객체 검출을 위한 레이블 할당 방법의 비용 함수 연구

권용혜, 장시예, 남지인, 전세윤
마크애니

{yhkwon, syjang, jinam, syjeon}@markany.com

A Study on the Cost Function of Label Assignment Method for Object Detection

Yonghye Kwon, Siye Jang, Ji-in Nam, Saeyun Jeon
MarkAny

요 약

본 논문에서는 객체 검출 방법의 레이블 할당 방법에 요구되는 최적의 비용 함수를 탐색하고 제안한다. 일대일 레이블 할당 방법을 기반으로 최적의 비용 함수를 탐색하고, 일대다 레이블 할당 방법에 해당 비용 함수를 적용했을 때의 일반화 성능을 보인다. 제안된 비용 함수를 이용하여 학습된 객체 검출 모델은, 기존 객체 검출 방법들과 다르게, 앵커 박스, Feature Pyramid Networks, Deformable Convolution 를 이용하지 않고 PASCAL VOC 2007 데이터셋에서 83.04 mAP 에 달하는 검출 성능을 보였다.

Ⅰ. 서 론

객체 검출은 사람 재식별 및 추적[1], 자율 주행 등 다양한 컴퓨터 비전 어플리케이션에서 요구될 수 있는 기술이다. 이러한 이유로 활발한 연구가 진행되고 있다. 최근 객체 검출 분야에서 레이블 할당(Label Assignment) 방법이 주목받고 있다 [2, 3]. 레이블 할당 방법은 학습 과정에서 정답(Ground-truth) 바운딩 박스들을 토대로, 모델로부터 예측된 바운딩 박스들을 전경(Foreground 혹은 Positive Sample)과 배경(Background 혹은 Negative Sample)으로 분류하는 방법이다. 분류 이후, 전경으로 분류된 바운딩 박스들은 추론 과정에서 정답 바운딩 박스들을 예측할 수 있도록 학습이 되고, 배경으로 분류된 바운딩 박스들은 추론 과정에서 배경으로 분류되도록 학습이 된다. 특히, 레이블 할당 방법은 추론 과정에서 부하 없이 검출 성능을 향상시킬 수 있는 장점이 있다. 본 논문에서는 기존 레이블 할당 방법의 비용 함수(Cost Function)에 대한 다양한 실험을 통해 객체 검출 모델 학습에 적합한 비용 함수를 제안한다.

Ⅱ. 비용 함수 탐색을 위해 설계된 모델 및 비용 함수

본 논문에서는 레이블 할당 방법의 비용 함수에 따른 객체 검출 모델의 성능을 평가하기 위하여, 기존에 객체 검출 모델의 성능 향상을 위해 제안된 앵커 박스 [4], FPN(Feature Pyramid Networks) [5], Deformable Convolution [6] 등 비용 함수를 제외하고 성능에 영향을 미칠 수 있는 요소를 최소화한 앵커 프리 객체 검출 모델을 설계하였다. 실험을 위해 설계된 객체 검출 모델은 그림 1. 과 같다. C_f는 백본 모델의 출력 피쳐맵의 채널 수, C_n 은 클래스 수를 의미한다. 백본 모델의 출력 피쳐맵은 백본 모델의 마지막 레이어로부터 반환되는 피쳐맵을 의미한다.

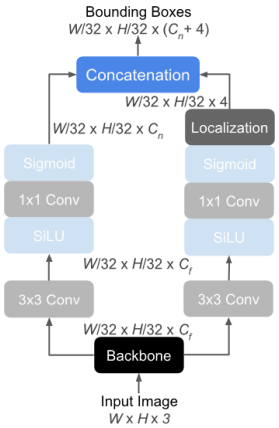

그림 1. 비용 함수 탐색을 위해 설계된 모델

제안하는 방법의 객체 위치 추정(Localization) 방법으로는 FCOS 의 위치 추정 방법을 채택하였다 [7]. 레이블 할당 방법은 DETR 에서 제안된 일대일 레이블 할당 방법을 채택하였다 [2].

$$C = \lambda_{cls}C_{cls} + \lambda_{iou}C_{iou} + \lambda_{l1}C_{l1} + \lambda_{center}C_{center} \quad (1)$$

C 는 레이블 할당 방법에 사용되는 비용을 의미한다. C_{cls} 는 예측된 바운딩 박스들과 클래스와 정답 바운딩 박스들간의 클래스 분류에 대한 손실(Loss) 값으로, Binary Cross Entropy 함숫값이다. C_{iou}는 예측된 바운딩 박스들과 정답 바운딩 박스들간의 객체 위치 추정에 대한 비용으로, UnitBox [8]에서 제안된 IoU(Intersection over Union) 손실 함숫값과 동일하다. C_{l1} 은 예측된 바운딩 박스들과 정답 바운딩 박스들의 좌표간 L1 손실 함숫값과 동일하다. C_{center} 는 바운딩 박스가 최종적으로 추론되는 피쳐맵의 각 피쳐의 정규화된 좌표값과 정답 바운딩 박스의 중심 좌표값간의

L1 손실 함숫값과 동일하다. 이는 기존 객체 검출 모델의 레이블 할당 방법이 대게 객체의 중심점에서 객체를 검출하도록 학습시켰을 때 높은 성능을 보인 결과 [7] 를 사전 지식으로써 반영한 것이다. λ_{cls}, λ_{iou}, λ_{l1}, λ_{center} 는 각각의 비용에 대한 가중치 값으로 0 이상의 값을 가질 수 있다. 손실 함수는 비용 함수와 유사하게 설계하였다.

$$L = L_{cls\_fg} + L_{cls\_bg} + L_{iou} + \lambda_{l1} L_{l1} \quad (2)$$

L 은 손실 함숫값으로 L_{cls_fg}, L_{iou}, L_{l1} 은 레이블 할당 방법에 의해 매칭된 바운딩 박스간의 비용 함숫값 C_{cls}, C_{iou}, C_{l1} 과 동일한 손실 함수를 이용해 값을 계산하도록 설계했다. L_{cls_fg} 의 경우 목푯값으로써 매칭된 정답 바운딩 박스와의 IoU 값을 할당하였다. L_{cls_bg} 의 경우 정답 바운딩 박스들과 매칭되지 않은 예측된 바운딩 박스들의 클래스별 컨피던스 값을 최소화하기위한 손실 함수의 값으로써 클래스 불균형 문제를 완화하기 위하여 손실 함수로 Focal Loss [9]를 사용하였다.

Ⅲ. 실험

레이블 할당 방법의 비용 함수에 따른 성능을 측정함에 있어, 학습 데이터로는 PASCAL VOC 2007 및 2012 데이터셋 학습 데이터와 검증 데이터를 사용했고, 성능 평가에는 PASCAL VOC 2007 테스트 데이터를 사용했다 [10]. 모델의 입력 이미지 크기를 512x512, 러닝 레이트를 0.000125, 배치 사이즈를 32 로 설정하여 70 에폭만큼 Adam 옵티마이저 [11]를 사용하여 학습시켰다. 러닝 레이트 조절 기법으로 Cosine Decay 기법을 적용하였다. 학습에는 RTX 3090 GPU 2 대를 사용하였다. 어그먼테이션은 Horizontal Flip, Translation, Random Crop, Resize 어그먼테이션을 사용했다. 백본 모델로는 ImageNet [12] Pre-trained ResNet18 [13]과 ConvNext [14]를 사용했다. 검출 성능은 mAP(Mean Average Precision)를 척도로 비교하였다.

λ_{cls}	λ_{iou}	λ_{l1}	λ_{center}	mAP
1	0	0	0	1.01
1	1	0	0	**71.94**
1	1	1	0	71.55
1	1	1	1	71.36

표 1. 비용 함수에 따른 객체 검출 성능

표 1. 은 비용 함수에 따른 검출 성능을 기록한 표이다. 이때 ResNet18 모델을 백본 모델로 사용하였다. 실험 결과 클래스 분류 비용과 함께 IoU 비용만을 계산하여 레이블을 할당했을 때 가장 높은 성능을 달성함을 확인할 수 있었다.

Methods	mAP
Ours-ResNet18 [13]	74.51
Ours-ConvNext-Tiny [14]	**83.04**
HSD [15]	83.00
Tencent-YOLOv3 [16]	79.60

표 2. 탐색된 비용 함수 기반의 일대다 레이블 할당 방법 및 백본 모델 별 객체 검출 성능 및 기존 방법간 성능 비교

표 2. 는 탐색된 비용 함수의 일반화 성능을 확인하기 위한 실험 결과로, 일대다 레이블 할당 방법 및 백본 모델 별 객체 검출 성능을 기록한 표이다. 일대다 레이블 할당 방법은 SimOTA 방법 [2]을 채택했다. 이때, 센터 샘플링 [7] 및 정답 바운딩 박스 외부에 위치한 피쳐로부터 추론된 바운딩 박스들에 대한 매칭 제약 조건은 제거하였다. ConvNext [14]의 경우 입력 이미지 사이즈를 640x640 로 설정하여 모델을 학습시켰다. 실험 결과, 앵커 박스, Deformable Convolution, FPN 등의 기법을 적용하지 않고 83.04 mAP 를 달성하였다. 이러한 결과는 레이블 할당 방법과 비용 함수의 중요성을 보인다.

Ⅳ. 결론

본 논문에서는 최근 객체 검출 분야에서 주목 받고 있는 레이블 할당 방법의 비용함수에 대한 다양한 실험을 진행하고, 최적의 비용함수를 탐색 및 제안하였다. 또한, 실험을 통해 일대일 매칭 방법을 기반으로 탐색된 비용 함수를 일대다 매칭 방법에 적용하고, 백본 네트워크별 검출 성능을 측정하였다. 실험을 통해 탐색된 비용 함수를 이용해 학습된 객체 검출 모델은 PASCAL VOC 데이터셋에 대해 83.04 mAP 를 달성하였다.

ACKNOWLEDGMENT

본 연구는 정보통신산업진흥원(NIPA)의 2022 년 AI 융합 국민 안전 확보 및 신속대응 지원 사업(실증환경의 학습데이터 구축에 기반하고 수배자 검색, 실종·가출자 대응과 위험예측이 가능한 인공지능 영상검색과 대상물 이동경로 추적 솔루션 개발, R-20220330-014563)의 지원을 받아 수행된 연구임.

참 고 문 헌

[1] Yifu Zhang, Chunyu Wang, Xinggang Wang, Wenjun Zeng, and Wenyu Liu, “FairMOT: On the Fairness of Detection and Re-Identification in Multiple Object Tracking,” International Journal of Computer Vision 129, pp. 3069-3087.

[2] Nicolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, and Sergey Zagoruyko, “End-to-End Object Detection with Transformers,” arxiv, 2020, https://arxiv.org/abs/2005.12872.

[3] Zheng Ge, Songtao Liu, Feng Wang, Zeming Li, and Jian Sun, YOLOX: Exceeding YOLO Series in 2021," arxiv, 2020, https://arxiv.org/abs/2107.08430.

[4] Joseph Redmon and Ali Farhadi, “YOLO9000: Better, faster, stronger,” Computer Vision and Pattern Recognition Conference, 2017.

[5] Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. “Feature pyramid networks for object detection,” Computer Vision and Pattern Recognition Conference, 2017.

[6] Xizhou Zhu, Han Hu, Stephen Lin, and Jifeng Dai, “Deformable ConvNets v2: More Deformable, Better

Results," Computer Vision and Pattern Recognition Conference, 2019.

[7] Zhi Tian, Chunhua Shen, Hao Chen, and Tong He, "FCOS: Fully Convolutional One-Stage Object Detection," International Conference on Computer Vision, 2019.

[8] Jiahui Yu, Yuning Jiang, Zhangyang Wang, Zhimin Cao, and Thomas Huang Shamir, "UnitBox: An Advanced Object Detection Network," 24th ACM Multimedia conference, 2016.

[9] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár, "Focal Loss for Dense Object Detection," International Conference on Computer Vision, 2017.

[10] Mark Everingham, Luc Van Gool, Christopher K. I. Williams, John Winn, and Andrew Zisserman, "The Pascal Visual Object Classes (VOC) Challenge," International Journal of Computer Vision 88, pp. 303-308.

[11] Diederik P. Kingma, and Jimmy Ba, "Adam: A Method for Stochastic Optimization," International Conference for Learning Representations 2015.

[12] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, Li Fei-Fei, "ImageNet: A large-scale hierarchical image database," Computer Vision and Pattern Recognition Conference, 2009.

[13] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep Residual Learning for Image Recognition", Computer Vision and Pattern Recognition Conference, 2016.

[14] Zhuang Liu, Hanzi Mao, Chao-Yuan Wu, and Christoph Feichtenhofer, "A ConvNet for the 2020s," Computer Vision and Pattern Recognition Conference, 2022.

[15] Jiale Cao, Yanwei Pang, Jungong Han, and Xuelong Li, "Hierarchical Shot Detector", International Conference on Computer Vision, 2019.

[16] https://github.com/Tencent/ObjectDetection-OneStageDet.