드론 데이터를 활용한 딥러닝 기반 다중 객체 추적

주상현 남해운
한양대학교

{tkdgus014, hnam}@hanyang.ac.kr

Multi-Object Tracking based on Deep Learning using Drone Data

Sanghyun Ju, Haewoon Nam
Hanyang Univ.

요 약

최근 다중 객체 추적 기술이 다양한 분야에서 사용되고, 드론 기술이 빠르게 발달함에 따라 딥러닝을 활용한 연구와 개발이 활발하게 이루어지고 있다. 드론에서 촬영된 영상은 공중에서의 촬영으로 인해 작은 크기의 객체를 효과적으로 탐지해야한다. 본 논문에서는 이를 위해사람, 자동차, 자전거 등 다양한 크기의 객체를 효과적으로 탐지할 수 있는 딥러닝 기반 모델을 선정하고 이에 적합한 학습 전략을 제시한다. 이는 드론을 활용한 다중 객체 추적 기술이 감시, 재난 대응, 교통 관리 등 다양한 실용적 분야에서 적용할 수 있음을 보여줄 수 있을 것이다.

Ⅰ. 서 론

최근 다중 객체 추적 기술은 군사, 상업, 보안, 재난 대응 등과 같은 다양한 분야에서 사용되고 있다. 특히, 드론 기술의 급속한 발달로 데이터셋 취득이 가능해졌고, 이를 이용한 딥러닝 연구가 활발하게 이루어지고 있다. 드론을 활용한 다중 객체 추적은 기동성이 뛰어나고, 다양한 환경에서 사용될 수 있어 접근하기 어려운 지역에서도 효과적으로 객체를 추적할 수 있다. 또한, 공중에서 촬영하는 드론은 객체의 움직임을 지속적으로 추적할 수 있으며, 비교적 저렴한 비용으로 운용할 수 있는 장점이 있다.

그러나 드론으로 촬영한 비디오의 객체들은 종종 이미지 전체에 비해 상대적으로 작은 사이즈를 가지는 문제를 가지고 있다. 이러한 작은 객체들은 제한된 픽셀 수와 낮은 해상도로 인해 정보가 부족하여, 정확한 식별과 분류에 어려움을 가진다. 이는 특히 드론 영상에서의 객체 탐지와 추적 알고리즘의 성능에 영향을 미칠 수 있다.

본 논문은 드론으로 촬영한 고해상도 데이터(VisDrone Dataset)을 활용하여, 작은 객체를 효과적으로 탐지할 수 있는 적절한 딥러닝 모델을 선정한다. 그리고 데이터 증강 기법을 활용하여 적합한 학습전략을 제안한다.

Ⅱ. 본론

1. 알고리즘 구조

본 논문에서 보여주는 방법은 객체 탐지를 위해 RetinaNet[1]을 사용하고, 추적을 위해 ByteTrack[2]을 활용한다. RetinaNet은 빠른 속도와 높은 정확도를 제공하는 One-Stage 탐지 모델로 그림1과 같이 나타낼 수 있다. 이는 다양한 크기의 객체를 효과적으로

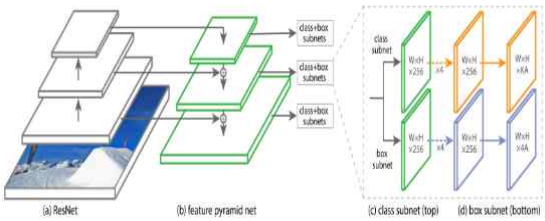

그림1. RetinaNet 구조

탐지하기 위해 특징 피라미드 네트워크를 사용한다. 백본 네트워크는 ResNet-50을 사용했으며, 손실함수로 Focal Loss 함수를 사용한다. 이 함수는 기존의 Cross-Entropy 함수의 변형으로, 쉽게 탐지되는 객체보다 어렵게 탐지되는 객체에 더 많은 가중치를 부여함으로써 불균형한 클래스 분포 문제를 해결하고, 작은 객체 탐지의 정확도를 높이는데 기여한다.

추적 모델인 ByteTrack은 탐지된 객체들 간의 매칭을 통해 이전 프레임과 현재 프레임 사이의 연속성을 추적한다. ByteTrack의 특징은 높은 Confidence Score를 가진 객체뿐만 아니라 낮은 Confidence Score를 가진 객체에 대해서도 매칭을 수행한다는 점이다. 이는 가려진 객체들의 추적을 가능하게 한다. 추적 모델은 탐지 결과를 기반으로 추적하기 때문에 추적 모델의 성능은 탐지 모델의 성능에 크게 의존한다. 따라서 본 논문은 RetinaNet의 활용과 학습 전략을 통해 탐지 성능을 높여 추적 성능을 높였다.

2. 탐지 모델 학습

학습에 사용된 데이터는 공개 데이터셋인 VisDrone Dataset의 다중 객체 추적(MOT) 데이터셋을 사용했다. VisDrone MOT Dataset은 드론을 활용하여 공중에서 촬영한 비디오 데이터로 구성되어 있으며, 실제 도시와

그림 2. RetinaNet+MixUp(좌)와 Ground Truth(우) 비교

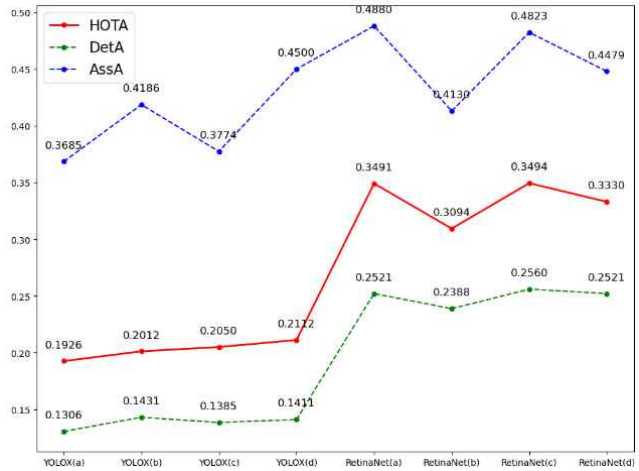

그림 3. YOLX와 RetinaNet 성능 비교 (a는 기본, b는 Mosaic, c는 MixUp, d는 Mosaic+Mixup을 적용시킨 모델이다)

농촌 지역 등 다양한 시나리오를 포함한다. 학습용 데이터는 총 56개의 비디오로 구성되어 있으며, 총 24,918장의 이미지로 이루어져 있다. 또한, 성능 평가를 위한 테스트 데이터셋에는 17개의 비디오가 포함되어 있으며, 총 6,635장의 이미지로 구성된다. 클래스는 pedestrian, people, car, truck 등 총 10개의 클래스로 구분되어 있다. 그 중 pedestrian, car, van, truck, bus의 클래스에 집중하여 학습을 진행하였다. 이를 통해 실상황의 복잡한 환경에서 발생하는 다양한 시나리오에 대해 강인한 추적 성능을 가질 수 있다.

드론 기반 비디오 데이터는 공중에서 촬영된 영상으로, 지상의 객체들이 상대적으로 작게 나타나는 특성을 가진다. 이러한 작은 객체들을 효과적으로 탐지하는 것이 본 논문의 도전 과제이다. 이를 위해 Mixup[3]과 Mosaic[4] 두 가지 데이터 증강 기법을 통해 탐지 성능을 향상시켰다. MixUp은 다양한 이미지를 선형적으로 혼합해 새로운 학습 데이터를 생성하였다. Mosaic은 여러 이미지를 하나의 이미지로 붙여 작은 객체들이 포함된 새로운 장면을 생성하였다. 이를 통해 작은 객체들을 효과적으로 학습할 수 있도록 하였다.

3. 실험 결과

성능 평가를 위해 DetA, AssA, HOTA 라는 세 가지 주요 지표를 사용했다. DetA는 프레임 내에서 객체가 얼마나 정확하게 탐지되었는지를 나타내며, 실제 객체와 탐지된 객체 간의 일치도를 평가한다. AssA는 프레임 간 객체 매칭의 정확도를 측정하며, 추적된 객체의 경로가 시간에 걸쳐 얼마나 일관되게 유지되는지를 나타낸다. HOTA는 DetA와 AssA의 조화평균으로 전반적인 성능을 평가한다.

그림2는 MixUp데이터 증강 기법을 적용한 RetinaNet의 추론 결과를 보여준다. Ground Truth와 비교하여 객체와 카메라 사이의 거리가 멀어질수록 객체가 작아져 일부 객체를 놓치는 현상이 관찰되었다. 또한, 빠른 화면 전환 시에도 일부 객체의 추적이 누락되는 현상이 발견되었다.

성능 비교를 위해 기존에 실시간 탐지 모델로 사용하던 YOLOX[6] 모델과 결과를 비교하였다. YOLOX는 높은 정확도와 함께 실시간 성능이 좋은 모델이다. 그림3을 보게되면 RetinaNet은 전반적으로 YOLOX에 비해 높은 성능을 보여주었다. 특히 Mosaic과 MixUp을 사용했을 때에도, 동일한 데이터 증강 기법을 사용한 YOLOX와 비교하여 좋은 성능을 보여주었다.

Ⅲ. 결론

본 논문에서는 드론 데이터셋을 활용한 딥러닝 기반 다중 객체 인식을 보여준다. 탐지 모델로 RetinaNet을 사용하고 데이터 증강 기법을 통해 작은 객체 탐지에 초점을 맞추어 기존의 YOLOX보다 높은 성능을 보여주고 조금의 성능 향상을 달성하였다. 이러한 접근 방식은 드론 영상에서의 객체 탐지와 추적 기술에 발전을 기여하며, 향후 보안, 감시, 재난 대응 등 다양한 응용 분야에 중요한 영향을 미칠 수 있는 잠재력을 가지고 있다.

ACKNOWLEDGMENT

이 논문은 정부(과학기술정보통신부)의 재원으로
한국연구재단의 지원을 받아 수행된 연구임
(No.2022R1A2C1011862)

참 고 문 헌

[1] Lin, T. Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2017). Focal loss for dense object detection. In Proceedings of the IEEE international conference on computer vision (pp. 2980-2988).

[2] Zhang, Y., Sun, P., Jiang, Y., Yu, D., Yuan, Z., Luo, P., ... & Wang, X. Bytetrack: Multi-object tracking by associating every detection box. arXiv 2021. arXiv preprint arXiv:2110.06864.

[3] Bewley, A., Ge, Z., Ott, L., Ramos, F., & Upcroft, B. (2016, September). Simple online and realtime tracking. In 2016 IEEE international conference on image processing (ICIP) (pp. 3464-3468). IEEE.

[4] Zhang, H., Cisse, M., Dauphin, Y. N., & Lopez-Paz, D. (2017). mixup: Beyond empirical risk minimization. arXiv preprint arXiv:1710.09412.

[5] Wei, Z., Duan, C., Song, X., Tian, Y., & Wang, H. (2020). Amrnet: Chips augmentation in aerial images object detection. arXiv preprint arXiv:2009.07168.

[6] Ge, Z., Liu, S., Wang, F., Li, Z., & Sun, J. (2021). Yolox: Exceeding yolo series in 2021. arXiv preprint arXiv:2107.08430.