

## YOLO 기반 열린 환경에서의 대조학습을 통한 강건한 객체 탐지 기술

이일규, 서영주

포항공과대학교

xignos3108@postech.ac.kr, yjsuh@postech.ac.kr

## Robust Open-Set Object Detection Based on YOLO Using Contrastive Learning Approach

Ilkyu Yi and Young-Joo Suh

Pohang University of Science and Technology (POSTECH)

## 요약

객체 탐지는 컴퓨터비전 분야에서 활발히 연구되는 분야로서 인공지능의 발전에 따라 성능과 정확도에서 많은 발전을 이루었다. 그러나 대부분의 객체 탐지 모델은 닫힌 환경(closed-set)을 전제로 하므로 훈련 시 학습되지 않은 클래스의 데이터가 입력되는 열린 환경(open-set)에서 심각한 성능저하 혹은 예기치 못한 결과를 야기할 수 있다. 본 논문에서는 현재 활발히 활용되는 모델인 YOLO를 기반으로, 대조학습의 접근을 통해 열린 환경에서도 강건한 객체 탐지 모델을 제안하고 성능을 평가한다.

## I. 서론

많은 인공지능 기반 분류 문제에서 주어진 보기 내의 결과를 확률로서 도출하는 것은 잘 알려진 쉬운 문제이다. 그러나 보기 내에 올바른 정답이 존재하지 않음을 판단하는 것은 전혀 다른 문제[1]로서 인공지능 모델은 이러한 경우, 높은 확률로 주어진 보기 내에서 오답을 도출하게 된다.

객체 탐지는 컴퓨터비전에서 잘 알려진 연구 분야로, 최근 눈에 띄는 성과와 발전으로 인해 자율주행을 비롯한 많은 분야에서 응용되고 있다. 많은 객체 탐지 모델은 닫힌 환경을 전제로 하는데, 이는 모델이 훈련 데이터셋과 동일한 클래스로 구성된 데이터셋에서 평가되는 것을 의미한다. 그러나 실제로는 모델이 훈련되지 않은 클래스에 속하는 객체, 즉 ‘알려지지 않은(unknown)’ 클래스의 객체를 포함하는 이미지와 마주치게 될 가능성이 높으며, 이러한 열린 환경에서 모델은 높은 확률로 훈련 데이터셋에 포함된 ‘알려진(known)’ 클래스 내에서 오답을 도출하거나 혹은 처음부터 객체를 인식하지 못하는 등 심각한 성능저하가 나타날 수 있다. 즉, 열린 환경에서 모델이 ‘알려지지 않은’ 객체를 ‘알려지지 않은’ 객체로서 인식하는 것은 매우 중요한 문제이며 이를 열린 환경에서의 객체 탐지(open-set object detection) 문제[2]라고 한다.

본 논문에서는 기저공간에서 데이터 간 거리를 학습하는 대조학습(contrastive learning)을 활용하여 열린 환경에서의 객체 탐지 문제를 해결하고자 하였으며, 기본 모델로는 가장 널리 활용되는 객체 탐지 모델인 YOLO[3]를 활용하였다.

## II. 본론

## 1. 모델 구조

본 논문은 열린 환경의 ‘알려지지 않은’ 객체는 모델에게 ‘알려진’ 객체와 비교하여 저밀도 기저영역에 분포한다는 타당한 가정에 기반하여, 기저공간에서 고밀도/저밀도 영역을 분리하여 ‘알려지지 않은’ 객체를 ‘알려지지 않은’ 클래스로서 인지하는 것을 목표로 한다. 이를 위해 대조학습에 기반

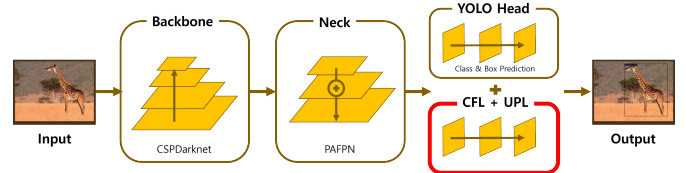


그림 1 제안된 대조학습 기반 모델

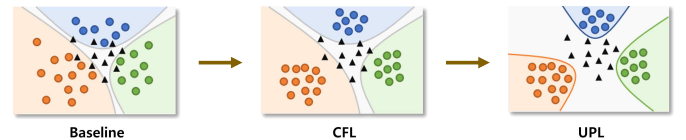


그림 2 CFL과 UPL 개념도[4]

한 추가적인 Contrastive Feature Learner(CFL)과 Unknown Probability Learner(UPL)[4] 요소를 활용한다. 제안된 대조학습 기반 모델의 전반적인 구조는 그림 1에 나타나 있고, 제시한 CFL과 UPL의 동작 과정은 그림 2와 같다.

YOLO(You Only Look Once)는 현재 가장 활발히 활용되는 객체 탐지 모델 중 하나로서 자율주행, 고장검출 등의 다양한 사례에 응용되고 있다. 본 논문에서는 YOLOv5s[3]를 기본 모델로 활용하여 YOLO 모델의 성능 향상을 위한 연구에 기여하고자 하였다.

1) Contrastive Feature Learner(CFL): CFL은 동일한 클래스의 데이터 간 거리를 줄이는 한편 서로 다른 클래스의 데이터 간 거리를 키우는 것을 목적으로 하여 대조학습을 통해 ‘알려지지 않은’ 클래스가 위치하는 저밀도 영역 크기를 넓히게 된다. CFL에 활용되는 손실함수는 식 (1)과 같이 표현된다. 이때  $c_i$ 는  $i$ 번째 이미지에 관한 클래스를 의미하고  $\tau$ 는 스케일링 하이퍼파라미터(temperature hyper-parameter),  $M(c_i)$ 는 클래스  $c_i$ 의 메모리뱅크(memory bank)를 나타내며,  $A(c_i) = M/M(c_i)$ 이다[4].

$$L_{IC} = \frac{1}{N} \sum_{i=1}^N L_{IC}(z_i)$$

$$L_{IC}(z_i) = \frac{1}{|M(c_i)|} \sum_{z_j \in M(c_i)} \log \frac{\exp(z_i \circ z_j / \tau)}{\sum_{z_k \in A(c_i)} \exp(z_i \circ z_k / \tau)} \quad (1)$$

2) Unknown Probability Learner(UPL): UPL은 예측의 불확실성을 기반으로 ‘알려지지 않은’ 클래스의 확률을 최적화하며, ‘알려진’ 클래스의 클러스터를 중심으로 저밀도 영역을 추가적으로 분할한다. 따라서 저밀도 영역의 ‘알려지지 않은’ 객체는 학습된 ‘알려지지 않은’ 객체의 확률로 쉽게 식별할 수 있다. UPL에 활용되는 손실함수는 식 (2)와 같이 표현된다. 손실함수는 크로스 엔트로피(cross-entropy)에 기반하였고, 이때  $u$ 는 ‘알려지지 않은’ 클래스를,  $w(p_{c^*})$ 는  $L_{UP}$ 에 관한 가중요소를 의미한다[4].

$$L_{UP} = -w(p_{c^*}) \log(p_u'),$$

$$p_u' = \frac{\exp(s_u)}{\sum_{j \in C, j \neq c^*} \exp(s_j)}, \quad w(p_{c^*}) = (1 - p_{c^*})^a p_{c^*} \quad (2)$$

## 2. 실험설정 및 성능평가 기준

본 논문에서는 열린 환경에서의 객체 탐지에 관한 실험설정 및 성능평가 표준으로 활용되는 Dhamijia et al.의 2020년 연구[2]에서 제안된 열린 환경에서의 객체 탐지 프로토콜을 기준으로 활용하였다.

1) 실험설정: 컴퓨터비전 분야에서 주로 활용되는 PASCAL VOC(VOC)[5]와 MS COCO(COCO)[6] 데이터셋에 대하여, VOC의 trainval 데이터셋을 활용하여 닫힌 환경 훈련을 수행한다. 한편 20개의 VOC 클래스와 VOC에 포함되지 않은 60개의 추가적인 클래스를 추출하여 제시한 모델에 관한 열린 환경에서의 성능을 평가한다.

이때 성능평가에 활용되는 VOC-COCO-{T1, T2} 데이터셋을 정의한다. 먼저 T1의 경우 열린 환경의 클래스를 점진적으로 증가시켜 각각 {20, 60}개의 비 VOC 클래스를 포함하는  $n=5000$ 의 VOC 검증 이미지와  $\{n, 3n\}$ 개의 COCO 이미지로 두 개의 공동 데이터셋을 구축한다. T2의 경우 ‘Wilderness Ratio(WR)’를 높여  $n$ 개의 VOC 검증 이미지와  $\{0.5n, n, 2n, 4n\}$ 개의 COCO 이미지로 4개의 공동 데이터셋을 구성한다. 이때 WI는 ‘알려지지 않은’ 이미지에 대한 ‘알려진’ 이미지의 비율을 의미한다[2].

2) 성능평가 기준: 열린 환경에서 객체 탐지 모델의 성능평가를 위해 Wilderness Impact(WI)와 Absolute Open-Set Error(AOSE)라는 지표를 추가적으로 활용한다. WI는 ‘알려진’ 클래스로 잘못 분류된 ‘알려지지 않은’ 클래스의 객체의 정도를 의미하는 것으로, 아래의 식 (3)과 같이 정의되며, 이때  $P_k$ 와  $P_{k \cup u}$ 는 각각 닫힌 환경과 열린 환경에서 모델의 정밀도(precision)를 의미한다. 또한 AOSE의 경우 잘못 분류된 ‘알려지지 않은’ 객체의 수를 나타낸다.

$$Wilderness Ratio = \left( \frac{P_k}{P_{k \cup u}} - 1 \right) * 100 \quad (3)$$

## 3. 실험결과

본 논문에서 제안된 방식을 활용하여 표 1과 같은 결과를 도출할 수 있었다. 앞서 언급한 것과 같이 준비한 총 6개의 서로 다른 검증 공동 데이터셋에 대하여 성능평가 기준으로 활용한 WI와 AOSE 수치 모두 기본 모델과 비교하여 열린 환경에서 개선된 모습을 보였다.

WI와 AOSE에 대하여 각각 평균 7.14%와 5.9%의 성능향상을 확인할 수 있었다. 특히 VOC-COCO-60과 VOC-COCO-20000과 같이 ‘알려지지 않은’ 이미지의 비율이 높은 강한 열린환경에서 제시한 모델이 우수한 성능을 보이는 것을 확인할 수 있었다.

## III. 결론

본 논문에서는 대조학습 기반의 접근방식을 통해 열린 환경에서 객체 탐

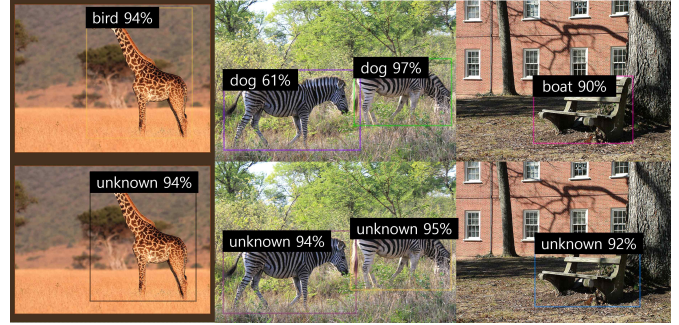


그림 3 실험결과 예시. 위는 기본 모델 아래는 제안된 모델의 결과를 나타냄

Method	VOC-COCO-40		VOC-COCO-60		VOC-COCO-2500	
	WI	AOSE	WI	AOSE	WI	AOSE
Baseline	1.74	55575	3.13	203521	1.31	27227
Ours	1.72	52470	3.03	186397	1.33	26404
Method	VOC-COCO-5000		VOC-COCO-10000		VOC-COCO-20000	
	WI	AOSE	WI	AOSE	WI	AOSE
Baseline	2.06	54926	2.98	113515	3.68	229572
Ours	1.93	53081	2.56	105656	3.07	209913

표 1 실험결과

지 모델인 YOLO의 성능을 개선하는 연구를 진행하였다. 모델에게 ‘알려지지 않은’ 데이터는 기저공간의 저밀도 영역에 존재한다는 가정을 활용하여, 대조학습 기반 학습요소인 CFL과 UPL을 추가함으로써 모델에게 ‘알려진’ 클래스의 클러스터를 축소하는 한편 ‘알려지지 않은’ 데이터가 위치하는 영역의 크기를 확장하는 방식으로 열린 환경에서 ‘알려지지 않은’ 객체를 ‘알려지지 않은’ 객체로서 잘 인식되도록 하였다.

본 연구를 통해 높은 연구가치가 있는 컴퓨터비전 분야의 객체 탐지 문제에 관하여 열린 환경에서의 강건성에 관하여 재고하고, 이를 대조학습 기반의 접근을 통해 해결하고자 하였다. 또한, 가장 널리 활용되는 객체 탐지 모델인 YOLOv5를 기본 모델로 활용함으로써 YOLO 모델의 성능향상에 관한 연구에 기여하였다.

한편, 실험 과정에서 활용한 기본 모델 선택에 따라 결과가 유동적이고 안정성이 확보되지 못한 경우가 종종 발생하였으며 이러한 문제는 추가적인 연구를 통해 개선하고자 한다.

## ACKNOWLEDGMENT

이 논문은 2023년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원(No.2019-0-01906,인공지능대학원지원(포항공과대학교))과 2023년도 정부(교육부)의 재원으로 한국연구재단의 지원(No. 2022R1A6A1A03052954, 기초연구사업)을 받아 수행한 연구 과제입니다.

## 참고 문헌

- [1] Abhijit Bendatle et al., "Towards open set deep networks" In proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2016
- [2] Akshay Dhamijia et al., "The Overlooked elephant of object detection: Open set" In proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pages 1021-1030, 2020
- [3] Glen Jocher et al., YOLOv5 by Ultralytics, 5 2020
- [4] Jiaming Han et al., "Expanding low-density latent regions from open-set object detection" In proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 9591-9600, 2022
- [5] Everingham et al., "The PASCAL Visual Object Classes (VOC) Challenge" International Journal of Computer Vision, 88(2), 303-338, 2010
- [6] Lin et al., "Microsoft COCO: Common Objects in Context" CoRR, abs/1405.0312, 2014