

# Gedas Bertasius

<https://gedasbertasius.com> | [gedas@cs.unc.edu](mailto: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

### Facebook AI Research (FAIR)

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

## Research Interests

---

- **Computer Vision:** video recognition, video-language modeling, egocentric vision, sport analytics.
- **Machine Learning:** multi-modal learning, large-scale representation learning, transfer learning.

## Education

---

### University of Pennsylvania

Philadelphia, PA, USA

*PhD in Computer Science*

2014-2019

- Advisor: Jianbo Shi

### Dartmouth College

Hanover, NH, USA

*Bachelor of Arts in Computer Science & Mathematics*

2010-2014

- Advisor: Lorenzo Torresani

## Honors and Awards

---

- SONY Faculty Innovation Award 2022
- UNC Junior Faculty Development Award 2022
- Meta Research Award 2022
- CSSA Excellence in Teaching Award 2022
- SONY Faculty Innovation Award 2021
- Best CVPR 2020 Paper Nominee 2020
- CVPR 2020 EPIC-Kitchens Object Detection Challenge Winner 2020
- Outstanding CVPR, ICCV Reviewer Award 2018, 2019
- NSF IGERT Doctoral Fellowship 2015
- First Prize in John Kemeny Thesis Competition 2014
- Rufus Choate Scholar 2014

## Top Five Cited Papers

---

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](#)] (Citations: 1845)
2. 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](#)] (Citations: 588)
3. Object Detection in Video with Spatiotemporal Sampling Networks  
**G. Bertasius**, L. Torresani, J. Shi  
In European Conference on Computer Vision (ECCV) 2018. [[pdf](#)] (Citations: 275)
4. Semantic Segmentation with Boundary Neural Fields  
**G. Bertasius**, J. Shi, and L. Torresani  
In Computer Vision and Pattern Recognition (CVPR) 2016. [[pdf](#)] (Citations: 237)
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](#)] (Citations: 190)  
**CVPR Best Paper Nominee**

## Peer-Reviewed Conference Publications

---

36. 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](#)]
35. 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](#)]
32. Ego-Exo4D: Understanding Skilled Human Activity from First- and Third-Person Perspectives  
K. Grauman, A. Westbury, L. Torresani, K. Kitani, J. Malik, **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. 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. 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 Nominee**
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. Nogueiras, 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 [Visual Recognition Using 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.

## Mentoring

---

### UNC Graduate Advisees

- Feng Cheng (UNC Ph.D.)
- Yan-Bo Lin (UNC Ph.D.)
- Md Mohaiminul Islam (UNC Ph.D.)
- Ce Zhang (UNC Ph.D.)
- Yue Yang (UNC Ph.D. w/ Dan Szafir)
- Soumitri Chattopadhyay (UNC Ph.D.)
- Yulu Pan (UNC MS)

### UNC Undergraduate Advisees

- Levi Harris (Class of 2024)
- Myles Mason (Class of 2024)
- Taixi Lu (Class of 2024)
- Lauren Feldman (Class of 2025)
- Vish Ravichandran (Class of 2025)
- Meliora Ho (Class of 2025)
- Jack Geng (Class of 2025)
- Wei Shan (Class of 2025)
- Feihong He (Class of 2025)
- Jeff Zhuo (Class of 2025)
- Xinyu (Zoey) Liu (Class of 2025)
- Oluwatumininu Oguntola (Class of 2026)

### UNC Alumni

- Avinash Madasu (UNC MS)
- Maitrey Gramopadhye (UNC MS)

### Facebook AI Interns

- Xudong Lin (Summers 2020, 2021)
- Yang Wang (Summer 2019)

## Professional Service

---

### Workshops

- The lead organizer of the Transformers for Vision workshop (CVPR 2022, CVPR 2023, CVPR 2024).

### 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

### **Doctoral Thesis Committees**

- Yifeng Shi (UNC Ph.D.)
- Yubo Zhang (UNC Ph.D.)
- Jisan Mahmud (UNC Ph.D.)
- Mykhailo Shvets (UNC Ph.D.)
- Thanh Vu (UNC Ph.D.)
- Peirong Liu (UNC Ph.D.)
- Dongxu Zhao (UNC Ph.D.)
- Hastings Greer (UNC Ph.D.)
- Lin Tian (UNC Ph.D.)
- Adyasha Maharana (UNC Ph.D.)
- Somnath Basu Roy Chowdhury (UNC Ph.D.)
- Boqi Chen (UNC Ph.D.)

## **Invited Talks**

---

### **Generative Video Models for Diverse Multimodal Applications**

- Invited Talk at UNC's School of Data Science and Society Generative AI Workshop Mar 2024

### **Building a Performant Text-to-Video Retrieval Framework with Limited Resources**

- Invited Talk at SONY Nov 2023

### **Computer Vision Models for Data-Driven Human Behavior Understanding**

- Invited Talk at Carolina Data Science Seminar Jun 2023

### **Toward Efficient and Scalable Audiovisual Video Models**

- Keynote Talk at CVPR 2023 Sight and Sound Workshop Jun 2023

### **Evaluating and Benchmarking Models for Long Video Understanding**

- Keynote Talk at CVPR 2023 Long Video Understanding Workshop Jun 2023

### **Computer Vision for Advanced Basketball Analytics**

- Invited Talk at Portland Trail Blazers Dec 2022

### **Understanding Complex Human Activities in Long-range Videos**

- Keynote Talk at CVPR 2022 EPIC Workshop Jun 2022

### **Long-range Video Understanding using Transformers**

- Invited Talk at Samsung AI Center Apr 2022
  - Invited Talk at VGG Seminar, University of Oxford Apr 2022
- Is Space-Time Attention All You Need for Video Understanding?**
- Keynote Talk at CVPR 2021 ActivityNet Workshop Jun 2021
  - Invited Talk at Facebook Video Understanding Reading Group May 2021
  - Invited Talk at Facebook AI Transformers Workshop Apr 2021
- Visual Perception Models for Semantic Video Understanding**
- Invited Talk at KUIS AI Center Apr 2021
  - Invited Talk at Brown University Mar 2021
- Video Understanding with Modern Language Models**
- Invited Talk at MIT Mar 2021
  - Invited Talk at UMass Amherst University Mar 2021
- Visual Perception Models for Multi-Modal Video Understanding**
- Invited Talk at 2d3d.ai Feb 2021
  - Invited Talk at Princeton University Dec 2020
- Understanding Human-Object Interactions in Video**
- Invited Talk at Columbia University Nov 2020
- COBE: Contextualized Object Embeddings from Narrated Instructional Video**
- ECCV20 Egocentric Perception, Interaction and Computing Workshop, Glasgow, Scotland Aug 2020
- Classifying, Segmenting, and Tracking Object Instances in Video with Mask Propagation**
- CVPR20 Oral Presentation, Seattle, WA Jun 2020
- Embodied Visual Perception Models for Human Behavior Understanding**
- Invited Talk at SRI International, Princeton, NJ May 2018
- First-Person Vision for Data-Driven Basketball Analytics**
- Invited Talk at Philadelphia 76ers, Philadelphia, PA Feb 2018
- Am I a Baller? Basketball Performance Assessment from First-Person Videos**
- ICCV17 Egocentric Perception, Interaction and Computing Workshop, Venice, Italy Oct 2017
- Unsupervised Learning of Important Objects from First-Person Videos**
- ICCV17 Egocentric Perception, Interaction and Computing Workshop, Venice, Italy Oct 2017
- Human Behavior Understanding Using First-Person Cameras**
- GRASP Deep Learning Seminar, Philadelphia, PA Sep 2017
- Convolutional Random Walk Networks for Semantic Image Segmentation**



- Penn Research in Machine Learning, Philadelphia, PA Sep 2017

### **First-Person Action Object Detection with EgoNet**

- RSS 2017 Oral Presentation, Cambridge, MA Jul 2017
- Robotics Collaborative Technology Alliance (RCTA) Meeting, Philadelphia, PA May 2017

### **Semantic Segmentation with Boundary Neural Fields**

- Complex Scene Perception Seminar, Philadelphia, PA Nov 2016

### **DeepEdge: A Multi-Scale Bifurcated Deep Network for Top-Down Contour Detection**

- Complex Scene Perception Seminar, Philadelphia, PA Nov 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).