

위성 이미지 객체 탐지에서 앵커 박스의 개수가 객체 탐지 성능에 미치는 영향

오인호¹, 임종민², 차수빈³, 유여은⁴, 김광수²

¹성균관대학교 시스템경영공학과

²성균관대학교 소프트웨어학과

³성균관대학교 인공지능학과

⁴성균관대학교 인공지능융합전공

inho2007@skku.edu, jm.lim@skku.edu, soobin99@skku.edu, annie2675@g.skku.edu, kim.kwangsus@skku.edu

The Impact of the Number of Anchor Boxes on Object Detection Performance in Satellite Image Object Detection

Inho Oh¹, Jongmin Lim², Soobin Cha³, Yoeun Yoo⁴, Kwangsus Kim²

¹Department of System Management Engineering, Sungkyunkwan University

²Department of Computer Science and Engineering, Sungkyunkwan University

³Department of Artificial Intelligence, Sungkyunkwan University

⁴Applied Artificial Intelligence, Sungkyunkwan University

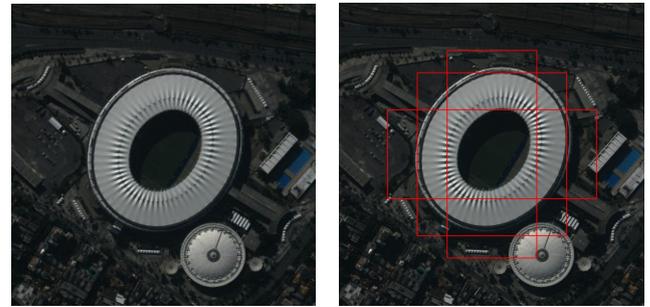
요약

위성 이미지가 가지고 있는 다양한 정보는 현재 여러 사회문제의 진단과 해결을 위해 사용되고 있다. 위성 정보 서비스 분야 시장 규모는 매년 크게 증가하고 있으며, 위성 이미지의 정보를 효율적으로 분석, 활용하기 위해서는 딥러닝 모델이 많이 활용된다. 한 이미지 내에 다양한 크기의 객체가 존재하는 위성 이미지의 특성 상, 위성 이미지 객체 탐지를 위해서는 훈련 단계에서의 앵커 박스 설정이 중요하다. 본 논문에서는 훈련 단계에서의 앵커 박스의 개수에 따른 모델의 성능 지표를 비교한다. 이를 통해 앵커 박스의 개수가 증가할수록 모델의 성능이 향상하는 것을 알 수 있다.

1. 서론

위성 이미지에서 얻을 수 있는 정보는 현재 다양한 사회문제의 진단과 해결을 위해 사용되고 있다.[1] 예를 들면, 위성 이미지는 항공 이미지와는 달리 접근 불가 지역을 아무런 외교적 마찰 없이 촬영할 수 있어, 타국의 핵 시설이나 군사 활동등을 감시할 수 있고, 매우 넓은 지역을 짧은 시간내에 촬영할 수 있어 지진과 같은 대형 재난사고시 신속히 상황을 파악하는데 사용될 수 있다. 또한, 위성 이미지는 전 지구를 주기적으로 촬영할 수 있어 전 세계에서 일어나는 변화를 관찰하는 데 유용하다. 위성 정보 서비스 분야 시장 규모는 매년 크게 증가하고 있으며 다양한 공공, 민간 서비스 수요에 대응하기 위해 위성 개발 또한 확대되고 있다.[2]

이러한 위성 이미지를 활용한 객체 탐지는 다양한 응용 분야에서 중요한 역할을 한다. 예를 들어, 교통 모니터링, 군사적 감시, 환경 변화 분석 등에서 위성 이미지는 고해상도의 데이터로 다양한 객체를 식별하고 분석할 수 있는 중요한 자료이다. 그러나 위성 이미지의 특성상 실제 객체의 크기가 이미지에 반영되어 클래스 간의 크기 불균형이 심각하다.[3] 즉, 한 이미지 내에 다양한 크기의 객체들이 존재한다. 그림 1 (가)에서 큰 크기의 객체인 스타디움과 중간 크기의 객체인 수영장, 작은 크기의 객체인 소형승용차가 한 이미지 내에 존재하는 것을 확인할 수 있다. 다양한 크기의 객체들을 동시에 탐지하기 위해서는 훈련 단계에서 여러 종류의 앵커 박스로 학습하는 과정이 필요하다. 앵커 박스는 다양한 크기와 비율의 객체를 탐지하기 위해서 사전에



(가)

(나)

그림 1. 위성 이미지와 위성 이미지에 대한 앵커 박스

정의된 바운딩 박스이다. 즉, 더 많은 앵커 박스를 생성해서 학습하면, 성능이 향상된다. 이 가설을 실험을 통해 검증해 보겠다. 그림 1 (나)는 스타디움을 검출하기 위한 앵커 박스 예시이다.

본 논문에서는 훈련 단계에서의 앵커 박스의 개수에 따른 객체 탐지 모델의 성능 지표를 비교하고, 앵커 박스의 개수가 증가할수록, 객체 탐지의 성능이 향상되는 것을 확인한다.

2. 본론

2.1 데이터

본 논문에서는 한국지능정보사회진흥원이 운영하는 AI 통합 플랫폼인 AI-Hub 에서 제공하는 위성영상 객체 판독 AI 데이터셋 중 관심 객체 검출 데이터셋을

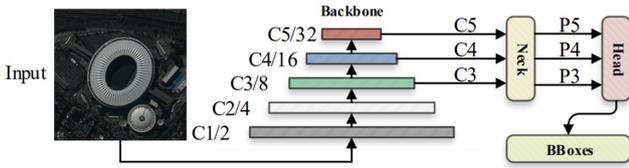


그림 2. 위성 이미지를 입력으로 사용하는 YOLOv5 구조

활용한다.[4] 관심 객체 검출 데이터셋은 인공위성 아리랑 3 호/3A 호에서 수집한 고해상도 광학 영상으로 구축된 이미지 데이터로 훈련용 이미지 1746 개와 검증용 이미지 219 개로 구성되어 있다. 각 이미지는 1024×1024 픽셀의 해상도를 가진다. 이미지 내의 관심 객체는 총 20 종의 클래스를 가지고, 데이터셋에는 총 50 만개 이상의 객체가 존재한다.

2.2 모델 및 실험

위성 이미지 객체 탐지를 위해서 본 논문에서는 YOLO v5[5]을 실험 모델로 선정하였다. YOLO 시리즈는 객체 식별과 바운딩 박스 위치 선정을 동시에 수행하는 대표적인 One-stage 모델이다.[6] 그 중 YOLOv5 은 Backbone 에서 이미지의 특징을 추출하여 피쳐맵 (C3, C4, C5)을 생성한다. Neck 부분에서는 FPN[7]의 up-sampling 을 통해 상위 계층의 정보를 하위 계층으로 전달하여 새로운 피쳐맵(P3, P4, P5)을 생성하고, PAN[8]을 통해 하위 계층의 정보를 상위 계층으로 전달하여 생성된 피쳐맵을 강화한다.[9] 그리고 마지막 Head 부분에서 3 개의 출력 피쳐맵에 대해서 최종 예측을 수행하는데, 각 피쳐맵 당 3 개의 앵커 박스를 생성하여 총 9 개의 앵커 박스를 생성한다. 생성된 앵커 박스는 Autoanchor 기능을 통해 데이터셋에 맞게 앵커 박스의 크기를 최적화한다. 그림 2 에 위성 이미지를 입력으로 사용하는 YOLOv5 의 간단한 구조가 나타나 있다.

본 논문의 실험에서는 각 피쳐맵 당 생성하는 앵커 박스의 수를 조절하여, 전체 앵커 박스의 개수가 객체 탐지에 미치는 영향을 알아보려고 한다. 훈련 단계에서 사전 학습된 YOLOv5 모델로 훈련을 진행하고, 검증 단계에서 성능 지표 AP, AP₅₀ 을 도출한다. 이때, 피쳐맵 당 생성하는 앵커 박스의 수가 1 개, 2 개, 3 개, 즉, 생성하는 총 앵커 박스의 수가 3 개, 6 개, 9 개인 경우를 비교하기 위해 총 3 번의 실험을 진행한다.

2.3 결과

표 1 은 훈련 단계에서의 앵커 박스 개수에 따른 검증용 데이터셋에 대한 모델의 성능 지표 결과이다. 앵커 박스의 개수가 증가할수록, 성능 지표 값이 좋아지는 것을 알 수 있다. 따라서, 더 정확한 객체 탐지를 위해서는 훈련 단계에서 더 많은 앵커 박스가 필요하다. 이를 통해 가설이 검증됐다.

앵커 박스 개수	AP	AP ₅₀	GFLOPs	FPS
3 개	23.0	41.8	47.9	14
6 개	29.6	51.7	48.0	11
9 개	35.1	57.5	48.1	10

표 1. 앵커 박스 개수에 따른 성능 지표 결과

3. 결론

본 논문에서는 위성 이미지 객체 탐지를 할 때, 훈련 단계에서의 앵커 박스의 개수를 증가시키면, 모델의 성능이 좋아지는 것을 확인하였다. 하지만, 표 1 을 보면

훈련 단계에서의 앵커 박스의 개수가 증가하면 GFLOPs 가 증가하고, FPS 가 감소한다. 즉, 추가적인 계산을 유발하여 모델의 연산량이 증가하고, 추론 속도에 영향을 미치게 된다. 따라서 훈련 단계에서 적절한 앵커 박스의 개수를 설정하는 것이 중요하다. 후속 연구로, 데이터셋의 특징과 모델의 목적에 부합하는 최적의 앵커 박스 개수를 도출하는 방법론에 대해 연구하고자 한다.

ACKNOWLEDGMENT

이 연구는 과학기술정보통신부의 재원으로 한국지능정보사회진흥원의 지원을 받아 구축된 "위성영상 객체탐지"을 활용하여 수행된 연구입니다. 본 연구에 활용된 데이터는 AI 허브(aihub.or.kr)에서 다운로드 받으실 수 있습니다.

참고 문헌

- [1] Jae-San Kim, & Soo-Jin Lee (2024). Improvement of Object Detection Performance in Satellite Images using Image Segmentation and Up-scaling. *The Journal of Korean Institute of Information Technology*, 22(6), 21-29, 10.14801/jkiit.2024.22.6.21
- [2] KIM SE HYOUNG, Kang Ju Young, & CHAE JUNG WOO (2022). Research Trends and Datasets Review using Satellite Image. *스마트미디어저널*, 11(1), 17-30.
- [3] Joo-Chan Park, Sung-Bin Son, Seon-Hoon Lee, Jun-Uk Jung, Yong-Jun Park, & Heung-Seon Oh (2021). Deep Ensemble based Object Detection from Aerial Images. *Journal of Institute of Control, Robotics and Systems*, 27(12), 944-952, 10.5302/J.ICROS.2021.21.0135
- [4] <https://aihub.or.kr/>
- [5] Ultralytics. YOLOv5. GitHub repository, <https://github.com/ultralytics/yolov5>, 2020.
- [6] Redmon, J. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*.
- [7] Lin, T. Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. (2017). Feature pyramid networks for object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2117-2125).
- [8] Liu, S., Qi, L., Qin, H., Shi, J., & Jia, J. (2018). Path aggregation network for instance segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 8759-8768).
- [9] Liu, H., Sun, F., Gu, J., & Deng, L. (2022). SF-YOLOv5: A Lightweight Small Object Detection Algorithm Based on Improved Feature Fusion Mode. *Sensors (Basel, Switzerland)*, 22(15), 5817.