

# YOLOv4 기반의 공장 근로자 안전관리를 위한 학습 데이터 구축과 모델 학습

이태준<sup>1</sup> · 조민우<sup>1</sup> · 송지호<sup>1</sup> · 황철현<sup>2</sup> · 정희경<sup>1\*</sup>

<sup>1</sup>배재대학교 · <sup>2</sup>경북대학교

## Construction of Training Data and Model Training for YOLOv4-based Factory Operation Safety Management

Taejun Lee<sup>1</sup> · Minwoo Cho<sup>1</sup> · Jiho Song<sup>1</sup> · Chulhyun Hwang<sup>2</sup> · Heokyung Jung<sup>1,\*</sup>

<sup>1</sup>Paichai University · <sup>2</sup>kyungbok University

E-mail : marlrero@kakao.com / kcjmw1208@kakao.com / akir35188@gmail.com /

cheolhyun.hwang@gmail.com / hkjung@pcu.ac.kr

### 요 약

산업안전보건연구원에 따르면 2019년 산업재해자 수가 109,242명으로 2018년에 비해 6.8% 증가하였다. 이러한 산업 안전보건 분야는 질병보다 사고가 더 자주 발생하고 있다. 이러한 상황에서 정부와 기업은 건설 시공 분야에서 ICT 기반 현장 안전사고 예방 핵심 기술 개발이 논의되고 있는 실정이다. 이러한 분야에서 최근 컴퓨터 비전과 인공지능을 활용한 기술들이 많이 사용되고 있다. 본 논문에서는 공장 근로자들의 안전관리를 위한 학습 데이터를 구축하고 YOLOv4를 기반으로 모델을 학습시켰다. 이를 통해 공장에서 근로자들의 위험 상황을 예측하는 초기 연구로써 활용할 수 있을 것으로 사료된다.

### ABSTRACT

According to the Institute for Occupational Safety and Health, the number of industrial injuries in 2019 was 109,242, an increase of 6.8% from 2018. In this situation, the government and companies are discussing the development of core technologies for preventing safety accidents on site based on ICT in the field of construction and construction. In these fields, technologies using computer vision and artificial intelligence have recently been widely used. In this paper, we built training data for safety management of factory workers and trained a model based on YOLOv4. It is believed that this can be used as an initial study to predict the risk situation of workers in factories.

### 키워드

Annotation tool, Industrial safety, Object Detection, YOLOv4

### 1. 서 론

2019년 산업안전보건연구원이 발간한 산업재해 통계 현황을 보면 산업재해자 수는 109,242명으로 사고 재해자 수가 86.1%를 차지하고 있다. 이 중 사고 사망자 수는 855명으로 업종별로 건설업, 제

조업, 서비스업 순으로 많았고 사고 형태별로 떨어 짐, 끼임, 부딪힘, 깔림 등의 순으로 많았다. 50인 미만 사업장에서 사망자 수가 77.2%를 차지하고 있다[1].

이에 따라 정부에서는 산업재해 예방을 위해 드론, 3D 스캐너, VR, 센서 데이터를 등을 이용하여 안전사고를 예방하기 위한 로드맵을 세우고 있다 [2]. 본 논문에서는 공장의 근로자에 대한 안전관리

\* corresponding author

를 위한 학습 데이터를 구축하고 YOLOv4를 이용해 모델 학습을 진행하였다.

## II. 본 론

### 2.1 객체 탐지(Object Detection)

객체 탐지는 이미지나 동영상에서 사람이나 사물을 찾는 것을 말한다. 객체 탐지에는 CNN(Convolutional Neural Network)이 활용되는데 CNN은 이미지에서 행렬 곱 연산 대신 컨볼루션(convolution) 연산을 통해 특징을 추출하는 컨볼루션 계층과 특징 맵의 크기를 줄이는 풀링(pooling) 계층으로 구성되어 있다[3]. 특히 풀링 계층은 이미지의 크기를 줄이면서 특징을 그대로 유지하는 장점이 있지만, 원본 이미지의 크기가 줄어들어 위치 정보를 잃는다는 단점이 존재한다[4]. 이는 객체 탐지나 의미론적 분할(semantic segmentation), 인스턴스 분할(instance segmentation) 분야의 애플리케이션에서 치명적인 단점이 되며 이를 해결하고 정확도를 높이거나 빠르게 탐지하도록 다양한 모델이 나오고 있다.

객체 탐지는 객체의 위치를 찾아야 하는 지역화(localization)와 객체를 구별하는 분류(classification) 작업으로 구성되어 있다. 이를 동시에 수행하는 방식이 One-stage 방식이고 순차적이며 거칠고 미세한(coarse-to-fine) 방식으로 진행되는 방식이 Two-stage 방식이 있다[5]. One-stage 방식은 YOLO, SSD, RetinaNet 등이 있고, Two-stage 방식은 R-CNN, Fast R-CNN, Faster R-CNN 등이 있다.

### 2.2 YOLOv4

YOLO(You Only Look Once)는 One-stage 방식의 모델로 구조를 단순화하여 빠르게 동작할 수 있도록 하는 모델이다[6]. YOLOv4로 누구나 빠른 속도로 훈련하여 효율적이고 적합한 최신 기법들을 적용한 모델이다. 다양한 데이터 증강 기법과 규제, 손실 함수를 실험하였고 구조적으로 수용장(receptive field)을 증가시키거나 어텐션 모듈(attention module)을 적용했을 때, 특징을 통합하기 위한 여러 방안에 대해서도 실험하였다. 또한, 다양한 배치 정규화(batch normalization) 기법과 활성화 함수, 후처리 과정에 대해서도 실험하였다.

YOLOv4는 컨볼루션 계층 내에 작은 그룹을 적용한 53개의 계층이 있는 DarkNet을 백본(backbone)으로 사용하였으며 수용장을 늘리기 위해 SPP(Spatial Pyramid Pooling)를 적용하였다. 또한, 다양한 크기(multi-scale)의 객체를 추출하기 위해 PANet을 사용하였고 머리(Head) 부분에 YOLOv3를 적용하였다[7].

기존의 모델들은 ReLU 활성화 함수를 많이 사용했지만, YOLOv4는 연속적으로 미분 가능한 함수인 Mish를 사용하였다. 그림 1의 왼쪽 그래프는

ReLU와 Mish 함수를 비교한 것이고 오른쪽 그래프는 ReLU와 Mish의 도함수를 비교한 것이다.

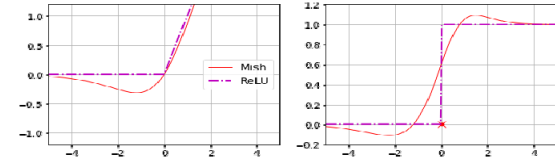


그림 1. Graph of ReLU and Mish functions

### 2.3 학습 데이터 구축과 모델 학습

학습 데이터는 총 120장의 샘플 이미지로 공장 CCTV로 직접 촬영한 데이터를 사용했으며 어노테이션 툴(annotation tool)은 CVAT(Computer Vision Annotation Tool)를 이용하였다. CVAT는 Django 웹 기반의 이미지, 비디오에서 객체를 라벨링(labeling)하는 툴로 오픈 소스이다[8].

훈련 데이터와 테스트 데이터는 100:20으로 나뉘었으며 객체는 근로자, 물체(공장 복도에 있는 사물), 위험지역(밀링이나 절삭 기계와 같은 근로자가 다루는 기계)으로 분류하였다. 그림 2는 CVAT로 이미지를 라벨링하는 화면이다.

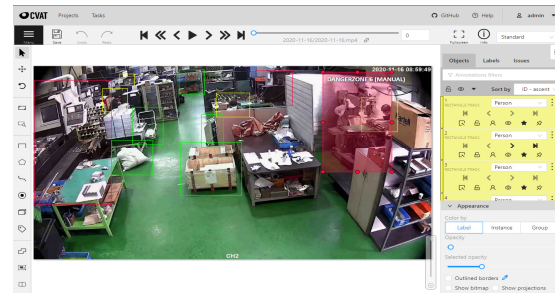


그림 2. Image labeling using CVAT

크기는 512×512에서 416×416을 입력 이미지로 사용하였고 초기 학습률(learning rate)은 0.0013, 모멘텀(momentum)은 0.949, 학습률 감소는 0.0005, 배치(batch) 사이즈는 16, 데이터 증강 기법으로 모자이크(mosaic)를 사용하였다. 이 외의 IoU 손실이나 NMS(Non-Maximum Suppression) 설정, 각 계층의 보폭(stride)과 필터의 개수 등은 YOLOv4 기본값을 사용하였다. 표 1은 학습한 하드웨어 사양이다.

표 1. System specification

OS	Ubuntu 20.04.1 LTS
CPU	Intel i7-10700 (2.9Ghz, Core x16)
GPU	Nvidia Geforce GTX 1660
RAM	16GB
Storage	SSD 250GB & HDD 4TB

### 2.4 학습 결과

훈련 반복 횟수(epoch)는 6,000번 진행했다. 평균 손실 값은 0.3684가 나왔고 학습 결과는 그림 3과 같다. 그림 3의 x축은 파란색 선은 손실 값이고 빨간색 선은 mAP(mean Average Precision)를 의미한다. 최종 mAP는 81%이다.

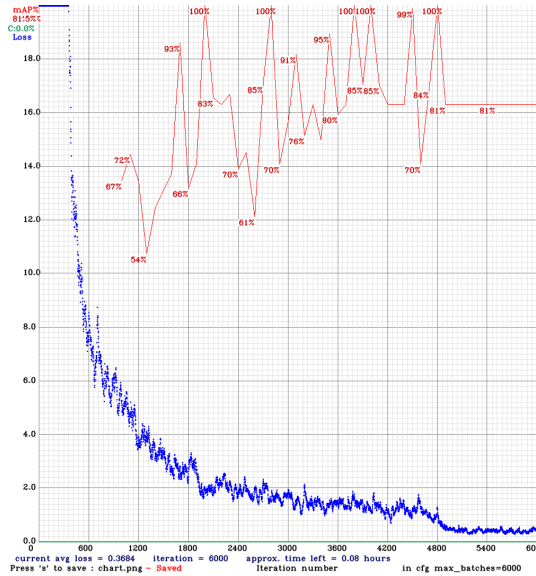


그림 3. Graph of mAP and loss value

### III. 결 론

산업재해는 매년 꾸준히 발생하고 있고 이에 따라 본 논문에서는 산업재해 예방을 위한 학습 데이터의 일부를 이용하였다. YOLOv4 모델 학습을 진행하여 mAP 81%의 결과를 얻었다. 이를 통해 근로자들의 위험 상황을 예측하는 초기 연구로써 활용할 수 있을 것으로 사료한다. 하지만 적은 수의 데이터를 이용해 학습을 진행하였다는 점과 위험 감지를 위한 의미론적인 추론 결과를 보이지 못했다는 점에서 아쉬움이 있었다. 향후 연구에서 이러한 문제를 보완하는 연구가 필요할 것으로 보인다.

### Acknowledgement

This work was supported by the Korea Medical Device Development Fund grant funded by the Korea government (the Ministry of Science and ICT, the Ministry of Trade, Industry and Energy, the Ministry of Health & Welfare, the Ministry of Food and Drug Safety) (Project Number: P0015365).

### References

- [1] Occupational Safety & Health Reserch Institutio ns(OSHRI). 2019 Industrial Accident Statistics [Internet]. Available: <https://oshri.kosha.or.kr/oshri/publication/industryDisasterStatistics.do?mode=list&&articleLimit=10&article.offset=10>.
- [2] Ministry of Land, Infrastructure and Transport. “Smart Construction Technology Road-map for Innovation in Construction Productivity and Enhancement of Safety,” Oct. 2018.
- [3] I. Goodfellow, Y. Bengio and A. Courville, “Convolutional Networks,” in *Deep Learning*, The MIT Press, ch. 9, pp. 330-372, 2017.
- [4] S. J. Park, J. H. Han and Y. S. Moon, “Efficient Deep Neural Network Architecture based on Semantic Segmentation for Paved Road Detection,” *Journal of the Korea Institute of Information and Communication Engineering*, vol. 24, no. 11, pp. 1437-1444, Nov. 2020.
- [5] Z. Zou and Z. Shi, “Object Detection in 20 Years: A Survey,” In *arXiv preprint arXiv:1905.05055*, 2019.
- [6] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, “You Only Look Once: Unified, Real-Time Object Detection,” In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 779-788, 2016.
- [7] A. Bochkovskiy, C.-Y. Wang and H.-Y. M. Liao, “YOLOv4: Optimal Speed and Accuracy of Object Detection,” In *arXiv preprint arXiv:2004.10934*, 2020.
- [8] Openvinotoolkit. CVAT Github [Internet]. Available: <https://github.com/openvinotoolkit/cvat>.