2024년도 한국멀티미디어학회 추계학술발표대회 논문집 제27권 2호

위장 객체 탐지를 위한 데이터 증강 기법

정영훈*, 김유경**, 서정일*
*동아대학교 컴퓨터공학과
**한국전자통신연구원 디지털융합연구소
e-mail : yeonghun.jeong@m4ml.re.kr, yk.kim@etri.re.kr, jeongilseo@dau.ac.kr

Data augmentation method for camouflage object detection

YeongHun Jeong*, Yookyung Kim**, JeongIl Seo*
*Dept of Computer Engineering, Dong-A University
**Digital Convergence Research Laboratory, ETRI

요 약

딥러닝 기반의 영상 처리 기술이 발전하면서 객체 탐지나 추적 기능을 통한 무인 장비의 자동화된 상황 인식 및 대응 능력이 크게 향상되고 있다. 그러나, 인식하기 쉬운 일반적인 객체에 비해 위장된 객체는 주변 환경과 색상, 질감, 패턴 등이 유사하여 배경과 객체를 구별하기 어렵다. 본 연구에서는 위장 객체 탐지를 위한 데이터 증강 기법을 제안한다. 제안한 데이터 증강 기법을 사용하여 위장된 객체와 유사한 배경에서도 객체를 식별할 수 있도록 고안하였다. 제안 방법의 평가는 기존 데이터 증강 기법과 비교하였고, 그 결과 AP_{50}에서는 96.7%, $AP_{50:95}$에서는 67.8%의 정확도를 보여 효과적인 위장 객체 탐지 성능을 확인하였다.

1. 서론

현대전에서 무인화된 무기 체계의 비율이 증가하고 있다. 비무장 지대의 과학화 감시 체계나, 드론과 같은 무인 장비에는 고해상도 카메라 모듈이 장착되어 실시간 영상 수집과 분석을 수행한다. 최근 이러한 무기 체계에 컴퓨터 비전 기술이 접목되면서, 객체 탐지나 추적 기능을 통한 자동화된 상황 인식 및 대응 능력이 크게 향상되고 있다. 이에 따라 감시와 정찰뿐만 아니라, 공중 공격과 표적 추적 등 다양한 군사적 임무에서 무인 무기 체계가 활발히 활용되고 있다.

위장 객체는 인식하기 쉬운 일반적인 객체와 다르게 주변 환경과 색상, 질감, 패턴 등이 매우 유사하게 설계되어 있다. 이러한 특성들 때문에 위장 객체는 배경에 자연스럽게 녹아 들어 쉽게 눈에 띄지 않으며, 이는 기존의 객체 탐지 알고리즘이 배경과 객체를 구별하는 것을 어렵게 만든다. 따라서 기존 객체 탐지 모델을 위장 객체 탐지에 특화된 데이터셋으로 재학습하거나 fine-tuning을 해야 한다.

위장 객체에 대한 데이터셋은 일반 객체 탐지 데이터셋에 비해 상대적으로 양이 부족하다는 어려움이 있다. 데이터가 부족할 경우, 딥러닝 모델은 다양한 상황에서의 일반화 성능이 저하될 수 있으며, 이는 실전 환경에서 탐지 성능이 낮아진다. 객체 탐지 모델을 학습하기 위한 특정 도메인의 데이터의 양이 부족할 때 모델의 성능을 향상시키기 위해 데이터 증강 기법을 사용하는 것이 매우 중요하다.

본 연구는 위장 객체 탐지 모델의 성능을 높이기 위한 최신 데이터 증강 기법인 인페인팅 기법을 제안한다[2]. 이 모델은 위장된 객체 탐지 모델을 위한 데이터셋의 부족을 해결하기 위해, 데이터의 다양성을 극대화하고 데이터셋의 양을 증가시키기 위한 데이터 증강 방식을 도입했다. YOLOv7의 강력한 실시간 객체 탐지 성능을 활용하면서도, 인페인팅을 통해 위장된 객체와 유사한 특징을 가진 배경으로 이미지를 복원하거나 보완함으로써 탐지 성능을 향상시킨다[3]. 제안 모델은 기존 탐지 모델이 가지는 한계를 극복하고, 다양한 환경에서 위장 객체를 더 정확하고 신속하게 탐지할 수 있는 가능성을 제시한다.

본 논문의 구성은 다음과 같다. 2장에서는 기존 데이터 증강 기법과 인페인팅 기법을 소개하고 객체 탐지 알고리즘에 대해 설명한다. 3장에서는 제안한 데이터 증강 기법에 대한 내용을 다룬다. 4장에서는 제안한 데이터 증강 기법과 기존 데이터 증강 기법의 성능을 검증하기 위해 YOLOv7 pretrained model에 fine-tuning하여 성능을 비교한다.

2. 관련연구

2.1 기존 데이터 증강 기법

데이터 증강 기법은 데이터의 핵심 특징을 유지하면서, 노이즈를 더하거나 이미지를 변환하여 데이터셋을 확장하는 방법을 말한다. 이를 통해 더욱 노이즈에 강인한 모델을 얻을 수 있다. 이러한 기법에는 랜덤하게 이미지 자르기, 회전, 밝기 조절, 노이즈 추가 등이 있으며, 이를 적용

하여 원본 이미지와 유사한 특징을 가지면서 생성된다.

2.2 인페인팅 기법

인페인팅 기법은 이미지에서 손실되거나 가려진 부분을 주변의 픽셀 정보나 학습된 패턴을 바탕으로 자연스럽게 복원하는 기술이다. 인페인팅 기법에는 Patch based Approach, Diffusion based Approach, Deep learning based Approach가 있다.

먼저, Patch based Approach는 이미지의 결손된 부분을 주변의 비슷한 패치를 복사하여 채우는 방법이다. 주로 텍스처가 중요한 이미지에서 사용되며, 손상된 부분과 유사한 패턴을 가진 영역을 찾아와 자연스럽게 채운다. 다음으로, Diffusion based Approach는 이미지의 결손된 부분을 주변의 픽셀 값을 기반으로 매끄럽게 확산시켜 채우는 방법이다. 경계선과 색상이 자연스럽게 이어지도록 복원하는 데 집중하며, 주로 간단한 구조나 색상 변화가 적은 이미지에서 사용된다. 마지막으로, Deep learning based Approach는 딥러닝 모델을 사용하여 결손된 부분을 복원하는 방법이다. 대규모 데이터셋으로 훈련된 신경망이 결손된 영역을 학습된 패턴에 기반하여 예측하고 복원한다. 복잡한 구조나 세부 사항을 자연스럽게 채워 넣을 수 있다.

2.3 객체 탐지

객체 탐지 알고리즘은 주요 객체를 탐지하고 해당 객체를 중심으로 경계 박스를 표시하여 구분한다. 최근에는 CNN기반 얼굴 인식, 이미지 분류 등과 같은 다양한 분야에서 활용되고 있다. 딥러닝 기반 객체 탐지에는 대표적으로 One Stage Detectors와 Two Stage Detectors 방법이 있다.

먼저, One Stage Detectors는 한 번의 단계로 객체의 경계 상자와 클래스 레이블을 직접적으로 예측하는 모델이다. 대표적인 One Stage Detectors로는 YOLO (You Only Look Once) 와 SSD(Single Shot MultiBox Detector) 등이 있다[4]. 다음으로, Two Stage Detectors는 객체 탐지를 위해 두 단계 과정을 거친다. 첫 번째 단계에서는 이미지 내의 관심 영역의 후보 세트를 생성한다. 두 번째 단계에서는 이러한 후보 영역에 대해 실제 객체의 경계 상자와 클래스 레이블을 예측한다. 대표적인 Two Stage Detectors로는 R-CNN, Faster R-CNN, Mask R-CNN 등이 있다.

3. 인페인팅 기반 데이터 증강 기법

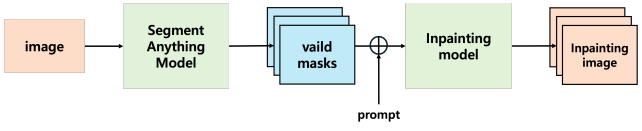


Fig. 1. 제안 방법의 구조도

Fig. 1은 제안하는 인페인팅 데이터 증강 기법의 구조를 나타낸다. 이 기법은 원본 이미지를 입력으로 받아 SAM (Segment Anything Model)을 통해 마스크 후보군을 생성한다[5]. 마스크 후보군 중 객체의 경계를 정확하게 표현한 마스크를 선정한다. 선정한 마스크와 프롬프트를 Stable Diffusion 모델에 입력하여 배경을 다양한 방식으로 변형한다[6].

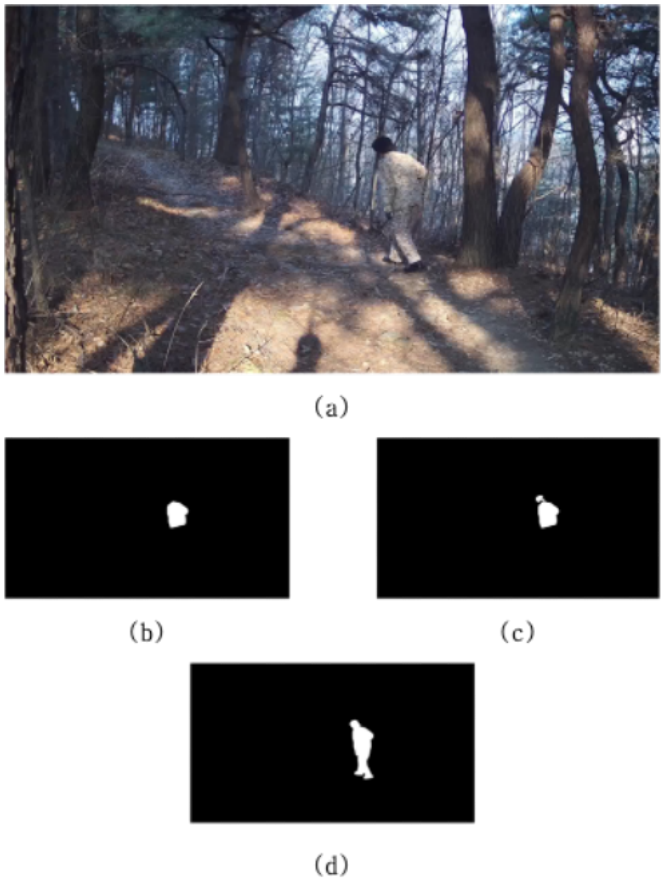


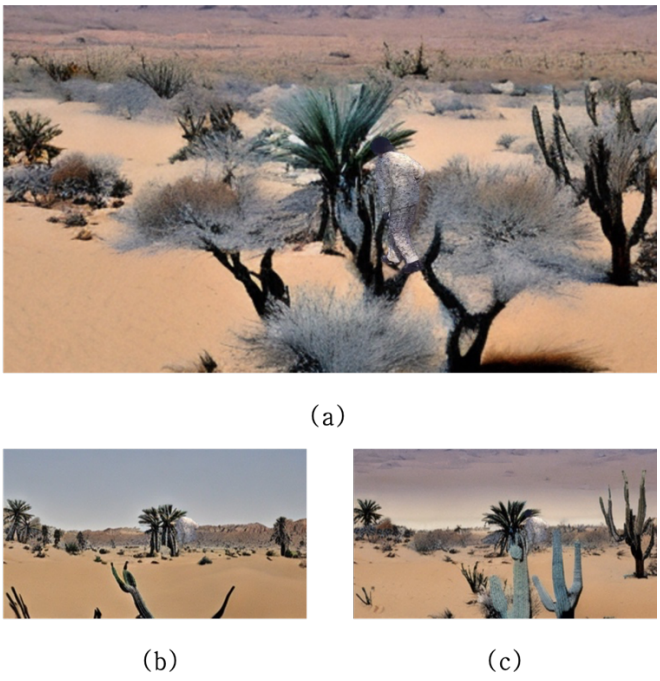
Fig. 2. 원본 이미지와 마스크

원본 이미지에서 특정 객체를 포함하는 3개의 마스크 후보군을 생성하기 위해 SAM을 적용한다. SAM은 사용자가 지정한 좌표를 통해 이미지 내 객체의 경계를 추정하여 마스크를 생성한다. 이 과정은 수동으로 경계를 설정하는 기존 방법보다 더 효율적이며, 다양한 마스크 후보를 생성할 수 있어 데이터의 다양성을 높인다. Fig. 2은 SAM으로 생성된 마스크 후보이다. Fig. 2의 (a)는 원본 이미지이며, Fig. 2의 (b)부터 (d)는 (a)를 이용해 생성된 마스크이다. 이를 보면, 마스크의 영역이 점점 커지는 것을 확인할 수 있다. 다양한 크기의 마스크를 이용해 원하는 객체를 정교하게 분리하고, 객체의 크기나 형태에 맞춰 배경을 자연스럽게 교체하거나 조정할 수 있다. 마스크의 크기를 조절함으로써 이미지 속 특정 객체를 더욱 정밀하게 선택할 수 있다.

Fig. 3는 각 마스크 별로 인페인팅 기법으로 생성된 이

미지들이다. Fig. 3의 (b)와 (c)는 Fig. 2의 (b)와 (c)를 이용하여 생성된 이미지들이다. 객체의 마스크가 객체 전체를 포함하여 생성되지 않은 것을 확인할 수 있다. 반면 Fig. 3의 (a)는 Fig. 2의 (d)로 부터 생성된 이미지이다. 객체의 마스크가 배경 교체를 수행하는데 적합한 것으로 판단된다. 본 연구에서는 인페인팅 할 때 배경을 교체하기 때문에 Fig. 2의 (d)를 선택하였다.

제안 기법은 배경을 변형하여 원본 이미지와 큰 차이를 만들어낸다. 이는 위장된 객체와 배경 간의 경계가 모호한 이미지를 생성하여, 모델이 다양한 위장 환경에 적응할 수 있도록 돕는다. 이를 통해 모델의 학습 범위가 확장되고, 실제 위장 상황에서 더 높은 성능을 발휘할 수 있을 것으로 기대된다.



(a)

(b) (c)

Fig. 3. 인페인팅 기법으로 생성한 이미지

4. 실험

4.1. 실험 환경

학습에 사용한 벤치마크 데이터셋은 ETRI에서 제공받은 데이터셋과 데이터 증강 기법으로 생성된 이미지를 적용하였다. 이미지 학습에 활용한 데이터 개수는 Table 1과 같다. train data는 총 4,200장, validation data는 525장, test data는 525장으로 8:1:1 비율로 구성하였다. Table 2는 인페인팅 기법을 통해 생성한 이미지의 배경 종류다. 원하는 객체는 고정시키고 주위 배경을 교체하였으며, 위장 객체 탐지 성능을 높이기 위해 객체와 구분하기 어려운 환경으로 생성하였다.

드론과 같은 무인 장비에서 효율적으로 동작하는 객체 탐지 모델을 선택하는 과정에서, 본 연구는 YOLOv7 모델을 채택했다. YOLOv7은 기존 모델들에 비해 높은 정확도를 유지하면서도 경량화된 구조 덕분에 추론 속도가 빠르고, 실시간 처리가 가능하다는 장점이 있다. 특히, 본 연구에서는 모델의 정확도를 높이기 위해 다양한 데이터셋을 활용한 fine-tuning 기법을 적용하였으며, 이를 통해 성능을 개선하면서도 추가적인 추론 비용을 발생시키지 않도록 최적화하였다. 이로써 실시간 응용에서의 효율성을 유지하면서도 드론과 같은 무인 장비의 탐지 정확도를 효과적으로 향상시킬 수 있다.

Table 1. 학습 데이터 구성 (이미지 개수)

	train	val	test
original image	3,406	439	439
inpainted image	794	86	86
total	4200	525	525

Table 2. 장소별 인페인팅 데이터셋 구성 (이미지 개수)

	mountain	desert	snow mountain
train	359	257	192
val	43	21	12
test	43	21	12

4.2. 실험 결과 분석

본 연구에서는 YOLOv7 pretrained model에 ETRI에서 제공 받은 데이터셋으로 fine-tuning한 모델을 기준으로 선택했다. 기존 데이터 증강 방식으로 생성한 데이터셋과 인페인팅 기법으로 생성한 데이터셋을 YOLOv7 base 모델을 fine-tuning하여 성능 비교를 진행하였다.

성능 지표로는 AP_{50}(Average Precision)과 $AP_{50:95}$를 사용했다. AP는 IoU(Intersection over Union)에 따라 모델의 정확도를 평가하며, 각 클래스별로 얻은 Precision-Recall 곡선의 면적을 계산하여 산출된다. AP_{50}은 객체의 위치가 50% 이상 겹칠 때를 기준으로 평가한 AP이다. $AP_{50:95}$은 객체의 위치가 GT와 비교하여 50% 이상 95% 이하로 겹칠 때의 AP들을 평균으로 구한 값이다.

Table 3은 제안한 방법이 다른 방법에 비해 정확도가 높은 것을 보여준다. test 이미지 525장으로 각 방법의 AP_{50}과 $AP_{50:95}$를 평가하였다. AP_{50}에서 base 모델은 81.6%, 기존 데이터 증강 기법으로 fine-tuning한 모델은 96.5%가 나왔으며, 인페인팅 기법으로 fine-tuning한 모델은 91.9%가 나왔다. 다른 방법에 비해 제안한 방법은 96.7%가 나왔다. 다른 모델에 비해 약 1% 정도 성능이 향상되었다. $AP_{50:95}$에서는 base 모델은 46.5%, 기존 데이터 증강 기법으로 fine-tuning한 모델은 67.2%가 나왔으며, 인페인팅 기법으로 fine-tuning한 모델은 60.3%가 나왔다. 다른

Table 3. 기존 데이터 증강 기법과 인페인팅 기법 비교

Techniques	Resolution (px)	AP_{50}	$AP_{50:95}$
base	640	81.6%	46.5%
flip, resize, etc	640	96.5%	67.2%
inpainting	640	91.9%	60.3%
inpainting + flip, resize, etc (Ours)	**640**	**96.7%**	**67.8%**

방법에 비해 제안된 기법으로 fine-tuning한 모델은 67.8%로 다른 모델에 비해 약 1~2% 정도 성능이 향상되었다. 실험 결과를 분석하였을 때, 인페인팅 기법과 기존 데이터 증강 기법을 통해 생성된 데이터셋으로 fine-tuning한 YOLOv7 모델이 위장된 객체를 탐지하는데 적합한 것을 알 수 있다. 이는 객체와 유사한 특징의 다양한 배경을 추가적으로 학습시켜서 인페인팅 기법이 위장 객체 탐지 성능을 향상시키는 것을 확인 할 수 있다.

5. 결론

본 연구에서는 위장 객체 탐지를 위한 데이터 증강 기법을 제안하였다. 제안된 방법을 통해 위장된 객체 탐지에 특화된 다양한 환경의 데이터셋을 생성하였다. YOLOv7 pretrained model을 fine-tuning하여 다양한 환경에서 위장된 객체를 잘 탐지하는 것을 확인하였다. 본 논문에서 제안하는 인페인팅 기법은 부족한 데이터셋으로도 데이터증강을 통해 다양한 이미지를 만들며, 위장된 객체의 특징과 유사한 배경으로 교체하여 객체 탐지 모델의 성능을 향상 시킬 수 있다.

제안하는 방법에서는 객체는 고정하고 배경만 교체하였다. 향후 연구로 인페인팅 모델을 위장 객체 탐지 데이터셋에 맞게 재학습하여 객체를 교체하거나 추가하는 실험도 진행하여 여러 형태의 이미지를 생성이 가능하도록 하는 연구가 필요하다.

감사의 글

본 과제(결과물)는 2024년도 교육부의 재원으로 한국연구재단의 지원을 받아 수행된 지자체-대학 협력기반 지역혁신 사업과(2023RIS-007), 2022년도 정부(방위사업청)의 재원으로 국방기술진흥연구소의 지원(No. 21-302-B00-002, 수풀지역 극복 자율비행 드론 개발)을 받아 수행된 연구임

참고문헌

[1] T.N. Le, T.V Nguyen, Z. Nie, M. T. Tran, A. Sugimoto, "Anabranch network for camouflaged object segmentation," *Computer vision and image understanding*, DOI: 10.1016/j.cviu.2019.04.006.

[2] N.M. Salem, "A Survey on Various Image Inpainting Techniques," Future Engineering Journal, Vol. 2, Issue 2, pp: 1-18, 2021.

[3] C.Y. Wang, A. Bochkovskiy, H.Y. Mark Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," arXiv Preprint, arXiv: 2207.02696, 2022.

[4] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, " You only look once: Unified, real-time object detection," Proceedings of the IEEE conference on computer vision and pattern recognition. DOI: 10.1109/CVPR.2016.91.

[5] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, et al, "Segment anything," In Proceedings of the IEEE/CVF International Conference on Computer Vision, DOI: 10.1109/ICCV51070.2023.00371

[6] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, B. Ommer, "High-resolution image synthesis with latent diffusion models," In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, DOI: 10.1109/CVPR52688.2022.01042