

딥러닝 기반의 장애물에 가려진 전장 객체에 대한 복원 성능 분석

한태준, 조영현, 김규리, 이윤서, 이광일*

국립한국해양대학교

{wnsdl0122, peakso_, k20221336, foxes1017}@g.kmou.ac.kr, *leeki@kmou.ac.kr

An Analysis of Deep Learning-based Restoration Performance for Warfare Objects Occluded by Obstacles

Tae Joon Han, Younghyeon Cho, Gyuri Kim, Yunseo Lee, and Kwangil Lee*

National Korea Maritime and Ocean University

요약

본 논문에서는 전장 환경을 고려하여 직접 데이터셋을 구축하고 이를 활용하여 다양한 복원 모델의 성능을 분석한다. 먼저, 전장 가상 환경 시뮬레이터를 활용하여 EO 및 IR 영상 데이터를 직접 수집한다. 전장 객체 종류, 영상 시점, 장애물 유무 등 다양한 요소들을 포함하는 복잡한 전장 상황을 고려한다. 또한, 다양한 공개 데이터셋을 병합하여 복원 모델의 일반화 성능을 높인다. 마지막으로 실험을 통해 구축한 데이터셋을 기반으로 복원 모델들의 복원 성능과 처리 속도를 비교하여 국방 분야 활용에 적합한지 분석한다.

I. 서론

전시 상황에서 정확한 적군 식별은 작전 성공의 핵심 요소이다. 그러나, 전장 환경은 복잡하고 변화가 빠르며, 다양한 장애물과 기상 조건은 객체 인식의 신뢰성과 정확성을 저하시킨다. 이는 상황 인식 및 의사결정에 부정적인 영향을 미쳐 막대한 피해로 이어질 수 있다. 따라서, 전장 환경에서도 빠르고 정확한 식별을 위해, 영상 복원 기법을 활용한 연구가 요구된다.

한편, 군사적 보안성, 정보의 기밀성, 그리고 특수 환경에 대한 데이터 획득의 어려움으로 인해 공개된 데이터셋이 제한적이다. 일반적인 공개 데이터셋과 달리, 국방 분야의 데이터는 높은 복잡성, 다양한 센서 조건, 실제 전장 환경 요소들을 반영하며, 데이터에 이러한 특성이 없으면 연구 개발의 신뢰성과 확장성이 심각한 제약이 발생할 수 있다. 또한, 모델의 학습과 평가 과정에서 일반화 성능을 보장하기 어렵고, 실제 적용 시 성능 저하가 우려된다.

앞선 분석에 차안하여, 본 논문에서는 특수한 전장 환경이 고려된 새로운 전장 데이터셋을 구축하고 이를 활용하여 다양한 복원 모델의 성능을 분석한다. 먼저, 데이터 수집을 위해 전장 가상 환경 시뮬레이터를 이용한다. EO 및 IR 영상 데이터를 다루며 전장 객체 종류, 영상 시점, 장애물 유무 등을 고려한다. 또한, 복원 모델의 일반화 성능을 높이기 위해 다양한 공개 데이터셋을 이용하여 데이터셋을 구축한다. 구축한 데이터셋을 기반으로 복원 모델들을 학습시키고, 복원 성능과 처리 속도를 평가하여 실시간성과 신뢰성이 요구되는 국방 분야에 적합한지 분석한다.

II. 데이터셋 구축

국방 분야의 영상 복원 연구의 어려움 중 하나는 적절한 학습용 데이터를 확보하는 것이다. 실제 전장 환경에서 촬영된 EO/IR 영상은 군사 기밀

과 보안상의 이유로 외부 접근이 제한되며, 다양한 상황을 포괄할 수 있을 만큼 충분한 수량을 확보하기도 어렵다. 이러한 한계를 해결하기 위해, 본 논문에서는 전장 환경 시뮬레이터를 활용하여 학습용 데이터를 직접 생성하였다.

시뮬레이션에는 총 6종의 주요 전장 객체(고정고사포, 장갑차, 곡사포, 방사포, 군인, 탱크)를 포함하였으며, 객체마다 장애물에 가려진 장면과 가려지지 않은 장면을 모두 구성하였다. 무인기(unmanned aerial vehicle) 활용 시의 시점과 유사한 관측 환경을 재현하기 위해 카메라가 지면으로 부터 약 45도의 시야각을 유지하도록 설정하여 영상을 획득하였다. 이를 통해 다양한 거리 및 상황에서 촬영하여 균형 있는 EO 및 IR 데이터를 확보하였다. 해당 방식으로 수집한 데이터는 EO 영상 11,960장, IR 영상 12,256장으로 구성되어 있으며, 세부 항목은 표 1에 정리되어 있다.

시뮬레이션 데이터만으로는 복원 모델이 다양한 상황에 대한 일반화 성능을 기대하기 어렵기 때문에 여러 공공 데이터셋을 추가하여 학습에 활용하였다. 복원 모델들의 학습을 위해 YouTube-VOS, DAVIS 등 다수의 대규모 비디오 데이터셋을 도입하였으며, 실제 도로 주행 영상, 자율 주행 드론 영상, RGBT234, VOT2016 등 복잡한 장면이 포함된 EO/IR 영상 데이터를 포함하였다. 이를 통해 다양한 객체 크기, 움직임, 장애물 조건에 대해 장인하게 복원할 수 있도록 모델의 학습 다양성을 확보하였다. 학습에 사용된 총 데이터 수는 EO 영상 약 91만 6천 장, IR 영상 약 14만 7천 장에 달하며, 전체 구성은 표 2에 요약되어 있다. 이러한 대규모 학습 데이터는 복원 모델이 전장 상황뿐만 아니라 일반적인 상황에서도 높은 복원 성능을 발휘할 수 있는 기반이 된다.

III. 실험 결과

본 논문에서는 2장에서 자체적으로 구축한 데이터셋을 이용하여 복원 모델들을 각각 처음부터 학습시켰다. 학습을 위해 사용된 하드웨어는

NVIDIA A100 80GB GPU, AMD EPYC 7763 CPU, 1TB RAM으로 구성되었으며, 소프트웨어 환경은 Ubuntu 20.04, NVIDIA CUDA 12.5.1, Python 3.10.12, PyTorch 2.3.0을 사용하였다.

다양한 전장 환경에서 전장 객체가 장애물에 가려지는 영상을 이용하여 복원 모델의 성능 분석을 진행하였다. 복원 성능은 SSIM[5]을 통해 정량적으로 평가하였고, 실험을 통해 얻은 각 모델의 성능 지표는 표 3을 통해 확인할 수 있다.

표 3에 따르면, E2FGVI 모델의 SSIM은 0.9967로 가장 높은 복원 품질을 기록하였다. 이와 유사한 SSIM을 가진 FcF는 E2FGVI보다 장당 처리 속도가 약 2.67배 빠른 것을 확인할 수 있다. 다른 두 복원 모델인 DSTT와 STTN은 속도 면에서는 DSTT가 유리하고, 복원 성능 면에서는 STTN이 유리하지만, FcF이 모든 측면에서 앞서는 것을 볼 수 있다.

앞선 실험 결과를 통해 FcF는 복잡한 전장 환경의 영상에서 우수한 복원 성능과 빠른 복원 추론 속도 달성이 가능한 것을 확인하였다. 하지만 장당 소요시간은 181.19ms 즉, 5.52fps에 해당한다.

IV. 결론

본 논문에서는 전장 환경을 고려하여 직접 데이터셋을 구축하고, 이를 활용하여 다양한 복원 모델의 성능을 비교 분석하였다. 먼저, 전장 가상 환경 시뮬레이터를 활용하여 EO 및 IR 영상 데이터를 직접 수집하였다.

표 1. 전장 환경 시뮬레이터로 수집한 전장 객체 6종 학습 데이터

Table 1. Learning data for 6 types of battlefield objects collected with battlefield environment simulator.

전장 객체	세부 객체	EO 데이터	IR 데이터
고정고사포	ZPU-4	1087	1136
장갑차	BTR-80	1087	1072
곡사포	MO-120RT	1079	1083
방사포	M901	1087	1082
방사포	SA-6	1093	1081
군인	SA7-prone	1085	1093
군인	SA7-standing	1086	1087
군인	PK74-prone	1079	1076
군인	PK74-standing	1081	1363
탱크	T-69	1108	1093
탱크	T-72	1088	1090
총계		11,960	12,256

표 2. 사용된 복원 모델 전체 학습 데이터셋 구성

Table 2. Composition of the data set used to train the restoration model.

데이터셋	EO	IR
전장 환경 시뮬레이터 생성 데이터	11,960	12,256
YouTube-VOS	469,887	-
DAVIS	62,370	-
드론 자율항법을 위한 영상 및 센서 데이터	17,876	-
UA-DERAC	140,132	-
UAVDT	76,284	-
RGBT234	116,949	116,949
VOT2016	21,356	18,500
총계	916,814	147,705

복잡한 전장 상황을 반영할 수 있도록 전장 객체 종류, 영상 시점, 장애물 유무 등 다양한 요소들을 이용하였다. 또한, 복원 모델의 일반화 성능을 높이기 위해 YouTube-VOS, DAVIS 등 다양한 공개 데이터셋을 병합하여 대규모 학습용 영상 데이터셋을 구성하였다. 구축한 데이터셋을 기반으로 네 가지 복원 모델을 각각 학습시키고, SSIM과 장당 소요시간을 기준으로 성능을 비교하였다. 구축한 데이터셋을 이용한 실험 결과는 FcF 모델이 우수한 SSIM과 가장 빠른 추론 속도를 보여주었다.

본 논문을 통해, 전장 환경과 유사한 조건에서 수집된 EO/IR 영상 기반 복원 성능을 종합적으로 평가할 수 있는 실험 환경을 제시하였으며, 복원 모델의 실제 적용 가능성에 대한 실증적 근거를 마련하였다. 이러한 결과는 향후 국방 분야에서의 실시간 상황 인식 시스템에 복원 기술을 접목할 수 있는 기초 자료로 활용될 수 있을 것이다.

한편, 일반적으로 모델의 실시간 처리 기준은 30fps으로 고려한다. 하지만 FcF는 5.52fps로 실시간성을 확보하기에는 어려움이 있다. 향후 연구에서는 FcF 모델을 이용하여 실시간성을 확보하기 위해 연구를 진행할 예정이다. 또한, 좀 더 다양한 전장 환경을 고려할 수 있도록 구축한 데이터셋을 보강할 계획이다.

ACKNOWLEDGMENT

This work was supported by Korea Institute for Advancement of Technology(KIAT) grant funded by the Korea Government(MOTIE) (P0024163, Development of RIC(Regional Innovation Cluster)).

참 고 문 헌

- [1] J. Jain, Y. Zhou, N. Yu, and H. Shi, “Keys to better image inpainting: Structure and texture go hand in hand,” arXiv:2208.03382, 2022.
- [2] R. Liu, H. Deng, Y. Huang, X. Shi, L. Lu, W. Sun, X. Wang, J. Dai, and H. Li, “Decoupled spatial-temporal transformer for video inpainting,” arXiv:2104.06637, 2021.
- [3] Y. Zeng, J. Fu, and H. Chao, “Learning joint spatial-temporal transformations for video inpainting,” In ECCV, pp. 528–543, 2020.
- [4] Z. Li, C. Lu, J. Qin, C. Guo, and M. Cheng, “Towards an end-to-end framework for flow-guided video inpainting,” arXiv:2204.02663, 2022.
- [5] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: From error visibility to structural similarity,” IEEE Transactions on Image Processing, vol. 13, no. 4, pp. 600–612, 2004.

표 3. 전장 객체 복원 실험 결과(SSIM)

Table 3. Battlefield object restoration experiment results(SSIM).

복원 모델	장당 소요시간(ms)	SSIM
FcF [1]	181.19	0.9930
DSTT [2]	279.66	0.9777
STTN [3]	305.81	0.9840
E2FGVI [4]	483.60	0.9967