Yash Pansari
Statement of Purpose
MS in Computer Science

I had a mortal enemy in middle school — Blackjack 247. I loved math, and I eventually realized that I was probabilistically doomed to lose long-term unless I could count cards. However, the game was lightning-fast, so I learned Python, wrote a bot to help, and won. Amused by the wonders of programming, I scoured the internet for my next mission, until I encountered Pranav Mistry's Ted Talk on *SixthSense*, a "truly seamless" computer interaction suite using only a webcam and projector. Inspired, I dreamed of building a similar interface, but effortless to use regardless of user or circumstance. Today, with Humane's AI pin, Apple's Vision Pro, and Meta's AI-infused Ray Bans, I see the world quickly advancing toward Mistry's ideas of augmented reality-based computer interaction, but now with massive scope for assistive computer vision (CV). From semantic segmentation and visual annotation for disabled persons to vision-correcting displays for higher-order aberrations unfixable via conventional lenses, the possibilities are truly endless. Hence, as a CV researcher/engineer for one of these products, I aspire to make this technology not just usable, but truly effortless for everyone. To this end, I hope to use Stanford's MS program to expand my knowledge of advanced CV and further explore these ideas under professors such as Serena Young or Jiajun Wu, whose work has continued to inspire me throughout my journey.

When I first entered Berkeley, I had a growing interest in applied CV but no technical exposure to machine learning besides Andrew Ng's *Deep Learning* (Coursera). Thus, I built my foundations with Machine Learning @ Berkeley's NMEP — *accelerated ML classics for exceptional undergraduates*. Quickly absorbing the content, I was selected to lecture in NMEP the following semester, but I also realized that deeper ML theory needed mathematical maturity. So I took *Real Analysis* to rederive everything I knew from the axioms of mathematics, and *Abstract Algebra* to maintain this rigor in unfamiliar universes. As a mathematician, I now better understood all models I encountered — be it the Hilbert space of random variables in *Random Processes*, higher dimensional constraint surfaces in *Optimization Models*, or hidden feature embeddings in *Modern Computer Vision* (audited class).

Eager to use this knowledge on real-world problems, I joined Lawrence Berkeley National Laboratory, working on crack detection for the Large Hadron Collider. I began by researching UNETs for microscope image segmentation and Vision Transformers for 3D damage tracing. However, these architectures required vast datasets and significant inference time, while even milliseconds of lag here could leave millions in damage undetected. Therefore, I incorporated unsupervised techniques like contouring, edge detection, and flood-fill, gaining an appreciation for image processing and the foundations of CV — the backbone of Pranav Mistry's original pre-ML prototypes, and a field I hope to further explore via *Computational Imaging* (CS488I) and *Computer Graphics* (CS248A) at Stanford. Moving forward, I was now curious to work with cutting-edge deep learning-based CV.

Hence, I accepted a contract project for Vyrill (AI-based consultancy) involving sentiment and demographic extraction from videos. I reviewed pre-training, sparsification, and compute optimization techniques for facial recognition tasks, building on works including "Heavy-Lifting Pre-training for Multi-facial Recognition" (a 2018 project by former Stanford MS students). Leveraging these, we fit an ensemble of state-of-the-art facial recognition models within Vyrill's business constraints. Leading a team of five, I also learned to effectively modularize large codebases and gathered industry-level ML Systems knowledge, managing to successfully reimplement the company's flagship product from scratch within

one semester. Working in industry gave me fresh perspectives on management and entrepreneurship, solidifying my aspirations of eventually starting my own company to bring my contributions in equitable technology to remote areas at an affordable cost in developing nations like my own.

Looking to further expand my horizons by exploring ML in other industries over the summer, I interned at Goldman Sachs as a Quantitative Researcher. Similar to computer vision, financial models encode real relationships within their embeddings and build inferences from these associations. However, finance has little open-source research and extreme competition, so I had to constantly innovate new architectures and rigorously understand their real-world implications. Exploring ML behaviors such as adversarial policies and embedding spaces in finance allowed me to understand them more fundamentally, and build more explainable models overall. Thus, quant finance expanded my intuition in ways orthogonal to a traditional CS education, making me a better researcher and computer scientist in general.

Having built my skills in theory, research, and industry, I revisited my goals via CV research for gesture-based computer input at Professor Brian Barsky's Assistive Technology Lab. Here we worked to create more equitable technology by eliminating physical contact and allowing disabled persons to use computers just as effortlessly without touch, aim, or hand stability. While researching assistive CV, I also explored intersections with medical technology in papers by Serena Young such as "Deep learning-enabled Medical Computer Vision" (2021), and shared these interests as a CV lecturer in Medical Technology @ Berkeley's education program. Following this, I read "A Computer Vision System for… Patient Mobilization Activities in the ICU″ (2019) and dreamed of using 3D vision for gait analysis, fall detection, and emergency prediction/response ubiquitously using mobile technology. Also inspired by more general works such as "Learning to See Physics via Visual De-animation" (2017), I hope to research these technologies under Serena Young or Jiajun Wu at Stanford, leveraging their world-class expertise in scene understanding to explore various assistive technology applications.

Overall, an MS from Stanford with a specialization in Visual Computing would expand my basis within CV techniques and ML research while building a robust breadth in Modelling and Systems. Specifically, I look forward to taking *Representation Learning in CV* (CS331B) and *Deep Learning for CV* (CS231n) to deepen my knowledge of 3D reconstruction, scene understanding, and natural language representation for multimodal applications, with *Visual Computing Systems* (CS348K) to better optimize practical implementations of these technologies.

At Stanford, I'm ready to approach advanced computer vision with a quickly growing passion for assistive technology, and the same curiosity that pushed me to beat Blackjack 247.