IDAP INSTITUTE OF DEFENSE ACQUISITION PROGRAM

JAMS JOURNAL OF ADVANCES IN MILITARY STUDIES

2025, Vol. 8, No. 1, 1-13.

https://doi.org/10.37944/jams.v8i1.277

Comparative analysis on few-shot models performance for improving object detection in the military Domain

Kim, Junsub* · Choi, Dongnyeok**

ABSTRACT

The application of Object Detection (OD) techniques in the military and defense domain is often restricted by stringent security requirements and limited data availability. To overcome these challenges, the present study investigates the potential of Few-Shot Object Detection (FSOD) for military applications. A military vehicle image dataset, composed of real-world defense imagery, was constructed for this purpose. Four representative object detection models—YOLO, DETR, GLIP, and CD-ViTO—were fine-tuned under 1-shot, 5-shot, and 10-shot conditions. The model performance was evaluated using mean Average Precision (mAP). Notably, the CD-ViTO model's cross-domain generalization capability was further examined by comparing its performance on this military dataset against public benchmarks previously used in FSOD studies. Experimental results demonstrate that CD-ViTO achieved superior mAP scores, highlighting the viability of FSOD for efficient and accurate object detection in military and defense applications.

Keywords : object detection, military domain, images of military vehicles, few-shot learning, model performance evaluation

* (First author) Funzin, Associate Researcher, junsub.kim@funzin.co.kr, https://orcid.org/0009-0008-3223-4529.

** (Corresponding author) Funzin, Principal Researcher, luke.dn.choi@funzin.co.kr, https://orcid.org/0000-0006-3383-1179.

Ⅰ. 서론

인공지능 기반 객체 탐지(Object Detection, OD)는 이미지 내에서 특정 객체의 종류와 위치를 자동으로 탐지하는 기술로 최근 다양한 산업 분야에서 빠르게 확산하면서 국방 분야도 여러 응용(전장 상황 인식, 정찰 자동화, 표적 인식 등) 영역에서 주목받고 있다. 다만, 국방 분야 연구는 분야의 특성상 민감 정보의 보안성이나 제한된 접근성으로 실제 대규모 주석 데이터셋을 구축하기 어려워 데이터 수집과 활용의 제약 문제가 있고, 기존 대용량 학습 기반 객체 탐지 모델을 효과적으로 적용하기 어려운 상황이다. 예를 들어 CNN 기반이나 트랜스포머 기반의 대표적 객체 탐지 모델은 수십만 장 이상의 학습 데이터가 필요하여 군사 도메인에 현실적으로 적용하는 것이 쉽지 않다. 따라서 객체 탐지・분류 정확도 향상을 위한 계량적 연구 접근이 부족하여 실시간 탐지 및 분석 시스템의 최적화를 위한 연구 설계가 필요하다(Ryu, Park, & Kim, 2024).

한편, 객체 탐지 기술의 확장으로 Open-Vocabulary Object Detection(OVOD) 개념이 등장하였다. 이는 사전 정의된 클래스 집합에 국한되지 않고, 텍스트 설명을 기반으로 새로운 객체 탐지 방식이다. 대표적인 OVOD 모델인 GLIP(Li et al., 2022) 및 Grounding DINO(Liu et al., 2024)는 대규모 텍스트-이미지 사전학습을 통해 제로샷(zero shot) 탐지 성능을 확보하며, 자연어 기반 객체 인식을 실현하였다. 특히 OVOD는 탐지 클래스의 제약을 줄인다는 점에서 국방 분야에 잠재적으로 활용 가능성이 있으나, 여전히 군사 도메인에서의 유효성 검증은 미비하다.

상기한 제약 상황을 해소하는 대안으로 소량 데이터만을 학습하여 새로운 객체를 탐지하는 FSOD(Few-shot Object Detection) 기법을 고려할 수 있다. 최근 다양한 비전 트랜스포머(Vision Transformer: ViT) 기반의 구조, 언어-시각 결합 방식이 등장하고 있으나, 군사(국방 및 방위산업) 분야에서 해당 기술의 유효성을 검증하는 실증 연구가 부족한 실정이다. 이 도메인의 Few-Shot 학습 연구는 대부분 이미지 분류(Classification)에 중점되어 있으며(Kang, 2022; Yuk, Oh, Jeong, 2024), 이미지 내 객체의 위치까지 탐지하는 OD 문제에 Few-Shot을 적용한 학술적 연구가 부족하다. 이미지 분류는 이미지 전체에 하나의 클래스 레이블을 예측하지만, 객체 탐지는 여러 객체의 존재 여부와 각 객체의 위치를 함께 추론하는 복잡한 구조적 특성이 있다.

따라서 본 연구는 국방 환경에 특화된 소량의 군사 데이터셋을 활용하여 FSOD 기반 객체 탐지 모델의 성능을 비교하고, 그 적용 가능성을 실험적으로 검토하는 것을 목표로 한다. 이를 위해, 다양한 구조적 특성을 갖는 대표적인 객체 탐지 모델을 소량의 학습 데이터(1-shot, 5-shot, 10-shot) 조건에서 학습시켜 그 성능을 비교하였다. 특히, 일부 모델의 경우에 기존 다른 도메인에서 우수 성능을 보였던 학습 방식을 군사 데이터에 적용하여 실제 군사 환경에서도 유사한 효과를 기대할 수 있을지를 함께 분석하였다. 이를 통해 국방 분야에서 소수의 데이터만으로도 객체 탐지 가능 여부를 실증적으로 평가하고, FSOD 기술의 실질적인 적용 가능성을 탐색하고자 한다.

Ⅱ. 관련 연구

2.1 Object Detection(OD) 및 Open-Vocabulary Object Detection(OVOD)

객체 탐지(Object Detection, OD)는 이미지 내 객체 종류와 위치 탐지 기술로 CNN 기반의 YOLO 모델, Transformer 기반의 DETR(Detection Transformer) 모델이 대표적이다. YOLO 모델은 기존 객체 탐지(R-CNN 계열) 방식의 제약을 해소하기 위한 개선 모델로 전체 이미지를 한 번에 탐지하고, 단일 신경망으로 모든 객체의 위치와 종류를 실시간 처리하여 동시에 예측할 수 있다(Remon et al., 2016). 그래서 해당 모델은 객체 탐지의 속도와 정확도 향상이 가능하며 YOLOv5, YOLOv8 등으로 발전하면서 실시간 비전 분야에서 활용되고 있다(e.g., Alawi & Mohammed, 2024; Ryu, Park, & Kim, 2024). 예를 들어 Baek et al.(2024) 연구는 실제 유도무기의 적외선 영상 내 객체 탐지의 신속성과 동적 영상에서 일관되고 정확한 검출을 위한 이미지 전처리 목적으로 YOLO 모델을 활용하였다.

반면, DETR 모델(Carion et al., 2020)은 Transformer 아키텍처를 활용해 기존 객체 탐지 모델의 복잡한 수작업 구성 요소를 제거하고 end-to-end 단일 네트워크로 객체 간 관계를 모델링하여 복잡한 장면에서도 뛰어난 성능의 장점이 있다. 이런 장점에도 불구하고, 해당 모델은 사전 정의된 고정 클래스에 국한된 탐지만 가능하고, 소형 객체에 대한 탐지 성능이 저조하며 실시간 탐지에 제약이 존재한다. 이에 Alawi & Mohammed(2024) 연구는 야전에서 주변 환경과 유사 색상・패턴을 활용한 군용 위장 표적 탐지의 성능 향상을 위해 기존 DETR 모델 개선을 시도하였다. 이들은 End-to-End Transformer 기반 구조로 다양한 이미지의 해상도 특징을 융합하고, 다중 Transformer 모델 구조를 제안하였다.

개방 어휘 객체 탐지(Open-Vocabulary Object Detection, OVOD)는 사전 정의된 객체 종류 외에 자연어 설명을 기반으로 이미지 내 객체를 탐지하는 기술이다. Li et al.(2022)이 제안한 GLIP(Grounded Language-Image Pretraining)은 텍스트-이미지 간 의미적 매칭으로 대규모 사전학습을 통해 zero-shot 기반 탐지가 가능하다. 최근 발표된 Grounding DINO 모델은 DETR 기반 구조에 텍스트 조건부 객체 탐지를 통합하고, 다양한 데이터셋 대상으로 사전학습이 되어 자연어 쿼리에 따른 객체 위치의 정밀 예측이 가능한 구조를 갖추고 있으나(Liu et al., 2023) 사전학습에 포함되지 않은 도메인이나 객체에 관한 탐지 성능이 제한된다는 한계도 있다.

2.2 Few-Shot Object Detection(FSOD) 및 Few-Shot Learning의 군사 분야 (국방・방위산업) 응용

FSOD는 객체 종류별 소량의 데이터(학습 자료)만을 활용하여 새로운 객체 카테고리를 탐지한다. Meta R-CNN(Wang et al., 2020)은 RoI(Region of Interest)[1] 수준의 메타러닝 구조를 도입하여 소수

샘플로 효과적인 객체 탐지가 가능하며, Frustratingly Simple FSOD(Wang, Huang, Darrell, Gonzalez, & Yu, 2020)은 단순 fine-tuning만으로 기존 메타러닝 기법을 능가하는 성능을 보였다. 또한, FS-DETR (Bulat, Guerrero, Martinez, & Tzimiropoulos, 2023)은 DETR 구조에 시각적 프롬프트를 적용하여 테스트 시 추가 학습 없이도 새로운 클래스 탐지가 가능하였다. 최근 Vision Transformer 기반의 개방형 탐지 구조를 확장한 CD-ViTO(Fu et al., 2024) 모델은 CD-FSOD(Cross-domain few-shot object detection) 모델보다 향상 성능을 나타냈다.

상기한 객체 탐지 모델은 학습 가능한 소량 객체 데이터로 객체의 고유 특징을 포착할 수 있어 군사 분야에서 객체의 분류 정확성을 높이는 데 적합하다(e.g., Park, Lee, Choi, Kim, Jeong, & Paik, 2024). 실제 군사 분야는 군사적 비밀이나 기밀성으로 대규모 데이터가 확보가 현실적으로 어려운 실정이다(Baek et al., 2024). 이런 실증연구 수행의 제약을 해소하기 위해 Kang(2022)은 전차 전투원의 생존력 및 전투수행능력 향상에 필요한 미식별 전차를 학습하지 않은 데이터 분류가 가능한 메타러닝 기반 학습모델을 제안하였다. Few-Shot Learning 방식을 적용한 결과, 학습에 사용되지 않은 총 103종의 다른 전차 데이터를 탐지하여 58% 수준의 분류 정확성의 성능을 보였다. 더 나아가 Yuk et al.(2024) 경우, 실제 전차와 전투기 이미지 데이터(오픈소스 데이터셋 및 인터넷 이미지 수집)를 사용하여 Prompt Learning에 기반한 Few-Shot Learning 방식으로 분류 정확성을 높이고자 연구를 수행하였다.

Ⅲ. 연구방법 및 분석절차

3.1 연구 대상 및 데이터셋 구성

선행연구는 주로 이미지 분류의 목적에 초점을 두어 객체 탐지에서 위치 추정의 복잡성이 높은 객체 탐지 환경에서 정확성을 확보하는 실험 연구가 부족하다. 그래서 본 연구는 군사 분야의 소량 객체 이미지 대상으로 탐지 성능을 확보하기 위해 Few-Shot 학습 기반 객체 탐지 기법을 적용하고자 한다.

본 연구는 제약적인 연구 환경을 고려하여 객체 탐지용으로 공개된 군사 데이터셋 Roboflow의 Russian Military Vehicles 데이터[2]를 이용한다. 해당 데이터셋은 총 10개 클래스의 군용 차량 이미지로 탱크, 군용 트럭, 장갑차 등 다양한 군사 장비 이미지를 포함하며(총 993장), 어노테이션 형식은 COCO format(JSON)을 제공한다. 즉, 데이터셋은 Roboflow 플랫폼에서 공개된 원천 이미지를 기반

1) 이미지 내에 객체가 존재할 가능성이 높은 영역을 의미하며, 객체 탐지 시 이 영역에 집중하여 객체의 클래스 분류와 정밀화 수준을 향상시킴.

2) 해당 데이터넷은 전체 클래스에서 약 120-150개 샘플이 존재하며, 본 연구는 클래스당 수 장 내외의 이미지로 학습이 이루어지는 Few-shot 학습 환경을 전제하므로 사용 데이터셋의 전체 분포 차이가 실험 결과에 직접적인 영향이 크지 않다고 판단하여 분석에 활용함.

으로 제공된 주석 정보를 기준으로 구축한다. 또한 데이터 수가 제한된 Few-Shot 학습 특성을 반영하고, 학습 적합성을 평가하고자 모델 성능향상의 이미지 증강은 별도로 적용하지 않는다. 상기한 절차를 통해 본 연구의 전체 데이터는 학습(15.0%, 149장), 검증(42.5%, 422장), 테스트(42.5%, 422장)로 나눈다. 연구에 사용된 전체 데이터셋의 형태와 각 실험 조건은 Table 1과 같이 구성한다.

해당 실험의 Few-Shot 학습 환경은 기존 연구(Bulat et al., 2023; Li et al., 2022)를 참고하여 클래스 별로 각각 1-shot, 5-shot, 10-shot 조건에 맞추어 학습 데이터를 구성한다. 이 데이터셋은 각 클래스별로 지정된 샷 수에 해당하는 어노테이션(annotation)만을 포함하고, 나머지 전체 데이터는 1:1 비율로 검증과 테스트 데이터를 무작위 분할한다. 그리고 최종 성능 평가는 테스트 데이터 상에서 수행하며, 학습과 평가 간 데이터 중복이 없도록 설계한다.

〈Table 1〉 Summary table of experimental data sets

Item	Description
Total # of images	993
# of classes	10
Annotation format	COCO format(json)
Data split ratio	train=149(15.0%), val=422(42.5%), test=422(42.5%)
Experimental setting	Number of images used for 10(1-shot), 44(5-shot), and 83(10-shot)

본 연구의 실험용 학습 데이터셋의 예시를 살펴보면(Figure 1), 각 이미지는 사전 정의된 객체 클래스에 해당하는 바운딩 박스(Bounding Box)와 함께 표시되며, 이 박스는 객체 위치와 크기를 정확하게 나타낸다. 이를 통해 하나의 객체만을 포함하는 이미지와 다중 객체를 포함하는 이미지로 구성된다.

〈Figure 1〉 Example images of training dataset

3.2 모델 구성 및 선정

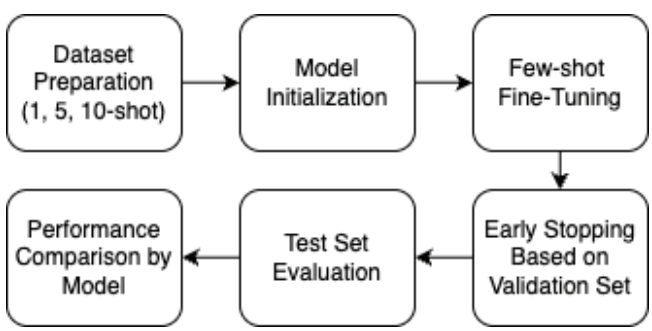
본 연구는 서로 다른 구조적 특성의 네 가지 모델을 선정 후에 비교 실험을 수행하였다. 선정한 모델의 종류와 각 모델을 선정한 이유는 다음 Table 2와 같다.

〈Table 2〉 Comparison of different object detection models

Model	Rationale
YOLOv8m (Jocher et al., 2023)	· A representative CNN-based object detection model · Offers a good balance between speed and accuracy
DETR (Carion et al., 2020)	· Transformer-based object detection model · Effectively models complex object relationships, making it suitable as a structurally different comparison baseline
GLIP (Li et al., 2022)	· An OVOD model based on text-image matching · Supports both zero-shot inference and few-shot fine-tuning
CD-ViTO (Fu et al., 2024)	· A ViT-based model specifically designed for FSOD · A recent model achieving high detection performance with limited data

3.3 학습 및 평가 절차

본 연구는 선정한 모든 모델(Table 2)을 동일한 실험 환경과 일관된 학습 전략을 적용하여 학습을 수행한다. 모든 모델은 각 모델별로 공개된 사전학습 가중치를 기반으로 초기화하고, Few-Shot 학습의 특성상 과적합(overfitting)을 방지하기 위해 조기 종료(Early Stopping)를 공통으로 적용하였다(e.g., Wu, Cong, Huang, Ju, Jiang, & Chen, 2025). 구체적으로는 검증 데이터의 성능 개선이 10회 연속으로 이루어지지 않으면 학습 진행이 종료되도록 설계한다.



〈Figure 2〉 Workflow for model training and evaluating in object detection

모델 성능은 객체 탐지 모델의 성능을 종합적으로 평가하는 대표적 지표인 mAP(mean Average

Precision)로 평가한다. mAP는 클래스별로 계산된 AP(Average Precision)의 평균값으로 정밀도(Precision)–재현율(Recall) 곡선을 기반으로 산출한다.[3] 본 연구는 다양한 IoU(Intersection over Union) 임계값을 측정한 AP 평균값인 mAP@[0.5:0.95]의 엄격한 성능 평가지표를 활용한다. IoU는 모델이 예측한 바운딩 박스와 실제 객체 바운딩 박스 간에 겹치는 영역을 비율로 나타낸 수치이며, 그 값이 클수록 위치 예측의 정확도가 높아진다고 판단할 수 있다. mAP@[0.5:0.95]는 IoU 임계값을 0.5부터 0.95까지 0.05 간격으로 조정하며 계산된 총 10개 AP의 평균값이다. 이 방식은 단일 임계값만을 사용하는 경우(mAP@0.5)보다 다양한 난이도의 정확도 요구 조건에서 모델성능을 정밀하게 평가할 수 있다.

Ⅳ. 분석결과

4.1 모델별 Few-shot 학습 성능의 비교

<Table 3>은 1-shot, 5-shot, 10-shot 조건에서 모델별 mAP의 비교표이다. 실험 결과, 1-shot 조건은 GLIP 모델을 fine-tuning 했을 때, 가장 높은 mAP가 나타났다. 이는 사전학습을 통해 습득한 텍스트-이미지 정렬 능력이 소량의 데이터에서도 빠른 적응과 일반화 학습이 효과적으로 작용하였다는 것을 알 수 있다. 그러나 GLIP Zero-shot 조건에서는 객체 탐지 성능이 0에 가까워 실질적으로 탐지가 이루어지지 않았으며, 이는 Open-Vocabulary Object Detection(OVOD) 모델이 Fine-tuning 없이 새로운 객체 탐지에 한계가 있음을 의미한다.

〈Table 3〉 Comparison of mean average precision (mAP) in Different Models with few-shot training conditions

Evaluation metric : mAP[0.5:0.95]

Model \ # of Shots	1-Shot	5-Shot	10-Shot
YOLOv8m	2.4	14.5	15.4
DETR	0.0	0.5	0.7
GLIP(Zero-Shot)	0.0		
GLIP(Few-shot Fine-tuning)	**9.51**	31.12	37.36
CD-ViTO	8.96	**32.46**	**45.20**

CD-ViTO 모델에서 5-shot(32.46), 10-shot(45.20) 조건을 보면 mAP가 가장 우수한 성능이 나타났

3) 정밀도는 탐지한 객체 중 정답인 비율, 재현율은 실제 객체 중 모델이 올바르게 탐지한 비율을 의미함.

다. 이런 비교분석의 결과, 해당 모델이 소량 샘플에서도 클래스 간의 차이를 효과적으로 학습한 것으로 FSOD에 특화된 구조적 설계의 장점을 입증한 것이다. 한편, 대용량 학습 데이터에 최적화된 모델(YOLOv8m, DETR)은 소량의 학습 예시만 주어지는 Few-shot 환경에서 탐지 성능이 크게 저하되는 경향을 발견하였다. DETR 모델의 분석결과를 살펴보면, 해당 모델은 객체 간 관계를 Transformer 구조에 기반하여 학습하므로 충분한 학습 데이터 확보가 어려운 분석 환경에서 안정적인 탐지 성능 확보에 한계가 있다.

4.2 CD-ViTO 모델의 데이터셋 별 FSOD 성능 비교

CD-ViTO는 다양한 도메인 간 전이 학습 상황을 고려한 Cross Domain Few-shot Object Detection (CD-FSOD) 문제해결을 위한 제안 모델로 다양한 특수 목적 데이터셋에 대해 탐지 성능을 평가한다. 본 실험은 국방 도메인의 Russian Military Vehicles 데이터셋을 포함하여 CD-ViTO 모델의 일반화 성능을 비교분석하였다.

해당 CD-ViTO 모델을 mAP[0.5:0.95] 기준으로 다양한 데이터셋에서 성능비교 결과(Table 4), 국방 도메인에서도 Clipart1k와 유사한 성능을 보였으며, NEU-DET, UODD와 같은 다른 특수 도메인보다 훨씬 우수한 성능을 보였다. 특히, 10-shot 조건에서 Clipart1k(44.3)과 비교하여 상대적으로 높은 성능을 기록하였다(45.2). 이는 국방 데이터의 구조가 일반적인 자연 이미지와 달리 특이성이 존재하여 CD-ViTO가 효과적으로 적용되어 실증적으로 성능 검증한 결과이다.

〈Table 4〉 FSOD performance of the CD-ViTO model with different datasets

Evaluation metric : mAP[0.5:0.95]

# of shots / Dataset	1-Shot	5-Shot	10-Shot
ArTaxOr	21.0	47.9	60.5
Clipart1k	17.7	41.1	44.3
DIOR	17.8	26.9	30.8
DeepFish	20.3	22.3	22.3
NEU-DET	3.6	11.4	12.8
UODD	3.1	6.8	7.0
Russian Military Vehicles	**9.0**	**32.5**	**45.2**

Ⅴ. 결론 및 논의

5.1 실험 결과의 요약 분석

본 연구는 Few-shot 학습 환경(1-shot, 5-shot, 10-shot)에서 다양한 객체 탐지 모델의 성능 차이를 분석하였다. 분석결과, CD-ViTO는 5-shot 및 10-shot 조건에서 가장 높은 모델 성능(mAP)을 나타냈으며, 이는 소수의 학습 예시만으로도 객체 특성을 파악하는 학습 능력이 있다고 볼 수 있다. GLIP의 경우, 사전학습(pre-training)을 통해 이미지와 텍스트 사이의 의미적 연관성을 학습한 구조로 fine-tuning만으로도 빠르게 성능 향상이 가능하다. 실제 1-shot 조건에서 모든 모델 중 가장 높은 mAP를 기록하였으며, 이는 GLIP의 텍스트 기반 질의 처리 능력이 소량 데이터 상황에서 효과적으로 작용했음을 보여준다. 반면에 zero-shot 조건은 사전학습에 포함되지 않은 객체탐지가 이루어지지 않아 성능이 0에 가까웠으며, Fine-tuning 없이 OVOD 모델이 새로운 객체의 실질적 감지가 제한된다는 것을 의미한다.

본 연구는 국방 도메인에서 Few-Shot Object Detection(FSOD)의 적용 가능성을 실증적으로 검토하였다는 점에서 학술적 의미가 있다. 특히, 데이터 수집이 어려운 군사 환경에서도 FSOD 접근이 현실적 대안이 될 수 있음을 다양한 모델 기반 실험을 통해 확인하였다. 기존 FSOD 연구는 주로 일반 도메인이나 비군사 이미지 탐지에 초점을 두었으나 본 연구는 데이터 확보 제약이 높은 특수한 도메인(군사 분야)의 영상을 활용하여 모델 성능을 정량적으로 검증했다는 점에서 연구의 의의가 높다. 이런 연구 결과를 토대로 향후 해당 분야의 연구에서 데이터 수급이 제한된 국방 환경에 FSOD 모델 활용의 가능성과 분석방법을 제시하였다는 점에서 실무적인 시사점이 있다. 실제 본 연구의 실험에서 CD-ViTO 모델은 소량의 학습 데이터만으로도 국방 데이터에 대해 안정적인 탐지 성능을 보여주었으며, 기존 Cross Domain-Few Shot Object Detection(CD-FSOD) 실험에서 사용된 데이터셋과 유사한 수준의 결과를 기록하였다.

이런 학술 · 실무적 시사점에도 불구하고 다음과 같은 연구의 한계점이 존재한다. 본 연구는 국방 도메인의 실제 데이터셋(Russian Military Vehicles) 활용하였다는 점에서 실제 전장 환경을 가정한 연구의 실재성(fidelity)을 높이고자 노력하였으나 해당 데이터셋의 특성을 보면 군용 차량 객체에 한정된 이미지로 연구결과의 일반화가 제한된다. 그래서 후속 연구는 다양한 무기체계를 대상으로 객체 유형 탐지의 모델성능을 검증하는 분석이 필요하다. 더 나아가 향후 실험설계는 객체 탐지와 관련된 실제 운용환경의 특징(영상 촬영 환경, 해상도, 배경 잡음 등)을 고려할 수 있도록 변수를 사전식별하고 반영해야 한다(e.g., Ryu, Park, & Kim, 2024).

끝으로 군사 분야의 연구는 실전 상황에서 발생할 수 있는 변수를 모델에 충분히 반영하는 설계를 시도하고 있으나, 실시간성이 요구되는 응용 환경(전장의 감시 · 정찰 현장)은 추가적인 fine-tuning이 현실적으로 어려운 실정으로 향후 군사 분야 연구(국방 및 방위산업)는 재학습 없이

실시간 활용이 가능한 FSOD 모델 설계를 지속적으로 개선할 필요가 있다.

Acknowledgements

This work was conducted with the support of Funzin Co., Ltd.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Author contributions

Conceptualization: KJ and CD, Literature review: KJ and CD, Resources and Data curation: KJ, Investigation and Methodology: KJ and CD, Writing (Original Draft): KJ, Project administration and Supervision: CD.

Reference

Alawi, A. E. B., & Mohammed, H. M. (2024, August). *The Role of YOLOv8 in Enhancing Strategic Military Equipment Detection*. In 2024 4th International Conference on Emerging Smart Technologies and Applications (eSmarTA) (pp. 1-5). IEEE. https://doi.org/10.1109/eSmarTA62850.2024.10638856

Bae, J. Y., Park, D. H., Shin, H. J., Yoo, Y. S., Kim, D. W., Hur, D. H., Bae, S. H., Cheon, J. H., & Bae, S. H. (2024). Research on Local and Global Infrared Image Pre-Processing Methods for Deep Learning Based Guided Weapon Target Detection. *Journal of The Korea Society of Computer and Information, 29*(7), 41-51. https://doi.org/10.9708/jksci.2024.29.07.041

Baek, J. Y., Park, D. H., Shin, H. J., Yoo, Y. S., Kim, D. W., Hur, D. H., Bae, S. H., Cheon, J. H., & Bae, S. H. (2024). Research on Local and Global Infrared Image Pre-Processing Methods for Deep Learning Based Guided Weapon Target Detection. *Journal of The Korea Society of Computer and Information, 29*(7), 41-51. https://doi.org/10.9708/jksci.2024.29.07.041

Bulat, A., Guerrero, R., Martinez, B., & Tzimiropoulos, G. (2023). *Fs-detr: Few-shot detection transformer with prompting and without re-training*. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 11793-11802). https://openaccess.thecvf.com/content/ICCV2023/html/Bulat_FS-DETR_Few-Shot_DEtection_TRansformer_with_Prompting_and_without_Re-Training_ICCV_2023_paper.html

Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., & Zagoruyko, S. (2020, August). *End-to-end object detection with transformers*. In European conference on computer vision (pp. 213-229). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-58452-8_13

Fu, Y., Wang, Y., Pan, Y., Huai, L., Qiu, X., Shangguan, Z., ... & Jiang, X. (2024, September). *Cross-domain few-shot object detection via enhanced open-set object detector*. In European Conference on Computer Vision (pp. 247-264). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-73636-0_15

Jocher, G., Chaurasia, A., & Qiu, J. (2023). YOLO by Ultralytics. Retrieved from https://github.com/ultralytics/ultralytics

Kang, S. H. (2022). *Research of Unidentified Tank Classification Using Few-Shot Learning*. Summer conference of Korean Institute of Communications and Information Sciences. 392-393. https://www.dbpia.co.kr/journal/articleDetail?dbid=edspia&text=Full+Text+%28DBPIA%29&nodeId=NODE11107752&an=edspia.NODE11107752

Li, L. H., Zhang, P., Zhang, H., Yang, J., Li, C., Zhong, Y., ... & Gao, J. (2022). *Grounded language-image pre-training*. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 10965-10975). https://openaccess.thecvf.com/content/CVPR2022/html/Li_Grounded_Language-Image_Pre-Training_CVPR_2022_paper.html?ref=blog.roboflow.com

Liu, S., Zeng, Z., Ren, T., Li, F., Zhang, H., Yang, J., ... & Zhang, L. (2024, September). *Grounding dino: Marrying dino with grounded pre-training for open-set object detection*. In European Conference on Computer Vision (pp. 38-55). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-72970-6_3

Madan, A., Peri, N., Kong, S., & Ramanan, D. (2024). Revisiting few-shot object detection with vision-language models. *Advances in Neural Information Processing Systems, 37*, 19547-19560. https://doi.org/10.48550/arXiv.2312.14494

Park, C., Lee, S., Choi, H., Kim, D., Jeong, Y., & Paik, J. (2024, January). *Enhancing defense surveillance: Few-shot object detection with synthetically generated military data*. In 2024 International Conference on Electronics, Information, and Communication (ICEIC) (pp. 1-2). IEEE. https://doi.org/10.1109/ICEIC61013.2024.10457124

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). *You only look once: Unified, real-time object detection*. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 779-788). https://www.cv-foundation.org/openaccess/content_cvpr_2016/html/Redmon_You_Only_Look_CVPR_2016_paper.html

Ryu, H., Park, M., & Kim, D. Y. (2024). Object prediction and detection of ground-based weapon with an improved YOLO11 approach: Focusing on assumptions underlying operational environments and UAV-captured features related to PLZ-05 Self-Propelled Howitzer. *Journal of Advances in Military Studies, 7*(3), 13-30. https://doi.org/10.37944/jams.v7i3.256

Wang, X., Huang, T. E., Darrell, T., Gonzalez, J. E., & Yu, F. (2020). *Frustratingly simple few-shot object detection*. arXiv preprint arXiv:2003.06957.

Wu, G., Cong, L., Huang, C., Ju, Y., Jiang, J., & Chen, C. (2025, January). *Meta-Learning Framework for Effective Few Shot Time Series Prediction*. In 2025 IEEE 5th International Conference on Power, Electronics and Computer Applications (ICPECA) (pp. 18-22). IEEE. https://doi.org/10.1109/ICPECA63937.2025.10928851

Yuk, T. K., Oh, S. H., Hwang, S. I., & Jeong, K (2024). Effective Few-Shot Learning for Military Vehicles Image Classification Using Prompt-Based Learning. *Journal of Convergence Security, 24*(5), 189-194. https://doi.org/10.33778/kcsa.2024.24.5.189

원 고 접 수 일 2025년 03월 28일
원 고 수 정 일 2025년 04월 16일
게 재 확 정 일 2025년 05월 12일

IDAP 광운대학교 방위사업연구소 INSTITUTE OF DEFENSE ACQUISITION PROGRAM

JAMS JOURNAL OF ADVANCES IN MILITARY STUDIES

2025, Vol. 8, No. 1, 1-13.

https://doi.org/10.37944/jams.v8i1.277

국방 객체 탐지를 위한 Few-shot Object Detection(FSOD) 모델 성능의 비교

김준섭* · 최동녁**

국문초록

국방 분야는 민감 정보에 대한 보안성과 데이터 수집의 제약으로 대규모 학습 기반의 객체 탐지(Object Detection, OD) 기술 적용이 현실적으로 한계가 있다. 본 연구는 이런 문제를 해결하기 위해 Few-shot Object Detection(FSOD) 기법을 국방 도메인에 적용하여 모델 성능을 정량적 지표로 검증한다. 이를 위한 실험은 실제 군사 영상(군용 차량 이미지) 데이터셋을 구축하고, 기존 객체 탐지 모델(YOLO, DETR, GLIP, CD-ViTO)을 대상으로 1-shot, 5-shot, 10-shot 조건에서 fine-tuning을 수행하였다. 각 모델의 성능은 객체 탐지 분야의 대표 지표인 평균 정밀도(mean Average Precision, mAP) 기준으로 비교하였으며, 특히 CD-ViTO 모델은 기존 Cross-Domain FSOD 연구에서 사용된 공개 데이터셋과의 비교하여 분석하였다. 실험 결과, CD-ViTO는 군사 데이터셋에서도 높은 mAP 성능을 달성하였으며, 이를 통해 FSOD 기법이 국방 객체 탐지 가능성을 실증적으로 검증하였다는 점에서 연구의 의의가 높다.

주제어 : 객체 탐지, 국방 도메인, 군용 차량 이미지. 퓨샷 학습, 모델 성능 평가

* (제1저자) ㈜펀진, 전임연구원, junsub.kim@funzin.co.kr, https://orcid.org/0009-0008-3223-4529.

** (교신저자) ㈜펀진, 수석연구원, luke.dn.choi@funzin.co.kr, https://orcid.org/0000-0006-3383-1179.