

Research on Local and Global Infrared Image Pre-Processing Methods for Deep Learning Based Guided Weapon Target Detection

Jae-Yong Baek*, Dae-Hyeon Park*, Hyuk-Jin Shin*, Yong-Sang Yoo*, Deok-Woong Kim*,
Du-Hwan Hur*, SeungHwan Bae**, Jun-Ho Cheon**, Seung-Hwan Bae***

*Ph.D. candidate, Vision & Learning Lab, Inha University, Incheon, Korea

*Ph.D. candidate, Vision & Learning Lab, Inha University, Incheon, Korea

*M. S. candidate, Vision & Learning Lab, Inha University, Incheon, Korea

*Ph.D. candidate, Vision & Learning Lab, Inha University, Incheon, Korea

*M. S. candidate, Vision & Learning Lab, Inha University, Incheon, Korea

*M. S. candidate, Vision & Learning Lab, Inha University, Incheon, Korea

**Senior Researcher, Performance analysis Team, LIGNEX1, Pangyo, Korea

**Senior Researcher, Performance analysis Team, LIGNEX1, Pangyo, Korea

***Associate Professor, Vision & Learning Lab, Dept. of Computer Engineering, Inha University, Incheon, Korea

[Abstract]

In this paper, we explore the enhancement of target detection accuracy in the guided weapon using deep learning object detection on infrared (IR) images. Due to the characteristics of IR images being influenced by factors such as time and temperature, it's crucial to ensure a consistent representation of object features in various environments when training the model. A simple way to address this is by emphasizing the features of target objects and reducing noise within the infrared images through appropriate pre-processing techniques. However, in previous studies, there has not been sufficient discussion on pre-processing methods in learning deep learning models based on infrared images. In this paper, we aim to investigate the impact of image pre-processing techniques on infrared image-based training for object detection. To achieve this, we analyze the pre-processing results on infrared images that utilized global or local information from the video and the image. In addition, in order to confirm the impact of images converted by each pre-processing technique on object detector training, we learn the YOLOX target detector for images processed by various pre-processing methods and analyze them. In particular, the results of the experiments using the CLAHE (Contrast Limited Adaptive Histogram Equalization) shows the highest detection accuracy with a mean average precision (mAP) of 81.9%.

▶ **Key words:** Guided weapons, Infrared imaging, Deep learning, Object detection, Image pre-processing

• First Author: Jae-Yong Baek, Dae-Hyeon Park, Hyuk-Jin Shin, Yong-Sang Yoo, Deok-Woong Kim, Du-Hwan Hur, Corresponding Author: Seung-Hwan Bae (Inha University)

*Jae-Yong Baek (jy1213@inha.edu), Vision & Learning Lab, Inha University

*Dae-Hyeon Park (saintPalite2221@inha.edu), Vision & Learning Lab, Inha University

*Hyuk-Jin Shin (shin0528@inha.edu), Vision & Learning Lab, Inha University

*Yong-Sang Yoo (cpp0094@inha.edu), Vision & Learning Lab, Inha University

*Deok-Woong Kim (k5000plus@inha.edu), Vision & Learning Lab, Inha University

*Du-Hwan Hur (gjenghks@inha.edu), Vision & Learning Lab, Inha University

**SeungHwan Bae (seunghwan.bae@lignex1.com), Performance analysis Team, LIGNEX1

**Jun-Ho Cheon (zeitgeisterc@gmail.com), Performance analysis Team, LIGNEX1

***Seung-Hwan Bae (shbae@inha.ac.kr), Vision & Learning Lab, Dept. of Computer Engineering, Inha University

• Received: 2024. 04. 25, Revised: 2024. 06. 21, Accepted: 2024. 06. 26.

[요 약]

본 논문에서는 적외선 이미지에서 딥러닝 물체 탐지를 사용하여 유도무기의 표적 탐지 정확도 향상 방법을 연구한다. 적외선 이미지의 특성은 시간, 온도 등의 요인에 의해 영향을 받기 때문에 모델을 학습할 때 다양한 환경에서 표적 객체의 특징을 일관되게 표현하는 것이 중요하다. 이러한 문제를 해결하는 간단한 방법은 적절한 전처리 기술을 통해 적외선 이미지 내 표적 객체의 특징을 강조하고 노이즈를 줄이는 것이다. 그러나, 기존 연구에서는 적외선 영상 기반 딥러닝 모델 학습에서 전처리 기법에 관한 충분한 논의가 이루어지지 못했다. 이에, 본 논문에서는 표적 객체 검출을 위한 적외선 이미지 기반 훈련에 대한 이미지 전처리 기술의 영향을 조사하는 것을 목표로 한다. 이를 위해 영상과 이미지의 전역(global) 또는 지역(local) 정보를 활용한 적외선 영상에 대한 전처리인 Min-max normalization, Z-score normalization, Histogram equalization, CLAHE (Contrast Limited Adaptive Histogram Equalization)에 대한 결과를 분석한다. 또한, 각 전처리 기법으로 변환된 이미지들이 객체 검출기 훈련에 미치는 영향을 확인하기 위해 다양한 전처리 방법으로 처리된 이미지에 대해 YOLOX 표적 검출기를 학습하고, 이에 대한 분석을 진행한다. 실험과 분석을 통해 전처리 기법들이 객체 검출기 정확도에 영향을 미친다는 사실을 알게 되었다. 특히, 전처리 기법 중에서도 CLAHE 기법을 사용해 실험을 진행한 결과가 81.9%의 mAP (mean average precision)을 기록하며 가장 높은 검출 정확도를 보임을 확인하였다.

▶ **주제어:** 유도무기, 적외선 영상, 딥러닝, 객체 검출, 이미지 전처리

I. Introduction

유도무기의 운용을 위해 정확한 검출과 인식은 필수적이다. 특히, 동적인 영상 환경에서 객체를 정밀하게 검출하는 능력은 유도무기의 정밀도를 향상하는 중요한 과제다. 이러한 맥락에서, 유도무기에 활용되는 객체 검출 알고리즘의 선택은 매우 중요하다. 이 문제에 대한 해결책 중 하나는 최근 주목받는 딥러닝(Deep Learning) 기반 객체 검출기 [1-3]를 활용하는 것이다. 딥러닝 객체 검출기는 전통적인 기법들이 사용하는 feature (예를 들어, Histogram of Oriented Gradient for human detecting, 이하 HOG)보다 컨볼루션 신경망(Convolution Neural Network, CNN)에서 추출되는 강력한 feature를 활용하여 더 높은 검출 성능을 제공한다.

하지만, 자율주행과 같이 통상적인 RGB 영상을 활용해 객체 검출을 수행하는 경우와 달리, 유도무기에서는 객체 검출기는 적외선(Infrared, IR) 영상이 주로 활용된다. 적외선 영상은 통상적으로 사용되는 RGB 영상에 비해서 물체의 열 감지를 통해 객체를 검출하여 야간, 연막, 악천후 등의 열악한 환경에서 정보획득이 유리하기 때문이다. 이 때문에, 감시 및 탐색을 위한 추적 장비와 유도무기 등 군사 사용 수요에 대응하기 위한 목적으로 활용된다 [4, 5]. 따라서, 적외선 영상 환경에서 효율적으로 학습할 수 있도록 객체 검출기를 최적화하는 것이 중요하다. 특히, 적외선 영상은 열적 특성을 통해 영상 내 객체 정보를 표현한다.

그렇기에, 배경(background)과 객체 간의 온도 차이가 작을 경우, 적외선 영상에서 학습 및 추론 대상 객체의 특징이 명확하게 드러나지 않을 수 있다. 따라서, 딥러닝 기반 객체 검출기를 활용해 유도무기의 표적 검출 정확도를 높이기 위해서는 다양한 환경에서 표적 객체의 특징을 잘 표현하는 방법을 개발하는 것이 필수적이다.

기존 연구에서는 적외선 영상에서 객체 검출 성능을 향상하기 위해 객체 검출기의 네트워크 구조를 변경 [6, 7]하거나 다중 스케일 feature를 융합하는 모델 [8, 9]을 제안하였다. 그러나 해당 연구들에서는 상대적으로 더 간단히 표적 객체의 특징을 강조하여 정확도를 향상시킬 수 있는 이미지 전처리에 관한 논의가 부족하였다. 더불어, 기존의 적외선 영상 데이터 셋을 다룬 연구에서도 이와 같은 사항이 충분히 논의되지 않았다 [10-14].

따라서, 본 연구에서는 주로 기존 연구와 달리 유도무기 운용에 특화된 객체 검출 모델 학습을 위한 이미지 전처리 기법에 관한 분석 연구를 수행한다. 본 연구에서는 유도무기 내 적외선 영상을 Fig. 1과 같이 시뮬레이션한 14-bit 합성 영상 데이터 셋을 활용하여 다양한 데이터 유형과 문제에 대해 높은 성능을 발휘하며, 이해와 구현하기 쉬운 네 가지 주요 전처리 기법 (min-max normalization, Z-score normalization, histogram equalization [15], Contrast Limited Adaptive Histogram Equalization



Fig. 1. The examples of infrared synthetic image dataset for guided missile environments

(이하, CLAHE [16])) 적용한 결과를 비교 분석한다. 특히, 본 연구의 핵심은 전역적 정보를 활용하는 Z-score normalization과 지역적 정보를 활용하는 min-max normalization, histogram equalization, CLAHE의 특성을 기반으로 각 전처리 기법의 영향을 분석하는 데 있다. 또한, 본 논문에서는 위 전처리 기법을 적용한 이미지를 활용해 YOLOX 객체 검출기를 학습하고, 이를 통해 객체 검출 성능을 비교 분석한다. 각 전처리 기법으로 변환된 이미지를 통해 얻은 검출 결과를 분석하여 어떠한 전처리 기법이 유도무기 환경 적외선 영상에서 가장 효과적인지 평가하고 이에 대한 분석을 제공한다.

본 논문의 주요 기여 사항은 다음과 같다.

- 적외선 영상 상에서 기존 이미지 전처리 기법들의 한계점을 유도무기 환경 적외선 영상 데이터 셋을 통해 분석을 수행하였다.
- 유도무기 환경 적외선 합성 영상 데이터 셋을 통해 YOLOX [2] 객체 검출기를 학습 및 성능을 도출하였으며, 이미지 전처리 기법들과의 비교 실험을 수행하였다.

본 논문의 구성은 다음과 같다. 2장에서는 이미지 전처리에 사용될 수 있는 알고리즘에 대한 설명을 수행한다. 3장에서는 적외선 영상에 기반하여 기존 이미지 전처리 기법 적용 결과를 보이고 그에 대한 분석을 수행한다. 4장에서는 각 전처리 기법을 통해 전환된 이미지를 통한 객체 검출 정확도 비교 실험에 관한 구성 및 결과에 대해 분석한 내용을 보인다. 마지막으로 5장에서는 결론 및 향후 연구과제에 관해 설명한다.

II. Preliminaries

본 장에서는 이미지 전처리에 사용되는 표준화 및 대비 향상 알고리즘들의 대표적인 사례와 IR Object Detector를 소개한다.

1. Normalization method for global information

1.1 Z-score normalization

Z-score normalization은 딥러닝 데이터 전처리에서 통상적으로 활용되는 기법으로써, 데이터 전체의 평균(μ)과 표준편차(σ)와 같은 전역적 정보를 활용해 이미지의 평균을 0, 표준편차를 1로 조절하는 기법이다. Z-score normalization은 아래와 같이 정의된다.

$$x_z = \frac{x - \mu}{\sigma} \quad (2)$$

이미지를 전처리한다는 측면에서는 min-max normalization과 유사한 기법이지만, 영상 전체의 평균값과 표준편차를 이용하여 전처리를 수행하는 만큼 이상치(outlier)에 대한 영향을 상대적으로 적게 받는다. 그러나, 영상의 전역적 정보를 통해 전처리를 수행하기 때문에 영상의 도메인이 일정하지 않거나 극단적으로 비대칭적인 분포를 보이는 경우 이미지의 중요한 특징이나 패턴을 왜곡할 수 있다 [17].

2. Normalization methods for local information

2.1 Min-max normalization

Min-max normalization은 이미지의 지역적 특성을 통해 이미지를 [0, 1]의 범위로 조절하는 기법이다. 해당 기법은 아래와 같이 정의된다.

$$x_z = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (1)$$

이때, x 는 입력 픽셀값, x_{\max} 과 x_{\min} 는 각각 이미지 내 최대/최소 픽셀값을 의미한다. 해당 기법은 쉽게 구현 가능하며 직관적이기에 자주 활용되는 전처리 기법이다. 그러나, 이미지의 최대/최소값을 활용하는 해당 기법의 특성상 극단값과 같은 이상치에 취약하다는 문제점이 존재한다 [18].

2.2 Histogram equalization

Histogram equalization은 이미지 처리에서 널리 활용되는 기법으로 이미지의 대비(contrast)를 향상하기 위해 픽셀 강도 분포를 조정하는 기법이다. 이때, 해당 기법은 이미지 내의 픽셀값만을 활용하기에 지역적인 전처리 기

법에 속한다. Histogram equalization은 이미지의 히스토그램을 변형하여 모든 픽셀 강도(intensity) 값이 거의 같은 빈도로 나타나도록 하는 것을 목적으로 한다. 이를 위해 먼저 이미지의 히스토그램을 계산하고, 누적 분포(cumulative distribution function)를 통해 누적 히스토그램을 계산한다. 이후, 누적 히스토그램값을 정규화한 후 각 픽셀을 새로운 강도 값으로 매핑한다. 본 기법은 이미지의 전체적인 명암 대비를 개선하여 저대비 이미지에서 세부적인 정보를 더 잘 부각할 수 있게 한다. 그러나, 해당 기법은 히스토그램에서 픽셀값의 균일한 분포 변환을 목적으로 하므로, 픽셀값이 과도하게 변함으로 영상 개선하지 못하는 문제가 발생할 수 있다. 또한, 발생빈도가 낮은 픽셀값의 경우 해당 값이 소실되게 되어 이미지 내 세부 정보가 손실되는 현상이 발생할 수 있다는 문제가 있다 [19-20].

2.3 CLAHE

CLAHE는 이미지를 일정 크기의 패치로 분할하여, 각 패치 단위로 histogram equalization을 수행하는 기법이다. 이때, histogram equalization 수행 시 대비 증가에 대한 제한을 두어 과도하게 이미지 특징이 대비되지 않도록 한다. 그 후, 각 분리된 이미지를 보정을 통해 자연스럽게 병합하여 이미지 특징 대비를 증가시키는 방법이다. 해당 기법은 이미지 패치 내의 지역적인 정보를 통해 영상 대비 향상을 수행하고, 대비 증가를 제한하므로 노이즈의 과도한 증폭을 histogram equalization 대비를 방지할 수 있다. 또한, 고대비 지역과 저대비 지역의 세부 정보를 모두 향상할 수 있으며, 자연스러운 이미지 외관을 유지하면서 세부적인 대비를 개선할 수 있다. 그러나, 해당 기법은 이미지를 패치 단위로 각각 histogram equalization을 적용하므로 처리 시간이 상대적으로 더 길다. 또한, 하이퍼 파라미터(hyper-parameter)인 대비 증가 제한 임계값(threshold)과 패치의 크기에 따라 처리 결과의 품질이 달라질 수 있다는 문제점이 존재한다 [20].

3. IR Object Detector

적외선 검출기는 컴퓨터 비전 분야에서 중요한 역할을 하고 있다. 적외선 탐지 기술은 군사, 보안, 의료, 산업 등 다양한 응용 분야에서 사용되고 있다. 특히, 적외선은 가시광선으로는 확인하기 어려운 환경에서도 유용한 정보를 검출할 수 있어 야간 감시, 열상 진단 등에서 필수적인 기술로 자리잡고 있다. 그러나 적외선 영상 도메인의 특성상 노이즈가 많이 발생하고, 해상도가 낮으며, 온도 변화에

민감하다 [22]. YOLO-IR [6]은 YOLO 모델에 적외선 영상 내에서 feature 표현과 모델의 표현력을 개선할 수 있는 어텐션 모델을 제안하였다. Multi-scale dual attention (이하 MDA) [9]는 다중 스케일 듀얼 어텐션 프레임워크를 제안하여 적외선 영상 내 객체 검출을 수행하였다. 하지만, 기존 연구는 유도무기의 표적 검출에 적합한 전처리 기법은 연구가 부족한 상황이다. 따라서, 본 논문에서는 해당 문제를 해결하기에 적합한 전처리 기법에 대해서 논하고자 한다.

III. Analysis of Infrared Image Pre-Processing Application Results

본 장에서는 유도무기 환경 합성 적외선 영상 데이터 셋을 설명한다. 또한, 해당 데이터 셋에 기반을 둔 적외선 영상 상에서의 이미지 전처리 기법의 효율성을 비교 분석한다.

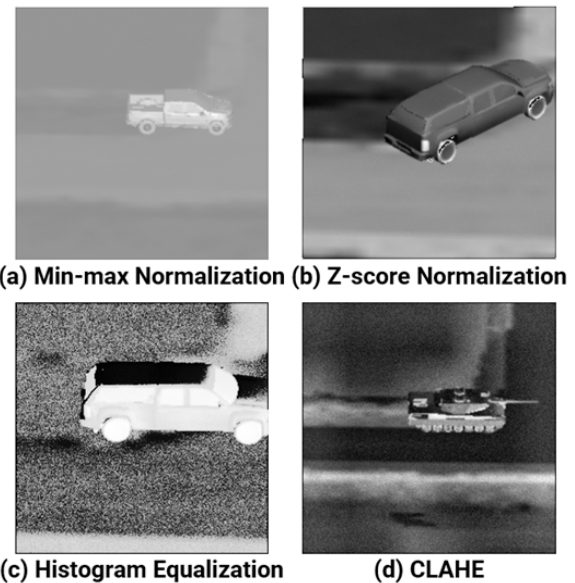


Fig. 2. The results of applying Z-score normalization, histogram equalization, and CLAHE techniques to the infrared synthetic image dataset for guided weapon environments.

1. Infrared Synthetic Image Dataset for Guided Missile Environments

본 절에서는 본 논문에서 활용하는 유도무기 환경을 시뮬레이션 한 14-bit 적외선 영상 데이터 셋에 관하여 설명한다. 본 데이터 셋은 유도무기의 운용 시나리오를 가정하여 생성된 합성 적외선 영상 데이터 셋이다. 본 데이터 셋은 다양한 환경에서 표적이 존재하는 시나리오를 담았다.

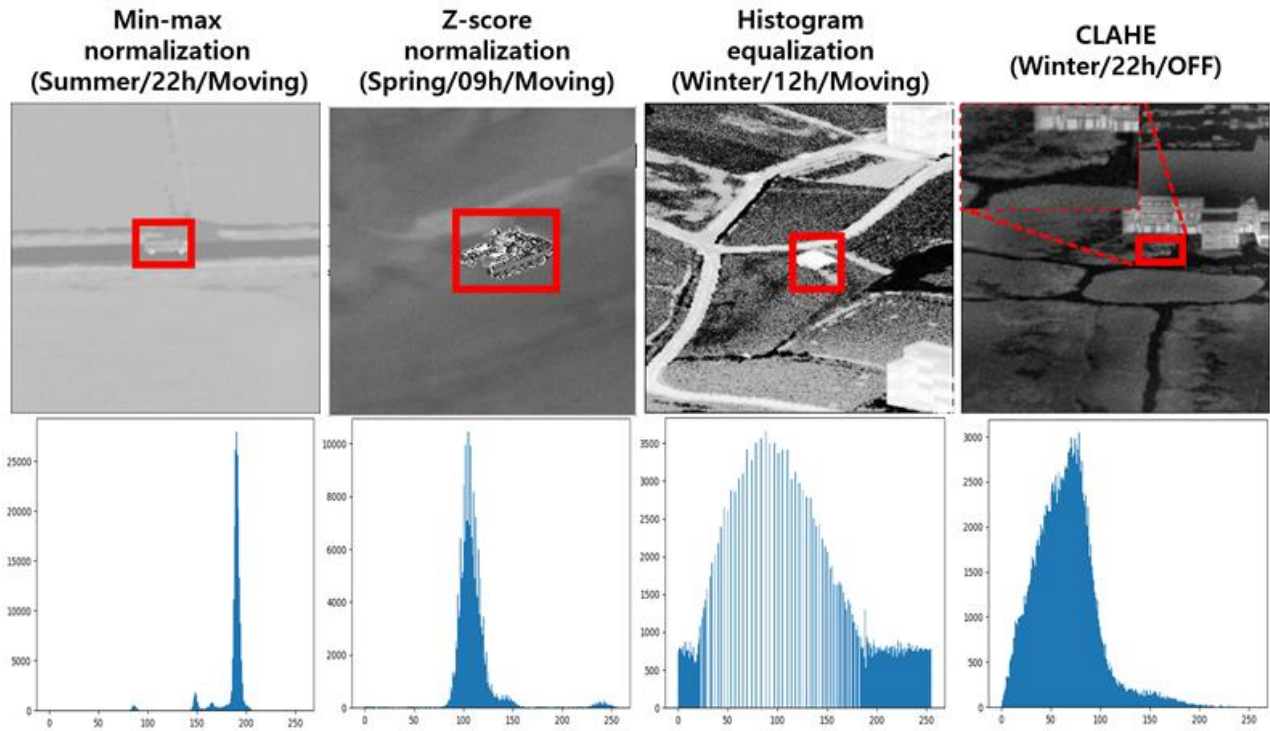


Fig. 3. Failure cases and image histograms for each image pre-processing method on an infrared synthetic image dataset for guided missile environments. We highlight the objects by red bounding box. The graph below represents the histogram for each image. The x-axis indicates the intensity value of the image, while the y-axis represents the count.

먼저 본 데이터 셋에서는 3종의 전차(Tank A, Tank B, Tank C)와 3종의 차량(Vehicle A, Vehicle B, Vehicle C), 총 6종의 클래스로 구성되어 있다. 각 클래스는 유도 무기에서 검출해야 하는 표적을 의미한다. 그리고 해당 데이터 셋은 각기 다른 기온 환경을 가진 봄, 여름, 겨울 계절 환경에서 각각 다른 시간대인 9시, 12시, 19시, 22시에 유도무기가 운용되는 시나리오를 포함하고 있다. 본 데이터 셋은 표적의 상태를 대기(idle), 엔진 off 상태, 이동 상태(전면, 측면 방향)로 구분한다. 또한, 유도무기는 1.5km 부터 5.0km의 범위에서 운용되는 시나리오를 포함하고 있다.

2. Analysis of IR Image Pre-Processing Methods

본 절에서는 유도무기 환경 기반 적외선 영상에 기존 영상 전처리 기법을 적용한 후 그 결과를 분석하고자 한다. 먼저, 3.1절에서 서술한 바와 같이, 본 데이터 셋은 다양한 환경을 반영하기 위해 다양한 계절 및 시간대에서의 환경을 반영한 영상이 존재한다. 다양한 환경(예. 계절 및 시간)에 운용할 수 있는 전장 특성상 유도무기의 객체 검출 모델은 객체를 강건하게 검출해야 한다. 이를 위해서는 1장에서 서술한 바와 같이 표적 객체의 특징이 주변 배경과 구분될 수 있도록 데이터 상에 잘 반영될 수 있도록 해야

한다. 특히, 딥러닝 네트워크의 경우 입력 데이터에 대해 매우 민감하게 반응한다. 따라서, 입력 데이터의 작은 변화에도 네트워크의 가중치가 민감하게 조정될 수 있으며, 이는 학습 과정에 중대한 영향을 미칠 수 있다 [22].

Fig. 2에서는 각 이미지 전처리 기법이 적용된 후 대상 객체의 특징이 이미지 상에서 잘 표현된 경우를 볼 수 있다. 그러나, 2장에서 언급한 바와 같이 각 기법의 한계점으로 인해 특정 적외선 영상에서는 이러한 적절한 처리가 어려울 수 있다. Fig. 3은 각 이미지 전처리 기법이 적외선 영상에서 표적 객체의 특징을 충분히 표현하지 못한 실패 사례(failure case)를 보여주고 있다. 해당 그림에서 각 표적 객체는 붉은색 바운딩 박스로 표시되어 있다.

먼저 min-max normalization의 경우 표적 객체와 배경 간 픽셀값이 비슷하고, 구분하기 어려운 것을 확인할 수 있다. 이는 해당 이미지의 히스토그램에서 특정 구간에 극단적으로 픽셀값이 몰려있음을 확인할 수 있다. 특히, Fig. 3과 같이 여름/이동 상황과 같이 배경과 객체 온도가 유사할 경우 이러한 문제가 발생할 가능성이 크다.

Z-score normalization을 사용했을 때, min-max normalization과 비교하면 객체와 배경 간의 구분이 더 명확하게 이루어 짐을 확인할 수 있다. 특히, min-max normalization이 적용된 이미지의 히스토그램과 비교했

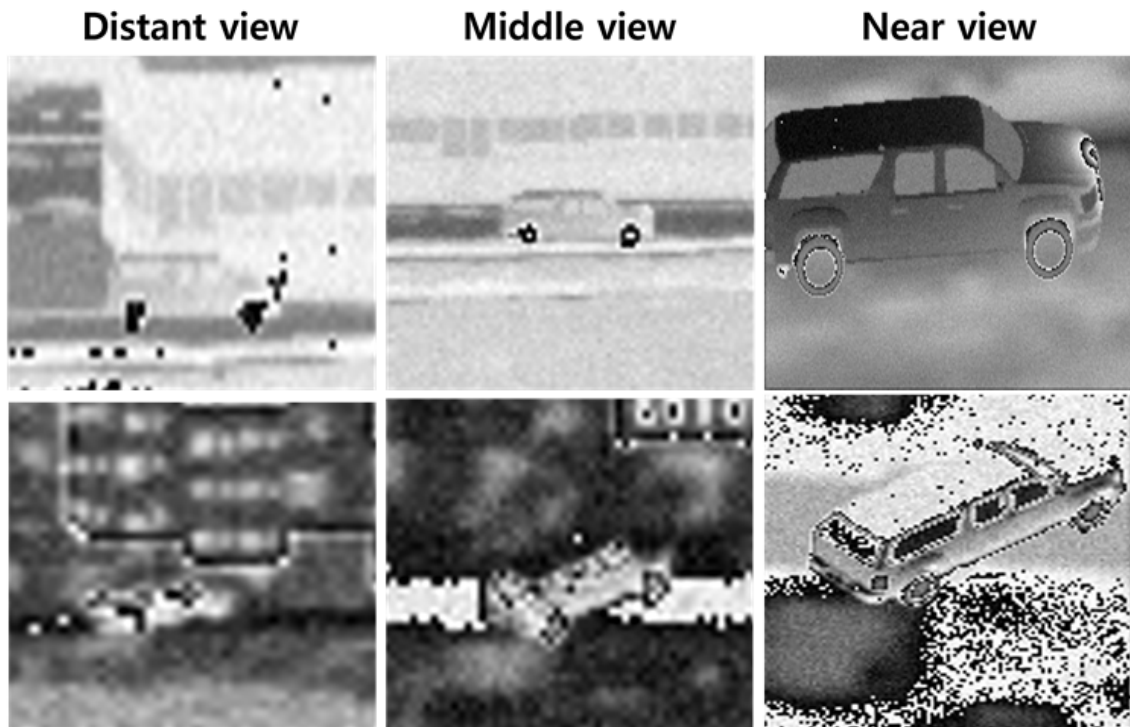


Fig. 4. Different view points of Infrared images from guided weapon environments. These images are transformed via Z-score normalization. It can be observed that each target object shows a different color domain in various view point domains (distant view, middle view, and near view). Specially, in the case of distant and middle views, the color domains of the background and the target objects are similar.

을 때, Z-score normalization이 적용된 이미지의 히스토그램은 특정 구간에 집중되지 않고 더 고르게 분포되는 것을 볼 수 있다. 그러나, 유도무기의 경우, Fig. 4에서 보듯이 유도무기 발사 시점부터 시작하여 표적에 점점 접근하기 때문에 다양한 진입 각도와 표적과 유도무기 사이의 거리가 하나의 영상에 존재한다. 이러한 상황에서 이미지의 도메인을 적절히 나누어 Z-score normalization을 수행하는 것은 어렵다. 특히, Fig. 4에서 보이듯이 원경(distant view)과 중경(middle view)에서는 표적 객체와 배경이 비슷한 색상 도메인을 나타내며, 근경에서야 표적과 배경 간의 구분이 명확히 드러난다. 결론적으로, 이와 같은 유도무기 환경에서는 다양한 도메인이 하나의 영상 데이터 내에 존재함으로써 Z-score normalization의 적용이 타 데이터 셋과 달리 어려울 수 있다.

Fig. 3의 histogram equalization 전환 결과와 다른 기법들과 비교했을 때 객체와 배경 간의 대비는 상대적으로 구분되는 것을 확인할 수 있다. 그러나, Fig. 3의 histogram equalization 적용한 결과를 보면, 전차의 외형적 특성은 배경과 대비되어 잘 드러나지만, 전차가 전부 흰색으로 나타나면서 표적의 세부적인 특성이 소실되었는 것을 확인할 수 있다. 이는, histogram equalization 기법의 특성상 발생빈도가 잦은 배경 부분에 히스토그램

변환함수의 기울기가 급격히 증가하고, 발생빈도가 낮은 표적 객체의 부분에서는 급격히 감소하기 때문이다 [20].

CLAHE의 경우 이미지를 특정 크기의 패치로 나누어 histogram equalization을 수행하는 해당 기법의 특성상, 지역적인 세부 특징을 표현하기에 유리한 기법이다. 특히, 이미지 히스토그램도 타 기법 대비 분포가 고른 것을 확인할 수 있다. 그러나, Fig. 3의 CLAHE 수행 결과의 경우 배경과 표적 객체의 구분이 어려울 정도로 명암 대비가 발생하지 않은 경우를 확인할 수 있다. 이는 CLAHE의 성능을 결정하는 두 매개변수인 이미지 패치 크기와 대비 제한 값이 적절하지 않아 발생할 수 있다. 이러한 관점에서는 각 환경에 알맞은 최적의 매개변수 값을 찾는 것이 문제일 수 있다. 두 번째로는 이미지 상의 표적 객체와 배경이 비슷한 색상 도메인을 가질 때 대비 효과가 미미할 수 있기 때문이다 [23].

결론적으로, 본 절에서는 유도무기 환경 적외선 영상에서 min-max normalization, Z-score normalization, histogram equalization, CLAHE 기법들을 통해 적외선 영상을 전처리하였을 때 그 효율성과 영향에 대해 분석을 수행하였다. 이를 통해 각 기법의 장·단점이 명확히 드러났으며, 특히 CLAHE의 경우 지역적 세부 특징을 보존함에 가장 유효함을 확인할 수 있었다. 그러나, CLAHE는

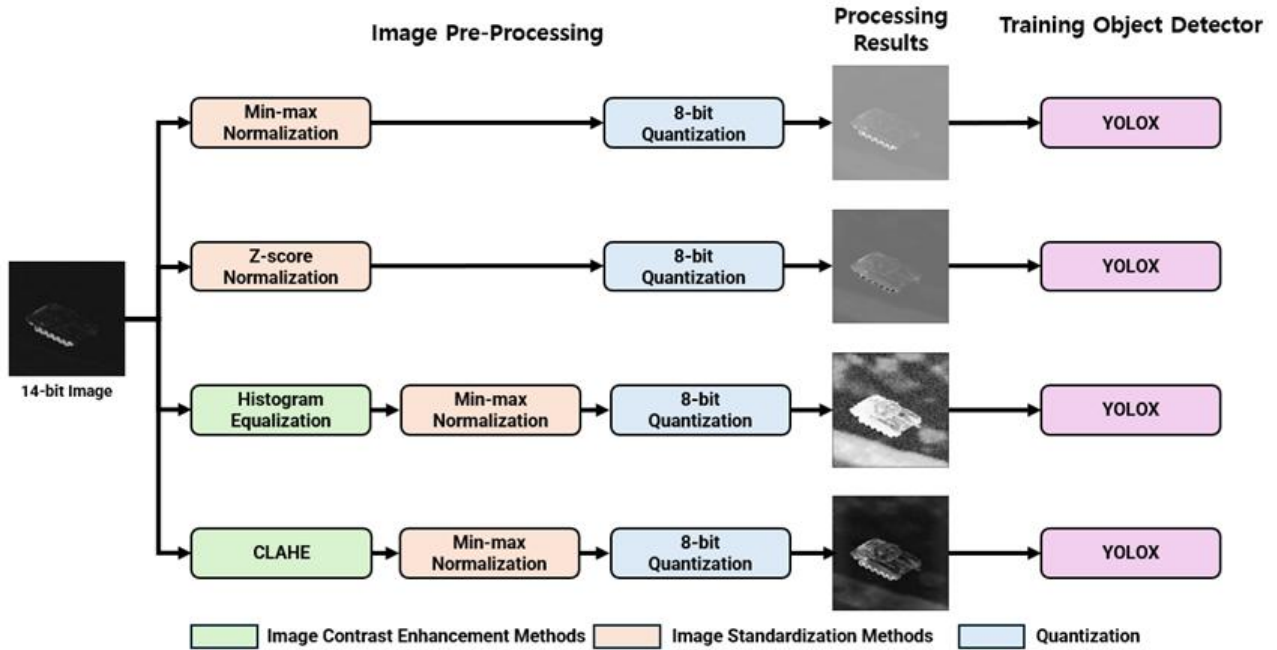


Fig. 5. The process for IR image pre-processing for training the object detector. We first apply min-max normalization, Z-score normalization, CLAHE, and histogram equalization to 14-bit images. Each processed image by histogram equalization and CLAHE is then standardized using min-max normalization, followed by 8-bit quantization.

Fig. 3에서 확인할 수 있듯이 일부 조건에서는 배경과 객체의 구분이 어려울 정도로 명함 대비가 발생하지 않는 문제를 나타냈다. 각 기법이 제공하는 장점에도 불구하고, 이들은 특정 환경에서 제한점이 있음을 확인하였다.

IV. Experiment results

본 장에서는 제안한 객체 검출기의 적외선 영상 학습을 위한 데이터 전처리 기법의 효과를 보이기 위한 실험 결과를 보인다.

1. YOLOX

본 논문에서는 유도무기 환경 적외선 영상 내 객체 검출을 수행하기 위한 객체 검출기로서 YOLOX 모델을 활용하였다. YOLOX는 대표적인 one-stage 객체 검출기인 YOLO (You Only Look Once) [24] 를 개량한 모델이다. YOLO 모델은 이미지 전체를 단 한 번만 보고 객체를 검출할 수 있는 효율적이고 상대적으로 빠른 속도의 객체 검출기로서 산-학 분야에서 많이 활용되었다. 하지만, 고정된 앵커(anchor) 박스 크기 및 비율에 의존하는 한계점이 존재한다. 이러한 문제는 특히, 다양한 크기 및 형태의 객체가 존재하는 환경에서는 모델 성능이 저하될 수 있다.

YOLOX 모델은 기존 객체 검출 모델이 사용하는 앵커(anchor) 기반 객체 검출기의 단점을 극복하기 위해 앵커

프리(anchor free) 기반의 객체 검출 알고리즘을 활용하였다. 앵커 프리 기반 객체 검출기는 사전에 정의된 앵커 박스를 활용하지 않고, 이미지 내의 임의의 위치에서 객체의 위치 및 클래스 추론을 수행한다.

기존 YOLO 모델 대비 YOLOX 모델의 특징적인 구조적 차이점은 decoupled head 구조다. 기존 YOLO 모델의 경우 객체의 검출을 위한 클래스 분류(classification)과 지역화(localization) 태스크를 하나의 결합된 헤드 네트워크(head network)를 통해 처리하였다. 그러나, 본 YOLOX 모델에서는 클래스 분류 및 지역화를 위한 헤드 네트워크를 분리하여 처리하도록 설계하였다. 분리된 헤드 네트워크 구조를 활용하여 YOLOX는 두 네트워크가 각 태스크에 최적화된 학습 및 추론을 할 수 있도록 설계되었다.

2. Implementation Details

본 논문에서는 4.1절에서 설명한 YOLOX 객체 검출기를 활용하여 min-max normalization, Z-score normalization, histogram equalization, CLAHE 전처리 기법으로 생성된 영상을 학습하였다. YOLOX 객체 검출기 학습은 데이터 셋은 3.1절에서 설명한 유도무기 환경 적외선 영상 데이터 셋을 활용하였다. 본 데이터 셋의 1,952개의 영상(학습: 1,172개, 테스트: 780개)에서 10프레임 간격으로 샘플링을 수행해 데이터를 확보하여 학습 시 96,017장의 이미지를 학습에 65,209장의 이미지를 테스트에 활용하였다. 이때, 이미지는 Fig. 5와 같이 생성 데

Table 1. Comparison detection results between various image pre-processing methods. We highlight the best record with **red bold**

Method	mAP	mAP ₅₀	mAP ₇₅	mAP _{small}	mAP _{medium}	mAP _{large}
Min-max normalization	64.3	76.8	71.3	52.9	80.0	81.8
Z-score normalization	56.6	72.5	65.0	43.5	72.8	73.9
Histogram equalization	75.7	89.6	86.8	64.1	89.6	92.4
CLAHE	82.6	95.0	91.7	71.8	93.6	94.4

이터 셋인 14-bit 영상에 각 전처리 기법을 적용한 후 대부분의 딥러닝 모델이 8-bit 데이터 셋으로 학습하기 때문에, 같은 학습 환경을 만들기 위해서 데이터 셋에 대해서 8-bit quantization을 진행하였다. 이때, histogram equalization과 CLAHE의 경우 14-bit 영상에서 각 기법을 적용한 뒤, min-max normalization을 적용하였다. min-max normalization도 8-bit 데이터 셋과 동일한 학습 환경을 위해 진행하였다. 또한, CLAHE 기법의 경우 대비 제한 값을 2.0으로 설정하였으며, 패치 크기는 4×4 로 설정하였다. 이는 파라미터를 튜닝하며 가장 높은 성능을 달성한 값으로 설정하였다.

YOLOX 객체 검출기 학습 시 25 epoch 학습을 수행하였으며, 학습률(learning rate)은 0.01로 설정하였다. 또한, weight decay는 0.0005로 설정하였다. 이 값은 YOLOX 설정 기본값으로, 기존에 보편적으로 사용되고 가장 높은 성능을 달성한 설정값을 사용하였다. 모든 코드는 Python 언어를 통해 구현되었으며, 객체 검출기 모델은 Pytorch로 구현되었다. 본 실험은 Ubuntu 22.04.01 환경에서 실험이 진행되었으며, NVIDIA GeForce 3090 GPU를 활용하여 수행되었다.

3. Benchmark Results between Pre-processing Methods

본 절에서는 min-max normalization, Z-score normalization, histogram equalization, CLAHE를 통해 전처리 수행한 이미지를 통해 YOLOX 객체 검출기를 학습한 결과에 관한 결과를 분석한다. 학습 결과는 Table 1에서 확인할 수 있다. 먼저, 전역적 정보를 활용해 이미지 전처리를 수행하는 Z-score normalization은 56.6 mAP로 가장 낮은 성능을 보인다. 이는 3장 2절에서 설명한 바와 같이 영상 내 다양한 표적과 유도무기 사이의 거리 및 진입 각도가 존재하는 유도무기 적외선 영상 특성상 Z-score normalization이 적용되어도 데이터 셋의 전역적인 정보를 전부 반영하는 만큼, 다양한 도메인이 존재한다. 사용한 합성 데이터 셋 내에 존재하는 다양한 객체, 유도무기 사이의 거리 및 진입 각도와 진입 각도에 따라 바

뀌는 객체의 세부적인 정보가 덜 반영되었기 때문에 다른 전처리에 비해서 성능이 낮게 나타난 것으로 사료된다.

이미지의 지역적 정보를 활용하는 min-max normalization, histogram equalization, CLAHE의 경우 각각 64.3, 75.7, 81.9의 높은 mAP 성능을 보여준다. 이는, 적외선 영상의 전처리 과정에서 이미지의 지역적 특성을 잘 반영해야 하는 것이 표적 객체의 특징을 잘 표현할 수 있다는 것을 보여준다. 이 중, min-max normalization은 3장 2절에서 상술한 배경/객체 간 대비가 잘 이루어지지 않는 문제점 때문에, 이미지의 대비를 강조하는 나머지 두 기법 대비 낮은 정확도를 보여주고 있다. 특히, 작은 객체(mAP_{small})상에서의 정확도가 41.6 mAP로 여타 비교 대상 기법들 가운데서 가장 낮은 성능을 보인다. Histogram equalization과 CLAHE의 경우 두 기법 모두 이미지의 대비를 향상을 목적으로 하는 기법이지만, 전자의 경우 3장 2절에서 보인 바와 같이 이상치에 인한 노이즈 발생 문제와 객체의 세부 특성을 잘 표현하지 못하는 경우가 있어 CLAHE보다 낮은 정확도를 보인다. CLAHE는 지역적인 정보를 통해 객체의 특성을 상대적으로 잘 표현하는 특성이 있어 모든 부분에서 가장 높은 정확도를 보인다.

본 절에서는 YOLOX 객체 검출기를 통한 비교 실험을 통해 가장 지역적인 정보를 활용해 배경과 객체 간의 대비를 명확하게 해주며, 객체의 특성을 잘 표현하는 CLAHE 기법이 유도무기 환경 적외선 영상에서 가장 적합한 전처리 기법임을 보였다. 그러나, 3장 2절에서 지적한 바와 같이 CLAHE 역시 특정 환경에서 배경과 객체 간 구분이 어려운 경우가 발생할 수 있기에 영상의 환경을 잘 고려하여 전처리 기법을 선정해야 한다.

V. Conclusions

본 논문에서는 유도무기의 표적 검출을 위해 적외선 영상을 기반으로 한 객체 검출기 학습에 중점을 두고, 다양한 이미지 전처리 기법의 효과를 연구 및 분석하였다. 구

체적으로, 전역적 정보를 활용하는 Z-score normalization과 지역적 정보를 활용하는 min-max normalization, histogram equalization, CLAHE 기법을 설명하고, 이들이 적외선 영상 환경에 적용된 결과를 분석하였다. 이어서, YOLOX 객체 검출기를 활용하여 각 전처리 기법의 실제 학습 효과를 정량적으로 평가하였다. 해당 실험을 통해, 영상의 도메인이 일정하지 못한 유도무기 환경의 적외선 영상에서는 이미지의 지역적 특성을 활용하는 기법들이 전역적인 정보를 활용하는 Z-score normalization보다 효과적임을 확인하였다. 특히, 이미지의 대비와 객체 특성을 개선하는 CLAHE 기법이 가장 높은 검출 정확도를 달성하였음을 확인하였다. 그러나, CLAHE 기법도 특정 환경에서는 객체와 배경 간의 대비가 부족하여 학습에 부정적인 영향을 미칠 수 있다는 한계를 확인하였다. 향후 연구에서는 이러한 전역 및 지역적 전처리 기법들의 장점을 결합할 수 있는 새로운 알고리즘을 개발하여 다양한 환경에서 객체의 특징을 더욱 효과적으로 표현할 수 있도록 하는 것을 목표로 한다. 또한, 영상은 해상도 측면에서 화소값이 정수에서 실수로 변환되면 모델의 정확도가 높아지는 경향이 존재하기 때문에, 이를 반영한 연구를 진행하는 것을 목표로 한다. 해당 연구들은 유도무기의 자동 표적 검출 성능을 향상하는 데 이바지 할 것으로 기대된다.

ACKNOWLEDGEMENT

This research is performed based on the cooperation with Inha University-LIG Nex1 Cooperation.

REFERENCES

- [1] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, Piotr Dollar, "Focal Loss for Dense Object Detection," Proceedings of the IEEE International Conference on Computer Vision, pp.2980-2988, Venice, Italy, October 2017. DOI: 10.1109/ICCV.2017.324
- [2] Ge Zheng, Liu Songtao, Feng Wang, Zeming Li, Jian Sun, "YOLOX: Exceeding YOLO Series in 2021," arXiv preprint arXiv:2107.08430, pp.1-7, July 2021. DOI: <https://doi.org/10.48550/arXiv.2107.08430>
- [3] Hao Zhang, Feng Li, Shilong Liu, Lei Zhang, Hang Su, Jun Zhu, Lionel M. Ni, Heung-Yeung Shum "DINO: DETR with Improved DeNoising Anchor Boxes for End-to-End Object Detection," International Conference on Learning Representations, pp.1-23, March 2022. DOI: <https://doi.org/10.48550/arXiv.2203.03605>
- [4] Janghoon Kim. "A Study on Sustainable Competitive Advantage of Defense Venture Companies: Focusing on the Case of i3system Inc.," Seminar and Discussion of the Korea Trade Association, pp.213-229, Chungnam University, Korea, March 2017.
- [5] Akula, Aparna, Ripul Ghosh, and H. K. Sardana. "Thermal imaging and its application in defence systems," AIP conference proceedings, Kolkata, India, pp.333-335, March 2012. DOI: <https://doi.org/10.1063/1.3643540>
- [6] Zixuan Zhang, Jiong Huang, Gawen Hei, Wei Wang, "YOLO-IR-Free: An Improved Algorithm for Real-Time Detection of Vehicles in Infrared Images," Sensors, Vol.23, pp.1-23, 2022. DOI: <https://doi.org/10.3390/s23218723>
- [7] Du, Shuangjiang, et al. "FA-YOLO: An improved YOLO model for infrared occlusion object detection under confusing background," Wireless Communications and Mobile Computing, Vol.2021, pp.1-10, October, 2021. DOI: <https://doi.org/10.1155/2021/1896029>
- [8] Lixin Zhang, Qingrong Nan, Shengging Bian, Tao Liu, Zhengguang Xu, "Real-time segmentation method of billet infrared image based on multi-scale feature fusion," Scientific Reports, Vol.18, pp.1-7, October 2022. DOI: <https://doi.org/10.1038/s41598-022-09233-6>
- [9] Yang, Guang, et al. "A Multi-scale Information Integration Framework for Infrared and Visible Image Fusion," arXiv preprint arXiv:2312.04328, December 2023. DOI: <https://doi.org/10.48550/arXiv.2312.04328>
- [10] Zhengzheng Tu, Yan Ma, Zhun Li, Chenglog Li, Jieming Xu, Yongtao Liu, "RGBT Salient Object Detection: A Large-Scale Dataset and Benchmark," IEEE Transactions on Multimedia, Vol.25, pp.4163-4176, May 2022. DOI: 10.1109/TMM.2022.3171688
- [11] Soonmin Hwang, Jaesik Park, Namil Kim, Yukyung Choi, In So Kweon, "Multispectral Pedestrian Detection: Benchmark Dataset and Baseline," Proceedings of the Conference on Computer Vision and Pattern Recognition, pp.1037-1045, Boston, USA, June 2015 DOI: 10.1109/CVPR.2015.7298706
- [12] Chenglong Li, Xinyan, Yijuan Lu, Nan Zhao, Jin Tang, "RGB-T object tracking: Benchmark and baseline," Pattern Recognition, Vol.96, pp.1-12, December 2019. DOI: <https://doi.org/10.1016/j.patcog.2019.106977>
- [13] Chenglong Li, Wanlin Xue, Yaqing Jia, Zhichen Qu, Bin Luo, Jin Tang, Dengdi Sun, "LasHeR: A Large-Scale High-Diversity Benchmark for RGBT Tracking," IEEE Transactions on Image Processing, Vol.31, pp.392-404, April 2021. DOI: 10.1109/TIP.2021.3130533

- [14] Jianqiang Xia, Dianxi Shi, Ke Song, Linna Song, XiaoLei, "Unified Single-Stage Transformer Network for Efficient RGB-T Tracking," arXiv preprint arXiv:2308.13764, August 2023. DOI: arXiv:2308.13764
- [15] R. Gonzalez, R. Wood, "Digital Image Processing, 3rd ed.," Pearson Education, U.K., 2009
- [16] S.M. Pizer, R.E. Johnston, J.P. Ericksen, B.C. Yankaskas, K.E. Muller, "Contrast-limited adaptive histogram equalization: speed and effectiveness," Proceedings of the First Conference on Visualization in Biomedical Computing, GA, USA, pp.337-345, May 1990, DOI: 10.1109/VBC.1990.109340
- [17] Huang, Lei, et al. "Normalization techniques in training dnns: Methodology, analysis and application," IEEE transactions on pattern analysis and machine intelligence, Vol.45, pp.10173-10196, August 2023. DOI: 10.1109/TPAMI.2023.3250241
- [18] Kandanaarachchi, Sevvandi, et al. "On normalization and algorithm selection for unsupervised outlier detection," Data Mining and Knowledge Discovery, Vol.34, pp.309-354, November 2019. DOI: <https://doi.org/10.1007/s10618-019-00661-z>
- [19] Jae-Won Lee, Sung-Hoon Hong, "Histogram Equalization based on Differential Compression for Image Contrast Enhancement," JOURNAL OF BROADCAST ENGINEERING, Vol.19, pp.96-108, January 2014 DOI: 10.5909/JBE.2014.19.1.96
- [20] Byong-Seok Min, Tea-Kyung Cho, "A Novel Method of Determining Parameters for Contrast Limited Adaptive Histogram Equalization," Journal of the Korea Academia-Industrial cooperation Society, Vol.14, pp.1378-1387, March 2013. DOI: 10.5762/KAIS.2013.14.3.1378
- [21] Hou, Fujin, et al. "Review on infrared imaging technology," Sustainability, Vol.14, pp.1-26, September 2022. DOI: <https://doi.org/10.3390/su14181116>
- [22] Shu, Hai, and Hongtu Zhu. "Sensitivity analysis of deep neural networks," Proceedings of the AAAI Conference on Artificial Intelligence, pp.4943-4950, Hawaii, USA, July 2019. <https://doi.org/10.1609/aaai.v33i01.33014943>
- [23] Hyunji Cho, Heewon Kye, "EThe Clip Limit Decision of Contrast Limited Adaptive HistogramEqualization for X-ray Images using Fuzzy," Journal of Korea Multimedia Society, Vol.18, No.7, pp.806-817, July 2015. DOI: <http://dx.doi.org/10.9717/kmms.2015.18.7.806>
- [24] Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," Proceedings of the Conference on Computer Vision and Pattern Recognition, pp.779-788, Las Vegas, USA, June 2016. DOI: <https://doi.org/10.48550/arXiv.1506.02640>

Authors



Jae-Yong Baek received the M.S degree in Computer Science and Engineering from Incheon Nation University, and worked as a software engineer that develop the software for object detection, lane detection and

semantic segmentation in autonomous vehicle startup. He is currently pursuing the Ph.D. degree with the Department of Electronic Computer Engineering at Inha University, Korea. His current research interests are object detection, generative model, quantization and optimization for edge AI.



Dae-Hyeon Park received the B.S degree in Computer Engineering from Inha University in 2020 and the M.S degree with the Department of Electronic Computer Engineering at Inha University in 2023.

He is currently pursuing the Ph.D. degree with the Department of Electronic Computer Engineering at Inha University, Korea and His current research interests are single/multi-object tracking, multi-modal learning, real-time system and self-attention mechanism.



Hyuk-Jin Shin received the B.S. degree with Department of Computer Science, Chung-Buk National University, South Korea. He is currently pursuing the M.S. - Ph.D. integrated course degree.

His research interests are object detection and continual learning.



Yong-Sang Yoo received the BS degree in Computer Science and Engineering from Incheon National University in 2021, and is currently pursuing the MS and PhD integrated degree with the Department of

Electronic Computer Engineering at Inha University, Korea. His current research interests include object detection, machine learning, continual learning, on-board AI and deep learning.



Deok-Woong Kim received the B.S. degree with Department of Industrial engineering, Public administration, Inha University, South Korea. He is currently pursuing the M.S. - Ph.D. integrated course degree.

His research interests are knowledge distillation, quantization, data-free and on-board AI.



Du-Hwan Hur received the B.S. degree in Computer Engineering from Hanbat National University in 2021, and is currently pursuing the M.S. - Ph.D. integrated course degree with the Department of Electronic Computer

Engineering at Inha University, Korea. His current research interest is objection detection, multi-object tracking, and deep learning.



SeungHwan Bae received the B.S and M.S degrees in Medical IT Engineering and Future Convergence Technology from Soonchunhyang University, Korea, in 2019 and 2021, respectively.

SeungHwan Bae joined the Senior Researcher of the Performance analysis Team at LIGNEX1, Pangyo, Korea, in 2022. He is currently a Senior Researcher in the Performance Analysis Team, LIGNEX1. He is interested in AI and Image Processing.



Jun-Ho Cheon received the B.S. degree in Computer Science and Engineering from Busan University of Foreign Studies, Korea, in 2012. Jun-Ho Cheon joined the Senior Researcher of the Performance analysis Team

at LIGNEX1, Pangyo, Korea, in 2022. He is currently a Senior Researcher in the Performance Analysis Team, LIGNEX1. He is interested in parallel computing, detection and tracking.



Seung-Hwan Bae received the BS degree in information and communication engineering from Chungbuk National University, in 2009 and the MS and PhD degrees in information and communications from the Gwangju

Institute of Science and Technology (GIST), in 2010 and 2015, respectively. He was a senior researcher at Electronics and Telecommunications Research Institute (ETRI) in Korea from 2015 to 2017. He was an assistant professor in the Department of Computer Science and Engineering at Incheon National University, Korea from 2017 to 2020. He is currently an Associate Professor with the Department of Computer Engineering at Inha University, His research interests include object tracking, object detection, generative model learning, continual learning, on-device ML, etc.