SPARK 데이터셋 기반 우주 객체 인식을 위한 RGB-Depth 융합 전략의 정량적 비교

장선우, 이은규*

인천대학교

{wkdtjsdn77, eklee}@inu.ac.kr

Revisiting RGB-Depth Fusion Strategies for Satellite and Debris Classification in the SPARK Dataset

Sun Woo Jang and Eun-Kyu Lee*

Incheon National University

요 약

SPARK는 RGB와 Depth 이미지로 구성된 우주 객체 인식용 데이터셋으로, 약 15만 장의 이미지가 11개의 클래스로 분류되어 있고, 모두 현실적인 시뮬레이션 환경에서 생성되었다. 해당 데이터셋을 공개한 연구에서는 무작위 초기화, 특징 추출, 미세 조정 방식 비교, 센서 노이즈 영향 분석, 그리고 다중 모달 우주선 인식 등 다양한 실험을 수행하였다. 그중, 다중 모달 인식 실험은 이미지 크기를 줄인 예비 평가이며, RGB만 학습했을 때와 Depth 이미지를 함께 학습했을 때의 성능도 간단히 언급만 하고 넘어갔다. 본 연구는 SPARK 데이터셋을 기반으로 총 다섯 가지융합 전략을 적용하여, RGB 단독 학습과 융합 구조 간의 성능 차이를 보다 정량적이고 체계적으로 비교한다. 특히, 파편과 위성 클래스의 분리 성능을 중심으로, 융합 방식이 우주 객체 인식 성능에 미치는 영향을 평가하였다. 실험 결과, 모든 융합 방식이 RGB 단독 입력보다 우수한 성능을 보였으며, Gated Fusion과 Late Fusion은 파편 감지와 위성 분류 성능 모두에서 가장 우수한 결과를 기록하였다. 본 연구는 융합 전략의 선택이 실질적인 성능에 영향을 미친다는 점을 정량적으로 보여주며, 향후 우주 객체 인식 시스템에서의 기초적 기준을 제공할 수 있다.

Ⅰ. Introduction

우주 객체 인식 및분류는 중요한 기술로 부상하고 있다. 이를 위한 데이터셋인 SPARK[1]는 약 15만 장의 시뮬레이션 기반 RGB 및 Depth(깊이) 이미지를 포함하고 있으며, 이를 활용한 다양한 분류 실험을 통해 우주 환경에서의 객체 인식 성능을 측정할 수 있다.

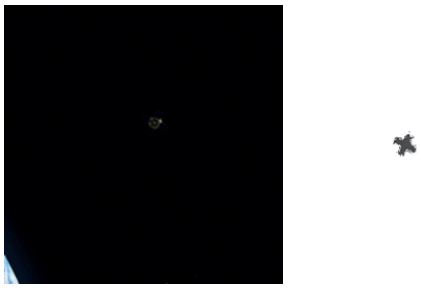

그림 1: SPARK 데이터셋의 예시 이미지. 왼쪽은 RGB 이미지, 오른쪽은 해당 객체의 Depth 이미지

그러나 SPARK를 제안한 기존 연구에서는 RGB 단독 학습과 RGB-Depth 융합 학습 간 성능 차이를 단순 수치로만 제시하고 있으며, 이미지 크기를 제한한 예비 학습으로 진행되었다. 또한, 기존 연구들은 대부분 하나의 융합 구조에 집중하거나, 융합 전략 간 비교 없이 성능 향상을 보고하는 데 그치고 있다.

본 연구는 이러한 문제를 바탕으로, 다양한 수준의 융합 구조를 동일한 조건 하에서 비교 분석함으로써 어떤 방식의 융합 방식이 SPARK 데이터셋에서 더 효과적인지 정량적인

결과로 나타낼 것이다. 구체적으로는 다음과 같은 융합 방식을 사용한다:

- RGB 단독 학습
- Early Fusion: 입력 단계에서 RGB와 Depth를 채널 차원에서 융합 (concat)
- Late Fusion: RGB와 Depth 이미지의 특징을 각각 추출한 후 융합 (concat)
- Cross Attention Fusion: Late Fusion 방식의 하나로, RGB를 쿼리, Depth를 키/밸류로 사용하는 구조적 융합
- Gated Fusion: RGB와 Depth의 특징 중요도를 동적으로 조절

본 연구의 실험은 다음과 같은 측면에서 의의를 갖는다. 기존 SPARK 논문에서는 RGB와 Depth 융합의 효과를 간략한 수치로만 비교하였으나, 본 연구는 다양한 융합 전략의 성능을 정량적 지표로 분석하여 각 전략 간 차이가 실제 성능에 어떤 영향을 미치는지를 객관적으로 평가한다. 또한, 단순히 이어붙이는 concat 방식 외에도 Cross Attention, Gated Fusion 같은 융합 기법을 비교하여, 서로 다른 데이터 간 파편 검출이나 위성 분류에 어떤 성능 향상을 가져오는지 실험적으로 검증한다.

Ⅱ. Background and Related Works

최근 RGB와 Depth와 같은 다양한 환경의 데이터를 융합하여 객체 인식 성능을 향상시키려는 연구가 활발히 진행되고 있다.

이 중 Early Fusion과 Late Fusion은 널리 사용되는 방식들이다. Early Fusion은 입력 단계에서 두 데이터를 채널 기준으로 이어 붙이고, Late Fusion은 RGB와 Depth 이미지의 특징을 각각 추출한 후 융합한다. [2], [3]에서는 이러한 Early Fusion과 Late Fusion 구조를 모두 구현하고 정량적으로 비교하였다. 본 연구에서는[3]의 융합 방식을 기반으로 Early Fusion과 Late Fusion 구조를 사용한다.

또한, 융합 과정에서 단순히 이어 붙이는 방식 외에도 CrossAttention[4]과 Gated Fusion[5] 같은 융합 방식도 도입했다. Cross Attention은 RGB 특징을 Query로, Depth 이미지의 특징을 Key와 Value로 설정하여 서로 다른 데이터 간의 상호작용을 학습한다. Gated Fusion은 RGB와 Depth 이미지의 특징 간 상대적 중요도를 동적으로 조절하여 융합하는 방식으로, SPARK 연구에서도 사용되었다.

기존 연구들은 RGB와 Depth의 융합을 포함한 다양한 전략을 통해 객체 인식 성능 향상을 시도해왔으며, 각각의 논문은 융합 위치, 구조 등에 따라 다양한 접근 방식을 보였다.

이처럼 기존 연구들을 보면 RGB와 Depth 융합 구조에 따라 다양한 전략을 제안하고 있으며, 융합 위치, 방식 등에 따라 성능이 달라진다는 점을 알 수

있다. 본 연구는 이러한 다양한 융합 전략을 동일한 조건에서 실험하여 각 방식의 실질적인 성능 차이를 정량적으로 분석해 나타낸다.

Ⅲ. Experiments and Results

본 연구는 모델 정확도와기존 SPARK 연구의 Task1에서 정의된 지표를 사용해 각각의 융합 전략의 성능을 비교한다. 사용한모델은 EfficientNet-B0이고, RGB 단독 모델은배치 크기, 나머지는 32 배치 크기로 20에폭 반복 훈련을 진행했다. 각 융합 전략의 지표는 3개의 시드로 평균을 낸 결과이다.

F_2-score(debris)는 파편(debris) 클래스에 대해 정밀도(precision)와 재현율(recall)을 이용해 계산된다. 이는 파편객체를 놓치지 않는 것이 중요하다는 점을 반영한다. Accuracy(Satellite)은 파편을 제외한 나머지클래스에 대해 계산된 정확도이다. 정상 위성 객체를 얼마나 정확하게 구분했는가를 측정한다. Performance는 두 값을 단순히 더한 것으로, 파편과 위성 인식 성능을 종합평가한다. 이 3가지 성능 지표는 SPARK task1[1]에서 채택한 평가 방식이고 본 연구에서는 여기에 전체적인 정확도(Accuracy) 수치를 같이 비교할 것이다.

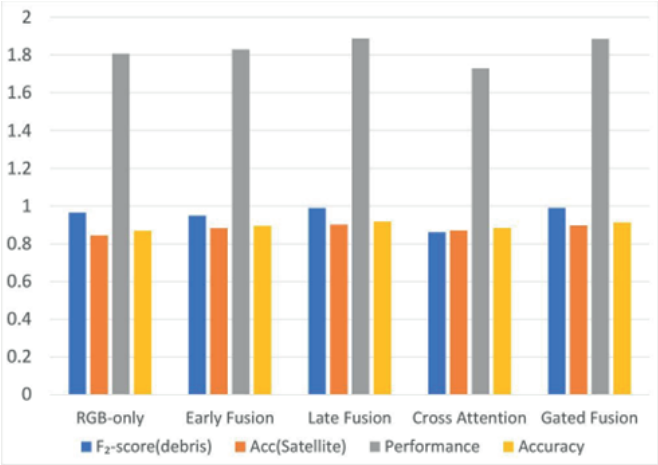

그림 2: 융합 전략 별 성능 비교 그래프

실험 결과, 모든 융합 구조가 RGB 단독 모델보다 파편 감지 및 위성 분류 성능에서 우수한 결과를 보였다. 특히 Late Fusion과 Gated Fusion의 성능이 돋보였는데, 각각 98.79% 및 98.98%의 F_2-score(debris)를 기록하며 파편 객체를 높은 확률로 놓치지 않는다는 것을 보였고, 위성 정확도 역시 90.1% 및 89.61%로 가장 우수하였다. 반면, Cross Attention 방식은 상대적으로 낮은 파편 감지 성능(86.05%)을 보였으며, 이는 attention이 세밀한 조정에따라 파편과 같은 클래스에 민감할 수 있음을 시사한다. 전체적인 구분 정확도 또한, Late Fusion과 Gated Fusion에서의 값이 가장 높은 값을 기록했다. RGB만 단독으로 썼을 때는 가장 낮은 정확도인 86.82%를 기록하면서 Depth 정보 없이 학습된 모델의 한계를 보여준다. Early Fusion은 전체적으로중간 정도의 성능을 보였는데, 간단한 구조 및 개념을 가졌지만 이정도면 효과적인 성능이라고 할 수 있다.
이와 같은 정량적 비교를 통해 다양한 융합 전략 간의 구조적 특성과 클래스별 민감도를 평가할수 있었다.

Ⅲ. Conclusion

본 연구는 SPARK 데이터셋을 기반으로, 다양한 RGB-Depth 데이터 융합 방식이 우주 객체 분류 성능에 어떤 영향을미치는지 정량적으로 분석하였다. RGB 단독 학습, Early Fusion, Late Fusion, Cross Attention, Gated Fusion의 총 5가지 융합 전략을 비교하였다.
실험 결과, 모든 융합 방식이 RGB 단독 학습보다 우수한 성능을 보였으며, 특히 Late Fusion과 Gated Fusion이 대부분의 지표에서 우수한 성능을 기록하였다. 이러한 결과는 RGB-Depth 융합 방식에 따라 성능이 상당한 차이를 보일 수 있음을 정량적으로 보여주고, 파편과 같은 위험 객체 분류의 성능 향상을 위해 어떤 융합 전략을 고르는지가 중요하다고 할 수있다.
향후 연구에서는 다양한백본 및 학습 전략을 적용하여 본 연구의 결과가 일반화 가능한지 검증하는 일이 중요할 것으로 보인다.

ACKNOWLEDGMENT

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학 ICT 연구센터사업의 연구결과로 수행되었음. (IITP-2025-RS-2023-00259061)
교신저자: 이은규(eklee@inu.ac.kr)

참 고 문 헌

[1] M. A. Musallam, K. Al Ismaeil, O.Oyedotun, M. D. Perez, M. Poucet, and D. Aouada, “SPARK: Spacecraft recognitionleveraging knowledge of space environment,” *Proc.IEEE Int. Conf. on Image Processing (ICIP)*, 2021.

[2] Tanguy Ophoff, Kristof Van Beeck,and Toon Goedem, "Exploring RGB+Depth fusion for real-time objectdetection," *Sensors*, vol. 19,no. 4, pp. 866, Feb. 2019.

[3] G. Tziafas and H. Kasaei,"Early or late fusion matters: Efficient RGB-D fusion in visiontransformers for 3D object recognition," *arXiv:2210.00843v2*, Mar. 2023.

[4] B. Xuyang, H. Zeyu, Z. Xinge, H.Qingqiu, C. Yilun, F. Hongbo, and T. Chiew-Lan, "TransFusion: RobustLiDAR-Camera Fusion for 3D Object Detection with Transformers," *arXiv:2203.11496v1*, Mar.

[5] O. K. Oyedotun, D. Aouada, and B.Ottersten, “Learning to fuse latent representations for multimodal data,” *Proc. IEEE Int. Conf. on Acoustics, Speechand Signal Processing (ICASSP)*, 2019.