Sayan Deb Sarkar

https://sayands.github.io
sdsarkar@stanford.edu ◇ LinkedIn ◇ GitHub ◇ Google Scholar

EDUCATION

Since 2024 PhD in 3D Computer Vision, *Stanford University*, United States
Advised by Prof. Iro Armeni, Gradient Spaces Research Group.

2022 - 2024 MSc in Computer Science, *ETH Zürich*, Switzerland
Advised by Prof. Marc Pollefeys, Computer Vision And Geometry Group. GPA: $5.48/6.0$

2016 - 2020 B.Tech in Information Technology, *Manipal Institute of Technology*, India
Relevant Coursework: Data Structures, Operating Systems. GPA: $9.16/10.0 \approx top$ 1%

EXPERIENCE

Summer 2025 Research Scientist Intern at **Microsoft Spatial AI Lab**, *Zurich*, Switzerland
Designed a scalable and efficient tokenization method for Video LLMs using codec information to enable longer context and faster video understanding. Patent pending [B].
Mentor: Prof. Marc Pollefeys

Autumn 2023 Research Intern at **Qualcomm XR Labs**, *Amsterdam*, Netherlands
Optimized SLAM algorithms for real-time performance for extented reality applications & improved tracking in adversarial scenarios. Patent filed [A].
Mentor: Dr. Marco Manfredi

2021 - 2022 Computer Vision Research Engineer at **Mercedes-Benz R & D**, *Bangalore*, India
Developed deep learning models for driver monitoring and head position estimation in multi-purpose camera systems for the Maybach S-Class under the Interior Assist program.

2022 - 2024 Research Student at **CVG, ETH Zürich**, *Zürich*, Switzerland
3D scene graph alignment in static and dynamic environments, leverage the graph matching to enable embodied agent tasks like map reuse, 3D localization and registration.
Paper published at ICCV '23 [4].
Supervisor: Dr. Dániel Béla Baráth, Dr. Ondrej Miksik & Prof. Iro Armeni

2020 - 2021 Research Engineer at **ICG, TU Graz**, *Graz*, Austria
Joint 3D hand + object pose estimation in close interaction scenarios and indoor 3D scene understanding estimation using Monte Carlo Tree Search on noisy RGB-D scans.
Paper published at CVPR '22 [3], CVPR '21 [2] & ECCV '20 [1].
Supervisor: Dr. Shreyas Hampali, Dr. Mahdi Rad & Prof. Vincent Lepetit

PUBLICATIONS

[7] GuideFlow3D: Optimization-Guided Rectified Flow For Appearance Transfer, *in* **NeurIPS 2025**.
Sayan Deb Sarkar, Sinisa Stekovic, Vincent Lepetit, and Iro Armeni

[6] SGAligner++: Cross-Modal Language-Aided 3D Scene Graph Alignment, *in arXiv 2025*.
Binod Singh*, **Sayan Deb Sarkar***, and Iro Armeni
[Paper] [Project Page]

[5] CrossOver: 3D Scene Cross-Modal Alignment, *in* **CVPR 2025** [Highlight, top 3%].
Sayan Deb Sarkar, Ondrej Miksik, Marc Pollefeys, Dániel Béla Baráth, and Iro Armeni
Featured: Open Robotics.
[Paper] [Project Page]

[4] SGAligner: 3D Scene Alignment with Scene Graphs, *in* **ICCV 2023**.
Sayan Deb Sarkar, Ondrej Miksik, Marc Pollefeys, Dániel Béla Baráth, and Iro Armeni
Featured: Computer Vision News, Learn OpenCV Blog.
[Paper] [Project Page]

[3] Keypoint Transformer: Solving Joint Identification in Challenging Hands and Object Interactions for Accurate 3D Pose Estimation, *in* **CVPR 2022** [Oral, top 4.1%].
Shreyas Hampali, **Sayan Deb Sarkar**, Mahdi Rad, and Vincent Lepetit
[Paper] [Project Page]

[2] Monte Carlo Scene Search For 3D Scene Understanding, *in* **CVPR 2021**.
Sinisa Stekovic*, Shreyas Hampali*, **Sayan Deb Sarkar**, Chetan Srinivasa Kumar, Friedrich Fraundorfer, and Vincent Lepetit
[Paper] [Project Page]

[1] General 3D Room Layout from a Single View by Render-And-Compare, *in* **ECCV 2020**.
Sinisa Stekovic, Shreyas Hampali, Mahdi Rad, **Sayan Deb Sarkar**, Friedrich Fraundorfer, and Vincent Lepetit
[Paper] [Project Page]

PATENTS

[B] Efficient Video Tokenization for Foundation Multi-modal Models
US patent, pending in 2025, by Microsoft.

[A] Learned Occlusion Modeling For Simultaneous Localization and Mapping
US patent, filed in 2024, by Qualcomm. [Patent Link]

INVITED TALKS

June 2025 Scalable Cross-Modal 3D Scene Understanding. *Imagine Labs, ENPC ParisTech*

ACADEMIC SERVICES

Reviewing CVPR, ECCV, ICCV, NeurIPS. ICRA
Organization CV4AEC Workshop@CVPR ('23 & '24)

TEACHING

Winter 2025 Computer Vision For The Built Environment [Course Website]

TECHNICAL SKILLS

Programming Python, C++, Java, JavaScript
Tools Pytorch, Tensorflow, Blender, OpenCV, mySQL, Node.js, Django, mongoDB

EXTRA CURRICULAR

2022 Co-founder, *CORD.ai*
Built and led a core team of 14 to establish a 350+ member community focused on democratizing AI, reducing barriers for young independent researchers, and fostering collaboration.

2020 Technical Head, *defeatCOVID*
Non-profit organisation, aimed at tracking the spread of COVID-19 using a mobile-based heat map interface.