

Prabhdeep Singh Sethi

E-mail: prabhdes@andrew.cmu.edu | Website: bit.ly/prab | LinkedIn: [prabhdeep1999](https://www.linkedin.com/in/prabhdeep1999) | Github: [Prabhdeep1999](https://github.com/Prabhdeep1999) | Mob: (412) 589-8023
Objective: Seeking Full-Time Computer Vision Roles starting December 2024

EDUCATION

Carnegie Mellon University Pittsburgh, PA
Master of Science in Computer Vision (Robotics Institute, School of Computer Science) (GPA: 4.11/4) Dec 2024
Coursework: Advanced Computer Vision, 3D Vision, Multimodal Learning, Reinforcement Learning, Geometric Methods in Vision

Government College of Engineering, Nagpur Nagpur, India
Bachelor of Engineering in Computer Science (GPA: 9.5/10) Aug 2021
Coursework: Operating Systems, Data Structures & Algorithms, Artificial Intelligence, Databases, Object Oriented Programming

PUBLICATIONS

S Jain*, A Kuthiala*, **PS Sethi**, P Saxena, "StyleSplat: 3D Object Style Transfer with Gaussian Splatting" Oct 2024
NeurIPS 2024, Creative AI Track (In review) [Project Page]

- Introduced the first fast, localized style transfer method for 3D scenes using Gaussian splatting.

PS Sethi*, A Agrawal*, CMS Lezcano*, I Heredia*, "Listen Then See: Video Alignment with Speaker Attention" Feb 2024
Proceedings of CVPR Workshops, 2024 [Project Page]

- Developed cross-modal alignment using audio as bridge between video and language, achieving 82.06% SOTA on Social IQ 2.

R Zawar, **PS Sethi**, R Roy, "Jensen-Shannon Divergence in Safe Multi-Agent RL" Dec 2023
ICLR, Tiny Paper Track, 2024 [Paper]

- Proposed a safer alternative in MACPO with JS divergence, achieving higher rewards and reduced costs in multi-agent systems.

EXPERIENCE

Apple Sunnyvale, CA
Computer Vision Research Intern, Vision Pro Team May 2024 - Aug 2024

- Developed an automated pipeline for immersive environments using SOTA 3D reconstruction for AR/VR applications.
- Led the end-to-end proof of concept development, integrating research-driven optimizations, and created a data capture app.

Wobot Intelligence New Delhi, India
Computer Vision Engineer-II Feb 2022 - Aug 2023

- Led a 6 member team to deliver vehicle and person Re-Identification solutions, serving 1M+ cameras and 10,000+ customers.
- Implemented an attribute-based fuzzy search with local and global attention. Further utilized VAE for dimensionality reduction. This approach reduced false IDs by 65% and improved Rank-1 of ReID by 35% in our multi-camera object tracking algorithm.

Solar Industries India Ltd. (Research and Development Lab) Nagpur, India
Senior Computer Vision Researcher Aug 2021 - Jan 2022

- Designed the pipeline for product inspection of critical military parts by using a Vision Transformer (ViT) for object detection, achieving 96.5% mAP for detecting 9 such parts. Deployed models using Nvidia Triton for enhanced operational efficiency.
- Developed Overspeeding & ANPR solutions with YOLOv4, PaddleOCR, DeepSORT for tracking & speed calc.

RESEARCH

Debiasing Question Dependency in VQA models (Multimodal Learning)
Advisor: [Dr Louis-Philippe Morency](#), Multicomp Lab Sep 2024 - Current

- Addressed VQA bias causing LLM reliance on question cues by designing a late fusion strategy to integrate video & transcript.
- Increased video and transcript dependency by 5% and a 3.5% overall improvement across datasets compared to baselines.

Generalizable Sparse view 3D Object Reconstruction (3D Gaussian Splatting) [Project Page] Jan 2024 - Current
Advisor: [Dr Shubham Tulsiani](#), Physical Perception Lab

- Developed a single-forward pass network to predict 3D Gaussian attributes for objects in sparse (4-view) and unposed settings.
- Enhanced generalization by using transformer blocks to capture latent 3D representations for scalable 3D modeling.

PROJECTS

GIF Tune (Video Generation) [Project Page] Feb 2024 - Apr 2024

- Developed GIF-Tune, a one-shot tuning strategy enabling continuous text-to-GIF synthesis from text prompts.
- Enhanced the output using depth-conditioned Stable Diffusion and 3D temporal attention layers from a single text-GIF pair.

SKILLS

Languages and Frameworks: Python, C++, Swift, Bash, SQL; PyTorch, PyTorch3D, TensorFlow, Numpy, OpenCV, Flask
Tools and Platforms: Docker, Triton, Git, DeepStream, TensorRT, AIMET, AWS, Azure