FPT UNIVERSITY

Automated Detection of Personal Protective Equipment for Construction Safety

Technical Report - DPL302m (2025)

Full Name	Student ID
Le Nguyen Gia Hung	SE194127
Huynh Quoc Viet	SE194225
Vo Tan Phat	SE194484
Ngo Hoai Nam	SE194190

Lecturer: Mr. Ho Le Minh Toan

Ho Chi Minh city, September 27, 2025

Contents

1 Project Overview **2**
1.1 Description . . . 2
1.2 Expected Outcomes . . . 2

2 Dataset Report **3**
2.1 Introduction . . . 3
2.2 Dataset Sources . . . 3
2.3 Data Visualization . . . 3
2.4 Dataset Description . . . 4
2.5 Data Annotation and Processing . . . 4
2.6 Dataset Statistics . . . 5
2.7 Challenges and Limitations . . . 5
2.8 Integration Plan . . . 5

3 Methodology **6**
3.1 Overall Framework . . . 6
3.2 Object Detection (YOLOv8) . . . 6
3.3 Pose Estimation . . . 7
3.4 Compliance Evaluation . . . 7
3.5 Evaluation Protocol . . . 7
3.6 Extensibility . . . 7

1 | Project Overview

1.1 | Description

Ensuring that workers consistently wear Personal Protective Equipment (PPE) such as helmets, reflective vests, gloves, and safety boots is essential for preventing accidents on construction sites. However, manual inspection is time-consuming and prone to human error, while sensor-based approaches are often costly and intrusive.

This project aims to develop an AI-powered computer vision system capable of automatically detecting whether construction workers are fully compliant with PPE requirements. By leveraging modern deep learning models—particularly the YOLO family of real-time object detectors—the system can recognize multiple PPE items under diverse site conditions, including challenging lighting, occlusion, and crowded scenes.

The project will make use of both open-source PPE datasets (e.g., CHV, Roboflow PPE, Construction-PPE) and custom-collected site-specific data to improve robustness and generalization. Model performance will be evaluated using standard object detection metrics such as mAP, IoU, precision, and recall, with an emphasis on balancing detection accuracy and real-time inference speed.

1.2 | Expected Outcomes

- A lightweight yet accurate detection model for helmets, vests, gloves, and boots.
- A dataset benchmark tailored for PPE compliance in construction sites.
- A prototype monitoring system capable of issuing real-time alerts and generating safety reports.

By automating PPE compliance monitoring, the proposed solution will help reduce accident risks, improve inspection efficiency, and support the creation of safer construction environments.

2 | Dataset Report

2.1 | Introduction

Personal Protective Equipment (PPE)—such as helmets, vests, gloves, and boots—is critical for ensuring worker safety on construction and industrial sites. This project leverages multiple open-source datasets to develop a robust object detection model for monitoring PPE compliance.
The datasets vary in scale, annotation format, and class definitions. Consequently, they require significant preprocessing, including cleaning, and unification into a common YOLOv8 format. This report provides a comprehensive summary of all datasets used, their key characteristics, and the preprocessing methodologies applied.

2.2 | Dataset Sources

The following table summarizes the datasets integrated into this project. The links in the table are clickable for easy access.

Table 2.1: Summary of Datasets Used in This Project

Dataset	Link	Format	Size	Images	Notes
CHVG Dataset	Link	TXT	429MB	1,699	8 PPE classes (helmet colors, vest, glasses, body, head)
Harvard Dataverse PPE	Link	XML	262MB	7,063	Multi-class PPE annotations
SoDaConstruction	Link	TXT	163MB	1,559	Provided in YOLO format
Deteksi APD	Link	TXT	112MB	3,958	PPE dataset, Indonesian context
Mendeley PPE	Link	TXT	229MB	2,286	Provided with YOLO labels
Hard Hat Detection	Link	XML	1.28GB	5,000	Helmet-focused dataset
PPE Kit Detection	Link	TXT	174MB	1,416	Multiple PPE kit items
SH17 PPE Dataset	Link	TXT	13GB	8,095	75,994 bbox, 17 PPE classes
HuggingFace PPE	Link	COCO	2.1GB	11,978	COCO-format annotations
Total	–	–	**19.9GB**	**43,054**	9 datasets

2.3 | Data Visualization

To better understand the dataset composition, sample images were reviewed. Figure 2.1 shows an example image from one of the datasets, illustrating typical scenarios and object classes.

Figure 2.1: A samples image from the aggregated dataset.

2.4 | Dataset Description

- **Image Count:** Approximately 43,000 images sourced from nine distinct datasets.
- **Annotation Formats:** A mix of YOLO TXT, Pascal VOC XML, and COCO JSON, all of which will be unified into the YOLOv8 format.
- **Original Classes:** The number of categories ranges from 1 to 17 PPE-related classes.
- **Planned Classes (Post-Remapping):** Helmet, Vest, Gloves, and Boots.

2.5 | Data Annotation and Processing

- **Annotation Tools:** LabelImg, CVAT, and Roboflow.
- **Format Conversion:** Scripts to convert XML and COCO annotations to YOLO TXT.
- **Data Cleaning:** Procedures to remove corrupted/duplicate images and unify the labeling schema.
- **Class Unification:** Remapping of dataset-specific classes into the four standardized categories.
- **Augmentation (Planned):** Application of flip, rotation, brightness/contrast shifts, and mosaic augmentation techniques.

2.6 | Dataset Statistics

Table 2.2: Planned Unified Class Distribution (Estimated)

Class	Image Count	Bounding Box Count
Helmet	35,210	81,550
Vest	24,880	54,120
Gloves	6,540	8,980
Boots	7,150	9,630

2.7 | Challenges and Limitations

- **Format Inconsistency:** Significant variations in annotation formats across the datasets.
- **Class Imbalance:** A high prevalence of helmets compared to a scarcity of gloves and boots.
- **Limited Diversity:** A lack of images captured in adverse conditions such as night, rain, or with occlusions.
- **Potential Overlap:** The possibility of duplicate images existing across different datasets.

2.8 | Integration Plan

- **Primary Datasets:** The SH17 and HuggingFace PPE datasets will serve as the foundation due to their large size and multi-class nature.
- **Supporting Datasets:** CHVG, Harvard, Mendeley, SoDaConstruction, Deteksi APD, and PPE Kit Detection will be used to supplement the primary sources.
- **Specialized Dataset:** The Hard Hat Detection dataset will be leveraged specifically for helmet detection.
- All datasets will be merged, consistently annotated, and normalized into the YOLOv8 format for model training.

3 | Methodology

We propose a modular pipeline for PPE (Personal Protective Equipment) compliance monitoring that combines high-performance object detection (**YOLOv8**) with **human pose estimation**. This design allows us to (i) detect required PPE instances (e.g., helmet, vest, gloves, boots), and (ii) assess whether they are worn *correctly* using pose-aware rules. The approach is intentionally general and extensible so that new PPE classes, rules, and deployment targets (e.g., edge devices) can be incorporated later with minimal refactoring.

3.1 | Overall Framework

The end-to-end system consists of four stages: *data preprocessing*, *object detection*, *pose estimation*, and *compliance evaluation*. Inputs may be images or video frames captured in diverse construction/industrial environments; outputs include per-worker compliance status and violation tags suitable for logging, alerting, and analytics.

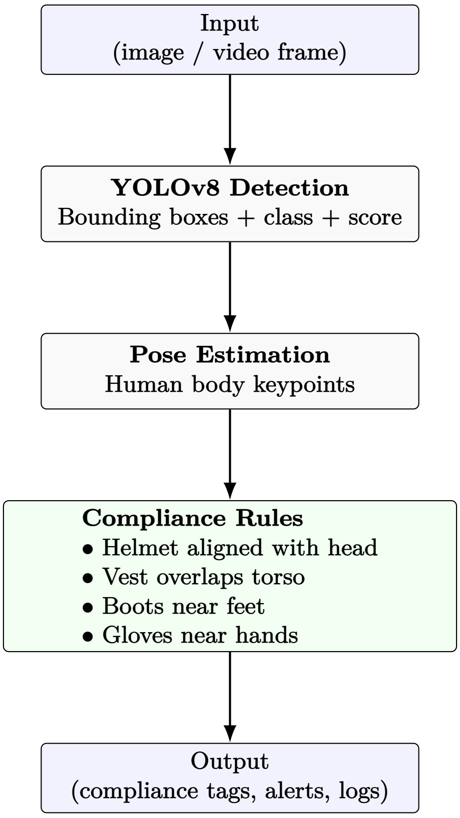

Figure 3.1: Pipeline overview: YOLOv8 detection + pose estimation + rule-based compliance checking.

3.2 | Object Detection (YOLOv8)

We adopt **YOLOv8** for object detection thanks to its accuracy–speed trade-off and mature open-source tooling. All datasets are unified to YOLO format to streamline training and evaluation.

- **Input & Preprocessing:** images/frames are resized and normalized per YOLOv8 requirements. Data augmentation such as flips, rotations, brightness/contrast adjustments, and mosaic is applied to improve robustness.
- **Training:** the unified dataset is used to train the model, ensuring consistency across multiple sources.
- **Output:** per-frame detections b_i, c_i, s_i with bounding boxes b_i, class $c_i \in \{\text{helmet, vest, gloves, boots}\}$, and confidence s_i.

3.3 | Pose Estimation

While detection provides information on PPE presence, pose estimation adds semantic context for correct usage.

- **Keypoints:** head, torso, shoulders, elbows, wrists, hips, knees, ankles.
- **Association:** PPE detections are aligned with corresponding body regions (e.g., helmets with head, boots with feet).
- **Robustness:** occlusion and motion blur can be mitigated through temporal smoothing and contextual constraints.

3.4 | Compliance Evaluation

Detection and pose outputs are combined using simple rule-based logic:

- **Helmet:** must overlap the head region.
- **Vest:** must cover torso region.
- **Gloves/Boots:** must appear close to wrist/ankle keypoints.
- **Outputs:** per-worker compliance labels (compliant / missing / improper).

3.5 | Evaluation Protocol

Two evaluation layers are considered:

- **Detection metrics:** mAP@0.5, mAP@0.5:0.95, precision, recall, and runtime performance.
- **Compliance metrics:** per-class F1, overall compliance accuracy, and violation detection rate.

3.6 | Extensibility

The pipeline is modular and can be expanded:

- **New PPE classes:** goggles, masks, or ear protection can be integrated by adding annotations and retraining.
- **Temporal analysis:** video-based tracking can ensure continuous compliance monitoring.
- **Deployment:** lightweight versions of the pipeline can be optimized for real-time use on edge devices.