

합성 적외선 영상을 활용한 딥러닝 기반 군장비 탐지

이기영^{*,1)} · 김덕윤^{*,1)} · 김유경¹⁾ · 차지훈^{**,1)}

¹⁾ 한국전자통신연구원 자율비행연구실

Deep Learning-based Detection of Military Equipment with Synthetic Infrared Images

Giyoung Lee^{*,1)} · Deokyun Kim^{*,1)} · Yookyung Kim¹⁾ · Jihun Cha^{**,1)}

¹⁾ *Autonomous UAV Research Section, Electronics and Telecommunications Research Institute, Korea*

(Received 4 November 2025 / Revised 17 April 2026 / Accepted 20 April 2026)

Abstract

Recent advances in deep learning-based object detection have highlighted the potential for applications in military domains. However, the scarcity of infrared(IR) training data remains a major limitation. In this work, we compare to representative EO-to-IR generation approaches and evaluate both image synthesis quality and downstream detection performance. Experimental results show that preserving spatial consistency with EO annotations, which means ensuring that EO-derived annotations remain valid by preventing hallucinated or misplaced objects in the generated IR images, is more important for object detection than photorealistic image generation. These findings suggest that our approach offers a practical training strategy for military IR object detection when real IR data are scarce, particularly by preserving spatial consistency between EO images and generated IR images.

Key Words : Deep Learning(심층 학습), Military Equipment Detection(군사 장비 탐지), Synthetic Infrared Image(합성 적외선 영상)

1. 서론

최근 딥러닝 기반 객체 탐지 기술은 군사 분야에도 빠르게 확산되고 있으며, 비정형적인 전장 환경에서 실시간 상황 인식과 신속한 의사결정을 지원하는 핵

심 요소로 주목받고 있다. 특히 탱크, 장갑차, 군 수송 차량 등 주요 지상 장비의 위치와 종류를 신속하게 탐지·식별하는 능력은 감시·정찰(Surveillance and Reconnaissance) 체계의 중추적 기능이자, 지상 전력 운용의 자동화와 지능화를 뒷받침하는 필수 기술로 평가된다.

일반적으로 딥러닝 기반 객체 탐지기는 지도 학습(supervised learning) 방식을 따르며, 학습 이미지에 대응하는 객체의 위치와 종류에 해당하는 정답 어노테

* Equal contribution

** Corresponding author, E-mail: jihun@etri.re.kr

Copyright © The Korea Institute of Military Science and Technology

이션(annotation)과 모델의 예측값 사이의 차이를 손실 함수(loss function)를 통해 계산하고, 이를 최소화하도록 역전파(backpropagation)를 수행하여 모델을 최적화한다. 지금까지의 객체 탐지 연구는 주로 벤치마크 데이터셋을 기반으로 모델 아키텍처 설계 및 손실 함수 개선을 통해 성능 향상을 추구해 왔다. 그러나 최근에는 데이터의 품질과 다양성이 성능에 결정적인 영향을 미친다는 점이 부각되면서, 데이터셋 구축이 연구의 핵심 축으로 자리 잡고 있다.

군사 작전 상황에서 적외선(IR, Infrared) 영상은 물체의 열 특성을 반영하기 때문에 조명 변화나 위장 환경에서도 강인한 탐지가 가능하다는 장점이 있다. 그러나 IR 영상은 고가의 센서 비용, 군사적 기밀성, 제한적인 촬영 조건 등의 제약으로 인해 가시광선 기반의 전자광학(EO, Electro-Optical) 영상에 비해 대규모 확보가 어렵다. 이러한 IR 데이터의 희소성은 단순한 모델 아키텍처 개선만으로는 극복하기 어려운 구조적 한계로 작용하며, 실제 전장 환경에서 멀티모달 학습의 잠재력을 충분히 발휘하는 데 주요 걸림돌이 되고 있다.

자율주행 및 감시 분야에서는 EO-IR 영상 쌍을 제공하는 공개 데이터셋이 보고된 바 있다. 예를 들어, KAIST 멀티스펙트럼 보행자 데이터셋^[1]은 주·야간 환경에서 수집된 EO-IR 영상 쌍을 제공하며, 보행자 검출을 위한 주요 벤치마크로 활용된다. 또한 FLIR ADAS 데이터셋^[2]은 다양한 도로 및 환경 조건에서 촬영된 주·야간 EO-IR 영상 데이터를 포함하고 있어 자율주행 보조 시스템 개발에 특화되어 있다. 그러나 이러한 데이터셋은 주로 보행자나 차량과 같은 민간 객체 위주로 구성되어 있으며, 군사 장비에 대한 IR 영상은 거의 포함되어 있지 않다. 결과적으로 군사 도메인에서는 EO 데이터조차 민간 분야만큼 풍부하지 않을 뿐 아니라, IR 데이터는 더욱 희소하여 학습 데이터의 다양성이 현저히 부족한 실정이다.

이러한 한계를 극복하기 위한 대안으로 최근에는 EO 영상을 활용한 IR 영상 합성 연구^[3,4]가 주목받고 있다. Özkanoglu et al.^[3]은 GAN 기반 EO-to-IR 변환 기법을 제안하여 가시광 영상으로부터 적외선 영상을 생성하였으며, Ran et al.^[4]은 Diffv2IR 모델과 IR-500K 데이터셋을 제안하여 EO-to-IR 변환 성능을 향상시켰다. 또한 Parmar et al.^[5]은 다양한 image-to-image translation 문제를 위한 one-step 영상 변환 구조를 제안하였다.

본 연구에서는 EO 영상만 확보된 상황을 가정하고,

이러한 접근들을 EO-to-IR 변환 문제에 적용하여 합성 IR 생성 성능을 비교·분석하여, 탐지 성능 측면에서 가장 효과적인 생성 모델을 선정하였다. 선정된 기법을 활용해 군 도메인 합성 IR 이미지를 생성하고 이를 군사 장비 탐지기의 학습에 적용함으로써, 실제 IR 영상이 부족한 상황에서도 합성 데이터가 IR 도메인 탐지 성능 향상에 기여할 수 있음을 실험적으로 검증하였다. 본 접근은 (1) EO 영상의 기존 어노테이션을 그대로 활용하여 추가 주석 비용을 절감하고, (2) 실제 IR 데이터가 제한적인 경우에도 학습 데이터의 다양성을 확보할 수 있으며, (3) EO 데이터만 존재하는 상황에서도 합성 IR 이미지를 생성하여, 실제 IR 데이터 없이도 IR 이미지에 대한 객체 탐지 성능을 확보할 수 있음을 보여준다.

2. EO-to-IR 합성을 위한 생성 모델 비교

2.1 실험 구성

EO 기반 IR 합성 기법을 평가하기 위해, EO-IR 영상 쌍을 제공하는 공개 데이터셋인 FLIR ADAS 데이터셋^[2]을 활용하였다. 비교 대상 생성 모델로는 Diffv2IR^[4]과 Pix2Pix-Turbo^[5]를 선정하였다. Pix2Pix-Turbo는 latent diffusion 기반의 single-step image-to-image translation 모델로, adversarial loss를 결합하여 입력 영상에 포함된 객체들의 공간적 배치와 구조를 유지하면서 대상 도메인의 시각적 특성을 반영할 수 있다. EO-IR 대응 데이터셋을 활용하여 EO 영상으로부터 IR 영상을 합성할 수 있으며, 주·야간 변환과 같은 EO 내 다양한 도메인 전환에서 성능을 입증한 바 있다. 반면, Diffv2IR은 DDPM(Denoising Diffusion Probabilistic Model) 기반 접근으로, Vision-Language 이해를 결합하고 Progressive Learning Module을 통해 IR 파장 영역을 점진적으로 학습하며, 대규모 IR-500K 데이터셋을 통해 기존 기법 대비 우수한 품질을 보고하였다.

본 연구에서 Pix2Pix-Turbo 모델은 IR-500K 데이터셋을 이용하여 학습하였다. 모든 EO 영상에 대해 공통적으로 “An infrared image”라는 텍스트 프롬프트를 사용하여 학습하였다. 야간 영상의 경우 EO 영상에서 어두운 객체가 IR 영상에서는 상대적으로 뚜렷하게 나타나는 특성이 있으므로, 주·야간 특성을 구분하여 시간 정보를 프롬프트에 포함하였다. Pix2Pix-Turbo with time 모델은 주간에는 “daytime infrared

image”, 야간에는 “nighttime infrared image” 프롬프트를 사용하여 시간대별 분포 차이를 반영하도록 학습되었다.

한편, Diffv2IR 모델은 EO 영상과 함께 SAM (Segment Anything Model)^[6]으로부터 생성된 분할 마스크(segmentation mask), 그리고 BLIP^[7]으로 생성한 텍스트 프롬프트를 조합한 조건을 입력으로 사용한다. 구체적으로 “turn the visible image of {BLIP 생성 프롬프트} into infrared”라는 조건부 텍스트를 EO 영상 및 분할 마스크와 함께 제공하여 IR 영상을 합성하였다. 또한 본 연구에서 활용한 데이터셋은 객체 바운딩 박스 정보를 포함하고 있으므로, 이를 SAM의 입력으로 제공하여 객체 중심 분할 마스크를 생성한 경우와, 바운딩 박스를 사용하지 않고 전체 이미지를 대상으로 전역 분할 마스크를 생성한 경우를 각각 비교하였다. 전자는 객체 위치에 대한 사전 지식을 반영한 접근이며, 후자는 사전 정보 없이 SAM의 결과만을 활용하는 접근이다. 이를 통해 바운딩 박스 기반 사전 지식이 Diffv2IR 모델의 IR 변환 성능에 미치는 영향을 체계적으로 분석하고자 하였다.

2.2 실험 결과

본 절에서는 EO-IR 변환 합성 품질과, 변환된 IR 영상을 객체 탐지에 활용했을 때의 성능을 함께 평가하였다. 변환된 IR 영상의 품질을 정량적으로 평가하기 위해 시각적 유사도(FID, Fréchet Inception Distance), 픽셀 유사도(PSNR, Peak Signal-to-Noise Ratio), 구조 보존력(SSIM, Structural Similarity Index Measure)을 지표로 사용하였다. FID는 생성 영상과 실제 영상을 Inception-v3 특징 공간에서 비교하여 두 영상 간 분포적 차이를 측정하는 지표로 낮을수록 우수하며, PSNR과 SSIM은 각각 픽셀 단위 복원 품질과 구조적 유사성을 평가하는 지표로 높을수록 원본 영상과의 유사도가 높음을 의미한다. EO-IR 영상이 1:1로 매칭되어 있는 FLIR ADAS 데이터셋^[2]의 테스트셋을 기준으로 평가하였다. 정량적 평가 결과는 Table 1에 제시하였으며, Fig. 1은 합성 결과의 시각적 품질을 예시로 나타낸다.

Table 1의 정량적 평가 결과에서 Diffv2IR은 FID와 SSIM 지표에서 상대적으로 우수한 성능을 보였으며, 이는 대응하는 IR 영상과 유사한 시각적 특성과 구조 정보를 효과적으로 재현하였음을 의미한다. 반면

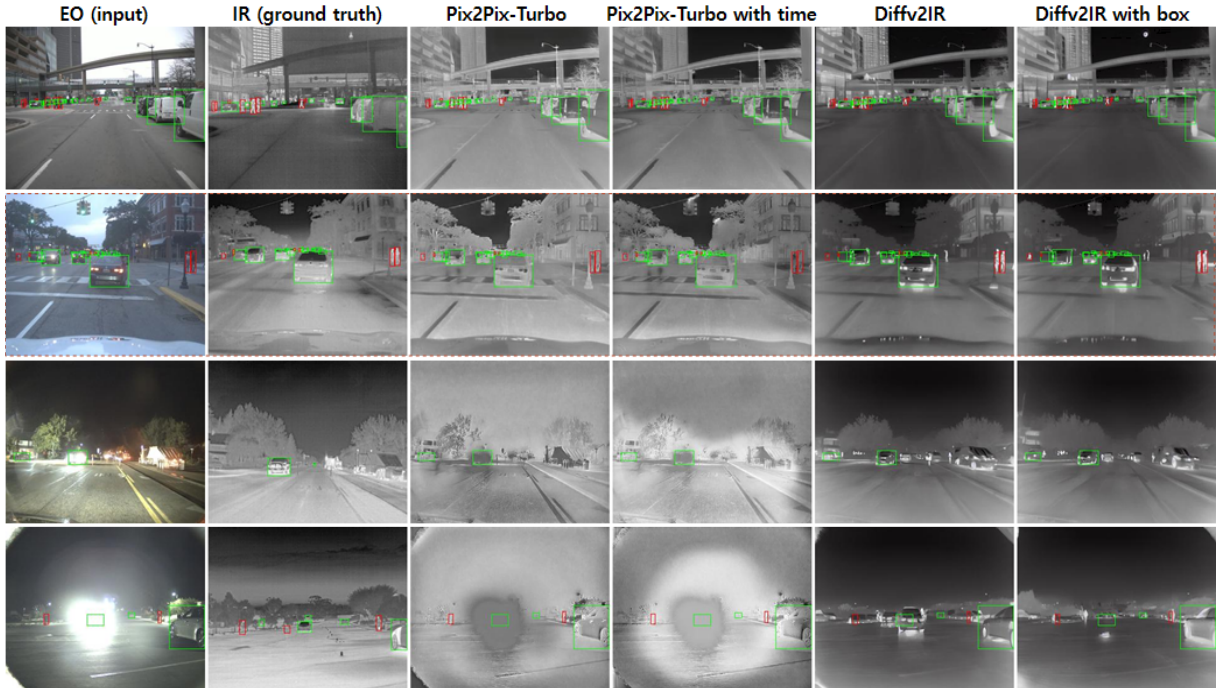


Fig. 1. Qualitative results of synthetic IR generation on the FLIR ADAS dataset(EO input, IR ground truth, and model outputs). Green boxes indicate vehicles and red boxes indicate persons

Pix2Pix-Turbo는 PSNR에서 높은 수치를 기록하여 대응하는 IR 영상과의 픽셀 단위 유사성이 상대적으로 높게 나타났다. 한편, Fig. 1의 정성적 결과에서 Diffv2IR은 열 패턴과 객체 경계를 자연스럽게 재현하였으나, 일부 장면에서는 EO 영상에 존재하지 않는 객체가 생성되는 hallucination 현상이 관찰되었다. Pix2Pix-Turbo 모델이 생성한 IR 영상에서는 hallucination 현상이 거의 나타나지 않았다.

객체 탐지 실험에서는 합성 IR 영상을 이용해 YOLOX-m 모델^[8]을 학습하고, 실제 IR 테스트셋에 대해 탐지 성능을 평가하였다. Table 2는 그 결과를 보여준다. 실험 결과 Pix2Pix-Turbo with time 모델을 이용해 생성한 합성 IR 이미지로 학습한 탐지 모델이 가장 높은 탐지 성능을 기록하였으며, Diffv2IR 계열 모델은 상대적으로 낮은 성능을 보였다. 이는 본 연구의 문제 설정과 같이 EO 데이터셋의 바운딩 박스 어노테이션을 합성 IR 영상에 전이하여 활용하는 경우, 생성 영상의 시각적 품질보다는 EO 영상과 생성된 IR 영상 간 객체 수준의 공간적 정확성이 탐지 성능에 더 직접적인 영향을 미칠 수 있음을 보여준다. 따라서 본 연구에서는 다운스트림 탐지 성능을 기준으로 Pix2Pix-Turbo with time 모델을 최종 생성 모델로 선정하였다.

Table 1. Quantitative evaluation of IR synthesis on FLIR ADAS test set

model	FID↓	PSNR↑	SSIM↑
Pix2Pix-Turbo	199.44	13.44	0.4355
Pix2Pix-Turbo with time	183.66	13.17	0.4436
Diffv2IR	157.04	12.56	0.5293
Diffv2IR with box	161.21	12.70	0.5403

Table 2. Quantitative detection results for FLIR ADAS dataset(YOLOX-m)

trained data	AP@0.5		mAP@0.5
	person	vehicle	
real IR	0.769	0.807	0.788
Pix2Pix-Turbo	0.557	0.594	0.575
Pix2Pix-Turbo with time	0.583	0.592	0.588
Diffv2IR	0.277	0.407	0.342
Diffv2IR with box	0.385	0.431	0.408

3. 합성 IR 영상을 활용한 객체 탐지 성능 평가

3.1 실험 구성

공개 EO-IR 벤치마크 데이터셋인 KAIST 멀티스펙트럼 보행자 데이터셋^[1]과 FLIR ADAS 데이터셋^[2]을 활용하여, 합성 IR 영상을 학습 데이터셋에 포함시켰을 때 객체 탐지 성능에 미치는 영향을 평가하였다. 앞선 생성 모델 비교 실험에서 가장 우수한 것으로 선정된 Pix2Pix-Turbo with time 모델을 이용하여 각 데이터셋의 EO 영상에 대응하는 IR 영상을 합성하였다. 이후 EO 영상만으로 학습한 모델과 EO 영상과 합성 IR 영상을 함께 학습한 모델의 성능을 비교함으로써, 합성 IR 데이터의 기여를 정량적으로 분석하였다.

탐지기로는 YOLOX-m^[8], YOLOv11-m^[9], EfficientDet-D3^[10]의 세 가지 객체 탐지 모델을 선정하였다. 모든 실험은 PyTorch 프레임워크 기반에서 수행하였으며, 학습 조건은 모델 간 공정한 비교를 위해 동일하게 설정하였다. MS-COCO 데이터셋으로 학습된 공개 사전학습 가중치(pretrained weights)를 이용해 전이 학습을 진행하였고, 배치 크기는 24, 학습률(learning rate)은 $1e-2$ 로 설정하였다. 최적화는 확률적 경사 하강법(SGD, Stochastic Gradient Descent)을 사용하였으며, 학습률 조정은 StepLR 스케줄러를 적용하였다. 학습은 총 200 epoch 동안 진행하였으며, 학습 과정에서 검증 데이터셋에서 가장 우수한 성능을 보인 모델을 최종 평가에 사용하였다.

합성 IR 데이터의 활용 효과를 체계적으로 분석하기 위해, 다음과 같이 다섯 가지 학습 데이터셋 구성을 설계하였다.

- ① EO: EO 학습 데이터만 사용
- ② IR: 실제 IR 학습 데이터만 사용
- ③ EO+IR: EO·IR 학습 데이터 동시 사용
- ④ SynIR: 합성 IR 학습 데이터만 사용
- ⑤ EO+SynIR: EO·합성 IR 학습 데이터 동시 사용

3.2 성능 지표

평가 지표는 mean Average Precision(mAP)을 사용하였으며, IoU threshold 50 % 이상 조건에서 계산한 mAP@0.5 값을 기준으로 모델 간 성능을 비교하였다. 모든 성능 평가는 YOLOv5 Metric을 활용하여 일관되게 측정하였다.

3.3 데이터 전처리 및 증강

모델의 일반화 성능을 향상시키기 위해 다음과 같은 데이터 증강 기법을 적용하였다.

- Random Horizontal Flip($p = 0.5$): 좌우 반전을 통해 다양한 시점 변화에 대응할 수 있도록 하였다.
- Random Rotation($p = 0.5$): $\pm 10^\circ$ 범위 내에서 임의 회전을 적용하여 시점 변화에 대한 강건성을 확보하였다.
- Random Hue & Saturation(EO 영상에만 적용): 색조(Hue)와 채도(Saturation)를 조절하여 다양한 조명 및 색상 조건을 반영하였다. Hue와 Saturation의 변동 범위는 $max_delta = 0.2$ 로 설정하였다.
- Random Brightness & Contrast($p = 0.2$): 밝기와 대비를 동시에 변형하여 다양한 광학적 조건을 모사하였다. 밝기와 대비의 변동 범위는 각각 ± 0.2 로 설정하였다.
- Inverse 변환(IR 영상에만 적용, $p = 0.5$): IR 영상의 밝기 값을 반전(inversion)하여 밝은 영역과 어두운 영역의 표현을 반전하였다. 이를 통해 다양한 열 표현 분포에 대한 강건성을 향상시키고자 하였다.
- Mosaic Augmentation($p = 0.5$): 네 장의 이미지를 하나로 결합하는 방식으로, 작은 객체나 밀집된 객체 상황에서 탐지 성능을 강화하기 위해 활용하였다. 단, EfficientDet 모델에서는 학습 안정성을 고려하여 적용하지 않았다.

또한 모델 입력 전 단계에서 영상을 텐서 형태로 변환하고, 정규화(Normalization)를 수행하였다. 정규화 과정에서는 평균(mean)과 표준편차(std)를 각각 $mean = [0.485, 0.456, 0.406]$, $std = [0.229, 0.224, 0.225]$ 로

설정하여 사전 학습된 모델의 입력 분포와 일관성을 유지하였다. 모든 입력 영상은 640×640 크기로 고정하였으며, 원본 비율을 유지한 상태에서 리사이즈한 후 남는 영역은 RGB 값 (128, 128, 128)로 패딩하였다. IR 영상은 단일 채널을 회색조로 읽은 뒤 3채널로 복제하여 EO 영상과 동일한 입력 차원으로 맞추었다.

3.4 KAIST 데이터셋의 실험 결과

KAIST 멀티스펙트럼 보행자 데이터셋은 EO와 IR 영상에 대해 사람(person), 사람들(people), 자전거 탄 사람(cyclist)의 세 가지 객체 라벨을 제공한다. 본 연구에서는 탐지 범위를 단순화하고 데이터 불균형을 완화하기 위해 이를 모두 단일 사람(person) 클래스로 통합하였다. 또한 프레임 간 중복으로 인한 학습 편향을 줄이기 위해 학습 데이터는 3 프레임마다, 테스트 데이터는 20 프레임마다 추출하였다. 전체 학습 데이터의 10 %에 해당하는 세 개의 비디오 시퀀스는 무작위로 선택하여 검증(validation)용으로 사용하였다. 데이터셋의 이미지 수와 타깃 개수에 대한 세부 통계는 Table 4에 요약하였다.

Table 3은 KAIST 데이터셋에서 합성 IR과 실제 EO-IR 데이터를 활용했을 때의 탐지 성능(AP@0.5)을 비교한 결과이다. EO 단독 학습은 EO 테스트셋에서는 양호했으나 IR에서는 성능이 저하되었고, IR 단독 학습은 반대로 EO에서 낮은 성능을 보였다. 두 모달리티를 함께 활용한 EO+IR 학습은 EO와 IR 테스트셋 모두에서 가장 높은 성능을 기록했으며, EO 탐지와 IR 탐지 모두에서 상호 보완 효과가 확인되었다. 한편, EO와 합성 IR을 함께 사용한 EO+SynIR 학습 역시 EO+IR 학습에 준하는 성능을 보여, 실제 IR 데이

Table 3. Quantitative detection results for KAIST dataset(AP@0.5)

trained data	YOLOX-m			YOLOv11-m			EfficientDet-D3		
	EO	IR	EO + IR	EO	IR	EO + IR	EO	IR	EO + IR
EO	0.551	0.198	0.379	0.440	0.146	0.301	0.524	0.165	0.345
IR	0.302	0.538	0.422	0.060	0.525	0.304	0.399	0.537	0.467
EO + IR	0.558	0.541	0.550	0.491	0.528	0.509	0.536	0.519	0.527
SynIR	0.262	0.465	0.363	0.075	0.405	0.262	0.341	0.402	0.370
EO + SynIR	0.555	0.479	0.516	0.493	0.434	0.462	0.535	0.456	0.494

터가 제한적인 상황에서 합성 IR이 효과적인 대안이 될 수 있음을 입증하였다.

Table 4. Statistics of target classes and images in the KAIST dataset

		train	val	test
EO/IR	# Videos	22	3	16
	# Images	9,296	1,619	1,142
	# Targets person	21,030	5,161	2,743

3.5. FLIR ADAS 데이터셋의 실험 결과

FLIR ADAS 데이터셋은 자율주행 기반 도로 환경에서 촬영된 EO/IR 영상으로 총 15개 객체 라벨을 포함한다. 본 연구에서는 객체 수가 충분한 사람(person)과 자동차(vehicle) 두 개 클래스를 대상으로 실험을 수행하였다. 데이터셋의 이미지 수와 타겟 개수에 대한 통계는 Table 5에 제시하였다.

Table 5. Statistics of target classes and images in the FLIR ADAS dataset

			train	val	test
EO	# Images		9,635	1,004	3,614
	# Targets	person	35,007	3,223	11,278
		vehicle	71,281	7,285	30,893
IR	# Images		10,223	1,096	3,493
	# Targets	person	50,478	4,470	12,323
		vehicle	73,623	7,133	30,517

Table 6의 실험 결과, KAIST 데이터셋과 동일하게 EO 단독 학습은 EO 테스트셋에서만 우수했으나 IR 도메인에서는 성능 저하를 보였고, IR 단독 학습은 반대의 양상을 나타냈다. EO와 IR을 함께 학습한 경우 두 테스트셋 모두에서 가장 높은 성능을 기록하였으며, 이는 서로의 학습 데이터로부터 상호 보완적인 학습 효과를 얻어, 일반화 능력을 향상시킨 결과로 볼 수 있다. 또한 EO+SynIR 학습 역시 EO+IR에 준하는 성능을 보여, 합성 IR 데이터가 실제 IR 데이터를 일정 부분 대체할 수 있음을 확인하였다. 이러한 결과는 서로 다른 모달리티 데이터가 상호 보완적으로 작용하여 단일 모달리티 학습 대비 더 높은 일반화 성능을 달성할 수 있음을 뒷받침한다.

3.6 합성 적외선 영상 데이터셋의 효과 분석

KAIST와 FLIR ADAS 데이터셋 실험 결과, EO와 실제 IR을 함께 학습한 EO+IR이 가장 높은 성능을 보였다. EO+SynIR은 EO+IR에는 미치지 못했으나, 실제 IR이 없는 상황에서는 EO 단독 학습 대비 뚜렷한 이점을 제공하였다. 특히 EO 테스트셋에서는 유사하거나 더 나은 성능을, IR 테스트셋에서는 뚜렷한 성능 개선을 확인하였다. 이는 합성 IR이 EO 학습 데이터의 한계를 보완하고 일반화 성능 향상에 기여함을 보여준다.

또한 EO와 IR은 상호 보완적으로 작용하여 특징 표현을 강화하며, 합성 IR 역시 EO 라벨과의 정합성을 유지함으로써 다운스트림 탐지 성능 개선에 기여하였다. 따라서 합성 IR은 실제 IR의 완전한 대체제는 아니지만, IR이 부족하거나 존재하지 않는 환경에서 EO 단독 학습 대비 실질적이고 비용 효율적인 대안이 될 수 있다.

Table 6. Quantitative detection results for FLIR ADAS dataset(mAP@0.5)

trained data	YOLOX-m			YOLOv11-m			EfficientDet-D3		
	EO	IR	EO + IR	EO	IR	EO + IR	EO	IR	EO + IR
EO	0.627	0.382	0.489	0.643	0.215	0.413	0.488	0.257	0.368
IR	0.412	0.788	0.610	0.396	0.801	0.611	0.307	0.692	0.505
EO + IR	0.642	0.791	0.717	0.659	0.797	0.730	0.511	0.701	0.610
SynIR	0.474	0.588	0.523	0.535	0.592	0.548	0.412	0.526	0.465
EO + SynIR	0.613	0.601	0.603	0.636	0.569	0.581	0.458	0.514	0.486

4. 합성 IR 기반 군사 장비 탐지 성능 평가

4.1 실험 구성

본 절에서는 인터넷을 통해 수집한 군사 장비 EO 영상 데이터셋을 활용하여 합성 IR 영상이 군사 장비 탐지에 미치는 효과를 검증하였다. 데이터셋은 ImageNet, OpenImagesV7, Roboflow 등 공개 리소스로부터 탱크, 장갑차, 군용 차량이 포함된 이미지를 수집하고, 이를 사람(person), 탱크(tank), 군용 차량(vehicle) 세 개 클래스로 구분하여 LabelStudio^[11]를 활용해 레이블링하였다. 실제 IR 영상은 확보 가능한 수량이 매우 제한적이어서, 테스트 단계에서만 성능 검증용으로 활용하였다. 데이터셋의 이미지 수와 클래스별 분포는 Table 7에 요약하였다.

앞선 생성 모델 비교 실험에서 선정된 Pix2Pix-Turbo with time 모델로 EO 영상으로부터 합성 IR 영상을 생성하였으며, Fig. 2는 그 예시를 제시한다. 합성된 IR 영상은 EO 라벨과 공간적으로 정합성을 유지하면서 실제 IR 영상과 유사한 특성을 보여준다.

Table 7. Statistics of target classes and images in our military equipment dataset

			train	val	test
EO	# Images		6,646	758	1,506
	# Targets	person	7,711	868	1,763
		tank	6,537	905	1,910
		vehicle	3,919	250	460
IR	# Images		-	-	893
	# Targets	person	-	-	739
		tank	-	-	795
		vehicle	-	-	264

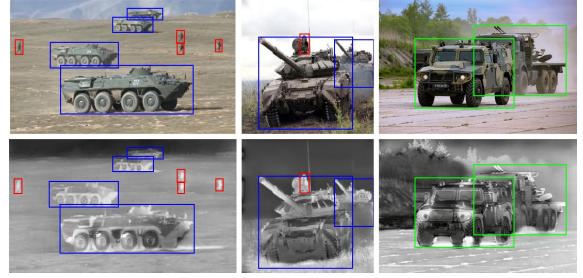


Fig. 2. Examples from our military dataset: EO(top) and synthetic IR(bottom). Red indicates person, blue indicates tank, and green indicates military vehicle

4.2. 실험 결과

탐지기는 Section 3과 동일하게 YOLOX-m, YOLOv11-m, EfficientDet-D3 구조를 사용하였으며, EO 단독 학습과 EO+합성 IR 학습을 비교하였다. Table 8은 그 결과를 정량적으로 제시한다. EO 단독 학습 모델은 EO 테스트셋에서는 높은 성능을 기록했으나, 실제 IR 테스트셋에서는 성능이 급격히 저하되었다. Fig. 3의 검출 결과 예시에서도 IR 영상에서 객체를 놓치거나 잘못 탐지한 것을 확인할 수 있다. 이는 EO 데이터만으로는 야간·열 기반 상황에서 나타나는 군사 객체의 특성을 충분히 포착하지 못하여, 은폐된 병력이나 장비를 식별하는 데 한계가 있음을 시사한다.

반면 합성 IR 데이터로 학습한 모델은 EO 테스트셋에서 다소 성능 저하가 있었으나, IR 테스트셋에서는 안정적인 성능을 유지하였다. 특히 EO+SynIR 학습은 EO 테스트셋에서 약간의 성능 손실이 있었으나, IR 테스트셋에서는 뚜렷한 개선을 달성하였다. EO 단독 학습 대비 IR 탐지 성능이 크게 향상되었으며, 종합 성능(mAP@0.5) 역시 가장 높은 수준에 도달하였다. 이는 합성 IR 데이터가 EO 데이터만으로는 부족한 부분을 효과적으로 보완하고, 군사 객체 탐지의 일반화 성능을 강화하는 데 기여함을 입증한다.

Table 8. Quantitative detection results for our military equipment dataset(mAP@0.5)

trained data	YOLOX-m			YOLOv11-m			EfficientDet-D3		
	EO	IR	EO + IR	EO	IR	EO + IR	EO	IR	EO + IR
EO	0.882	0.499	0.777	0.852	0.372	0.716	0.856	0.611	0.775
SynIR	0.742	0.711	0.730	0.676	0.616	0.652	0.789	0.725	0.768
EO + SynIR	0.862	0.719	0.819	0.833	0.651	0.773	0.847	0.639	0.779

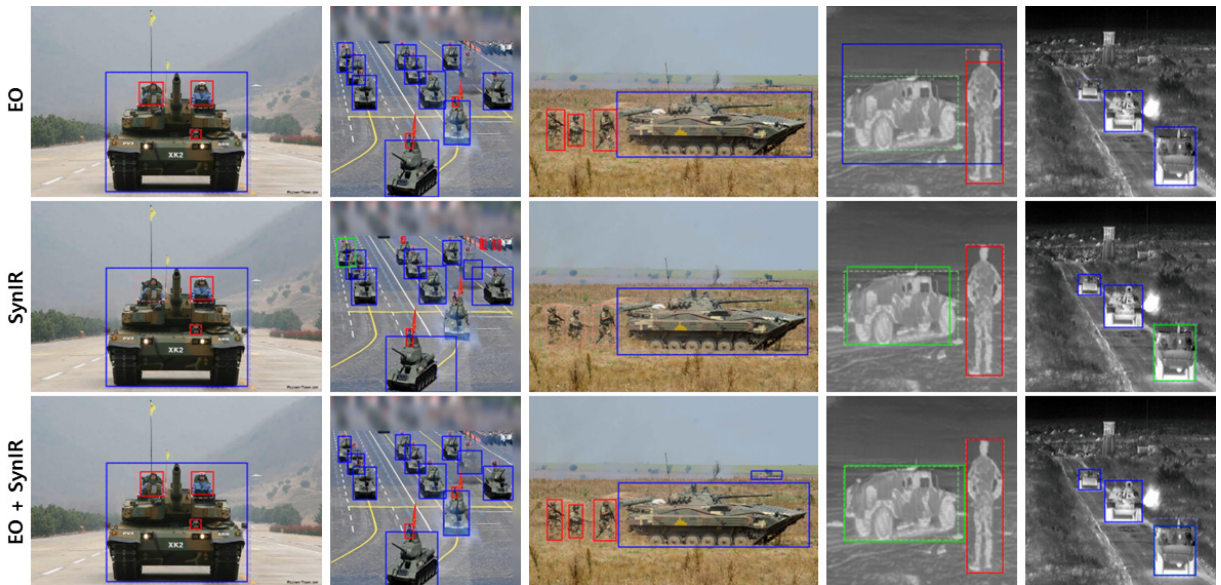


Fig. 3. Visual comparison of detection results: (top) YOLOX-m trained on EO data, (middle) trained on synthetic IR data, (bottom) trained on EO and synthetic IR data. Red indicates person, blue indicates tank, and green indicates military vehicle, with solid boxes for detections and dashed boxes for ground truth.

5. 결론

본 연구는 군사 장비 탐지에서 적외선(IR) 데이터 부족 문제를 해결하기 위해, 전자광학(EO) 영상을 기반으로 합성 IR 데이터를 생성해 학습에 활용하는 방안을 제안하였다. 기존 생성 모델들이 사실적인 영상 생성에 집중해 온 것과 달리, 본 연구는 탐지 성능 향상 측면에서는 EO 주석과 합성 결과 간 정합성이 더 중요함을 규명하였다. 또한 EO와 합성 IR을 함께 활용할 경우 EO 단독 학습 대비 IR 성능이 유의미하게 개선되는 것을 확인하였다. 이러한 결과는 실제 IR 데이터 수집이 어려운 상황에서 EO 데이터와 합성 IR 데이터를 함께 활용하는 전략이 IR 도메인 탐지 성능을 향상시키는 실질적인 대안이 될 수 있음을 보여준다.

후 기

이 논문은 2022년 정부(방위사업청)의 재원으로 국방기술진흥연구소의 지원을 받아 수행된 연구임(KRIT-CT-22-018, 수폴지역 극복 자율비행 드론 개발).

References

- [1] S. Hwang, J. Park, N. Kim, Y. Choi and I. So Kweon, "Multispectral pedestrian detection: Benchmark dataset and baseline," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition(CVPR), pp. 1037-1045, 2015.
- [2] FLIR Thermal Dataset, [Online]. Available: <https://www.flir.com/oem/adas/adas-dataset-form/>
- [3] M. A. Özkanoglu and S. Ozer, "InfraGAN: A GAN architecture to transfer visible images to infrared domain," Pattern Recognition Letters, Vol. 155, pp. 69-76, 2022.
- [4] L. Ran, L. Wang, G. Wang, P. Wang and Y. Zhang, "DiffV2IR: visible-to-infrared diffusion model via vision-language understanding," arXiv preprint arXiv: 2503.19012, 2025.
- [5] G. Parmar, T. Park, S. Narasimhan and J. Y. Zhu, "One-step image translation with text-to-image models," arXiv preprint arXiv:2403.12036, 2024.
- [6] A. Kirillov, et al., "Segment anything," Proceedings of the IEEE/CVF International Conference on

- Computer Vision(ICCV), 2023.
- [7] J. Li, et al., “BLIP: Bootstrapping language-image pre-training for unified vision-language understanding and generation,” Proceedings of the International Conference on Machine Learning(ICML), 2022.
- [8] Z. Ge, et al., “YOLOX: Exceeding YOLO series in 2021,” arXiv preprint arXiv:2107.08430, 2021.
- [9] G. Jocher and J. Qiu, “Ultralytics YOLO11,” GitHub Repository: <https://github.com/ultralytics/ultralytics>, Version 11.0.0, 2024.
- [10] M. Tan, R. Pang and Q. V. Le, “EfficientDet: Scalable and efficient object detection,” Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR), pp. 10781-10790, 2020.
- [11] Label Studio, [Online]. Available: <https://labelstud.io/>