

# 터널 내 딥러닝 객체인식 오탐지 데이터의 반복 재학습을 통한 자가 추론 성능 향상 방법에 관한 연구

이규범<sup>1</sup> · 신휴성<sup>2\*</sup>

<sup>1</sup>학생회원, 한국건설기술연구원 미래스마트건설연구본부 학생연구원 / 과학기술연합대학원대학교 (UST) 스마트도시건설융합 석박통합과정

<sup>2</sup>정회원, 한국건설기술연구원 미래스마트건설연구본부 선임연구위원

## A study on improving self-inference performance through iterative retraining of false positives of deep-learning object detection in tunnels

Kyu Beom Lee<sup>1</sup> · Hyu-Soung Shin<sup>2\*</sup>

<sup>1</sup>Student Researcher, Dept. of Future & Smart Construction Research, Korea Institute of Civil Engineering and Building Technology / Ph.D. Student, Smart City and Construction Convergence, University of Science & Technology (UST)

<sup>2</sup>Senior Research Fellow, Dept. of Future & Smart Construction Research, Korea Institute of Civil Engineering and Building Technology

\*Corresponding Author : Hyu-Soung Shin, [hyushin@kict.re.kr](mailto:hyushin@kict.re.kr)

### OPEN ACCESS

Journal of Korean Tunnelling and Underground Space Association  
26(2)129-152(2024)  
<https://doi.org/10.9711/KTAJ.2024.26.2.129>

eISSN: 2287-4747  
pISSN: 2233-8292

Received February 7, 2024  
Revised March 4, 2024  
Accepted March 11, 2024



This is an Open Access article distributed under the terms of the Creative Commons Attribution

Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright © 2024, Korean Tunnelling and Underground Space Association

### Abstract

In the application of deep learning object detection via CCTV in tunnels, a large number of false positive detections occur due to the poor environmental conditions of tunnels, such as low illumination and severe perspective effect. This problem directly impacts the reliability of the tunnel CCTV-based accident detection system reliant on object detection performance. Hence, it is necessary to reduce the number of false positive detections while also enhancing the number of true positive detections. Based on a deep learning object detection model, this paper proposes a false positive data training method that not only reduces false positives but also improves true positive detection performance through retraining of false positive data. This paper's false positive data training method is based on the following steps: initial training of a training dataset - inference of a validation dataset - correction of false positive data and dataset composition - addition to the training dataset and retraining. In this paper, experiments were conducted to verify the performance of this method. First, the optimal hyperparameters of the deep learning object detection model to be applied in this experiment were determined through previous experiments. Then, in this experi-

ment, training image format was determined, and experiments were conducted sequentially to check the long-term performance improvement through retraining of repeated false detection datasets. As a result, in the first experiment, it was found that the inclusion of the background in the inferred image was more advantageous for object detection performance than the removal of the background excluding the object. In the second experiment, it was found that retraining by accumulating false positives from each level of retraining was more advantageous than retraining independently for each level of retraining in terms of continuous improvement of object detection performance. After retraining the false positive data with the method determined in the two experiments, the car object class showed excellent inference performance with an AP value of 0.95 or higher after the first retraining, and by the fifth retraining, the inference performance was improved by about 1.06 times compared to the initial inference. And the person object class continued to improve its inference performance as retraining progressed, and by the 18th retraining, it showed that it could self-improve its inference performance by more than 2.3 times compared to the initial inference.

**Keywords:** Tunnel CCTV-based accident detection, Precision, Recall, Deep learning-based object detection model, False positive data training method for object detection

## 초 록

터널 내 CCTV를 통한 딥러닝 객체인식 적용에 있어서 터널의 열악한 환경조건, 즉 낮은 조도 및 심한 원근현상으로 인해 오탐지가 대량 발생한다. 이 문제는 객체인식 성능에 기반한 영상유고시스템의 신뢰성 문제로 직결되므로 정탐지 향상과 더불어 오탐지의 저감 방안이 더욱 필요한 상황이다. 이에 본 논문은 딥러닝 객체인식 모델을 기반으로, 오탐지 데이터의 재학습을 통해 오탐지의 저감뿐만 아니라 정탐지 성능 향상도 함께 추구하는 오탐지 학습법을 제안한다. 본 논문의 오탐지 학습법은 객체인식 단계를 기반으로 진행되며, 학습용 데이터셋 초기학습-검증용 데이터셋 추론-오탐지 데이터 정정 및 데이터셋 구성-학습용 데이터셋에 추가 후 재학습으로 이어진다. 본 논문은 이에 대한 성능을 검증하기 위해 실험을 진행하였으며, 우선 선행 실험을 통해 본 실험에 적용할 딥러닝 객체인식 모델의 최적 하이퍼파라미터를 결정하였다. 그리고 본 실험에서는 학습영상 포맷을 결정하기 위한 실험, 반복적인 오탐지 데이터셋의 재학습을 통해 장기적인 성능향상을 확인하기 위한 실험을 순차적으로 진행하였다. 그 결과, 첫 번째 본 실험에서는 추론된 영상 내에서 객체를 제외한 배경을 제거시키는 경우보다 배경을 포함시키는 경우가 객체인식 성능에 유리한 것으로 나타났으며, 두 번째 본 실험에서는 재학습 차수별 독립적으로 오탐지 데이터를 재학습시키는 경우보다 차수마다 발생하는 오탐지 데이터를 누적시켜 재학습 시키는 경우가 지속적인 객체인식 성능 향상 측면에서 유리한 것으로 나타났다. 두 실험을 통해 결정된 방법으로 오탐지 데이터 재학습을 진행한 결과, 차량 객체 클래스는 1차 재학습 이후부터 AP값이 0.95 이상 우수한 추론 성능이 발현되었으며, 5차 재학습까지 초기 추론 대비 약 1.06배 추론성능이 향상되었다. 보행자 객체 클래스는 재학습이 진행됨에 따라 지속적으로 추론 성능이 향상되었으며, 18차 재학습까지 초기 추론대비 2.3배 이상 추론성능이 자가 향상될 수 있음을 보였다.

**주요어:** 터널 영상유고시스템, 정밀도, 인식률, 딥러닝 객체인식 모델, 객체인식 오탐지 학습법

## 1. 서론

터널 영상유고시스템은 국내에서 정차, 역주행, 보행자 및 화재로 정의되는 유고상황을 터널 CCTV로부터 인

지할 수 있는 시스템을 말하며, 터널 관리자에게 유고상황 알람 전달을 통해 터널 내 사고 발생시 신속한 대처를 할 수 있도록 한다(MOLIT, 2021). 그러나, 터널 CCTV로부터 촬영되는 영상은 낮은 조도 및 먼지가 쌓이기 쉬운 환경으로 인해 영상 노이즈가 발생되며, 터널은 공간적인 제약이 존재하므로 낮은 CCTV 설치위치로 인해 심한 원근현상과 차량간 겹침현상이 유발된다(Frías-Velázquez et al., 2011). 이러한 악조건으로 인해, 개방된 외부 도로상에서 운용되는 일반 영상유고시스템을 그대로 터널에 적용할 경우, 상당한 수의 오류경보와 매우 낮은 유고상황 감지성능을 보인다.

이 문제를 해결하기 위해, Lee는 영상의 픽셀구조를 기하학적으로 변환시킬 수 있는 역 원근변환법으로 통해 변환된 영상을 활용함으로써 원근현상 문제를 해결하였으며, 촬영지점으로부터 200 m 떨어진 거리까지도 일관된 인식성능을 달성하였다(Lee and Shin, 2022). 이 영상은 본 논문에서 변환영상이라고 정의되며, 변환영상을 바탕으로 한 객체인식 알고리즘으로 터널 영상유고시스템의 개발 및 유고상황의 인지가 가능함을 보였다.

여기서, 해당 터널 영상유고시스템이 유고상황을 정확하게 인지하는 경우 정탐지로 명명하며, 그렇지 않은 경우는 오탐지로 명명한다. 그러나, 상기 서술한 열악한 영상환경으로 인해 정탐지 뿐만 아니라 비슷한 모양의 오탐지가 대량 발생한다(MOLIT, 2015). 이러한 정탐지 및 오탐지의 예시는 Fig. 1에서 확인할 수 있다.

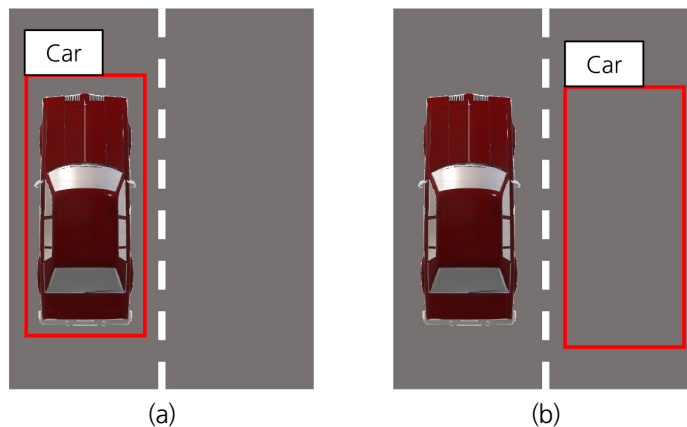


Fig. 1. True and false example for tunnel CCTV (a) true (b) false

Fig. 1은 터널 CCTV 영상에서 볼 수 있는 정탐지 및 오탐지 예시를 나타낸다. Fig. 1과 같이 영상에서 목표 인식대상을 의미하는 Ground truth (GT) 차량 1대가 보이는 경우에 대하여, Fig. 1(a)와 같이 차량 1대를 직사각형 형태의 경계박스 및 차량 객체 클래스를 정확하게 인식한 경우를 정탐지로 정의할 수 있다. 그리고 Intersection over union (IOU)이라는 지표를 통해 GT 경계박스과 추론된 경계박스 사이의 겹침값을 비교하며, 일정한 문턱값(Threshold)을 통해 참/거짓을 판별한다. 오탐지는 (1) GT 경계박스과 비교했을 때, IOU 문턱값 미만인 경우 또는 (2) IOU 문턱값은 넘었지만 차량 객체 클래스를 정확하게 인식하지 못한 경우, (3) 1개의 GT 경계박스에 대하여 정탐지를 2개 이상 중복 인식하는 경우, 가장 높은 IOU값을 가진 정탐지 객체를 제외한 나머지 객체들에 대

한 경우 오탐지로 정의되며, Fig. 1(b)는 (1)에 해당한다고 볼 수 있다.

이러한 설명을 바탕으로, 상기 서술했던 문제를 해결하기 위해 터널 영상유고시스템에서 발생한 오탐지 데이터를 전부 배경 객체 클래스로 정정하여 학습시키는 오탐지 학습법이 제시되었다(Lee and Shin, 2019). 이를 통해 하루에 0~40개정도 발생된 화재 유고상황 오탐지가 0~3개로 발생빈도가 감소하였으며, 결과적으로 터널 관리자 입장에서 본 시스템의 신뢰도가 향상되었다. 그러나, 해당 연구는 일괄적으로 배경 객체로 변경 및 재학습을 진행하였으므로 다른 유고상황에 대한 감지성능의 하락 우려가 존재한다.

본 논문은 발생한 오탐지 데이터에 대하여 상기 서술했던 오탐지 학습법에서 조건에 따라 오탐지, 정탐지 또는 제거를 선택하여 정정된 오탐지 데이터셋을 학습시키는 개선된 방법을 제시하며, 이를 통해 오탐지 발생의 저감 뿐만 아니라 정탐지 성능의 향상도 추구하는 것을 목표로 한다. 이에 따라 단일 터널 현장 CCTV 영상을 기반으로 한 데이터셋을 제작하고, 객체인식 단계에서 실험을 수행하였다. 그리고 정량적으로 평가할 수 있는 객체인식 성능 지표를 바탕으로 실험을 진행하였다.

## 2. 이론

### 2.1 딥러닝 및 객체인식

딥러닝은 주어진 데이터를 학습하고 표현할 수 있는 알고리즘인 기계학습에서 진화된 분야이며, 기계학습보다 깊고 복잡하게 구성된 알고리즘이라고 할 수 있다(LeCun et al., 2015). 이 때, 데이터셋은 주어진 데이터의 모음으로 정의되는데, 입력값 및 출력값으로 구성되거나 입력값만 주어진 경우도 존재한다. 예를 들면, 이미지 분류는 데이터셋의 입력값을 정지영상, 출력값을 개, 사람, 고양이 등과 같은 하나의 분류값을 결과값으로 한다(Lu and Weng, 2007). 이 때, 딥러닝 모델은 수만 개에서 수십만 개 이상의 빅데이터로 이루어진 데이터셋의 학습을 통해 다양한 상황을 인식하는 것이 가능하다.

객체인식은 컴퓨터 비전 분야의 목표 중 하나로, 정지영상에서 목표 대상 물체를 의미하는 객체를 직사각형의 경계박스로 인식하는 목표를 말한다(Zou et al., 2023). 이 경우엔 입력값이 정지영상이 되며, 출력값은 하나 이상의 경계박스들 및 각 경계박스에 대응되는 객체 클래스들이다. 객체인식은 객체의 위치를 특정할 수 있으므로 의료 영상에서 질병 식별, 자율주행, 안면인식, CCTV 감시 등 다양한 분야에서 활용되고 있다.

객체인식 분야의 경우 Histograms of oriented gradients (HOG) (Dalal and Triggs, 2005), Deformable part models (DPM) (Felzenszwalb et al., 2008)과 같은 로직 및 기계학습 기반 알고리즘이 존재하였으나, 낮은 인식 성능 및 느린 추론속도로 인해 활용하기 어렵다. 그러나 Region-based convolutional neural network (R-CNN) (Girshick et al., 2014)의 등장으로, Faster R-CNN (Ren et al., 2015), Single shot multibox detector (SSD) (Liu et al., 2016), You only look once (YOLO) (Redmon et al., 2016) 등과 같은 Convolutional neural network (CNN) 기반 딥러닝 객체인식 모델이 등장하면서 상기 서술한 문제들을 획기적으로 해결할 수 있었다.

## 2.2 Recall, precision 및 AP

여기서 객체인식 알고리즘 인식결과는 4가지 분류로 나눌 수 있다. 먼저 목표 대상 객체를 인식한 경우를 정탐지(True positive, TP), 인식하지 못한 경우를 미탐지(False negative, FN)로 정의한다(Fawcett, 2006). 그리고 목표 대상 객체가 아닌데 정탐지로 인식한 경우를 오탐지(False positive, FP), 배경 객체로 인식한 경우를 배경탐지(True negative, TN)로 정의한다(Fawcett, 2006). 이 때, 객체인식 알고리즘의 성능 평가는 정밀도, 인식률 및 평균 정밀도와 같은 정탐지 성능 지표를 활용하여 진행된다.

정밀도(Precision)는 알고리즘이 정탐지라고 인식한 결과(정탐지+오탐지)들 중에서 실제로 정탐지 객체인 비율을 말하며, 식 (1)과 같이 나타낸다(Lewis, 1990). 즉, 정밀도가 높을수록 객체인식 알고리즘은 오탐지를 인식한 개수가 적어진다.

$$Precision = \frac{TP}{TP + FP} = \frac{TP}{Total\ predicted\ detection\ as\ true} \quad (1)$$

한편, 인식률(Recall)은 인식 대상 정탐지 객체들(정탐지+미탐지) 중에서 알고리즘이 정탐지 객체를 인식한 비율을 말하며, 식 (2)와 같이 나타낼 수 있다(Lewis, 1990). 다시 말하면 인식률이 높아질 수록 인식 대상 정탐지 객체들을 모두 탐지할 수 있다.

$$Recall = \frac{TP}{TP + FN} = \frac{TP}{Total\ positive} \quad (2)$$

그리고 객체인식 알고리즘의 성능 평가는 이 알고리즘이 해당 객체가 참일 가능성을 표현하는 신뢰점수(Confidence score)가 높은 객체부터 낮은 순서로 정렬한 다음, 순차적으로 참/거짓 판별 및 판별결과를 누적시키면서

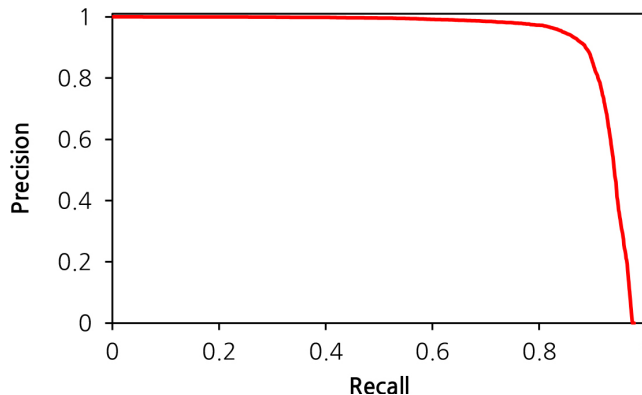


Fig. 2. Recall-precision curve example

인식률 및 정밀도를 계산한다(Sergey and Venu, 2013). 매 계산마다 인식률 및 정밀도를 그래프로 그리는 것이 가능하며, Fig. 2와 같이 인식률-정밀도 곡선(Recall-precision curve)이 된다(Fawcett, 2006). 이 곡선은 인식률 및 정밀도가 1에 가까워질수록 우수하다고 평가할 수 있으며, 서로 다른 알고리즘들이 동일한 객체인식 데이터셋을 추론하면서 산출되는 인식률-정밀도 곡선들을 비교하면서 어느 알고리즘이 더 우수한지 시각적으로 확인할 수 있다.

한편, 객체인식 분야에는 수많은 객체인식 알고리즘들이 존재하며, 수십개 이상의 알고리즘들에 대하여 인식률-정밀도 곡선으로 한 번에 비교하기 어렵다는 문제가 있다. 이 문제를 해결하기 위해, 평균 정밀도(Average precision, AP)라는 지표값이 도입되었다(Zhu, 2004). AP는 식 (3)과 같이 산출된 인식률-정밀도 곡선에서 정밀도의 평균 값으로 정의할 수 있으며, Fig. 2에서 그려진 인식률-정밀도 곡선의 면적으로도 정의할 수 있다. AP는 숫자값으로 알고리즘에 대한 객체인식 성능을 정량적으로 확인할 수 있다는 장점이 있으므로 딥러닝 객체인식 분야에서 널리 사용된다.

$$\text{Average precision} = \frac{1}{N} \sum_{i=1}^N P_i = \int_1^N P(R) dR \quad (3)$$

한편, 객체인식 알고리즘은 신뢰점수가 1에 가까울수록 정밀도가 높아지는 경향이 있지만 인식률이 낮아지며, 반대로 0에 가까울수록 정밀도가 낮아지지만 인식률이 높아진다. 즉, 정밀도와 인식률은 서로 반비례하는 관계에 있다고 볼 수 있으며, Table 1과 같이 신뢰점수 문턱값이 0.1에서 0.6으로 증가할수록 정밀도가 높아지는 동시에 인식률이 낮아지는 경향을 확인할 수 있다. 다시 말하면 객체인식 알고리즘은 현장에 적용될 때, 인식률-정밀도 지점(Recall-precision point)을 비교하면서 현장 상황에 맞는 정밀도 및 인식률을 가질 수 있는 적절한 신뢰점수 문턱값을 선정해야 한다.

**Table 1.** Relationship between confidence score threshold and recall-precision point

Confidence score threshold	Recall-precision point	
	Recall	Precision
0.1	0.908	0.811
0.2	0.885	0.906
0.3	0.858	0.941
0.4	0.831	0.96
0.5	0.786	0.974
0.6	0.715	0.984

### 2.3 오탐지 학습법

본 논문에서 제안하는 오탐지 학습법은 딥러닝 객체인식 모델을 통해 추론된 데이터를 이용한다. 이 때, 경계박스 좌표를 조작하지 않고 객체 클래스의 수정 또는 경계박스 제거를 통해 오탐지 데이터를 정정한 후, 학습용 데이터셋에 추가하여 재학습시키는 방법을 말한다. 현존하는 데이터셋의 레이블링은 정지영상에서 경계박스 좌표를 직접 입력해야 하므로 많은 시간이 소요되지만, 제안된 오탐지 학습법에서는 이 과정을 대신하여 객체 클래스의 변경 또는 제거만 이루어지므로 데이터셋 제작에 상당한 시간을 절약할 수 있다.

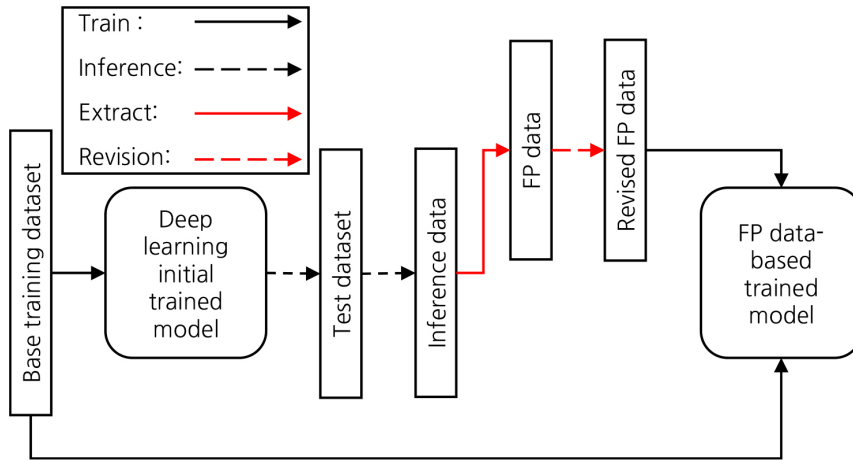


Fig. 3. Basic process about false positive data training method

Fig. 3은 본 논문에서 제안하는 오탐지 학습법의 재학습 과정을 순서도로 표현하였으며, 각 데이터의 흐름을 학습, 추론, 추출 및 정정으로 구분 및 기호로 표시하였다. 먼저 본 프로세스는 정탐지 데이터만 포함된 기본 학습용 데이터셋(Base training dataset)을 학습함으로써 딥러닝 초기학습 모델(Deep learning initial trained model)을 얻을 수 있다. 그 다음, 초기학습 모델은 검증용 데이터셋(Test dataset)을 추론하여 추론 데이터(Inference data)를 얻는데, 정지영상 기준으로 오탐지 객체가 하나 이상 포함된 오탐지 데이터(FP data)만 추출한다. 이 때, 오탐지 데이터에서 오탐지 객체에 대하여 (1) IOU 문턱값을 기준으로 완전히 객체를 인식하지 않은 경우 제거 (2) 완전히 배경을 인식한 경우, 가상객체 즉, Not-객체 클래스로 변경 (3) 다른 정탐지 객체 클래스인 경우, 해당 클래스로 변경이라는 선택지들을 통해 정정한다. 이 과정을 통해 정정된 오탐지 데이터들을 바탕으로 정정된 오탐지 데이터셋(Revised FP dataset)을 구성한다. 마지막으로 기본 학습용 데이터셋과 정정된 오탐지 데이터셋을 학습시킴으로써 오탐지 데이터셋 기반 재학습 모델(FP data-based trained model)을 얻는다.

### 3. 선행 실험

본 논문은 딥러닝 객체인식 모델을 사용하므로 본 실험을 하기에 앞서 선행실험을 통해 본 실험에 영향을 끼치는 하이퍼파라미터를 결정한다.

#### 3.1 선행 실험 개요

본 논문에서 사용한 딥러닝 객체인식 모델은 YOLO V7이다(Wang et al., 2023). 이 모델은 추론과정 중에서 추가적인 계산비용을 발생시키지 않으면서 성능을 향상시키는 기법을 적용하여 YOLO V4 (Bochkovskiy et al., 2020) 대비 모델 파라미터의 수를 43% 줄일 수 있었고, 0.5 IOU 기준 AP값이 1.5% 향상되었다.

이 모델은 기존 Faster R-CNN, YOLO V4, SSD와 같은 객체인식 모델과 비교했을 때 동등 또는 높은 객체인식 성능을 보임과 동시에 추론속도가 161 FPS로 Faster R-CNN 및 SSD의 추론속도가 각각 7~15 FPS, 20~50 FPS인 것에 비해 3~20배 이상 빠르다(Wang et al., 2023). 이러한 속도 측면의 장점은 반복적인 재학습을 진행할 본 논문의 실험에서 시간적 측면에 장점이 있으므로 YOLO V7을 실험 대상 객체인식 모델로 사용하였다.

#### 3.2 실험 대상 데이터셋 조건 및 학습 환경

본 논문에서 사용된 실험 대상 데이터셋 현황은 Table 2와 같다. 먼저, 데이터셋의 제작은 단일 터널 현장에서 촬영된 CCTV 영상으로부터 발췌하였으며, 차량(Car) 및 보행자(Person)을 객체인식 대상 클래스로 한다. 학습용 데이터셋과 검증용 데이터셋은 서로 시간적으로 완전히 독립된 데이터셋들이다. Table 2에서 학습용 데이터셋은 정상상황 영상과 유고상황 영상이 포함되어 있으므로 보행자 대비 차량의 비중이 약 6.2배 크며, 약 20,300장의 정지영상을 학습대상으로 한다. 한편, 검증용 데이터셋은 순수하게 유고상황 영상을 기반으로 구성되어 있으므로 학습용 데이터셋보다 보행자의 비중이 더 크며, 약 5,300장의 정지영상을 검증 대상으로 한다.

**Table 2.** Total deep learning dataset status

Dataset type	Number of images	Number of objects		
		Car	Person	Total
Training	20,264	39,015	6,302	45,317
Test	5,379	13,761	4,173	17,934

딥러닝 객체인식 모델 학습에 사용된 하드웨어는 AMD RYZEN 5955WX, 128GB RAM, NVIDIA RTX 4090 GPU 4개를 사용하였으며, Ubuntu 20.04, Python 3.10, Pytorch 2.0.0 버전의 소프트웨어를 활용하였다. 이 때, 선행 실험과 본 실험의 학습환경은 동일하다. 그리고 정지영상에 대한 Batch size는 96, Learning rate는 0.01, 정지영상 크기는 640 x 640으로 설정하여 딥러닝 객체인식 모델의 학습을 진행하였다.



### 3.3 선행학습

본 실험에 앞서 딥러닝 하이퍼파라미터의 결정을 위해 선행학습을 진행하였다. 데이터셋 전체를 학습한 횟수는 Epoch로 정의되는데(Brownlee, 2022), 적절한 학습 성능 및 비용을 Epoch마다 손실함수(Loss function)의 비교를 통해 결정할 수 있다.

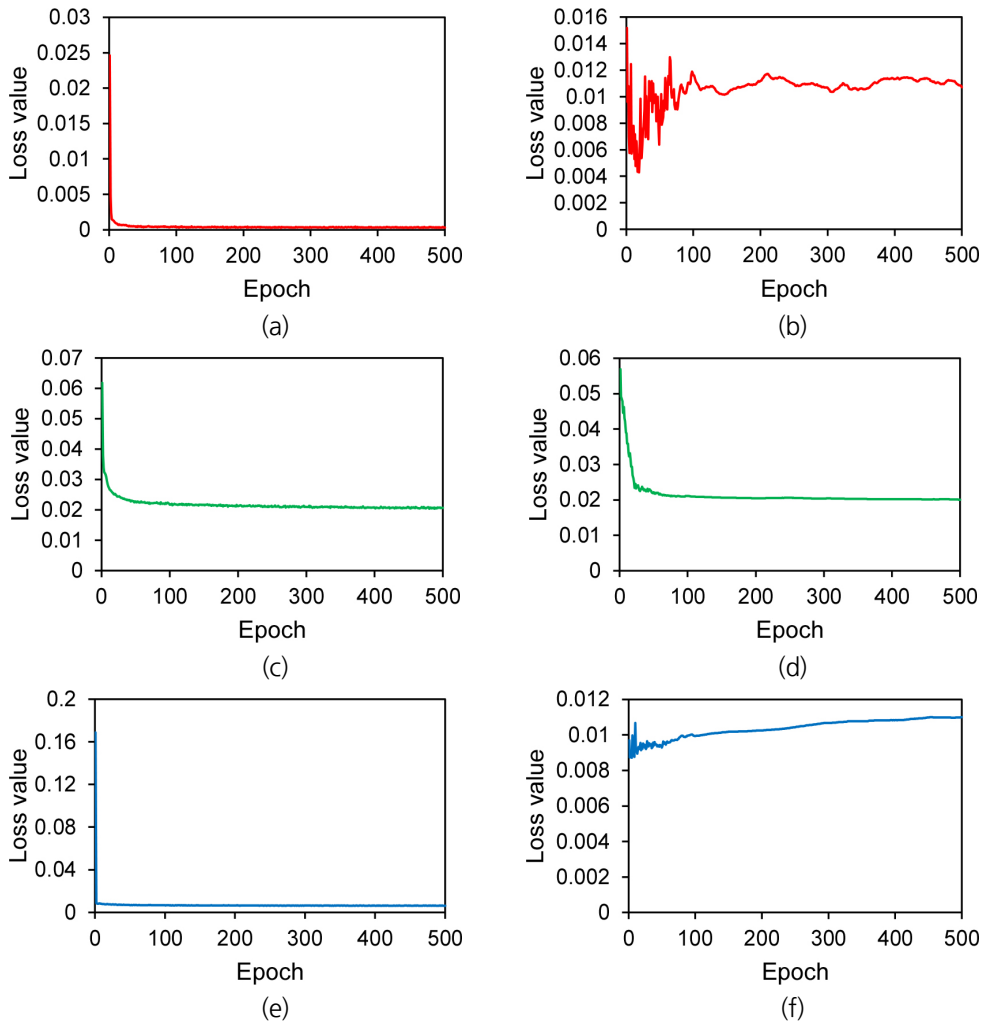
선행학습을 위한 데이터셋의 현황은 Table 3과 같다. 선행학습은 학습용 데이터셋 전체를 학습시키는 본학습 이전에 하이퍼파라미터의 결정을 위한 학습이므로 Table 2의 학습용 데이터셋에서 정지영상의 경우, 임의로 8:2 비율로 나눈 다음 Table 3과 같이 선행학습용 데이터셋(Pre-training) 및 선행학습 후 확인용 데이터셋(Validation)이 된다.

**Table 3.** Deep learning dataset status for pre-training

Dataset type	Number of images	Number of objects		
		Car	Person	Total
Pre-training	16,211	31,326	5,110	36,436
Validation	4,053	7,689	1,192	8,881

상기 서술된 데이터셋을 바탕으로 500 epoch의 선행학습을 진행하였다. 이 때, 딥러닝 객체인식 모델의 학습은 약 15시간이 소요되었다. 그 다음, Fig. 4와 같이 학습용 데이터셋, 확인용 데이터셋에 대한 손실함수 추이를 각각 확인하였다. YOLO V7 모델의 손실함수는 Classification loss, Regression loss 및 Objectness loss 3개로 구성되며, 해당 모델의 갱신 과정에서 3개의 손실함수값들을 한 번에 반영하는 Multi task loss의 형태이다. Fig. 4의 (a), (c), (e)는 학습용 데이터셋, Fig. 4의 (b), (d), (f)는 확인용 데이터셋에 대한 손실함수에 대한 추이이다. 그리고 Fig. 4의 (a)와 (b)는 Classification loss, Fig. 4의 (c)와 (d)는 Regression loss, Fig. 4의 (e)와 (f)는 Objectness loss에 대한 추이이다.

먼저 학습용 데이터셋에 대한 손실함수 추이를 확인해 보면, Classification loss와 Regression loss는 50 epoch 부근에서 수렴되었으며, Objectness loss는 약 10 epoch에서 수렴된다. 그리고 확인용 데이터셋의 경우, Classification loss는 하강-상승 과정을 거치다가 50 epoch에서 잠시 안정된 다음, 다시 하강-상승 과정을 거치면서 100 epoch에서 안정된다. Regression loss는 50 epoch까지 낮아지다가 50 epoch 이후에서 수렴하는 경향이 있었으며, Objectness loss는 50 epoch 부근에서 안정되었다가 이후에서는 점점 손실값(Loss value)이 상승하는 경향이 있었다. 모든 손실함수의 추이를 고려했을 때, 50 epoch 부근 이후에서 수렴하는 경향이 있으므로 본학습의 Epoch는 50 epoch로 결정하였다. 이 때, 본 학습의 Epoch는 오탐지 데이터셋을 추가한 다음 재학습을 진행할 때에도 동일한 값을 사용한다.



**Fig. 4.** Loss function for pre-training (a) classification loss for training dataset (b) classification loss for validation dataset (c) regression loss for training dataset (d) regression loss for validation dataset (e) objectness loss for training dataset (f) objectness loss for validation dataset

### 3.4 선행 추론

선행 추론 실험은 선행 학습 진행 후, 딥러닝 모델 추론에 영향을 끼치는 Non-maximum suppression (NMS) IOU 문턱값 및 NMS 신뢰점수 문턱값을 결정하기 위한 실험이다. 이 때, NMS는 서로 중복된 경계박스들에 대하여 가장 높은 신뢰점수값을 제외한 나머지 경계박스들을 제거하는 후처리 프로세스를 말한다(Neubeck and Van Gool, 2006). 이 때, NMS IOU 문턱값은 중복된 경계박스로 인식하는 IOU 문턱값을 말한다.

그리고 NMS confidence score 문턱값은 추론된 경계박스들 중에서 문턱값 이하의 신뢰점수를 가진 경계박스들을 일괄적으로 제거하는 역할을 한다. 객체인식 알고리즘은 정지영상 한 장당 수만 개 정도의 경계박스를 추론

하는데, 거의 대부분의 인식된 경계박스들은 배경 탐지 또는 오탐지에 해당되며 신뢰점수가 0에 가깝다. 그러나 객체인식 분야는 정탐지 인식 및 오탐지 저감에 초점을 맞추고 있으므로 적절한 NMS 신뢰점수 문턱값의 결정을 통해 배경 탐지 및 오탐지 경계박스들을 최대한 제거한다.

YOLO V7은 학습 및 추론 전주기 프로세스가 진행된 이후에 NMS가 적용되므로 epoch 결정 이후에 NMS IOU 문턱값 및 NMS score 문턱값을 결정하게 된다. 즉, 두 파라미터들은 학습에 영향을 미치는 파라미터들이 아니므로 본 학습을 통해 사용되며 Table 2에 있는 학습용 데이터셋과 검증용 데이터셋을 활용한다.

선행 추론 실험은 본 학습 진행 후, 먼저 Table 4와 같이 NMS IOU 문턱값을 변화시키면서 검증용 데이터셋을 추론하여 평균 AP값(Mean AP value, mAP)을 비교하였다. 딥러닝 객체인식 모델의 본학습은 50 epoch로 진행되었으며, 약 2시간이 소요되었다. NMS IOU 문턱값은 0.45부터 0.85까지 값을 0.1 증가시키면서 AP값을 확인하였는데 0.55에서 가장 높은 mAP를 보였다. 이 때, 보행자에 대한 AP값은 NMS IOU 문턱값이 0.45일 때보다 0.001 낮지만 거의 비슷한 수준으로 볼 수 있으며, 차량에 대한 AP값이 더 높은 것까지 고려하면 NMS IOU 문턱값이 0.55일 때 객체인식 성능이 더 높다고 볼 수 있다. 이후 NMS IOU 문턱값이 증가할수록 AP값이 감소하는 경향이 있었으므로 NMS IOU 문턱값은 가장 높은 AP값을 보이는 0.55가 적절하다고 할 수 있다.

**Table 4.** AP value for NMS IOU threshold decision

NMS IOU	Average precision (AP)		Mean average precision (mAP)
	Car	Person	
0.45	0.925	0.445	0.685
0.55	0.928	0.444	0.686
0.65	0.927	0.437	0.682
0.75	0.922	0.418	0.67
0.85	0.907	0.377	0.642

그리고 Fig. 5는 NMS 신뢰점수 문턱값을 결정하기 위해 해당 값을 0.1에서 0.6까지 값을 0.1씩 증가시키면서 인식률-정밀도 지점 변화추이를 비교한 것이며, 차량과 보행자 객체들에 대하여 각각 추이를 확인하였다. NMS IOU 문턱값은 상기 결정된 바와 같이 0.55로 설정하여 추론하였다. 이 때, 본 논문의 목적은 검증용 데이터셋에서 추론된 오탐지 데이터를 바탕으로 재학습시켜 성능향상을 도모하는 것이기 때문에 높은 정밀도보다 상대적으로 낮은 정밀도에서 발생하는 오탐지 데이터를 얻는 것이 실험 목적에 부합된다.

이를 고려하면 Fig. 5의 인식률-정밀도 점 추이는 NMS 신뢰점수 문턱값이 감소할수록 인식률은 높아지면서 정밀도가 감소하는 추이를 나타낸다. 인식률-정밀도 추이가 0.2에서 0.1로 감소할 때, 인식률의 증가치는 비교적 일정한데 정밀도의 감소폭이 커진다. 이 때, 두 객체 클래스에 대한 정밀도의 감소폭이 0.6부터 0.2까지 일정한 것을 고려하면, NMS 신뢰점수 문턱값은 0.2로 설정하는 것이 타당할 것으로 판단된다.

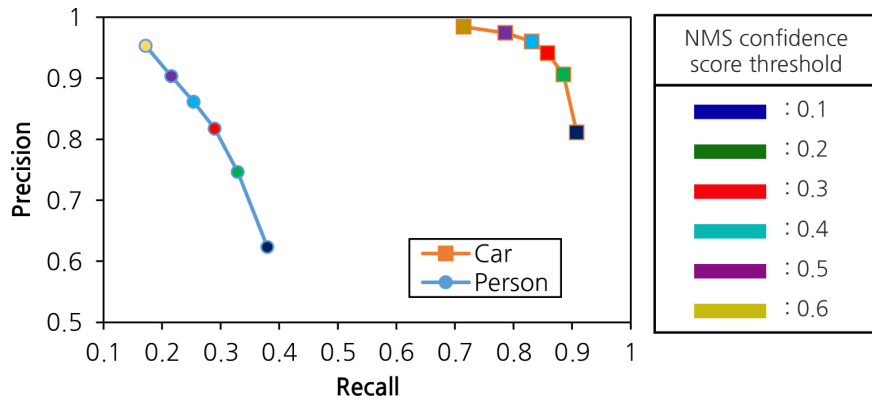


Fig. 5. Recall-precision point for NMS confidence score threshold decision

## 4. 본 실험

### 4.1 본 실험 개요

본 실험은 선행 실험을 통해 얻은 학습 및 추론 하이퍼파라미터를 바탕으로 진행되었다. 본 실험은 2개의 실험으로 구성되며, 첫 번째 본 실험은 오탐지 학습 영상 포맷을 결정하며, 미탐지 영향을 확인하기 위한 실험이다. 학습 영상 포맷은 객체를 제외한 배경에 대한 포함 여부이며, 오탐지 데이터 재학습 이후 객체인식 성능의 비교를 통해 결정한다. 두 번째 본 실험은 앞서 결정된 오탐지 학습영상 포맷을 바탕으로 진행되며, 오탐지 학습법의 지속적인 객체인식 성능 향상효과를 검토한다. 이 때, 반복적인 오탐지 데이터 재학습 프로세스에서는 재추론이 진행될 때 마다 발생하는 오탐지 데이터를 과거의 데이터를 배제하고 신규로 구성하여 학습할지, 또는 반복 재학습마다 과거의 오탐지 데이터를 계속 누적시켜서 재학습을 수행할 지에 따른 객체인식 성능 향상효과를 비교할 것이다.

### 4.2 1차 본 실험: 오탐지 학습영상 포맷 결정

앞서 서술한 대로 첫 번째 본 실험은 오탐지 데이터를 정정할 때, 객체를 제외한 배경 포함여부에 따른 재학습 성능을 확인하기 위한 실험이며, 이를 위해 Fig. 6과 같이 두 경우로 나누어서 진행하였다.

Fig. 6(a)는 배경 영상을 포함한 경우이며 case 1이 된다. 그리고 Fig. 6(b)는 배경 영상을 회색으로 처리하여 제거한 경우이며 case 2가 된다. Fig. 6에서 보여지는 예시는 레이블링 대상 객체가 모두 포함되어 있지만 Fig. 5에서 확인할 수 있듯이 NMS 신뢰점수 문턱값이 0.2일 때 보행자의 인식률은 약 0.33, 차량은 0.89 정도로 검증용 데이터셋을 완벽하게 인식한 상태가 아니다. 그렇기 때문에 추론된 데이터는 미탐지가 다수 포함된 상태이며, 미탐지 객체에 대한 처리여부에 따른 객체인식 성능을 확인하는 것이 본 절의 실험 목표이다.

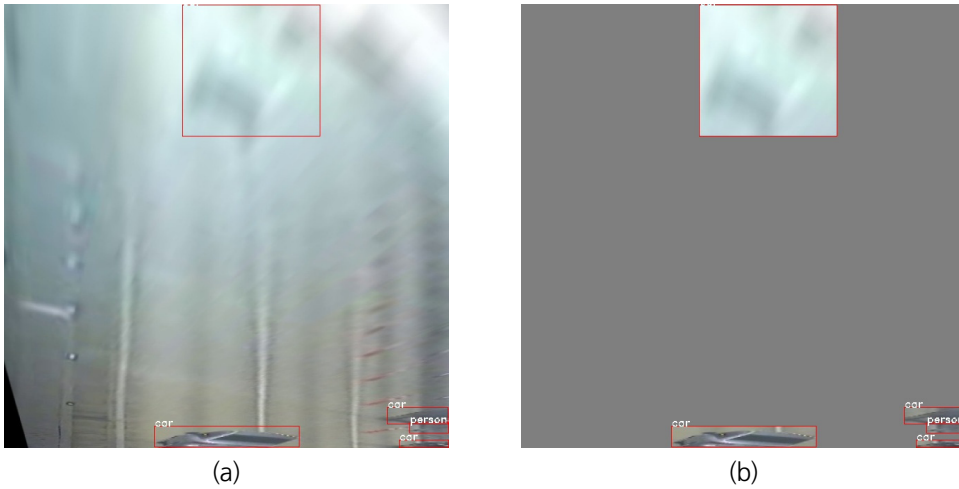


Fig. 6. Example images for each main experiment 1 case (a) case 1: include background image (b) case 2: exclude background image

Table 5는 초기학습 모델을 바탕으로 확인용 데이터셋으로 추론된 데이터에서 오탐지 데이터가 포함된 정지 영상들만 수집한 다음 정정된 오탐지 데이터를 포함시킨 재학습 데이터셋 현황이다. Table 5에서 정정된 오탐지 데이터셋은 정지영상 기준으로 기본 학습용 데이터셋 대비 약 6.7%의 비중을 차지한다. 그리고 차량 객체수는 약 8.5%, 보행자 객체수는 약 8.3%의 비중을 차지하면서 기본 학습용 데이터셋에 포함되어 재학습이 진행된다. 한편, 배경을 대상으로 하는 가상객체들, 즉 Not-car 및 Not-person 객체는 총 정탐지 객체수 대비 1% 미만의 낮은 비중으로 학습되지만, 배경 오탐지는 CCTV 영상에서 비슷한 배경을 반복적으로 발생시키므로 이를 학습시키면 배경 오탐지의 발생을 낮출 수 있을 것으로 기대된다.

Table 5. Deep learning retraining dataset status

Sub dataset type	Number of images	Car	Person	Not-car	Not-person	Total
Base training dataset	20,264	39,015	6,302	0	0	45,317
Revised FP dataset	1,362	3,347	521	213	36	4,117
Total	21,626	42,362	6,823	213	36	49,434

두 경우에 대한 재학습 실험결과는 Table 6에서 확인할 수 있다. Table 6은 배경 영상 포함여부에 따른 두 경우를 각각 재학습시킨 후, 검증용 데이터셋을 추론하여 각 클래스마다 AP값, 인식률 및 정밀도를 측정된 결과이며, 초기학습 모델의 추론 결과도 포함시켰다. Table 6을 보면, case 1의 지표들은 case 2보다 모든 지표에서 더 높은 값을 기록하고 있다. 이 때, case 2는 AP값 및 차량 객체 클래스에 대한 정밀도를 보았을 때 초기 학습모델보다 향상되었지만 case 1보다 떨어지며, 차량 객체 클래스에 대한 인식률은 오히려 떨어진다. 다시 말하면 비록 미탐지

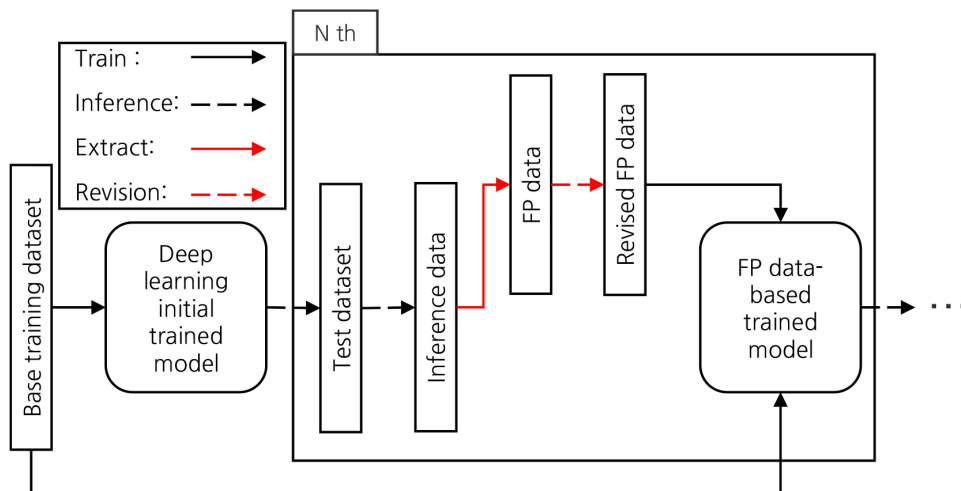
된 객체가 배경으로써 학습됨에도 불구하고, 영상 배경을 전부 제거한 경우보다 더 높은 객체인식 성능을 가지게 됨을 의미한다. 따라서 첫 번째 본 실험의 결과를 바탕으로, 두 번째 본 실험에서는 배경 영상을 포함한 학습영상 포맷으로 재학습을 진행하였다.

**Table 6.** AP and recall-precision point results for main experiment 1

Experiment case	Average precision (AP)		Car		Person	
	Car	Person	Recall	Precision	Recall	Precision
Initial trained model	0.928	0.444	0.885	0.906	0.329	0.746
Case 1	0.953	0.64	0.908	0.958	0.466	0.918
Case 2	0.935	0.533	0.879	0.94	0.433	0.746

### 4.3 N차 재학습 프로세스 누적여부 실험

터널 CCTV에서 보이는 영상환경은 단기적으로 보면 비교적 일정한 환경을 유지하나, 수년 이상의 시간이 지났을 때 터널 내 조명의 교체 및 새로운 차량의 출시 등의 요인으로 인해 조금씩 영상환경이 변화하게 된다. 그렇기 때문에 오탐지 학습법을 통해 재학습된 모델은 현장에 대한 객체인식 성능이 단기적으로 개선되지만, 시간이 지날수록 그 성능이 점진적으로 하락할 가능성이 있다. 두 번째 본 실험은 이 문제에 대응하기 위해 현장에서 꾸준히 발생하는 오탐지 데이터를 바탕으로, 오탐지 학습법 전주기 프로세스를 반복적으로 진행했을 때, 객체인식 성능의 향상여부를 확인하는 실험이다. 이 때, 두 번째 본 실험은 세부적인 프로세스를 두 경우로 나누어서 실험을 진행하였는데, 첫 번째는 각 차수별 독립적으로 오탐지 데이터를 정정 및 재학습하는 경우, 두 번째는 오탐지



**Fig. 7.** The process about Nth false positive data training method for case 1 of main experiment 2: do not accumulate FP dataset

학습법을 진행하였을 때 각 차수별 오탐지 데이터를 누적시키면서 정정 및 재학습하는 경우이다. 각 경우의 프로세스는 Fig. 7 및 Fig. 8로 표현하였다.

Fig. 7은 두 번째 본 실험의 case 1에 대한 오탐지 학습법 프로세스를 나타낸 것이다. 이 프로세스는 Fig. 3에서 보였던 기본적인 프로세스와 동일하며, 각 차수별 오탐지 데이터는 독립적이므로 다른 차수 오탐지 데이터에 영향을 주지 않는다.

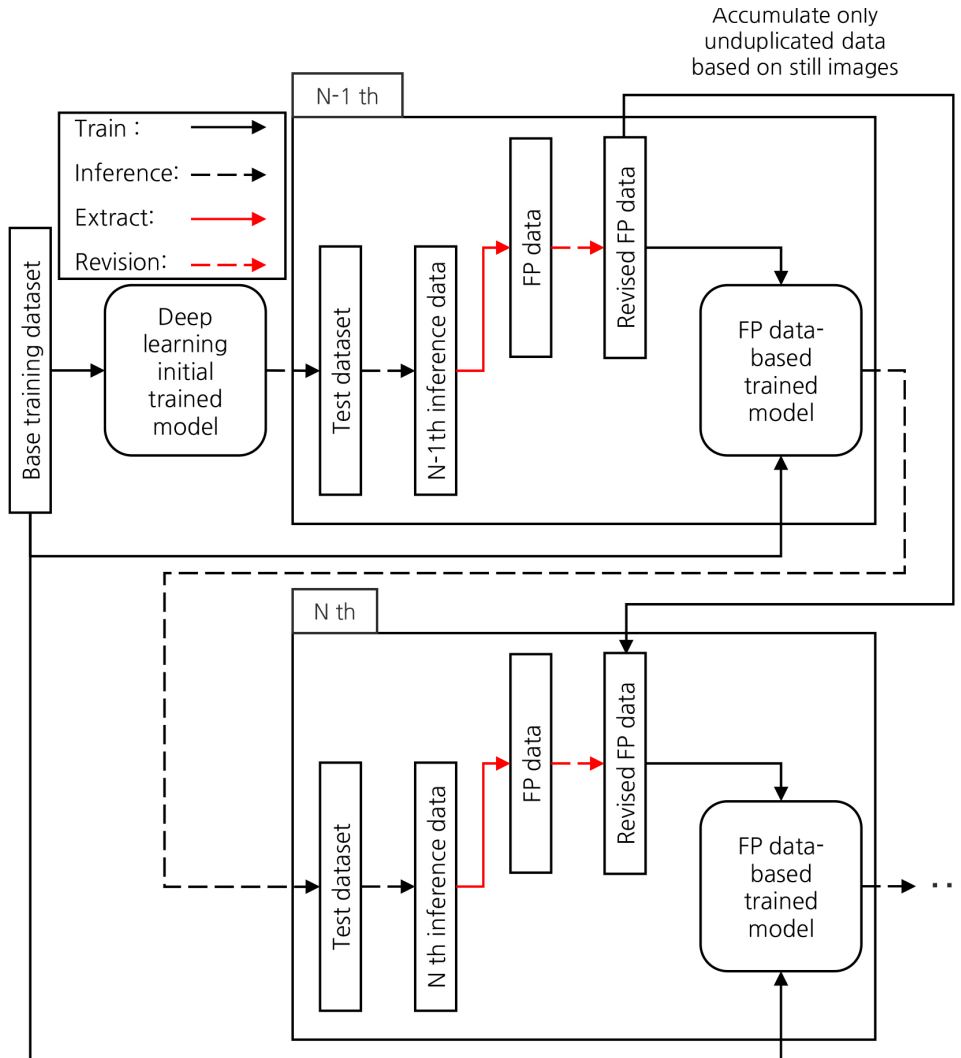


Fig. 8. The process about Nth false positive data training method for case 2 of main experiment 2: accumulate FP dataset

한편, Fig. 8과 같이 본 실험의 case 2 프로세스는 기본적으로 Fig. 3과 동일하지만, N차 재학습에서 오탐지 데

이터 정정 및 데이터셋 구성을 진행할 때, N-1차 재학습 데이터를 가져와 누적시킨다. 이 때, 정지영상 기준으로 N차와N-1차 데이터는 서로 겹치지 않는다. 이러한 경우들을 바탕으로, 두 번째 본 실험은 각각 5차까지 매 차수마다 오탐지 학습법 전주기 프로세스를 진행하였으며, 이에 대한 오탐지 데이터셋의 변화는 Table 7 및 Table 8에서 확인할 수 있다.

Table 7은 두 번째 본 실험의 case 1에 대하여 5차 재학습까지 진행했을 때, 2차 재학습 데이터셋부터 각 항목마다 변화된 데이터 현황을 나타낸다. 이 때, 1차 재학습 데이터셋은 해당 데이터셋의 현황을 나타내며, 2차 재학습 데이터셋 부터는 이전 차수 재학습 데이터와 비교했을 때 증감수치를 나타낸다.

Fig. 7에서 설명한 바와 같이, 각 차수별 오탐지 데이터는 독립적이므로 1차부터 4차 재학습 과정에서 데이터의 변화가 정지영상 기준 186개에서 706개 정도로 변화가 큰 것을 알 수 있다. 이에 따라, 오탐지 데이터셋에서 정지영상 개수의 변화는 각 객체 클래스마다 객체 수 변화에 영향을 미치며, 결과적으로 객체인식 성능의 변동에 영향을 미칠 것이다.

**Table 7.** Change status of retraining dataset for case 1 of main experiment 2 until 5th retraining

Sub dataset type	Number of images	Car	Person	Not-car	Not-person	Total
Base training dataset	20,264	39,015	6,302	0	0	45,317
1st FP dataset	1,362	3,347	521	213	36	4,117
2nd FP dataset	-706	-1,653	-275	-139	-18	-2,085
3rd FP dataset	406	925	264	199	21	1,409
4th FP dataset	-186	-413	-116	-154	-8	-691
5th FP dataset	-65	-170	68	-11	3	-110

**Table 8.** Change status of retraining dataset for case 2 of main experiment 2 until 5th retraining

Sub dataset type	Number of images	Car	Person	Not-car	Not-person	Total
Base training dataset	20,264	39,015	6,302	0	0	45,317
1st FP dataset	1,362	3,347	521	213	36	4,117
2nd FP dataset	285	709	146	21	9	885
3rd FP dataset	189	492	92	6	33	623
4th FP dataset	115	324	78	7	4	413
5th FP dataset	104	283	74	18	0	375

한편, 동일한 조건에서 case 2에 대한 오탐지 데이터셋의 변화는 Table 8과 같다. 이 경우, 오탐지 데이터는 Fig. 8의 과정과 같이 재학습 차수에 따라 정지영상 기준 104~285개의 범위에서 안정적으로 누적된다. 이러한 경향성은 차량 및 보행자 객체도 영향을 받으며, 오탐지 재학습 차수에 따라 객체인식 성능이 안정적으로 향상될 것이다.



상기 서술 및 정리된 오탐지 데이터셋을 바탕으로, Fig. 9는 2차 본 학습의 case 1에 대한 재학습 진행 후, 객체 인식 성능 변화추이 결과를 인식률-정밀도 지점으로 표현하였다. 각 객체 클래스별로 지점 변화 추이를 확인하였을 때, 차량 객체 클래스는 이미 인식률 및 정밀도가 0.9 이상 값을 가지는 상태이므로 재학습의 진행에 따른 인식률-정밀도 점의 변화추이가 거의 없는 상태이다. 한편, 보행자 객체 클래스는 초기학습 때 낮은 인식률(0.329) 및 정밀도(0.746)를 가진 상태에서 재학습의 진행에 따라 두 지표가 향상되는 것을 확인할 수 있다. 그리고 Table 7과 같이 재학습 차수에 따른 오탐지 정지영상 및 객체수의 큰 변동폭이 인식률-정밀도 지점의 변화추이에 영향을 미치며, 재학습 진행에 따라 정밀도는 상승 및 하강을 반복하는 경향을 보인다. 그리고 인식률은 초기~3차 재학습까지는 큰 폭으로 상승한 반면 3~5차 재학습까지 인식률의 상승폭이 둔화된 것을 Fig. 9에서 확인할 수 있다.

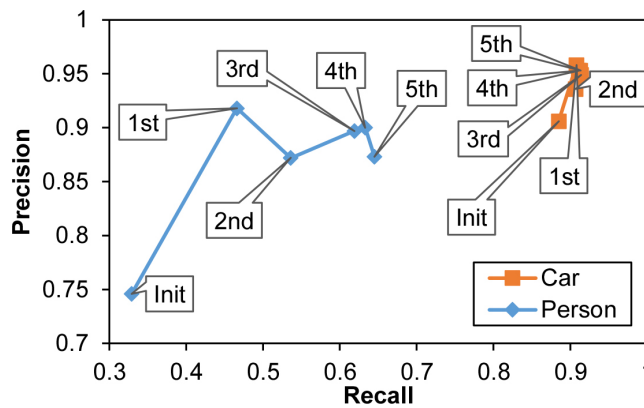


Fig. 9. Changes in recall-precision point as retraining progresses for case 1 of main experiment 2

Case 2의 경우, Fig. 10과 같이 case 1보다 더 안정적인 인식률-정밀도 지점 변화추이를 보인다. 차량 객체 클래스는 Fig. 9와 마찬가지로 인식률-정밀도 지점의 변화추이가 거의 없었으며, 보행자 객체 클래스는 case 1처럼

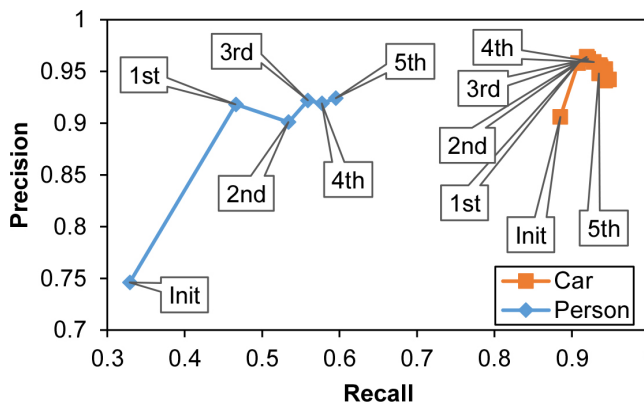


Fig. 10. Changes in recall-precision point as retraining progresses for case 2 of main experiment 2

인식률의 상승폭이 크지 않지만, 정밀도의 변화 폭이 case 1보다 작은 것을 확인할 수 있다. 비록 5차 재학습까지 진행했을 때, 보행자의 인식률 값은 case 2보다 case 1이 높다. 그러나 안정적으로 상승하는 인식률 및 일정한 정밀도를 보이는 case 2가 현장 적용성에 더 적합하다고 판단된다.

Fig. 11은 차량 객체에 대하여, 초기학습부터 5차 재학습 과정 진행에 따라 변화되는 객체인식 성능을 시각적으로 보여주는 예시이다. Fig. 11과 같이 파란색 경계박스는 GT 경계박스이며, 초록색 경계박스는 오탐지를 나타낸다. 그리고 적색 경계박스는 정탐지에 해당한다. 이를 바탕으로 초기학습 단계부터 추론영상을 확인해 보면, 초기학습 모델은 객체인식 모델이 영상 하단의 GT 객체 2개를 한 번에 인식하려고 하면서 오탐지 1개가 발생한 것을 볼 수 있다. 그런데, 1차 재학습 모델부터는 영상 하단의 GT 객체 2개를 정탐지로 인식하기 시작하였으며, 3차 재학습 모델에서 영상 중간의 GT 객체도 정탐지로 인식하는 모습을 보였다.

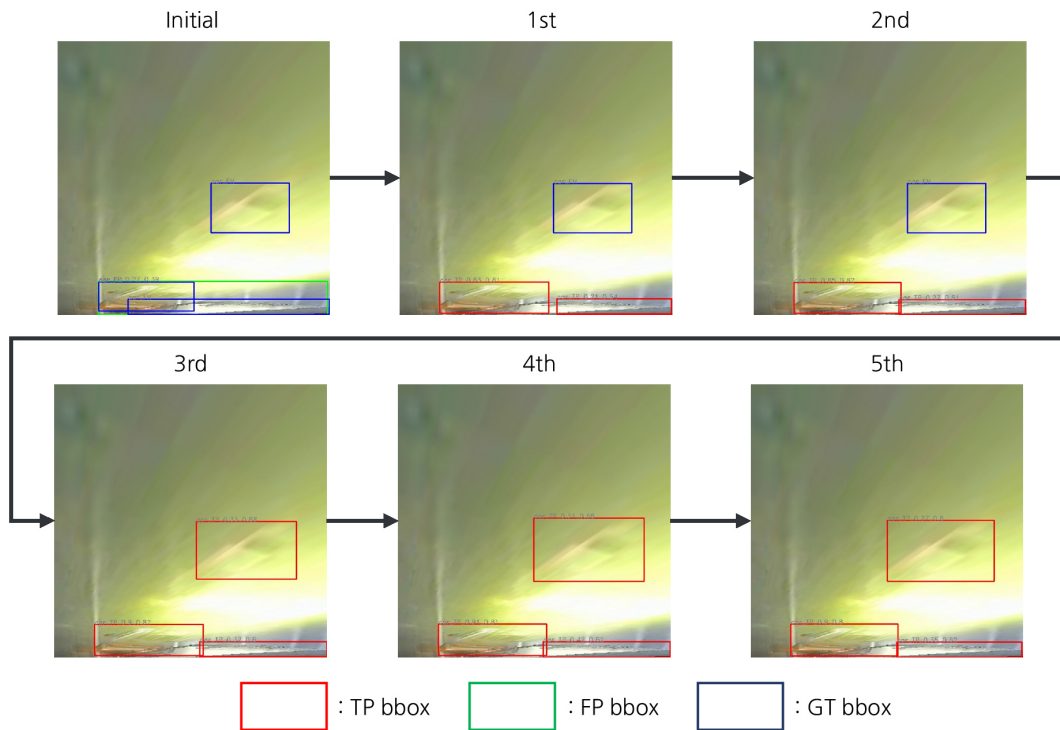


Fig. 11. Changes in Car class inference image example as retraining progresses

그리고 Fig. 12는 초기학습부터 5차 재학습 과정 진행에 따라 변화되는 객체인식 성능을 시각적으로 보여주는 예시이며 차량 객체를 제외하고 보행자에 대한 시각적인 변화에 초점을 맞추었다. 경계박스 종류에 대한 설명은 Fig. 11과 동일하다. 이 때, 초기학습 단계에서는 GT 객체 3개를 모두 인식하지 못하는 상태였지만 1차 재학습 단계부터 모두 정탐지로 인식하였다. 이후 5차 재학습까지 재학습 모델은 지속적으로 3개의 정탐지 객체인식을 유지하였다.

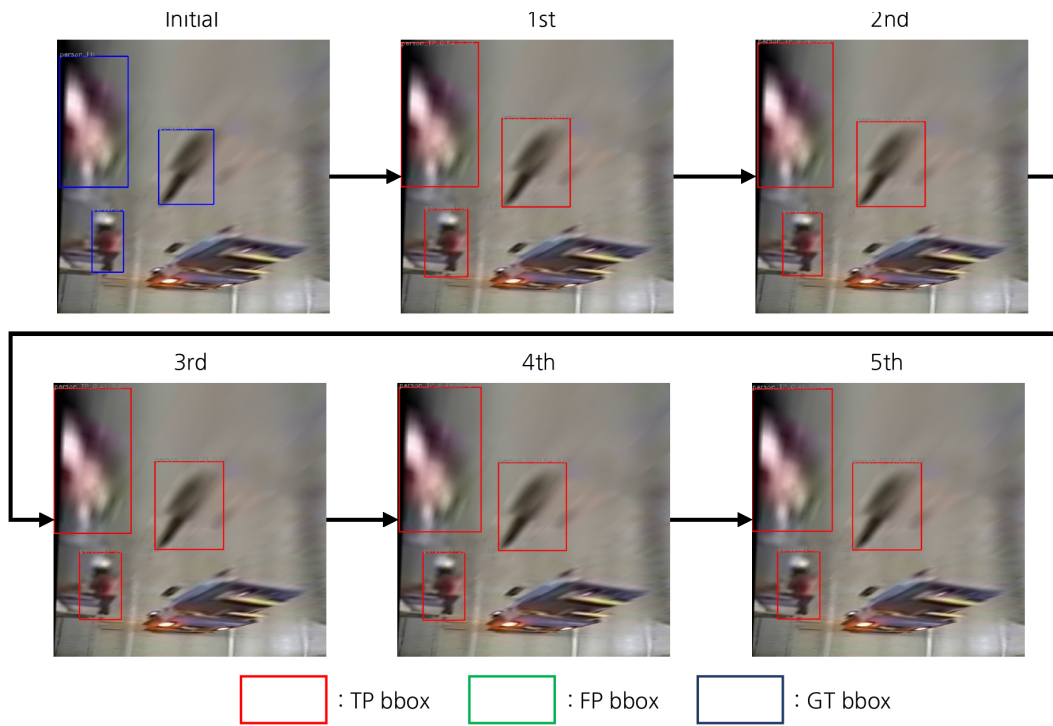


Fig. 12. Change in Person class inference image example as retraining progresses

두 번째 본 실험의 case 2 방법은 안정적인 정밀도 유지 및 지속적으로 인식률의 향상을 보였다. 이러한 결과가 추가 반복 오탐지 재학습을 통해 계속 성능이 향상될 수 있는지를 확인하기 위하여, case 2 방법으로 오탐지 학습법 전주기 프로세스를 18차까지 진행하였다. Table 9는 case 2 방법으로 18차 재학습까지 진행하였을 때의 오탐지 데이터셋의 변동 추이를 보여준다. Table 9에 나타난 바와 같이, 정지영상 개수의 변동폭은 2~7차 재학습 단계에서 정지영상 기준 104~285개를 기록하고 있으며, 8차 이후부터 28~80개의 증가폭을 보인다. 즉, 8차 재학습 단계부터는 객체인식 모델의 성능이 안정화되면서 새로운 오탐지의 발생이 줄어들게 되었다고 해석할 수 있다.

Table 9. Change status of retraining dataset for case 2 of the 2nd main experiment 2 until 18th retraining

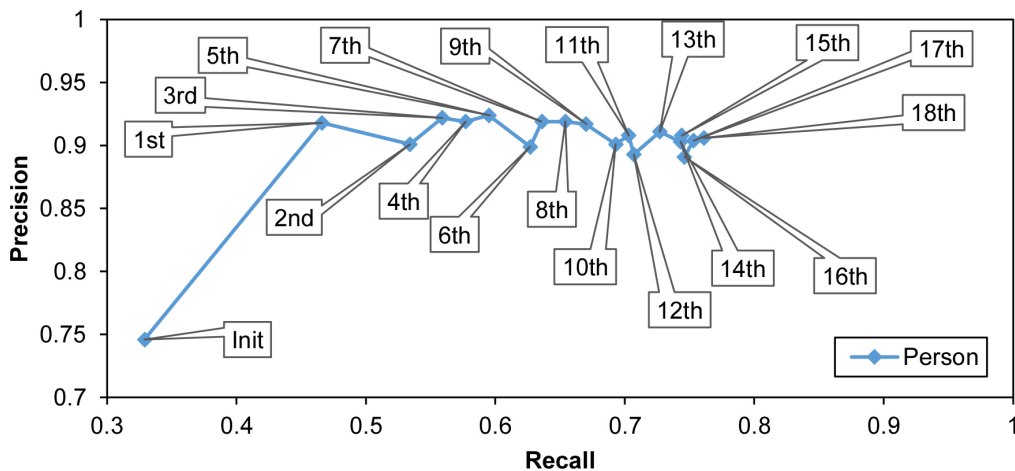
Sub dataset type	Number of images	Car	Person	Not-car	Not-person	Total
Base training dataset	20,264	39,015	6,302	0	0	45,317
1st FP dataset	1,362	3,347	521	213	36	4,117
2nd FP dataset	285	709	146	21	9	885
3rd FP dataset	189	492	92	6	33	623
4th FP dataset	115	324	78	7	4	413
5th FP dataset	104	283	74	18	0	375
6th FP dataset	115	291	66	44	9	410

**Table 9.** Change status of retraining dataset for case 2 of the 2nd main experiment 2 until 18th retraining (continued)

Sub dataset type	Number of images	Car	Person	Not-car	Not-person	Total
7th FP dataset	129	365	119	0	6	490
8th FP dataset	79	182	68	19	7	276
9th FP dataset	80	224	57	14	4	299
10th FP dataset	48	136	69	9	7	221
11th FP dataset	65	160	75	3	16	254
12th FP dataset	41	110	43	8	-8	153
13th FP dataset	72	146	52	0	23	221
14th FP dataset	45	136	58	5	-1	198
15th FP dataset	42	128	54	15	6	203
16th FP dataset	34	98	42	13	-4	149
17th FP dataset	42	131	33	2	15	181
18th FP dataset	28	76	48	0	4	128

한편, 가상객체인 Not-car와 Not-person은 재학습 진행에 따른 객체의 증가폭이 일정치 않다. 그러나, 가상객체인 Not-객체는 배경을 재학습한 다음 재학습 차수에서 추론시키면 해당 배경 오탐지와 비슷한 배경을 전부 Not-객체로 인식되면서 오탐지 발생을 억제시킨다. 이러한 영향으로 인해 새로운 배경 오탐지 객체 발생 - Not-객체 재학습 - 배경 오탐지 억제가 재학습 차수 진행에 따라 반복되는 것으로 볼 수 있다.

Fig. 13은 보행자 객체 클래스에 대하여, case 2 프로세스로 오탐지 데이터 재학습을 18차까지 진행하였을 때 인식률-정밀도 지점의 변화 추이를 그래프로 표현한 것이다. 앞에서 서술한 것처럼, case 2는 예상대로 18차 재학습



**Fig. 13.** Changes in recall-precision point for Person class as retraining progresses about case 2 of main experiment 2 until 18th retraining

까지도 인식률이 꾸준히 증가하는 모습을 보였으며, 정밀도의 변화는 큰 변동폭 없이 0.9 인근에서 안정적으로 유지되는 것을 확인할 수 있다. 결과적으로, case 2 방법은 초기학습 때 0.33의 낮은 인식률에서 0.77까지 2배 이상 인식률이 크게 향상되었으며, 그 과정에서 일정한 정밀도를 유지할 수 있었다.

#### 4.4 결과 분석

오탐지 학습법의 검증을 위한 두 개의 본 실험을 진행하였으며, 각 결과는 다음과 같다.

먼저 첫 번째 본 실험 결과는 오탐지 데이터를 정정할 때, 미탐지된 객체를 포함하더라도 배경을 포함시키는 경우가 그렇지 않은 경우보다 더 높은 정밀도를 보였다. 특히 차량보다 더 작은 크기를 가진 보행자 객체는 주변 배경에 민감하게 영향을 받는 것을 확인할 수 있다. 이러한 결과는 딥러닝 객체인식 모델이 배경이 제거된 오탐지 데이터를 학습할 때, 주변 배경과 연결성이 없는 객체를 학습하므로 객체 및 배경에 대한 판별능력과 정밀도가 떨어지기 때문이다.

두 번째 본 실험 결과의 경우, 재학습 진행에 따른 인식률-정밀도 지점의 변화 추이 일관성은 과거 차수의 오탐지 데이터를 누적시켜 학습하는 것이 각 차수마다 독립적으로 학습시키는 것보다 더 일관성 있게 추론 성능이 향상됨을 확인할 수 있었다. 오탐지 데이터는 독립적으로 학습시킬 경우, 오탐지 다량발생-재학습-오탐지 저감-재학습-오탐지 다량발생으로 사이클이 이어졌으며, 인식률은 오탐지 데이터를 누적시킨 경우보다 높았지만 정밀도 측면에서 일관적인 성능을 유지하지 못하였다.

반면에, 오탐지 데이터를 누적시키면서 학습하는 경우, 이전 차수의 오탐지 데이터는 현재 차수의 오탐지 데이터셋에 일부 남아있는 상황이 되므로 인식률-정밀도 지점의 변화 추이에서 정밀도의 큰 변동없이 꾸준히 인식률이 향상된다. 이러한 특징은 현장적용성 및 신뢰성 측면에서 오탐지 데이터를 독립적으로 학습시키는 경우보다 더 우수하다고 볼 수 있다.

마지막으로 두 번째 본 실험은 오탐지 데이터를 누적시키면서 학습하는 경우에 대하여, 장기적인 재학습에 따른 인식률-정밀도 지점의 변화 추이 일관성을 검증하기 위해 지점의 변화가 미미해진 18차까지 재학습을 진행하였다. 그 결과, 보행자 객체 클래스에 대한 객체인식 성능은 18차 재학습까지 일정한 정밀도를 유지하였고, 인식률은 계속적으로 향상됨을 보였다. 이러한 결과는 상기 서술했던 현장 적용성 및 신뢰성을 뒷받침한다.

### 5. 결론

본 논문은 터널 영상유고시스템에 적용되는 딥러닝 객체인식 모델에서 발생하는 오탐지 데이터의 확인 및 해당 데이터의 재학습을 통해 딥러닝 객체인식 모델의 인식률을 지속적으로 향상시킬 수 있는 학습법을 제안하였다. 이 방법은 오탐지 데이터를 대상으로 반복적인 추론-오탐지 확인 및 정정-재학습 과정을 거치면서 오탐지의 저감 및 정탐지에 대한 인식률의 향상이 가능함을 의미한다. 즉, 인식률의 증대를 위해 추가적인 영상확보와 레이블링 작업을 통해 GT 학습용 데이터셋을 확장시켜 학습시키는 정형적인 방법과 더불어, 학습모델의 추론단계에서

발생하는 오탐지 데이터를 재학습 함으로써 오탐지 저감과 정탐지율 향상 효과를 얻을 수 있다. 이때, 오탐지 결과의 재학습을 위한 오탐지 데이터셋 구성에 있어서, 비용과 시간이 많이 소요되는 오탐지 영상데이터의 경계박스 편집 작업은 하지 않으나 오탐지 여부의 판단과 이에 따른 오탐지 객체 클래스의 정정 정도의 작업은 요구된다.

오탐지 데이터를 정정하는 과정에서 객체를 제외한 배경을 재학습 영상에 포함시킬 것인지에 대한 결정을 위한 비교실험, 반복적인 재학습 과정에서 과거 차수의 추론 오탐지 데이터를 누적시킬 것인지를 결정하기 위한 비교실험을 진행하였으며, 이러한 실험적 연구를 통해 다음과 같은 결론을 얻을 수 있었다.

본 논문의 오탐지 학습법은 반복적인 재학습을 통해 객체에 대한 인식을 향상에 기여할 수 있으며, 경계박스 좌표의 조정없이 객체의 클래스만 변경 또는 제거시키므로 레이블링 비용 및 시간이 절감된다. 이러한 장점은 터널 영상유고시스템의 유고상황 인지성능의 향상 및 오탐지 데이터 저감으로 이어질 수 있다.

이러한 결론을 바탕으로 터널 영상유고시스템에서 학습시간 단축과 인식을 향상을 위한 오탐지 학습법 후속 연구가 진행될 예정이며, 터널 관리자에게 빈번한 오탐지 발생없이 정확한 유고상황 인지 및 알람을 통해 시스템의 신뢰도를 높일 수 있을 것으로 기대할 수 있다.

## 감사의 글

본 연구는 과학기술정보통신부 한국건설기술연구원 주요사업사업의 일환으로 수행되었다(과제번호 202300 81-001, 극한건설 환경 구현 인프라 및 TRL6 이상급 극한건설 핵심기술 개발).

## 저자 기여도

이규범은 원고 작성, 데이터 수집 및 데이터 분석을 하였고, 신후성은 연구 기획 및 설계, 원고 검토를 수행하였다.

## References

1. Bochkovski, A., Wang, C.Y., Liao, H.Y.M. (2020), “Yolov4: Optimal speed and accuracy of object detection”, arXiv preprint arXiv:2004.10934.
2. Brownlee, J. (2022), Difference between a batch and an epoch in a neural network, <https://machinelearningmastery.com/difference-between-a-batch-and-an-epoch> (December 27, 2023).
3. Dalal, N., Triggs, B. (2005), “Histograms of oriented gradients for human detection”, Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, Vol. 1, pp. 886-893.
4. Fawcett, T. (2006), “An introduction to ROC analysis”, Pattern Recognition Letters, Vol. 27, No. 8, pp. 861-874.

5. Felzenszwalb, P., McAllester, D., Ramanan, D. (2008), "A discriminatively trained, multiscale, deformable part model", Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, pp. 1-8.
6. Frías-Velázquez, A., Nino-Castaneda, J.O., Jelača, V., Pižurica, A., Philips, W. (2011), "A mathematical morphology-based approach for vehicle detection in road tunnels", Applications of Digital Image Processing XXXIV, Vol. 8135, pp. 560-567.
7. Girshick, R., Donahue, J., Darrell, T., Malik, J. (2014), "Rich feature hierarchies for accurate object detection and semantic segmentation", Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, pp. 580-587.
8. LeCun, Y., Bengio, Y., Hinton, G. (2015), "Deep learning", Nature, Vol. 521, No. 7553, pp. 436-444.
9. Lee, K.B., Shin, H.S. (2019), "Effect on self-enhancement of deep-learning inference by repeated training of false detection cases in tunnel accident image detection", Journal of Korean Tunnelling and Underground Space Association, Vol. 21, No. 3, pp. 419-432.
10. Lee, K.B., Shin, H.S. (2022), "A study for improvement of far-distance performance of a tunnel accident detection system by using an inverse perspective transformation", Journal of Korean Tunnelling and Underground Space Association, Vol. 24, No. 3, pp. 247-262.
11. Lewis, D.D. (1990), "Representation quality in text classification: An introduction and experiment", Speech and Natural Language: Proceedings of a Workshop Held at Hidden Valley, Pennsylvania, pp. 288-295.
12. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.Y., Berg, A.C. (2016), "SSD: Single shot multibox detector", Proceedings of the 14th European Conference on Computer Vision (ECCV 2016), Amsterdam, pp. 21-37.
13. Lu, D., Weng, Q. (2007), "A survey of image classification methods and techniques for improving classification performance", International journal of Remote sensing, Vol. 28, No. 5, pp. 823-870.
14. MOLIT (2015), Study on revision of installation and operation guideline for hazard mitigation facilities of road tunnels, Ministry of Land, Infrastructure and Transport, pp. 223-236.
15. MOLIT (2021), Guideline of installation and management of disaster prevention facilities on road tunnels (established rule of MOLIT No. 336), Ministry of Land, Infrastructure and Transport, Korea.
16. Neubeck, A., Van Gool, L. (2006), "Efficient non-maximum suppression", Proceedings of the 18th International Conference on Pattern Recognition (ICPR'06), Vol. 3, pp. 850-855.
17. Redmon, J., Divvala, S., Girshick, R., Farhadi, A. (2016), "You only look once: unified, real-time object detection", Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, pp. 779-788.
18. Ren, S., He, K., Girshick, R., Sun, J. (2015), "Faster R-CNN: Towards real-time object detection with region proposal networks", Proceedings of the Advances in Neural Information Processing Systems, Montreal, pp. 91-99.
19. Sergey, T., Venu, G. (2013), Handbook of Statistics, Vol. 31, Elsevier, Amsterdam, pp. 151-175.
20. Wang, C.Y., Bochkovskiy, A., Liao, H.Y.M. (2023), "YOLOv7: Trainable bag-of-freebies sets new state-

- of-the-art for real-time object detectors”, Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, pp. 7464-7475.
21. Zhu, M. (2004), “Recall, precision and average precision”, Working Paper, Department of Statistics and Actuarial Science, University of Waterloo, Vol. 2, No. 30, pp. 1-11.
  22. Zou, Z., Chen, K., Shi, Z., Guo, Y., Ye, J. (2023), “Object detection in 20 years: A survey”, Proceedings of the IEEE, Vol. 111, No. 3, pp. 257-276.