

비지역적 특징 융합을 이용한 물체 검출 기법

최준호, 이민규, *송병철
 인하대학교
 *bcsong@inha.ac.kr

Object Detection Method with Non-local Feature Fusion

Jun Ho Choi, Min Kyu Lee, *Byung Cheol Song
 Inha University

요 약

최근 딥러닝 기반의 다양한 물체 검출 알고리즘이 제안되어 높은 성능을 보이고 있다. 본 논문은 이러한 딥러닝 기반 물체 검출의 성능을 향상시키기 위해 입력영상에서 추출된 특징 지도를 강화하는 비지역적 특징 융합과, 이를 이용한 물체 검출 기법을 제안한다. 제안 기법은 입력영상에서 CNN 을 통해 추출한 특징 지도를 비지역적 특징 강화 블록을 통해 강화한다. 해당 블록 내에서 입력된 특징 지도는 먼저 여러 리셉티브 필드를 갖는 특징 지도로 분기된다. 그리고 분기된 특징 지도들은 비지역적 특징 융합 모듈에 의해 융합되어 강화된다. 이러한 과정을 통해 강화된 특징 지도는 비지역적 문맥 정보가 강화된 특성을 가지며, 해당 특징 지도를 이용하여 최종적으로 물체 검출을 수행한다. Pascal VOC 공인 데이터셋을 통한 실험 결과, 제안 기법은 기존 비교 기법 대비 향상된 검출 성능을 보인다.

1. 서론

영상에서 다중 물체의 위치 정보와 클래스 정보를 획득하는 물체 검출은 컴퓨터 비전의 가장 중요한 요소 분야 중 하나이다. 최근에는 딥러닝 기반의 다양한 물체 검출 알고리즘이 제안되어 가장 높은 성능을 보이고 있다. 딥러닝 기반 물체 검출은 크게 두가지 방식으로 나뉜다. 하나는 영역 제안 단계와 물체 검출 단계로 구성되어 있는 다중 단계 검출 방식이고, 다른 하나는 별도의 영역 제안 없이 추출된 특징 지도에서 곧바로 물체를 검출하는 단일 단계 검출 방식이다. 다중 단계 검출 방식의 대표적인 기법은 Faster R-CNN [1]이 있으며, 단일 단계 검출 방식에는 SSD [2]가 있다.

한편, 이러한 딥러닝 기반 물체 검출 기법의 검출 성능을 향상시키기 위해 다양한 접근들이 존재했다. 대표적으로 FPN [3]는 CNN 의 단계별 특징 지도를 융합함으로써 낮은 레벨과 높은 레벨 특징 지도 사이의 의미적 간격을 줄여 물체 검출 성능을 향상시키고자 하였다. 또한, 최근에는 RFBNet [4]이 제안되어 특징 지도를 다양한 리셉티브 필드를 갖는 특징 지도로 분기한 뒤 다시 융합하는 방식으로 특징 지도를 강화하여 검출 성능을 향상시켰다. 그러나 이러한 기법들은 특징 지도의 지역적 정보는 강화할 수 있으나 비지역적 문맥 정보는 고려하지 않고 있다. 기존에 비지역적 문맥 정보를 고려한 연구는 대표적으로 Non-local Neural Networks [5]가 있었으나 이는 단일 특징 지도가 입력되어 스스로 비지역적 특징점 사이의 상관 정도를 이용해 특징을 강화하는 self-attention 기법이므로 고정된 리셉티브 필드를 갖는 특징점 간의 상관 정도만 이용한다는 한계점이 존재한다.

이러한 문제를 해결하기 위하여 본 논문은 영상에

존재하는 여러 크기의 비지역적 문맥 정보를 통해 특징 지도를 강화하는 새로운 비지역적 특징 융합과, 이를 이용한 높은 검출 성능의 물체 검출 기법을 제안한다. 제안 기법은 CNN 에서 출력된 특징 지도를 비지역적 특징 강화 블록을 통해 강화하여 물체 검출을 수행한다. 비지역적 특징 강화 블록 내에서 특징 지도는 다양한 리셉티브 필드를 갖는 특징 지도로 분기되고, 분기된 특징 지도들은 다시 비지역적 특징 융합 모듈을 거쳐 비지역적 문맥 정보가 강화된 형태로 출력된다. Pascal VOC 공인 데이터 셋을 이용한 실험 결과, 제안 기법은 SSD 대비 3.2%, RFBNet 대비 0.5% 향상된 검출 성능을 보인다.

2. 제안하는 검출기 구조

제안하는 검출기의 전체적인 구성도는 그림 1 과 같다. 먼저 CNN 을 통해 입력 영상에 대한 특징을 추출한다. 그리고 추출된 특징 지도에 대해 비지역적 특징 강화 블록을 거쳐 특징 강화를 수행한다. 해당 블록은 특징 지도를 서로 다른 리셉티브 필드를 갖는 여러 특징 지도로 분기하는 단계, 그리고 이를 이용하여 비지역적 특징 융합을 수행하는 단계로 구성되어 있다. 비지역적 특징 강화 블록을 거쳐 강화된 특징

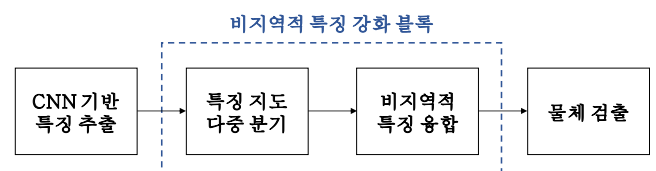


그림 1. 제안하는 검출기의 전체 구성도

지도에 대해 최종적으로 물체 검출을 수행한다.

제안 기법은 단일 단계 검출 방식으로서 기본적으로 SSD 와 동일한 검출 구조를 갖는다. CNN 기반 특징 추출 단계에서는 VGG-16 특징 추출 신경망에 추가적인 컨볼루션 레이어를 결합한 형태의 기본 신경망을 사용하고, 기본 신경망에서 단계별로 6 개의 서로 다른 크기의 특징 지도를 추출하여 물체 검출에 이용한다. 물체 검출 단계에서는 각 특징 지도에 대해 미리 정의한 디폴트 박스 별로 클래스 분류 및 박스 위치 보정을 수행한다. 비지역적 문맥 정보를 이용한 특징 강화가 요구되는 물체는 대부분 영상에서 작게 표시되는 물체에 해당하므로 제안 기법은 디폴트 박스 크기가 가장 작은 첫번째 특징 지도에 비지역적 특징 강화 블록을 적용하였다. 또한, 다른 특징 지도에는 RFBNet 에서 제안한 RF 블록을 적용하였다.

3. 비지역적 특징 융합

제안하는 비지역적 특징 융합은 그림 2 의 비지역적 특징 강화 블록을 통해 이루어진다. 입력된 특징 지도는 그림 2 와 같이 서로 다른 리셉티브 필드를 갖는 특징 지도로 분기된다. 이때 r 은 컨볼루션의 dilation rate 를 뜻하며, shortcut 은 단순히 입력과 동일한 채널로 출력하는 1×1 컨볼루션을 뜻한다.

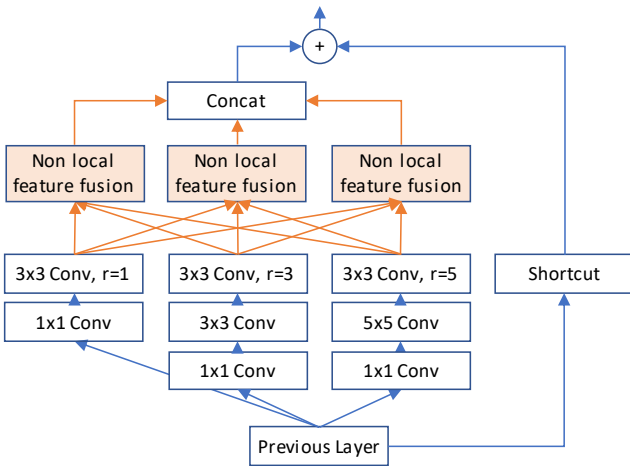


그림 2. 제안하는 비지역적 특징 강화 블록

분기된 특징 지도는 비지역적 특징 융합 단계를 거치게 되는데 이 단계는 그림 3 과 같다. 서로 다른 리셉티브 필드를 갖는 특징 지도들에 대해 한번의 1×1 컨볼루션에 의한 임베딩을 거친 후 2 차원 행렬 형태로 재배열하여 기준이 되는 특징 지도 A 에 대해 $B \cdot A^T$ 형태의 내적을 수행한다. 이는 두 특징 지도의 좌표 별 특징 간의 상관정도를 구하는 과정이다. 모든 조합에 의해 출력된 행렬은 모두 더해져 Softmax 연산을 통해 좌표 단위의 확률값으로 변환된다. 이는 기준 특징 지도를 임베딩한 다른 분기와 곱해지고 다시 한번 1×1 컨볼루션을 거친 후 기준 특징 지도와 더해진다. 이러한 과정을 통해 최종적으로는 기준 특징 지도의 각 좌표별 특징점이 여러 리셉티브 필드를 갖는 특징 지도들의 모든 좌표의 특징점들과 갖는 상관 정도가 기준 특징 지도에 반영되어 비지역적 특징이 강화된 특징 지도를 얻는다.

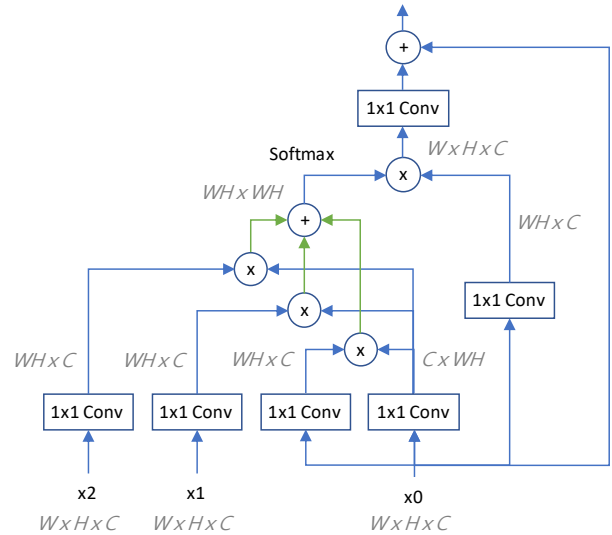


그림 3. 제안하는 비지역적 특징 융합 구조

4. 실험 결과

제안 기법을 평가하기 위해 Pascal VOC 물체 검출 공인 데이터셋을 이용하여 실험을 진행하였다. 테스트에는 VOC 2007 테스트셋을 사용하였으며 학습은 VOC 2007 과 VOC 2012 를 모두 사용하였다. Optimizer 는 SGD 를 사용하였고 learning rate 는 RFBNet 과 동일한 schedule 로 적용하였다. Mini-batch size 는 32 로 설정하였으며, GPU 는 Titan X, 딥러닝 프레임워크는 PyTorch 를 사용하였다.

실험 결과는 표 1 과 같다. 제안 기법은 mAP 기준으로 SSD 대비 3.2%, RFBNet 대비 0.5% 높은 성능을 보인다. SSD 수치는 논문에서 발췌하였으며, RFBNet 은 제안 기법과 동일한 환경에서 실험을 진행하였다. 실험 결과, 제안 기법은 기존 비교 기법 대비 향상된 성능을 보이는 것을 확인하였다.

Method	Input	Backbone	Data	mAP(%)
SSD [2]	300 x 300	VGG	07+12	77.2
RFBNet [4]	300 x 300	VGG	07+12	79.9
Proposed	300 x 300	VGG	07+12	80.4

표 1. Pascal VOC 2007 데이터셋에 대한 검출 성능 비교

5. 결론

본 논문은 비지역적 특징 융합을 통해 CNN 특징 지도를 강화하여 향상된 검출 성능을 갖는 물체 검출 기법을 제안하였다. 제안 기법은 공인 데이터셋을 통한 검출 성능 실험에서 기존 비교 기법 대비 높은 성능을 보여준다. 비지역적 특징 융합은 영상의 비지역적 문맥 정보를 강화하기 때문에 물체 검출을 비롯한 컴퓨터 비전의 다양한 분야에 적용될 수 있을 것으로 예상된다. 향후 비지역적 특징 융합의 최적화된 구조에 대한 연구를 진행할 것이다.

6. 감사의 글

본 논문은 산업통상자원부의 산업기술혁신사업으로 지원된 연구결과이며 [10073154, 인간 내면상태의 인식 및 이를 이용한 인간친화형 인간-로봇 상호작용 기술 개발], 2019 년도 정부(산업통상자원부)의 재원으로 한국산업기술진흥원의 지원을 받아 수행된 연구이며 (N0002428, 2019 년 산업전문인력역량강화사업), 한국연구재단의 지원을 받아 수행된 연구임(2016R1A2B4007353).

7. 참고문헌

- [1] S. Ren, K. He, R. Girshick and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," Advances in neural information processing systems, pp. 91-99, 2015.
- [2] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Fu and A.C. Berg, "Ssd: Single shot multibox detector," European conference on computer vision, pp. 21-37, 2016.
- [3] T. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan and S. Belongie, "Feature pyramid networks for object detection," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2117-2125, 2017.
- [4] S. Liu and D. Huang, "Receptive field block net for accurate and fast object detection," Proceedings of the European Conference on Computer Vision (ECCV), pp. 385-400, 2018.
- [5] X. Wang, R. Girshick, A. Gupta and K. He, "Non-local neural networks," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7794-7803, 2018.