

드론영상기반 실시간 객체인식 시스템 설계 및 구현 (Design and Implementation of Real-time Object Detection based on Drone Video Streaming Data)

정철우, 천승만*, 석수영
(재)경북IT융합산업기술원

(Cheol-Woo Jung, Seung-Man Chun, and Soo-Young Suk)
(Gyeongbuk Institute of IT Convergence Industry Technology)

Abstract : This paper proposed the real-time object detection (car and people) system based on drone's video data by utilizing YOLOv3(You Only Look Once) for the disaster situation analysis, disaster relief, etc. To do this, the real-time video streaming data collected by the drone's camera via LTE, was analysed by the machine learning model (YOLOv3, DarkNet) in the server. Also, the machine learning model learned the labeling data (10,000) included car and people by the pre-training method. We newly developed the labeling data collection method. The proposed system was developed and evaluated. In the performance evaluation, the object detection accurate rate of car and people showed over than 96 percents.

Keywords : YOLOv3, DarkNet, Object Detection, Drone, Machine Learning

1. 서론

드론은 과거 군사적 용도를 위해 개발되었으나 현재 재난 감시, 농업, 물류 방송 등에 다양한 산업 분야에서 활용되고 있으며 또한 레저용으로도 대중화 및 보편화되고 있는 추세이다. 이는 드론의 낮은 유지비용, 신속성 등과 같은 장점을 통해 산/학/연에서는 인공지능, 빅데이터, IoT 등을 적용하여 다양한 연구가 진행되고 있다. 특히 드론은 화재, 교통사고 현장과 같이 긴급구조가 필요한 재난현장에서 신속한 구조 활동을 수행하기 위해 미리 상황을 판단하고 이를 출동 중인 소방관, 경찰관들에게 현장 상황을 증계함으로써 신속한 구조 활동을 할 수 있다[1].

최근에는 머신러닝 기반 드론을 활용하여 실시간 객체를 인식할 수 있는 다양한 연구가 시도 중에 있다[2]. 객체인식을 하는 다양한 기법들이 있지

만 특히 머신러닝 기법 중 YOLOv3(You Only Look Once)은 빠른 속도의 장점이 있기 때문에 실시간 영상에 객체를 인식하기 위해 특화된 기법이다[3]. 이는 드론 영상에서의 객체인식 정확도와 성능의 속도를 검증하는 Vis-Drone Challenge 2018에서 YOLOv3 모델이 가장 우수한 속도를 보였으며, 객체인식 수준에서도 준수한 성능을 보였다. 그러나 기존 YOLOv3에서 제공하는 모델은 그림 1에서도 확인할 수 있듯이 오검출되는 결과를 확인할 수 있다. 해당 영상에서는 자동차를 'Cell Phone'으로 인식하는 오검출이 발생하였다. 기존의 모델에는 다양한 환경을 고려하여 객체를 학습하기 때문에 얻고자 하는 객체 데이터를 얻기에는 어렵다.

본 논문에서는 YOLOv3 기법을 활용하여 실시간 드론영상에서 객체를 인식하는 방법을 제안하고자 한다. 본 논문에서 구현한 시스템은 교통사고 현장이라는 가상의 시나리오 속에서 자동차와 사람에 대한 두 개의 클래스로 이루어진 객체 검출을 수행한다. 이를 위해 드론을 통하여 자동차와 사람을 촬영하여 영상 데이터를 구축하였으며 객체 검출 분야의 YOLOv3 모델의 기반인 DarkNet 프레임워크에 맞게 적용하여 교통사고 현장에서 사용가능한 맞춤형 모델을 생성한다.

* 교신저자(Corresponding Author)

천승만 : (재)경북IT융합산업기술원

본 연구는 산업통상자원부와 한국산업기술진흥원의 “창의산업거점지원사업”(과제번호 P0003329)으로 수행된 연구결과입니다.

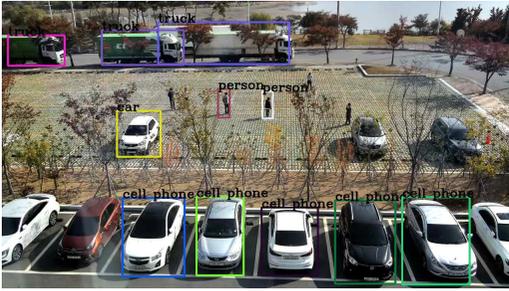


그림 1. 기존 모델의 오검출 결과
Fig. 1. Incorrect Detection results in Original Model

II. 본 론

그림 2는 실시간 드론 영상기반 객체인식기술에 대한 구조를 나타낸다. 연구에 사용된 드론에는 Jetson TX2보드와 e-con system사의 See3 4K Cam 카메라 그리고 LTE 통신모듈을 탑재하였다. 또한 실시간으로 영상을 전송하기 위해서 Wowza 스트리밍 서버를 이용하였고, 이를 통하여 카메라에서 촬영되는 영상을 LTE 통신을 이용하여 서버로 전송된다. 서버로 전송된 영상은 학습된 YOLOv3 모델을 통하여 실시간 객체인식이 되고, 결과 값을 사용자에게 각 프레임별 사진 및 동영상으로 제공한다.

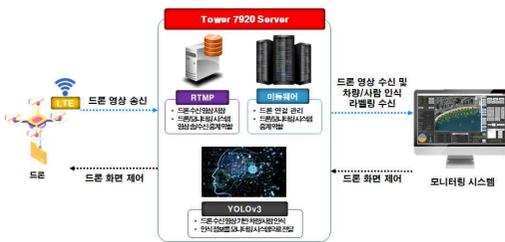


그림 2. 제안 시스템 구조
Fig. 2. Proposed system structure

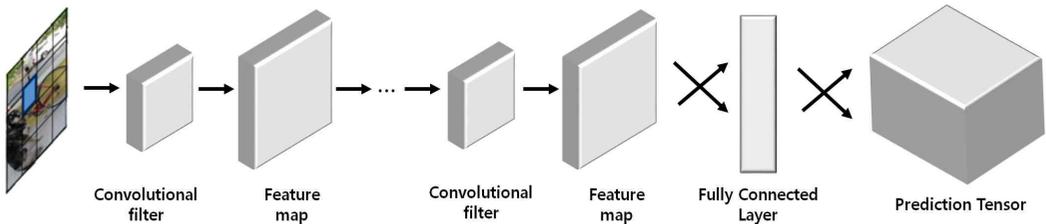


그림 3. YOLOv3의 네트워크 구조
Fig. 3. Structure of YOLOv3 network

1. YOLOv3 객체인식 기법

YOLOv3 모델에서는 각 입력 이미지를 $(S \times S)$ 그리드로 나눈다. 각 그리드 셀(Cell) 당 B개의 바운딩 박스를 예측한다. 이 때 바운딩 박스의 위치는 물체의 중심점(x,y)와 물체의 길이, 높이(w,h) 값으로 표현된다. 각 바운딩 박스는 그 위치 예측에 대한 신뢰도 값을 가지므로 하나의 바운딩 박스에 대해 5개의 예측 값(x, y, w, h, confidence score)이 필요하다.[4] 이를 CNN(Convolution Neural Network) 구조의 신경망을 통해 특징이 추출되며, 완전 연결 노드(Fully connected layer)를 통해 최종적으로 그림 3과 같이 예측 텐서(Prediction tensor)가 출력된다. 예측 텐서는 $(S \times S)$ 의 크기와 $(B \times 5 + C)$ 의 길이를 가진다.[5]

2. YOLOv3 Pre-training 방법

기존의 YOLOv3 모델에서 원하는 데이터만을 검출하도록 Pre-training 방법을 적용하였다. 우선 드론을 통하여 자동차와 사람을 촬영하여 영상 데이터를 수집하였고, 다양한 자동차와 사람의 검출을 위해 각각의 다양한 영상을 수집하여 이미지로 변환하고 등장하는 객체마다 직사각형 박스로 위치를 표시하고 2가지 클래스(자동차, 사람)로 라벨링을 하였다. 객체인식의 라벨링은 수작업으로 수행해야 하므로 많은 시간을 소비한다. 이를 개선하기 위해서 YOLOv3의 결과물을 기반으로 한 라벨링 데이터 생성 프로그램을 만들어 사용하였다. 최초의 라벨링 데이터들을 만들고 학습을 시킨 후 비슷한 결과물이 나오면 그 이후 작업은 결과물을 다시 라벨링 데이터로 변환하여 준다. 그림 4에서는 본 연구에서 사용한 라벨링 방법에 대해 소개한다. 먼저 Labelimg 툴을 활용하여 라벨링 데이터를 직접 생성하고, 원본이미지 및 라벨링데이터를 확보한다. 이후 머신러닝 모델을 통하여 데이터를 학습하고 정확도를 검증한다. 정확도가 60% 이상이 출력되면 모델을 활용하여 결과 데이터를 생성하고, 결과 테

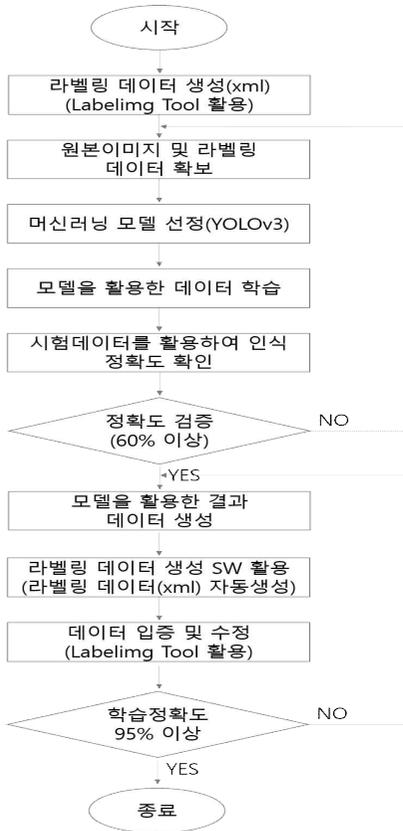


그림 4. 라벨링 데이터 생성방법
Fig. 4. Generate labeling data

이더를 라벨링 데이터 생성 SW에 활용하여 라벨링 데이터를 추출한다. Labeling 툴을 통하여 결과를 입증 및 수정하는 작업을 진행한다. 이를 이용하여 기존에 직접 수작업으로 진행하던 라벨링 데이터 수집의 속도가 기존의 10/1정도로 줄어들었다.

수집한 이미지 중 10%는 시험(test) 데이터로 두고, 나머지 90%는 학습(train) 및 검출(validation) 데이터로 나누었다. <표 1>은 최종적으로 구축된 학습 및 시험 데이터이다. 하나의 이미지에 여러 개의 객체가 존재할 수 있으므로 장소별 이미지 개수 및 각 클래스에 해당하는 자동차와 사람의 개체 개수를 표시하였다. YOLOv3 신경망 학습은 NVIDIA Titan V 2개의 GPU 환경에서 수행되었다. Batch Size를 10으로 하여 약 400 epochs를 수행하였다. 이후 학습을 시켜 나온 가중치 파일을 획득하여 실제 촬영한 드론영상 및 사고 현장 영상에 대해 실험하였다.

표 1. 학습 데이터 (단위: 개수)
Table 1. Construction data

장소	이미지	자동차	사람
주차장1	743	4,916	2,972
주차장2	1,913	21,043	15,304
공터1	401	2,005	3,609
공터2	1,816	3,632	23,608
사고현장1	854	12,810	8,540
사고현장2	2,271	24,981	29,523
기타	384	1,536	3,456
총계	8,382	70,923	87,012

3. YOLOv3을 이용한 객체 검출결과

앞서 학습한 방법을 통하여 모델을 통하여 그림 1의 환경과 동일하게 실험을 진행했을 때 그림 5와 같이 'Cell phone'으로 표시하던 기존의 모델과는 달리 정확하게 'Car'로 인식하는 것을 확인할 수 있었다. 검출을 하지 못하는 자동차와 사람의 경우에도 모두 예측하는 것을 확인할 수 있었다. <표 2>는 영상 객체인식을 다양한 환경에서 시도하여 영상 객체 인식률을 비교하였다. 인식률의 평균은 95% 정도를 보였다.

III. 결 론

본 논문에서는 드론영상에서 YOLOv3 모델 기반의 DarkNet 프레임워크를 활용하여 실시간으로 객체를 검출하였다. 기존의 YOLOv3 모델에서는 오검출되던 부분들의 Pre-training을 진행하여 이를 개선하였다. 이를 위해 교통사고 현장이라는 가상의 시나리오에서 사용가능한 모델을 학습하기 위하여

표 2. 영상 객체 인식률
Table 2. Image object detection rate

회차	육안		머신러닝		인식률
	사람	자동차	사람	자동차	
1	5	14	5	14	100%
2	7	12	7	11	94%
3	3	11	3	11	100%
4	6	9	5	9	94%
5	5	13	4	12	88%
인식률 평균					95.2%

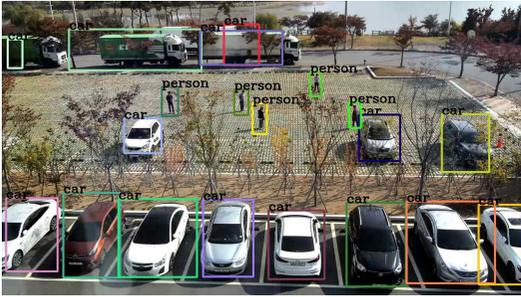


그림 5. 학습된 모델의 검출 결과

Fig. 5. Detection results in trained models 자동차와 사람의 영상 촬영을 다양한 장소 실시하여 데이터 구축하였다. 추가적으로 다양한 객체를 추가하기 위해서는 다양한 라벨링 데이터의 수를 증가시켜 추가적으로 학습이 가능하다.

이를 이용하여 재난 안전분야에 드론을 특화하여 활용 가능할 것이라고 생각한다. 향후 연구에서는 화재현장 영상을 기반으로 객체인식 모델을 학습시켜 다양한 응급현장에서 사용할 수 있는 기술을 개발할 예정이다.

참 고 문 헌

- [1] 임수연, "재난 안전 현장에서의 드론(drone) 활용", 과학기술정책연구원, 제 25권, 제 6호, 16-19쪽, 2015.
- [2] C. Aker and S. Kalkan, "Using deep networks for drone detection", In 2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), pp. 1-6, 2017.
- [3] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection", 2016 IEEE Conference on Computer Vision and Pattern Recognition(CVPR), 2016.
- [4] 김철연, 한택진, 윤일로, 이운진, 이지영, 최경현, 원정인, 김영민, "YOLOv3을 이용한 CCTV 영상에서의 쓰레기 검출", 한국컴퓨터종합학술대회, 785-787쪽, 2018.
- [5] 김진수, 조경호, "RGB 영상 및 LiDAR 포인트 클라우드 합성을 통한 YOLOv3 기반 실시간 객체 탐지", 한국정보기술학회, 제 17권, 제 8호, 93-105쪽, 2019.