GIST develops the world's highest-performance AI robot phage model that can collaborate with people: With just one inference, any object can be grasped accurately and stably in various environments... Expected to revolutionize industrial sites

- Department of AI Convergence Professor Kyoobin Lee's team, supports various types of prompt inputs to accurately grasp unlearned objects in complex environments according to the worker's intention... and even a grasping function based on worker gaze tracking

- Expected to be used in various fields such as home and service robots... Scheduled to be presented at the world's largest robot conference, 'IEEE International Conference on Robotics and Automation' in May 2025

▲ (From left) Department of AI Convergence Professor Kyoobin Lee, PhD student Sangjun Noh, Jongwon Kim, Raeyoung Kang, and Dongwoo Nam from the integrated master's and doctoral program (PhD program), and Seunghyeok Back, senior researcher at the Korea Institute of Machinery and Materials

In order for robots to accurately and stably grasp objects in industrial settings, sophisticated grasping technology is essential. However, existing grasping technology requires new model learning depending on the type of object or environmental changes, and has limitations in that it relies on limited data sets.

The Gwangju Institute of Science and Technology (GIST, President Kichul Lim) announced that Professor Kyoobin Lee's research team in the Department of AI Convergence has developed the world's highest-performance innovative robot grasping model 'GraspSAM' that goes beyond simple automation and considers collaboration with workers.

GraspSAM, an innovative AI model that considers collaboration with workers, supports various types of prompt inputs such as points, boxes, and text, and is designed to accurately predict the grasping point of

an object with just one inference. Through this, it has overcome the limitations of existing models and has the ability to stably grasp unlearned objects in complex environments.

Existing deep learning-based phage models had limitations in that they had to learn separate AI models depending on the environment and situation.

To solve this problem, the research team introduced SAM (Segment Anything Model), a general-purpose image segmentation model developed by Meta, Facebook's parent company, to enable robot phage output for the first time.

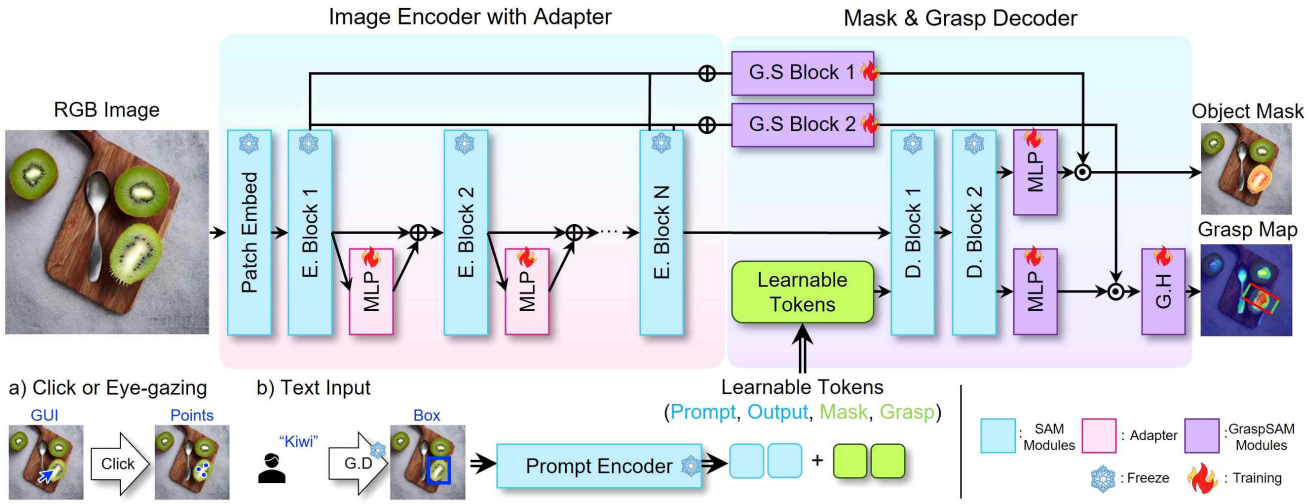
GraspSAM, developed by the research team, is an innovative model that predicts the grasping point of an object with minimal fine-tuning by utilizing SAM's powerful object segmentation ability. To this end, SAM was optimized for phage point inference by applying the Adapter technique and Learnable Tokens technique*.

* SAM (Segment Anything): A general-purpose image segmentation model developed by Meta AI (formerly Facebook). The biggest feature of this model is that it can segment anything (Segment Anything), as its name suggests. Existing image segmentation models were often optimized for specific datasets or specific objects, but SAM is based on a large pre-learned dataset, so it shows excellent performance even on new images or unseen objects.

* adapter and learnable tokens: These two concepts are widely used in efficient fine-tuning of large models. Their roles are slightly different, but their goal is the same: to improve performance with the least amount of resources. First, the adapter is a method that learns by adding only a small network (module) while keeping the parameters of the large model almost fixed. It can learn new data without significantly changing the existing model. Learnable Tokens add learnable embedding tokens to the existing input data (text, image patches, etc.) and pass them to the model. It is used when you want to emphasize context or additional information.

GraspSAM supports prompt-based input and is designed to immediately adapt to various environments, objects, and situations through simple point, box, and text input provided by the user.

This allows the robot to easily grasp a wider variety of objects and dramatically expands the scope of application in industrial settings by predicting the gripping point with just one operation.



▲ GRASPSAM full pipeline: A network that inputs an RGB image and various types of prompts (dots, boxes, or language) and outputs a mask of the target object to be grasped and a phage map for grasping the corresponding object.

GraspSAM achieved the highest level (SOTA, State of the Art)* performance on the famous phage benchmark datasets 'Grasp-anything' and 'Jacquard'. In addition, the experimental results confirmed that the robot can stably perform phage tasks even in complex real environments.

* SOTA (State of the Art): It is mainly used in the fields of artificial intelligence (AI) and machine learning (ML), and a SOTA model generally refers to a model that records the best performance on a benchmark dataset or provides the most efficient and accurate results in a specific task.

In particular, it successfully implemented the function of performing phage tasks in line with the worker's gaze by combining it with eye tracking technology, suggesting wide potential for use in various fields as well as industrial sites. More detailed information can be found on the official GraspSAM website*.

* https://gistailab.github.io/graspsam/

Professor Kyoobin Lee said, "The GraspSAM model enables intuitive interaction between robots and users, and it is expected to be widely used in various fields such as household robots and service robots as it demonstrates excellent grasping ability even in complex environments as well as industrial sites."

This research, supervised by Professor Kyoobin Lee of the GIST Department of AI Convergence and conducted by Ph.D. candidate Sangjun Noh, was supported by the Ministry of Trade, Industry and Energy and the Ministry of Science and ICT. The research results are scheduled to be presented in May 2025 at the IEEE ICRA (International Conference on Robotics and Automation), the world's most prestigious academic society in the field of robotics.

GIST
Since 1993