

무인 항공기를 이용한 밀집영역 자동차 탐지

서창진
상명대학교 정보보안공학과

Vehicle Detection in Dense Area Using UAV Aerial Images

Chang-Jin Seo

Department of Information Security Engineering, Sangmyung University

요약 본 논문은 최근 물체탐지 분야에서 실시간 물체 탐지 알고리즘으로 주목을 받고 있는 YOLOv2(You Only Look Once) 알고리즘을 이용하여 밀집 영역에 주차되어 있는 자동차 탐지 방법을 제안한다. YOLO의 컨볼루션 네트워크는 전체 이미지에서 한 번의 평가를 통해서 직접적으로 경계박스를 예측하고 각 클래스의 확률을 계산하고 물체 탐지 과정이 단일 네트워크이기 때문에 탐지 성능이 최적화되며 빠르다는 장점을 가지고 있다. 기존의 슬라이딩 윈도우 접근법과 R-CNN 계열의 탐지 방법은 region proposal 방법을 사용하여 이미지 안에 가능성이 많은 경계박스를 생성하고 각 요소들을 따로 학습하기 때문에 최적화 및 실시간 적용에 어려움을 가지고 있다. 제안하는 연구는 YOLOv2 알고리즘을 적용하여 기존의 알고리즘이 가지고 있는 물체 탐지의 실시간 처리 문제점을 해결하여 실시간으로 지상에 있는 자동차를 탐지하는 방법을 제안한다. 제안하는 연구 방법의 실험을 위하여 오픈소스로 제공되는 Darknet을 사용하였으며 GTX-1080ti 4개를 탑재한 Deep learning 서버를 이용하여 실험하였다. 실험결과 YOLO를 활용한 자동차 탐지 방법은 기존의 알고리즘 보다 물체탐지에 대한 오버헤드를 감소 할 수 있었으며 실시간으로 지상에 존재하는 자동차를 탐지할 수 있었다.

Abstract This paper proposes a vehicle detection method for parking areas using unmanned aerial vehicles (UAVs) and using YOLOv2, which is a recent, known, fast, object-detection real-time algorithm. The YOLOv2 convolutional network algorithm can calculate the probability of each class in an entire image with a one-pass evaluation, and can also predict the location of bounding boxes. It has the advantage of very fast, easy, and optimized-at-detection performance, because the object detection process has a single network. The sliding windows methods and region-based convolutional neural network series detection algorithms use a lot of region proposals and take too much calculation time for each class. So these algorithms have a disadvantage in real-time applications. This research uses the YOLOv2 algorithm to overcome the disadvantage that previous algorithms have in real-time processing problems. Using Darknet, OpenCV, and the Compute Unified Device Architecture as open sources for object detection. a deep learning server is used for the learning and detecting process with each car. In the experiment results, the algorithm could detect cars in a dense area using UAVs, and reduced overhead for object detection. It could be applied in real time.

Keywords : Vehicle Detection, Deep Learning, Darknet, YOLOv2, Object Detection

1. 서론

지능형교통시스템(ITS : intelligent transport system)은 교통체계의 효율성과 안전성을 제고하기 위하여 기존

의 교통체계에 전자, 정보, 통신, 제어 등의 지능형 기술을 접목시킨 차세대 교통체계이다. 지능형 교통시스템은 도로의 변화하는 환경을 실시간으로 분석하여 현재의 도로환경을 개선하는 방향으로 교통량의 흐름을 조절하는

본 논문은 상명대학교 2016년 교내연구비 지원에 의해 수행되었음.

*Corresponding Author : Chang-Jin Seo(Sangmyung Univ.)

Tel: +82-41-550-5445 email: cjseo@smu.ac.kr

Received February 5, 2018

Revised (1st February 19, 2018, 2nd February 22, 2018)

Accepted March 9, 2018

Published March 31, 2018

시스템이다. 이러한 지능형교통시스템의 구축을 위해서는 도로의 상황을 실시간으로 파악할 수 있는 데이터 수집기능이 선행되어야 한다. 현재 도로에 설치되어 있는 기계식 루프검지기는 설치 및 유지보수의 문제를 가지고 있으며 고정형 무선센서를 이용한 방법은 일정한 탐지영역만을 가진다는 단점을 가지고 있다. 최근 무인항공기와 고해상도 카메라(UHD : ultra high definition)의 발전으로 저렴한 비용으로 항공 이미지를 얻을 수 있게 되었다[1-3].

최근까지 자동차의 항공 영상을 얻기 위하여 인공위성(satellites), 비행기(air-craft), 헬리콥터 등을 이용하였으나 영상 획득을 위한 비용이 과다하고 시간 및 날씨의 변화에 실시간으로 대응하지 못하는 단점을 가지고 있다. 최근 저고도 항공 영상을 이용하여 자동차를 탐지하는 분야에 대한 연구가 활발히 진행되고 있고 그 응용분야도 매우 다양하게 적용되고 있다. 특히, 무인항공기(UAVs : unmanned aerial vehicles)를 이용한 자동차 탐지에 대한 연구가 최근에 활발히 진행되고 있다. 무인항공기를 이용하는 방법은 저렴한 비용으로 초고화질 항공 영상을 사용할 수 있게 됨으로 변화하는 도로 환경(기상의 변화, 조명의 변화 등)에서도 손쉽게 항공 영상을 획득할 수 있게 되었다[4-5]. 본 논문에서는 실시간 물체 탐지에 최근 주목을 받고 있는 YOLO를 이용하여 저고도 항공 영상에 존재하는 자동차를 실시간으로 탐지하는 방법을 제안한다.

논문의 구성은 다음과 같다. 2장에서는 제안하는 연구의 배경이 되는 R-CNN, YOLO에 대하여 살펴보면, 3장에서는 제안하는 연구방법, 4장에서는 제안하는 연구방법의 실험결과, 5장에서는 결론 및 향후연구 과제를 제시한다.

2. 관련 연구

UAV 항공이미지를 이용하여 물체를 탐지하고 추적하는 최근의 연구를 살펴보면 항공 영상에 존재하는 사람, 자동차, 군사적 목표물을 찾기 위하여 다양한 형태로 연구가 진행되고 있다[6]. 항공 영상은 대부분 대용량 데이터를 사용하기 때문에 물체를 탐지하는데 많은 계산시간이 필요한 단점을 가지고 있다. 그러므로 탐지 알고리즘의 성능이 물체탐지 및 추적기의 성능에 많은 영향을

주게 된다. 영상에 존재하는 물체를 탐지하기 위하여 복잡한 알고리즘을 채택하면 탐지성능은 좋아지나 실시간 분석능력이 저하되고 간단한 알고리즘을 채택하면 실시간 처리 능력은 좋아지나 탐지성능이 저하되는 문제점을 가지고 있다[7-8]. 결국, 영상에 존재하는 다양한 물체의 특징정보를 어떻게 추출 하는가에 따라 전체 시스템의 성능에 영향을 주게 된다.

2.1 R-CNN

Ross Girshick는 object proposal과 CNN(convolution neural network)을 결합시켜 물체를 인식하고 탐지하는 R-CNN(region with convolutional neural network)을 제안하였다. R-CNN의 기본 원리는 입력 이미지로부터 약 2,000개의 proposal 영역을 생성하고, selective search 알고리즘을 이용하여 후보 영역을 생성한다. 선정된 후보 영역은 CNN의 입력 크기에 맞게 warping과 crop을 사용하여 이미지의 크기를 변경하고 CNN에 데이터로 입력하고 CNN의 최종 출력 단계에서 후보영역에 대한 특징 벡터를 생성하고 linear SVM을 이용하여 해당 영역을 분류한다. R-CNN은 CNN을 물체 탐지에 적용하여 기존의 SIFT나 HOG에 비하여 뛰어난 성능을 보이고 있다[9-11]. 그러나 R-CNN과 같이 region proposal 방법을 이용한 물체 탐지 방법은 연산량이 많기 때문에 처리속도가 느려 실시간으로 사용하기에는 적합하지 않은 단점이 있다.

Faster R-CNN는 그림 1과 같은 구조를 가지고 있다. R-CNN계열의 검출 네트워크들은 이미지에서 물체가 있을 것 같은 ROI(region of interest)를 구성하고 후보로 선정된 ROI들은 분류기에 의해 클래스 분류가 이루어지고 경계박스(bounding box)를 찾는 구조로 되어 있다. region proposal network는 이미지에서 물체가 있을 확률이 있는 영역을 선정하여 여기에 어떤 클래스가 존재하는지 확인할 수 있도록 제안을 하는 구조이다. 일반적으로 하나의 이미지에서 300개 정도의 후보를 선정하는 것으로 알려져 있다. 표 1과 같이 Faster R-CNN(RPN+ZF, shared)는 학습과정에 2,000개, 테스트 과정에서는 300개의 영역을 선정한다[12]. 이러한 방식은 제안된 영역이 어떤 클래스인지 분류하기 위해 분류기의 많은 레이어를 통과해야 한다. 기존의 R-CNN 계열의 물체 탐지 방법이 느린 이유는 proposal의 수도 많고, 처리해야 하는 과정에서 오버헤드도 크기 때문이다.

클래스가 있는지에 대한 정보가 포함되어 있다. 그림 4의 오른쪽 영상은 네트워크 최종 출력을 이용해 생성한 정보로 네트워크가 직접 생성한 것이다. YOLO 네트워크는 영상을 7×7 그리드로 나눈다. 각 그리드에서 중심을 그리도 안쪽으로 하면서 크기가 일정하지 않는 경계박스를 2개씩 생성한다. 그리드 셀이 7×7=49개 이므로 총 98개의 경계박스가 생성되는 구조를 가지고 있다. 이 정보들 중에서 경계박스 안쪽에 어떤 물체가 있을 것 같다고 확신(confidence score)이 높을수록 박스를 굵게 생성한다. 최종 단계에서는 굵은 경계박스들만 남기고 얇은 경계박스를 지우며 남은 후보 경계박스를 NMS(non-maximal suppression) 알고리즘을 이용하여 최종 오브젝트를 선정하면 그림 4의 우측 이미지와 같이 3개만 남은 구조로 구성되어 있다.

3. 제안하는 연구방법

YOLO는 실시간 물체 검출기로 물체의 경계박스를 나타내고 동시에 클래스를 분류하는 기능을 수행한다. 본 논문에서는 YOLO의 최신버전인 YOLOv2를 이용하여 실험에 사용하였다[15]. YOLOv2는 성능과 속도를 모두 개선시켜 현재 SSD(single shot multibox detector)보다 성능이 뛰어난 것으로 알려져 있다.

YOLOv2는 네트워크의 크기를 조절하여 FPS(frame per second)와 mAP(mean average precision)를 균형 있게 조절할 수 있다. YOLOv2는 그림 5와 같이 성능향상을 위하여 10가지 추가적 방법을 도입함으로써 성능이 향상 되었다[15].

	YOLO								YOLOv2
batch norm?		✓	✓	✓	✓	✓	✓	✓	✓
hi-res classifier?			✓	✓	✓	✓	✓	✓	✓
convolutional?				✓	✓	✓	✓	✓	✓
anchor boxes?					✓	✓	✓	✓	✓
new network?						✓	✓	✓	✓
dimension priors?							✓	✓	✓
location prediction?								✓	✓
passthrough?								✓	✓
multi-scale?								✓	✓
hi-res detector?								✓	✓
VOC2007 mAP	63.4	65.8	69.5	69.2	69.6	74.4	75.4	76.8	78.6

Fig. 5. The path from YOLO to YOLOv2[15]

3.1 실험환경 구축

본 논문에서는 지상에 주차되어 있는 자동차를 무인항공기로 탐지하기 위하여 다음과 같은 실험환경을 구축하였다. CNN 모델을 구축하기 위하여 본 논문에서는 오

픈소스로 제공 되어지는 YOLOv2 Linux 버전(https://github.com/AlexeyAB/darknet)을 사용하였다[16]. YOLOv2는 빠른 물체 탐지 속도를 유지하면서 Faster R-CNN에 비하여 우수한 인식률을 보이고 있다. 그리고 제안하는 연구를 위하여 OpenCv 3.2, CUDA 8.0, Xeon E5-2650 4-CPU, GTX-1080TI 4-GPU를 탑재한 SDX-4185 Deep Learning서버를 사용하여 그림 6과 같은 환경에서 C와 CUDA기반의 프레임워크인 darknet을 이용하여 실험하였다.

3.2 학습 데이터 구성

실험에 사용한 영상은 DJI사의 Phantom3 Professional을 이용하여 교내 캠퍼스에 주차 및 이동하는 자동차를 고도 50m 상공에서 촬영하였다. 실험 영상의 크기는 1920×1080 Full HD 크기로 구성하였다. 무인항공기 영상을 이용하여 지상의 자동차를 탐지하기 위해서는 먼저 학습과정을 거쳐야 한다. 컨볼루션 네트워크에 대한 학습을 위해서 지상의 자동차 영상에 대한 학습이 필요하다. 그림 7은 학습에 사용한 자동차 영상이다.

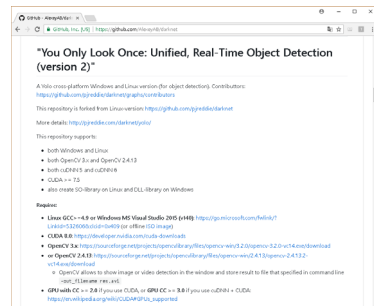


Fig. 6. The experiment environment[16].

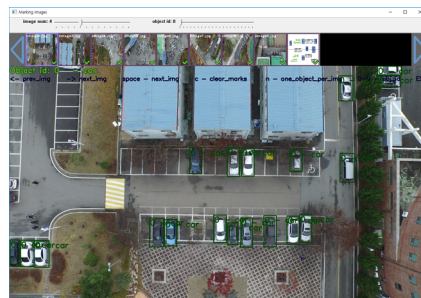


Fig. 7. The training data set by YOLO Marker[15].

5. 결론 및 향후과제

본 논문에서는 지상의 자동차를 학습시킨 컨볼루션 네트워크를 이용하여 지상의 자동차를 실시간으로 탐지하는 모델을 제안하였다. 실험 결과 컨볼루션 네트워크의 정밀도가 각 실험영상에서 84.06, 91.05, 87.67이며 최대 42FPS로 물체를 탐지하고 실시간으로 동작하는 것을 확인할 수 있었다. 하지만 실험에 사용한 영상의 수가 적어 실험의 정확성을 판단하기에는 다소 어려움이 있지만 물체탐지 속도는 다른 알고리즘에 비하여 현재 최고의 성능을 보임을 알 수 있었다.

향후 연구 과제는 탐지된 자동차의 종류(승용, 트럭, 승합, 버스)에 대한 분류가 필요할 것으로 보이며 이동 물체에 대한 추적개념의 알고리즘을 도입하면 보다 우수한 성능의 물체 탐지 및 추적 시스템에 적용할 수 있을 것으로 판단된다.

References

- [1] K. Kozempel and R. Reulke, "Fast Vehicle Detection and Tracking in Aerial Image Bursts", in *ISPRS City Models, Roads and Traffic(CMRT)*, Paris, France, vol. 38, no. 3/W4, pp. 175-180, 2009.
- [2] J. Leitloff, S. Hinz, and U. Stilla, "Vehicle extraction from very high resolution satellite images of city areas," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 7, pp. 2795-2806, 2010.
DOI: <https://doi.org/10.1109/TGRS.2010.2043109>
- [3] W. Yao, M. Zhang, S. Hinz, and U. Stilla, "Airborne traffic monitoring in large areas using lidar data," *Int. J. Remote Sens.*, vol. 33, no. 12, pp. 3930 - 3945, 2012.
DOI: <https://doi.org/10.1080/01431161.2011.637528>
- [4] D. Lenhart, S. Hinz, J. Leitloff and U. Stilla, "Automatic Traffic Monitoring Based On Aerial Image Sequences", *Pattern Recognition and Image Analysis*, vol. 18, no. 3, pp. 400-405, 2008.
DOI: <https://doi.org/10.1134/S1054661808030061>
- [5] A. C. Holt, E. Y. W. Seto, T. Rivard and G. Peng, "Object-Based Detection and Classification of Vehicles from High-Resolution Aerial Photography", *Photogrammetric Engineering and Remote Sensing (PE& RS)*, vol. 75, no. 7, pp. 871-880, 2009.
DOI: <https://doi.org/10.14358/PERS.75.7.871>
- [6] M. Elmiktay and T. Stathaki, "Car Detection in High-Resolution Urban Scenes Using Multiple Image Descriptors", in *Proc. Of International Conference on Pattern Recognition (ISPR)*, Stockholm, Sweden, pp. 4299-4304, 2014.
DOI: <https://doi.org/10.1109/ICPR.2014.737>
- [7] T. Moranduzzo and F. Melgani, "Detecting Cars in UAV Images with a Catalog-Based Approach," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 10, pp. 6356-6367, 2014.
DOI: <https://doi.org/10.1109/TGRS.2013.2296351>
- [8] S. Tuermer, F. Kurz, P. Reinartz and U. Stilla, "Airborne Vehicle Detection in Dense Urban Areas Using HoG Features and Disparity maps," *IEEE Journal of Selected Topics in Applied Earth Observation and Remote Sensing*, vol. 6, no. 6, pp. 2327-2337, 2013.
DOI: <https://doi.org/10.1109/JSTARS.2013.2242846>
- [9] S. Liao, X. Zhu, Z. Lei, L. Zhang and S. Z. Li, "Learning Multi-scale Block Local Binary Patterns for Face Recognition," *ICB 2007*, pp. 828-837, 2007.
DOI: https://doi.org/10.1007/978-3-540-74549-5_87
- [10] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *IEEE Computer Society Conf. Computer Vision and Pattern Recognition (CVPR)*, San Diego, CA, USA, vol. 1, IEEE Computer Society, pp. 886-893, 2005.
DOI: <https://doi.org/10.1109/CVPR.2005.177>
- [11] X. Chen and Q. Meng, "Vehicle Detection from UAVs by Using SIFT with Implicit Shape Model", in *IEEE International Conference on Systems, Man, and Cybernetics*, pp. 3139-3144, 2013.
DOI: <https://doi.org/10.1109/SMC.2013.535>
- [12] Shaoqing Ren, Kaiming He, Ross Girshick, Jian Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Volume: 39, Issue: 6, pp. 1137-1149, June 2017.
DOI: <https://doi.org/10.1109/TPAMI.2016.2577031>
- [13] Mahyar Najibi, Mohammad Rastegari, Larry S. Davis, "G-CNN: An Iterative Grid Based Object Detector", *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2369-2377, 2016.
DOI: <https://doi.org/10.1109/CVPR.2016.260>
- [14] Redmon Joseph, Divvala Santosh, Girshick Ross, Farhadi Ali, "You Only Look Once: Unified, Real-Time Object Detection", *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 779-788, 2016.
- [15] Joseph Redmon, Ali Farhadi, "YOLO9000: Better, Faster, Stronger", *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7263-7271, 2017.
DOI: <https://doi.org/10.1109/CVPR.2017.690>
- [16] Alexey, "Yolo-v2 Windows and Linux version", <https://github.com/AlexeyAB/darknet>

서 창 진(Chang-Jin Seo) [정회원]



- 1999년 2월 : 부산대학교 대학원 멀티미디어 석사
- 2003년 2월 : 부산대학교 대학원 멀티미디어 박사
- 2000년 3월 ~ 2013년 2월 : 성덕대학교 컴퓨터계열 교수
- 2013년 3월 ~ 현재 : 상명대학교 정보보안공학과 교수

<관심분야>

Object Detection, Target Tracking, Artificial Vision, Multimedia, E-Learning, Security