

YONI : 객체 탐지를 위한 개선된 합동 훈련

김 호¹ · 이수원² · 서영건^{2*}¹경상대학교 컴퓨터과학과 학부생²경상대학교 컴퓨터과학과 교수

You Only Need Images : An Improved Joint Training for Object Detection

Ho Kim¹ · Suwon Lee² · Yeong Geon Seo^{2*}¹Bachelor's Course, Dept. of Computer Science, Gyeongsang Nat'l University, Gyeongnam 52828, Korea²Professor, Dept. of Computer Science, Gyeongsang Nat'l University, Gyeongnam 52828, Korea

[요 약]

딥러닝을 이용한 객체인식 방법은 나날이 발전하여 오늘날 매우 빠른 속도와 성능을 보이고 있다. 다만 그러한 딥러닝 모델을 충분히 학습시키려면 상당한 양의 데이터가 필요한데, 이는 개인이 소자본으로 사용하기에 부적절하다. 이전에도 데이터셋 부족 문제를 해결하기 위한 여러 가지 방법이 제안되었으나, 비용을 획기적으로 줄이거나 성능을 개선하기는 어려웠다. 이에 본 연구에서는 기존 방법의 성능 향상을 위해, 두 종류의 데이터셋을 각각 학습하는 YONI(You Only Need Images) 방법을 제안한다. 제안 방법은 먼저, 학습하려는 카테고리의 공통점 카테고리를 기존의 잘 알려진 데이터셋을 이용해 학습한다. 다음으로, 학습한 모델이 사용자가 만든 데이터셋의 box annotation 작업을 수행한다. 마지막으로, 학습하려는 카테고리를 사용자 데이터셋을 이용해 학습한다. 이 학습 방법이 기존의 방법처럼 비싼 box annotation 작업을 생략하면서도, box annotation 정보를 사용하지 않아 발생하는 성능 손실을 개선함으로써 기존의 방법보다 성능이 우수함을 보인다.

[Abstract]

The object recognition method using deep learning has been evolving day by day, showing very high speed and performance. To fully train such a deep learning model, however, a significant amount of data is required, which is inappropriate for individuals to use as small capital. In the past, various methods have been proposed to solve the dataset shortage problem, but it has been difficult to drastically reduce costs or improve performance. In this paper, we propose YONI method for learning two types of data sets to improve the performance of the existing methods. First, the method learns the common categories of the categories we want to learn using existing well-known datasets. Next, the trained model performs box annotation on the dataset created by the user. Finally, it uses the user dataset to learn the categories we want to learn. While this learning method omits expensive box annotation work like the existing method, it shows better performance than existing method by improving performance loss caused by not using box annotation information.

색인어 : 합동 훈련, 객체 탐지, 딥러닝, YONI, 박스 주석**Key word** : Joint Training, Object Detection, Deep Learning, YONI, box annotation<http://dx.doi.org/10.9728/dcs.2020.21.1.11>

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.

Received 06 November 2019; Revised 01 December 2019

Accepted 23 January 2020

***Corresponding Author; Yeong Geon Seo**

Tel: +82-55-772-1392

E-mail: young@gnu.ac.kr

1. 서론

딥러닝은 나날이 발전하고 있으며, 기존의 객체인식 방법들보다 우수한 방법들이 딥러닝을 활용해 만들어지고 있다. 오늘날, Faster R-CNN[1], YOLO[2], SSD[3], RetinaNet[4] 등 훌륭한 객체인식 방법들이 딥러닝을 통해 만들어지고 있다. YOLOv3[2]의 경우, 영상 하나에서 객체들을 추출하는데 겨우 22ms의 연산 시간이 필요하다. 그야말로 혁신이 아닐 수 없다. 다만 딥러닝 모델은 수준급의 학습을 위해선 엄청난 양의 데이터가 요구된다. 데이터가 많을수록 딥러닝 모델이 안정적으로 학습되기 때문이다. 때문에 딥러닝 모델을 사용하려면 데이터가 많을수록 좋다. 만약 학습한 모델이 공사장이나 의류기기, 독버섯 분류 등 위험한 현장에 투입되려면, 모델이 큰 실수를 벌이지 않도록 매우 안정적으로 학습시켜야 한다. 그럴수록 필요로 하는 데이터의 양이 많아진다. 이러한 점은 개인이, 소자본으로 딥러닝 모델을 사용하기 어려운 이유이다.

본 논문에서는 객체인식 분야의 데이터셋 부족 문제를 분석하고, 그 중 비용이 많이 드는 작업을 대체하기 위해 연구하였다. 여기서 데이터셋 annotation 작업이 많은 비용이 소요됨을 보이고, 딥러닝 모델을 이용해 annotation 작업을 대체하는 YONI(You Only Need Images)라는 새로운 객체인식 모델의 학습방법을 제안한다. 또한 YONI 방법의 사용은 일부 데이터의 box annotation 정보를 생략 가능한 합동 훈련(Joint Training)에 비해 성능이 개선됨을 보인다. 논문의 구성은 2장에서 기존의 비용 절감 방법을 살펴보고, 3장에서 새로운 학습 방법인 YONI 방법에 대해 서술하며, 4장에서 제안한 방법의 유효성을 검증하겠다.

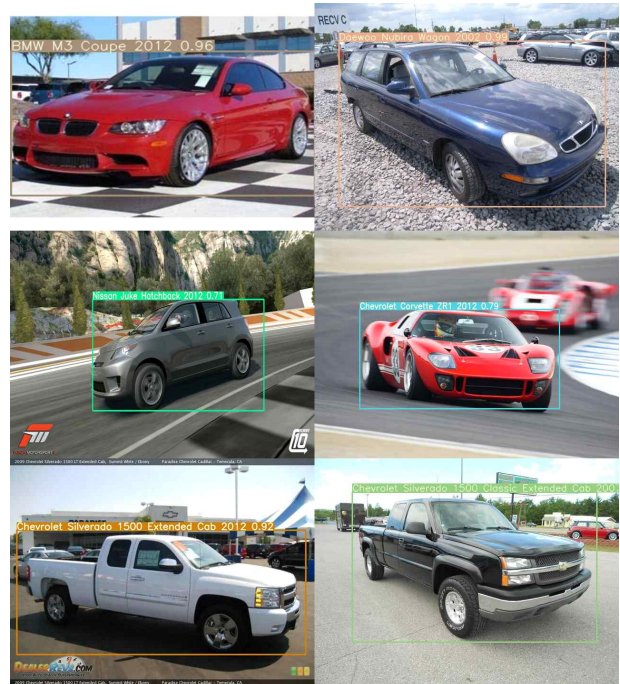


그림 1. YONI를 사용하여 영상 속 자동차 탐지와 분류 예
 Fig. 1. Example of detecting and classifying cars in video using YONI

II. 기존의 연구 방법

일반적으로 기계학습을 위한 객체인식 데이터셋을 구축할 때 주로 영상 수집, 영상 annotation, 데이터셋 분할 등의 단계를 거친다[5][6]. 이 중 비용이 가장 많이 드는 단계는 영상 annotation으로, 주요 데이터셋의 구축 비용 중 가장 큰 부분을 차지한다. 때문에 이 단계에서 발생하는 비용을 줄이기 위한 여러 가지 방법이 활용되고 있다.

2-1 크라우드소싱(Crowdsourcing)

AMT(Amazon's Mechanical Turk) 등의 서비스를 통해 불특정 다수가 annotation에 참여하는 방법이다. AMT는 많은 인력을 필요로 하는 작업을 의뢰하면 불특정 다수가 참여하고, 이에 따른 보수를 제공하는 온라인 플랫폼이다. AMT의 경우, 참여 인원이 많을수록 비용은 증가하지만 결과물 품질이 좋아지고, 경우에 따라 전문가들의 작업 결과보다 우수하기도 하다[7]. 현재 널리 사용되는 데이터셋들이 이 방법을 통해 비용을 절감하고 있다[6]. 그럼에도 [6]의 경우 총 작업 시간이 70k 이상이 걸려 적은 자본으로는 시도하기 부적절한 측면이 있다.

2-2 과도한 클릭(Extreme Clicking)

[8]의 경우, 과도한 클릭[9]이라는 방법을 통해 비용을 절감했다. 이 방법은 영상 속 객체의 bbox(bounding box)를 annotation 하기 위해, 숙련된 작업자들이 각 객체의 top, bottom, left, right의 네 점을 찍는 방법이다. 여기서 네 점은 bbox 하나를 의미한다. [7]의 데이터 셋에서는 bbox 하나에 보통 42초가 필요했지만, 위 방법을 이용한 경우 7.4초가 필요하니 작업 속도가 꽤 빨라졌고, 절감된 시간만큼 비용도 절감할 수 있었다. 하지만 이 방법을 사용하기 위해서는 먼저 작업자들을 훈련시켜야 하고, 때문에 개인 혹은 소규모 인원이 모여 대규모 데이터 셋을 구축하기엔 부적절하다. 그림 2는 과도한 클릭 작업의 흐름도를 나타내며, 초록 영역(중앙 점선)에서 작업자들이 일정 수준의 훈련 성과를 보이면 과란 영역(우측 점선)에서 실제 annotation 작업을 수행한다.

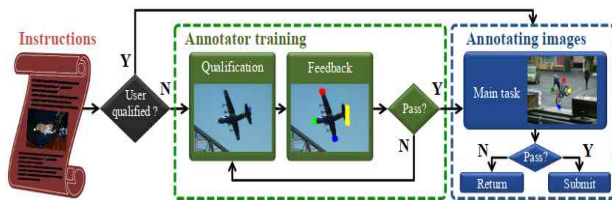


그림 2. 과도한 클릭의 작업 흐름도
Fig. 2. Flow chart of Extreme clicking

2-3 합동 훈련(Joint Training)

yolo9000의 경우, 객체인식 데이터 셋과 분류 데이터 셋을 학습에 동시에 활용한다[10]. 이를 위해 yolo9000은 WordNet을 응용하였다. WordNet은 일종의 말뭉치 데이터 베이스로, 트리가 아닌 방향 그래프의 구조를 취한다[11]. WordNet의 뿌리 노드를 "Physical object"라고 정의하고, 아래로 내려갈수록 개념이 구체화되는 트리 구조의 WordTree를 사용하였다. yolo9000에서는 탐지 데이터 셋인 COCO[12]의 WordTree와 분류 데이터 셋의 WordTree를 결합하고, 두 데이터 셋의 영상을 학습에 사용하였다. 이를 위해, 각 카테고리의 하위 카테고리들에 softmax 함수를 취해 조건부 확률의 꼴이 되도록 하였다. 따라서 각 카테고리의 확률 값은 해당 카테고리의 조건부 확률에 상위 카테고리의 조건부 확률을 모두 곱한 것과 같다. 여기서 뿌리 노드인 "Physical object"의 확률 값은 1이다. 다만 분류 데이터 셋의 영상을 학습할 때는 분류에 해당하는 오차에만 역전파 알고리즘을 수행한다.

이 방법을 채택한 yolo9000 모델은 ImageNet[13]을 분류 데이터 셋으로, COCO를 탐지 데이터 셋으로 학습하여, 총 200 종류 객체들을 19.7mAP(mean Average Precision)[PASCAL VOC=5]의 성능으로 추론할 수 있었다. 그 중 COCO 데이터 셋에도 존재하는 44 종류를 제외하

면 16.0mAP의 성능을 보였다. 위 방법을 사용하지 않고 COCO 데이터 셋만 학습한 경우는 총 80 종류 객체들을 21.6 mAP의 성능으로 추론하였다. 식 1은 Norfolk terrier 카테고리의 확률 값을 구하는 예시를 보이고 있다.

$$\begin{aligned}
 Pr(Norfolkterrier) &= Pr(Norfolkterrier|terrier) \\
 &* Pr(terrier|huntingdog) \\
 &* \dots * \\
 &* Pr(mammal|animal) \\
 &* Pr(animal|physicalobject)
 \end{aligned}
 \tag{1}$$

기존의 비용 절감 방법들은 여전히 높은 비용이 요구되거나 성능을 다소 희생시켜야 한다. 특히 과도한 클릭과 같이 수작업을 개선시키는 방식은 비용을 획기적으로 낮출 수 없다. yolo9000과 같은 합동 훈련 방법은 분류 데이터 셋을 병행하여 비용을 크게 낮출 수 있으나 성능이 떨어지는 단점이 있다. 이에 본 연구는 합동 훈련 방법의 성능 하락 원인을, 한 모델에 탐지 데이터 셋과 분류 데이터 셋을 동시에 학습해, 분류 데이터 셋의 누락된 box annotation 정보가 학습에 부정적인 영향을 주기 때문으로 가정하였다. 이를 입증하기 위해, 두 차례에 걸쳐 각각 탐지 데이터 셋과 분류 데이터 셋을 학습하는 YONI 방법을 제안한다.

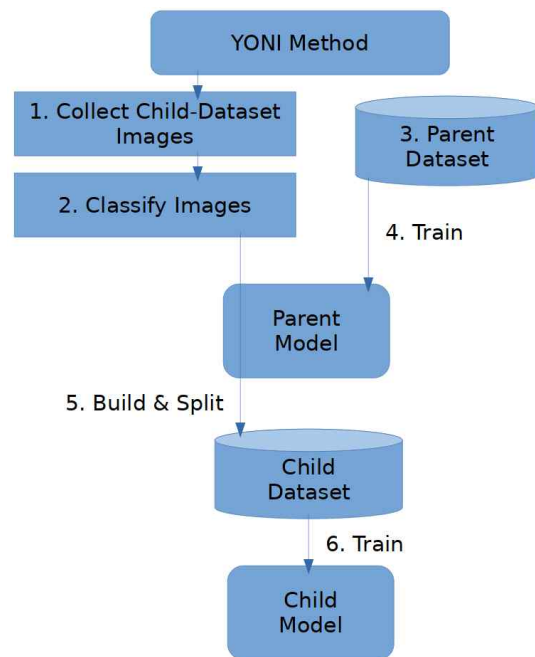


그림 3. YONI 방법의 전체적인 수행 과정
Fig. 3. Processing steps of the proposed method, YONI

III. 개선된 합동 훈련(YONI)

YONI 방법은 학습할 카테고리에 유의미한 공통점이 있다고 가정한다. 자세한 수행 단계는 다음과 같다.

3-1 차일드 데이터 셋 영상 수집

첫 번째 과정은 학습할 카테고리에 해당하는 영상을 수집한다. 예를 들어 가위바위보 손 모양을 학습시키고 싶다면, "가위", "바위", 그리고 "보"에 해당하는 영상을 수집한다. 영상은 온라인에서 수집하거나, 동영상 촬영하고 프레임 분할하여 만들 수 있다. 혹은 기존의 분류기 데이터 셋을 활용할 수도 있다. 다행인 점은, 수집한 영상 중 카테고리와 관련 없는 것들은 추후 필터링 되므로 데이터의 오염이 다소 있더라도 상관없다는 점이다 - 데이터는 많을수록 좋다. 다만, 한 영상에는 반드시 하나의 객체만 등장해야 한다. 이 단계의 결과물은 차일드 데이터 셋(child dataset)의 영상이다. 그 어떤 카테고리 정보도, box annotation 정보도 제공되지 않은 상태이다. 그림 4는 차일드 데이터 셋 영상의 예를 보이고 있으며, 카테고리 정보가 누락된 단순한 영상의 뭉치이다.



그림 4. 차일드 데이터 셋 영상의 예
Fig. 4. Examples of child data set image

3-2 차일드 데이터 셋 영상 분류

이 단계는 해당하는 카테고리 대로 영상을 분류한다. 만약 영상을 수집할 때 미리 카테고리를 지정해 놓으면 이 단계는 생략할 수 있다. 특히 동영상 촬영을 통해 영상을 수집할 때, 한 동영상에 한 카테고리만 등장시키면 모든 프레임의 카테고리가 한 카테고리로 특정되므로 보다 수월히 작업할 수 있다. 이 단계를 통해, 차일드 데이터 셋 영상은 카테고리 정보를 획득하고, 차일드 데이터 셋은 분류 데이터 셋의 특징을 가진다. 그림 5는 차일드 데이터 셋 분류의 예를 보이고 있으며, 각 영상이 상응하는 카테고리 정보를 가지므로, 분류 데이터 셋의 성질을 가진다.



"Ford Mustang Convertible 2007" "Dodge Durango SUV 2012"

그림 5. 차일드 데이터 셋 분류의 예
Fig. 5. Examples of classifying child data set

3-3 페어런트 데이터 셋 획득

제한한 방법을 사용하기 위해서 학습할 카테고리들이 공통점이 있고, 이 공통점 카테고리가 기존의 잘 알려진 데이터 셋에 존재한다고 가정한다. 예를 들어, 자동차 중 "경차", "승용차", "화물 트럭" 카테고리를 학습시키고 싶다면, 이들 카테고리의 공통점 카테고리인 "자동차"가 어떤 잘 알려진 데이터 셋에 존재한다고 가정한다. 만약 해당하는 카테고리가 없다면 공통점을 더 일반화한다 - "자동차" 위의 "탈것" 카테고리를 찾는다. 이 단계에서 페어런트 데이터 셋(parent dataset)을 획득한다. 이 때, 페어런트 데이터 셋은 영상 속의 각 객체에 대해 box annotation 정보가 카테고리 정보를 모두 포함하는 객체탐지 데이터 셋이어야 한다. 그림 6은 페어런트 데이터 셋의 예를 보이고 있으며, 영상은 https://ai.stanford.edu/~jkrause/cars/car_dataset.htm [14]에서 가져 온 것이다.



그림 6. 페어런트 데이터 셋의 예
Fig. 6. Examples of parent data set

3-4 페어런트 모델의 학습

임의의 객체인식 모델에 페어런트 데이터 셋을 학습시킨다. 이 때 사용하는 모델은 공통점 카테고리 내 모든 객체를 탐지하며, 모든 학습 단계가 끝나면 더 이상 필요하지 않다. 때문에 추론 성능보단 성능에 중점을 두어 모델을 선택한다. 이 단계에서 사용하는 모델은 페어런트 모델(parent model)이라 하겠다. 페어런트 모델은 학습이 끝나면 페어런트 카테고리의 객체들을 탐지할 수 있으나, 차일드 데이터 셋을 사용하지 않았기 때문에 객체를 차일드 카테고리로 분류할 수 없다. 그림 7은 객체의 영역과 카테고리를 추정하도록 학습하는 예를 보이고 있다.

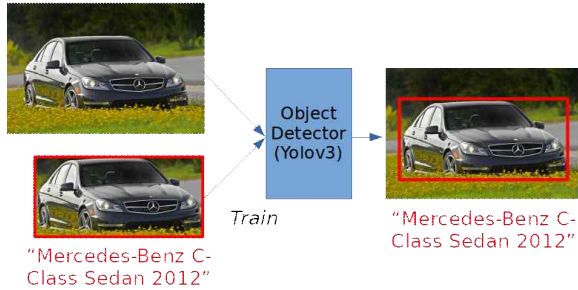


그림 7. 객체의 영역과 카테고리를 추정하도록 학습 하는 예
Fig. 7. Example of training to estimate the category and region of an object

3-5 차일드 데이터 셋의 구성 및 분리

페어런트 모델을 이용해 차일드 데이터 셋 영상에서 객체를 찾는다. 찾은 객체는 차일드 데이터 셋 예box annotation 정보로 포함된다. 이때 한 영상에서 두 개 이상의 객체가 검출되거나, 혹은 하나도 검출되지 않는다면 그 영상은 차일드 데이터 셋에서 제외한다 - 반드시 하나의 객체만 검출되어야 한다. 데이터 셋을 만들 때 반드시 한 객체만 등장하기 때문 이어서, 이들을 걸러내어 데이터 셋을 오염시키지 않는다. 이 과정을 통과한 모든 영상은 객체의 box annotation 정보와 차일드 카테고리 정보가 추가되고, 차일드 데이터 셋은 객체 인식 데이터 셋의 특징을 가진다. 차일드 데이터 셋 영상을 모두 통과시킨 후, 훈련용 데이터 셋과 평가용 데이터 셋을 분리한다. 여기서는 훈련용과 평가용의4:1의 비율로 나눈다. 그림 8은 차일드 데이터 셋 영상의 객체 영역을 페어런트 모델을 통해 추정한 예를 보이고 있으며, 각 영상이 상응하는 객체 영역과 카테고리 정보를 가지므로, 객체 인식 데이터 셋의 성질을 가진다.

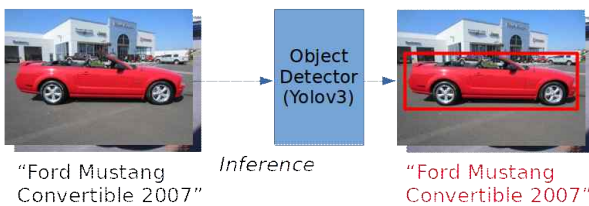


그림 8. 차일드 데이터 셋 영상의 객체 영역을 페어런트 모델을 통해 추정한 예
Fig. 8. Example of estimating the object region of a child data set image using a parent model

3-6 차일드 모델의 학습

임의의 객체 인식 모델에 차일드 데이터 셋을 학습시킨다. 이 때 사용하는 모델은 성능 평가 등에 최종적으로 사용할 모델이므로 모델의 사용 환경을 고려하며 선택한다. 이 단계에서 사용하는 모델은 하위 모델(child model)이라 하겠다. 최

종 결과물인 하위 모델은 학습이 끝나면 하위 카테고리의 객체들을 탐지할 수 있다. 그림 9는 생성한 객체 영역 정보를 포함하여 학습하는 예를 보이고 있다.

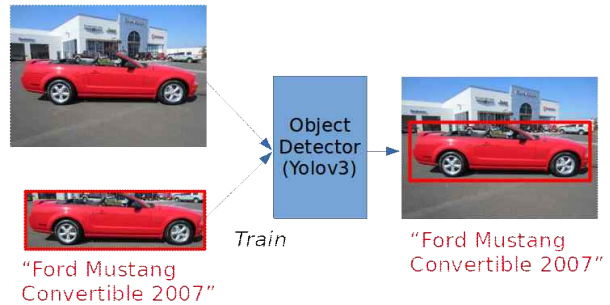


그림 9. 생성한 객체 영역 정보를 포함하여 학습하는 예
Fig. 9. Example of training by including the created object region information

IV. 실험 및 평가

제안된 방법을 평가하기 위해, 데이터 셋의 box annotation 정보를 모두 사용한 경우 (이하 Ground Truth) 와, 합동 훈련 방법을 사용한 경우, 제안한 방법을 사용한 경우의 평가 mAP 점수를 구했다. 학습에 사용할 데이터 셋으로 Cars Dataset[14]을 채택했다. 실험에는 yolov3 모델을 사용하며, 해당 모델의 소스 출처는 <https://github.com/ultralytics/yolov3> 이다. 실험에 사용된 컴퓨팅 환경은 다음과 같다.

- OS: ArchLinux 2019.09.01.
- CPU: Intel Core i7-8700K CPU @ 3.70GHz
- VGA: Nvidia GTX 1080 Ti
- Memory: 64GB

Cars Dataset 은 총 196 개의 차종과 8144 장의 훈련 데이터, 8041 장의 평가 데이터가 수록되어 있다. 훈련용과 평가용 모두 차량의 box annotation 정보가 주어져 있다. 다만 평가 데이터 셋은 훈련용과 달리 차량의 카테고리 정보가 누락된 상태이다. Cars Dataset에는 평가 데이터 셋의 카테고리가 없으므로, 이를 공통점 카테고리를 학습하는데 사용하기 위해 페어런트 데이터 셋으로 사용한다. 따라서 제안한 방법을 사용하는 실험군은 Cars Dataset을 다음과 같이 분할하였다.

- 페어런트 학습 데이터 셋: 원본 평가 데이터 셋의 80%
 - 페어런트 평가 데이터 셋: 원본 평가 데이터 셋의 20%
 - 차일드 학습 데이터 셋: 원본 학습 데이터 셋의 80%. 단, box annotation 정보는 버림.
 - 차일드 평가 데이터 셋: 원본 학습 데이터 셋의 20%
- 합동 훈련 방법을 사용한 경우, Cars Dataset을 다음과 같이 분할하였다.

- 학습 데이터 셋 중 객체 인식 데이터 셋: 원본 평가 데이터셋의 100%
- 학습 데이터 셋 중 분류 데이터 셋: 원본 학습 데이터 셋의 80%. 단, box annotation 정보는 버림.
- 평가 데이터 셋: 원본 학습 데이터 셋의 20%

Ground Truth의 경우, Cars Dataset을 다음과 같이 분할하였다.

- 학습 데이터 셋: 원본 학습 데이터 셋의 80%
- 평가 데이터 셋: 원본 학습 데이터 셋의 20%

실험 결과는 표 1과 표 2와 같다. 입력 해상도에 관계없이, YONI 방법을 사용한 경우가 합동 훈련 방법을 사용한 경우보다 mAP 점수가 높았다. 입력 해상도가 320x320 일 때, YONI 방법을 사용한 경우는 Ground Truth에 비해 약 15.7~16.2%의 점수 차가 있는 반면, 합동 훈련 방법을 사용한 경우는 약 24.3%의 점수 차가 있었다. 입력 해상도가 416x416 일 때, YONI 방법을 사용한 경우는 Ground Truth에 비해 약 4.2%의 점수 차가 있는 반면, 합동 훈련 방법을 사용한 경우는 약 15.6%의 점수 차가 있었다.

추론 속도에서는, 합동 훈련 방법이 YONI 방법보다 약 0.8%~1.8%의 우위를 보였다. Ground Truth와 비교하여 YONI 방법의 추론 속도는 입력 해상도가 320x320 일 때, 거의 일치했고, 입력 해상도가 416x416일 때 0.8%의 우위를 보였다. 결국 모든 해상도에서 YONI 방법의 mAP 점수가 합동 훈련 방법의 경우보다 높다. 즉, YONI 방법의 성능 손실이 합동 훈련 방법의 경우보다 적다. 따라서, 제안한 방법은 학습 데이터로 제작 비용이 싸지만 box annotation 정보가 없는, 분류 데이터 셋을 포함하여 발생하는 성능 하락 문제를 발견하고 이를 개선하였다는 결론을 내렸다. 그러나, YONI 방법은 합동 훈련 방법과 마찬가지로, box annotation 정보를 모두 사용하여 학습한 결과보다는 낮은 성능을 보였다. 이는 페어런트 모델이 추론한 객체 영역이 잘못된 경우가 있지만, 차일드 모델은 그 결과를 신뢰하고 학습하기 때문이다.

그림 11은 제안 방법으로 실행한 차종 인식의 성공한 예를 보이고 있으며 좌우의 차는 같은 차종이다. 그림 12는 인식에 실패한 예를 보이고 있으며, 3행의 자동차와 1행의 자동차가 같은 차종으로 분류하고 있으나, 실제로는 3행과 2행이 같은 차종이다. 실패한 예를 보면 사람의 눈으로도 쉽게 분류하기 어려운 영상이다.

표 1. 자동차 차종 인식 실험 중 성능 평가 결과

Table 1. Performance evaluation results of the car model recognition experiment

Model		mAP		
Parent Model	Child Model / Control Model	yolov3 - Ground Truth	yolov3 - Joint Training	yolov3 - YONI (Ours)
yolov3 320x320	yolov3 320x320	74.1	56.1	62.1
yolov3 416x416	yolov3 320x320	74.1	56.1	62.5
yolov3 416x416	yolov3 416x416	81.2	68.6	77.8

표 2. 자동차 차종 인식 실험 중 추론 속도 평가 결과

Table 2. Inference speed evaluation results of the car model recognition experiment

Model		Inference time(ms)		
Parent Model	Child Model / Control Model	yolov3 - Ground Truth	yolov3 - Joint Training	yolov3 - YONI (Ours)
yolov3 320x320	yolov3 320x320	11.0	10.8	11.0
yolov3 416x416	yolov3 320x320	11.0	10.8	11.0
yolov3 416x416	yolov3 416x416	13.0	12.8	12.9

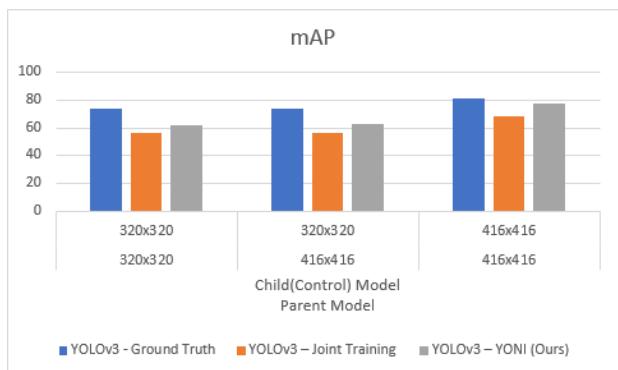


그림 10. 차종 인식 실험 결과의 시각화

Fig. 10. Visualization of the result of cars model recognition

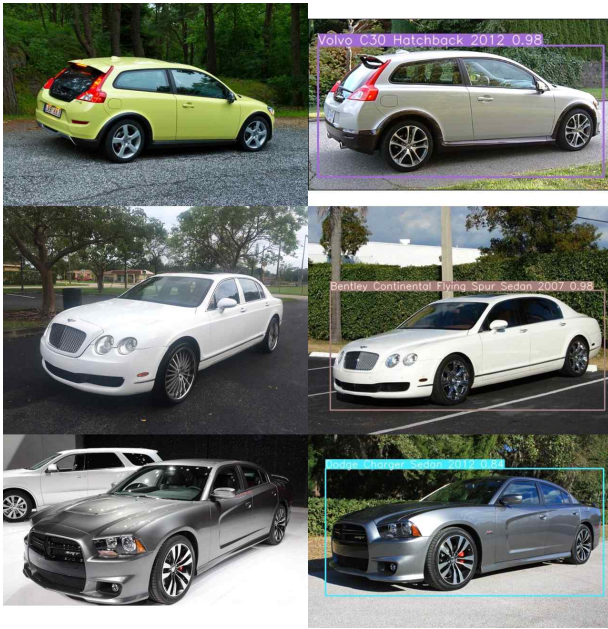


그림 11. 제안 방법으로 자동차 차종 인식의 예
Fig. 11. Examples of cars model recognition with the proposed method



그림 12. 제안 방법으로 자동차 차종 인식에 실패한 예
Fig. 12. Examples of cars model recognition with the proposed method incorrectly



그림 13. 페어런트 모델이 차일드 데이터 셋 영상에서 객체 영역을 잘 못 추론한 예
Fig. 13. Examples of parent model inferring object region from child data set image incorrectly

V. 결론

본 연구에서는 객체인식 분야의 데이터 셋 부족 문제를 해결하기 위해 YONI 방법을 제안하였다. YONI는 잘 알려진 데이터 셋을 활용해 공통점 카테고리를 학습하는 페어런트 모델과, 임의의 세부 카테고리를 페어런트 모델의 도움을 받아 학습하는 차일드 모델의 2개 모델을 사용함으로써 세부 카테고리의 데이터 셋 제작 비용을 절감한다. 자동차 차종 인식에서 yolov3 모델을 사용했을 때, 416x416의 입력 해상도에서 합동 훈련 방법은 약 15.6%의 점수 차를 보였으나, 제안한 방법은 그 보다 적은 4.2%의 점수 차를 보여, 제안한 방법이 box annotation 정보가 누락되어 발생하는 성능 하락 문제를 개선하였다는 결론을 내렸다. 다만 YONI 방법을 사용하기 위해선, 합동 훈련 방법의 경우와 같이 학습하려는 카테고리의 공통점 카테고리가 잘 알려진 데이터 셋에 이미 존재해야 한다. 또한, 총 2번의 학습이 필요하기에 학습에 소요되는 시간이 길다는 단점이 있다. 때문에 성능에 지장을 주지 않으며 학습 시간을 단축하는 방법이 요구된다.

감사의 글

본 연구는 문화체육관광부 및 한국콘텐츠진흥원의 2019년도 문화기술 연구개발 지원사업으로 수행되었음.
 [No. 1375026932, 사용자 참여형 문화공간 콘텐츠를 위한 AR 플랫폼 기술개발]

참고문헌

- [1] S. Ren, K. He, R. Girshick, and J. Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *arXiv preprint arXiv:1506.01497*, 2015.
- [2] Redmon and A. Farhad, “Yolov3 : An Incremental Improvement“, *arXiv preprint arXiv:1804.02767*, 2018.
- [3] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, “Ssd: Single Shot Multibox Detector”, *European conference on computer vision*, pp. 21–37. Springer, 2016.
- [4] T. Y. Lin, P. Goyal, R. Girshick, K. He, and P. Doll, “Focal Loss for Dense Object Detection”, *arXiv preprint arXiv:1708.02002*, 2017.
- [5] M. Everingham, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman, “The Pascal Visual Object Classes (voc) Challenge”, *International journal of computer vision*, Vol. 88, No. 2, pp. 303–338, 2010.
- [6] T.Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ra-manan, P. Doll and C. L. Zitnick, “ Microsoft coco: Common Objects in Context”, *European conference on computer vision*, pp. 740–755. Springer, 2014.
- [7] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, “ImageNet Large Scale Visual Recognition Challenge”, *International Journal of Computer Vision (IJCV)*, 2015.
- [8] A. Kuznetsova, H. Rom, N. Alldrin, J. Uijlings, I. Krasin, J. Pont-Tuset, S. Kamali, S. Popov, M. Mallocci, and T. Duerig, “The Open Images Dataset v4: Unified Image Classification, Object Detection, and Visual Relationship Detection at Scale”, *arXiv preprint arXiv:1811.00982*, 2018.
- [9] D. P. Papadopoulos, J. R. R. Uijlings, F. Keller, and V. Ferrari, “Extreme Clicking for Efficient Object Annotation”, *CoRR*, abs/1708.02750, 2017.
- [10] Joseph Redmon and Ali Farhadi, “YOLO9000 : Better, Faster, Stronger”, <http://pjreddie.com/yolo9000/>, Dec 2016.
- [11] WordNet : A Lexical Database for English, <http://wordnet.princeton.edu>.
- [12] COCO dataset, <http://cocodataset.org>.
- [13] ImageNet, <http://image-net.org>.
- [14] Cars Dataset, https://ai.stanford.edu/~jkrause/cars/car_dataset.html



김호(Ho Kim)

2017년~현재 : 경상대학교 컴퓨터과학과 재학

관심분야 : 머신 러닝, SLAM, 영상 인식



이수원(Suwon Lee)

2002년 : 한국과학기술원(공학석사)
2017년 : 한국과학기술원(공학박사)
2018년~현재 : 경상대학교 컴퓨터과학과 조교수

관심분야 : 증강현실, 컴퓨터비전



서영건(Yeong Geon Seo)

1987년 : 경상대학교 전산과 학사
1997년 : 숭실대학교 전산과 박사
1989년~1992년 : 삼보컴퓨터
1997년~현재 : 경상대학교 컴퓨터과학과 교수

2014년~현재 : 경상대학교 대학원 문화융복합학과 교수

2011년~현재 : 경상대학교 공학연구원 멤버

관심분야 : 의료 영상 처리, 머신 러닝, SLAM, 영상 인식, 컴퓨터 네트워크, 컴퓨팅 사고