Manu Gaur

Research Assistant, **IIIT-Hyderabad**

+91 9811899789 manugaurdl.github.io manugaurwork@gmail.com

Education

Delhi Technological University, India — 08/2019 – 05/2023
B.Tech in Applied Physics (GPA: 8.5/10)

Experience

CVIT, IIIT Hyderabad, *India* — 09/2023 - Present
Research Assistant with Dr. Makarand Tapaswi

Amazon Research, *India | International ML Team* — 02/2023 – 07/2023
Applied Scientist Intern with Dr. Vinay Kumar Verma

University of Technology Sydney, *Remote* — 01/2022 – 01/2023
Research Intern with Dr. Mukesh Prasad

Delhi Technological University, *India* — 09/2022 – 05/2023
Research Intern (Bachelor Thesis) with Dr. Dinesh Kumar Vishwakarma

Taaza, *India* — 06/2022 – 07/2022
Machine Learning Intern with Mr. Sandeep Raheja

Publications

[1] **No Detail Left Behind: Revisiting Self-Retrieval for Fine-Grained Image Captioning**
Manu Gaur, Darshan Singh S, Makarand Tapaswi
Transactions on Machine Learning Research, 2024 — Arxiv | Website

[2] **Detect, Describe, Discriminate: Moving Beyond VQA for MLLM Evaluation**
Manu Gaur, Darshan Singh S, Makarand Tapaswi
ECCV 2024 Workshop on Emergent Visual Abilities and Limits of Foundation Models — Arxiv | Website

[3] **Self-Supervised Ensembled Learning for Autism Spectrum Classification.**
Manu Gaur, Kunal Chaturvedi, Dinesh Kumar Vishwakarma, Mukesh Prasad
Journal of Research in Autism Spectrum Disorders, Elsevier, 2023 — ScienceDirect

Other Featured Projects

MLLM inversion: Text-conditioned visual representations — 05/2025 – Present
Advisors: Dr. Yuki Asano, Dr. Makarand Tapaswi — IIIT Hyderabad

* Steer visual features to attend to objects of interest by conditioning encoders with text embeddings.
* Instilling text understanding in visual encoders through grounding at the ViT feature level.
* Resulting representations are more suitable for visual reasoning and dense downstream tasks.

Guiding MLLMs with Policy Gradients — 12/2023 – 01/2024
Advisors: Dr. Makarand Tapaswi — IIIT Hyderabad

* Reproduced ClipCap's (a simple MLLM) captioning performance via next-token prediction training on COCO.
* Maximized different rewards (CIDEr, SR) with REINFORCE for better retrieval and captioning performance.
* Both model and optimization code *written from scratch* in PyTorch.

Contextualized Visual Compatibility with GNNs 03/2023 – 07/2023
Advisors: Dr. Vinay Kumar Verma, Prateek Sircar Amazon Research

* Trained GNNs to model visual compatibility between fashion items based on aesthetics.
* Curated Fashion and Furnishing datasets, each with 5.3M and 8.2M compatible groups of segmented images.
* GNN encoder generates contextualized multimodal product embeddings by conditioning on compatible items.
* During inference, relational information captured by GNN is encoded in new nodes through similarity edges.

Self-Supervised Learning for Modelling Fashion Compatibility 03/2023 – 04/2023
Advisors: Dr. Vinay Kumar Verma Amazon Research

* Learnt shape-invariant, yet style-variant representations through self-distillation of visually compatible items.
* Employed Triplet loss with hard negatives using color, occasion, gender and product information.
* Implemented a non-contrastive region-matching objective for improved perfromance and faster convergence.

Label-Efficient ADHD Classification using 4D rs-fMRI 09/2022 – 03/2023
Advisors: Dr. Dinesh Kumar Vishwakarma Delhi Technological University

* Reduced redundant features within adjacent frames by sampling across different temporal neighbourhoods.
* Self-supervised visual features are learnt using self-distillation across multiple views.
* Transformer models temporal relations across the time-series of spatial features and optimizes cross-entropy.

Constrained Vehicle Routing Optimization 06/2022 – 07/2022
Advisors: Mr. Sandeep Raheja Taaza

* Worked on vehicle routing problem with pickup/delivery services and time constraints for a NEMT client.
* Implemented pointer network and stored geocoding data with geometry projections and spatial references.
* Extended PostGIS database to provide geospatial routing and other functionalities using pgRouting library.

Featured Positions

Machine Learning Summer School, **Amazon** 07/2022
Computer Vision and AI Summer School, **CVIT**, **IIIT-H** 07/2022 – 08/2022
Reviewer: CVPR'25

Technical Skills

Languages: Python, Java, C, MATLAB, SQL(Postgres)
Frameworks: Pytorch, DGL Tensorflow/Keras, PostGis

Featured Coursework

Mathetmatics: Probability and Statistics (5^{th} Sem., DTU); MIT RES-6-012: Intro. to Probability; MIT-OCW: Linear Algebra; Mathetmatics for ML (Imperial College London); Computational Methods (4^{th} Sem., DTU)
Programming: Algorithms I and II (4^{th} Sem., DTU); Database Management Systems (6^{th} Sem., DTU)
Machine Learning: Machine Learning (5^{th} Sem., DTU); Deep Learning Specialization (Deeplearning.ai); Introduction to Reinforcement Learning (University College London)
Computer Vision: Computer Vision (8^{th} Sem., DTU); CS231n: DL for Computer Vision (Stanford University); Introduction to Self-Driving Cars, State Estimation and Localization (University of Toronto)