

Gedas Bertasius

<https://gedasbertasius.com> | gedas@cs.unc.edu | [Google Scholar Profile](#)

Last updated: September 2026

Appointments

University of North Carolina at Chapel Hill

Assistant Professor (tenure-track), Department of Computer Science 2021 – present

Facebook AI Research (FAIR)

Postdoctoral Researcher. Mentor: Lorenzo Torresani. 2019 – 2021

Facebook AI Research (FAIR)

Research Intern. Mentors: Lorenzo Torresani, Christoph Feichtenhofer, Du Tran. 2018

Research Overview

The long-term goal of my research is to develop machines that understand human behavior from video. My group builds models that perceive long, multimodal recordings (video perception), infer the goals, skills, and strategies that shape behavior and anticipate what may happen next (behavioral reasoning), and turn these inferences into coaching, task assistance, and robotic action (guiding action). Our current focus, Strategic Video Intelligence, unifies perception, causal reasoning, simulation, and agentic synthesis in dynamic multi-agent settings, where inferences about goals, decisions, and consequences are testable.

Highlights

- **Over 10,000 citations** on Google Scholar. 59 refereed publications (41 since joining UNC in 2021), including papers at ICML, NeurIPS, CVPR, ICCV, ECCV, ACL, EMNLP, COLM, and RSS.
- **TimeSformer** (ICML 2021) was the first video architecture built exclusively on self-attention. It is among the five most cited ICML 2021 papers, with over 4,000 citations and over 1.5 million model downloads.
- **NSF CAREER Award** (2026) for perceptual AI agents that learn to see, reason, and act from video.
- **CVPR 2020 Best Paper Award Finalist** for MaskProp, a mask propagation approach to video instance segmentation that ranked first on the YouTube-VIS leaderboard at publication and won the CVPR 2020 EPIC-Kitchens detection challenge.
- **Egocentric Vision Distinguished Paper Award** (2025) for Video ReCap, a recursive captioning model for hour-long video, and the accompanying Ego4D-HCap dataset.
- Approximately **\$3.4 million** in research support attributable to my role since joining UNC in 2021.
- Co-founder and lead organizer of the **CVPR Transformers for Vision workshop**, 2022 – 2026.

Honors and Awards

NSF CAREER Award	2026
Sony Focused Research Award	2024, 2025
Egocentric Vision Distinguished Paper Award	2025
1st Place, CVPR EgoSchema Challenge	2025
1st Place, CVPR Multi-Discipline Lecture Understanding Challenge	2025
ECIR Best Student Paper Award	2023
Meta Research Award	2022, 2023
UNC Junior Faculty Development Award	2022
Excellence in Teaching Award	2022
Sony Faculty Innovation Award	2021, 2022
CVPR Best Paper Award Finalist	2020
1st Place, CVPR EPIC-Kitchens Object Detection Challenge	2020

Education

University of Pennsylvania

PhD in Computer Science. Advisor: Jianbo Shi.

Philadelphia, PA, USA

2014 – 2019

Dartmouth College

Bachelor of Arts in Computer Science. Advisor: Lorenzo Torresani.

Hanover, NH, USA

2010 – 2014

Selected Publications

(my graduate student advisees are underlined; my undergraduate advisees are denoted by *)

1. Is Space-Time Attention All You Need for Video Understanding?
G. Bertasius, H. Wang, L. Torresani
In International Conference on Machine Learning (ICML) 2021. [\[pdf\]](#)
Top-5 Most Cited ICML 2021 Paper (>4,000 citations, >1.5M model downloads)
2. SVI-Bench: A Dynamic Microworld for Strategic Video Intelligence
Y. Pan, H. Yi, S. Ha, M. Islam, B. Zhang*, L. Torresani, **G. Bertasius**
In European Conference on Computer Vision (ECCV) 2026. [\[pdf\]](#)
3. Video ReCap: Recursive Captioning of Hour-Long Videos
M. Islam, N. Ho*, X. Yang, T. Nagarajan, L. Torresani, **G. Bertasius**
In Computer Vision and Pattern Recognition (CVPR) 2024. [\[pdf\]](#)
Egocentric Vision Distinguished Paper Award
4. ExAct: A Video-Language Benchmark for Expert Action Analysis
H. Yi, Y. Pan, F. He*, X. Liu*, B. Zhang*, O. Oguntola*, **G. Bertasius**
In Neural Information Processing Systems (NeurIPS), Datasets and Benchmarks Track 2025. [\[pdf\]](#)
5. SiLVR: A Simple Language-based Video Reasoning Framework
C. Zhang, Y.-B. Lin, Z. Wang, M. Bansal, **G. Bertasius**
In Transactions on Machine Learning Research (TMLR) 2026. [\[pdf\]](#)
1st Place, CVPR Multi-Discipline Lecture Understanding Challenge
6. Classifying, Segmenting, and Tracking Object Instances in Video with Mask Propagation
G. Bertasius, L. Torresani
In Computer Vision and Pattern Recognition (CVPR) 2020 (Oral). [\[pdf\]](#)
CVPR Best Paper Award Finalist

Research Funding

Total awarded research support attributable to my role: approximately \$3.4 million across 16 awards. Sole PI on 14.

16. **Singapore Technologies Engineering: Agentic Video Models for Efficient Analysis of Long Multimodal Videos** (09/01/2026 – 03/01/2029).
Sole PI. Awarded Amount: \$738K.
15. **National Science Foundation (NSF) CAREER Award: CAREER: Learning to See, Reason, and Act: Language-Grounded Video Intelligence for Perceptual AI Agents** (09/01/2026 – 08/31/2031).
Sole PI. Awarded Amount: \$584K.
14. **Sony Focused Research Award: PlayFormer: A Hierarchical Multimodal Foundation Model for Understanding and Assisting Strategic Gameplay** (09/01/2026 – 08/31/2027).
Sole PI. Awarded Amount: \$150K.
13. **Laboratory for Analytic Sciences (LAS): The AI Video Analyst: Agentic, Interpretable and Multimodal Framework for Deep Video Analysis** (01/01/2026 – 12/31/2026).
Sole PI. Awarded Amount: \$170K.

12. **Sony Focused Research Award:** *A Hierarchical Video Model and Dataset for Hour-Long Multimodal Video Understanding* (09/01/2025 – 08/31/2026).
Sole PI. Awarded Amount: \$150K.
11. **Laboratory for Analytic Sciences (LAS):** *A Memory-augmented Audiovisual Transformer for Object and Activity Localization in Long Video* (01/01/2025 – 12/31/2025).
Sole PI. Awarded Amount: \$160K.
10. **Laboratory for Analytic Sciences (LAS):** *Continually Improving Video Models for Text-based Analysis of Long Videos* (01/01/2024 – 12/31/2024).
Sole PI. Awarded Amount: \$157K.
9. **Meta Research Award:** *Proficiency Estimation in Physical Human Activities using Egocentric Cameras* (06/01/2023).
Sole PI. Awarded Amount: \$80K.
8. **Office of Naval Research (ONR):** *Generalizable and Transparent Vision-Language Modeling* (04/01/2023 – 03/31/2027).
Co-PI. PI: Mohit Bansal. Awarded Amount: \$900K (total); attributable share: \$450K.
7. **Sony Faculty Innovation Award:** *State-Space Diffusion Models for Text-Guided Audiovisual Video Generation* (08/01/2023 – 08/01/2024).
Sole PI. Awarded Amount: \$100K.
6. **UNC Junior Faculty Development Award:** *Computer Vision Models for Improving Physical Therapy* (01/01/2023 – 01/01/2024).
Sole PI. Awarded Amount: \$10K.
5. **Laboratory for Analytic Sciences (LAS):** *Parameter-Efficient Transfer Learning for Large-Scale Text-to-Video Retrieval* (01/01/2023 – 12/31/2023).
Sole PI. Awarded Amount: \$125K.
4. **Amazon Research Award:** *Zero-Shot Person Search Using Unlabeled Videos* (08/01/2022).
Sole PI. Awarded Amount: \$20K.
3. **Meta Research Award:** *Human Activity Recognition using Egocentric and Exocentric Cameras* (07/01/2022).
Sole PI. Awarded Amount: \$100K.
2. **Sony Faculty Innovation Award:** *Multimodal Video Representation Learning for Video-and-Language Tasks* (08/01/2022 – 08/01/2023).
Sole PI. Awarded Amount: \$100K.
1. **NSF-NIH Smart and Connected Health (SCH):** *An Augmented Reality Neurorehabilitation System for Monitoring and Management of Motor Symptoms of Parkinson’s Disease* (08/01/2022 – 08/01/2026).
Co-PI. PI: Henry Fuchs. Other Co-PIs: Mike Lewek, Dan Szafir, Danielle Szafir. Awarded Amount: \$1.2M (total); attributable share: \$300K.

Preprints

(my graduate student advisees are underlined; my undergraduate advisees are denoted by *)

4. Searching Videos as Trees: Self-Correcting Agents for Grounded Long Video QA
C. Zhang, Z. Wang, Y. Pan, O. Oguntola*, P. Wagh*, Q. Wu, H. Wakaki, M. Bansal, **G. Bertasius**
arXiv preprint, 2026. [\[pdf\]](#)
3. WatchAct: A Benchmark for Behavior-Grounded Robot Manipulation
B. Li, C. Zhang, Y. Fang, Y. Yang, S. Li, M. Ding, **G. Bertasius**
arXiv preprint, 2026. [\[pdf\]](#)
2. LiLo-VLA: Compositional Long-Horizon Manipulation via Linked Object-Centric Policies
Y. Yang, S. Cheng, Y. Fang, H. Bharadhwaj, M. Ding, **G. Bertasius**, D. Szafir
arXiv preprint, 2026. [\[pdf\]](#)

1. V2M-Zero: Zero-Pair Time-Aligned Video-to-Music Generation
Y.-B. Lin, J. Casebeer, L. Mai, A. Mahapatra, **G. Bertasius**, N. Bryan
arXiv preprint, 2026. [\[pdf\]](#)

Peer-Reviewed Publications

(my graduate student advisees are underlined; my undergraduate advisees are denoted by *)

59. TimeBlind: A Spatio-Temporal Compositionality Benchmark for Video LLMs
B. Li, K. Zhao, C. Zhang, C. Mitra, J. Nyandwi, **G. Bertasius**
In Empirical Methods in Natural Language Processing (**EMNLP**) 2026. [\[pdf\]](#)
58. SVI-Bench: A Dynamic Microworld for Strategic Video Intelligence
Y. Pan, H. Yi, S. Ha, M. Islam, B. Zhang*, L. Torresani, **G. Bertasius**
In European Conference on Computer Vision (**ECCV**) 2026. [\[pdf\]](#)
57. EgoMemReason: A Memory-Driven Reasoning Benchmark for Long-Horizon Egocentric Video Understanding
Z. Wang, Y. Zhang, S. Yu, C. Zhang, Z. Zhao, J. Yoon, H. Lee, **G. Bertasius**, M. Bansal
In Conference on Language Modeling (**COLM**) 2026. [\[pdf\]](#)
56. DocSLM: A Small Vision-Language Model for Long Multimodal Document Understanding
T. Hannan, D. Mallios, P. Pathak, F. Sardari, T. Seidl, **G. Bertasius**, M. Fayyaz, S. Sengupta
In Computer Vision and Pattern Recognition (**CVPR Findings**) 2026. [\[pdf\]](#)
55. SiLVR: A Simple Language-based Video Reasoning Framework
C. Zhang, Y.-B. Lin, Z. Wang, M. Bansal, **G. Bertasius**
In Transactions on Machine Learning Research (**TMLR**) 2026. [\[pdf\]](#)
1st Place, CVPR Multi-Discipline Lecture Understanding Challenge
54. Zero-Shot Audio-Visual Editing via Cross-Modal Delta Denoising
Y.-B. Lin, K. Lin, Z. Yang, L. Li, J. Wang, C. Lin, X. Wang, **G. Bertasius**, L. Wang
In Winter Conference on Applications of Computer Vision (**WACV**) 2026 (Oral). [\[pdf\]](#)
53. TimeRefine: Temporal Grounding with Time Refining Video LLM
X. Wang, F. Cheng, Z. Wang, H. Wang, M. Islam, L. Torresani, M. Bansal, **G. Bertasius**, D. Crandall
In Winter Conference on Applications of Computer Vision (**WACV**) 2026. [\[pdf\]](#)
52. Enhancing Visual Planning with Auxiliary Tasks and Multi-token Prediction
C. Zhang, Y. Song, R. Desai, M. Iuzzolino, J. Tighe, **G. Bertasius**, S. Kottur
In Winter Conference on Applications of Computer Vision (**WACV**) 2026. [\[pdf\]](#)
51. ReAgent-V: A Reward-Driven Multi-Agent Framework for Video Understanding
Y. Zhou, Y. He, Y. Su, S. Han, J. Jang, **G. Bertasius**, M. Bansal, H. Yao
In Neural Information Processing Systems (**NeurIPS**) 2025. [\[pdf\]](#)
50. ExAct: A Video-Language Benchmark for Expert Action Analysis
H. Yi, Y. Pan, F. He*, X. Liu*, B. Zhang*, O. Oguntola*, **G. Bertasius**
In Neural Information Processing Systems (**NeurIPS**), Datasets and Benchmarks Track 2025. [\[pdf\]](#)
49. Video-RTS: Rethinking Reinforcement Learning and Test-Time Scaling for Efficient and Enhanced Video Reasoning
Z. Wang, J. Yoon, S. Yu, M. Islam, **G. Bertasius**, M. Bansal
In Empirical Methods in Natural Language Processing (**EMNLP**) 2025. [\[pdf\]](#)
48. BOSS: Benchmark for Observation Space Shift in Long-Horizon Task
Y. Yang, L. Zhao, M. Ding, **G. Bertasius**, D. Szafir
In IEEE Robotics and Automation Letters (**RA-L**) 2025. [\[pdf\]](#)
47. ReBot: Scaling Robot Learning with Real-to-Sim-to-Real Robotic Video Synthesis
Y. Fang, Y. Yang, X. Zhu, K. Zheng, **G. Bertasius**, D. Szafir, M. Ding
In International Conference on Intelligent Robots and Systems (**IROS**) 2025. [\[pdf\]](#)

46. BIMBA: Selective-Scan Compression for Long-Range Video Question Answering
M. Islam, T. Nagarajan, H. Wang, **G. Bertasius**, L. Torresani
 In Computer Vision and Pattern Recognition (**CVPR**) 2025. [\[pdf\]](#)
1st Place, CVPR EgoSchema Challenge
45. BASKET: A Large-Scale Video Dataset for Fine-Grained Skill Estimation
Y. Pan, C. Zhang, **G. Bertasius**
 In Computer Vision and Pattern Recognition (**CVPR**) 2025. [\[pdf\]](#)
44. VideoTree: Adaptive Tree-based Video Representation for LLM Reasoning on Long Videos
 Z. Wang, S. Yu, E. Stengel-Eskin, J. Yoon, F. Cheng, **G. Bertasius**, M. Bansal
 In Computer Vision and Pattern Recognition (**CVPR**) 2025. [\[pdf\]](#)
43. ReVisionLLM: Recursive Vision-Language Model for Temporal Grounding in Hour-Long Videos
 T. Hannan, M. Islam, J. Gu, T. Seidl, **G. Bertasius**
 In Computer Vision and Pattern Recognition (**CVPR**) 2025. [\[pdf\]](#)
42. VMAs: Video-to-Music Generation via Semantic Alignment in Web Music Videos
Y.-B. Lin, Y. Tian, L. Yang, **G. Bertasius**, H. Wang
 In Winter Conference on Applications of Computer Vision (**WACV**) 2025 (Oral). [\[pdf\]](#)
41. DAM: Dynamic Adapter Merging for Continual Video QA Learning
F. Cheng, Z. Wang, Y.-L. Sung, Y.-B. Lin, M. Bansal, **G. Bertasius**
 In Winter Conference on Applications of Computer Vision (**WACV**) 2025. [\[pdf\]](#)
40. A Simple LLM Framework for Long-Range Video Question-Answering
C. Zhang, T. Lu*, M. Islam, Z. Wang, S. Yu, M. Bansal, **G. Bertasius**
 In Empirical Methods in Natural Language Processing (**EMNLP**) 2024. [\[pdf\]](#)
39. Augmented Reality Demonstrations for Scalable Robot Imitation Learning
Y. Yang, B. Ikeda, **G. Bertasius**, D. Szafrir
 In International Conference on Intelligent Robots and Systems (**IROS**) 2024. [\[pdf\]](#)
38. Mementos: A Comprehensive Benchmark for Multimodal Large Language Model Reasoning over Image Sequences
 X. Wang, Y. Zhou, X. Liu, H. Lu, Y. Xu, F. He*, J. Yoon, T. Lu*, **G. Bertasius**, M. Bansal, H. Yao, F. Huang
 In Association for Computational Linguistics (**ACL**) 2024. [\[pdf\]](#)
37. Propose, Assess, Search: Harnessing LLMs for Goal-Oriented Planning in Instructional Videos
M. Islam, H. Wang, F.-J. Chu, K. Kitani, **G. Bertasius**, X. Yang
 In European Conference on Computer Vision (**ECCV**) 2024 (Oral). [\[pdf\]](#)
36. Siamese Vision Transformers are Scalable Audio-visual Learners
Y.-B. Lin, **G. Bertasius**
 In European Conference on Computer Vision (**ECCV**) 2024. [\[pdf\]](#)
35. 4Diff: 3D-Aware Diffusion Model for Third-to-First Viewpoint Translation
F. Cheng, M. Luo, H. Wang, A. Dimakis, L. Torresani, **G. Bertasius**, K. Grauman
 In European Conference on Computer Vision (**ECCV**) 2024. [\[pdf\]](#)
34. RGNet: A Unified Clip Retrieval and Grounding Network for Long Videos
 T. Hannan, M. Islam, T. Seidl, **G. Bertasius**
 In European Conference on Computer Vision (**ECCV**) 2024. [\[pdf\]](#)
33. Video ReCap: Recursive Captioning of Hour-Long Videos
M. Islam, N. Ho*, X. Yang, T. Nagarajan, L. Torresani, **G. Bertasius**
 In Computer Vision and Pattern Recognition (**CVPR**) 2024. [\[pdf\]](#)
Egocentric Vision Distinguished Paper Award

32. Ego-Exo4D: Understanding Skilled Human Activity from First- and Third-Person Perspectives
K. Grauman, ... , M. Islam, W. Shan*, O. Oguntola*, J. Zhuo*, ... **G. Bertasius**, M. Wray
In Computer Vision and Pattern Recognition (**CVPR**) 2024 (Oral). [\[pdf\]](#)
Co-led the proficiency estimation benchmark.
31. LoCoNet: Long-Short Context Network for Active Speaker Detection
X. Wang, F. Cheng, **G. Bertasius**, D. Crandall
In Computer Vision and Pattern Recognition (**CVPR**) 2024. [\[pdf\]](#)
30. Unified Coarse-to-Fine Alignment for Video-Text Retrieval
Z. Wang, Y.-L. Sung, F. Cheng, **G. Bertasius**, M. Bansal
In International Conference on Computer Vision (**ICCV**) 2023. [\[pdf\]](#)
29. SimpleClick: Interactive Image Segmentation with Simple Vision Transformers
Q. Liu, Z. Xu, **G. Bertasius**, M. Niethammer
In International Conference on Computer Vision (**ICCV**) 2023. [\[pdf\]](#)
28. VindLU: A Recipe for Effective Video-and-Language Pretraining
F. Cheng, X. Wang, J. Lei, D. Crandall, M. Bansal, **G. Bertasius**
In Computer Vision and Pattern Recognition (**CVPR**) 2023. [\[pdf\]](#)
27. Vision Transformers are Parameter-Efficient Audio-Visual Learners
Y.-B. Lin, Y.-L. Sung, J. Lei, M. Bansal, **G. Bertasius**
In Computer Vision and Pattern Recognition (**CVPR**) 2023. [\[pdf\]](#)
26. Efficient Movie Scene Detection using State-Space Transformers
M. Islam, M. Hasan, K. Athrey, T. Braskich, **G. Bertasius**
In Computer Vision and Pattern Recognition (**CVPR**) 2023. [\[pdf\]](#)
25. Improving Video Retrieval Using Multilingual Knowledge Transfer
A. Madasu, E. Aflalo, G. Stan, S.-Y. Tseng, **G. Bertasius**, V. Lal
In European Conference on Information Retrieval (**ECIR**) 2023. [\[pdf\]](#)
Best Student Paper Award
24. ECLIPSE: Efficient Long-range Video Retrieval using Sight and Sound
Y.-B. Lin, J. Lei, M. Bansal, **G. Bertasius**
In European Conference on Computer Vision (**ECCV**) 2022 (Oral). [\[pdf\]](#)
23. Long Movie Clip Classification with State-Space Video Models
M. Islam, **G. Bertasius**
In European Conference on Computer Vision (**ECCV**) 2022. [\[pdf\]](#)
22. TALLFormer: Temporal Action Localization with a Long-memory Transformer
F. Cheng, **G. Bertasius**
In European Conference on Computer Vision (**ECCV**) 2022. [\[pdf\]](#)
21. Learning to Retrieve Videos by Asking Questions
A. Madasu, J. Oliva, **G. Bertasius**
In ACM Multimedia (**ACM MM**) 2022. [\[pdf\]](#)
20. Learning To Recognize Procedural Activities with Distant Supervision
X. Lin, F. Petroni, **G. Bertasius**, M. Rohrbach, S. F. Chang, L. Torresani
In Computer Vision and Pattern Recognition (**CVPR**) 2022. [\[pdf\]](#)
19. Long-Short Temporal Contrastive Learning of Video Transformers
J. Wang, **G. Bertasius**, D. Tran, L. Torresani
In Computer Vision and Pattern Recognition (**CVPR**) 2022. [\[pdf\]](#)
18. Is Space-Time Attention All You Need for Video Understanding?
G. Bertasius, H. Wang, L. Torresani
In International Conference on Machine Learning (**ICML**) 2021. [\[pdf\]](#)
Top-5 Most Cited ICML 2021 Paper (>4,000 citations, >1.5M model downloads)

17. Vx2Text: End-to-End Learning of Video-Based Text Generation From Multimodal Inputs
X. Lin, **G. Bertasius**, J. Wang, S. F. Chang, D. Parikh, L. Torresani
In Computer Vision and Pattern Recognition (**CVPR**) 2021. [\[pdf\]](#)
16. Supervoxel Attention Graphs for Long-Range Video Modeling
Y. Wang, **G. Bertasius**, T. H. Oh, A. Gupta, M. Hoai, L. Torresani
In Winter Conference on Applications of Computer Vision (**WACV**) 2021. [\[pdf\]](#)
15. COBE: Contextualized Object Embeddings from Narrated Instructional Video
G. Bertasius, L. Torresani
In Neural Information Processing Systems (**NeurIPS**) 2020. [\[pdf\]](#)
14. Classifying, Segmenting, and Tracking Object Instances in Video with Mask Propagation
G. Bertasius, L. Torresani
In Computer Vision and Pattern Recognition (**CVPR**) 2020 (Oral). [\[pdf\]](#)
CVPR Best Paper Award Finalist
13. Attentive Action and Context Factorization
Y. Wang, V. Tran, **G. Bertasius**, L. Torresani, M. Hoai
In British Machine Vision Conference (**BMVC**) 2020. [\[pdf\]](#)
12. Learning Temporal Pose Estimation from Sparsely-Labeled Videos
G. Bertasius, C. Feichtenhofer, D. Tran, L. Torresani, J. Shi
In Neural Information Processing Systems (**NeurIPS**) 2019. [\[pdf\]](#)
11. Object Detection in Video with Spatiotemporal Sampling Networks
G. Bertasius, L. Torresani, J. Shi
In European Conference on Computer Vision (**ECCV**) 2018. [\[pdf\]](#)
10. Egocentric Basketball Motion Planning from a Single First-Person Image
G. Bertasius, A. Chan, J. Shi
In Computer Vision and Pattern Recognition (**CVPR**) 2018. [\[pdf\]](#)
9. Am I a Baller? Basketball Performance Assessment from First-Person Videos
G. Bertasius, H. S. Park, S. X. Yu, J. Shi
In International Conference on Computer Vision (**ICCV**) 2017. [\[pdf\]](#)
8. Unsupervised Learning of Important Objects from First-Person Videos
G. Bertasius, H. S. Park, S. X. Yu, J. Shi
In International Conference on Computer Vision (**ICCV**) 2017. [\[pdf\]](#)
7. Convolutional Random Walk Networks for Semantic Image Segmentation
G. Bertasius, L. Torresani, S. X. Yu, J. Shi
In Computer Vision and Pattern Recognition (**CVPR**) 2017. [\[pdf\]](#)
6. First-Person Action Object Detection with EgoNet
G. Bertasius, H. S. Park, S. X. Yu, J. Shi
In Robotics: Science and Systems (**RSS**) 2017 (Oral). [\[pdf\]](#)
5. Local Perturb-and-MAP for Structured Prediction
G. Bertasius, Q. Liu, L. Torresani, J. Shi
In Artificial Intelligence and Statistics (**AISTATS**) 2017. [\[pdf\]](#)
4. Semantic Segmentation with Boundary Neural Fields
G. Bertasius, J. Shi, L. Torresani
In Computer Vision and Pattern Recognition (**CVPR**) 2016. [\[pdf\]](#)
3. Automatic Lymph Node Segmentation Using Holistically-nested Networks and Structured Optimization
I. Nogues, L. Lu, X. Wang, H. Roth, **G. Bertasius**, N. Lay, J. Shi, Y. Tsehay, R. M. Summers
In Medical Image Computing and Computer Assisted Intervention (**MICCAI**) 2016.

2. High-for-Low, Low-for-High: Efficient Boundary Detection from Deep Object Features and its Applications to High-Level Vision
G. Bertasius, J. Shi, L. Torresani
In International Conference on Computer Vision (ICCV) 2015. [\[pdf\]](#)
1. DeepEdge: A Multi-Scale Bifurcated Deep Network for Top-Down Contour Detection
G. Bertasius, J. Shi, L. Torresani
In Computer Vision and Pattern Recognition (CVPR) 2015. [\[pdf\]](#)

Open-Source Software and Datasets

Public releases from my group, listed by associated publication. Every release is linked from gedasbertasius.com/publications.

29. **SVI-Bench** (2026). Benchmark, dataset, and evaluation code for strategic video intelligence across basketball, soccer, and hockey. Released with SVI-Bench, ECCV 2026.
github.com/texaser/svi-bench · huggingface.co/datasets/mvp-group/svi-bench · svi-bench.github.io
28. **VTs** (2026). Code, training data, and pretrained model for self-correcting agentic long-video question answering. Released with Searching Videos as Trees, 2026.
github.com/CeeZh/VTs · huggingface.co/datasets/ceezh/VTs_data · huggingface.co/ceezh/VTs-Qwen3-VL-8B
27. **EgoMemReason** (2026). Benchmark and evaluation code for memory-driven reasoning over week-long egocentric video. Released with EgoMemReason, COLM 2026.
github.com/Ziyang412/EgoMemReason · huggingface.co/datasets/Ted412/EgoMemReason
26. **WatchAct** (2026). Benchmark and dataset for behavior-grounded robot manipulation. Released with WatchAct, 2026.
github.com/Baiqi-Li/WatchAct · huggingface.co/datasets/BaiqiL/WatchAct
25. **TimeBlind** (2026). Benchmark for spatio-temporal compositionality in video LLMs. Released with TimeBlind, EMNLP 2026.
github.com/Baiqi-Li/TimeBlind · huggingface.co/datasets/BaiqiL/TimeBlind
24. **SiLVR** (2026). Code for language-based video reasoning with strong reasoning LLMs. Released with SiLVR, TMLR 2026.
github.com/CeeZh/SILVR
23. **ExAct** (2025). Benchmark, dataset, evaluation code, and public leaderboard for expert action analysis. Released with ExAct, NeurIPS 2025 Datasets and Benchmarks Track.
github.com/Texaser/Exact · huggingface.co/datasets/Alexhimself/ExAct · texaser.github.io/exact_project_page
22. **Video-RTS** (2025). Code for data-efficient reinforcement learning and test-time scaling for video reasoning. Released with Video-RTS, EMNLP 2025.
github.com/Ziyang412/Video-RTS
21. **ReBot** (2025). Code for real-to-sim-to-real robotic video synthesis. Released with ReBot, IROS 2025.
github.com/yuffish/rebot
20. **BIMBA** (2025). Code and pretrained model for selective-scan compression in long-range video question answering. Released with BIMBA, CVPR 2025.
github.com/md-mohaiminul/BIMBA · huggingface.co/mmiemon/BIMBA-LLaVA-Qwen2-7B
19. **BASKET** (2025). Dataset of 4,477 hours of basketball video covering 32,232 players across 21 leagues, with skill annotations, plus training code. Released with BASKET, CVPR 2025.
github.com/yulupan00/BASKET
18. **VideoTree** (2025). Code for adaptive tree-based video representation for LLM reasoning. Released with VideoTree, CVPR 2025.
github.com/Ziyang412/VideoTree
17. **ReVisionLLM** (2025). Code for recursive temporal grounding in hour-long video. Released with ReVisionLLM, CVPR 2025.
github.com/Tanveer81/ReVisionLLM

16. **DAM** (2025). Code for dynamic adapter merging in continual video QA. Released with DAM, WACV 2025.
github.com/klauscc/DAM
15. **LLoVi** (2024). Code and extracted captions for EgoSchema, NExT-QA, NExT-GQA, and IntentQA. Released with A Simple LLM Framework for Long-Range Video Question-Answering, EMNLP 2024.
github.com/CeeZh/LLoVi
14. **AVSiam** (2024). Code for scalable audio-visual learning with Siamese vision transformers. Released with Siamese Vision Transformers are Scalable Audio-visual Learners, ECCV 2024.
github.com/GenjiB/AVSiam
13. **RGNet** (2024). Code for unified clip retrieval and grounding in long video. Released with RGNet, ECCV 2024.
github.com/Tanveer81/RGNet
12. **Video ReCap and Ego4D-HCap** (2024). Code, pretrained models, and the Ego4D-HCap dataset of hierarchical captions for hour-long video. Released with Video ReCap, CVPR 2024.
github.com/md-mohaiminul/VideoRecap
11. **Ego-Exo4D Proficiency Estimation Benchmark** (2024). Benchmark I co-led within the multi-institution Ego-Exo4D dataset of skilled human activity. Released with Ego-Exo4D, CVPR 2024.
ego-exo4d-data.org
10. **VindLU** (2023). Code and pretrained models for video-and-language pretraining. Released with VindLU, CVPR 2023.
github.com/klauscc/VindLU
9. **LAVISH** (2023). Code for parameter-efficient audio-visual adaptation of vision transformers. Released with Vision Transformers are Parameter-Efficient Audio-Visual Learners, CVPR 2023.
github.com/GenjiB/LAVISH
8. **ECLIPSE** (2022). Code for efficient long-range audio-visual video retrieval. Released with ECLIPSE, ECCV 2022.
github.com/GenjiB/ECLIPSE
7. **TALLFormer** (2022). Code for temporal action localization with a long-memory transformer. Released with TALLFormer, ECCV 2022.
github.com/klauscc/TALLFormer
6. **ViS4mer** (2022). Code for state-space video models for long movie clip classification. Released with Long Movie Clip Classification with State-Space Video Models, ECCV 2022.
github.com/md-mohaiminul/ViS4mer
5. **TimeSformer** (2021). Code and pretrained checkpoints for the first video architecture built exclusively on self-attention. Over 1.5 million model downloads. Released with Is Space-Time Attention All You Need for Video Understanding?, ICML 2021.
github.com/facebookresearch/TimeSformer
4. **HowTo100M_BB** (2020). Pseudo-annotations of HowTo100M for contextualized object embeddings. Released with COBE, NeurIPS 2020.
3. **PoseWarper** (2019). Code and models for temporal pose estimation from sparsely labeled video. Ranked first on the PoseTrack multi-frame leaderboard at release. Released with Learning Temporal Pose Estimation from Sparsely-Labeled Videos, NeurIPS 2019.
github.com/facebookresearch/PoseWarper
2. **BNF** (2016). Code for semantic segmentation with boundary neural fields. Released with Semantic Segmentation with Boundary Neural Fields, CVPR 2016.
github.com/gberta/BNF_globalization
1. **HFL** (2015). Code for efficient boundary detection from deep object features. Released with High-for-Low, Low-for-High, ICCV 2015.
github.com/gberta/HFL_code

Teaching

- Instructor for Vision Transformers (COMP 669), UNC Chapel Hill, Fall 2026.
- Research Study Assignment (research leave), Spring 2026.
- Parental leave, Fall 2025.
- Instructor for Vision Transformers (COMP 590/790), UNC Chapel Hill, Spring 2025.
- Instructor for Video Recognition (COMP 590/790), UNC Chapel Hill, Fall 2024.
- Instructor for Visual Recognition with Transformers (COMP 590/790), UNC Chapel Hill, Spring 2024.
- Instructor for Video Recognition (COMP 590/790), UNC Chapel Hill, Fall 2023.
- Parental leave, Spring 2023.
- Instructor for Self-Supervised Visual Representation Learning (COMP 590/790), UNC Chapel Hill, Fall 2022.
- Instructor for Visual Recognition Using Transformers (COMP 790), UNC Chapel Hill, Spring 2022.
- Instructor for Advanced Topics in Video Understanding (COMP 790), UNC Chapel Hill, Fall 2021.
- Co-Instructor for Vision and Learning (CIS 680), UPenn, Fall 2018.
- Teaching Assistant for Vision Intelligence and Machine Learning (MicroMasters Program), UPenn, Summer 2017.
- Teaching Assistant for Artificial Intelligence (CIS 521), UPenn, Spring 2016.
- Teaching Assistant for Machine Learning (CIS 520), UPenn, Fall 2015.

Research Advising

Doctoral Advisees

10. Jr-Jen Chen, 2026 – present
9. Baiqi Li, 2025 – present
8. Seongsu Ha, 2025 – present
7. Yulu Pan, 2024 – present
6. Han Yi, 2024 – present
5. Yue Yang (co-advised with Dan Szafir), 2023 – present
4. Ce Zhang, 2023 – present
3. Yan-Bo Lin, 2021 – 2025 (Research Scientist at Bosch Center for Artificial Intelligence)
2. Md Mohaiminul Islam, 2021 – 2025 (Machine Learning Scientist at Netflix)
1. Feng Cheng, 2021 – 2024 (Member of Technical Staff at SpaceXAI)

Masters Advisees

4. Jack Fan, 2024 – 2025
3. Levi Harris, 2024 – 2025
2. Avinash Madasu, 2021 – 2023
1. Maitrey Gramopadhye, 2021 – 2022

UNC Undergraduate Advisees

Eight are co-authors on refereed publications listed above.

16. Lanye Wang, 2026 – present
15. Eric Pan, 2026 – present
14. Mairui Li, 2026 – present
13. Pranav Wagh, 2025 – 2026
12. Ben Zhang, 2025 – 2026
11. Oluwatumininu Oguntola, 2022 – 2026 (Honors thesis, 2026)
10. Xinyu (Zoey) Liu, 2023 – 2025
9. Vish Ravichandran, 2023 – 2024

8. Myles Mason, 2023 – 2024
7. Jack Geng, 2023 – 2024
6. Lauren Feldman, 2023 – 2024
5. Feihong He, 2022 – 2025
4. Jeff Zhuo, 2022 – 2025
3. Wei Shan, 2022 – 2025
2. Meliora Ho, 2022 – 2024
1. Taixi Lu, 2022 – 2024

Facebook AI Interns

2. Xudong Lin (Summers 2020, 2021)
1. Yang Wang (Summer 2019)

Doctoral Thesis Committees

21. Maitrey Gramopadhye (UNC Ph.D.)
20. Arushi Rai (University of Pittsburgh Ph.D.)
19. Luchao Qi (UNC Ph.D.)
18. Michael Matena (UNC Ph.D.)
17. Romy Luo (University of Texas at Austin Ph.D.)
16. Yi-Lin Sung (UNC Ph.D.)
15. Jaemin Cho (UNC Ph.D.)
14. Jialu Li (UNC Ph.D.)
13. Yilin Liu (UNC Ph.D.)
12. Boqi Chen (UNC Ph.D.)
11. Somnath Basu Roy Chowdhury (UNC Ph.D.)
10. Adyasha Maharana (UNC Ph.D.)
9. Lin Tian (UNC Ph.D.)
8. Hastings Greer (UNC Ph.D.)
7. Dongxu Zhao (UNC Ph.D.)
6. Peirong Liu (UNC Ph.D.)
5. Thanh Vu (UNC Ph.D.)
4. Mykhailo Shvets (UNC Ph.D.)
3. Jisan Mahmud (UNC Ph.D.)
2. Yubo Zhang (UNC Ph.D.)
1. Yifeng Shi (UNC Ph.D.)

Professional Service

Workshops

- Co-founder and lead organizer, Transformers for Vision workshop, CVPR 2022 – 2026.
- Organizer, Long Video Understanding workshop, CVPR 2026.
- Organizer, Computer Vision with Small Data workshop, CVPR 2026.
- Organizer, Multimodal Video Agents workshop, CVPR 2025.

Area Chair

- Computer Vision and Pattern Recognition (CVPR)
- Neural Information Processing Systems (NeurIPS)

Conference and Journal Reviewing

- Computer Vision and Pattern Recognition (CVPR)
- International Conference on Computer Vision (ICCV)

- European Conference on Computer Vision (ECCV)
- Neural Information Processing Systems (NeurIPS)
- International Conference on Machine Learning (ICML)
- International Conference on Learning Representations (ICLR)
- International Conference on Robotics and Automation (ICRA)
- IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)

Committees at UNC

- Graduate Admissions (2021 – present)
- Faculty Search (2021 – 2025)
- Diversity, Equity, and Inclusion (2021 – 2022)

Invited Talks

49. Distinguished Lecture Series at Oracle AI, “From Watching to Understanding: Reasoning and Agency in Video Models”, July 2026.
48. Invited Talk at CVPR Vision for Intelligent Task Assistants workshop, “From Perception to Agency: The Cognitive Stack for Video Task Assistants”, Denver, CO, June 2026.
47. Invited Talk at WACV Computer Vision for Winter Sports workshop, “From Perception to Agency in Strategic Video Understanding”, Tucson, AZ, March 2026.
46. Invited Talk at Ai2, “From Perception to Agency in Strategic Video Understanding”, Seattle, WA, February 2026.
45. Spotlight Talk at UNC AI Core Day, “From Pixels to Coaches: AI That Sees, Understands and Teaches”, Chapel Hill, NC, February 2026.
44. Invited Talk at UNC Vision and Multimodal AI Seminar, “Paradoxes of Language-Driven Video Understanding”, Chapel Hill, NC, January 2026.
43. Guest Lecture at UT Austin’s Computer Vision course, “From Pixels to AI Coaches”, Austin, TX, November 2025.
42. Invited Talk at the ICCV Skilled Activity Understanding workshop, “Beyond Recognition: Challenges in Expert-Level Skill Analysis”, Honolulu, HI, October 2025.
41. Guest Lecture at the University of Michigan Advanced Vision Course, “Efficient Video Understanding”, Ann Arbor, MI, October 2025.
40. Invited Talk at the Video AI Symposium, “Paradoxes of Language-Driven Video Understanding”, Paris, France, September 2025.
39. Invited Talk at the CVPR Multimodal Video Agents Workshop, “SiLVR: A Simple Language-based Video Reasoning Framework”, Nashville, TN, June 2025.
38. Invited Talk at the CVPR Area Chair Workshop, “Rethinking Video Understanding: When Language Isn’t Enough”, June 2025.
37. Invited UNC Community Talk at the William and Ida Friday Center, “From Cooking to Basketball: Skill Learning in the Age of AI”, Chapel Hill, NC, June 2025.
36. Invited Seminar Talk at Twelve Labs, “Complex Video Understanding using Video-Language Models”, March 2025.
35. Invited Seminar Talk at Twelve Labs, “Toward Sports-based Video Understanding using Multimodal LLMs”, March 2025.
34. Invited Talk at the Data Science in Sports workshop, “Toward Sports-based Video Understanding using Multimodal LLMs”, Tübingen, Germany, March 2025.
33. Guest Lecture at UT Austin’s Computer Vision course, “Complex Video Understanding using Language ... and Video”, February 2025.
32. Invited Talk at the Video-Language Models workshop, “Complex Video Understanding using Language ... and Video”, Vancouver, Canada, December 2024.

31. Invited Talk at the UNC School of Data Science and Society Generative AI Workshop, "Generative Video Models for Diverse Multimodal Applications", Chapel Hill, NC, March 2024.
30. Invited Talk at Sony, "Building a Performant Text-to-Video Retrieval Framework with Limited Resources", November 2023.
29. Invited Talk at the Carolina Data Science Seminar, "Computer Vision Models for Data-Driven Human Behavior Understanding", June 2023.
28. Keynote Talk at the CVPR Sight and Sound Workshop, "Toward Efficient and Scalable Audiovisual Video Models", Vancouver, Canada, June 2023.
27. Keynote Talk at the CVPR Long Video Understanding Workshop, "Evaluating and Benchmarking Models for Long Video Understanding", Vancouver, Canada, June 2023.
26. Invited Talk at the Portland Trail Blazers, "Computer Vision for Advanced Basketball Analytics", December 2022.
25. Keynote Talk at the CVPR EPIC Workshop, "Understanding Complex Human Activities in Long-range Videos", New Orleans, LA, June 2022.
24. Invited Talk at the Samsung AI Center, "Long-range Video Understanding using Transformers", April 2022.
23. Invited Talk at University of Oxford, "Long-range Video Understanding using Transformers", April 2022.
22. Keynote Talk at the CVPR ActivityNet Workshop, "Is Space-Time Attention All You Need for Video Understanding?", June 2021.
21. Invited Talk at the Facebook Video Understanding Reading Group, "Is Space-Time Attention All You Need for Video Understanding?", May 2021.
20. Invited Talk at the Facebook AI Transformers Workshop, "Is Space-Time Attention All You Need for Video Understanding?", April 2021.
19. Invited Talk at the KUIS AI Center, "Visual Perception Models for Semantic Video Understanding", April 2021.
18. Invited Talk at Brown University, "Visual Perception Models for Semantic Video Understanding", March 2021.
17. Invited Talk at MIT, "Video Understanding with Modern Language Models", March 2021.
16. Invited Talk at UMass Amherst, "Video Understanding with Modern Language Models", March 2021.
15. Invited Talk at 2d3d.ai, "Visual Perception Models for Multi-Modal Video Understanding", February 2021.
14. Invited Talk at Princeton University, "Visual Perception Models for Multi-Modal Video Understanding", December 2020.
13. Invited Talk at Columbia University, "Understanding Human-Object Interactions in Video", November 2020.
12. Talk at the ECCV Egocentric Perception Workshop, "COBE: Contextualized Object Embeddings from Narrated Instructional Video", August 2020.
11. Oral Presentation at CVPR, "Classifying, Segmenting, and Tracking Object Instances in Video with Mask Propagation", June 2020.
10. Invited Talk at SRI International, "Embodied Visual Perception Models for Human Behavior Understanding", Princeton, NJ, May 2018.
9. Invited Talk at the Philadelphia 76ers, "First-Person Vision for Data-Driven Basketball Analytics", Philadelphia, PA, February 2018.
8. Talk at the ICCV Egocentric Perception Workshop, "Am I a Baller? Basketball Performance Assessment from First-Person Videos", October 2017.
7. Talk at the ICCV Egocentric Perception Workshop, "Unsupervised Learning of Important Objects from First-Person Videos", October 2017.
6. Talk at the GRASP Deep Learning Seminar, "Human Behavior Understanding Using First-Person Cameras", Philadelphia, PA, September 2017.

5. Talk at Penn Research in Machine Learning, “Convolutional Random Walk Networks for Semantic Image Segmentation”, Philadelphia, PA, September 2017.
4. Oral Presentation at RSS, “First-Person Action Object Detection with EgoNet”, July 2017.
3. Talk at the Robotics Collaborative Technology Alliance (RCTA) Meeting, “First-Person Action Object Detection with EgoNet”, Philadelphia, PA, May 2017.
2. Talk at the Complex Scene Perception Seminar, “Semantic Segmentation with Boundary Neural Fields”, Philadelphia, PA, November 2016.
1. Talk at the Complex Scene Perception Seminar, “DeepEdge: A Multi-Scale Bifurcated Deep Network for Top-Down Contour Detection”, Philadelphia, PA, November 2015.

Basketball Honors

- NCAA Division I basketball player, Dartmouth College (2010 – 2014).
- Professional basketball player, Akademija Vilnius, Lithuania (2007 – 2009).
- Member, Lithuania U16, U18, and U19 national basketball teams (2006 – 2009).
- Silver medalist, FIBA U18 European Basketball Championship (2008).