

GIST, 인공지능 오디오 분야 연구력 입증

국제 AI 음향 인식 대회 「IEEE DCASE 2024 챌린지」 1위

- 전기전자컴퓨터공학부 김홍국 교수 연구팀, 언어질의 기반 오디오 소스 분리 (Language-Queried Audio Source Separation) 부문서 1위 달성
- 거대언어모델(LLM) 기반 데이터 증강의 새로운 기법 제시 성과



▲ (왼쪽부터) 전기전자컴퓨터공학부 김홍국 교수, 이도현 학생, 송윤아 학생

광주과학기술원(GIST, 총장 임기철)이 인공지능(AI) 기반 음성 신호처리 분야에서 탁월한 연구 성과로 주목받고 있다.

오디오 지능을 연구하는 전기전자컴퓨터공학부 김홍국 교수 연구팀은 국제 음향 장면 및 이벤트 탐지 분류(DCASE, IEEE AASP Challenge on Detection and Classification of Acoustic Scenes and Events) 경진대회에서 '언어질의 기반 오디오 소스 분리(Language-Queried Audio Source Separation) 부문' 1위와 '실내 음향 이벤트 탐지 부문' 3위의 성과를 거뒀다.

국제전기전자공학회(IEEE) 산하 신호처리소사이어티(AASP)가 2013년부터 개최하는 세계적인 대회로, 인공지능(AI)을 이용해 소리를 듣고 상황을 판단하는 음향 인식 기술을 겨룬다. 올해는 4월 1일부터 6월 15일까지 약 3개월간 108개의 우수 기관 및 대학이 참가해 음향 장면 인식, 기계 이상 진단 확인 등 10개 분야에서 경쟁을 펼쳤다.

오디오지능연구실(AiTeR) 학생들로 구성된 'GIST-AunionAI' 팀(석박통합과정 송윤아·석박통합과정 이도현 학생, 김홍국 교수)은 김홍국 교수의 창업 회사인 오니온에이아이(AunionAI)의 지원을 받아 'DCASE 챌린지 2024'의 '언어질의 기반 오디오 소스 분리 과제 (Language-Queried Audio Source Separation, Task 9)' 부문에서 1위를 달성하며 연구 성과의 우수성을 입증했다.

또한, 한화비전 연구원과 함께 구성된 **GIST-HanwhaVision 팀**(GIST: 석사과정 손상원 학생, 석박통합과정 박종연 학생, 김홍국 교수, HanwhaVision: 노승인 상무, 임정은 수석연구원, Sulaiman Vesal)은 **음향 이벤트 탐지 과제(DCASE Task 4)**에서 **3위를 차지했다.**

‘언어질의 기반 오디오 소스 분리(LASS) 기술’은 사용자가 입력한 텍스트에 따라 이에 맞는 오디오 신호를 분리하는 기술이다. 텍스트 질의를 통해 오디오 소스를 분리 생성함으로써 언어와 오디오를 연결하는 생성형 AI 모델 개발의 기초를 제공하며 **자동 오디오 편집, 멀티미디어 콘텐츠 검색, 증강 청취 등 다양한 응용 분야에서 활용될 수 있다.**

이번 경진대회에서 ‘GIST-AunionAI’ 팀은 다양한 오디오 지능을 표현할 수 있는 AI 기술들을 접목하여 고성능 **언어질의 기반 오디오 소스 분리 기술**을 개발했다. **▲ LLM(거대언어모델) 기반 프롬프트 기술 및 데이터 증강 기술 ▲사전 학습 훈련 모델*과 기존 모델의 추론 결과 융합 기술** 그리고 **▲AI 능력 향상을 위한 앙상블 기술**을 통해 AI 모델을 개선하였다.

* 사전 학습 훈련 모델 : 대량의 데이터 셋에 대해 학습된 대용량의 모델

‘실내외 음향 이벤트 탐지 기술’은 실내외 환경에서 발생할 수 있는 진공청소기 소리, 설거지하는 소리, 차량 소리 등 **27종류의 다양한 소리를 AI를 통해 탐지 및 구별하는 기술이다.** 카메라를 통해 처리하는 데 제한이 있는 상황에서 소리만으로 음향 이벤트를 탐지할 수 있다는 강점이 있어 실내외 상황 감시, 차량 모니터링 등 다양한 어플리케이션에 활용될 수 있다.

GIST-HanwhaVision 팀은 다양한 오디오 지능을 표현할 수 있는 AI 기술들을 접목하여 고성능 **실내외 음향 이벤트 탐지 기술**을 개발했다. **▲보조 분류기 기반 모델 학습 기술 ▲다양한 입력 특성 추출 기술** 등을 통해 AI 모델을 개선함으로써 우수한 성과를 거뒀다.

김홍국 교수는 “GIST 연구실과 (주)오니온에이아이, 한화비전이 협업하여 얻은 성과로서 개발된 AI 모델이 연구실에 머물지 않고 **사업화로 나아갈 수 있다는 가능성에 그 의의가 매우 크다**”며 “특히, LLM 기반 오디오 생성 및 인식 AI 모델을 지속적으로 개선하는 노력과 함께 이를 다양한 분야에 적용하여 **편리하고 안전한 삶을 위한 기술 발전에 기여하겠다**”고 말했다.

‘GIST-AunionAI’ 팀 학생들은 “김홍국 교수님의 지도와 아낌없는 지원 덕분에 국제 경진대회에서 좋은 성과를 얻을 수 있었다”고 말하며, “이번 성과에 만족하지 않고 오디오 지능에 대한 AI 모델을 꾸준히 발전시킬 수 있도록 연구에 박차를 가하겠다”고 다짐했다.

한편, GIST 오디오지능연구실(AiTeR, 지도교수 김홍국)에서는 음성 및 오디오와 관련된 다양한 AI 모델을 연구하고 있으며, 언어질의 기반 음성 소스 분리뿐만 아니라 음향 이벤트 탐지, 음성 합성, 음성 잡음 제거, 음성 인식, 이상 상황 감지, 다국어 인식 및 번역 등 다양한 연구를 국내 산업체와 대학, 연구기관 및 미국 매사추세츠공대(MIT) 등 해외 연구기관과 공동으로 진행하고 있다.

‘언어질의 기반 오디오 소스 분리’는 MIT 국제 공동 연구로 GIST 과학기술혁신사업단의 ‘실용화 연구개발사업’, 연구개발특구진흥재단의 ‘지역의 미래를 여는 과학기술 프로젝트’의 지원으로 수행되었으며, ‘음향 이벤트 탐지 연구’는 한화비전과 과학기술정보통신부 및 정보통신기획평가원의 ‘미디어 콘텐츠 음성 언어 현지화 기술개발 사업’의 지원으로 수행되었다.