

Mask R-CNN을 이용한 물체인식 및 개체분할의 학습 데이터셋 자동 생성

Automatic Dataset Generation of Object Detection and Instance Segmentation using Mask R-CNN

조현준¹·김다윗²·송재복[†]
HyunJun Jo¹, Dawit Kim², Jae-Bok Song[†]

Abstract: A robot usually adopts ANN (artificial neural network)-based object detection and instance segmentation algorithms to recognize objects but creating datasets for these algorithms requires high labeling costs because the dataset should be manually labeled. In order to lower the labeling cost, a new scheme is proposed that can automatically generate a training images and label them for specific objects. This scheme uses an instance segmentation algorithm trained to give the masks of unknown objects, so that they can be obtained in a simple environment. The RGB images of objects can be obtained by using these masks, and it is necessary to label the classes of objects through a human supervision. After obtaining object images, they are synthesized with various background images to create new images. Labeling the synthesized images is performed automatically using the masks and previously input object classes. In addition, human intervention is further reduced by using the robot arm to collect object images. The experiments show that the performance of instance segmentation trained through the proposed method is equivalent to that of the real dataset and that the time required to generate the dataset can be significantly reduced.

Keywords: Object Detection, Instance Segmentation, Deep Learning, Dataset Generation

1. 서 론

최근 로봇 분야는 인공지능 기술과의 융합을 통해 많은 발전을 이루고 있다^{1,2}. 특히, 이미지 상의 물체의 종류와 위치를 검출하는 물체인식(object detection) 기술과 형상까지도 검출하는 개체분할(instance segmentation) 기술은 로봇 분야에서 다양한 작업의 수행을 위해서 활용되고 있다. 그러나 물체인식과 개체분할을 위한 데이터셋의 생성 비용이 높으므로, 이들 기술과 로봇을 융합한 기술은 아직 잘 활용되지 못하고 있다.

데이터셋의 생성 과정 중 레이블링(labeling) 단계에 가장 많은 비용이 소모되므로, 이를 줄임으로써 물체인식 및 개체

분할의 활용도가 더욱 높아질 수 있다. 레이블은 데이터가 의미하는 것을 표시하는 일종의 정답이고, 이와 같은 정답을 표시하는 작업을 레이블링이라고 한다. 물체인식의 경우, 이미지 상의 모든 대상 물체들의 종류와 위치를 표시하는 것을 레이블링 작업이라고 한다. 이와 같은 레이블링 작업에 대해, 공개된 이미지 데이터셋 중 하나인 Open Images Dataset v4에서는 하나의 물체에 대한 레이블링 작업을 위해 42초가 걸렸다고 명시하고 있다³. 이와 같이 물체인식과 개체분할은 큰 레이블링 비용이 요구되므로, 이를 감소시키기 위한 연구가 필요하다.

레이블링 비용의 감소를 위한 연구로, 크게 4종류의 연구가 수행되고 있다. 첫 번째로, 실제 이미지와 물체의 종류만으로 물체의 위치를 검출하는 방법이다⁴. 이 방식은 경계박스(bounding box)를 학습하지 않음에도, 물체가 있을 가능성이 높은 위치를 찾아냄으로써 레이블링 비용을 줄이는 데 도움을 줄 수는 있지만, 사람이 레이블링하는 것에 비해서 부정확하다. 두 번째로, 사람이 개입하는 능동 학습(active learning)을

Received : Dec. 10. 2018; Revised : Jan. 11. 2019; Accepted : Jan. 15. 2019

※ This work was supported by IITP grant funded by the Korea Government MSIT. (No. 2018-0-00622)

1. Student, Korea University, Seoul, Korea (jhj0630@korea.ac.kr)

2. Student, Korea University, Seoul, Korea (ekdh1rta@korea.ac.kr)

† Professor, Corresponding author: Mechanical Engineering, Korea University, Seoul, Korea (jbsong@korea.ac.kr)

이용하여, 이미지에서 물체가 있을 가능성이 높은 곳을 사람에게 보여주고, 경계박스를 그리는 작업 자체는 사람이 진행하는 방식이다⁵⁾. 이는 첫 번째 방식에 비해 정확한 레이블링이 가능하지만, 레이블링 비용 절감 효과는 첫 번째 방식보다 작게 된다. 세 번째로, 학습을 통해 여러 경계박스를 사람에게 제시해주고, 사람은 사용 가능 여부만을 판단하는 방식이다⁶⁾. 이는 레이블링 비용이 대폭 줄어드는 장점이 있지만, 학습 대상 물체가 변경되면 레이블링을 위한 알고리즘도 재학습해야 하는 단점이 있다. 네 번째로, 시뮬레이션에서 조명, 물체들의 색, 형태 등을 변화시키면서 다양한 환경에서의 데이터를 얻는 방법이다⁷⁾. 이는 다양한 환경에서 데이터를 얻을 수 있으며 자동화가 가능하지만, 생성된 데이터로 학습 시에 실제 환경과의 차이로 인해 실제 환경에서 잘 적용될지는 불확실한 단점이 있다. 또한, 위 4가지 방법들은 모두 물체인식 알고리즘에는 적용 가능하지만, 개체분할 알고리즘에는 적용이 불가능하다.

본 연구의 목표는 처음 보는 물체도 인식할 수 있도록 학습시킨 개체분할 알고리즘을 이용하여, 레이블링 작업을 자동으로 수행함으로써 데이터셋 생성을 위해 필요한 레이블링 비용을 대폭 절감할 수 있을 뿐만 아니라, 물체인식 및 개체분할 알고리즘 모두에 적용 가능한 방법을 제안하는 것이다. 먼저 대상 물체를 로봇을 이용하여 다양한 각도에서의 이미지를 얻고, 처음 보는 물체의 마스크를 얻을 수 있도록 학습한 개체분할 알고리즘을 이용하여 배경을 제거한다. 배경이 제거된 물체 이미지를 다양한 배경 이미지와 합성하며, 다양한 환경을 반영하기 위해 여러 후처리 기법들을 사용한다. 마지막으로, 합성 시에 사용한 물체의 종류와 위치 정보를 통해, 레이블링 작업을 자동으로 수행한다.

본 논문의 기여는 다음과 같다. 첫째, 학습데이터의 구조에 변화를 주어, 처음 보는 물체도 인식할 수 있도록 Mask R-CNN을 학습하였다. 둘째, 데이터의 레이블링 작업에 필요한 시간을 줄임으로써, 데이터셋 생성 시간을 절감하였다. 셋째, 물체인식과 개체분할 알고리즘 모두에 적용할 수 있다.

본 논문은 다음과 같이 구성된다. 2장에서는 데이터 생성과 관련된 전반적인 내용을 서술하고, 3장에서는 배경 제거 알고리즘을 학습하기 위한 학습데이터셋의 구성 방법에 대해 서술한다. 4장에서는 생성된 데이터셋으로 학습한 결과와 실제 데이터셋으로 학습한 결과를 비교한다. 마지막으로, 5장에서는 결론을 서술한다.

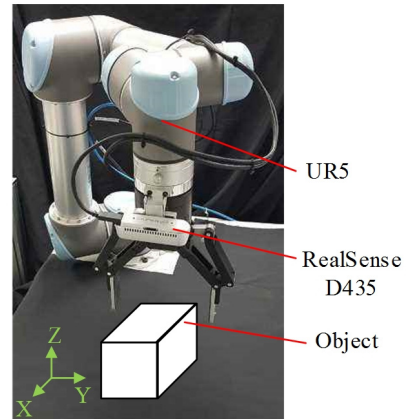
2. 데이터셋 생성 방법

데이터셋의 생성은 물체 이미지 수집(object image acquisition),

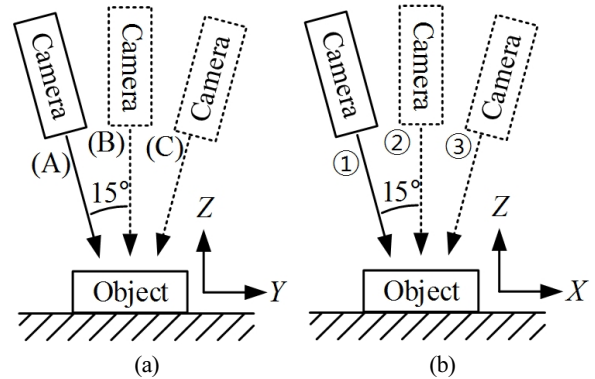
배경 제거(background subtraction), 이미지 합성(image synthesis), 자동 레이블링(automatic labeling)의 순서로 진행된다. 물체 이미지 수집은 로봇을 이용하여 진행되며, 다양한 각도에서 물체의 이미지를 얻는다. 배경 제거에서는 처음 보는 물체의 형상도 찾을 수 있는 Mask R-CNN을 이용하여, 물체 이미지에서 배경을 제거하고, 물체의 마스크를 얻는다. 이미지 합성에서는 배경이 제거된 물체 이미지와 따로 준비된 배경 이미지를 합성하여, 새로운 이미지를 생성한다. 마지막으로, 물체 이미지의 마스크를 이용하여 레이블링 작업을 자동으로 수행한다. 아래에 각 단계에 대해서 자세히 설명하기로 한다.

2.1 물체 이미지 수집

본 연구에서는 [Fig. 1]과 같이 로봇의 말단에 카메라를 장착하여, 물체의 RGB 이미지를 얻는다. 이때, [Fig. 2]와 같이 로봇의 말단을 X축 및 Y축에 대해서 15°씩 기울이는 자세 (A), (B), (C)와 자세 ①, ②, ③의 조합으로 만들 수 있는 9가지 자세에서 물체의 이미지를 얻는다.



[Fig. 1] Environment for data collection



[Fig. 2] End-effector poses for taking object images

2.2 배경 제거

개체분할 알고리즘의 학습 시에 2종류의 레이블을 준비한다. 한 세트는 각 학습 대상 물체에 대해 종류와 위치를 레이블링한 것이고, 다른 세트는 모든 학습 대상 물체를 하나의 종류와 각각의 위치로 레이블링한 것이다. 이와 같이 준비한 후, 두 레이블을 함께 사용하여 학습하면, 처음 보는 물체도 인식할 수 있도록 학습한 개체분할 알고리즘을 이용하여 물체가 존재하는 영역을 나타내는 마스크를 얻을 수 있으며, 이를 이용하여 처음 보는 물체 이미지라도 배경을 제거할 수 있다. 이에 대해서는 3장에서 구체적으로 설명한다. 이와 같은 방법을 기반으로, [Fig. 3]과 같이 RGB 이미지와 마스크를 이용하여, 마스크가 있는 영역만을 남기고 나머지는 삭제함으로써, 배경을 제거한 물체 이미지를 얻을 수 있으며, [Fig. 4]에서 배경이 제거된 물체 이미지의 예시를 볼 수 있다.

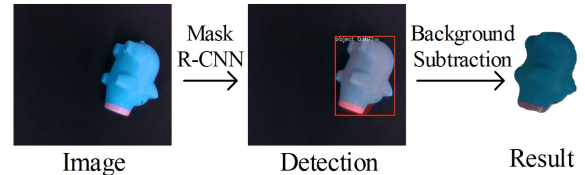
2.3 이미지 합성

앞서 획득한 물체 이미지에는 배경이 없으므로, 이미지를 합성하기 위해 배경 이미지가 필요하다. 배경 이미지를 [Fig. 5]와 같이 최대한 다양한 환경에서 수집함으로써, 생성될 합성 데이터셋에 큰 다양성을 부여한다. 이미지 합성 과정은 [Fig. 6]과 같이, 물체 이미지에 회전 및 크기 변경 과정을 거친 후에 배경과 합성한다. 이때, 물체 이미지의 합성 위치와 물체 이미지의 크기를 이용하여 점 A와 점 B의 좌표인 (x_{min}, y_{min}) , (x_{max}, y_{max}) 를 알 수 있는데, 이는 자동 레이블링 단계에서 이미지의 레이블링을 위해 사용된다. [Fig. 6]에서는 하나의 배경 이미지에 하나의 물체 이미지만 합성하였으나, 실제로는 여러 개의 물체를 같은 방식으로 합성하여 데이터셋의 다양성 확보에도움을 줄 수 있다.

2.4 자동 레이블링

일반적으로 물체인식과 개체분할에 사용되는 데이터셋의 레이블은 구조가 다르므로, 어느 것을 사용할지에 따라 자동 레이블링의 진행 방식도 다르게 된다. 먼저 물체인식의 경우에는, 경계박스의 위치를 표시하는 파라미터 4개와 경계박스 내부의 물체의 종류가 필요하다. 이미지 합성 과정에서 [Fig. 6]의 점 A와 B의 좌표를 알 수 있으므로, 이 두 점으로 경계박스를 표현할 수 있다. 또한, 합성한 물체의 종류는 이미지 합성 시에 선택할 수 있으므로, 자동 레이블링이 가능하다. 개체분할의 경우에는, 물체의 형상을 나타내는 마스크가 필요하다. RGB 이미지의 합성은 [Fig. 6]와 같은 방식으로 수행되고, 이때의 물체의 위치는 점 A와 B를 통해 알 수 있다. 마스크는 이

진 이미지(binary image)이며, 이때의 물체의 마스크는 물체 이미지의 배경을 0, 물체를 1로 설정함으로써 얻을 수 있으며, [Fig. 6]와 동일한 과정을 통해 마스크를 얻을 수 있다. 다만, 이미지 생성 시에는 RGB 이미지이지만, 마스크 생성 시에는 이진 이미지라는 차이점이 있다. [Fig. 7]은 물체마다 레이블링된 마스크의 예시이다.



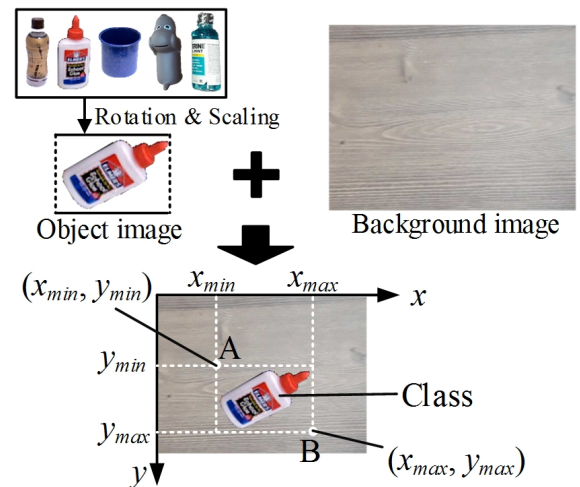
[Fig. 3] Background subtraction using Mask R-CNN



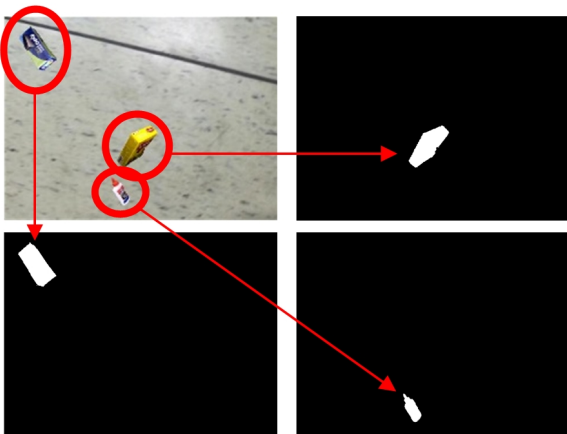
[Fig. 4] Examples of object images without background



[Fig. 5] Examples of background images



[Fig. 6] Synthesizing new images



[Fig. 7] Generated dataset of instance segmentation

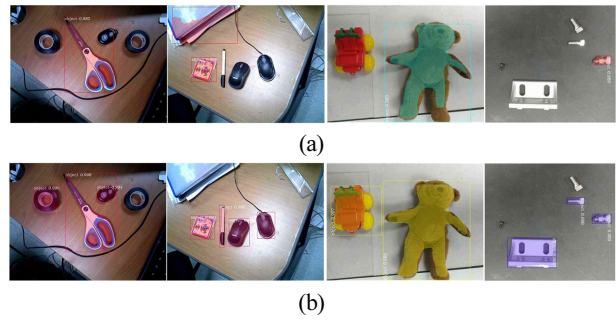
3. 배경 제거를 위한 Mask R-CNN의 학습

3.1 Mask R-CNN^[8]

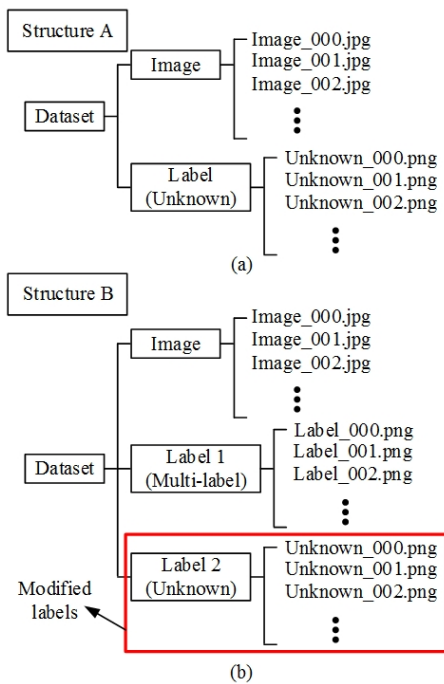
Mask R-CNN은 개체분할 알고리즘의 일종으로, [Fig. 8]과 같이 물체인식을 수행할 뿐만 아니라 개체의 마스크를 얻을 수 있는 딥러닝 기반 알고리즘이다. 비슷한 역할을 할 수 있는 의미분할(semantic segmentation)에 비해, 같은 종류의 물체가 겹쳐 있어도, 서로 다른 개체로 분할 가능한 장점이 있다.

3.2 배경 제거를 위한 데이터셋 구조

배경 제거를 위한 Mask R-CNN은 처음 보는 물체라도 마스크를 추정할 수 있어야 하는데, 데이터셋의 구조를 변경하여 [Fig. 9]와 같이 기존보다 더 좋은 성능을 보일 수 있도록 하였다. 개체분할 알고리즘의 데이터셋은 이미지, 이미지 상의 물체의 종류 및 마스크 등으로 구성된다. 따라서 처음 보는 물체를 인식하기 위한 데이터를 구성한다면, 일반적으로 데이터셋은 이미지와 [Fig. 10(a)]과 같이 Image와 Label로 구성되지만, 본 연구에서는 [Fig. 10(b)]과 같이 데이터셋을 구성하였다. 이때, Label 2는 Label 1의 모든 물체의 종류를 unknown이라는 종류 하나로 변경한 새로운 레이블이다.



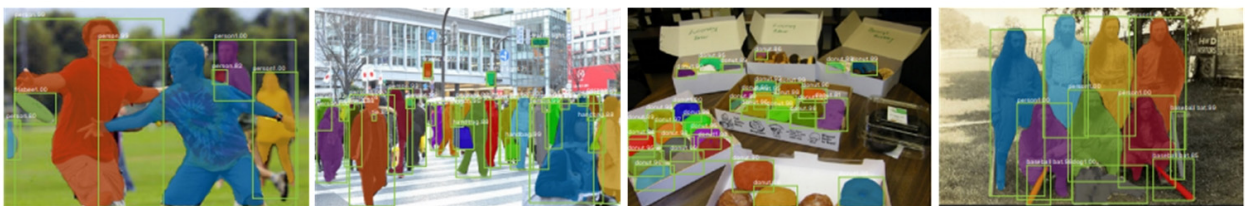
[Fig. 9] Comparison of performance to recognize unknown objects trained by dataset of (a) structure A and (b) B in [Fig.10]



[Fig. 10] Modified dataset structure:
(a) General dataset structure and (b) modified dataset structure

3.3 학습 세부 사항

[Fig. 11]과 같이 Amazon Robotics Challenge^[9]에 사용되었던 물건들과 그 외 몇 가지 물체를 더 추가하여 unknown 클래스까지 포함하여 총 45종의 물체를 학습하였다. 기 학습 모델



[Fig. 8] Examples of Mask R-CNN

[Table 2] Results of object detection and instance segmentation using Mask R-CNN (unit: %)

Real dataset	Object	Water_bottle	Scissors	Bottle_opener	Duct_tape	Baseball	Dog_bowl	Glue
	IoU	92.0	90.4	90.8	93.8	89.0	95.8	89.5
	AP	20.0	85.0	60.0	61.7	56.0	92.5	90.0
	Object	Utility_brush	Knit_glove	Remote_controller	Toothpaste	Brush	Water_gun	Cup
	IoU	86.6	91.7	91.6	88.9	85.1	87.7	88.5
	AP	75.2	90.0	68.3	72.5	13.2	70.0	40.0
	Object	Glue_sticks	Toothbrush	Board_eraser	Wine_glass	Cosmetics	Mean	
	IoU	86.5	90.1	90.4	79.9	88.5	89.3	
	AP	85.0	74.5	88.4	93.4	70.0	68.7	
Synthetic dataset	Object	Water_bottle	Scissors	Bottle_opener	Duct_tape	Baseball	Dog_bowl	Glue
	IoU	85.5	92.5	89.7	90.9	88.4	93.3	77.2
	AP	30.0	70.0	50.0	62.5	95.0	85.0	10.0
	Object	Utility_brush	Knit_glove	Remote_controller	Toothpaste	Brush	Water_gun	Cup
	IoU	82.7	85.7	90.9	84.0	92.7	86.0	90.4
	AP	34.6	82.5	95.0	95.0	63.6	63.0	80.0
	Object	Glue_sticks	Toothbrush	Board_eraser	Wine_glass	Cosmetics	Mean	
	IoU	65.0	78.5	79.8	57.4	76.4	83.5	
	AP	65.4	63.4	77.7	33.3	50.0	63.5	

[Table 3] Results of object detection using YOLO v3 (unit: %)

Real dataset	Object	Water_bottle	Scissors	Bottle_opener	Duct_tape	Baseball	Dog_bowl	Glue
	AP	90.9	96.0	97.0	79.0	66.4	84.4	98.4
	Object	Utility_brush	Knit_glove	Remote_controller	Toothpaste	Brush	Water_gun	Cup
	AP	86.2	70.9	55.9	67.2	88.3	90.9	96.5
	Object	Glue_sticks	Toothbrush	Board_eraser	Wine_glass	Cosmetics	Mean	
AP	90.7	5.4	90.4	97.2	93.6	81.3		
Synthetic dataset	Object	Water_bottle	Scissors	Bottle_opener	Duct_tape	Baseball	Dog_bowl	Glue
	AP	55.0	95.8	52.2	97.9	98.5	97.6	85.6
	Object	Utility_brush	Knit_glove	Remote_controller	Toothpaste	Brush	Water_gun	Cup
	AP	68.1	98.0	95.9	94.7	45.1	98.9	94.6
	Object	Glue_sticks	Toothbrush	Board_eraser	Wine_glass	Cosmetics	Mean	
AP	78.8	68.1	72.8	26.25	89.1	76.7		

[Table 2]의 평균 IoU를 보면, 5.8% 정도의 차이가 있다. 이를 통해, 물체인식 외에도 개체분할에 대해서도 합성 데이터셋을 이용한 성능이 실제 데이터셋을 이용하였을 때와 유사함을 알 수 있다.

4.2.2 YOLO v3를 이용한 물체 인식 실험 결과

제한한 방법을 통해 얻은 합성 데이터셋이 타 알고리즘에도 사용 가능하다는 점을 보이기 위해 YOLO v3에 앞선 실험과 동일한 데이터셋을 사용하여 학습한 결과를 [Table 3]에서 볼 수 있다. YOLO v3는 개체분할 알고리즘이 아닌 물체인식

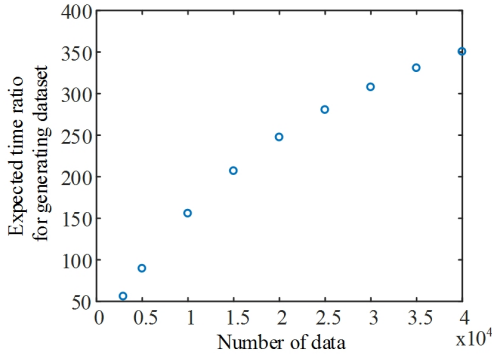
알고리즘이므로, IOU가 0.5 이상인 영역만을 대상으로 mAP를 계산하였다. [Table 3]에서 YOLO v3에서도 약 4.6%의 mAP 차이가 있는 것을 알 수 있다.

4.2.3 데이터셋 수집 효율

실제 데이터셋과 합성 데이터셋의 레이블링 비용을 비교하기 위하여 동일 양의 데이터 수집 시에 소요되는 시간을 측정하였다. [Table 4]는 각각 3,053장의 실제 데이터셋과 합성 데이터셋의 수집 시에 소요되는 시간이며, 총 소요시간을 비교하면, 합성 데이터셋이 실제 데이터셋 보다 56배 가량 빠르게

[Table 4] Time for obtaining real and synthetic datasets for 3,053 images

Real	Taking images	Labeling	Total time
Time (s)	18,318	54,954	73,272
Synthetic	Taking object images	Image synthesis & labeling	Total time
Time (s)	1,186	122.1	1,308



[Fig. 13] Ratio with respect to the number of data

생성할 수 있다는 것을 알 수 있다.

실제 데이터 수집에서의 레이블링 시간은 이미지 양에 따라 정비례 관계로 증가한다. 그러나 합성 데이터셋은 데이터를 많이 생성하더라도, 물체의 종류가 늘어나지 않으면 물체 이미지 수집이 불필요하므로, PC를 통해 자동으로 수행하는 이미지 합성 및 레이블링 시간만 늘어난다. 이 차이로 인해 데이터를 많이 만들수록 합성 데이터셋의 효율이 증가하는 효과가 나타난다. 이를 식으로 나타내면 다음과 같다.

$$\text{Expected time ratio for generating dataset} = \frac{cx}{ax+b} \quad (5)$$

여기서 a 는 한 장의 합성 이미지 생성 및 레이블링 작업에 필요한 시간이며, b 는 합성 데이터의 물체 이미지를 얻는 데 걸리는 시간으로 물체의 종류에 비례한다. c 는 실제 이미지 한 장을 찍고 레이블링 작업까지 수행하는데 필요한 시간이며, x 는 총 데이터의 수이다. [Table 3]에서의 데이터를 식 (5)에 적용하여 그래프로 나타내면 [Fig. 13]과 같다. 이때, 데이터의 수가 40,000개라고 하면 효율이 약 350배 가량 차이가 나는 것을 볼 수 있다.

4.3 실험 결과 해석

물체인식 및 개체분할 알고리즘을 합성 데이터셋을 이용하

여 학습하면, 실제 데이터셋으로 학습하는 것과 5% 정도의 차이가 나타난다. 그러나 이를 이용하면, 데이터 수집 효율은 약 3,000장의 이미지로 구성된 데이터셋을 생성할 때 약 56배 빠르게 수집할 수 있으며, 데이터 양이 증가할수록 그 효율이 증가하는 것을 볼 수 있다. 즉, 데이터를 많이 수집할수록 합성 데이터셋을 이용하여 학습하는 것이 레이블링 비용을 줄일 수 있다. 또한, 제안한 방법을 Mask R-CNN 외에도 물체인식의 대표적인 알고리즘인 YOLO v3에 적용하여도 비슷한 결과를 얻을 수 있다는 점을 보임으로써, 제안한 방법이 다양한 알고리즘에 적용 가능함을 증명하였다.

추가적으로 수치화하기 어렵지만, 사람의 노동력도 줄일 수 있다. 실제 데이터셋의 이미지를 수집하기 위해 지속적으로 장소를 바꾸며 사진을 찍어야 하며, 일반적으로 사진을 찍기 쉬운 디지털 카메라나 휴대폰 카메라를 이용하여 이미지를 수집한다면, 인공신경망이 처리하기에는 너무 큰 사이즈의 이미지이므로 이미지의 크기를 변경하는 과정이 필요하다. 그러나 제안한 방법을 이용하면, 로봇 말단의 카메라를 통해 이미지를 수집하고, 이는 이미 사용하기 적당한 크기의 이미지이므로 이미지의 크기 변경도 필요 없으며, 단순히 물체만 교체 해주며 클릭 몇 번으로 대량의 이미지를 수집하는 것이 가능하다. 이와 같이 사람의 노동력을 줄이는 것도 레이블링 비용 절감에 기여한다.

5. 결론

본 연구에서는 데이터셋의 구조를 조작하여 학습한 Mask R-CNN을 이용하여, 이미지의 생성 및 레이블링 작업을 자동화함으로써, 물체인식 및 개체분할 알고리즘에 사용할 수 있는 데이터셋 생성을 위한 레이블링 비용을 대폭 줄일 수 있는 방법을 제안하였다. 실험을 통해, 물체인식 알고리즘과 개체분할 알고리즘을 실제 데이터셋으로 학습하였을 때보다 제안한 방법으로 생성된 데이터셋으로 학습하였을 때, AP 기준으로 약 5%의 차이가 나타났다. 그러나 데이터셋 생성 효율은 약 3,000장의 이미지를 생성할 때 약 56배 높았고, 생성하는 데이터셋의 양이 늘어날수록 효율이 높아짐을 보였다. 또한, 개체분할 알고리즘인 Mask R-CNN 외에도 물체인식 알고리즘인 YOLO v3에 제안한 방법이 적용 가능함을 실험을 통해 보임으로써, 제안한 방법이 다양한 알고리즘에 적용 가능함을 보였다. 따라서 제안한 알고리즘을 사용하여, 물류 현장과 같이 다양한 물체를 다루는 분야에서 물체인식 및 개체분할에 필요한 데이터 생성 비용을 대폭 줄임으로써, 물체인식과 개체분할의 활용도를 높일 수 있다.

사 사

이 논문은 2018년도 과학기술정보통신부의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임. (No. 2018-0-00622)

References

- [1] S. Levine, P. Pastor, A. Krizhevsky, J. Ibarz, and D. Quillen, "Learning Hand-Eye Coordination for Robotic Grasping with Large-Scale Data Collection," *The International Journal of Robotics Research (IJRR)*, vol. 37, no. 4-5, pp. 421-436, 2018.
- [2] J. Mahler, J. Liang, S. Niyaz, M. Laskey, R. Doan, X. Liu, J. A. Ojea, and K. Goldberg, "Dex-net 2.0: Deep learning to plan robust grasps with synthetic point clouds and analytic grasp metrics," *arXiv:1703.09312 [cs.RO]*, 2017.
- [3] A. Kuznetsova, H. Rom, N. Alldrin, J. Uijlings, I. Krasin, J. Pont-Tuset, S. Kamali, S. Popov, M. Mallocci, T. Duerig, and V. Ferrari, "The Open Images Dataset V4: Unified image classification, object detection, and visual relationship detection at scale," *arXiv:1811.00982 [cs.CV]*, 2018.
- [4] M. Oquab, L. Bottou, I. Laptev, and J. Sivic, "Is object localization for free? -Weakly-supervised learning with convolutional neural networks," *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA, pp. 685-694, 2015.
- [5] A. Yao, J. Gall, C. Leistner, and L. V. Gool, "Interactive object detection," *2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Providence, RI, USA, pp. 3242-3249, 2012.
- [6] D. P. Papadopoulos, J. R. R. Uijlings, F. Keller, and V. Ferrari, "We don't need no bounding-boxes: Training object class detectors using only human verification," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, pp. 854-863, 2016.
- [7] Y. Wu, Y. Wu, G. Gkioxari, and Y. Tian, "Building generalizable agents with a realistic and rich 3D environment," *arXiv:1801.02209 [cs.LG]*, 2018.
- [8] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," *2017 IEEE International Conference on Computer Vision (ICCV)*, Venice, Italy, pp. 2980-2988, 2017.
- [9] J. Leitner, A. W. Tow, N. Sunderhauf, J. E. Dean, J. W. Durham, M. Cooper, M. Eich, C. Lehnert, R. Mangels, C. McCool, P. T. Kujala, L. Nicholson, T. Pham, J. Sergeant, L. Wu, F. Zhang, B. Upcroft, and P. Corke, "The ACRV picking benchmark: A robotic shelf picking benchmark to foster reproducible research," *2017 IEEE International Conference on Robotics and Automation (ICRA)*, Singapore, Singapore, pp. 4705-4712, 2017.
- [10] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft COCO: Common Objects in Context," *European Conference on Computer Vision (ECCV)*, Zurich, Switzerland, pp.740-755, 2014.
- [11] M. Everingham, S. M. A. Eslami, L. V. Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The Pascal Visual Object Classes Challenge: A Retrospective," *International Journal of Computer Vision (IJCV)*, vol. 111, no. 1, pp.98-116, 2015.



조 현 준

2016 고려대학교 기계공학과(학사)
2016~현재 고려대학교 기계공학과(석박사 통합과정)

관심분야: 로봇 제어, 딥 러닝



김 다 윷

2018 한국항공대학교 항공우주학과(학사)
2018~현재 고려대학교 기계공학과(석사과정)

관심분야: 로봇 제어, 딥 러닝



송 재 복

1983 서울대학교 기계공학과(공학사)
1985 서울대학교 기계공학과(공학석사)
1992 Mechanical Engineering,
Massachusetts Institute of
Technology, Cambridge(공학박사)
1993~현재 고려대학교 기계공학부 정교수

관심분야: 다양한 머니플레이터의 개발, 이동로봇의 주행, 지능로봇 시스템의 설계 및 제어

