

# Adaptive Cross-Attention Gated Network for Radar-Camera Fusion in BEV Space

Ji-Yong Lee\*, \*\*, Jae-Hyeok Lee\*, Dong-oh Kang\*

\*Visual Intelligent Lab., ETRI, Daejeon, Republic of Korea

\*\*The University of Texas at Austin, Austin, TX 78712, USA

jiyong.lee@etri.re.kr, heuyklee@etri.re.kr, dongoh@etri.re.kr

**Abstract**— Fusing multimodal sensors for 3D object detection has been extensively researched in the field of autonomous driving. However, existing multimodal sensor fusion methods still struggle to provide reliable detection across different modalities under diverse environmental conditions. Specifically, straightforward methods like summation or concatenation in radar-camera fusion may lead to spatial misalignment and fail to localize objects in complex scenes. To address this, we propose Adaptive Cross-Attention Gated Network (ACAGN) to enhance radar-camera fusion capabilities in Bird’s-Eye View (BEV) space. Our approach integrates a deformable cross-attention and an adaptive gated network mechanism. The deformable cross-attention aligns radar and camera features from BEV with greater spatial precision, handling variations between those features effectively. Meanwhile, the adaptive gated network dynamically filters and prioritizes the most relevant information from each sensor. This dual approach improves stability and robustness of detection, as demonstrated through extensive evaluations on the nuScenes dataset.

**Keywords**— Autonomous driving, multi-modal sensor fusion, millimeter-wave radar, multi-view camera, 3D object detection



**Ji-Yong Lee** received his B.S. degree in Electrical and Computer Engineering at the University of Texas at Austin in 2021 and is currently pursuing a master's degree in Artificial Intelligence. He is currently a research associate at the Electronics and Telecommunications Research Institute (ETRI), Daejeon, South Korea. His research interests include machine learning, knowledge distillation, and multi-modal sensor fusion.



**Jae-Hyeok Lee** received the Ph.D. degree in electrical engineering from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea, in 2024. He is currently a researcher at the Electronics and Telecommunications Research Institute (ETRI), Daejeon, South Korea. His current research interests include machine learning, computer vision, 3D vision, and interpretable deep learning.



**Dong-oh Kang** received his B.S. degree in electronic engineering from Yonsei University, Korea, in 1994, and his M.S. degree and Ph.D. in electronic engineering from Korea Advanced Institute of Science and Technology (KAIST), Korea, in 1996 and 2001, respectively. Since 2001, he has been working at Electronics and Telecommunications Research Institute (ETRI) and is currently a principal researcher of the Visual Intelligence Research Section. His research interests include multimodal knowledge representation, knowledge graph completion, and self-growing intelligent agents.