

논문 2022-17-35

# Bounding Box CutMix와 표준화 거리 기반의 IoU를 통한 재활용품 탐지 (Recyclable Objects Detection via Bounding Box CutMix and Standardized Distance-based IoU)

이 해 진, 정 희 철\*  
(Haejin Lee, Heechul Jung)

**Abstract** : In this paper, we developed a deep learning-based recyclable object detection model. The model is developed based on YOLOv5 that is a one-stage detector. The deep learning model detects and classifies the recyclable object into 7 categories: paper, carton, can, glass, pet, plastic, and vinyl. We propose two methods for recyclable object detection models to solve problems during training. Bounding Box CutMix solved the no-objects training images problem of Mosaic, a data augmentation used in YOLOv5. Standardized Distance-based IoU replaced DIoU using a normalization factor that is not affected by the center point distance of the bounding boxes. The recyclable object detection model showed a final mAP performance of 0.91978 with Bounding Box CutMix and 0.91149 with Standardized Distance-based IoU.

**Keywords** : Deep learning, Convolutional neural network (CNN), Object detection, Recyclable object, Data augmentation, IoU

## I. 서 론

국내 쓰레기 발생량은 매년 꾸준히 증가하고 있다. 특히 COVID-19로 인해 언택트 서비스 수요가 증가하면서 이와 관련된 일회용 플라스틱 쓰레기, 택배 포장 쓰레기의 발생량이 급증하였다. 이렇게 발생한 쓰레기 대부분은 매립되거나 소각된다 [1]. 쓰레기 매립으로 발생하는 환경 문제를 해결하기 위한 다양한 친환경 기술과 정책이 제안되었지만 대부분 원활한 생활폐기물 분리수거를 필요로 한다. 매년 도시별로 증가하고 있는 생활폐기물을 효율적으로 처리하기 위해서는 생활폐기물 배출기준에 따른 자동분리·수거 솔루션을 개발하고 이를 활용한 시스템을 갖추어야 한다. 이러한 기술은 기존 생활폐기물 분리수거 비용을 대폭 절감하고 쓰레기 매립과 환경 문제를 해결하는 자동 분리수거 시스템의 기반이 될 것으로 기대된다. 본 논문에서는 사물탐지 딥러닝 기술을 통해 재활용품을 검출하고 생활폐기물 배출기준에 따라 7가지 클래스로 분류하는 모델을 개발하였다. 그리고 학습 과정에서 발생하는 문제점을 해결하고 더 높은 성능을 보이는 2가지 방식을 제안한다.

객체 탐지 (object detection)는 이미지에 나타나는 물체에 대해 각각 클래스와 바운딩 박스 (bounding box)를 얻는

작업이다. 객체 탐지 모델은 모델 구조에 따라 크게 1단계 검출기 (detector)와 2단계 검출기로 나눌 수 있다. 1단계 검출기는 backbone에서 얻은 이미지의 특징 맵 (feature map)을 통해 localization과 분류 (classification) 문제를 동시에 해결한다. 1단계 검출기의 대표적인 모델로는 YOLO [2]가 있다. 반면에, 2단계 검출기는 backbone외에 Region Proposal Network (RPN) [3]를 사용하여 localization 문제를 해결한 뒤에 분류 문제를 해결한다. 이 때문에 1단계 검출기보다 추론 속도가 느리고 모델 크기가 크지만, 더 높은 정확도를 가진다. 2단계 검출기의 대표적인 모델에는 Mask R-CNN [4]이 있다. 본 논문에서는 임베디드 탑재와 추론 속도를 고려하여 1단계 검출기 모델 중 좋은 성능을 보이는 YOLOv5 [5]를 사용하여 개발하였다.

## II. Bounding Box CutMix

YOLOv5 [5]는 데이터 증강 (data augmentation) 기법 중 하나인 Mosaic [6]를 기본적으로 사용한다. Mosaic는 여러 이미지를 모서리가 맞게 뭉쳐 하나의 학습 이미지로 나타내는 방식으로, 작은 객체 또는 큰 데이터셋을 학습하는 효과가 있다. 하지만, 재활용품 데이터셋을 사용할 때, 그림 1과 같이 학습 이미지에 객체가 없다는 문제가 발생한다. 따라서 해당 문제를 해결하고 성능을 향상하는 방법인 Bounding Box CutMix를 제안한다. 해당 기법은 대한전자공학회 학술대회 [7]에서 처음으로 제안하였고 본 논문에서

\*Corresponding Author (heechul@knu.ac.kr)

Received: Jul. 11, 2022, Revised: Aug. 2, 2022, Accepted: Oct. 11, 2022.

H.J. Lee: Kyungpook National University (M.S.)

H.C. Jung: Kyungpook National University (Assist. Prof.)

※ 본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학CT연구센터 지원사업 (IITP-2022-2020-0-01808\*)의 연구결과로 수행되었으며, 또한 정부 (과학기술정보통신부)의 지원으로 한국연구재단 (2020R1C1C100742311)의 지원을 받아 수행된 연구임.

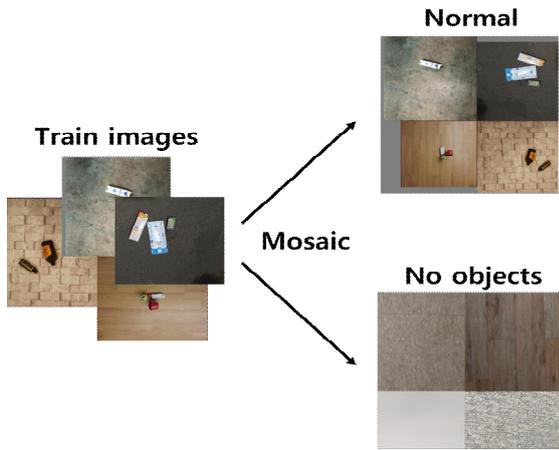


그림 1. Mosaic 사용 시 발생하는 학습 이미지에 객체가 없는 문제 예시  
 Fig. 1. Example of no-object problem with train images when using Mosaic

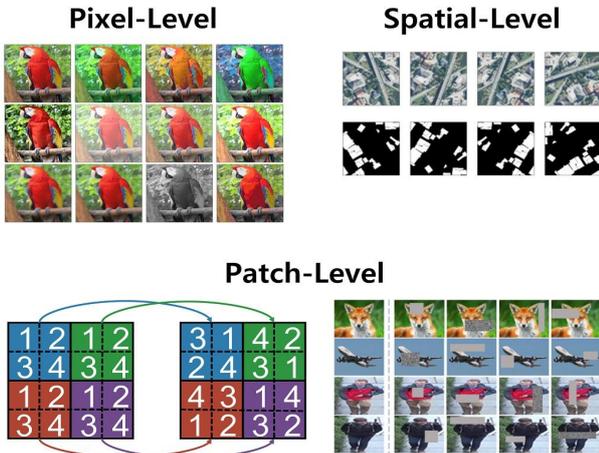


그림 2. One Image Data Augmentation 예시  
 Fig. 2. Example of One Image Data Augmentation

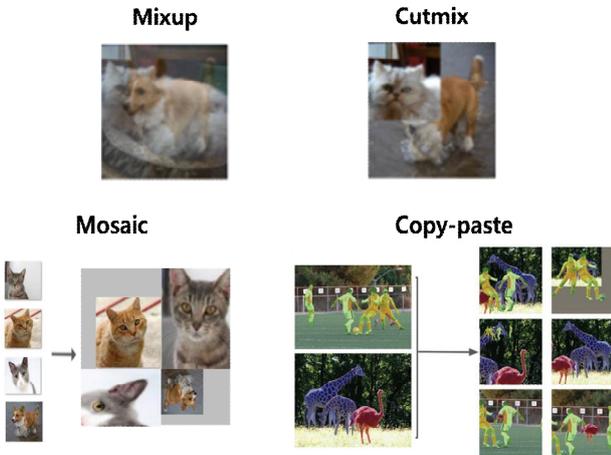


그림 3. Two Images Data Augmentation 예시  
 Fig. 3. Example of Two Images Data Augmentation

는 학술대회 논문 내용을 바탕으로 기존 성능을 향상하고 추가 실험을 진행하였다.

1. Related Works

최근까지 이미지 분류 성능을 향상하기 위해 다양한 데이터 증강 기법이 개발되었다. 객체 탐지 또는 세분화 (segmentation) 작업은 이미지 분류를 기반으로 개발된 데이터 증강 방법을 활용한다. 대부분의 데이터 증강 기법은 모델의 성능을 높여주지만, 일부 데이터 증강 기법은 객체 탐지에 적합하지 않다. 데이터 증강 기법은 크게 하나의 이미지를 사용한 One Image Data Augmentation과 두 개 이상의 이미지를 사용한 Two Images Data Augmentation으로 나눌 수 있다.

1.1 One Image Data Augmentation

One Image Data Augmentation은 하나의 이미지만 사용하는 데이터 증강 기법을 의미한다. 첫 번째로 픽셀 수준 (pixel-level)의 데이터 증강이 있다. 픽셀에 해당하는 강도 (intensity) 값을 임의로 변경하여 이미지의 대비나 색상을 변경할 수 있다. 공간 수준 (spatial-level) 데이터 증강은 rotation, flip, perspective 등 이미지에 기하학적 변환을 적용하는 방법이다. 마지막으로 패치 단위로 이미지를 나누고 특정 패치에 변화를 주는 패치 수준 (patch-level)의 데이터 증강이 있다. 대표적인 방법으로는 CutOut [8]과 PatchShuffle [9]이 있다. One Image Data Augmentation의 예시는 그림 2에서 볼 수 있다.

1.2 Two Images Data Augmentation

One Image Data Augmentation과 달리, Two Images Data Augmentation은 두 개 이상의 이미지를 사용한 데이터 증강 기법을 의미한다. Mixup [10]은 이미지에 투명도를 추가하여 두 이미지를 겹치는 방법이다. Mixup을 사용하는 경우 레이블 (label)이 섞여 나타나므로 soft-labeling이 필요하다. CutMix [11]는 한 이미지를 자른 다음 다른 이미지에 붙여넣는 방법이다. Mosaic [6]는 세 개의 추가 이미지를 무작위로 선택하고 모서리가 맞닿도록 뭉쳐 하나의 학습 이미지로 나타내는 방법이다. Copy-paste [12]는 이미지에 나타나는 객체 마스크 (mask)를 다른 이미지에 붙여넣는 세분화 작업을 위해 개발된 데이터 증강 기법이다. 그림 3에서 Two Images Data Augmentation의 예시를 볼 수 있다.

2. Proposed Method

객체 탐지에서 바운딩 박스 좌표와 클래스는 학습 레이블로 사용된다. 따라서 바운딩 박스에 대한 정보는 데이터 증강 기법을 적용한 후에도 여전히 보존되는 것이 중요하다. 공간 수준의 데이터 증강의 경우 기하학적 변환을 통해 이미지가 변형됨에 따라 바운딩 박스의 형태도 변형된다. 이때 바운딩 박스가 객체를 제대로 둘러싸지 않는 경우가 발생할 수 있다. 정확한 바운딩 박스 좌표가 필요한 모델에는 이러한 데이터 증강 기법을 사용하는 것은 좋지 않다. CutOut [8]과 CutMix [11]처럼 패치 단위로 이미지가 변형

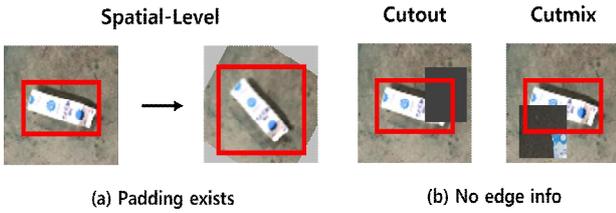


그림 4. 객체 탐지에 부적절한 데이터 증강 기법 예시  
Fig. 4. Example of inappropriate data augmentation for object detection

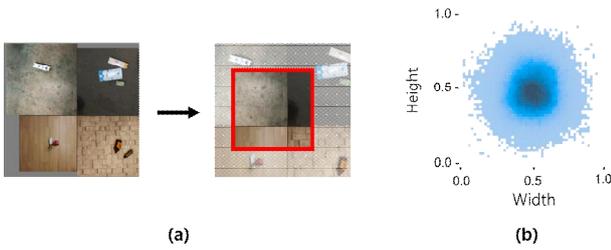


그림 5. 학습 이미지에 객체가 없는 이유  
Fig. 5. Reason why there are no objects in train images

될 경우 특정 객체가 패치에 가려지게 되면서 객체의 테두리에 대한 정보가 사라지게 되고 이러한 데이터는 학습에 부적절하다. 그림 4에서 객체 탐지에 부적절한 데이터 증강 기법 예시를 볼 수 있다.

2.1 학습 이미지에 객체가 없는 문제

YOLOv5 [5]는 초기 학습의 안정화와 큰 데이터셋 학습 효과를 위해 기본적으로 Mosaic을 사용한다. 그러나 그림 1과 같이 객체가 없는 학습 이미지를 생성하는 경우가 발생한다. 일반적으로 객체 탐지 모델 학습에서는 FP (false positive)를 줄이기 위해 데이터셋에 0~10%의 배경 이미지를 포함하는 경향이 있다. 그러나 재활용품 데이터셋의 경우 동일한 객체에 다른 배경을 가진 이미지가 충분히 많으므로 해당 방식을 고려하지 않고 성능을 더 올릴 수 있는 여지가 있다. 그림 5의 (b)는 재활용품 데이터셋의 대부분 객체가 이미지 중앙에 위치한다는 것을 보여준다. 또 Mosaic 적용 후 그림 5의 (a)와 같이 이미지에 scale, shift 추가로 적용되어 객체가 없는 학습 이미지가 생성된다. Mosaic를 사용하지 않으면, 해당 문제를 쉽게 해결 가능하지만 생성된 학습 이미지에서 scale, shift를 적용하여 다양한 데이터를 생성해 큰 데이터셋을 학습하는 효과가 있기 때문에 Mosaic를 그대로 사용하면서 문제를 해결할 수 있는 방법이 필요하다.

2.2 Bounding Box CutMix

학습 이미지에 객체가 없는 문제를 해결하기 위해 바운딩 박스를 활용한 데이터 증강 기법인 Bounding Box CutMix를 개발하였다. 해당 데이터 증강 기법은 재활용품 데이터셋의 대부분 객체가 이미지 중심에 위치한다는 특성을 활용하여 학습 이미지의 객체 주변에 학습 데이터셋의 객체를 임의로 추출하여 삽입하는 방법이다. Bounding Box

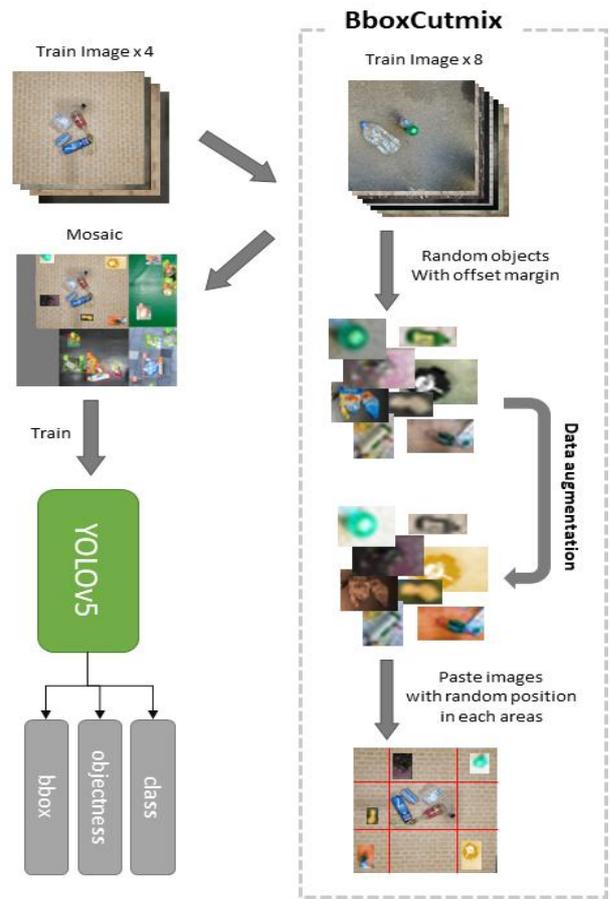


그림 6. Bounding Box CutMix 수행 과정  
Fig. 6. Procedure of Bounding Box CutMix

CutMix는 다음 절차에 따라 최종 학습 이미지를 생성한다. 먼저 학습 이미지에 나타나는 모든 객체를 둘러싸는 가장 작은 상자를 기준으로 이미지를 총 9 영역으로 나누어 객체 주변에 8개 영역을 생성한다. 학습 데이터셋에서 8개의 이미지를 무작위로 선택하고 각 이미지에서 객체 하나를 무작위로 추출한다. 이때 바운딩 박스 그대로 추출하는 경우 모델이 학습 중 바운딩 박스를 쉽게 찾아 테스트에서 성능이 떨어지는 것을 확인하였고 이를 방지하기 위해 margin offset을 추가하여 추출했다. 또 학습 데이터셋에서 중복된 객체를 사용하여 발생할 수 있는 과적합을 방지하기 위해 추출된 객체에 대해 flip, rotate, shift, scale, brightness, contrast, saturation 등 One Image Data Augmentation을 추가로 적용했다. 마지막으로 추출된 8개의 객체는 앞에서 나누었던 이미지의 8개의 영역에 각각 삽입된다. 그림 6에서 Bounding Box CutMix을 활용한 학습 이미지 생성 과정을 확인할 수 있다.

III. 표준화 거리 기반의 IoU

재활용품 데이터셋의 바운딩 박스 크기 분포를 분석한 결

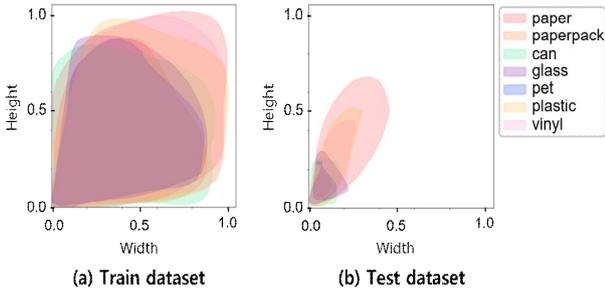


그림 7. 재활용품 데이터셋의 바운딩 박스 크기 분포  
Fig. 7. Size distribution of bounding boxes in Recyclable objects dataset

과, 그림 7과 같이 학습 데이터셋과 테스트 데이터셋 사이에 차이가 있다는 것을 알 수 있다. 이 경우 바운딩 박스의 교차 영역 (intersection)의 크기가 동일하더라도 IoU 임계값 (threshold)에 따라 background 또는 객체로 잘못 분류될 수 있기 때문에 IoU 값의 최적화가 필요하다.

### 1. Related Work

IoU는 두 바운딩 박스가 겹치는 정도를 의미한다. NMS [13] 또는 예측된 바운딩 박스 수를 줄이기 위한 임계값에 사용되는 중요한 성능 지표이며, ground truth의 바운딩 박스와 일치 정도를 비교하기 위해 사용된다. IoU는 두 바운딩 박스가 겹치는 정도만 나타내기 때문에 서로 겹치지 않는 경우 바운딩 박스 사이의 거리 또는 크기 비율을 직접적으로 알기 어렵다. 따라서 IoU 값은  $R_{GIoU}$  [14],  $R_{DIoU}$ ,  $R_{CIoU}$  [15]와 같은 regulator를 추가하여 보정하게 된다. 최종 IoU 값은 IoU 값에서 regulator 값을 빼서 얻을 수 있다.

#### 1.1 GIoU

GIoU [14]는 ground truth의 바운딩 박스와 겹치지 않는 경우 예측한 바운딩 박스의 위치에 관계없이 IoU 값이 0으로 고정되는 문제를 해결하기 위해 제안되었다. 두 개의 바운딩 박스를 포함하는 최소 크기 박스를 기준으로 ground truth와 예측한 바운딩 박스를 제외한 영역의 크기를 의미한다. GIoU는 아래 수식 1과 같이 계산된다.

$$R_{GIoU} = \frac{|C - (A \cup B)|}{|C|}. \quad (1)$$

수식 1에서  $C$ 는 두 개의 바운딩 박스를 포함하는 최소 크기의 상자를 가리킨다. 교차 영역의 크기가 0일 때, IoU 손실 (loss) 값은 1을 가졌지만 바운딩 박스의 대략적인 거리를 나타내는  $R_{GIoU}$ 를 통해 1과 2 사이의 값을 가지게 된다.

#### 1.2 DIoU & CIoU

두 바운딩 박스 사이의 교차 영역의 크기가 0일 때 GIoU를 통해 대략적인 거리를 계산할 수 있었다.  $R_{GIoU}$  값은 바운딩 박스의 위치를 고려하지 않기 때문에 두 바운딩 박스의 거리가 달라지더라도 같은 IoU 값을 가지는 문제가 발생한다. 이 문제를 해결하기 위해 DIoU [15]가 제안되었다.

DIoU는 두 바운딩 박스의 중심 좌표 거리를 나타낸다. DIoU의 수식은 다음과 같다.

$$R_{DIoU} = \frac{\rho^2(b, b^{gt})}{c^2}. \quad (2)$$

수식 2에서  $b$ 와  $b^{gt}$ 는 예측된 바운딩 박스와 ground truth 바운딩 박스의 중심좌표를 나타내고  $\rho$ 는 두 바운딩 박스의 중심좌표 간 거리를 나타낸다.  $c$ 는 두 바운딩 박스를 포함하는 최소 크기 박스의 대각선 길이를 나타내고, 무한히 커질 수 있는 중심 좌표의 거리를 정규화하는데 사용한다. 중심 좌표의 거리는 대각선의 길이를 초과할 수 없으므로,  $R_{DIoU}$ 는 항상 0과 1 사이의 값을 갖는다. Zhaohui [15]는 DIoU 뿐만 아니라 CIoU도 제안했다. DIoU는 단순히 바운딩 박스의 중심 좌표 거리를 계산하기 때문에 바운딩 박스 크기에 대한 비교는 이루어지지 않는다. 이를 위해 CIoU는 가로, 세로 비율과  $\arctan$ 를 사용하여 각도를 구하고 ground truth와의 차이를 계산한다. CIoU에 대한 방정식은 다음과 같다.

$$R_{CIoU} = av \begin{cases} a = \frac{\nu}{(1 - IoU) + \nu} \\ \nu = \frac{4}{\pi^2} (\arctan \frac{w^{gt}}{h^{gt}} - \arctan \frac{w}{h})^2 \end{cases}. \quad (3)$$

$\nu$ 는 두 바운딩 박스의 가로, 세로 비율을 나타내는 각도 값의 차이를 나타낸다. 이 때  $\arctan$  값은  $\frac{\pi}{2}$ 와  $-\frac{\pi}{2}$  사이 값을 가지기 때문에 제곱값에 대해  $\frac{4}{\pi^2}$ 로 정규화한다.  $a$ 는 IoU 값에 따라  $\nu$  값을 조절하는 trade-off 파라미터로 IoU 값이 큰 박스에 대해 페널티를 주는 역할을 한다.  $R_{CIoU}$ 는 오로지 바운딩 박스 크기만을 비교하기 때문에 바운딩 박스 간의 거리를 측정하는  $R_{DIoU}$ 와 함께 사용된다.

### 2. Proposed Method

YOLOv5에서는 IoU 지표 중 DIoU와 CIoU를 사용한다. 앞에서 말했듯이 재활용품 데이터셋에서 학습 데이터셋과 테스트 데이터셋 간의 바운딩 박스 크기 분포 차이가 크기 때문에 IoU 최적화가 필요했고  $R_{DIoU}$ 를 구하는 수식 2에서 문제점을 발견하였다. Ground truth의 바운딩 박스 중심 좌표가 원점으로 설정하고 예측한 바운딩 박스의 중심 좌표과의 거리를 벡터로 나타낼 때 이루는 각도를  $\theta$ 라고 가정한다. 그리고 중심 좌표 간 거리를  $d$ , 각 바운딩 박스의 가로, 세로 길이를  $(w, h)$ ,  $(w', h')$ 로 가정한다면  $R_{DIoU}$ 에 대한 수식은 다음과 나타낼 수 있다. 그림 8은 수식 4의 각 요소 값을 이해하기 쉽게 나타냈다. 수식 4에서 정규화 역할을 하는 대각선 길이가 중심 좌표 간 거리에 영향받을 것을 알 수 있다. 이때 정규화 인자인 대각선 길이로 인해 두 가지 문제가 발생한다.

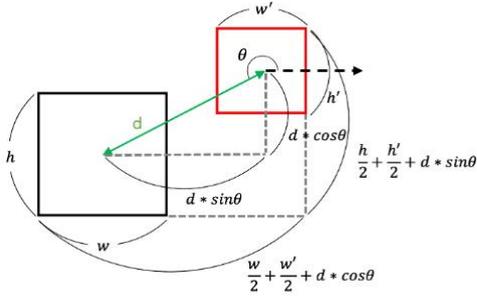


그림 8. 중심좌표 거리와 각도에 따른 각 요소 값  
Fig. 8. Each values according to the center coordinate distance and angle

$$R_{DIoU} = \frac{(d \cdot \cos\theta)^2 + (d \cdot \sin\theta)^2}{\left(\frac{w}{2} + \frac{w'}{2} + d \cdot \cos\theta\right)^2 + \left(\frac{h}{2} + \frac{h'}{2} + d \cdot \sin\theta\right)^2} \quad (4)$$

2.1 바운딩 박스 위치 문제

그림 9는  $R_{DIoU}$ 의 바운딩 박스 위치 문제 예시를 보여준다. 두 경우 모두 같은 중심좌표 간 거리를 가지지만 예측된 바운딩 박스의 위치가 다른 것을 알 수 있다. 중심 좌표 간 거리는 동일하지만 ground truth의 바운딩 박스 중심 좌표와의 각도에 따라 대각선의 길이가 달라져  $R_{DIoU}$  값이 다르다는 문제가 발생한다.

2.2 바운딩 박스 스케일 문제

예측된 바운딩 박스의 크기에 따라서도 값이 달라지는 문제가 발생한다. 그림 10은  $R_{DIoU}$ 의 바운딩 박스 스케일 문제 예시를 보여준다. 세 가지 경우 모두 중심 좌표 간 거리는 동일하지만 바운딩 박스 크기에 따라 대각선의 길이 다르고 결과적으로 각각 다른  $R_{DIoU}$  값을 가지게 된다.

2.3 표준화 거리 기반의 IoU

바운딩 박스 위치 문제와 스케일 문제는 대각선 길이를 정규화 인자로 사용함으로써 발생한다. 따라서, 유클리드 거리를 기반으로 하는 DIoU 대신 각 차원 별 거리를 해당 차원 변수들의 분산으로 정규화하는 표준화 거리를 이용한 IoU를 제안한다. 분산을 사용하는 표준화 거리는 수식 5처럼 계산된다.

$$Standardized\ distance = \sqrt{\left(\frac{x_1 - x_2}{\sigma_1}\right)^2 + \left(\frac{y_1 - y_2}{\sigma_2}\right)^2} \quad (5)$$

$d_x$ 를  $x$  축 중심좌표 간 거리라고 가정한다면  $x$  축에 해당하는 요소는  $d_x$  와 0이기 때문에  $\sigma_1^2$ 는 다음과 같이 나타낼 수 있다.

$$\sigma_1^2 = \frac{\left(d_x - \frac{d_x}{2}\right)^2 + \left(0 - \frac{d_x}{2}\right)^2}{2} = \left(\frac{d_x}{2}\right)^2 \quad (6)$$

수식 7에서 보이는 것처럼  $d_x$  값이 두 바운딩 박스의 가로 길이 이 평균과 같다면 표준화 거리의  $x$  성분 값은 4가 된다. 또한  $d_x$  값이 평균보다 작으면 4보다 작고 크면 4보다

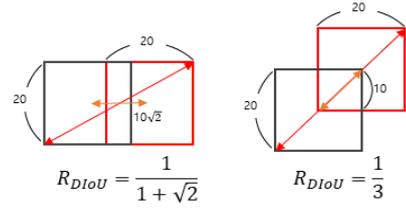


그림 9. DIoU의 바운딩 박스 위치 문제 예시  
Fig. 9. Example of bounding box location problem on DIoU

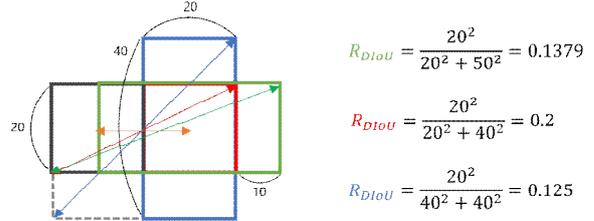


그림 10. DIoU의 바운딩 박스 스케일 문제 예시  
Fig. 10. Example of bounding box scale problem on DIoU

큰 값을 얻게 된다.  $y$  성분 값도 두 바운딩 박스의 세로 길이 평균에 대해 똑같이 적용된다.

$$\sigma_1^2 = \left(\frac{w+w'}{4}\right)^2 \begin{cases} SD_x < 4, & d_x < \frac{w+w'}{2} \\ SD_x = 4, & d_x = \frac{w+w'}{2} \\ SD_x > 4, & d_x > \frac{w+w'}{2} \end{cases} \quad (7)$$

수식 7에서  $SD_x$ 는 표준화 거리의  $x$  성분을 의미한다.  $d_x$  값을 두 바운딩 박스의 가로 길이 평균으로 대체하면 거리에 따라 바운딩 박스 중첩 여부를 구분할 수 있고 가로 길이 이 평균보다 작은 경우 0과 1사이의 값으로 정규화할 수 있다. 수식 7의  $\sigma_1^2$ 과  $\sigma_2^2$ 을 수식 5에 대입한다면 다음과 같이 정리할 수 있다.

$$Standardized\ distance^2 = \frac{\rho^2(b_x, b_x^{gt})}{\left(\frac{w+w'}{4}\right)^2} + \frac{\rho^2(b_y, b_y^{gt})}{\left(\frac{h+h'}{4}\right)^2} \quad (8)$$

유도된 표준화 거리는 수식 8처럼 각  $x, y$  성분이 가로, 세로 평균과 같을 때 4의 값을 가진다. 이 경우, 각 성분이 평균보다 작은 경우 (바운딩 박스가 서로 겹치는 경우)를 기준으로 0과 1사이의 값을 가지기 위해 4로 나누어 정규화를 한다. 최종적으로 본 논문에서 제안하는 표준화 거리 기반의 IoU는 수식 9로 나타낼 수 있다.

$$R_{SDIoU} = \frac{1}{4} \left( \frac{\rho^2(b_x, b_x^{gt})}{\left(\frac{w+w'}{4}\right)^2} + \frac{\rho^2(b_y, b_y^{gt})}{\left(\frac{h+h'}{4}\right)^2} \right) = \frac{\rho^2(b_x, b_x^{gt})}{\left(\frac{w+w'}{2}\right)^2} + \frac{\rho^2(b_y, b_y^{gt})}{\left(\frac{h+h'}{2}\right)^2} \quad (9)$$

IV. 실험 결과

1. 데이터셋 및 실험 환경

2020년 AI 그랜드 챌린지 [16]에서 제공한 452개의 이미지를 테스트 데이터셋으로 사용했고 종이, 종이 상자, 캔, 유리, 페트병, 플라스틱, 비닐 총 7개 클래스로 구성된다. 7만 8680개의 이미지를 추가로 수집해 학습 데이터셋으로 사용하였다. 추가로 표준화 거리기반의 IoU 실험에서는 PASCAL VOC [17] 데이터셋을 사용하여 제안된 IoU 지표의 보편성을 확인하였다. PASCAL VOC 데이터셋에서 VOC 2007 테스트 데이터만 테스트 데이터셋으로 사용하였고 나머지 VOC 2007과 VOC 2012의 모든 학습 및 테스트 데이터는 학습 데이터셋으로 사용하였다.

모든 실험은 8개의 NVIDIA RTX 8000 48GB GPU를 통해 학습 및 추론을 진행했고, 학습 모델은 YOLOv5 모델 중 P6 아키텍처가 적용된 X6 모델을 이용하였다 [5]. 하이퍼 파라미터는 YOLOv5 [18]에서 제공하는 COCO scratch 파라미터를 사용하였다. 배치 크기는 32로 설정하고 학습은 총 30번 진행하였다. Bounding Box CutMix 실험에서는 Mosaic 미적용, Mosaic, Mixup, copy-paste와 비교하고 추가로 여러 기법을 동시에 적용하여 성능비교를 하였다. 특히 copy-paste는 객체 탐지에서 사용하기 위해 마스크 대신 바운딩 박스를 사용하도록 수정하였다. 표준화 거리 기반의 IoU 실험에서도 YOLOv5가 CIoU를 기본값으로 사용하기 때문에 CIoU를 이용한 실험도 수행되었다. 또한, 수식 9의 가로와 세로 평균 대신 RMS (root mean square)를 사용한 표준화 거리 기반 IoU도 추가로 실험하였다.

2. 실험 결과

2.1 Bounding Box CutMix

표 1에서 Bounding Box CutMix에 대한 실험결과를 확인할 수 있다. Mosaic의 경우 mAP@0.5는 0.9109의 성능을 보였고, Bounding Box CutMix의 경우 0.91967로 다른 방법을 사용한 실험보다 우수한 성능 향상을 보여주었다. 여러 기법을 동시에 사용하는 경우 성능이 떨어지는 경우가 있었지만 테스트 loss는 안정적으로 감소하여 학습 모델이 학습 데이터셋에 과적합된 것이 아닌 테스트 데이터셋과의 분포 차이로 인한 것으로 확인하였다. Copy-paste의 데이터 증강 결과는 Bounding Box CutMix와 유사하지만 성능에서 차이를 보인다. 학습 이미지에 copy-paste와 Bounding Box CutMix를 적용한 예시를 그림 11에서 확인할 수 있다. 두 방법 모두 임의의 위치에 무작위로 추출된 객체를 삽입한다는 점에서 동일하지만, Bounding Box CutMix는 margin offset을 추가해 바운딩 박스를 찾기 어렵게 만들고, 객체를 삽입 후에도 기존 이미지에 존재하던 객체들의 테두리 정보가 그대로 유지되는 것을 알 수 있다. 이러한 특징들은 Bounding Box CutMix가 copy-paste보다 좋은 성능을 보이는 이유로 볼 수 있다.

표 1. 데이터 증강 기법에 따른 mAP 비교

Table 1. mAP comparison for data augmentation methods

Method	mAP@0.5	mAP@0.5:0.95
No Mosaic	0.89508	0.77857
Base(Mosaic)	0.91090	0.79401
+ Mixup	0.91700	0.80490
+ copy-paste	0.91668	0.79978
+ Mixup + copy-paste	0.91291	0.79952
+ BboxCutMix	0.91967	0.80758
+ BboxCutMix + Mixup	0.91890	<b>0.80798</b>
+ BboxCutMix + copy-paste	<b>0.91978</b>	0.80622
+ All	0.90412	0.79141

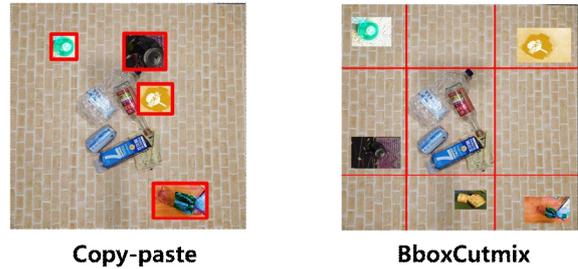


그림 11. copy-paste와 Bounding Box CutMix 예시

Fig. 11. Example of copy-paste and Bounding Box Cutmix

표 2. IoU 지표에 따른 mAP 비교

Table 2. mAP comparison for IoU metrics

Method	mAP@0.5	mAP@0.5:0.95
Base (CIoU)	0.91090	0.79401
DIoU	0.90234	0.78660
SDIoU (mean)	<b>0.91149</b>	<b>0.79759</b>
SDIoU (RMS)	0.90825	0.79313

표 3. IoU 지표에 따른 mAP 비교 (PASCAL VOC)

Table 3. mAP comparison for IoU metrics (PASCAL VOC)

Method	mAP@0.5	mAP@0.5:0.95
DIoU	0.88758	0.68262
SDIoU (mean)	0.88900	<b>0.68573</b>
SDIoU (RMS)	<b>0.89029</b>	0.68335

2.2 표준화 거리 기반의 IoU

재활용품 데이터셋에 대한 IoU 지표 실험 결과는 표 2에서 확인할 수 있다. DIoU를 사용한 경우 CIoU에 비해 mAP@0.5는 0.90234로 낮은 성능을 보였으며, 본 논문에서 제안한 방법인 표준화 거리 기반의 IoU를 사용한 경우 mAP@0.5에서 0.91149로 CIoU보다 더 높은 성능을 보였다. 추가로 RMS를 사용한 표준화 거리 기반의 IoU는 CIoU보다 낮지만 DIoU보다 높은 성능을 보였다.

PASCAL VOC 데이터셋에 대한 실험 결과는 표 3에서 확인할 수 있다. 재활용품 데이터셋에 비해 두드러지는 성

능 향상이 없지만 RMS를 사용하는 경우 mAP@0.5는 0.89029로 DIoU보다 0.003 높은 성능을 보여주었고 이를 통해 표준화 거리 기반의 IoU의 보편성을 확인하였다.

## V. 결론

자체적으로 수집한 재활용품 데이터셋과 1단계 검출기 모델인 YOLOv5를 사용하여 재활용품을 탐지하고 7가지 클래스로 분류하는 객체 탐지 모델을 개발하였다. 본 논문에서는 모델 학습 중에 발생한 문제들을 해결하고 기존 기법들보다 높은 성능을 보인 두 가지 방법을 제안한다.

첫 번째로 Bounding Box CutMix는 Mosaic를 사용하면 생성되는 학습 이미지에서 객체가 없는 문제를 해결한다. 특히 Bounding Box CutMix는 기존 copy-paste와 달리 모든 객체의 테두리 정보를 그대로 보존하고 무작위로 추출된 객체 이미지에 margin offset을 추가해 바운딩 박스를 쉽게 찾는 것을 방지하였다. 추가로 추출된 객체에 적용된 데이터 증강은 다양한 데이터를 학습하는 효과를 가진다.

표준화된 거리 기반 IoU는 DIoU에서 발견된 바운딩 박스 위치 문제와 스케일 문제를 해결한다. 각 차원의 분산을 정규화 인자로 사용하는 표준화 거리를 사용하였다. 특히 예측된 바운딩 박스의 위치와 크기에 따른 regulator 값 비교는 표준화 거리 기반의 IoU가 바운딩 박스 위치에 상관없이 동일한 값을 갖는다. 추가로 PASCAL VOC 데이터셋에 대한 실험을 통해 제안된 IoU 지표의 보편성을 보였다.

결과적으로 Bounding Box CutMix와 표준화 거리 기반의 IoU를 통해 학습에서 발생한 문제들을 해결하고 기존 방법보다 더 높은 성능을 도달하였다.

## References

- [1] P. Kim, D. B. Im, "Understanding Hybrid Governance: The Case of Waste Recycle Management," *The Korean Governance Review*, Vol. 27, No. 2, pp. 29-56, 2020 (in Korean).
- [2] J. Redmon, S. K. Divvala, R. B. Girshick, A. Farhadi, "You only look once: Unified, Real-time Object Detection," In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 779-788, 2016.
- [3] S. Ren, K. He, R. B. Girshick, J. Sun, "Faster R-CNN: Towards Real-time Object Detection with Region Proposal Networks," *Advances in Neural Information Processing Systems*, Vol. 28, 2015.
- [4] K. He, G. Gkioxari, P. Dollár, R. B. Girshick, "Mask R-CNN," In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2961-2969, 2017.
- [5] Ultralytics, "YOLOv5," <https://github.com/ultra-lytics/yolov5>, 2022.
- [6] A. Bochkovskiy, C. Wang, H. M. Liao, "Yolov4: Optimal Speed and Accuracy of Object Detection," *arXiv preprint, arXiv:2004.10934*, 2020.
- [7] H. J. Lee, H. C. Jung, "Implementation of Recyclable Trash Object Detection Model Based on YOLOv5," In *Proceedings of the Institute of Electronics and Information Engineers*, pp. 1066-1069, 2021 (in Korean).
- [8] T. Devries, G. W. Taylor, "Improved Regularization of Convolutional Neural Networks with Cutout," *arXiv preprint, arXiv:1708.04552*, 2017.
- [9] G. Kang, X. Dong, L. Zheng, Y. Yang, "Patchshuffle Regularization," *arXiv preprint, arXiv:1707.07103*, 2017.
- [10] H. Zhang, M. Cissé, Y. N. Dauphin, D. Lopez-Paz, "Mixup: Beyond Empirical risk Minimization," *arXiv preprint, arXiv:1710.09412*, 2017.
- [11] S. Yun, D. Han, S. J. Oh, S. Chun, J. Choe, Y. Yoo, "Cutmix: Regularization Strategy to Train Strong Classifiers with Localizable Features," In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 6023-6032, 2019.
- [12] G. Ghiasi, Y. Cui, A. Srinivas, R. Qian, T. Lin, E. D. Cubuk, Q. V. Le, B. Zoph, "Simple Copy-paste is a Strong Data Augmentation Method for Instance Segmentation," In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2918-2928, 2021.
- [13] N. Bodla, B. Singh, R. Chellappa, L. S. Davis, "Improving Object Detection with one line of code," *arXiv preprint, arXiv:1704.04503*, 2017.
- [14] H. Rezatofighi, N. Tsoi, J. Y. Gwak, A. Sadeghian, I. Reid, S. Savarese, "Generalized Intersection Over Union: A Metric and A loss for Bounding box Regression," In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 658-666, 2019.
- [15] Z. Zheng, P. Wang, W. Liu, J. Li, R. Ye, D. Ren, "Distance-iou Loss: Faster and Better Learning for Bounding box Regression," In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 34, No. 07, pp. 12993-13000, 2020.
- [16] IITP, "2020 AI Grand Challenge," <https://ai-challenge.kr>, 2020.
- [17] M. Everingham, L. Gool, C. K. Williams, J. Winn, A. Zisserman, "The Pascal Visual Object Classes (voc) Challenge," *International Journal of Computer Vision*, Vol. 88, No. 2, pp. 303-338, 2010.
- [18] Ultralytics, "YOLOv5 pre-trained weights of COCO dataset," <https://github.com/ultralytics/yolov5/releases/tag/v6.1>, 2022.

**Haejin Lee (이 해 진)**

2020 Computer and Science Engineering from  
Kyungpook National University (B.S.)

2022 Artificial Intelligence from Kyungpook  
National University (M.S.)

## Career:

2022 Captos, AI Research Engineer

Field of Interests: Deep Learning & Computer Vision

Email: seareale@gmail.com

**Heechul Jung (정 희 철)**

2007 Internet from Sejong University (B.S.)

2010 Information and Communication Engineering  
from the Gwangju Institute of Science and  
Technology (M.S.)

2018 Electrical Engineering from the Korea  
Advanced Institute of Science and  
Technology (Ph.D.)

2019~ Artificial Intelligence from Kyungpook National University  
(Assistant Professor)

## Career:

2019 Senior Researcher, Hyundai Motor Company

2019 Assistant Professor, Kyungpook National University

Field of Interests: Deep Learning & Computer Vision

Email: heechul@knu.ac.kr