

카메라의 이동 속도와 영상 객체 인식 정확도 간의 상관관계 분석

손본, 김준우, 최정운, 이향원
건국대학교 컴퓨터공학부

bono0208@konkuk.ac.kr, kjw5841@naver.com, 99jenny0614@naver.com,
leehw@konkuk.ac.kr

Correlation Analysis between the Speed of Camera and the Accuracy of Video Object Recognition

Sohn Vaughn, Kim Junu, Choi Jeongyun, Hyang-Won Lee,
Konkuk Univ.

요 약

본 논문에서는 강화학습 모델의 학습을 위하여 영상을 촬영하는 카메라를 장착한 이동체의 속도에 따른 객체 인식의 정확도에 대한 상관관계를 분석한다. 연구를 위한 이동체는 Raspberry Pi 4B, 4WD Car-Rpi 를 사용했으며 객체 인식을 위한 모델은 사전 학습된 YOLOv5 를 사용했다. 선형 회귀 분석 결과, 이동체의 속도가 빨라질수록 객체 인식의 정확도(mAP)가 낮아지는 경향을 확인할 수 있었다.

I. 서 론

최근 자율주행기술이나 로봇 관련 분야의 발전에 따라 스스로 환경을 인식하여 행동하는 지능형 로봇에 대한 관심이 증가하고 있다. 복잡한 실제 환경을 인식하고 해석하여 적절한 행동을 수행하기 위해서는 DNN(Deep Neural Networks) 모델의 사용이 요구된다. 그러나 제한된 컴퓨팅 능력을 갖는 엣지 디바이스에서는 이러한 계산 집약적인 일을 처리하기에는 한계를 보인다. 반면에 모든 계산을 클라우드에 전송하여 처리하는 방법은 엣지 디바이스의 자원을 사용하지 않아 잉여 컴퓨팅 자원이 발생하고 또한 네트워크의 대역폭이나 지연시간에 따라 성능이 크게 달라질 수도 있다.

이러한 단점들을 개선하기 위하여 DNN 의 전체 연산 중 일부는 엣지 디바이스에서 처리하고 나머지 연산은 클라우드에서 진행하는 연산 오프로딩(computation offloading)에 대한 연구가 활발히 진행되고 있다[1].

본 연구는 연산 오프로딩 전략을 사용하여 DNN 모델의 분할 위치를 결정하고 높은 객체 인식 정확도를 유지하면서 주어진 경로를 따라 최대한 빨리 원위치로 되돌아오는 무인 이동체를 위한 강화학습 모델을 만드는 것을 목표로 한다. 이때, 강화학습 모델의 학습을 위해서 무인 이동체의 속도에 따라 객체 인식 정확도가 어떻게 변화하는지를 분석해야 할 필요성이 대두되었다.

따라서 본 연구에서는 영상을 촬영하는 카메라를 장착한 이동체의 속도에 따른 YOLOv5 모델의 영상 객체 인식 정확도 간의 상관관계를 분석하고자 한다.

II. 본론

본 연구에서 이용한 이동체는 Raspberry Pi 4B 와 4WD Car-Rpi [2]를 이용하여 제작되었다. 이동체가 촬영한 영상 데이터는 사전 학습된 YOLO v5 모델[3]을 이용하여 객체 인식을 진행하였으며 모델이 예측한

데이터와 실제 정답 클래스를 표기한 라벨링 데이터를 비교하여 객체 인식 정확도를 측정하였다.

YOLO(You Only Look Once)는 객체 인식을 위해 고안된 신경망 모델이다. 객체 인식 분야에서 R-CNN 계열 모델들이 바운딩 박스를 찾는 과정과 객체의 클래스를 분류하는 과정을 분리해서 수행하던 것과 달리 YOLO 모델은 2 개의 문제를 하나의 회귀 문제로 통합하여 단일 신경망 모델로 나타낸다. 이러한 구조적 특징에 따라 YOLO 는 단순하고 빠르며 실시간 객체 인식에도 사용할 수 있다[4]. 본 연구팀은 무인 이동체의 스트리밍 영상에 실시간 객체 인식을 적용하기 위하여 YOLOv5 모델을 채택하였다.

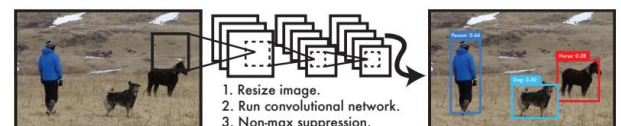


그림 1. YOLO 의 객체 인식 방법 [4]

객체 인식의 정확도를 측정하기 위한 성능지표로서 mAP(mean Average Precision)를 사용하였다. AP(Average Precision)은 객체를 인식했는지(P) 하지 못했는지(N)를 결정하는 confidence threshold 를 변경시키면서 얻은 recall 과 precision 의 변화를 나타내는 P-R curve 의 면적을 의미한다. recall 과 precision 은 서로 간에 trade-off 가 존재하기 때문에 두 값이 모두 클수록 모델의 성능이 더 좋다고 평가할 수 있다. mAP 는 각각의 클래스별로 계산된 AP 값에 대한 평균으로 계산된다. 사전 학습된 객체 인식 모델에서 다른 입력 데이터를 주었을 때, mAP 의 차이를 통하여 객체 인식의 정확도를 비교할 수 있다. mAP 을 계산하기 위해 YOLO v5 의 val.py 파일을 사용했다[3].

YOLOv5 가 인식할 수 있는 80 개의 클래스(COCO dataset) 중에서 bottle, umbrella, keyboard, mouse,

sports ball, tv 에 해당하는 객체를 배치하여 실험환경을 구성하였다. 동일한 실험 환경에서 이동체의 속도만 변경시켜가며 영상을 촬영해 실험에 사용할 데이터를 얻었다. 객체 인식 정확도를 측정하기 위하여 촬영한 영상의 프레임별로 직접 각 객체에 바운딩 박스를 그린 뒤 해당하는 클래스를 부착하여 라벨링 데이터를 제작하였다. 그림 2는 동일한 프레임에서 YOLOv5 모델이 예측한 결과(a)와 실제 라벨링 데이터(b)를 비교한다.

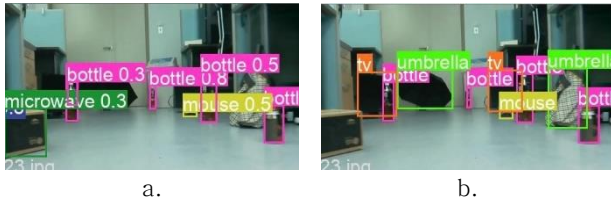


그림 2. pred, label 비교

이때, 이동체의 속도가 느릴 때 촬영했던 영상은 속도가 빠를 때 촬영한 영상에 비해 영상의 길이가 더 길다. 영상의 길이가 길어지면 유사한 프레임이 계속 반복되게 된다. 이런 반복되는 특정 프레임에 대한 예측 결과 때문에 전체 영상에 대한 객체 인식 정확도에 편향이 발생할 수 있다. 따라서 각각의 속도별 영상이 비슷한 수의 프레임을 가지도록 배속한 영상을 사용하여 실험을 진행했다. 동일한 간격마다 하나의 프레임을 선택하여 새로운 배속 영상을 생성했다. 표 1은 속도별 원본 영상과 배속 영상의 프레임 개수를 나타낸다.

	10	20	30	40	50	60	70	80	90	100
원본	515	331	241	188	154	134	114	105	92	85
배속	85	82	80	86	78	90	89	88	86	85

표 1. 속도별 프레임 개수

III. 실험 결과

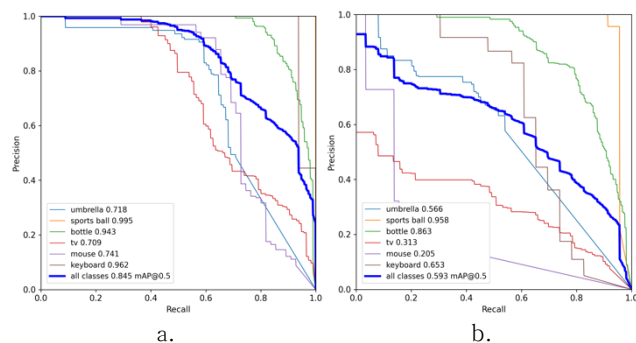


그림 3. 속도별 P-R curve

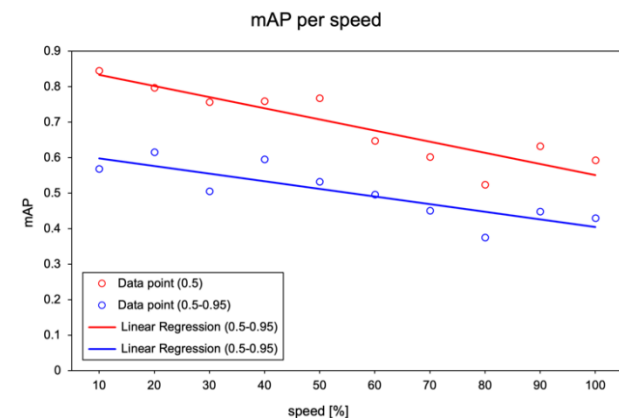


그림 4. 속도별 mAP

그림 3은 이동체의 속도가 10일 때(a)와 100일 때(b)의 P-R curve를 비교한다. 속도의 단위는 라즈베리 파이에 연결된 DC 기어박스 모터가 낼 수 있는 최대 파워에 대한 비율(%)이다. 속도가 느릴 때 P-R curve의 면적이 속도가 빠를 때 P-R curve의 면적보다 더 큰 것을 확인할 수 있다. 그림 4는 이동체의 속도가 10-100으로 변화할 때 mAP의 변화를 나타낸다. 독립변수는 이동체의 속도이고 종속변수는 mAP이다. mAP 0.5의 선형 회귀식은 0.808의 결정계수를 가지며 mAP 0.5-0.95의 선형 회귀식은 0.705의 결정계수를 가진다.

선형 회귀 분석 결과, 이동체의 속도가 증가할수록 객체 인식의 정확도가 점차 감소하는 경향성을 띠는 것을 확인할 수 있다. 이는 이동체의 속도가 증가할수록 카메라의 흔들림이 커져서 동적인 모션을 가진 프레임의 수가 증가하기 때문이라고 추측할 수 있다. 또한 연구에서 객체 인식 모델로 사용한 YOLOv5는 영상의 객체를 인식할 때, 연속된 프레임 간의 정보를 사용하지 않고 독립된 하나의 프레임으로서 처리하기 때문에 이동체의 속도에 대한 영향을 더 크게 받았을 것이다.

단, 제시한 회귀모델은 특정한 실험 환경을 기준으로 분석한 결과이기 때문에 인식의 대상이 되는 객체가 적은 환경에서는 회귀식의 기울기가 완만하게 변화하는 등의 차이가 발생할 수 있다.

추후 연구에서는 본 연구에서 분석한 회귀 모델을 활용하여 높은 객체 인식 정확도를 유지하고 DNN 모델의 연산 오프로딩 위치를 결정하며 주행하는 무인 이동체를 위한 강화학습 모델의 학습에 활용할 예정이다.

IV. 결론

본 논문에서는 카메라를 장착한 이동체의 속도에 따른 YOLOv5 모델의 객체 인식 정확도에 대한 상관관계를 선형 회귀 모델로 분석하였다. 본 연구에서는 유사한 환경에서 촬영된 비교적 적은 수의 데이터를 사용하여 상관관계를 분석했지만 다양한 환경에서 촬영한 더 많은 데이터를 사용하여 분석한다면 더 정확하고 신뢰도 있는 회귀모델을 얻을 수 있으리라 기대한다.

ACKNOWLEDGMENT

본 연구는 2020년도 산업통상자원부(MOTIE)와 한국에너지기술연구원(KETEP)의 지원을 받아 수행되었음 (No. 20204010600220). 그리고 2021년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행되었음.(No. 2021R1A2C2012801)

참 고 문 헌

- [1] Y. Kang *et al.*, "Neurosurgeon: Collaborative Intelligence Between the Cloud and Mobile Edge", ACM SIGPLAN Notices, vol. 52, no.4, pp.615-629, 2017
- [2] 4WD Car-RPi, <http://www.yahboom.net/study/4wd-Pi>
- [3] YOLOv5, <https://github.com/ultralytics/yolov5>
- [4] Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi, "You Only Look Once: Unified, Real-time Object Detection", IEEE Conference on Computer Vision and Pattern Recognition, pp.779-788, 2015