Journal of KIIT. Vol. 24, No. 6, pp. 45-59, Jun. 30, 2026. pISSN 1598-8619, eISSN 2093-7571
http://dx.doi.org/10.14801/jkiit.2026.24.6.45

VLM 기반 장면 맥락 인지와 포즈 정렬 ROI를 결합한 정책 중심 PPE 준수 판별 프레임워크

박수진*[1], 이규혁*[2], 이세연*[3], 오동현**, 김건우***

Policy-Centric PPE Compliance Determination Framework with Scene-Aware VLM and Pose-Aligned ROI

Su-Jin Park*[1], Gyu-Hyeok Lee*[2], Se-Yeon Lee*[3], Dong-Hyeon Oh**, and Gun-Woo Kim***

본 논문은 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구결과(RS-2026-25476256) 및 산업통상자원부 및 한국산업기술기획평가원(KEIT)의 연구비 지원(RS-2025-02633048)에 의한 연구결과임

요 약

산업 현장에서의 개인보호구 착용은 작업자의 안전을 지키는 핵심 요소이나, 기존 자동 모니터링 연구들은 현장 상황과 무관하게 단순 착용 여부만을 판별하는 한계를 가지고 있다. 본 논문은 시각적 위험 단서를 해석해 요구 개인보호구 목록을 생성하고, 포즈 기반 신체 부위 ROI에서 착용 증거를 검증하는 정책 중심 개인보호구 준수 판별 프레임워크를 제안한다. 제안 프레임워크는 포즈 추정, 신체 부위 ROI 생성, VLM 기반 요구 추론, 준수 판정으로 구성되며, 관절 키포인트 기반 회전 ROI로 다중 작업자 환경에서의 부위별 증거 수집 안정성을 높인다. 실험 결과, SH17 데이터셋에서 전체 준수 판정 완전 일치 정확도는 81.27%를 기록했으며, 상체 ROI 최적화를 통해 안전조끼 커버리지 재현율을 11.95%에서 70.83%로 향상시켜 실무 적용 가능성을 확인하였다.

Abstract

Personal protective equipment compliance is essential for worker safety in industrial environments. However, most existing automated monitoring studies only determine whether PPE is worn, regardless of site-specific conditions. This paper proposes a policy-centric PPE compliance framework that infers required PPE from visual risk cues and verifies wearing evidence using pose-based body-part ROIs. The framework consists of pose estimation, ROI generation, VLM-based requirement inference, and compliance determination. Keypoint-based rotated ROIs improve part-level evidence collection in multi-worker environments. On the SH17 dataset, the proposed framework achieved 81.27% end-to-end exact-match accuracy, and upper-body ROI optimization improved safety-vest coverage recall from 11.95% to 70.83%.

Keywords

ppe detection, object detection, vision-language model, CLIP, pose-aligned

* 경상국립대학교 컴퓨터공학과
- ORCID[1]: https://orcid.org/0009-0008-2573-3901
- ORCID[2]: https://orcid.org/0009-0004-3679-4540
- ORCID[3]: https://orcid.org/0009-0006-5365-4245
** ㈜유니드 기업부설연구소
- ORCID: https://orcid.org/0009-0009-7243-2674
*** 경상국립대학교 컴퓨터공학과 교수(교신저자)
- ORCID: https://orcid.org/0000-0001-5643-4797

· Received: Mar. 04, 2026, Revised: Apr. 30, 2026, Accepted: May 03, 2026
· Corresponding Author: Gun-Woo Kim
Dept. of ComputerScience and Engineering, College of IT Engineering, Gyeongsang National University, JinJu, Korea
Tel.: +82-55-772-3323, Email: gunwoo.kim@gnu.ac.kr

Ⅰ. 서 론

산업 현장에서는 추락, 낙하물 충돌, 절단, 화학물질 노출 등 다양한 위험 요인이 상존하며, 이러한 환경에서 작업자의 개인보호구(PPE, Personal Protective Equipment) 미착용 또는 부적절한 착용은 사고 발생 가능성을 높일 뿐 아니라 사고 발생 시 피해 범위를 급격히 확대시킬 수 있다[1]-[3]. 이에 따라 많은 국가에서 산업안전 관련 법·규정은 작업 유형과 위험 요인에 따라 필요한 PPE 착용을 의무로 명시하고 있으며 현장 안전관리 체계에서 PPE 준수는 가장 기본적이면서 핵심적인 예방 수단으로 다뤄진다[4]. 미국(OSHA), 유럽(EU), 일본 등 주요 국가의 산업안전 규정은 작업 환경 및 위험 요인에 대응하는 PPE 제공·착용 의무를 제도적으로 규정하고 있어 다양한 산업 현장에서 PPE 준수 여부를 체계적으로 관리할 필요를 지속적으로 강조하고 있다[5]-[7]. 그러나 대규모 작업장이나 다수 작업자가 동시에 활동하는 환경에서 PPE 착용을 상시 점검하는 것은 쉽지 않으며 규정 준수 여부를 지속적으로 모니터링할 수 있는 자동화된 감시·판별 기술의 필요성이 점차 커지고 있다.

전통적인 PPE 점검 방식은 안전 관리자가 현장을 순회하며 육안으로 확인하거나 담당자가 CCTV를 통해 수동으로 모니터링하는 형태가 일반적이다. 이러한 방식은 인력과 시간이 많이 소요되며 감시 범위가 넓어질수록 누락 가능성이 증가하고 담당자의 피로도 및 주의력 저하로 인한 실수에 취약하다는 한계를 가진다. 특히 PPE 준수 여부는 단순히 보호구가 보이는지를 확인하는 문제에 그치지 않고 장면의 작업 환경과 위험 요인에 따라 어떤 보호구가 요구되는지를 함께 판단해야 하므로 수작업 점검만으로는 일관된 기준을 적용하기 어렵다. 본 논문에서 PPE 준수 판별(PPE compliance assessment)은 특정 보호구의 단순 존재 여부나 착용 여부를 탐지하는 문제를 넘어, 입력 장면의 작업 환경과 위험 요인에 기반하여 요구 PPE를 추론하고 각 작업자가 해당 요구 PPE를 적절한 신체 부위에 착용하였는지를 검증함으로써 최종 준수 여부를 판단하는 문제로 정의한다. 따라서 본 논문에서 말하는 올바른 착용은 이미지 내 PPE가 검출되었는지 여부가 아니라, 현장 상황에 따라 요구되는 PPE 종류가 작업자의 해당 신체 부위에 요구 수준에 맞게 착용되었는지를 의미한다. 이러한 상황 기반 요구 조건은 장면마다 달라질 수 있으므로, 다양한 환경에서 확장 가능하면서도 일관된 기준으로 PPE 준수 여부를 판별할 수 있는 시스템이 요구된다.

컴퓨터 비전 기반 PPE 준수 탐지 연구는 주로 객체 탐지(Object detection)를 이용하여 작업자와 PPE를 검출한 뒤 두 객체 간의 거리·포함 관계·IoU 등 공간적 관계를 기반으로 착용 여부를 판단하는 방식으로 발전해왔다. 그러나 실제 산업 현장 이미지는 배경이 복잡하기 때문에 이미지 전체에서 PPE를 직접 탐지할 경우 오탐지 또는 미탐지가 증가할 수 있다[8]-[11]. 특히 다중 작업자가 등장하는 환경에서는 검출된 PPE를 특정 작업자에게 정확히 귀속시키는 매칭이 불안해져 다른 사람의 PPE를 잘못 연결해 착용 여부를 판단하는 문제가 발생할 수 있다.

더 나아가 기존 접근법의 상당수는 PPE가 보이는가 또는 착용으로 추정되는가에 초점을 두는 경우가 많아 현장 환경의 위험 요인에 따라 어떤 PPE가 요구되는지를 결정한 뒤 요구 목록에 대해 착용 준수를 검증하는 체계적 프레임으로 확장되기 어렵다는 한계를 가진다. 따라서 현장 PPE 준수 판단은 상황에 따른 요구 PPE 결정과 착용 증거 기반 준수 판별을 결합한 형태로 정식화할 필요가 있다.

이러한 한계를 해결하기 위해 본 논문에서는 포즈 추정 모델(Pose Estimation Model)[12]을 기반으로 작업자의 신체 부위를 안정적으로 추출하고 추출된 신체 ROI에서 PPE 착용 증거를 포착하여 착용 여부를 판별하며, 비전-언어 모델(VLM, Vision-Language Model)[13]을 활용하여 장면의 작업 맥락과 위험 요인을 해석함으로써 상황 기반 요구 PPE 목록을 생성하고 이를 착용 판별 결과와 결합해 최종 준수 여부를 판단하는 모듈형 프레임워크를 제안한다. 포즈 추정은 관절 키포인트를 통해 인체 구조 단서를 제공하므로 다중 인원 및 부분 가림 환경에서도 머리·상체·발과 같은 신체 부위 단위 ROI를 구성하는 데 유리하다. 또한 VLM 기반 장면 이

해를 결합함으로써 고정된 체크리스트가 아닌 현장 상황 적응형 요구 PPE를 생성할 수 있어 단순 착용 탐지를 넘어 안전 정책 중심의 준수 판별 체계를 구축할 수 있다.

위 연구의 기여는 다음과 같이 요약할 수 있다. 첫째, 포즈 기반 신체 부위 ROI 생성과 ROI 중심 증거 결합을 통해 다중 인원 및 복잡한 배경에서 PPE 착용 판별 안정성을 높이는 착용 판별 모듈을 제시한다. 둘째, VLM을 활용한 장면 맥락 기반 위험 요인 추론을 통해 현장 상황 적응형 요구 PPE 생성 방식을 제안한다. 셋째, 요구 PPE 목록과 착용 판별 결과를 결합하는 모듈형 파이프라인을 설계하고 데이터셋 기반 정량 평가를 통해 각 모듈의 기여와 오류 전파 특성을 분석한다.

본 논문의 구성은 다음과 같다. 2장에서는 PPE 검출 및 준수 판단, 포즈 기반 ROI/앵커링, 장면 이해 기반 정책·요구 PPE 추론 등 관련 연구를 정리한다. 3장에서는 제안하는 모듈형 프레임워크의 전체 구조와 각 모듈의 역할 및 세부 사항을 기술한다. 4장에서는 데이터셋과 라벨링 방법, 평가 지표를 소개하고 모듈별 성능 및 end-to-end 성능 결과를 보고한다. 마지막 5장에서는 결과에 대한 논의와 한계 및 향후 연구 방향을 제시한다.

II. 관련 연구

2.1 객체 탐지 기반 PPE 준수 여부 판별

산업 현장의 PPE 준수 모니터링에서 가장 널리 사용되는 접근법은 객체 탐지를 기반으로 작업자와 PPE를 검출하는 방식이다. 기존 연구들은 Faster R-CNN 계열과 같은 Two-Stage 탐지기 또는 SSD/YOLO 계열과 같은 One-Stage 탐지기를 활용하여 헬멧, 안전조끼, 안전화 등 PPE 객체를 탐지하고 탐지 결과를 기반으로 준수 여부를 판단하는 파이프라인을 구성해왔다[14]-[17]. 특히 YOLO 계열의 경량·실시간 탐지기가 산업 현장 모니터링의 실용 요구성에 부합하여 PPE 탐지 백본으로 빈번히 채택되는 경향이 보고된다. 객체 탐지 기반 접근법은 구현이 직관적이고 PPE 클래스 확장이 용이하다는 장점이 있으나 산업 현장 이미지의 특성상 탐지 성능이 장면 조건에 민감하게 변동될 수 있다. 예컨대 PPE는 작업자 대비 상대적으로 크기가 작고 부분적으로 노출되거나 다른 객체에 의해 가려지는 경우가 많아 탐지 오차가 누적되기 쉽다. 또한 다중 작업자가 동시에 등장하는 장면에서는 검출된 PPE를 특정 작업자에게 귀속시키는 단계가 성능 병목으로 작용될 수 있으며 이 단계의 불안정성은 최종 준수 판단 오류로 직접 연결된다. 이러한 한계를 완화하기 위해 최근 연구는 소형 객체 탐지 성능 강화, 사람과 PPE 간 관계 추정의 안정화, 실시간 처리 가능성을 동시에 고려하는 방향으로 확장되고 있다. 예를 들어 X. Li et al.[18]의 OAM-YOLO는 YOLO 계열 탐지기를 기반으로 소형 PPE 탐지를 강화하는 구조를 제안하고 탐지와 더불어 구조 단서를 결합하는 방식으로 준수 판별의 안정성을 개선하고자 하였다. 또한 One-Stage 프레임워크에서 작업자·PPE와 더불어 관절 키포인트를 함께 추정하는 멀티태스크 구조처럼 탐지 결과에 인체 구조 정보를 결합하여 관계 판별의 불확실성을 줄이려는 시도도 보고된다.

2.2 포즈 추정 기반 PPE 준수 여부 판별

객체 탐지 단독 접근법에서 빈번히 발생하는 오류는 다중 인원 환경에서의 사람과 PPE 간 귀속 혼동과 소형 객체 및 부분 가림 PPE의 불안정한 검출이 준수 판별 단계로 전파되는 문제로 요약될 수 있다. 이에 따라 최근 PPE 준수 연구는 포즈 추정 모델을 활용하여 인체 구조 단서를 명시적으로 도입하고 신체 부위 단위에서 PPE 증거를 검증하는 방향으로 확장되고 있다. 포즈 기반 접근은 작업자 바운딩 박스 수준의 포괄적인 단위가 아니라 작업자의 관절 키포인트를 이용해 머리/상체/발 등 부위 단위 ROI를 구성하고 해당 영역에서 PPE 단서를 확인함으로써 배경 잡음을 줄이고 관계 판별을 명확히 하는데 목적이 있다. 즉, 누가 무엇을 착용했는가 라는 문제를 사람 박스 중심의 매칭 규칙으로 처리하기보다 인체 구조에 정렬된 영역에서 증거를 수집하도록 유도하여 사람과 PPE 간 관계 매칭의

불확실성을 줄이는 방식이다. A. M. Vukicevic et al.[19]은 2D 포즈로부터 신체 랜드마크를 추출한 뒤 부위별 관심 영역을 정의하고 각 부위 단위 분류를 통해 PPE 착용 여부를 판별하는 접근을 제시하였다. 또한 R. Xiong et al.[20]은 사람–PPE 관계의 불확실성을 핵심 병목으로 보고 관절 키포인트를 공간 앵커로 사용해 머리/상체/발에 정렬된 신체 부위별 ROI를 구성한 뒤 PPE 착용 준수 여부 판별을 수행하였다. 이처럼 포즈 기반 PPE 연구는 부위 ROI 기반 판별, 앵커링 기반 관계 안정화라는 흐름으로 정리할 수 있으며 공통적으로 객체 탐지 단독 접근법의 취약점을 인체 구조 단서로 보완한다는 점에서 의의가 있다.

종합하면, 기존 PPE 준수 판별 연구는 객체 탐지 기반 PPE 검출과 포즈 기반 신체 부위 ROI 판별을 중심으로 발전해왔다. 객체 탐지 기반 방법은 구현이 직관적이고 실시간 적용 가능성이 높지만 복잡한 배경, 소형 PPE, 부분 가림, 다중 작업자 환경에서 검출 및 귀속 오류가 발생할 수 있다. 포즈 기반 방법은 관절 키포인트를 활용하여 이러한 관계 판별의 불확실성을 완화하나, 주로 사전에 정의된 PPE 클래스의 착용 여부를 확인하는 데 초점을 둔다. 따라서 기존 연구만으로는 입력 장면의 작업 환경과 위험 요인에 따라 요구 PPE를 동적으로 판단하고 해당 요구 PPE를 기준으로 준수 여부를 검증하는 데 한계가 있다. 본 연구는 이러한 필요성에 따라 VLM 기반 장면 맥락 이해를 통해 상황별 요구 PPE를 추론하고 포즈 기반 신체 부위 ROI에서 착용 증거를 검증한 뒤, 두 결과를 결합하여 최종 PPE 준수 여부를 판단하는 모듈형 프레임워크를 제안한다.

III. 제안하는 방법

본 장에서는 산업 현장 이미지 I가 주어졌을 때 작업자별 PPE 준수 여부를 산출하는 모듈형 프레임워크를 제안한다. 제안 프레임워크는 먼저 장면의 작업 맥락과 위험 단서를 바탕으로 현장 상황에서 요구되는 PPE 목록을 추론한다. 이후 작업자의 신체 부위 단위에서 요구 PPE의 착용 증거를 검증하고 두 결과를 결합하여 최종 준수 여부를 판단하는 구조를 가진다. 본 논문에서는 작업자의 신체 부위를 머리, 상체, 발의 3개 파트로 정의하고 각 파트에 대응되는 PPE 후보를 각각 헬멧, 안전조끼, 안전화로 설정한다. 제안하는 프레임워크는 장면 맥락 기반 요구 PPE를 결정하는 단계와 신체 부위 ROI 기반으로 PPE 착용 여부를 판별하는 단계를 분리하여 설계한다. 이를 통해 산업 현장별 요구 조건의 변화와 착용 증거의 불확실성을 체계적으로 다룰 수 있다. 결과적으로 본 프레임워크는 요구-착용-준수의 연결 구조를 명시적으로 포함하며 현장 상황 적응형 PPE 준수 판별을 목표로 한다.

3.1 전체 프레임워크 구조

그림 1은 제안하는 모듈형 프레임워크의 전체 구조를 나타낸다. 제안 방법은 입력 이미지 I로부터 작업자별 요구 PPE와 착용 증거를 추출한 뒤 최종 준수 여부를 산출한다. 첫 번째로, 포즈 추정 기반 관절 키포인트 추출 모듈은 YOLO 계열의 포즈 추정 모델을 사용하여 이미지 내 다중 작업자를 탐지하고 각 작업자에 대한 2D 관절 키포인트를 추정한다. 두 번째로, 관절 키포인트 기반 신체 부위 ROI 추출 모듈은 추정된 관절 키포인트를 이용하여 작업자별 머리, 상체, 발 영역을 안정적으로 구성한다. 세 번째로, VLM 기반 장면 맥락 인지 및 요구 PPE 생성 모듈은 장면의 작업 맥락과 위험 단서를 해석하여 헬멧, 안전조끼, 안전화 각각에 대해 요구 여부를 출력한다. 마지막으로, 신체 부위 ROI 기반 PPE 착용 판별 및 준수 판정 모듈은 각 신체 부위 ROI에서 PPE 착용 증거를 통합하여 작업자별 착용 여부를 판별하고, VLM 기반 장면 맥락 인지 및 요구 PPE 생성 모듈의 요구 PPE와 결합하여 최종 준수 여부를 산출한다. 요약하면 본 논문은 포즈 기반 신체 부위 ROI와 같이 구조 단서를 통해 착용 증거 추출을 안정화하고, VLM 기반 이미지 속 장면 이해를 통해 상황 적응형 요구 조건을 생성한 뒤 두 결과를 결합하여 산업 장면에서도 일관된 기준으로 PPE 준수 여부를 판단하도록 설계한다.

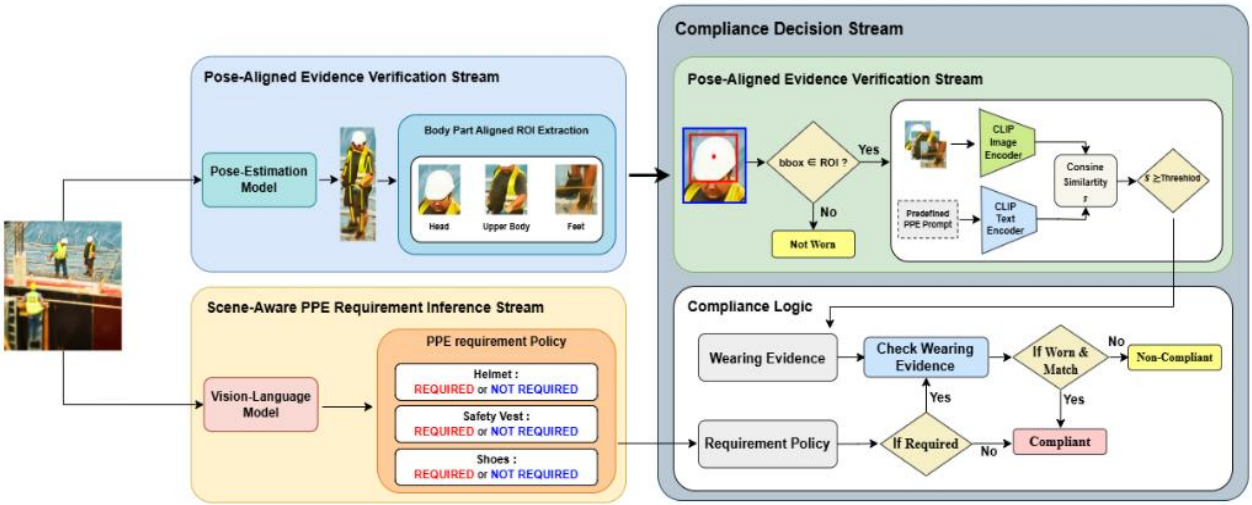

그림 1. 제안한 정책 중심 개인보호구(PPE) 준수 판별 프레임워크 개요
Fig. 1. Overview of the proposed policy-centric PPE compliance determination framework

3.2 포즈 추정 및 관절 키포인트 추출

본 모듈의 목적은 입력 이미지 I 내에서 활동 중인 다수의 작업자를 탐지하고 각 작업자의 신체 구조 단서를 제공하기 위해 2D 관절 키포인트를 추정하는 것이다. 이를 위해 탐지와 포즈 추정을 동시에 수행하는 One-stage 구조의 YOLOv8-Pose[21]를 백본으로 채택한다. 해당 모듈은 각 작업자에 대해 COCO 데이터셋 표준인 17개의 2D 관절 키포인트를 출력한다. 이는 이후 관절 키포인트 기반 신체 부위 ROI 추출 모듈에서 머리, 상체, 발 각 신체 부위별 ROI 패치를 생성하기 위한 핵심 앵커 정보로 활용된다. 모델의 출력은 작업자 바운딩 박스 B_{person}와 각 관절 키포인트 i의 좌표 (x_i, y_i) 및 키포인트 신뢰도 c_i를 포함한다. 가림 현상이나 오검출로 인한 노이즈를 완화하기 위해 키포인트 신뢰도 임계값을 $k_p = 0.3$을 적용하여 $c_i < k_p$ 인 키포인트는 후속 ROI 구성 단계에서 제외하거나 보수적으로 처리한다.

3.3 관절 키포인트 기반 신체 부위 ROI 추출

신체 부위 ROI 추출 모듈은 이전 단계에서 획득한 관절 키포인트 집합 $K = \{(x_i, y_i, c_i)\}_{i=1}^{17}$를 활용하여 머리, 상체, 발 부위의 이미지 패치를 생성한다. 산업 현장 장면에서는 배경이 복잡하고 PPE가 소형이거나 부분적으로 가려지는 경우가 많으므로, 본 논문에서는 신체 구조 단서로 정렬된 신체 부위 ROI를 구성하여 후속 착용 증거 검증의 신뢰성을 확보한다. 특히 본 모듈은 신체 방향성을 반영하기 위해 각 부위를 회전 박스 형태로 정의하여 정렬한 후 이미지 패치로 추출한다. 이때 각 신체 부위별 ROI 패치를 생성하기 위한 유효 관절 키포인트가 부족한 경우 ROI를 생성하지 않고 출력에서 제외한다. 생성된 신체 부위 ROI는 이후 PPE 착용 증거 통합 및 준수 판정 모듈의 입력으로 사용된다.

3.3.1 귀·코 앵커 기반 머리 ROI 추출

머리 영역은 양 귀의 중점 $P_{mid-ear}$를 기준으로 생성한다. 두 귀가 모두 가시적으로 나타나는 경우 ROI의 한 변 길이 l_h와 회전 각도 θ_h는 아래의 식 (1)과 (2)와 같이 결정된다.

$$l_h = r_a \cdot \| P_{lear} - P_{rear} \|_2 \tag{1}$$

$$\theta_h = \text{atan2}(y_{lear} - y_{rear},\ x_{lear} - x_{rear}) \tag{2}$$

이때 박스의 상향 방향을 보정하기 위해 상향 법선 벡터 n의 부호를 보정한다. 구체적으로 목 위치 또는 코 위치를 참조점을 사용하여 참조 벡터 d_{ref}를

구성하고 $n \cdot d_{ref}$의 부호가 일관되도록 n을 선택한다. 헬멧 상단 영역을 충분히 포함하기 위해 중심점 C_{head}는 식 (3)과 같이 이동하여 설정한다.

$$C_{head} = P_{mid-ear} + n \cdot \left(\frac{l_h}{2} \right) \tag{3}$$

또한 작업자가 뒤통수를 향하고 있거나 코가 부분적으로 가려진 상황을 대비하여 코의 y좌표가 목 위치보다 특정 비율 이상 낮은 경우 코를 참조점에서 제외하는 예외 처리를 적용한다.

3.3.2 어깨·엉덩이 앵커 기반 상체 ROI 추출

안전조끼 판별을 위한 상체 ROI는 어깨 중점 P_{mid-sh}과 엉덩이 중점 $P_{mid-hip}$을 잇는 상체 구조를 기준으로 구성된다. 어깨와 엉덩이가 모두 가시적으로 나타나는 경우 상체 ROI의 높이 h_u와 폭 w_u는 아래의 식 (4), (5)와 같이 정의한다.

$$h_u = 1.10 \cdot \| P_{mid-sh} - P_{mid-hip} \|_2 \tag{4}$$

$$w_u = 1.15 \cdot max(d_{sh}, d_{hip}) \tag{5}$$

여기서 d_{sh}와 d_{hip}는 각각 좌/우 어깨 및 좌/우 엉덩이 간 거리이며 상수 1.10과 1.15는 키포인트 위치 오차 및 자세 변화로 인해 ROI 축소 현상을 방지하고 조끼 영역을 충분히 확보하기 위한 여유 마진을 의미한다. 상체 ROI의 중심은 조끼 영역을 포함하도록 어깨 중점에서 하향 방향으로 $h_u / 2$만큼 이동하여 배치한다. 엉덩이 관절이 가려진 경우에는 아래의 식 (6)과 같이 어깨 너비 d_{sh}를 기준으로 확장 계수 s_b를 적용해 높이를 근사한다.

$$h_u = s_b \cdot d_{sh} \tag{6}$$

이와 같은 구성은 부분 가림이나 자세 변화가 큰 장면에서도 상체 영역이 과도하게 축소되는 현상을 완화한다.

3.3.3 무릎·발목 앵커 기반 발 ROI 추출

안전화를 위한 발 ROI는 발목 중점 $P_{mid-ank}$를 앵커로 하며 무릎에서 발목으로 향하는 하향 벡터 n을 방향 축으로 설정한다. 발 ROI의 폭 w_f는 발목 간 거리 기반으로 정의하고 높이 h_f는 아래의 식 (7)과 같이 발목과 무릎 간 거리 $d_{ank-kne}$에 비례하되 과도한 확장을 방지하기 위해 상한을 적용한다. 이때 상수 2.2는 폭 대비 높이의 최대 비율을 제한함으로써 키포인트 오검출이나 원근 왜곡 시 ROI가 발 영역을 벗어나 다리 위쪽으로 과도하게 확장되는 것을 방지한다.

$$h_f = \min(r_{f_h} \cdot d_{ank-kne},\ 2.2 \cdot w_f) \tag{7}$$

신발은 발목보다 아래에 위치하는 경우가 일반적이므로 중심점은 식 (8)처럼 C_{feet}에 down-shift δ를 적용해 하향 편향을 부여한다.

$$C_{feet} = P_{mid-ank} + n \cdot (h_f \cdot (0.5 + \delta)) \tag{8}$$

무릎 키포인트가 누락된 경우에는 수직 하향 방향을 기본 축으로 사용하여 보수적으로 ROI를 구성한다.

3.4 VLM 기반 장면 맥락 인지 및 요구 PPE 생성

본 모듈은 입력 이미지 I로부터 장면 내 작업 맥락과 시각적 위험 단서를 해석한다. 단순 객체의 존재 여부를 확인하는 수준을 넘어, 신체 부위별 현장 상황 기반 요구 PPE 목록을 생성하는데 목적이 있다. 이를 위해 장면 이해 및 위험 단서를 식별할 수 있는 백본으로 Qwen2.5-VL[22]을 채택하고, 이미지 정보와 텍스트 지시문을 결합한 추론을 통해 헬멧, 안전조끼, 안전화 각각의 요구 여부를 산출한다. 또한 요구 PPE 생성 과정에서는 라벨 누수 가능성을 완화해야 한다. 이를 위해 시스템 프롬프트 수준에서 실제 착용 여부가 아닌 환경적 위험 요인만을 근거로 판단하도록 제약을 부여한다. 따라서 본 모듈

은 착용 검출이 아니라 장면 기반 요구 조건 도출을 수행하도록 설계된다. 모듈의 출력은 후속 모듈과의 연동 및 데이터 일관성을 위해 구조화된 JSON 형태로 생성된다. 각 PPE별 요구 여부와 함께 판단 근거를 나타내는 증거 태그를 포함하여 설명 가능성을 확보한다. 또한 장면 맥락이 모호하거나 단서가 불충분한 경우에는 불확실성 플래그를 함께 제공하여 보수적인 정책 운영이 가능하도록 지원한다.

3.5 신체 부위 ROI 기반 PPE 착용 판별 및 준수 판정

프레임워크의 마지막 단계는 신체 부위 ROI 기반 PPE 착용 판별 및 준수 판정 모듈이다. 이 모듈은 관절 키포인트 기반 신체 부위 ROI 추출 모듈에서 생성한 신체 부위 ROI와 실제 PPE 객체 정보를 결합한다. 이를 통해 각 작업자의 부위별 PPE 착용 여부를 확인한다. 이후 착용 판별 결과를 VLM 기반 장면 맥락 인지 요구 및 PPE 생성 모듈에서 도출된 현장 상황 기반 요구 PPE 벡터와 대조하여 작업자별 최종 준수 여부를 산출한다. 본 모듈은 신체 부위 ROI 내 착용 여부 판별, CLIP 기반의 종류 일치 검증, 최종 준수 판정의 3단계 논리 구조로 동작한다.

3.5.1 신체 부위 내 PPE 착용 판별

부위별 착용 여부는 관절 키포인트 기반 신체 부위 ROI 추출 모듈의 신체 폴리곤 P_{part}와 PPE 바운딩 박스 $B_{ppe}=[x_{\min}, y_{\min}, x_{\max}, y_{\max}]$ 간의 공간적 포함 관계를 통해 판정한다. 먼저 PPE 박스의 중심점 c를 아래의 식 (9)과 같이 정의한다.

$$c = \left(\frac{x_{\min}+x_{\max}}{2}, \frac{y_{\min}+y_{\max}}{2} \right) \tag{9}$$

이후 점-다각형 검사를 수행하여 $c \in P_{part}$ 여부를 확인한다. 중심점이 신체 부위 폴리곤 내부에 포함될 경우 해당 부위에 PPE가 착용된 것으로 간주한다. 이때 단순 공간 포함 관계에 의한 오탐지를 방지하기 위해 본 모듈은 CLIP(Contrastive Language-Image Pre-training)을 활용한 시각-언어 정합성 검증을 수행한다. 착용된 것으로 판정된 신체 부위 ROI 패치 I_{part}에 대해 CLIP 이미지 인코더로 시각 임베딩 z_{img}를 추출하고, 헬멧, 안전조끼, 안전화를 대표하는 사전 정의 텍스트 프롬프트로부터 텍스트 임베딩 z_{txt}를 생성한다. 두 임베딩 간의 코사인 유사도 s는 아래의 식 (10)과 같이 계산된다.

$$s = \cos(z_{img}, z_{txt}) = \frac{z_{img} \cdot z_{txt}}{\|z_{img}\| \, \|z_{txt}\|} \tag{10}$$

유사도 s가 임계값 이상인 경우 관찰된 착용 증거가 목표 PPE의 의미적 정의와 일치한다고 판단한다.

3.5.2 최종 준수 판정 로직

최종 준수 여부는 식 (11)과 같이 정의된다. 이때 해당 PPE가 요구되지 않는 상황에서는 착용 여부와 무관하게 준수로 판정하며, 요구되는 상황에서는 착용 및 종류 일치가 모두 성립할 때에만 준수로 판정한다.

$$compliant = \begin{cases} 1, & required = 0 \\ worn \wedge match, & required = 1 \end{cases} \tag{11}$$

이러한 단계적 검증 구조는 장면 맥락 기반의 안전 정책과 신체 부위 기반의 시각적 증거가 논리적으로 일치하는 경우에만 준수를 승인함으로써 복합적인 산업 현장에서도 신뢰성 있는 모니터링 결과를 제공한다.

이상의 절차를 종합한 정책 중심 PPE 준수 판별 과정은 알고리즘 1에 요약하였다. 알고리즘 1은 입력 이미지로부터 포즈 추정, 신체 부위 ROI 생성, 장면 기반 요구 PPE 추론, 부위별 착용 증거 검증, 그리고 최종 준수 판정에 이르는 전체 처리 절차를 단계적으로 나타낸다. 이를 통해 제안 프레임워크의 각 모듈이 어떤 순서로 연동되며 최종 준수 여부가 어떻게 산출되는지를 보다 명확하게 확인할 수 있다.

알고리즘 1. 정책 중심 PPE 준수 판별 알고리즘 흐름
Algorithm 1. Algorithmic flow of policy-centric PPE compliance determination

Algorithm 1 Policy-Centric PPE Compliance Determination

Require: Input image I, PPE object boxes B_{ppe}
Ensure: Worker-level PPE compliance result C

Detect worker instances and estimate COCO-17 keypoints from I.
Infer the required PPE vector $R = [r_{\text{helmet}}, r_{\text{vest}}, r_{\text{shoes}}]$.
for each detected worker w **do**
  Construct pose-aligned ROIs for the head, upper body, and feet.
  for each PPE class $k \in \{\text{helmet}, \text{safety-vest}, \text{shoes}\}$ **do**
    Select the corresponding body-part ROI $P_{w,k}$.
    Compute the PPE center c_k and set $\text{wearing}_{w,k} \leftarrow \mathbf{1}[c_k \in P_{w,k}]$.
    if $\text{wearing}_{w,k} = 1$ **then**
      Compute $\text{match}_{w,k}$ using CLIP-based image-text matching.
    else
      Set $\text{match}_{w,k} \leftarrow 0$.
    end if
    if $r_k = 0$ **then**
      Set $\text{compliant}_{w,k} \leftarrow 1$.
    else
      Set $\text{compliant}_{w,k} \leftarrow \text{wearing}_{w,k} \wedge \text{match}_{w,k}$.
    end if
  end for
end for
return Part-level and worker-level compliance results C.

Ⅳ. 실 험

본 장에서는 제안한 모듈형 프레임워크의 성능을 검증하기 위한 데이터셋 구성 및 평가 지표를 기술하고 각 모듈과 전체 시스템의 성능 분석 결과를 보고한다.

4.1 데이터셋 및 라벨 정의

본 논문은 산업 현장 개인보호구 탐지용 공개 데이터셋인 SH17[23]을 기반으로 실험을 수행한다. SH17은 산업 환경 이미지와 함께 작업자 및 주요 PPE 객체에 대한 바운딩 박스 주석을 제공한다. 본 논문에서는 원본 주석 중 작업자 및 핵심 PPE 관련 바운딩 박스 라벨을 활용하며, 제안하는 프레임워크 평가를 위해 이미지 단위 요구 PPE 라벨을 별도로 구축하여 결합한다. 실험에 사용된 재구성 데이터는 총 7,753장의 이미지와 25,621개의 객체 라벨로 구성된다. 핵심 PPE 클래스의 분포는 헬멧 927개, 안전조끼 530개, 안전화 4,560개로 나타나며 클래스 간 불균형이 존재한다. 특히 안전화는 상대적으로 많은 객체 라벨을 포함하는 반면, 안전조끼는 표본 수가 적어 ROI 커버리지 및 착용 판별 성능이 작은 검출 오류나 ROI 누락에도 민감하게 변동될 수 있다. 또한 본 연구의 준수 판정은 단순 PPE 객체 검출이 아니라 장면 기반 요구 여부와 신체 부위 기반 착용 여부를 결합하여 산출되므로, 요구되는 사례와 요구되지 않는 사례의 분포 역시 최종 성능 해석에 영향을 줄 수 있다. 따라서 실험 결과는 데이터셋의 클래스 분포와 요구/비요구 사례의 불균형을 함께 고려하여 해석할 필요가 있다.

요구 PPE 라벨링은 이미지 내 작업자의 실제 착용 여부와 무관하게 해당 장면에서 무엇이 요구되는지를 결정하는 것을 목표로 한다. 따라서 각 PPE별 요구 여부는 이미지에서 관찰 가능한 시각적 위험 단서에 기반한 일관된 안전 정책으로 정의한다. 특히 라벨 누수를 방지하기 위해 작업자의 현재 PPE 착용 상태를 요구 판단의 근거로 사용하지 않으며 오직 환경적 위험 요소만을 근거로 판단한다. 또한 장면 해석의 모호함을 처리하기 위해 불확실성 플래그를 도입하여 사전 정의한 위험 단서가 2개 이상 명확히 관찰되는 경우 해당 PPE가 요구된 것으로 라벨링하고, 위험 단서가 관찰되지 않거나 일상 환경임이 명확한 경우에는 요구되지 않는 것으로 처리한다. 반면 위험 단서가 1개만 부분적으로 관찰되거나 장면 자체가 작업 현장인지 확신하기 어려운 경우에는 판단을 유보한 불확실 사례로 분류하여 이후 학습 및 평가에서 보수적으로 다룬다.

각 PPE의 요구 여부를 판단하기 위한 시각적 위험 단서는 사전 정의된 체크리스트를 기준으로 수행한다. 헬멧의 경우 비계나 철근 등 건설 현장 구조물, 상부 작업 위험, 중장비 근접, 산업 설비 밀집, 안전 표지 등 낙하·충돌 위험과 연관된 단서를 중심으로 정의하고, 안전조끼 차량 통행 환경, 물류 이동, 야간·비와 같은 가시성이 낮은 환경, 안전 통제선 존재 여부 등 가시성 확보 필요성에 중점을 두어 판단한다. 안전화는 흙이나 자갈 등 거친 지면, 자재 적재 및 파편 존재, 산업용 작업 바닥, 중량물 취급 등 발 보호의 필요성과 직결되는 물리적 단서를 기준으로 정의하였다. 마지막으로 판정 근거의 설명 가능성을 확보하기 위해 *evidence* 태그를 함께 기록하였으며, 요구되는 PPE에 대해 최소한의

근거를 사전 정의된 어휘로 부여한다. 라벨링은 현장 상황 파악 및 위험 단서 식별, PPE 요구 여부 매핑의 3단계 루틴으로 나눠서 수행한다.

4.2 평가 지표

본 연구는 제안한 모듈형 프레임워크의 단계별 기여와 오류 전파를 분석하기 위해 모듈 단위 지표와 End-to-End 지표를 구분하여 평가한다. 특히 SH17 기반 재구성 데이터는 PPE 클래스별 객체 수가 불균형하고 준수 판정 과정에서도 요구되는 사례와 요구되지 않는 사례의 분포가 서로 다르다. 이러한 특성으로 인해 전체 Accuracy만으로는 특정 PPE 클래스의 성능 저하나 요구 상황에서의 판정 실패를 충분히 설명하기 어렵다. 따라서 본 논문에서는 전체 정확도와 함께 클래스별 Precision, Recall, F1, 요구 상황에 한정한 Required-only Recall, 그리고 세 부위의 동시 정답 여부를 평가하는 Exact Match를 함께 사용하여 성능을 해석한다.

먼저 포즈 추정 기반 관절 키포인트 추출 모듈은 Object Keypoint Similarity(OKS)-AP/AR로 관절 키포인트 위치 정확도와 검출 성능을 측정한다. 관절 키포인트 기반 신체 부위 ROI 추출 모듈은 생성된 신체 부위 ROI가 실제 PPE 영역을 얼마나 안정적으로 포함하는지를 중심으로 평가한다. 이를 위해 GT PPE 바운딩 박스 중심점이 ROI 폴리곤 내부에 포함되는 비율인 GT Coverage Recall과 전체 ROI 중 GT 중심점을 포함하는 유효 ROI의 비율인 Pseudo Precision을 사용한다. VLM 기반 장면 및 맥락 인지 요구 PPE 생성 모듈은 이미지 단위의 헬멧, 안전조끼, 안전화 요구 여부 예측을 다중 라벨 분류로 보고 Precision, Recall, F1과 세 항목을 동시에 맞춘 Exact Match Accuracy를 사용한다. 마지막으로 신체 부위 ROI 기반 PPE 착용 판별 및 준수 판정 모듈은 ROI 기반 착용 탐지 성능을 나타내는 Worn Coverage Recall과 요구 조건과 착용 판별 결과를 결합하여 산출되는 Compliance Accuracy를 적용한다. 다만 본 연구의 준수 로직에서는 해당 PPE가 요구되지 않는 경우 착용 여부와 무관하게 준수로 판정되므로, 요구되지 않는 사례가 다수 포함될 경우 Compliance Accuracy가 과대평가될 수 있다. 이를 보완하기 위해 실제로 PPE가 요구되는 상황에 한정하여 올바르게 준수 여부를 판정했는지를 측정하는 준수 판정 재현율(Required-only recall)을 별도로 제시한다. End-to-End 성능은 한 이미지에서 모든 작업자에 대해 머리, 상체, 발 3개 부위의 준수 여부가 정답과 완전히 일치하는 End-to-End Exact Match로 측정한다.

4.3 모듈 및 전체 프레임워크 성능 결과

4.3.1 포즈 추정 기반 관절 키포인트 추출

포즈 추정 기반 관절 키포인트 추출 모듈은 아래의 표 1과 같이 COCO OKS 기반 평가에서 $AP_{0.50:0.95} = 0.784$, $AR_{0.50:0.95} = 0.839$를 기록하며 전반적으로 우수한 키포인트 추출 성능을 보였다. 세부적으로 AP_{50} / AP_{75}는 0.904 / 0.821, AR_{50} / AR_{75}는 0.938 / 0.865로 나타나 비교적 엄격한 기준에서도 안정적인 포즈 추정이 가능함을 확인하였다. 처리 속도는 평가 스크립트 기준 $5.85 img/s$ ($170.82 ms/img$)로 측정되었다. 다만 인물 크기 조건에 따라 AP / AR 저하가 관찰되어 거리, 가림, 블러 등 장면 조건에 따른 포즈 품질 변동이 이후 ROI 추출 단계의 안정성에 영향을 줄 수 있음을 시사한다.

표 1. 포즈 추정 기반 관절 키포인트 추출
Table 1. Worker pose estimation performance

Metric	0.50:0.95	@0.50	0.75
AP	0.784	0.904	0.821
AR	0.839	0.938	0.865

4.3.2 신체 부위 ROI 커버리지

관절 키포인트 기반 신체 부위 ROI 추출 모듈은 생성된 부위 ROI가 실제 PPE 영역을 포함하는지를 GT Coverage Recall과 Pseudo Precision으로 평가하였다. 전체 커버리지는 0.311로 나타나 ROI가 PPE 중심점을 포함하는 비율이 제한적임을 확인하였다. 부위별로는 머리-헬멧 0.6722, 상체-안전조끼 0.1195,

발-안전화 0.2606으로 집계되었으며, 특히 상체 ROI의 안전조끼 커버리지 부족이 주요 성능 저하 요인으로 확인되었다. 이 결과는 데이터셋의 클래스 분포와도 관련된다. 안전조끼 객체는 헬멧 및 안전화에 비해 표본 수가 적고 작업자 자세나 가림에 따라 상체 영역이 부분적으로만 관찰되는 사례가 포함될 수 있으므로 ROI가 조금만 축소되어도 GT 중심점을 포함하지 못할 가능성이 커진다.

또한 Pseudo Precision은 머리 0.0534, 상체 0.0049로 매우 낮았는데, 이는 실제 PPE 위치와 무관한 ROI가 다수 생성되어 불필요한 연산 부하가 발생했음을 의미한다. 따라서 ROI 커버리지 결과는 단순히 ROI 생성 알고리즘의 기하학적 정확도뿐 아니라, PPE 객체 크기, 신체 부위별 가림 정도, 클래스별 표본 수 차이에 의해 함께 영향을 받는 것으로 해석할 수 있다. 처리 속도는 $2.28\,img/s$ $(438.07\,ms/img)$로 측정되어, 본 모듈이 전체 파이프라인에서 중요한 병목 구간임을 확인하였다.

표 2. 신체 부위 ROI 추출 성능 분석
Table 2. Analysis of body part ROI extraction performance

PPE	Coverage recall	Pseudo precision
Helmet	0.6722	0.0534
Safety-vest	0.1195	0.0049
Shoes	0.2606	0.2612
Overall	0.3110	-

4.3.3 장면 기반 요구 PPE 추론

VLM 기반 장면 맥락 인지 및 요구 PPE 생성 모듈은 각 PPE의 요구 여부 예측을 다중 라벨 분류로 평가하였다. 실험 결과, 모든 클래스에서 Recall이 상대적으로 높게 나타났으나 Precision이 낮아 과다 예측 경향이 두드러졌다. 구체적으로 헬멧의 F1은 0.5936, 안전조끼는 0.2393, 안전화는 0.1841로 세 항목 모두 Precision이 낮아 전반적으로 낮은 F1을 기록하였다. 세 항목의 요구 여부를 동시에 정확히 맞춘 Exact Match Accuracy(EM-Acc)의 경우 0.7984로 측정되었다. 이러한 결과는 VLM 기반 요구 PPE 추론 모듈이 위험 상황을 놓치지 않기 위해 보수적으로 요구 PPE를 예측하는 경향을 보였음을 의미한다. 즉, Recall이 높다는 점은 실제 요구되는 PPE를 폭넓게 포착한다는 장점으로 해석할 수 있으나 Precision이 낮다는 점은 요구되지 않는 장면에서도 PPE가 필요하다고 판단하는 사례가 많음을 의미한다. 특히 안전조끼와 안전화는 장면 내 교통, 물류 이동, 거친 지면, 자재 적재 등 위험 단서가 부분적으로만 관찰되어도 요구 PPE로 예측될 가능성이 있어 과다 예측이 발생할 수 있다. 따라서 본 모듈의 성능은 Exact Match Accuracy(EM-Acc)만으로 해석하기보다 클래스별 Precision과 Recall을 함께 고려하여 안전성 중심의 보수적 예측과 오탐 증가 사이의 균형으로 해석해야 한다. 추론 속도는 $1.38\,img/s$ $(722.18\,ms/img)$로 측정되었다.

표 3. 장면 기반 요구 PPE 추론 성능
Table 3. Scene-Based Required PPE Inference Performance

PPE	Prec	Recall	F1	EM-Acc
Helmet	0.4473	0.8822	0.5936	-
Safety-vest	0.1374	0.9274	0.2393	-
Shoes	0.1029	0.8701	0.1841	-
Macro avg.	0.2292	0.8932	0.3390	0.7984

4.3.4 착용 판별 및 준수 판정

신체 부위 ROI 기반 PPE 착용 판별 및 준수 판정 모듈은 Worn Coverage Recall(WCR)에서 머리 0.8146, 상체 0.2415, 발 0.4721로 측정되었다. 이는 머리 부위의 헬멧 착용 증거는 비교적 안정적으로 포착되지만 상체와 발 부위에서는 착용 증거 수집 능력이 제한적임을 보여준다. 특히 상체의 낮은 WCR은 앞선 ROI 커버리지 실험에서 확인된 안전조끼 ROI 포함 부족과 직접적으로 연결된다.

전체 Compliance Accuracy는 헬멧 0.9492, 안전조끼 0.9013, 발 0.8427로 나타났으며, 모든 부위에서 비교적 높은 값을 보였다. 그러나 본 연구의 준수 로직에서는 특정 PPE가 요구되지 않는 경우 착용 여부와 무관하게 준수로 처리되므로, 요구되지 않는 음성 사례가 다수 포함될 경우 Compliance Accuracy는 실제 요구 상황에서의 판정 성능보다 높게 나타날 수 있다. 이에 따라 요구되는 상황에 한정한

Required-only Recall을 함께 분석하였다. Required-only Recall은 헬멧 0.5684, 안전조끼 0.1167, 발 0.2616으로 나타났으며, 특히 안전조끼와 발 부위에서 큰 성능 저하가 관찰되었다. 이는 전체 정확도는 높더라도 실제 보호구가 필요한 장면에서 착용 증거를 충분히 포착하지 못하면 준수 판정이 실패할 수 있음을 보여준다. 따라서 본 연구에서는 Compliance Accuracy를 전체적인 시스템 일치율로 해석하되, 실무적 안전 관점에서는 Required-only Recall을 핵심 보완 지표로 함께 고려해야 한다.

표 4. PPE 착용 판별 및 준수 판정 성능
Table 4. Performance of PPE wearing detection and compliance determination

PPE	WCR	Comp-acc	Req-only rec
Helmet	0.8146	0.9492	0.5684
Safety-vest	0.2415	0.9013	0.1167
Feet	0.4721	0.8427	0.2616
E2E	-	0.8127	-

4.3.5 End-to-End 성능 및 병목 분석

전체 파이프라인 평가는 총 7,617장을 대상으로 수행되었으며, 포즈 추정 단계에서 사람을 탐지하지 못한 79장은 ROI 생성 불가로 평가에서 누락되었다. 통합 성능 측면에서 세 부위 준수의 완전 일치 성능은 0.8127로 측정되었다. 처리량은 $2.18 img/s$ (총 $3{,}498.60s$기준)이며, 주요 병목은 관절 키포인트 기반 신체 부위 ROI 추출 모듈의 로케이팅 및 패치 추출 과정에서 발생하였다. 반면 착용 판별 및 준수 판정 모듈의 순수 추론은 $404.61 img/s$로 매우 경량화되어, 전체 지연은 주로 ROI 추출과 데이터 흐름에 의해 결정됨을 확인할 수 있었다.

E2E(End-to-End Exact Match) 0.8127은 세 부위의 준수 여부가 모두 일치해야 정답으로 인정되는 엄격한 지표라는 점에서 의미가 있다. 그러나 앞선 모듈별 결과를 함께 고려하면, 해당 성능은 요구되지 않는 음성 사례의 영향으로 일부 높게 나타날 수 있으며, 실제 요구 PPE가 존재하는 상황에서는 ROI 커버리지와 착용 증거 수집 성능이 더 중요한 병목으로 작용한다. 특히 안전조끼와 발 부위의 낮은 Required-only Recall은 최종 End-to-End 성능을 제한하는 주요 원인으로 해석된다. 종합하면 전체 프레임워크 성능은 첫째, 상체 및 발 ROI 커버리지 부족으로 인한 착용 증거 누락, 둘째, VLM 기반 요구 PPE 추론의 과다 예측 경향, 셋째, 데이터셋 내 요구/비요구 사례 및 클래스 분포 불균형에 의해 영향을 받는다. 따라서 본 논문은 End-to-End Exact Match와 함께 모듈별 지표 및 Required-only Recall을 병행하여 결과를 해석한다.

4.4 성능 개선 실험: 신체 부위 ROI 최적화

앞선 모듈별 성능 분석에서 확인한 바와 같이, 전체 프레임워크 성능을 제한하는 주요 병목은 관절 키포인트 기반 ROI 추출 모듈의 신체 부위 ROI 커버리지 부족이다. 특히 상체 ROI의 GT Coverage Recall이 낮게 나타나 안전조끼 착용 증거를 충분히 포함하지 못하며, 이는 후속 착용 판별 및 준수 판정 단계의 Required-only Recall 저하로 직접 연결된다. 이를 완화하기 위해 본 절에서는 ROI 추출 로직과 주요 파라미터를 조정하였다. 구체적으로 상체 ROI는 엉덩이 관절이 가려지거나 결측인 상황에서도 상체 영역이 과도하게 축소되지 않도록 어깨 너비 기반 확장 계수 s_b를 재조정하고, 상체 패치가 조끼 영역을 충분히 포함하도록 기하학적 마진을 확보하였다. 또한 발 ROI는 원거리 인물 또는 이미지 하단 가림으로 인한 신발 누락을 완화하기 위해 down-shift 비율 δ를 0.06으로 조정하고 높이 상한을 재설정하여 과대·과소 추출을 동시에 억제하였다.

ROI 최적화 후 성능 평가는 표 5의 개선 전·후 비교로 요약된다. 상체 ROI의 GT Coverage Recall은 0.1195에서 0.7083으로 크게 향상되어 상체 구조를 보수적으로 보정하는 ROI 생성 로직이 안전조끼 착용 증거 확보에 직접적으로 기여함을 확인하였다. 발 부위 역시 0.2606에서 0.3821로 개선되었고 전체 커버리지는 0.3110에서 0.4837로 증가하였다. 또한 Pseudo Precision은 모든 부위에서 유의미하게 증가하여 실제 PPE 위치를 포함하는 유효 ROI의 비중이 확대되었으며, 특히 상체 부위는 0.0049에서 0.0725로 개선되어 ROI 과생성 및 비관련 패치 생성 문제가 완화되었음을 시사한다.

이 개선 결과는 데이터 불균형 환경에서 특히 중요하다. 안전조끼처럼 표본 수가 적고 초기 ROI 커버리지가 낮은 클래스에서는 소수의 ROI 누락이 클래스별 성능과 Required-only Recall에 큰 영향을 줄 수 있다. 따라서 상체 ROI 커버리지 향상은 단순한 위치 추출 성능 개선을 넘어 희귀 클래스에 대한 착용 증거 수집 안정성을 높이고 요구 상황에서의 준수 판정 신뢰도를 개선하는 기반으로 해석할 수 있다. 다만 ROI 커버리지 향상이 곧바로 최종 준수 판정의 모든 오류를 제거하는 것은 아니며 VLM 요구 PPE 추론의 과다 예측, 포즈 추정 실패, PPE 종류 매칭 오류가 여전히 후속 오류로 남을 수 있다. 이에 따라 향후 연구에서는 ROI 최적화와 함께 요구 PPE 추론의 정밀도 개선 및 클래스 불균형을 고려한 학습·평가 전략을 함께 고도화할 필요가 있다.

표 5. ROI 최적화 전·후 성능 비교
Table 5. Performance comparison before and after ROI optimization

PPE	Metric	Before	After
Helmet	GT Cov Rec	0.6722	0.6636
	Pseudo Prec	0.0534	0.1604
Safety -vest	GT Cov Rec	0.1195	0.7083
	Pseudo Prec	0.0049	0.0725
Feet	GT Cov Rec	0.2606	0.3821
	Pseudo Prec	0.2612	0.4245
Overall	GT Cov Rec	0.3110	0.4837

V. 결론 및 향후 과제

본 논문에서는 산업 현장의 복잡한 장면 맥락을 해석하여 요구 PPE 목록을 생성하고 작업자의 신체 구조 단서에 정렬된 ROI를 통해 착용 증거를 검증하는 모듈형 프레임워크를 제안하였다. 제안된 시스템은 단순 객체 탐지를 넘어 장면 기반 요구 조건 도출과 부위 단위 착용 증거 검증을 결합함으로써 정책 중심의 준수 판정 체계를 구성한다.

본 연구의 핵심 기여는 다음과 같이 정리할 수 있다. 첫째, PPE 준수 판별 문제를 단순 착용 여부 탐지가 아니라 장면 기반 요구 PPE 추론과 신체 부위 기반 착용 증거 검증의 결합 문제로 정의하였다. 이를 통해 기존 PPE 탐지 중심 접근이 다루기 어려웠던 현장 상황별 요구 조건을 준수 판정 과정에 명시적으로 반영하였다. 둘째, 포즈 추정 기반 신체 부위 ROI 생성 방식을 도입하여 다중 작업자 및 복잡한 배경 환경에서 사람과 PPE 간 귀속 관계를 신체 구조 단서에 기반해 안정화하였다. 특히 머리, 상체, 발 부위에 대응되는 ROI를 구성함으로써 헬멧, 안전조끼, 안전화의 착용 증거를 부위 단위로 검증할 수 있도록 하였다. 셋째, VLM 기반 장면 맥락 이해를 활용하여 이미지 내 위험 단서를 해석하고 상황별 요구 PPE를 생성하는 모듈을 설계하였다. 이를 통해 고정된 체크리스트 방식이 아니라 입력 장면의 작업 환경과 위험 요인에 따라 요구 PPE를 동적으로 추론할 수 있는 가능성을 제시하였다. 넷째, 요구 PPE 추론 결과와 착용 증거 검증 결과를 결합하는 모듈형 준수 판정 파이프라인을 구성하고, 모듈별 성능과 End-to-End 성능을 분리하여 분석함으로써 오류 전파와 병목 요인을 정량적으로 확인하였다.

실험 분석 결과, 전체 프레임워크 성능을 제한하는 핵심 병목은 신체 부위 ROI 추출 모듈의 커버리지 한계로 확인되었다. 특히 초기 설정에서 상체 ROI의 GT Coverage Recall이 0.1195로 낮아 안전조끼 착용 증거 확보가 어렵다는 점이 관찰되었다. 이에 따라 어깨 너비 기반 확장 계수 조정 및 기하학적 마진 확보를 포함한 ROI 최적화를 수행한 결과, 상체 ROI 커버리지는 0.7083으로 크게 개선되었다. 이는 ROI 생성 로직의 정교화가 후속 준수 판정 성능 개선에 직접적인 기반이 됨을 보여준다. 또한 장면 기반 요구 PPE 생성 단계에서는 높은 재현율로 위험 상황을 폭넓게 포착하는 경향이 확인된 반면, 정밀도 저하에 따른 과다 예측 경향이 함께 관찰되어 실제 현장 적용 시 오검출을 줄이기 위한 정밀도 중심의 개선이 필요함을 시사한다. 최종적으로 통합 프레임워크는 End-to-End Exact Match 성능 0.8127을 기록하여 정책 기반 PPE 준수 판정의 자동화 가능성을 정량적으로 확인하였다.

본 연구의 한계와 향후 확장 가능성은 다음과 같다. 첫째, 본 연구는 SH17 기반 재구성 데이터셋을 중심으로 평가되었으며 헬멧, 안전조끼, 안전화

의 세 가지 PPE를 주요 대상으로 설정하였다. 따라서 다양한 산업군과 작업 조건에서의 일반화 성능 검증은 추가로 필요하다. 건설 현장, 제조 현장, 물류 환경, 도로 작업 환경 등은 요구되는 PPE 종류와 위험 단서가 서로 다를 수 있으므로 향후에는 산업 도메인별 정책 기준을 반영한 데이터셋 확장과 교차 도메인 평가를 수행할 필요가 있다. 또한 장갑, 보안경, 마스크, 안전벨트 등 추가 PPE 항목으로 적용 범위를 확장함으로써 보다 다양한 현장 안전관리 시나리오에 대응할 수 있도록 개선할 필요가 있다. 둘째, 포즈 추정 기반 ROI 구성은 복잡한 배경 및 다중 인원 장면에서 신체 부위 중심의 착용 증거 수집을 가능하게 하지만, 원거리 작업자, 부분 가림, 심한 자세 변화가 존재하는 경우 키포인트 품질 저하가 ROI 품질로 전파될 수 있다. 따라서 소형 및 원거리 인물 대응이 강화된 포즈 모델, 키포인트 누락을 보완하는 다중 ROI 후보 생성, 시간적 추적 기반 보정 전략 등을 도입할 필요가 있다. 셋째, VLM 기반 요구 PPE 추론은 위험 상황을 높은 재현율로 포착하는 장점이 있으나, 일부 장면에서는 부분적인 위험 단서만으로도 PPE 요구를 과다 예측하는 경향이 관찰되었다. 향후에는 evidence 기반 필터링 규칙 강화, 불확실성 점수 활용, 산업 환경별 정책 프롬프트 분리, 소규모 도메인 적응 학습 등을 통해 정밀도를 개선할 필요가 있다. 또한 생성된 evidence 태그를 활용해 준수 및 미준수 판정 근거를 제시하는 설명 가능한 AI로 확장하는 연구도 병행되어야 한다. 넷째, 현재 프레임워크는 이미지 단위 분석을 중심으로 평가되었기 때문에 실제 CCTV 기반 실시간 모니터링 환경으로 확장하기 위해서는 영상 프레임 간 작업자 추적, 시간적 일관성 검증, 중복 경고 억제, 파이프라인 병렬화 및 데이터 흐름 최적화가 필요하다. 특히 실시간 적용에서는 딥러닝 추론 속도뿐 아니라 ROI 로케이팅, 패치 추출, 모듈 간 입출력 변환 과정의 지연을 함께 줄이는 시스템 수준의 최적화가 요구된다. 또한 기존 객체 탐지 기반 PPE 판별 방식 및 단순 ROI 기반 방식과의 추가 정량 비교를 통해 제안 프레임워크의 구조적 차별성과 실용적 효과를 보다 폭넓게 검증할 필요가 있다.

종합하면, 본 연구는 단순 착용 탐지를 넘어 안전 정책과 시각적 증거를 논리적으로 결합한 통합 판별 체계를 제시하였으며 장면 맥락 기반 요구 PPE 추론과 포즈 정렬 ROI 기반 착용 증거 검증의 결합 가능성을 실험적으로 확인하였다. 향후 다양한 산업 도메인, 추가 PPE 항목, 영상 기반 실시간 모니터링 환경으로 확장한다면 산업 안전 관리의 자동화 및 지능화를 위한 실용적 기술 기반으로 발전할 수 있을 것으로 기대된다.

References

[1] M. Kim, "Ministry of Employment and Labor field inspections find about 10,000 cases of not wearing protective equipment", CBS NoCut News, Dec. 2021. https://www.nocutnews.co.kr/news/5669352. [accessed: Feb. 23, 2026]

[2] R. Sehsah, A.-H. El-Gilany, and A. M. Ibrahim, "Personal protective equipment (PPE) use and its relation to accidents among construction workers", La Medicina del Lavoro, Vol. 111, No. 4, pp. 285-295, Aug. 2020. https://doi.org/10.23749/mdl.v111i4.9398.

[3] A. J. Al-Bayati, A. T. Rener, M. P. Listello, and M. Mohamed, "PPE non-compliance among construction workers: An assessment of contributing factors utilizing fuzzy theory", Journal of Safety Research, Vol. 85, pp. 242-253, Jun. 2023. https://doi.org/10.1016/j.jsr.2023.02.008.

[4] Ministry of Labor, "Deployment of ‘Construction Safety Watchers’ at Small and Medium Construction Sites", Korea Policy Briefing, Mar. 2010. https://www.korea.kr/news/policyNewsView.do?newsId=148691133. [accessed: Feb. 23, 2026]

[5] Occupational Safety and Health Administration (OSHA), "1926.28 - Personal protective equipment", OSHA, https://www.osha.gov/laws-regs/regulations/standardnumber/1926/1926.28. [accessed: Feb. 23, 2026].

[6] European Union, "Personal protective equipment",

EUR-Lex, Sep. 2016. https://eur-lex.europa.eu/EN/legal-content/summary/personal-protective-equipment.html. [accessed: Feb. 23, 2026].

[7] International Labour Organization (ILO), "Personal protective equipment", International Labour Organization, https://www.ilo.org/topics-and-sectors/occupational-safety-and-health-guide-labour-inspectors-and-other/personal-protective-equipment. [accessed: Feb. 23, 2026]

[8] V. Isailovic, A. Peulic, M. Djapan, M. Savkovic, and A. M. Vukicevic, "The compliance of head-mounted industrial PPE by using deep learning object detectors", Scientific Reports, Vol. 12, No. 1, Art. no. 16347, Sep. 2022. https://doi.org/10.1038/s41598-022-20282-9.

[9] M. I. B. Ahmed, et al., "Personal Protective Equipment Detection: A Deep-Learning-Based Sustainable Approach", Sustainability, Vol. 15, Art. no. 13990, Sep. 2023. https://doi.org/10.3390/su151813990.

[10] S. Malaikrisanachalee, N. Wongwai, and E. Kowcharoen, "ESPCN-YOLO: A High-Accuracy Framework for Personal Protective Equipment Detection Under Low-Light and Small Object Conditions", Buildings, Vol. 15, No. 10, Art. no. 1609, May 2025. https://doi.org/10.3390/buildings15101609.

[11] D. Liu, K. Wang, Q. Dai, J. Liu, and Y. Liu, "Working-at-high operation safety protection recognition based on target detection and spatial relationship", Scientific Reports, Vol. 15, No. 1, Art. no. 35191, Nov. 2025. https://doi.org/10.1038/s41598-025-19048-w.

[12] C. Zheng, W. Wu, C. Chen, T. Yang, S. Zhu, J. Shen, N. Kehtarnavaz, and M. Shah, "Deep Learning-based Human Pose Estimation: A Survey", ACM Computing Surveys, Vol. 56, No. 1, pp. 1-37, Aug. 2023. https://doi.org/10.1145/3603618.

[13] J. Zhang, J. Huang, S. Jin, and S. Lu, "Vision-Language Models for Vision Tasks: A Survey", arXiv preprint arXiv:2304.00685v2, pp. 1-24, Feb. 2024. https://doi.org/10.48550/arXiv.2304.00685.

[14] Q. Fang, H. Li, X. Luo, L. Ding, H. Luo, T. M. Rose, and W. An, "Detecting non-hardhat-use by a deep learning method from far-field surveillance videos", Automation in Construction, Vol. 85, pp. 1-9, Jan. 2018. https://doi.org/10.1016/j.autcon.2017.09.018.

[15] F. Zhafran, E. S. Ningrum, M. N. Tamara, and E. Kusumawati, "Computer Vision System Based for Personal Protective Equipment Detection, by Using Convolutional Neural Network", Proc. 2019 International Electronics Symposium (IES), Surabaya, Indonesia, pp. 516-521, Sep. 2019. https://doi.org/10.1109/ELECSYM.2019.8901664.

[16] J. Wu, N. Cai, W. Chen, H. Wang, and G. Wang, "Automatic detection of hardhats worn by construction personnel: A deep learning approach and benchmark dataset", Automation in Construction, Vol. 106, Art. no. 102894, 2019. https://doi.org/10.1016/j.autcon.2019.102894.

[17] L. López, J. Suárez-Ramírez, M. Alemán-Flores, and N. Monzón, "Automated PPE compliance monitoring in industrial environments using deep learning-based detection and pose estimation", Automation in Construction, Vol. 176, Art. no. 106231, Aug. 2025. https://doi.org/10.1016/j.autcon.2025.106231.

[18] X. Li, M. Hu, B. Li, and R. Tong, "OAM-YOLO: A real-time small object detection framework for PPE compliance monitoring in industrial environments", Process Safety and Environmental Protection, Vol. 204, Art. no. 108058, Dec. 2025. https://doi.org/10.1016/j.psep.2025.108058.

[19] A. M. Vukicevic, M. Djapan, V. Isailovic, D. Milasinovic, M. Savkovic, and P. Milosevic, "Generic compliance of industrial PPE by using

deep learning techniques", Safety Science, Vol. 148, Art. no. 105646, Apr. 2022. https://doi.org/10.1016/j.ssci.2021.105646.

[20] R. Xiong and P. Tang, "Pose guided anchoring for detecting proper use of personal protective equipment", Automation in Construction, Vol. 130, Art. no. 103828, Oct. 2021. https://doi.org/10.1016/j.autcon.2021.103828.

[21] D. Maji, S. Nagori, M. Mathew, and D. Poddar, "YOLO-Pose: Enhancing YOLO for Multi Person Pose Estimation Using Object Keypoint Similarity Loss", arXiv preprint arXiv:2204.06806, pp. 1-10, Apr. 2022. https://doi.org/10.48550/arXiv.2204.06806.

[22] S. Bai et al., "Qwen2.5-VL Technical Report", arXiv preprint arXiv:2502.13923, pp. 1-23, Feb. 2025. https://doi.org/10.48550/arXiv.2502.13923.

[23] H. M. Ahmad and A. Rahimi, "SH17: A dataset for human safety and personal protective equipment detection in manufacturing industry", Journal of Safety Science and Resilience, Vol. 6, No. 2, pp. 175-185, Jun. 2025. https://doi.org/10.1016/j.jnlssr.2024.09.002.

저자소개

박 수 진 (Su-Jin Park)

2022년 3월 : 경상국립대학교 경영정보학과(경영학사)
2026년 3월 ~ 현재 : 경상국립대학교 컴퓨터공학과 석사과정
관심분야 : 인종지능, 멀티모달, Meme Detection, Deepfake

이 규 혁 (Gyu-Hyeok Lee)

2020년 3월 : 경상국립대학교 컴퓨터공학과(공학사)
관심분야 : 인종지능, 데이터 분석, 자연어 처리

이 세 연 (Se-Yeon Lee)

2023년 3월 ~ 현재 : 경상국대학교 컴퓨터공학과 학사과정
관심분야 : 인공지능, 컴퓨터비전, 이미지 생성, 대규모 언어모델

오 동 현 (Dong-Hyeon Oh)

2011년 2월 : 경남과학기술대학교 컴퓨터공학과(공학사)
2013년 2월 : 경남과학기술대학교 창업학(창업학석사)
2015년 2월 : 경남과학기술대학교 컴퓨터메카트로닉스융합 (박사수료)
2011년 9월 ~ 현재 : ㈜유니드 사내 기술이사 및 기업부설연구소 연구전담부서 연구소장
관심분야 : IoT, 인공지능, 시스템아키텍쳐

김 건 우 (Gun-Woo Kim)

2006년 12월 : 호주뉴캐슬대학교 컴퓨터공학과(공학사)
2007년 9월 : 호주뉴캐슬대학교 컴퓨공학과(공학석사)
2017년 8월 : 한양대학교 컴퓨터공학과(공학박사)
2021년 9월 ~ 현재 : 경상국립대학교 컴퓨터공학과 부교수
관심분야 : 인공지능, 시멘틱 헬스케어, 데이터마이닝