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, GPA: 4.0 / 4.0 — *Aug. 2023 – Present*
- Advisor: Prof. Gedas Bertasius

Brown University — Providence, RI, US
M.S. in Computer Science, GPA: 3.75 / 4.0 — *Aug. 2021 – May 2023*
- Advisor: Prof. Chen Sun

Southeast University — Nanjing, Jiangsu, China
B.S. in Computer Science, GPA: 3.75 / 4.0, Rank: Top 10% — *Aug. 2016 – Jun. 2020*

Research Interest

I'm broadly interested in Computer Vision, Multimodal learning and Robotics. Currently, I'm mainly working on **video understanding**, with a focus on leveraging **foundation models** (LLMs, VLMs, etc.) to solve multiple video understanding tasks.

Academic Experience

Research Assistant — Aug. 2023 – Present
UNC-Chapel Hill, Advisor: Prof. Gedas Bertasius — *Chapel Hill, NC*
- Proposed SiLVR, a training-free language-based video reasoning framework that transforms video and speech into language and applies LLMs for video reasoning. SiLVR achieves state-of-the-art performance on multiple video reasoning benchmarks such as Video-MMMU and EgoLife. (**arXiv 2025)** paper link
- Curated a large-scale video dataset for fine-grained basketball skill estimation with over 4000 hours and 32K basketball players. Benchmarked multiple state-of-the-art video models on the proposed dataset. (**CVPR 2025)** paper link
- Proposed a framework for long-range videoQA by decomposing videoQA into short-term visual captioning and long-range language modeling. Achieved SOTA zero-shot videoQA 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*
- Represent the video 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 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*
- Augmented 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, CrossTask and Ego4D. (**arXiv, 2025**) link

Machine Learning Engineer Intern Mar. 2021 – July 2021
QCraft *Beijing, China*

- Worked on 3D Multi-Object Tracking for vehicles and pedestrians by fusing 2D and 3D appearance feature (from ResNet and MLP) and 2D motion feature (from Kalman Filter).

Machine Learning Engineer Intern Feb. 2020 – July 2020
Momenta *Suzhou, China*

- Generated pseudo labels for 2D facial key points by fusing 2D detection from multi-view cameras. Bootstrapped the model by training on pseudo labels.

Publication

- *Ce Zhang*, Yan-Bo Lin*, Ziyang Wang, Mohit Bansal, Gedas Bertasius.*
  SiLVR: A Simple Language-based Video Reasoning Framework **(arXiv 2025)** 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 **(arXiv 2025)** 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

Services

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

Technical Skills

Languages: Python, C/C++, Go
Deep Learning Frameworks: PyTorch, TensorFlow
Other: Slurm, Git, Docker