

# 쿨백-라이블러 발산을 이용한 후보영역 추출 네트워크와 객체 검출

서건석, 유재영, 최재석, 박노준  
서울대학교 융합과학기술부

e-mail : *geonseoks@snu.ac.kr*, *yoojy31@snu.ac.kr*, *jaeseok.choi@snu.ac.kr*,  
*nojunk@snu.ac.kr*

Region Proposal Network by using KL-Divergence in Object Detection

Geon-Seok Seo, Jae-Young Yoo, Jae-seok Choi, No-jun Kwak  
Graduate School of Convergence Science and Technology  
Seoul National University

## Abstract

The traditional region proposal network that using the deep neural network(DNN) is composed of two loss function that classifies objects and regress the box coordinates offset. However, it does not consider the relationship between object score and uncertainty of box coordinates. In this paper, we propose the region proposal network considering the uncertainty of box coordinates. To learn the network, we uses KL-Divergence which represents the distance between two probability distributions. We apply it to the existing two-stage object detection framework and show that it can improve the performance of existing method.

## I. 서론

과거부터, 객체 검출은 컴퓨터 비전 분야에서 중요한 문제였다. 이는 현재 자율주행, 얼굴인식, 스마트 팩토리 등 여러 산업 분야에서 사용되고 있다. 과거에는 특징기반 알고리즘을 사용하여 연구가 수행되었는데, 최근에는 DNN을 사용하여 높은 성능을 나타내는

객체검출 알고리즘들이 존재한다[6, 7]. 객체검출 알고리즘은 크게 일 단계 검출 방법(one-stage)과 이 단계 검출 방법(two-stage)으로 나뉜다. 이 단계 검출 방법은 먼저 객체가 있을만한 후보영역을 추출하고, 그 다음에 해당 후보영역에서 어떤 종류의 객체가 있는지 한번 더 검출하는데 전자의 상황에서 후보영역 추출 네트워크(region proposal network : RPN)가 사용된다. 이 단계 검출 방법은 초기에 후보영역 추출을 위해 신경망을 사용하지 않는 Selective Search[4] 나 EdgeBoxes[5]를 사용하였다. 그 이후 Faster R-CNN[1]이 RPN을 제안하며 성능과 속도를 끌어 올렸다.

후보영역 추출 네트워크 학습에는 2가지 손실 함수가 사용된다. 첫 번째는 객체가 있는지 없는지를 학습하는 이진 교차 엔트로피(binary-cross-entropy)함수이고, 두 번째는 박스 좌표의 차이를 줄이는 회귀(regression)함수이다.

일반적으로 사람이 그리는 정답 박스 또한 불확실성이 존재하는데, 기존의 후보영역 추출의 학습에는 이것이 고려되지 않는다. 즉, RPN의 학습은 후보영역(RoI : Region of Interest)의 객체 스코어와 박스 좌표의 분산간의 관계를 고려하지 않는다. 본 논문에서는 박스 좌표의 분산을 고려한 후보영역 추출 네트워크를 제안

한다. 제안한 후보영역 추출 네트워크를 객체검출 알고리즘에 적용하여 그 성능을 확인하였다.

## II. 본론

### 2.1 방법

후보영역 추출 네트워크를 학습하기 위해선 정답 박스 학습을 위한 샘플들이 필요하다. 우리는 정답 박스의 경우 Dirac delta 함수로, 틀린 박스의 경우 표준편차가 큰 가우시안 분포로 타겟을 정하였다. 타겟이 정해지면, 네트워크는 정답박스와 anchor간의  $x, y, w, h$  각 좌표의 차이인 평균과 표준편차를 학습한다. 학습을 위해 우리는 두 확률 분포간의 거리를 나타내는 함수인 쿨백-라이블러 발산(KL-Divergence)를 손실 함수로 사용하였다. 학습을 통해 RoI의  $x, y, w, h$ 의 표준편차가 구해지면, 이들의 곱으로 클래스 스코어를 정의하였다.

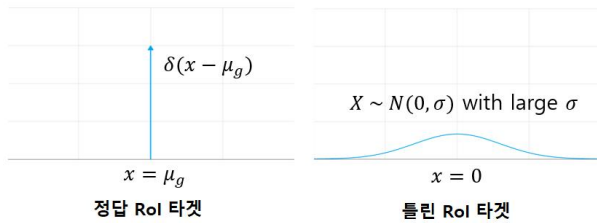


그림 1. RoI 종류에 따른 타겟 확률 분포

$$L = D_{KL}(P(x)||Q(x)),$$

$$= \int P(x) \log P(x) dx - \int P(x) \log Q(x) dx$$

그림 2. 손실 함수

## III. 구현 및 결과

데이터 셋은 PASCAL Visual Object Classes (VOC)[3]의 VOC2007 데이터 셋을 사용하였다. VOC 데이터 셋의 경우, 20가지 객체 종류가 있으며 총 학습 데이터는 5011장, 테스트 데이터는 4952장이 있다. 딥러닝 라이브러리는 Pytorch[2]를 사용하였으며, 이하세팅은 기존의 [1]과 같게 하였다. 실험은 원래 이미지의 비율을 그대로 유지하며 이미지의 작은 사이즈가 600으로 하였다. 타겟 가우시안 분포의  $\sigma$ 는 1/3로 하였다.

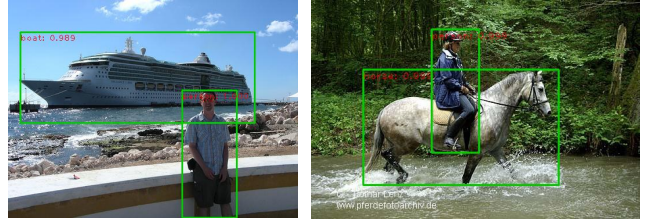


그림 3. Faster R-CNN에 적용 결과

후보영역	기본 네트워크	이미지 사이즈 (작은)	mAP
RPN	VGG-16	600	70.1
쿨백-라이블러 발산	VGG-16		<b>72.1</b>
RPN	ResNet-101		75.2
쿨백-라이블러 발산	ResNet-101		<b>75.6</b>

표 1. VOC 2007에서 mAP

## IV. 결론

본 논문에서는 객체를 분류하는 작업과 박스 좌표를 예측하는 작업으로 이루어져 있는 기존의 후보영역 추출 네트워크를 두 확률 분포간의 거리 함수인 쿨백-라이블러 발산을 이용하여 하나의 작업으로 정의하여 문제를 해결하였다. 각각의 샘플들은 타겟 분포의 평균과 표준편차를 학습한다. 쿨백-라이블러 발산을 사용함으로써, 네트워크는 정답박스와 차이의 표준편차를 예측할 수 있는 장점이 생기고 이를 객체 스코어로 사용하였다. 제안한 알고리즘을 이 단계 검출 알고리즘에 적용하였을 때 성능이 향상된 것을 확인할 수 있었다.

## ACKNOWLEDGMENT

본 연구는 과학기술정보통신부 및 정보통신기술진흥센터의 정보통신, 방송연구개발 사업의 일환으로 진행되었음 (2017-0-00306)

## 참고문헌

- [1] S. Ren, K. He, R. Girshick, and J. Sun. Faster R-CNN: Towards real-time object detection with region proposal networks. In NIPS, 2015.
- [2] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. Automatic differentiation in pytorch. In NIPS-W, 2017.
- [3] Mark Everingham, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. The pascal visual object classes (voc) challenge. *International journal of computer vision*, 88(2):303 - 338, 2010.
- [4] J. R. Uijlings, K. E. van de Sande, T. Gevers, and A. W. Smeulders. Selective search for object recognition. *IJCV*, 2013.
- [5] C. L. Zitnick and P. Dollar. Edge boxes: Locating object proposals from edges. In *ECCV*, 2014.
- [6] H. Law and J. Deng, “Cornersnet: Detecting objects as paired keypoints,” in *ECCV*, September 2018.
- [7] Xin Lu, Buyu Li, Yuxin Yue, Quanquan Li, and Junjie Yan. Grid r-cnn. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2019.