X-ray 영상 데이터를 통한 자기 지도 학습 기반 객체 탐지의 영역 간극에 따른 성능차이 검증

윤의현, *이재구
*국민대학교

*jaekoo@kookmin.ac.kr

Verifying Performance Differences in Object Detection Based on Self-Supervised Learning Using X-ray Image Data According to Domain Gap

Euihyun Yoon, *Jaekoo Lee
College of Computer Science, Kookmin University.

요 약

최근 심층 학습(deep learning)은 다양한 분야에서 폭넓게 개발되어 활용되고 있다. 그중 자기 지도 학습은 정답 값이 없는 방대한 데이터로 사전 학습하여 정답 값을 사용한 지도 학습 대비 조금 낮은 성능을 낸다. 대부분의 자기 지도 학습 방식은 정답 값이 없는 ImageNet 을 사전 훈련하여 전이 학습을 진행할 때 높은 성능을 보인다. 본 논문에서는 자기 지도 학습 방법으로 사전 훈련된 모델을 엑스레이(X-ray) 이미지에 전이 학습 할 때 사전 훈련에 사용된 데이터집합에 따른 성능 차이를 확인했다. 자기 지도 학습의 경우 이미지의 모양에 집중하므로 데이터집합의 객체 분포에 따라 전이 학습 시 영역 간극이 존재하였다. 이러한 영역 간극이 적은 데이터로 사전 학습된 모델은 전이 학습 데이터의 25%만 사용해도 mAP(mean average precision)가 30% 높은 성능을 보이며 지도학습 대비 mAP 가 평균 0.5% 이하의 성능 차이를 낸다. 본 논문에서는 자기 지도 학습을 통해 엑스레이 이미지에 전이 학습 할 경우 객체 분포에 따른 영역 간극이 적은 모델의 사용을 제안한다.

Ⅰ. 서 론

최근 인공지능 기술은 다양한 분야에서 연구, 개발되어 실생활에도 폭넓게 활용되고 있다. 그 중 컴퓨터 비전(computer vision) 분야의 과업을 살펴보면 이미지 분류, 이미지 생성, 객체 탐지 등 이 존재 한다.

지도학습은 정답 값을 토대로 학습을 진행하는 방식으로 데이터집합에 정답 값을 필수로 요구한다. 반면에 자기 지도 학습은 사전에 정답 값없이 방대한 데이터로 학습을 진행하고 학습된 추출기(extractor)를 토대로 과업에 맞게 전이 학습 하는 방식이다. 이런 자기 지도 학습의 가장 큰 이점은 큰 비용이 소모되는 정답 값이 없이도 학습을 진행할 수 있다는 점이다.

대부분의 자기 지도 학습 방식은 ImageNet[1]데이터 집합에 사전 학습된 추출기가 CoCo[2] 데이터집합으로 사전 학습한 추출기보다 객체탐지 과업에 전이학습 시 더 높은 성능을 보인다[3].

본 논문에서는 지도학습과 자기 지도 학습, 준 지도 학습을 엑스레이(X-ray) 이미지에 대해 객체 탐지 전이 학습 성능을 비교하였으며 각각 ImageNet 데이터집합과 CoCo 데이터집합으로 사전 학습된 모델을 사용했다. 실험을 통하여 사전 학습에 사용된 데이터내 객체 분포에 따른 [그림 1]의 엑스레이 사진(x-ray image)에서의 영역(domain) 간극을 확인하였다.

Ⅱ. 본론

초기의 자기 지도 학습 방식은 사전 과업(pretext task)을 지정하여 학습을 진행했다[4]. 점점 자기 지도 학습 방식이 연구됨에 따라 기존의 사전 과업을 학습하는 방식보다 높은 성능을 내는 대조 학습(contrastive learning) 방식이 연구되고 있다[6].

대조 학습 방식을 순서대로 살펴보면 다음과 같다. 하나의 이미지로부터 서로 다르게 변형 (augmentation) 된 이미지 두 개를 하나의 쌍으로 생성한다. 해당 쌍을 긍정적인 쌍(positive pair)으로 두고 나머지 이미지들로부터 나온 두 개의 쌍을 부정적인 쌍(negative

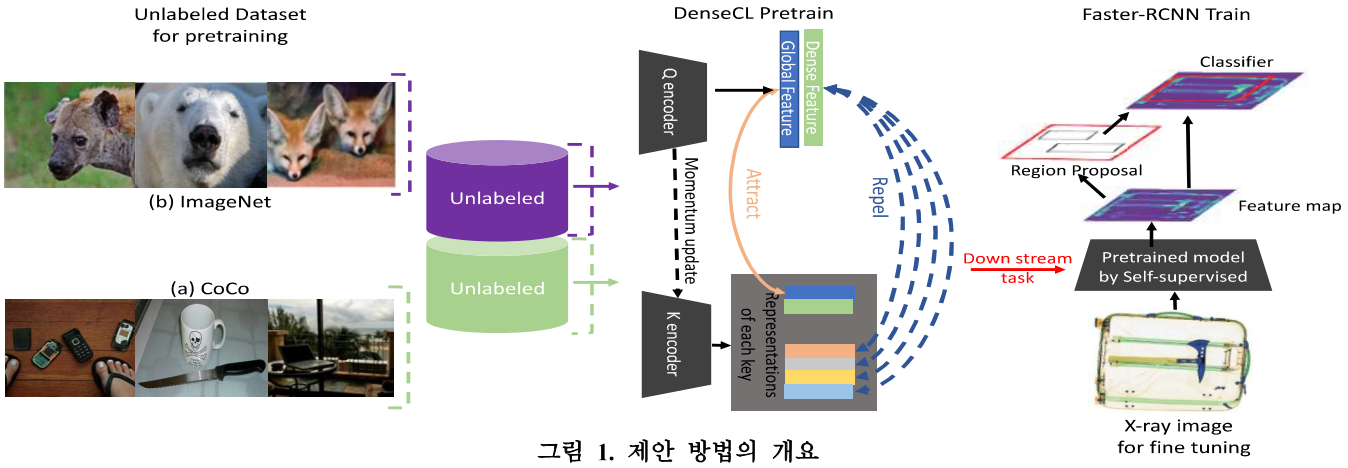

그림 1. 제안 방법의 개요

표 1. 데이터 비율 별 지도학습 대비 성능, 사전 데이터 집합에 따른 전이학습 성능 대비

	(a) from CoCo (pretrained) to X-ray (finetuned)			(b) from ImageNet (pretrained) to X-ray (finetuned)			(a)CoCo 와 (b)ImageNet 의 전이성능차이		
평가지표 / 데이터	mAP	mAP_{50}	mAP_{75}	mAP	mAP_{50}	mAP_{75}	mAP	mAP_{50}	mAP_{75}
25%	**64.5(-22.5)**	**81.3(-17.6)**	**73.4(-22.3)**	49.5(-37.5)	66.2(-32.7)	57.9(-37.8)	**30.3%**	22.8%	26.7%
50%	**77.4(-9.6)**	**90.9(-8.0)**	**85.8(-9.9)**	72.1(-14.9)	85.7(-13.2)	81.8(-13.9)	7.3%	6.0%	3.8%
75%	**85.3(-1.7)**	**97.8(-1.1)**	**94.2(-1.5)**	80.3(-6.7)	95.4(-8.2)	90.7(-5.0)	6.2%	2.5%	3.8%
100%	**87.2(+0.2)**	**98.3(-0.6)**	**95.4(-0.3)**	84.7(-2.3)	97.5(-1.4)	93.2(-2.5)	2.9%	0.8%	2.3%

pair)으로 두고 학습을 진행한다. 대조 학습은 긍정적인 쌍의 코사인 유사도는 가깝게 하고 부정적인 쌍의 코사인 유사도는 멀게 학습한다. 이런 방식으로 학습된 추출기는 사용자의 과업에 맞게 전이 학습을 통해서 사용된다.

본 논문에서 사용한 자기 지도 학습 모델은 DenseCL[3]이며 MoCo[5]의 전역 정보만 보는 문제점을 개선한 모델이다. MoCo 는 전역 풀링 층(global pooling layer)을 사용하여 전역적인 정보만 보게 된다. 반면에 DenseCL 은 1x1 합성곱 층(convolution layer)을 사용하여 세밀한 특징(dense feature) 보게 된다. 이러한 학습의 장점은 객체 탐지같이 객체의 세밀한 부분의 특징을 파악하는 과업에서 더욱 좋은 성능을 보여 주었다[3].

Ⅲ. 실험 및 고찰

본 논문에서 평가를 위해 사용한 객체 탐지 모델은 Faster-RCNN[6]으로 합성곱 신경망(convolutional neural networ)에서 나온 특징 맵(feature map) 을 토대로 지역 제안(region proposal)과 분류(classification)를 하는 모델이다. 전이 학습의 성능은 합성곱 신경망을 사전에 자기 지도로 학습한 모델을 사용하며 지역 제안 네트워크와 분류기만 학습하여 평가한다. 지도학습 성능은 Faster-RCNN 을 처음부터 정답 값을 토대로 학습하여 평가한다.

데이터집합 은 AI-hub 에 있는 위해 물품 엑스레이[7] 사진이며 학습에 20,000 장과 평가에 4,000 장 을 사용하였다. 또한 준 지도 학습 의 성능을 측정하기 위해 데이터를 각각 25, 50, 75 비율로 나누어서 사용 하였다. 해당 데이터의 객체 종류는 칼, 총, 도끼, 노트북 등이 존재한다.

ImageNet 데이터 집합의 경우 [그림 1]과 같이 1000 개의 종류 중 대략 300~400 개 정도가 동물에 대한 종류이다. CoCo 데이터집합의 경우 [그림 1]과 같이 91 개의 종류 중 오직 10 개의 종류만 동물이고 나머지는 망치, 칼, 가위 등 가정과 주방에서 사용되는 이미지가 많이 포함돼있다. 본 논문에서 전이학습에 사용한 데이터의 경우 [그림 1]과 같이 위해 물품에 관한 엑스레이 이미지로 객체의 분포가 CoCo 데이터에 존재하는 객체 종류 와 더욱 유사하다.

[표 1]에서 CoCo 로 사전 학습된 모델의 성능을 보면 전의 학습의 경우 평균 0.5%p 의 차이만 난다. 전체 데이터의 75%를 사용하여 전이 학습한 경우 전체 데이터로 지도 학습한 모델보다 mAP50 기준 최소 1.1%p 의 차이로 준 지도 학습 시에도 지도 학습과 유사한 성능을 확인할 수 있다. 또한 [표 1]의 가장 오른쪽을 보면 영역 간극에 따른 성능 대비를 확인할 수 있으며 mAP 기준 최대 30% 의 성능 차이가 난다.

이러한 결과를 토대로 사전 학습 데이터 집합에 따른 영역 간극을 확인하였다. 만약 사전 학습 데이터 내 객체 형상에 따른 영역 간극이 최소화된 모델을 사용한다면 [표 1]과 같이 데이터의 정답 값 없이 학습한 자기 지도 학습 방법이 지도 학습에 준하는 성능을 보임을 확인 할 수 있다.

Ⅳ. 결론

본 논문에서는 실험을 통하여 자기 지도 학습을 통해 엑스레이 이미지에 전이 학습을 할 때 영역 간극이 적은 모델을 사용한다면 정답 값 없이 학습한 자기 지도 학습이 정답 값을 사용한 지도 학습과 유사한 성능을 내며 전체 데이터의 일부분만 사용해도 준수한 성능을 볼 수 있었다.

또한 자기 지도 학습을 통해 엑스레이 이미지에 전이 학습 할 경우 데이터의 영역 간극이 적은 모델을 선택하기 위해선 전이 학습 데이터의 객체 형상과 더욱 유사한 사전 학습 모델을 사용해야 함을 제안한다.

ACKNOWLEDGMENT

이 논문은 2022 년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임(No.RS-2022-00167194,미션 크리티컬 시스템을 위한 신뢰 가능한 인공지능)

참 고 문 헌

[1] Deng, Jia, et al. "Imagenet: A large-scale hierarchical. image database." 2009 IEEE conference on computer vision and pattern recognition. Ieee, 2009.

[2] Tsung-Yi Lin., Michael Maire, et al, “Microsoft COCO: Common Objects in Context”, Computer Vision – ECCV 2014

[3] Wang, Xinlong, et al. "Dense contrastive learning for self-supervised visual pre-training." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021.

[4] Noroozi, Mehdi, and Paolo Favaro. "Unsupervised. learning of visual representations by solving jigsaw puzzles." European conference on computer vision. Springer, Cham, 2016.

[5] He, Kaiming, et al. "Momentum contrast for unsupervised visual representation learning." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020.

[6] Ren, Shaoqing, et al. "Faster r-cnn: Towards real-time. object detection with region proposal networks." Advances in neural information processing systems 28 2015

[7] Hazardous Goods X-ray Image, AI-Hub, Available: https://aihub.or.kr/aihubdata/data/view.do?currMenu=115&topMenu=100&aihubDataSe=realm&dataSetSn=233