

# Mask R-CNN에 의한 자동차 탐지에서 학습 영상 화면 축척과 촬영계절이 정확도에 미치는 영향 분석

## Analysis of the Effect of Learned Image Scale and Season on Accuracy in Vehicle Detection by Mask R-CNN

최주영<sup>1)</sup> · 원태연<sup>2)</sup> · 어양담<sup>3)</sup>  
Choi, Jooyoung · Won, Taeyeon · Eo, Yang Dam

### Abstract

In order to improve the accuracy of the deep learning object detection technique, the effect of magnification rate conditions and seasonal factors on detection accuracy in aerial photographs and drone images was analyzed through experiments. Among the deep learning object detection techniques, Mask R-CNN, which shows fast learning speed and high accuracy, was used to detect the vehicle to be detected in pixel units. Through Seoul's aerial photo service, learning images were captured at different screen magnifications, and the accuracy was analyzed by learning each. According to the experimental results, the higher the magnification level, the higher the mAP average to 60%, 67%, and 75%. When the magnification rates of train and test data of the data set were alternately arranged, low magnification data was arranged as train data, and high magnification data was arranged as test data, showing a difference of more than 20% compared to the opposite case. And in the case of drone images with a seasonal difference with a time difference of 4 months, the results of learning the image data at the same period showed high accuracy with an average of 93%, confirming that seasonal differences also affect learning.

Keywords : Deep Learning, Mask R-CNN, Remote Sensing, Object Detection

### 초 록

본 연구에서는 딥러닝 객체탐지 기법의 정확도 향상을 위해 항공사진과 드론 영상을 대상으로 확대율 조건과 계절요인이 탐지정확도에 미치는 영향을 실험을 통해 분석하였다. 딥러닝 객체탐지기법 중 빠른 학습 속도와 높은 정확도를 나타내는 Mask R-CNN을 사용하여 탐지대상인 자동차를 픽셀 단위로 탐지하고자 하였다. 서울시 항공사진서비스를 통해 화면 확대 레벨을 달리하며 학습 영상을 캡처하고 각각을 학습하여 정확도를 분석하였다. 실험결과에 따르면 확대 레벨이 높아질수록 mAP 평균이 60%, 67%, 75%로 높아졌다. 데이터 세트의 train, test 데이터의 확대율을 엇갈려서 배치한 경우에는 확대율이 매우 낮은 경우를 제외하고 저배율의 데이터를 train 데이터로, 고배율의 데이터를 test 데이터로 배치하였을 때 높은 mAP로 반대의 경우보다 20% 이상 차이를 보였다. 그리고 4개월의 시차로 계절적 차이를 두고 촬영한 드론 영상의 경우, 같은 시기 영상자료 학습결과가 평균 93%로 높은 정확도를 나타내어 계절적 차이도 학습에 영향을 주는 것을 확인되었다.

핵심어 : 딥러닝, Mask R-CNN, 원격탐사, 객체탐지

Received 2022. 01. 06, Revised 2022. 02. 07, Accepted 2022. 02. 17

1) Dept. of Environmental Health Science, Dept. of Technology Fusion Engineering, Undergraduate student, Konkuk University, Seoul, Korea (E-mail: qzwsx0116@konkuk.ac.kr)

2) Dept. of Advanced Technology Fusion, M.S. student, Konkuk University, Seoul, Korea (E-mail: teadone@konkuk.ac.kr)

3) Corresponding Author, Member, Dept. of Civil and Environmental Engineering, Professor, Konkuk University, Seoul, Korea (E-mail: eoandrew@konkuk.ac.kr)

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

## 1. 서론

객체탐지(object detection)는 영상이나 동영상에서 대상 객체의 종류와 그 위치를 정확하게 찾는 기술로써 기존에는 배경과 객체를 분리하고 획득된 영상을 흑백 영상 혹은 영상 이진화, 특징점 추출 등 영상 내에서 해당 특징을 검출하는 다소 번거로운 방법으로 연구가 진행됐다(Lee *et al.*, 2013). 이러한 복잡한 과정의 문제점을 해결하고 다양한 환경 아래(내)에서의 탐지정확도 향상을 위해 최근에는 인공지능(AI) 분야가 꾸준히 연구되었으며 영상 딥러닝 기술이 대두하고 있다(Gu, 2018).

딥러닝에 의한 탐지정확도에 미치는 요인 중 하나가 학습데이터이다. 학습데이터는 다양성이 특징이기 때문에 데이터의 양과 오류율 등 다양한 조건들이 분류 및 탐지정확도에 영향을 미칠 수 있다. Kim *et al.* (2018)은 기계학습과 딥러닝 모두에서 훈련자료가 증가할수록 분류 성능이 상승하는 경향을 보이는 것을 밝혔다. 또한, 다량의 훈련자료를 수집하더라도 분광학적 구분이 낮게 나타나는 일종의 잡음 효과가 있을 때 분류 성능이 저하될 수 있다고 하였다(Kim *et al.*, 2018). 이와 함께 Song *et al.* (2021)은 항공사진의 오류율을 낮추기 위해 품질검사 자동화의 필요성을 제시하며 학습데이터로 항공사진과 위성영상을 추가 활용하며 항공사진 자료의 부족 시 탐지정확도 향상 방안을 연구하였다.

항공 영상을 활용하여 차량의 객체탐지 자동화는 교통정보를 반영하는 GeoAI 기반 도시 환경 구현에서 필수적이다. 객체탐지에 유용한 Mask R-CNN을 이용하여 차량을 검출하는 연구가 진행되었으며 제한된 양의 학습데이터에 따른 탐지정확도 향상을 위한 추가적인 연구의 필요성을 제시하였다(Youn *et al.*, 2019). 빠르게 변화하는 특징이 있는 도시 지역에서 시간의 경과와 관계없이 객체를 효과적으로 탐지하는 것은 중요하다. Jeong *et al.* (2008)은 항공, 위성 영상은 자연환경, 시간적인 변화에 취약한 특성을 언급하며 항공 레이저 스캐너로부터 획득 가능한 LiDAR 데이터를 활용하여 시간적인 변화를 탐지하는 연구를 진행하였다. Kim *et al.* (2021)은 항공사진 데이터의 취득 시기에 따른 분류 결과의 변동성을 소개하며 이를 개선하는 방안을 연구하였다. 동일 지역의 계절적 차이를 고려한 다중 시계열 학습으로 분류 정확도를 향상시켜 계절에 따른 탐지 오차를 줄였다(Kim *et al.*, 2021).

항공 영상은 공개된 자료가 많지 않고 수집과 레이블링 비용 부담이 있어 정부 차원의 제재가 이루어지고 있으나 현재까지 현실적으로 충분한 자료가 확보되지 않고 있다(Park *et*

*al.*, 2021). 이를 보완하기 위해 항공 영상과 위성 영상의 이미지 단위의 공개 자료의 활용도 고려해야 하고 이에 대한 유효성도 살펴볼 필요가 있다. 따라서 본 연구에서는 원격탐사 영상자료로서 원자료 대신 접근성이 쉬운 공개포털에서 캡처한 영상으로 학습을 진행하였으며 제한된 데이터 조건에서의 탐지정확도 향상에 중점을 두어 분석하였다. 항공 영상에서 Mask R-CNN을 활용하여 차량의 객체를 탐지하며 공개 자료에서 조절 가능한 축척(확대율) 변화와 조합에 따른 탐지정확도의 차이를 분석한다. 또한, 드론으로 촬영된 영상의 시간적인 변화가 탐지정확도에 미치는 영향을 확인한다.

## 2. Mask R-CNN 에 의한 원격탐사 영상 객체 탐지

컴퓨터 비전 분야에서 이미지를 분류하고, 객체를 탐지하는 것은 항상 중요한 문제였으며, 이미지 분류를 낮은 부하와 높은 정확도로 효율적인 분류를 진행하기 위해 기존의 FCN(Fully Connected Network) 대신, 합성곱 필터(Convolutional filter)와 풀링(Pooling)으로 특징을 추출하는 신경망인 CNN(Convolutional Neural Network)이 연구되었다(Lecun *et al.*, 1989). 이후, CNN은 ILSVRC(ImageNet Large Scale Visual Recognition Challenge)라는 이미지 넷 이미지 인식 대회에서 정확도를 개선하기 위한 많은 연구가 진행되었고, 이 대회에서 높은 정확도와 넓은 범용성을 지닌 다양한 CNN들이 개발되어 많은 성과를 얻었다. 따라서 이미지로부터 객체 인식에 뛰어난 성능을 검증받아 딥러닝 분야에서 다방면으로 활용되고 있다(Simonyan and Zisserman, 2015; Szegedy *et al.*, 2015; Russakovsky *et al.*, 2015).

ILSVRC를 통해 CNN이 개선되었지만, 딥러닝의 비전 분야에서 객체탐지는 해결하기 어려운 과제였다. 초기의 객체탐지 모델인 R-CNN은 2013년 발표되었지만, 느린 탐지 속도와 낮은 정확도로 인하여 실시간 객체탐지에 실용적이지 못하였고, 이러한 단점을 개선하기 위해 MR-CNN, AttentionNet, Fast R-CNN, Faster R-CNN, YOLO(You Only Look Once)와 같은 객체탐지 딥러닝 기법이 연구되었다(Girshick *et al.*, 2014; Ren *et al.*, 2015; Redmon *et al.*, 2016).

이들 중 실험에 사용할 Mask R-CNN은 R-CNN의 여러 단점을 보완한 알고리즘으로, 기존 R-CNN 모델과 다른 구조를 지니고 있다. Mask R-CNN은 Faster R-CNN에 픽셀 단위의 segmentation mask를 예측하는 mask branch를 추가하여 기존의 Faster R-CNN이 수행하지 못했던 pixel-to-

pixel alignment를 가능하게 한다. 또한, CNN을 통과하며 RoI Pool 영역의 위치에 생기는 오차를 쌍선형 보간법을 이용하는 RoI align을 통해 mask 정확도를 Faster R-CNN보다 높게 픽셀 위치 추출이 가능하다. 이들을 통해 기존 R-CNN에서 발생하는 비효율적인 학습 과정을 GPGPU (General-Purpose computing on Graphics Processing Units)로 효율적인 학습이 가능하도록 개선되었으며, 객체별로 구분이 가능한 세그멘테이션인 Instance Segmentation이 가능하게 되었다. 이러한 개선들을 통해 빠른 학습, 예측 그리고 높은 정확도를 획득할 수 있었다(He *et al.*, 2017). 이 Mask R-CNN 기법을 활용하여 항공 영상 데이터로 도로 위의 균열을 검출하는 자동검출시스템을 연구하며 기존의 인력, 시간, 비용 문제점을 해결하는 데 사용되고 있다(Lee *et al.*, 2019). 본 논문의 실험에서는 픽셀 단위로 영역을 탐지하고 오차를 줄여 정확도를 높일 수 있는 Mask R-CNN을 이용하여 축척 및 시간 변화에 따른 자동차 객체의 탐지정확도를 알아보려고 한다.

### 3. 실험 및 분석

#### 3.1 실험자료 및 지역

실험자료는 확대율에 의한 탐지정확도 분석에 사용된 서울시 항공사진서비스(<https://aerogis.seoul.go.kr/>) 데이터(Fig. 1)와 계절 차이에 의한 탐지정확도 영향분석을 위한 직철 촬영한 드론 영상이 있다. 항공사진 데이터는 서울특별시를 대상으로 무작위로 임의의 지역을 캡처하여 자료를 수집한다. 원하는 지역과 기간을 설정해준 후, Fig. 1에서 보는 것처럼 오른쪽의 확대 레벨이 표시되어있는 숫자를 조절해서 캡처를 진행하였다.

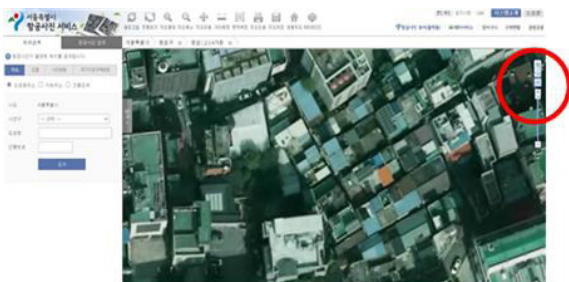


Fig. 1. Seoul metropolitan city aerial photo website (<https://aerogis.seoul.go.kr/>) - magnification rate

자동차를 대상으로 실험을 진행하였으며 탐지 항목에 따

라 선명도, 확대율 등의 실험 영향인자를 미리 설정하여 데이터 수집 시, 효율적인 지도 선정에 고려하도록 한다. 탐지모델 학습데이터는 오픈소스 데이터 주석 도구인 VIA (VGG Image Annotator)를 이용하여 라벨링 하였다. 라벨링이란 학습 및 테스트에 사용할 데이터에 정답을 지정하는 작업으로, 객체탐지에서의 라벨링은 Fig. 2와 같이 목표대상을 사각형, 다각형 등의 도형으로 지정하는 과정을 일컫는 말이다. 준비된 사진 자료와 라벨링을 통해 얻은 json 파일을 이용하여 Mask R-CNN을 지도학습시킨다. 이후 학습된 데이터를 활용하여 test 폴더에 있는 이미지에서 객체탐지를 한다. 이를 통해 자동차가 존재하는 범위를 픽셀 단위로 인식한 예측 이미지가 생성된다. 생성된 예측 이미지를 토대로 객체를 얼마나 정확히 탐지했는지 mAP (Mean Average Precision, %)로 탐지정확도를 수치화한다. mAP는 합성곱 신경망(CNN)의 정확도 성능을 평가하는 단위로 인식 알고리즘의 성능을 하나의 값으로 표현한 AP (Average Precision)의 평균을 지칭한다. 따라서 이 mAP가 높을수록 그 알고리즘의 성능이 전체적으로 우수함을 의미한다.



Fig. 2. Vehicle labeling using via tool

드론 영상의 경우 Fig. 3과 같이 서울시 광진구의 건국대학교 공과대학을 중심으로 촬영하여 축척 계절의 변화에 따른 정확도 차이를 살펴보았다. 이때 촬영영역이나 이차록지를 설정하고 사전답사하여 드론 촬영에 문제가 생기지 않도록 하였고 Fig. 3의 좌측 아래 20.01.30 오후 두 시 50m의 고도에서 'Phantom 3 Pro' 드론을 사용하여 촬영한 영상 자료와 Fig. 3의 우측 위 20.05.28 오후 두 시 53m의 고도에서 'Phantom 4 Pro V2' 기종의 드론을 사용하여 촬영한 데이터를 사용하였다.





Fig. 3. Entire experiment area – drone experiment

### 3.2 항공영상의 확대율에 대한 실험

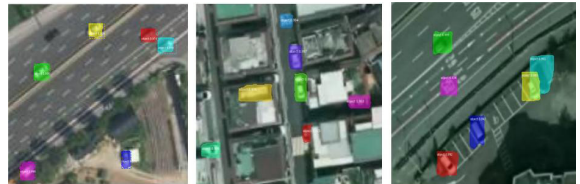
서울시 항공사진서비스 화면에서 화면 확대 레벨을 9, 10, 11로 변화를 주어 캡처하여 데이터를 수집한다. 확대율의 경우, 육안으로 사람이 객체를 인식 가능한 범위에서 설정하였으며, 항공사진서비스 화면상에서 육안으로 인식이 어려운 확대율인 9부터 상대적으로 인식이 쉬운 11까지로 선정하였다. 각 확대 레벨마다 200개의 캡처 영상을 확보하고 학습은 Train, Val, Test 별로 각각 128장, 32장, 40장으로 배치한다. 자동차를 라벨링하고 자동차와 모양이 확연히 다른 트럭이나 버스들은 라벨링에서 제외한다. 라벨링이 완료된 데이터들은 Train, Val, Test 각각 그림에 해당하는 자료들을 묶어 json 파일로 저장하며 각각의 파일들을 해당 폴더에 배치하였다.

딥러닝 모델의 경우, Mask R-CNN용 COCO 데이터 세트 v1.0을 통해 전이학습을 진행하였다. 이때 FPN (Feature Pyramid Network)의 head를 확장하여 Mask R-CNN의 head를 구성하였으며, 특징 추출의 경우는 ResNet-101 모델을 사용하여 진행하였다. COCO란 Common Objects in Context의 줄임말로, 약 330,000개의 이미지를 포함하고 있는 데이터 세트다. 약 330,000개의 이미지 중, 약 200,000개의 이미지는 라벨링이 되어있으며, 150만 개의 개체 인스턴스가 존재한다. 사전에 학습된 모델을 사용하여 실제로 학습, 실험에 사용할 수 있는 데이터양의 한계를 보완하였다.

학습은 30 epoch를 진행하고 학습이 완료되면 logs 폴더가 생성되며 학습을 통해 얻은 가중치들은 이 logs 폴더에 저장된다. 학습 후, test 폴더에 각각의 예측 이미지가 생성된 것을 확인할 수 있다. 또한, 딥러닝 학습 정도에 따른 편향된 결과를 배제하기 위해 각각의 epoch마다 예측, 평가 과정을 거쳐 정확도의 평균을 계산하였다.

Table 1. mAP(%) - Analysis of accuracy between magnification rates

Magnification Level	mAP average (%)	Maximum mAP (%)	Minimum mAP (%)	Deviation
9	60	89	23	66
10	67	92	37	55
11	75	93	61	31



(a) Magnification Level 9 (b) Magnification Level 10 (c) Magnification Level 11

Fig. 4. Predicted image

Table 1 및 Fig. 3은 각각의 확대 레벨 9, 10, 11에 대한 mAP(%)의 정량적인 평가 결과이다. Table 1에 따르면, 확대 레벨이 9, 10, 11로 높아짐에 따라 maximum mAP는 89%, 92%, 93%로 증가하고 minimum mAP 또한 23%, 37%, 61%로 증가하는 양상을 띤다. 이는 deviation이 66, 55, 31로 명확하게 차이를 보이는 것을 확인할 수 있다. 평가 척도인 mAP average는 각 확대 레벨에 따라 60%, 67%, 75%로 증가하며 오탐지의 가능성이 줄어들을 의미한다. 이는 객체탐지 시, 데이터 세트에 높은 확대 레벨의 입력자료를 활용하는 것이 객체탐지 mAP 향상에 효과적임을 나타낸다. 특히, 확대 레벨 10에서 11로의 변화는 9에서 10으로의 변화와 비교하면 mAP의 편차가 약 24% 감소하고 mAP(%)의 최솟값이 24% 증가하며 확연한 차이를 보여준다.

Fig. 3은 학습을 통해 얻은 가중치 모델을 테스트 데이터 세트 이미지에 적용하여 생성된 예측 이미지 자료이다. 데이터 세트의 확대 레벨이 9로 가장 낮은 Fig. 3(a)의 경우 Figs. 3(b) and 3(c)보다 장애물들을 자동차로 오인하여 탐지하거나 버스나 트럭 등의 객체를 탐지하는 오류를 빈번히 확인할 수 있다. 배경과 구분이 어려운 검은색의 차량은 mAP가 현저히 낮았다. 세 가지의 확대 레벨 모두 대체로 도로에 있는 자동차는 탐지가 원활하게 되었지만, 주택가가 몰려있거나 주위 장애물들이 밀집된 장소에서는 탐지의 오류율이 높았다. 가장 확대 레벨이 높았던 Fig. 3(c)은 Figs. 3(a) and 3(b)보다 버스나 장애물을 탐지하는 오류율이 적었고 배경과 구분이 어려운 검은

색의 차량도 탐지 가능성을 확인했다. 이는 확대 레벨이 높은 데이터 세트의 영상자료들이 좀 더 미세한 라벨링이 가능하고 이로 인해 더욱 명확히 탐지할 수 있으므로 제한된 조건에서 정확도 성능을 높일 수 있었다고 판단된다.

### 3.3 항공영상의 학습 영상 서로 다른 배율 학습 적용 결과

트레이닝에 사용되는 데이터 세트의 배율에 따른 정확도 차이를 알아보기 위하여 (Train, Val, Test) 확대 레벨을 6가지로 조합하였다. 실험 3.2와 동일하게 육안으로 사람이 자동차 객체를 인식 가능한 9, 10, 11배율의 데이터를 캡처하여 수집하였다. 수집한 각기 다른 배율의 데이터를 train과 test 데이터 세트에 위치시킴으로써 확대율, 학습, mAP 사이의 관련성을 파악해보고자 하였다. 저배율로 학습시키고 고배율로 테스트를 진행한 (9, 10, 10), (10, 11, 11), (9, 11, 11) 데이터 세트와 고배율로 학습시키고 저배율로 테스트한 (11, 10, 10), (10, 9, 9), (11, 9, 9) 데이터 세트로 나누어 실험을 진행하였다. 트레이닝에 사용하는 영상과 결과를 테스트해보는 영상의 확대 레벨을 서로 다른 확대 레벨로 엮갈려 배치하였다.

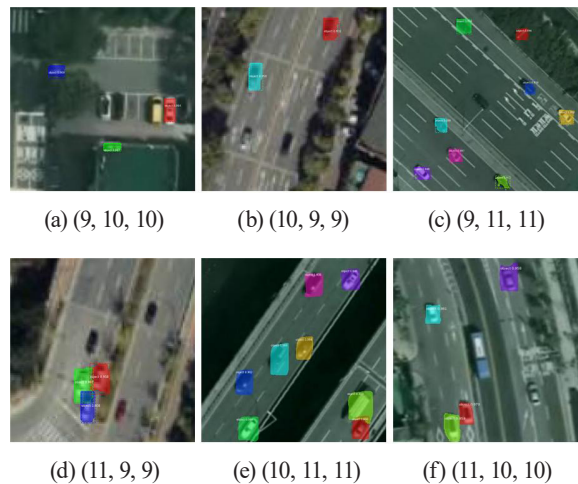
대체로 연구자들은 딥러닝 학습에서 데이터 세트를 세 가지 혹은 두 가지로 나누는데, 두 가지로 나눌 때는 train과 test로, 세 가지로 나눌 때는 train, test, validation으로 나눈다. train 데이터는 딥러닝 모델을 학습시키기 위한 데이터이며, test 데이터는 최종적인 정확도를 평가하기 위한 데이터이다. validation 데이터는 학습 도중에 모델을 평가하기 위한 데이터로, test 데이터와 validation 데이터가 유사해야 실험 수행자가 학습이 옹게 진행되는지를 확인할 수 있다. 따라서 본 실험에서는 test 데이터와 validation 데이터를 같은 확대 레벨을 적용하였다.

Table 2는 테스트한 결과에 대한 정량적인 표이다. Tables 2(b) and 2(c)에서 저배율로 학습시키고 고배율로 테스트한 것이 반대의 경우보다 mAP(%)가 Table 2(b)는 31%, Table 2(c)는 22% 차이를 보이며 확연히 높은 것을 확인할 수 있다. Table 2(c)에서 확대레벨 10으로 학습시킨 데이터 세트의 mAP(%) 예측값이 18 epoch에서는 100%를 기록하며 모든 자동차를 탐지하는 결과가 나오기도 했다. 그리고 mAP(%)의 평균이 86%로 높은 수치를 보였다. 반면, Table 2(a)의 경우 (9, 10, 10)이 mAP(%) 평균 27%, (10, 9, 9)는 51%로 Tables 2(b) and 2(c)와는 반대 결과가 확인된다. (9, 10, 10) 데이터 세트는 mAP(%)가 27%로 낮은 수치를 기록하였고 예측 이미지 Fig. 4(a)와 같이 전반적으로 탐지가 원활히 이뤄지지 않았다. 3.2 확대율 실험결과에서 가장 성능이 좋은 고배율의 11 확대 레벨 데이터가 포함되어있는 Tables 2(b) and

2(c)와 비교하면 미세하게 객체를 라벨링 하기 어려울 뿐 아니라, 작은 크기의 객체를 확대해서 라벨링하고 학습을 진행하다 보니 화질이 좋지 않아 정확성이 떨어진 것으로 예상된다. 특히, 확대 레벨 9의 경우, 저배율로 인하여 Figs. 4(b) and 4(d)에서 알 수 있듯이 객체가 눈으로도 구분이 어려운 이미지도 존재했기에 mAP(%)가 27%로 현저히 낮게 평가되었음을 예상한다.

**Table 2. Comparison of mAP(%) by magnification rate**

(a) Dataset Magnification level 9, 10				
Magnification Level	mAP average (%)	Maximum mAP (%)	Minimum mAP (%)	Deviation
(9, 10, 10)	27	62	10	52
(10, 9, 9)	51	69	25	44
(b) Dataset Magnification level 9, 11				
Magnification Level	mAP average (%)	Maximum mAP (%)	Minimum mAP (%)	Deviation
(9, 11, 11)	80	93	58	35
(11, 9, 9)	49	63	22	41
(c) Dataset Magnification level 10, 11				
Magnification Level	mAP average (%)	Maximum mAP (%)	Minimum mAP (%)	Deviation
(10, 11, 11)	86	100	64	36
(11, 10, 10)	64	77	42	35



**Fig. 5. Predicted image**

3.2와 3.3 실험 모두 Fig. 5와 같은 오류를 확인할 수 있었다. 예측 이미지를 보면 대체로 자동차가 다른 물체에 반절 이상 가려져 보이지 않을 때는 탐지가 불가능했고 흰색 자동차와 비교하면 검은색 자동차의 mAP가 현저히 떨어졌다. 물체가 지표면 도로와 정확히 분리되어 보이는 흰색 차보다 검은색 차의 경우 도로와의 경계가 뚜렷하지 않고, 그림자와 물체의 차이가 크지 않아 제한된 데이터의 양으로는 한계를 보인다. 따라서 전반적인 mAP average가 90% 이상인 높은 측정치를 기록하지 못했다고 판단된다. 데이터 세트를 구축할 때, 방사 보정과 같은 전처리과정이 선행된다면 분류 정확도 향상에 이바지할 수 있을 것이다.

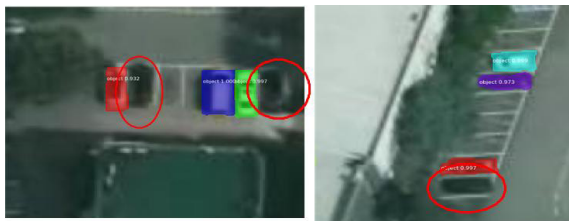


Fig. 6. Examples of detection error

### 3.4 드론 영상을 이용한 계절영향 분석 실험

객체탐지를 dataset 1과 dataset 2 두 개로 나누어 진행하여 비교해보고자 한다. Table 3과 같이 dataset 1의 train data는 2020.01.30에 촬영한 드론 영상을 사용하고 test, validation data는 2020.05.28. 드론 촬영 영상을 사용하여 4개월의 시간 간격을 두었다. dataset 2는 test, val, train 모두 2020.05.28에 촬영한 드론 영상을 사용하였다.

Table 3. Drone experimental dataset

	data set 1	data set 2
data	Test, Val : 20.05.28 drone images.	Test, Val, Train : 20.05.28 drone images.
	Train : 20.01.30 drone images.	
<b>Test : Train : Val (sheet)</b>	20 : 44 : 16	20 : 64 : 16

실험지역 주위의 자동차들을 Fig. 6과 같이 사각형(rectangle)으로 라벨링하고 가능한 확대하여 사각형 안에 자동차가 가득 차도록 모양을 만든다. 그리고 자동차와 모양이 확연히 차이가 나서 오류를 발생시킬 가능성이 있는 화물차나 트럭 등을 라벨링

대상에서 제외하도록 한다.



(a) 20.01.30 data (b) 20.05.28 data

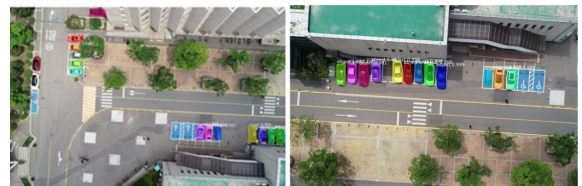
Fig. 7. Image labeling

확대율과 정확도 실험과 마찬가지로 기계 지도 학습을 통해 트레이닝 시킨다. 학습을 통해 가중치가 입력된 logs 파일을 이용하여 예측 이미지와 mAP(%)를 구하여 탐지결과의 신뢰성을 판별한다.

Table 4. Comparison of mAP(%)

	mAP average (%)	Maximum mAP (%)	Minimum mAP (%)	Deviation
data set 1	77	89	53	36
data set 2	93	97	86	10

Table 4를 보면, 학습을 진행하여 테스트자료에 대한 mAP(%) 값은 dataset 1과 dataset 2의 차이가 명확히 보인다. dataset 1의 경우 학습을 과거의(20.01.30) 드론 촬영 영상으로 진행하였기 때문에 모두 같은 20.05.28에 촬영한 영상을 사용한 dataset 2와 비교하면 mAP(%)가 낮음을 볼 수 있다. dataset 2의 예측 mAP 평균값(%)이 92%로 77%인 dataset 1과 비교하면 약 15% 정도 높은 것을 수치로 확인할 수 있다. dataset 2의 mAP(%)는 약 85~100%인 상위권에 계속해서 머물며 높은 수치를 보인다. 그리고 dataset 2의 편차는 10으로 굉장히 낮은 수치를 보이지만 dataset 1의 경우 편차가 36으로 차이가 있다.



(a) dataset 1 (b) dataset 2

Fig. 8. Detection result



Fig. 7(a)에서 dataset 1의 탐지된 결과 이미지를 보면, 실험 지역 주위의 자동차 탐지가 대체로 잘 되긴 하였지만, 간혹 탐지가 안 된 것을 확인할 수 있다. 반면에 Fig. 7(b)에서 볼 수 있듯이 dataset 2의 경우에는 거의 모든 자동차를 탐지하였다.

이를 통해 같은 해상도의 값을 가진 사진 자료를 함께 사용하면 학습능력을 높일 수 있음을 확인할 수 있다. 그리고 촬영 시기의 변화는 주위 환경의 변화를 의미한다. 시간이 지나 계절이 변화함에 따라 식생도 차이가 생긴다. 식생의 확연한 차이는 객체를 탐지하는데 정확도의 감소로 이어진다. 그러므로 탐지정확도를 높이기 위해서는 같은 시기, 환경에서 촬영된 영상 데이터를 사용하는 것이 바람직하다.

#### 4. 결론

본 연구는 '서울시 항공사진서비스'에서 제공하는 항공사진과 촬영한 드론 영상을 대상으로 제한된 양의 데이터에서 탐지정확도를 높이는 방법을 실험한 결과 다음과 같은 결론을 얻을 수 있었다. 항공사진을 활용하여 딥러닝 객체탐지 학습자료 구성에서 확대 레벨이 탐지정확도에 미치는 영향을 분석하였다. 사람이 객체를 인식 가능한 범위의 확대 레벨 9, 10, 11 데이터 이미지를 이용하였고, 확대 레벨이 높은 데이터 세트를 사용할수록 대상 객체 특성이 잘 반영되어 mAP 평균값이 60%, 67%, 75%로 증가하는 것으로 나타났다. 학습자료 구성에서 훈련자료와 테스트자료의 영상 확대 레벨을 다르게 했을 경우는 낮은 확대 레벨의 훈련 데이터와 높은 확대 레벨의 테스트 데이터의 구성이 반대의 경우에 비해 mAP 평균값이 20% 이상 높았다. 이는 수집과 레이블링 비용이 비싸 많은 데이터의 확보, 수집이 어려운 항공 영상의 한계를 공개 자료의 추가적인 활용을 통해 보완하고 학습자료 활용범위를 넓힐 수 있을 것으로 기대된다. 높은 확대 레벨의 데이터 혹은 고해상도의 훈련자료가 부족한 경우 학습데이터, 테스트 데이터의 축척(확대율) 조절과 적절한 조합을 통해 탐지정확도 향상에 이바지할 수 있을 것이다.

또한, 드론 영상으로 계절적인 차이가 탐지정확도에 미치는 영향을 분석하였다. 학습데이터와 테스트 데이터에 약 4개월의 차이를 둔 데이터 세트보다 같은 시기의 데이터 세트를 이용한 실험이 약 15% 높은 mAP 값을 보였다. 계절적인 차이가 식생의 차이로 이어지고 분광학적인 특성의 차이로 연결된다고 예상한다. 이는 계절적인 국토의 변화를 탐지하거나 탐지정확도 향상이 필요할 때 기본 영향요인 자료로써 도움을 줄 수 있다. 추후 연구에서 학습자료 구성 시 지역적 구분 외에 계절적 구분에 관한 연구가 방사 보정을 포함한 전처

리 기술 적용과 함께 진행된다면 탐지 오류율을 더욱 감소시킬 것으로 기대된다.

#### References

- Girshick, R., Donahue, J., Darrell, T., and Malik, J. (2014), Rich feature hierarchies for accurate object detection and semantic segmentation, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 24-27 June, Columbus, Ohio, pp. 580-587.
- Gu, H.I. (2018), Artificial intelligence and deep learning trends, *The Korean Institute of Electrical Engineers*, Vol. 67, No. 7, pp. 7-12. (in Korean)
- He, K., Gkioxari, G., Dollár, P., and Girshick, R. (2017), Mask R-CNN, *IEEE International Conference on Computer Vision (ICCV)*, 22-29 October, Venice, Italy, pp. 2961-2969.
- Jeong, J.Y., Cho, W.S., Chang, H.J., and Jeong, J.W. (2008), A Study on Object-based Change Detection Using Aerial LiDAR Data, *The Korean Society of Remote Sensing Spring Conference-2008*, 21 March, Korea, pp. 95-100.
- Kim, J., Song, Y.H., and Lee, W.K. (2021), Accuracy analysis of Multi-series Phenological Landcover Classification Using U-Net-based Deep Learning Model –Focusing on the Seoul, Republic of Korea–, *Korean Journal of Remote Sensing*, Vol. 37, No. 3, pp. 409-418. (in Korean with English abstract)
- Kim, Y.S., Kwak, G.H., Lee, K.D., Na, S.I., Park, C.W., and Park, N.W. (2018), Performance evaluation of machine learning and deep learning algorithms in crop classification: impact of hyper-parameters and training sample size, *Korean Journal of Remote Sensing*, Vol. 34, No. 5, pp. 811-827. (in Korean with English abstract)
- Lecun, Y., Boser, B., Denker, J.S., Henderson, D., Howard, R.E., Hubbard, and W., Jackel, L.D., (1989), Backpropagation applied to handwritten zip code recognition, *Neural Computation*, Vol. 1, No. 4, pp. 541-551.
- Lee, M.H., Nam, K.W., and Lee, C.W. (2019), Crack Detection on the Road in Aerial Image using Mask R-CNN, *Journal of the Korea Industrial Information Systems Research*, Vol. 24, No. 3, pp. 23-29. (in Korean with English abstract)

- Lee, S.G., Park, Y.S., Lee, G.S., Lee, J.Y., and Lee, S.H. (2013), An automatic object extraction method using color features of object and background in image, *The Journal of Digital Policy & Management*, Vol. 11, No. 12, pp. 459-465. (in Korean with English abstract)
- Park, J.C., Son, S.B., Lee, S.H., Jung, J.U., Park, Y.J., and Oh, H.S. (2021), Deep Ensemble based Object Detection from Aerial Images, *Journal of Institute of Control, Robotics and Systems*, Vol. 27, No. 12, pp. 944-952. (in Korean with English abstract)
- Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016), You only look once: unified, real-time object detection, *29th IEEE Conference on Computer Vision and Pattern Recognition*, 27-30 June, Las Vegas, Nevada, USA, pp. 779-788.
- Ren, S., He, K., Girshick, R., and Sun, J. (2015), Faster R-CNN: towards real-time object detection with region proposal networks, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 39, No. 6, pp. 91-99.
- Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A.C., and Fei-Fei, L. (2015), Imagenet large scale visual recognition challenge, *International Journal of Computer Vision*, Vol. 115, No. 3, pp. 211-252.
- Simonyan, K. and Zisserman, A. (2015), Very deep convolutional networks for large-scale image recognition, *Computer Vision and Pattern Recognition*, Cornell University, Ithaca, NY, USA.
- Song, J.Y., Won, T.Y., Jo, S.M., Eo, Y.D., Park, S.Y., Shin, S.H., Park, J.S., and Kim, C.J. (2021), Comparative experiment of cloud classification and detection of aerial image by deep learning, *Journal of the Korean Society of Surveying, Geodesy, Photogrammetry and Cartography*, Vol. 39, No. 6, pp. 351-360. (in Korean with English abstract)
- Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A. (2015), Going deeper with convolutions, *2015 IEEE Conference on Computer Vision and Pattern Recognition*, 7-12 June, Boston, MA, USA, pp. 1-9.
- Youn, H.J., Lee, M.H., Jeong, Y.S., Lee, C.W., Lee, H.S., and Jo, J.W. (2019), A Study on Car Detection in Road Surface Using Mask R-CNN in Aerial Image, *Korea Institute of Information and Communication Engineering Spring Conference-2019*, 23 May, Korea, pp. 71-73.