Name: BlackBox Robotics

Tag: Creating the Scale AI of humanoid robotic tasks

Thesis:

I am creating BlackBox Robotics, the Scale AI for humanoid robotic tasks. We are helping humanoids learn to be human by creating structured learning problems and providing the associated sensory data for these tasks.

A few months back I had the chance to read Fei-Fei Li's autobiography, The Worlds I See. Her dreams and visions have imparted themself in my own, and they largely serve as inspiration for BlackBox Robotics. I will first take some time to outline this vision.

Fei-Fei draws the curious point that in the process of human evolution, the eyes developed before the brain. In a more general pattern of nature, cognition is a product of sensory data - not the other way around.

Before AlexNet popularized the CNN in 2012, we first saw ImageNet from Fei-Fei Li and her lab. Before AlphaFold launched in 2018, protein structural data was curated en masse for projects like the Protein Data Bank. And before Waymo first began operating in San Francisco in 2022, we first saw map reconstruction of the city en masse.

Data precedes the intelligence that processes it because the development of reasoning capabilities is a natural adaptation to the mere existence of organized input data. We must see before we know what we see. If this is not the case, is there any reason the human brain does not have an atomic or at least molecular level of understanding of our surroundings? Our reasoning is confined by the sensory data it has evolved to process.

In this sense, we will need data curation en masse for a variety of tasks that define what it means to be human before we see the development of true humanoid reasoning capabilities. I have a strong conviction that the coming of this robotic intelligence will bring about the greatest age of prosperity ever known to humankind. I am willing to dedicate my early career toward the fruition of this horizon. All of this is the North Star both for myself and for BlackBox Robotics.

—

We will be starting with a subset of human tasks: working as an employee in a restaurant. The protocol for designing the learning task is as follows. First, find restaurant partners (or any retail vendors with onsite size < 3000 sqft). Then, install the "BlackBox" device onsite (camera +

LiDAR + ultrasonic multi-sensor). Onsite partners will be paid monthly in exchange for BlackBox data captured. From here, we will curate, label, and anonymize the input data and output a timeline of agents performing labelled jobs, delivered through a 4D map of the site that can be used by foundation model companies to design structured learning tasks.

The technical innovation needed will start with these items: designing of the blackbox multi-sensor device, performing 4D reconstruction from sensor fusion of blackbox sensor data, segmentation of agents within the 4D reconstruction, and task labelling of the jobs performed by the agents within the reconstruction.

From this list, I have decided to start with the third task as I believe it will be the most crucial step before onboarding onsite partners with BlackBox installation. I have attached a GitHub repository below with my current progress on this problem.