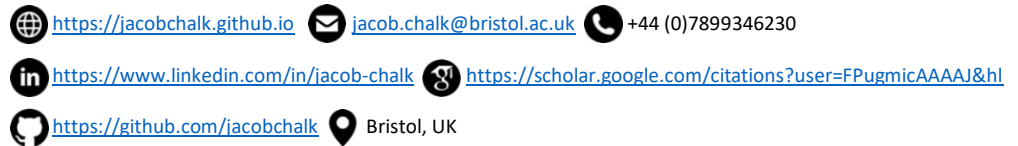


Jacob Chalk



Profile

Final-year PhD researcher in Computer Vision at the University of Bristol, supervised by Prof. Dima Damen. My research focuses on multi-modal and 3D video understanding, particularly in egocentric settings. I'm currently a PhD Intern at NAVER LABS Europe, working on 3D Object Tracking with the Visual Representation Learning team.

Expertise

Generative AI

- Integrated Multi-Modal Large Language Models (MLLMs) and Vision-Language Models (VLMs) for complex video reasoning tasks, including spatial, temporal, and gaze-conditioned queries [1].
- Designed gaze-primed object interaction tasks using 3D scene geometry, and benchmarked LLM capabilities on gaze-based predictive video question answering [1].

Computer Vision & 3D Vision

- Developed methods for 3D scene understanding and object tracking from egocentric video [1, 2].
- Proposed a multi-modal, long-context approach for audio-visual action recognition and detection [4].
- Co-led the creation of EPIC-Sounds, delivering a scalable, quality-controlled audio annotation pipeline for egocentric video. Designed and currently maintain public benchmarks and hosted challenges adopted by academia and industry. Recognised with an ICASSP 2023 oral presentation and the EgoVis Distinguished Paper Award [5].
- Maintain open-source codebases across multiple projects, enabling reproducibility and adoption [2, 4, 5].
- Ongoing research on 3D multi-object tracking as part of the Visual Representation Learning team at NAVER LABS Europe.

Technical Skills

- **Languages & Frameworks:** Python (primary), C++, JavaScript, PyTorch, TensorFlow, FastAPI
- **Tools & Systems:** NumPy, Pandas, OpenCV, Docker, Git, W&B, large-scale GPU clusters (JADE II, BlueCrystal4, Isambard AI)
- **Current focus:** CUDA optimisation, ONNX model deployment, Kubernetes, and high-performance inference with TensorRT

Research Experience

NAVER LABS Europe

Industrial Internship - Visual Representation Learning Team

Conducting early-stage research on 3D object tracking, supervised by [Diane Larlus](#).

Grenoble, France
February 2025 – Current

Education

University of Bristol

PhD in Computer Vision

Current Thesis Title: Multi-modal Annotations & Representation Learning in Egocentric Videos, supervised by [Prof. Dima Damen](#).

Bristol, UK

September 2021 – Current

University of Bristol

MEng in Computer Science - *First Class Honours*

Dissertation Title: Video GANS for Human-Object Interactions, supervised by [Prof. Dima Damen](#).

Bristol, UK

September 2017 – September 2021

Selected Publications & Awards

1. T. Perrett, A. Darkhalil, S. Sinha, O. Emara, S. Pollard, K. Parida, K. Liu, P. Gatti, S. Bansal, K. Flanagan, **J. Chalk**, Z. Zhu, R. Guerrier, F. Abdelazim, B. Zhu, D. Moltisanti, M. Wray, H. Doughty, and D. Damen, "HD-EPIC: A Highly-Detailed Egocentric Video Dataset" in [CVPR 2025](#). [\[Page\]](#) | [\[Paper\]](#)
2. C. Plizzari, S. Goel, T. Perrett, **J. Chalk**, A Kanazawa, D Damen, "Spatial Cognition from Egocentric Video: Out of Sight, Not Out of Mind" in [3DV 2025](#). [\[Page\]](#) | [\[Paper\]](#)
3. J. Huh, **J Chalk**, E. Kazakos, D. Damen, and A. Zisserman, "EgoVis 2022/23 Distinguished Paper Awards" awarded in [First Joint Egocentric Vision Workshop \(CVPR 2024\)](#).
4. **J. Chalk**, J. Huh, E. Kazakos, A. Zisserman, and D. Damen, "TIM: A Time Interval Machine for Audio-Visual Action Recognition" in [CVPR 2024](#). [\[Page\]](#) | [\[Paper\]](#)
5. J. Huh, **J Chalk**, E. Kazakos, D. Damen, and A. Zisserman, "EPIC-SOUNDS: A Large-Scale Dataset of Actions that Sound." in [ICASSP 2023](#). [\[Page\]](#) | [\[Paper\]](#)
6. Top 5 Third Year MEng Computer Science/Computer Science with Maths Student at the [University of Bristol 2020](#).