Research Statement

Nikhil Pandey

Computer Vision, Machine Learning, and Robotics.

The first instance when I seriously considered a career in computer vision was while I was performing image processing research and development in the third year of undergrad college. For the last 4 years, I have been implementing and modifying research coming out of conferences like CVPR and NIPS to fit the industrial and infrastructural limitations of this part of the world. Post my undergrad, I joined Occipital Tech. While, on one hand, the chaotic environment of an early-stage startup helped me to develop rapid prototyping skills, on the other hand, it constantly challenged my mathematical understanding. Tasks like object localization, segmentation, and motion assessment are a common feature of my daily work. Over the next couple of years, I made extensive use of transfer learning and model compression techniques to train and fit visual perception models on SBCs like Jetson TX2. Due to the deep integration of visual perception modules with the hardware of robots, we were able to produce state-of-the-art optical sorting machines that cost a fraction of the preexisting machines while outperforming them. My designs for robotic perception helped in the generation of two patents last year.

My pursuit of curiosity has shaped my research in the field of computer vision. Over the last few years, as I have started working in more complex fields, I am facing questions that are still unanswered. Developing efficient and reliable perception pipelines has been my primary focus. Unlike the cloud where I am not bound by computing power, achieving perfection is not so straightforward in robotic perception. By robot perception, I am referring to the embodiment of human-like intelligence in robots or agents for exploration and decision-making. Robots have not been able to meet anything comparable to a human-level perception even with multimodal sensors and better computers. In my research experience, I have often encountered a state of the art model tuned and tweaked to near-perfect performance on a standard dataset, only to fall flat as soon as it is introduced to a real-world environment. I believe this is due to underspecification and data shift. As per my latest experiment, I trained 9 ResNet object detection models with different initial weights. All of them reached a similar metric of performance while training and testing on a standard dataset. But during testing on the augmented validation set, their performances varied widely. I believe there is something wrong with the way we evaluate the neural network models and the assumptions we make. The perception system's ability to adapt becomes an imperative problem to solve because, in regards to a robot, the only factor that varies in time and space is its environment. I am deeply interested in expanding my knowledge and work on this particular topic and other topics with a similar nature.

Furthermore, Robots usually have limited availability of onboard computing and energy. For a person who learned coding on a $35 raspberry pi, the value of efficient programming has been deeply imbibed in my mind. Perhaps this is why it makes me uncomfortable to see that we have taken the computing resources for granted. I do not mean to infer that the resources themselves are limited, rather it's the access to those resources that go overlooked. It is natural to celebrate when a model improves its mean-average-precision by a couple of percentages after training on thousands of GPUs for weeks because in any case, the generalization of machine learning models is a big deal. But what we don't seem to take into account is that such models also create an entry barrier for individuals and organizations that don't have that many resources. From my experience in the startup ecosystem, I can say that the cost of training a language model like GPT3 is often unaffordable for many aspiring entrepreneurs. Besides, we seem to be completely oblivious to the carbon emissions that have been incurred to gain those improvements. I cannot consciously consider such developments as steps in the right direction if even inadvertently they promote economic disparity and environmental destruction. Hence, my ultimate aim is to design and create energy-efficient paradigms of learning that are accurate enough for real-world use, fast enough for deployment on mobile processors, and accessible enough for a student in a third-world country to tinker.