https://doi.org/10.7840/kics.2026.51.1.184

실시간 전장 환경의 EO/IR 영상 복원 및 객체 인식을 위한 통합 시스템 개발

이 광 일◆°, 김 재 환*, 이 창 은*

Development of an Integrated System for Real-Time EO/IR Image Restoration and Object Recognition in Battlefield Environments

Kwangil Lee◆°, Jaehwan Kim*, Chang-Eun Lee*

요 약

현대 군사 기술은 빠르게 변화하며 높은 불확실성을 갖는 전장에서 실시간 상황 인식이 가능한 기술을 요구한다. 본 논문에서는 장애물에 가려진 전장 객체를 실시간으로 복원하고 인식하는 영상 복원-인식 통합 시스템을 제안한다. 제안된 시스템은 DSTT (decoupled spatial-temporal transformer) 기반의 복원 모듈과 RT-DETR 기반의 인식 모듈로 구성된다. 복원 모듈의 시공간적 연속성과 빠른 추론 속도를 달성하기 위해 온라인 복원 기법과 메모리 뱅크 기법을 적용한다. 또한, 인식 모듈의 빠른 처리 속도를 위해 TensorRT 최적화를 적용한다. 마지막으로, 전장 가상 환경 시뮬레이터를 이용하여 데이터 세트를 구축하고, 이를 기반으로 실험 결과를 통해 본 논문의 주요 목표인 향상된 인식 정확도와 20 fps 이상의 추론 속도를 확인하여 제안하는 시스템의 타당성을 검증한다.

키워드 : 실시간, 객체 복원, 객체 인식, EO/IR 영상, 트랜스포머, 전장 환경 가상 시뮬레이터, 전장 환경 데이터 세트

Key Words : Real-time, Object restoration, Object recognition, EO/IR images, Transformer, Battlefield virtual environment simulator, Battlefield environment dataset

ABSTRACT

Modern military technology demands real-time situational awareness capabilities in highly dynamic and unpredictable battlefield environments. In this paper, we propose an integrated image restoration and recognition system that restores and recognizes battlefield objects occluded by obstacles in real time. The proposed system consists of a restoration module based on the decoupled spatial-temporal transformer (DSTT) and a recognition module based on RT-DETR. To achieve spatiotemporal continuity and fast inference speed of the restoration module, the online restoration and memory bank techniques are applied. In addition, the recognition module is optimized using TensorRT to enable high-speed processing. Finally, a dataset is constructed using a battlefield virtual environment simulator, and experimental results are presented to verify the feasibility of the proposed system by showing improved recognition accuracy and inference speed of more than 20 fps, which are the main goals of this paper.

※ 이 논문은 2024년 정부(방위사업청)의 재원으로 국방기술진흥연구소의 지원을 받아 수행된 연구임("실시간 복합 전장정보 상황인지 기술", 21-107-E00-009-02)
◆° First and Corresponding Author : DMASTA, leeki@kmou.ac.kr, 종신회원
* Electronics and Telecommunications Research Institute (ETRI), jh.kim@etri.re.kr; celee@etri.re.kr, 정회원
논문번호 : 202504-101-E-RU, Received April 30, 2025; Revised July 21, 2025; Accepted July 31, 2025

Ⅰ. 서 론

현대 군사 기술은 인공지능, 빅데이터, 첨단 센서 기술의 급격한 발전에 따라 더욱 정교하고 복잡한 형태로 진화하고 있다. 다양한 군사 작전에서 인공지능 및 첨단 센서 기술의 융합은 필수적인 요소가 되었고, 이러한 기술들은 전장 상황 인식을 실시간으로 수행할 수 있는 능력이 요구된다[1,2]. 특히, 신속하고 정확한 전략적 의사결정이 군사 작전의 성패를 결정하기 때문에 전장 객체의 위치와 움직임 등 전장 환경에 대한 실시간 정보 확보는 전략적 우위를 확보하기 위해 필수적이다.

전장은 복잡하고 빠르게 변화하는 환경이므로, 적의 위협을 조기에 탐지하고 신속하게 대응하기 위해 수집된 데이터를 지연 없이 처리하는 기술이 요구된다. 데이터 처리 지연은 군사적 피해로 직결되는 문제이기 때문에, 실시간성은 현대 군용 시스템의 필수 요건으로 자리 잡고 있다. 또한, 전장 환경은 다양한 불확실성과 비정형적 요소를 내포하고 있어, 다양한 장애물과 기상 조건 등의 영향으로 인해 객체 인식과 데이터 분석의 신뢰성과 정확성이 저하될 수 있다.

이러한 문제를 극복하기 위해 데이터 복원 기법과 객체 인식 기술을 결합하여 전장 환경에서 신뢰도 높은 정보를 실시간으로 제공할 수 있는 통합 시스템 구축이 필요하다. 특히 전장 환경에서는 실시간성이 중요한 기술적 요건이기 때문에, 복원과 인식이 빠르고 정밀하게 동작해야 한다.

한편, 컴퓨터 비전 분야에서는 인공지능을 활용한 객체 인식, 객체 추적, 복원 등 다양한 작업에 관한 연구가 활발히 진행되고 있다[3-6,19]. [7]에서는 두 단계 객체 검출기 (two-stage object detector)의 검출 헤드 (detection head)에 추가할 수 있는 플러그인 (plugin) 모듈을 제안하였고, 부분적으로 가려진 객체의 재현율을 향상하여 가려진 객체의 인식 성능을 개선했다. [8]에서 제안된 ProPainter는 동영상 인페인팅 (video inpainting) 작업에서 프레임 간 전파 (propagation)와 렌더링 (rendering) 성능을 향상했다. 이 모델은 새로운 아키텍처를 통해 YouTube-VOS 데이터 세트에서 기존 최첨단 모델을 능가하는 성능을 달성하며, 비디오 복원 작업의 최첨단 성능 모델로 자리 잡았다. 하지만, 낮은 추론 속도로 인해 실시간 처리에는 한계가 존재한다.

최근 복원 분야 연구에서는 DSTT (decoupled spatial-temporal transformer)[9]가 제안되었다. 이 모델은 공간적 전파와 시간적 전파가 분리된 트랜스포머 블록으로 구성함으로써 기존 STTN (spatial-temporal transformer network)[10] 대비 계산 효율성과 복원 성능을 동시에 향상했다. 또한, 프레임 간의 시공간 정보를 정교하게 전달하고 실시간 처리가 가능하다는 점에서 전장 환경에서 가려진 객체를 복원하는 데 적합한 모델이다.

이와 동시에 객체 인식 분야에서는 RT-DETR (real time DETR)[11]이 주목받고 있다. RT-DETR은 어텐션 기반의 실시간 end-to-end 객체 검출기로, AIFI (attention-based intra-scale feature interaction)와 CCFF (CNN-based cross-scale feature fusion) 모듈로 구성된 효율적인 하이브리드 인코더를 통해 멀티스케일 특징을 빠르게 처리한다. 또한, 불확실성 최소화 쿼리 선택 (uncertainty-minimal query selection) 기법을 도입하여 탐지 정확도를 향상했다. 한편, 후처리 과정이 없어서, 추가적인 비최대 억제 (non-maximum suppression) 파라미터 설정이 없고, 재학습 없이 추론 속도를 유연하게 조절할 수 있어 실시간 응용에 적합한 모델이다.

하지만, 객체 복원 및 인식 기술의 발전에도 불구하고, 장애물로 인해 가려진 객체를 복원한 후 이를 기반으로 인식하는 통합적 접근에 관한 연구는 아직 미흡한 실정이다. 대부분의 연구는 객체 복원 또는 인식 중 하나에 집중하고 있으며, 두 작업을 유기적으로 결합하려는 시도는 제한적이다. 그러나 실시간 처리가 필수적인 복잡한 군사 영상 환경에서는 객체 복원과 인식이 긴밀하게 연계되어야 하며, 이를 통해 높은 처리 속도와 정확도를 동시에 만족할 수 있는 기술이 필요하다.

앞선 분석에 착안하여, 본 논문에서는 DSTT 기반 복원 모델과 RT-DETR 기반 인식 모델을 통합하여, 장애물에 의해 가려진 전장 객체를 복원하고 인식하는 영상 복원-인식 통합 시스템을 제안한다. 제안된 시스템은 전장 가상 환경 시뮬레이터로 생성된 EO/IR 영상과 시멘틱 분할 완료된 이미지를 기반으로 장애물 마스크 데이터를 생성한다. 생성된 장애물 마스크 데이터와 EO/IR 영상을 복원 모듈을 통해 장애물에 가려진 전장 객체를 복원하여 복원 완료된 이미지를 생성한다. 또한, 실시간으로 복원하는 환경을 고려하여 기존의 복원 모델과 다르게 현재 프레임을 기준으로 과거의 프레임에 대한 정보만 복원 과정에서 사용한다. 인식 모듈은 복원 완료된 이미지를 입력받고 이미지에 존재하는 전장 객체를 인식하여 결과를 출력한다.

본 논문의 주요 기여는 다음과 같다.

- 본 논문에서는 장애물에 의해 가려진 전장 객체를 인식하기 위한 영상 복원-인식 통합 시스템을 제안하며, 원본 영상 대비 향상된 인식 정확도와 20 fps 이상의 추론 속도라는 실시간 지향의 목표 성능을

달성하는 것을 주요 목표로 한다. 제안된 시스템은 DSTT 기반의 복원 모듈과 RT-DETR 기반의 인식 모듈로 구성되며, 복원 모듈은 현재 프레임과 과거 프레임만을 활용하여 계산량을 줄이고, 온라인 방식으로 복원 수행이 가능하다. 또한, 어텐션 연산의 효율성을 높이기 위해 메모리 뱅크 기법을 적용하여 중복 연산을 최소화한다. 인식 모델은 자체 구축한 데이터 세트로 전이 학습한 후, TensorRT를 이용하여 최적화하고 높은 연산 효율을 확보한다. 이러한 구성은 객체가 가려진 상황에서도 원본 영상 대비 향상된 인식 정확도를 달성함과 동시에, 실시간 응용이 가능한 수준의 추론 속도를 가능하게 한다.

- 전장 가상 환경 시뮬레이터를 이용하여 사전에 정의한 전장 상황 시나리오를 기반으로 전장 객체에 대한 데이터 세트를 구축한다. 인식 모델은 구축한 데이터 세트를 이용하여 전이 학습을 한 이후, TensorRT를 적용한다. 복원 모델은 구축한 데이터 세트와 YouTube-VOS, DAVIS와 같은 공용 데이터 세트를 이용하여 처음부터 학습한다. 구축한 데이터 세트를 기반으로 학습한 모델들을 이용하여 가려진 전장 객체를 복원하고, 인식 결과를 출력한다.
- 제안된 시스템은 전장 가상 환경 시뮬레이터에서 생성한 EO/IR 영상을 이용하여 원본 영상과 복원 영상의 인식 정확도를 비교 분석한다. 또한, 실시간 응용이 가능한 수준으로 추론되는 것을 보장하기 위해, 본 논문에서는 사용한 복원 모델과 다른 복원 모델을 비교 분석하고 인식 모듈의 추론 속도를 분석하여 제안된 시스템 설계 타당성을 검증한다.

Ⅱ. 관련 연구

2.1 DSTT

DSTT는 영상 인페인팅을 위한 트랜스포머 기반 모델로, 실시간 처리에 유리하도록 설계된 구조를 갖는다. 모델의 전체 구조는 그림 1에서 확인할 수 있다. 기존 STTN이 공간과 시간 전파를 동시에 처리하는 반면, DSTT는 이를 분리하여 각각 독립적인 트랜스포머 블록으로 구현함으로써 계산 효율성을 높이고 보다 정교한 시공간 정보 전파를 가능하게 한다. 시간 전파 블록은 프레임 간 객체 움직임을, 공간 전파 블록은 정적 배경 텍스처를 각각 모델링하며, 두 블록은 교차 스택되어 전체 시공간 상호작용을 효과적으로 반영한다. 이러한 시공간 분리 구조는 멀티 헤드 셀프 어텐션

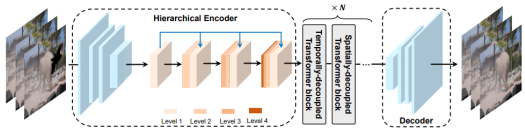

그림 1. DSTT 모델 아키텍처 개요
Fig. 1. DSTT architecture overview

(multi-head self-attension, MSA)의 계산 복잡도도 크게 줄이는데, 기존의 $O(t^2s^4n^2c)$에서 $O(t^2s^2n^2c)$와 $O(ts^4n^2c)$로 나누어 처리된다. 여기서, t는 토큰 수, s는 프레임 수, n는 이미지 패치 수, c는 특징 맵의 채널 수이다. DSTT의 이러한 설계는 실험 결과에서 높은 성능을 보이며 실시간 적용 가능성을 증대시킨다.

또한, DSTT는 계층적 인코더 구조를 도입해 다양한 수준의 공간 정보를 통합한다. 입력 이미지를 다단계 컨볼루션으로 처리해 생성된 특징 맵은 채널 방향으로 결합되며, 이를 통해 작은 디테일부터 큰 패턴까지 포괄하는 종합적인 복원이 가능하다. 이러한 설계는 복잡한 시각 정보를 손실 없이 학습할 수 있도록 하며, 실제 실험에서도 높은 성능과 빠른 처리 속도를 동시에 달성함을 입증하였다.

한편, 트랜스포머 기반 복원 모델인 DSTT는 우수한 복원 성능을 보장하지만, 복원 과정에서 각 프레임마다 계산 복잡도가 높은 어텐션 연산을 수행하므로, 많은 계산 자원을 요구하며 이로 인해 추론 속도가 저하될 수 있다. 또한, DSTT는 기존 복원 모델들과 마찬가지로 복원 성능 향상을 위해 현재 프레임뿐만 아니라 이전 및 이후 프레임까지 모두 활용한다. 그러나, 이후 프레임은 실시간 처리 환경에서는 활용할 수 없으므로, DSTT를 실시간 운용 시스템에 적용하는 데에는 어려움이 있다.

2.2 RT-DETR

[11]에서 제안된 RT-DETR 모델은 어텐션 메커니즘을 기반으로 한 실시간 end-to-end 객체 검출기이다. RT-DETR 모델은 백본과 효율적인 하이브리드 인코더(efficient hybrid encoder)와 보조 예측 헤더 (auxiliary prediction header)를 가진 디코더로 구성된다. 모델 아키텍처의 전체적인 구성은 그림 2에서 확인할 수 있다.

이전 연구[12]에서 DETR 모델의 학습 수렴 속도와 성능을 향상하기 위해 멀티스케일 특징을 도입하였지만, 입력 시퀀스의 길이가 증가하면 여전히 인코더의 계산 병목 현상이 발생하고 DETR의 실시간성을 보장하지 못했다. RT-DETR에서는 이전 연구의 문제를 해결하기 위해 스케일 내 상호작용과 스케일 간 융합을

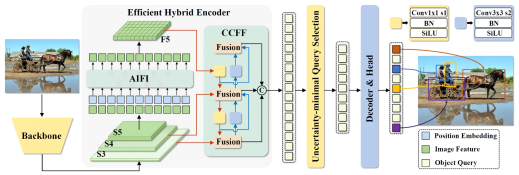

그림 2. RT-DETR 모델 아키텍처 개요
Fig. 2. RT-DETR architecture overview

분리하여 AIFI (attention-based intra-scale feature interaction) 모듈과 CCFF (CNN-based cross-scale feature fusion) 모듈로 구성된 효율적인 하이브리드 인코더를 제안했다. 제안된 인코더는 멀티스케일 특징을 처리하여 일련의 이미지 특징으로 변환한다. 또한, 불확실성 최소화 쿼리 선택 (uncertainty-minimal query selection)은 객체 탐지에서 최적의 쿼리 선택을 통해 정확도를 향상하는 방법으로서 분류 점수와 IoU (intersection over union) 점수를 같이 고려하여 인코더의 출력 중에서 불확실성이 적고 높은 신뢰도를 가진 특징을 선별해 디코더의 초기 쿼리로 사용한다. 마지막으로 RT-DETR은 재학습을 하지 않고 유연하게 추론 속도를 조정할 수 있으며, NMS 임곗값을 설정하는 복잡함이 없으므로 실용성도 높다는 특징이 있다. 또한, TensorRT를 이용하여 사전 학습된 RT-DETR 모델을 최적화하여 추론 속도를 향상했다.

Ⅲ. 모델 설계 및 데이터 세트 구축

3.1 영상 복원-인식 통합 시스템

본 논문에서 제안하는 영상 복원-인식 통합 시스템의 설계 목표는 장애물에 가려진 전장 객체 원본 영상 대비 향상된 인식 정확도와 20 fps 이상의 추론 속도를 동시에 달성하는 것이다. 그림 3에서는 제안한 시스템의 전체적인 구조를 확인할 수 있다. 시스템은 원본 동영상과 시멘틱 정보를 입력으로 받고, 전장 객체 복원 모듈을 통해 장애물에 가려진 전장 객체를 복원한다. 이후, 인식 모듈에서는 복원이 완료된 프레임을 바탕으로 전장 객체를 인식하는 과정을 수행한다.

객체 복원 단계에서는 입력받은 원본 영상과 시멘틱 정보를 활용하여 가려진 영역을 복원한다. 본 논문에서는 나무를 장애물로 인식하고 해당 영역에 대한 마스크를 생성한다. 장애물 마스크는 영상 내 가려진 영역을 식별하고 해당 픽셀을 복원할 때 사용된다. 또한, 복원 모듈에는 추론 속도 향상을 위해 온라인 복원 기법과 메모리 뱅크 기법을 적용한다.

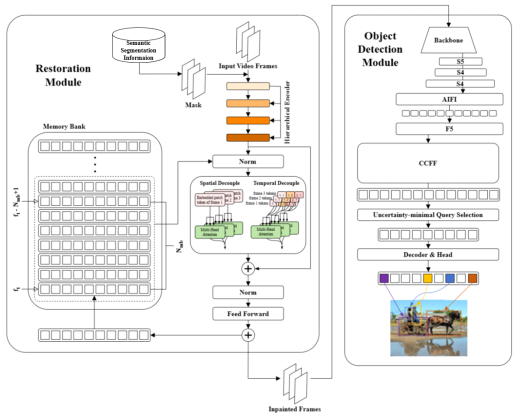

그림 3. 제안된 통합 시스템 아키텍처 개요
Fig. 3. Proposed integrated system architecture overview

3.1.1 온라인 복원 기법

기존 복원 모델들은 복원 성능을 높이기 위해 현재 프레임과 이전 프레임뿐만 아니라 이후 프레임까지 활용하는 방식이 일반적이다. 그러나, 이러한 방식은 실시간 처리에 필요한 낮은 지연 시간을 확보하기 어려울 뿐만 아니라, 본 논문에서 고려하는 전장 환경과 같은 실제 운용 상황에서는 이후 프레임에 접근이 불가능하므로 적용에 한계가 있다.

따라서, 본 논문에서는 이러한 전장 환경을 고려하여, 이후 프레임은 사용하지 않고, 현재 프레임과 이전 프레임만을 사용하는 온라인 복원 기법을 제안한다. 이 방식은 프레임 간의 시공간적 연속성은 활용하며, 실제 운용 환경에서 이후 프레임 즉, 미래 프레임에 접근할 수 없는 온라인 처리 조건을 만족시키며, 동시에 연산 지연을 최소화할 수 있는 장점이 있다.

3.1.2 메모리 뱅크 기법

본 논문의 주요 설계 목표 달성을 위해 제안된 통합 시스템의 복원 모듈은 트랜스포머 기반의 어텐션 구조로 설계되었으며, 이는 시공간적 정보를 효과적으로 활용하여 정확한 객체 복원을 가능하게 한다. 그러나, 어텐션 연산은 높은 계산 복잡도를 가지며, 복원 과정에서 반복적으로 수행되면 추론 속도 저하로 이어질 수 있다.

이러한 문제를 해결하고, 빠른 복원 추론 속도를 확보하기 위해, 본 논문에서는 메모리 뱅크 기법을 제안한다. 메모리 뱅크는 이전에 계산된 어텐션 값을 저장하고, 이후 복원 시 해당 값을 재사용함으로써 중복 연산을 제거하고 전체 연산량을 효율적으로 감소시킨다. 이를 통해 복원 모듈의 추론 속도를 개선하고, 실시간 처리에 적합한 연산 구조를 구현할 수 있다.

또한, 메모리 자원의 효율적인 사용을 위해, 일정 time step 이후 사용되지 않는 어텐션 값은 메모리에서 해제되도록 하여, 제한된 하드웨어 자원 환경에서도 안정적인 시스템 운용이 가능하여지도록 하였다.

객체 인식 단계에서는 입력받은 복원 완료된 프레임을 이용하여 백본 네트워크를 통해 멀티스케일 특징을 얻는다. 앞서 얻은 멀티스케일 특징은 효율적인 하이브리드 인코더를 통해 일련의 이미지 특징들로 변환된다. 이후, 불확실성 최소화 쿼리 선택을 통해 분류 점수와 IoU 점수를 함께 고려하여 불확실성이 낮은 특징을 선택하여 디코더에 입력하여 전장 객체를 인식하고, 인식된 객체는 경계 상자 (bounding box) 및 클래스 정보와 함께 생성된다. 구축한 데이터 세트를 통해 전이 학습을 진행하였고, 빠른 추론 속도를 위해 TensorRT 최적화를 적용하여 인식 모듈의 속도와 효율성을 향상했다.

3.2 전장 환경 데이터 세트 구축

실제 전장 환경에 대한 데이터 세트는 수량이 적고 접근이 제한적이며, 다양한 상황을 포괄하기 어렵기 때문에 본 연구의 목적에 적합하지 않은 경우가 많다. 이에 따라, 본 연구에서는 전장 가상 환경 시뮬레이터를 활용하여 EO/IR 기반의 전장 영상을 직접 수집하였다.

수집된 영상은 장애물의 유무에 따라 구성되며, 장애물은 나무로 제한하였다. 전장 객체는 고정고사포, 장갑차, 곡사포, 방사포, 군인, 탱크로 구성하였으며, 무인기 (unmanned aerial vehicle) 촬영 영상과 유사한 형태를 구현하기 위해, 시뮬레이션 환경에서 시점을 지면과 약 45도의 각도로 설정하여 영상을 획득하였다. 해당 시점 설정은 무인기 비행 중 촬영 시야를 모사하기 위함이다. 4.2 절에서 구축한 데이터 세트의 자세한 내용을 확인할 수 있다.

Ⅳ. 실험 및 결과

4.1 실험 환경

실험은 다음과 같은 하드웨어 및 소프트웨어 환경에서 진행되었다. 하드웨어 환경은 AMD EPYC 7763 CPU, NVIDIA A100 80GB GPU, 512GB RAM으로 구성되었으며, 소프트웨어 환경은 Ubuntu 20.04 운영체제와 Python 3.10.12, PyTorch 2.3.0을 사용하였다. 또한, NVIDIA CUDA 12.5.1과 NVIDIA TensorRT 10.2.0.19를 포함하여 다양한 딥러닝 및 비디오 처리 라이브러리(cython, opencv-python 4.8.0.74, lapx, pycuda, filterpy, portalocker, fvcore, omegaconf)를 활용하였다.

4.2 데이터 세트 구성

실험에 사용된 학습 데이터는 전장 가상 환경 시뮬레이터를 통해 수집된 전장 객체 6종 데이터와 기타 복원 및 인식 모델을 위한 다양한 영상 데이터를 포함한다. 표 1은 시뮬레이터에서 수집된 데이터의 세부 정보를 나타내며, 표 2와 표 3은 각각 복원 모델인 DSTT와 인식 모델인 RT-DETR 학습에 사용된 데이터 세트 구성을 보여준다.

본 연구에서 이용하는 데이터 세트는 크게 두 가지가

표 1. VR-Forces로 수집한 전장 객체 6종 학습 데이터
Table 1. Learning data for 6 types of battlefield objects collected with VR-Forces

전장 객체	세부 객체	EO 데이터	IR 데이터
고정 고사포	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 데이터 세트 구성
Table 2. Composition of the data set used to train the restoration model

데이터 세트	EO 데이터	IR 데이터
VR-Forces	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

표 3. 사용된 인식 모델 학습 데이터 세트 구성
Table 3. Composition of the data set used to train the recognition model

전장 객체	세부 종류	EO 데이터		IR 데이터	
		복원 전	복원 후	복원 전	복원 후
고정고사포	ZPU-4	1262	840	1083	858
장갑차	BTR-80	1325	568	1194	693
곡사포	MO-120RT	751	514	735	523
방사포	SA-6, M901	858	609	843	668
군인	SA7, PK74	1099	887	1566	938
탱크	T-69, T-72	1318	758	1193	721
총계		6613	4176	6614	4401

있다. 하나는 시뮬레이터로 수집한 데이터이고, 다른 하나는 복원 성능을 올리기 위한 추가 데이터 세트이다.

먼저, 전장 가상 환경 시뮬레이터를 이용하여 수집한 데이터는 표 1과 같다. 전장 객체는 총 6종으로 고정고사포, 장갑차, 곡사포, 방사포, 군인, 탱크로 구성되어 있으며, 세부 객체로는 ZPU-4, BTR-80, MO-120RT, M901, SA-6, SA7-prone, SA7-standing, PK74-prone, PK74-standing, T-69, T-72로 총 11종이다. 여기서 SA7-prone과 PK74-prone은 해당 무기를 소지하고 엎드려 있는 상태의 군인들을 의미하고, SA7-standing과 PK74-standing은 해당 무기를 소지하고 일어서 있는 상태의 군인들을 의미한다. 각각의 객체들은 장애물에 가려진 상황과 가려지지 않은 상황을 고려하고, 실제 무인기 비행 중 촬영 시야를 반영하여 지면과의 약 45도 각도와 다양한 거리에서 영상을 수집하였다. 앞선 조건들을 고려하여 EO 영상과 IR 영상에서 각각 세부 객체 별로 1,000개 이상의 샘플을 얻었다.

다음으로, 복원 모델인 DSTT를 학습하기 위해 시뮬레이터를 통해 수집한 데이터 외에도 YouTube-VOS 및 DAVIS와 같은 공개 데이터 세트를 추가로 활용하였다. 도로 차량 영상, 드론 자율항법 영상 등 실제 무인기 기반 정찰 시야와 유사한 데이터를 포함하여, 모델이 복잡한 상황에서도 손상된 영역을 효과적으로 복원할 수 있도록 다양한 장면에서 수집된 시각적 데이터를 학습에 사용하였다. 복원 모델 학습에는 총 916,814장의 EO 영상과 총 147,705장의 IR 영상이 사용되었다.

한편, 인식 모델인 RT-DETR의 학습에는 복원 전과 복원 후의 영상을 모두 포함한 데이터 세트가 구성되었다. 동일한 전장 객체에 대해 복원 전후 데이터를 병렬로 구성하여, 복원 처리가 인식 정확도에 어떤 영향을 주는지 평가할 수 있도록 하였다. 또한, 각 영상에 대한 데이터 주석 (data annotation)은 경계 상자 방법을 이용하여 전장 객체 클래스 정보와 경계 상자 좌표 정보, 경계 상자 넓이 값 등을 작성하였다. 여기서, 경계 상자 좌표 정보는 경계 상자의 좌상단 x값과 y값, 너비 값, 높이 값이 작성되었다.

이처럼 다양한 환경과 조건에서 수집된 영상 데이터를 활용하여, 본 연구는 복원과 인식 모델 모두에 대해 강인하고 일반화된 성능을 확보할 수 있도록 하였다.

4.3 객체 복원 실험

객체 복원 모듈의 성능을 평가하기 위해 본 논문에서 제안하는 시스템의 DSTT*의 복원 결과를 기존의 영상 복원 기법들과 비교하였다. 실험은 다양한 전장 환경을 시뮬레이션하여 복잡한 배경과 나무 장애물이 존재하는 조건에서 복원 성능을 분석하는 방식으로 진행되었다. 표 4는 이 실험을 정량적으로 분석한 결과를 나타낸다.

본 논문에서는 기존의 대표적인 영상 복원 기법들과 DSTT*의 성능을 비교하였다. 비교 대상으로는 FcF (fourier coarse-to-fine inpainting)[13], E2FGVI (end-to-end framework for flow-guided video

표 4. 전장 객체 복원 모델 성능 비교
Table 4. Comparison of performance of battlefield object restoration models

복원 모델	SSIM ↑	PSNR ↑	LPIPS ↓	fps ↑
DSTT[1)]	0.882	24.012	0.213	110.07
FcF[13]	0.939	31.558	0.069	12.36
E2FGVI [14]	0.952	34.580	0.051	7.98
STTN[10]	0.881	26.748	0.275	9.06

1) 기존 DSTT에 온라인 복원 기법과 메모리 뱅크 기법을 적용한 모델

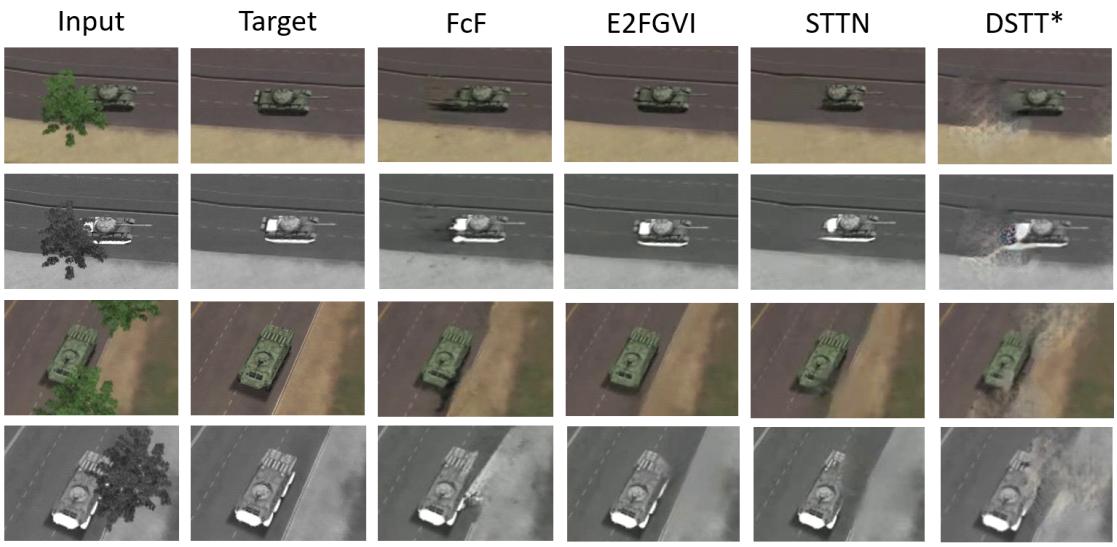

그림 4. 전장 객체 영상 복원 결과의 시각적 비교
Fig. 4. Visual comparison of battlefield object image restoration results

inpainting)[14] 및 STTN을 선정하였다.

복원 성능은 SSIM[15], PSNR, LPIPS[16]의 세 가지 복원 품질 지표와 함께, 실시간 확보를 위한 모델의 처리 속도(fps) 또한 주요 성능 지표로 포함하여 정량적으로 평가하였다. 빠른 추론 속도를 얻기 위해 온라인 복원 기법과 메모리 뱅크 기법을 적용한 DSTT* 모델은 110.07 fps의 빠른 처리 속도를 기록하며, 현재 프레임과 이전 프레임만을 사용하는 구조에도 불구하고, SSIM은 0.882로 준수한 성능을 보였다.

한편, 가장 높은 SSIM 0.952, PSNR 34.680, LPIPS 0.051을 기록한 E2FGV는 복원 품질 측면에서는 우수한 성능을 보였으나, 처리 속도는 7.98 fps로 낮아 실시간 처리 기준인 30 fps를 충족하지 못하였다. FcF 또한 SSIM 0.939, PSNR 31.558, LPIPS 0.069로 우수한 복원 품질을 기록하였지만, 12.36 fps에 그쳐 실시간 처리에는 제약이 있었다. 반면, STTN은 모든 복원 품질 지표에서 온라인 복원 기법과 메모리 뱅크 기법을 적용한 DSTT*와 비슷한 성능을 기록하였지만, 처리 속도가 9.06 fps에 불과하여 전체적으로 열세를 보였다.

앞선 분석 결과를 통해 온라인 복원 기법과 메모리 뱅크 기법을 적용한 DSTT*는 높은 SSIM을 유지하면서도 빠른 처리 속도를 가지기 때문에 실시간 영상 복원에 적합함을 입증하였다.

또한, 그림 4를 통해 복원 모델들의 복원 결과를 확인할 수 있으며, 위에서부터 순서대로 탱크 EO, 탱크 IR, 장갑차 EO, 장갑차 IR 영상의 복원 결과를 보여준다. 분석 결과, EO 영상이 IR 영상보다 복원된 영역의 번짐이 적게 나타났으며, 이는 EO 영상에서 전장 객체와 배경 간의 색 차이가 더 크기 때문으로 해석된다. E2FGVI는 Target 이미지와 가장 유사한 복원 결과를 보여주며 가장 우수한 시각 품질을 나타냈다. 반면, STTN은 탱크 EO 영상과 탱크 IR 영상, 장갑차 IR 영상에서 전장 객체의 일부가 소실되는 품질 저하가 관찰되었다. 제안된 DSTT*는 E2FGVI에 비해 세부 질감 복원은 부족했으나, STTN과 달리 전장 객체의 전체적인 형태와 경계가 안정적으로 유지되었으며, 배경과 객체 간의 구분이 명확하게 확인되었다. 특히, 복원된 영상에서 객체의 외곽선이 유지되면서도 배경과 자연스럽게 연결되어, 후속 객체 인식 단계에서 검출이 용이한 형태로 복원이 이루어졌다는 점이 확인되었다.

실험 결과 제안된 시스템의 DSTT*는 프레임 간의 시공간적 연속성을 유지하면서도 배경과 객체의 세부 사항을 효과적으로 복원하는 것을 확인하였다. 분석된 다른 모델들과 비교했을 때 가장 빠른 처리 속도를 기록하면서도 일정 수준의 복원 성능을 유지하였다는 점에서, 온라인 복원 기법과 메모리 뱅크 기법을 적용한 DSTT*는 본 논문에서 제안하는 통합 시스템의 실시간 영상 복원에 가장 적합한 모델로 판단하였다.

4.4 복원 모듈 소거 실험

제안하는 복원 모듈에서 온라인 복원 기법과 메모리 뱅크 기법이 각각 성능에 미치는 영향을 분석하기 위해 소거 실험을 수행하였다. 기준 모델 (baseline model)은

표 5. 복원 모듈 소거 실험 결과
Table 5. The results of the ablation study on restoration modules

기준 모델	온라인 기법	메모리 뱅크	SSIM ↑	PSNR ↑	fps ↑
✓	×	×	0.896	33.635	12.93
✓	✓	×	0.895	25.542	27.57
✓	✓	✓	0.882	24.012	110.07

두 기법을 모두 적용하지 않은 DSTT로 설정하였으며, 이를 기준으로 각 기법의 도입 여부에 따른 복원 품질과 처리 속도를 비교하였다. 표 5는 그 결과를 정량적으로 나타낸다.

기준 모델은 SSIM 0.896, PSNR 33.635를 기록하며 복원 품질 면에서는 준수한 성능을 보였으나, 처리 속도는 12.93 fps로 낮아 실시간 복원에는 적합하지 않았다. 온라인 기법만 적용한 경우, SSIM과 PSNR은 각각 0.895, 25.542로 기준 모델에 비해 소폭 감소하였고, 처리 속도는 27.57 fps로 향상되었지만, 여전히 실시간 처리 기준인 30 fps에는 미치지 못하였다.

반면, 두 기법을 모두 적용한 모델에서는 SSIM과 PSNR이 각각 0.882, 24.012로 다소 낮아졌지만, fps는 110.07로 대폭 상승하여 실시간 처리가 가능해졌다. 이러한 결과는 제안하는 온라인 복원 기법과 메모리 뱅크 기법이 실시간성 확보에 각각 기여하며, 두 기법을 병행하여 적용했을 때 시너지 효과를 발휘해 실시간 처리 조건을 만족시키는 최적의 성능을 구현함을 보여준다. 온라인 기법은 미래 프레임 의존성을 제거하면서도 시공간적 연속성을 유지하여 프레임 간 일관성을 확보하고, 메모리 뱅크 기법은 반복되는 어텐션 연산의 중복을 제거해 연산 효율성을 크게 향상시켰다. 이러한 두 기법의 결합은 품질 저하라는 대가가 다소 있긴 하나, 단독 적용 시보다 훨씬 높은 처리 속도를 확보해 실시간 처리가 필요한 실제 운용 환경에서 복원 품질과 속도의 균형을 달성하는 데 기여한다는 점에서 의의가 크다.

4.5 객체 인식 실험

표 6은 본 논문에서 사용하는 객체 인식 모듈인 RT-DETR, YOLOv7[17], DINO[18], Deformable-DETR[12]을 비교한 결과를 나타낸다. 인식 성능은 정량적 지표인 mAP와 fps를 기준으로 평가하였다. Deformable-DETR은 다른 모델들과 비교했을 때, EO 영상에서 0.773 mAP, IR 영상에서 0.763 mAP로 가장 낮은 인식 정확도를 보였다. DINO는 Deformable-DETR 보다 높은 mAP를 기록하였으나, EO 영상에서 11.2 fps, IR 영상에서 11.4 fps로 가장 낮은 추론 속도를 나타냈으며, RT-DETR과 YOLOv7에 비해 10배 이상의 차이를 보였다. 이러한 결과는 Deformable-DETR과 DINO 두 모델 모두 실시간 처리 기준인 30 fps를 만족하지 못해, 실시간 응용에 제약이 존재한다. 한편, YOLOv7과 RT-DETR은 mAP 측면에서 유사한 성능을 보였으나, EO 영상에서 RT-DETR이 더 높은 인식 정확도를 기록하였고, 추론 속도는 YOLOv7이 더 우수하였다.

표 6. 전장 객체 인식 모델 성능 비교
Table 6. Comparison of performance of battlefield object recognition models

인식 모델	mAP (0.5:0.95)	mAP (0.5)	mAP (0.75)	fps
DINO-EO	0.780	0.993	0.905	11.2
YOLOv7-EO	0.816	0.993	0.936	329.0
RT-DETR-EO	0.821	0.992	0.940	229.7
Deformable-DETR-EO	0.773	0.988	0.900	27.1
DINO-IR	0.780	0.990	0.925	11.4
YOLOv7-IR	0.808	0.995	0.947	293.4
RT-DETR-IR	0.806	0.994	0.944	229.0
Deformable-DETR-IR	0.763	0.990	0.912	27.2

본 논문은 복잡한 전장 환경에서의 운용을 고려한 영상 복원-인식 통합 시스템을 제안하고 있으며, 이러한 환경에서는 인식 정확도와 추론 속도 간의 균형이 중요하다. 앞선 비교 분석을 통해, 트랜스포머 기반의 RT-DETR이 높은 인식 정확도와 빠른 추론 속도를 동시에 만족하며, 본 논문의 시스템에 적합한 인식 모듈임을 실증적으로 보여준다.

또한, 객체 인식 모듈은 원본 영상과 복원된 영상을 기반으로 RT-DETR 기반의 인식 성능을 평가하였다. 실험은 다양한 전장 객체와 나무 장애물이 포함된 EO/IR 영상에서 수행되었으며, 인식 정확도와 실시간 처리 속도를 분석하였다. 전장 객체 인식 실험에 관한 결과는 표 7에 나타내었다.

인식 성능은 mAP를 통해 정량적으로 평가되었다. 복원된 EO 영상에서 mAP(0.5:0.95)는 54.32%, mAP(0.5)는 29.50%, mAP(0.75)는 53.59%의 성능 향상을 확인하였다. 복원된 IR 영상에서의 mAP(0.5:0.95)는 15.80%, mAP(0.5)는 6.65%, mAP(0.75)는 18.30%의 성능 향상을 확인하였다. 복원된 EO/IR 영상이 원본 EO/IR 영상과 비교를 통해 인식 성능이 전체적으로 향상되는 것을 확인하였다.

또한, TensorRT를 이용한 모델 최적화로 인해 처리

표 7. 전장 객체 인식 실험 결과 (mAP)
Table 7. Battlefield object recognition experiment results (mAP)

평가지표	mAP (0.5:0.95)	mAP (0.5)	mAP (0.75)
원본 EO	0.532	0.766	0.612
복원 EO	0.821	0.992	0.940
원본 IR	0.696	0.932	0.798
복원 IR	0.806	0.994	0.944

속도가 32.3 fps에서 229.7 fps로 단축되었으며, 이를 통해 실시간 객체 인식의 효율성이 크게 향상했다.

앞선 인식 성능 분석과 처리 속도 향상을 통해 RT-DETR은 본 논문에서 제안한 복원-인식 통합 시스템의 실시간 객체 인식에 가장 적합한 모델로 판단하였다.

4.6 결과 분석

앞선 실험을 통해 본 논문에서 제안한 복원-인식 통합 시스템의 성능을 평가하였다. DSTT 기반의 복원 모듈은 자체 구축한 데이터 세트로 처음부터 학습시켰으며, 온라인 복원 기법과 메모리 뱅크를 적용하여 처리 속도는 110.07 fps가 측정되었고, 평균 SSIM 0.8814를 달성하였다. 이는 시공간적으로 연속성을 가지고 우수한 품질로 복원할 수 있으며, 매우 빠른 속도로 복원 추론이 수행될 수 있음을 의미한다. 인식 모듈은 RT-DETR을 사용하여 자체 구축한 데이터 세트로 전이 학습을 진행하였다. 또한 TensorRT 기반 최적화를 적용하여 처리 속도가 229.7 fps로 크게 향상되었으며, 복원된 영상에 대해 EO/IR 영상 모두에서 mAP 향상을 기록하여 인식 정확도 역시 개선된 것을 확인하였다. 특히, 제안하는 시스템은 복원 모듈과 인식 모듈의 각 처리 속도만 측정하는 것이 아닌, 전체 시스템 기준으로 실측한 결과 30.67 fps로 측정되었다. 이는, 실시간 처리 기준인 30 fps를 초과 달성한 결과로, 제안한 시스템이 실시간성을 안정적으로 보장할 수 있음을 입증하였다.

종합적으로, 본 논문에서는 향상된 인식 정확도와 20 fps 이상의 추론 속도라는 주요 목표를 모두 달성하였으며, 특히 추론 속도는 목표를 초과 달성하여 실시간성을 보장하였다.

V. 결 론

본 논문에서는 장애물에 의해 가려진 전장 객체를 복원하고 인식하는 복원-인식 통합 시스템을 제안하였다. 이를 위해 전장 가상 환경 시뮬레이터와 다양한 공개 데이터 세트를 활용하여 전장 환경 데이터 세트를 구축하였고, 이를 기반으로 복원 모델과 인식 모델을 각각 학습시켰다. 제안한 시스템의 복원 모듈은 온라인 복원 기법과 메모리 뱅크를 적용하여 기존 복원 기법 대비 시공간적 연속성을 가지며 빠른 처리 속도를 달성하였다. 또한, 인식 모듈은 복원된 EO/IR 영상에 대해 mAP 성능이 향상됨을 확인하였고, TensorRT 최적화를 적용함으로써 빠른 추론 속도를 확보하였다. 마지막으로, 전체 시스템의 추론 속도 결과가 실시간 처리 기준을 초과 달성하여 실시간성을 안정적으로 보장할 수 있음을 입증하였다. 실험 결과들을 통해 본 논문의 주요 목표를 모두 달성하여 제안한 복원-인식 통합 시스템의 타당성을 검증하였다.

향후 연구에서는 통합 최적화를 통해 전체 시스템의 추론 속도를 개선하여 실시간성을 더욱 향상할 예정이다. 또한, 특정 복잡한 장면에서 성능 저하가 발생할 가능성이 있으므로, 이러한 한계를 보완할 수 있는 개선 방안을 모색할 계획이다. 한편, 현재 구축된 전장 환경 데이터 세트는 장애물에 가려진 전장 객체 상황에 집중되어 있어, 실제 다양한 전장 환경을 충분히 반영하기에는 한계가 있다. 이에 따라 향후 연구에서는 나무 외에도 다양한 유형의 장애물을 포함해 데이터의 일반화를 높이고, 전장 환경 조건을 세분화하여 데이터 세트를 보강할 예정이다. 아울러, 실제 EO/IR 센서에서 관측되는 노이즈, 해상도 변화, 대기 왜곡 등 다양한 열화 요소들도 고려함으로써, 실제 운용 환경을 반영한 데이터 세트로 확장할 계획이다. 이러한 보강된 데이터 기반의 추가 실험을 통해 시스템의 성능을 더욱 향상하고, 실 환경에서도 안정적으로 동작할 수 있도록 연구를 진행할 계획이다.

References

[1] C.-E. Lee, J. Baek, J. Son, and Y.-G. Ha, "Deep AI military staff: Cooperative battlefield situation awareness for commander's decision making," *J. Supercomputing*, vol. 79, no. 6, pp. 6040-6069, 2023. (https://doi.org/10.1007/s11227-022-04882-w)

[2] M. Alzboon, M. Alqaraleh, and M. S. Al-Batah, "AI in the sky: Developing real-time UAV recognition systems to enhance military security," *Data Metadata*, vol. 3, p. 417, 2024.

[3] H. Zhang, K. Liu, Z. Gan, and G.-N. Zhu, "UAV-DETR: Efficient end-to-end object

detection for unmanned aerial vehicle imagery," *arXiv preprint arXiv:2501.01855*, 2025.

[4] L. D. Bella, Y. Lyu, B. Cornelis, and A. Munteanu, "HybridTrack: A hybrid approach for robust multi-object tracking," *arXiv preprint arXiv:2501.01275*, 2025.

[5] W. Quan, J. Chen, Y. Liu, D. M. Yan, and P. Wonka, "Deep learning-based image and video inpainting: A survey," *IJCV*, pp. 1-34, 2024.

[6] Y. Cui, W. Ren, and A. Knoll, "Omni-kernel network for image restoration," in *AAAI*, vol. 38, no. 2, pp. 1426-1434, 2024.

[7] G. Zhan, W. Xie, and A. Zisserman, "A tri-layer plugin to improve occluded detection," *arXiv preprint arXiv:2210.10046*, 2022.

[8] S. Zhou, C. Li, K. C. Chan, and C. C. Loy, "ProPainter: Improving propagation and transformer for video inpainting," in *ICCV*, pp. 10477-10486, 2023.

[9] 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 preprint arXiv:2104.06637*, 2021.

[10] Y. Zeng, J. Fu, and H. Chao, "Learning joint spatial-temporal transformations for video inpainting," in *ECCV*, pp. 528-543, 2020.

[11] Y. Zhao, W. Lv, S. Xu, J. Wei, G. Wang, Q. Dang, Y. Liu, and J. Chen, "DETRs beat YOLOs on real-time object detection," in *CVPR*, pp. 16965-16974, 2024.

[12] X. Zhu, W. Su, L. Lu, B. Li, X. Wang, and J. Dai, "Deformable DETR: Deformable transformers for end-to-end object detection," *arXiv preprint arXiv:2010.04159*, 2020.

[13] J. Jain, Y. Zhou, N. Yu, and H. Shi, "Keys to better image inpainting: Structure and texture go hand in hand," *arXiv preprint arXiv:2208.03382*, 2022.

[14] Z. Li, C.-Z. Lu, J. Qin, C.-L. Guo, and M.-M. Cheng, "Towards an end-to-end framework for flow-guided video inpainting," *arXiv preprint arXiv:2204.02663*, 2022.

[15] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600-612, 2004.

[16] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *CVPR*, pp. 586-595, 2018.

[17] C. Wang, A. Bochkovskiy, and H. M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," in *CVPR*, pp. 7464-7475, 2023.

[18] H. Zhang, F. Li, S. Liu, L, Zhang, H, Su, J. Zhu, L. M. Ni, and H. Shum, "DINO: DETR with improved denoising anchor boxes for end-to-end object detection," *arXiv preprint arXiv:2203.03605*, 2022.

[19] T. J. Han, Y. Cho, G. Kim, Y. Lee, and K. Lee, "An analysis of deep learning-based restoration performance for warfare objects occluded by obstacles," in *KICS*, pp. 1222-1223, 2025.

이 광 일 (Kwangil Lee)

1993년 2월 : 충남대학교 컴퓨터과학과 졸업
1996년 8월 : 충남대학교 컴퓨터과학과 석사 졸업
2001년 2월 : 충남대학교 컴퓨터과학과 박사 졸업
2000년 2월~2002년 1월 : 미국 국립표준과학기술원(NIST) 연구원
2002년 4월~2004년 8월 : 미국 매릴랜드대학교 연구원
2004년 9월~2006년 1월 : 미국 텍사스주립대학교 연구원
2006년 5월~2017년 2월 : 한국전자통신연구원 책임연구원
2017년 3월~현재 : 한국해양대학교 인공지능공학부 교수
2022년 5월~현재 : ㈜디매스타 대표이사
<관심분야> Military AI, 상황인지, 객체인식 및 복원, 생성형 AI, 자율운항선박
[ORCID:0000-0002-8307-9003]

김 재 환 (Jaehwan Kim)

2002년 2월 : 인하대학교 컴퓨터공학과 졸업
2005년 2월 : 포항공과대학교 컴퓨터공학과 석사 졸업
2006년~현재 : 한국전자통신연구원 책임연구원
<관심분야> Military AI, 머신러닝, 컴퓨터 비전
[ORCID:0000-0002-7320-8211]

이 창 은 (Chang-Eun Lee)

1996년 : 한양대학교 전자공학과 졸업
1998년 : 한양대학교 전자공학과 석사 졸업
2017년 : 충남대학교 정보통신학과 박사 졸업
1998년~1999년 : LG 산전 연구원
1999년~2001년 : 엘지오티스엘리베이터 연구원
2001년~현재 : 한국전자통신연구원 책임연구원
<관심분야> Military AI, 인공지능, 로봇, 상황인지
[ORCID:0000-0002-4711-1088]