Pranav Saxena

✉ pranavsaxena2003@gmail.com | 🔗 pranav-saxena.github.io | in pranav-saxena13 | ⌾ Pranav-Saxena

EDUCATION

Birla Institute Of Technology And Science, Pilani   Goa, India
Bachelor of Engineering in Electronics and Communication   *Oct 2022 – May 2026*

EXPERIENCE

Visiting Researcher   Jun 2025 – Present
Carnegie Mellon University   *Pittsburgh, USA*
- Pursuing my undergraduate thesis at Carnegie Mellon University's Robotics Institute under the supervision of Dr. Wenshan Wang and Dr. Ji Zhang.
- Working on 3D Object Representation Learning.
- Previously worked on Vision-Language Navigation for Outdoor Scenarios and co-organized the CMU-VLA Challenge.

Research Intern   May 2024 – Present
National University of Singapore   *Remote*
- Research intern at MARMoT Lab, NUS under the supervision of Prof. Guillaume Sartoretti.
- Worked on 3D Scene Graph Generation using Vision-Language Models.
- Designed a generalized framework for Open-Vocabulary 3D Gaussian Splatting, eliminating the need for a per-scene autoencoder.
- Developed a Next-Best-View Policy for 3D Reconstruction using Neural Radiance Fields (co-supervised by Prof. Marija Popović).

PUBLICATIONS AND PREPRINTS

Pranav Saxena, Nishant Raghuvanshi, Neena Goveas. *"UAV-VLN: End-to-End Vision-Language Guided Navigation for UAVs"*. European Conference on Mobile Robots 2025 (Oral). [Paper]

Pranav Saxena, Avigyan Bhattacharya, Ji Zhang, Wenshan Wang. *"LLM-RG: Referential Grounding in Outdoor Scenarios using Large Language Models"*. Human-Aware Embodied AI Workshop, IROS 2025. [Paper]

Pranav Saxena, Jimmy Chiun. *"ZING-3D: Zero-Shot Incremental 3D Scene Graphs via Vision-Language Models"*. preprint. [Paper]

Pranav Saxena. *"Gen-LangSplat: Generalized Language Gaussian Splatting with Pre-Trained Feature Compression"*. preprint. [Paper]

PROJECTS

3D Object Representation Learning   Oct 2025 – Present
Carnegie Mellon University
- Creating a generalized 3D encoder that can be used to represent 3D objects.
- The learned embeddings could be used for various downstream tasks like 3D Object Segmentation, 3D Shape Completion, 3D Captioning.

Vision Language Navigation for Outdoor Scenarios | Paper, Code   Jun 2025 – Present
Carnegie Mellon University
- Developed a zero-shot hybrid referential grounding pipeline for outdoor driving scenarios.
- Leveraged off-the-shelf Vision-Language Models (VLMs) for detailed visual understanding and Large Language Models (LLMs) for symbolic reasoning, without end-to-end retraining.
- Achieved 64.72% and 67.91% accuracy on val and test subsets of the Talk2Car dataset, respectively.
- Currently developing a benchmark dataset to further advance referential grounding and multimodal reasoning in dynamic outdoor scenes.

3D Scene Graph Generation using VLMs | [Paper]   Apr 2025 – Sept 2025
National University of Singapore (NUS)
- Developed a zero-shot pipeline for incremental 3D scene graph generation using Vision-Language Models (VLMs), enabling open-vocabulary reasoning and geometric grounding.
- Achieved 97% node labels accuracy on scenes in the Replica dataset and 96% on scenes in the HM3D dataset.
- Subsequently extended the framework for object-goal navigation tasks.

Efficient 3D Open Vocabulary Gaussian Splatting | [Paper], [Code]   Dec 2024 – May 2025
NUS, TU Delft
- Developed a generalized framework for efficient language-grounded 3D Gaussian Splatting eliminating per-scene autoencoder training and improving cross-scene generalization.
- Implemented a cross-scene generalized autoencoder trained on ScanNet for CLIP feature compression into a 16-dimensional latent space.
- Achieved 84.4% localization accuracy on the LERF dataset and 93.3% mIoU on the 3D-OVS dataset, with a nearly $2\times$ reduction in training time.

Next-Best-View Policy for NeRFs | [Code]   May 2024 – Nov 2024
NUS, TU Delft, Purdue University
- Developed an attention-based Next-Best-View (NBV) policy for Neural Radiance Fields to select the most informative viewpoints for 3D reconstruction.
- Implemented a Vision Transformer (ViT) that encodes multi-view image embeddings and candidate poses to predict ΔPSNR gain, guiding view selection.
- Integrated the framework with pixelNeRF to compute ground-truth ΔPSNR via differentiable rendering and supervise policy training under limited measurement budget.

3D Human Pose Estimation using WiFi signals | [Report]   Jan 2025 – Apr 2025
BITS Pilani, Goa
- Developed a WiFi-based 3D human pose estimation pipeline leveraging Channel State Information (CSI).
- Implemented phase sanitization techniques like unwrapping, wavelet denoising, and NLMS filtering to suppress hardware-induced phase noise and multipath distortions.
- Validated the framework on the MM-Fi dataset, achieving improved 3D joint localization and pose fidelity using sanitized CSI signals.

TECHNICAL SKILLS

Languages: Python, C/C++, SQL
Frameworks: ROS, ROS2, PyTorch, TensorFlow, CUDA, Keras, Flask
Tools: Git, Docker, Arduino IDE, RViZ, Gazebo

COURSE WORK

Artificial Intelligence | Applied Stochastic Process | Digital Image Processing | Artificial Intelligence for Robotics | Game Theory and Its Applications | Statistical Inference and Applications | Computer Programming | Linear Algebra | Calculus | Probability & Statistics

HONORS AND AWARDS

Indian Olympiad Qualifier in Astronomy   May 2022
Qualified Part-1 (NSEA). Among top 264 students in India to be shortlisted for Part-2 (INAO). Received Certificate of Merit from Indian Association of Physics Teachers.

Pre-Regional Mathematics Olympiad   Sept 2019
Qualified Pre-Regional Mathematics Olympiad organized by HBCSE. Among the top 300 students in my state to get qualified.

LEADERSHIP

Organizer, CMU-VLA Challenge
Lead organizer of the CMU-VLA Challenge 2025, held as part of IROS 2025. Managed the overall organization and evaluation of the challenge.

Founder and Vice-Captain, Team Artemis
Founded Team Artemis, BITS Pilani-Goa's autonomous drone team. Led the automation subdivision. Ranked 3rd out of 100 teams in Phase-I in nationwide competition, SAE Aerothon'24. Won Best Mission Award in final round.