"Robots can accurately recognize objects they see for the first time at a glance!" GIST develops real-time AI vision technology

- Department of AI Convergence Professor Kyoobin Lee's team develops AI technology that automatically corrects unlearned object recognition results through error estimation... Overcoming the limitations of AI vision with technology that precisely corrects misdetected objects in real time

- Expected to be utilized in various fields such as robot vision, autonomous driving, and industrial automation... Scheduled to be presented in May at the IEEE International Conference on Robotics and Automation (ICRA), the world's most prestigious conference in the field of robotics

▲ (Clockwise from the left on the back row) Professor Kyoobin Lee of the Department of AI Convergence, PhD student Jemo Maeng, Hyundai Motor Company Robotics Lab Senior Researcher Sungho Shin, Korea Institute of Machinery and Materials Senior Researcher Seunghyeok Back, integrated PhD students Joosoon Lee and Sangbeom Lee, and PhD student Kangmin Kim

In order for robots to work smoothly in new environments, the ability to quickly and accurately recognize objects they see for the first time is essential. However, existing AI vision technology can only recognize objects included in pre-learned data, which has the limitation of low recognition rates for unlearned objects.

Korean researchers have developed an innovative artificial intelligence (AI) technology to solve this problem, and its use in various fields such as robot vision and autonomous driving is expected.

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 an AI technology that refines the recognition results of unlearned objects* through error estimation.

This technology is expected to significantly improve the visual recognition capabilities of robots, as it has the function of deleting or adding objects that are misdetected in real time.

* unlearned objects: Objects not used in the learning process (objects not included in the learning data)

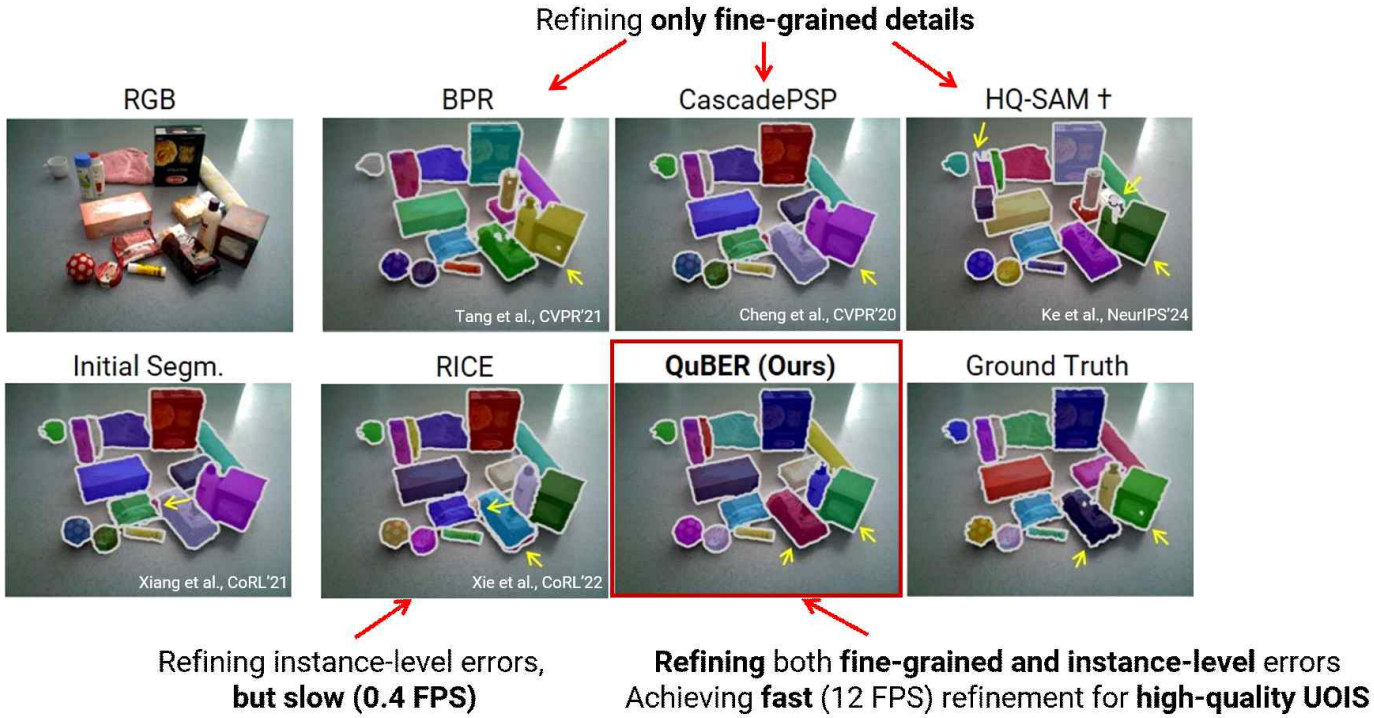
There are some limitations to the currently widely used AI vision technology. The Mask R-CNN (2018)* model can only recognize objects that have been learned in advance, so it lacks the ability to distinguish new objects. The Segment Anything (2023)* model can detect arbitrary objects based on user input (clicks, boxes, text, etc.), but has a problem that accuracy is low in complex environments without human assistance.

To solve this, technologies that correct errors based on images and initial prediction data have been developed. However, existing technologies have the disadvantage of being slow to operate even if they can only correct detailed parts or correct incorrect recognition.

* Mask R-CNN (He et al. TPAMI'18): This is the most widely used model for object segmentation tasks, and has a structure that predicts object region candidates (Region of Interest) and performs classification, bounding box prediction (Bounding Box Regression), and segmentation region prediction (Segmentation Mask Prediction) based on the region. Many object segmentation models are being developed based on this paper.

* Segment Anything (Kirillov et al. ICCV'23): This model was developed due to the need for a foundation model in the computer vision field (e.g. GPT in the natural language field). It was trained with a large-scale dataset of millions of items, and is a model that can segment arbitrary objects in real time based on various user inputs (clicks, boxes, texts, etc.).

To solve this problem, the research team developed the 'QuBER' model that applies fast and accurate error correction technology.



▲ Comparison between the developed model QuBER and existing models. Existing unlearned object segmentation refinement models have limitations in that they can only refine the detailed parts of objects, or they can refine at the object level, but the execution time is too long. QuBER, the model developed by our research team, has the advantage of being able to refine at the object level, such as adding or deleting objects from the detailed parts of objects, while also having a fast execution speed.

The QuBER model is characterized by analyzing the 'Quadruple Boundary Error' using RGB-D (Red · Green · Blue-Depth, color+depth) images and initial prediction data to improve the accuracy of object

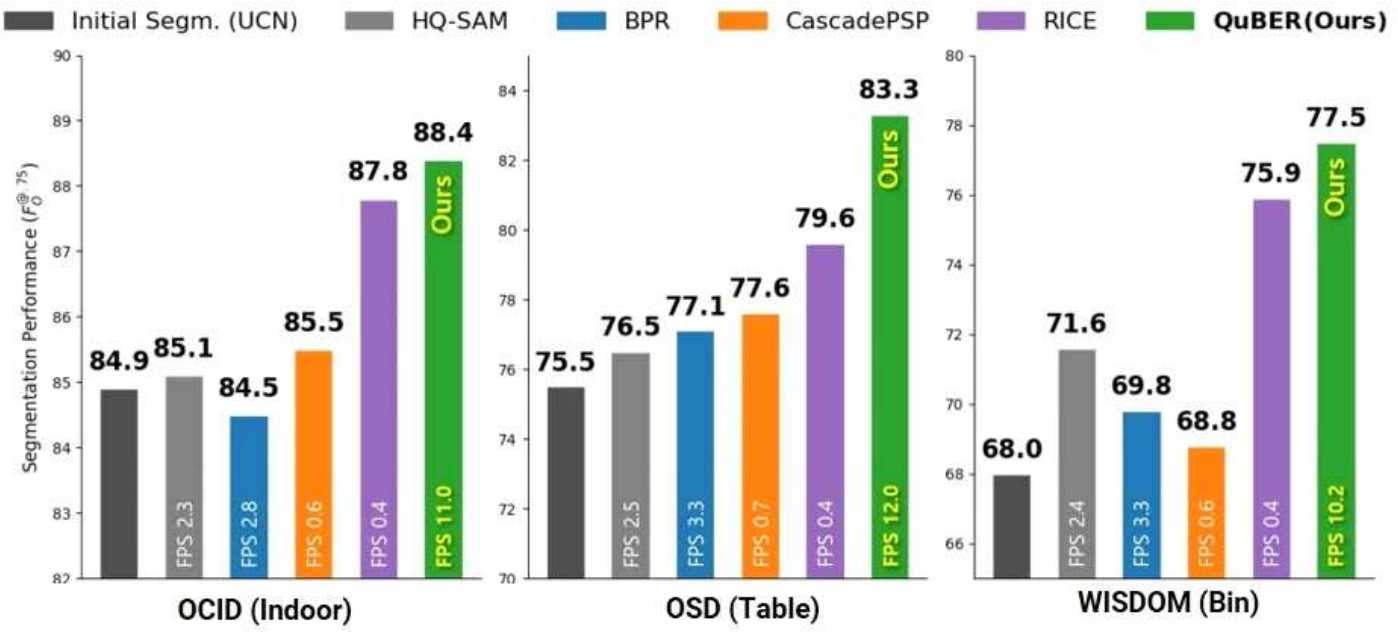
recognition. This technology can quickly and accurately recognize unlearned objects seen for the first time in real time.

In this way, the Quadruple Boundary Error is drawing attention as a key technique for improving the accuracy of AI vision technology.

This technology efficiently corrects errors in object recognition by analyzing the difference between the initial prediction value of AI and the actual data (Ground Truth) into four boundary criteria: ▲True Positive boundary (correctly detected boundary), ▲False Negative boundary (boundary that should have been detected but was missed), ▲False Positive boundary (boundary that was incorrectly detected), and ▲True Negative boundary (boundary that accurately undetected a part that should not have been detected).

The QuBER model developed by the research team showed fast and accurate segmentation ability even in situations where recognition is difficult due to many occluded objects, and recorded the world's highest level of accuracy.

▲ 88.4% recognition rate in indoor environment (OCID), ▲ 83.3% accuracy in objects on table (OSD), and ▲ 77.5% accuracy in complex objects in a box (WISDOM).



▲ Performance comparison graph between untrained object segmentation refinement models. From the left, this is a graph comparing the performance of untrained object segmentation refinement models for the OCID, OSD, and WISDOM datasets. It can be seen that the QuBER model (Ours) shows higher performance than existing models.

Professor Kyoobin Lee said, “Through this study, we have confirmed the possibility that robots can accurately and efficiently recognize objects they see for the first time,” and expressed his expectations, saying, “This technology will be applied to various robot tasks and play a large role in the development of robots that operate stably in new environments.”

This research, supervised by Professor Kyoobin Lee of the GIST Department of AI Convergence and conducted by Dr. Seunghyeok Back while he was a doctoral student, was supported by the Ministry of Trade, Industry and Energy and the Ministry of Science and ICT, and is scheduled to be presented in May 2025 at the IEEE International Conference on Robotics and Automation (ICRA), the world's most prestigious academic society in the robotics field.