The field of artificial intelligence for medical applications is advancing rapidly. Recent developments in machine learning and computer vision have the potential to fundamentally change fields ranging from safe and efficient acquisition of medical images to precise detection and rapid analysis of conditions in complex medical modalities. I strongly believe that a Ph.D. in Electrical Engineering from Stanford would train me to effectively contribute to signal processing, (physics-informed) machine learning, and computer vision as applied to medical data and healthcare settings.

To date, my primary research experience has been in machine perception, learning, and reasoning. It began when I was awarded the Indian Academy of Sciences fellowship to work as a student researcher at the Indian Institute of Science as a sophomore. I developed a first-of-its-kind fully convolutional neural network for saliency prediction in images and incorporated a novel Location Biased Convolutional layer to model location-dependent patterns like center-bias. Working on this project (published at IEEE Transactions on Image Processing) motivated me to further explore ML and computer vision. The following summer, I worked as a research intern at the Big Data Experience Lab, Adobe. Our team proposed and created a novel consumer-targeting system through modeling the rich data from Augmented Reality (AR) systems. I worked with multiple tools and technologies including statistical modeling, a structure-transcending method for evaluating the stylistic similarity between 3D shapes, and color compatibility. This process convinced me of the power and possible future applications of AI algorithms in creating value for the users, and resulted in three international patents as well as a publication in ISMAR'17.

After graduation, I joined a newly minted team at Adobe that was establishing itself as a research group. My projects were comprehensive in scope and excitingly open-ended, exposing me to many problems and areas within the broad umbrella of ML and computer vision. Having developed an avid interest in AR during my internship, I subsequently forayed into this area and personally undertook several initiatives which were aimed at improving product recommendations in AR-based retail apps. I also led projects in fashion commerce like (a) image-based virtual try-on for fashion where I proposed a multi-stage generative framework and employed a novel dueling triplet loss method to improve texture transfer, (b) modeling visual cues for fashion compatibility, outfit recommendation, and style extraction using graph convolutional network, and (c) visual similarity search. The works have been highlighted in several conferences (WACV'20, WACV'20, ICCV'19, CVPR'19, CVPR'19).

After Adobe, I decided to go for a Master's in order to pursue advanced research in ML and computer vision. In my literature review, I read published works by Stanford on topics such as representation learning, scene understanding, generative modeling, etc., and these drove me to apply to Stanford. In my Master's, I have been exceedingly fortunate to work with Prof. Stefano Ermon in ML and computer vision in the following directions: data augmentation, self-supervised representation learning for images and videos, generative modeling, and computational sustainability. I completed my Master's thesis under Prof. Stefano Ermon and Prof. Marshall Burke on *Combining Machine Learning and Satellite Imagery for Sustainability Challenges* which lead to publications in IJCAI'20, AAAI'20, ICLR'21, ICCV'21 (with some works still under review).

I currently work as a Research Engineer at Google Research, focusing on topics like Camera Autofocus and Neural Structured Learning. My personal professional desire is to expand my research's scope and develop machine learning algorithms with a primary focus on computer vision to enable new capabilities in biomedicine and healthcare. I developed my nascent interests in this area during my bachelor's degree where I worked with Prof. Pabitra Mitra on leveraging GANs for semi-supervised learning on large-scale fundus imaging modality for sample efficient vessel segmentation as part of my Bachelor's thesis. I 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. I also received the **Best Undergraduate Thesis** award in my department for this work.

In a paper at ICCV'21, I led a project that exploited the spatio-temporal structure of remote sensing data by leveraging spatially aligned images over time to construct temporal positive pairs in contrastive learning based self-supervised learning (SSL) and geo-location to design pre-text tasks. I observed that though conventional data augmentation strategies for SSL have seen great success on traditional vision datasets like ImageNet, they are sub-optimal for remote sensing owing to their different characteristics. Our scheme for creating positive pairs provided more complex similarity cues to the model compared to what random transformations can offer. This insight motivated me to explore SSL for medical imaging (Chest X-Rays and Diabetic Retinopathy). I found that conventional data augmentation schemes led to a suboptimal performance. The different characteristics of medical image data thus suggest promising avenues for research in this domain. I hope to develop new ways to leverage domain knowledge inherently present in medical volumes like MRI 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 devising localized objectives to learn distinctive local-level representations within an image useful for segmentation tasks.

In another work (ICLR'21), to enable a wider range of augmentations, I explored negative data augmentation (NDA) for natural images and videos that intentionally create out-of-distribution samples. We theoretically show that such negative out-of-distribution samples provide information on the support of the data distribution, and can be leveraged for generative

modeling and representation learning in images and videos. I plan to continue this line of research to investigate NDA in medical imaging modalities by leveraging their important characteristics along with rigorous analysis of its theory and limitations.

There are multiple faculty at Stanford whose research inspired me and many opportunities that excite me in future work with them. Prof. Serena Yeung showed that models that capture implicit hierarchical relationships between subvolumes in 3D biomedical images with self-supervised hyperbolic representations are better suited for unsupervised segmentation of such modalities. I am interested in studying the robustness of this method (to learn domain-invariant features for segmentation) to unforeseen data distribution shifts during deployment, e.g. change of image appearances or contrasts caused by different scanners, unexpected imaging artifacts, etc. Can we also apply ideas from latent space data augmentation in this method for generating hard examples for SSL and studying its effect on model generalization and robustness under limited data settings?

In addition to interpretation tasks such as classification, segmentation, or detection applied to widely used volumetric medical imaging modalities like MRI, I am inspired to work on developing ML algorithms for computational imaging inverse problems in MRI or CT reconstruction. Recent works at Prof. Jonn Pauly's, Prof. Lei Xing's, and Prof. Akshay Chaudhari's groups in solving inverse problems in medical imaging are a perfect fit for my research interests. I have identified several projects that match my interests with these groups, such as applying physics-driven data augmentations for consistency training for accelerated MRI reconstruction in which they leverage domain knowledge of the forward MRI data acquisition process and MRI physics for improved data efficiency and robustness to clinically-relevant distribution drifts. This work served as a deciding factor for me to apply to the EE program where I can work through my understanding of the physical principles involved in biomedical computational imaging via courses (EE169, EE369C, EE469B) and hands-on research with the aforementioned groups. Another project that piqued my interest was based on an implicit neural representation learning methodology with prior embedding (NeRP) to reconstruct a computational image from sparsely sampled measurements. The use of implicit functions for capturing the information in a scene is an exciting new breakthrough that has been producing very impressive results and is ripe for a lot more progress. Another question worth investigating is can we develop data-efficient, robust methods for solving semantic tasks (on natural or bio-medical images) using implicit representations; tasks to which image-centric methods are traditionally applied: detection, semantic segmentation, classification, etc. Can supplanting image-centric methods with neural scene representations as the representation for vision potentially increase accuracy and consistency with fewer samples?

The prospect of working on interpretation of high-dimensional medical videos, like assessing cardiac motion in cardiac ultrasound videos and human activity and behavior understanding from videos in healthcare settings, also drives me. Prof. Serena Yeung and Prof. James Zou have led interesting works in this direction. I am interested in investigating the use of recently popular Video Vision Transformers for the analysis of medical videos along with ways to effectively regularize the model to be able to train on comparatively small datasets owing to the general paucity of data in healthcare settings. In a potential project with Prof. Chaudhari (& Prof. Hargreaves), we discussed combining imaging methods (providing structural information) with functional information that can be extracted from videos of individuals walking to study knee injury recovery and osteoarthritis.

Despite major advances brought by deep learning, computer vision is far from being solved. There is a lack of strong generalization across tasks in computer vision. At this point, the field is fractured into a collection of separate tasks (classification, detection, segmentation, tracking, captioning, etc.), each of which requires a substantial amount of specialized labeled data and model engineering, even when using pre-trained convolutional backbones. This is in stark contrast with humans who can perform a wide variety of vision tasks and generalize to new tasks given very limited task-specific labeled data. Given a few glimpses of an unfamiliar object, humans can recognize it under changing conditions, detect or segment it, track it over time, describe it in text, estimate its attributes, etc., -- all this with just a few coarse "labels" and no explicit supervision for the other tasks. Can we build artificial systems with similar capabilities? One encouraging example comes from natural language processing (NLP), which ∼5 years ago was in a state similar to contemporary computer vision. NLP systems were fragmented into many low-level tasks (morphology, parsing, etc.) and were trained using custom annotations. Recently NLP has moved on to more universal end-to-end solutions, driven by (i) successful scaling of pre-training algorithms, and (ii) the use of a unifying sequence-to-sequence API and, more recently, a unifying text-to-text API. As a result, a single NLP model can now generalize to an endless variety of text-defined tasks. Can we achieve a similar transition in computer vision? Achieving a grand universal model can lead to SOTA performance on both "natural" and "non-natural" tasks along with improvements in model uncertainty, calibration, and out-of-distribution (OOD) detection on images from the medical domain.

Inspired by the research problems I encountered, I am thus interested in (a) medical computational imaging by incorporating the physics and geometry priors of imaging model to build a unified machine learning framework with an aim to improve robustness and reduce data requirements (b) interpretation of medical image data leveraging their important characteristics in self-supervised/unsupervised learning frameworks, (c) video-based analysis of human motion and other 3+ dimensional imaging datasets, (d) devising new clinically driven and data-driven metrics that can mitigate discordance between existing quantitative and qualitative metrics (thus having true downstream utility). The EE program will equip me with the skills allowing me to develop algorithms with physical conservation laws and mathematical symmetries built into the networks themselves. So far, I have been extremely fortunate to have worked with inspiring collaborators on rewarding projects, and I wish to advance further to the unknown ground. A Ph.D. in Electrical Engineering at Stanford would allow me to contribute to the challenges I am passionate about with unmatched support and prepare me for a career as a researcher in academia or industry.