Journal of Innovative Image Processing (ISSN: 2582-4252)
www.irojournals.com/iroiip/

CSOD-24: Construction Site Object Detection Dataset for Safety Monitoring at Construction Site using Deep Learning

Meenakshi N. Shrigandhi[1], Sachin R. Gengaje[2]

[1-2]Department of Electronics Engineering, Walchand Institute of Technology, Solapur, India
E-mail: [1]mapasnur@witsolapur.org, [2]srgengaje@witsolapur.org

Abstract

Monitoring the use of personal protective equipment (PPE) and worker proximity to heavy machinery are two areas where ensuring safety compliance on construction sites continues to be difficult. The lack of dynamic ambient circumstances, comprehensive annotations, and real-time video data in existing datasets restricts their applicability to real-world situations. In order to fill in these gaps, this work presents CSOD-24, a video dataset intended for construction site object detection and safety monitoring. The dataset includes 100 ten-second video clips (16.6 minutes total), covering four major classes: "Dump Truck", "Worker with Helmet", "Worker without Helmet" and "Excavator". The videos were recorded at 10 frames per second (fps) and annotated in .txt, .json, and .xml formats. This dataset supports the development and validation of algorithms for automated safety compliance monitoring, object detection, and tracking in dynamic construction environments. The CSOD-24 dataset address these challenges, enabling a robust foundation for advancing computer vision-based safety monitoring, thereby contributing to reduced workplace hazards and improved operational efficiency.

Keywords: Safety Monitoring, Construction Equipment, Personal Protective Equipment, Object Detection, Classification, Proximity Detection.

1. Introduction

The construction industry faces unique safety and operational challenges, particularly when it comes to monitoring the presence and actions of both workers and machinery on site. Safety compliance, including the use of personal protective equipment (PPE) [7], [21], [22], [27] like helmets, is essential to reducing accidents, as is the effective identification and management of heavy equipment [4], such as excavators and loaders. Traditional safety

DOI: https://doi.org/10.36548/jiip.2025.1.009
Received: 29.02.2025, received in revised form: 28.03.2025, accepted: 13.04.2025, published: 25.04.2025

monitoring methods often rely on manual inspections, which are time-consuming and can be inconsistent due to human error. As a result, computer vision and deep learning technologies [16], [15], [20] have become integral to enhancing safety and efficiency on construction sites by automating the detection and tracking of workers and equipment [2].

Automated video-based monitoring systems, driven by Convolutional Neural Networks (CNNs) [3], [5] and other deep learning models [1], offer promising solutions for real-time identification of workers and machinery. However, deploying these models in real-world construction environments presents several challenges like i) occlusion in which workers may be partially or fully obscured by equipment, materials, or other workers, leading to missed detections ii) lighting variations like harsh sunlight, shadows, and artificial lighting conditions affect the visibility and accuracy of detections iii) high dust levels: construction sites often have airborne dust and debris, which can blur video frames and reduce model performance iv) dynamic background clutter in which moving machinery, temporary structures, and varying environmental conditions add complexity to object identification.

Despite advancements in deep learning, existing datasets for construction site monitoring often lack comprehensive real-time video annotations, focus primarily on static images, or do not account for worker-equipment proximity detection. To address these gaps, this study introduces CSOD-24, a video dataset specifically designed for helmet detection, worker-equipment interaction analysis, and safety monitoring in dynamic construction environments.

In this study, a new video dataset: CSOD-24 is presented to specifically support the detection of workers (with and without helmets) and key construction machinery, including excavators and dump trucks. The dataset consists of 100 ten-second video clips, each at a resolution of 10 frames per second (fps), resulting in a total of 10,000 video frames. The selection of 10-second videos enables the recording of significant worker-equipment interactions and offers enough temporal context to identify mobility patterns and safety infractions. In order to balance computational efficiency with capturing sufficient temporal resolution and guarantee that important events like wearing a helmet and receiving proximity alerts are detected without an excessive number of duplicated frames, a frame rate of 10 frames per second was chosen. Each frame in a total of 10,000 video frame is annotated with bounding boxes in three standard formats: .txt, .json, and .xml, allowing for broad compatibility with various machine learning frameworks. In contrast to other datasets, CSOD-24 offers

comprehensive annotations for real time video-based safety monitoring, facilitating more precise hazard identification and worker compliance checks.

This dataset addresses several specific needs in the field of construction safety. First, by distinguishing between workers who are wearing helmets and those who are not, it supports automated compliance checks, an essential function for real-time monitoring systems aimed at preventing injuries. Second, the inclusion of machinery classes, such as excavators and dump trucks, facilitates object detection systems that can alert operators and safety personnel when workers enter hazardous zones near operating equipment [32], [25]. These features make the dataset a versatile tool for training and evaluating models that can operate effectively in real-world construction environments.

Ultimately, the development and deployment of such datasets have significant implications for the future of construction safety. With the ability to automatically detect and classify on-site objects, deep learning models trained on this dataset can support a wide range of applications, from worker compliance monitoring to collision prevention and equipment tracking. By providing an annotated, application-specific dataset, this study contributes a valuable resource for advancing computer vision-based safety systems in construction, potentially reducing the incidence of workplace injuries and enhancing operational efficiency.

2. Literature Survey

Safety monitoring on construction sites is a critical aspect of managing and reducing hazards in high-risk environments. Given the complex and dynamic nature of construction activities, workers are exposed to numerous potential dangers, including falls [13],[14], heavy machinery accidents [29],[30],[31], and structural failures. Effective safety monitoring ensures compliance with regulations, identifies potential risks early, and helps in implementing preventive measures to protect workers and equipment's. Recent advancements, such as remote sensing, artificial intelligence, and UAVs, have greatly enhanced safety protocols by enabling real-time monitoring and data collection. These technologies support proactive decision-making and enhance the safer work environments, ultimately minimizing accidents and improving overall productivity on construction sites. The authors in [11] addresses a significant gap in publicly available labelled data for rebar counting on construction sites using UAV imagery, which is essential for training deep learning models. Unlike previous datasets, it includes comprehensive annotations and uses diverse augmentation techniques to improve

model robustness. Additionally, it is adaptable for other construction tasks, such as estimating rebar diameter or shape classification, making it highly versatile and beneficial for automated inspection research.

The synthetic datasets for construction site safety monitoring reveals a need to streamline the labor-intensive processes of dataset generation and manual labeling. Traditional methods, which often treat image generation and labeling as separate steps, face issues with time efficiency, dataset realism, and privacy compliance. The research [10] proposes a novel, automated approach that integrates synthetic image generation and labeling to closely resemble real-world construction scenarios by simulating different lighting, angles, and asset configurations. Tested with object detection algorithms, this method shows promising improvements in dataset accuracy and automation. However, further work is needed to expand object diversity and validate the model on real-world data for broader applicability across diverse construction safety contexts.

The dataset presented in [9] includes 1,046 images captured from four static cameras placed around a construction site, covering eight object classes typical in environments, such as (e.g., excavators, dump trucks, cranes, and personnel). The images, collected during morning hours, were manually annotated with bounding boxes to support object detection and classification tasks in construction monitoring, making this dataset valuable for training neural networks aimed at improving safety and operational efficiency. Despite its strengths, the dataset has limitations that could be addressed in future work. It is limited to static images from a single location, which may reduce its generalizability. The lack of dynamic tracking data restricts its use in real-time safety applications, where continuous monitoring of worker proximity to machinery is essential. Additionally, the absence of safety-specific labels, such as helmet usage or risk-related behaviours, limits its utility for targeted safety assessments. Addressing these gaps could broaden the dataset's applicability for advanced safety monitoring and compliance in diverse construction environments.

The "Moving Objects in Construction Sites" (MOCS) dataset [12] addresses the need for large-scale, diverse data for detecting objects in construction environments. It includes 41,668 images collected from 174 sites, featuring 2,22,861 annotated instances across 13 categories (e.g., workers, excavators, cranes). Annotations include bounding boxes and pixel-level masks for precise detection and segmentation. The dataset's diversity encompasses various lighting, weather, occlusion, and viewpoint conditions, making it robust for training

Deep Neural Networks (DNNs). Benchmarks with 15 DNN-based detectors show high performance in detecting objects under challenging scenarios. The dataset supports safety monitoring, productivity analysis, and automation tasks like hard hat detection and hazard prevention. However, limitations include a focus on images rather than videos and missing annotations for structural elements. Expanding categories and global coverage are recommended for future work. MOCS offers a valuable tool for advancing object detection and automation in construction.

The authors in [8] included the use of YOLO-based models, particularly YOLOv5, known for its efficiency in real-time detection of workers, PPE, and equipment, essential for maintaining site safety. Edge computing has been integrated into recent solutions to process data on-site, reducing network load and ensuring faster response times. Image cropping techniques are commonly applied to focus on specific regions of interest, thereby enhancing the visibility of small objects. Additionally, feature fusion and contextual cues are employed to differentiate workers and equipment from complex site backgrounds. Despite these advancements, challenges remain in scaling detection systems for large, multi-scale sites, dynamically adjusting detection focus for various object sizes, and handling diverse lighting and angles. These gaps point to the need for more adaptive, flexible SOD approaches that can perform reliably in real-world construction environments.

The study [7] utilizes deep learning to improve construction safety by detecting personal protective equipment (PPE), specifically hard hats and high-visibility vests, using YOLOv3, YOLOv4, and YOLOv7 algorithms. A newly developed dataset containing 11,000 images with 88,725 labelled PPE items significantly enhances detection accuracy, achieving 97% mean average precision (mAP) and a real-time processing speed of 25 frames per second (FPS). YOLOv7 outperforms its predecessors with better speed and precision, making it well-suited for real-time compliance monitoring to ensure workers adhere to PPE regulations, thus reducing risks of occupational injury. However, the study notes challenges in detecting multiple PPE items in crowded or complex scenes and scaling the system to varied construction environments, which present opportunities for further research.

The DATS_2022 dataset [6] addresses the need for comprehensive training data customized to unstructured Indian traffic conditions. It comprises over 10,000 high-resolution images captured using smartphones, covering diverse environments like urban roads, highways, rural areas, and hilly terrains in Maharashtra. Unique features include stray animals,

unregulated traffic, and varied weather and lighting conditions. The images are annotated using tools like LabelImg, with outputs in XML (Pascal VOC), TXT (YOLO), and JSON (Create ML) formats, enabling direct application in machine learning. Unlike traditional datasets, it reduces redundancy by systematically extracting frames. It supports research in autonomous navigation, driver-assistance systems, and IoT-based intelligent transport, while aligning with technologies like 5G and V2X communication. This dataset, is freely accessible through Mendeley,which is a valuable resource for enhancing object detection models in complex traffic scenarios, addressing a gap in region-specific data for India. While relatively smaller in size, DATS_2022 offers periodic updates to remain relevant and impactful.

Table 1 provides a summary of existing datasets and their corresponding algorithms used for construction site safety monitoring. It outlines key attributes such as the number of images or instances, annotation formats, implemented algorithms, applications, and the limitations of each dataset. This overview highlights the gaps in current datasets, including limited diversity, static images, and narrow focus on specific safety elements, which the CSOD-24 dataset aims to address comprehensively

Table 1. Summary of Datasets and Algorithms for Construction Site Safety Monitoring

Dataset Name	Images/ Instances	Annotations	Algorithms Implemented	Applications	Limitations
UAV Rebar Dataset [11]	13,974 images; 19,034 rebar instances	Bounding boxes in VOC XML format	Faster R-CNN, YOLO variants (ResNet, MobileNetV3, EfficientNetV2, etc.)	Automated rebar counting, rebar shape classification, rebar diameter estimation	Limited to five sites, lacking diversity in conditions like rain or occlusion. Only supports counting tasks; diameter and spacing estimation are excluded.
Synthetic Construction Dataset (S, M, SM) [10]	S: 6400, M: 2880, SM: 9280	Automatic labeling in YOLOv4 format	Faster R-CNN, RetinaNet, YOLOv4	Object detection for machinery and workers on sites	Synthetic dataset lacks real-world complexities like unpredictable weather, occlusions, and lighting variability, reducing generalizability.
Manually Classified Construction Site Dataset [9]	1,046 images captured by 4 static cameras	Bounding boxes in .txt format for 8 object classes	Not specified	Object detection and classification of construction	Dataset is small-scale and lacks diversity in weather, viewpoints, and lighting conditions, which may limit

				machinery and workers	model generalization.
Moving Objects in Construction Sites (MOCS) [12]	41,668 images; 2,22,861 annotated instances	Bounding boxes and masks for 13 categories	Faster R-CNN, YOLO, Mask R-CNN, PointRend	Object detection and segmentation of moving objects in construction sites	Limited diversity across global regions as data is collected only from China and Pakistan. Lacks annotations for structural elements or detailed machine keypoints, limiting advanced applications.
PPE Compliance Detection Dataset [7]	11,000 images; 88,725 labels	Bounding boxes (XML format)	YOLOv3, YOLOv4, YOLOv7	Real-time detection of PPE compliance (helmets, vests)	Limited to detection of only helmets and high-visibility vests; other PPE types like gloves or boots are not considered, reducing its comprehensiveness.
Small Object Detection (SOD) Dataset [8]	2,99,655 images; 5,75,913 instances	Bounding boxes for workers and equipment	YOLOv5	Small object detection for safety monitoring at construction sites	The dataset struggles with detecting extremely small objects (appearing as dots) and lacks data diversity in imaging angles, limiting robustness in real-world scenarios.

Existing datasets for construction site safety monitoring lack real-world diversity, dynamic video-based interactions, and comprehensive annotations for safety behaviors like helmet compliance [23], [26], [28] and worker proximity to machinery [31], [29]. These limitations hinder the development of robust machine learning models for real-time monitoring in dynamic construction environments. CSOD-24 dataset overcomes these gaps by providing a well-annotated, video-based dataset under diverse environmental and operational conditions, enabling the development of advanced safety monitoring systems.

Following the literature survey, this study provides a comprehensive overview of the methodology employed in creating the CSOD-24 dataset, specifically designed for object detection in construction site environments. The dataset creation process encompasses data collection techniques utilizing diverse cameras and settings, detailed annotation procedures for generating bounding boxes in multiple formats, and the classification of objects into four key categories. This comprehensive approach ensures the dataset is robust, versatile, and

customized to meet the requirements of safety monitoring and equipment identification in dynamic construction scenarios.

3. Dataset Construction

This video dataset was created to support object detection for safety monitoring at construction sites, specifically targeting two essential safety conditions: (1) whether workers are wearing safety helmets and (2) their proximity to excavators. By capturing these specific behaviours, the dataset is designed to facilitate the development of automated monitoring systems aimed at reducing accidents and enforcing safety compliance in dynamic construction environments. Figure. 1 explains the process of dataset creation

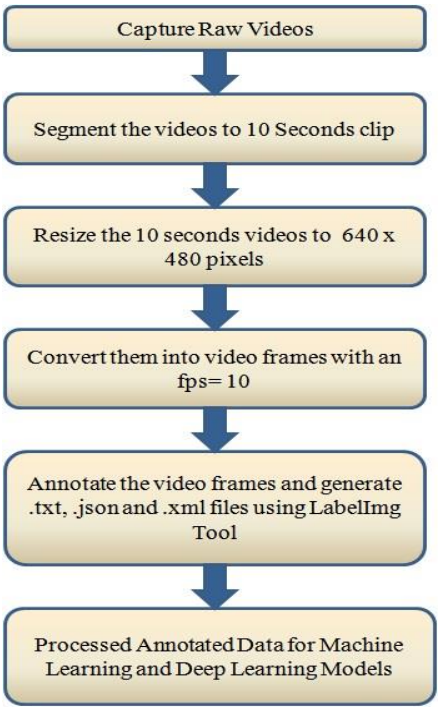

Figure 1. Process of Dataset Creation

3.1. Data Collection

Data for this dataset was collected across a variety of construction sites, including small-scale projects in the city of Solapur, Maharashtra and larger construction projects in the city of Pune, Maharashtra. This approach ensured coverage of diverse safety scenarios across different types of working environments. Videos were recorded using high-resolution personal devices, including a Samsung Galaxy A22 smartphone with a 48MP camera and a Panasonic

Handicam, without the involvement of professional photographers. This setup enabled high-quality videos suitable for detecting safety equipment compliance and monitoring worker proximity to heavy machinery. Clear, high-quality video from these devices was useful for tracking worker proximity to large machinery and determining helmet compliance. Different recording devices, however, may introduce bias in the quality of the video since detection accuracy may be impacted by differences in resolution, color accuracy, and compression artifacts. In order to counteract this, pre-processing methods like resolution normalization were used to guarantee uniformity throughout all videos.

Data collection was carried out under various lighting and weather conditions to enhance the dataset's robustness and adaptability. The focus was on capturing footage that prominently featured workers and excavators across a range of contexts, including active machinery operation, material handling, and worker movement. By capturing workers in different positions and activities, the dataset effectively covers a wide range of scenarios, supporting the development of models for proximity risk assessment and helmet compliance detection. The dataset specifically focuses on videos collected from building construction sites, despite the existence of various other types of construction sites. This ensures a consistent context for developing and evaluating safety monitoring models.

The primary purpose of this dataset is to facilitate safety monitoring of workers at construction sites through video analysis. The dataset is categorized into four classes: worker with helmet, worker without helmet, excavator, and dump truck. While the primary focus is on the three core classes: worker with helmet, worker without helmet, and excavator. The dump truck class is included because excavators are frequently surrounded by dump trucks during construction activities. Table 2 show the class object Ids and the corresponding classes.

Table 2. Class Object IDs and Corresponding Classes

Object ID	Class
0	Worker without Helmet
1	Worker with Helmet
2	Excavator
3	Dump Truck

3.2 Data Preprocessing

The raw video footage, initially recorded at a resolution of 1920 x 1080 pixels, underwent several pre-processing steps to optimize it for safety monitoring applications. First, the videos were segmented into 10-second clips using Windows Movie Maker. This segmentation allowed for the creation of multiple short videos, capturing various worker actions within the same scene and adding diversity to the dataset. By segmenting the footage in this way, the research obtained a total of 100 distinct 10-second videos, each representing different activities or perspectives within similar contexts. The selection of 10-second videos enables the recording of significant worker-equipment interactions and offers enough temporal context to identify mobility patterns and safety infractions. In order to balance computational efficiency with capturing sufficient temporal resolution and guarantee that important events like wearing a helmet and receiving proximity alerts are detected without an excessive number of duplicated frames, a frame rate of 10 frames per second was chosen. Following segmentation, each 10-second clip was resized to 640 x 480 pixels, significantly reducing the file size. This resizing not only conserved storage space but also enhanced processing efficiency, making the data easier to handle without sacrificing important visual details necessary for detecting worker safety compliance and proximity to machinery.

To prepare the dataset for detailed analysis, each resized 10-second video was then converted into individual frames at a rate of 10 frames per second (fps), resulting in a comprehensive collection of 10,000 frames. These frames provide a well-structured dataset that can be used to train and validate machine learning models aimed at monitoring worker safety and compliance. Together, these preprocessing steps ensured that the dataset is not only manageable in size but also contains sufficient variety and granularity to support robust safety monitoring applications. Figure 2 illustrates example frames extracted from different videos, showcasing the variety in worker activities and contexts captured in the dataset.

Figure 2. Frames Extracted from Different Videos

3.3. Annotation of the Video Frames Using LabelImg

The annotation of video frames was carried out using LabelImg, an open-source annotation tool widely used for object detection datasets. This process was carefully designed to ensure that the dataset supports diverse deep learning frameworks. Annotations were generated in three formats: .txt (YOLO), .xml (Pascal VOC), and .json (Create ML), making the dataset versatile and compatible with a range of machine learning models. Figure 3 shows the LabelImg annotation Tool while performing the annotation of a video frame containing worker with helmet.

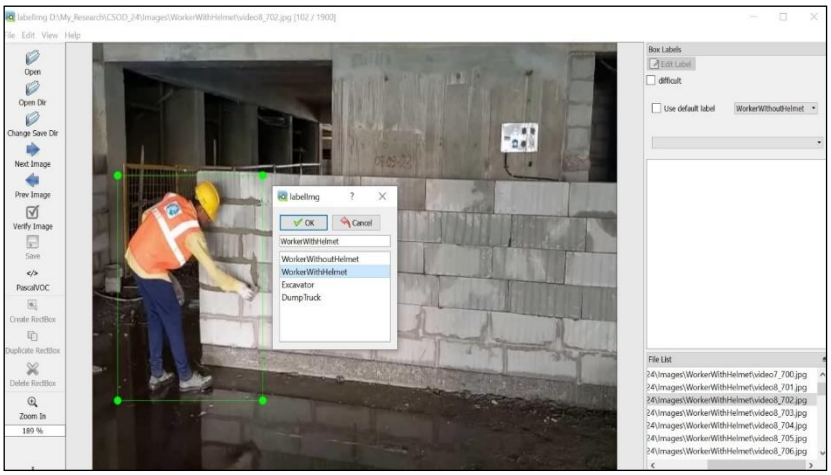

Figure 3. LabelImg Annotation Tool while Performing the Annotation of a Video Frame Containing Worker with Helmet

The step-by-step process is detailed below

3.3.1 Preparing the Dataset

Extracted frames from the videos were saved in a dedicated directory in .jpg format for compatibility with LabelImg. A consistent naming convention was followed (e.g., video1_001.jpg) to maintain order and ensure traceability during annotation. The directory structure was organized to separate frames into categories, scenarios, and sequences as required.

3.3.2 Setting Up LabelImg for Annotation

LabelImg was installed and configured to support multiple formats. The dropdown menu of the tool was used to choose the annotation format. For YOLO (.txt): The tool was set to save annotations in YOLO's bounding box format. Each annotation file includes the object class, normalized bounding box coordinates, and dimensions relative to the image size. For Pascal VOC (.xml): The annotations were saved in XML format, containing metadata such as image size, object name, and pixel coordinates of the bounding box. For Create ML (.json): JSON annotations included class labels, bounding box coordinates, and other attributes compatible with Apple's machine learning framework.

3.3.3 Annotation Workflow

Each frame was opened in LabelImg, and the relevant objects were identified. For each object, a bounding box was manually drawn by clicking and dragging the mouse to enclose the object. The bounding box's placement ensured that the object was entirely contained within the box while minimizing surrounding empty space. After drawing the bounding box, the appropriate class label (e.g., WorkerWithHelmet, WorkerWithoutHelmet Excavator, DumpTruck) was selected from the predefined list. The annotations were saved in the selected format, generating a separate annotation file for each frame.

3.3.4 Annotation in Multiple Formats

The annotation process was repeated for each frame to generate annotations in all three formats. Each .txt file contains a line for each object with the format

```
<class_id> <x_center> <y_center> <width> <height>
```

All coordinates are normalized (range: 0–1) relative to the image dimensions. Figure 4 shows the .txt format containing two classes: worker with helmet and excavator.

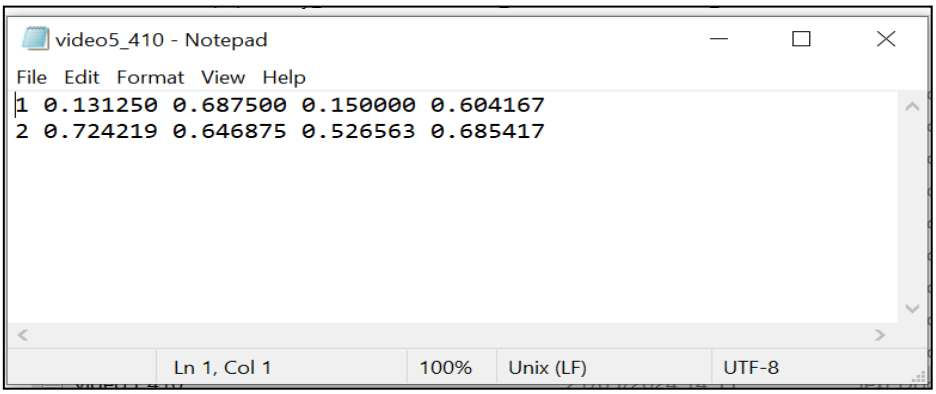

Figure 4. .txt Annotation File

.json file contains a list of annotations in a structured JSON format, detailing image name, class labels, bounding box dimensions. Figure 5 shows the .json format containing two classes: worker with helmet and excavator

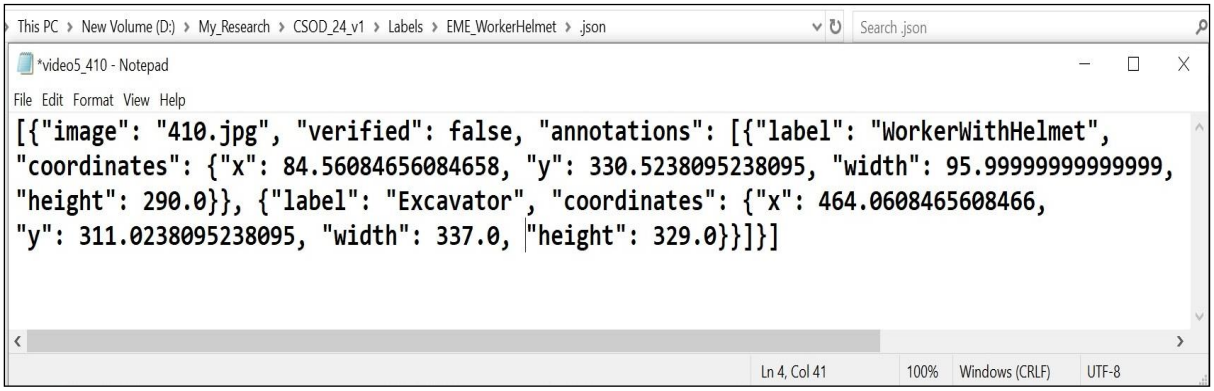

Figure 5. .json Annotation File

Each .xml file contains structured information in XML format, including: Image properties (size, path) and object properties (class name, bounding box coordinates in pixels) as shown in Figure 6 containing two classes: worker with helmet and excavator.

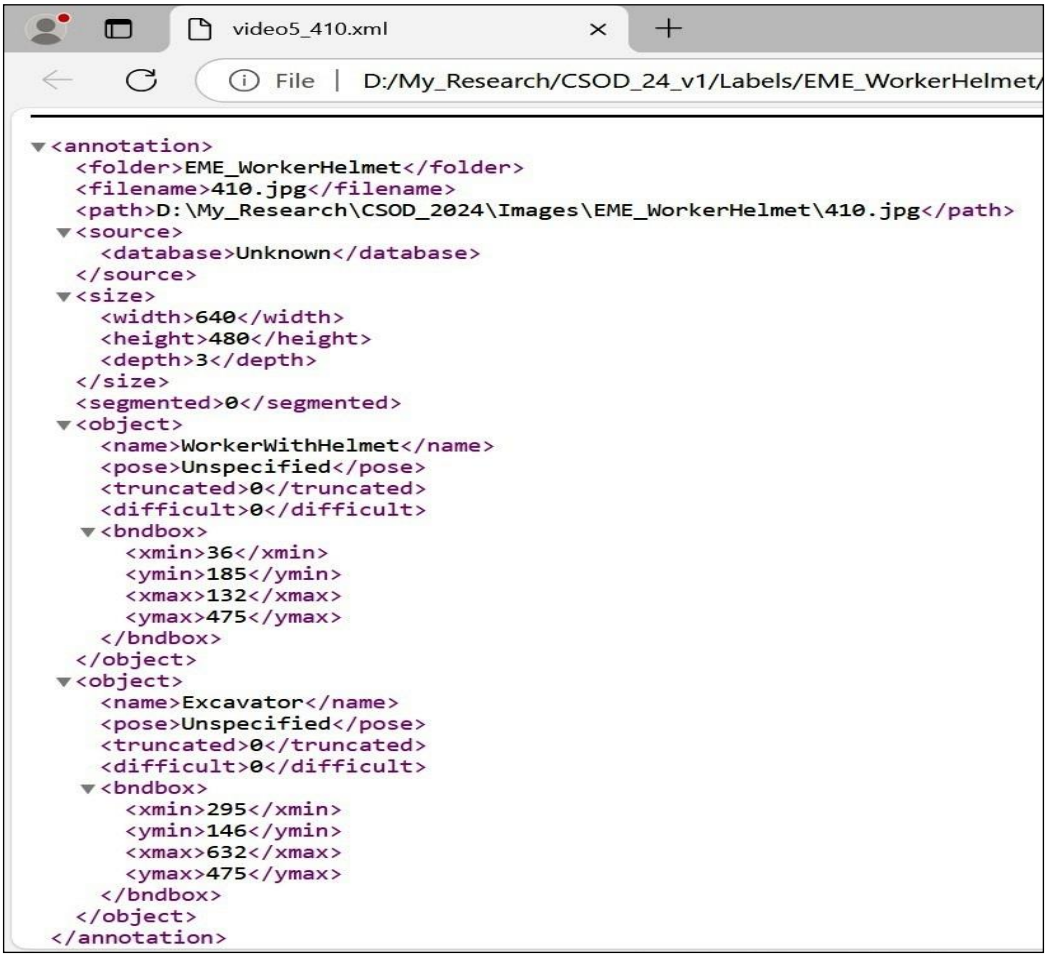

Figure 6. .xml Annotation File

3.3.5 Dataset Statistics and Distribution

The dataset, comprising 10,000 video frames, is organized into several folders to support detailed analysis and annotations. A folder named images contains the 10,000 video frames, while another folder named labels includes three subfolders: .txt, .json, and .xml, each providing 10,000 annotations corresponding to the 10,000 frames. Additionally, the labels folder contains a classes.txt file, listing the defined classes used in the annotations. A separate folder holds 100 video files, each with a duration of 10 seconds, providing raw footage for further exploration and validation. This comprehensive structure ensures flexibility in processing, allowing users to work with their preferred annotation format and analyze the data effectively.

The inclusion of annotations in multiple formats (.txt, .json, .xml) facilitates compatibility with various machine learning frameworks, making the dataset versatile for training object detection and classification models. The presence of both image and video data

enhances the scope for temporal analysis, that are essential for understanding dynamic interactions on construction sites. These organizational features streamline the dataset's usability for safety monitoring applications.

This dataset has been further analyzed to uncover patterns and trends, with the results visualized through various charts, including pie and bar charts. These charts illustrate the proportional distribution of annotated frames from small and large construction projects, as well as the occurrence of different object combinations and scenarios captured in the dataset, providing valuable insights into safety compliance and worker-machine interactions.

3.3.5.1 Safety Compliance in Small vs. Large Projects

The dataset reveals distinct patterns in safety compliance between small and large construction projects. In small projects, where informal practices often prevail, a significant proportion of workers are observed without helmets, reflecting limited adherence to safety protocols. This non-compliance contributes to a higher frequency of unsafe scenarios, especially as workers frequently operate in close proximity to excavation equipment without adequate protective gear. In contrast, large projects enforce stringent safety measures, making it mandatory for all workers to wear helmets and restricting access to excavation zones. As a result, frames from large projects predominantly depict safe scenarios, with workers maintaining a safe distance from excavators and adhering to helmet compliance. This dichotomy highlights the disparity in safety standards between small and large-scale operations and emphasizes the essential need for enhanced safety awareness and enforcement in smaller construction projects.

3.3.5.2 Distribution of Frames by Project Type

The pie chart in Figure 7 illustrates the proportional distribution of annotated frames collected from small and large construction projects. Out of the total 10,000 frames, 45% (4,500 frames) were captured from small projects, where workers often lacked helmets and worked closer to excavation zones. In contrast, 55% (5,500 frames) were obtained from large projects, which enforce stricter safety regulations, including mandatory helmet use and restricted access to excavation areas. The chart highlights the significant contribution of both project types to the dataset, emphasizing the variability in safety practices across different construction scales.

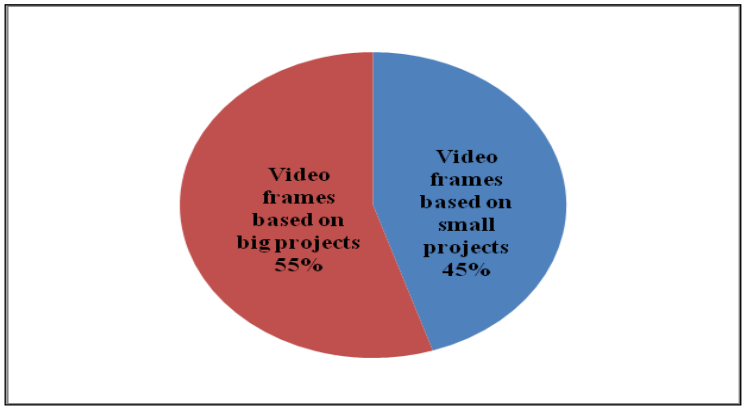

Figure 7. Proportional Distributions of Annotated Frames from Small and Large Construction Projects

3.3.5.3 Proportional Distribution of Annotated Frames

The bar chart in Figure 8 illustrates the distribution of annotated video frames based on different object combinations and scenarios. The dataset includes 2,100 frames containing only an excavator, making it a frequently observed single-class scenario, followed by 1,900 frames showing only a worker with a helmet. Frames with only a worker without a helmet are equally common, with 2,100 instances. Combinations of objects are also well-represented: 1,600 frames feature both an excavator and a worker with a helmet, while 2,100 frames capture an excavator alongside a worker without a helmet. Additionally, 200 frames depict scenarios where both a worker without a helmet and a worker with a helmet are present. This distribution emphasizes the dataset's diversity and its focus on scenarios involving interactions between workers and machinery, which are essential for safety monitoring applications.

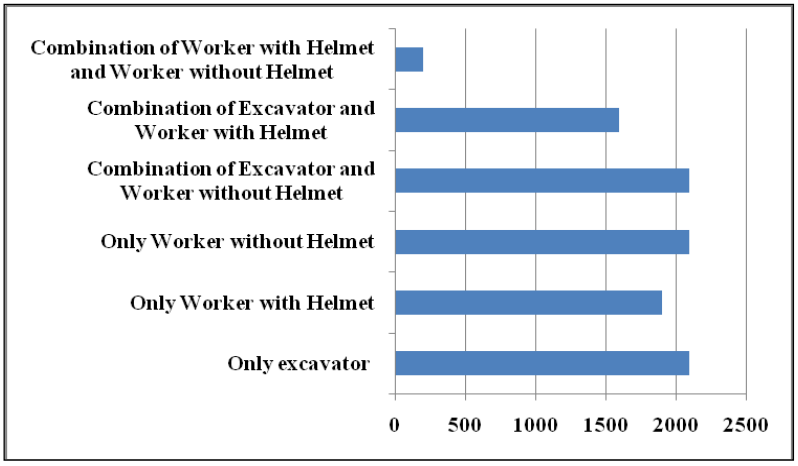

Figure 8. Distribution of Video Frames Across Different Instances

3.3.5.4 Dataset Statistics for Object Classes

The CSOD-24 dataset comprises four key object classes with a substantial number of annotated instances: Worker Without Helmet, Worker With Helmet, Excavator, and Dump Truck. A detailed breakdown reveals that the dataset includes 8,957 instances of Worker Without Helmet, 11,066 instances of Worker With Helmet, 6,066 instances of Excavators, and 1,003 instances of Dump Trucks. This distribution highlights a significant emphasis on worker-related annotations, which are essential for safety compliance monitoring. The dataset's diversity ensures robust training and evaluation of object detection models across varying construction site scenarios. Figure 9 presents a bar chart illustrating the distribution of these instances across the four classes.

Figure 9. Distribution of Object Classes in the Dataset

Here is the bar chart illustrating the distribution of object instances across the four classes in the CSOD-24 dataset. This visualization complements the paragraph discussing dataset statistics and provides a clear representation of the emphasis on worker-related annotations for safety compliance monitoring.

3.3.5.5 Key Characteristics of the Dataset

To provide a structured overview of the dataset, Table 3 summarizes its key characteristics, including the number of videos and frames, resolution, annotation formats, and

the environmental conditions considered during data collection. The dataset is designed to support helmet detection and worker-excavator proximity monitoring in construction environments.

Table 3. Key Characteristics of the CSOD-24 Dataset

Feature	Description
Dataset Name	CSOD-24
Total Videos	100 ten-second video clips
Total Frames	10,000 frames
Frame Rate.	10 frames per second (fps)
Resolution	640 × 480 pixels (resized from 1920 × 1080)
Number of Classes	4 (Worker without Helmet, Worker with Helmet, Excavator, Dump Truck)
Annotation Formats	.txt (YOLO), .json (COCO), .xml (Pascal VOC)
Recording Devices	Samsung Galaxy A22 (48MP), Panasonic Handicam
Video Collection Locations	Solapur (small construction sites), Pune (large construction projects)
Environmental Variability	Captured under different lighting and weather conditions
Dataset Purpose	Helmet detection and worker-equipment proximity monitoring

The annotations for the dataset were carried out solely by the author, ensuring consistency and accuracy across all 10,000 images. A systematic approach was followed, adhering to predefined labeling guidelines for the four classes: Worker Without Helmet, Worker With Helmet, Excavator, and Dump Truck. To maintain high-quality annotations, periodic self-review sessions were conducted, where a subset of labeled images was re-evaluated to identify and correct any discrepancies. This comprehensive process helped ensure uniformity in the dataset and minimized annotation errors. The number of bounding boxes per image varied based on the scene complexity. On average, each image contained 2 to 5 bounding boxes, with some images featuring a single object and others capturing multiple workers and machinery in dynamic construction site settings.

3.3.5.6 Comparison with Existing Datasets

The CSOD-24 dataset addresses important challenges in existing construction safety datasets by providing a well-annotated, video-based collection of dynamic interactions between workers and machinery. Unlike most existing datasets that rely on static images or synthetic data, CSOD-24 captures real-world scenarios under diverse environmental conditions such as lighting, weather, and occlusions, ensuring better generalizability. With 10,000 annotated video frames across four key classes: workers with helmet, workers without helmet, excavators, and dump trucks, CSOD-24 enables real-time monitoring of safety compliance and proximity risks. Its annotations in multiple formats (.txt, .json, and .xml) make it versatile for various machine learning frameworks. Furthermore, the dataset focuses on dynamic temporal data, making it highly suitable for training and evaluating models for automated safety monitoring, a domain often overlooked in existing datasets. Table 4 provides a comparative analysis between the CSOD-24 dataset and existing datasets used for construction site safety monitoring. It highlights the limitations of the existing datasets, such as lack of diversity, focus on static images, or specific safety elements, and contrasts them with the strengths of CSOD-24. By addressing these limitations, CSOD-24 stands out as a comprehensive and robust dataset customized for real-time safety compliance and worker-machine interaction analysis.

Table 4. Comparison of CSOD-24 Dataset with Existing Datasets

Dataset	Limitations of Existing Dataset	Strengths of CSOD-24
Synthetic Construction Dataset [10]	Synthetic data lacks real-world variability, reducing generalizability.	Real-world data captured under diverse conditions, ensuring robust model performance.
Manually Classified Construction Site Dataset [9]	Limited to static images from a single site; lacks diversity in conditions and object interactions.	Video-based dataset with 10,000 frames collected from multiple sites, ensuring diversity.

MOCS Dataset [12]	Limited to static images	Incorporates dynamic video data from Indian construction sites
Small Object Detection Dataset [8]	Struggles with extremely small objects and lacks scenario diversity.	Focuses on medium-sized objects and captures diverse real-world scenarios.

In future work, the CSOD-24 dataset will serve as a valuable resource for implementing a wide range of deep learning [17],[18],[19],[24] and machine learning algorithms aimed at enhancing construction site safety monitoring. Currently, the dataset has been validated using the YOLOv8 algorithm, demonstrating its effectiveness in detecting and classifying key objects with high accuracy. The model was trained for 30 epochs over a period of 2.415 hours on a Tesla T4 GPU, achieving impressive performance metrics with an overall mAP@0.5 of 0.982 and mAP@0.5:0.95 of 0.794. Class-wise performance included Worker Without Helmet (0.689), Worker With Helmet (0.666), Excavator (0.892), and Dump Truck (0.928) for mAP@0.5:0.95. These results validate the robustness of CSOD-24 and establish a strong baseline for future algorithmic enhancements.

4. Conclusion

The CSOD-24 dataset represents a significant advancement in the domain of construction site safety monitoring and automated compliance systems. This research outlines the creation of a comprehensive annotated dataset comprising 10,000 video frames derived from 100 ten-second video clips, with labels provided in three widely used formats: .txt, .json, and .xml. These features ensure broad compatibility with various machine learning frameworks, facilitating the training of object detection, classification, and proximity analysis models. The dataset's focus on four important classes: "worker with helmet", "worker without helmet", "excavator" and "dump truck", directly addresses the core safety challenges prevalent in dynamic construction environments. A key contribution of this study is the detailed analysis of safety compliance, emphasising the differences between small and large construction projects. The dataset captures these variations, emphasizing the need for enhanced safety enforcement in smaller projects, where non-compliance is more prevalent. Additionally, the proportional distribution of frames and object scenarios, visualized through charts, emphasizes

the dataset's diversity and relevance for real-world applications. The research further discusses the robust data preprocessing and annotation workflow, ensuring high-quality data representation and usability. The inclusion of both static frames and video clips enhances the dataset's versatility, supporting temporal analysis for dynamic safety scenarios, such as proximity monitoring and equipment tracking. In the future, machine learning and deep learning algorithms can be applied to the CSOD-24 dataset for advanced object detection and proximity detection between workers and machinery, such as excavators. These applications will enable automated systems to identify unsafe behaviours, provide real-time alerts, and prevent potential accidents on construction sites. By focusing on these essential safety aspects, the dataset paves the way for innovative solutions aimed at reducing workplace hazards and improving operational efficiency. Future efforts may expand the dataset by incorporating additional classes, more complex scenarios, and diverse environmental conditions, thereby broadening its applicability across global construction settings. The CSOD-24 dataset serves as a foundation for developing intelligent safety monitoring systems, marking a step forward in utilizing machine learning and computer vision to transform safety practices in the construction industry. This work emphasizes the transformative potential of technology in creating safer and more efficient construction environments.

References

[1] Mingpu Wang, Gang Yao, Yang Yang, Yujia Sun, Meng Yan, Rui Deng, “Deep learning-based object detection for visible dust and prevention measures on construction sites”, Developments in the Built Environment, Volume 16, December 2023, 100245, ISSN 2666-1659, https://doi.org/10.1016/j.dibe.2023.100245.

[2] Shrigandhi, M. N., and S. R. Gengaje. "Systematic literature review on object detection methods at construction sites." In International Conference on Expert Clouds and Applications, pp. 709-724. Singapore: Springer Nature Singapore, 2022.

[3] Thalange, A.V., Shrigandhi, M.N., Konapure, R.R., Ankaskar, V.N. (2023). “Performance Analysis of American Sign Language Using Wavelet Transform and CNN”. In: Reddy, V.S., Prasad, V.K., Wang, J., Rao Dasari, N.M. (eds) Intelligent Systems and Sustainable Computing. ICISSC 2022. Smart Innovation, Systems and Technologies, vol 363. Springer, Singapore. https://doi.org/10.1007/978-981-99-4717-1_3.

[4] Weili Fang, Lieyun Ding, Botao Zhong, Peter E.D. Love, Hanbin Luo, "Automated detection of workers and heavy equipment on construction sites: A convolutional neural network approach", Advanced Engineering Informatics, Sciencedirect, Volume 37, 2018, Pages 139-149, ISSN 1474-0346, https://doi.org/10.1016/j.aei.2018.05.003.

[5] Seda Yeşilmen, Bahadır Tatar, "Efficiency of convolutional neural networks (CNN) based image classification for monitoring construction related activities: A case study on aggregate mining for concrete production", Case Studies in Construction Materials, Volume 17, December 2022, e01372, ISSN 2214-5095, https://doi.org/10.1016/j.cscm.2022.e01372.

[6] Bhakti A Paranjape, Apurva A Naik, "DATS_2022: A Versatile Indian Dataset for Object Detection in Unstructured Traffic Conditions", Data in Brief, Sciencedirect, Volume 43, 2022, 108470, ISSN 2352-3409, https://doi.org/10.1016/j.dib.2022.108470.

[7] Lo Jye-Hwang, Lin Lee-Kuo & Hung Chu-Chun. (2022), "Real-Time Personal Protective Equipment Compliance Detection Based on Deep Learning Algorithm", Sustainability 2023, 15, 391. https://doi.org/10.3390/su15010391.

[8] Siyeon Kim, Seok Hwan Hong, Hyodong Kim, Meesung Lee, Sungjoo Hwang, "Small object detection (SOD) system for comprehensive construction site safety monitoring", Automation in Construction, Sciencedirect, October 2023, 105103, ISSN 0926-5805, https://doi.org/10.1016/j.autcon.2023.105103.

[9] Alexandre Del Savio, Ana Luna, Daniel Cárdenas-Salas, Mónica Vergara, Gianella Urday, "Dataset of manually classified images obtained from a construction site", Data in Brief, Sciencedirect, Volume 42, June 2022, 108042, ISSN 2352-3409, https://doi.org/10.1016/j.dib.2022.108042.

[10] Ari Yair Barrera-Animas, Juan Manuel Davila Delgado, "Generating real-world-like labelled synthetic datasets for construction site applications", Automation in Construction, Sciencedirect, Volume 151, July 2023, 104850, ISSN 0926-5805, https://doi.org/10.1016/j.autcon.2023.104850.

[11] Seunghyeon Wang, Ikchul Eum, Sangkyun Park, Jaejun Kim, "A labelled dataset for rebar counting inspection on construction sites using unmanned aerial vehicles", Data in Brief,

Sciencedirect, Volume 55, August 2024, 110720, ISSN 2352-3409, https://doi.org/10.1016/j.dib.2024.110720.

[12] An Xuehui, Zhou Li, Liu Zuguang, Wang Chengzhi, Li Pengfei, Li Zhiwei, “Dataset and benchmark for detecting moving objects in construction sites”, Automation in Construction, Sciencedirect, Volume 122, 2021, 103482, ISSN 0926-5805, https://doi.org/10.1016/j.autcon.2020.103482.

[13] Weili Fang, Lieyun Ding, Hanbin Luo, Peter E.D. Love, “Falls from heights: A computer vision-based approach for safety harness detection”, Automation in Construction, Sciencedirect, Volume 91, 2018, Pages 53-61, ISSN 0926-5805, https://doi.org/10.1016/j.autcon.2018.02.018.

[14] Weili Fang, Botao Zhong, Neng Zhao, Peter E.D. Love, Hanbin Luo, Jiayue Xue, Shuangjie Xu, “A deep learning-based approach for mitigating falls from height with computer vision: Convolutional neural network”, Advanced Engineering Informatics, Sciencedirect, Volume 39, 2019, Pages 170-177, ISSN 1474-0346, https://doi.org/10.1016/j.aei.2018.12.005.

[15] Xiyu Wang , Nora El-Gohary, “Few-shot object detection and attribute recognition from construction site images for improved field compliance”, Automation in Construction, Sciencedirect, August 2024, ISSN 0926-5805, https://doi.org/10.1016/j.autcon.2024.105539.

[16] Xu, Jiayi, and Wei Pan. "Deep learning-based object detection for dynamic construction site management." Automation in Construction 165 (2024): 105494.

[17] Li Liu, Wanli Ouyang, Xiaogang Wang, Paul Fieguth, Jie Chen, Xinwang Liu,• Matti Pietikäinen, “Deep Learning for Generic Object Detection: A Survey”, International Journal of Computer Vision, 128, 261–318 (2020). https://doi.org/10.1007/s11263-019-01247-4.

[18] Jixiu Wu, Nian Cai, Wenjie Chen, Huiheng Wang, Guotian Wang, “Automatic detection of hardhats worn by construction personnel: A deep learning approach and benchmark dataset”, Automation in Construction, Sciencedirect, Volume 106, 2019, 102894, ISSN 0926-5805, https://doi.org/10.1016/j.autcon.2019.102894.

[19] Zdenek Kolar, Hainan Chen, Xiaowei Luo, “Transfer learning and deep convolutional neural networks for safety guardrail detection in 2D images”, Automation in Construction, Sciencedirect, Volume 89, May 2018, Pages 58-70, ISSN 0926-5805, https://doi.org/10.1016/j.autcon.2018.01.003.

[20] Hyojoo Son, Hyunchul Choi, Hyeonwoo Seong, Changwan Kim, “Detection of construction workers under varying poses and changing background in image sequences via very deep residual networks”, Automation in Construction, Sciencedirect, Volume 99, March 2019, Pages 27-38, ISSN 0926-5805, https://doi.org/10.1016/j.autcon.2018.11.033.

[21] Wang Z, Wu Y, Yang L, Thirunavukarasu A, Evison C, Zhao Y. “Fast Personal Protective Equipment Detection for Real Construction Sites Using Deep Learning Approaches”. Sensors. 2021; 21(10):3478. https://doi.org/10.3390/s21103478.

[22] Ferdous M, Ahsan SMM. 2022. “PPE detector: a YOLO-based architecture to detect personal protective equipment (PPE) for construction sites”. PeerJ Computer Science 8:e999 https://doi.org/10.7717/peerj-cs.999.

[23] Wei Yang, Guang-Le Zhou, Zhi-Wei Gu, Xiao-Dan Jiang and Zhe-Ming Lu. “Safety Helmet Wearing Detection Based On An Improved Yolov3 Scheme.” , International Journal of Innovative Computing, Information and Control, Volume 18, Number 3, ISSN 1349-4198, June 2022.

[24] Jinwoo Kim, Jeongbin Hwang, Seokho Chi, JoonOh Seo, “Towards database-free vision-based monitoring on construction sites: A deep active learning approach”, Automation in Construction, Volume 120, 2020, 103376, ISSN 0926-5805, https://doi.org/10.1016/j.autcon.2020.103376.

[25] Haosen Chen, Lei Hou, Guomin (Kevin) Zhang, Shaoze Wu, “Using Context-Guided data Augmentation, lightweight CNN, and proximity detection techniques to improve site safety monitoring under occlusion conditions”, Safety Science, Volume 158, 2023, 105958, ISSN 0925-7535, https://doi.org/10.1016/j.ssci.2022.105958.

[26] Yanman Li, Jun Zhang, Yang Hu, Yingnan Zhao, and Yi Cao, “Real-time Safety Helmet-wearing Detection Based on Improved YOLOv5”, Computer Systems Science & Engineering, 2022, DOI: 10.32604/csse.2022.028224.

[27] Jiaqi Li, Xuefeng Zhao, Guangyi Zhou, Mingyuan Zhang, “Standardized use inspection of workers' personal protective equipment based on deep learning”, Safety Science, Volume 150, 2022, 105689, ISSN 0925-7535, https://doi.org/10.1016/j.ssci.2022.105689.

[28] An Q, Xu Y, Yu J, Tang M, Liu T, Xu F. “Research on Safety Helmet Detection Algorithm Based on Improved YOLOv5s”, Sensors. 2023; 23(13):5824. https://doi.org/10.3390/s23135824.

[29] Yong YP, Lee SJ, Chang YH, Lee KH, Kwon SW, Cho CS, Chung SW, “Object Detection and Distance Measurement Algorithm for Collision Avoidance of Precast Concrete Installation during Crane Lifting Process”, Buildings. 2023; 13(10):2551. https://doi.org/10.3390/buildings13102551.

[30] Seong J, Kim H-s, Jung H-J, “The Detection System for a Danger State of a Collision between Construction Equipment and Workers Using Fixed CCTV on Construction Sites”, Sensors. 2023; 23(20):8371. https://doi.org/10.3390/s23208371.

[31] Jiaqi Li, Qi Miao, Zheng Zou, Huaguo Gao, Lixiao Zhang, Zhaobo Li, "A Review of Computer Vision-Based Monitoring Approaches for Construction Workers’ Work-Related Behaviors," in IEEE Access, vol. 12, pp. 7134-7155, 2024, doi: 10.1109/ACCESS.2024.3350773.

[32] Shuai Tang, Dominic Roberts, Mani Golparvar-Fard, “Human-object interaction recognition for automatic construction site safety inspection”, Automation in Construction, Volume 120, 2020, 103356, ISSN 0926-5805, https://doi.org/10.1016/j.autcon.2020.103356.