

"AI that recognizes household chores with only sound" GIST-Hanwha Vision wins 1st place in international AI acoustic recognition contest

- GIST Audio Intelligent Technology Lab-Hanwha Vision, designed an AI model for detecting sounds occurring in the room
- AI acoustic recognition contest 「DCASE Challenge 2023」, won first place in indoor acoustic event detection category



(From left) GIST student Yuna Song, Professor Hong Kook Kim, Jiwon Kim, student Sangwon Son (from the top of the screen) Hanwha Vision Researcher Ilhoon Song, Senior Researcher Jeongeun Lim

The GIST (Gwangju Institute of Science and Technology, Acting President Raekil Park) joint research team won the first place in the indoor acoustic event detection category on the 1st, beating the world's leading universities and research institutes in the international AI acoustic recognition competition 「DCASE Challenge 2023」*.

* DCASE Challenge: International Challenge on Detection and Classification of Acoustic Scenes and Events

This global competition was held since 2013 by the Signal Processing Society (AASP) under the Institute of Electrical and Electronics Engineers (IEEE) for sound recognition technology that uses artificial intelligence (AI) to listen to sounds and judge situations.

The 'GIST-Hanwha Vision Team' (GIST Jiwon Kim, Sangwon Son, Yuna Song, Hanwha Vision Researcher Ilhoon Song, Senior Researcher Jeongeun Lim) composed of students from the Audio Intelligent Technology Lab led by GIST School of Electrical Engineering and Computer Science Professor Hong Kook Kim and researchers at the Hanwha Vision AI Research Center (Director Seung-in Roh) proved its excellence in research performance by achieving first place in the single model category and second place in the ensemble model category in the indoor acoustic event detection category.

In the indoor acoustic event detection category, AI technology detects and distinguishes 10 sounds, such as vacuum cleaners, washing dishes, dogs barking, and water running, that occur in the indoor environment and ranks them.

The GIST-Hanwha Vision team has significantly improved performance by incorporating AI technology that can implement various audio intelligence. Specifically, ▲ semi-supervised learning technology that utilizes artificial intelligence-inferred correct answers for learning, ▲ technology that fuses the inference results of a pre-learning training model* with the inference results of an existing model, ▲ refining data to improve performance technology, and ▲ an ensemble technology that combines various techniques that improve the performance of artificial intelligence was applied.

If this technology is used, it is expected that it will be used in various applications such as indoor monitoring and AI speakers as it can detect what is going on by listening to the sound even in situations where that cannot be seen with a camera.

* Pre-learning training model: a large-capacity model trained on a large-scale dataset

Professor Hong Kook Kim said, "The experience and technology gained through this competition will be applied to the detection of acoustic events in CCTV developed by Hanwha Vision. Furthermore, we will strive to develop more efficient and user-friendly services, such as technology for detecting voice sections and sound events in social media contents."

A GIST student said, "Thanks to the discussions with Hanwha Vision researchers and the guidance of Professor Hong Kook Kim, we were able to obtain good results. Not satisfied with this achievement, we will accelerate our research so that we can steadily develop an AI model for audio intelligence."

Meanwhile, GIST Audio Intelligent Technology Lab (AiTeR) is researching an AI model for voice and sound recognition. In addition, various researches such as acoustic event detection, voice synthesis, voice noise removal, voice recognition, abnormal situation detection, and biometric information processing are being jointly conducted with domestic and foreign industry-academia research institutes such as MIT.

This research was carried out with the support of the media content voice language localization technology development project of Hanwha Vision, the Ministry of Science and ICT, and the Information and Communication Planning and Evaluation Institute.