다중 위성 영상 데이터를 활용한 연합 학습 기반의 객체 탐지 연구

김성철, 김성용, 조규태
LIG 넥스원

sungcheol.kim@lignex1.com, sungyong.kim@lignex1.com, kyutae.cho2@lignex1.com

Federated Learning for Fine-Grained Vehicle Detection Across Multi-Platform Satellite Imagery

Kim Sung Cheol, Kim Sung Yong, Cho Kyu Tae
LIG NEX1

요 약

위성 영상 기반 객체 탐지 기술은 국방 및 다양한 산업 분야에서 핵심적인 역할을 한다. 그러나 플랫폼별 데이터 수집의 제약과 데이터 분산 문제로 인해 성능 향상에 어려움이 있다. 본 연구는 항공, 지상, 해상 플랫폼에서 수집된 데이터를 통합한 새로운 다중 플랫폼 데이터 셋을 제안하고, 폐쇄망 환경에서도 적용 가능한 연합 학습 기반 객체 탐지 프레임워크를 설계하였다. 실험 결과, 폐쇄망 환경에서의 학습 안정성을 확인할 수 있었고 향후 위성 체계 분야에 도움이 될 것이라 기대한다.

Ⅰ. 서 론

최근 몇 년간 위성 체계 및 무인 시스템을 활용한 다양한 산업 분야에서 객체 탐지 기술 연구가 활발히 진행되고 있다. 객체 탐지는 이미지 혹은 비디오 내에서 특정 물체를 탐지하고 분류하는 기술로, 경계 감시, 표적 식별, 전장 상황 인식 등 다양한 군사 작전에서 핵심적인 역할을 한다. 대표적인 딥러닝 기반 탐지 모델로 YOLO 가 있다.

위성 영상 분석은 항공, 지상, 해상 등 다양한 플랫폼에서 환경에서 적용되고 있다. 그러나 현실적으로 각 플랫폼에서 수집되는 데이터의 수는 제한적이다. 또한 위성 영상 데이터는 여러 기관 및 국가에 분산되어 저장되는 경우가 많아 데이터 접근성과 통합에 어려움이 존재한다. 이에 따라 연합 방식 기반의 연구가 제안되었다. 연합 학습이란, 원본 데이터를 직접 공유하지 않고 모델을 학습하는 방식으로 데이터 프라이버시를 보호할 수 있는 방법이다.

본 연구는 항공, 지상, 해상 플랫폼으로부터 수집된 다양한 센서 데이터를 활용하여 새로운 다중 플랫폼 데이터 셋을 제안한다. 또한 폐쇄망 환경에서도 적용 가능한 연합 학습 기반 객체 탐지 프레임워크를 설계하였다. 제안한 프레임워크는 CNN 및 Transformer 기반 탐지 모델을 포함하여 비교 실험을 수행하였다. 실험 결과, 제안된 연합 학습 모델이 개별 로컬 학습 모델에 비해 효과적임을 확인하였다.

Ⅱ. 본론

2.1 데이터 구성

본 연구는 기존의 공개 데이터 셋(Mar20 [1], ShipRSImageNet [2], DroneVehicle [3])을 통합하여 새로운 다중 플랫폼 객체 탐지 데이터 셋인 MetaFleet 을 구축하였다. 이때, 연합 학습 시나리오를 고려하여, 클래스 별 인스턴스 수가 700 개 미만인 이미지는 제외하여 데이터를 재구성하였다. MetaFleet 은 항공(aircraft), 해상(ship), 지상(ground) 세 가지 플랫폼을 포함하여, 총 20 개의 세분화된(fine-grained) 클래스로 구성되어 있다. 본 연구에서 제안한 데이터 셋은 Table 1.과 같다.

Datasets	Platform	Images (Train / Val)	Instances (Train / Val)
Mar20 [1]	Aircraft	2,689/ 769	15,577 / 4,527
ShipRSImageNet [2]	Ship	2,200/ 545	10,860 / 3,103
DroneVehicle [3]	Ground	17,948 / 1,469	297,716 / 22,462
MetaFleet	Aircraft, Ship, Ground	22,835 / 2,788	302,898 / 28,478

Table 1. 제안 데이터 셋 (MetaFleet)

제안 데이터 셋은 단일 플랫폼(항공, 해상, 지상) 데이터를 통합하여 학습 데이터의 도메인 다양성을 확보하였다. 또한 세분화된 클래스 데이터를 활용함으로써, 기존의 대분류(coarse-grained) 데이터 셋에 비해 더 정밀한 성능 분석이 가능하도록 구성하였다.

2.2 연합 학습 프레임워크

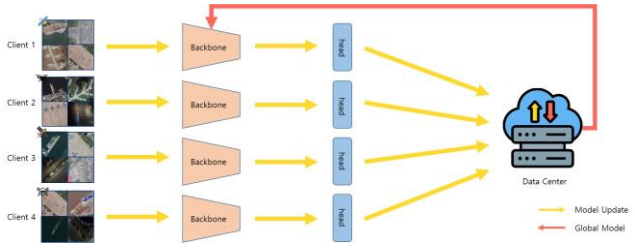

Figure 1. 연합 학습 프레임워크

Figure 1.은 폐쇄망 환경에서의 연합 학습 기반의 객체 탐지 프레임워크를 나타낸다. 제안한 데이터 셋을 네 명의 클라이언트에게 분배하였으며, 각 클라이언트는 항공, 지상, 해상 플랫폼에서 수집된 데이터를 활용하여 로컬 모델을 독립적으로 학습한다. 이때, 데이터는 비독립적이며 동일 분포를 따르지 않는 Non-IID 특성을 가진다. 본 연구는 CNN 기반의 YOLOv5, YOLOv8, YOLOv11 및 Transformer 기반의 D-fine 모델을 활용하여 학습을 수행하였다. 학습 후, 각 클라이언트는 로컬 모델의 업데이트를 데이터 센터(Data Center)로 전송하며, 데이터 센터에서는 FedAvg 알고리즘을 통해 글로벌 모델의 가중치를 업데이트한다. 이후 업데이트 된 글로벌 모델의 가중치는 각 클라이언트의 로컬 모델로 전달된다. 본 프레임워크는 폐쇄망 환경에서 설계되었으며, 모든 과정은 내부 네트워크에서 수행하였다.

Ⅲ. 실험

3.1 실험 세팅

본 연구에서는 연합 학습 환경을 구성하여 총 4 개의 클라이언트(client)를 설정하였으며, 각 클라이언트가 보유한 데이터는 Non-IID 분포를 따른다. 모든 클라이언트가 매 통신 라운드에 참여하도록 참여 비율을 1.0 으로 설정하였다. 각 클라이언트가 소유한 로컬 모델의 학습 에폭 수는 10, 20, 40 으로 설정하였다. 파라미터 집계 방식은 FedAvg(Federated Averaging) 알고리즘을 적용하였다.

객체 탐지 실험 세팅의 경우, YOLOv5, YOLOv8, YOLOv11, D-fine 아키텍처를 활용하였다. 이미지 크기는 1024x1024 로 고정하였으며, optimizer 는 Adamw 를 사용하였다. 학습률을 3e-4 로 설정하였고, 배치 크기는 8 로 고정하였다. 모델 성능 평가는 mAP50, mAP50-95 지표를 활용하였다.

3.2 실험 결과

Datasets	Client	Model	Centralized (%)	Federated (%)			
				Method	E10	E20	E40
MetaFleet	Ⅲ	YOLOv5n	81.7	FedAvg	72.6	76.2	76.9
		YOLOv8n	83.4		73.2	78.0	78.6
		YOLOv11n	83.8		73.5	78.1	79.0
		D-fine-n	85.9		83.5	83.9	84.7

Table 2. Local Epoch(E)에 따른 객체 탐지 성능 비교

Table 2.는 연합 학습 환경에서 로컬 에폭(Local Epoch, E)의 수에 따른 알고리즘 별 객체 탐지 성능 비교 결과를 나타낸다. 실험 결과, 모든 모델에서 로컬 에폭 수의 증가에 따라 연합 학습 성능이 점진적으로 향상되었으며, 특히 학습 초기 단계에서 나타나는 성능 불안정성이 완화하는 경향을 보였다. 이는 로컬 에폭 수의 증가가 클라이언트 내에서 반복적인 모델 업데이트를 수행함으로써, 각 통신 라운드 시 글로벌 모델이 보다 안정적이고 일관된 방향으로 수렴하도록 돕는다는 점을 시사한다.

Training	mAP50 / mAP50-95 (%)			
Model	YOLOv5n	YOLOv8n	YOLOv11n	D-fine-n
Centralized	81.7 / 61.6	83.4 / 63.7	83.8 / 64.1	85.9 / 67.1
Client 1	75.8 / 55.4	76.9 / 56.4	78.1 / 57.6	78.9 / 59.3
Client 2	75.7 / 55.2	76.8 / 56.7	77.5 / 57.1	78.6 / 59.5
Client 3	75.1 / 54.6	77.6 / 57.2	77.9 / 57.4	78.8 / 59.2
Client 4	74.9 / 54.8	76.8 / 56.6	76.8 / 56.8	77.7 / 58.6
Client Avg	75.4 / 55.0	77.1 / 56.8	77.6 / 57.3	78.5 / 59.2
FedAvg	76.9 / 56.1	78.6 / 58.2	79.0 / 68.8	84.7 / 65.3

Table 3. 로컬 클라이언트 및 연합 학습 기반의 객체 탐지 따른 객체 탐지 성능 비교

Table 3.은 로컬 클라이언트에서 개별적으로 학습한 모델(Local)과 클라이언트 간 파라미터만을 공유하는 연합 학습(FedAvg)를 비교 결과를 의미한다. 실험 결과, 연합 학습 방식으로 학습된 모델이 로컬 학습 대비 효과적임을 확인하였다. 특히 D-fine-n 의 경우, 로컬 모델과 6.2%의 성능 차를 보였다.

Ⅳ. 결론

본 연구에서는 항공, 해상, 지상 플랫폼에서 수집된 데이터를 통합하여 세분화된 데이터셋을 제안하고, 폐쇄망 환경에서의 연합 학습 기반 객체 탐지 프레임워크를 설계하였다. 실험 결과 데이터의 수가 적은 경우에는 로컬 에폭의 수를 길게 설정하여 로컬 모델의 가중치가 수렴하도록 하는 방식이 연합 학습에 유의미한 결과를 가져오는 것으로 확인하였다. 또한 연합 학습 방식으로 학습된 모델이 로컬 학습 대비 효과적임을 실험적으로 확인하였다. 향후 폐쇄망 환경에서도 이러한 접근 방식이 위성 영상 기반 객체 탐지 기술 발전에 기여할 수 있을 것으로 기대한다.

참 고 문 헌

[1] WENQI, Y. U., et al. MAR20: A benchmark for military aircraft recognition in remote sensing images. National Remote Sensing Bulletin, 2024, 27.12: 2688-2696.

[2] ZHANG, Zhengning, et al. ShipRSImageNet: A large-scale fine-grained dataset for ship detection in high-resolution optical remote sensing images. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2021, 14: 8458-8472.

[3] SUN, Yiming, et al. Drone-based RGB-infrared cross-modality vehicle detection via uncertainty-aware learning. IEEE Transactions on Circuits and Systems for Video Technology, 2022, 32.10: 6700-6713.