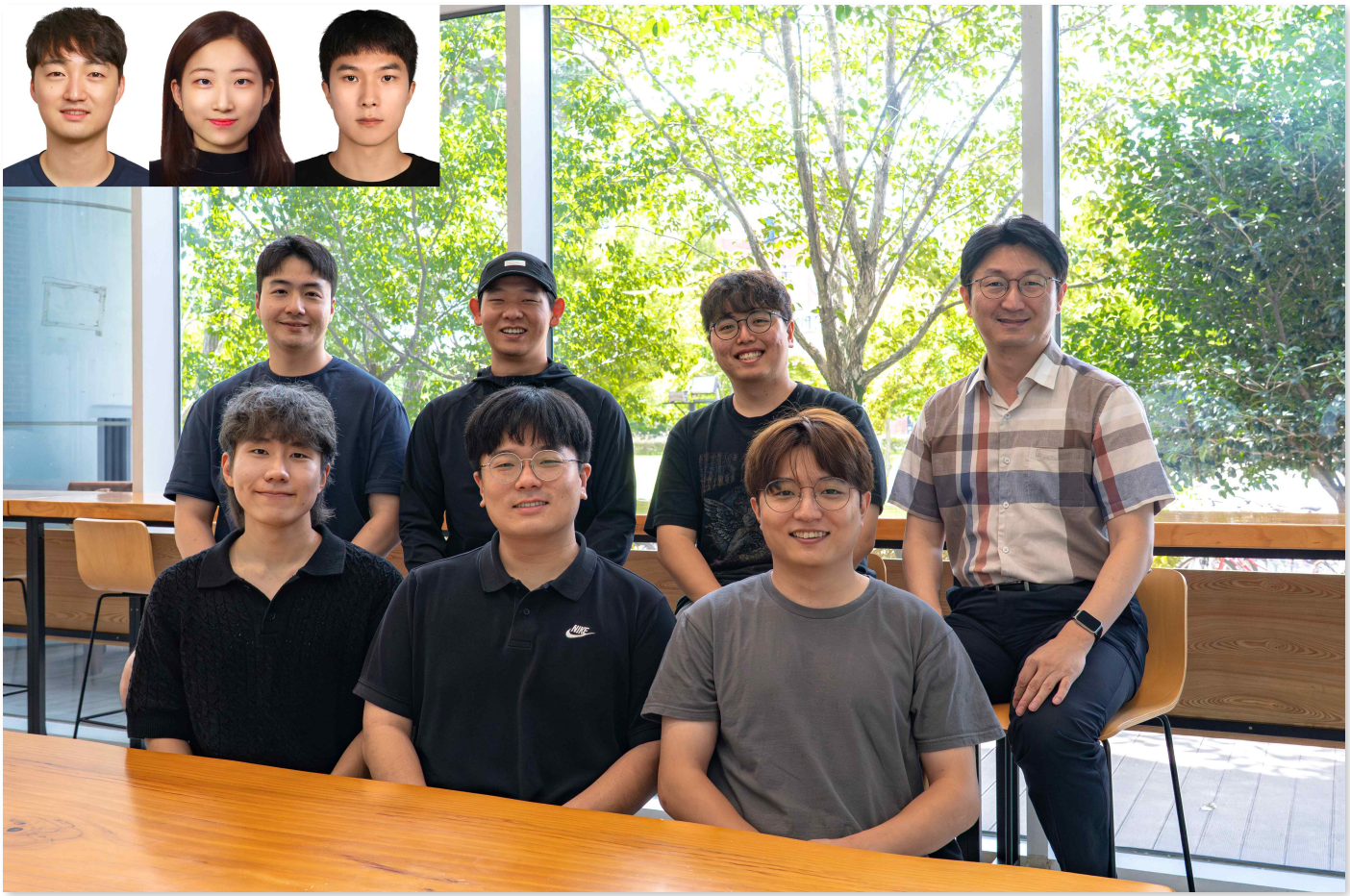
"In the age of physical AI, robots can accurately grasp objects even in cluttered real-world environments" GIST research team unveils the world's largest dataset for training robots to grasp objects in real-world environments

- Professor Kyoobin Lee's team from the Department of AI Convergence has developed and released "GraspClutter6D," a system that accurately reflects the complexities of real-world living and industrial environments, enabling robots to learn to manipulate objects reliably in the real world

- The results show improved performance compared to existing datasets, with the grasping success rate increasing from 77.5% to 93.4% in simple environments and from 54.9% to 67.9% in complex environments... The results were published in the international journal 《IEEE Robotics and Automation Letters》



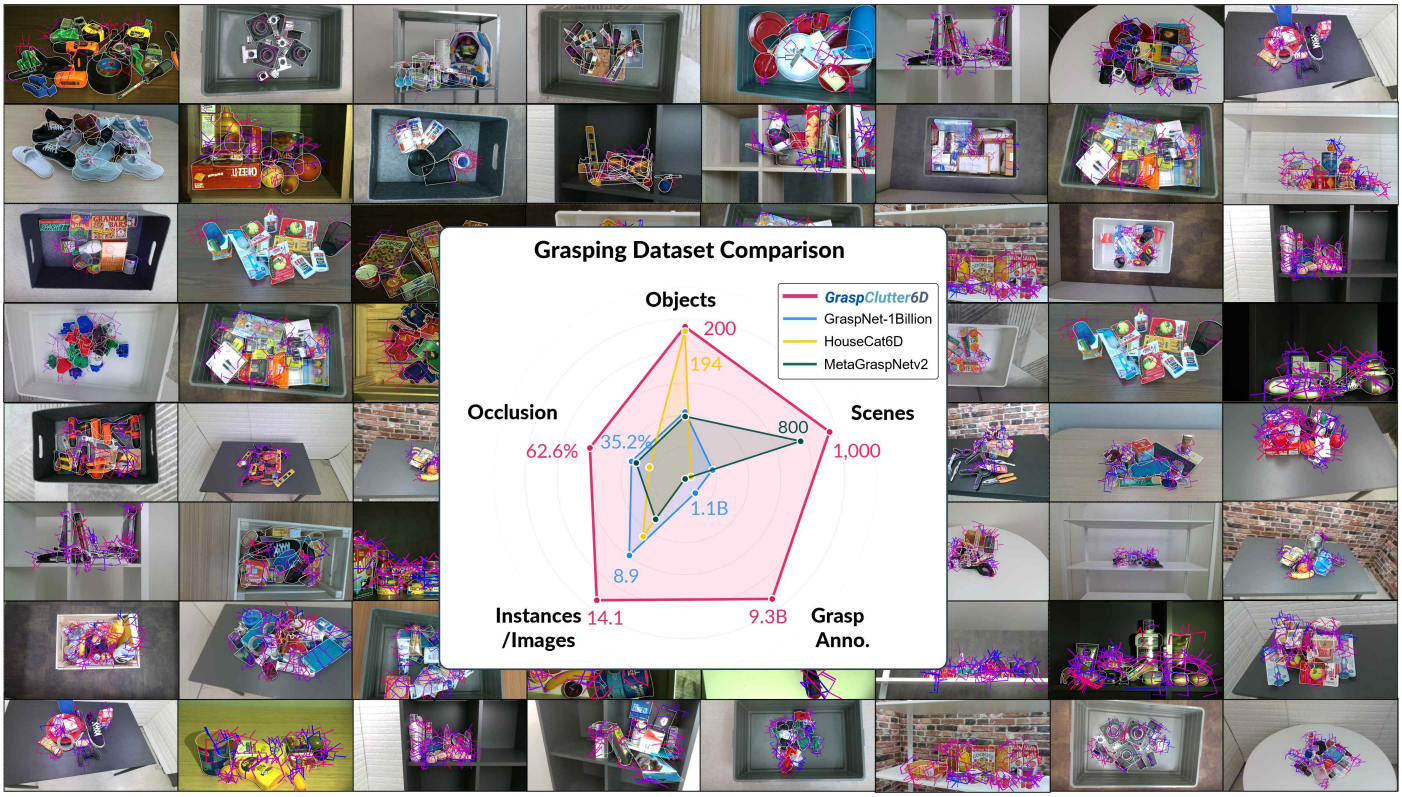
▲ (From top left) Dr. Seunghyeok Back (Korea Institute of Machinery and Materials), GIST Department of AI Convergence Ph.D. candidate Heeseon Rho, and master's candidate Youngjin Lee (clockwise from left on back row) GIST Department of AI Convergence Ph.D. students Geonhyup Lee, Sangjun Noh, Sangbeom Lee, and Professor Kyoobin Lee; and Ph.D. students Raeyoung Kang, Joosoon Lee, and Kangmin Kim

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 built and released "GraspClutter6D," the world's largest robot grasping dataset that precisely reflects the complexity of real-world environments.

This dataset lays the foundation for building a "Robot Foundation Model*" that overcomes the limitations of existing robot AI, which operates only in simple and organized situations, and operates reliably in real-

world environments with entangled objects. It is also expected to contribute significantly to the advancement of "Physical AI," a field that has recently attracted global attention. This research was conducted in collaboration with the Korea Institute of Machinery and Materials (KIMM).

* Robot Foundation Model: A general-purpose robot AI model trained on large-scale robot sensor data, it is characterized by its ability to perform a variety of tasks without being restricted to specific objects or environments.



▲ Example images from the GraspClutter6D dataset and comparison with existing datasets. Average number of objects: 14.1 and 62.6%.

With an occlusion rate of 62.6%, it provides data with approximately twice the level of difficulty compared to existing datasets.

Robot grasping is one of the most basic yet challenging tasks. In particular, in situations where objects overlap and are occluded, such as retrieving items from a warehouse or organizing items at home, it is difficult for robots to accurately recognize and grasp objects reliably.

Recent advances in deep learning technology have significantly improved robot grasping performance. However, previous training datasets have primarily assumed organized and simple environments, limiting their applicability to real-world situations. As a result, robot performance has deteriorated significantly in real-world situations where multiple objects are entangled or the background is diverse.

For example, the widely used dataset "GraspNet-1Billion"* has an average of less than 9 objects per scene and a occlusion rate of only about 35%, failing to adequately reflect real-world situations.

* GraspNet-1Billion: A dataset developed for robots learning to grasp objects, it contains approximately 1 billion robot grasping pose data.

To overcome these limitations, the research team built "GraspClutter6D," an ultra-large dataset that precisely recreates real-world living and industrial environments.

The team created 75 diverse environments, including boxes, shelves, and tables, and mounted four RGB-D (color + depth) cameras* on a robot arm to collect a total of 52,000 images from 1,000 scenes.

The dataset includes ▲ high-quality 3D models of 200 real objects ▲ 736,000 six-dimensional (D) object poses * ▲ 9.3 billion six-dimensional (D) grasping poses of robots *. This is a scale that dwarfs any existing public dataset.

* RGB-D Camera: A sensor device capable of simultaneously acquiring color information (RGB) and depth information (D). It is used in various fields such as robotics, augmented reality, and 3D reconstruction. While conventional cameras only recognize the color and brightness of an object, RGB-D cameras use infrared or structured light to measure the distance (depth) between the object and the camera, enabling them to accurately determine the position and shape of an object in 3D space.

* 6-dimensional (D) grasping posture: This represents the position and orientation of a robot in 3D space, a key concept in robotic manipulation and 3D perception. Specifically, the 6-dimensional posture consists of three position coordinates (x, y, z) and three rotation angles (roll, pitch, yaw), defining exactly where and how the robot rotates to grasp in space.

The research team also released benchmark results evaluating the performance of the latest AI recognition and grasping models along with the dataset.

▲ State-of-the-art object segmentation ▲ 6-dimensional pose estimation ▲ Pickup detection methods were evaluated. While existing AI technologies significantly degrade in complex environments with many overlapping objects, the AI model trained on the "GraspClutter6D" dataset demonstrated a marked performance improvement in real-world robotic object picking experiments.

In a simple environment (5 objects), the pickup success rate improved by 15.9 percentage points, from 77.5% to 93.4%. In a complex environment (15 objects), the pickup success rate improved by 13.0 percentage points, from 54.9% to 67.9%.

▲ Overview of the GraspClutter6D dataset. Collected using a multi-camera capture system in various environments, including shelves, desks, and boxes, it contains 6D poses for 730,000 objects and 9.3 billion grasp poses.

This demonstrates that "GraspClutter6D" is not simply a large dataset, but a "realistic dataset" that faithfully reflects real-world environments.

Professor Kyoobin Lee stated, "This achievement not only faithfully reproduces complex situations encountered in industrial and home environments for the first time, but also provides a crucial foundation for physical AI research aimed at enabling robots to learn and act in the real world. He added, "This will allow us to further advance the use of robots in diverse fields, including logistics, manufacturing, and home services."

"GraspClutter6D" and related tools are freely available to researchers worldwide on a public website*, allowing anyone to utilize them for robotic grasping research, 6-dimensional object pose estimation, and general-purpose robot foundation model training.

* https://sites.google.com/view/graspclutter6d

This research, supervised by Professor Kyoobin Lee of the Department of AI Convergence at GIST and conducted by Senior Researcher Seunghyeok Back of the Korea Institute of Machinery and Materials, was supported by the Ministry of Trade, Industry and Energy's Robot Industry Core Technology Development Program and the Industrial Technology Alchemist Project.

The results of this research were published in the international journal 《IEEE Robotics and Automation Letters》 on August 20, 2025. The research team also plans to present their findings at the IEEE International Conference on Robotics and Automation (ICRA), the world's most prestigious robotics conference, scheduled to be held in Vienna, Austria in June next year. It contains 6D poses and 9.3 billion grasping poses.

GIST
Since 1993