

Gedas Bertasius

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

Appointments

University of North Carolina at Chapel Hill

Assistant Professor in the Department of Computer Science 2021-

Facebook AI Research (FAIR)

Postdoctoral Researcher 2019-2021
Mentor: Lorenzo Torresani

Facebook AI Research (FAIR)

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

Research Overview

My research group develops foundational models for **multimodal video understanding**, enabling machines to perceive, reason about, and interact with complex **video, audio, and language data**. Moving beyond passive perception, we investigate the spatiotemporal abstractions necessary for AI to **understand intricate human behaviors over long horizons**. We deploy these models across high-impact domains—including egocentric vision, robotics, AI assistants, and sports analytics—to redefine the boundaries of machine perception and human-AI interaction.

Education

University of Pennsylvania

PhD in Computer Science Philadelphia, PA, USA 2014-2019
Advisor: Jianbo Shi

Dartmouth College




Bachelor of Arts in Computer Science & Mathematics Hanover, NH, USA 2010-2014
Advisor: Lorenzo Torresani

Honors and Awards

NSF CAREER Award	2026
Sony Focused Research Award	2024, 2025
Egocentric Vision Distinguished Paper Award	2025
Meta Research Award	2022, 2023
UNC Junior Faculty Award	2022
CSSA Excellence in Teaching Award	2022
Sony Faculty Innovation Award	2021, 2022
CVPR Best Paper Award Finalist	2020
NSF IGERT Doctoral Fellowship	2015

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 (>4000 Citations)**
2. 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**
3. 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. [[pdf](#)]
Co-led the proficiency estimation benchmark
4. Long Movie Clip Classification with State-Space Video Models
M. Islam, **G. Bertasius**
In European Conference on Computer Vision (**ECCV**) 2022. [[pdf](#)]
5. Classifying, Segmenting, and Tracking Object Instances in Video with Mask Propagation
G. Bertasius, L. Torresani
In Computer Vision and Pattern Recognition (**CVPR**) 2020. [[pdf](#)]
 **CVPR Best Paper Award Finalist**

Research Grants, Sponsorship, Gifts



Total funding (own share): ~ \$2.7 million

15. **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).
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).

Peer-Reviewed Conference Publications

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

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 at CVPR 2025 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 Applications of Computer Vision (WACV) 2026. [\[pdf\]](#)
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 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 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 D&B 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 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 Winner at CVPR 2025 Ego4D 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 Applications of Computer Vision (**WACV**) 2025. [[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 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. Szafir
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. [[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. [[pdf](#)]
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. 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. 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. [[pdf](#)]
24. Long Movie Clip Classification with State-Space Video Models
M. Islam, **G. Bertasius**
In European Conference on Computer Vision (**ECCV**) 2022. [[pdf](#)]
23. TALLFormer: Temporal Action Localization with a Long-memory Transformer
F. Cheng, **G. Bertasius**
In European Conference on Computer Vision (**ECCV**) 2022. [[pdf](#)]
22. 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**) 2022. [[pdf](#)]



Best Student Paper Award

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](#)]
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 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. [[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, and 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, and J. Shi
In Robotics: Science and Systems (**RSS**) 2017. [[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, and 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. [[pdf](#)]
 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, and 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, and L. Torresani
In Computer Vision and Pattern Recognition (**CVPR**) 2015. [[pdf](#)]

Teaching

-
- 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.

- 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 Machine Learning (CIS 520), UPenn, Fall 2015.

Research Advising

Doctoral Advisees

9. Baiqi Li, 2025-present
8. Seongsu Ha, 2025-present
7. Yulu Pan, 2024-present
6. Han Yi, 2024-present
5. Yue Yang (w/ Dan Szafir), 2023-present
4. Ce Zhang, 2023-present
3. Yan-Bo Lin, 2021-2025
2. Md Mohaiminul Islam, 2021-2025
1. Feng Cheng, 2021-2024

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

13. Pranav Wagh, 2025-present
12. Ben Zhang, 2025-present
11. Oluwatumininu Oguntola 2022-present
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

- The lead organizer of the Transformers for Vision workshop (CVPR 2022, CVPR 2023, CVPR 2024, CVPR 2025, CVPR 2026).
- Organizer of the Multimodal Video Agents workshop (CVPR 2025, CVPR 2026).
- Organizer of the Computer Vision with Small Data workshop (CVPR 2026).

Area Chair

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

Conference & Journal Reviewing

- Computer Vision and Pattern Recognition (CVPR)
- International Conference on Computer Vision (ICCV)
- European Conference on Computer Vision (ECCV)
- Neural Information and 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

- Diversity, Equity and Inclusion (DEI)
- Faculty Search
- Graduate Admissions

Invited Talks

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 Agent Workshop, “SiLVR: A Simple Language-based Video Reasoning Framework”, CVPR, June, 2025.

38. Invited Talk at the CVPR Area Chair Workshop, “Rethinking Video Understanding: When Language Isn’t Enough”, CVPR, 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”, Austin, TX, February, 2025.
32. Invited Talk at the Video-Language Models workshop, “Complex Video Understanding using Language ... and Video”, NeurIPS, 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”, CVPR, June, 2023.
27. Keynote Talk at the CVPR Long Video Understanding Workshop, “Evaluating and Benchmarking Models for Long Video Understanding”, CVPR, 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”, CVPR, June, 2022.
24. Invited Talk at the Samsung AI Center, “Long-range Video Understanding using Transformers”, April, 2022.
23. Invited Talk at the VGG Seminar, “Long-range Video Understanding using Transformers”, University of Oxford, April, 2022.
22. Keynote Talk at the CVPR ActivityNet Workshop, “Is Space-Time Attention All You Need for Video Understanding?”, CVPR, 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 University, “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”, ECCV, August, 2020.
11. Oral Presentation at CVPR, “Classifying, Segmenting, and Tracking Object Instances in Video with Mask Propagation”, CVPR, 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”, ICCV, October, 2017.
7. Talk at the ICCV Egocentric Perception Workshop, “Unsupervised Learning of Important Objects from First-Person Videos”, ICCV, 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”, RSS, Jul., 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.

Other Activities

- Member of U-16, U-18, U-19 Lithuanian National Basketball Teams.
- Silver medalist at FIBA U-18 European Basketball Championship.
- NCAA Division 1 Player at Dartmouth College.
- #1 Ranked Basketball Player in Eastern Connecticut (Class of 2010).
- Most Valuable Player in Lithuanian Students League (2005, 2006).