강건한 객체탐지 구축을 위해 Pseudo Labeling 을 활용한 Active Learning

김채윤 [1], 이상민 [2]
[1] 광운대학교 인공지능응용학과 석사과정
[2] 광운대학교 인공지능응용학과 교수

2022115409@kw.ac.kr, smlee5679@gmail.com

Active Learning with Pseudo Labeling for Robust Object Detection

ChaeYoon Kim[1], Sangmin Lee[2]
[1]Dept. of Artificial Intelligence Applications, Kwangwoon University
[2]Dept. of Artificial Intelligence Applications, Kwangwoon University

요 약

딥러닝 기술의 발전은 고품질의 대규모 데이터에 크게 의존한다. 그러나, 데이터의 품질과 일관성을 유지하는 것은 상당한 비용과 시간이 소요된다. 이러한 문제를 해결하기 위해 최근 연구에서 최소한의 비용으로 최대의 성능을 추구하는 액티브 러닝(active learning) 기법이 주목받고 있는데, 액티브 러닝은 모델 관점에서 불확실성(uncertainty)이 높은 데이터들을 샘플링 하는데 중점을 둔다. 하지만, 레이블 생성에 있어서 여전히 많은 시간적, 자원적 비용이 불가피한 점을 고려할 때 보완이 불가피 하다. 본 논문에서는 의사-라벨링(pseudo labeling)을 활용한 준지도학습(semi-supervised learning) 방식과 학습 손실을 동시에 사용하여 모델의 불확실성(uncertainty)을 측정하는 방법론을 제안한다. 제안 방식은 레이블의 신뢰도(confidence)와 학습 손실의 최적화를 통해 비용 효율적인 데이터 레이블 생성 방식을 제안한다. 특히, 레이블 데이터의 품질(quality) 및 일관성(consistency) 측면에서 딥러닝 모델의 정확도 성능을 높임과 동시에 적은 데이터만으로도 효과적인 학습이 가능할 수 있는 메커니즘을 제안한다.

1. 서론

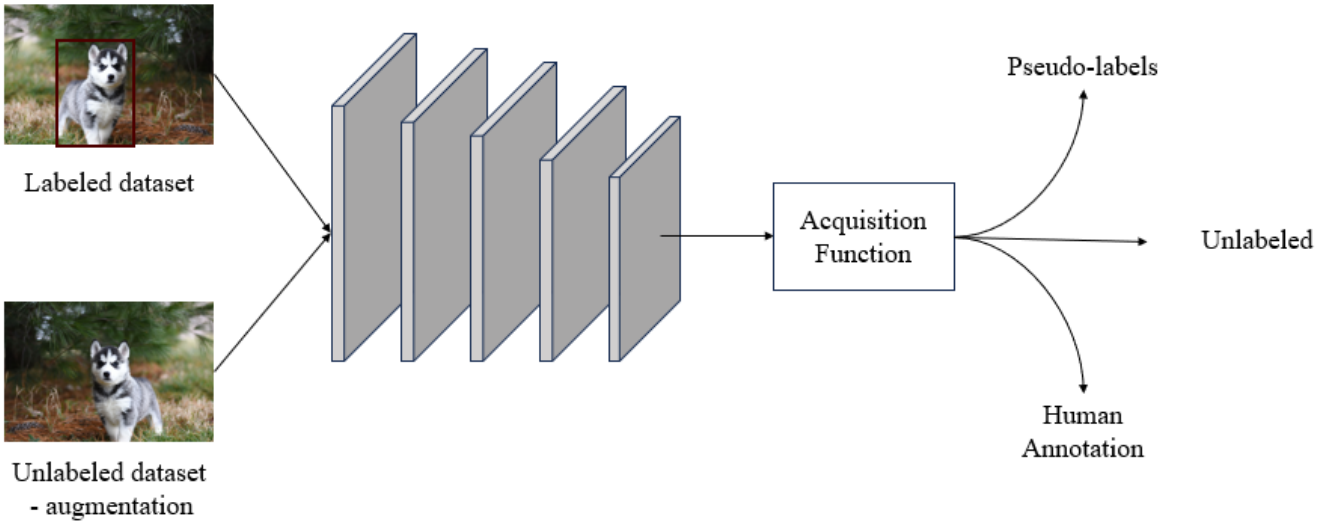
딥러닝 알고리즘은 많은 양의 데이터를 기반으로 복잡한 패턴들을 학습하기 때문에, 딥러닝 기술 발전과 성능은 대규모의 레이블링 된 데이터셋에 의존하여 발전해왔다. 그러나, 낮은 품질의 데이터 세트는 모델이 잘못된 패턴을 학습할 가능성이 있기 때문에, data 의 quality 와 consistency 가 딥러닝 알고리즘의 성능을 결정짓는 중요한 요소로 작용한다. 그럼에도 불구하고, train dataset 을 만들 때, data 의 quality 와 consistency 을 유지하면서 train dataset 을 생성하고 관리하는 것은 상당한 자원을 필요로 한다. 특히, 의료 영상 등, 전문성이 필요한 분야에는 소모되는 자원이 증가된다. 따라서, 이런 문제를 해결하기 위해 최근 연구들에서 액티브 러닝(active learning) 연구가 제안되었다.

액티브 러닝은 모델이 학습에 가장 유익한 샘플들을 선별하고, 이를 통해 최소비용으로 최대 성능을 추구하는 방식이다. 이러한 접근법은 모델의 변화를 크게 유도하는 데이터를 정확하게 샘플링 할 수 있는지에 중점을 둔다. 그러나, 기존의 액티브 러닝 방식은 모든 클래스에 대해 동일한 신뢰도(confidence)를 부여한다는 점, 클래스 불균형 문제를 반영하지 못하는 점, 그리고 distribution shift 문제에 취약하다는 문제점을 갖는다.

따라서, 본 연구에서는 이런 문제를 해결하기 위해 개별 class 마다 차별적인 신뢰도(distinctive confidence)를 반영할 수 있도록 pseudo-labeling 과 loss prediction module 을 도입한 액티브 러닝 전략을 제안한다. pseudo-labeling 은 신뢰도가 높은 예측에 대해 자동으로 라벨을 부여하는 준지도학습 방식으로, 클래스의 편향성을 방지할 수 있다. 그러나, 초기 단계에서는

많은 양의 불확실성 데이터로 인해 labeling 비용이 효율적이지 않을 수 있다.[1]

따라서, 본 연구에서는 네트워크의 견고성과 일관성에 기반하여 신뢰도가 낮거나 성능이 낮은 클래스의 객체까지 고려할 수 있는 손실 예측 기법인 loss prediction 모듈을 도입하여 효율성을 높이고자 한다.



(그림 1) architecture overview

2. 방법

일반적인 액티브 러닝 방법론은 모든 클래스에 대해 동일한 confidence 로 학습되기 때문에 데이터셋의 구성이 클래스 간에 불균형 하거나 다양한 패턴을 보이는 경우, 특정 클래스에 낮은 정확도를 보이거나 특정 클래스에 distribution shift 가 발생할 수 있다. 본 연구에서는 각 클래스에 대해 다른 confidence 를 반영할 수 있도록 준지도 학습(semi-supervised learning) 전략과 손실 예측 모듈을 활용한 액티브 러닝 전략을 제안한다. 본 연구에서 제안된 모델의 아키텍처(architecture)는 그림 1 과 같다.

2-1) Consistency loss

consistency loss 은 모델 예측과 실제 값 사이의 일관성을 측정하는 데 사용된다. 이는 original image 와 horizontal flip 이 적용된 image 에 대해 detector 가 예측한 class 와 bounding box 의 예측 결과를 비교하여 계산한다. 이를 위해, Eq. 1 에 나타낸 바와 같이 모든 매칭된 예측 쌍(matched pairs of predictions)에 대해 class consistency loss 와 Bbox 의 IoU 에 해당하는 손실을 구하기 위한 localization loss 를 평균을 내어 계산한다.

예를 들어, original image 와 horizontal flip 이 적용된 image 에 대한 두 예측값이 매우 유사하다면, 모델은 해당 이미지에 대해 강건함(robustness)을 가진 것으로 판단한다. 반대로 두 예측값이 상이할 경우 inconsistent 한 데이터로 판단하여 실제 레이블을 확인할 수 있는 오라클(oracle)의 확인을 요청한다. 판단이 어려운 데이터에 대해서는 최종적으로 human annotation 통해 모델의 성능을 향상시킨다.

$$L_{con} = E[L_{conC}(c_i', \hat{c_i})] + E\left[L_{conL}\left(b_i', \hat{b_i}\right)\right] \quad (1)$$

2-1-1) Class inconsistency loss

Class inconsistency loss 는 $c_i', \hat{c_i}$ 가 주어진 상황에서, KL(Kullback-Leibler) divergence 을 사용하여 모델의 inconsistency 를 계산한다.

Eq. 2 에 나타낸 바와 같이 KL divergence 를 이용하여 두 예측 결과의 차이값을 구해 inconsistency 를 판단한다. 만약 이전에 예측한 결과와 새롭게 예측한 결과가 일치하지 않는다면, 이것을 'inconsistent'하다고 정의하며, 'inconsistency'가 클수록 모델 훈련에 가치 있는 샘플로 판단할 가능성이 크다.

$$\mathrm{L_{conC}}(\mathrm{c_i'}, \hat{\mathrm{c_i}}) = \frac{1}{2}[\mathrm{KL}(\mathrm{c_i'}, \hat{\mathrm{c_i}}) + \mathrm{KL}(\hat{\mathrm{c_i}}, \mathrm{c_i'})] \quad (2)$$

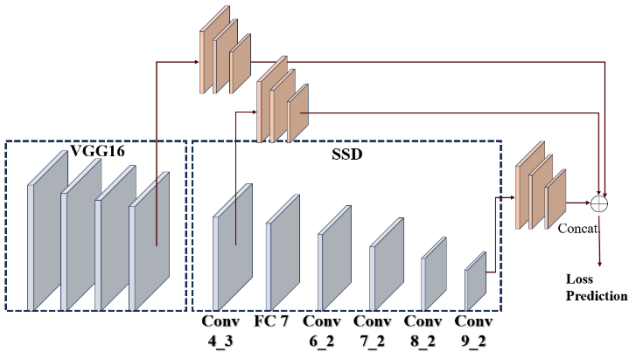
2-1-2) Bounding box loss

localization inconsistency loss 은 예측된 바운딩 박스(bounding box)가 실제 객체와 얼마나 잘 일치하는지 측정하는 방법이다. 이를 위해 original image 에 horizontal flipping augmentation 을 적용하였다. original image 와 augmented image 를 모델에 입력하고, 모델이 예측한 원본 이미지의 바운딩 박스 위치와 반전된 이미지의 바운딩 박스 위치 간의 IoU(Intersection over Union)를 계산함으로써 localization inconsistency loss 를 계산한다. 이 때, unlabeled data 는 horizontal flip 을 적용하기 때문에 Eq. 3 에서 보인 바와 같이 예측된 x 좌표값 $\hat{x}$에 대해서 negation 을 적용한다. 이를 통해 물체의 위치를 나타내는 바운딩 박스의 중심점과 크기에 대한 차이를 계산하도록 한다. 이러한 localization inconsistency loss 를 통해 모델이 객체의 위치를 정확하게 예측하도록 도와주며, 모델의 일관성을 향상시킬 수 있다.

$$L_{conL}(b_i', \hat{b_i}) = \frac{1}{4}\left(\left||\delta x_i' - (-\delta \hat{x}i)\right|\right|^2 + \left||y0'i - \hat{y}0_i\right|\right|^2 + \left||w_i' - \widehat{w_i}\right|\right|^2 + \left|\left|h_i' - \hat{h_i}\right|\right|^2\Big) \quad (3)$$

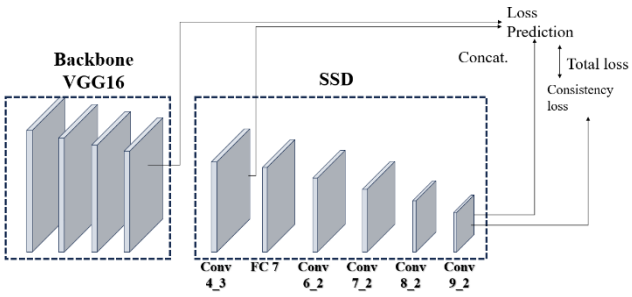
2-2) Loss Prediction Module

Loss Prediction Module 은 레이블이 없는 샘플에 대한 손실을 예측하기 위함이다. 이것은 학습 과정에서 가장 많은 정보를 제공할 가능성이 있는 샘플들을 선택하고, 이러한 샘플들에 레이블링 작업을 진행함으로써 효율적인 학습을 돕는다. 예를 들어, 10 개의 샘플이 있다면, 각 샘플에 대한 손실 값을 계산하고 이를 기준으로 정렬한다. 그런 다음 상위 K 개의 샘플(즉, 가장 큰 손실 값을 가진 샘플들)을 선택하여 전문가 사람에 의한 레이블링 한다. 손실 예측 값을 구하는 방법은 그림 2 와 같이 feature extractor 에서 나오는 특징 h 들을 GAP(Global Average Pooling)과 FC(Fully Connected Layer), 그리고 ReLU(Rectified Linear Unit) 연산을 통해 차원 축소를 한 후, 추출된 특징들을 결합하여 사용한다. 결합된 특징은 모델이 각 샘플에 대한 손실 값을 예측하는데 사용된다.



(그림 2) loss prediction module architecture

2-3) Acquisition function



(그림 3) Acquisition function

$$L_{total} = L_{con}(\hat{y}, y) + \lambda \cdot L_{loss}(\hat{l}, l) \quad (4)$$

최종 손실 함수는 Eq. 4 와 같다. 여기서 $L_{con}(\hat{y}, y)$ 은 Eq.1 에서 보인 바와 같이 consistency 를 판단하기 위한 손실 함수이며, L_{loss} 는 손실 예측에 대한 손실을 나타낸다. 여기서, λ 는 두 항 사이의 상대적인 가중치를 조절하는 hyperparameter 이다. 간단한 MLP(Multi-Layer Perceptron) 구조를 가진 손실 예측 모듈이 전반적인 경향성을 파악하여 손실 값을 정확하게 예측하는 것은 불가능하다. 그러나 loss prediction module 의 주요 목적은 정확한 손실 값을 계산하는 것이 아니라, 데이터 포인트들 중에서 어느 것이 가장 많은 정보를 가지고 있는지, 즉, 어떤 샘플의 손실 값이 가장 큰지 판별하는 것이다. 이런 접근 방식으로 인해 불확정성이 큰 샘플들에 대해선 더욱 주목하게 되고, 이를 통해 학습 과정에서 중요한 역할을 하는 샘플들을 선별하여 효율적으로 학습할 수 있다.

3. 결론 및 향후 연구

연구의 결과로는, 학습이 반복됨에 따라 모델의 성능이 향상되는 것을 확인하였으나 데이터 수와 반복 횟수를 줄일 수 있는지에 대한 검증이 필요하다. 이를 위해 추후 연구에서는 Iteration Step 마다 K%의 데이터를 다르게 Sampling 하는 방법을 고려하여 비교 실험을 진행하고자 한다.

참고문헌

[1] Elezi, I., Yu, Z., Anandkumar, A., Leal-Taixe, L., & Alvarez, J. M. “Not all labels are equal: Rationalizing the labeling costs for training object detection.” In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 14492-14501), 2022

[2] Donggeun Yoo and In So Kweon. “Learning loss for active learning”. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 93–102), 2019

[3] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C Berg. “Ssd: Single shot multibox detector.” In ECCV, 2016.