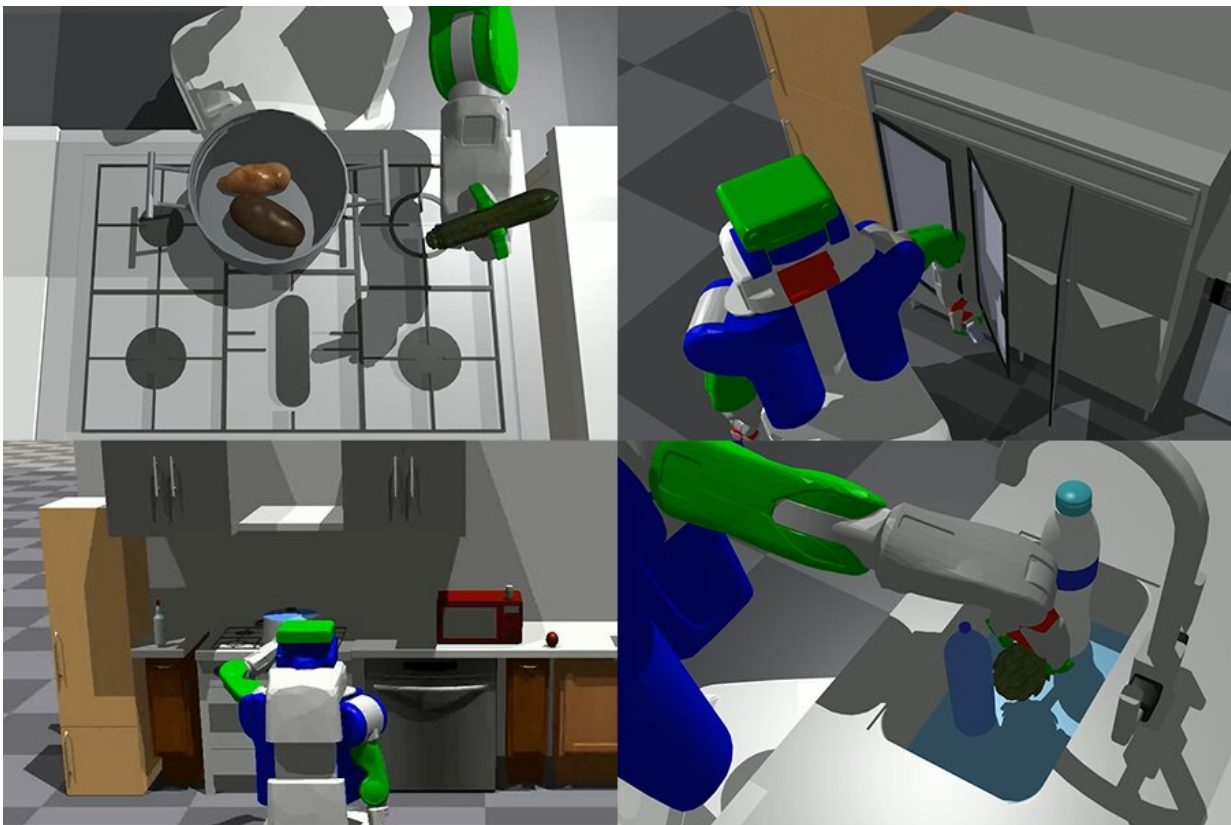


AI helps household robots cut planning time in half

July 17 2023, by Rachel Gordon



Credit: MIT Computer Science & Artificial Intelligence Lab

Your brand new household robot is delivered to your house, and you ask it to make you a cup of coffee. Although it knows some basic skills from previous practice in simulated kitchens, there are way too many actions

it could possibly take—turning on the faucet, flushing the toilet, emptying out the flour container, and so on. But there's a tiny number of actions that could possibly be useful. How is the robot to figure out what steps are sensible in a new situation?

It could use [PIGINet](#), a new system that aims to efficiently enhance the problem-solving capabilities of household robots. Researchers from MIT's Computer Science and Artificial Intelligence Laboratory (CSAIL) are using machine learning to cut down on the typical iterative process of task planning that considers all possible actions. PIGINet eliminates task plans that can't satisfy collision-free requirements, and reduces planning time by 50%–80% when trained on only 300–500 problems.

Typically, robots attempt various task plans and iteratively refine their moves until they find a feasible solution, which can be inefficient and time-consuming, especially when there are movable and articulated obstacles. Maybe after cooking, for example, you want to put all the sauces in the cabinet. That problem might take two to eight steps depending on what the world looks like at that moment. Does the robot need to open multiple cabinet doors, or are there any obstacles inside the cabinet that need to be relocated in order to make space? You don't want your robot to be annoyingly slow—and it will be worse if it burns dinner while it's thinking.

Household robots are usually thought of as following predefined recipes for performing tasks, which isn't always suitable for diverse or changing environments. So, how does PIGINet avoid those predefined rules? PIGINet is a [neural network](#) that takes in "Plans, Images, Goal, and Initial facts," then predicts the probability that a task plan can be refined to find feasible motion plans.

In simple terms, it employs a transformer encoder, a versatile and state-of-the-art model designed to operate on data sequences. The input

sequence, in this case, is information about which task plan it is considering, images of the environment, and symbolic encodings of the initial state and the desired goal. The encoder combines the task plans, image, and text to generate a prediction regarding the feasibility of the selected task plan.

Keeping things in the kitchen, the team created hundreds of simulated environments, each with different layouts and [specific tasks](#) that require objects to be rearranged among counters, fridges, cabinets, sinks, and cooking pots. By measuring the time taken to solve problems, they compared PIGINet against prior approaches. One correct task plan may include opening the left fridge door, removing a pot lid, moving the cabbage from pot to fridge, moving a potato to the fridge, picking up the bottle from the sink, placing the bottle in the sink, picking up the tomato, or placing the tomato. PIGINet significantly reduced planning time by 80% in simpler scenarios and 20%–50% in more complex scenarios that have longer plan sequences and less training data.

"Systems such as PIGINet, which use the power of data-driven methods to handle familiar cases efficiently, but can still fall back on 'first-principles' planning methods to verify learning-based suggestions and solve novel problems, offer the best of both worlds, providing reliable and efficient general-purpose solutions to a wide variety of problems," says MIT Professor and CSAIL Principal Investigator Leslie Pack Kaelbling.

PIGINet's use of multimodal embeddings in the input sequence allowed for better representation and understanding of complex geometric relationships. Using image data helped the model to grasp spatial arrangements and object configurations without knowing the object 3D meshes for precise collision checking, enabling fast decision-making in different environments.

One of the major challenges faced during the development of PIGINet was the scarcity of good training data, as all feasible and infeasible plans need to be generated by traditional planners, which is slow in the first place. However, by using pretrained vision language models and data augmentation tricks, the team was able to address this challenge, showing impressive plan time reduction not only on problems with seen objects, but also zero-shot generalization to previously unseen objects.

"Because everyone's home is different, robots should be adaptable problem-solvers instead of just recipe followers. Our key idea is to let a general-purpose task planner generate candidate task plans and use a deep learning model to select the promising ones. The result is a more efficient, adaptable, and practical household robot, one that can nimbly navigate even complex and dynamic environments. Moreover, the practical applications of PIGINet are not confined to households," says Zhutian Yang, MIT CSAIL Ph.D. student and lead author on the work.

"Our future aim is to further refine PIGINet to suggest alternate task plans after identifying infeasible actions, which will further speed up the generation of feasible task plans without the need of big datasets for training a general-purpose planner from scratch. We believe that this could revolutionize the way robots are trained during development and then applied to everyone's homes."

"This paper addresses the fundamental challenge in implementing a general-purpose [robot](#): how to learn from past experience to speed up the decision-making process in unstructured environments filled with a large number of articulated and movable obstacles," says Beomjoon Kim Ph.D. '20, assistant professor in the Graduate School of AI at Korea Advanced Institute of Science and Technology (KAIST).

"The core bottleneck in such problems is how to determine a high-level task plan such that there exists a low-level motion plan that realizes the

high-level plan. Typically, you have to oscillate between motion and task planning, which causes significant computational inefficiency. Zhutian's work tackles this by using learning to eliminate infeasible task plans, and is a step in a promising direction."

Their research was presented at the conference [Robotics: Science and Systems](#), held July 10–14 in Korea.

Provided by MIT Computer Science & Artificial Intelligence Lab

Citation: AI helps household robots cut planning time in half (2023, July 17) retrieved 6 May 2024 from <https://techxplore.com/news/2023-07-ai-household-robots.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.