YOLOv12 기반 RGB-IR 멀티모달 융합을 이용한 차량 내 객체 검출에 관한 연구

김현덕, 이상헌, 손명규, 김준광
대구경북과학기술원
hyunduk00@dgist.ac.kr, pobbylee@dgist.ac.kr, smk@dgist.ac.kr, kjk1208@dgist.ac.kr

A Study on the Multimodal RGB-IR Fusion for In-Cabin Object Detection using YOLOv12

Hyunduk Kim, Sang-Hoen Kee, Myoung-Kyu Sohn, Junkwang Kim
Daegu Gyeongbuk Institute of Science & Technology (DGIST)

요 약

최근 차량 내 안전성과 편의성 향상을 위해 in-cabin 객체 검출 기술에 대한 연구가 활발히 이루어지고 있다. 기존의 단일 모달 기반(RGB 또는 IR) 접근 방식은 조명 변화나 반사, 저조도 환경 등 다양한 주행 조건에서 인식 성능이 불안정하다는 한계를 가진다. 이에 본 연구에서는 이러한 문제를 해결하기 위해 YOLOv12 기반의 RGB - IR 멀티모달 객체 검출 모델을 제안한다. 제안된 모델은 입력 단계에서 RGB 및 IR 영상을 동시에 받아 Early Fusion 기반의 Gated Fusion 모듈을 통해 두 모달리티의 특징을 효과적으로 융합한다. Gated Fusion은 학습 가능한 게이트를 통해 RGB와 IR 정보의 상대적 중요도를 동적으로 조정함으로써, 조명 변화나 환경적 요인에 강건한 특징 표현을 학습할 수 있도록 한다. 융합된 특징은 YOLOv12 백본을 통해 다단계 특징 추출 및 객체 검출을 수행한다. 실험은 차량 내부 객체 인식을 위한 공개 데이터셋인 SVIRO를 이용하여 수행하였다. 실험 조건은 RGB 단일 모달, IR 단일 모달, 그리고 RGB - IR 멀티모달 입력으로 나누어 비교하였으며, 제안된 융합 방식의 유효성을 검증하였다. 본 연구는 멀티모달 융합을 통한 조명 변화에 강인한 객체 검출의 가능성을 제시하며, 향후 실시간 차량 내 탑승자 모니터링 시스템으로의 확장 가능성을 보인다.

Ⅰ. 서 론

최근 차량 내 안전성과 편의성 향상을 위한 탑승자 모니터링 시스템(in-cabin monitoring system) 연구가 활발히 진행되고 있다. 이러한 시스템은 차량 내부의 탑승자 상태나 좌석 점유 여부를 인식하여 안전벨트 착용 감지, 아동 방치 방지, 에어백 제어 등 다양한 운전자 보조 기능으로 확장될 수 있다.

기존의 많은 연구들은 RGB 카메라를 이용한 단일 모달 기반 접근 방식을 사용해 왔다. 그러나 RGB 영상은 조명 변화나 햇빛 반사, 야간 환경 등 조도에 민감하다는 한계를 지닌다. 예를 들어, 실내 조명이 약하거나 외부 빛이 창문을 통해 직접 유입될 경우 객체의 윤곽이나 질감 정보가 손실되어 정확한 검출이 어렵다. 이러한 문제를 해결하기 위해 적외선(IR, Infrared) 센서가 주목받고 있다. IR 영상은 가시광에 비해 조명 변화의 영향을 덜 받고, 야간 환경에서도 안정적인 감지가 가능하다.

하지만 IR 영상은 텍스처 정보가 부족하고, 물체 간의 명암 대비가 낮아 객체의 세부 분류에는 불리하다. 따라서 RGB와 IR 각각의 장점을 결합한 멀티모달 기반 객체 인식이 새로운 연구 흐름으로 대두되고 있다. 두 모달리티는 상호 보완적인 정보를 제공하므로, 이를 적절히 융합하면 다양한 조명 환경에서도 강건한 인식 성능을 확보할 수 있다 [1, 2].

본 연구에서는 차량 내부 객체 검출을 위한 멀티모달 기반 YOLOv12 구조를 제안한다. 제안된 모델은 RGB 및 IR 입력을 동시에 받아, Early Fusion 기반의 Gated Fusion 모듈을 통해 두 모달리티의 특징을 효과적으로 통합한다. Gated Fusion은 학습 가능한 게이트를 통해 RGB 및 IR 특징의 중요도를 동적으로 조정함으로써, 환경 변화에 따라 적응적으로 정보 융합이 가능하다. 융합된 특징은 YOLOv12 백본 네트워크로 전달되어 객체 검출을 수행한다.

Ⅱ. 본론

YOLO(You Only Look Once) 시리즈는 객체 검출 분야에서 높은 정확도와 빠른 추론 속도를 동시에 달성한 대표적인 단일 단계(object-level, single-stage) 검출기이다. 최근 발표된 YOLOv12는 이전 버전 대비 백본(Backbone)과 넥(Neck) 구조가 크게 개선되어, 경량화와 정밀도를 모두 향상시킨 모델이다 [3]. 세부적으로, YOLOv12는 효율적인 특징 추출을 위해 A2C2F(Adaptive Additive Cross-scale Fusion) 블록을 도입하여, 다양한 스케일의 피처맵 정보를 효과적으로 통합한다. 또한 C3k2 모듈을 통해 지역적 특징(Local feature)과 전역적 문맥(Global context)을 균형 있게 학습하도록 설계되었다. 넥 부분에서는 다단계 업샘플링과 피처 결합을 통해 중간 계층의 정보를 활용하여, 작은 객체나 세부 영역에 대한 검출 성능을 강화한다. 이러한 구조적 개선을 통해 YOLOv12는 기존 YOLO 계열보다 더 적은 파라미터 수로 높은 검출 성능을 유지하며, 다양한 환경에서의 실시간 객체 인식에 적합하다.

본 연구에서는 YOLOv12를 기반으로 한 RGB+IR 멀티모달 in-cabin 객체 검출 모델을 제안한다. 제안된 모델의 전체 구조는 Fig. 1에 도시되어 있으며, 크게 Early Fusion 모듈, YOLOv12 Backbone, Neck, 그리고 Detection Head로 구성된다. 먼저, 입력으로 주어진 RGB(3채널) 영상과 IR(1채널) 영상을 Early Fusion 단계에서 융합한다. 이때 단순한 채널 결합(concatenation) 대신, Gated Fusion 방식을 도입하여 두 모달리티의 상대적 중요도를 학습적으로 조절하도록 설계하였다.

Early Fusion 모듈은 RGB와 IR 각각에 대해 Instance Normalization을 수행하여 조명이나 감도 차이에 의한 통계적 불균형을 보정한 뒤, 두 모달을 채널 방향으로 결합한다. 이어서 합성된 특징에 대해 두 단계의 합성곱 연산과 Batch Normalization, Sigmoid 활성화를 적용하여 RGB와 IR 각

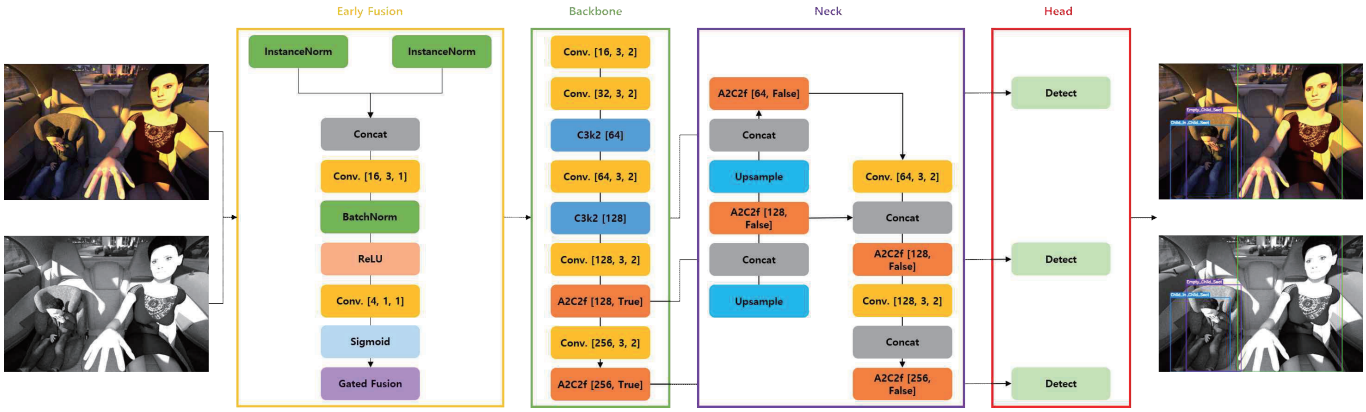

그림 1. 제안한 멀티모달 차량 내 객체 검출 알고리즘 구조도

각에 대한 게이트 맵(gate map)을 생성한다. 이 게이트는 RGB 및 IR 피처의 공간적 가중치를 조절하며, 네트워크가 상황에 따라 조명에 강한 IR 정보 혹은 질감이 풍부한 RGB 정보를 선택적으로 반영하도록 돕는다. 즉, Gated Fusion은 환경 변화에 따른 모달리티 의존성을 동적으로 조정함으로써, 단일 모달 기반 모델이 가지는 취약점을 완화한다.

Gated Fusion을 통해 생성된 융합 피처맵(fused feature)은 YOLOv12의 백본으로 전달되어 다단계 특징 추출 및 통합 과정을 거친다. 이후 넥(Neck) 구조에서는 상하위 피처맵 간의 skip connection과 업샘플링을 통해 다중 스케일 정보를 결합하며, 최종적으로 Detection Head에서 객체의 위치와 클래스(예: Adult, Child, Child Seat 등)를 예측한다.

이러한 방식은 RGB 또는 IR 단일 입력에서도 동작 가능하지만, 두 입력을 함께 사용할 때 가장 높은 표현력을 발휘한다. 특히, 실내 조명이 불안정한 환경이나 강한 빛 반사가 존재하는 상황에서도 안정적인 검출 성능을 기대할 수 있다.

본 연구에서는 제안한 멀티모달 객체 검출 모델의 성능을 검증하기 위해 SVIRO (Synthetic Vehicle Interior Rear-seat Occupancy) 데이터셋 [4]을 사용하였다. SVIRO는 차량 내부 환경을 사실적으로 모사하기 위해 3차원 모델링 기반으로 생성된 합성 영상 데이터셋으로, 다양한 차량 모델과 좌석 구성, 조명 조건을 포함하고 있다. 각 차량의 뒷좌석 영역에는 성인, 아동, 유아, 아동용 카시트, 유아용 카시트, 일상적 물체, 빈 좌석 등 여러 객체가 배치되어 있으며, 이러한 구성이 차량 내부 상황 인식을 위한 실험에 적합하다. 이 데이터셋은 RGB 영상과 함께 적외선(IR) 영상, 깊이 맵, 세그멘테이션 마스크 등 다양한 모달리티를 제공하여 멀티모달 학습 실험이 가능하다. 특히, 낮과 밤, 강한 조명, 부분 그림자 등 다양한 조명 환경에서 촬영된 장면을 포함하고 있어, 실제 차량 내부 조명 변화에 따른 모델의 강건성을 평가하기에 적합하다. 또한 차량 모델별로 훈련 세트와 테스트 세트가 명확히 구분되어 있어, 학습 차량과 다른 내부 구조를 가진 새로운 차량 환경에서도 모델의 일반화 성능을 검증할 수 있다.

표 1은 RGB, IR, 그리고 RGB+IR 멀티모달 입력에 대한 객체 검출 성능을 비교한 결과이다. 실험 결과, RGB+IR 멀티모달 모델이 모든 지표에서 가장 우수한 성능을 보였다. 특히 mAP@50 기준에서 RGB only 모델의 0.715, IR only 모델의 0.723에 비해 멀티모달 모델은 0.766을 기록하여, 약 5%p 이상의 성능 향상을 달성하였다. 또한, 조명 변화에 영향을 받는 RGB 입력과 텍스처 정보가 부족한 IR 입력의 단점을 상호 보완함으로써, 전반적인 검출 신뢰도가 개선되었다. 이러한 결과는 제안한 Gated Fusion 기반 멀티모달 융합 방식이 RGB의 질감 정보와 IR의 조명 강건성을 효과적으로 결합하여, 차량 내부 환경에서도 안정적인 객체 인식을 가능하게 함을 보여준다.

표 1. 객체 검출 성능 비교

Modality	Accuracy			
	Precision	Recall	mAP50	mAP50-95
RGB	0.491	0.828	0.715	0.576
IR	0.687	0.647	0.723	0.568
RGB+IR	0.729	0.654	0.766	0.61

Ⅲ. 결론

본 연구에서는 차량 내부 환경에서의 객체 인식을 위해 YOLOv12 기반 RGB - IR 멀티모달 검출 모델을 제안하였다. 제안된 모델은 Early Fusion 기반의 Gated Fusion 모듈을 통해 RGB와 IR 영상의 상호 보완적 정보를 효과적으로 융합하여, 조명 변화나 반사 등 다양한 환경에서도 안정적인 객체 검출이 가능하도록 설계되었다. 공개 데이터셋을 이용한 실험 결과, 멀티모달 융합이 차량 내 객체 검출의 정확도와 강건성 향상에 유효함을 확인 하였으며, 향후 실시간 in-cabin 모니터링 시스템으로의 확장을 목표로 한다.

ACKNOWLEDGMENT

본 연구는 2025년 과학기술정보통신부 지원하는 DGIST 기관고유사업 (25-IT-01)의 지원을 받아 수행된 연구임.

참 고 문 헌

[1] F. Erlik Nowruzi, W.A. El Ahmar, R. Laganiere, & A. H. Ghods, “In-vehicle occupancy detection with convolutional networks on thermal images,” In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. 2019.

[2] S. Vadidar, A. Kariminezhad, C. Mayr, L. Kloeker, & L. Eckstein, “Robust environment perception for automated driving: A unified learning pipeline for visual-infrared object detection,” In IEEE Intelligent Vehicles Symposium. pp. 367-374, 2022.

[3] Y. Tian, Q. Ye, & D. Doermann, “Yolov12: Attention-centric real-time object detectors,” arXiv preprint arXiv:2502.12524. 2025.

[4] S. D. D. Cruz, O. Wasenmuller, H. P. Beise, T. Stifter, & D. Stricker, “Sviro: Synthetic vehicle interior rear seat occupancy dataset and benchmark,” In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 973-982, 2020.