

# Siddhant Bansal

PhD Candidate in Computer Science

✉ [siddhant.bansal@bristol.ac.uk](mailto:siddhant.bansal@bristol.ac.uk)

🌐 [sid2697.github.io](https://github.com/sid2697)

🎓 [Google Scholar](#)

📍 Bristol, UK

## Profile

---

Six years of research experience (MSc+PhD) in Video Understanding. Third year PhD researcher in Computer Vision at the University of Bristol, supervised by [Prof. Dima Damen](#). My research focuses on understanding hand-object interactions in two- and three-dimensions, particularly in egocentric setting. I am also a guest scientist at Max Planck Institute for Intelligent Systems working with [Prof. Michael Black](#). Previously, supervised by [Prof. C V Jawahar](#) as part of my 3-years masters in video understanding.

## Expertise

---

### Egocentric / Embodied Video Understanding

- Worked on **3D-aware egocentric perception** with interaction-timeline reconstruction and geometry-constrained reasoning over object's pose across time (useful for manipulation understanding and state tracking). [**P1**]
- Contributed to 3D-grounded egocentric datasets to enable **3D-consistent evaluation** beyond 2D metrics. [**P2**]
- Proposed learning frameworks for **procedure learning** from untrimmed egocentric video using self-supervised and graph-based modelling of temporal structure. [**P6, P7**]

### Generative AI & Multimodal Foundation Models (VLMs / MLLMs)

- Built and benchmarked multimodal evaluation suites for video-language foundation models, targeting spatio-temporal reasoning, activity understanding, and grounding in egocentric video. [**P2, P3, P5, P8**]
- Released large-scale instruction-style QA supervision for egocentric interaction understanding (millions of QA pairs) along with VLM to ground in egocentric view. [**P3**]

### Datasets, Benchmarks & Evaluation at Scale (Ego4D, Ego-Exo4D, HD-EPIC)

- Contributed to **large-scale public datasets** and maintained benchmark + baseline pipelines to drive adoption across academia/industry. [**P5, P8**]
- Built **task definitions** + **metrics** that make model progress measurable (procedure learning, HOI reference/QA, multi-view understanding), accelerating iteration for foundation-model training/eval. [**P2, P3, P7, P8**]

## Publications & Preprints [**P**]

---

(Author order as in papers; \* indicates equal contribution )

1. **Reconstructing Objects along Hand Interaction Timelines in Egocentric Video.** Zhifan Zhu, Siddhant Bansal, Shashank Tripathi, Dima Damen. **CVPR-W 2026**. [paper](#) | [project](#) | [code](#) | [results video](#)
2. **HD-EPIC: A Highly-Detailed Egocentric Video Dataset.** Toby Perrett\* *et al.* (incl. Siddhant Bansal\*). **CVPR 2025**. [paper](#) | [project](#) | [code+data](#)
3. **HOI-Ref: Hand-Object Interaction Referral in Egocentric Vision.** Siddhant Bansal, Michael Wray, Dima Damen. *arXiv* 2024. [paper](#) | [project](#) | [code+data](#)
4. **An Outlook into the Future of Egocentric Vision.** Chiara Plizzari\*, Gabriele Goletto\*, Antonino Furnari\*, Siddhant Bansal\*, Francesco Ragusa\*, Giovanni M. Farinella, Dima Damen, Tatiana Tommasi. **IJCV 2024**. [paper](#)
5. **Ego-Exo4D: Understanding Skilled Human Activity from First- and Third-Person Perspectives.** Kristen Grauman *et al.* (incl. Siddhant Bansal\*). **CVPR 2024**. [paper](#) | [project](#) | [code](#)
6. **United We Stand, Divided We Fall: UnityGraph for Unsupervised Procedure Learning from Videos.** Siddhant Bansal, Chetan Arora, C. V. Jawahar. **WACV 2024**. [paper](#) | [project](#)
7. **My View is the Best View: Procedure Learning from Egocentric Videos.** Siddhant Bansal, Chetan Arora, C. V. Jawahar. **ECCV 2022**. [paper](#) | [project](#) | [code](#)
8. **Ego4D: Around the World in 3,000 Hours of Egocentric Video.** Kristen Grauman *et al.* (incl. Siddhant Bansal). **CVPR 2022 (Oral; Best paper finalist)**. [paper](#) | [project](#) | [code](#)

9. **Improving Word Recognition using Multiple Hypotheses and Deep Embeddings.** Siddhant Bansal, Praveen Krishnan, C. V. Jawahar. **ICPR 2020.** [paper](#) | [project](#) | [code](#)
10. **Fused Text Recogniser and Deep Embeddings Improve Word Recognition and Retrieval.** Siddhant Bansal, Praveen Krishnan, C. V. Jawahar. **DAS 2020 (Oral).** [paper](#) | [project](#) | [code](#)

## Education

### University of Bristol, United Kingdom

PhD in Computer Science

*Apr 2023 – Present*

Advisor: Prof. Dima Damen

### IIT Hyderabad (CVIT), India

MS (by Research) in Computer Science

*Jan 2020 – Mar 2023*

CGPA: 9.2/10.0

**Thesis:** I-Do, You-Learn: Techniques for Unsupervised Procedure Learning using Egocentric Videos ([PDF](#))

**Advisors:** Prof. C. V. Jawahar; Prof. Chetan Arora

### Vishwakarma Government Engineering College, Ahmedabad, India

BE in Electronics and Communication

*Jun 2015 – Jul 2019*

CPI: 8.01/10.0

**Capstone:** Automatic Garbage Detection and Collection (Top 4% SSIP projects; full government funding)

## Research Experience

### PhD Researcher, University of Bristol

Advisor: Prof. Dima Damen

*Apr 2023 – Present*

- **HD-EPIC (CVPR 2025):** contributed to a highly-detailed kitchen egocentric dataset with **41 hours** of unscripted recordings, **3D digital twins**, dense multi-modal annotations, and a **26.6K** VQA benchmark for evaluating VLMs.
- **HOI-Ref (VLM for hand-object interaction referral):** proposed a novel referral task for egocentric images; curated HOI-QA with 3.9M QA pairs; trained VLM4HOI (Llama2-7B based) demonstrating strong gains from egocentric fine-tuning.
- **Ego-Exo4D (CVPR 2024):** contributed to a large consortium dataset for skilled activity understanding with synchronised ego/exo video and multi-modal sensors; supported benchmarks and open-source release practices for reproducible research.
- Contributed to **ROHIT** (Reconstructing Objects along Hand Interaction Timelines; **CVPR-W 2026**) by modeling temporal constraints over object hand-interaction timelines and evaluating constrained optimisation-based pose propagation on egocentric datasets.

### Guest Scientist, Max Planck Institute for Intelligent Systems (MPI-IS), Tübingen

Host: Prof. Michael J. Black

*Sep 2025 – Present*

- Working on hand-object 3D reconstruction (*under review*).

### MS Researcher, Center for Visual Information Technology (CVIT), IIT Hyderabad

Advisors: Prof. C. V. Jawahar; Prof. Chetan Arora

*Jan 2020 – Mar 2023*

- **EgoProceL (ECCV 2022):** created a large-scale procedure-learning dataset (**62 hours**, 130 subjects, 16 tasks) and proposed **Correspond-and-Cut** self-supervised learning using key-step temporal correspondences.
- **UnityGraph (WACV 2024):** proposed a graph-based procedure learning framework that leverages intra-/inter-video context; improved performance on ego procedure learning (+3.6% on EgoProceL).
- **Ego4D (CVPR 2022 Oral):** contributed to the Hands & Objects benchmark effort; built baseline training/evaluation pipelines and supported data governance workflows (IRB, consent, de-identification).
- **Word recognition & multilingual OCR (ICPR/DAS 2020):** improved word recognition via multiple hypotheses + deep embeddings; built OCR for Hindi/Tamil/Telugu and explored semi-supervised CRNN training with CTC loss.

### Research Intern, IIT Gandhinagar

Advisor: Prof. Shanmuganathan Raman

*Mar 2019 – Aug 2019*

- Cultural heritage 3D modeling: point-cloud acquisition and alignment (ICP) and point-cloud completion with autoencoders.

### AI Intern, Meditab Software Inc. (India)

- Evolutionary layout optimization for robotic systems (reported 75% reduction in travel time).

*Sep 2018 – Mar 2019*

### AI Research Intern, Bennett University, India

- Collected a dataset of footprints (180 volunteers) and trained a deep model for footprint-based human classification.

*Jun 2018 – Jul 2018*

### Data Analyst Intern, Bioscan Research, Ahmedabad, India

*Jun 2018 – Jul 2018*

- Built a Python+SQLite patient-tracking tool and signal extraction utilities for an onsite intracranial-bleeding scanning tool.

## Academic Service & Leadership

---

- **Workshop organiser:** EgoVis @ CVPR 2026; EgoVis @ CVPR 2025; EgoVis @ CVPR 2024; Ego4D+EPIC Workshop @ CVPR 2023; Ego4D Workshop @ ECCV 2022.
- **Reviewer:** CVPR2022–2026; ICCV2023&2025 (**Outstanding Reviewer**); ECCV2022&2024; WACV2023–2024; ICRA2025.
- **Journal reviewer:** IEEE TPAMI; IJCV; CVIU.

## Awards & Funding

---

- **Outstanding Reviewer**, ICCV 2025.
- **Google Summer of Code (GSoC)**, 2022: Enhancing DBpedia with image-based querying.
- **Standing Committee Admission** award for MS by Research at IIIT Hyderabad.
- **Student Startup and Innovation Policy:** Top 4% project selected for full funding (Automatic Garbage Collection).
- Supported by a charitable donation to University of Bristol from Meta.

## Technical Skills

---

Programming: Python (primary), SQL, Bash  
DL/ML: PyTorch; HuggingFace/Transformers; TensorFlow/Keras (prior); training & evaluation of VLMs/LLMs;  
CV/Video: OpenCV; video understanding; HOI reasoning; 3D geometry (point clouds, ICP)  
Tooling: Git, Linux, Conda, L<sup>A</sup>T<sub>E</sub>X; Docker; multi-GPU training (`torchrun`)