

적외선 영상의 전처리 기법의 학습 유무에 따른 객체 탐지 딥 러닝 모델의 성능 비교

배재현*, 강병진, 김대현, 백경훈
LIG 넥스원

jaehyun.bae@lignex1.com

Comparison of Object Detection Deep Learning Model Performance Based on the Presence and Absence of Preprocessing Methods for Infrared Image

Bae Jaehyun*, Kang Byungjin, Kim Daehyeon, Baek Kyounghoon
LIGNex1

요약

본 논문은 객체 탐지 딥 러닝 모델의 입력으로 사용할 적외선 영상에 세 가지의 영상 전처리 기법을 적용한 영상을 각각 학습한 이후, 학습을 하지 않은 전처리 영상을 모델의 입력으로 사용했을 때의 성능 차이를 확인하였다. YOLOX 객체 탐지 모델에서 세 가지 기법 중 한 개의 전처리 기법만 활용해야 한다면 Min-max 기법이 가장 범용적으로 활용할 수 있다는 것을 실험적으로 확인하였다.

I. 서론

적외선 영상은 일반적인 RGB 3 개의 채널을 가지는 영상과 다르게 한 개의 픽셀이 14bit 1 채널을 가진다. 따라서, 딥 러닝에 적외선 영상을 활용하기 위해서는 14bit 영상을 8bit 표현으로 변경하는 영상 평탄화 및 정규화의 전처리 과정이 필수적이다[1]. 이 과정에서 전처리 기법이 딥 러닝 모델에 어떠한 영향을 미치는지 이전 연구에서 다루지 못했다. 본 논문에서는 객체 탐지 딥 러닝 모델 중 YOLOX 모델을 활용하여 전처리 기법에 대한 성능 분석을 했던 선행연구[2]를 이어, 학습과 추론에서 동일한 전처리 기법이 아닌 상이한 기법을 적용했을 때 탐지 성능이 유사한지 실험적으로 검증하였다.

II. 본론

본 논문에서 사용한 영상 데이터는 LIG 넥스원 용인하우스 앞에서 획득하였다. 총 9,180장의 영상을 수집하여 차량과 사람 두 가지의 종류를 레이블링하여 학습 및 추론을 수행하였다. 학습과 검증, 테스트 셋의 비율은 약 6:1:3이다. 이는 획득한 데이터의 특징이 시각적으로 다르거나 거리에 따른 객체의 크기를 상이하게 하기 위함이다.

데이터 획득에 사용한 카메라는 아이쓰리시스템의 TE-EV1을 사용하였다. LWIR(Long wave IR)의 파장대역을 가지며, 영상의 해상도는 640x480이다.

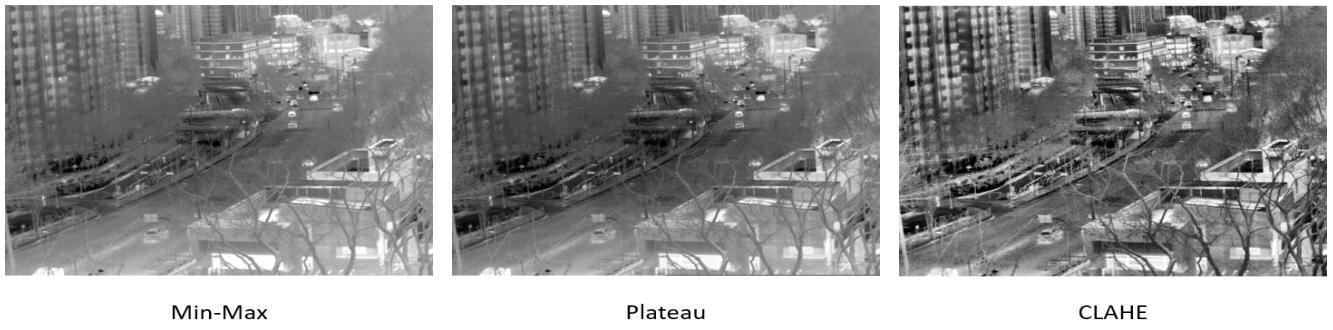
본 논문에서 사용한 전처리 기법은 총 세 가지로 min-max, plateau, CLAHE(Contrast limited adaptive histogram equalization)이다. 그림 1은 획득한 영상에 각각의 전처리를 적용했을 때의 예시를 보여준다.

min-max 기법은 작은 차량의 지붕 등에서 방사되는 강한 열이 최댓값으로 식별되어 영상이 포화되는 경우가 있어 이상치 제외를 위해 영상의 최댓값과 최솟값을 화소값의 평균과 표준편차로 가공하여 사용하였다. Plateau 기법에서 가장 중요한 파라미터는 문턱값이다. 문턱값을 어떤 값으로 적용하는지에 따라 전처리에 적용하는 화소값의 기준치가 변경되기 때문이다. 본 논문에서는 기준치를 이미지의 크기에 0.001을 곱하고 0.5를 더해준 값을 적용하였다. CLAHE는 공개 코드를 활용하였으나 14bit 영상이기 때문에 코드를 일부 수정하여 사용하였다. 구현에 활용한 코드는 아래와 같다.

<https://github.com/joshdoe/opencv-clahe/blob/master/clahc.cpp>

본 논문에서 사용한 모델은 범용적으로 사용되고 있는 YOLOX이다. 학습은 YOLOX의 저자가 추천하는 파라미터를 적용하였으며, 1024x768의 영상을 640x640으로 리사이징하여 네트워크의 입력으로 사용하였다. 배치 크기는 8, 학습 횟수는 300회를 수행하였다. 모델의 크기는 파라미터의 개수에 따라 나뉘는데, 가장 작은 모델은 YOLOX-S(small)를 적용하여 9 M개라는 비교적 적은 파라미터 수를 가질 때의 성능을 검증하였다.

그림 2는 본 논문의 실험 과정을 설명한다. 14bit의 적외선 영상을 세 가지의 전처리 기법을 통해 8bit로 변환하여 YOLOX에 학습 영상으로 사용하여 각 전처리 기법의 영상에 대한 가중치 파일을 생성한다. 이후, 각 전처리 기법을 학습하지 않은 영상을 입력 영상으로 사용하여 상이한 전처리 기법이 성능에 얼마나 영향을 미치는지 확인하였다. 예를 들어, min-max로 학습한 가중치로 추론할 때는 plateau와 CLAHE로 전처리를



Min-Max

Plateau

CLAHE

Fig. 1. The example of the preprocessed Images

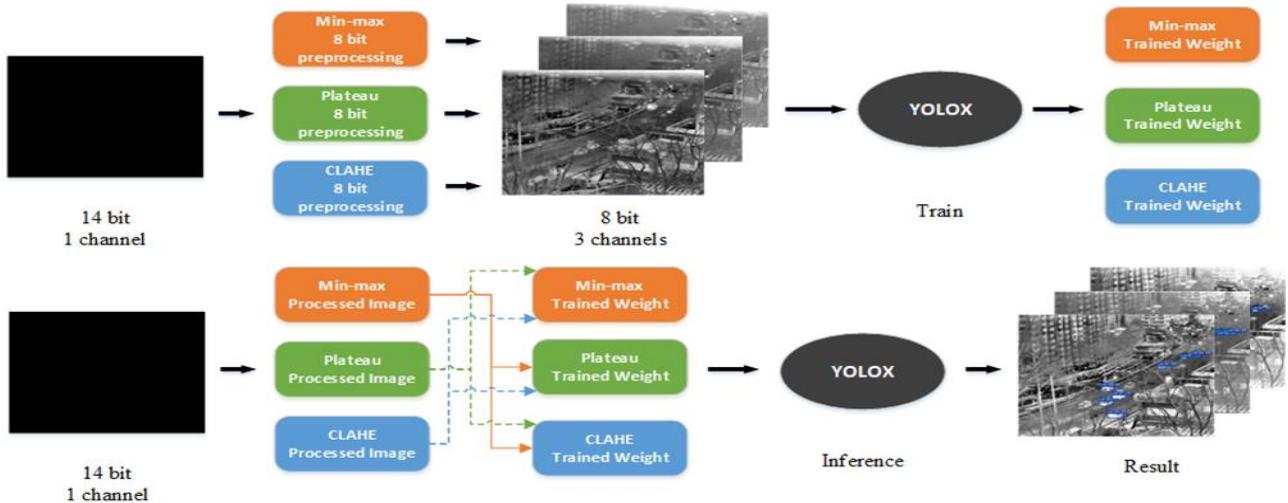


Fig. 2. Framework of the experiment: Train phase(upper) and inference phase(lower)

Table 1. mAP_{50} of each preprocessing methods

추론영상 가중치파일	Min-max	Plateau	CLAHE
Min-max	0.81 ($\Delta 0$)	0.81 ($\Delta 0.1$)	0.71 ($\Delta 0.1$)
Plateau	0.635 ($\Delta 0.135$)	0.77 ($\Delta 0.14$)	0.63 ($\Delta 0.14$)
CLAHE	0.65 ($\Delta 0.155$)	0.68 ($\Delta 0.125$)	0.805

수행한 영상을 입력으로 사용하였다.

표 1은 각 전처리 기법에 따른 mAP를 보여준다. 행은 가중치 파일, 즉 학습에 사용한 전처리 기법이고, 열은 추론에 사용한 전처리 기법이다. 각 mAP 아래에는 가중치 파일을 기준으로 타 전처리 기법을 추론했을 때 얼마나 차이가 나는지 차이 값을 표시하였다. 따라서, 대각선은 학습과 추론 과정에서 같은 전처리 기법을 적용한 값이다. 기준 딥 러닝의 특성과 마찬가지로 학습과 추론에 같은 전처리 기법을 사용했을 때 각각 가장 높은 성능을 보여주었다. 특이하게 min-max 기법으로 학습을 수행하고 다른 기법으로 추론을 진행했을 때 학습-추론에서 같은 기법을 적용한 수치를 제외하고 가장 좋은 성능을 보였다. 하지만 plateau 나 CLAHE 기법에서는 상이한 전처리 기법을 사용한 경우 최대 약 15.5%의 성능 저하가 발생하였다. CLAHE 기법은 영상 전체가 아니라 8x8 등의 그리드 내에서 정규화를 수행하기 때문에 영상 전체의 정보를 활용하는 min-max 와 plateau 와 크게 다를 수 있다. 실험 결과를

통해 사용하는 데이터 셋에 따라 다를 수 있으나 여러 전처리 기법을 학습하기 어려운 상황일 경우 min-max 기법을 사용하는 것이 mAP 측면에서 가장 좋은 성능을 보여줄 수 있음을 시사한다.

III. 결론

본 논문에서는 객체 탐지 딥 러닝 모델인 YOLOX에 학습 및 추론 과정에서 서로 상이한 영상 전처리 기법을 사용했을 때 모델의 성능이 얼마나 달라지는지 확인해보았다. 실험 결과를 통해 min-max 기법을 학습 과정에서 사용하고 추론 과정에서 타 전처리 기법을 적용하더라도 성능 손실이 가장 적은 것을 확인할 수 있었다. 추후에는 학습 과정에서 여러 전처리 기법을 섞어서 사용했을 때 성능 변화가 있는지 검증할 예정이다.

참 고 문 헌

- [1] B.J Kang, J. Bae, D. Kim, & K. Baek. "Performance Evaluation of Deep Learning Based Object Detection Model on Image Infrared Preprocessing". JKSSAS, 52(12), 1055-1061
- [2] J. Bae, B. J. Kang, D. Kim, & K. Baek. "Performance Evaluation of LWIR Image Detection Using Fine-tuning of YOLOX Model". JCROS, 30(7), 685-690.