

이종 특징 맵을 이용한 교차 어텐션 기반 소형 객체 검출 기법

Tiny Object Detection Method Based on Cross Attention of Heterogeneous Feature Maps

권승찬, 김도연, 임규일, 한영준*, 김성신*

Seungchan Kwon, Doeon Kim, Gyuil Lim, Youngjun Han, Seongshin Kim

송실대학교 지능시스템학과

E-mail: {kwonsc36, ilsln205, edu9059}@soongsil.ac.kr

요 약

컴퓨터 비전 분야에서 객체 탐지는 다양한 응용 분야에서 활용되고 있는 핵심적인 기술이다. 그러나 기존 CNN 기반 객체 탐지 방법들은 일반적인 크기의 객체에 대해서 우수한 성능을 보이지만, 소형 크기의 객체(Tiny object) 탐지에서는 상대적으로 저조한 성능을 보인다. 이러한 문제는 Deep network 내에서 특징 맵의 Down sampling 과정을 거치면서 소형 객체의 특징 정보가 소실되기 때문이다. 본 연구에서는 이러한 정보 손실 문제를 완화하기 위해 분할 신경망(Segmentation network)의 특징 맵을 이용한 교차 어텐션 기반 소형 객체 탐지 방법을 제안한다. 그 결과 AI-TOD 데이터셋에서 CNN 기반 모델들 기준 SOTA 를 달성함으로써 본 논문에서 제안하는 기법의 우수성을 확인하였다.

키워드 : 객체 탐지, Tiny object, Feature map fusion, Multi-task attention

1. 서 론

딥러닝이 발전하면서 객체 탐지에 대해서도 많은 연구가 진행되었다. Tiny object detection은 객체 탐지의 하위 분야로, 일반적인 객체 탐지 모델은 객체가 64 Pixel보다 큰 반면 Tiny object는 8-16 Pixel을 탐지하는 분야다. 이는 위성 및 항공 영상 분석, 의료 영상 분석, 드론 영상, 생물학 및 생태학 연구 등의 다양한 시나리오에서 이론적 및 실용적 중요성을 가지고 있다.

기존 객체 탐지 모델들로 소형 객체를 탐지하기 어려운 이유는 작은 객체의 제한적이고 왜곡된 정보로부터 적절한 표현을 학습하는 본질적인 어려움이라고 볼 수 있다.

따라서 본 논문에서는 해당 문제들을 해결하기 위해 첫째, 세그멘테이션 모델이 가지고 있는 높은 공간적 표현력을 객체 탐지 모델과 결합하여 공간적 표현력을 증가시키고자 했다. 둘째, 세그멘테이션 모델의 특징 맵과 객체 탐지 모델의 특징 맵을 효과적으로 결합할 수 있는 Cross Attention Heterogeneous Feature map(CAHF) 방법을 제안한다.

AI-TOD 데이터 셋에서 CNN 기반 모델 중 State-Of-The-Art(SOTA)를 달성하면서 SOTA 트랜스포머 모델과 비슷한 수준의 mAP(0.5-0.95)를 달성하였다. 이를 통해, 제안하는 모델이 작은 객체 탐지에 적합한 구조라는 것을 보여주었다.

2. 이론적 배경

2.1 Faster R-CNN

Faster R-CNN 은 객체 탐지 분야의 딥러닝 모델로, RPN 과 Fast R-CNN 으로 구성된다. RPN 은 이미지에서 객체가 존재할 가능성이 있는 영역을 제안하며, Fast R-CNN 은 이를 기반으로 클래스 분류와 박스 회귀를 수행한다. CNN 기반 영역 제안을 통해 빠른 속도와 정확도를 동시에 달성하며 객체 탐지의 대표 모델로 자리 잡았다.

2.2 Unet plus plus(Unet++)

U-Net 은 인코더-디코더 구조를 통해 이미지의 특징을 추출한 후 복원하는 세그멘테이션 모델이다. U-Net++ 은 U-Net 에 여러 개의 Skip connection 을 추가해서 모든 단계에서의 의미 정보를 혼합하여 Semantic-gap 을 줄인다. 이를 통해 특징 맵의 지역적, 공간적 정보가 최대한 유지되도록 한다.

2.3 Dual path Attention Module(DAM)

DAM 은 Remote-sensing 분야에서 사용되는 소형 객체 정보를 효과적으로 추출하고, 배경 노이즈를 억제하기 위한 모듈이다. DAM 은 Spatial attention 브랜치를 통해 중요한 공간 정보를 추출하고 Channel attention 브랜치를 통해 쓸모없는 채널 정보를 억제한다. 이를 통해 중요한 정보는 강조된 특징 맵을 얻게 된다.

감사의 글 : 본 연구는 과학기술정보통신부 및 정보통신기획평가원의 SW 중심대학사업의 지원을 받아 수행되었음(2024-0-00071)

3. 제안 방법

약 8픽셀 정도의 소형 객체에 대한 정보는 네트워크를 통과하면서 점차 손실되고, 컨볼루션 과정에서 배경이나 다른 객체의 특징에 의해 영향을 받기 쉽다. 따라서 네트워크가 후속 작업에 필요한 정보를 효과적으로 포착하기 어려워진다. 이를 해결하기 위해 Spatial attention과 Channel Focus Attention으로 이루어진 CAHF 기법을 제안하여 소형 객체의 공간적 정보를 보존하고, 유의미한 채널 특징을 선별한다.

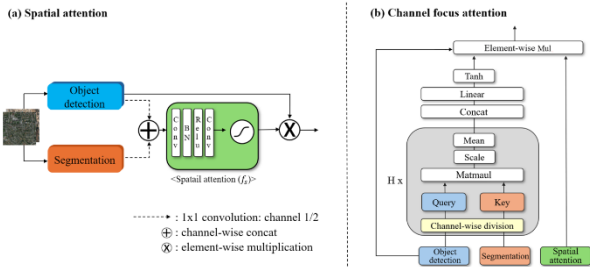


그림 1. (a) Spatial attention (b) Channel focus attention

3.1. Spatial attention

Spatial attention ($Attention_s$)은 세그멘테이션 인코더의 특징 맵(F_s)과 객체 탐지 백본의 특징 맵(F_o)을 융합하여 소형 객체의 공간 정보를 추출하는 모듈이다. 두 가지 태스크의 특징 맵을 융합하는 방법은 다음과 같다.

$$Attention_s(F_o, F_s) = f_s\{Conv'(F_o) \oplus Conv'(F_s)\} \quad Eq(1)$$

위 수식(Eq(1))의 $Conv'$ 은 커널 크기가 1 인 Convolution layer로 채널 수를 1/2로 줄이는 연산이다. Spatial attention은 그림 1. (a)에서 자세하게 설명한다.

Spatial attention은 기존 객체 탐지 모델의 특징 맵의 배경 노이즈를 줄이고, 탐지하고자 하는 객체에 대한 공간적 정보를 보완해 준다. 이는 복잡한 배경 속에서 소형 객체를 정확하게 분할할 수 있도록 한다.

3.2. Channel focus attention

Channel focus attention ($Attention_{CF}$)은 그림 1. (b)와 같이 F_o 와 F_s 간 관계를 학습하여 각 채널의 유의미한 특징을 매핑하는 방법이다. 이를 수식으로 나타내면 Eq(2)와 같다.

$$Attention_{CF}(F_o, F_s) = \text{Linear} \left(\text{Mean} \left(\frac{F_o F_s^T}{\sqrt{d_o}} \right) \right) \quad Eq(2)$$

Channel focus attention에서는 입력 특징 맵 간 내적을 통해 각 채널들의 상관 행렬을 생성한다. 이때 행렬의 값이 커지는 것을 방지하기 위해 특징 맵의 크기(d_o)로 스케일링한다. 그 후 행을 기준으로 평균을 구해 상관 행렬을 1차원 벡터로 변환한다. 이 과정에서 채널별로 모델이 집중해야 할 정보를 추출한다.

$$CAHF(F_o, F_s) = Attention_s(F_o, F_s) * Attention_{CF}(F_o, F_s) * F_o \quad Eq(3)$$

최종적으로 $Attention_s$ 와 $Attention_{CF}$ 를 통해 두 가지 태스크의 융합된 특징 맵의 공간적, 채널적 정보를 효과적으로 어텐션 한다. 이는 배경의 노이즈를 줄이면서 채널의

유의미한 정보에 집중하게 하여 모델의 소형 객체 탐지 능력이 향상되게 한다. 특히 복잡한 배경에서도 소형 객체를 더 정확하게 탐지할 수 있게 한다.

4. 실험 및 결과

본 논문에서 실험은 AI-TOD 데이터 셋을 사용하였다. AI-TOD는 소형 객체 탐지 연구를 위해 만들어진 항공 위성 이미지로, 객체의 크기는 약 12.8 픽셀 정도이며 배, 비행기 등을 포함한 9 개의 클래스로 이루어져 있다.

성능 측정을 위한 지표는 mAP(0.5)와 mAP(0.5-0.95)를 사용하였다. 비교한 모델로는 Faster R-CNN 과 CAHF, CNN 기반 SOTA 성능을 비교하였다. 옵티마이저는 Adam을 사용하였고, 학습률은 $3e-4$ 로 설정하였다. GPU는 NVIDIA RTX 3090을 사용하였다.

Method	mAP	mAP(0.5-0.95)	Epoch
Faster R-CNN	40.03%	18.72%	20
CAHF	43.8%	23.1%	20
DetectoRS	32.8%	14.8%	12

표 1. AI-TOD 데이터셋에 대한 모델 성능 비교

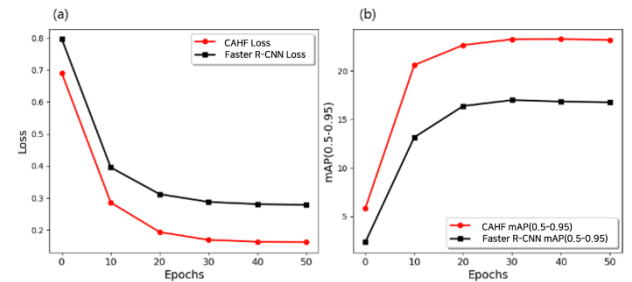


그림 3. (a) Faster R-CNN과 CAHF의 loss, (b) mAP(0.5-0.95) 비교

먼저 표 1을 보면 Faster R-CNN 모델과 비교하여 mAP(0.5)와 mAP(0.5-0.95)가 각각 3.77%, 4.38%의 성능이 향상되었고 CNN 기반 SOTA 객체 탐지 모델인 DetectoRS와 비교했을 때 각각 11%, 8.3% 증가했다. 그림 3.에서 Loss의 감소 양상과 mAP(0.5-0.95)의 증가 차이를 확인할 수 있다.

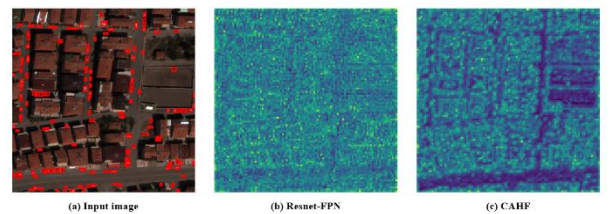


그림 4. CAHF 적용 전후를 비교한 특징 맵 시각화 (a) 입력 이미지 (b) 적용 전 특징 맵 (c) 적용 후 특징 맵

결론

본 논문은 객체 탐지와 세그멘테이션의 특징 맵을 동시에 활용하는 Cross Attention Heterogeneous Feature map(CAHF)을 사용하여 소형 객체 탐지 능력을 극대화하였다. CAHF는 Spatial attention과 Channel focus attention이라는 2 가지 모듈로 구성되어 있다. Spatial attention은 배경 노이즈를 제거하고 Channel attention은 두 가지 태스크에 대한 특징 맵의 상관관계를

통해 유의미한 채널 가중치를 만들어낸다. 결과적으로 CAHF 을 이용해 소형 객체의 제한적인 정보에 대한 표현력을 보충하여 객체 탐지 성능을 높일 수 있었다.

위 방법의 효과를 입증하기 위해 항공 이미지 데이터 셋 AI-TOD 에서 실험한 결과, CNN 기반 SOTA 모델보다 우수한 성능을 달성하였다. 또한 특징 맵을 시각화하여 비교함으로써 일반 객체 탐지 모델에 비해 배경과 객체가 명확하게 분리됨을 정성적으로 확인하였다. 이를 통해 본 연구는 Tiny object detection 분야에서 세그멘테이션 모델의 활용 가능성을 보여주었고, CNN 기반 모델의 성능을 다시 한번 증명하였다.

참 고 문 헌

- [1] J. Wang, W. Yang, H. Guo, R. Zhang and G.-S. Xia, "Tiny Object Detection in Aerial Images," 2020 25th International Conference on Pattern Recognition (ICPR), Milan, Italy, 2021, pp. 3791-3798, doi: 10.1109/ICPR48806.2021.9413340.
- [2] Shaoqing Ren, Kaiming He, Ross Girshick, Jian Sun, 2015, Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks
- [3] Zhou, Zongwei et al. "UNet++: A Nested U-Net Architecture for Medical Image Segmentation." Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support : 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, held in conjunction with MICCAI 2018, Granada, Spain, S... 11045 (2018): 3-11 .
- [4] Lu, Xiaocong & Ji, Jian & Xing, Zhiqi & Miao, Qiguang. (2021). Attention and Feature Fusion SSD for Remote Sensing Object Detection. IEEE Transactions on Instrumentation and Measurement. PP. 1-1. 10.1109/TIM.2021.3052575.