

Agneet Chatterjee

Ph.D. Student (2nd Year)
Computer Science
Arizona State University

Email: agneet@asu.edu

[Website](#)
[Google Scholar](#)
[LinkedIn](#)

EDUCATION

Arizona State University

Ph.D. in Computer Science

Tempe, AZ
Spring 2023 –

- **Advisors:** Dr. Chitta Baral, Dr. Yezhou Yang
- **CGPA:** 4.0
- **Awards:** SCAI Doctoral Fellowship, SCAI Conference Award, Engineering Graduate Fellowship

Jadavpur University

B.E (Hons) Computer Science

Kolkata, India
2015-2019

RESEARCH INTERESTS

I work at the intersection of computer vision and natural language processing. My current research interests are :

- Improving spatial reasoning abilities in generative vision-language models.
- Solving low-level vision tasks with natural language guidance.

Publication Venues: CVPR NAACL

TECHNICAL SKILLS

Languages : Python, C++, JAVA, R, MATLAB

Libraries & Tools : PyTorch, OpenCV, NumPy, Matplotlib, Kornia, Diffusers, Version Control, GIT, Scikit-Learn, TensorBoard, Diffusers, Linux

PUBLICATIONS

1. On the Robustness of Language Guidance for Low-Level Vision Tasks: Findings from Depth Estimation [CVPR 2024](#)

[Agneet Chatterjee](#), [Tejas Gokhale](#), [Chitta Baral](#), [Yezhou Yang](#)

2. Accelerating LLM Inference by Enabling Intermediate Layer Decoding

[NAACL 2024 \(Findings\)](#)

[Neeraj Varshney](#), [Agneet Chatterjee](#), [Mihir Parmar](#), [Chitta Baral](#)

SELECTED PROJECTS

1. Improving Spatial Consistency in Text-to-Image Models with Spatially Focused Captions

Under Review, 2023

- * We create SPRIGHT, the first spatially focused, large scale vision-language dataset by re-captioning ~6 million images from 4 widely used existing datasets. To demonstrate the efficacy of SPRIGHT, we fine-tune baseline Stable Diffusion models on a small subset of our data and achieve performance gains across multiple spatial reasoning benchmarks while improving the corresponding FID and CMMD scores.
- * We achieve state-of-the-art performance on spatial relationships by developing an efficient training methodology; specifically, we optimize over a small number (<500) of images which consists of a large number of objects, and achieve a 41% improvement over our baseline model.
- * Through multiple ablations and analysis, we present our diverse findings related to spatial relationships: the impact of long captions, the trade-off between spatial and general captions, layer-wise activations of the CLIP text encoder and improvements over attention relevancy maps.

2. Enabling Spatial Fidelity in Vision-Language Models with Rendering Tools

Under Review, 2023

- * We develop the REVISION framework, a 3D rendering pipeline that is guaranteed to generate spatially accurate synthetic images, given an input text prompt. An extendable framework, REVISION currently accomodates 100+ assets across 11 spatial relationships and 3 diverse backgrounds.
- * Utilizing images from REVISION, we improve upon the spatial shortcomings of existing T2I models in an efficient and training-free manner, achieving competitive performance across multiple benchmarks. Over baseline methods, we achieve a 58.6% improvement on the VISOR Score and attain a spatial score of 0.336 on T2I-CompBench.
- * We also create the REVISION benchmark to evaluate the spatial reasoning abilities of Multimodal Large Language Models (MLLMs). Evaluating 5 state-of-the-art models across 16 question types, we reveal their shortcomings in reasoning over complex spatial questions and their vulnerability to adversarial perturbations.

EXPERIENCE

Salesforce

Software Engineer 2

June 2019 - July 2022

Hyderabad, India

- * Lead developer on a 15-person global team in the QTC domain, aiming to unlock a revenue of \$1B in India by FY27.
- * Spearheaded the development and deployment of a full-stack application to automate 70k monthly emails, creating a modular and configurable microservice-based subscription model that was adopted organization-wide, with potential earnings of \$7M+.
- * Developed a scalable, OAuth 2.0-secure, data extraction solution utilising Mulesoft and Heroku, facilitating 35M+ transactions and achieving a 200x performance improvement over the preceding reporting-based solution.

Samsung R&D Institute

Research Intern

May 2018 – July 2018

Bangalore, India

- * As part of the the Graphics and VR team, worked on Physically Based Rendering (PBR) of 3D models creating a C# and Java-based application using Unity's Rendering API's.

OPEN-SOURCE CONTRIBUTIONS

SunPy

Open-source Contributor

2017

- * Python-based implementation of a Multi-Scale Gaussian Normalisation Algorithm, as well as refactoring of existing codebase, released as part of v0.8. [🔗](#)

HONORS AND AWARDS

- SCAI Doctoral Fellowship, 2024 (\$2,000)
- SCAI Conference Award, 2024 (\$500)
- Engineering Graduate Fellowship, 2023 (\$2,000)
- 1st position - 2020, Summer Hackathon (Salesforce)
- INSPIRE Scholarship - 2020, Govt. of India

SERVICE

- Reviewer for CVPR workshop (Open-Domain Retrieval Under a Multi-Modal Setting)
- Project Mentor, CSE 576 - Natural Language Processing, ASU

RESEARCH AREAS

Computer Vision
Text-to-Image Generation

Natural Language Processing
Reasoning & Robustness

Diffusion Models

Multimodal Large Language Models