Nakul Sharma

0xnakul.sh@gmail.com | Google Scholar | LinkedIn | GitHub

Education

Indian Institute of Technology Jodhpur Dec 2020 – May 2024
Bachelor of Technology in Artificial Intelligence and Data Science *CGPA: 8.08/10*
Director's Prize for Best Academic Innovation Work

Disha Delphi Public School, Kota Apr 2016 – Mar 2020
Class XII: Physics, Chemistry, Mathematics, English and Computer Science *Grade: 96.0%*

Publications

- Shreya Shukla*, **Nakul Sharma***, Manish Gupta, and Anand Mishra. *PatentLMM: Large Multimodal Model for Generating Descriptions for Patent Figures*, in Proceedings of AAAI 2025. (*: equal contribution)
- **Nakul Sharma**, Aditay Tripathi, Anirban Chakraborty and Anand Mishra. *Sketch-guided Image Inpainting with Partial Discrete Diffusion Process*, in Proceedings of CVPR Workshops 2024.
- **Nakul Sharma**, Abhirama Subramanyam Penamakuri, and Anand Mishra. *Contrastive Multi-View Textual-Visual Encoding: Towards One Hundred Thousand-Scale One-Shot Logo Identification*, in Proceedings of ICVGIP 2022.

Experience

ML Research Engineer Jan 2024 – Present
SpreeAI *Remote*
- Led the development of state-of-the-art in-house virtual try-on models using Stable Diffusion 2 & 3.5 as backbone.
- Experimented and ablated different model designs and data choices at the scale of millions of data points and billions of parameters.
- Currently working on the pose-transfer problem, incorporating garment information for better synthesis.

Machine Learning Intern – Think Tank Team May 2023 – Aug 2023
Samsung Research America *Mountain View, CA*
- Worked on LLMs for automation of electronic devices: synthesized data from internal automation config files and leveraged LoRA for fine-tuning LLMs for the task.
- Proposed a method to distill knowledge from language models to vision models, outperforming CLIP for zero-shot classification on CIFAR-10 and CIFAR-100 dataset when pre-trained with MS-COCO dataset.

Computer Vision Research Intern May 2022 – Jul 2022
Bosch India
- Worked on the task of lost cargo identification using driving videos, positioned as self-supervised outlier detection in low-shot and zero-shot settings.
- Proposed additional loss in the SimCLR framework to capture the feature space of outliers, which improved the recall by 5% over baselines in low-shot setting.

Open Source Contributor Feb 2021 – Mar 2021
ML-DL-implementation | GitHub | PyPI
- Developed the following deep learning components for the library and integrated them into the existing codebase: tensors and autograd (automatic differentiation) engine.
- Developed modules for use in the optimizers and Neural Networks that would be useful for customizing models.

Research Projects

Exploring VAE encoders in Multi-modal LLMs Jan 2025

Self-exploration

- Experimented with VAE encoders as vision encoders in LLaVA-1.5 framework. The study included VAEs from StableDiffusion 2, Stable Diffusion XL and Flux models. Concluded that traditional CLIP vision encoders produce more discriminative features than VAEs for LLMs to work on.

Almost Free Lunch for Imbalanced Classification using Vision Models July 2024 - Aug 2024

Self-exploration

- Developed an efficient framework based on latent feature oversampling using foundation vision models. The proposed method trains a simple logistic regression model that can be trained in 5 minutes and outperforms state-of-the-art approaches on the imbalanced CIFAR-10 and CIFAR-100 datasets.

Description Generation for Patent Figures May 2023 – Aug 2024

Supervisor: Dr. Anand Mishra, IIT Jodhpur & Dr.Manish Gupta, Microsoft Bing

- Curated a specialized dataset PatentDesc-355K tailored to the unique nuances of patent images for the description generation task. The dataset contains $\sim 355K$ patent images along with their brief and detailed descriptions. We proposed PatentLMM, a multi-modal LMM for the task which comprises of a multi-model encoder for patent images trained using weakly supervised objectives on 900k patent images, and a domain adapted LLaMA LLM.

Sketch-guided Image Inpainting July 2022 – Aug 2023

Supervisor: Dr. Anand Mishra, IIT Jodhpur & Dr. Anirban Chakraborty, IISc Banglore

- Curated a dataset from MS-COCO dataset for the novel task of object-level sketch-guided image inpainting. Proposed Partial Discrete Diffusion formulation for the task aligning forward and reverse diffusion process. The proposed approach outperformed SOTA techniques including the ControlNet in the guided image inpainting literature by significant margins margins.

One-Shot Logo Identification Sept 2021 – May 2022

Supervisor: Dr. Anand Mishra, IIT Jodhpur | *Project Website*

- Worked on One Shot Logo Identification problem in the open-set setting using Deep Contrastive Learning where the goal is to learn good representations of logo images for open-set retrieval. Formulated a supervised contrastive representation learning method utilizing text embeddings from logos, and extended the proposed setup for end-to-end one-shot detection in natural scene images, and to 100k scale for unseen logos.

Achievements

- Best poster presentation award at UG Research Day, IIT Jodhpur for one-shot logo recognition work.
- 2nd position out of the 22 participating IITs for the High-Prep event of "Model Extraction for Video Classification" in the Inter-IIT Tech Meet 10.0.
- 4th position in the NLP Hackathon organized by Samsung on the problem statement related to the Intent Identification and Slot classification of the Hinglish dialogues.
- 100/100 Score in ABU Robocon 2020 Stage-I as part of the IIT Jodhpur Contingent.
- Among top 0.7% candidates who appeared for the Joint Entrance Exam 2020 (JEE-Main 2020).

Relevant Projects

Sketch-guided Object Localization | Course Project Mar 2023 – April 2023

- Explored and implemented baselines for the task of sketch-guided object localization.
- Experimented and adapted the UP-DETR framework for the task to formulate a semi-supervised approach that uses sketchified version of the images for training.

Unsupervised Adversarial Attack Mitigation through Denoising | Course Project Mar 2023 – April 2023

- Formulated a denoising approach for adversarial attack mitigation using dual encoder architecture.
- Exploited networks trained through SimCLR framework to adapt the previous SOTA approach unsupervised.

Model Extraction Attack For Video Classification | Inter-IIT Tech Meet 10.0 Feb 2022 – Mar 2022

- Led the efforts for Bosch's High Prep Statement of "Model Extraction Attack For Video Classification" where the goal was to extract the Swin-T and MoViNet-A2 models for video classification on the Kinetics dataset in greybox and blackbox setting.
- Adapted the Data Free Model Extraction approach to work with videos and formulated various strategies for the greybox attack approach.

Aspect Based Sentiment Analysis Dec 2021

- Implemented various methods using models ranging from SVM to BERT for the task of classification of the sentiment expressed by a phrase in app reviews.
- Deployed the best performing method (based on DistilBERT) to the Amazon AWS Lambda service in a serverless container and served it through an API endpoint.

craneGPT | GitHub Jul 2021

- Finetuned the GPT-2 transformer model available in the HuggingFace library on the transcripts of a popular T.V. sitcom, F.R.I.E.N.D.S. using Jax+Flax to generate new scripts for the show.
- Explored the architecture and functioning of large language models and studied the bias embodied in them through the training data and the effect of data used for finetuning on the pre-induced bias.

Face Recognition at scale | GitHub May 2021 – Jul 2021

- The project aimed to enable the students of IIT Jodhpur to retrieve their photos from a large gallery using only a few of their facial images.
- Analyzed a few pre-trained facial detectors for performance v/s latency trade-off along with retrieval strategies and integrated them into the rest of the pipeline.

Technical Skills

General Skills: Machine Learning, Deep Learning, Natural Language Processing & Computer Vision
Languages: Python, LaTeX, C++, SQL
Frameworks & Libraries: PyTorch, PyTorch Lightning, Jax, Flax, Keras, NumPy, HuggingFace, pandas, scikit-learn, Flask, Weights & Biases, HuggingFace transformers
Development: HTML, CSS, Git

Relevant Coursework

Institute Courses: Data Structures and Algorithms, Design and Analysis of Algorithms, Principles of Computing Systems, Signals and Systems, Calculus, Linear Algebra and Differential Equations, Probability, Statistics and Stochastic Processes, Pattern Recognition and Machine Learning, Artificial Intelligence, Optimization for Machine Learning, Data Engineering, Advanced Machine Learning, Computer Vision, Natural Language Processing, Dependable AI, Speech Processing.
MOOCs: Neural Networks and Deep Learning, Improving Deep Neural Networks, Structuring Machine Learning Projects, Convolutional Neural Networks [Courses from the Deep Learning Specialization, Coursera].
Online Courses: Deep Learning for Computer Vision (CS231n, Stanford), Natural Language Processing with Deep Learning (CS224n, Stanford).

Teaching and Services

Reviewer for CVPR	*2025-*
Co-Reviewer for ACL ARR	*2024-*
Reviewer for CVPR NTIRE Workshop	*2024*
Technical Head, Prometeo @ IIT Jodhpur	*2023*
Society Coordinator, AI Society @ IIT Jodhpur	*2022-2023*
Coordinator, Google Developer Student Club @ IIT Jodhpur	*2022-2023*
Teaching Assistant, Introduction to Machine Learning	*2022*