

전장 상황 무인 시스템 상황인지 향상을 위한 Top-down 변환 기반 합성 데이터 생성 및 학습 기법

A Synthetic Data Generation and Learning Framework Using Top-down View Transformation for Situation Awareness Enhancement in Unmanned Battlefield Systems

김 예 린¹·백 재 욱²·전 성 우²·이 창 은[†]

Yerin Kim¹, Jaekuk Baek², Sungwoo Jun², Chang-Eun Lee[†]

Abstract: Military robots and unmanned systems require training data with diverse viewpoints for robust object detection and situational awareness. Top-down view images are difficult to acquire, resulting in limited viewpoint diversity. In this paper, we propose a pipeline that transforms ground view images into top-down view images while preserving structural consistency. The framework employs FLUX.2 with Low-Rank Adaptation (LoRA) for parameter-efficient viewpoint adaptation. Instead of updating the entire network, we apply LoRA only to the Cross-Attention layers, which are critical for viewpoint conditioning, while freezing the pretrained network. We demonstrate that this approach effectively performs viewpoint transformation in data-limited military domains using only a small-scale training dataset. In addition, we employ conditional generation with ground view inputs, prompts, and target images. Our approach adapts a large pretrained model using only small datasets to generate structurally consistent top-down synthetic images. Structural consistency analysis based on LightGlue correspondences and CLIPPER-based geometric filtering yields a maximum inlier ratio of 52.7%. Finally, training with the synthetic data improves the mean Average Precision (mAP) by 6.5 percentage points over the original-data-only baseline.

Keywords: Military Robot, Situation Awareness, Synthetic Data Generation, Multi-Robot Systems

1. 서 론

군사 환경에서 운용되는 무인 시스템은 제한된 관측 시간 내 복잡한 전장 장면을 신속히 해석하고 임무 수행을 위한 의사결정을 지원해야 한다. 지상·공중 등 멀티 시점 데이터를 획득하는

멀티로봇 시스템은 협업 기반 실시간 상황인지 구현을 위해 데이터 변환 및 통합이 필수적이다. 정찰·감시 임무에서 지상 로봇은 건물·수목에 가려진 근접 객체를 지상 시점에서, 공중 로봇은 동일 지역을 상부에서 광역으로 관측한다. 이때 두 로봇의

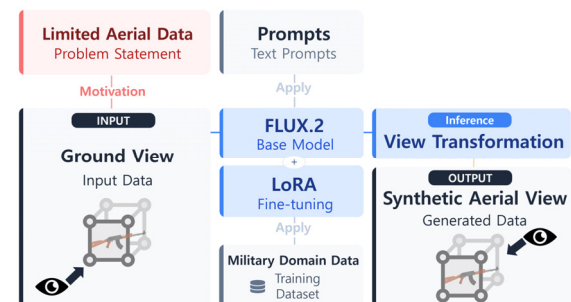
Received : Mar. 23. 2026; Revised : Jul. 6. 2026; Accepted : Jul. 10. 2026

※ This work was supported by Korea Research Institute for defense Technology planning and advancement (KRIT) grant funded by the Korea government (DAPA (Defense Acquisition Program Administration)) (No. 21-107-E00-009-02, "Realtime complex battlefield situation awareness technology").

1. Student Researcher, Electronics and Telecommunications Research Institute (ETRI); Artificial Intelligence Major, University of Science and Technology (UST), Daejeon, Korea (imyeslin@etri.re.kr)

2. Senior Researcher, Electronics and Telecommunications Research Institute (ETRI), Daejeon, Korea (jubaek, swjun@etri.re.kr)

† Professor, Corresponding author: Electronics and Telecommunications Research Institute (ETRI); Artificial Intelligence Major, University of Science and Technology (UST), Daejeon, Korea (celeee@etri.re.kr)



[Fig. 1] FLUX.2-LoRA based Military View Transformation Framework

관측 시점이 상이하여 동일 객체에 대한 탐지 결과를 직접 정합하기 어렵다. [Fig. 1]과 같이 지상-공중 시점 변환 파이프라인을 지상 로봇에 적용하면 지상로봇의 관측을 공중 로봇과 동일한 Top-down 표현 공간으로 변환하여 지상 로봇이 놓친 사각지대 정보를 상호 보완할 수 있다. 제안 기법은 단일 무인 시스템의 데이터 증강을 넘어 멀티로봇 간 시점 차이를 해소하는 협업 상황인지의 전처리 모듈로 활용될 수 있다.

이러한 상황인지 성능은 다양한 군사 작전 조건을 반영한 학습 데이터에 의존한다. 동일 객체에 대해서도 관측 시점 변화에 따라 기하학적 형태와 시각적 단서가 변화하므로 모델 학습을 위해서는 시점 다양성에 대한 표본화가 요구된다. 공중 시점 데이터는 지상 시점에 비해 확보가 어려워 학습 데이터의 지상 시점 편향을 초래한다. 최근 생성 모델을 활용한 합성 데이터 생성은 이러한 데이터 수급 제약을 보완하는 대안으로 주목받고 있다^[1,2]. 특히 군사 도메인은 위장·은폐 객체의 시점별 외형 변화 등 일반 도메인과 구별되는 고유한 제약이 존재하며 이러한 조건에서의 시점 변환 기반 합성 데이터 생성에 관한 연구는 상대적으로 부족하다^[3-5].

그러나 기존의 생성 모델 및 시점 변환 연구들은 실제 군사 환경에 적용하기에 다음과 같은 한계를 지닌다. 첫째, 관측 시점 변화에 수반되는 기하학적·구조적 변화를 충분히 모사하기 어렵다. 둘째, 군사 환경 데이터는 양이 제한적인 경우가 많아 대규모 생성모델을 full fine-tuning하는 접근은 비용 및 효율 측면에서 현실적 부담이 크다. 셋째, 생성된 합성 데이터가 후속 객체 탐지에 실제로 활용되기 위해서는 단순한 시각적 변환을 넘어 객체의 구조적 정보와 공간적 배치가 보존되어야 한다. 결론적으로, 학습 데이터의 시점 다양성을 확장하고 멀티로봇 간 시점 정합에 활용하기 위해 제한된 데이터만으로도 파라미터를 효율적으로 조정하여 지상 시점 이미지를 공중 시점으로 변환하는 방법론이 필수적이다.

본 논문의 기여는 다음과 같다.

- FLUX.2와 Low-Rank Adaptation (LoRA)를 결합한 지상-공중 시점 변환 파이프라인 설계: 제한된 도메인 데이터 환경에서도 모델을 효율적으로 미세조정 할 수 있는 최적의 학습 및 추론 전략을 제시한다.
- 생성 데이터의 유효성 검증 프레임워크 제안: 변환된 공중 시점 데이터가 실제 상황인지에 미치는 영향을 구조적 일관성 지표와 후속 객체 탐지 성능을 통해 평가한다.
- 멀티로봇 협업 기반 상황인지 시나리오 적용: 지상 로봇의 관측 데이터를 공중 시점으로 변환하고 이를 공중 로봇의 탐지 결과와 융합함으로써, 사각지대를 상호 보완하는 실시간 상황인지 멀티로봇 시나리오에서의 활용 가능성을 제시한다.

2. 관련 연구

2.1 2D Image-to-Image Translation and View Transformation-based Synthesis

Image-to-image 변환 기법에서 Pix2Pix는 paired 데이터를 기반으로 조건부 GAN을 적용하여 입력 영상과 출력 영상 간의 대응 관계를 학습하는 대표적 방법이다^[1]. CycleGAN은 paired 데이터 없이도 도메인 간 변환이 가능하도록 cycle-consistency를 도입함으로써 데이터 수집이 제한된 환경에서의 적용 가능성을 확대한다^[2]. 그러나 시점 변환에는 구조 보존과 공간적 일관성 확보가 추가로 요구된다. 그러나 기존 cross-view synthesis 연구는 주로 도시·위성 영상 등 일반 도메인에서 수만 장 규모의 학습 데이터로 진행되었다. 군사 도메인과 같이 극소량 데이터 환경에서의 적용 가능성을 검증한 연구는 아직 부족하다. 본 연구는 이러한 데이터 제약 환경에서 Cross-Attention 계층에만 선택적으로 LoRA를 적용하여 사전학습 모델의 일반 표현 능력은 유지하면서 시점 변환 조건만을 효율적으로 반영하는 전략을 제안한다.

2.2 2D-to-3D Novel View Synthesis

지상 시점에서 공중 시점에서의 변환은 기본적으로 2차원 영상 간 변환 문제에 해당하지만, 시점 변화의 폭이 크다는 점에서 3차원 기하 정보를 고려하는 연구들과도 관련된다. Neural Radiance Fields (NeRF) 계열 연구는 연속적인 radiance field를 학습함으로써 새로운 시점의 영상을 합성할 수 있는 가능성을 제시하였으며^[6] 이후 Instant-NGP는 해시 인코딩을 도입하여 학습 효율을 크게 향상시켰다^[7]. Zip-NeRF는 안티앨리어싱을 적용하여 렌더링 품질을 향상시키고 최근에는 3D Gaussian Splatting이 빠른 학습 속도와 실시간 렌더링 성능을 바탕으로 주목받고 있다^[8,9]. Zero-1-to-3는 diffusion 기반 프레임워크를 통해 단일 입력 영상으로부터 새로운 시점의 이미지를 생성하였고 LRM은 이를 3D 재구성으로 확장하였으며 Wonder3D는 cross-domain diffusion을 활용하여 텍스처 정보를 포함한 3차원 형상 생성을 제시한다^[10-12]. Diffusion은 단일 지상 이미지를 입력으로 사용하여 텍스트 유도를 바탕으로 공중 시점 이미지를 생성하는 방법을 제안하였으며 확산 모델이 Ground to Top-down 합성 문제에 적용될 수 있음을 보여준다^[3]. DreamFusion은 score distillation을 이용한 3차원 생성 방법을 제안한다^[13]. 이 연구들은 다중 시점 정보가 추가로 요구되는 경우가 많다. 이는 3D 시뮬레이션 환경에 의존하며 실제 영상의 시각적 특성을 반영하기 어렵다는 한계가 있다.

2.3 Parameter-Efficient Fine-Tuning

대규모 생성 모델을 특정 도메인에 적용하기 위해서는 일반적으로 많은 계산 자원과 학습 비용이 요구된다. LoRA는 사전학습 가중치를 고정한 상태에서 low-rank 행렬을 추가로 학습하는 방식으로 학습해야 하는 파라미터 수를 크게 줄이는 대표적인 Parameter-Efficient Fine-Tuning (PEFT) 기법이다^[14]. 제한된 데이터와 자원을 사용하는 환경에서도 비교적 효율적인 도메인 적응이 가능하다. 본 연구에서는 이러한 LoRA의 특성을 활용하여 제한된 군사 도메인 데이터만으로도 지상 시점 영상을 공중 시점 영상으로 변환할 수 있도록 생성 모델을 적응시킨다.

2.4 Rationale for Using FLUX.2 in Limited-Data Environments

FLUX.2는 고품질 이미지 생성 능력과 시점 변환 적용 가능성을 갖추고 있다. 제안 방법은 지상 시점 이미지로부터 공중 시점 이미지 생성 모델로 FLUX.2를 사용하고 효율적으로 적응시키기 위해 제한된 군사 도메인 데이터를 LoRA fine-tuning하여 적용한다^[14-16]. FLUX.2와 LoRA를 결합한 공중 시점 변환 파이프라인을 구성하고 제한된 데이터 환경에서도 시점 다양성을 보완할 수 있는 합성 데이터 생성 방안을 제시한다.

3. 방법론 제안

3.1 Overview of the Proposed Framework

전장 상황에서 무인 시스템의 상황인지 성능 향상을 위해 Ground view를 Top-down view로 변환하는 합성 데이터 생성 프레임워크는 FLUX.2 기반 생성 모델에 LoRA를 결합한 파인튜닝 방식으로 구성된다. 학습 단계에서는 Ground view 이미지와 이에 대응하는 target 이미지를 이용하여 시점 변환 특성을 학습하고 데이터 수를 조절하면서 LoRA 가중치를 최적화한다. 추론 단계에서는 학습된 LoRA 가중치가 적용된 FLUX.2 모델에 새로운 Ground view 이미지를 입력하여 Top-down view 이미지를 생성한다. 입력 이미지는 VAE encoder를 통해 잠재 표현으로 변환된 후 FLUX.2 Transformer block에서 프롬프트와 함께 처리되며 시점 변환 정보가 반영된 잠재 표현으로 갱신된다. 이후 변환된 잠재 표현은 VAE decoder를 통해 복원되어 최종적인 Top-down view 이미지로 생성된다.

3.2 Fine-tuning

3.2.1 Top-down View Synthesis Architecture

제안 방법은 Ground view 이미지를 입력으로 받아 Top-down view 이미지로 변환하는 조건부 생성 구조를 기반으로 한다. 입력 이미지는 먼저 VAE encoder를 통해 잠재공간의 표현으로 압축되며 이는 다음과 같이 표현할 수 있다.

$$z_i = E_{VAE}(x_i^g). \quad (1)$$

식 (1)에서 x_i^g 는 i 번째 Ground view 입력 이미지, $E_{VAE}(\cdot)$ 는 VAE encoder, z_i 는 입력 이미지의 잠재 표현(latent representation)을 의미한다. 이후 잠재 표현 z_i 는 프롬프트와 함께 FLUX.2의 반복적 Transformer block에 입력되며 Cross-Attention에 적용된 LoRA 모듈을 통해 목표 시점 정보가 반영된 잠재 표현으로 변환된다. 이 과정은 다음과 같이 나타낼 수 있다.

$$\tilde{z}_i = F_{FLUX2}(z_i, p_i; \phi_{LoRA}). \quad (2)$$

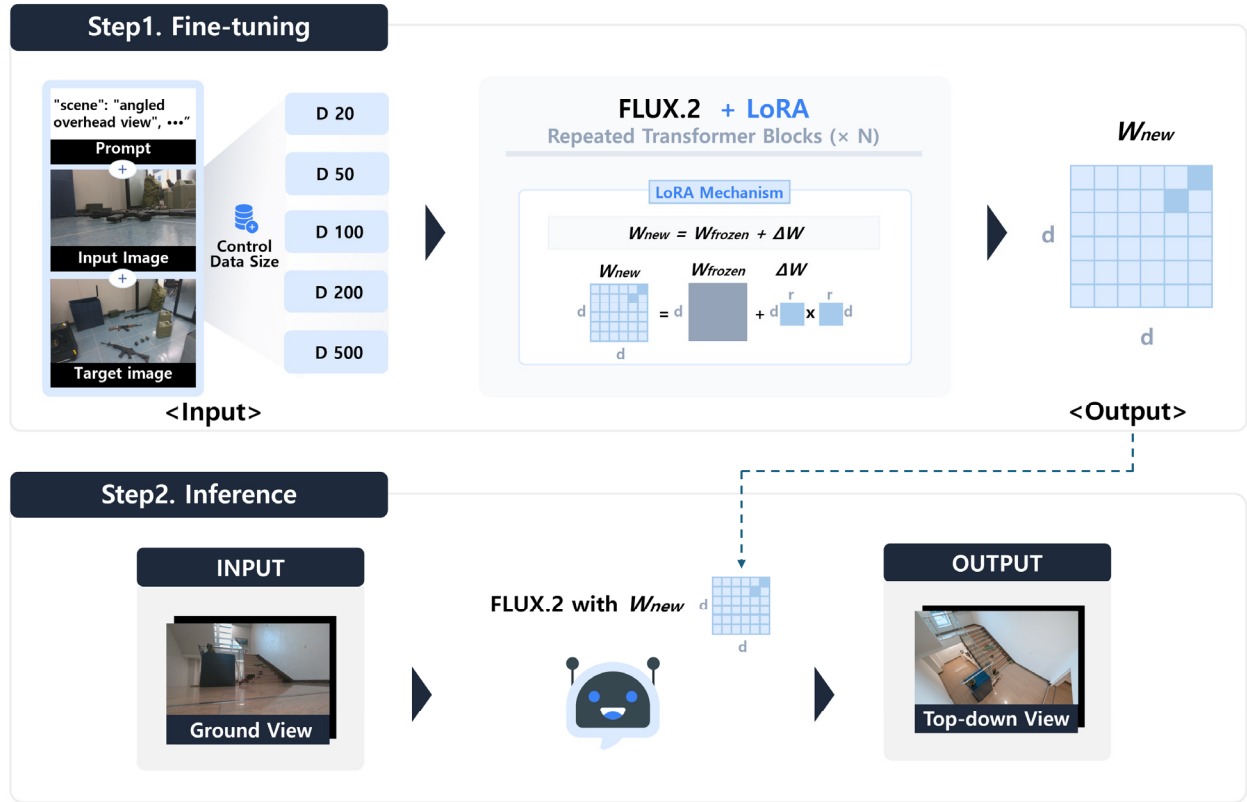
식 (2)에서 $F_{FLUX2}(\cdot)$ 는 FLUX.2 Transformer block, p_i 는 Top-down 시점 변환을 유도하는 프롬프트, ϕ_{LoRA} 는 LoRA 파라미터, $\tilde{z}_i$ 는 시점 변환 정보가 반영된 잠재 표현을 의미한다. 이 과정에서 텍스트 조건은 입력 장면을 단순 복원하는 것이 아니라 공중 시점으로 재구성되도록 유도하는 역할을 수행한다. 최종적으로 변환된 잠재 표현은 VAE decoder를 통해 이미지로 복원되며 이를 통해 Top-down view 합성 결과를 생성한다.

$$\hat{x}_i^t = D_{VAE}(\tilde{z}_i). \quad (3)$$

식 (3)에서 $D_{VAE}(\cdot)$ 는 VAE decoder이며 $\hat{x}_i^t$ 는 생성된 Top-down view 합성 이미지이다.

3.2.2 Parameter-Efficient Fine-tuning with LoRA

대규모 생성 모델의 전체 파라미터를 직접 미세조정하는 방식은 높은 연산 비용과 메모리 사용량을 요구한다. 이를 해결하기 위해 본 연구에서는 LoRA 기반의 파라미터 효율적 미세조정 기법을 적용한다. FLUX.2의 기본 가중치는 고정한 상태로 유지하고 LoRA 가중치만 학습되도록 구성하여 제한된 학습 데이터 환경에서도 효과적인 학습이 가능하다. 이를 통해 기존 사전학습 모델의 일반적인 표현 능력은 유지하면서도 Ground view로부터 변환에 필요한 특성만을 반영하여 객체의 구조적 정보 손실 없이 기하학적 일관성을 안정적으로 유지할 수 있다.



[Fig. 2] Proposed framework for top-down view generation using FLUX.2 and LoRA

3.2.3 Cross-Attention-based View Conditioning

FLUX.2의 Transformer block은 Self-Attention, Cross-Attention, Feed-Forward Network (FFN)로 구성된다. Self-Attention은 입력 이미지의 잠재 표현 내부에서 공간적·의미적 관계를 학습하여 장면의 전반적인 구조 정보를 유지하는 역할을 수행하며 FFN은 각 토큰 표현을 비선형적으로 변환하여 특징 표현력을 보강한다. 반면 Cross-Attention은 이미지의 잠재 표현과 외부 조건 정보 사이의 상호작용을 담당하는 계층으로 프롬프트에 포함된 의미 정보를 잠재 표현에 반영하는 핵심 역할을 수행한다.

본 연구에서는 이러한 구조적 특성을 고려하여 사전학습된 일반 표현 능력을 최대한 유지하면서도 시점 변환에 필요한 조건 정보만을 효과적으로 반영하기 위해 Self-Attention과 FFN은 고정하고 시점 조건 반영에 핵심적인 Cross-Attention 계층에만 LoRA를 적용한다. 이는 전체 네트워크를 모두 업데이트하는 방식에 비해 학습해야 하는 파라미터 수를 크게 줄일 수 있으며, 제한된 학습 데이터 환경에서도 과적합 위험을 낮추면서 효율적인 미세조정이 가능하다는 장점이 있다.

학습은 [Fig. 2]와 같이 D20, D50, D100, D200, D500의 데이터 규모 조건에서 수행하였다. 여기서 D는 fine-tuning에 사용된 학습 이미지 쌍의 개수를 의미하며 제한된 데이터 환경에서

의 시점 조건 반영 특성을 비교하기 위한 설정이다. 여기서 프롬프트는 단순한 부가 설명이 아니라, 생성 결과의 목표 관측 시점을 명시하는 조건 신호로 작용한다. 예를 들어 high-angle, aerial view, drone shot과 같은 표현은 모델이 동일한 장면을 보다 높은 시점, 상부 관측 관점, 혹은 감시·정찰 관점에서 재구성하도록 유도한다.

Cross-Attention은 입력 이미지의 원래 장면 정보와 프롬프트가 제공하는 목표 시점 정보를 정렬하고 결합하는 역할을 수행하며 Ground view 이미지로부터 Top-down view 이미지로의 방향성 있는 시점 변환이 가능해진다. 이를 통해 모델은 객체의 존재 여부와 상대적 배치, 장면 구성과 같은 핵심 의미 정보를 보존하면서도 이를 Top-down view에 적합한 형태로 재구성하게 된다. 이러한 미세조정 방식은 원본 이미지의 구조적 왜곡이나 시각적 손실을 최소화한다. 각 데이터 규모별 생성 품질을 비교 분석하여 군사 도메인에서의 데이터 수집 비용 대비 최적의 시점 변환 효율성을 검증할 수 있다. 이렇게 확보된 고품질의 합성 데이터는 최종적으로 멀티로봇 객체 탐지 모델의 학습에 적용되어 상황인지 시스템의 시점 편향 문제를 효과적으로 해소하는 데 기여한다. 결과적으로 지상 로봇과 공중 로봇 간의 시야를 동기화하여 시각지대 없는 통합적인 상황인지 능력을 극대화할 수 있다.

3.3 Inference

3.3.1 Cross-Attention-based View Conditioning

추론 단계에서는 학습된 LoRA 가중치를 적용한 FLUX.2 모델에 새로운 Ground view 이미지를 입력하여 대응되는 Top-down view 합성 이미지를 생성한다. 입력 이미지는 학습 단계와 동일하게 VAE encoder를 통해 잠재 표현으로 변환되고, FLUX.2 Transformer block에서 프롬프트와 함께 처리되어 시점 변환이 수행된다. 이 과정에서 Cross-Attention 계층에 적용된 LoRA 가중치는 입력 이미지의 잠재 표현과 프롬프트 간의 상호작용을 조정하여 목표 시점 조건을 반영한다. 학습 과정에서 획득한 LoRA 파라미터는 Ground view와 Top-down view 간의 시점 대응 관계를 인코딩하고 있으며, 이를 통해 추론 시 새로운 입력 이미지에 대해서도 일관된 시점 변환이 가능하다.

프롬프트는 단순한 텍스트 설명이 아닌 생성 조건을 명시하는 제어 신호로 작용하며, “high-angle oblique aerial view”, “realistic drone photography”와 같은 시점 관련 표현을 포함하여 모델이 공중 관측 특성을 반영하도록 유도한다. Self-Attention과 FFN 계층은 사전학습 상태로 고정되어 있어 원본 이미지의 의미적 내용과 객체 정보는 보존하면서도 Cross-Attention을 통해 시점만을 선택적으로 변환하는 구조적 분리가 이루어진다.

3.3.2 Synthetic Data Generation

생성된 Top-down view 이미지는 실제 수집이 어려운 공중 시점 데이터를 보완하는 합성 데이터로 활용된다. 이러한 데이터는 객체 탐지 모델 학습에 적용되어 시점 다양성을 확장하고 일반화 성능을 향상시키는 데 기여한다. 생성된 이미지는 객체 간 상대적 위치 관계와 장면 구조를 유지하고 있어 단순한 시각적 변환을 넘어 실제 학습 데이터로 활용 가능한 구조적 정보를 포함한다.

구체적으로 생성된 Top-down view 이미지는 다음과 같은 방식으로 객체 탐지 학습 데이터로 활용된다. 첫째, 시점 다양성 확보 측면에서 기존 학습 데이터가 지상 시점에 편향된 경우 합성 데이터를 통해 공중 시점 표본을 보충함으로써 모델이 다양한 관측 조건에서 강건하게 동작하도록 한다. 둘째, 데이터 증강 측면에서 Flip, Elastic Transform, Affine Transform 등의 기하학적 변환과 InstructPix2Pix 기반 스타일 변환을 추가로 적용하여 생성 데이터의 시각적 다양성을 확대한다^[17]. 셋째, 라벨 효율성 측면에서 시점 변환이 객체의 구조적 위치를 보존하므로 원본 Ground view 이미지에 대한 annotation을 Top-down view 합성 이미지에도 그대로 적용할 수 있어 추가적인 라벨링 비용 없이 학습 데이터를 확장할 수 있다. 특히 구

조적 일관성 지표로 품질이 보장된 합성 데이터만을 학습에 활용한다.

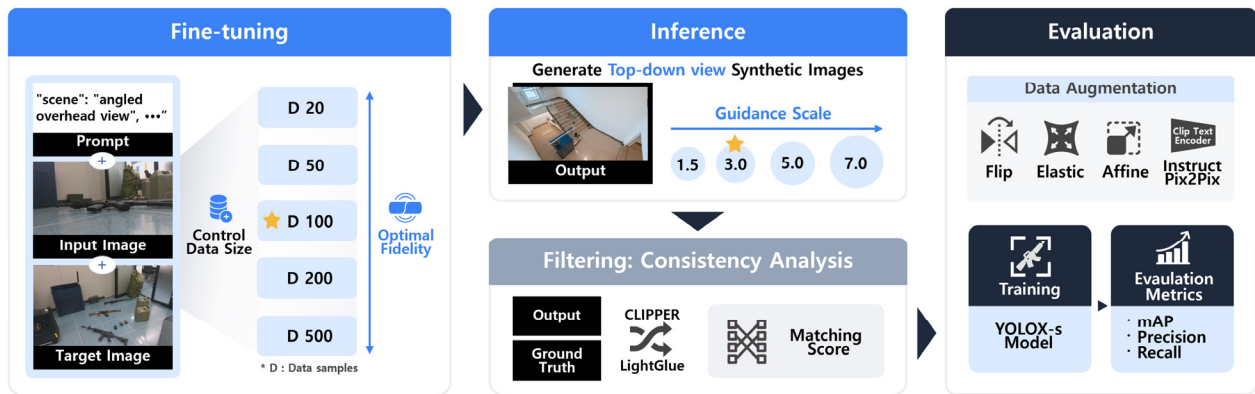
이러한 특성을 바탕으로 생성된 합성 데이터는 단순히 학습 표본 수를 늘리는 것을 넘어, 실제 수집이 제한적인 공중 시점 조건에서의 모델 일반화 능력을 향상시키는 효과적인 데이터 증강 수단으로 기능한다. 특히 군사 도메인과 같이 데이터 수집에 시간적·공간적 제약이 큰 환경에서는 이러한 합성 데이터 기반 접근이 실용적 대안이 될 수 있다.

3.4 Application to Multi-Robot Collaborative Situation Awareness

제한한 시점 변환 파이프라인이 멀티로봇 시스템의 협업 상황인지에서 활용되는 방안을 구체적 임무 상황을 중심으로 기술한다. 정찰·감시 임무에서 지상 로봇은 건물·수목에 의해 가려진 근접 객체를 Ground view로 관측하는 반면, 공중 로봇은 동일 지역을 상부에서 광역으로 관측한다. 이때 두 로봇의 관측 시점이 상이하여 동일 객체에 대한 탐지 결과를 직접 정합·융합하기 어렵다. 이러한 시점 불일치 문제를 해결하기 위해서는 시점 정렬이 필수적이며 본 연구에서는 지상 시점을 공중 로봇의 Top-down view로 변환함으로써 두 로봇의 탐지 결과를 일관된 좌표계에서 융합 가능하도록 한다.

먼저 공중 로봇은 Top-down view 관측 이미지에 대해 객체 탐지를 수행하여 기준 상황인지 맵을 구성한다. 다음으로 각 지상 로봇이 획득한 Ground view 이미지를 입력으로 하여 학습된 FLUX.2-LoRA 시점 변환을 통해 공중 로봇과 동일한 Top-down view 이미지로 변환한다. 이 과정에서 생성 모델 기반 시점 변환은 기하학적 왜곡이나 구조 정보 손실을 수반할 수 있으므로, 변환된 이미지의 신뢰성을 보장하기 위한 품질 검증이 요구된다. 이후 구조적 일관성 지표를 통해 원본 Ground view 이미지와 변환된 Top-down view 이미지 간의 기하학적 대응 관계를 정량적으로 평가하고, 일관성이 낮은 결과는 잘못된 변환으로 인한 오탐지 및 상황인지 오류를 방지하기 위해 배제한다. 신뢰도가 확보된 변환 이미지에 대해서만 공중 로봇의 Top-down view 기준으로 객체 탐지를 수행한다.

이를 통해 공중 로봇이 가림 및 해상도 한계로 탐지하지 못한 사각지대 객체를 지상 로봇의 변환 결과로 보완할 수 있으며 단일 시점 편향과 탐지 누락을 줄여 협업 상황인지의 범위를 확장한다. 따라서 제안 기법은 지상 로봇과 공중 로봇이 공통의 Top-down view 표현 공간에서 관측을 공유하는 멀티로봇 협업 상황인지의 구성요소로 활용한다.



[Fig. 3] Overall experimental pipeline: from LoRA fine-tuning to object detection evaluation

4. 실험

4.1 Overall Experimental Pipeline

본 연구의 전체 실험 절차는 [Fig. 3]과 같이 총 다섯 단계로 구성된다. 먼저 FLUX.2-klein-9B 기반 LoRA 시점 변환 모델을 데이터 규모별로 학습하여 가중치를 획득한다. 이후 학습된 가중치를 이용하여 입력 Ground view 이미지로부터 Top-down view 이미지를 생성한다. 생성된 이미지는 일관성 분석 및 데이터 증강 과정을 거쳐 학습 데이터로 구성되며 최종적으로 YOLOX-s 객체 탐지 모델을 통해 성능을 평가한다^[18]. 본 실험은 Ubuntu 환경에서 NVIDIA RTX A6000 GPU 4장을 사용한다. 실험 환경의 드라이버 버전은 535.86.10, CUDA 버전은 12.2이며 생성모델 기반 LoRA 학습과 객체 탐지 학습은 Python 및 PyTorch 기반 환경에서 수행한다. 학습 데이터는 군사 도메인 데이터셋으로 야외 및 건물 주변 등 다양한 전장 환경을 포함하여 직접 촬영한다.

4.2 Fine-tuning

본 연구에서는 FLUX.2-klein-9B 모델을 기반으로 LoRA 방식의 image-to-image fine-tuning을 수행한다. 입력(source) 이미지와 목표(target) 이미지 쌍을 이용하여 지상 시점 이미지를 공중 시점 이미지로 변환하도록 학습한다. 학습 데이터의 규모가 LoRA 모델의 성능에 미치는 영향을 분석하기 위해 학습 이미지 수에 따라 D20, D50, D100, D200, D500의 다섯 가지 실험 설정을 정의한다. 각 설정은 각각 20장, 50장, 100장, 200장, 500장의 학습 이미지를 사용한다. 프롬프트는 공중 시점 변환에 필요한 시점 정보를 일관되게 반영할 수 있도록 구성한다.

[Table 1] LoRA Fine-tuning Hyperparameters

Parameter	Value
Rank	2
Alpha	2
Learning rate	1e-4
Image resolution	256×256
Optimizer	8-bit Adam
Batch size	2
Gradient accumulation steps	64
Max train steps	500

LoRA fine-tuning 하이퍼파라미터는 [Table 1]과 같이 지정하여 학습 데이터 수에 따른 성능 변화만을 독립적으로 분석한다.

4.3 Inference

추론 단계에서는 학습된 LoRA 기반 시점 변환 모델을 이용하여 입력 Ground view 이미지 500장으로부터 Top-down view 이미지를 생성한다. Guidance Scale은 Classifier-Free Guidance에 기반한 파라미터로 텍스트 프롬프트 조건의 반영 강도를 조절한다. 값이 높을수록 프롬프트에 명시된 시점 조건이 생성 결과에 강하게 반영되며 값이 낮을수록 사전학습 모델의 자체적인 생성 분포에 가까운 결과가 산출된다. 본 실험에서는 1.5, 3.0, 5.0, 7.0의 4개 설정으로 구성하여 총 2,000장을 생성한다.

공중 시점 변환을 유도하기 위한 프롬프트는 "high-angle oblique aerial view, realistic drone photography, drone shot, 45 degree oblique view, about 3m away"를 사용한다. 해당 프롬프트는 상부 관측 시점, 무인 시스템 촬영 관점, 그리고 비스듬한 고각 시점을 동시에 반영하도록 설계되었으며 생성 이미지가 목표 시점의 특성을 안정적으로 반영하도록 유도한다.

4.4 Filtering: Consistency Analysis

생성된 이미지의 구조적 일관성과 시점 변환 품질을 정량적으로 평가하기 위해 Ground Truth 이미지와의 대응 관계 기반 유사도 분석을 수행한다. 입력 이미지와 생성 이미지 간 local feature correspondence를 추정하기 위해 LightGlue 기반 feature matching을 적용하여 대응 후보를 생성한다^[19]. 이후 생성된 대응 관계 중 기하학적으로 일관되지 않은 outlier를 제거하기 위해 CLIPPER 기반 geometric consistency filtering을 수행한다^[20]. 이를 통해 시점 변환 과정에서 원본 장면의 기하학적 구조가 얼마나 보존되었는지를 정량적으로 측정할 수 있다.

기하학적으로 타당한 대응 집합을 바탕으로 다음의 세 가지 consistency metric을 정의하여 활용한다. 첫째, Inlier Ratio는 LightGlue로 추정된 전체 대응점 중 CLIPPER 기반 기하학적 필터링을 통과한 inlier의 비율로 값이 높을수록 생성 이미지와 Ground Truth 간의 기하학적 일관성이 우수함을 의미한다. Avg Matches는 생성 이미지와 Ground Truth 이미지 간 LightGlue가 추정된 평균 특징점 대응 수로, 값이 높을수록 두 이미지 간 공유되는 구조적 정보가 많음을 나타낸다. Learned Perceptual Image Patch Similarity (LPIPS)는 사전학습된 심층 신경망 기반의 시각적 유사도 지표로 값이 낮을수록 두 이미지 간 시각적 유사성이 높음을 의미한다. 이는 후속 학습 단계에서 구조 보존이 우수한 생성 결과를 선별하기 위한 필터링 기준으로 활용된다. 이 과정은 생성 데이터의 품질을 사전에 점검하여 학습 데이터 구성의 신뢰성을 높이기 위한 절차이다.

4.5 Evaluation

생성 및 증강된 데이터를 이용하여 YOLOX-s 객체 탐지 모델로 데이터 4,000장으로 학습하고 200장씩 평가와 테스트를 수행한다. 본 연구에서는 제안 기법의 실효성을 입증하기 위해 합성 데이터를 배제한 총 400장의 실제 군사 도메인 공중 시점 (real Top-down view) 이미지를 평가 및 테스트 데이터셋으로 구축한다. 이는 가상 데이터를 통해 미세조정된 모델이 domain gap을 성공적으로 해소하여 실제 전장과 같은 복잡한 환경에서도 강건한 상황인지 능력을 확보할 수 있는지를 객관적으로 실증하기 위한 설계이다.

평가 지표로는 mean Average Precision (mAP), Precision, Recall을 활용하여 생성 데이터의 실제 활용 가능성을 종합적으로 검증한다. 생성 데이터가 실제 학습 데이터로 활용될 때 탐지 성능에 미치는 영향을 분석하기 위해 원본 데이터와 합성 데이터의 비율을 100:0, 80:20, 50:50, 20:80, 0:100으로 구분하여 학습 구성을 설계한다.

특히 생성 이미지의 시각적 품질 자체보다 실제 객체 탐지 학습 과정에서 유의미한 학습 데이터로 기능할 수 있는지를 중점적으로 평가한다. 생성된 공중 시점 이미지의 다양성 확보를 위해 Flip, Elastic, Affine Transform 및 InstructPix2Pix 기반 증강을 적용한다. 이때 모든 증강은 객체의 구조적 위치 관계를 보존하는 범위 내에서 지정한다. 원본 라벨을 그대로 재할용할 수 있도록 구성하였으며 이를 통해 라벨 일관성을 유지하면서 학습 표본의 다양성을 확대한다.

5. 실험 결과 분석

5.1 Qualitative Comparison

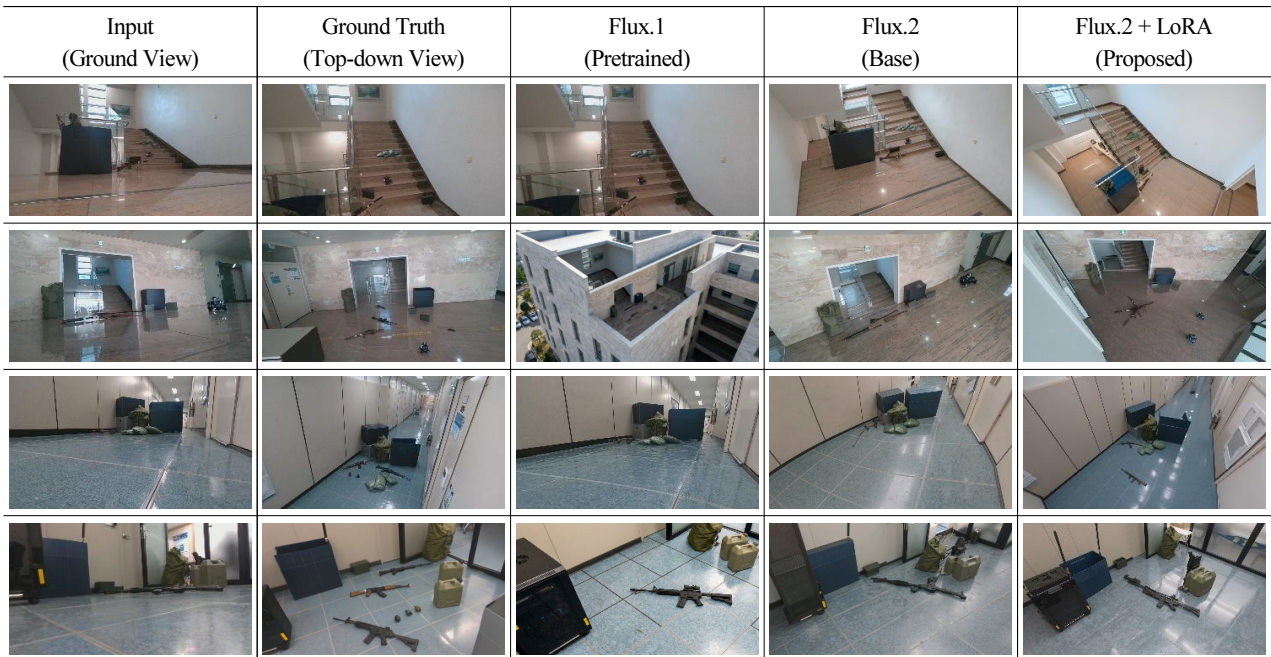
[Table 2]는 target sample에 대한 Top-down view 생성 결과를 정성적으로 비교한 것이다. 비교 대상은 입력 Ground-view 이미지, Ground-truth Top-down 이미지와 함께 다음과 같이 정의한다. Flux.1(Pretrained)은 FLUX 2 모델의 이전 버전을 사전 학습 상태 그대로 시점 변환에 적용한 결과이며 Flux.2(Base)는 본 연구에서 사용하는 FLUX.2-klein-9B 모델을 fine-tuning 없이 프롬프트만으로 시점 변환을 시도한 결과이다. Flux.2+LoRA (Proposed)는 제안 방법으로, FLUX.2-klein-9B에 Cross-Attention 기반 LoRA fine-tuning을 적용한 결과이다. Flux.1(Pretrained)은 FLUX 아키텍처의 세대별 발전에 따른 성능 차이를 확인하기 위한 참조 모델로 포함한다.

먼저 사전학습 상태의 비교 모델은 시점 변환에 한계를 보였다. Flux.1(Pretrained)은 Ground view의 외형적 특징은 일부 반영하였으나 상부 관측 시점에서의 전환이 불완전하여 객체의 형태 왜곡과 공간 배치의 부정확성이 두드러졌다. Flux.2(Base)는 세대 발전에 따라 전반적인 생성 품질은 향상되었으나, fine-tuning 없이 프롬프트만으로 시점을 유도할 경우 목표 시점이 충분히 반영되지 않거나 배경 구조가 불안정하게 생성되는 사례가 관찰된다. 이는 대규모 사전학습 모델이라 하더라도 군사 도메인의 시점 변환 조건을 프롬프트만으로는 안정적으로 반영하기 어렵다는 점을 보여준다.

반면 제안 방법은 다양한 객체가 존재하는 장면에서도 객체 간 배치 형태를 유지하여 구조적 일관성을 확보한다. 생성 이미지에서는 장면의 객체 구조도 비교적 안정적으로 유지되어 실제 Top-down view에 가까운 시각적 구성을 형성하였으며 시점 변환 과정에서 요구되는 구조보존과 공간적 정합성이 제안 방법에서 보다 효과적으로 확보된다.

또한 Cross-Attention 계층에만 LoRA를 적용함으로써 사전 학습 모델의 일반적인 장면 표현 능력은 유지하면서도 목표 시점 조건을 선택적으로 반영하여 제한된 학습 데이터 환경에서도 과도한 왜곡 없이 안정적인 변환이 가능하다.

[Table 2] Qualitative Comparison of Top-down View Generation for Target Samples



이와 같은 정성적 결과는 생성 이미지가 단순한 시각적 변환을 넘어 후속 객체 탐지 데이터셋 구성에 필요한 구조적 정보와 시각적 신뢰성을 충분히 갖추고 있음을 시사한다. 종합적으로 Flux.2+LoRA는 제한된 군사 도메인 데이터 환경에서도 합성 학습 데이터 생성 방법으로서의 활용 가능성을 보여준다.

5.2 Filtering: Consistency Analysis

본 실험에서 제안하는 기법의 LoRA 학습 데이터 크기 및 프롬프트 구성 변화에 따른 Top-down 변환 결과를 비교한다. 생성 이미지와 Ground truth 이미지 간의 정량적 비교를 위해 CLIPPER 기반 유사도 분석과 LightGlue 기반 특징점 매칭을 수행하였으며 Inlier Ratio, Avg Matches, LPIPS를 consistency metric으로 활용한다. [Table 3]에서 Guidance가 낮을수록 LPIPS가 낮아 지각적 유사도는 높으나 특징점 대응 수(Avg Matches)가 적고 Guidance가 높을수록 특징점 대응은 증가하나 지각적 유사도가 감소한다. 이는 Guidance Scale이 높을수록 프롬프트 조건이 강하게 반영되어 시점 변환은 명확해지나, 원본 이미지와의 지각적 거리가 증가하기 때문으로 해석된다. 따라서 본 연구에서는 원본의 기하학적 일관성을 유지하는 최적의 합성 이미지를 선별하여 후속 학습에 활용한다.

본 실험에서는 구조적 일관성(Inlier Ratio)이 가장 높은 D=100, Guidance=3.0 조건의 합성 데이터를 후속 객체 탐지 실험에 사용한다. 제안한 방법이 장면의 구조 정보를 보존하여 시점 변환을 수행할 수 있음을 확인한다. 데이터 개수별 비교 결과에서

[Table 3] Structural Consistency Analysis of Synthetic Data

Data Size	Guidance	Inlier Ratio	Avg Matches	LPIPS
Synthetic (D20)	1.5	0.501	35.74	0.698
	3.0	0.494	65.31	0.710
	5.0	0.519	71.22	0.721
	7.0	0.520	69.94	0.729
Synthetic (D50)	1.5	0.531	46.28	0.667
	3.0	0.507	68.94	0.692
	5.0	0.508	76.36	0.710
	7.0	0.513	74.42	0.722
Synthetic (D100)	1.5	0.509	55.15	0.672
	3.0	0.527	84.84	0.699
	5.0	0.512	92.18	0.716
	7.0	0.525	89.07	0.725
Synthetic (D200)	1.5	0.499	38.33	0.672
	3.0	0.506	60.91	0.699
	5.0	0.501	66.18	0.716
	7.0	0.492	65.12	0.725
Synthetic (D500)	1.5	0.506	36.71	0.684
	3.0	0.521	58.41	0.704
	5.0	0.517	65.98	0.719
	7.0	0.479	64.96	0.730

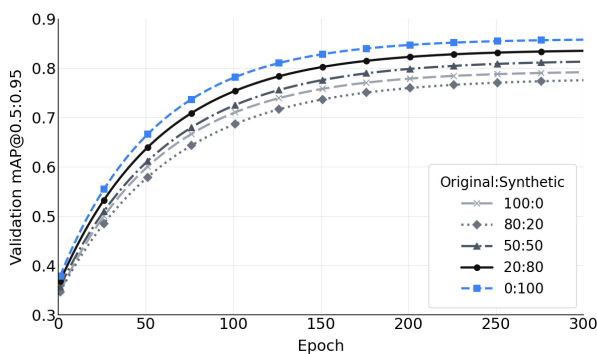
D가 증가함에 따라 D=100까지는 Inlier Ratio와 Avg Matches가 전반적으로 향상되는 경향을 보였으나, D=200 이상에서는 오히려 수치가 하락하는 현상이 관찰된다. 이는 학습 데이터 규모가 일정 수준을 초과할 경우, 제한된 도메인 내 LoRA의 저차원 파라미터 공간에서의 과적합 가능성에 기인하는 것으로 해석된다. 따라서 구조적 일관성 측면에서 D=100이 가장 효과적인 결과를 나타낸다. 또한 프롬프트 비교 실험에서는 “top-down view”, “looking straight down” 등 시점 제약을 명확히 포함한 프롬프트가 구조적 일관성을 향상시키는 반면, “wide shot”, “high altitude”와 같이 촬영 조건을 강조한 프롬프트에서는 객체 스케일 변화 및 배경 왜곡이 발생된다. 이는 프롬프트 설계가 생성 결과의 안정성과 구조 보존에 영향을 미친다는 것을 보여준다.

5.3 Quantitative Results

[Table 4]는 원본 데이터와 합성 데이터의 비율에 따른 YOLOX-s 객체 탐지 성능을 비교한 결과를 나타낸다. 합성 데이터의 비율이 증가할수록 mAP과 Recall이 전반적으로 향상되는 경향을 보였으며 0:100 조건에서 가장 높은 성능을 기록하였다. 이는 제안한 top-down view 기반 합성 데이터가 객체 탐지 학습에 효과적으로 활용될 수 있음을 시사한다.

[Table 4] YOLOX-s Performance According to the Original-to-Synthetic Training Data Ratio

Training Data (Original : Synthetic)	mAP @0.5:0.95	Precision	Recall
100:0	0.795	0.781	0.828
80:20	0.780	0.766	0.815
50:50	0.817	0.804	0.845
20:80	0.838	0.823	0.866
0:100	0.860	0.847	0.887



[Fig. 4] YOLOX-s Validation mAP Curves by Original-to-Synthetic Training Data Ratio

[Fig. 4]는 원본 데이터와 합성 데이터 비율에 따른 YOLOX-s의 epoch별 validation mAP 곡선을 나타낸다. 0:100 조건이 전 구간에서 가장 높은 mAP와 가장 우수한 수렴 경향을 보였고 20:80 조건도 비교적 높은 성능을 나타냈다. 반면 100:0 및 80:20 조건은 상대적으로 낮은 성능을 보였다. 이는 합성 데이터 비율이 높은 구성이 객체 탐지 학습에 효과적임을 보여준다. 0:100 조건에서 가장 높은 성능이 도출된 결과는 다음과 같이 해석할 수 있다. 첫째, 원본 데이터(100:0)는 지상 시점 위주로 구성되어 있어 공중 시점 평가 데이터셋에 대한 domain gap이 존재하는 반면, 합성 데이터(0:100)는 Top-down 시점으로 변환된 이미지로 구성되어 평가 데이터의 시점 분포와 보다 잘 정합된다. 둘째, 합성 데이터 생성 과정에서 적용된 Flip, Elastic, Affine Transform 및 InstructPix2Pix 기반 증강이 시각적 다양성을 확대하여 과적합을 방지하고 일반화 성능을 향상시켰다. 셋째, 합성 데이터는 시점 변환 파이프라인의 구조 보존 특성에 기반하여 원본 라벨을 재활용하는 방식으로 생성되어 라벨 일관성이 유지된다. 본 결과는 합성 데이터만의 학습이 모든 조건에서 최적임을 주장하는 것이 아니라, 본 실험의 특수한 조건에서 나타난 경향으로 해석되어야 하며 평가 데이터 구성에 따른 편향 가능성에 대해서는 향후 추가 검증이 필요하다.

5.4 Discussion

LoRA 기반 Top-down view 변환 합성 데이터는 제한된 실제 데이터 환경에서 무인시스템의 상황인지 성능 향상에 효과적으로 활용된다. 또한 consistency metric과 객체 탐지 성능 사이에는 일정한 연관 경향이 나타나며 생성 데이터의 구조적 일관성이 학습 성능에 영향을 미칠 가능성을 시사한다. 원본 데이터와 합성 데이터의 혼합 비율에 따라 성능 차이가 발생하고 본 실험에서는 합성 데이터 비율이 높은 조건이 비교적 우수한 탐지 성능을 나타낸다.

제안 프레임워크의 각 구성 요소별 기여를 분석하면 다음과 같다. 첫째, Cross-Attention 계층에만 LoRA를 적용한 설계는 [Table 2]에서 Flux.2(Base) 대비 Flux.2+LoRA(Proposed)의 정성적 우위로 확인되며 이는 시점 조건 반영에 Cross-Attention이 핵심적 역할을 수행함을 시사한다. 둘째, LightGlue+CLIPPER 기반 필터링은 [Table 3]의 Inlier Ratio를 통해 기하학적으로 일관된 대응점만을 선별하는 효과가 검증되었다. 합성 데이터의 기여도를 검증하기 위한 Ablation study 결과, 실제 데이터만으로 학습한 baseline 조건(100:0)은 mAP 0.795에 머무른 반면, 제안하는 합성 데이터를 적용한 경우(0:100)는 mAP 0.860으로 향상되어 합성 데이터의 구조적 품질이 탐지 성능 향상에 기여함을 알 수 있다. 결과적으로, 전장 유사 환경에서의 객체 탐지 성능 향상은 제안 기법의 군사적 실용성을 보여준다.

6. 결론 및 향후 연구 방향

본 연구는 전장 환경 무인 시스템의 상황인지 성능 향상을 위한 시점 다양성 확보 문제를 다룬다. 공중 시점 데이터 부족이 학습 데이터 편향과 일반화 성능 저하로 이어질 수 있다는 점에 주목하여 지상 시점 이미지를 공중 시점 이미지로 변환하는 구조적 일관성 기반 합성 데이터 생성 파이프라인을 제안한다. 이를 통해 제한된 실험데이터 환경에서 시점 다양성을 확장하고 비용 효율적인 학습 데이터 확보 가능성을 제시한다.

향후 연구에서는 제안 방법의 신뢰성과 실제 적용 가능성을 더욱 높일 필요가 있다. 군사 도메인 적용을 위한 품질 필터링과 활용 기준을 포함한 합성 데이터 관리 체계 정립이 요구된다. 제안한 시점 변환 파이프라인은 서론에서 제시한 정찰·감시 시나리오와 같이 멀티로봇시스템에서 지상 로봇과 공중 로봇 간 시점 데이터를 상호 변환·보완하는 협업 상황인지 프레임워크로 확장될 수 있다.

References

- [1] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-Image Translation with Conditional Adversarial Networks," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, pp. 1125-1134, 2017, DOI: 10.1109/CVPR.2017.632.
- [2] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks," in *IEEE International Conference on Computer Vision (ICCV)*, Venice, Italy, pp. 2223-2232, 2017, DOI: 10.1109/ICCV.2017.244.
- [3] D. Kothandaraman, T. Zhou, M. Lin, and D. Manocha, "Aerial Diffusion: Text Guided Ground-to-Aerial View Translation from a Single Image using Diffusion Models," in *SIGGRAPH Asia 2023 Technical Communications*, Sydney, Australia, 2023, DOI: 10.1145/3610543.3626177.
- [4] A. Arrabi, X. Zhang, W. Sultani, C. Chen, and S. Wshah, "Cross-View Meets Diffusion: Aerial Image Synthesis with Geometry and Text Guidance," in *2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, Tucson, AZ, USA, 2025, [Online], https://openaccess.thecvf.com/content/WACV2025/html/Arrabi_Cross-View_Meets_Diffusion_Aerial_Image_Synthesis_with_Geometry_and_Text_WACV_2025_paper.html.
- [5] J. Ye, J. He, W. Li, Z. Lv, Y. Lin, J. Yu, H. Yang, and C. He, "Leveraging BEV Paradigm for Ground-to-Aerial Image Synthesis," in *IEEE/CVF International Conference on Computer Vision (ICCV)*, Honolulu, HI, USA, pp. 28451-28461, 2025, [Online], https://openaccess.thecvf.com/content/ICCV2025/html/Ye_Leveraging_BEV_Paradigm_for_Ground-to-Aerial_Image_Synthesis_ICCV_2025_paper.html.
- [6] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, "NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis," in *European Conference on Computer Vision (ECCV)*, Glasgow, UK, pp. 405-421, 2020, DOI: 10.1007/978-3-030-58452-8_24.
- [7] T. Müller, A. Evans, C. Schied, and A. Keller, "Instant Neural Graphics Primitives with a Multiresolution Hash Encoding," *ACM Transactions on Graphics*, vol. 41, no. 4, Article 102, 2022, DOI: 10.1145/3528223.3530127.
- [8] J. T. Barron, B. Mildenhall, D. Verbin, P. P. Srinivasan, and P. Hedman, "Zip-NeRF: Anti-Aliased Grid-Based Neural Radiance Fields," in *IEEE/CVF International Conference on Computer Vision (ICCV)*, Paris, France, pp. 19697-19705, 2023, DOI: 10.1109/ICCV51070.2023.01804.
- [9] B. Kerbl, G. Kopanas, T. Leimkühler, and G. Drettakis, "3D Gaussian Splatting for Real-Time Radiance Field Rendering," *ACM Transactions on Graphics*, vol. 42, no. 4, Article 139, 2023, DOI: 10.1145/3592433.
- [10] R. Liu, R. Wu, B. Van Hoorick, P. Tokmakov, S. Zakharov, and C. Vondrick, "Zero-1-to-3: Zero-shot One Image to 3D Object," in *IEEE/CVF International Conference on Computer Vision (ICCV)*, Paris, France, pp. 9298-9309, 2023, DOI: 10.1109/ICCV51070.2023.00853.
- [11] X. Long, Y.-C. Guo, C. Lin, Y. Liu, Z. Dou, L. Liu, Y. Ma, S.-H. Zhang, M. Habermann, C. Theobalt, and W. Wang, "Wonder3D: Single Image to 3D using Cross-Domain Diffusion," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, USA, pp. 9970-9980, 2024, DOI: 10.1109/CVPR52733.2024.00951.
- [12] Y. Hong, Y. Zhang, C. Pan, Z. Yang, and L. Liu, "LRM: Large Reconstruction Model for Single Image to 3D," in *The Twelfth International Conference on Learning Representations (ICLR)*, Vienna, Austria, 2024, [Online], <https://openreview.net/forum?id=slU8vvsFF>.
- [13] B. Poole, A. Jain, J. T. Barron, and B. Mildenhall, "DreamFusion: Text-to-3D using 2D Diffusion," in *The Eleventh International Conference on Learning Representations (ICLR)*, Kigali, Rwanda, 2023, [Online], <https://openreview.net/forum?id=FjNys5c7VyY>.
- [14] E. J. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, and W. Chen, "LoRA: Low-Rank Adaptation of Large Language Models," in *The Tenth International Conference on Learning Representations (ICLR)*, Virtual, 2022, [Online], <https://openreview.net/forum?id=nZeVKeeFYf9>.
- [15] Black Forest Labs, Introducing FLUX.2, [Online], <https://bfl.ai/blog>, Accessed: Mar. 11, 2026.
- [16] Black Forest Labs, FLUX.2-klein-9B, Hugging Face Model Card, [Online], <https://huggingface.co/black-forest-labs/FLUX.2-klein-9B>, Accessed: Mar. 11, 2026.
- [17] T. Brooks, A. Holynski, and A. A. Efros, "InstructPix2Pix: Learning To Follow Image Editing Instructions," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Vancouver, Canada, pp. 18392-18402, 2023, DOI: 10.1109/CVPR52729.2023.01764.

- [18] Z. Ge, S. Liu, F. Wang, Z. Li, and J. Sun, "YOLOX: Exceeding YOLO Series in 2021," *arXiv:2107.08430*, 2021, DOI: 10.48550/arXiv.2107.08430.
- [19] P. Lindberger, P.-E. Sarlin, and M. Pollefeys, "LightGlue: Local Feature Matching at Light Speed," in *IEEE/CVF International Conference on Computer Vision (ICCV)*, Paris, France, pp. 17627-17638, 2023, DOI: 10.1109/ICCV51070.2023.01616.
- [20] P. C. Lusk, K. Fathian, and J. P. How, "CLIPPER: A Graph-Theoretic Framework for Robust Data Association," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*, Xi'an, China, pp. 13828-13834, 2021, DOI: 10.1109/ICRA48506.2021.9561069.



김 예 린

2019 한밭대학교 정보통신공학과(공학사)
 2025~현재 UST ETRI 스쿨 인공지능 전공
 (석사과정)
 2025~현재 한국전자통신연구원(학생연구원)

관심분야: Military AI, 인공지능, 로봇, 데이터 전처리



백 재 욱

2015 한양대학교 융합전자공학부(공학사)
 2020 한국과학기술원 전기 및 전자공학과
 (공학박사)
 2020 한국과학기술원 정보전자연구소
 (박사후연구원)
 2020~현재 한국전자통신연구원(선임연구원)

관심분야: Military AI, 인공지능, 로봇, 상황인지



전 성 우

1998 연세대학교 환경에너지공학부(이학사)
 2004 충북대학교 전산학과(이학석사)
 2007 UST 컴퓨터소프트웨어공학(박사수료)
 2011~현재 한국전자통신연구원(선임연구원)

관심분야: Military AI, 인공지능, MUM-T, 상황인지



이 창 은

1996 한양대학교 전자공학과(공학사)
 1998 한양대학교 전자공학과(공학석사)
 2017 충남대학교 기계공학과(공학박사)
 1998~1999 LG 산전(연구원)
 1999~2001 엘지오티스엘리베이터(연구원)
 2001~현재 한국전자통신연구원(책임연구원)
 2025~현재 UST ETRI 스쿨 인공지능 전공(교수)

관심분야: Military AI, 인공지능, 로봇, 상황인지