Journal of Korea Robotics Society (2025) 20(3):396-407
https://doi.org/10.7746/jkros.2025.20.3.396 ISSN: 1975-6291 / eISSN: 2287-3961

TCAT: 제조 환경 내 인간-로봇 협업을 위한 시간 일관성 기반 적응형 객체 추적 기법

TCAT: A Temporal Consistency-based Adaptive Object Tracking Method for Human-Robot Collaboration in Manufacturing Environments

황 도 경[1] · 김 용 국[1] · 이 예 준[1] · 김 민 규[1] · 정 의 정[†]
Do-Kyung Hwang[1], Yong-Kuk Kim[1], Ye-Jun Lee[1], Min-Gyu Kim[1], Eui-Jung Jung[†]

Abstract: This paper proposes an enhanced tracking method for stable object recognition in Human-Robot Collaboration (HRC) manufacturing environments. Existing methods often struggle with occlusions by workers' hands, robotic interference, and dynamic pose variations. To address these challenges, we introduce Temporal Consistency-based Adaptive Tracking (TCAT), which leverages temporal consistency analysis and adaptive weighting strategies. TCAT quantifies temporal consistency by analyzing detection confidence scores across three consecutive frames and employs sigmoid-based adaptive weighting to balance spatial IoU and appearance feature similarities effectively. We validate TCAT on a custom dataset comprising 100,000 images of five vacuum cleaner components, integrating it with state-of-the-art object detectors such as YOLOv10-12 and RT-DETR. Experimental results demonstrate that TCAT significantly outperforms baseline methods, achieving a mAP@0.5 of 85.7% with YOLOv12—a 3.4 percentage-point improvement over ByteTrack—while maintaining a real-time processing speed of 15 FPS on an NVIDIA Jetson Orin Nano. The proposed method shows robust tracking performance and computational efficiency, which is suitable for practical HRC applications.

Keywords: Temporal Consistency, Adaptive Tracking, Human-Robot Collaboration, Manufacturing Environment, Object Tracking, YOLO

1. 서 론

제조업 환경에서의 인간-로봇 협업(Human-Robot Collaboration, HRC)은 디지털 전환, 스마트 팩토리 고도화, 생산 유연화 등의 흐름에 따라 핵심적인 기술 주제로 자리 잡고 있다[1-4]. 특히 소량 다품종 생산과 같은 불규칙한 제조 환경에서는 단순 반복 작업이 아닌, 작업자의 동적 판단을 지원하거나 보완할 수 있는 로봇의 역할이 요구된다[5,6]. 이러한 환경에서 로봇이 주변 물체를 인식하고, 그 위치와 형태를 정확하게 파악하는 것은 협업 효율성과 안전성을 동시에 만족시키기 위한 전제 조건이다[7,8].

최근 객체 인식 연구에서는 YOLOv10, 11, 12, RT-DETR 등 다양한 심층 학습 기반 모델들이 제안되어 높은 정확도를 보이고 있다[9-12]. 특히 Transformer 구조의 도입과 다중 스케일 특징 추출 등의 발전을 통해 일반적인 환경에서의 객체 인식 성능은 크게 향상되었다[13,14]. 그러나 이러한 발전에도 불구하고, 제조 현장의 특수한 환경에서는 여전히 안정적인 인식 성능을 보장하기 어려운 상황이다. 실제 제조 환경에서는 작업자의 손이나 로봇 암에 의한 일시적 가려짐, 부품의 급격한 자세 변화, 조명 조건의 변화 등 다양한 외란 요소들이 복합적으로 발생하며, 이는 객체 인식의 연속성과 안정성을 크게 저하시키는 원인이 된다[15].

Received : May. 28. 2025; Revised : Jul. 16. 2025; Accepted : Jul. 17. 2025
※ This work was supported by the Technology Innovation Program (RS-2024-00417663) funded By the Ministry of Trade Industry & Energy (MOTIE, Korea).
1. Researcher, Human-Robot Interaction research center, KIRO, Pohang, Korea (dokyung, ykkim0625, banily07, mingyukim@kiro.re.kr)
† Principal Researcher, Corresponding author: HRI Center, KIRO, Pohang, Korea (ejjung@kiro.re.kr)

이러한 문제를 해결하기 위해 다양한 객체 추적 알고리즘들이 적용되어 왔다. ByteTrack, BoTSORT 등의 최신 추적 방법들은 객체의 외관 유사도와 공간적 연속성을 활용하여 일시적 가려짐이나 인식 실패 상황을 보완하고자 하였다[16,17]. 특히 IOU (Intersection over Union) 기반의 매칭과 칼만 필터를 활용한 움직임 예측 등을 통해 연속적인 객체 추적을 시도하였다[18]. 그러나 이러한 접근법들은 주로 공간적 특징과 외관 정보에 의존하고 있어, 제조 환경의 복잡한 외란 상황에서는 여전히 한계를 보인다. 특히 급격한 자세 변화나 장시간의 가려짐이 발생하는 경우, 객체의 외관이나 위치 정보만으로는 안정적인 추적이 어려운 문제가 존재한다[19,20]. 아울러, 최근에는 이러한 한계를 극복하기 위해 시간적 정보를 활용하는 연구들이 제안되고 있다. 연속된 프레임에서의 객체 특징 변화를 모델링하거나[21,22], 장단기 시간 의존성을 고려한 추적 방법 등이 연구되었다[23,24]. 그러나 이러한 접근법들은 대부분 복잡한 모델 구조나 높은 계산 비용을 요구하여 실시간 처리가 필요한 제조 현장에 직접 적용하기에는 제약이 있다[25,26].

또한, 기존의 객체 인식 연구들은 주로 일반적인 환경이나 고정된 조건에서의 성능 향상에 초점을 맞추어 왔다[27-29]. 특히 COCO, Pascal VOC와 같은 일반적인 데이터셋을 기반으로 한 모델들은 제조 현장의 특수한 조건들을 충분히 반영하지 못하는 한계를 보인다[30]. 최근 YOLOv10-12나 RT-DETR과 같은 최신 모델들이 높은 정확도와 실시간성을 보여주고 있으나, 제조 환경의 고유한 도전 과제들에 대한 체계적인 대응은 여전히 부족한 실정이다[9-12].

본 연구에서는 이러한 한계들을 극복하기 위해, 실제 제조 환경 중 하나인 청소기 제조 공정을 실험적 모델로 채택하여, 커스텀 모델 학습과 추적 방법 분석을 위한 커스텀 데이터 셋을 구축하고, 시간적 일관성 기반 적응형 추적 방법(Temporal Consistency-based Adaptive Tracking, TCAT)을 제안한다. TCAT은 기존 추적 방법들과 달리 객체 인식 결과의 시간적 패턴을 분석하고, 이를 기반으로 검출과 추적 결과의 가중치를 동적으로 조절한다. 특히 시간적 일관성을 정량화하는 효율적인 지표를 도입하고, 이를 통해 제조 현장의 다양한 외란 상황에서도 안정적인 객체 추적이 가능하도록 한다. 또한 기존 추적 방법들의 장점을 유지하면서도 계산 효율성을 고려한 경량화된 구조를 제안함으로써, 실제 제조 환경에서의 실시간 적용 가능성을 높였으며, 주요 기여는 다음과 같다:

- 실제 청소기 제조 공정의 5종 부품을 대상으로 한 10만장 규모의 대규모 커스텀 데이터셋 구축 및 체계적인 증강 기법 적용 사례 제시
- 객체 인식 결과의 시간적 일관성을 정량화하고 이를 기반으로 한 적응형 가중치 조절 메커니즘 제안
- 기존 객체 인식 모델과 결합한 추적 방법 대비 향상된 객체 추적 성능 검증
- 실제 제조 현장을 모사한 HRC 워크셀 환경에서의 구현 및 실증적 성능 평가

2. 관련 연구

제조업에서의 객체 인식 기술은 그 활용 범위가 지속적으로 확장되고 있다. Muller et al.는 자동차 조립 라인에서의 HRC 시스템 구현을 통해 작업자의 위치와 도구를 실시간으로 추적하였으나, 고정된 조명 조건과 제한된 작업 영역에서만 검증이 이루어졌다[31]. Zhang et al.는 다품종 소량 생산 환경에서 로봇의 유연한 부품 파지를 위한 객체 인식 시스템을 개발하였으나, 단일 종류의 외란(조명 변화)만을 고려하여 실제 제조 현장의 복잡한 상황을 완전히 반영하지 못했다[32]. 이러한 한계를 인식한 Wang et al.은 제조 현장의 다양한 외란 요소들이 객체 인식의 정확도에 미치는 영향을 체계적으로 분석하였으며, 특히 조명 변화와 부분 가려짐 현상이 성능 저하의 주요 원인임을 밝혔다[33].

이러한 문제들을 해결하기 위해 객체 검출 분야에서는 다양한 발전이 이루어져 왔다. 특히 단일 신경망 기반의 YOLO 계열은 실시간 처리가 가능한 구조로 산업 현장에서 널리 활용되어 왔다. YOLOv5를 시작으로 v7, v8 등을 거치며 꾸준한 성능 개선이 이루어졌으며, 최근에는 YOLOv10-12를 통해 Transformer의 self-attention 구조를 부분적으로 도입함으로써 복잡한 장면에서도 높은 정확도를 확보할 수 있게 되었다[9-11]. 특히 YOLOv12에서 제안된 Cross-Scale Fusion과 Dynamic Head 구조는 작은 객체나 중첩된 상황에서도 우수한 성능을 보여준다[11]. 한편, RT-DETR은 YOLO와는 다른 접근 방식으로, Transformer 기반의 인코더-디코더 구조를 통해 객체 간의 관계성을 자연스럽게 모델링할 수 있도록 설계되었다[12]. 그러나 이러한 최신 모델들도 제조 현장의 특수성을 고려한 최적화가 이루어지지 않아, 실제 적용 시에는 기대만큼의 성능을 발휘하지 못하는 경우가 많다[9-12]. 객체 인식 모델의 성능 개선을 위해, 인식 모델과 같이 결합하여 사용할 수 있는 다양한 추적 알고리즘들이 연구되어왔다. ByteTrack은 기존 방식들과 달리 신뢰도가 낮은 검출 결과도 추적에 활용함으로써 일시적 가려짐 상황에서의 성능을 크게 개선하였으며[16], BoTSORT는 여기에 Re-ID 특징과 광학 흐름 정보를 추가로 활용하여 더욱 강건한 추적을 가능하게 하였다[17]. 그러나 이러한 방법들은 여전히 외관이나 움직임 정보의 연속성에 크게 의존하고 있어, 제조 환경에서 발생하는 급격한 변화 상황에서는 추적 실패가 빈번히 발생하는 문제가 있다[16,17]. 최근에는 이러한 한계를 극복하기 위해 시간적 정보를 적극적으로 활용하는 연구들이 제안되고 있다. Sun et al.는

Transformer 구조를 활용하여 시간적 의존성을 명시적으로 모델링하였으며[34], Wang et al.는 시간 윈도우 내의 특징 변화 패턴을 학습하는 방식을 제안하였다[35]. 그러나 이러한 접근법들은 대부분 복잡한 모델 구조를 필요로 하여 실시간 처리가 요구되는 제조 환경에 직접 적용하기에는 제약이 있다[36].

본 연구에서는 제조 환경의 복잡한 외란 상황에서도 안정적인 객체 추적이 가능하도록, 기존 ByteTrack의 낮은 신뢰도 검출 결과 활용 전략과 BoTSORT의 외관 특징 매칭 방식을 기반으로 하되, 시간적 패턴 분석을 통한 적응형 가중치 조절 메커니즘을 새롭게 제안한다. 이는 기존 추적 방법들의 시공간적 특징 활용은 유지하면서도, 실시간 처리가 가능한 효율적인 구조로 시간적 일관성을 통합한다. 제안된 방법의 검증을 위해 실제 청소기 제조 공정의 5종 부품을 대상으로 한 10만장 규모의 데이터셋을 구축하였으며, HRC 워크셀 환경에서의 실험을 통해 그 효과를 정량적으로 분석하였다.

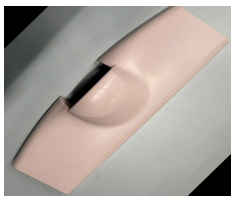
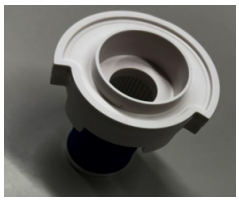
3. 실험 환경 구축

본 연구에서는 제조업 환경을 실험적으로 재현하기 위해 청소기 제조 공정에서 실제 활용되는 5종의 부품(Cover dust, Body brush, Connector brush Cover brush, Filter body)을 [Table 1]과 같이 인식 대상으로 선정하고, 해당 부품들로 HRC 작업을 하는 실제 환경과 유사한 워크셀 환경을 구축하였다.

3.1 협업 실험 워크셀 구성

협업 실험 워크셀의 객체 인식 대상 부품들이 배치되는 작업 테이블 위에는 Intel RealSense D435i RGB-D 카메라가 고정 설

[Table 1] Five target objects for recognition

Cover dust	Body brush	Connector brush
		
Cover brush	**Filter body**	
		

[Fig. 1] Experimental workcell configuration for object recognition: Intel RealSense D435i camera installation with workspace coverage optimization

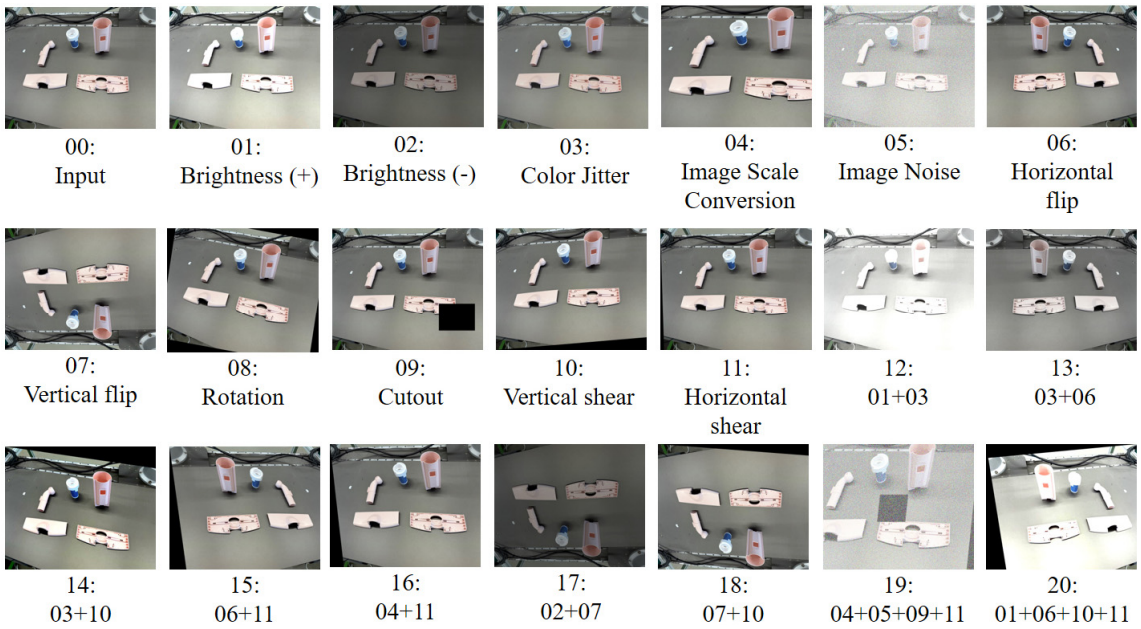
치되었으며, 카메라의 FOV (field of view)는 70~80cm 거리에서 테이블 전면을 충분히 커버할 수 있도록 조정되었다. 카메라는 USB 3.0 인터페이스를 통해 실시간 RGB 프레임(640×480 해상도, 30FPS)과 동기화된 Depth map을 제공하였다. 또한 향후 확장 가능성을 고려하여 협동 로봇을 좌측에 배치하였으며, 로봇 제어기는 별도 제어 모듈을 통해 좌표 데이터를 기반으로 작업할 수 있도록 [Fig. 1]과 같이 구성하였다.

3.2 커스텀 데이터셋 구축

커스텀 모델 학습을 위해, 구축된 데이터셋은 기존 공개 데이터셋이 아닌, 실제 청소기 제조 도메인에 특화된 데이터셋으로 직접 구축되었다. 초기에 대상 객체인 부품 5종을 다양한 배경, 조도, 각도에서 촬영하여, 총 10,000장을 구축하였으며, 촬영된 원본 RGB 이미지에 대해서는 제조 환경의 특성을 반영한 다음 [Table 2]와 같은 체계적인 데이터 증강 기법들이 적용되었으며, [Fig. 2]는 각 증강 기법들이 적용된 결과를 보여준다. 00번은 원본 이미지이며, 01-11번은 [Table 2]에 정의된 증강 기법의 단일 적용 결과, 12-20번은 여러 증강 기법들을 조합한 결과이다. 증강 이후 총 100,000장의 학습 이미지와 이에 대응하는 YOLO 형식(.txt) 및 COCO 형식(.json)의 바운딩 박스 레이블을 생성하였다. 클래스 분포는 5종 부품이 1:1:1:1:1 비율로 유지하였다.

3.3 데이터 분할 및 학습 설정

전체 데이터는 Train/Valid/Test를 8:1:1로 분할하였으며, 모



[Fig. 2] Data augmentation examples applied to the custom manufacturing dataset: Images 00-11 show single augmentation techniques, and images 12-20 show combinations of multiple methods, as defined in Table 2

[Table 2] Data augmentation techniques applied for custom dataset enhancement

Categories	Methods	Purpose and effect
Color Jitter	Color Jitter (Hue, saturation, brightness adjustment)	Enhancement of model robustness against lighting and color variations
Noise	Gaussian noise	Enhancement of durability against image quality changes due to sensor noise
	Random blur	
Flip	Horizontal	Enhancement of robustness against pose changes and spatial orientation variations
	Vertical	
Image Scale Conversion	Random Scale	Enhancement of scale invariance and multi-scale adaptation capability
	Image Resize	
	Crop	
Cutout	Object Cutout (Partial masking)	Enhancement of capability to handle partial object masking scenarios
Shear	Horizontal (±10°)	Enhancement of robustness against various viewpoint changes
	Vertical (±10°)	
Rotation	Between ± 15%	Enhancement of robustness against various viewpoint changes
Brightness	Between ± 25%	Enhancement of robustness against various illumination changes

든 모델은 동일한 분할 조건에서 학습되었다. 실험에 사용된 모델은 YOLOv10-12, RT-DETR 총 4종이며, 각 모델은 사전학습(Pre-trained) 가중치를 기반으로 Fine-tuning 되었다. 모든 실험은 동일 GPU 환경(NVIDIA RTX 3090 Ti*2)에서 수행되었으며, 학습 Epochs는 모델별 200회, batch size는 32로 고정하고, Valid 지표는 mAP@0.5, mAP@[0.5:0.95]이 사용되었다.

4. TCAT: A Temporal Consistency-based Adaptive Object Tracking Method

4.1 TCAT 개요

본 장에서는 제조 환경의 복잡한 상황에서도 안정적인 객체 추적이 가능한 시간적 일관성 기반 적응형 추적 방법(TCAT)을 제안하며, 제안된 방법은 연속된 프레임에서의 객체 인식 결과 패턴을 분석하여 각 객체의 시간적 일관성을 정량화하고, 이를 기반으로 검출과 추적 결과의 가중치를 동적으로 조절한다. [Table 3]은 제안된 TCAT의 전체 구조 흐름을 보여준다. 객체 검출기로부터 얻은 결과는 시간적 일관성 분석 모듈을 통해 처리되며, 산출된 일관성 점수를 기반으로 적응형 가중치가 결정된다. 이후 상태 전이 모델링과 매칭 최적화 과정을 거쳐 최종 추적 결과가 생성되는 구조로 동작한다.

4.2 시간적 일관성 기반 신뢰도 분석

객체 인식 결과의 시간적 일관성을 정량화하기 위해, 연속된 프레임에서의 신뢰도 점수 패턴을 분석한다. 기존 추적 방법들이 단일 프레임의 신뢰도 점수만을 고려했던 것과 달리, TCAT

[Table 3] Pseudocode for Temporal Consistency-based Adaptive Tracking (TCAT)

Algorithm 1. Steps for temporal consistency-based tracking

Input: Frame sequence at time *t*
Input: Object detection results *D(t)*
Output: Tracked object states *O*

Initialize: Temporal weights $\alpha=0.5, \beta=0.3, \gamma=0.2$
Initialize: Circular buffer B of size 3
for time *t* **do**
  Update temporal buffer B with *D(t)*
  for i = 1, ..., N **do**
    Compute $TC(i,t) = \alpha \cdot C(i,t) + \beta \cdot C(i,t-1) + \gamma \cdot C(i,t-2)$
    Compute $w(i,t) = \sigma(1.5 \cdot TC(i,t))$
    Update $S(i,t) = w(i,t) \cdot D(i,t) + (1-w(i,t)) \cdot T(i,t)$
  end for
  for i = 1, ..., N **do**
    Compute $\Delta TC(i,t) = |TC(i,t) - TC(i,t-1)|$
    Update $P(i,t) = exp(-0.8 \cdot \Delta TC(i,t))$
  end for
  for each pair *(i,j)* **do**
    Set $\omega = 0.7 \cdot min(TC(i,t), TC(j,t-1))$
    Update $M(i,j) = \omega \cdot IoU(i,j) + (1-\omega) \cdot cos(f(i), f(j))$
  end for
  Apply Hungarian matching to optimize associations
  Update tracking states using optimized matches
  Apply TensorRT optimization for real-time processing
  Update final object trajectories *O*
end for

은 최근 3개 프레임의 신뢰도 점수 변화를 종합적으로 분석한다. 이는 제조 환경에서 발생하는 일시적인 가려짐이나 객체의 움직임으로 인한 순간적인 신뢰도 저하와 효율적인 연산 트레이드 오프 관계를 적절히 고려하여 설정한 프레임 개수이며, 시간 *t*에서의 객체 *i*에 대한 시간적 일관성 점수 TC(i,t)는 식 (1)과 같이 정의된다:

$$TC(i,t) = \alpha \cdot C(i,t) + \beta \cdot C(i,t-1) + \gamma \cdot C(i,t-2) \quad (1)$$

여기서 $C(i,t)$는 시간 t에서 객체 *i*의 검출 신뢰도 점수이며, α, β, γ는 시간 가중치로 $\alpha + \beta + \gamma = 1$을 만족한다. 현재 프레임의 신뢰도에 더 높은 가중치를 부여하되, 이전 프레임들의 정보도 적절히 반영하기 위해 본 연구에서는 경험적 분석을 통해 $\alpha = 0.5, \beta = 0.3, \gamma = 0.2$로 설정하였다. 이러한 가중치 설정은 실제 제조 환경에서의 다양한 외란 패턴 분석을 통해 결정되었으며, 특히 시간적 일관성 점수는 단순한 가중 평균이 아닌, 객체의 상태 변화 패턴을 효과적으로 포착할 수 있도록 설계되었다. 예를 들어, 작업자의 손에 의한 일시적 가려짐이 발생할 경우, 급격한 신뢰도 저하가 발생하더라도 이전 프레임들의 정보가 적절히 반영되어 객체의 존재 여부를 안정적으로 판단할 수 있다. 이는 기존의 단일 프레임 기반 방식들이 가려짐 상황에서 객체를 쉽게 놓치는 문제를 효과적으로 해결한다.

4.3 적응형 조절 메커니즘

TCAT은 시간적 일관성 점수를 기반으로 두 결과의 가중치를 동적으로 조절하는 메커니즘을 도입한다. 이는 객체의 상태가 안정적일 때는 현재 프레임의 검출 결과를, 불안정할 때는 이전 프레임들의 추적 결과를 더 신뢰하는 방식으로 동작하며, 시간 t에서의 최종 객체 상태 $S(i,t)$는 식 (2)와 같이 계산된다:

$$S(i,t) = w(i,t) \cdot D(i,t) + (1-w(i,t)) \cdot T(i,t) \quad (2)$$

여기서 $D(i,t)$는 현재 프레임의 검출 결과, $T(i,t)$는 추적 결과이며, $w(i,t)$는 적응형 가중치로 다음 식 (3)과 같이 결정된다:

$$w(i,t) = \sigma(\lambda \cdot TC(i,t)) \quad (3)$$

σ는 시그모이드 함수이며, λ는 가중치 변화의 민감도를 조절하는 파라미터이다. 시그모이드 함수를 사용함으로써 가중치의 급격한 변화를 방지하고 안정적인 전이가 가능하다. λ값은 실험적 분석을 통해 1.5로 설정되었으며, 이는 시간적 일관성 점수의 변화에 대해 적절한 수준의 민감도를 제공한다.

4.4 상태 전이 모델링

객체의 시간적 연속성 상태 변화를 더욱 정교하게 모델링하기 위해 상태 전이 확률을 도입한다. 이는 연속된 프레임에서 객체의 상태가 얼마나 자연스럽게 변화하는지를 평가하며, 특히 급격한 상태 변화가 발생할 때 이를 효과적으로 필터링하는 역할을 하며, 시간 *t*에서 객체 *i*의 상태 전이 확률 $P(i,t)$는 식 (4)와 같이 정의된다:

$$P(i,t) = exp(-\delta \cdot \Delta TC(i,t)) \quad (4)$$

여기서, $\Delta TC(i,t)$는 연속된 프레임 간 시간적 일관성 점수의 변화량이며, δ는 전이 확률의 감쇄 계수이다. 큰 변화량에 대해 지수적으로 감소하는 확률값을 산출함으로써, 급격한 상태 변화를 자연스럽게 억제한다. δ값은 0.8로 설정되었으며, 이는 실제 제조 환경에서 발생하는 상태 변화의 특성을 고려하여 결정되었다.

4.5 매칭 전략 최적화

객체 간 매칭 과정에서는 공간적 유사도와 외관 특징을 시간

적 일관성에 따라 동적으로 결합한다. 이는 기존 추적 방법들이 단순히 IoU나 외관 유사도만을 고려하던 한계를 극복하고, 제조 환경의 복잡한 상황에서도 안정적인 매칭을 가능하게 하며, 객체 i와 j 사이의 매칭 점수 $M(i,j)$는 식 (5)과 같다:

$$M(i,j) = \omega \cdot IoU(i,j) + (1-\omega) \cdot cos\_sim(f(i), f(j)) \quad (5)$$

여기서 $IoU(i,j)$는 바운딩 박스 간 교차 비율, $f(i)$와 $f(j)$는 각각의 외관 특징 벡터, $cos_sim(f(i), f(j)) = f(i) \cdot f(j) / (\|f(i)\| \cdot \|f(j)\|)$는 코사인 유사도, ω는 두 항의 상대적 중요도를 조절하는 파라미터이다. 특히 ω는 시간적 일관성 점수에 따라 식 (6)과 같이 동적으로 조절된다:

$$\omega = \rho \cdot min(TC(i,t), TC(j,t)) \quad (6)$$

여기서 ρ는 조절 계수로써, 0.7로 설정되었다. 이러한 동적 가중치 조절을 통해 시간적 일관성이 높은 객체들 간의 매칭에서는 공간적 정보에, 일관성이 낮은 경우에는 외관 특징에 더 높은 가중치가 부여된다.

4.6 최종 매칭 및 구현 최적화

본 절에서는 4.2-4.5절에서 제안한 시간적 일관성 기반 메커니즘을 통합하여 구현한 전체 추적 파이프라인의 구성과, 실시간 적용을 위한 최적화 기법 및 연산 성능 검증 내용을 기술한다.

4.6.1 헝가리안 매칭 최적화

식 (5)로 계산된 매칭 점수 행렬을 입력으로 하여 헝가리안 알고리즘을 통한 전역 최적 할당을 수행한다. 기존 ByteTrack의 순차적 매칭과 달리, 모든 검출-추적 쌍의 조합을 동시에 고려하여 전체 시스템 관점에서 최적인 매칭을 보장한다. 이는 특히 여러 객체가 교차하거나 가려짐이 발생하는 복잡한 상황에서 ID 스위치를 효과적으로 방지한다.

4.6.2 추적 상태 업데이트

매칭 결과에 따라 각 추적 객체의 상태를 체계적으로 갱신한다. 성공적으로 매칭된 객체들은 현재 검출 정보로 위치, 크기, 외관 특징을 업데이트하며, 매칭에 실패한 객체는 현재 프레임에서는 업데이트되지 않으며, 일정 프레임 동안 추적 상태를 유지한 후 소멸 처리된다. 새롭게 나타난 검출 결과는 연속된 프레임에서의 안정성을 확인한 후 새로운 추적 ID를 할당받는다.

4.6.3 실시간 처리 최적화

실시간 처리 요구사항을 만족하기 위해 TCAT의 각 모듈에는 다음과 같은 최적화 기법들이 적용되었다. 먼저, 시간적 일관성 점수 계산을 위한 순환 버퍼를 도입하여 이전 프레임들의 정보를 효율적으로 관리한다. 버퍼 크기는 3으로 설정되어 메모리 사용량을 최소화하면서도 필요한 시간적 정보를 모두 포함할 수 있다. 매칭 과정에서는 행렬 연산 최적화를 통해 병렬 처리 효율을 높였으며, 특히 객체 간 유사도 계산 단계(식 5의 IoU 및 cosine similarity 연산)를 중심으로 TensorRT 최적화를 적용하여, 연산 병목을 줄이고 Edge device에서의 실시간 처리 성능을 향상시켰다. 또한 외관 특징 벡터의 계산 결과를 캐싱하여 중복 연산을 방지하였다.

4.6.4 최종 궤적 생성

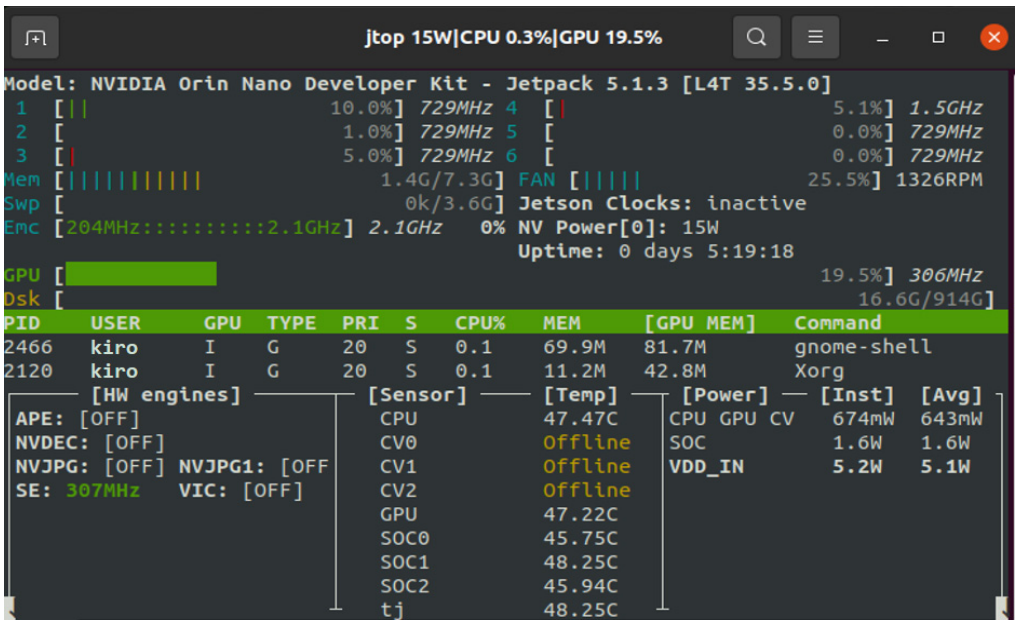
업데이트된 객체의 추적 정보를 바탕으로, 각 객체의 ID 기반 궤적 데이터 (trajectory O)가 구성된다. 이 궤적은 시간 순으로 정렬된 바운딩 박스 위치, 추적 ID, 시간적 일관성 점수 $TC(i,t)$, 상태 전이 확률 $P(i,t)$ 정보를 포함하며, 각 객체가 장면 내에서 어떤 이동/변화를 겪었는지 구조화된 형태로 추적된다.

4.6.5 연산 성능 검증

제안된 TCAT의 연산량 성능 검증을 위해, Edge device에서 실행 결과를 관찰하였으며, 실험은 NVIDIA Jetson Orin Nano Developer Kit (Jetpack 5.1.3)에서, YOLO 모델의 입력 해상도인 640×480와 Batch size를 32로 설정하여 실시간 처리 환경을 모사하였다. 시스템 모니터링 결과를 아래 [Fig. 3]에서 확인할 수 있듯이, GPU 사용률은 평균 19.5% (306 MHz), CPU 사용률은 0.3% (729 MHz) 수준을 유지하였으며, 메모리 사용량은 약 1.4 GB로 측정되었다. 특히 GPU 온도는 47.22℃로 안정적인 수준을 유지하였다. 최적화 구현 결과 TCAT 적용에 따른 연산 부하는 평균 약 0.5 FPS 내외로, 기존 파이프라인 대비 실시간 성능에는 큰 영향을 주지 않는 수준의 처리 속도 감소가 발생하였다.

5. 실 험

본 장에서는 제안된 TCAT의 성능을 정량적, 정성적 두 가지 측면으로 성능을 확인한다. 먼저, 정량적 검증에서는 학습 과정에서 사용한 Valid 평가 지표를 그대로 Test 평가 지표로 사용하였으며, 추가로 TCAT를 기존 모델과 함께 사용했을 때, 프레임 딜레이 여부를 같이 검증하였다. 정량적 검증 측면에서는 TCAT를 기존 객체 인식 및 추적 모델과 함께 사용했을 때, 인식 성능과 프레임 딜레이의 변화이고, 정성적 검증 측면에서는 제조 현장의 주요 외란 요소인 작업자와 로봇의 부품 파지로 인한 객체 가려짐, 부품의 자세 변화 상황에서의 성능을 중점적으로 확인하였다.



[Fig. 3] System monitoring results: on NVIDIA Jetson Orin Nano show GPU (19.5%), CPU (0.3%), and memory consumption (1.4 GB) usage while maintaining 15 FPS for real-time tracking

5.1 정량적 성능 평가

제안된 TCAT의 성능을 평가하기 위해 3장에서 구축한 데이터셋의 Test set (10%)과 구축된 커스텀 데이터로 학습된 커스텀 객체 인식 모델 YOLOv10-12와 함께 사용할 수 있는 추적 모델 ByteTrack/BoTSORT의 조합, 그리고 RT-DETR을 포함한 총 7가지 상황에 대해 결과를 확인하였다. 평가 지표로는 mAP@0.5, mAP@[0.5:0.95]두 가지를 사용하였으며, 연산량 검증을 위해 FPS도 함께 측정하였다. 그리고 [Table 4]는 각 모델 구성에 대한 TCAT 적용 전후의 성능을 비교한 결과를 보여준다. TCAT을 적용하지 않은 경우, YOLOv12와 ByteTrack의 조합이 mAP@0.5 기준 **82.3%**로 가장 우수한 성능을 보였다. 4.2절에서 제안한 시간적 일관성 점수 $TC(i,t)$가 최근 3개 프레임의 신뢰도 점수를 $\alpha=0.5, \beta=0.3, \gamma=0.2$의 가중치로 결합하여

[Table 4] Performance comparison of different proposed tracking methods with and without TCAT. Check marks (✓) indicate TCAT application

Case	Methods	TCAT (Proposed)	mAP@0.5	mAP@[0.5:0.95]	FPS (on Jetson Orin nano Dev. Kit)
1	YOLOv10 + Bytetrack	-	80.1	75.3	11.8
	YOLOv10 + Bytetrack	✓	**83.4 (+3.3)**	**77.8 (+2.5)**	**11.2 (-0.6)**
2	YOLOv11 + Bytetrack	-	81.5	76.2	9.8
	YOLOv11 + Bytetrack	✓	**84.2 (+2.7)**	**78.5 (+2.3)**	**9.3 (-0.5)**
3	YOLOv12 + Bytetrack	-	82.3	77.1	8.7
	YOLOv12 + Bytetrack	✓	**85.7 (+3.4)**	**79.8 (+2.7)**	**8.3 (-0.4)**
4	YOLOv10 + Botsort tracker	-	79.8	74.9	11.5
	YOLOv10 + Botsort tracker	✓	**83.1 (+3.3)**	**77.4 (+2.5)**	**11.0 (-0.5)**
5	YOLOv11 + Botsort tracker	-	81.2	75.8	9.5
	YOLOv11 + Botsort tracker	✓	**83.9 (+2.7)**	**78.1 (+2.3)**	**9.1 (-0.4)**
6	YOLOv12 + Botsort tracker	-	82.0	76.8	8.5
	YOLOv12 + Botsort tracker	✓	**85.2 (+3.2)**	**79.4 (+2.6)**	**8.1 (-0.4)**
7	RT-DETR	-	81.5	76.8	4.8
	RT-DETR	✓	**84.2 (+2.7)**	**78.9 (+2.1)**	**4.2 (-0.6)**

계산되므로, 일시적인 신뢰도 저하 상황에서도 약 3~4%p의 성능 향상이 이론적으로 예측된다. 실제 실험 결과에서도 TCAT 적용 시 YOLOv12+ByteTrack 조합이 **85.7%**까지 성능이 향상되어 이론값과 유사한 **3.4%p**의 개선을 보였다. BoTSORT를 활용한 경우, Re-ID 특징을 추가로 사용함에도 불구하고 ByteTrack 대비 약간 낮은 성능을 보였다. 그러나 TCAT 적용 시에는 4.5절의 매칭 전략 최적화를 통해 시각적 특징과 시간적 일관성이 $\rho=0.7$의 비율로 결합되면서, YOLOv12+BoTSORT 조합에서도 **85.2%**의 높은 성능을 달성하였다. 이는 TCAT의 적응형 가중치 조절 메커니즘이 기존 추적 방법들의 장점을 살리면서도 시간적 일관성이라는 새로운 관점을 효과적으로 통합할 수 있음을 보여준다.

처리 속도 측면에서는 NVIDIA Jetson Orin Nano 환경에서 YOLOv10 기반 모델이 11 FPS 수준로 가장 빠른 성능을 보였고, YOLOv11과 v12는 모델 크기 증가로 인해 각각 9 FPS, 8 FPS 수준의 처리 속도를 달성하였다. RT-DETR은 Transformer 구조임에도 최적화된 구현을 통해 기존 대비 **0.6 FPS**가 감소한 처리 속도를 확인하였다.

5.2 Ablation Study

제안된 TCAT의 각 구성 요소가 전체 성능에 미치는 영향을 정량적으로 분석하기 위해 ablation study를 수행하였다. 실험은 YOLOv12+ByteTrack 조합을 기준으로 하며, 각 구성 요소를 순차적으로 제거하여 성능 변화를 관찰하였다.

5.2.1 구성 요소별 기여도 분석

[Table 5]는 TCAT의 각 구성 요소별 성능 기여도를 보여준다. 기본 ByteTrack 대비 시간적 일관성 점수(TC)만 적용했을 때 2.1%p의 성능 향상을 보였으며, 적응형 가중치 조절(AW)을 추가했을 때 추가로 0.8%p 향상되었다. 매칭 전략 최적화(MS)까지 모두 적용했을 때 최대 3.4%p의 성능 향상을 달성하였다.

[Table 5] Ablation study results for TCAT components

Components	TC	AW	MS	mAP@ 0.5(±)	FPS(±)
Baseline(v12+ ByteTrack)	-	-	-	82.3	8.7
+Temporal Consistency	✓	-	-	84.4 (+2.1)	8.5 (-0.2)
+Adaptive Weighting	✓	✓	-	85.2 (+2.9)	8.4 (-0.3)
+ Matching Strategy (Full TCAT)	✓	✓	✓	**85.7 (+3.4)**	8.3 (-0.4)

[Table 6] Effect of temporal window size (based on Full TCAT)

Window size (Frames)	mAP@ 0.5(±)	FPS(±)	E
1	82.3	8.7	0.0
2	84.5(+2.2)	8.5(-0.2)	2.0
3	85.7(+3.4)	8.3(-0.4)	**3.0**
4	85.8(+3.5)	7.7(-1.0)	2.5
5	86.2(+3.9)	7.1(-1.6)	2.3

[Table 7] Analysis of temporal weight parameters (based on Full TCAT)

$\alpha(t)$	$\beta(t-1)$	$\gamma(t-2)$	mAP@ 0.5(±)	Characteristics
0.7	0.2	0.1	82.3	Current frame focused
0.5	**0.3**	**0.2**	**85.7(+3.4)**	**Balanced weighting**
0.3	0.4	0.3	84.8(+2.5)	Past frame focused
0.33	0.33	0.33	84.2(+1.9)	Uniform weighting

5.2.2 시간 윈도우 크기 영향 분석

TCAT 구성 요소 별 성능 기여도 확인에 이어, 시간적 일관성 계산에 사용되는 프레임 수가 성능에 미치는 영향을 분석하였으며, [Table 6]은 그 결과를 나타낸다. [Table 6] 결과들을 수치적으로 개선 정도를 확인하기 위해, mAP 대비 FPS 성능을 식 (7)과 같이 계산하여 효율성을 확인하였다. 결과에서 3개 프레임을 연산에 사용했을 때, 효율성 지표가 최고 값을 보여 성능과 속도의 최적 균형적임을 정량적으로 확인하였다.

$$E(Efficiency)=\Delta mAP@0.5+\Delta FPS \tag{7}$$

5.2.3 구성 요소별 기여도 분석

다음으로, 식 (1)의 시간 가중치 α(Current frame, t), β(t-1), γ(t-2)의 조합이 성능에 미치는 영향을 [Table 7]과 같이 분석하였다. 분석 결과, 현재 프레임 정보인 $\alpha(t)$와, 과거 프레임 정보인 $\beta(t-1),\gamma(t-2)$를 *1:1* 조합 균형인$\alpha(t):\beta(t-1)\gamma(t-2): = 0.5: 0.3:0.2$ 가 최적 성능을 보였다.

5.3 정성적 성능평가

5.3.1 비외란 상황에서의 추적 성능 확인

제안된 TCAT의 정성적 성능을 확인하기 위해, 정량적 비교에서 가장 우수한 성능을 보인 YOLOv12+ByteTrack 조합을 기준으로 TCAT 적용 유, 무에 따른 추적 결과를 비교하였다. 협업 워크셀 환경에서 로봇 단독 파지, 작업자 단독 파지, 인간-로봇 협동 파지 상황별로 각 객체의 인식 및 추적 성능을 [Fig. 4]와 같이 관찰하였다. TCAT의 가장 큰 특징은 시간적 일관성을

[Fig. 4] Comparison of object tracking results with and without TCAT in various task scenarios: (a) Robot grasp, (b) Human grasp, and (c) Collaborative grasp. TCAT maintains stable tracking performance even under occlusions by worker's hands and pose changes

기반으로 이전 프레임들의 객체 정보를 활용하여 현재 프레임의 추적 신뢰도를 보다 안정적으로 보완한다는 점이다. 비교 실험의 구성으로, TCAT는 최근 3개의 프레임을 사용하여 시관 일관성을 곗산하는 점을 고려하여, 정성적인 비교 실험 또한 3개의 연속 프레임 이미지들을 활용하였다. 결과를 살펴보면, TCAT를 적용하지 않았을 때 객체 인식 후, 연속적인 추적이 실패하는 상황이 존재하지만, TCAT를 적용했을 땐, 시관 일관성 정보를 기반으로 인식 후 안정적인 객체 추적이 이뤄짐을 확인할 수 있다. 특히 협동 파지 상황 [Fig. 4(c)]에서 이러한 장점이 두드러진다. 작업자의 양손에 의해 객체가 크게 가려지는 상황에서도, TCAT는 이전 프레임들에서 축적된 시간적 일관성 정보를 바탕으로 객체의 존재를 지속적으로 추적할 수 있음을 확인하였다. 또한 [Fig. 4(a)]의 Body brush 추적 결과에서는 로봇이 파지한 상태로 객체의 형태가 크게 틀어진 상황에서도 TCAT 적용 시 안정적인 객체 추적이 가능했다. 이는 4.2절에서 제안한 시간적 일관성 점수가 객체의 상태 변화를 효과적으로 분석하기 때문이다.

5.3.2 외란 상황에서의 추적 성능 확인

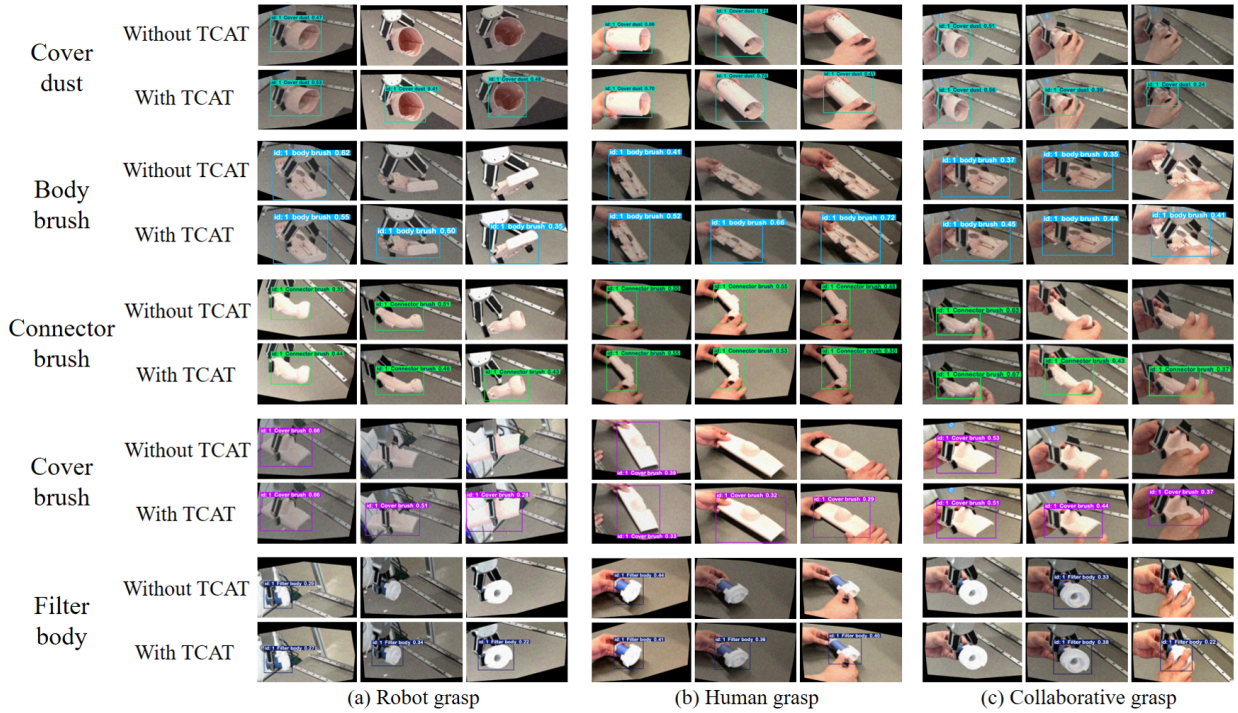
제조 환경에서 발생 가능한 복합적인 외란 상황에 대한 TCAT의 강건성을 검증하기 위해 추가적인 정성적 성능 평가를 수행하였다. [Fig. 4]의 기본 시나리오에 더하여, 실제 제조 현장에서 동시에 발생할 수 있는 다양한 외란 조건을 시뮬레이션하기 위해, 4가지 증강 기법을 복합적으로 적용하였으며, 적용된 외란 효과는 다음과 같다:

- 조명 외란(±25%): 작업 환경의 조명 조건 변화를 모사하기 위한 밝기 조절
- 색상 왜곡(색조, 채도 각 ±20%): 카메라 설정이나 조명 색온도 변화로 인한 색조 및 채도 변화
- 노이즈(가우시안 $\sigma=15$, 블러 커널 $3*3$): 실제 제조 현장의 노이즈와 진동으로 인한 블러 효과
- 스케일 변화(회전 ±10%, Shear ±10%): 카메라 각도 변화나 객체 움직임으로 인한 회전 및 전단 변형

[Fig. 5]는 외란 효과를 적용한 이미지에 대해, TCAT를 적용했을 때와 적용하지 않았을 때 추적 결과를 보여준다. 결과에서 확인할 수 있듯이 조명 변화 상황, 노이즈가 있는 상황에서 TCAT를 적용하지 않았을 때 비해, 상대적으로 안정적인 추적 결과를 확인할 수 있다. 아울러 이미지 스케일 변화에 대해서도 강건한 객체 추적 결과를 확인할 수 있었다.

6. 결 론

본 연구에서는 제조 환경의 복잡한 상황에서도 안정적인 객체 추적을 목적으로 시간적 일관성 기반 적응형 추적 방법(TCAT)을 제안하였다. TCAT은 연속된 프레임에서의 객체 인



[Fig. 5] Combined disturbance evaluation (brightness ±25%, color ±20%, noise σ=15, rotation ±15°, shear ±5~10%) across three scenarios: (a) Robot grasp, (b) Human grasp, (c) Collaborative grasp. A comparison between systems without TCAT and those with TCAT demonstrates that TCAT achieves superior robustness in recognition and tracking under severe disturbance conditions

식 결과 패턴을 분석하여 각 시간적 일관성을 정량화하고, 이를 기반으로 검출과 추적 결과의 가중치를 동적으로 조절한다. 제안된 방법의 검증을 위해 실제 청소기 제조 공정의 5종 부품을 대상으로 한 대규모 커스텀 데이터셋 10만장을 구축하였으며, 최신 객체 인식 모델을 활용하여 정량적, 정성적 성능을 확인하였다. 실험 결과, YOLOv12와 ByteTrack 조합에서 mAP@0.5 기준 최고 3.4%p의 성능 향상을 달성하였으며, 특히 양손 가림과 같은 상황과 조명, 이미지 노이즈, 스케일 변화 상황에서도 안정적인 추적 성능을 유지할 수 있음을 확인하였다. 향후 연구에서는 정확도-연산량 간 트레이드 오프 관계의 개선과 더불어, 실질적인 HRC의 실험 및 검증과 다양한 제조업 도메인에서의 시간적 일관성 분석을 위한 효율적인 최적화 방안에 대한 연구를 진행할 예정이다.

References

[1] Z. Kemény, J. Váncza, L. Wang, and X. V. Wang, "Human–robot collaboration in manufacturing: a multi-agent view," *Advanced Human-Robot Collaboration in Manufacturing*, 1st ed. Springer, 2021, ch. 1, pp. 3-41, DOI: 10.1007/978-3-030-69178-3_1.

[2] J. Fan, Y. Yin, T. Wang, W. Dong, P. Zheng, and L. Wang, "Vision-language model-based human-robot collaboration for smart manufacturing: a state-of-the-art survey," *Frontiers of Engineering Management*, vol. 12, no. 1, pp. 177-200, Jan., 2025, DOI: 10.1007/s42524-025-4136-9.

[3] M. Shaaban, A. Carfì, and F. Mastrogiovanni, "Digital twins for human-robot collaboration: a future perspective," *arXiv:2311.02421*, 2023, [Online], https://arxiv.org/abs/2311.02421.

[4] Y. Lu, X. Xu, and L. Wang, "Smart manufacturing process and system automation – A critical review of the standards and envisioned scenarios," *Journal of Manufacturing Systems*, vol. 56, pp. 312-325, Jul., 2020, DOI: 10.1016/j.jmsy.2020.06.010.

[5] J. Wan, X. Li, H.-N. Dai, A. Kusiak, M. Martínez-García, and D. Li, "Artificial intelligence-driven customized manufacturing factory: key technologies, applications, and challenges," *arXiv:2108.03383*, 2020, [Online], https://arxiv.org/abs/2108.03383.

[6] K. Darwish, F. Wanderlingh, B. Bruno, E. Simetti, F. Mastrogiovanni, and G. Casalino, "Flexible human-robot cooperation models for assisted shop-floor tasks," *arXiv:1707.02591*, 2017, [Online], https://arxiv.org/abs/1707.02591.

[7] L. M. Amaya-Mejía, N. Duque-Suarez, D. Jaramillo-Ramírez, and C. Martinez, "Vision-based safety system for barrierless human-robot collaboration," *arXiv:2208.02010*, 2022, [Online], https://arxiv.org/abs/2208.02010.

[8] N. Robinson, B. Tidd, D. Campbell, D. Kulic, and P. Corke, "Robotic vision for human-robot interaction and collaboration: a survey and systematic review," *arXiv:2307.15363*, 2023, [Online], https://arxiv.org/abs/2307.15363.

[9] A. Wang, H. Chen, L. Liu, K. Chen, Z. Lin, J. Han, and G. Ding, "YOLOv10: real-time end-to-end object detection," *arXiv:2405.*

14458, 2024, [Online], https://arxiv.org/abs/2405.14458.

[10] R. K. Hanam and M. Hussain, "YOLOv11: an overview of the key architectural enhancements," *arXiv:2410.17725*, 2024, [Online], https://arxiv.org/abs/2410.17725.

[11] Y. Tian, Q. Ye, and D. Doermann, "YOLOv12: attention-centric real-Time object detectors," *arXiv:2502.12524*, 2025, [Online], https://arxiv.org/abs/2502.12524.

[12] Y. Zhao, W. Lv, S. Xu, J. Wei, G. Wang, Q. Dang, Y. Liu, and J. Chen, "DETRs beat YOLOs on real-time object detection," *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, USA, pp. 16965-16974, 2024, DOI: 10.1109/CVPR52733.2024.01605.

[13] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, "End-to-end object detection with transformers," *European Conference on Computer Vision (ECCV)*, Online, pp. 213-229, 2020, DOI: 10.1007/978-3-030-58452-8_13.

[14] J. Beal, E. Kim, E. Tzeng, D. H. Park, A. Zhai, and D. Kislyuk, "Toward transformer-based object detection," *arXiv:2012.09958*, 2020, [Online], https://arxiv.org/abs/2012.09958.

[15] K. Saleh, S. Szenasi, and Z. Vamossy, "Occlusion handling in generic object detection: a review," *arXiv:2101.08845*, 2021, [Online], https://arxiv.org/abs/2101.08845.

[16] Y. Zhang, P. Sun, Y. Jiang, D. Yu, F. Weng, Z. Yuan, P. Luo, W. Liu, and X. Wang, "ByteTrack: multi-object tracking by associating every detection box," *European Conference on Computer Vision (ECCV)*, Tel Aviv, Israel pp. 1-21, 2022, DOI: doi.org/10.1007/978-3-031-20047-2_1.

[17] N. Aharon, R. Orfaig, and B.-Z. Bobrovsky, "BoT-SORT: robust associations multi-pedestrian tracking," *arXiv:2206.14651*, 2022, [Online], https://arxiv.org/abs/2206.14651.

[18] A. Bewley, Z. Ge, L. Ott, F. Ramos, and B. Upcroft, "Simple online and realtime tracking," *IEEE International Conference on Image Processing (ICIP)*, Phoenix, AZ, USA, pp. 3464-3468, 2016, DOI: 10.1109/ICIP.2016.7533003.

[19] X. Zhao, Y. Zhang, and Y. Zhou, "Pose tracking and object reconstruction based on occlusion relationships in complex environments," *Applied Sciences*, vol. 14, no. 20, pp. 9355-9375, Oct. 2024, DOI: 10.3390/app14209355.

[20] L. Zhong, X. Zhao, Y. Zhang, S. Zhang, and L. Zhang, "Occlusion-aware region-based 3D pose tracking of objects with temporally consistent polar-based local partitioning," *IEEE Transactions on Image Processing*, vol. 29, pp. 5065-5078, Feb., 2020, DOI: 10.1109/TIP.2020.2973512.

[21] Z. Song, R. Luo, L. Ma, Y. Tang, Y.-P. P. Chen, J. Yu, and W. Yang, "Temporal coherent object flow for multi-object tracking," *AAAI Conference on Artificial Intelligence*, vol. 39, no. 7, pp. 6978-6986, Apr., 2025, DOI: 10.1609/aaai.v39i7.32749.

[22] Y. Han and K. Huang, "ACTrack: adding spatio-temporal condition for visual object tracking," *arXiv:2403.07914*, 2024, [Online], https://arxiv.org/abs/2403.07914.

[23] R. Gao and L. Wang, "MeMOTR: long-term memory-augmented transformer for multi-object tracking," *IEEE/CVF International Conference on Computer Vision (ICCV)*, Paris, France, pp. 9901-9910, 2023, DOI: 10.1109/ICCV51070.2023.00908.

[24] J. Cai, M. Xu, W. Li, Y. Xiong, W. Xia, Z. Tu, and S. Soatto, "MeMOT: multi-object tracking with memory," *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, New Orleans, Louisiana, USA, pp. 8090-8100, 2022, DOI: 10.1109/CVPR52688.2022.00792.

[25] H. M. Ahmad, A. Rahimi, and K. Hayat, "Capacity constraint analysis using object detection for smart manufacturing," *arXiv: 2402.00243*, 2024, [Online], https://arxiv.org/abs/2402.00243.

[26] H. M. Ahmad and A. Rahimi, "Deep learning methods for object detection in smart manufacturing: a survey," *Journal of Manufacturing Systems,* vol. 64, pp. 181-196, Jul., 2022, DOI: 10.1016/j.jmsy.2022.06.011.

[27] M. Ahmed, K. A. Hashmi, A. Pagani, M. Liwicki, D. Stricker, and M. Z. Afzal, "Survey and performance analysis of deep learning based object detection in challenging environments," *Sensors*, vol. 21, no. 15, pp. 5116-5145, Jul., 2021, DOI: 10.3390/s21155116.

[28] S. Liang, W. Wang, R. Chen, A. Liu, B. Wu, E.-C. Chang, X. Cao, and D. Tao, "Object detectors in the open environment: challenges, solutions, and outlook," *arXiv:2403.16271*, 2024, [Online], https://arxiv.org/abs/2403.16271.

[29] A. Guleria, K. Varshney, G. Shweta, and S. Jindal, "A systematic review: object detection," *AI & Society*, pp. 1-18, Apr., 2025, DOI: 10.1007/s00146-025-02372-0.

[30] Y. Li, J. Li, W. Lin, and J. Li, "Tiny-DSOD: lightweight object detection for resource-restricted usages," *arXiv:1807.11013*, 2018, [Online], https://arxiv.org/abs/1807.11013.

[31] R. Muller, M. Vette, and M. Scholer, "Robot workmate: a trustworthy coworker for the continuous automotive assembly line and its implementation," *Procedia CIRP*, vol. 44, pp. 263-268, 2016, DOI: 10.1016/j.procir.2016.02.077.

[32] L. Xia, J. Lu, Y. Lu, Y. Fan, Z. Wang, and H. Zhang, "Vision Ai+Ar: integrating augmented reality and small object detection for enhanced production logistics in flexible workshop," *SSRN Electronic Journal*, Aug., 2024, [Online], https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4925306.

[33] A. Wang, Y. Sun, A. Kortylewski, and A. Yuille, "Robust object detection under occlusion with context-aware CompositionalNets," *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Online, pp. 12645-12654, 2020, DOI: 10.1109/CVPR42600.2020.01266.

[34] P. Sun, J. Cao, Y. Jiang, R. Zhang, E. Xie, Z. Yuan, C. Wang, and P. Luo, "TransTrack: multiple object tracking with transformer," *arXiv:2012.15460*, 2021, [Online], https://arxiv.org/abs/2012.15460.

[35] N. Wang, W. Zhou, J. Wang, and H. Li, "Transformer meets tracker: exploiting temporal context for robust visual tracking," *arXiv:2103.11681*, 2021, [Online], https://arxiv.org/abs/2103.11681.

[36] P. Liao, F. Yang, D. Wu, J. Yu, W. Zhao, and D. Zhang, "Fast TrackTr: towards real-time multi-object tracking with transformers," *arXiv:2411.15811*, 2024, [Online], https://arxiv.org/abs/2411.15811.

황 도 경

2015 경상국립학교 메카트로닉스공학과(학사)
2020 부산대학교 로봇융합전공(전자공학과) (석사)
2020-2024 POSTECH 인공지능연구원 연구원
2024~현재 한국로봇융합연구원 주임연구원

관심분야: Object Detection, Image processing, VLM/VLA

김 민 규

2012 University of Tsukuba(공학박사)
2015~현재 한국로봇융합연구원 책임연구원
2015~현재 한국로봇융합연구원 인간로봇 상호작용연구센터 센터장

관심분야: HRI, Data Analytics, Artificial Intelligence

김 용 국

2015 한국항공대학교 항공우주공학과(학사)
2017 한국항공대학교 항공우주 및 기계공학과 (석사)
2023 한국항공대학교 항공우주 및 기계공학과 (박사)
2023~현재 한국로봇융합연구원 선임연구원

관심분야: Human-Robot Collaboration, RL

정 의 정

2013 한양대학교 전기전자제어계측학과(박사)
2014~2015 Carnegie Mellon University, The Robotics Institute, Postdoctoral Fellow
2015~현재 한국로봇융합연구원 책임연구원

관심분야: 모바일 로봇, 자율주행, 로봇 기구학, 객체 인식

이 예 준

2015 동아대학교 전자공학과(학사)
2018 경북대학교 전자공학부(석사)
2019~현재 경북대학교 미래자동차공학과 (박사수료)
2018~현재 한국로봇융합연구원 주임연구원

관심분야: Computer Vision, Vision AI, HRI