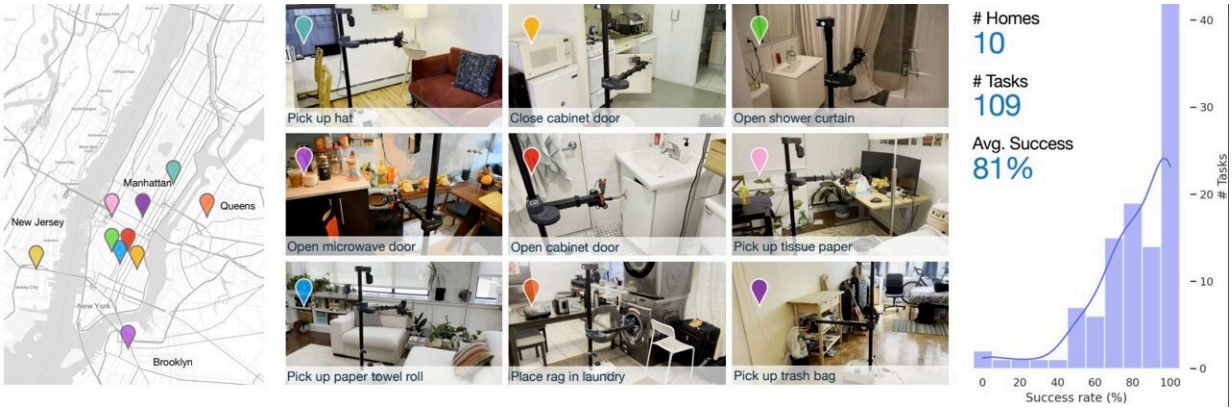
Dobb-E: A framework to train multi-skilled robots for domestic use

December 18 2023, by Ingrid Fadelli



Credit: Shafiullah et al.

Roboticists have been trying to develop robots that can tackle various everyday house chores, such as washing dishes or tidying up, for several years. However, so far none of the robots created has been commercialized adopted on a large scale.

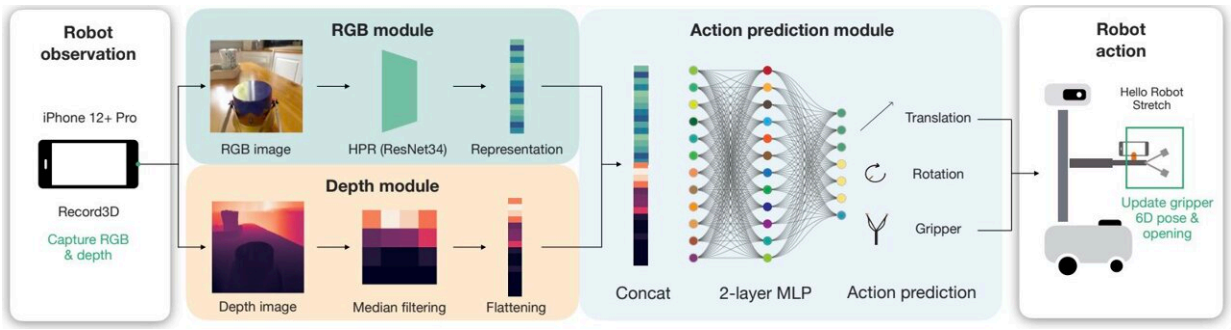
Researchers at New York University recently introduced Dobb-E, a new framework specifically designed to effectively train mobile robots on domestic tasks, ultimately contributing to their future widespread use. This framework, outlined in a paper pre-published on the server *arXiv*, could be applied to various robots designed to assist humans in their homes.

"This paper was born from our vision of introducing robots into the average American household in the near future," Lerrel Pinto, co-author of the paper, told Tech Xplore. "We already have specialist 'robots' in our homes, such as a dishwasher or a laundry machines, but a generalist robot that can learn how to complete each home-work and how it can help best in that situation has been a distant goal for all too long now."

The recent work by Pinto and his colleagues had a series of primary objectives, all of which would need to be met for robots to be successfully integrated into household environments. Firstly, the team wished to devise an efficient approach that would allow users to rapidly teach robots new skills.

This approach should also ensure the robots' safety, ensuring that they do not cause any damage while they are learning to complete new tasks. Finally, it should not be particularly demanding for end-users, thus ensuring it simplifies rather than complicates their life.

"To achieve efficiency, we relied on data-driven learning systems, the success of which are readily visible in the largest machine learning models in deployment today," Pinto said. "We achieve safety by formulating our system around learning with supervision from the user rather than learning from trial-and-error. Finally, we developed an ergonomic demonstration collection tool, enabling us to gather task-specific demonstrations in unfamiliar homes without direct robot operation."



Credit: Shafiullah et al.

The Dobb-E framework has four key components, namely a data collection tool, a pre-trained model, a diverse dataset and a deployment scheme. The first of these components, called the "Stick," is designed to simplify the collection of data, leveraging a user's smartphone.

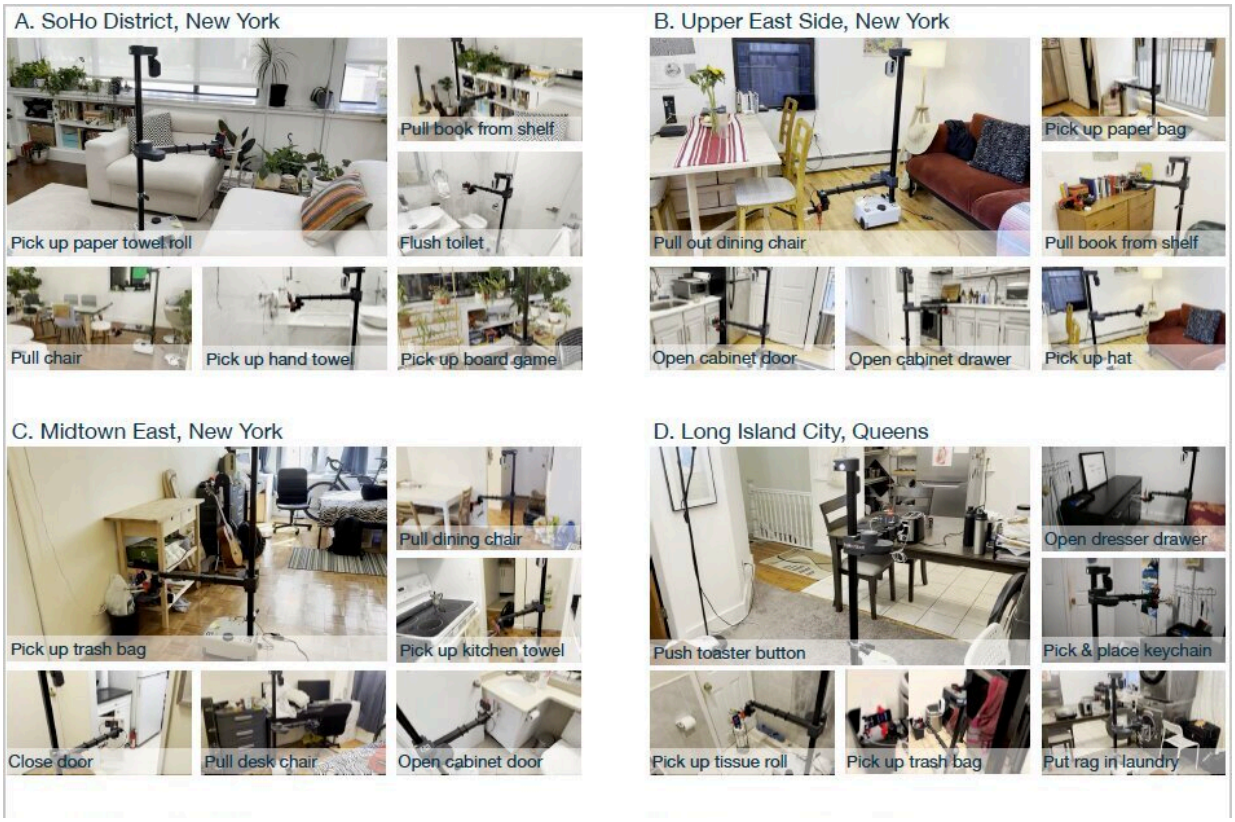
"The Stick is our cheap but ergonomic data collection tool that we build out of a reacher-grabber tool, 3D printed mounts and an iPhone pro," Pinto said. "A simple list of ingredients makes this tool accessible and cheap, while the mounted iPhone lets us record high-resolution video, depth, and movement information from the demonstrations."

Using the Stick data collection tool, Pinto and his colleagues compiled a new dataset for training domestic robots, which they called the Homes of New York (HoNY) dataset. This dataset contains footage collected using their smartphone-based setup in 216 home environments in New York.

Unlike other datasets for robot training developed in the past, the HoNY dataset focuses on a more diverse set of scenes and robot behaviors. In addition, the Stick collection tool allowed them to compile an order of magnitude more scenes than those featured in previous datasets.

The third component of the Dobb-E system is a pre-trained perception model. This model was trained on the HoNY dataset, employing a self-supervised learning approach.

"We train Home Pretrained Representations (HPR) as our visual recognition model using the HoNY dataset and a state-of-the-art self-supervised learning algorithm, MoCo-v3," Pinto explained. "HPR lets our method scale across the vastly different scenes in different homes."



Credit: Shafiullah et al.

Pinto and his colleagues assessed the potential of their data collection tool, HoNY dataset and pre-trained visual recognition model in a series

of experiments in real home environments. In these experiments, they deployed their trained algorithm on the Hello Robot Stretch, a promising multi-function mobile home robot.

Remarkably, the robot was taught to complete 109 different household tasks. For each of these tasks, the researchers fine-tuned their model with five minutes of new video data on average.

"The most exciting result of this paper is the confirmation that with our current level of technology we can build learned robotic agents that can address a wide range of tasks in a similarly large range of homes," Pinto said.

"Dobb-E is a cutting-edge research project, but by scaling this project up and building proper scaffolding around it, we can expect it to grow into the first steps towards a general home assistant that can assist senior citizens, people affected by disabilities, or just busy parents. However, getting there from where we are needs a lot more work both in improving the capabilities, and [polishing] to make it more user friendly."

The Dobb-E framework is a great contribution to ongoing efforts aimed at enabling the widespread deployment of multi-function household robots. The team's initial experiments yielded very promising results, while also highlighting some of the key factors affecting the performance of home robots.

In the future, this recent work could inform the development of increasingly advanced domestic robot systems. Pinto and his colleagues have publicly released their data collection tool, dataset and pre-trained model, thus other research teams could soon utilize them or adapt them as part of their own studies.

"While Dobb-E addresses learning low-level skills in a variety of different scenarios, it does not address the need for a higher-level planner or policy that ties these skills together to accomplish a variety of tasks in homes," Pinto added.

"One of our future directions should be looking into chaining skills to complete meaningful, long-horizon tasks in homes. Another direction that we want to investigate is to improve upon the sensors on the Stick and the robot, and potentially iterating on our depth sensing, adding more camera views, and adding more sensor modalities such as touch and sounds."

More information: Nur Muhammad Mahi Shafiullah et al, On Bringing Robots Home, *arXiv* (2023). DOI: 10.48550/arxiv.2311.16098