

<http://dx.doi.org/10.17703/JCCT.2022.8.5.735>

JCCT 2022-9-92

## YOLOv5 및 OpenPose를 이용한 건설현장 근로자 탐지성능 향상에 대한 연구

### A Study on the Improvement of Construction Site Worker Detection Performance Using YOLOv5 and OpenPose

윤영근\*, 오태근\*\*

Younggeun Yoon\*, Taekeun Oh\*\*

**요약** 건설업은 사망자 수가 가장 많이 발생하는 산업이며, 다양한 제도 개선에도 사망자는 크게 줄어들지 않고 있다. 이에 따라, CCTV 영상에 인공지능(AI)을 적용한 실시간 안전관리가 부각되고 있다. 건설현장의 영상에 대한 AI를 적용한 근로자 탐지연구가 진행되고 있지만, 건설업의 특성상 복잡한 배경 등의 문제로 인해 성능 발현에 제한이 있다. 본 연구에서는 근로자의 탐지 및 자세 추정에 대한 성능 향상을 위해 YOLO 모델과 OpenPose 모델을 융합하여, 복잡 다양한 조건에서의 근로자에 대한 탐지 성능을 향상시켰다. 이는 향후 근로자의 불안전안 행동 및 건강관리 측면에서 활용도가 높을 것으로 예상된다.

**주요어** : 건설업, 근로자 탐지, 자세 추정, 딥러닝, YOLO, OpenPose

**Abstract** The construction is the industry with the highest fatalities, and the fatalities has not decreased despite various institutional improvements. Accordingly, real-time safety management by applying artificial intelligence (AI) to CCTV images is emerging. Although some research on worker detection by applying AI to images of construction sites is being conducted, there are limitations in performance expression due to problems such as complex background due to the nature of the construction industry. In this study, the YOLO model and the OpenPose model were fused to improve the performance of worker detection and posture estimation to improve the detection performance of workers in various complex conditions. This is expected to be highly useful in terms of unsafe behavior and health management of workers in the future.

**Key words** : Construction, Work Detection, Pose Estimation, Deep Learning, YOLO, OpenPose

#### 1. 서 론

최근 국내의 건설현장은 노령화로 인한 내국인 숙련 인력이 감소하고 있으며, 무경력, 비전문 내국인 및 외국인들의 유입에 따라 안전관리의 변화가 필요한 시점이다. 고용노동부에서 발표한 산업재해 사고사망 현황

(표 1)에서 2021년 전체 사고사망자는 828명이며, 건설업의 사고사망자는 50.4%(417명)로 2020년 대비 41명 감소하였지만, 여전히 전 산업의 절반을 차지하고 있다 [1]. 공사금액별로는 50억 미만의 소규모 건설현장에서 건설업 전체 사고사망자의 71.5%가 발생하였다 [1]. 대기업 시공사의 경우 중대재해처벌법 시행 등에 따라

\*정회원, 인천대학교 안전공학과 연구원 (제1저자)

\*\*정회원, 인천대학교 안전공학과 교수 (교신저자)

접수일: 2022년 8월 30일, 수정완료일: 2022년 9월 5일

게재확정일: 2022년 9월 9일

Received: August 30, 2022 / Revised: September 5, 2022

Accepted: September 9, 2022

\*\*Corresponding Author: thoh@inu.ac.kr

Dept. of safety engineering, Incheon National Univ, Korea

현장의 안전관리를 더욱 강화하고 있는 반면에 중소규모 현장은 근로자의 낮은 인식과 관리자의 부재에 따른 사각지대에서 지속적인 후진국형 사고가 발생하고 있으므로 고질적 안전문제 해결을 위한 새로운 안전관리 방안의 도입이 필요하다.

표 1. 2017-2021 유형별 사망자 통계

Table 1. Fatalities statistics by types in 2017-2021

| 구 분             | 2017 | 2018 | 2019 | 2020 | 2021 |
|-----------------|------|------|------|------|------|
| 업무상<br>사고사망자(명) | 964  | 971  | 855  | 882  | 828  |
| 건설업             | 506  | 485  | 428  | 458  | 417  |
| 제조업             | 209  | 217  | 206  | 201  | 184  |
| 기타업종            | 249  | 269  | 221  | 223  | 227  |
| 사고사망만인율(‰)      | 0.53 | 0.52 | 0.51 | 0.46 | 0.43 |

현재 국내외에서 4차 산업혁명 시대를 맞이하여 다양한 스마트 건설기술(BIM, 드론, ICT 등)과 스마트 건설안전기술(IoT 기반의 근로자 안전관리)에 대한 연구를 활발히 진행되고 있다 [2]. 이 중 건설현장 근로자의 안전관리를 위한 스마트 건설안전기술은 아직 초기 단계로 민간기업의 주도하에 현장의 기술 도입 필요성에 대한 의견을 수렴하여 기술개발이 진행되고 있다. IoT를 활용한 다양한 스마트 건설안전기술 중에서 CCTV에 의한 안전관리가 효과가 있는 것으로 분석되었으며 [3], 스마트 건설안전기술 보급업체들은 인공지능 스타트업과 협업하여 사각지대 해소 및 건설현장 관리자들의 부재 시 실시간 안전관리가 가능하도록 인공지능(AI) 기술을 탑재한 CCTV를 개발 및 보급하고 있다. 하지만, 근로자(객체) 탐지기술은 건설현장의 복잡, 다양한 배경하에 제대로 된 성능을 발휘하지 못하고 있으며, 이러한 문제 해결을 위한 다양한 방법의 연구가 필요한 시점이다.

본 연구에서는 건설현장의 CCTV를 통해 수집되는 영상 데이터에 대해 근로자의 탐지 및 자세 추정 딥러닝 모델의 현장 적용성 저하 요인을 해결하고자 한다. 먼저, 영상분석을 위한 최신 딥러닝 모델 중 YOLO 모델과 OpenPose 모델의 일반적인 상황에서 성능을 테스트하고, 건설현장 영상에 적용을 통한 한계점 분석 및 문제 해결을 위한 모델 융합을 통해 복잡, 다양한 배경 조건에 따른 탐지 성능 저하 문제를 해결하기 위한 융합 절차와 개선된 모델을 제시하고자 한다. 이를 통해

중소규모 건설현장의 관리자 부재 등에 따른 사각지대에 대한 관리방안으로 AI CCTV 도입을 통해 해결될 수 있으며, 향후 근로자의 불안정한 행동 및 작업자세 관리를 통한 건강관리 측면의 연구와 연계될 수 있을 것으로 기대된다.

## II. 연구방법

### 1. 딥러닝 방법

#### 1.1 YOLO

YOLO(You Only Look Once) 알고리즘[4]은 영상 내에 객체 탐지를 위해 사용되는 딥러닝 방법 중 한 가지이다. 여러 방법 중에서 YOLO 방법은 높은 프레임 레이트(FPS) 성능을 가지므로 실시간 객체 탐지에 적합한 알고리즘으로 알려져 있다. 이 중 가장 최근에 제안된 YOLOv5의 4가지 모델의 성능은 그림 1과 같이 구분된다 [5]. 본 연구에서는 처리속도가 빠르고 다른 모델에 비해 연산과정이 다소 적은 YOLOv5s 모델이 사용되었다.

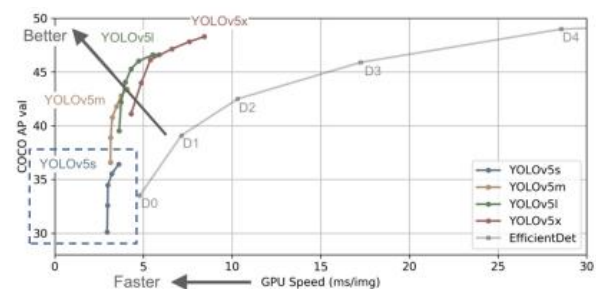


그림 1. YOLOv5 모델별 성능

Figure 1. Performance of YOLOv5 models

#### 1.2 OpenPose

오픈포즈(OpenPose)는 인간의 자세를 예측(Human Pose Estimation)하는 한 가지 방법으로 사람의 얼굴, 몸, 손과 발 등의 관절을 찾아내는 방법이다. 과거의 포즈 추정 방법은 Top-Down 방식으로 사람(객체)을 검출하고 검출된 사람의 자세를 추정하는 방식이었지만, 최신의 OpenPose 방법은 영상 내의 모든 사람의 신체 부위(코, 왼쪽 팔꿈치 등)를 먼저 탐지하고 Bottom-up 방식으로 신체 부위의 연결 가능한 쌍을 기반으로 관절을 조립하여 자세를 추정하는 방법으로 발전하였다 [6]. 영상에서 신체 부위를 식별하기 위해 OpenPose는 입력

이미지에서 신체 부위에 대한 히트맵 및 위치 및 방향 정보로 구성되어 모션을 표현할 수 있는 2차원 좌표계인 part affinity field(PAF)를 예측하는 사전 훈련된 신경망을 사용한다 [7]. 과거에는 출력 관절이 15개로 시작하여, 18개, 25개까지 관절을 세분화하는 방향으로 알고리즘이 개선되고 있다. 본 연구에서는 COCO 데이터 세트를 활용한 18개의 관절을 탐지하여 자세를 추정하는 모델이 활용되었다.

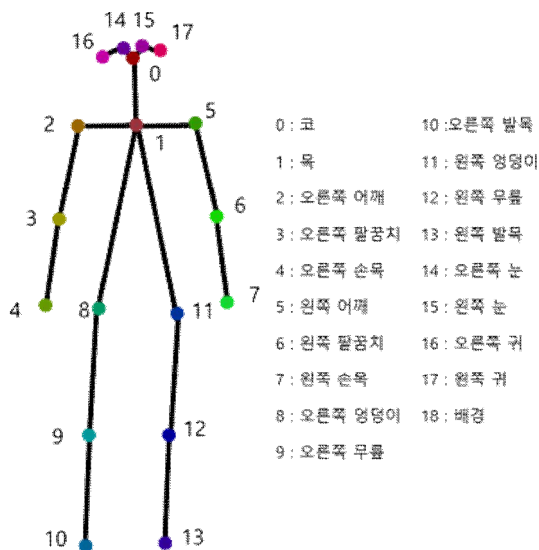


그림 2. OpenPose 학습 데이터 세트(COCO 18parts)  
Figure 2. OpenPose training dataset (COCO 18parts)

## 2. 연구 방법 및 절차

본 연구는 그림 3과 같이 총 4단계의 절차에 따라 수행되었다. 1단계는 영상 데이터 수집 단계로 00건설의 현장 협조를 통해 수집된 실제 현장 작업 영상과 AI 허브에서 (주)미디어그룹사람과숲이 제공한 고소작업 현장 실시간 영상 자료(떨어짐 사고 묘사) 50장을 수집 및 활용하였다. 2단계는 YOLOv5s 및 OpenPose 상용 모델을 활용하여 일반 영상에 대한 기본적인 성능을 검증하는 단계이다. 촬영자의 포즈(정면, 옆면), 영상 촬영 각도, 마스크 착용 등과 같은 영상을 활용하여 일반적인 성능을 테스트하였다. 3단계는 동일한 모델을 활용하여 건설현장의 영상에 대해 검증을 하는 단계이다. 일반 영상과 비교할 때 건설현장 영상에서 성능이 저하되는 복잡 다양한 배경 및 다수의 근로자를 포함하는 영상에 대해 대표적 영상 3개를 선정하고 예측 오류를 일으키는 여러 요인을 도출하였다. 4단계는 3단계에서

도출된 문제점의 해결을 위해 YOLOv5s 모델과 Open Pose 모델을 융합하는 단계이다. 딥러닝 모델에 따라 law data로 읽어 들이는 Pixel의 차이가 있으므로 이를 조율하여 융합모델의 탐지 결과를 원본 영상에 대입하여 최종 결과를 도출한다.

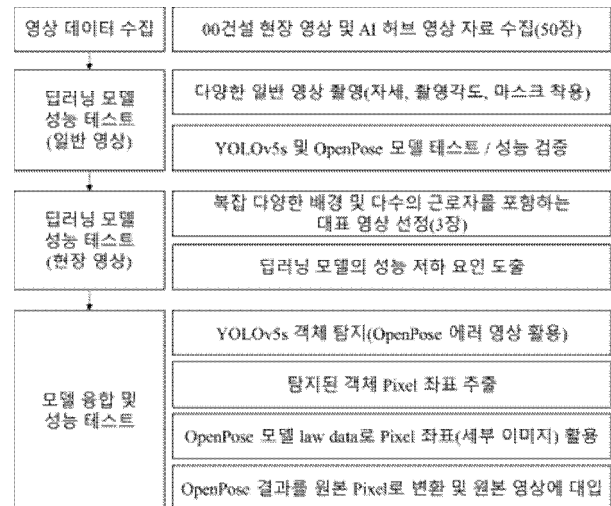


그림 3. 영상 분석 및 모델 융합 절차  
Figure 3. Image analysis and model fusion procedure

## III. 연구결과

### 1. YOLOv5s 및 OpenPose의 일반 성능 테스트

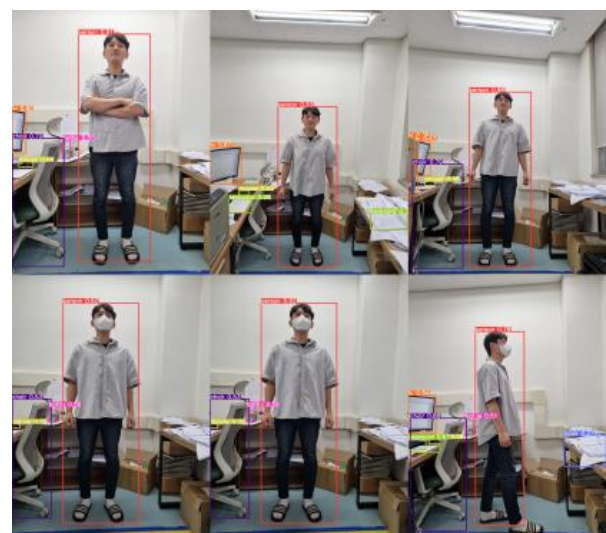


그림 4. YOLOv5s의 객체 탐지 성능(일반 영상)  
Figure 4. Object detection performance of YOLOv5s(normal video)



그림 4에서 YOLOv5s를 이용하여 일반 영상에 대한 촬영자 탐지를 통해 객체 탐지 성능을 확인하였다. 마스크 착용 여부, 촬영 각도 등과는 무관하게 90% 이상의 정확도를 가졌으며, 옆 모습의 경우 78%의 정확도를 보이는 것으로 분석되었다. 사람에 대한 객체 탐지의 경우 다양한 분야에서 방대한 양의 영상을 통해 연구되고 있어 높은 정확도를 가지는 것으로 분석된다.

그림 5는 OpenPose를 이용하여 일반 영상에 등장하는 촬영자의 자세를 추정한 결과이다. 촬영 각도, 마스크 착용 등의 영상에서도 문제없이 모든 관절을 적절히 추정하는 것을 볼 수 있다. 하지만, 옆 모습의 경우 0(코), 14(오른쪽 눈), 15(왼쪽 눈), 16(오른쪽 귀), 17(왼쪽 귀)에 대한 관절 조립이 미흡한 것으로 분석되었다.



그림 5. OpenPose의 자세 추정 성능(일반 영상)  
Figure 5. Posture estimation performance of OpenPose(normal video)

일반 영상에 대한 2가지 모델의 검증을 통해 충분히 활용가능 함을 확인하였지만, 건설현장 영상의 경우 일반영상 보다 배경 조건이 복잡함에 따라 성능이 저하될 수 있다. 따라서, 해당 결과에 기반하여 건설현장 영상의 탐지 정확도를 분석하고, 건설현장 근로자의 탐지에 부합하도록 모델을 개선할 필요가 있다.

## 2. 건설현장 영상에 대한 성능 검증 및 한계점 도출

그림 6은 건설현장 영상에 대해 YOLOv5s 및 Open Pose 모델을 통한 탐지 결과가 적절한 경우에 대한 샘플이다. 많은 영상 데이터 중 대부분의 정면 영상은 근로자

탐지성능이 70% 이상이었으며, 자세 추정의 경우 가려진 부위를 제외하고는 대부분의 관절을 탐지하는 것으로 분석되었다. 하지만, 옆모습의 경우에는 일반 영상 분석과 동일하게 목 윗부분 부터는 탐지가 안되는 경우가 발생한다. 현장의 분진 등에 의해 근로자 대부분이 안면마스크를 착용함에 따른 코 등에 대한 특징 추출이 미흡하여 발생한 것으로 사료된다.



그림 6. 적절한 근로자 탐지 및 자세 추정 결과(건설 현장 영상)  
Figure 6. Appropriate worker detection and posture estimation results (construction site video)



그림 7. 부적절한 근로자 탐지 및 자세 추정 결과(건설 현장 및 AI 허브 영상)

Figure 7. Inappropriate worker detection and posture estimation results (construction site and AI hub video)

그림 7은 자세 추정 결과에 오류가 있는 영상에 대한 샘플이다. 이는 얼굴 부분을 제외한 다른 관절 부위가 탐지가 안되거나 배경조건에 의해 엉뚱한 곳에서 관절이 탐지되는 경우를 보여준다. 분석 결과, 안전고리, 하늘 배경, 철근, 자재 등에 대한 빛 또는 조명에 의한 빛 반사 등의 요인으로 인해 탐지 성능이 저하되는 것으로 사료된다.

CCTV 영상을 안전관리에 활용하는 경우는 넓은 범위를 한번에 탐지 및 식별하기 위함이므로 지엽적인 부분만 관찰하는 것은 활용도가 낮으므로 넓은 범위의 영상에 대한 그림 7의 오류를 해결하여 성능을 향상시킬 필요가 있다.

### 3. 제안한 방법에 따른 개선 결과

앞선 분석을 통해 빛이나 조명에 의한 주변의 빛 반사 등으로 인해 특징추출 성능이 저하되어 OpenPose의 자세 추정에 오류가 있음이 확인되었다. 하지만, YOLOv5s 모델의 경우에는 보다 정확하게 근로자(객체)를 탐지하는 것으로 분석되었다. 따라서, 본 연구를 통해 그림 3의 마지막 단계와 같은 모델 융합 절차를 제안하고자 한다. 이 방식은 Top-Down 방식으로 볼 수 있지만, 탐지성능이 높은 객체 탐지 모델과 융합한다는 점에 있어서 차별성이 있으며, Bottom-up 방식의 OpenPose 모델만을 사용하는 것보다 정확도가 더 높을 것으로 예상된다.

YOLOv5s를 통해 객체를 탐지하고, 해당 객체의 boxes에 대한 좌표를 추출하여 개별로 OpenPose의 raw data로 활용하여 분석을 진행하였다. 본 방법을 통해 예측한 결과가 그림 8에 제시되었다. 기존의 단일 모델만을 사용한 경우 그림 7과 같이 근로자의 관절 미식별, 빛 반사가 강한 곳에서 탐지 등과 같은 예러가 발생했지만, 제안한 방법을 통해서 배경조건을 사전에 제외함으로써 모든 관절을 정확하게 탐지할 수 있는 방법으로 모델이 개선되었다. 정량적으로 첫 번째 영상은 다수의 근로자 중 중앙에 위치한 근로자의 미탐지된 관절 3개가 적절히 탐지되어, 융합모델에서 탐지된 관절의 수/단일 OpenPose 모델에서 탐지된 관절의 수를 비교할 때 5% 정도의 정확도가 향상되었다. 복잡한 배경 조건에 따라 근로자 이외의 물체를 근로자로 탐지하는 두 번째 및 세 번째 영상의 경우 영상 내 근로자 수/단일 OpenPose 모델, 융합모델로 탐지된 객체 수를 비교

하였을 때 단일 모델은 40%, 25%의 정확도이지만, 융합모델은 100%의 정확도를 보여 충분히 활용 가능할 정도로 개선된 것으로 판단된다. 2가지의 딥러닝 모델을 융합하는 것은 방대한 연산량을 요구할 수 있지만, 이를 고려하여 실시간 탐지에 가장 많이 활용되며, 탐지성능이 높고 연산량이 낮은 YOLOv5s 모델과 융합함으로써 해당 문제를 다소 해결할 수 있다.



그림 8. 융합 모델의 자세 추정 개선 결과  
Figure 8. Pose estimation improvement result of fusion model

본 연구에서는 상용모델을 활용하여 최소한의 파라미터 수정을 통해 근로자의 자세 추정에 대해 관절의 탐지 정확도를 높이는 방안의 연구 위주로 진행하였으며, 해당 결과를 활용하여 근로자의 정상적인 자세와 불안정한 자세, 사고 발생 시 기절한 경우 등에 대한 탐지가 가능할 것으로 예상되며, 이 외에도 여러 분야에서 추정된 자세를 활용한 다양한 연구가 가능할 것으로 예상된다.

#### IV. 결 론

건설업의 사고사망자 수를 줄이기 위해서는 기존의 안전관리 방안 외에도 최신 기술을 도입하여 현장의 안전사고를 사전에 파악하는 것이 요구되는 상황이다. 본 연구에서는 실시간 안전관리를 위해 활발히 연구되고 있는 분야인 영상 데이터에 대한 딥러닝 모델을 이용하여 건설현장 근로자의 탐지 및 자세 추정에 대해 분석하였다. 분석 결과 건설현장의 배경 조건에 따라 성능을 저하시키는 요인을 파악하고, 이에 대한 개선을 위해 융합모델을 제안하였다. 본 연구 결과는 다음과 같이 요약될 수 있다.

- YOLOv5s 모델 단일로 건설현장 영상의 근로자를 탐지한 결과 70% 이상의 정확도를 보였다. 해당 모델은 영역을 구분하여 예측함에 따라 배경조건의 제한이 다소 적으며, 근로자에 대한 머리, 팔, 발 등의 끝단을 정확하게 예측하는 것으로 분석되었다.

- OpenPose 모델로 건설현장 영상을 추정한 결과 빛 또는 조명에 의한 빛 반사 부분에서 강한 특징이 추출되어 오류가 발생하는 것으로 분석되었다.

- Bottom-up 방식의 OpenPose 모델은 멀티 추정에 용이하지만, 건설현장의 다양한 배경에는 적합하지 않으므로 배경조건에 제한을 덜 받고 탐지 속도가 빠른 YOLOv5s 모델과의 융합 절차를 제안하였으며, 융합 모델은 단일 OpenPose 모델보다 높은 정확도로 건설현장 영상에서 자세를 추정할 수 있었다.

본 연구에서는 복잡하고 다양한 조건의 배경을 가지는 건설현장 영상에서 근로자의 자세 추정에 대한 탐지 성능을 향상시키는 모델을 제안하였다. 개발된 모델과 방법론은 건설산업에서 향후 근로자의 작업 시 불안정한 행동 예측 및 반복작업 시 보건관리 등 다양하게 추가 연구가 가능할 것으로 예상되며, 타 산업인 제조업 등에서

인간공학적 측면에서 활용도가 높을 것으로 예상된다.

#### References

- [1] Ministry of Employment and Labor, "2017-2021 Industrial Accident and Death Statistics", 2022.
- [2] Ministry of Land, Infrastructure and Transport, "Smart construction technology development project", 2019.
- [3] Y. S. Kim, T. K. Oh, C. S. Kim, N. E. Lee, C. S. Hong, S. Y. Lee, Y. G. Yoon, "A Study on the Actual Condition Analysis and Activation Plan of Smart Construction Safety Technology by the Survey", Journal of the Korean Society of Safety, Vol. 37, No. 1, pp. 30-40, 2022.
- [4] 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, pp.779-788, 2016.
- [5] Ultralytics. YOLOv5 [Internet] <https://github.com/ultralytics/yolov5>
- [6] Cao, Zhe, Gines Hidalgo, Tomas Simon, Shih-En Wei, and Yaser Sheikh. "OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields." ArXiv:1812.08008 [Cs], May 30, 2019. <https://arxiv.org/abs/1812.08008>.
- [7] Osokin, Daniil. "Real-Time 2D Multi-Person Pose Estimation on CPU: Lightweight OpenPose." ArXiv:1811.12004 [Cs], November 29, 2018. <https://arxiv.org/abs/1811.12004>.

※ 이 논문은 2021년 및 2022년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(No. 2021R1A6A3A01086763, No. 2022R1I1A1A01061658).