Avishkar Saha

avishkar.saha@gmail.com | avishkarsaha.com/ | avishkarsaha | avishkar-saha | google-scholar

Summary

I'm a founding **Research Engineer** at **MireloAI** (https://www.mirelo.ai), where I build **foundation models for video and sound**. My work at Mirelo involves the complete machine learning pipeline — from large-scale data collection/curation through model architecture design to distributed training of multimodal foundation models. I'm particularly interested in building models that can understand and generate content across multiple modalities.

Previously, I interned at AWS in Tübingen and completed a PhD in Computer Vision and Machine Learning at the University of Surrey's CVSSP, under the supervision of Prof. Richard Bowden and Prof. Chris Russell.

My PhD paper on mapping for autonomous driving won the **best paper award** at the International Conference on Robotics and Automation 2022 (ICRA).

Education

University of Surrey — Centre for Vision, Speech and Signal Processing (CVSSP) — *Surrey, U.K*

PhD in Computer Vision and Machine Learning — *Jan.2020 - Jan.2024*

- I developed **deep learning** methods for **3D scene understanding** and **generative models for graphs**.
- **Supervisors:** Prof. Richard Bowden, Prof. Chris Russell & Dr. Oscar Mendez

University of Bath — *Bath, U.K*

MSc in Machine Learning & Autonomous Systems · Grade: 79.8% — *Sep.2018 - Sep. 2019*

- **Thesis title:** "Style invariant graph representations for object recognition"
- **Supervisor:** Prof. Peter Hall

University of Bath — *Bath, U.K*

MEng in Structural Engineering · Grade: First Class Honours, 1:1 — *Sep.2014 - Jun. 2018*

- **Thesis title:** "Spatial ordering with multi-objective particle optimization"

Work Experience

MireloAI — *London, UK*

Research Engineer, First Employee — *Oct. 2023 - Present*

- Building generative models for video-to-sound and text-to-music.
- I led the development of our video-to-sound model, which is now available on **fal.ai:** https://blog.fal.ai/mirelo-now-available-on-fal/
- https://www.mirelo.ai/

Amazon (AWS) — *Tübingen, Germany*

Applied Scientist, Intern — *Dec. 2022 - June 2023*

- Worked on object-centric reasoning in computer vision, resulting in a **NeurIPS submission**.
- Supervisory team: Thomas Brox, Francesco Locatello and Chris Russell.

University of Surrey — *Surrey, U.K*

Teaching Assistant — *Feb. 2020 - June. 2021*

- Teaching assistant for various machine learning modules.

BuroHappold — *Bath, U.K*

Engineering Software Developer — *Jun. 2017 - Sep. 2017*

- Developed a **software toolkit for generating 3D structures**. This was initially implemented for a global team of 10-15 engineers.
- Wrote a **multi-objective particle swarm optimiser** to optimise the 3D structure of Qatar Foundation Stadium, resulting in reduced material usage by 15%.

University of Bath *Bath, U.K*

Lead Ambassador, Structural Engineering Department *Oct. 2016 - May 2017*

- Gave presentations to prospective engineering students on every Open Day, and then ran half-day mini design projects with the students to give them a flavour of the course's major engineering projects.

BuroHappold *Bath, U.K*

Engineering Software Developer *Jun. 2016 - Sep. 2016*

- Implemented **genetic and dynamic mesh relaxation algorithms** to design key 3D structures in the new Tottenham Hotspurs Stadium in London.

Populous *London, U.K*

Computational Designer *Mar. 2014 - Sep. 2015*

- Developed **geometric algorithms for panelling complex surfaces**, which were used to construct the new Tottenham Hotspurs Stadium in London.

BuroHappold *Bath, U.K*

Engineering Software Developer, Intern *Feb. 2013 - Feb. 2014*

- Developed a **software toolkit for analyzing 3D structures**.

Awards

Outstanding Paper Award *Philadelphia, U.S.A*

International Conference of Robotics and Automation (ICRA) *2022*

Basil Spence Award *Bath, U.K*

University of Bath *2018*

Publications

1. (under review) Avishkar Saha, Francesco Locatello, Thomas Brox and Chris Russell. *One step at a time: time-varying causal structures for physical systems*. **Submitted to a major Machine Learning conference**.

2. (under review) Avishkar Saha, Oscar Mendez, Chris Russell and Richard Bowden. *Learning Stable Topologies for Point Cloud Labelling*. **Submitted to a major Computer Vision conference**.

3. Avishkar Saha, Oscar Mendez, Chris Russell and Richard Bowden. ***Learning Adaptive Neighborhoods for Graph Neural Networks***. **ICCV 2023**.

4. Avishkar Saha, Oscar Mendez, Chris Russell and Richard Bowden. ***"The Pedestrian next to the Lamppost" Adaptive Object Graphs for Better Instantaneous Mapping***. **CVPR 2022**.

5. Avishkar Saha, Oscar Mendez, Chris Russell and Richard Bowden. ***Translating Images into Maps*** . **ICRA 2022**. **Oustanding Paper Award.**

6. Avishkar Saha, Oscar Mendez, Chris Russell and Richard Bowden. ***Enabling spatio-temporal aggregation in Birds-Eye-View Vehicle Estimation.*** **ICRA 2021**.

7. James Ross, Oscar Mendez, Avishkar Saha, Mark Johnson and Richard Bowden. ***BEV-SLAM: Building a Globally-Consistent World Map Using Monocular Vision***. **IROS 2022**.

Presentations

International Conference on 3D Vision 2022 *Prague, Czech Republic*

Tutorial: BEV mapping and addressing its shortcomings *September 2022*

- I conducted a half-day tutorial based on our work in mapping images to birds-eye-view for autonomous vehicles.

Amazon *Tubingen, Germany*

Invited talk: sparse representations for scene understanding *July 2022*

- I gave a talk on the use of graphs for 3d object detection and mapping, based off our CVPR 2022 paper.

Computer Vision and Pattern Recognition Conference 2022 *New Orleans, U.S.A*

Presentation: Adaptive object graphs for better instantaneous mapping *June 2022*

- I presented our paper *""The Pedestrian next to the Lamppost" Adaptive object graphs for better instantaneous mapping"*, which proposed a graph-based method for object localization.

International Conference on Robotics and Automation 2022 *Philadelphia, U.S.A*

Presentation: Translating Images into Maps *May 2022*

- For our Outstanding Paper Award, I presented our work *"Translating Images into Maps"*.

Wayve *London, U.K*

Invited talk: image-to-birds eye view mapping *January 2022*

- I gave a talk on our work in mapping images to birds-eye-view for autonomous vehicles.

Skills

Programming languages	Python, C#, JAVA, C++
Core ML libraries	PyTorch, Tensorflow, PyTorch Lightning, NumPy
HPC	AWS, GCP, SLURM
Operating Systems	Unix, Linux
Computer Graphics	Rhino, Maya, Blender

Service to the scientific community

Peer Review Service

1. Conference of Computer Vision and Pattern Recognition (CVPR) 2024
2. Transactions on Image Processing (TIP) 2023
3. Transactions on Pattern Analysis and Machine Intelligence (TPAMI) 2023
4. Conference of Computer Vision and Pattern Recognition (CVPR) 2023
5. International Conference of Computer Vision (ICCV) 2023
6. British Machine Vision Conference (BMVC) 2021

Teaching

1. Teaching assistant for Natural Language Processing COM3029 — University of Surrey (2021)
2. Teaching assitant for Machine Learning and Data Mining — University of Surrey (2020)

Software

`avishkarsaha/translating-images-into-maps` Official PyTorch code for my ICRA 2022 paper (400 stars).

References

Prof. Richard Bowden, University of Surrey *Surrey, U.K*

r.bowden@surrey.ac.uk

Prof. Chris Russell, University of Oxford *Oxford, UK*

chris.russell@oii.ox.ac.uk