Soumyadip Sarkar

soumya997.sarkar@gmail.com +91-7047289997 soumyadip-sarkar soumya997 soumya997.github.io

Education

Institute of Engineering and Management, Kolkata 2018 - 2022
Bachelor of Technology in Electrical Engineering; CGPA: 8.46/10 *Kolkata, West Bengal*

Experience

Jio Reality Labs - Jio Platforms Limited (JPL) May 2025 - Present
Robotics Software Engineer, Computer Vision and AI *Navi Mumbai*

- Fine-tuned the **NVIDIA GR00T N1.6** VLA model for a 7-DOF humanoid dexterous arm **bottle pick-and-place** task using real and simulated (*Isaac Sim*) data via a custom ROS bag–to–LeRobot pipeline. Calibrated joint gains for sim-to-real consistency. Explored data distribution strategies (perturbation, recovery) (∼12%), camera configurations, domain randomization (∼3%), and **DAgger** (∼7%)—improving closed-loop task success from **65%** to **87%**.
- Deployed fine-tuned GR00T N1.6 as an offboard ZMQ server on an RTX 4090, achieving ∼**100 ms** round-trip latency on robot; optimized with TensorRT to ∼**50 ms** model latency. Built an **evaluation framework** (per Gao et al.) across 20 scenes covering visual, behavioral perturbations with automated success metrics and logging.
- Developed and deployed an end-to-end **Voice Agent** on a humanoid robot. Integrated wake-word detection (Sensory SDK), Silero VAD, and **OpenAI Realtime API** for streaming STT/LLM/TTS. ROS 2 pub-sub and Behavior Tree action servers for **voice to gesture** or locomotion execution, achieving ∼**4 s** e2e latency on hardware.
- Built an optimized on-device **Voice-to-Action** pipeline on NVIDIA AGX Orin 64GB, benchmarking 8 STT models (Whisper variants, Parakeet, Conformer via **NVIDIA Riva**); deployed faster-whisper large-v3 (**12% WER**, Indian English) at ∼**0.6 RTF**, RoBERTa action classifier at ∼**50ms**, and PocketTTS at ∼0.1 RTF; cloud STT selectively invoked for action queries.
- Benchmarked different **Visual** and **LiDAR SLAM** pipelines (VSLAM, RTAB-Map, HDL-Localization, GLIM, FAST-LIO2) on a bipedal humanoid in sim + real, with APE/RPE evaluation (evo) against ground truth. Mitigated gait-induced drift via Madgwick IMU fusion + ArUco; TF publishers, and pointcloud→occupancy-grid pipeline.

DeepVidya AI Private Limited (OpenCV University) November 2023 - April 2025
Computer Vision Engineer *Bengaluru*

- Collaborated with **Garry Bradsky** and team on **indoor re-localization** project, optimizing the **Gaussian Splatting** pipeline to evaluate relocalization results. Reduced processing time from ∼**10** to ∼**2 min** for 3D reconstruction and ∼**4 min** for 3DGS training using Pose Refinement (**+12% SSIM**). Benchmarked various 3DGS methods—including **3DGS MCMC + Bilateral Grid**, TamingGS, and ScaffoldGS on the ScanNet++ dataset.
- Built an automated attendance system by training **MobileFaceNet** and **ResNet50-IR** with ArcFace loss on the AgeDB and Bollywood Celeb datasets. Later, integrated the **AWS Rekognition API** for face identification improving accuracy by $\geq$ **10%**. Tested under diverse conditions, and deployed the application on AWS EC2.

Autonomous Last mIle VEhicle (ALIVE), Infosys CAI, IIIT Delhi July 2022 – October 2023
Research Engineer under Dr. Saket Anand and Dr. Sanjit K. Kaul *New Delhi*

- Designed and developed **ADAS** features including **FCW** and **TLW** with integrated **AEBS**, compliant with AIS standards. Used **YOLOv5** for detection, tracking, and a GMM classifier (96%), along with kinematic modeling. Validated in Carla and deployed on NVIDIA Orin.
- Worked on an **XR test-bed** for autonomous vehicle testing with an RViz-based interface for spawning and visualizing static and dynamic obstacles. Developed a **C++ ROS** package to publish obstacle positions and integrated **obstacle projections into the camera image** stream.
- Developed a pipeline to generate **High-Definition Maps** for the IIITD campus using **LiDAR maps** to create **Lanelet maps**. Used **YOLOPV2** road segmentation and **plane estimation** for point selection, and developed a method to embed key map attributes such as stop-line locations, traffic-light positions, and road slope information.
- Trained **YOLOP** on IDD, BDD and collected IIITD Campus data, with extensive data augmentation and SORT tracker, achieving +0.2 mAP on traffic light detection. Implemented a **C++ ROS** Traffic Light Following package, integrated it into the **Planning stack Behaviour Tree**, and validated end-to-end on **Carla** and **E2O** platforms after porting to **Nvidia Orin**.

Projects

HuBMAP + HPA: Multi-organ FTU Segmentation | *Pytorch, timm*

- Tackled **5-class semantic segmentation** of imbalanced medical tissue from 351 cross-source high-res images; built a custom augmentation pipeline and WandB experiment tracking.
- Progressed from U-Net/U-Net++ baselines → **Swin-Tiny + UpperNet (0.7 LB)** → **CoaT 3-fold (0.75 LB)**, with mixed-precision and gradient accumulation for high-res training.

- Boosted score with stain normalization, mask transpose, HPAHuBMAP pixel-size matching, focal loss as auxiliary, **SWA fold ensembling** + **multi-model TTA**; annotated +30 HuBMAP images via labelme.

Tensorflow-GBR Underwater Starfish Detection | *Python, Pytorch, Ultralytics*

- Initially used **FasterRCNN with default ResNet50 FPN backbone**, then tried various backbones and hyperparameters with geometric and color augmentation techniques, including ResNet101, ResNet50, EfficientB3, and SwinTransformer backbones.
- Introduced a CLAHE based underwater image enhancement technique for better training.
- Switched to **ultralytics yolov5s6 and yolov5m6 on video based splitting**. I tried different hyper parameters in that and different training image resolutions. Also tried yolov5 model freezing.
- **Tricks and Post-processing:** Along with yolov5, tracking was doing a better job increasing the CV/LB, using inference time high resolution worked, **TTA and WBF** was used for ensemble. Using classification on bbox helped increase LB.

Landscape Generation Using Diffusion Model | *Pytorch*

- Scratch implemented Denoising Diffusion Probabilistic Model (**DDPM**) using Pytorch. Two variants of Unet are implemented for DDPM, simple CNN based Unet and **ViT** based Unet. I trained this for generating landscape images.
- I further experimented with Classifier Free Guidance(CFG), on both the models. This **conditional DDPM** setup was trained on CIFAR10 dataset. Experimenting with Linear Attention and Flash Attention.

Object Insertion in Gaussian Splatting | *Python, COLMAP, gsplat, Rerun*

- **Reduced COLMAP reconstruction time 10×** by extracting features at 512×512 and upscaling keypoints for 1024×1024 undistortion — enabling fast SfM on low-res while training 3DGS on high-res.
- Applied **mask-guided 3DGS training** with Floater-Free **MCMC + Bilateral Grid** for large-scene reconstruction; used **KNN-based outlier removal** to clean Gaussian models. composed final scenes by placing trained Gaussian models at target locations.

Technical Skills

Languages:	Python, C++, Bash, SQL
ML / DL:	PyTorch, TensorFlow, Transformers, MCP, RAG, LoRA, LeRobot, TensorRT, NVIDIA GR00T N1.6, Nvidia Riva
Robotics:	ROS2, PCL, Eigen, Ceres-Solver, Behaviour Trees, Lanelet2
3D Vision:	NeRF Studio, gsplat, COLMAP, hloc, OpenCV
Simulators:	Isaac Sim, Carla, Rerun
Edge & Cloud:	Jetson Orin, AWS EC2, AWS Rekognition
Domains:	Computer Vision (2D/3D), Robot Learning, Reinforcement Learning, SLAM, Gaussian Splatting

Achievements & Responsibilities

- **Conducted Session on Radiance Field (NeRF) and Gaussian Splatting in OpenCV Live**
- **4x Kaggle Expert, Competitions Expert**
- **57th - Silver Medal** in Kaggle TF-GBR
- **62th - Bronze Medal** in Kaggle HuBMAP+HPA
- **214th rank -top 15 Pct.** in Kaggle Sartorius- CIS
- 7th at Octahacks:3 hackathon out of 456 teams
- CORD.ai Community Manager

Interests

- Technical Blogging, ML Competitions
- Running, Table Tennis, Badminton, Cricket