

실내 사람 객체 인식을 위한 데이터셋 구축 연구

김명섭, 박세호, *이경택
한국전자기술연구원

myeongseopkim@keti.re.kr, sehohpark@keti.re.kr, *ktechlee@keti.re.kr

A Study on the dataset for Indoor Human Object Recognition

Myeongseop Kim, Se ho Park, Kyung-Taek Lee
Korea Electronics Technology Institute

요 약

본 논문은 실내 사람 객체 인식을 위한 데이터셋을 제안하고 이를 검증한다. 사람 객체 인식의 경우 다양한 데이터셋과 모델을 통해 성능향상이 가속화되고 있지만 다양한 환경에서 목적에 맞게 적용하기 어렵다. 따라서, 본 논문은 Nvidia 에서 제공하는 데이터셋 제작 툴을 사용해 VOC 형식의 실내 사람 객체 데이터셋(own dataset)을 구축하고 타 데이터셋을 검증하기 위해 보편적인 데이터셋과 비교한다. 제안된 데이터셋을 검증하기 위해 Nvidia Jetson Nano 임베디드 디바이스 환경에서 MobileNet V2+SSDLite 네트워크를 활용해 3 종 데이터셋에 대해 성능을 평가하고 PennFudan 데이터셋으로 학습한 사람 객체 인식모델과 비교하여 더 뛰어난 학습 결과를 보이며 데이터셋을 검증하였다.

I. 서론

최근 인간 중심의 인터페이스, 스마트공장, 복합데이터처리 기술의 발전으로 사람 객체 인식에 대한 연구의 중요성이 커지고 있다. 특히, 실내 환경변화를 예측하는 AI 모델을 만들기 위해 복합 센서데이터와 상관관계가 깊은 실내 사람 객체를 수집하는 기술은 매우 중요하다[1]. 하지만, 실내용 인공지능 기반 산업의 확대되고 있음에도 불구하고 공인 사람 검출 데이터셋은 실외 데이터로 구성되어 있다.

대표적인 사람 검출 데이터셋인 펜실베이니아 대학교와 푸단 대학교에서 공동으로 제작한 PennFudan 데이터셋[2]는 실외 사람 데이터 세트로 구성되어 실내용 사람 검출을 위한 신경망 학습 데이터베이스로 활용하는 것에 한계가 있다.

위 문제를 해결하기 위해, 본 논문은 실내에서 사람 객체를 인식하기 위한 데이터셋을 구축 및 제안한다. 본 연구는 자체적으로 수집한 환경뿐만 아니라 사람 객체 인식 연구에서 대중적으로 사용되는 데이터셋에 대해서도 성능을 검증하여 보편적인 데이터셋으로의 증명을 목적으로 한다. 특히, 검증을 위해 딥러닝용 임베디드 디바이스 Nvidia Jetson Nano 에서 성능을 확인하고 타 모델과 비교해 구축한 데이터 세트를 검증한다.

II. 본론

2.1 Own Dataset

실내 인원 데이터를 수집하기 위해 NVIDIA 에서 제공하는 카메라 캡처 툴을 사용해 사람 객체에 대한

데이터셋(Own dataset)을 구축한다[3]. 실내 중앙에 설치된 Logitech C930 webcam 를 사용해 각각의 View 에 대해 1280x720 크기의 RGB 이미지를 수집하고 카메라 캡처 툴을 이용해 VOC 형식으로 사람의 Annotation 을 제작한다. Own dataset 은 이미지 프레임에 위치한 다수의 사람에 대한 Bounding Box 정보와 Class 를 포함한다. 사람 객체는 전신, 상반신 모두 포함되었으며 데이터셋은 총 1,433 장의 이미지로 구성되며 3,459 개의 사람에 대한 Bounding Box 를 포함한다.

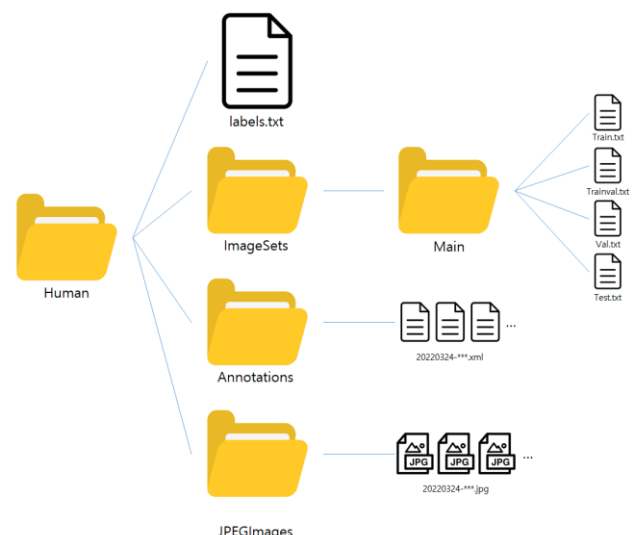


Figure 1. Own dataset structure

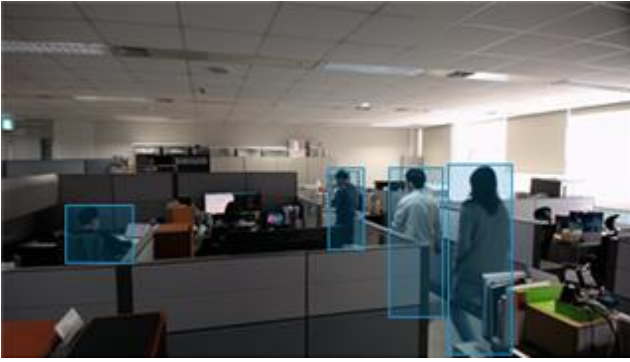


Figure 2. Own dataset

2.2 MobileNet V2 network 의 구성 및 학습

본 논문은 다양한 공간에서 쉽게 동작할 수 있는 환경 변화 예측모델을 제작하는 것을 상위 목적으로 하기 때문에 실내 사람 객체 데이터세트는 저가형 하드웨어에서 동작해야 한다. 그러므로, 이 연구는 데이터세트를 검증하기 위해 Linux os 기반 Nvidia Jetson Nano 보드에서 객체 인식 연구에서 높은 성능과 가벼운 메모리 사용량을 보이는 MobileNet V2 네트워크 사용해 학습을 진행했다[4]. 본 연구에서 사람 객체를 학습하기 위해 SSD 객체 검출 네트워크의 Convolution layer 를 모두 MobileNet V1 에서 사용한 separable convolution 으로 교체시킨 SSDLite 를 사용했다. 또한, SSDLite 의 Backbone 을 MobileNet V2 으로 교체하여 MobileNet V2 + SSDLite 네트워크를 이용해 mAP 를 비교했다.

본 논문에서 제작한 Own dataset 모델의 성능을 평가하기 위해 사람 객체 인식을 위해 대표적으로 사용되는 3 종류의 데이터세트(Own dataset, Aihub, PennFudan[2])에 대해 mAP 를 확인했다. 또한, PennFudan 데이터세트를 이용해 제작한 모델과 비교해 데이터세트의 적절성 여부를 검토했다. 학습은 동일한 조건에서 진행했다(learning rate=0.01, momentum=0.9, weight decay=5e-4, gamma=0.1, initial learning rate=0.001, scheduler=cosine, batch-size=32, epochs=100).

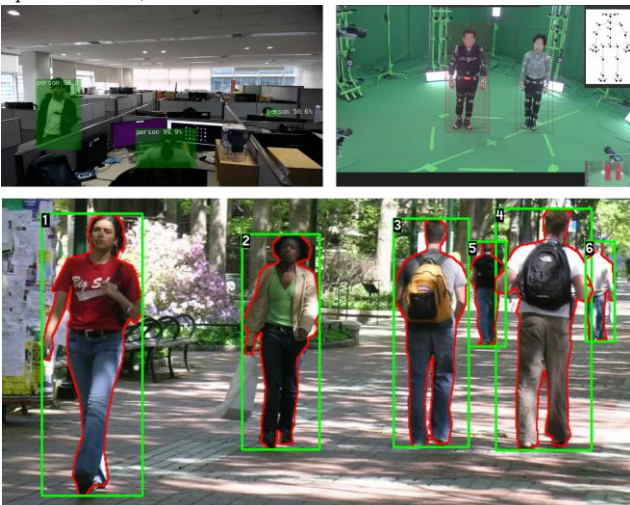


Figure 3. (Left-top) Own dataset, (Right-top) Aihub, (Down) PennFudan

2.3 성능비교결과

2.2 에서 언급한 MobileNet V2+ SSDLite 네트워크를 이용해 사람 객체 인식을 위해 본 연구에서 제작한 Own dataset 을 학습했다. Table1 과 같이 Own dataset 을

포함한 3 종의 테스트환경에서 mAP 를 비교했을 뿐만 아니라 PennFudan 데이터세트를 이용한 학습한 모델과 비교했다.

Table 1. Human object recognition benchmark

MobileNet V2+SSDLite		
Train data	Test dataset	mAP
Own dataset	Own dataset	0.62
	Aihub	0.79
	PennFudan	0.80
PennFudan	Own dataset	0.19
	Aihub	0.76
	PennFudan	0.95

Own dataset 으로 학습한 AI 모델은 3 종의 데이터세트(Own dataset, Aihub, PennFudan)를 테스트했을 경우 모든 데이터세트에서 높은 mAP 를 보였다. 그러나 PennFudan 데이터세트로 학습한 AI 모델은 실내 사람 객체로 제작한 Own dataset 에서 낮은 mAP 를 보이며 실내에 적합하지 않음을 시사했다.

III. 결론

본 논문은 실내 환경변화를 예측하는 모델에 필요한 사람 객체를 수집하고 데이터세트를 검증 및 제안했다. 데이터세트를 검증하기 위해 제작한 AI 모델로 3 종의 테스트 환경에서 mAP 를 확인하고 대표적인 데이터세트와 성능을 비교했다. 결과적으로, 제작한 실내 사람 객체 데이터세트는 모든 테스트 환경에서 좋은 성능을 보이며 범용성을 띄었다. 특히, PennFudan 데이터세트로 학습한 모델은 실내 환경에 적합성이 떨어졌지만 Own dataset 는 실내 환경에 적합한 데이터세트임을 보였다.

ACKNOWLEDGMENT

이 논문은 산업통상자원부와 한국산업기술평가관리원의 지원을 받아 수행한 연구 결과임. (No.20010525, 플렉시블 재구성이 가능한 지능형 복합 환경센서 기술 개발)

참 고 문 헌

- [1] 김명섭(Myeong Seop Kim), 박세호(Se Ho Park), and 이경택(Kyung-Taek Lee). "LSTM 기반 온도, 조도, CO2 데이터를 활용한 실내 인원 수 예측 연구." 한국통신학회 학술대회논문집 2021.6 (2021): 1169-1170.
- [2] Wang, Liming, et al. "Object detection combining recognition and segmentation." Asian conference on computer vision. Springer, Berlin, Heidelberg, 2007.
- [3] Dustin Franklin. 2020. camera-capture. <https://github.com/dusty-nv/camera-capture/tree/48c47cdb778c844d4786b7dd08dcc31a2e715867>. (2022).
- [4] Sandler, Mark, et al. "Mobilenetv2: Inverted residuals and linear bottlenecks." Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.