My primary research interests lie in the intersection of machine perception, learning, and reasoning, mainly from the perspective of vision. These interests have drawn me towards the following research directions: data augmentation, representation learning, and generative modeling. I have been working with Prof. Stefano Ermon from the onset of my master's degree at Stanford, and I have been fortunate to conduct some interesting work in these directions. These initial steps and my future plans are outlined below.

Data augmentation strategies for synthesizing new data in a way that is consistent with an underlying task are extremely effective in both supervised and unsupervised learning. They allow for the incorporation of prior knowledge (inductive bias) about the properties of typical samples from the underlying data distribution. Many recent methods for unsupervised or self-supervised learning involve training models to be invariant to different "views", or transformed versions of an input. However, these data augmentation/transformation schemes are limited by their ability to produce only additional "positive" examples of how a task should be solved. I am excited about overcoming these limitations by specifying an even wider range of inductive biases that are not limited by just how a task should be solved. My impetus for pursuing this research direction came from my experience during a project (submitted to ICLR 2021) where I worked with Prof. Ermon to leverage an alternative and complementary source of prior knowledge that specifies how a task should not be solved. To enable a wider range of augmentations, we explored negative data augmentation strategies (NDA) that intentionally create out-of-distribution samples and show that such samples provide information on the support of the data distribution, and can be leveraged for improved generative modeling using GANs and representation learning in images and videos (using contrastive learning framework for self-supervised representation learning). I plan to continue this line of research to further understand the empirical abilities and limitations of such negative data augmentation along with their application to other domains such as natural language processing.

I also enjoy eclectic applications of machine learning to areas like remote sensing, medical imaging, etc. Contrastive learning based self-supervised representation learning requires a good design of data augmentation to work well. Despite the success of data augmentation, the influence of different augmentation schemes has been less studied. My first-hand experience with the critical role of data augmentation in contrastive learning coupled with their lack of application in remote sensing motivated my subsequent research in this largely unexplored domain of satellite imagery. In a paper submitted to CVPR 2021, I co-led a project that seeks to exploit the spatio-temporal structure of remote sensing data by leveraging spatially aligned images over time to construct temporal positive pairs in contrastive learning and geo-location to design pre-text tasks. We noted that though conventional data augmentation strategies have seen great success on traditional vision datasets like Imagenet, they are sub-optimal for remote sensing owing to their different characteristics and our scheme for creating positive pairs provided more complex similarity cues to the model compared to what random transformations can offer. I am interested in extending this work by developing new ways of designing more geo-aware negative sampling strategy for contrastive learning. Through this direction, I hope to further explore my nascent interests in contrastive learning and data augmentation.

I received my undergraduate degree in computer science from the Indian Institute of Technology Kharagpur, where I worked with Prof. Pabitra Mitra on leveraging GANs for semi-supervised learning on large-scale fundus imaging modality for vessel segmentation as part of my Bachelor's thesis. We extended the concept of GANs to a multi task learning setup wherein a discriminator-classifier network differentiates between fake/real examples and also assigns correct class labels. This can be understood as a form of data-augmentation where we use fake images as an effective form of weak supervision in addition to real labeled data. We achieved comparable performance (sometimes even better) with recent CNN based techniques while using up to 9 times less labeled training data. This work ended up being accepted in the Workshop on Computer Vision for Microscopy Image Analysis at CVPR 2017. I also received the best undergraduate thesis award in my department.

Prior to starting my master's at Stanford, I worked at Adobe for two years where I undertook several projects both in research and software development. In a paper accepted at WACV 2020 (and CVPR 2020 Workshops), I co-led a project on a multi-stage generative framework for image-based virtual try-on for fashion commerce where I employed a duelling triplet loss strategy to improve the texture transfer process. We pit the output obtained from the network with the current weights (anchor) against that from the network with weights from the previous phase (negative), and push it towards the ground-truth (positive). This can be thought of as an online hard negative mining method where we generate negative samples for the triplet loss from an older version of the model itself. This scheme provided significant boost in performance and I believe that neural models can be a powerful tool for data augmentation or negative sample generation. One direction that I am excited about is investigating the use of generative models to replace heuristic data augmentation schemes by producing input-conditioned transformations to be used as data augmentation for various learning systems.

I am also interested in exploring how we can mitigate the high training cost of pre-training on large-scale datasets. This is important for transfer learning via fine-tuning from large scale datasets to small scale, domain specific datasets. In a paper submitted to CVPR 2021, we attempt to address this by proposing efficient target dataset conditioned filtering methods to remove less relevant samples from the large-scale pre-training dataset. Unlike prior work, we focus on efficiency, adaptability, and flexibility in addition to performance. Additionally, we discover that lowering image resolutions in the pre-training step offers a great trade-off between cost and performance. Examining how performance on a target dataset changes as we pre-train on more domain similar images can lead to new insights for transfer learning. Additionally, I also plan to study other important factors in transfer learning beyond domain similarity.

I wish to further extend my experience in representation learning, data augmentation, and generative modeling to various topics. Specifically, I want to conduct extensive research in self-supervised representation learning for video understanding and 3D computer vision. Self-Supervised learning and data augmentation for point clouds in 3D computer vision are less studied and open room for investigation. Due to the nature of geometric information in 3D data, transformations for point clouds typically have a large number of parameters including geometric distance, operation strength, sampling probability, etc., and certain image augmentation techniques, such as color shifting, simply wouldn't apply to monochromatic 3D data. I am also excited by the prospects of self-supervised representation learning for activity discovery which carries great promise for a variety of video-based real-world applications e.g. human activity and behavior understanding in healthcare settings.

At Stanford, I really enjoyed my time being part of Prof. Stefano Ermon's lab for the past four quarters during which I developed my nascent interests in data augmentation, generative modeling and representation learning. I am also excited by the prospect of working with Prof. Chris Re whose recent work on Model Patching piqued my interest in which they propose a framework that exploits data augmentation for improving robustness and encouraging the model to be invariant to subgroup differences. I am also interested in the kind of research that Prof. Serena Yeung is doing in the space of healthcare with a primary focus on computer vision. I thoroughly enjoyed her class, "CS271: AI for Healthcare", during which I explored self-supervised representation learning for medical imaging (Chest X-Rays and Diabetic Retinopathy) and found that conventional data augmentation schemes led to suboptimal performance. The different characteristics of medical datasets and my findings suggest promising avenues for research in this domain. I am interested in developing new ways to leverage domain knowledge inherently present in medical volumes like MRI datasets in defining the positive and negative pairs of images in a contrastive learning framework for learning good global level representations. I am also interested in collaborating with Prof. Juan Carlos Niebles on self-supervised learning for videos. I am intrigued by the prospect of learning representations for videos directly from naturally occurring pairing of visual and textual data (like Youtube titles, Instagram caption, etc.) in a contrastive learning concept. Following the work of these groups has led me to see a clear fit for my skills and interests at Stanford, and I am confident that it is a great place for me to pursue a Ph.D. I am confident I will do well in the program and that my academic and professional experiences will hold me in good stead. I hope to learn, share and grow as a member of your vibrant scholastic community.

Apart from my academic and professional experiences, I also quite enjoy sharing my ideas with people. To this effect, I have volunteered as a teacher at eVidyaloka where I taught Maths and Science to rural students in Jharkhand via Skype, I was a Student Academic Mentor where I mentored a group of five freshmen during my time at IIT Kharagpur. As a Technovation Mentor, I have mentored five high-school girls to build a business plan and mobile app addressing a community problem. As a Master's student, the diverse student body at Stanford enabled me to understand and engage in current issues and deeper societal questions. Due to these experiences, I have had the privilege of making friends from a wide variety of cultural backgrounds which has, in turn, helped open my mind to many new ideas. In fact, this is an additional motivation for me to continue my stay at Stanford as a PhD student as I know your institution as a diverse student body and I will be able to grow as a person through the multi-cultural experience I will continue to have at your campus.

Google Scholar: `https://scholar.google.com/citations?user=gIlnMF8AAAAJ`