480 Banbury Road, Oxford OX2 8EN, UK
(+44) 7771001032
piyush@robots.ox.ac.uk
bpiyush.github.io
Scholar | GitHub | LinkedIn

Piyush Bagad

Final-year Oxford DPhil Student in multimodal video understanding advised by Prof. Andrew Zisserman at VGG. My work spans time-sensitive video understanding, generation, and physics inference from multimodal data (audio, tactile). Seeking Research Internship for 2027.

Education

2023–Present **DPhil in Computer Vision**, *University of Oxford*, United Kingdom.
Advised by Prof. Andrew Zisserman; expected completion: October 2027

2021–2023 **MSc in Artificial Intelligence**, *University of Amsterdam*, The Netherlands.
GPA: 9.35/10, *cum laude*; advised by Prof. Cees Snoek

2015–2019 **BS in Mathematics & Scientific Computing**, *Indian Institute of Technology Kanpur*, India.
GPA: 9.30/10

Research Highlights

TARA: Multimodal Video Retrieval (ECCV '26)
- Address nuanced (temporally sensitive, negated and multimodal) queries in video retrieval.
- Repurpose an MLLM as an embedding model using text-only contrastive training with targeted hard negatives.
- Achieves state of the art across nuanced retrieval benchmarks while reducing the video–text modality gap.

RBI: Recognition by Imagination
- Study action recognition from sparse motion stimuli without appearance cues inspired by Johansson's psychophysical experiments.
- Fine-tune a video generator to *imagine* realistic videos from keypoints alone, then *recognize* them with an MLLM.
- Outperforms frontier MLLMs and generalizes to novel actions, datasets and psychophysical stimuli.

Sound of Water: Physics from audio-visual learning (TPAMI '25)
- Infer physical properties of a container–liquid system solely from the sound of pouring.
- Train a physics-inspired audio model using synthetic sounds and visual co-supervision.
- Estimates container geometry, shape and level of water from sound alone, generalizing to in-the-wild YouTube samples.

Selected Publications

ECCV '26 Adapting MLLMs for Nuanced Video Retrieval. **Bagad** & Zisserman. (📄 🐙 🖵)
arXiv '26 Recognition by Imagination: A Generative Approach to Form-from-Motion. **Bagad**, Tapaswi & Zisserman. (📄)
NeurIPS '25 Chirality in Action: Time-Aware Video Representation Learning by Latent Straightening. **Bagad** & Zisserman. (📄 🐙 🖵)
TPAMI '25 The Sound of Water: Inferring Physical Properties from Pouring Liquids. **Bagad** et al. (📄 🐙 🖵)
CVPR '23 Test of Time: Instilling Video-Language Models with a Sense of Time. **Bagad** et al. (📄 🐙 🖵)

Technical Skills and Coursework

Skills Python, PyTorch, DDP; end-to-end ML pipelines for training and evaluating deep neural networks
Research Video understanding, generation, representation learning; multimodal fusion, cross-modal retrieval, audio-visual learning
UvA Deep Learning 1 & 2; Computer Vision 1 & 2
Online CS336: LMs from Scratch (Stanford Online); Flow Matching (MIT)

Research and Industry Experience

2019–2021 **Research Fellow**, *Wadhwani Institute for AI*.
Built monocular-video methods, a 3D neonatal model and a synthetic-data pipeline for newborn weight estimation. Also, developed a method to screen for COVID-19 signals from cough sounds.

2018–2019 **Research Intern**, *Adobe Research India*.
Developed GAN representations for privacy-preserving data sharing and utility analysis.

Selected Awards

2023–2026 Outstanding Reviewer: CVPR 2023/2025, ECCV 2024 and NeurIPS 2025
2023 KHMW Thesis Prize and Amsterdam AI Thesis Prize; Amsterdam Merit Scholarship
2021 Best Paper Award, ML Reproducibility Challenge