A visual-linguistic framework that enables open-vocabulary object grasping in robots

August 1 2024, by Ingrid Fadelli

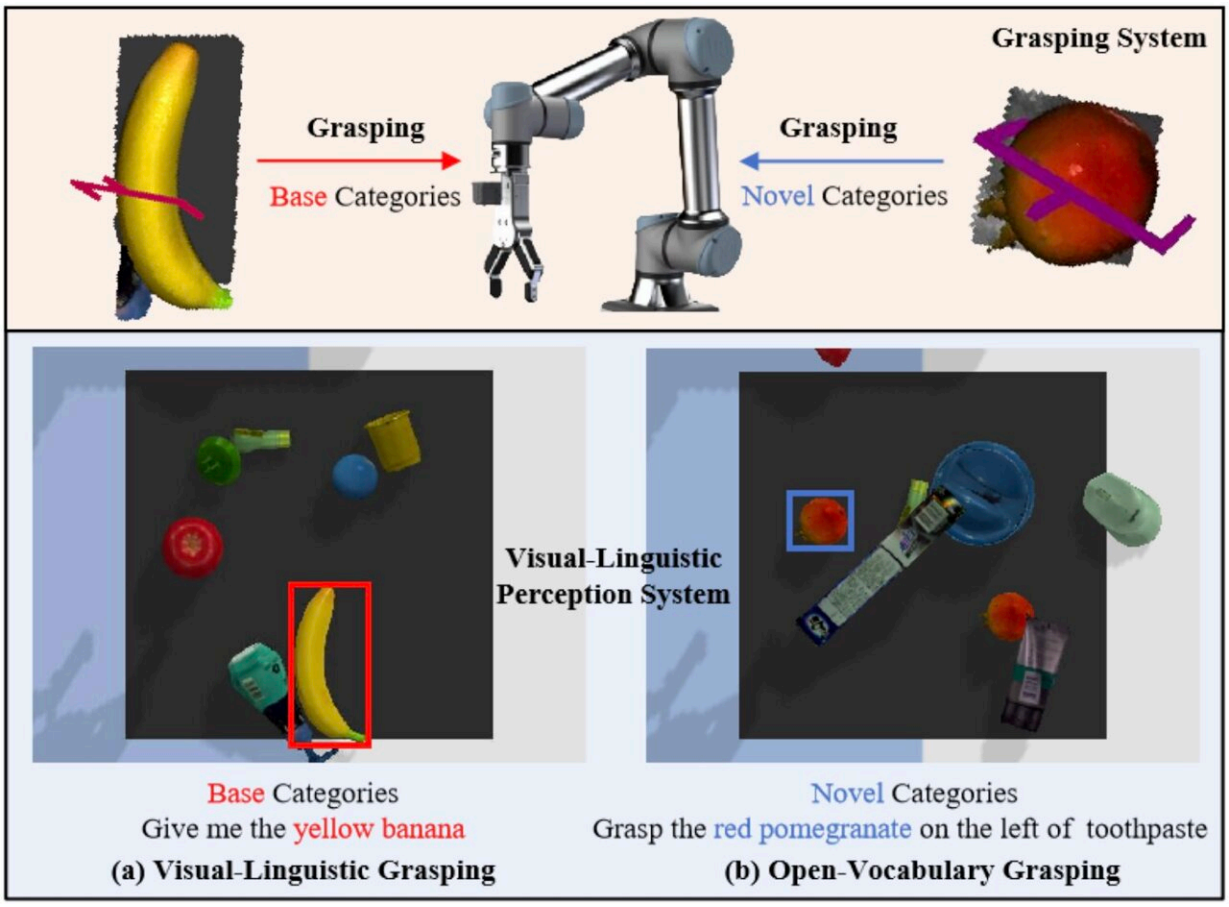

Diagram explaining what open-vocabulary grasping entails. Credit: Meng et al, *arXiv* (2024). DOI: 10.48550/arxiv.2407.13175

To be deployed in a broad range of real-world dynamic settings, robots

should be able to successfully complete various manual tasks, ranging from household chores to complex manufacturing or agricultural processes. These manual tasks entail grasping, manipulating and placing objects of different types, which can vary in shape, weight, properties and textures.

Most existing approaches to enable robotic object grasping and manipulation, however, only allow robots to successfully interact with objects that match or are very similar to those they encountered during training. This means that when they encounter a new (i.e., unseen before) type of object, many robots are unable to grasp it.

A team of researchers at Beihang University and University of Liverpool recently set out to develop a new approach that would overcome this key limitation of systems for robotic grasping. Their paper, posted to the *arXiv* preprint server, introduces OVGNet, a unified visual-linguistic framework that could enable open-vocabulary learning, which could in turn allow robots to grasp objects in both known and novel categories.

"Recognizing and grasping novel-category objects remains a crucial yet challenging problem in real-world robotic applications," Meng Li, Qi Zhao and their colleagues wrote in their paper. "Despite its significance, limited research has been conducted in this specific domain.

"To address this, we seamlessly propose a novel framework that integrates open-vocabulary learning into the domain of robotic grasping, empowering robots with the capability to adeptly handle novel objects."

The researchers' framework relies on a new benchmark dataset they compiled, called OVGrasping. This dataset contains 63,385 examples of grasping scenarios with objects belonging to 117 different categories, which are divided into base (i.e., known) and novel (i.e., unseen) categories.

"First, we present a largescale benchmark dataset specifically tailored for evaluating the performance of open-vocabulary grasping tasks," Li, Zhao and their colleagues wrote. "Second, we propose a unified visual-linguistic framework that serves as a guide for robots in successfully grasping both base and novel objects. Third, we introduce two alignment modules designed to enhance visual-linguistic perception in the robotic grasping process."

OVGNet, the new framework introduced by this team of researchers, is based on a visual-linguistic perception system trained to recognize objects and devise effective strategies to grasp them using both visual and linguistic elements. The framework includes both an image guided language attention module (IGLA) and a language guided attention module (LGIA).

These two modules collectively analyze the overall features of detected objects, enhancing a robot's ability to generalize its grasping strategies across both known and novel object categories.

The researchers evaluated their proposed framework in a series of tests run in a grasping simulation environment based on pybullet, using a simulated ROBOTIQ-85 robot and UR5 robotic arm. Their framework achieved promising results, outperforming other baseline approaches for robotic grasping in tasks that involved novel object categories.

"Notably, our framework achieves an average accuracy of 71.2% and 64.4% on base and novel categories in our new dataset, respectively," Li, Zhao and their colleagues wrote.

The OVGrasping dataset compiled by the researchers and the code for their OVGNet framework are open-source and can be accessed by other developers on GitHub. In the future, their dataset could be used to train other algorithms, while their framework could be tested in additional

experiments and deployed on other robotic systems.

More information: Li Meng et al, OVGNet: A Unified Visual-Linguistic Framework for Open-Vocabulary Robotic Grasping, *arXiv* (2024). DOI: 10.48550/arxiv.2407.13175