

딥러닝 기반 객체 탐지를 활용한 엔진 유압 제어 장치 호스 조립 공정 검사 자동화 시스템

김창영¹ · 조현건¹ · 최준혁² · 이수동^{1*}

¹울산대학교 산업경영공학부 / ²경희대학교 산업경영공학과

An Automated Inspection System for Hose Assembly of Engine Main Control Valves Using Deep Learning-based Object Detection

Changyeong Kim¹ · Hyungun Cho¹ · Junhyeok Choi² · Sudong Lee¹

¹Department of Industrial Engineering, University of Ulsan

²Department of Industrial and Management Systems Engineering, Kyung Hee University

In many manufacturing sites, the completion of the process is examined by a manual visual inspection, which is costly and prone to human errors. To overcome this, we propose an automated inspection system for hose assembly of engine main control valves using deep learning-based object detection. CCTVs capture the operator's real-time work process, and the hose connection points and work gears are automatically recognized to determine whether the assembly is completed at the end of the process. We trained a deep learning model based on the Normalized Gaussian Wasserstein Distance and YOLOv5 to detect small objects accurately. The images collected from multiple cameras at different angles were processed in parallel using multi-camera processing. The experimental results prove the applicability of our proposed method, as the object detection model detects all classes correctly.

Keywords: Automated Inspection, Object Detection, Deep Learning

1. 서론

제조 공정의 생산성을 높이기 위해서는 공정의 불량률을 낮춰야 한다. 작업자의 실수 혹은 망각으로 발생하는 전체 인적 오류는 제조 품질 불량률의 주요 원인 중 하나로, 조립 생산 시스템에서 발생하는 품질 결함의 약 70~90%는 직간접적으로 사람에게 의해 발생하는 것으로 알려져 있다(Yang *et al.*, 2012; Wang *et al.*, 1997). 이를 해결하기 위해 제조 공정에서의 인적 오류 모니터링에 관한 연구가 진행되어 왔으며(Lee *et al.*, 2006; Yoo *et al.*, 2008), 최근 딥러닝(deep learning) 기술의 발전에 따라 컴퓨터 비전과 공정 모니터링 기술을 결합한 연구가 주목받고 있

다(Malburg *et al.*, 2021).

본 연구는 건설 중장비용 엔진의 유압 제어 장치(main control valve, MCV) 호스 조립공정 검사에 딥러닝 기반 객체 탐지(object detection)를 적용하여, 인적오류로 인한 조립 누락을 자동으로 탐지하는 시스템을 제안한다. MCV는 굴착기 작동 동력인 유압을 제어하는 장치이며, 건설기계의 엔진 동력을 전환한 유압을 각 부에 공급하여 기계를 실제로 작동시키는 주요 부품이다. 본 연구에서는 MCV의 호스 조립 공정에서 발생할 수 있는 작업자의 인적오류에 주목하였다. MCV 조립 공정은 <Figure 1>과 같다.

첫 번째 단계로 작업자는 MCV(<Figure 1(a)>) 호스 연결부

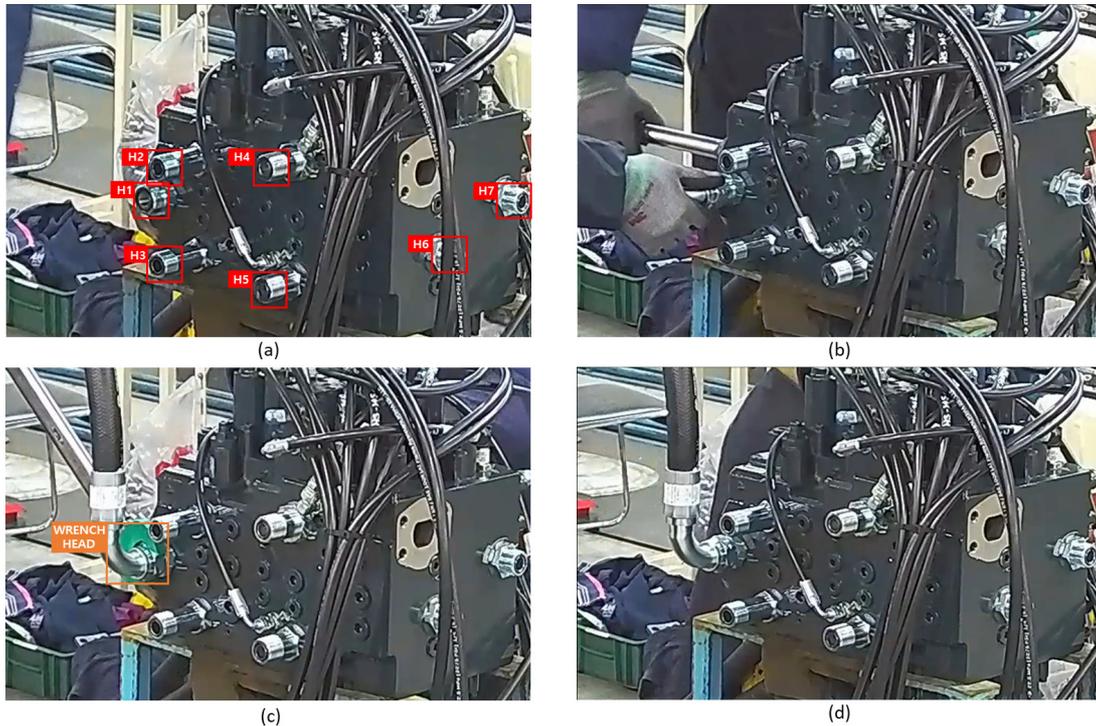


Figure 1. The Hose Assembly Process for an MCV

(connection point)에 수작업으로 호스를 체결(<Figure 1(b)>)한다. 이를 ‘가조립 상태’라 정의한다. 가조립 상태의 연결부를 렌치로 조이면(<Figure 1(c)>) ‘조립 완료 상태(<Figure 1(d)>’가 되며 조립공정이 최종 완료된다. 조립 완료된 제품은 다음 공정으로 넘어가기 전에 작업자의 육안검사를 거친다. 실제 작업 현장에서는 MCV 이외에 다양한 제품이 존재하며 작업 순서는 매뉴얼화 되어 있지 않다. 따라서, 본 연구에서 언급하는 렌치 체결 누락은 특정 공정의 순서와는 무관하게 발생할 수 있다. <Figure 2>는 <Figure 1>의 조립공정의 연결부를 확대한 그림이다. <Figure 2>의 가조립 상태(<Figure 2(b)>)와 조립 완료 상태(<Figure 2(d)>)는 외형이 유사하기 때문에 검사자의 육안으로 구분하기 어렵다. 또한, 육안검사의 특성상 다양한 통제 불가능한 요인에 의한 인적오류에 취약하다. 따라서, 본 연구에서는 딥러닝 기반 객체 탐지를 활용한 MCV 호스 조립 공정 검사 자동화 시스템을 개발하고자 한다. 제안하는 방법론의 주요 내용은 아래와 같다.

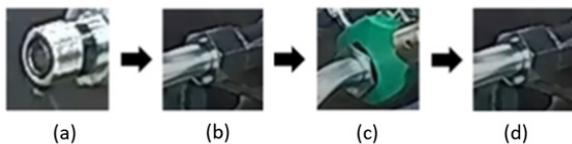
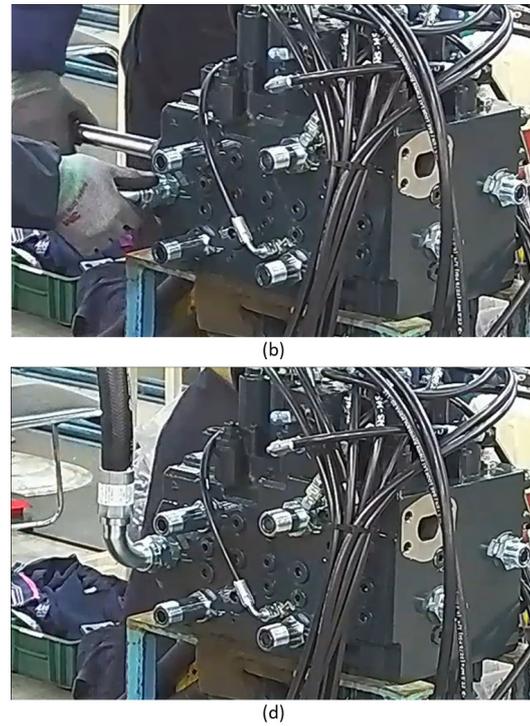


Figure 2. The Assembly Process at a Hose Connection Point

- 이미지상 외형이 동일한 가조립 상태(<Figure 2> (b))와 조립 완료 상태(<Figure 2> (d))를 자동으로 구분하기 위



한 작업자 행동 인식 기반의 딥러닝 객체 탐지 모델을 제안한다.

- 전체 이미지에서 매우 작은 면적을 차지하는 소형 객체인 호스 연결부를 정확하게 탐지하기 위해 Normalized Gaussian Wasserstein Distance (NGWD) 기반의 소형 객체 탐지 방법론을 적용하였다.
- 작업자 동선에 의한 촬영 간섭으로 발생하는 사각지대 문제를 해결하기 위해 여러 각도에서 촬영된 영상에 대한 객체 탐지 결과를 결합하는 멀티 카메라 프로세싱 (multi-camera processing)을 적용한다.

본 논문의 구성은 다음과 같다. 제2장에서는 연구배경 및 관련 연구를 소개한다. 제3장에서는 제안하는 MCV 호스 조립 공정 검사 자동화 시스템을 설명하고, 제4장에서 실험을 통해 제안방법론의 성능을 평가한다. 제5장에서는 결론 및 향후 연구 방향을 제시한다.

2. 연구배경 및 관련 연구

2.1 객체 탐지

객체 탐지는 이미지 내 객체를 감지하고 레이블을 부여하는 작업으로, 객체의 위치를 추론하는 바운딩 박스(bounding box)의 생성과 객체의 클래스를 분류하는 과정으로 이루어진다 (Tang *et al.*, 2023). 본 연구에서는 작업 상황을 카메라로 촬영

한 실시간 영상에 딥러닝 기반 객체 탐지 모델을 적용하여 각 연결부의 조립상태를 실시간으로 모니터링하고 작업 후 누락 여부를 자동으로 판단해 작업자에게 전달하기 위한 시스템을 제안한다.

객체 탐지를 위한 딥러닝 접근 방식은 두 가지로 나뉜다 (Rajeshwari *et al.*, 2019). 첫 번째 방식은 객체의 영역 추출과 분류를 단계적으로 해결하는 다단계(multi-stage) 접근 방식이다. 대표적인 연구로 RCNN (region-based convolutional neural networks), Fast RCNN, Mask RCNN 등이 있다. 영역 제안(region proposal), 특성 추출(feature extraction), 분류(classification)를 단계적으로 진행하는 이러한 방식은 계산 비용이 높다는 단점을 가지고 있다. 두 번째 접근법은 단일 단계(one-stage) 접근 방식이다. 다단계 접근 방식의 영역 제안 단계를 생략하고 이미지의 각 샘플 영역에서 객체 탐지를 직접 수행한다. YOLO(You Only Look Once) 계열의 방법론들이 단일단계 접근의 대표적인 예시이다. YOLO 기반 모델은 객체의 영역을 다차원적으로 분리하고 클래스 확률을 적용하여 하나의 회귀 문제로 처리한다는 점에서 기존의 다단계 접근 방식과 대조적이다(Girshick *et al.*, 2014; Girshick, 2015). 이러한 특징으로 인해 YOLO 모델의 객체 탐지 속도는 R-CNN 대비 약 1,000배, Fast RCNN 대비 약 100 배 빠르며(Redmon *et al.*, 2016; Rajeshwari *et al.*, 2019), 짧은 추론 시간을 요구하는 실시간 객체 탐지에 더 적합하다고 할 수 있다(Han, 2022). 본 연구에서는 실시간 객체 탐지를 위한 추론 속도를 확보하기 위해 YOLOv5 (Jocher, 2020)를 사용하였다. YOLOv5의 구조는 <Figure 3>과 같다.

2.2 멀티 카메라 프로세싱 기반 객체 탐지

객체 탐지 모델의 주요 제약 사항으로는 가려짐, 불충분한 조명, 물체 크기의 다양성, 시점 변화, 배경 잡음 등이 있다.

이러한 요소들은 실제 환경에서 모델이 물체를 정확하게 탐지하는 데 어려움을 줄 수 있다(Sood *et al.*, 2021). 본 연구의 대상인 MCV 조립 공정의 경우, 작업자가 조립을 진행하는 동안 카메라를 가리는 경우가 빈번히 발생한다. 작업자에 의해 가조립 상태 연결부의 조립 장면이 가려지는 경우, 조립 완료 상태를 정확하게 감지할 수 없다. 이를 해결하기 위해 장애물에 의한 영상 간섭을 상호 보완할 수 있는 여러 대의 카메라로 촬영한 영상을 통합하여 객체를 탐지하는 기술이 필요하다. 이를 ‘멀티 카메라 프로세싱(multi-camera processing) 기반 객체 탐지’라 정의한다. 관련 선행연구로 He *et al.*(2011)은 여러 카메라로부터 촬영된 영상에 Conditional Random Field(CRF)를 적용하여 다중 객체를 탐지하는 방법론을 제안하였다. 각 카메라뷰에서 얻은 정보를 CRF를 통해 통합하여 최적의 객체 레이블링을 수행하며, CRF 모델 내에서 명시적으로 가림 현상을 처리한다. 이는 객체 간의 관계와 가림 현상을 복잡한 그래프 구조로 모델링하여 정확한 탐지를 가능하게 한다. CRF 방법론의 경우에는 높은 계산 복잡성으로 인하여, 실시간 처리가 필요한 실제 공정 현장에 도입하기에 어려움이 있다. 본 연구에서는 멀티 카메라 설정을 적용하되 CRF 등 별도의 다중 객체 탐지 모델을 적용하는 대신, 파이썬 기반 멀티 프로세싱을 통해 각 카메라의 객체 탐지 결과를 병렬처리하고, 작업자가 렌치로 가조립 연결부를 조이는 행동을 감지하여 조립 완료 상태를 탐지할 수 있도록 하였다.

공정의 시작과 동시에 카메라 수와 같은 수의 서브 프로세스를 생성하고, 각 프로세스에서 병렬적으로 객체 탐지를 수행한다. 공정이 종료됨과 동시에 모든 서브 프로세스들은 종료되며, 각 서브 프로세스에서 도출된 결과들은 메인 프로세스에서 취합한다. 메인 프로세스에서는 각 객체 탐지 결과를 종합하여 최종적으로 공정 상태를 추론한다.

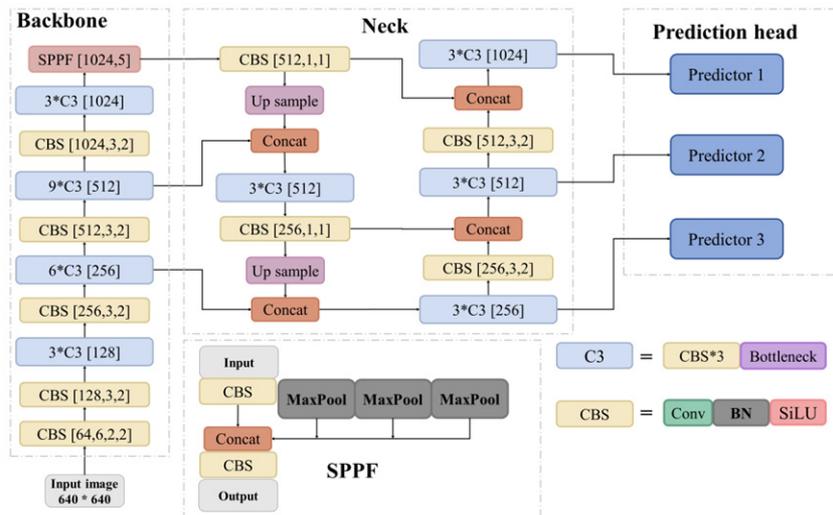


Figure 3. Structure of YOLOv5

3. 제안방법론

본 연구는 딥러닝 기반 객체 탐지를 활용한 엔진 유압 제어 장치 호스 조립 공정 검사 자동화 시스템을 제안한다. 제안 방법의 전체 구조는 <Figure 4>와 같다. 모델의 학습을 위해 호스 조립 공정 현장에 서로 다른 각도의 카메라 2대를 설치하고 각 카메라에서 동시 촬영된 영상을 수집한다. 수집된 영상의 프레임별 이미지에 연결부 좌표와 상태를 어노테이션하여 학습 데이터셋을 확보한다. 수집된 학습 데이터로 각 카메라별 YOLOv5 기반 객체 탐지 모델을 학습한다. 추론 단계에서는 학습에 사용한 2대의 카메라로 작업 영상을 촬영한다. 학습된 YOLOv5 모델을 통해 각 카메라에 대한 실시간 객체 탐지가 병렬적으로 수행된다. 이 과정에서 ID 할당 알고리즘(ID allocation algorithm)을 적용하여 각 객체를 추적하며, 런치 작업에 대한 규칙 기반 조립상태 전환 알고리즘(assembly status transition algorithm)을 통해 연결부의 ‘가조립 상태’와 ‘조립 완료 상태’를 분류한다. 작업이 종료되면 멀티 카메라 프로세싱을 통해 2대의 카메라에서 추론된 결과를 통합하여 최종 결과를 생성한다.

3.1 객체 탐지 및 추적

객체 탐지의 목적은 호스 체결 작업의 진행에 따른 호스 연결부 상태 변화를 실시간으로 탐지하는 것이다. <Figure 1>에서 확인할 수 있듯이 각 연결부는 MCV 전체 면적에 비해 매우 작은 면적을 차지하는 소형 객체로 정확한 딥러닝 기반 객체 탐지가 어렵다.

본 연구에서는 소형 객체 탐지 성능 개선을 위해 NGWD 기반의 객체 탐지 방법론을 제안한다. 객체 탐지 모델 학습의 목적 함수는 각 객체에 대한 실제(ground truth) 바운딩 박스(B_t)와 모델에 의해 예측된 바운딩 박스(B_p)의 차이로 정의된다. 일반적으로 B_t 와 B_p 가 겹치는 영역의 비율을 나타내는 Intersection over Union (IoU)이 널리 사용되며, 식 (1)과 같이 정의한다.

$$IoU = \frac{area(B_t \cap B_p)}{area(B_t \cup B_p)} \quad (1)$$

하지만 소형 객체의 경우 IoU의 경우 학습 초기 단계에서 바운딩 박스의 교집합 영역이 생성되지 않아 IoU 값이 0이 되어 기울기 소실 문제를 유발할 수 있다(Rezatofghi *et al.*, 2019). 또한, IoU는 소형 객체의 위치 변화에 대한 민감도가 높다는 한계점을 가지고 있다(Wang *et al.*, 2021). 본 연구에서 다루는 MCV 호스 조립 공정의 경우, 학습에 사용한 이미지의 원본 크기는 1920×1080인 반면 호스 연결부의 크기는 20×20 이하로 작아 소형 객체에 해당한다. 이러한 소형 객체의 정확한 탐지를 위해 Wang *et al.*(2021)이 제안한 NGWD 기반의 손실함수를 적용하였다. 객체 간 거리를 계산하기 위해 바운딩 박스의 위치 및 크기 정보를 직접 이용하는 IoU와 달리, NGWD는 각 바운딩 박스를 정규분포로 변환한 후 두 정규분포의 거리로 객체 탐지 정확도를 계산한다. 중심 좌표가 (c_x, c_y) 이고, 너비와 높이가 각각 w, h 인 바운딩 박스 $B = (c_x, c_y, w, h)$ 에 내접하는 타원을 정규분포 $N(\mu, \Sigma)$ 로 가정하며 μ 와 Σ 는 식 (2)와

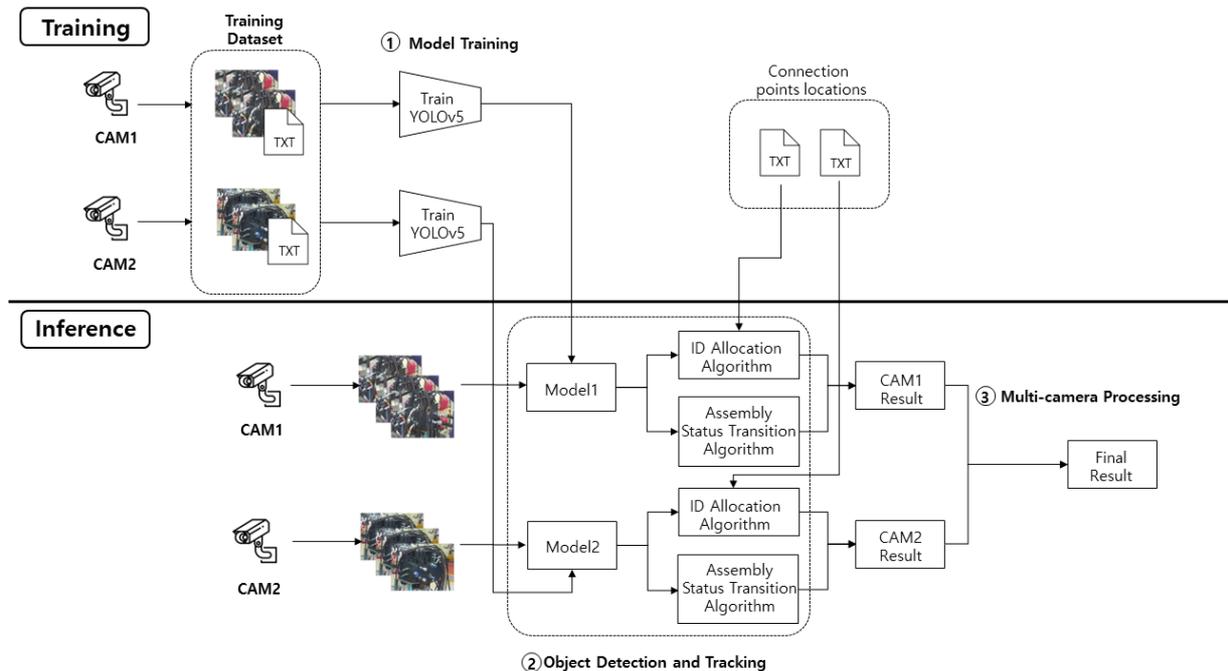


Figure 4. Overview of the Proposed Hose Assembly Inspection System

같이 정의한다.

$$\mu = \begin{bmatrix} c_x \\ c_y \end{bmatrix}, \Sigma = \begin{bmatrix} \frac{w^2}{4} & 0 \\ 0 & \frac{h^2}{4} \end{bmatrix} \quad (2)$$

두 개의 바운딩 박스 B_1, B_2 에 대응하는 두 정규분포 N_1 와 N_2 사이의 거리는 식 (3)과 같이 Wasserstein Distance, $W_2^2(N_1, N_2)$ 로 정의한다.

$$W_2^2(N_1, N_2) = \left\| \left[\left[c_{x,1}, c_{y,1}, \frac{w_1}{2}, \frac{h_1}{2} \right]^T, \left[c_{x,2}, c_{y,2}, \frac{w_2}{2}, \frac{h_2}{2} \right]^T \right] \right\|_2^2 \quad (3)$$

식 (3)의 $W_2^2(N_1, N_2)$ 을 정규화한 NGWD는 식 (4)와 같이 정의한다. 이때 $C = w \times h$ 는 바운딩 박스의 면적이다.

$$NGWD(N_1, N_2) = \exp\left(-\frac{\sqrt{W_2^2(N_1, N_2)}}{C}\right) \quad (4)$$

NGWD를 활용한 YOLOv5 기반 객체 탐지 모델에 의해 탐지된 각 연결부에 식별자를 할당하고 작업 진행에 따른 상태 변화를 추적하기 위한 객체 추적 알고리즘을 제안한다. 첫 번째로, 사용자가 사전에 정의된 각 호스 연결부의 중심 좌표를 직접 입력한다. 다음으로 객체 탐지 단계에서 탐지된 연결부들의 중심 좌표를 계산하고, 사전에 입력한 중심 좌표와의 거리를 계산하여 가장 가까운 연결부에 대한 식별자를 객체에 할당한다. 조립 공정의 시작 시점부터 종료 시점까지 촬영되는 프레임마다 위 순서를 반복하여 연결부의 식별자를 유지하며 상태 변화를 추적한다.

3.2 조립 완료 탐지

MCV 각 연결부의 ‘조립 완료 상태’ 전환 여부를 자동으로 탐지하기 위한 조립 상태 전환 알고리즘을 제안한다. 조립 완료 탐지의 목적은 3.1의 객체 탐지 및 추적 알고리즘에 의해 식별자가 할당된 각 연결부에 대하여 초기 상태의 연결부가 수작업에 의한 호스 체결을 거쳐 ‘가조립 상태’로 전환된 후, 렌치에 의한 조립 작업을 통해 ‘조립 완료 상태’로 전환되는 과정을 자동으로 탐지하고 조립 완료가 누락된 연결부 유무를 판정하는 것이다. 따라서, 이미지 기반 객체 탐지 모델은 촬영된 실시간 작업 영상으로부터 ‘조립 완료 상태’를 자동으로 탐지해야 한다. 조립 공정은 렌치로 연부를 조이는 작업에 의해 완료된다는 점에 착안하여 이미지 내의 렌치 작업을 거리, 면적, 시간 기준으로 탐지하는 알고리즘을 제안한다.

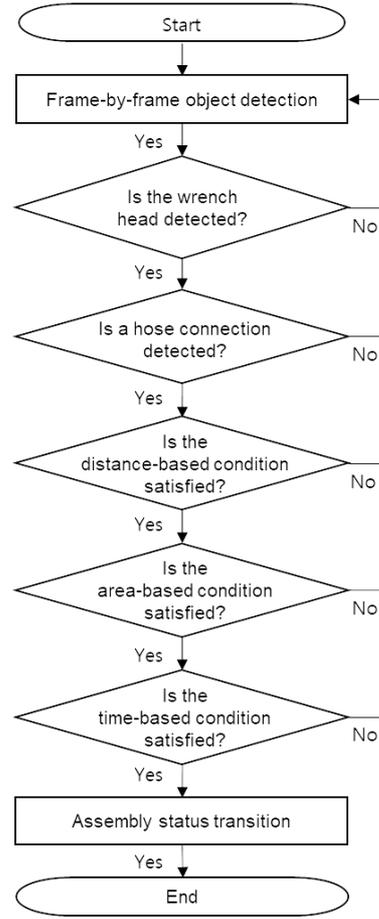


Figure 5. Assembly Status Transition Algorithm

먼저, 렌치 헤드와 연결부의 접촉 여부를 판단하기 위해 두 객체의 바운딩 박스 간 거리와 중첩 면적을 계산한다. 작업자가 연결부에 렌치 체결 작업을 진행하지 않았음에도 렌치를 들고 움직이는 과정에서 렌치 헤드와 연결부가 겹치는 경우가 발생할 수 있다. 이러한 경우에 대한 거리 및 면적 기반 조건의 한계를 보완하기 위해 렌치 헤드가 연결부에 체류한 시간에 대한 조건을 함께 고려한다. 거리, 면적, 시간에 의한 렌치 체결 작업 탐지를 통한 조립 완료 탐지 알고리즘은 <Figure 5>와 같다.

(1) 거리 기반 조건

렌치 헤드와 연결부 사이의 거리를 기반으로 연결부 렌치 체결 작업 진행 여부를 판단한다. 객체 탐지 모델에 의해 탐지된 연결부의 바운딩 박스 중심 좌표 $\mathbf{c} = (x_c, y_c)$ 와 렌치 헤드의 바운딩 박스 중심 좌표 $\mathbf{w} = (x_w, y_w)$ 사이의 유클리디안 거리 (Euclidean distance)가 임계값 ϵ 보다 작으면 렌치와 연결부 접촉의 필요조건을 만족한다고 가정한다.

$$d(\mathbf{c}, \mathbf{w}) = \sqrt{(x_c - x_w)^2 + (y_c - y_w)^2} < \epsilon \quad (5)$$

(2) 면적 기반 조건

작업자가 연결부에 렌치 체결 작업을 진행하지 않았음에도 렌치를 들고 움직이는 과정에서 렌치 헤드와 연결부의 거리가 가까워지는 경우가 발생할 수 있다. 이러한 경우에 대한 거리 기반 조건의 한계를 보완하기 위해 렌치 헤드와 연결부의 바운딩 박스가 겹치는 부분의 면적을 함께 고려한다. 연결부 바운딩 박스의 면적 대비 두 바운딩 박스의 교집합 면적의 비율 Intersection over Connection point (IoC)가 임계값 θ 보다 큰 경우 렌치와 연결부 결합의 필요조건을 만족한다고 가정한다.

(3) 렌치 체류 시간 조건

거리 및 면적 기반 조건을 통해 렌치와 연결부가 접촉되었는지 여부를 판단한 후 체류 시간(wrench stay time, wst)을 측정하여 체결 작업이 진행되었는지 판단한다. 위의 거리 및 면적 기반 조건을 만족하는 연속적인 이미지 프레임의 수가 임계값 τ 를 초과하는 경우 해당 연결부에 대한 렌치 체결 작업이 실제로 조립이 최종적으로 완료되었다고 판단한다.

3.3 멀티 카메라 프로세싱

작업자의 동선에 의한 영상 촬영 호스 연결부의 조립 작업 완료 여부를 실시간으로 판단하기 위해서 카메라를 통해 이미지가 취득되는 시점마다 연결부의 상태를 탐지하고 최신화해야 한다. 파이썬에서 제공하는 멀티 프로세싱 라이브러리 ‘multi processing’을 활용하여 각 영상에 대한 병렬적인 객체 탐지를 진행한다. 프로세스 스폰닝(spawning) 기반 병렬처리를 통해 메인 프로세스 아래에 카메라와 같은 수의 서브 프로세스를 생성하고, 각각 독립적으로 수집된 영상에 대한 ‘객체 탐지 및 추적’, ‘조립 완료 탐지’ 과정을 수행한다. 추론된 연결부의 상태는 실시간으로 기록되며, 공정이 종료된 후 메인 프로세서는 각 서브 프로세스 수행 결과를 취합한다. 1개 이상의 서브 프로세스에 의해 조립 완료로 추론된 연결부는 조립 완료로, 모든 서브 프로세스가 조립 미완료로 추론한 연결부는 최종 조립 미완료로 판정한다.

4. 실험

4.1 데이터

본 연구의 실험을 위해 조선 및 건설장비 기계부품 제조기업 D사와 협업하여 MCV 호스 조립 공정의 영상 데이터를 제공받았다. 총 2개의 카메라(CAM1, CAM2)를 사용하여 실제 작업 영상을 촬영하였다. CAM1은 MCV의 우측면을, CAM2는 좌측면을 촬영한 영상이다. 동일 공정에 대하여 학습에 사용할 영상 데이터와 검증에 사용할 영상 데이터를 구분하여 수집하였다. 학습용 영상은 객체 탐지 모델 학습에 활용되며, 검증용 영상은 학습된 모델을 기반으로 추론하여 성능을 검증한다.

Table 1. Description of Classes

Class	Description
0	Open
1	Closed
2	Wrench Head
3	Main Control Valves
4	Completed

CAM1과 CAM2에서 수집한 학습용 영상의 이미지 수는 각각 9,394장, 9,279장이다. 검증용 영상의 카메라별 이미지 수는 CAM1과 CAM2 각각 9,264장, 9,145장이다. 객체 탐지 모델 학습을 위해 각 연결부에 대한 레이블링을 수행하였으며, 각 클래스에 대한 설명은 <Table 1>과 같다. 두 카메라는 동일 시점에 촬영되었으나 종료 시점이 달라 수집된 이미지 수에 차이가 발생하였다.

Table 2. Class Distribution of the Training Data

CAM	Class	Train	Test
1	0	10,062	8,715
	1	19,159	12,597
	2	219	189
	3	7,140	6,185
	4	0	5,563
2	0	10,062	10,961
	1	5,694	6,205
	2	225	245
	3	6,331	6,899
	4	0	2,879

조립 작업의 진행에 따라 호스 연결부의 클래스는 0(open), 1(closed), 4(completed)의 순서로 변경된다. ‘클래스 4(completed)’는 학습 이미지 데이터 어노테이션에는 포함되지 않는 클래스이며, 조립상태 전환 알고리즘을 통해 가조립 상태에서 조립 완료 상태로 전환 될 때 ‘클래스 1(closed)’로부터 전환되어 생성되는 가상의 클래스이다. D사에서는 MCV 외에도 고유한 조립 단계를 가지는 실린더, 펌프, 유압 조작 밸브(RCV) 등의 공정이 진행된다. ‘클래스 3(MCV)’는 제안된 방법론의 대상인 MCV를 식별하는 역할을 한다.

<Table 2>는 학습 데이터와 검증 데이터의 클래스 분포를 나타내었다. 카메라의 시점에 따라 각 연결부가 작업자에 의해 가려지는 시간이 다르기 때문에 CAM1과 CAM2의 클래스별 이미지 수는 다르게 수집되었다.

Table 3. Number of Connection Points Captured by CAM1 and CAM2

CAM	Connection Point	Number
1	H1	2,392
	H2	8,504
	H3	1,005
	H4	6,612
	H5	4,699
	H6	6,243
	H7	4,684
2	H1	265
	H2	5,570
	H3	481
	H4	790
	H5	4,104

<Figure 1>에 나타난 바와 같이 하나의 MCV에는 총 7개의 호스 연결부가 존재하며, 촬영 각도 및 작업자 동선에 의한 간

섭을 고려하여 모든 연결부가 실시간으로 모니터링 될 수 있도록 두 대의 카메라 CAM1, CAM2를 설치하였다. 각 연결부가 두 카메라에 의해 탐지된 횟수는 <Table 3>과 같다. MCV 측면에 위치한 2개의 연결부 H6과 H7은 CAM1에 의해서만 모니터링된다.

4.2 모델 및 평가지표

본 연구에서는 실시간 객체 탐지를 위한 딥러닝 모델로 YOLOv5를 사용하였다. 모델 사이즈에 따른 다양한 모델 중 YOLOv5l를 사용하였다. Microsoft COCO 데이터셋(Lin *et al.*, 2014)에 사전 학습된 모델에 수집된 데이터를 학습하여 전이 학습을 수행하였다. 모델 학습을 위한 하이퍼파라미터는 학습 에포크 수 100, 배치 사이즈 16, 이미지 사이즈 640으로 설정하였다.

조립 완료 탐지를 위한 거리, 면적, 렌치 채류 시간 조건에 대한 하이퍼파라미터는 실험적으로 설정하였으며 최종 모델에는 $\epsilon = 100$, $\theta = 0.5$, $\tau = 3$ 으로 설정하였다. 객체 탐지 성능 평가를 위해 혼동행렬(confusion matrix)을 사용하였다. 혼동행렬은 분류 분석 결과를 True Positive(TP), False Positive(FP), True Negative(TN), False Negative(FN)로 나누어 집계한다.

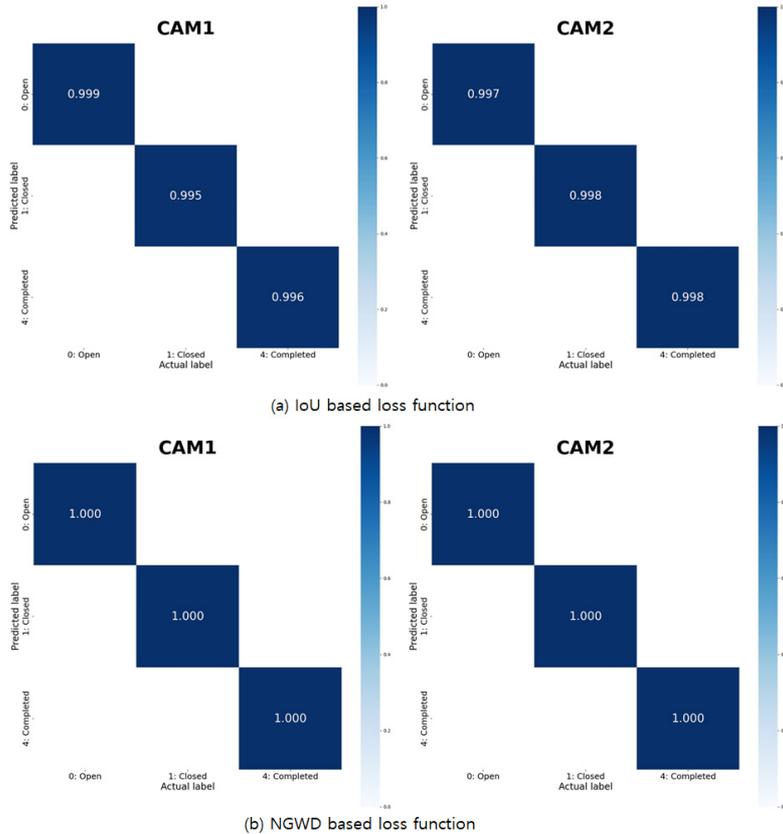


Figure 6. Confusion Matrices Comparing the Performance of (a) IoU-based and (b) NGWD-based loss function in hose connection point detection

4.3 실험 결과

<Figure 6>은 CAM1과 CAM2에 의해 탐지된 호스 연결부 상태의 실시간 추론 결과로 공정 진행에 따른 호스 연결부에 대한 클래스 0, 1, 4의 추론 정확도를 혼동행렬로 나타내었다. 모든 카메라에서 각 연결부에 대한 상태 탐지 정확도는 100%로, 이는 모든 객체가 정확히 탐지되었음을 의미한다. 소형 객체 탐지에 대한 NGWD (Normalized Gaussian Wasserstein Distance) 기반 손실함수의 효과를 검증하기 위해 IoU (Intersection over Union) 기반 손실함수와의 비교 실험을 수행하였다. 손실함수를 제외한 나머지 조건은 동일하게 유지한 상태에서 실험하여 두 결과를 도출하였고, <Figure 6>을 통해 본 연구에서 사용한

NGWD 기반 손실함수를 사용할 때 IoU 기반 손실함수에 비해 호스 연결부에 대한 객체 탐지 성능이 우수함을 확인하였다.

<Figure 6> 혼동행렬에 따르면, NGWD 기반 손실함수를 사용하는 경우 클래스 0, 1, 4를 100% 정확도로 분류한다. 본 연구에서 제안하는 조립 상태 전환 알고리즘은 가조립 상태(클래스 1)가 탐지된 이후부터 일련의 과정을 탐지하여 조건이 충족되면 '조립 완료 상태'로 전환하므로, 최종 조립 완료 여부를 감지하기 위해서는 클래스 1에 대한 정확도가 1이 되어야 한다.

<Figure 7>은 각 호스 연결부의 체결 상태를 실시간으로 판정하기 위한 객체 추적 정확도를 평가하였다. H1~H7을 촬영하는 CAM1과 H1~H5를 촬영하는 CAM2 모두 촬영된 영상 내의 모든 호스 연결부를 정확히 추적 및 탐지함을 확인할 수 있었다.

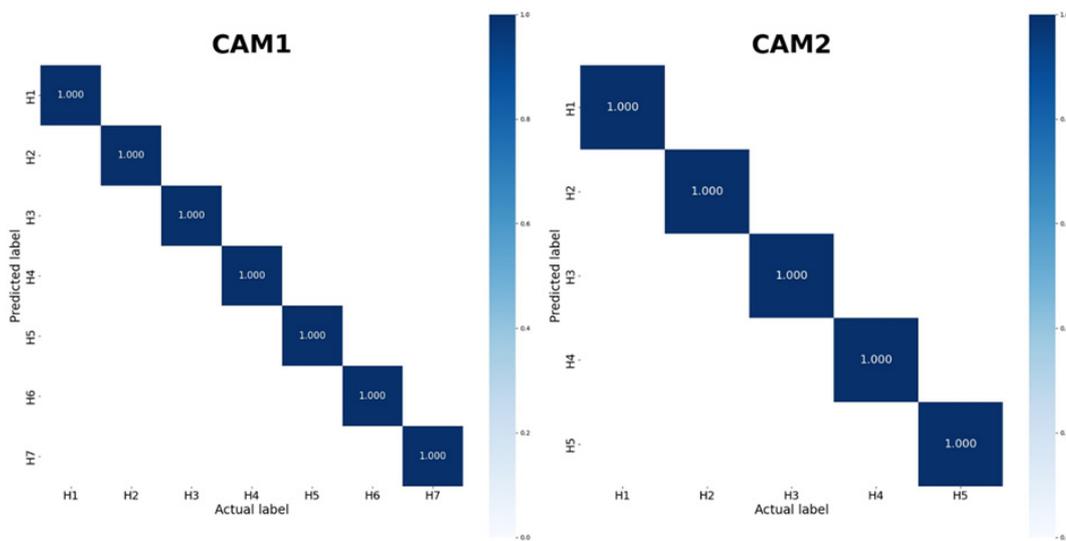


Figure 7. Confusion Matrices for Detection of the Hose Connection Points

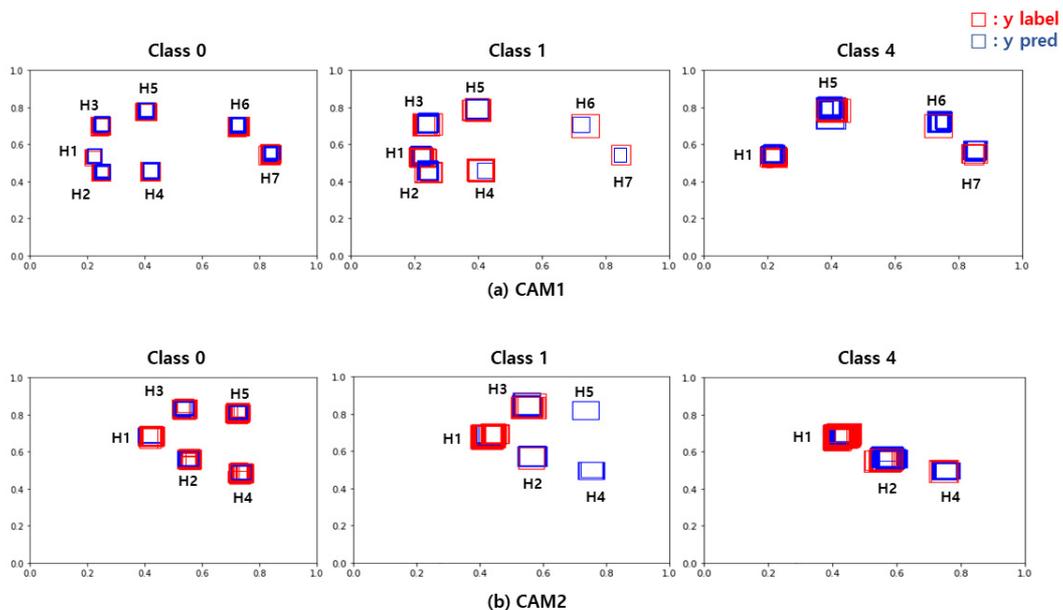


Figure 8. Visualization of Bounding Boxes for Hose Connection Points and Classes of (a) CAM1 and (b) CAM2



Figure 9. Results of the Automated Inspection by CAM1 and CAM2

<Figure 8>은 CAM1과 CAM2에 의해 탐지된 호스 연결부와 클래스의 프레임별 실시간 추론 결과를 바운딩 박스로 시각화한 것이다. <Figure 8(a)>과 같이 클래스 0, 1에서는 H1~H7, 클래스 4에서는 H1, H5, H6, H7이 탐지되었음을 확인했다. <Figure 8(b)>과 같이 클래스 0, 1에서는 H1~H5, 클래스 4에서는 H1, H2, H4가 탐지되었음을 확인했다. 검증용 영상 내의 모든 연결부는 체결 작업이 완료되었기 때문에 최종적으로는 조립 완료 상태인 클래스 4가 되어야 한다. 그러나 <Figure 8>을 통해, 작업자와 장애물의 한 가려짐 현상으로 나머지 호스 연결부가 조립 상태 전환 알고리즘을 통한 전환이 되지 못하였음을 확인했다.

<Figure 9>는 본 연구에서 제안하는 검사 자동화 시스템이 CAM1과 CAM2에 적용된 결과이다. <Figure 9(b)>의 H2를 보면 조립 상태 전환 알고리즘이 적용되어 '조립 완료 상태'로 전환된 것을 확인할 수 있지만, <Figure 9(a)>의 H2에서는 작업자 혹은 장애물로 인한 가림현상에 의해 조립 상태 전환 알고리즘이 적용되지 못한 것을 확인할 수 있다.

<Table 4>는 CAM1, CAM2의 객체 탐지 및 추적과 조립 상태 전환 알고리즘 결과를 멀티 카메라 프로세싱을 통해 병합하여 각 연결부의 조립 공정 완료 여부를 판정한 결과이다. 총 7개의 호스 연결부에 대하여 CAM1은 3개 연결부(H2, H3, H4), CAM2는 4개 연결부(H3, H5, H6, H7)의 조립 완료를 탐지하지 못하였다. H6와 H7은 사각지대로 인해 CAM2 영상에서는 한 번도 탐지되지 않았다. 멀티 카메라 프로세싱을 통해 두 카메라의 영상을 병합한 최종 결과에서는 H3의 조립 완료를 탐지하지 못하였다. 사각지대를 해소하기 위한 카메라의 추가 설치가 필요함을 확인할 수 있었다.

Table 4. The Inference Results for the Final Class of Each Connection Point by Multi-camera Processing

Result	H1	H2	H3	H4	H5	H6	H7
CAM1 result	4	1	1	1	4	4	4
CAM2 result	4	4	1	4	1		
Final result	4	4	1	4	4	4	4

5. 결론

본 연구에서는 건설 중장비용 엔진의 MCV 호스 조립 공정 검사에 딥러닝 기반 객체 탐지 기술을 적용하여, 인적오류로 인한 조립 누락을 자동으로 탐지하는 시스템을 개발하였다. 육안으로 구분이 어려운 조립 완료 상태를 탐지하기 위한 작업자 행동 인식 기반의 딥러닝 객체 탐지 알고리즘을 제안하였다. 소형 객체인 호스 연결부를 탐지하고 추적하기 위한 NGWD 기반 소형 객체 탐지 방법론과 촬영 각도가 다른 여러 영상의 결과를 병합하기 위한 멀티 카메라 프로세싱 기술을 활용하여 높은 추론 정확도를 달성하였다. 본 논문에서 제안하는 방법론은 작업자의 수작업에 대하여 육안 검사를 수행하는 다양한 제품의 제조 현장에 적용할 수 있는 일반화 가능한 기술로 그 파급효과가 크다.

추후 연구에서는 다음과 같은 한계점을 개선하고자 한다. 첫째, 더 많은 수의 카메라를 활용하여 사각지대를 해소할 수 있다. 본 연구에서는 작업 현장의 여건으로 인해 2대의 카메라만 사용하였지만, 다양한 각도의 카메라를 추가할 수 있다면 추론 정확도를 개선할 수 있다. 둘째, MCV 상의 호스 연결부를 도메인 전문가가 사전에 수기로 입력해야 한다. 개발된 기술을 다른 제품군 및 공정으로 확장 적용하기 위해서는 이 부분을 자동화하기 위한 연구가 필요하다.

참고문헌

- Girshick, R., Donahue, J., Darrell, T., and Malik, J. (2014), Rich feature hierarchies for accurate object detection and semantic segmentation, In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 580-587.
- Girshick, R. (2015), Fast r-cnn. In *Proceedings of the IEEE International Conference On Computer Vision*, 1440-1448.
- Han, J.-G. (2022), Webcam object recognition using YOLOv4 through object recognition algorithm analysis, *2022 Fall Conference of the Korean Institute of Electrical Engineers, Industrial Application Society*, 17-18.

- He, L., Boix, X., Ben Shitrit, H., and Fua, P. (2011), Conditional Random Fields for Multi-Camera Object Detection, In *Proceedings of the 2011 IEEE International Conference on Computer Vision*, 563-570. IEEE. doi:10.1109/ICCV.2011.6126319.
- Jocher, G. (2020), YOLOv5 by Ultralytics, v7.0, AGPL-3.0, <https://github.com/ultralytics/yolov5>, doi:10.5281/zenodo.3908559.
- Lee, J-H., Lee, S-H., Oh, H-O., Lee, W., and Park, S-B. (2006), Development of a Process Monitoring System for Real-time Process Management, *Proceedings of the Korean Institute of Industrial Management Systems Conference*, **2006**(3), 203-206.
- Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., ... and Zitnick, C. L. (2014), Microsoft coco: Common objects in context. In *Computer Vision—ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13*, Springer International Publishing, 740-755.
- Malburg, L., Rieder, M.-P., Seiger, R., Klein, P., and Bergmann, R. (2021), Object Detection for Smart Factory Processes by Machine Learning, *Procedia Computer Science*, **184**, 581-588.
- Rajeshwari, P., Abhishek, P., Srikanth, P., and Vinod, T. (2019), Object Detection: An Overview, *International Journal of Trend in Scientific Research and Development (ijtsrd)*, **3**(3), 1663-1665.
- Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016), You only look once: Unified, real-time object detection, In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 779-788.
- Rezatofighi, H., Tsoi, N., Gwak, J., Sadeghian, A., Reid, I., and Savarese, S. (2019), Generalized intersection over union: A metric and a loss for bounding box regression, In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 658-666.
- Sood, S., Singh, H., Malarvel, M., and Ahuja, R. (2021), Significance and Limitations of Deep Neural Networks for Image Classification and Object Detection, *2021 2nd International Conference on Smart Electronics and Communication (ICOSEC)*, IEEE, 7-9
- Tang, S., Zhang, S., and Fang, Y. (2023), HIC-YOLOv5: Improved YOLOv5 For Small Object Detection, arXiv preprint arXiv:2309.16393v2.
- Wang, J., Xu, C., Yang, W., and Yu, L. (2021), A Normalized Gaussian Wasserstein Distance for Tiny Object Detection, arXiv preprint arXiv:2110.13389v2.
- Wang, W., Zhang, D., and Cao, Q. (1997), Reliability Analysis of Human Error Identification in Man-Machine Systems, *System Engineering and Electronics*, **1997**(3). Southwest Jiaotong University, Chengdu 610031.
- Yang, L., Su, Q., and Shen, L. (2012), A novel method of analyzing quality defects due to human errors in engine assembly line, *2012 International Conference on Information Management, Innovation Management and Industrial Engineering*, IEEE, 20-21.
- Yoo, C., Choi, S. W., and Lee, I. B. (2008), Recent Research Trends of Process Monitoring Technology: State-of-the-Art, *Korean Chemical Engineering Research*, **46**(2), 233-247.

저자소개

김창영 : 울산대학교 산업경영공학부에서 2024년 학사학위를 취득하였다. 연구분야는 산업인공지능이다.

조현건 : 울산대학교 산업경영공학부를 재학 중이다. 연구분야는 산업인공지능이다.

최준혁 : 울산대학교 조선해양공학부에서 2024년 학사학위를 취득하고 경희대학교 산업경영공학과 석사과정에 재학 중이다. 연구분야는 산업인공지능이다.

이수동 : 한양대학교 산업공학과에서 2012년 학사, 포항공과대학교에서 2018년 박사학위를 취득하였다. SK하이닉스 책임, 한국원자력연구원 선임연구원을 역임하고 2020년부터 울산대학교 산업경영공학부 교수로 재직하고 있다. 연구분야는 산업인공지능이다.