

부분 폐색 상황에서의 열영상 탐지 성능 향상을 위한 데이터 증강 기법

공민중^{0, 1*}, 조수영^{0, 1*}, 박진선^{1, 2}

¹ 부산대학교 정보컴퓨터공학부

² 부산대학교 Digital-X AIoT 연구센터

{alswnddl02, whtndud101, jspark}@pusan.ac.kr

요 약

최근 군 병력 감축에 따라 전방 경계 작전의 효율성을 제고하기 위해 열영상 감시 체계에 인공지능 기반 객체 탐지 기술을 적용하려는 시도가 증가하고 있다. 그러나 야전 환경의 산악 지형과 수풀로 인한 객체 폐색(Occlusion)은 열영상의 탐지 성능을 저하시키는 주요 원인이다. 특히 국방 분야에서는 미탐지가 작전 실패로 이어질 수 있어 높은 Recall(재현율) 확보가 필수적이다. 이에 본 논문에서는 객체의 국소 특징 학습을 강화하는 Inpainting 기반 기하학적 증강과, 실제 가림 현상을 모사하는 Diffusion 기반 의미론적 증강을 결합한 하이브리드 기법을 제안하였다. 실험 결과, 제안 기법은 YOLOv11 베이스라인 대비 원본 테스트셋에서 Recall 이 약 4.9%p, 폐색 데이터셋에서 약 12.3%p 향상되어 다양한 가림 환경에서 미탐지 문제를 효과적으로 완화함을 확인하였다.

1. 서론

전 세계적인 병력 감축 기조에 따라 국방 경계 시스템은 인력 중심에서 Thermal Observation Device(TOD) 및 UAV 등을 활용한 유무인 복합 체계로 전환되고 있다[1]. 그러나 실제 야전 환경에서는 산악 지형과 장애물로 인해 표적이 가려지는 폐색 상황이 빈번하며, 색상 정보가 부재한 열영상의 특성상 이는 심각한 성능 저하를 야기한다[2]. 적군 식별과 같은 감시 정찰 임무에서 객체의 미탐지는 치명적인 결과를 초래할 수 있으므로, 제한된 정보만으로도 객체를 강건하게 탐지하는 기술이 요구된다.

기존 연구들은 CutOut[3], Random Erasing[4], GridMask[5]와 같은 픽셀 수준의 기하학적 증강을 통해 이 문제를 해결하고자 하였으나, 이는 단순한 정보 소실만을 유도할 뿐 실제 자연물의 복잡한 가림 패턴을 모사하기에는 한계가 있다. 이에 따라 최근 생성형 AI(Generative AI)를 활용한 증강 연구가 대두되고 있으나[6], 대부분 RGB 영상에 집중되어 있어 열화상 도메인의 텍스처와 노이즈 특성을 반영하지 못하는 실정이다.

이에 본 연구에서는 열영상 내 부분 폐색 객체에 대한 탐지 강건성을 확보하기 위해, Inpainting[7]과 diffusion 생성 모델인 FLUX.1-Fill[8]을 결합한 하이브리드 데이터 증강 기법을 제안한다. 특히 높은 프롬프트 추종성을 갖는 FLUX 모델을 활용하여 수풀이나 나뭇가지와 같은 장애물을 열화상

텍스처로 정교하게 생성함으로써, 단순 노이즈가 아닌 실제 야전 환경의 가림 현상을 효과적으로 모사하였다. 이러한 의미론적 생성 증강은 Inpainting을 통한 객체의 불완전한 형상 추론 학습과 상호 보완적으로 작용하여, 결과적으로 국방 분야의 핵심 지표인 Recall 성능을 개선하는 효과를 보였다.

2. 방법론

본 연구는 열화상 객체 탐지 모델의 일반화 성능과 폐색 강건성을 동시에 확보하기 위해, 원본 데이터, Inpainting 증강, FLUX 기반 생성형 증강을 결합한 하이브리드 데이터 증강 기법을 제안한다. 그림 1은 원본 데이터셋에 Inpainting 데이터셋, FLUX 모델을 통해 생성한 Occluded 데이터셋을 추가하여 YOLOv11을 Fine-tuning 하는 전체 파이프라인이다.

제안한 방법에서 원본 열화상 이미지는 두 개의 독립적인 증강 경로로 분기된다.

첫 번째 경로는 Inpainting 모델을 통해 객체의 국소 영역을 제거하여 데이터의 기하학적 다양성을 확보한다. 두 번째 경로는 FLUX 모델을 활용하여 원본 이미지 위에 가상의 장애물을 합성함으로써 의미론적 다양성을 부여한다.

최종적으로 원본 이미지와 두 경로를 통해 생성된 증강 이미지들이 하나의 복합 데이터셋으로 통합되며, 이는 YOLOv11 모델의 학습 데이터로 활용된다.

* Equal Contribution

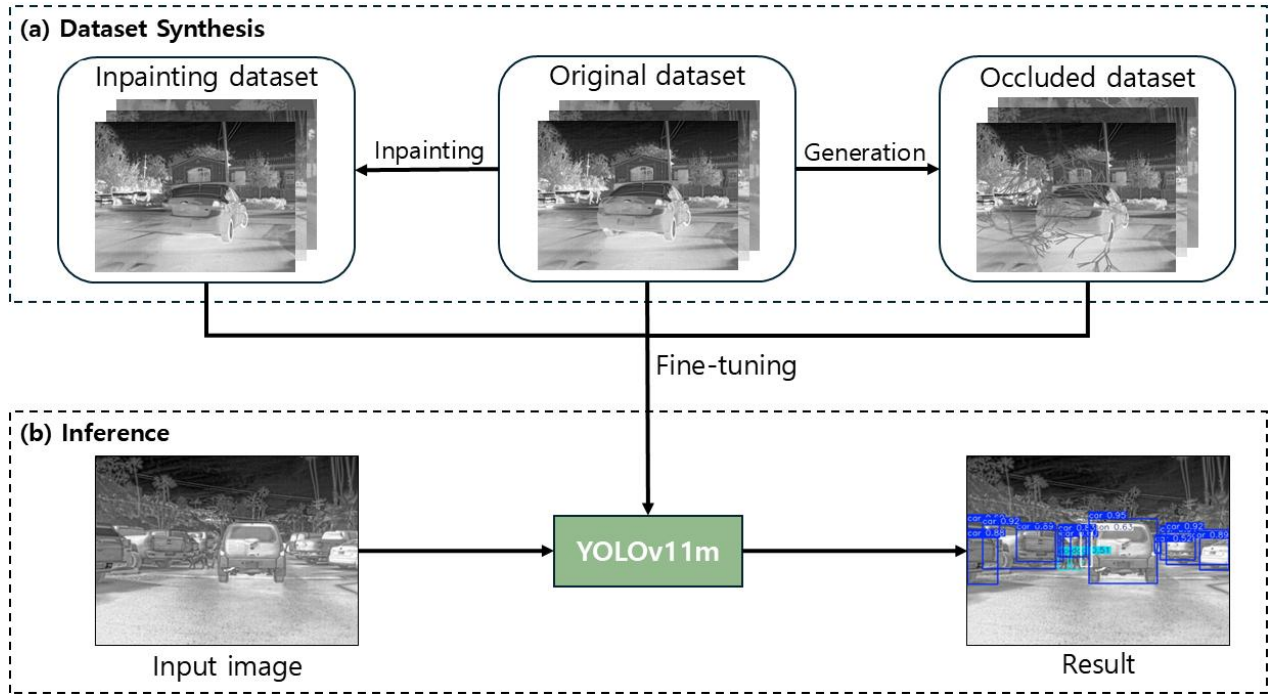


그림 1. Inpainting 과 FLUX 모델을 결합한 데이터의 증강 파이프라인

2.1 Inpainting 기반 증강

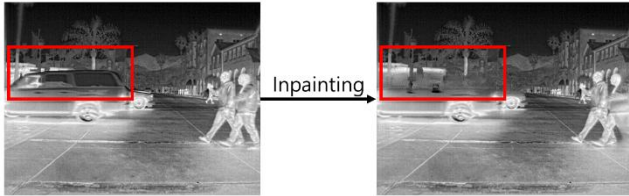


그림 2. Inpainting 기법 적용 결과

전통적인 CutOut 이나 Random Erasing 기법은 제거된 영역을 0(검은색)이나 무작위 노이즈로 채우기 때문에, 모델이 이를 비자연적인 표현으로 인식하여 학습을 방해할 소지가 있다. 이를 해결하기 위해 본 연구는 객체의 일부를 제거한 후 배경 정보를 자연스럽게 복원하는 Inpainting 기법을 도입함으로써, 모델이 객체의 전체 형상에 의존하지 않고 부분적인 단서와 주변 맥락을 활용해 객체를 추론하도록 설계하였다. 이 과정에서 효율적이고 안정적인 배경 복원을 위해 SimpleLaMa 모델을 채택하였다.

여기서 SimpleLaMa 는 LaMa(Large Masking Field)[9] 아키텍처를 기반으로, Fast Fourier Convolution(FFC)[10]을 통해 이미지의 전역적 수용 영역을 활용해 그림 2 와 같이 넓은 영역에서의 자연스러운 배경 질감을 생성한다.

학습 데이터셋은 $D = \{(I_i, B_i)\}_{i=1}^N$ 으로, I_i 는 i 번째 입력 열화상 이미지, $B_i = \{b_{i,j}\}_{j=1}^M$ 는 해당 이미지 내에 존재하는 M 개 객체의 bounding box 집합, N 은 전체 데이터 샘플의 수를 의미한다. 제안하는 증

강 기법은 무작위 마스크 생성 함수 M_s 을 통해 각 객체 $b_{i,j}$ 의 부분 영역을 선택하고, 입력 이미지와 마스크의 합집합을 받아 소거된 영역을 배경 맥락으로 자연스럽게 대체하는 인페인팅 모델 Φ_{LaMa} 로 복원하는 과정으로 정의된다. 최종 증강 이미지 I'_i 는 식 1 과 같이 단일 수식으로 표현된다:

$$I'_i = \Phi_{LaMa}(I_i, \cup_{j=1}^M k_{i,j} \cdot M_s(b_{i,j})) \quad (1)$$

여기서 $k_{i,j} \sim \text{Bernoulli}(P_{prob})$ 는 증강 적용 여부를 결정하는 확률 변수이며, P_{prob} 는 객체의 Inpainting 적용 여부를 제어하기 위한 확률 임계값으로, 학습 시 하이퍼파라미터로 입력된다.

2.2 FLUX 기반 생성형 증강

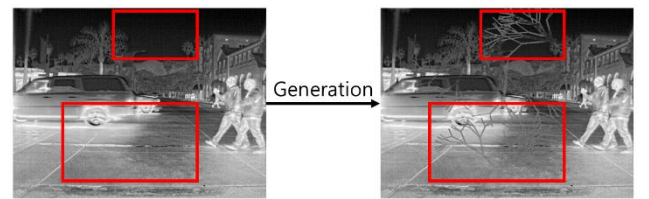


그림 3. 생성 기법 적용 결과

기존의 기하학적 증강 방식은 픽셀 정보를 단순히 소실시키거나 노이즈로 대체하기 때문에 시각적 불연속성을 초래하고 실제 자연환경의 가림 현상을 반영하는 데 한계가 있다. 이에 본 연구에서는 diffusion 모델인 FLUX 를 활용하여 그림 3 과 같이, 열화상 도메인의 특성을 유지하면서도 의미론적으로 자연스러운 합성 Occluded 데이터를 구축하였다.

우선 실제 야전 환경의 수풀이나 나뭇가지와 같은 비정형적 장애물을 모사하기 위해 프랙탈 트리(Fractal Tree) 알고리즘을 적용하여 불규칙하고 복잡한 패턴의 이진 마스크를 생성하였다. 이후 생성된 마스크 영역에 열화상 특유의 텍스처와 노이즈 패턴을 모사하기 위해 다음과 같은 텍스트 프롬프트를 적용하였다:

“thermal imaging texture, grayscale. View through a thicket of thin, dry, withered tree twigs and tangled branches in the foreground, covered in dense fog and thick mist. The atmosphere is hazy, with low visibility, obscuring the background. Atmospheric effect, heavy grain, noisy thermal sensor style.”

이처럼 짙은 안개와 나뭇가지, 센서 노이즈를 모사하는 구체적인 프롬프트를 통해 가시광선 데이터로 사전 학습된 생성 모델의 표현적 편향을 억제하고 원본 열화상 배경과의 시각적 이질감을 최소화하여 도메인 간 격차를 해소하였다.

3. 실험 결과 및 분석

3.1 실험 환경

실험은 Ubuntu 20.04 환경의 NVIDIA RTX A6000 GPU 를 사용하여 수행되었다. YOLOv11m 모델의 Fine-tuning 은 40 Epoch 과 배치 크기 64 로 설정하였으며, 상세한 학습 하이퍼파라미터는 표 1 과 같다.

표 1. 하이퍼파라미터 세팅

Hyperparameter	Value
Model	YOLOv11m
Input Resolution	640 × 640
Batch Size	64
Total Epochs	40
Optimizer	SGD
Initial Learning Rate (lr_0)	1e-4
Momentum	0.937
Weight Decay	5e-4
P_{prob}	0.8

베이스라인 모델로는 객체 탐지 모델인 YOLOv11m 을 적용하였으며, 데이터셋은 Roboflow FLIR Thermal Dataset (v27)[11]을 활용하였다. 훈련 데이터셋은 원본, Inpainting 증강, 생성형(FLUX) 증강의 세 가지를 혼합한 복합 데이터셋으로 구성되며 이를 YOLOv11m 모델 학습에 활용한다.

3.2 정량적 분석

제안하는 기법의 유효성을 검증하기 위해 먼저 베이스라인 모델과 제안 모델의 전체적인 성능 차이를 분석하고, 이후 세부적인 학습 데이터 구성에 따른 소거 연구 결과를 기술하였다.

일반적으로 강한 데이터 증강을 적용할 경우 학습 데이터의 분포가 변형되어 원본 데이터에 대한 성능이 저하되는 부작용이 발생하기 쉽다. 이를 확인하기 위해, 가려짐이 없는 Original Test Set 에 대하여 베이스라인 모델과, 제안하는 Ours(All Mixed) 모델의 성능을 비교하였다.

표 2. 원본 테스트셋에 대한 YOLOv11 과 제안 모델 성능 비교

Method	Recall	mAP50	mAP50-95
Baseline	0.820	0.895	0.580
Ours	0.869	0.909	0.561

표 2 에서 볼 수 있듯이, 제안하는 모델은 베이스라인 대비 Recall 이 0.820 에서 0.869 로 4.9%p 향상되었고, mAP50 이 0.895 에서 0.909 로 1.4%p 향상되었다. 이는 다양한 occluded 데이터를 학습하는 과정에서 모델이 객체의 특징을 더 견고하게 학습하여 일반적인 환경에서의 탐지 능력도 강화되었음을 의미한다. 반면, mAP50-95 는 0.580 에서 0.561 로 하락하였다. 이러한 현상의 원인은 본 연구에서 적용한 FLUX 기반 증강의 특성에서 찾을 수 있다. 학습 데이터 생성 시 적용된 짙은 안개(dense fog)와 센서 노이즈(heavy grain)가 포함된 프롬프트는 실제 환경을 모사하는 과정에서 객체의 외곽선 정보를 흐릿하게 만든다. 이로 인해 모델은 객체의 존재 유무를 판단하는 능력은 강화되었으나, 흐릿해진 경계 정보로 인해 객체의 정확한 테두리를 특징하는 단계에서 불확실성이 증가한 것으로 분석된다.

제안하는 기법이 실제 폐색 상황에서 얼마나 효과적이지, 그리고 각 증강 요소(Inpainting, FLUX)가 성능에 미치는 영향을 분석하기 위해 학습 데이터 구성을 세분화하여 교차 검증을 수행하였다. 우선 베이스라인(Original)을 설정하고, 이를 원본에 Inpainting 과 FLUX 를 각각 결합한 Orig.+Inp. 및 Orig.+FLUX 모델과 비교하였다. 나아가 제안 모델인 All Mixed(Ours)는 이들 모든 데이터를 통합하여 학습하도록 구성함으로써 각 증강 기법의 시너지 효과를 확인하고자 하였다. 아울러 각 모델의 일반화 성능과 강건성을 다각도로 평가하기 위해 테스트 데이터셋을 Original, Inpainting, FLUX 의 단일 속성 3 종과 이들이 혼합된 Orig.+Inp.과 Orig.+FLUX, 그리고 All Mixed 의 복합 속성 3 종을 포함한 총 6 가지 시나리오로 세분화하였다. 표 3 은 전체 테스트 시나리오에 대한 상세 결과이다.

표 3. 학습 데이터 및 테스트 데이터 조합에 따른 성능 비교 (R: Recall, mAP50, mAP50-95)

Test \ Train	Original			Inp.			FLUX			Orig.+Inp.			Orig.+FLUX			All Mixed		
	R	mAP 50	mAP 50-95	R	mAP 50	mAP 50-95	R	mAP 50	mAP 50-95	R	mAP 50	mAP 50-95	R	mAP 50	mAP 50-95	R	mAP 50	mAP 50-95
Original	0.820	0.895	0.580	0.773	0.836	0.506	0.674	0.779	0.487	0.794	0.865	0.542	0.752	0.838	0.534	0.757	0.837	0.524
Orig.+Inp.	0.852	0.925	0.582	0.795	0.880	0.555	0.721	0.808	0.495	0.829	0.903	0.568	0.769	0.869	0.539	0.778	0.873	0.544
Orig.+FLUX	0.846	0.885	0.577	0.780	0.853	0.519	0.763	0.835	0.521	0.815	0.869	0.548	0.811	0.860	0.548	0.803	0.856	0.536
All Mixed	0.869	0.909	0.561	0.848	0.892	0.541	0.797	0.818	0.488	0.862	0.900	0.549	0.806	0.864	0.522	0.821	0.873	0.528

Orig.+Inp. 모델과 Orig.+FLUX 모델이 각각 기하학적 형태 변화와 텍스처 및 문맥 변화에 국소적인 강점을 보인 것과 대조적으로, 제안 모델(All Mixed)은 Inpainting Test Set (Recall 0.848)과 FLUX Test Set (Recall 0.797) 모두에서 일관되게 향상된 성능을 달성했을 뿐만 아니라, All Mixed Test Set (mAP50 0.873, Recall 0.821)에 대해서도 가장 높은 탐지성능을 달성하였다. 결론적으로 제안 모델은 특정 폐색 유형에 편중되지 않고 다양한 가림 환경에 유연하게 대응하는 범용적인 탐지 강건성을 확보하였으며, 폐색 상황에서 객체 탐지 성능을 향상시키는 효과적인 접근임을 정량적으로 입증하였다.

다만, All Mixed 모델의 mAP50-95 수치가 각 단일 증강 모델 대비 소폭 낮게 나타난 현상은 주목할 만하다. 이는 서로 상이한 성격의 폐색 패턴이 bounding box 학습 과정에서 서로 충돌하였기 때문으로 분석된다. 구체적으로 Inpainting 증강은 객체의 일부가 완전히 소실된 상태이므로, 모델은 보이지 않는 영역까지 형상을 유추하여 박스를 넓게 잡으려는 경향을 갖는다. 반면, FLUX 증강은 얇은 나뭇가지나 안개 등의 노이즈가 섞인 상태이므로, 장애물을 배제하고 실제 객체가 보이는 영역에 맞춰 박스를 국한하려는 경향을 갖는다.

이처럼 영역을 복원하여 확장하려는 경향과 노이즈를 피해 범위를 좁히려는 경향이 혼재됨에 따라, 모델이 bounding box의 경계를 설정하는 데 있어 절충안을 찾게 되었고, 이로 인해 mAP50-95 성능이 다소 하락한 것으로 보인다.

결론적으로 제안 모델은 정밀도 측면의 미세한 감소는 있었으나, 작전상 중요한 높은 Recall을 확보하였을 뿐만 아니라 다양한 가림 환경에 유연하게 대응할 수 있는 범용적 탐지 능력을 입증하였다.

3.3 정성적 분석

그림 4는 베이스라인과 본 논문이 제안한 방법의 정성적 비교를 제공한다. 상단의 원본 테스트 데이터셋 결과를 살펴보면, 차량이 이미지의 가장자리에 위치하여 차체의 상당 부분이 프레임 밖으로 잘려나감과 동시에, 지나가는 보행자에 의해 추가적인 폐색까지 발생한 것을 확인할 수 있다.

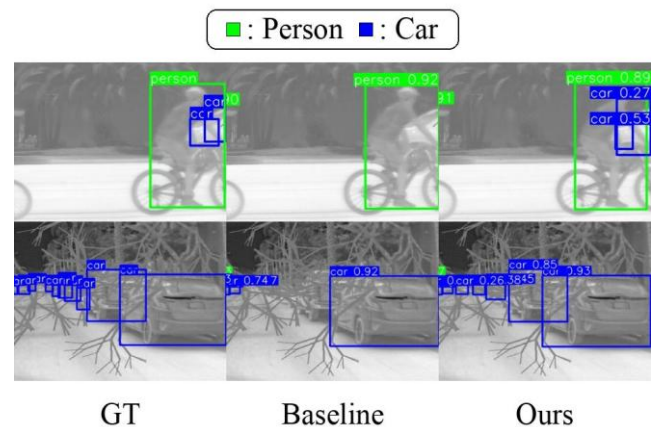


그림 4. 원본 테스트 데이터셋(위)과 FLUX 테스트 데이터셋(아래)에서 진행한 정성적 성능비교

이러한 복합적인 정보 손실 상황에서 베이스라인 모델은 탐지에 실패하였으나, 제안 모델은 화면에 드러난 극히 일부분의 형상만으로도 해당 객체가 차량임을 정확히 추론해냈다. 이는 Inpainting 학습을 통해 획득한 강건한 부분 특징 추론 능력, 촬영 앵글에 의한 자연적인 잘림 현상뿐만 아니라 보행자와 같은 장애물에 의한 폐색 상황에도 효과적으로 적용됨을 시사한다. 하단의 FLUX 테스트 데이터셋 결과에서는 두 모델의 강건성 차이가 더욱 극명하게 드러난다. 해당 예시는 FLUX 모델이 생성한 나무에 의해 차량의 열 신호 대부분이 가려진 상황이다. 베이스라인 모델은 나뭇가지와 수풀에 의해 발생한 노이즈 효과로 차량을 전혀 탐지하지 못하였으나, 제안 모델은 수풀 사이로 미세하게 드러난 차량의 국소적 특징을 포착하여 정확한 bounding box를 예측하였다.

4. 결론

본 연구에서는 산악 및 야전 환경에서 운용되는 TOD 및 UAV 열영상 감시 체계의 핵심 난제인 객체 폐색 문제를 해결하기 위해 Inpainting 기법과 생성형 모델 기반의 의미론적 합성을 결합한 데이터 증강 기법을 제안하였다.

이는 객체의 국소 영역을 제거하고 이를 배경 정보로 복원하는 Inpainting 기법을 통해 모델이 파편화된 특징만으로도 객체를 식별하도록 유도하였으며, Diffusion 모델인 FLUX 를 활용하여 열화상 도메인의 텍스처를 보존한 채 실제 수풀과 유사한 장애물을 합성함으로써 고난도의 폐색 상황을 모사하였다.

실험 결과, 제안한 기법은 베이스라인 대비 원본 테스트셋에서 Recall 은 0.869, mAP50 은 0.909 를 기록하며 일반화 성능을 향상시켰다.

특히, Inpainting 테스트셋과 FLUX 테스트셋에서 다른 기법들보다 높은 Recall 성능을 달성하여 두 증강 기법이 상호 보완적으로 작용함을 입증하였다. 정성적 분석에서도 제안 모델은 나뭇가지나 수풀에 의해 열 신호가 심각하게 가려지거나 화면 앵글에 의해 잘려 나간 객체에 대해서도 정확한 탐지를 수행하였다. 비록 실제 야전 환경의 짙은 안개와 노이즈를 학습하며 객체 외곽선 정보가 모호해짐에 따라 mAP50-95 는 소폭 감소하였으나, 객체의 존재 유무 판단이 중요한 국방 감시 체계 등의 다양한 응용을 고려할 경우 미탐지를 줄인 높은 Recall 성능은 본 연구의 높은 활용성을 의미한다.

향후 연구에서는 제안된 증강 모델을 경량화하여 실제 전방 감시 초소나 정찰 드론과 같은 엠티디바이스 환경에서도 실시간으로 동작 가능한 최적화 연구를 수행하고자 한다.

참고문헌

- [1] J. Cho, J. Park, and J. Park, "Object Detection Performance Analysis of Infrared Camera and Weapon System Application," Military Robotics Society, vol. 1, no. 1, pp. 1-8, 2022.
- [2] Z. Song, Y. Yan, Y. Cao, et al., "An infrared dataset for partially occluded person detection in complex environment for search and rescue," Scientific Data, vol. 12, art. no. 300, 2025.
- [3] T. DeVries and G. W. Taylor, "Improved regularization of convolutional neural networks with cutout," arXiv preprint arXiv:1708.04552, 2017.
- [4] Z. Zhong, L. Zheng, G. Kang, S. Li, and Y. Yang, "Random erasing data augmentation," in Proc. AAAI Conf. Artif. Intell., 2020, vol. 34, no. 7, pp. 13001-13008.
- [5] P. Chen, S. Liu, H. Zhao, and J. Jia, "GridMask Data Augmentation," arXiv preprint arXiv:2001.04086, 2020.
- [6] Y. Li et al., "A simple background augmentation method for object detection with diffusion model," in Proc. Eur. Conf. Comput. Vis. (ECCV), Cham, Switzerland: Springer Nature, 2024.
- [7] D. Pathak, P. Krahenbuhl, J. Donahue, T. Darrell, and A. A. Efros, "Context encoders: Feature learning by inpainting," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2016, pp. 2536-2544.
- [8] Black Forest Labs, "FLUX.1-Fill-dev Model Card," Hugging Face, 2024. [Online]. Available: <https://huggingface.co/black-forest-labs/FLUX.1-Fill-dev>. (Accessed: Dec. 21, 2025).
- [9] R. Suvorov et al., "Resolution-robust Large Mask Inpainting with Fourier Convolutions," in Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV), 2022, pp. 2149-2159.
- [10] L. Chi, B. Jiang, and Y. Mu, "Fast Fourier Convolution," in Proc. Adv. Neural Inf. Process. Syst. (NeurIPS), 2020, vol. 33, pp. 4479-4488.
- [11] Roboflow, "FLIR Thermal Dataset (v27)," Roboflow Universe, 2024. [Online]. Available: <https://universe.roboflow.com/thermal-imaging-0hwwfw/flir-data-set>. (Accessed: Dec. 21, 2025).