

SuMul 모듈을 이용한 퓨-샷 객체 검출 방법

임동주, 정순기

경북대학교 IT대학 컴퓨터학부

dddd1226@gmail.com, skjung@knu.ac.kr

Few-Shot Object detection with SuMul module

Dong Ju Im, Soon Ki Jung

Kyungpook National University School of Computer Science and Engineering

요 약

본 논문은 퓨-샷 객체 검출 방법을 연구했다. AI의 궁극적인 목표인 인간처럼 사고하고 행동하는 것은 하위 단계인 딥러닝도 적용되어야 한다. 퓨-샷의 목표는 AI의 목표와 비슷하다. 퓨-샷 러닝은 적은 데이터를 이용하여 훈련하고 인간처럼 구분할 수 있도록 학습한다. 과적합 및 정확도를 해결하기 위해 삼 네트워크 기반으로 연구하며 학습하는 방법을 학습하기 위해 퓨-샷 학습 방식인 에피소딕 훈련 방법을 이용하여 훈련 및 테스트를 진행한다. FSOD 네트워크를 참조하고 본 논문에서 제안된 SuMul 모듈을 이용하여 정확도를 확인한다.

1. 서 론

현존하는 객체 검출 모델은 많고 다양하다. 하지만 모든 객체 검출 모델은 두 종류 분류가 가능하다. 객체의 경계 상자와 분류를 동시에 진행하는 1-Stage Detector와 객체의 경계 상자를 먼저 얻어내고 그 정보를 바탕으로 객체의 분류를 진행하는 2-Stage Detector가 있다. 두 종류의 대표적인 모델로 1-Stage Detector에는 YOLO가 있고, 2-Stage Detector에는 Faster RCNN이 있다. 모든 객체 검출 모델은 각자의 장점이 있고 단점이 존재하지만, 하나의 공통점을 가지고 있다. 모두 훈련에 필요한 데이터가 많다는 점이다.

AI의 궁극적인 목표는 인간처럼 사고하고 행동하는 것이다. 하위 단계인 딥러닝은 조금의 정보만으로 구분할 수 있는 인간처럼 사고 모델을 가지고 있다. 이러한 문제점을 해결하기 위해서 학습하는 방법을 학습하여야 한다. 이것이 퓨-샷 러닝이며 이를 이용한 객체 검출을 적은 데이터를 이용한 객체 검출 방법이라 한다.

2. 관련 연구

2.1 퓨-샷 러닝 (Few-Shot Learning)

기존 딥러닝 학습에는 수백 혹은 천개 이상의 데이터를 이용하여 학습하여야만 과 적합 문제나 정확도 그리고 딥러닝 모델의 신뢰성을 확립할 수 있다. 하지만, 이러한 방식은 인간처럼 사고하고 행동한다는 AI의 의의와는 다르다. 실제 인간은 어린아이도 사물을 구분하는 방법을 가르쳐 주지 않아도 구분해 내는 것을 알 수 있다. 퓨-샷 러닝은 이러한 문제점을 해결하고자 설계되었다.

퓨-샷 러닝은 훈련에 사용되는 데이터의 수가 한정되어있다. 1장만을 이용해 훈련하는 원-샷 러닝, 최대 100개의 클래스와 각각 25장의 데이터를 이용하는 퓨-샷 러닝, 데이터를 사용하지 않는 제로-샷 러닝이 있다. 각 훈련 마다 각기 다른 훈련 방법이 존재하며 그중 퓨-샷 러닝은 일반적으로 메타러닝 방법을 이용하여 훈련 및 테스트를 진행한다.

이전의 경험으로부터 새로운 태스크를 학습하는 방법을 학습하는 것이 메타러닝이다. 퓨-샷 러닝 모델이 완전히 새롭게 주어진 데이터에서도 잘

작동하도록 만들기 위해서는 메타 러닝이 필요하다. 퓨-샷 러닝 태스크에서는 에피소딕 훈련 (episodic training) 방식으로 메타 러닝을 시도한다. 에피소딕 훈련은 퓨-샷 태스크와 유사한 형태의 훈련 태스크를 통해 모델 스스로 학습 규칙을 도출할 수 있게 함으로써 일반화 성능을 높일 수 있게 한다.[1]

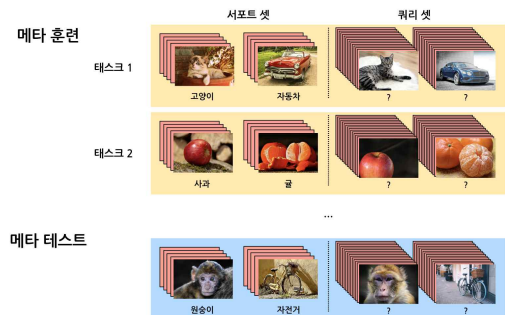


그림 1 에피소딕 훈련 방법[2]

2.2 삼 네트워크 (Siamese Neural Network)

삼쌍둥이라는 신체의 일부를 공유한다는 특징에서 착안된 네트워크로 동일한 두 네트워크의 구조가 서로 닮았으며 가중치를 공유한다.[3] 2005년 처음 구조가 발표되었으며 2015년에 Neural Network를 접목시킨 삼 네트워크가 발표되었다.

이 네트워크는 두 데이터의 특징을 추출하고 서로 비교하여 두 데이터가 같은 종류의 데이터인지 아닌지를 계산하고 결과를 도출해낸다. 적은양의 데이터를 사용하기 때문에 비교 군이 필요한 Few Shot을 이용한 네트워크에서 사용되고 있다. 본 논문에서도 해당 네트워크를 이용하여 연구를 진행한다.

3. 연구 내용

3.1 연구 방법

본 논문에서는 특징을 추출하기 위해 그림1과 같은 방법을 이용하였다.

기본적인 구조는 FSOD[4]에서 제안된 네트워크 구조를 사용한다. 비교 군으로 사용될 서포트 데이터(Support data)와 객체를 찾으려는 쿼리 데이터(Query data)가 같은 가중치를 공유하여 특징을 추출한다. 그리고 Attention RPN 모듈을 이용하여 추출된 쿼리 특징을 강화하여 얻어낼 수 있다. 그 다음 결과물을 이용하여 그림2의 SuMul 모듈에 적용하여 실험을 진행한다. 요소별 합과 곱을 이용하여 쿼리 특징에서 추출된 특징을 조금 더 강조하고 마지막에 쿼리 특징과 결과를 연결 및 FC작업과 활성화 함수 중 하나인 ReLu를 적용하여 찾고자 하는 특징만 추출해낸다.

실험은 총 3가지 항목을 진행한다. 첫 번째는 요소별 합과 쿼리 특징만을 이용하여 Concat, 두 번째는 요소별 곱과 쿼리 특징만을 이용하여 Concat, 마지막으로 SuMul를 이용하여 실험을 진행한다. 실험으로 확인할 것은 SuMul 모듈로 적용한 실험 결과가 정확도 향상에 기여를 하는지와 단일 결과에 비해 얼마나 정확도가 향상하는지가 이번 실험의 목표이다.

3.2 SuMul 모듈

퓨-샷 네트워크는 서포트를 데이터를 최대한 이용하여 쿼리 데이터에서 원하는 객체를 찾아내야 한다. 이를 위해 SuMul 모듈을 제안한다. SuMul 모듈은 Yang Xiao 논문의 Aggregator 모델을 참조하여 제작하였다.[5] 요소별 합성 합(Element Wise Sum)과 요소별 합성 곱(Element Wise Multiply)을 이용하여 객체의 특징을 극대화하기 위해 제안하였다. 쿼리 특징을 기준으로 서포트 특징을 요소별로 더하고, 곱한 결과를 원본 쿼리 특징과 Concat하여 결과를 도출해 낸다. 요소별 합은 세세한 부분의 특징을 검출할 수 있게 해주며, 요소별 곱은 강조된 부분을 더욱 강조하고 원본을 이용하여 기존에 추출된 특징에서 너무 벗어나지 않게 한다.

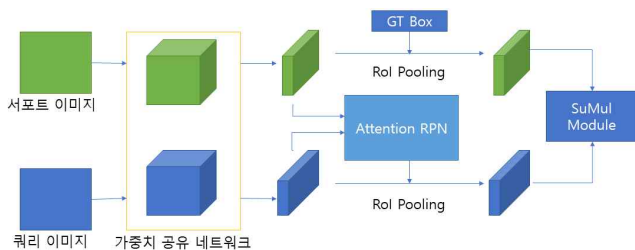


그림 2 FSOD 네트워크 구조에 SuMul 모듈을 합성

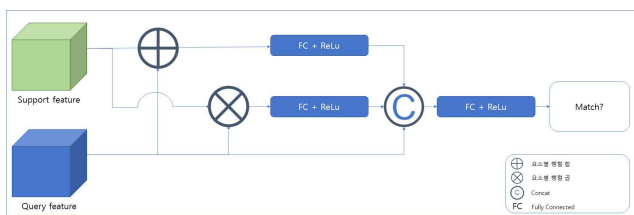


그림 3 제안된 SuMul 모듈

3.3 연구 결과

본 실험에서는 COCO 데이터셋에서 20개의 클래스와 그에 해당하는 이미지 5장씩 총 100장을 이용하여 훈련을 진행하고, 테스트 하려는 이미지에 해당하는 클래스를 가진 이미지를 이용하여 테스트를 진행했다. 테스트에 대한 정확도 척도는 AP(Average Precision)을 이용하였으며 결과는 표1과 같다.

3.4 연구 분석

테스트 결과 단순히 요소별 합성 합이나 곱을 따로 이용하는 것 보다 두

결과를 합성하여 이용하는 것이 결과가 좋게 나오는 것을 확인할 수 있다. 하지만, 기존 네트워크의 mAP는 11인 것을 생각하면 좋은 결과라 할 수 없다.

SuMul 모듈은 서포트와 쿼리의 요소별 합과 요소별 곱을 이용하기 위해 서포트 특징을 GAP(Global Average Pooling)을 이용하여 채널별 평균을 구한 뒤 해당 값을 쿼리에 적용하는 방식으로 진행하였다. 기존 네트워크의 경우 본 논문의 SuMul과는 다르게 GAP 없이 각 픽셀에 해당하는 결과들의 관계에 따라서 객체 검출을 진행하였기 때문에 더 나은 결과를 도출해 낸 것이다.

	mAP	AP50	AP75
Sum	5.93	13.92	4.12
Mul	6.80	15.76	5.01
Sum + Mul	7.59	16.41	6.37

표 1 테스트 결과

4. 결론 및 향후 연구

본 논문에서는 적은 데이터를 이용한 객체 검출 모델에서 특징을 더욱 극대화하기 위해 SuMul 모듈을 제안하였다. 제안된 모듈은 추출된 서포트 특징을 최대한으로 이용하기 위해 요소별 합과 곱을 이용하였고 추출된 원본 특징도 이용하여 원본에서 최대한 벗어나지 않게 만들었다. 기존 네트워크의 mAP가 11인 것을 생각한다면, 결코 좋은 결과라고 할 수 없다. SuMul 모듈의 서포트의 이용 방식을 GAP로 채널 별 평균을 구하는 것이 아닌, 다양한 네트워크를 참조 및 연구하여 더 나은 방향으로 개선할 것이다.

향후 연구는 추출된 특징을 이용하여 객체 검출을 극대화하는 방법이 아닌, 특징 추출부터 다시 설계하여 서포트를 최대한으로 이용하는 방법을 연구할 예정이다.

ACKNOWLEDGMENT

이 연구는 한국연구재단(NRF-2019R1A2C1010786) 사업의 지원을 받아 수행된 연구임

참 고 문 헌

- [1] Li, Da, et al. "Episodic training for domain generalization." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019.
- [2] 이수경, "1.퓨-샷 러닝(few-shot learning) 연구 동향을 소개합니다.," kakaobrain, 2019.11.06., (<https://www.kakaobrain.com/blog/106>).
- [3] Koch, Gregory, Richard Zemel, and Ruslan Salakhutdinov. "Siamese neural networks for one-shot image recognition." ICML deep learning workshop. Vol. 2. 2015.
- [4] Fan, Qi, et al. "Few-shot object detection with attention-RPN and multi-relation detector." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020.
- [5] Xiao, Yang, and Renaud Marlet. "Few-shot object detection and viewpoint estimation for objects in the wild." European Conference on Computer Vision. Springer, Cham, 2020.