

Research on Channel-Wise Preprocessing for Enhanced Infrared Object Detection

Jae-Uk Kim*, Byung-In Choi**

*Senior Engineer, Intelligence Software Team, Hanwha-Systems Co., Ltd., Korea

**Chief Engineer, Intelligence Software Team, Hanwha-Systems Co., Ltd., Korea

[Abstract]

In this paper, we address the limitation of single-channel infrared (IR) images, which are difficult to directly apply to RGB-based detection models. Previously, a single channel was often replicated into three channels; however, this approach may limit detection performance due to information redundancy. To overcome this limitation, we propose a method that replicates the single-channel IR image into three channels, with each channel processed using different preprocessing techniques, such as CLAHE (Contrast Limited Adaptive Histogram Equalization), Laplacian Filter, and Top-hat transform, to improve detection performance. In this study, we utilized the RT-DETRv2 detection model and the Anti-UAV300 dataset, using IR images sampled at 10-frame intervals for our experiments. By evaluating the effects of each preprocessing technique and deriving the optimal configuration, our method achieved a 2.2% improvement in mean Average Precision (mAP) over conventional methods. This confirms that our method enhances performance over simple replication, presenting a novel approach to improving object detection performance in IR imaging, with promising applications across various fields, particularly in disaster situations where infrared cameras are utilized, as well as in nighttime surveillance and reconnaissance.

▶ **Key words:** Deep learning, Object detection, Image Preprocessing, Infrared Imaging, Thermal Imaging

[요약]

본 논문에서는 단일 채널(Channel)로 구성된 적외선(Infrared, IR) 영상이 RGB 기반 탐지 모델에 직접 적용하기 어렵다는 한계를 해결하고자 하였다. 기존에는 단일 채널을 단순히 3채널로 복제하는 방식이 주로 사용되었으나, 이 방법은 정보 중복으로 인해 탐지 성능이 제한될 가능성이 있었다. 본 연구는 이러한 한계를 극복하기 위해 적외선 영상의 단일 채널을 3채널로 복제한 뒤, 각 채널에 CLAHE(Contrast Limited Adaptive Histogram Equalization), 라플라스 필터(Laplacian Filter), Top-hat 변환의 전처리 기법을 적용하여 탐지 성능을 높이는 방법을 제안한다. 본 연구에서는 RT-DETRv2 탐지 모델과 Anti-UAV300 데이터 세트(Dataset)를 사용해 10프레임(Frame) 간격으로 샘플링(Sampling)한 적외선 영상을 실험에 활용하였다. 이를 통해 각 전처리 기법의 효과를 평가하고 최적 구성을 도출한 결과, mAP(mean Average Precision) 성능이 기존 방식 대비 2.2% 향상되었다. 이는 단순 복제 방식보다 성능이 개선됨을 확인한 것으로, 본 연구는 적외선 영상 기반 객체 탐지 성능 개선을 위한 새로운 접근법을 제시하며, 향후 적외선 카메라가 사용되는 재난 상황, 야간 감시 정찰, 자율주행 차량의 저조도 환경에서의 객체 검출 분야에서 기여할 가능성이 클 것으로 기대된다.

▶ **주제어:** 딥러닝, 객체 검출, 영상 전처리, 적외선 영상, 열 영상

- First Author: Jae-Uk Kim, Corresponding Author: Byung-In Choi
- *Jae-Uk Kim (jaeuk.kim@hanwha.com), Intelligence Software Team, Hanwha-Systems Co., Ltd.
- **Byung-In Choi (byungin.choi@hanwha.com), Intelligence Software Team, Hanwha-Systems Co., Ltd.
- Received: 2024. 10. 17, Revised: 2024. 11. 19, Accepted: 2024. 11. 20.

I. Introduction

객체 검출(Object Detection)은 객체 종류를 분류하고, 위치와 크기를 경계 상자(Bounding Box)를 통해 표현하는 영상처리 분야 중 하나이다. 객체 검출은 자율주행, 산업용 로봇, 보안 시스템, 비디오 감시를 포함한 다양한 산업 분야에서 활용되고 있다. 특히, 높은 정확도를 요구하는 자율주행 및 보안 시스템 분야에서는 오검출로 인한 오류를 최소화하기 위해 다양한 환경 조건을 고려한 학습 데이터가 필요하다.

안개, 비, 어둠과 같은 저조도 환경에서 촬영된 가시광선 영상은 노이즈(Noise)가 많고 객체에 대한 정보량이 감소함에 따라 객체 검출의 정확도가 떨어지기 쉽다. Fig. 1은 동일한 관점(View Point)에서 촬영된 가시광선 및 적외선 영상 쌍에 대한 객체 검출 성능 예시를 보여준다. 가시광선 영상은 배경과 객체를 표현하는 많은 의미론적(Semantic) 정보를 제공하지만, 날씨나 시간과 같은 환경 조건에 크게 영향을 받아 객체 오검출의 원인이 될 수 있다. 반면, 적외선 영상은 열복사 에너지를 이용하기 때문에 열악한 (무월광, 흐림) 환경에서도 가시광선 기반 영상보다 강인하게 배경과 객체의 정보를 표현할 수 있다.

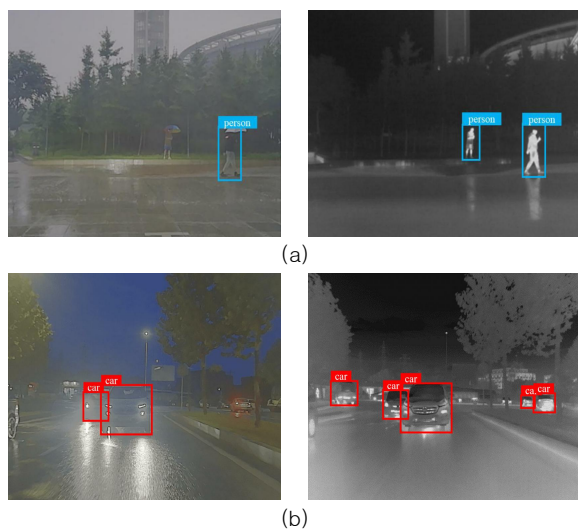


Fig. 1. Examples of object detection performance for visible and infrared image pairs:
 (a) Visible and infrared image pairs in cloudy weather (low illumination environment)
 (b) Visible and infrared image pairs in a night-time autonomous driving environment

적외선 영상은 열 에너지 기반의 정보로 객체를 감지할 수 있기 때문에, 가시광선 영상 기반 RGB 영상의 효율이 많이 떨어지는 밤이나 악천후와 같은 상황에서도 안정적으로

활용될 수 있다. 이는 자율주행차, 무인 항공기(Unmanned Aerial Vehicle, UAV), 군사 감시, 재난 구조, 보행자 검출 등의 분야에서 매우 유용하게 사용된다[1-3].

최근의 불안정한 국제 정세와 군사적 갈등의 증가로 인해 감시 및 정찰 기술의 중요성이 부각되고 있다. 특히, 다양한 환경에서 작전 및 임무 수행을 위해 효과적으로 정보를 수집할 수 있는 기술이 요구됨에 따라, 적외선 영상 기술이 현대 군사작전에서 필수적인 도구로 자리 잡고 있다. 적외선 영상은 가시광선 대신 객체나 생명체에서 방출되는 열을 감지하여 이미지를 생성하는 방식으로, 야간이나 악천후와 같은 가시성이 저하된 상황에서도 안정적인 감시와 정찰이 가능하다. 이는 낮과 밤을 가리지 않고 24시간 감시를 수행해야 하는 군사작전에서 매우 중요한 역할을 한다.

특히, 최근 전쟁 양상이 도시 전투나 복잡한 지형에서 야간 및 악천후에도 전개되면서 적외선 영상의 활용도가 더욱 높아지고 있다. 도심지나 밀집된 인구가 있는 지역에서는 일반적인 영상처리 알고리즘은 한계에 부딪힐 수 있으며, 이러한 상황에서 적외선 영상은 적의 움직임, 무기 체계, 차량 등을 효과적으로 식별할 수 있다. 또한, 추가적인 조명 없이 적외선 센서만으로 정보를 수집할 수 있어 감시 활동이 적에게 노출되는 위험을 최소화할 수 있다.

적외선 영상 기술은 단순한 감시에 그치지 않고, 유도무기 체계에서도 중요한 역할을 한다[4]. 특히, 미사일과 같은 유도 무기들은 목표물의 열 신호를 추적하는 적외선 센서를 통해 저가시성 환경에서도 높은 정밀도의 타격이 가능하다. 또한, 레이저 유도 무기 체계에서는 적외선 영상을 활용하여 목표를 탐지하고, 그 결과 적은 부수적 피해로도 효과적인 타격을 수행할 수 있다[5]. 최근 전력화된 고출력 레이저 무기 체계 역시 다양한 전장 조건에서 효과적으로 임무를 수행하기 위해 여러 적외선 센서를 사용하고 있어 적외선 센서가 필수 불가결한 센서로 자리잡고 있다[6].

가시광선 영상과 비교하여 적외선 영상의 객체 탐지는 적은 정보로 작업을 수행해야 하기 때문에, 탐지 성능을 최적화하는 것이 매우 중요하다. 그러나 적외선 영상은 주로 단일 채널로 구성되어 있어 RGB 기반의 탐지 모델에 적합하지 않다. 일반적으로 단일 채널을 3채널로 복제하여 RGB 입력으로 사용하는 방법이 존재하지만, 이 방식은 원본 데이터의 중복으로 인하여 객체 탐지 모델 탐지 성능에 제약이 될 수 있다. 기존 연구에서는 적외선 영상에서 객체 탐지 성능을 높이기 위해 전처리 기법을 단일 채널에 적용하는 방식이 주로 사용되었다. 반면, 본 연구는 단일 채널의 적외선 영상을 3채널로 복제한 후, 각 채널에

CLAHE, 라플라시안 필터, Top-hat 변환과 같은 상이한 전처리 기법을 각각 적용하여 다양한 객체 특성을 반영할 수 있도록 하였다. 이를 통해 본 연구는 기존 방식 대비 더 높은 탐지 성능을 기대할 수 있으며, 단일 전처리 기법의 제한을 극복하는 새로운 접근법을 제시한다. 본 연구는 이러한 한계를 극복하기 위해 3채널로 복제된 적외선 영상의 각 채널에 서로 다른 이미지 처리 기법을 적용하는 방안을 제안한다. 이를 통해 단일 채널로부터 얻을 수 있는 정보를 최대한 활용하는 방안을 제안하며, 탐지 모델을 통해 그 성능을 평가한다.

II. Preliminaries

2.1 Preprocessing

적외선 영상은 열 에너지를 기반으로 객체를 감지하기 때문에, 야간 감시, 자율주행, 군사 감시 등 다양한 분야에서 활용되고 있다. 반면 적외선 영상은 단일 채널로 구성되어 있어 RGB 기반 객체 탐지 모델을 바로 적용하기는 어렵다. 이러한 문제를 해결하기 위해 기존 연구들은 적외선 영상의 특성을 반영한 다양한 이미지 전처리 기법을 적용하여 탐지 성능을 향상시키고자 하였다.

2.1.1 CLAHE

CLAHE는 적외선 영상의 대조를 개선하는 데 사용되는 대표적인 기법이다[7,8]. 일반적인 히스토그램 평활화(Histogram Equalization) 기법과 달리 CLAHE는 국부적인 대조를 향상시키며, 대조가 과도하게 강화되지 않도록 제한을 두어 영상의 세부 정보 손실을 방지한다. 이 기법은 적외선 영상에서 객체와 배경 간의 대조가 약할 때, 객체의 경계를 강조해 탐지 성능을 향상시킨다.

히스토그램 평활화는 식 (1)과 같이 가로, 세로 $M \times N$ 화소의 영상에서 밝기 값 I 의 누적분포인 누적분포함수(Cumulative Distribution Function) cdf 를 계산한 뒤 누적분포함수의 분포를 평준화시키도록 화소의 값을 조절하여 각 픽셀의 값인 $h(x,y)$ 를 계산한다. L 은 정규화된 영상을 변환하는 스케일 팩터(Scale Factor)로 8비트 레벨 영상에서 256을 사용한다. cdf_{\min} 은 누적분포함수 최소값으로 히스토그램이 가장 낮은 부분을 0으로 분포시키는 역할을 수행하며 cdf_{\max} 는 누적분포함수의 최댓값으로 총 화소의 수와 같다. 계산 후 영상의 밝기 값을 정수형으로 변환하기 위해 반올림(round)을 수행한다.

$$\begin{aligned} hist(k) &= \sum_{x=1}^M \sum_{y=1}^N \delta(I(x,y) - k), \\ \delta(k) &= \begin{cases} 1 & \text{if } k=0 \\ 0 & \text{if } k \neq 0 \end{cases}, \\ cdf(k) &= \frac{1}{M \times N} \sum_{i=0}^k hist(i), \\ round(x) &= \begin{cases} [x] & \text{if } x - [x] < 0.5, \\ [x] & \text{otherwise} \end{cases}, \\ [x] &= n \text{ if } n \leq x < n+1, n \in \mathbb{Z}, \\ h(x,y) &= round \left(\frac{cdf(x,y) - cdf_{\min}}{cdf_{\max} - cdf_{\min}} \times (L-1) \right). \end{aligned} \quad (1)$$

CLAHE는 영상을 가로 i , 세로 j 개의 $T_{i,j}$ 의 타일로 나누고 히스토그램을 생성한다. 이후 cdf 를 계산하고 대조 제한값(Clip Limit) 이상의 히스토그램 기울기를 조절하여 히스토그램 평활화를 수행한다. 전체적인 CLAHE 알고리즘은 식 (2)와 같으며 분할 경계에 이중선형 보간(Bilinear Interpolation)을 적용한다.

$$clahe(x,y) = round \left(\frac{cdf_{T_{i,j}}((x,y)) - cdf_{\min, T_{i,j}}}{cdf_{\max, T_{i,j}} - cdf_{\min, T_{i,j}}} \times (L-1) \right). \quad (2)$$

2.1.2 Top-hat

Top-Hat 변환은 형태학적 연산을 사용하여 객체와 배경의 차이를 강조하는 기법으로, 주로 저대비 영상에서 객체를 부각시키는 데 사용된다[9]. 적외선 영상에서 객체와 배경의 온도 차이가 적을 때 이 기법이 유용하며, 객체의 형태적 특성을 강조함으로써 탐지 모델의 성능을 높일 수 있다.

$$\begin{aligned} T_w(f) &= f - f \circ b, \\ T_b(f) &= f \cdot b - f. \end{aligned} \quad (3)$$

입력 영상 f 에 대해 Top-Hat 변환의 수식은 식 (3)과 같으며, T_w , T_b 는 각각 밝은 영역을 강조하는 White Top-Hat 변환, 어두운 영역을 강조하는 Black Top-Hat 변환 결과이다. 연산 \circ 와 \cdot 는 모폴로지 연산으로 각각 침식(Erosion) 연산 후 팽창(Dilation) 연산을 수행하는 열림(Opening), 팽창 연산 후 침식 연산을 수행하여 노이즈 제거에 효과적인 닫힘(Closing) 연산이다. 이때 b 는 회색조 구조 요소를 뜻하며 대체로 박스 형태의 커널을 사용한다.

본 연구에서는 적외선 영상에서 상대적으로 온도가 높아 밝은 값으로 표현되는 무인 항공기를 효과적으로 검출하기 위해 White Top-Hat 변환을 사용하였다. 이는 원 영상에서 열림 연산을 통해 주변 배경을 제거한 후, 남은 밝은 부분을 강조하여 관심 객체를 강조하는 기법이다.

2.1.3 Laplacian Filter

마지막으로 라플라스 필터(Laplacian Filter) 기반 경계 강조 기법은 객체의 경계 부분을 더욱 선명하게 하고, 노이즈를 억제하여 더 명확한 경계를 제공하는 방법이다[10]. 이 기법은 적외선 영상에서 객체의 경계를 더욱 명확하게 드러내 탐지 성능을 개선하는 데 기여할 수 있다. 라플라스 필터는 식 (4)와 같이 영상 I 에 대해 가로 방향 x , 세로 방향 y 의 2차 미분으로 계산된다. 해당 수식을 마스크 형태로 미리 계산한 뒤 각 픽셀에 합성곱을 진행한다.

$$L(I) = \frac{\partial^2 I}{\partial x^2} + \frac{\partial^2 I}{\partial y^2}. \quad (4)$$

이러한 처리 방법들은 단일 채널을 3채널로 단순하게 복제하는 방법 대신, 각 채널에 다른 기법을 적용하여 다양한 객체 특성을 모델에 제공함으로써 객체 탐지 성능을 극대화하고자 한다.

2.2 Object Detection

2.2.1 YOLO

YOLO(You Only Look Once)는 실시간 객체 탐지에서 주로 사용되는 딥러닝 모델로, 한 번의 연산으로 이미지 전체에서 객체를 탐지하는 방식이다[11]. YOLO는 객체 탐지 속도와 정확도 측면에서 높은 성능으로, 객체 탐지 분야에서 광범위하게 활용되어 왔다. YOLO 알고리즘은 지금도 지속적으로 연구개발 되고 있으며, 최신 버전은 더욱 최적화되고 성능이 개선되었다.

YOLO 모델의 기본 구조는 Fig. 2와 같이 Backbone, Neck, Head로 구성되어 있다. Backbone은 입력 이미지를 처리하여 특징을 추출하는 부분으로, 주로 ResNet, CSPNet과 같은 네트워크가 사용된다. Neck은 Backbone에서 추출된 특징을 다중 스케일에서 통합하여, 다양한 계층적 특징의 결합 및 추출을 수행한다. Head는 마지막으로 객체의 위치와 클래스(Class) 예측을 담당하며, 사전 정의된 경계 상자의 유무에 따라 객체의 위치 추정을 수행한다.

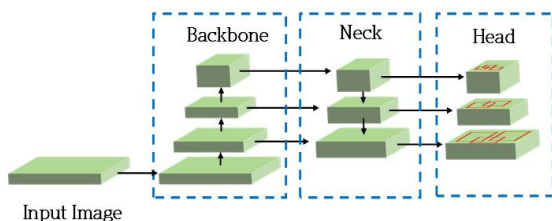


Fig. 2. YOLO architecture.

각 YOLO 버전은 기본 구조를 기반으로 성능 최적화를 위해 다양한 기법을 도입했다. YOLOv4에서는

CSPNet(Cross Stage Partial Networks)를 도입하여 연산 효율성을 높였고, YOLOv5에서는 모델의 경량화를 통해 더 빠른 속도와 적은 자원으로도 높은 성능을 유지할 수 있었다 [12]. YOLOv8에서는 YOLOX의 객체의 중심을 직접 예측하는 앵커 프리 방식을 도입해 경계 상자 예측 수를 줄여 NMS(Non-Maximum Suppression)을 가속화시켰다.

2.2.2 Detection Transformer

최근 객체 탐지 분야에서는 트랜스포머(Transformer) 기반 모델들이 주목받고 있다. 그중에서도 DETR(Detection Transformer)은 CNN(Convolutional Neural Network) 기반 객체 탐지 모델과 달리, 트랜스포머 구조를 사용해 객체 간의 상관관계를 학습하고 전역적인 정보 처리가 가능하다[13]. CNN 모델들이 주로 지역 정보를 중점적으로 처리하는 반면, DETR은 이미지 내의 모든 픽셀을 한 번에 처리하며 전역 문맥을 학습하는 데 강점을 지닌다. 복잡한 장면에서도 객체 간 상호작용을 학습할 수 있어 더 정밀한 탐지가 가능하다. 또한 DETR은 앵커 박스나 NMS(Non-Maximum Suppression) 같은 후처리 없이 객체를 탐지하는 특징을 가진다.

하지만 DETR은 대규모 데이터 세트(Dataset)에서 수렴 속도가 느리다는 문제가 있어 실시간 객체 탐지에는 한계가 있었다.

이 문제를 해결하기 위해 RT-DETR(Real-Time Detection Transformer)이 개발되었다. RT-DETR은 DETR의 장점을 유지하면서 실시간 처리를 가능하게 하기 위해 여러 최적화 기법을 적용했다. RT-DETR은 효율적인 트랜스포머 구조를 통해 학습과 추론 속도를 크게 개선했고, 위치 인코딩을 통해 객체의 위치 정보를 더 정밀하게 처리할 수 있도록 했다. 이를 통해 RT-DETR은 다양한 환경에서 실시간성을 만족하면서 객체 탐지 성능을 향상시켰다[14].

이후 등장한 RT-DETRv2는 이전 버전보다 학습과 추론 속도를 더욱 최적화한 모델이다. RT-DETRv2는 Deformable Attention Module를 활용하여 소형 객체 탐지 성능을 크게 개선했으며, 다양한 증강 기법의 활용으로 적은 양의 데이터로도 높은 성능을 유지할 수 있도록 학습 효율성을 더욱 높였다. 이로 인해 RT-DETRv2는 실시간 객체 탐지 분야에서 더욱 우수한 성능을 보여주며 발전한 모델로 평가받고 있다[15]. RT-DETR의 전체적인 구조는 Fig. 3과 같으며 하이브리드 인코더는 Attention 특징 추출을 위한 AIFI(Attention-based Intra-scale Feature Interaction), 다른 스케일의 특징 맵을 융합(Fusion)하는 CCFF(CNN-based Cross-scale Feature Fusion)로 구성되어 있다.

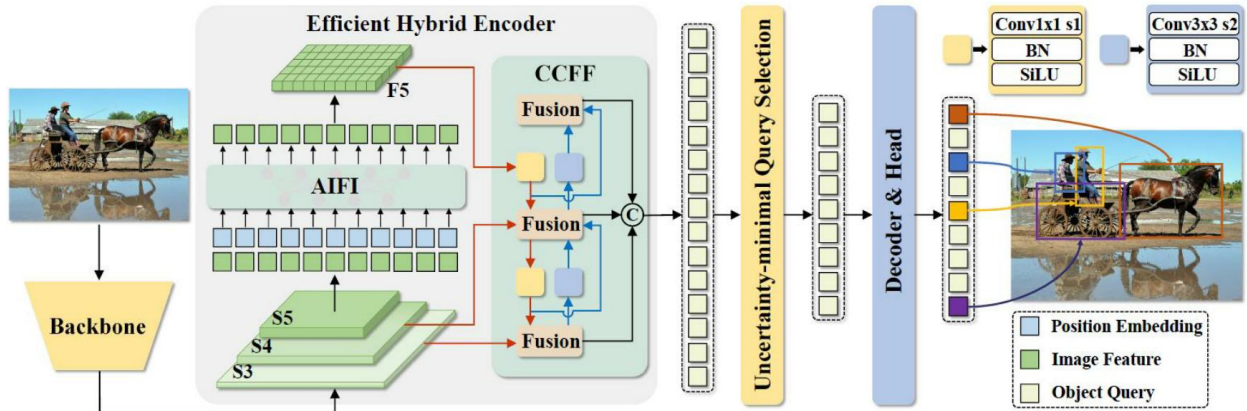


Fig. 3. RT-DETR architecture.

III. The Proposed Scheme

적외선 영상용 객체 검출 모델인 YOLO-IR-Free 모델이 발표되었으나[16], 데이터 세트의 수와 사용자 수가 적고 코드가 공개되지 않아 최신 기법들은 RGB 기반 탐지를 기준으로 활발히 발전되고 있다. 따라서, 기존 연구에서는 단일 채널을 3채널로 복제하여 RGB 형식에 맞추는 방식이 널리 사용되고 있다. 기존 연구에서는 적외선 영상에서 객체 탐지 성능을 높이기 위해 전처리 기법을 단일 채널에 적용하는 방식이 주로 사용되었다.

예를 들어, Back과 Z. Yuan은 단일 채널의 적외선 영상에서 대조 및 밝기를 개선하기 위해 히스토그램 전처리 기법을 도입해 탐지 성능을 향상시키는 방법을 연구하였다[4,17]. 이 연구는 적외선 영상의 대조를 개선해 객체 경계를 더욱 뚜렷하게 만드는 효과를 보여주었으나, 단일 채널에 단일 전처리 기법을 적용하는 방식이었기 때문에 정보 다양성 측면에서 한계가 존재하였다.

반면, 본 연구는 단일 채널의 적외선 영상을 3채널로 복제한 후, 각 채널에 CLAHE, 라플라시안 필터, Top-hat 변환과 같은 상이한 전처리 기법을 각각 적용하여 다양한 객체 특성을 반영할 수 있도록 하였다. CLAHE는 전체 이미지의 대조를 조정하여 어두운 영역의 세부 정보를 강화하고, 라플라스 필터는 이미지의 경계선을 강조해 객체의 형태를 보다 뚜렷하게 표현한다. 또한, Top-hat 변환은 배경과 객체의 차이를 부각하여 특정 특징을 더 선명하게 나타내는 역할을 한다. 이러한 개별 기법들이 각 채널에 적용됨으로써 단일 전처리 기법의 제한을 극복하고, 보다 다양한 객체의 특성을 반영해 탐지 성능을 높일 수 있는 새로운 접근법을 제시한다.

$$I' = \begin{pmatrix} \text{clahe}(I_R) \\ T_w(I_G) \\ I_B + \alpha \times L(I_B) \end{pmatrix}. \quad (5)$$

제안하는 방법은 식 (5)와 같이 단일 채널 적외선 영상에 세 가지 전처리 기법을 각각 적용한 후 결합하여 탐지 모델의 입력으로 제공하는 것으로 탐지 성능을 향상시키는 것이다. 여기서 I_R , I_B , I_G 는 각각 복제된 RGB 채널 영상이며 α 는 1로 설정하고 각 픽셀의 합을 255로 제한하여 경계 성분만을 강조하였다.

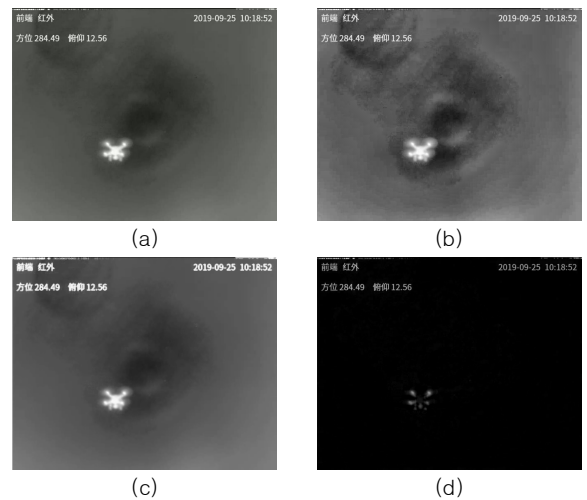


Fig. 4. Examples of image preprocessing techniques for infrared images:

- (a) Original infrared image of Anti-UAV dataset
- (b) CLAHE result
- (c) Laplacian edge enhancement result
- (d) Top-Hat result.

Fig. 4는 상기 방법론을 적용하여 적외선 영상의 채널을 확장한 결과이다. 이 방식은 적외선 영상의 특성을 고려하여 적절한 전처리를 적용하여, 단순할 수 있었던 객체의 정보를 확장함으로써 기존의 단순 복제 방식보다 더 나은

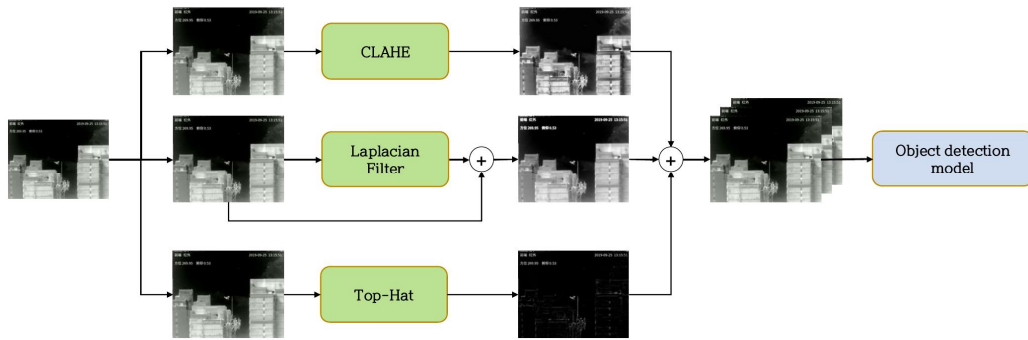


Fig. 5. Proposed Scheme Overview.

탐지 성능을 기대할 수 있다.

Fig. 5는 제안된 방식의 흐름을 보여준다. 먼저 적외선 영상의 단일 채널을 세 개로 복제한 후, 각 채널에 서로 다른 전처리를 적용한 뒤, 이 채널들을 결합해 객체 탐지 모델에 입력으로 전달한다. 제안된 방식은 세 가지 전처리 기법을 활용하여, 적외선 영상에서 객체 탐지 성능을 향상시키는 최적의 방법을 찾는 데 목적이 있다. 실험을 통해 각 전처리 기법이 제공하는 효과를 비교한 후, 최종적으로 탐지 성능을 극대화하는 방법을 선택하여 기존 방식보다 더 우수한 성능을 달성할 수 있을 것이다.

한 포맷(Format)인 COCO(Common Objects in Context) 포맷으로 변환하였다[20]. COCO 포맷은 각 객체의 위치와 크기를 경계 상자 형태로 라벨링 (Labeling)하여 탐지 모델이 쉽게 학습할 수 있도록 지원하며, 다양한 객체의 클래스 정보를 포함하고 있어 객체 탐지에 있어 표준적으로 활용된다. 실험에 사용된 Anti-UAV300 데이터 세트는 Fig. 6과 같이 다양한 환경에서의 영상으로 구성되어 있다.

IV. Experiment Results

4.1 Dataset

본 연구에서는 무인 항공기(UAV) 탐지를 목적으로 한 Anti-UAV300 데이터 세트를 사용하였다[18]. 이 데이터 세트는 다양한 환경에서 촬영된 무인 항공기 추적 장면을 포함하며, 고해상도의 적외선 및 가시광선 영상 쌍을 제공한다. Anti-UAV300 데이터 세트는 지상에서 적외선 센서를 통해 무인 항공기를 촬영한 영상을 제공한다[19]. 본 연구에서는 적외선 영상만을 추출하여 실험을 진행하였다. 적외선 영상은 열 에너지를 기반으로 객체를 감지할 수 있기 때문에, 무인 항공기 탐지와 같은 복잡한 환경에서도 유리한 특성을 가진다.

Anti-UAV300 데이터 세트는 연속적인 영상으로 구성되어 있기 때문에, 본 연구에서는 비디오에서 10프레임당 1프레임씩 추출하는 방식을 통해 학습 및 평가용 이미지를 구축하였다. 이 방법을 통해 영상의 연속성에서 발생할 수 있는 데이터 중복을 최소화하면서, 효율적인 학습을 위한 이미지 데이터를 확보할 수 있었다.

추출된 이미지들은 학습 및 평가용으로 구분되었으며, 최종적으로 14,972장의 학습 이미지와 8,548장의 테스트 이미지로 구성되었다. 또한, 모든 이미지는 객체 탐지에 적합

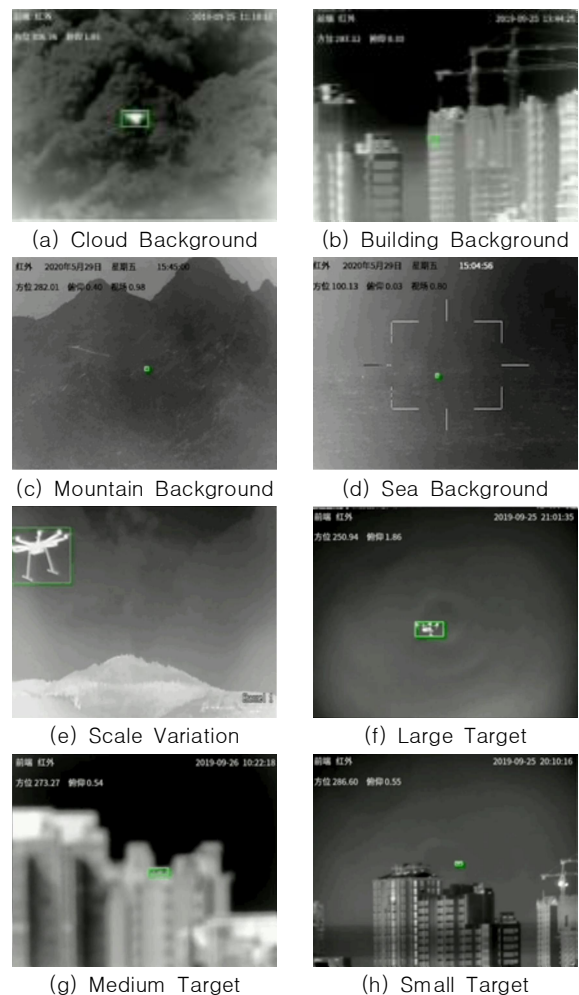


Fig. 6. Examples of Anti-UAV300 datasets.

4.2 Implement Details

본 연구에서는 적외선 영상의 특성을 유지하면서 모델의 탐지 성능을 극대화하기 위해 세 가지 전처리 기법을 적용하였다. 첫 번째로 CLAHE를 사용하여 대조를 개선하였고, 이때 대조 제한값은 2.0, 패치 크기는 8x8로 설정하였다. 두 번째로, 라플라스 필터를 적용해 경계 부분을 더욱 선명하게 강조하였으며, 5x5 커널을 사용하였다. 마지막으로 Top-Hat 변환은 객체와 배경 간 차이를 극대화하기 위해 15x15 커널을 사용하였다. 이러한 전처리 기법들은 적외선 영상의 대조를 극대화하여 탐지 성능을 향상시키는 데 목적이 있다.

객체 탐지 모델로는 실시간 감시 시스템에 적합한 RT-DETRv2 모델을 사용하였고, 측정 간 오차를 최소화하기 위해 무작위성이 포함된 증강 기법은 배제하고 실험을 수행하였다. 실험은 Ubuntu 20.04.6 환경에서 Nvidia Tesla V100 GPU를 활용하여 진행하였다. 실험은 RT-DETRv2의 ResNet-18 backbone 모델을 사용하였고 해당 모델에서 제안하는 하이브리드 인코더와 RTDETRtransformerv2 디코더를 사용하였다.

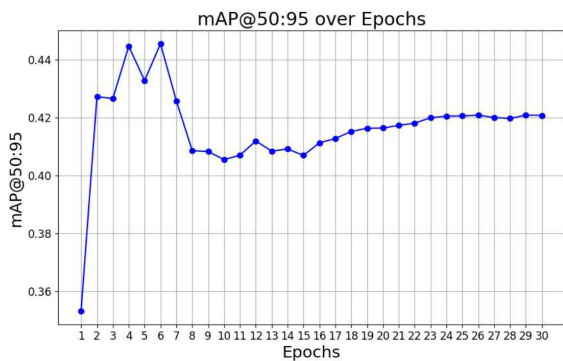


Fig. 7. mAP@50:95 over Epochs.

학습 epoch 수는 실험적으로 30 epoch로 설정하여 진행하였다. 위에서 서술한 데이터셋을 학습할 경우 Fig. 7과 같이 10 epoch 내에서 모델이 최적 성능에 도달하여 전반적으로 초기 epoch에서 빠르게 최적 성능을 달성하는 경향을 보였다.

Table 1. Experiment setup.

Experiment	Method
A	CLAHE
B	Laplacian Filter
C	Top-Hat
D	CLAHE + Laplacian Filter
E	Laplacian Filter + Top-Hat
F	CLAHE + Top-Hat
G	CLAHE + Laplacian Filter + Top-Hat

본 연구에서는 다양한 이미지 전처리 방법이 탐지 성능에 미치는 영향을 분석하기 위해 일련의 실험을 설계하였다. Table 1에 나타난 것처럼, 각 실험에서는 CLAHE, 라플라스 필터, Top-Hat 전처리 기법을 독립적으로 또는 조합하여 적용하였다. 실험 A는 CLAHE 알고리즘만을 사용하여 학습을 수행하였으며, 실험 B, C에서는 각각 라플라스 필터와 Top-Hat 변환을 독립적으로 적용하였다. 이후 실험 D, E, F, G에서는 이 세 가지 기법을 서로 조합하여 다양한 전처리 조합이 객체 탐지 성능에 미치는 영향을 분석하였다. 이러한 전처리 기법의 조합을 통해 개별 기법과 비교했을 때 성능 향상 여부 및 특정 조합이 작은 객체 탐지에 더 효과적인지를 분석하고자 하였다.

본 논문에서는 객체 탐지 성능을 평가하기 위해 많은 연구에서 사용하는 객체 탐지 성능 지표인 mAP 지표를 사용하였다. 실험 결과는 COCO 평가 지표(COCO eval)를 사용하여 평가하였으며, COCO 평가 지표는 IoU(Intersection over Union) 임계값에서 모델의 성능을 측정한다. 특히, COCO는 다양한 IoU 임계값(0.5에서 0.95까지 0.05 간격)에서 AP(평균 정밀도, Average Precision)를 계산하여 모델의 전반적인 성능을 평가하며, 이는 객체의 크기와 복잡한 장면에서도 성능을 종합적으로 평가하는 중요한 척도로 사용된다. 주요 성능 지표로는 mAP@50:95, mAP@50, mAP@75, mAP@s, mAP@m 지표를 포함하여 성능을 평가하였다. 여기서 s, m은 소형(Small), 중형(Medium)을 뜻한다. mAP@50:95는 IoU 임계값을 0.50에서 0.95까지 0.05 단위로 변화시키며 평균적인 성능을 측정한 값으로, 모델의 전반적인 성능을 평가하는 데 중요한 지표이다. 또한 mAP@50은 IoU가 0.50일 때의 정확도를, mAP@75는 IoU가 0.75일 때의 정확도를 나타낸다. 실험에 사용된 데이터 세트에는 대형(Large) 객체가 포함되지 않았기 때문에 대형 객체에 대한 지표는 실험 결과에서 제외하였다.

실험 결과는 Table 2와 같고 실험 F의 mAP@50:95가 0.457로 기존 모델 대비 2.2%의 상승으로 가장 높은 결과를 보였다. 이는 한 채널에서는 원본 영상을 사용하는 것이 효과적이며, 영상이 저장될 때 이미 적절한 대비가 적용되었기 때문으로 판단된다.

또한, 실험 A, B의 결과로 확인할 수 있듯이 한가지의 전처리 기법 적용 후에도 객체 탐지 성능 향상되었다. 이는 적절한 전처리 방법의 적용은 객체 탐지 성능 향상과 연관됨을 알 수 있다.

마지막으로 실험 G의 경우 mAP@50과 중간(m) 크기의 객체에서 검출율이 가장 높게 측정되었으며 IoU를 제외하

Table 2. Experiment results (The highest value is highlighted in red bold).

Experiment	mAP@50:95	mAP@50	mAP@75	mAP@s	mAP@m
Baseline	0.435	0.893	0.333	0.355	0.465
A	0.440 (↑ 0.005)	0.886 (↓ 0.007)	0.349 (↑ 0.016)	0.354 (↓ 0.001)	0.472 (↑ 0.007)
B	0.442 (↑ 0.007)	0.886 (↓ 0.007)	0.357 (↑ 0.024)	0.362 (↑ 0.007)	0.473 (↑ 0.008)
C	0.433 (↓ 0.030)	0.872 (↓ 0.021)	0.332 (↓ 0.001)	0.372 (↑ 0.017)	0.454 (↓ 0.011)
D	0.432 (↓ 0.030)	0.886 (↓ 0.030)	0.323 (↓ 0.010)	0.349 (↓ 0.006)	0.462 (↓ 0.006)
E	0.447 (↑ 0.012)	0.886 (0.0)	0.372 (↑ 0.039)	0.364 (↑ 0.007)	0.474 (↑ 0.009)
F	0.457 (↑ 0.022)	0.890 (↑ 0.007)	0.390 (↑ 0.057)	0.374 (↑ 0.017)	0.488 (↑ 0.023)
G	0.449 (↑ 0.014)	0.894 (↑ 0.010)	0.368 (↑ 0.035)	0.378 (↑ 0.023)	0.474 (↑ 0.009)

고는 가장 높은 검출율을 보였다. 이는 각 전처리 기법의 적용은 감시 장비의 역할에 따라 선택적으로 사용될 수 있음을 도시한다.

V. Conclusions

본 연구에서는 적외선 영상의 객체 탐지 모델 적용 시 채널 복제로 인한 데이터 중복 문제를 전처리 방법을 통해 해결하며 탐지 성능을 향상시킬 수 있는 방안을 살펴보았다. 구체적으로 CLAHE, 라플라스 필터, Top-Hat 변환을 설명하고, 각 조합을 복수 채널에 사용하여 객체 탐지 성능의 변화를 분석하였다. 실험 결과 CLAHE와 라플라스 필터를 적용한 경우 mAP@50:95가 0.457로 기존 모델 대비 2.2% 상승했으며, 한 가지 전처리만을 사용한 결과와 비교했을 때도 mAP@50:95가 1.5% 상승하였다. 이를 통해, 본 연구에서 제안한 전처리 방법이 적외선 영상에서 객체 탐지 성능을 높이는 효과적인 접근법임을 확인하였다.

대부분의 적외선 영상 센서는 RGB 영상과 달리 10~16 비트의 색 깊이(Bit Depth)를 가지므로, 객체 탐지 모델 적용 시 8비트 변환 과정에서 적절한 정규화 작업이 필요하다. 이 과정에서 본 연구에서 제안하는 채널별로 상이한 전처리 방법을 적용하면 적외선 영상 감시 시스템에서의 객체 탐지 성능을 더욱 향상시킬 수 있을 것으로 기대된다.

본 연구에서 제안하는 방법은 각 채널에서 중요한 특징점을 추출할 수 있기 때문에 적외선 카메라가 사용되는 다양한 응용 분야에서 유용하게 활용될 수 있다. 예를 들어, 재난 현장에서 적외선 카메라를 통해 구조 대상을 탐지하거나, 야간 감시 시스템에서 사람 또는 차량의 위치를 추적하는 데 효과적이다. 또한, 군사 작전에서의 야간 감시,

자율주행 차량의 저조도 객체 인식, 산업 설비의 이상 감지 등에서 본 연구에서 제안한 전처리 방법은 적외선 영상의 탐지 성능을 개선할 수 있을 것으로 기대된다. 이러한 다양한 응용 가능성은 본 연구가 제안한 방법이 적외선 기반 객체 탐지에 있어 새로운 기여를 제공할 수 있음을 시사한다.

향후 연구에서는 제안된 전처리 방법의 일반화를 위해 다양한 환경과 조건에서의 성능을 검증할 예정이다. 추가 데이터 세트와 다양한 객체 탐지 모델을 통해 본 방법의 범용성을 확인하고, 실시간 성능 개선을 위한 최적화 방법을 탐색하는 방향으로 연구를 계속할 것이다.

REFERENCES

- [1] A. Akula, Ripul Ghosh, and H. K. Sardana, "Thermal imaging and its application in defence systems," *American Institute of Physics*, Vol. 1391, No. 1, October 2011. DOI: 10.1063/1.3643540
- [2] R. S. Allison, J. M. Johnston, G. Craig, and S. Jennings, "Airborne optical and thermal remote sensing for wildfire detection and monitoring." *Sensors*, 16(8), 2016. DOI: 10.3390/s16081226.
- [3] M. Bertozzi, A. Broggi, and A. Fascioli, "Pedestrian Detection in Far Infrared Images based on the use of Probabilistic Templates," *Proceedings of the IEEE Intelligent Vehicles Symposium*, pp. 134-139, Istanbul, Turkey, July 2007. DOI: 10.1109/IVS.2007.4290135.
- [4] Baek, Jae-Yong, et al., "Research on Local and Global Infrared Image Pre-Processing Methods for Deep Learning Based Guided Weapon Target Detection," *Journal of The Korea Society of Computer and Information* Vol. 29, No. 7, pp.41-51, April 2024. DOI: 10.9708/jksoci.2024.29.07.041
- [5] Myung-Won Shin, "A Study on the Development of High Energy

- Laser(HEL) Weapon System," Journal of the Korea Academia-Industrial cooperation Society, Vol. 23, No. 5, pp.141-150, May 2024. DOI: 10.5762/KAIS.2022.23.5.141
- [6] Seung-Hyun Kang, "The Development Trend and Future Direction of Laser-Based Individual Combat Firearm Weapon Systems," Journal of the Korea Academia-Industrial cooperation Society, Vol. 25, No. 6, pp.544-552, July 2024. DOI: 10.5762/KAIS.2024.25.6.544
- [7] S. M. Pizer, E. P. Amburn, J. D. Austin, R. Cromartie, A. Geselowitz, T. Greer, B. ter Haar Romeny, J. B. Zimmerman, and K. Zuiderveld, "Adaptive histogram equalization and its variations," Computer vision, graphics, and image processing Vol. 39, No. 3, pp.355-368, September 1987. DOI : 10.1016/S0734-189X(87)80186-X
- [8] K. Zuiderveld, "Contrast Limited Adaptive Histogram Equalization," In Graphics Gems IV, Academic Press Professional, Inc., pp. 474-485, August 1994. DOI: 10.5555/180895.180940
- [9] J. Serra, "Image Analysis and Mathematical Morphology," Academic Press, January 1983. DOI: 10.1002/cyto.990040213
- [10] P. A. Mlsna, J. J. Rodriguez, "Gradient and Laplacian Edge Detection," In The Essential Guide to Image Processing, Academic Press, pp. 495-524, April 2009. DOI: 10.1016/B978-0-12-374457-9.00019-6
- [11] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 779-788, Las Vegas, USA, June 2016. DOI: 10.1109/CVPR.2016.91
- [12] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "Yolov4: Optimal Speed and Accuracy of Object Detection," arXiv preprint arXiv:2004.10934, April 2020. DOI: 10.48550/arXiv.2004.10934
- [13] N. Carion, et al. "End-to-end Object Detection with Transformers," In European Conference on Computer Vision, pp. 213-229, Cham, Switzerland, August 2020. DOI: 10.48550/arXiv.2005.12872
- [14] Y. Zhao, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov and S. Zagoruyko, "DETRs Beat Yolos on Real-Time Object Detection," In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1234-1245, 2024. DOI: 10.1109/CVPR46437.2024.00132
- [15] W. Lv, Y. Zhao, Q. Chang, K. Huang, G. Wang, and Y. Liu, "RT-DETRv2: Improved Baseline with Bag-of-Freebies for Real-Time Detection Transformer," arXiv preprint arXiv:2407.17140, 2024. URL: <https://arxiv.org/abs/2407.17140>
- [16] Y. Yang, Z. Jiang, B. Yang, and X. Xia, "YOLO-IR-Free: An Improved Algorithm for Real-Time Detection of Vehicles in Infrared Images." Sensors Vol. 23, No. 21, August 2023. DOI: 10.3390/s23218723
- [17] Z. Yuan, J. Zeng, Z. Wei, L. Jin, S. Zhao, X. Liu, Y. Zhang, G. Zhou, "CLAHE-based low-light image enhancement for robust object detection in overhead power transmission system," IEEE Transactions on Power Delivery, Vol. 38, pp. 2240-2243, April 2023. DOI: 10.1109/TPWRD.2023.3269206
- [18] P. Zhu, L. Wen, D. Du, X. Bian, H. Ling, Q. Hu, B. Cheng, C. Sun, Y. Lei, W. Hu, and L. Jiao, "Anti-UAV: A Large-Scale Benchmark for Vision-Based UAV Tracking," IEEE Transactions on Multimedia, Vol. 25, No. 2, pp. 486-500, April 2021. DOI: 10.1109/TMM.2021.3073459
- [19] J. Zhao, J. Li, L. Jin, J. Chu, Z. Zhang, J. Wang, J. Xia, K. Wang, Y. Liu, and Gulshad, Anti-Unmanned Aerial Vehicle (UAV), <https://github.com/ZhaoJ9014/Anti-UAV>
- [20] T. Y. Lin, M. Maire, S. Belongie, L. Bourdev, R. Girshick, J. Hays, P. Perona, D. Ramanan, C. L. Zitnic and P. Dollár, "Microsoft COCO: Common Objects in Context," In Proceedings of the European Conference on Computer Vision (ECCV), pp. 740-755, Zurich, Switzerland, September 2014. DOI: 10.1007/978-3-319-10602-1_48

Authors



Jae-Uk Kim received the B.S. and M.S. degrees in Robotics from Kwangwoon University, Korea, in 2015 and 2017, respectively. He is currently a Senior Researcher of the Intelligent S/W team

Hanwha Systems Co., Ltd., Seongnam, South Korea. His current research interests include object detection, object tracking.



Byung-In Choi received the B.S., M.S., and Ph.D. degrees in electronic engineering from Hanyang University in Seoul, South Korea, in 2001, 2003, and 2008, respectively. He is currently a Leader of the Intelligent S/W

team Hanwha Systems Co., Ltd., Seongnam, South Korea. His current research interests include object detection, object tracking, and super resolution.