

강인한 물체 검출을 위한 뉴럴 네트워크의 특징 분석

김준표, 김희제, 정유진, 박상진, *송병철

인하대학교 전기컴퓨터공학과

ogum1945@hanmail.net, gimeejae@gmail.com, feeljin8018@naver.com,

san9569@naver.com, *bcsong@inha.ac.kr

Feature Analysis of Neural Network for Robust Object Detection

Junpyo Kim, Heeje Kim, Yoojin Jung, Sangjin Park and Byung Cheol Song

Inha University

요 약

왜곡된 영상에 강인한 물체 검출은 자율 주행과 같은 안전에 치명적인 실생활 응용 분야에서 핵심 문제로 다뤄지고 있다. 이러한 이유로, 영상 처리 및 컴퓨터 비전 분야에서 강인한 물체 검출에 대한 연구가 활발하게 진행되고 있다. 본 논문에서 우리는 왜곡된 영상이나 이미지에서도 일정한 성능의 물체 검출을 위한 연구 [4]를 benchmark 하여 다양한 방법으로 변형된 데이터셋을 통해 학습한 모델을 성능과 feature map 측면에서 분석해 봄으로써, 향후 강인한 물체 검출에 있어 효과적인 성능 향상을 위한 intuition 을 제공하고자 한다.

1. 서론

물체 검출은 컴퓨터 비전 분야의 가장 기본적인면서 많이 활용되는 기술 중 하나로, 이미지에서 특정 범주의 모든 객체를 정확하게 식별하고 localize 하는 것을 목표로 한다. Convolutional Neural Network (CNN)의 등장으로 많은 CNN 기반 물체 검출 접근법이 제안되었고 물체 검출 알고리즘의 성능이 크게 향상되었다.

그러나, 현실에서의 물체 검출은 여전히 시점, 배경, 조명, 이미지 품질 등의 차이로 인해 학습 데이터와 테스트 데이터 간 도메인 이동 (domain shift)이 발생한다는 문제에 직면해 있다. 예를 들어, 자율 주행의 경우, 특정 자동차에 사용되는 카메라 유형 및 설정은 학습 데이터 수집에 사용되는 카메라의 것과 다를 수 있으며, 실제 자율 주행 시스템이 작동할 기상 조건에 비해, 학습 데이터는 일반적으로 가시성이 더 좋은 날씨에 수집된다. 이러한 도메인 이동은 물체 검출에 있어서 상당한 성능 저하를 초래하는 것으로 확인되었다 [1]. 더 많은 학습 데이터의

수집이 이러한 도메인 이동의 영향을 완화할 수 있지만, 추가적인 annotation 은 비용과 시간이 많이 소요되므로 이러한 해결법을 적용하기는 현실적으로 어렵다. 따라서, 존재하는 학습 데이터로부터 시각적으로 다른 새로운 도메인의 새로운 학습 데이터를 증강하여 강인한 물체 검출기를 학습하는 접근법이 바람직하다.

이러한 강인한 물체 검출 (robust object detection)을 위해, 몇 가지 유형으로 왜곡된 데이터셋으로 모델을 학습하여 학습에서 사용되지 않은 corruption 에 대한 robustness 를 달성하려는 시도가 있었으나, CNN 이 다양한 왜곡에 대한 학습을 받았음에도 불구하고 새로운 왜곡 유형에 잘 일반화되지 않는다는 것이 입증되었다 [2]. 반면, 다른 연구 [3]는 GAN 으로부터 stylized 된 ImageNet 데이터셋으로 모델을 학습시켜 모델이 질감을 무시하고 객체 모양에 초점을 맞추는 더 강한 편향을 가지게 하였으며 다양한 corruption 에 대해 모델의 robustness 가 증가함을 확인하였다.

[3]의 결과를 기반으로 본 논문에서는 COCO 2017 데이터셋으로부터 세 가지 유형의 학습 데이터 셋 (COCO-clean,

COCO-stylized, COCO-combined)을 구축하고, 각 데이터셋으로 학습한 faster-RCNN 모델을 각각 COCO 와 COCO-corrupted 데이터셋으로 평가하였다 [4]. 이후, feature map visualization 을 통해 각 데이터셋으로 학습한 object detector 의 feature map 을 분석하여 향후 연구에 대한 intuition 을 제공하고자 한다.

2. AdaIN 기반 style-transfer 를 통한 학습 데이터셋 변형

이미지 분류 시, 이미지의 내용을 다른 이미지의 스타일과 결합하는 style-transfer 기법 [5]을 적용하면, corruption 에 대하여 모델의 robustness 가 증가하는 것으로 나타났다 [3].

Style-transfer 기법 중 하나인 Adaptive Instance Normalization (AdaIN)은 AdaIN layer 를 통해 content feature 의 channel-wise mean 과 variance 를 style feature 와 align 함으로써 임의의 style 을 실시간으로 transfer 할 수 있다 [6].

본 논문에서는 이 기법을 기존 물체 검출 학습 데이터셋 (COCO-clean) 변형에 활용하여 두 가지 변형 학습 데이터셋 COCO-stylized, COCO-combined 을 구축하였다. 이때, style-transfer 를 위한 style 데이터셋은 Kaggle's Painter by Numbers 데이터셋이 사용되었다. COCO-stylized 는 모든 학습 데이터셋을 stylized 버전으로 교체한 데이터셋, COCO-combined 는 기존 학습 데이터셋에 stylized 버전 데이터셋을 1:1 로 섞은 데이터셋이다.

3. 다양한 corruption 기법을 활용한 테스트 데이터셋

Corruption 에 대한 모델의 성능을 평가하기 위해 기존 COCO 2017 테스트 데이터셋(COCO)에 corruption 을 가하여 COCO-corrupted 테스트 데이터셋 (COCO-C)을 구축하였다. 이때, corruption 은 Gaussian noise, Motion blur, Snow 등 15 가지 기법 [7] 중 한 가지 기법을 각 영상에 균등하게 적용하였으며, Corruption 의 강도인 severity 는 가장 약한 1 에서부터 가장 강한 5 까지의 다섯 가지 값 중 하나의 값을 각 이미지에 랜덤으로 적용하였다.

4. Feature map visualization

Feature map visualization 은 뉴럴 네트워크의 각 layer 에서 추출되는 feature map 을 시각화 하여, feature 가 각 layer 를 통과하면서 어떻게 변화하는지 확인하는 기법을 말한다 [8]. Inner feature 와 filter 패턴을 통해 해당 뉴럴 네트워크 모델이 특정 물체를 검출하는데 영향을 미친 feature 가 무엇인지, 특정 픽셀 영역이 어떻게 판단에 기여했는지 등을 확인하여 종합적으로 모델의 성능을 비교 및 분석할 수 있다.

5. 실험 결과

우리는 [4]의 실험 환경을 따라 총 6 개의 실험 과정을 거쳤다. 우리는 기존의 COCO 학습 데이터셋으로부터 세 가지 유형의 학습 데이터셋 (clean, stylized, combined)을 구축한 후 faster-RCNN [9]모델을 학습하고, 학습한 모델을 각각 COCO-clean 과 COCO-C 데이터셋으로 평가한다. 이후 우리는 faster-RCNN 모델의 backbone 인 ResNet50 을 통과한 feature map 을 테스트 데이터셋의 이미지와 함께 시각화 하여 물체 검출의 성능 평가 지표인 Average Precision (AP)과 함께 비교함으로써 성능 개선의 요인을 파악하고자 한다.

표 1. 다양한 학습 및 테스트 데이터셋에서 faster-RCNN 의 성능.

Train dataset \ Test dataset	COCO [AP]	COCO-C [AP]
	COCO-clean	37.3
COCO-stylized	22.5	12.3
COCO-combined	33.2	18.3

표 1 은 6 가지 실험 과정의 결과를 보여준다. 기존 학습 데이터셋과 stylized 버전 데이터셋을 1:1 로 섞은 COCO-combined 데이터셋이 강인한 물체 검출에 가장 효과적임을 나타낸다.



그림 1. COCO-clean 과 COCO-combined 데이터셋으로 학습한 모델로부터 추출된 feature map.

그림 1 은 COCO-combined 학습 데이터셋으로 학습한 모델의 feature map 이 왜곡된 영상에서 좀 더 강인한 물체 검출에 적합한 것을 입증한다. CNN 을 통해 생성된 feature map 이 region proposal network 을 거쳐 관심 영역을 추출하기 때문에, feature map 이 COCO-combined 학습 데이터셋으로 학습되었을 때 기존의 COCO 학습 데이터셋보다 검출하고자 하는 물체의 위치에 더 focus 하므로 성능 개선의 주요한 원인이 됨을 확인할 수 있다.

6. 결론

우리는 다양한 기후 환경에서도 안정적으로 작동할 수 있는 물체 검출 모델 [4]을 reproduce 하여 style-transfer 기법을 통해 구축된 데이터셋으로 물체 검출 모델을 학습한 뒤, mAP (mean average precision)와 feature map visualization 을 통해 학습한 모델이 다양한 corruption 에 대해 robustness 가 증가했음을 확인하였다. 특히, feature

map visualization 으로 inner feature 를 확인하는 과정을 통해, feature 가 검출하고자 하는 물체의 위치에 더 정확하게 mapping 될 수록 object detector 의 성능이 향상될 것임을 추론할 수 있었다. 하지만, 현재 robustness 는 아직 실제 적용까지 한계가 있으며 더 강인한 물체 검출 모델을 개발하거나 학습하는 방법이 필요하다. 단지 input 데이터의 변형 뿐만 아니라 feature level 의 augmentation 을 통해 보다 왜곡에 강인한 물체 검출 모델을 학습하는 것이 향후 연구로 남아있다.

감사의 글

본 연구는 한국연구재단 중견연구 (2022R1A2C2010095)의 연구비 지원을 받아 수행되었습니다.

참고문헌

- [1] S. Ren et al., "Faster R-CNN: Towards real-time object detection with region proposal networks", *Neural Information Processing Systems*, 2016.
- [2] R.Geirhos et al., "Generalisation in humans and deep neural networks", *Neural Information Processing Systems*, 2018.
- [3] R.Geirhos, et al., "ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness", *International Conference on Learning Representations*, 2019.
- [4] C. Michaelis et al., "Benchmarking Robustness in Object Detection: Autonomous Driving when Winter is Coming", *Machine Learning for Autonomous Driving Workshop at Neural Information Processing Systems*, 2019.
- [5] L. A. Gatys, A. S. Ecker and M. Bethge, "Image Style Transfer Using Convolutional Neural Networks," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2414-2423, 2016.

- [6] X. Huang and S. Belongie, "Arbitrary style transfer in real-time with adaptive instance normalization", *International Conference on Computer Vision*, pp.1501-1510, 2017.
- [7] D. Hendrycks and T. Dietterich. "Benchmarking neural network robustness to common corruptions and perturbations", *International Conference on Learning Representations*, 2019.
- [8] Z. Qin et al., "How convolutional neural networks see the world - a survey of convolutional neural network visualization methods", *Mathematical Foundations of Computing*, Vol.1, No.2, pp.149-180, May 2018.
- [9] Ren, Shaoqing, et al. "Faster r-cnn: Towards real-time object detection with region proposal networks." *Advances in neural information processing systems* 28 (2015).