

# 학습 데이터 개수에 따른 객체 탐지 정확도 분석

권대혁, 우동훈, 김준희, 안홍명, 양상진  
아이오테크아이

[dhgwon@iotech.com](mailto:dhgwon@iotech.com), [dhwoo@iotech.com](mailto:dhwoo@iotech.com), [kimjh@iotech.com](mailto:kimjh@iotech.com), [hman@iotech.com](mailto:hman@iotech.com),  
[sjyang@iotech.com](mailto:sjyang@iotech.com)

## Observation of object detection accuracy depending on the number of training data

Daehyeok Gwon, DongHoon Woo, Junhee Kim, Hongmyung Ahn, and Sangjin Yang  
IOTECHI

### 요 약

본 논문은 영상 내에서 객체 탐지를 위한 인공지능 모델을 학습할 때 학습에 사용되는 데이터 개수가 증가함에 따라 객체 탐지의 성능이 향상됨을 실험을 통해 확인하고 이를 통해 소량의 데이터를 활용하여 자동화 레이블링과 같은 솔루션을 구축함에 있어 필요한 데이터의 개수를 유추할 수 있는 결과를 제시한다.

### I. 서 론

인공지능(A.I. Artificial intelligence) 기술이 발전함에 따라 다양한 분야에서 이를 활용한 서비스를 제공하고 있거나 새로운 서비스를 제공하고자 하는 경우가 크게 늘어나고 있다. 딥러닝(Deep learning)으로 대표되는 인공지능 기술은 데이터 기반의 모델 훈련 방식을 활용한 것으로 서비스에 적합한 다수의 영상, 텍스트, 음성 등과 같은 데이터를 수집하고 이를 토대로 인공지능 모델을 훈련하는 방식이다.

상당수의 모델은 학습과정에서 데이터에 대한 정답을 제공하는 레이블링이 요구되며 높은 정확도를 얻기 위해서는 방대한 양의 데이터가 요구된다. 이를 위하여 많은 기업들은 사람을 고용하여 높은 비용과 시간을 소요하며 레이블링 과정을 수행하고 있다. 이러한 높은 비용과 시간을 절감하기 위하여 레이블링을 자동화하는 기술[1]을 활용할 수 있다. 하지만 레이블링을 자동화하기 위해서도 최소 1 회의 수동 레이블링 작업이 필요하며, 어느 정도의 데이터를 확보하여야 자동화 레이블링 모델이 사용자가 원하는 수준의 정확도를 보여주는지를 알기 어렵다. 따라서 본 논문에서는 학습한 데이터 개수에 따른 성능을 측정하여 자동화 레이블링을 수행하는 과정에서 필요 데이터 수를 대략적으로 유추할 수 있도록 한다.

### II. 본론

본 논문에서는 인공지능 모델에 학습하는 데이터의 개수에 따라 객체 인식의 정확도의 변화를 분석하였다. 모델은 YOLO v4[2]를 사용하였으며, 제공된 사전학습 가중치(yolov4.conv.137)에 직접 촬영한 영상을 사용하여 커스텀 학습을 진행하였다. 학습 영상과 테스트 영상은 각각 갤럭시 S10 Plus 와 아이폰 6S Plus 를 사용하였으며, 약 12m 정도 높이에서 차량이 양방향으로 통행하는 차도를 순차적으로 촬영하였다. 각 학습은 8000 epoch 로 설정하였으며, 그 외에는 학습에 사용된 이미지의 개수(객체의 개수)를 제외하고는 모두 동일한 설정을 사용하여 학습을 진행하였다. 표 1 은 테스트 별로 학습에 사용한 이미지의 개수 및 이미지에 포함된 총 객체의 개수를 나타낸 것이다. 표 1 에서 프레임의 개수는 상위 테스트의 이미지를 포함한 개수를 나타낸 것이다. 예로, T7 의 경우 T6 에서 사용된 125 개의 이미지와 추가 25 개의 이미지를 합하여 150 장의 이미지를 학습에 사용한 것이다.

표 1 학습에 사용한 데이터 개수

Test	프레임 개수	승용차	트럭	오토바이
T1	10	76	18	0
T2	25	190	28	0
T3	50	322	70	0
T4	75	500	92	4
T5	100	733	114	17
T6	125	921	131	25
T7	150	1215	145	25

학습 데이터 개수에 따른 객체 인식 정확도의 분석을 위해, 학습에 사용한 프레임과는 다른 이미지 50 장을 사용하여 테스트를 진행하였다. 정확도 측정은 식 2와 식 3과 같이 객체 탐지에서 주로 사용하는 지표인 정밀도(Precision)와 재현율(Recall)을 사용하여 측정하였다.

$$Precision = TP / (TP + FP) \quad (1)$$

$$Recall = TP / (TP + FN) \quad (2)$$

$TP$ 는 검출된 객체 중에서 올바르게 검출된 객체의 개수를 나타낸다.  $FP$ 는 검출된 객체 중에서 객체의 종류가 틀렸거나 객체가 없는 곳을 검출하는 것과 같이 잘못 검출된 객체의 개수를 나타낸다. 마지막으로  $FN$ 은 실제로는 객체가 존재하지만 검출되지 않은 객체의 개수를 나타낸다.

그림 1은 표 1의 데이터를 사용하여 학습된 가중치들을 활용하여 테스트 영상 50장에 대하여 객체 검출을 시도하였을 때 Precision과 Recall을 그래프로 나타낸 것이고 그림 2는 객체 인식을 수행한 화면을 나타낸다. 그림 1에서 세로 축은 Precision 및 Recall의 값을 나타내고 가로 축은 학습한 프레임 개수를 나타낸다.

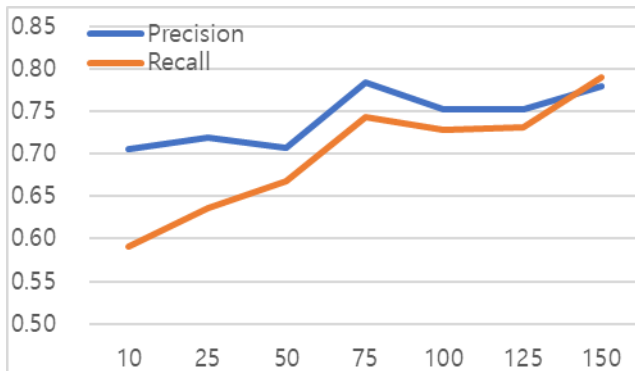


그림 1 학습 데이터 개수에 따른 Precision/Recall 그래프



그림 2 객체 탐지를 수행한 화면

그림 1과 같이 학습에 사용한 데이터가 늘어남에 따라 객체 인식의 정확도가 늘어나는 것을 확인할 수 있었다. Precision의 경우 학습 데이터가 일정 수 이상 늘어나는 경우, 어느정도 정확도에 반영이 되지만 비슷한 데이터양인 경우에는 큰 차이가 없는 모습을 보였다. 이는 객체 검출 과정에서 true로 탐지한 객체들이

실제로 true인 경우가 늘어 정밀도가 증가하였다는 것을 알 수 있다. Recall의 경우 학습 데이터가 늘어나면 대체적으로 늘어나는 경향을 보였으며, Precision 대비 더 큰 폭으로 증가하는 것을 확인할 수 있다. 이는 실제 영상의 객체를 보다 잘 탐지할 수 있게 되었음을 나타낸다. 실제로 T1, T2와 같이 매우 적은 수의 트럭 데이터를 학습한 경우, 대부분의 트럭을 잘 탐지하지 못하였으며, 탐지하더라도 승용차로 판단하는 경우가 많았다. 또한 오토바이의 경우, 전체 경우에서 탐지된 건이 존재하지 않았다. 따라서 최소한의 검출을 위해서는 학습 과정에서 해당 객체를 충분히 포함하도록 진행할 필요가 있다.

### III. 결론

본 논문에서는 학습에 사용하는 데이터 개수가 증가함에 따라 객체 탐지의 성능이 향상되는 것을 실험을 통해 확인하고 이를 통해 자동화 레이블링과 같은 솔루션을 구축함에 있어 필요한 데이터의 개수를 유추할 수 있도록 하였다. 또한 학습한 이미지의 개수와 그에 포함되어 있는 객체의 개수를 함께 제시하고 이에 따른 결과를 제시하여 객체 탐지를 위한 모델을 학습할 때, 최소한의 객체를 포함하여야 하는 것을 보였다. 추후 연구에서는 반자동/자동 레이블링 과정을 구축하고 이에 따른 시간 절감 효과 및 정확도를 분석할 것이다.

### 참 고 문 헌

- [1] 이용, 장래영, 박민우, 이건우, 최명석. “자동-레이블링 기반 영상 학습데이터 제작 시스템,” 한국콘텐츠학회논문지, 21(6), 701-715, 6월, 2021.
- [2] Bochkovskiy, Alexey, Chien-Yao Wang, and Hong-Yuan Mark Liao. “YOLOv4: Optimal Speed and Accuracy of Object Detection,” arXiv preprint arXiv:2004.10934, 2020.
- [3] Redmon, Joseph, et al. “You only look once: Unified, real-time object detection,” Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.