Yash Pansari
MEng. in EECS
Statement of Purpose

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 the disabled to vision-correcting displays for higher-order aberrations unfixable via conventional lenses, the possibilities are truly endless. Hence, as a CV 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 leverage Berkely's MEng program to expand my knowledge of advanced CV and further explore these ideas under faculty — such as Allen Yang, Jitendra Malik, or Angjoo Kanazawa — that have inspired me for years.

When I first entered Berkeley, I had a math background and a growing interest in applied CV but no technical exposure to machine learning, a field I knew would be central to my goals. 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 *Machine Learning* (taken with Professor Malik), or hidden feature embeddings in *Modern Computer Vision* (audited DeCal).

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 learned about optimized unsupervised algorithms for image processing tasks such as contouring, edge detection, flood-fill, etc., finally building a pipeline fundamentally inspired by foundational insights presented in John Canny's *A Computational Approach to Edge Detection* (IEEE'86) — the backbone of Pranav Mistry's original pre-ML prototypes. Next, I was curious to work with more advanced deep learning-based CV.

Thus, I accepted a contract project for Vyrill (AI-based consultancy) involving sentiment and demographic extraction from videos. I reviewed literature on pre-training, sparsification, and compute optimization techniques for facial recognition tasks, from classics (including *Robust Face Recognition via Sparse Representation*, Allen Yang, IEEE'08) to the cutting-edge (including *Caffe,* Trevor Darrell, 2014). 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 learned to modularize large codebases effectively and gained industry-level ML Systems knowledge, successfully reimplementing the company's flagship product from

scratch within one semester. Working in industry also gave me fresh perspectives on management and entrepreneurship, solidifying 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. To further hone the skills required for these endeavors, I look forward to exploring the Fung Institute's engineering leadership curriculum.

Looking to further expand my horizons by exploring ML in other industries over the summer, I interned at Goldman Sachs as a Quantitative Researcher. Like computer vision, financial models encode real relationships within embeddings and build their 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.

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. In doing so, I constantly read literature about CV applied to assistive and medical technology, and discussed these interests as a CV lecturer in Medical Technology @ Berkeley's education program. Learning about modern 3d vision for my research, I also read 3D pose tracking research including *Humans in 4D* (ICCV'23) by Professors Malik and Kanazawa, which inspired me to envision functionalities such as fall detection and emergency prediction/response, possibly performed on any device with a webcam. Additionally, works like Professor Malik's *Rich feature hierarchies for accurate object detection and semantic segmentation* (IEEE'14) could lead into visual annotation in AR, a field with countless assistive applications. As a Masters' student, I hope to explore these ideas and build upon this research under these professors, leveraging their expertise in 3D scene/pose analysis and semantic segmentation to explore assistive technology applications.

Overall, an MEng from Berkeley would expand my basis within CV techniques and ML research while building a robust breadth with classes like *Applications of Parallel Computers* (CS C267) focused on optimizing practical implementations of this technology. I also look forward to taking *Computer Vision* (CS280) to better understand motion tracking, 3D reconstruction, scene understanding, and segmentation. Coupled with 280, I'd take *Natural Language Processing* (CS288) and *Reinforcement Learning* (CS285) to explore multimodal models, wearable robotics, and the intuition behind hybrid architectures such as Vision Transformers. To teach this knowledge forward, I would lead NMEP for ML@Berkeley, and possibly the Poker @ Berkeley decal.

At Berkeley, I've created a space for myself at every lecture hall, every hackathon, and every research fair. Now, I'm ready to approach advanced computer vision with these resources behind me, a quickly growing passion for assistive technology, and the same curiosity that pushed me to beat Blackjack 247.