Arkanath Pathak

pathak.arkanath@gmail.com • arkanath.com

INTERESTS

Machine Learning research, especially interested in Multimodal Learning, Generative Modeling and Representation Learning

EXPERIENCE

Media Integrity, Google DeepMind | *New York (previously in Google Research and in San Francisco)* Nov 2020 - Present

- Research on developing methods to extract image-descriptive sentences from large amounts of text
  Key problems: *Multimodal retrieval embeddings, Multimodal LLM skill distillation, LLM in-context learning*
- Research on studying generative model image outputs and towards building efficient detectors
  Key problems: *GAN architectures, Diffusion architectures, Diffusion reconstruction*
- Researching diffusion models for text
  Key problems: *AR encodings, Text VAEs, Sampling methods*
- Research and engineering for inferring image context from web references, to build a tool for investigating misinformation
  Key problems: *Structured response from LLMs*

YouTube Brand Safety, Google | *Mountain View* Nov 2017 - Nov 2020

- Research on YouTube Brand Safety classifier
  Key problems: *Distribution shift & drift, Ensemble learning, Active learning*
- Research on video prediction with large neural networks
  Key problem: *Increasing model capacity*
- Research on robotic grasping
  Key problem: *3D shape prediction from images*

Cloud Support, Google | *Sydney* Sep 2016 - Nov 2017

- Implemented tool to automate the process for service escalation communications to customers

YouTube Personalization, Google (Internship) | *Bangalore* May 2015 - July 2015

- Exploratory research for improving YouTube personalization using clusters of videos

Indian Statistical Institute (Internship) | *Kolkata* May 2014 - June 2014

- Clustering of mixed data by integrating fuzzy, probabilistic and collaborative clustering

EDUCATION

Indian Institute of Technology Kharagpur | *Kharagpur* 2012 - 2016

Bachelor of Technology (B.Tech.) in Computer Science and Engineering, GPA 9.4/10

PUBLICATIONS

[1] AMMeBa: A Large-Scale Survey and Dataset of Media-Based Misinformation In-The-Wild
Nicholas Dufour, **Arkanath Pathak**, Pouya Samangouei, Nikki Hariri, Shashi Deshetti, Andrew Dudfield, Christopher Guess, Pablo Hernández Escayola, Bobby Tran, Mevan Babakar, Christoph Bregler
preprint, 2024

[2] Sequential training of GANs against GAN-classifiers reveals correlated "knowledge gaps" present among independently trained GAN instances
Arkanath Pathak, Nicholas Dufour
IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2023

[3] High Fidelity Video Prediction with Large Stochastic Recurrent Neural Networks
Ruben Villegas, **Arkanath Pathak**, Harini Kannan, Dumitru Erhan, Quoc V. Le, Honglak Lee
Neural Information Processing Systems (NeurIPS), 2019

[4] Learning 6-DOF Grasping Interaction via Deep Geometry-Aware 3D Representations
Xinchen Yan, Mohi Khansari, Yunfei Bai, Jasmine Hsu, **Arkanath Pathak**, Abhinav Gupta, James Davidson, Honglak Lee

IEEE International Conference on Robotics and Automation (ICRA), 2018

[5] A Two-Phase Approach Towards Identifying Argument Structure in Natural Language
Arkanath Pathak, Pawan Goyal, Plaban Bhowmick
3rd Workshop on Natural Language Processing Techniques for Educational Applications (NLPTEA), 2016

[6] Clustering of Mixed Data by Integrating Fuzzy, Probabilistic, and Collaborative Clustering Framework
Arkanath Pathak, Nikhil R. Pal
International Journal of Fuzzy Systems (IJFS), 2016

MISCELLANEOUS

- Attended ICML 2017, 2019, 2022, 2023, 2024, CVPR 2023, 2024, NeurIPS 2019.
- Completed Deep Multi-Task and Meta Learning course, offered by Stanford University, with grade A in fall 2020.
- Attended Bay Area Deep Learning school, held at Stanford University in 2016.
- Developer of various free-to-use applications listed here. Developed a personal assistant chatbot in 2015.
- Previously involved in competitive programming: Codeforces (achieved Candidate Master), TopCoder, CodeChef.
- Completed Machine Learning course on Coursera by Andrew Ng. in 2013 with 100% grade.
- Selected in National Standard Examination in Physics in 2011.
- Awarded the KVPY scholarship for basic sciences in 2010, 236 scholars selected nationally for this merit.
- Selected in Regional Mathematical Olympiad from Delhi region in 2009 and 2010 (secured rank 8).

I am applying for admission to the Ph.D. in Computer Science program. My research interests lie in Machine Learning and Artificial Intelligence, primarily in Multimodal Learning, Representation Learning, and Generative Modeling.

I completed my undergraduate study at IIT Kharagpur in 2016, and since then, I have worked at Google on several ML projects. Now, I am eager to return to academia because I am drawn to the intellectual challenge of fundamental research. I want to research solutions that address critical limitations in the field and push the boundaries of what's possible in the future. My experience in the industry has equipped me for a successful PhD for two primary reasons. First, I have gained extensive background tackling various machine learning problems, both general challenges and specific research aligned with my current interests. Second, I am highly self-motivated and proficient at identifying and pursuing promising avenues autonomously.

My foundation in research was laid through an internship in 2014, guided by Prof. Nikhil R. Pal at the Indian Statistical Institute. We formulated a new approach for clustering data with mixed feature types, leveraging insights from the collaborative clustering framework. The need to survey papers, formulate an approach, design experiments, and rigorously analyze results offered me an initial glimpse into the world of academic research. The sense of expertise at the end of the project fueled my passion for research. For my bachelor's thesis project, I worked in NLP on Argument mining (extracting argument structures from text articles). My main contribution was using word embeddings to train a classifier to detect argumentative support or attack relationships between text nodes. I was particularly fascinated by word embeddings and how they capture semantic relationships, which inspired my approach to this project.

After graduation, I joined Google Cloud in Australia for a temporary position due to US Visa issues. While in that role, I initiated a part-time engagement in open-ended research under the guidance of Prof. Honglak Lee, where I would work on the collaboration for one day every week. I contributed to the ICRA 2018 paper "Learning 6-DOF Grasping Interaction via Deep Geometry-Aware 3D Representations" and the NeurIPS 2019 paper "High Fidelity Video Prediction with Large Stochastic Recurrent Neural Networks." For the ICRA work, I was responsible for scaling up the 3D shape prediction network to higher-resolution images. For the NeurIPS work, I actively participated in experimentation from inception to completion.

In 2017, I joined the YouTube Brand Safety classifier modeling team. This large-scale deep multi-task classifier is trained on millions of labeled videos. I independently led research efforts in three directions. First, I designed an in-graph ensemble learning approach using classifier subnetworks and an ensemble layer. I devised an alternating schedule for training, optimizing subnetworks independently and then the ensemble layer, which outperformed a standard mean ensemble. Second, my clustering-based active learning approach prioritized labeling diverse clusters of unlabeled videos. Third, I discovered concept drift in the YouTube data. In addition to importance weighting, I trained ensembles on monthly data to learn from the drifting data distribution. I achieved significant performance gains through these efforts, which were recognized through a promotion to senior software engineer.

For my next and current role, I joined Google Research as part of the Media Integrity team. I

lead research for inferring contextual background for an image using the visual content and associated text articles. Prioritizing groundedness and interpretability, my initial approach employed two steps: identifying sentences that describe or discuss the image and then extracting contextual concepts from such sentences. Traditional image-text models are trained on tasks like captioning or VQA and don't capture these relationships. I designed a sequence model to detect the image-text alignment, where a transformer takes as input the pre-trained (and fine-tuned during training) CLIP encodings for both image and sentence semantics. Last year, I shifted my focus to using LLMs for image context extraction. To improve the training of my image-text alignment model, I used the multimodal Gemini LLMs to generate data for weak supervision. I'm now experimenting with using LLMs directly for the image context extraction task through Retrieval Augmented Generation (RAG). To that end, I'm developing a benchmark to measure the performance of retrieval embeddings in this context.

In my current role, I also make time to work on fundamental research on generative models. My research on GANs investigated the phenomenon of shared artifacts across outputs from independently initialized GAN instances. In my approach, a learned classifier is incorporated as an additional, frozen discriminator during GAN training to study the classifier's influence on GAN outputs. I found that a high-capacity GAN like StyleGAN2 would easily avoid the set of artifacts learned by the classifier and move to a new space of artifacts to be found in the GAN outputs. Surprisingly, this again led to a set of shared artifacts across new GAN instances, and this intriguing phenomenon persisted as we kept introducing new frozen discriminators. I published this work at CVPR 2023 titled "Sequential training of GANs against GAN-classifiers reveals correlated "knowledge gaps" present among independently trained GAN instances."

At Google, I chose roles that provided the autonomy to lead ML efforts and grow independently. Additionally, I found part-time collaborations to learn from working with experts. However, working in the industry does not provide the optimal environment for me to work on problems that would change the field fundamentally. In the long run, my ambition is to make groundbreaking contributions that redefine the landscape and shape the future of the field.

For my PhD, I want to focus on multimodal and representation learning, two closely related topics that would drive the next breakthroughs. Learning the most effective representations from multiple modalities that capture different world views would foster a new phase in machine learning. With the advent of LLMs, many research efforts nowadays focus on improving their extraordinary abilities. However, there are still limitations inherent to the current LLM design. At Google, I collaborated on a project on fine-tuning representations learned by autoregressive LLMs to train a masked diffusion model. Ideas like this could address some limitations in autoregressive modeling. Another direction I'd like to explore is whether we can comprehend and distill the skills acquired by large models and reduce the parametric world knowledge. I am eager to tackle these challenges during my PhD, and I am drawn to the strong ML research community at Columbia. My interests align with the recent research by Professors Carl Vondrick, Richard Zemel, and Zhou Yu. I'm particularly excited by Prof. Vondrick's research in multimodality and video prediction, and his work on representations, such as the interpretability of LLM representations for reasoning and the use of discrete token representations to improve vision transformers.