, US
Ph.D. student in Computer Science — *Aug. 2023 – Present*
- Advisor: Prof. Gedas Bertasius

Brown University — Providence, RI, US
M.S. in Computer Science — *Aug. 2021 – May 2023*
- Advisor: Prof. Chen Sun

Southeast University — Nanjing, Jiangsu, China
B.S. in Computer Science — *Aug. 2016 – June 2020*

Research Interests

I'm broadly interested in computer vision and multimodal AI, particularly **long-video understanding and reasoning**. My research focuses on developing intelligent agents that can efficiently perceive, search, and reason over long videos. Recently, I have been especially interested in **agentic video understanding, long-term multimodal memory, and adaptive information retrieval**, aiming to enable AI systems to effectively understand and reason over long-context multimodal information.

Academic Experience

Research Assistant — Aug. 2023 – Present
UNC-Chapel Hill, Advisor: Prof. Gedas Bertasius — *Chapel Hill, NC*
- Proposed DeLoVi, which retrieves evidence from ultra-long videos with pretrained decision models that score relevance without generating text. Achieved SOTA on EgoLifeQA, EgoMemReason, and MM-Lifelong; an evolved variant stays comparable with 34x fewer input tokens. (**arXiv 2026**) paper link
- Proposed VideoTreeSearch (VTS), a self-correcting agentic framework for grounded long-video QA. VTS represents a long video as an adaptive temporal tree and navigates it via discrete actions (zoom_in, zoom_out, shift, answer), enabling explicit backtracking and recovery. VTS outperforms prior agentic methods on grounded LVQA benchmarks (CG-Bench, Haystack-LVBench, Haystack-Ego4D) and transfers to general long-video QA (Video-MME, MLVU, LVBench). (**NeurIPS 2026**) paper link
- Proposed SiLVR, a training-free language-based video reasoning framework that transforms video and speech into language and applies LLMs for reasoning. Achieved SOTA on benchmarks such as Video-MMMU and EgoLife; won 1st place in Track 1B of the CVPR 2025 Multimodal Video Agent Challenge. (**TMLR 2026**) paper link
- Proposed LLoVi, a framework for long-range video QA that decomposes the task into short-term visual captioning and long-range language modeling. Achieved SOTA zero-shot video QA performance on EgoSchema, NExT-QA, IntentQA, and NExT-GQA. (**EMNLP 2024**) paper link

Research Assistant — Apr. 2022 – May 2023
Brown University, Advisor: Prof. Chen Sun — *Providence, RI*
- Represented videos with discretized action labels and utilized LLMs for reasoning. Achieved SOTA long-term action anticipation performance on Ego4D, EPIC-Kitchens-55, and EGTEA Gaze+. (**ICLR 2024**) paper link
- Extracted task-specific object-centric representations from pretrained models (e.g., GLIP). Utilized object-centric representations for long-term action anticipation. Achieved competitive results on Ego4D, 50Salads, and EGTEA Gaze+. (**WACV 2024**) paper link
- Condensed expert trajectory demonstrations into useful representations for policy learning. Achieved competitive performance on AntMaze, FrankaKitchen, and Locomotion. (**NeurIPS 2023**) paper link

Industry Experience

Research Scientist Intern — May 2026 – Aug. 2026
Meta FAIR, Mentor: Ramakanth Pasunuru — *Seattle, WA*
- Worked on parametric memory for long-document understanding. Conducted empirical analyses comparing explicit and parametric memory, and designed a pipeline that integrates both approaches.

Research Scientist Intern May 2025 – Aug. 2025

Google DeepMind, Mentor: Dahun Kim *Mountain View, CA*

- Worked on long-form video temporal grounding and question answering. Designed and implemented a training-free, multi-tier coarse-to-fine framework that significantly improves the temporal grounding ability of multimodal large language models (MLLMs), achieving both higher effectiveness and efficiency.

Research Scientist Intern May 2024 – Dec. 2024

Meta FAIR, Mentor: Satwik Kottur *Menlo Park, CA*

- Improved the long-horizon video-based planning ability of large VLMs. Designed multiple auxiliary tasks (e.g., goal prediction, state prediction) and trained the model on all tasks jointly. Leveraged multi-token prediction to model the structured action space. Achieved SOTA performance on COIN and CrossTask, and competitive performance on Ego4D. (**WACV 2026**) paper link

Publications

* denotes equal contribution

- *Ce Zhang, Baiqi Li, Seongsu Ha, Gedas Bertasius*
  Searching Ultra-Long Videos with Decision Models **(arXiv 2026)** paper link
- *Shuai Liu, Hongming Guo, Weiyang Jin, Kairui Hu, Shulin Tian, Yuhao Dong, Jinming Wu, Kaichen Zhang, Yujin Tang, Xinyang Han, Wenhao Chai, Yuqian Fu, Ce Zhang, Yuanhan Zhang, Lingdong Kong, Li Siyao, Jingkang Yang, Gedas Bertasius, Manling Li, Jiajun Wu, Mengdi Wang, Dawn Song, Ziwei Liu*
  Always-On Personal Intelligence: A Roadmap Towards Continuing, Self-Evolving Loops **(Preprints 2026)** paper link
- *Ce Zhang, Ziyang Wang, Yulu Pan, Oluwatumininu Oguntola, Pranav Wagh, Qiyu Wu, Hiromi Wakaki, Mohit Bansal, Gedas Bertasius*
  Searching Videos as Trees: Self-Correcting Agents for Grounded Long Video QA **(NeurIPS 2026)** paper link
- *Baiqi Li, Ce Zhang, Yu Fang, Yue Yang, Shangzhe Li, Mingyu Ding, Gedas Bertasius*
  WatchAct: A Benchmark for Behavior-Grounded Robot Manipulation **(arXiv 2026)** paper link
- *Ziyang Wang*, Yue Zhang*, Shoubin Yu, Ce Zhang, Zengqi Zhao, Jaehong Yoon, Hyunji Lee, Gedas Bertasius, Mohit Bansal*
  EgoMemReason: A Memory-Driven Reasoning Benchmark for Long-Horizon Egocentric Video Understanding **(COLM 2026)** paper link
- *Baiqi Li, Kangyi Zhao, Ce Zhang, Chancharik Mitra, Jean de Dieu Nyandwi, Gedas Bertasius*
  TimeBlind: A Spatio-Temporal Compositionality Benchmark for Video LLMs **(EMNLP 2026)** paper link
- *Ce Zhang, Yan-Bo Lin, Ziyang Wang, Mohit Bansal, Gedas Bertasius*
  SiLVR: A Simple Language-based Video Reasoning Framework **(TMLR 2026)** paper link
- *Ce Zhang, Yale Song, Ruta Desai, Michael Louis Iuzzolino, Joseph Tighe, Gedas Bertasius, Satwik Kottur*
  Enhancing Visual Planning with Auxiliary Tasks and Multi-token Prediction **(WACV 2026)** paper link
- *Yulu Pan, Ce Zhang, Gedas Bertasius*
  BASKET: A Large-Scale Video Dataset for Fine-Grained Skill Estimation **(CVPR 2025)** paper link
- *Ce Zhang*, Taixi Lu*, Md Mohaiminul Islam, Ziyang Wang, Shoubin Yu, Mohit Bansal, Gedas Bertasius*
  A Simple LLM Framework for Long-Range Video Question-Answering **(EMNLP 2024)** paper link
- *Qi Zhao*, Shijie Wang*, Ce Zhang, Changcheng Fu, Minh Quan Do, Nakul Agarwal, Kwonjoon Lee, Chen Sun*
  AntGPT: Can Large Language Models Help Long-term Action Anticipation from Videos? **(ICLR 2024)** paper link
- *Ce Zhang*, Changcheng Fu*, Shijie Wang, Nakul Agarwal, Kwonjoon Lee, Chiho Choi, Chen Sun*
  Object-centric Video Representation for Long-term Action Anticipation **(WACV 2024)** paper link
- *Zilai Zeng, Ce Zhang, Shijie Wang, Chen Sun*
  Goal-Conditioned Predictive Coding as an Implicit Planner for Offline Reinforcement Learning **(NeurIPS 2023)** paper link

Service

Conference Reviewer: NeurIPS 2026, CVPR 2026, NeurIPS 2025, ICCV 2025, CVPR 2025, ECCV 2024 (**Outstanding Reviewer Award**), ACL Rolling Review (May 2025, Dec. 2024, June 2024)
Journal Reviewer: TPAMI, TCSVT
Workshop Organizer: LOVEU @ CVPR 2026, T4V @ CVPR 2026, T4V @ CVPR 2025, T4V @ CVPR 2024