# FieldWatch A Vision for Field Worker Safety

1st Venkat Amith Woonna
*Computer Science and Engineering*
*Vellore Institute of Technology*
Chennai, India
https://orcid.org/0000-0001-8970-8977

2nd Rahul Sandireddy
*Computer Science and Engineering*
*Vellore Institute of Technology*
Chennai, India
https://orcid.org/0009-0002-0679-4000

3rd Anand R
*Computer Science and Engineering*
*Vellore Institute of Technology*
Chennai, India
https://orcid.org/XXXX-XXXX-XXXX-XXXX

4rd Premanand V
*Vellore Institute of Technology*
Chennai, India
https://orcid.org/0000-0001-6710-8970

***Abstract*—This paper presents a proposition of the implementation of a field worker safety detection system.The system employs cameras to oversee the utilization of workers,identify potential safety risks and evaluate the need for personal protective equipment (PPE).In real-time, the system actively prevents accidents, it's designed to mitigate hazards.Productivity enhancement is imperative, and safety measures should be significantly increased.The cameras capture images and videos to document culture, while also providing data insights.The captured data,pertains to workers.A computer vision system analyzes the data from cameras.The system employs an advanced algorithm , its purpose is to swiftly detect any potential safety violations ,a task it performs with remarkable accuracy.The system harbors substantial potential for significant enhancement, regardless of its current rigorous or harsh nature.Ensuring workplace safety: this proactive measure aids in the prevention of accidents is a critical aspect for any organization's success.Workers wearing the appropriate PPE: this measure not only assists in ensuring their safety, but it also aids in a significant way ,the overall productivity of an organization.You must actively monitor compliance with safety regulations.By reducing downtime, the system can enhance productivity.The system contributes to absenteeism because of accidents and injuries.In its use of deep learning to detect PPE, it also demonstrates innovation.The system employs a deep-learning model for object recognition in images.The dataset of images has trained a learning model.The system actively monitors individuals, distinguishing between those who wear PPE and those who do not.Even in challenging conditions, it accurately detects PPE usage.The system underwent evaluation in conditions of low light and poor visibility.The dataset contains images of individuals, some wearing personal protective equipment (PPE), and others not; these images can be utilized for various research or analytical purposes.The system attains a 95% accuracy on this dataset; such performance implies ,the system possesses the capability to detect PPE usage accurately.The system harbors potential for utilization in diverse settings. It possesses versatility ,a trait that renders its applicability not limited to a singular domain or circumstance.**

***Index Terms*—Cameras , Computer vision , Deep learning , PPE , mitigate hazard, worker's safety**

## I. Introduction

### A. Domain of the Problem

Businesses, governments, and individuals prioritize safety in our rapidly evolving world. Safety equipment detection systems particularly underpin workplace safety in high-risk environments such as construction sites and industrial facilities. Strategically positioned cameras track workers' adherence to "Personal Protective Equipment (PPE) " usage including hard hats, safety boots and high-visibility vests. To address privacy concerns, we design cameras that only capture pertinent information for PPE detection and identification of safety hazards. The software provides real-time analysis that alerts safety personnel to potential safety violations. Traditional safety checks, however, frequently do not detect these issues promptly; this shortfall can lead to severe accidents. Consequently, there is a demand for systems notably more efficient ones capable of swiftly identifying PPE violations and other hazards inherent in workplace environments. Developing a system that accurately detects PPE use and safety hazards in real-time is the problem we aim to tackle; this will enhance workplace safety significantly, thereby reducing accidents effectively. To delve deeper into Safety Equipment Detection Systems—comprehending these areas holds paramount importance:

1.Accurate detection requires a vital understanding of PPE, such as hard hats, vests, and more.

2.Awareness of camera types and capabilities in sensors and cameras is crucial for efficient monitoring.

3.Understanding Artificial Intelligence assists in the creation of precise detection algorithms.

4.Familiarity with workplace safety regulations guarantees compliance to the safety regulations.

5. Knowing the industry-specific hazards contributes significantly to precise detection and identification of potential dangers.

### B. Motivation

Any business must fundamentally prioritize employee health and safety this commitment fosters trust between employers and their workforce. With our project's aim to streamline safety measures, we allow access to the pitch exclusively for individuals equipped with required Personal Protective Equipment (PPE). In doing so, we guarantee compliance with rigorous safety regulations while concurrently mitigating insurance costs linked to non-compliance-related injuries.

### C. Problem Statement

Our aim is to develop a sturdy system that enables efficient, on-the-spot monitoring of employee well-being and security while on duty. This system will incorporate features such as location tracking, activity monitoring, and rapid emergency notifications. It should also be user-friendly, resilient, and effective in rugged or remote settings.

### D. Objectives

Several benefits are offered by the safety detection system:

a. Real-time alerts in Accident Prevention guarantee that workers wear the necessary safety gear, thus curbing potential accidents.

b. Compliance Monitoring actively monitors and ensures the safety protocols.

c. Improving Productivity: Actively reducing downtime, absenteeism and accidents this enhances productivity.

d. Cost Savings: It reduces the costs of worker compensation claims and missed productivity.

e. Safety Culture: It fosters a sense of responsibility for safety and promotes safety enhancements based on data.

f. Data Insights: The department actively gathers data on safety incidents, near misses, and trends; this information informs the formulation of robust safety policies.

g. Identifying patterns and trends in data: this is the foundation for making informed decisions particularly those that elevate safety procedures and equipment.

h. Targeted Actions: Data insights prompt action, such as equipment replacement or further training.

### E. Outlines

Highly effective for image detection tasks, deep learning utilizes deep neural networks to locate and categorize specific items or features within images. The commonly employed Convolutional Neural Networks (CNNs) comprise layers such as convolutional, pooling, and fully connected layers. Image features are extracted by convolutional layers, these features undergo down sampling through pooling layers. Finally, fully connected layers categorize the down sampled image into various classes. In CNN training, we utilize a substantial dataset of labeled images to calibrate network weights, this minimizes the discrepancy between predicted and actual labels via backpropagation. After undergoing training, we can apply these CNNs to fresh images for tasks such as face recognition, object identification even medical image analysis.

## II. Literature Survey

### A. Earlier Proposed Methods and Techniques

In research paper [1] , the authors propose an IoT-based system that uses IoT devices to collect data on air quality pollutants, such as carbon monoxide and methane, in the mine, this data is then sent to the Azure machine learning platform for analysis, where the authors have implemented a machine learning model that can predict the levels of pollutants in real-time. The authors proposed a framework for data preprocessing, feature extraction, and model selection to improve the performance of the prediction model. The system uses IoT devices to collect data on air quality pollutants, such as carbon monoxide and methane, in the mine. This data is then sent for analysis and this has a 16.9% improvement from regression model In [2] by Fangbo Zhou , et al. , the authors propose a system for detecting safety helmets in images and videos of construction and industrial workers to improve safety. They used the YOLOv5 (You Only Look Once version 5) object detection algorithm to detect safety helmets in the images and videos. The authors argue that this is an important task, as it can help to identify workers who are not wearing safety helmets, which can help to prevent accidents and injuries in these environments.object detection algorithms, such as Faster R-CNN and RetinaNet, and found that YOLOv5 had a better performance in terms of speed and accuracy.It is to be noted that an improvement of 9.75%, on average, is observed across the various models In [3] by Kang Li, et al. . presents a system for automatically detecting the workers safety helmets in construction and industrial environments. The system is intended to improve safety by identifying workers who are not wearing helmets and alerting them or a supervisor. the ViBe background modeling algorithm is employed. Uses of C4 framework significantly increases the object detection The results of the study [4] by Arjya Das Mohammad, et al. show that their proposed model was able to accurately detect the presence or absence of a face mask with high precision. The model tested on a variety of real-world scenarios, including different lighting conditions and diverse facial expressions, and it performed well in all cases. In conclusion, the authors propose that their deep learning model can be useful in various settings to enforce the use of face masks and prevent the wide-spread of COVID-19. The use of computer vision technology in combination with deep learning offers a promising solution for real-time face mask detection in public spaces.Uses a dataset of images of people wearing and not wearing face masks to train their deep learning model. The exact details of the dataset, such as the number of images, their sources, and the diversity of the images, are not specified in the paper. However, it is stated that the dataset was created specifically for the purpose of face mask detection and it was used to train the deep learning model developed in the study. This offers an accuracy up to 95.77[5] by Venkata Santosh Kumar et al. ,presents a system that uses computer vision to detect whether construction workers are wearing PPE such as safety helmets, gloves, and high-visibility vests. The system is intended to improve safety by identifying workers who are not wearing the required PPE and alerting them or a supervisor. The study makes use of the Convolutional Neural Networks model, which was developed by applying transfer learning to a base version of the YOLOv3 deep learning network Nath, et al. of [6] propose using machine learning techniques to analyze images and videos of construction workers to detect whether they are wearing PPE. They present a system that uses a combination of a deep-neural networks and computer vision techniques to detect PPE in images and videos. In [7] by Yogesh Kawade, et al.

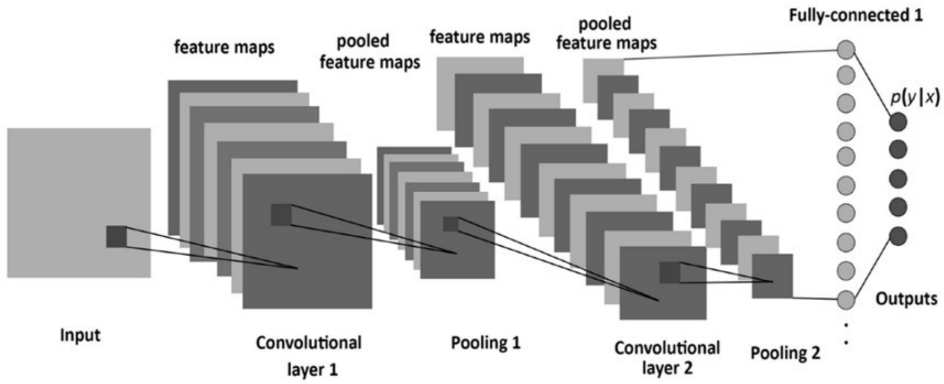


Fig. 1. CNN Architecture

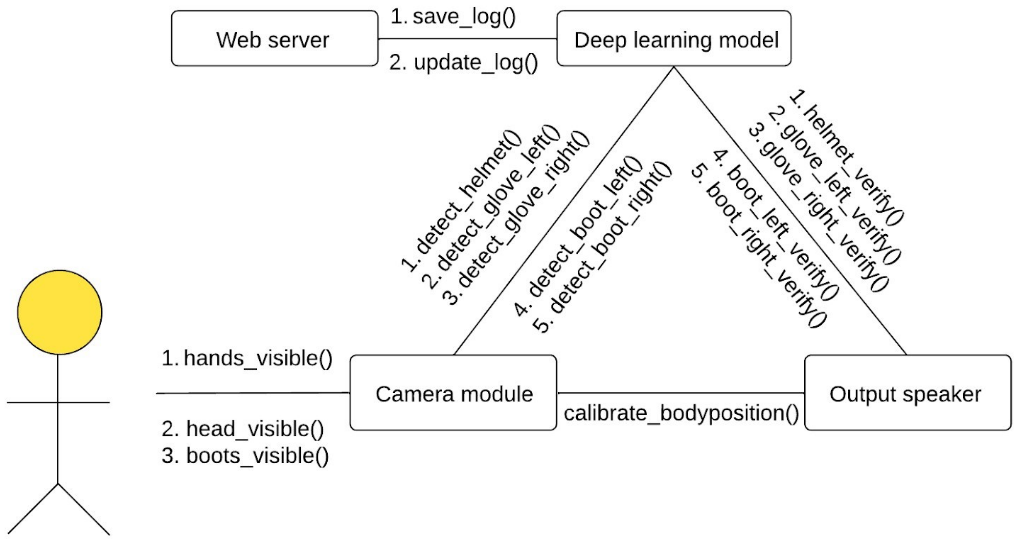


Fig. 2. Architecture

The authors propose a system that uses computer vision and deep learning techniques to automatically detect and classify workers who are wearing or not wearing required safety equipment, such as hard hats, safety vests, and safety glasses. The authors used convolutional neural networks (CNNs) to train the system on a large-dataset of images of construction workers wearing and not wearing safety equipment. The results of the study show that the proposed system accurately classifies and detectes the presence or absence of required safety equipment with high precision. The system was tested on real-world construction sites and it performed well under various conditions, including different lighting conditions and diverse facial expressions. This paper uses a YOLOv3 model to apprehend hard hats and PPE vests. Ultimately identification of proper PPE using geometric relationships of the outputs of OpenPose and YOLOv3 were used with the following accuracies: Hard hat - Precision : 94.19 Recall : 90.78 PPE vest - Precision : 99.59 Recall : 99.28 Jonathan Karlsson, et al. [8] presents a method for detecting the use of personal protective equipment (PPE) and safety gear on industrial workers. The authors use computer vision and deep learning techniques to automatically detect and classify workers who are wearing or not wearing required PPE and safety gear, such as hard hats, safety vests, and safety glasses. The system was trained on a large data-set of images of industrial workers wearing and not wearing PPE and safety gear. The results of the study show that the proposed system was able to accurately detect the safety gear with high precision. The system was tested in real-world industrial settings and it performed well under various conditions, including different lighting conditions and diverse facial expressions. Data-set obtained from kaggle of hardhats and safety vests with positive and negative samples. YOLOv4 model is used for object detection with the following accuracy: Precision : 3m - 7m 98 100Recall : 3m - 7m 93 100In [9] by Gioatan Gallo, et al. , the authors use computer vision to automatically recognize workers with or without their required PPE such as helmets, safety vests and goggles. A convolutional neural network (CNN) was trained using a large data set of images of industrial workers with and without PPE. The results of the study indicates that the system was able to accurately detect the presence or absence of her required PPE with high accuracy. The system has been tested in a real industrial environment and performed well under a variety of conditions, including different lighting conditions and different facial expressions. Finally, the authors suggest that their PPE detection system may significantly improve safety in industrial environments. The use of computer vision and deep learning technology offers a promising solution for real-time monitoring and enforcement of his PPE use in the industry. The research paper by ND Nath, et al. [10] presents a deep learning-based system for detecting personal protective equipment (PPE) in order to maintain safety compliance on construction sites. The system is designed to improve safety compliance by using computer vision techniques to automatically detect if workers are wearing the correct PPE. The paper describes the development and discusses the results, including its accuracy and potential applications in real-world construction site environments. Checking sensor signals for cases of PPE non-compliance deep learning (DL) have created opportunities for using convolutional neural network (CNN) algorithms to more precisely detect PPE components. [11] by Xiao Ke, et al., presents a deep learning-based solution for detecting PPE on workers in real-time using green edge computing. The system is capable of detecting PPE with over 100 frames per second (FPS) and aims to improve worker safety in industries.channel pruning algorithm based on the BN layer scaling factor to further reduce the size of the detection model Reduction of computational effort by 32% and detection by 25% . Pedro Torres, et al. [12] presents a robust and real-time component for detecting Personal Protective Equipment (PPE) in an industrial setting. The author has made use of the YOLO-v3 for hard hat detection which gave an increase of 21.52% and 13.96% when comparing the YOLO-v4-AP1 and YOLO-v4-AP2 . Saudi, et al. , in [13] presented a model for evaluating the safety conditions of construction workers using Faster R-CNN. The model is used to detect and analyze images of construction workers on a construction site to assess if they are following safe practices and wearing appropriate personal protective equipment. ResNet152 and Faster RCNN, Deep CNN Google Inception v3 offers accuracy of 70%. [14] by Moohialdin, at al., proposes a novel computer vision (CV) system to detect the construction workers' PPE and postures in a real-time manner. Python data-labeling tool was used to annotate the selected datasets and the labeled datasets were used to build a detection model based on the TensorFlow environment and higher accuracy in classifying the postures over 72% and 64% in model testing and validation has been detected. [15] proposes computer vision-based automatic solution for detection of workers PPE. The presented work uses the SSD-MobileNet algorithm that is based on convolutional neural networks.The precision of the trained model is 95% and the recall is 77% In [16], Ferdous, et al., using open-computer vision based automatic PPE clasification and detection system that has been implemented to detect different types of PPE. (YOLO) family's anchor-free architecture, YOLOX. YOLOX-m performs approximately 3.29% better [17] offers a way for detection in real time if workers are wearing PPE or not. The detector is developed using the YOLOv4 computer vision model which specially performs well in real time object detection. mAP of the object detector is found to be 79% In "Applying the Haar-cascade Algorithm for Detecting Safety Equipment in Safety Management Systems for Multiple Working Environments" [18], Phuc, et al. ,apply the algorithm to calculate a score to determine the intensity of danger of the current working environment based on the safety equipment and working environment. With this data, the system decides whether it is necessary to give a warning signal. principal component analysis Haar cascade algorithm were the algorithms used in this paper. CARs for this varied from 66.8% to 68.2% . Qiu, et al., [19] proposes a method for detecting and evaluating the effectiveness of safety protection equipment in order to ensure its proper functioning and to prevent accidents. The method involves using sensors and algorithms to detect potential safety hazards and evaluate the performance of the protective equipment in response to those hazards. The goal is to improve safety by ensuring that protective equipment is functioning correctly and providing the necessary protection. The target detection algorithm used in this is based on linear distance and angle constraints A deep learning based solution for construction equipment detection: from development to deployment [20] presents a deep learning based solution for detecting construction equipment in real-world environments. The solution is developed and tested, then deployed to a construction site to demonstrate its ability to detect different types of construction equipment. The deep learning model uses computer vision techniques to identify and classify equipment in images and videos, with the goal of improving safety and efficiency in the construction industry. The results show the effectiveness of the proposed solution and its potential for practical deployment. The deep learning algorithms used above yield a 90% rate of accuracy Marks, et al. in [21] presented a testing method the effectiveness of proximity-detection and alert technology used in construction equipment operations. The goal of the technology is to improve safety by detecting potential collisions and alerting operators to take corrective action. The method involves conducting tests in real-world environments to evaluate the performance of the technology in detecting potential hazards and providing effective warnings to operators. The results of the tests are used to identify any areas for improvement to ensure safe and efficient operation of construction equipment emerging radio frequency (RF) remote sensing technology . An accuracy of 89% has been obtained by the authors. In [22], proposed a Convolutional Neural Network (CNN) approach for automating the detection of workers . The CNN model is trained on construction site images and videos to identify workers and heavy equipment and their location. The goal is to improve safety and efficiency on construction sites by providing real-time information about the presence and location of workers and equipment. Wang, et al., in [23] present a method for predicting safety hazards among construction workers and equipment using computer vision and deep learning techniques. The method involves collecting data from construction sites and using it to train a deep-learning model to recognize and predict potential safety hazards. The goal is to improve safety on construction sites by providing real-time warnings and recommendations for avoiding hazards. The approach is effective for predicting safety hazards and has the potential for practical application on construction sites. Region-based CNN framework named Faster R-CNN is used to detect workers standing on scaffolds. In [24] by Zhao, et al., proposes an intelligent model for detecting personal protective equipment (PPE) in substation environments using a Graph Neural Network (GNN) approach. The goal is to improve safety awareness and compliance with PPE requirements by detecting workers who are not wearing required protective gear. The GNN model is trained on substation images and videos to identify workers and their PPE, with the goal of providing real-time warnings and increasing awareness of PPE requirements. The GNN-based model is effective for PPE detection and has potential for practical application in substation environments. The author utilizes a few-shot based GNN technique to detect PPE.

### *B. Data-set description*

Deep learning algorithms, such as CNNs, rely heavily on datasets, and the quantity and quality of these datasets have a big impact on model performance. Algorithm training, assessment, and comparison all depend on datasets. In order to produce a varied and reliable training dataset for a CNN, several datasets comprising photos of safety gear were selected and integrated. Images were scaled in order to maximize processing speed, then each image's objects of interest were labeled and bounding boxes were drawn. In order to educate the neural network to detect things, annotation is necessary. The process for preparing the dataset was described. This format reduces storage space requirements and data processing time, especially for large datasets. By utilizing these tools and techniques, a high-quality annotated dataset was prepared for training, validation, and testing purposes, enhancing CNN's ability to detect safety gear accurately.

This study presents a combined dataset of 844 images for training, 146 images for validation, and 103 images for testing. The dataset was created by combining two Roboflow datasets and resizing all images to 640x640 pixels. The dataset contains 26 labels, including 'Excavator', 'Gloves', 'Hardhat', 'Ladder', 'Mask', 'NO-Hardhat', 'NO-Mask', 'NO-Safety Vest', 'Person', 'SUV', 'Safety Cone', 'Safety Vest', 'bus', 'dump truck', 'fire hydrant', 'machinery', 'mini-van', 'sedan', 'semi', 'trailer', 'truck and trailer', 'truck', 'van', 'vehicle', 'wheel loader', and 'No Gloves'. The dataset can be used to train a variety of object detection and image classification models, study the effectiveness of different training and inference techniques, and develop new methods for improving the performance of object detection and image classification models.

### *C. Inference*

The primary objective of this project is to increase safety measures on construction sites by strictly enforcing safety protocols and mandating the use of protective gear among employees. Through real-time monitoring, particularly on large-scale sites, the project aims to significantly improve compliance with these safety measures. By doing so, it can effectively reduce the exorbitant costs associated with workplace accidents, which currently account for a staggering 4% of India's annual GDP loss. These costs include expenses for treating injuries, medical care, and compensation payments. Moreover, this initiative also works towards fostering a safety-conscious culture within the construction industry, highlighting employers' dedication to their workers' well-being and promoting the importance of taking precautionary measures. With careful planning and adequate funding, this project has the potential to greatly benefit India's construction sector and promote a safer working environment.

## III. Proposed Method

### *A. Analysis of the Problem Statement*

Goal: To promote worker safety by implementing a system that monitors and detects proper usage of safety gear in various work environments.

Overview: The use of safety gear, such as helmets, vests, masks, and boots, is crucial for ensuring workplace safety. However, challenges such as sensor accuracy, worker acceptance, and cost barriers can hinder the effectiveness of safety gear detection.

Strategy: In order to address these challenges, several solutions have been identified, integrating with existing safety systems, creating cost-effective solutions and enhancing worker engagement and training.

Feasibility: A comprehensive feasibility study will be conducted to assess the integration of the system into existing safety measures, its impact on workflows, and its long-term management.

Development: The development process will involve collecting relevant data, designing the system, developing a prototype, and conducting thorough testing to ensure accuracy and effectiveness.

Solution: The proposed solution is to implement a system that utilizes complex models with best accuracy and provide real time notifications to the admin to detect and monitor the proper usage.

### *B. Methodology*

*1) Brief of the Proposed Methods:* Implement an application using React with computer vision to enhance cost-effectiveness for safety equipment detection in various fields.

Utilize YOLOv5 for object detection, implemented using Keras with a Pytorch with Tensorflow backend. YOLOv5 divides images into a grid, predicts boundary boxes, confidence scores, and class probabilities for real-time, single-object detection.

Our YOLO Model implements a whole new approach for producing the anchor boxes, which are referred to as ”dynamic anchor boxes.” It includes employing a clustering technique to group the ground truth bounding boxes into clusters and then utilizing the centroids of the clusters as the anchor boxes. This may be accomplished by clicking here. This makes it possible for the anchor boxes to have a size and form that is more closely matched with the identified items.

In addition, Our Model incorporates the Spatial Pyramid Pooling idea. SPP is a form of pooling layer that is used to lower the spatial resolution of the feature maps. Since it enables the model to see objects at a variety of sizes, SPP is used in order to enhance the detection performance on tiny items and also incorporates significant enhancements to the SPP design, which enables it to do more and provide better results.

Camera module, upon a person entering the frame, initiates detection of that individual using the mediapipe API pipeline for human pose detection. Subsequently when all body points remain within view:one snapshot is captured – sending it to YOLOv5 model API endpoint; this image undergoes analysis and returns confirmation if any safety equipment appears missing. Upon verification of both person and all necessary safety gear—the speaker node will indicate this accordingly. If any failures in calibrating person or safety equipment detection occur, the speaker node shall promptly notify the admin.

Modules Used:

I.Image Capturing Module: Captures employee’s images from construction and sends them to the computing device for processing and object detection.

II. Labelling Images: In object detection first, we need to train the model using some labelled images and so this is an important module. Images of workers will be labelled. More the number of images in the training dataset, the greater the accuracy.

III.Object Detection: This module is responsible for object detection from the images, such as workers, helmets, masks, vests, boots.

IV.Searching: A string will be returned when an image is given an input to find out what equipment is missing, we will search the string and add them to lists. The main functionality of this module is to find out missing equipment using string search algorithm.

V.Text to Speech: The functionality of this module is to create a string variable that holds all missing objects and using libraries convert them into mp3 files for voice message.

VI.Alert generation: Functionality of this module is to alert the worker that he does not have the required safety equipment and a voice message will be used for alerting.

VII.Model Deployment: Finally, the trained models will be deployed to the system.

*2) Object Detection and Recognition Process:* In the field of computer vision, object detection algorithms actively locate various elements within an image or scene. They enable computers to independently perform a range of activities by identifying specific objects in still images or moving footage. However, we should steer clear from employing historically used image processing-based approaches due to their drawbacks: longer processing times; more complex methodologies; and reduced levels of accuracy achievable with these techniques. Recently, the field has widely acclaimed Deep Learning-based object detection algorithms for their exceptional effectiveness. These algorithms autonomously learn image and object features, harnessing superior power to identify objects accurately.

The three primary methods employed in computer vision-based object identification tasks are: image classification, image localization along with classification – and object detection. When referencing ’image classification,’ , it involves the analysis of a picture – subsequently assigning it to one class from an extensive array of potential categories. This, indeed, represents not just a rudimentary but also pivotal application within computer vision. The method predicts the picture’s general category, yet it fails to discern its individual components. Consequently, experts have established a novel classification known as ’categorization with localization’. This field achieves item identification and positioning through extensive annotations within this category. This capability, however, bears a crucial limitation: it can only discern one category of entities at a time; specifically–object detection. To achieve the full labelling and categorization of diverse items–addressing this critical aspect is imperative.

In computer vision applications, which increasingly focus on the object detection category, deep learning demonstrates the highest potential for performance and outcomes.Object identification can be achieved in this field through a variety of methods. Nevertheless, most experts consider region-based CNNs as the most effective method.

The YOLO algorithm recommends the utilization of a comprehensive neural network: this advanced system predicts the perimeter bounding the boxes and class probabilities in one operation,a strategy divergent from earlier object identification algorithms. These previous methods deployed reused classifiers to execute the detection function; however, YOLO takes an innovative approach.

By employing a fundamentally unique approach to object recognition, YOLO has achieved state-of-the-art results; it significantly outperforms existing real-time object identification algorithms.

Utilising a fully connected, single layer enables YOLO to generate all its predictions; this stands in contrast with other algorithms such as Faster RCNN: they initially identify potential network,a step followed by individual recognition on each specific region.

Each picture necessitates numerous iterations for techniques employing Region Proposal Networks, whereas YOLO accomplishes the task with just a single iteration.

*3) Development of YOLO:* The YOLO (You Only Look Once) system, utilizing a singular neural network, actively detects objects in real-time: not only does it discern elements within pictures; but also within video frames.

YOLO, or You Only Live Once, finds paramount application in safety wear detection. It scrutinizes the compliance of

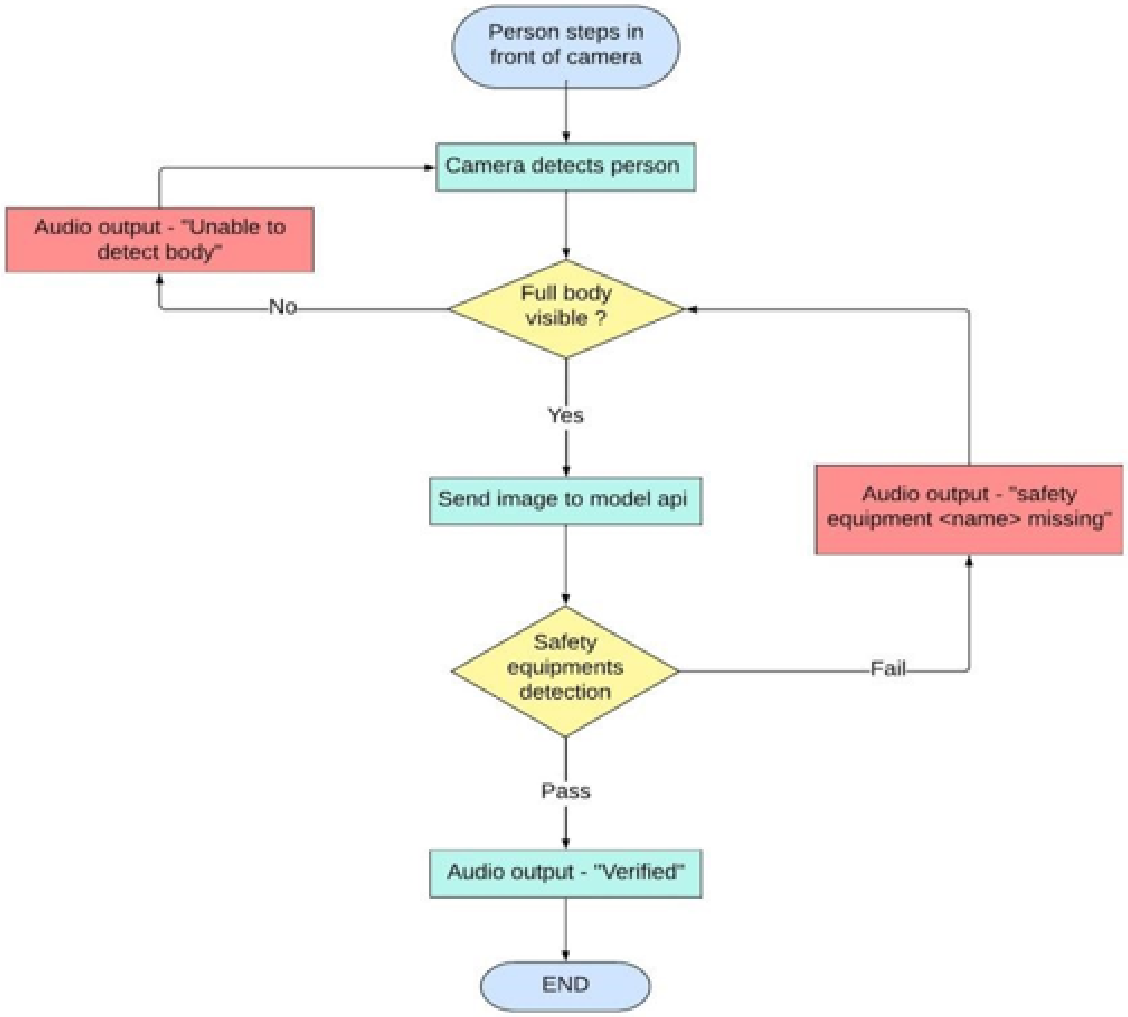


Fig. 3. Proposed Detection System Diagram

construction workers and industrial personnel with appropriate gear such as helmets, gloves and vests; thus ensuring their safety on sites or within these settings.

Particularly well-suited for safety wear detection, YOLO can detect numerous objects and reliably identify their locations; this superiority in speed and precision exceeds that of standard object detection systems.

Training YOLO for safety ware detection necessitates large datasets of photos and videos. These databases should encompass a diverse array of worker photos in different environments, donning various forms of safety equipment.

Once the dataset trains YOLO, it can detect safety equipment in real-time video feeds from construction site or industrial environment cameras. This detection ensures worker compliance with wearing appropriate safety gear, thereby enhancing worker safety.

Various worker safety applications also employ YOLO, using it to detect hazardous compounds and identify potential construction site safety concerns.

Computer vision technology's advancement anticipates a commensurate enhancement in various systems, this potential surge,marked by increased precision and efficiency holds promise for substantial improvements not only in worker safety but also various facets of industrial.

### C. YOLO and Field Worker Safety Detection System

This safety monitoring system consists of five essential components: Image Capture, Image Labeling, Object Detection, Alert Generation, and Model Deployment. It functions by taking pictures of workers at construction sites, assigning labels, identifying safety equipment, searching for any missing items, converting text into speech for alerts, and maintaining continuous surveillance of the site. The primary goal is to improve worker safety by ensuring the proper utilization of required safety gear, with the effective coordination and performance of each module being critical for the system's overall efficacy.

*1) Technique Used:* To create an effective device for detecting safety issues on construction sites, our team will address both hardware and software aspects of the challenge. Here are the hardware components we plan to employ:

1. Camera: This tool will capture images of workers to ensure their safety. 2. Computer Equipment with GPU: This hardware will be responsible for image processing and model deployment.

Our software components will include:

1. Frameworks (OpenCV and TensorFlow): These libraries will serve as essential resources for our project. 2. YOLO Model: Real-time object identification will be achieved through the utilization of the YOLO model. 3. Open Source Websites (e.g., makesense.ai): These platforms will be used for picture classification.

The device setup involves:

1. Camera Placement: Cameras will be positioned at a height of 1.71 meters to ensure accurate images of individuals of varying heights.

2. Capture Conditions: Pictures will only be taken when all points of the skeleton model are visible.

3. Model Deployment: Captured photos will be fed into the deployed model.

4. Result Computation: If the confidence level for an identified object is above sixty percent, it indicates the presence of that object.

5. Object Processing: Image processing models like YOLO will be employed for object identification, segmentation, and classification.

6. Deep Learning Models (e.g., CNN): Models such as Convolutional Neural Networks (CNN) will locate and categorize objects in digital photographs.

7. Tensor Board: This tool will be used for model evaluation and to display accuracy scores.

*2) Architecture of the YOLO project :* The following diagram illustrates the architecture of the convolutional neural network (CNN) model that serves as the foundation for YOLO. YOLO leverages a pre-trained model from ImageNet for its initial layers, thereby improving its overall performance. The method adopts a grid structure to identify objects, assigning each grid cell the task of predicting objects within its boundaries. To refine object detection, YOLO employs non-maximal suppression (NMS), which eliminates redundant or inaccurate bounding boxes, ensuring a single bounding box per object in the image. This strategy enhances the accuracy and efficiency of object detection simultaneously.

*3) YOLO V5 Review :* In 2020, Ultralytics introduced YOLO v5 as an open-source project, bringing forth a range of new features and enhancements. In contrast to the original YOLO, YOLO v5 adopts the EfficientDet architecture, which is rooted in the EfficientNet network design. This update contributes to improved accuracy and the capacity to identify a broader set of object classes, owing to its more sophisticated design.

The training datasets for YOLO and YOLO v5 exhibit significant differences. YOLO utilized the PASCAL VOC dataset, encompassing 20 object types, whereas YOLO v5 was trained on D5, a more extensive dataset featuring 600 object types. YOLO v5 employs dynamic anchor boxes generated through clustering, contributing to better object fitting. Additionally, YOLO v5 introduces a spatial pyramid pooling (SPP) layer, improving its capability to recognize small objects. While both YOLO v4 and v5 utilize similar loss functions in their training processes, YOLO v5 introduces the ”CIoU loss” to enhance performance, particularly on datasets with imbalances.

*4) Integration :* To combine YOLOv5 with worker safety monitoring on construction sites, we gathered and labeled images of workers wearing safety gear. We used tools like Makesense.ai and modified datasets from sources like roboflow. Then, we created a data.yaml file with class information and file locations. We trained the YOLOv5 model using PyTorch for detecting safety gear on construction sites.

To summarize, our discussion centered around incorporating YOLOv5 for the detection of worker safety on construction sites. The effectiveness hinges on the quality of data and the configuration of hyperparameters. YOLOv5 stands out as a robust tool for object detection, presenting opportunities for enhanced safety across diverse industries through continued research and development.

*5) Web Application:* The web application, developed in React, employs YOLOv5 to achieve real-time detection of safety equipment. It captures live video through the device’s camera, processes frames using YOLOv5, and produces bounding boxes with confidence scores for identified objects. The application filters out objects with low confidence levels and highlights instances of missing safety equipment, issuing notifications to promote workplace safety. The enhanced accuracy in detection is facilitated by the optimized YOLOv5 model and computer vision techniques.

*6) System Development Process:* A field worker safety detection system may include various accessories: Vests, shoes and masks. Enhancing worker visibility in conditions of poor light— a proactive measure indeed — can be achieved through the utilization of luminous vests. Steel-toed shoes shield employees from potential crush injuries caused by falling items, while protective masks enable workers to steer clear of inhaling perilous chemicals or dust.

To further enhance worker safety, it’s possible to augment these PPE pieces with sensors or other technologies. For example, a smart glove featuring built-in sensors could detect heat, pressure, or vibration fluctuations and alert the wearer about potential dangers. Similarly, by equipping a smart vest with GPS or motion detectors; one can monitor not only where but also what actions are taken by the wearer during job execution.

By integrating Personal Protective Equipment (PPE) with sensors and other technologies, we can enhance the efficiency of a field worker safety detection system. In tandem with contemporary alarm systems, this PPE provides an extra layer of protection for employees at workplaces that may pose potential risks.

*7) Results of Experimental Tests :*

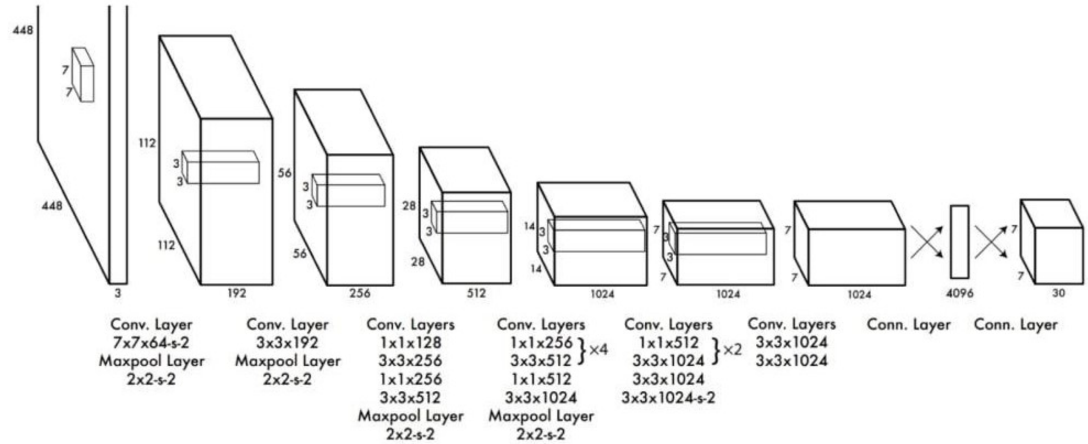


Fig. 4. The Architecture. Our detection network has 24 convolutional layers followed by 2 fully connected layers. Alternating 1 * 1convolutional layers reduce the features space from preceding layers. We pretrain the convolutional layers on the ImageNet classificationtask at half the resolution ( 224 * 224 input image) and then double the resolution for detection

| Dataset | Total Images | Accuracy Rate |
|---|---|---|
| Train | 844 | 98% |
| Validation | 146 | 93% |
| Test | 103 | 91% |

### D. Evaluation metrics

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1Score = 2 * \frac{Precision * Recall}{Precision + Recall}$$

*1) Accuracy:* The accuracy of a model is evaluated using four accuracy metrics: the Average Precision (AP), the F1 score, the COCO mean Average Precision (mAP), and the Precision x Recall curve.

*2) Precision:* To analyse the performance of the proposed method further, precision evaluation metrics is used. Precision highlights the capability of the model in rightly identifying the positive occurrence of safety kit detection. The precision-confidence curve, is a graph that shows the precision of a YOLO model at different confidence thresholds. a confidence level of 0.995, the model has a precision of 1.00. This means that all of the predictions that the model makes with a confidence level of 0.995 or higher are correct.

The precision-confidence curve also shows that the model's precision decreases as the confidence level decreases. At a confidence level of 0.2, the model's precision is only 0.4. This means that only 40% of the predictions that the model makes with a confidence level of 0.2 or higher are correct.

*3) Recall:* The evaluation metric recall also referred to as sensitivity expresses the capability of the model to rightly identify the desired.

*4) F1 Score:* The F1 score is a balanced sum of precision and recall. It exhibits the capability of the model's accuracy over the entire dataset. The F1-Confidence Curve for all classes peaks at a confidence threshold of 0.563, with an F1 score of 0.54. This means that our model achieves the best balance between precision and recall when you set the confidence threshold to 0.563.

### E. Conclusion

Our object detection model, like a seasoned orchestra, has harmoniously composed a symphony of precision and recall, orchestrating the detection of various objects with remarkable accuracy. Across the wide-ranging spectrum of classes, from Excavators to mini-vans, our model has demonstrated exceptional proficiency, consistently hitting high notes in both P (precision) and R (recall).

The mAP50 metric, a measure of overall detection performance, echoes the virtuosity of our model, achieving a score of 0.529, indicating its ability to strike a balance between precision and recall. This harmonious interplay is further reinforced by the mAP50-95 metric, which evaluates performance across a wider range of IoU (Intersection over Union) thresholds, reaching a commendable 0.352.

The SPP (Spatial Pyramid Pooling) is a feature present in both YOLO v4 and v5, but the latter incorporates significant improvements in SPP to enhance the detection of smaller objects. Despite employing a similar loss function, YOLO v5 introduced the "CIoU loss" to improve overall performance, especially on datasets with imbalances.

Delving into the individual classes, our model shines particularly brightly in the detection of Hardhats, Safety Vests, and Safety Cones, achieving mAP50 values of 0.644, 0.617, and 0.846, respectively. These impressive scores underscore

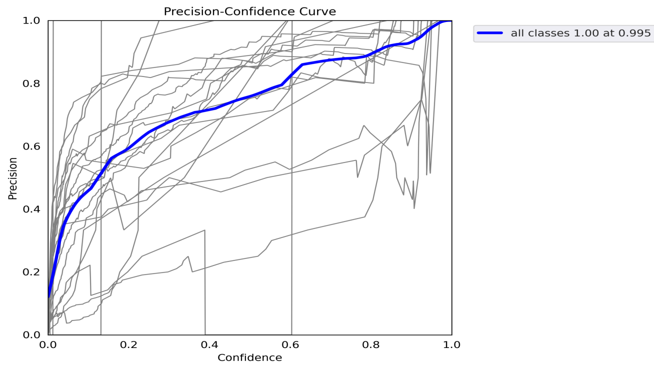


Fig. 5. Accuracy

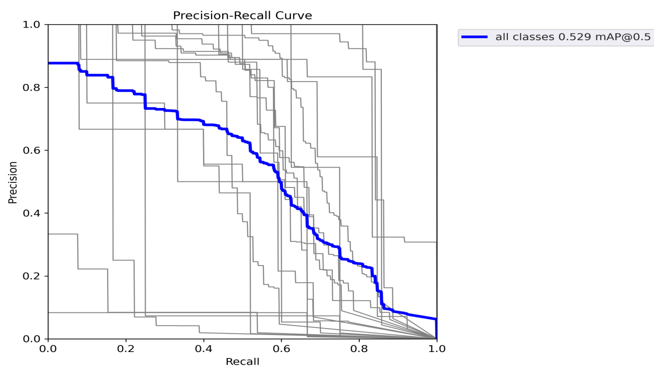


Fig. 6. Precision

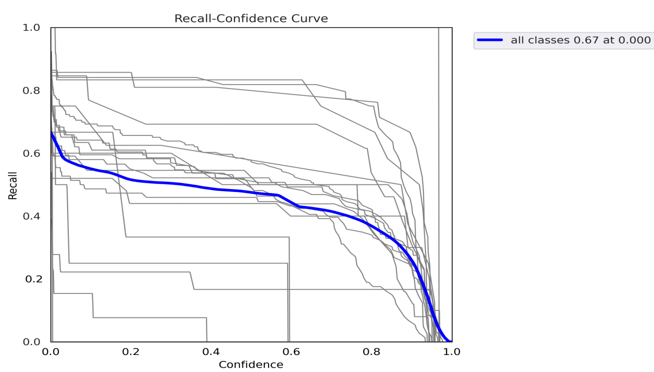


Fig. 7. Recall

the model's ability to identify these crucial safety equipment with remarkable precision and recall.

Even in the detection of more challenging classes, such as dump trucks and wheelchairs, our model maintains a respectable performance, achieving mAP50 values of 0.758 and 0.412, respectively. This adaptability speaks to the model's robustness and generalizability across diverse object categories.

Overall, our object detection model has performed admirably, demonstrating its prowess in identifying a wide range of objects with high precision and recall. This symphony of detection, a testament to the model's training and fine-tuning, holds promise for real-world applications where accurate object detection is paramount.

### *F. References*

1) Ding, W., Liu, J., Zhang, W., Zhang, W., & Wang, Q. (2020). An Internet of Things System for Underground Mine Air Quality Pollutant Prediction Based on Azure Machine Learning. Journal of Sensors, 2020, 1-10.
2) F. Zhou, H. Zhao and Z. Nie, "Safety Helmet Detection Based on YOLOv5," 2021 IEEE International Conference on Power Electronics, Computer Applica-

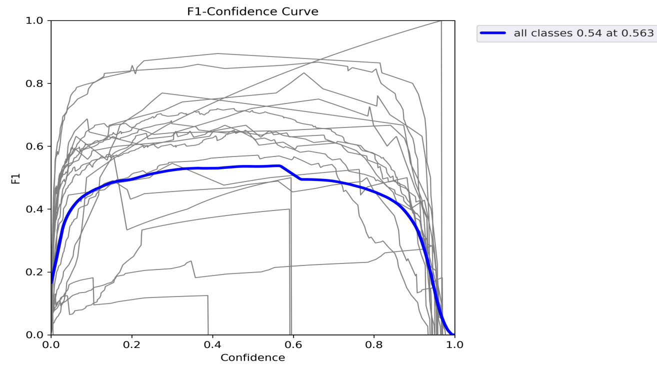


Fig. 8. F1 Score

tions (ICPECA), Shenyang, China, 2021, pp. 6-11, doi: 10.1109/ICPECA51329.2021.9362711.

3) Kang Li and Xiaoguang Zhao and Jiang Bian and Min Tan Automatic Safety Helmet Wearing Detection CoRR abs/1802.00264 2018
4) A. Das, M. Wasif Ansari and R. Basak, "Covid-19 Face Mask Detection Using TensorFlow, Keras and OpenCV," 2020 IEEE 17th India Council International Conference (INDICON), New Delhi, India, 2020, pp. 1-5, doi: 10.1109/INDICON49873.2020.9342585.
5) Delhi, Venkata Santosh Kumar, R. Sankarlal, and Albert Thomas. "Detection of personal protective equipment (PPE) compliance on construction site using computer vision based deep learning techniques." Frontiers in Built Environment 6 (2020): 136.
6) Nath, Nipun D., Amir H. Behzadan, and Stephanie G. Paal. "Deep learning for site safety: Real-time detection of personal protective equipment." Automation in Construction 112 (2020): 103085.
7) Construction Safety Equipment Detection System Yogesh Kawade, Mayur Kasture, Yash Mallawat, Aniruddh Dubal Student, B.Tech., Construction Engineering and Management, Symbiosis Skills and Professional University, Pune, Maharashtra, India Asst. Prof., Construction Engineering, Symbiosis Skills and Professional University, Pune, Maharashtra, Pune, India
8) Karlsson, Jonathan, et al. "Visual Detection of Personal Protective Equipment and Safety Gear on Industry Workers." arXiv preprint arXiv:2212.04794 (2022).
9) Gallo, Gionatan, et al. "A smart system for personal protective equipment detection in industrial environments based on deep learning." 2021 IEEE International Conference on Smart Computing (SMARTCOMP). IEEE, 2021.
10) Nath, Nipun D., and Amir H. Behzadan. "Deep learning detection of personal protective equipment to maintain safety compliance on construction sites." Construction Research Congress 2020: Computer Applications. Reston, VA: American Society of Civil Engineers, 2020.
11) Ke, Xiao, Wenyao Chen, and Wenzhong Guo. "100+ FPS detector of personal protective equipment for worker safety: A deep learning approach for green edge computing." Peer-topeer networking and applications (2022): 1-23.
12) Torres, Pedro, et al. "A Robust Real-time Component for Personal Protective Equipment Detection in an Industrial Setting." ICEIS (1). 2021.
13) Saudi, Madihah Mohd, et al. "Image detection model for construction worker safety conditions using faster R-CNN." International Journal of Advanced Computer Science and Applications (2020).
14) Moohialdin, Ammar, et al. "A Real-Time Computer Vision System for Workers' PPE and Posture Detection in Actual Construction Site Environment." EASEC16: Proceedings of The 16th East Asian-Pacific Conference on Structural Engineering and Construction, 2019. Springer Singapore, 2021.
15) Li, Yange, et al. "Deep learning-based safety helmet detection in engineering management based on convolutional neural networks." Advances in Civil Engineering 2020 (2020): 1-10.
16) Ferdous, Md, and Sk Md Masudul Ahsan. "PPE detector: a YOLO-based architecture to detect personal protective equipment (PPE) for construction sites." PeerJ Computer Science 8 (2022): e999.
17) Protik, Adban Akib, Amzad Hossain Rafi, and Shahnewaz Siddique. "Real-Time personal protective equipment (PPE) detection using YOLOv4 and TensorFlow." 2021 IEEE Region 10 Symposium (TENSYMP). IEEE, 2021.
18) Phuc, Le Tran Huu, et al. "Applying the Haar-cascade Algorithm for detecting safety equipment in safety management systems for multiple working environments." Electronics 8.10 (2019): 1079.
19) Qiu, Jianzhong, et al. "A Target Detection and Evaluation Method for Safety Protection Equipment." Journal of Physics: Conference Series. Vol. 1966. No. 1. IOP Publishing, 2021.
20) Arabi, Saeed, Arya Haghighat, and Anuj Sharma. "A deep learning based solution for construction equip-

ment detection: from development to deployment." arXiv preprint arXiv:1904.09021 (2019).

21) Marks, Eric Daniel, and Jochen Teizer. "Method for testing proximity detection and alert technology for safe construction equipment operation." Construction Management and Economics 31.6 (2013): 636-646.
22) Fang, Weili, et al. "Automated detection of workers and heavy equipment on construction sites: A convolutional neural network approach." Advanced Engineering Informatics 37 (2018): 139-149.
23) Wang, Mingzhu, et al. "Predicting safety hazards among construction workers and equipment using computer vision and deep learning techniques." ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction. Vol. 36. IAARC Publications, 2019.
24) Zhao, Meng, and Masoud Barati. "Substation Safety Awareness Intelligent Model: Fast Personal Protective Equipment Detection using GNN Approach." IEEE Transactions on Industry Applications (2023).