적외선 영상의 객체 탐지를 위한 전이 학습 기법

이명학, 유영준, 조영완, 이재구*
국민대학교 소프트웨어학부

*jaekoo@kookmin.ac.kr

Transfer Learning for Object Detection in Infrared Image

Myunghak Lee, Youngjun Yoo, Youngwan Jo, Jaekoo Lee*
College of Computer Science, Kookmin University.

요 약

객체 탐지(Object Detection)는 영상이 주어졌을 때 해당 영상에서 특정 객체를 찾아 경계 상자(Bounding Box)를 설정하고 분류하는 과업이다. 현재 대부분의 관련 연구는 가시광선을 이용한 일반 영상에 대해서만 이루어지고 있다. 하지만 가시광선 영상의 경우 환경의 영향에 민감하여 자율주행 등의 과업에서는 사용하기 부적절하다. 따라서 본 논문에서는 객체 탐지 과업을 가시광선 영상이 아닌 적외선 영상에 적용해야 할 필요성을 인지하였다. 그러나 적외선 영상 데이터 집합은 일반 가시광선 영상 데이터 집합보다 그 수와 다양성이 떨어지므로 학습이 불완전하게 이루어진다는 한계가 있다. 따라서 본 논문에서는 적외선 영상에서 객체 탐지를 하는 모델을 학습할 때에 가시광선 영상을 이용하여 부족한 데이터 집합을 보완하는 방법을 제안한다. 우리는 가시광선 영상으로 모델을 사전 학습(Pretrain) 시킨 후 이를 적외선 영상에 대해 전이학습(Transfer Learning)을 시키는 방법으로 부족한 데이터 집합을 보완하여 성능을 끌어 올릴 수 있었다.

Ⅰ. 서 론

객체 탐지(Object Detection) 과업은 다양한 용도로 사용할 수 있다[6, 7]. 그러나 현재 대부분의 객체 탐지 과업에 관한 연구는 가시광선 카메라를 이용해 얻은 RGB 영상에 대해서만 중점적으로 진행되고 있다. 가시광선 카메라의 경우 환경적 요인에 굉장히 민감하다는 한계가 있다. 예를 들어 [그림 1]에서 볼 수 있듯이 주변의 빛의 강도, 혹은 안개나 흙먼지 등의 요인에 의하여 영상의 질이 급격히 떨어지는 경우도 있다. 이러한 단점은 자율주행을 위한 객체 탐지 과업에서 더욱 두드러지게 된다. 차량의 경우 그 특성상 다양한 환경에 노출되게 되는데 이때 환경에 의존적인 가시광선 카메라로는 객체 탐지가 제대로 이루어지지 않는다. 따라서 다른 대책이 필요하며 그 중 하나가 적외선 카메라이다.

적외선 카메라는 평상시 모든 물체가 지속적으로 방출하는 적외선을 이용하므로 가시광선 카메라에 비해 주변 환경에 덜 민감하다는 장점이 있다. 그러나 적외선 영상에 관한 데이터 집합의 양은 가시광선 영상에 비하면 매우 부족하다. 또한 적외선 데이터의 경우 일반 가시광선 영상보다 수집하는데 더 큰 비용이 필요하므로 데이터 집합을 만드는 것이 더 어려워 학습이 불완전하게 이루어질 가능성이 있다. 따라서 본 논문에서는 이를 해결하고자 다른 도메인인 가시광선 영상을 이용한 전이 학습을 시도해보았고 실험 결과 [표 2]에서와 같이 유의미한 성능 향상을 볼 수 있었다.

그림 1. 가시광선 영상과 적외선 영상의 비교.

Ⅱ. 본론

우리는 실험의 신뢰성을 높이기 위하여 2 가지 모델을 사용하여 실험을 진행하였다. 이때 처음으로 사용한 Faster R-CNN[1] 모델의 경우 관심 영역(Region of Interest)을 구하는 문제와 해당 관심 영역에 대하여 객체를 탐지를 하는 문제를 따로 진행하는 반면 두번째 모델인 YOLO-v3[2]는 동시에 한단계로 진행하는 모델이다. 그리고 데이터 집합으로는 가시광선 데이터 집합인 PASCAL VOC[5]와 COCO[3], 그리고 적외선

데이터 집합	사람	자전거	차량
학습 (Train)	22,372	3,986	41,260
검증 (Validation)	5,779	471	5,432
테스트 (Test)	21,965	1,205	14,013

표 1. FLIR 데이터[4]의 각 클래스 간 객체의 개수

신경망 모델	학습에 이용한 데이터 집합	FLIR 테스트 집합에 대한 성능 지표			
		사람	자전거	차량	mAP
Faster R-CNN [1]	(a)Pascal VOC[5]	0.198	0.022	0.247	0.211
	(b)FLIR	0.325	0.331	0.404	0.355
	(c)Pascal VOC + FLIR	0.455	0.327	0.493	**0.465**
YOLO-v3 [2]	(d)COCO[3]	0.203	0.467	0.627	0.371
	(e)FLIR[4]	0.326	0.176	0.817	0.506
	(f)COCO + FLIR	0.355	0.442	0.874	**0.553**

표 2. 각 모델에 대한 학습에 사용한 데이터 집합 별 성능을 FLIR 데이터의 테스트 집합에 대한 성능

데이터 집합인 FLIR[4] 데이터 집합을 사용하였다. 그러나 FLIR 데이터 집합의 경우 그 양이 부족하며 [표 1]에서와 같이 각 클래스 간의 비율이 맞지 않는 문제가 있다. 또한, 학습 데이터가 모두 개별적인 것이 아니라 특정 지역에서만 찍은 몇몇 비디오 영상에서 프레임을 잘라와 사용한 것이므로 데이터의 다양성이 부족하다는 문제점도 있다. 따라서 적외선 데이터 집합만을 가지고 신경망을 훈련하면 [표 3]에서 볼 수 있듯이 과적합(Overfitting) 문제가 쉽게 일어난다는 문제점이 존재한다. 그러므로 본 논문에서는 이러한 문제를 해결하기 위해 다른 도메인인 가시광선 영상의 데이터 집합으로 사전 학습을 시키는 방법을 사용하였고 그 결과 유의미한 성능 향상을 이룰 수 있었다.

Ⅲ. 실험

본 논문에서는 사전학습 시에 처음에는 중추(Backbone) 네트워크를 동결시키고(Freeze) 실험하였으나 가시광선 영상과 적외선 영상의 도메인 간의 차이 탓에 유의미한 성능 향상을 확인하지 못하였다. 따라서 두 도메인의 차이를 줄이기 위하여 COCO 와 Pascal VOC 데이터 집합을 흑백으로 전환한 후 신경망을 사전 학습시켰으며 클래스의 수도 적외선 영상과 같게 맞춘 후 진행하였다. 그리고 이후 적외선 영상으로 전이 학습 시에도 중추 네트워크를 동결시키지 않고 학습률(learning rate)만 10 분의 1 로 줄인 뒤 학습을 진행하였다.

또한 가시광선 영상이 적외선 영상에 대한 신경망 학습에 어느 정도의 영향을 끼치는지 알아보기 위하여 첫 번째로 가시광선 영상만으로 학습한 모델이 실제 적외선 영상에서 어느 정도의 객체 탐지 성능을 가지는지 확인해 보았다. 그 결과 [표 2]의 (a)와 (d)에서 볼 수 있듯이 가시광선 데이터 집합만으로 학습한 모델에 적외선 데이터 집합을 테스트 집합으로 사용했을 때 Faster R-CNN 와 YOLO-v3 에서 mAP(mean Average precision) 수치가 각각 0.221 과 0.371 이라는 유의미한 성능을 보임을 입증할 수 있었다.

두 번째로는 적외선 영상만으로 학습한 모델과 가시광선 영상으로 사전학습을 한 모델의 차이를 알아보기 위한 실험을 진행하였다. 그 결과 [표 2]의 (b)와 (c) 그리고 (e)와 (f)의 비교에서 볼 수 있듯이 사전학습을 시킨 모델의 성능이 사전 학습을 하지 않은 모델에 비하여 각각 9.29%와 30.98% 씩 향상됨을 알 수 있었다.

신경망 모델	이용한 데이터 집합	mAP
YOLO-v3	FLIR	0.683
	COCO + FLIR	0.613

표 3. FLIR 데이터의 학습 데이터 집합에 대한 성능을 보면 [표 2]의 결과와는 다르게 사전학습을 하지 않은 모델의 성능이 지나치게 높게 나온다.

Ⅳ. 결론

본 논문에서는 적외선 영상에서의 객체 탐지 모델의 성능을 높이기 위하여 일반 가시광선 영상을 이용하여 사전학습을 시키는 방법을 사용하였다. 이 방법을 통해 Faster R-CNN 에서 0.11, YOLO-v3 에서 0.182 만큼의 mAP 수치 향상을 이루었다. 또한 향후 연구에서는 도메인 적응(Domain Adaptation) 방법을 통하여 전이학습의 성능을 향상시키는 방법을 조사할 것이다.

ACKNOWLEDGEMENT

이 성과는 2020 년도 정부(미래창조과학부)의 재원으로 한국연구재단의 지원(No. NRF2018R1C1B5086441)과 과학기술정보통신부 및 정보통신기획 평가원의 SW 중심대학지원사업(2016-0-00021)으로 수행된 연구임.

참 고 문 헌

[1] Joseph Redmon and Ali Farhadi. “Yolov3: An incremental improvement.” arXiv preprint arXiv:1804.02767, 2018.

[2] S. Ren, K. He, R. Girshick, and J. Sun. “Faster r-cnn: Towards real-time object detection with region proposal networks.” arXiv preprint arXiv:1506.01497, 2015.

[3] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollar, and C. L. Zitnick. “Microsoft coco: Common objects in context.” In European conference on computer vision, pages 740-755. Springer, 2014

[4] F. A. Group. Flir thermal dataset for algorithm training. “https://www.flir.in/oem/adas/adas-dataset-form/”, 2018.

[5] M. Everingham, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman. “The pascal visual object classes (voc) challenge”. International journal of computer vision, 88(2):303-338,2010

[6] Licheng Jiao, Fan Zhang, Fang Liu, Shuyuan Yang, Lingling Li, Zhixi Feng, Rong Qu “A Survey of Deep Learning-based Object Detection” arXiv preprint arXiv:1907.09408, 2019.

[7] Shervin Minaee, Yuri Boykov, Fatih Porikli, Antonio Plaza, Nasser Kehtarnavaz, Demetri Terzopoulos “Image Segmentation Using Deep Learning: A Survey” arXiv preprint arXiv:2001.05566,2020