

Efficient Detector 기반의 도메인 적응 물체 탐지 기법

김화랑^{1,2}, 임길택², 최두현¹¹ 경북대학교 대학원 전자전기공학부, ² 한국전자통신연구원

khr1393@knu.ac.kr, ktl@etri.re.kr, dhc@ee.knu.ac.kr

Domain Adaptive Object Detection Based on Efficient Detector

Hwa-Rang Kim^{1,2}, Kil-Taek Lim², Doo-Hyun Choi¹¹ Graduate School of Electronic and Electrical Engineering, Kyungpook National University,² Electronics and Telecommunications Research Institute

요약

본 논문은 기존에 학습되었던 모델이 새로운 도메인의 데이터에 대한 해석을 용이하게 할 수 있도록 하는 도메인 적응 기법에 대해 연구한다. 학습 데이터의 양이 충분하다 할지라도 실제 시험 현장에서는 기존의 학습 데이터와는 다른 도메인의 데이터를 처리해야 하는 경우가 있을 수 있기 때문에 도메인 적응 기법은 실제 현장에서 용이하게 쓰일 수 있다. 본 논문에서는 도메인 적응 기법을 적용하여 물체 탐지 기법에 대한 연구를 수행한다. 기본 모델은 최근까지도 물체 탐지에 우수한 성능으로 큰 영향력을 미치고 있는 efficient detector 를 사용하였으며, source 이미지와 target 이미지에 대한 특징맵 간의 차이를 줄이기 위해 Gradient Reversal Layer (GRL)을 이용하였다. 따라서 모델이 source 와 target 간의 비슷한 특징을 추출할 수 있으며 target 이미지에 대해서 일반적인 efficient detector 보다 좋은 성능을 보였다.

I. 서론

컴퓨터 비전 분야의 주요 분야 중 하나인 물체 탐지 분야는 물체의 범주 정보와 동시에 위치 정보까지 예측하며, 단순 분류 모델보다 더욱 복잡한 차원의 해석 능력을 요구한다. 따라서 단순 이미지 분류 작업보다 많은 양의 데이터가 요구된다. 하지만 많은 양의 데이터를 확보했다 할지라도 테스트되는 모든 도메인의 데이터를 학습할 수는 없다. 가령, 낮과 밤에 촬영한 도로 이미지로 차량 탐지 모델을 설계하여도 안개 낀 날, 눈 오는 날처럼 특수한 상황은 기존의 학습 데이터와 분포가 다르므로 잘 동작하지 않을 수 있다. 또한 특수한 상황에 대한 데이터는 구하기가 쉽지 않기 때문에 추가적으로 모델을 학습시키기도 쉽지 않다.

따라서 본 논문에서는 이러한 도메인 차이에 대한 성능 저하의 문제를 해결하기 위해 도메인 적응 기법을 사용한 물체 탐지 모델을 제안한다. 기존에 사용자가 다량 보유하고 있는 데이터를 source 데이터, 적응시키고자 하는 목표가 되는 데이터를 target 데이터라고 하며, 도메인 적응 기법을 적용하여 target 데이터의 수가 현저히 적을 경우에도 target 테스트 데이터에 대한 성능 개선을 기대할 수 있다.

제안하는 모델은 그 성능이 입증된 efficient detector [1]를 기반으로 하며, 도메인 적응을 위해 source 도메인과 target 도메인에 대해 최대한 비슷한 특징 추출할 수 있도록 하는 GRL [2]을 사용하였다. GRL 은 gradient 값을 반전시켜 고의적으로 학습이 잘 되지 않도록 한다. 이를 통해 모델은 source 와 target 도메인을 잘 구분하지 못하게 되며, source 데이터로부터 학습된 모델로 target 도메인에서도 효과를 볼 수 있게 된다.



그림 1. (a) Cityscapes 이미지 (b) Foggy Cityscapes 이미지

본 논문에서는 도메인 적응 분야에서 주로 쓰이는 Cityscapes [3]와 Foggy Cityscapes [4] 데이터로 실험하였다 (그림 1 참고).

II. 본론

그림 2 는 도메인 적응을 위해 Efficient Detector 모델에 GRL 을 추가한 것이다. 모델 중간에 특징맵들은 도메인 분류기를 통해 해당 특징맵이 source 도메인인지 target 도메인인지 분류되도록 학습될 수 있지만, 본 실험의 목적은 모델이 데이터의 도메인을 분류할 수 없도록 만드는 것이므로 입력 데이터가 GRL 을 통과하도록 하였다.

GRL 은 모델 학습 시 연산되는 역전파 알고리즘에서 gradient 를 반전시키기 때문에 손실함수의 최적화를 방해하는 역할을 한다. GRL 과 연결되는 도메인 분류는 특징맵 추출 단계인 backbone 과, backbone 뒤에 연결되는 neck 인 BiFPN 의 출력단에서 수행하며, backbone 에 연결되는 분류는 image 레벨의 손실값, neck 에 연결되는 분류는 instance 레벨의 손실값을 출력하며 각 손실함수는 식 1, 2 와 같다.

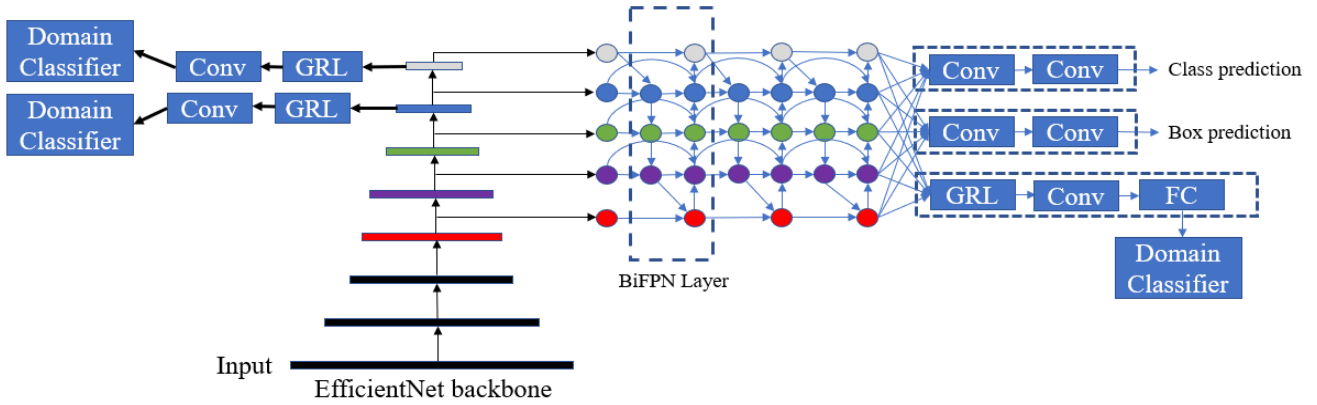


그림 2. 제안하는 Efficient Detector 기반의 도메인 적응 물체 탐지 모델 (Backbone 의 마지막 두 특징맵과 neck 의 출력에 GRL 과 도메인 분류기를 연결).

표 1. Foggy Cityscapes 이미지에 대한 평가 결과표.

	person	rider	car	truck	bus	train	mcycle	bicycle	mAP
EfficientDet	13.6	18.7	22.5	8.8	17.7	4.5	11.2	17.5	14.3
제안하는 모델	17.7	21.5	28.1	13.1	22.6	8.6	15.4	22.8	18.7

$$L_{img} = - \sum_{i,u,v} [D_i \log p_i^{(u,v)} + (1 - D_i) \log(1 - p_i^{(u,v)})] \quad (1)$$

$$L_{ins} = - \sum_i [D_i \log p_i + (1 - D_i) \log(1 - p_i)] \quad (2)$$

D_i 는 i 번째 batch 의 도메인 label 을 의미하며, source 는 0, target 은 1 로 정하였다. 식 1 에서 u 와 v 는 특징맵의 픽셀 좌표를 의미한다. 식 2 는 특징맵이 Fully Connected Layer 를 통과하므로 좌표 값을 고려하지 않는다. 최종 손실함수는 식 3 에서 정의된 바와 같으며 L_{det} 는 bounding box 예측 손실값과 class 예측 손실값을 더한 값이고 α 는 detection loss 와 domain loss 의 비율을 정해주는 hyperparameter 이다. 본 실험에서는 0.1 로 설정하였다.

$$L_{total} = L_{det} + \alpha(L_{img} + L_{ins}) \quad (3)$$

실험에 사용된 데이터셋은 도메인 적응 분야에서 널리 사용되고 있는 Cityscapes 와 Foggy Cityscapes 데이터셋이다. 학습 이미지는 2,975 장이며 테스트 이미지는 500 장이다. 본 실험에서는 Foggy Cityscapes 이미지를 label 이 없는 target 으로 간주하며, 학습 단계에서 target 이미지의 L_{det} 은 역전파되지 않는다.

표 1 은 실험 결과를 나타낸다. 테스트는 target 인 Foggy cityscapes 에 대해 진행하였으며, 도메인 적응 기법을 사용하지 않은 Efficient Detector 보다 제안하는 모델이 더 우수한 성능을 보여주고 있다.

III. 결론

도메인 적응을 위해 기존 모델에 GRL 과 도메인 분류기를 연결함으로써 target 데이터에 대한 label 없이도 target 의 물체 탐지 능력을 향상시킬 수 있었다.

따라서 추가적인 labeling 작업 없이 새로운 데이터에 대한 분석을 용이하게 할 수 있다.

ACKNOWLEDGMENT

This work was supported by Electronics and Telecommunications Research Institute (ETRI) grant funded by the Korean government. [21ZD1120, Development of ICT Convergence Technology for Daegu-GyeongBuk Regional Industry].

참 고 문 헌

- [1] M. Tan, R. Pang, and Q. V. Le, "Efficientdet:: Scalable and efficient object detection," In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020, pp. 10 781–10 790.
- [2] Y. Ganin and V. Lempitsky, "Unsupervised domain adaptation by back-propagation," In International conference on machine learning. PMLR, 2015, pp. 1180–1189.
- [3] Cordts, Marius, et al., "The cityscapes dataset for semantic urban scene understanding," In Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 3213–3223.
- [4] C.Sakaridis, D. Dai, and L. Van Gool, "Semantic foggy scene understanding with synthetic data," International Journal of Computer Vision, 2018, pp. 973–992.