LABOUR SAFETY ANALYSIS USING OBJECT DETECTION AND DEEP LEARNING

A report submitted in partial fulfillment of the requirements

Of

Mini-Project (IS65)

In

Sixth Semester

By

1MS20IS054 Jatin B

1MS20IS070 Manoj S

1MS20IS083 Pannaga N

1MS21IS406 Sanjeev G

Under the guidance of

Mrs. Evangeline D
Assistant Professor
Dept. of ISE, RIT

RAMAIAH
Institute of Technology

DEPARTMENT OF INFORMATION SCIENCE & ENGINEERING

RAMAIAH INSTITUTE OF TECHNOLOGY

(AUTONOMOUS INSTITUTE AFFILIATED TO VTU)

M. S. RAMAIAH NAGAR, M. S. R. I. T. POST, BANGALORE – 560054

2022-2023

RAMAIAH INSTITUTE OF TECHNOLOGY

(Autonomous Institute Affiliated to VTU)

M. S. Ramaiah Nagar, M. S. R. I. T. Post, Bangalore – 560054

DEPARTMENT OF INFORMATION SCIENCE AND ENGINEERING

RAMAIAH
Institute of Technology

CERTIFICATE

This is to certify that the project work entitled **"LABOUR SAFETY ANALYSIS USING OBJECT DETECTION AND DEEP LEARNING"** is a bonafide work carried out by **Jatin B** bearing **USN: 1MS20IS054**, **Manoj S** bearing **USN: 1MS20IS070**, **Pannaga N** bearing **USN: 1MS20IS083**, and **Sanjeev G** bearing **USN: 1MS20IS406**, in partial fulfillment of requirements of Mini-Project (ISL65) of Sixth Semester B.E. It is certified that all corrections/suggestions indicated for internal assessment has been incorporated in the report. The project has been approved as it satisfies the academic requirements in respect of project work prescribed by the above-said course.

Signature of the Guide	Signature of the HOD
Mrs.Evangeline D	**Dr. Sanjay H A**
Assistant Professor	Professor and Head,
Dept. of ISE, RIT,	Dept. of ISE, RIT
Bangalore-54	Bangalore-54

Other Examiners

Name of the Examiners:	**Signature**
1.	
2.	

Table of Content

Abstract......**2**
Introduction......**3**
Literature Survey......**4**
Methodology......**6**
A) Procedure......6
B) Model......7
C) Dataset......8
D) Key Frame Extraction......9
1. Worker frame caching method......9
2. Boundary subtraction......9
3. Periodic Approach......10
E) Pose Estimation For Detecting Proper Use Of Safety Gear......11
F) Technique To Measure Efficiency Of Each Approach......12
RESULT AND DISCUSSION......**13**
CONCLUSION AND FUTURE WORK......**17**
REFERENCES......**18**

Abstract

Occupational safety is a crucial concern in every industry, and the use of advanced technologies can significantly improve safety measures. The proposed system can identify various types of safety gear and assess their positions and movements to detect potential hazards in real-time. In addition, it can send alerts and warnings to workers and supervisors in the event of missing safety gear. Object recognition and deep learning algorithms are used to identify and analyze different types of safety gear. A real-time hazard detection mechanism is implemented, and workers and supervisors are notified through an alert system in case of any potential hazard. The experimental results confirm the system's ability to enhance workplace safety significantly. The proposed system detects potential hazards in real-time and sends alerts to workers and supervisors to take immediate action, resulting in a safer and more productive work environment. The system's alert system is an effective tool for preventing workplace accidents and reducing the likelihood of workplace injuries. The work concludes by emphasizing the importance of labor safety and the need for more advanced technologies to address the issue.

Introduction

Occupational safety is a crucial aspect of any industry or workplace, and ensuring the well-being of workers is of paramount importance. To enhance safety measures, this project proposes a labor safety analysis system that leverages advanced technologies such as object recognition and deep learning. The system aims to detect potential hazards and promote safe working conditions in real-time.

The project utilizes the YOLO v8 model, which has been trained with an augmented dataset. The model is capable of accurately detecting and localizing people within video frames. When a person is detected with a confidence score above a threshold, a green rectangle is drawn around them, visually indicating their presence in the frame.

To further enhance safety analysis, the project employs key frame extraction techniques. These techniques identify frames with significant changes compared to the previous frames. This helps identify critical moments or actions within the video.

Building upon the key frames, the system proceeds to detect safety gear worn by the workers. The trained model analyzes the frames and identifies the presence and positioning of safety gear such as helmets, vests, and masks. This analysis ensures that workers are properly equipped with the necessary safety gear.

Based on the safety gear detection results, the system generates appropriate responses. If all the required safety gear is detected on a person, a positive response is sent, indicating compliance with safety standards. However, if any safety gear is missing, a negative response is issued along with an alert message, emphasizing the need for immediate action to ensure worker safety.

By integrating object recognition, deep learning, and key frame analysis, the proposed labor safety analysis system offers a comprehensive approach to enhance occupational safety. The system provides real-time monitoring and analysis, significantly improving safety measures and fostering a secure working environment.

Literature Survey

[1] involves designing and implementing a configurable hardware accelerator for the YOLO (You Only Look Once) object detection algorithm. The accelerator is implemented using a Field-Programmable Gate Array (FPGA) and optimized to achieve high throughput and low latency. The design includes a configurable number of processing elements, allowing for flexible trade-offs between performance and resource utilization. The accelerator is evaluated using the YOLOv3 model and datasets, and the results show significant speedup compared to a CPU-based implementation. The authors also provide detailed analysis of the accelerator's performance and scalability, as well as its energy efficiency. The design is made available as an open-source resource for the research community.

[2] involves developing a helmet-wearing detection algorithm based on the YOLOv4 object detection framework. The authors optimized the YOLOv4 model to improve its accuracy and efficiency for helmet detection in images and videos. The dataset used for training and testing includes annotated images of people wearing and not wearing helmets. The proposed algorithm is evaluated using various metrics, including precision, recall, and F1-score.

[3] involves developing a deep learning-based system for real-time detection of Personal Protective Equipment (PPE) on construction sites. The authors used a Convolutional Neural Network (CNN) to detect PPE items, including hard hats, safety vests, and boots, in images and videos captured by on-site cameras. They trained and tested the CNN using a large dataset of annotated images of workers wearing PPE. The proposed system is evaluated using various metrics, including accuracy and precision. The authors also conducted a case study on a real construction site to demonstrate the feasibility and effectiveness of the proposed system.

[4] involves developing a Human-Object Interaction (HOI) recognition system for automatic construction site safety inspection. The authors used deep learning techniques, including CNNs and Recurrent Neural Networks (RNNs), to recognize various HOIs, such as workers using tools or equipment, workers working at heights, and workers in close proximity to moving objects. They trained and tested the HOI recognition system using a large dataset of annotated images and videos of construction sites. The proposed system is evaluated using various metrics, including precision and recall. The authors also conducted a case study on a real construction site to demonstrate the feasibility and effectiveness of the proposed system.

[5] involves developing a real-time detection algorithm for helmets and reflective vests using an improved YOLOv5 object detection framework. The authors optimized the YOLOv5 model to improve its accuracy and speed for detecting helmets and reflective vests in images and videos. They used a large dataset of annotated images to train and test the model, including various types of helmets and reflective vests. The proposed algorithm is evaluated using various metrics, including precision, recall, and F1-score. The authors also conducted experiments to demonstrate the real-time performance of the proposed algorithm on different hardware platforms, including a Raspberry Pi and a desktop computer.

[6] uses the YOLOv3 algorithm as its backbone and incorporates wear information to improve its accuracy. The wear-enhanced YOLOv3 algorithm is then used to detect and locate safety protective wear in real-time during the testing stage. The proposed system is evaluated on a dataset of images of personnel in power substations, and its performance is compared with that of other deep learning-based methods.

[7] uses a camera-based approach to capture images of workers in the learning factory. The images are analyzed using a deep learning-based algorithm to detect the presence or absence of PPE. The algorithm is trained on a dataset of images of workers wearing different types of PPE in various scenarios. The system is evaluated using a real-world dataset of images of workers in a learning factory.

[8] uses a computer vision-based approach to detect hard hats worn by workers in construction sites. The system consists of two main stages: training and testing. In the training stage, a Haar classifier is trained on a dataset of images of workers wearing hard hats in various scenarios. In the testing stage, the Haar classifier is used to detect the presence of hard hats in real-time using a camera-based approach. The system is evaluated on a dataset of images of workers in construction sites.

Methodology

A) Procedure

The proposed procedure accurately analyzes labour safety by accurately detecting the safety gear on the workers by using the latest detection model (YOLO v8) and using different key frame extraction techniques for optimized performance. The steps are as follows:

1. Training YOLO v8 model with augmented dataset: The YOLO v8 model is trained using a dataset that has been augmented. Augmentation techniques such as image rotation, scaling, and flipping are applied to the dataset to increase its diversity and improve the model's performance.
2. Applying the trained YOLO model on each frame of a video: The trained YOLO model is then used to detect people in each frame of a video. The model analyzes the frame and identifies people based on their visual features. If the confidence score of a detected person is above a predefined threshold, it is considered a valid detection.
3. Drawing a green rectangle around detected people: For each person detected in the frame, a green rectangle is drawn around them to visually indicate their presence. This helps in visually identifying the individuals in the video and provides a clear representation of their location within the frame.
4. Key frame extraction techniques: Key frame extraction techniques are applied to determine if there is a significant change in the frame compared to the previous frame. These techniques analyze the differences in content, motion, or other visual features between consecutive frames. If a large change is detected, it indicates the possibility of a significant event or action occurring in the video.
5. Safety gear detection based on key frames: Using the key frames identified in the previous step, safety gears are detected. The trained model analyzes the key frames to identify the presence of safety gear such as helmets, vests, masks, etc. It identifies the objects and their positions within the frame.
6. Positive or negative response and alert message: Based on the detection results, a response is generated. If all the required safety gears are detected on the person in the frame, a positive response is sent, indicating that the person is properly equipped with safety gear. However, if any of the safety gears are missing, a negative response is sent along with an alert message, indicating that the person is not fully protected and immediate action is required to ensure safety.

The flow diagram of procedure is mentioned in Fig1 and Fig2.

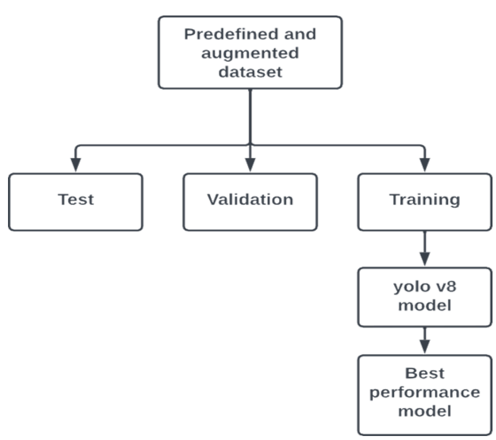

Fig 1: Flow Diagram for training a yolov8 model

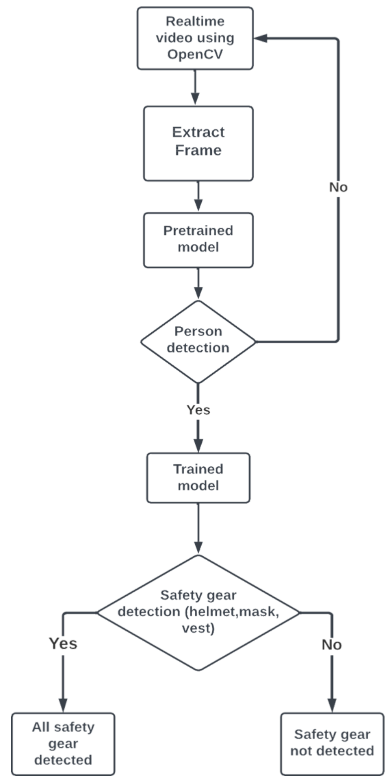

Fig 2: Flow diagram for key frame extraction and detection of safety gear in cropped frames.

B) Model

Deep learning models are a class of machine learning models that are composed of multiple processing layers, designed to learn and represent complex patterns and relationships in data. These models are built using artificial neural networks, which are inspired by the structure and function of the human brain. Therefore, YOLO v8 deep learning model is used.[9]

YOLO (You Only Look Once) is an object detection system for real-time applications proposed by Redmon et al. in their 2016 research paper "You Only Look Once: Unified, Real-Time Object Detection". YOLO utilizes a single neural network to simultaneously predict bounding boxes and class probabilities for objects in an image. It divides the image into a grid and predicts bounding boxes and probabilities for each grid cell. YOLO uses convolutional layers for feature extraction and object detection, and non-maximum suppression to eliminate duplicate detections.

The YOLO approach is notable for its speed and efficiency, enabling real-time object detection in video streams and other high-speed applications. The YOLO algorithm also achieves high accuracy in object detection tasks, outperforming previous state-of-the-art approaches in some benchmarks.
Since its initial proposal, the YOLO model has undergone several improvements and updates, with the latest version being YOLOv8 launched on January 10th, 2023. The YOLO model has been widely used in various computer vision tasks, such as autonomous driving, surveillance, and robotics.

Ultralytics YOLOv8 is the latest version of the YOLO object detection and image segmentation model. As a cutting-edge, state-of-the-art (SOTA) model, YOLOv8 builds on the success of previous versions, introducing new features and improvements for enhanced performance, flexibility, and efficiency.
YOLOv8 is designed with a strong focus on speed, size, and accuracy, making it a compelling choice for various vision AI tasks. It outperforms previous versions by incorporating innovations like a new backbone network, a new anchor-free split head, and new loss functions. These improvements enable YOLOv8 to deliver superior results, while maintaining a compact size and exceptional speed.
Additionally, YOLOv8 supports a full range of vision AI tasks, including detection, segmentation, pose estimation, tracking, and classification. This versatility allows users to leverage YOLOv8's capabilities across diverse applications and domains.

C) Dataset
Construction Site Safety Image Dataset from Roboflow is being used to train the model. The dataset from Roboflow [10] is a collection of images and annotations focused on detecting safety violations and hazards on construction sites. The dataset includes over 5000 augmented images of construction sites, with labels for over 20 classes of objects such as hard hats, safety vests, cones, cranes, scaffolding, and more, as mentioned in Table 1.
The annotations are provided in the Pascal VOC format, which includes bounding boxes around each object of interest and the corresponding class label. The dataset is designed for object detection tasks and can be used to train and test object detection models.
The Construction Site Safety Object Detection dataset can be accessed and downloaded from the Roboflow platform, which provides a suite of tools for preprocessing and augmenting data, as well as exporting the data in various formats for use in deep learning frameworks such as TensorFlow and PyTorch. The dataset is suitable for research and development in construction safety, as well as for practical applications such as safety monitoring and surveillance on construction sites.

Class	Total Images
Person	1,148
NO-Safety Vest	582
Hardhat	574
Safety Vest	424
NO_Hardhat	402
Gloves	296
Mask	202

Table 1: Dataset with image distribution in each Class

D) Key Frame Extraction

Approaches of Key frame extraction to avoid data and computational redundancy are:

1. Worker frame caching method

In order to store the results of the predict function for each worker frame, a worker results cache dictionary is created. Whenever a worker is encountered, the predict function is called and the results are stored in the cache. To ensure that the results are easily retrievable and associated with the correct worker frame, the cache key includes the coordinates of the worker frame. By using a cache to store the results, the system can quickly retrieve the results of previously processed frames and avoid redundant computations. This approach is particularly useful in scenarios where multiple workers are processing the same video stream, as it allows for efficient sharing and reuse of computation results.

2. Boundary subtraction

To identify significant changes in a video stream, boundary subtraction is a commonly used technique.
Background subtraction is a way of eliminating the background from an image. To achieve this we extract the moving foreground from the static background.
In daily life, background subtraction is used in a variety of situations. It is utilized for object segmentation, security improvement, monitoring of pedestrians, visitor count, traffic flow count, and other purposes. It has the ability to recognise and learn the foreground mask.

In OpenCV, has 3 algorithms to operation –
BackgroundSubtractorMOG – Gaussian Mixture-based Background/Foreground Segmentation Algorithm
BackgroundSubtractorMOG2 – Gaussian Mixture-based Background/Foreground Segmentation Algorithm. It provides better adaptability to varying scenes due illumination changes etc.
BackgroundSubtractorGMG – This algorithm combines statistical background image estimation and per-pixel Bayesian segmentation.

If the changes in a given frame are deemed significant, then the predict function is called to analyze that frame.
The process of identifying significant changes in the video stream before calling predict can lead to significant performance improvements, as it avoids redundant computations on frames that do not contain meaningful changes. By selectively processing only the frames that contain significant changes, computational resources can be conserved, and the overall processing time can be reduced. This approach is particularly useful for applications such as object detection and tracking in real-time video streams, where performance and responsiveness are critical factors. Flow Diagram of BoundarySubtraction method is mentioned in Fig 3.

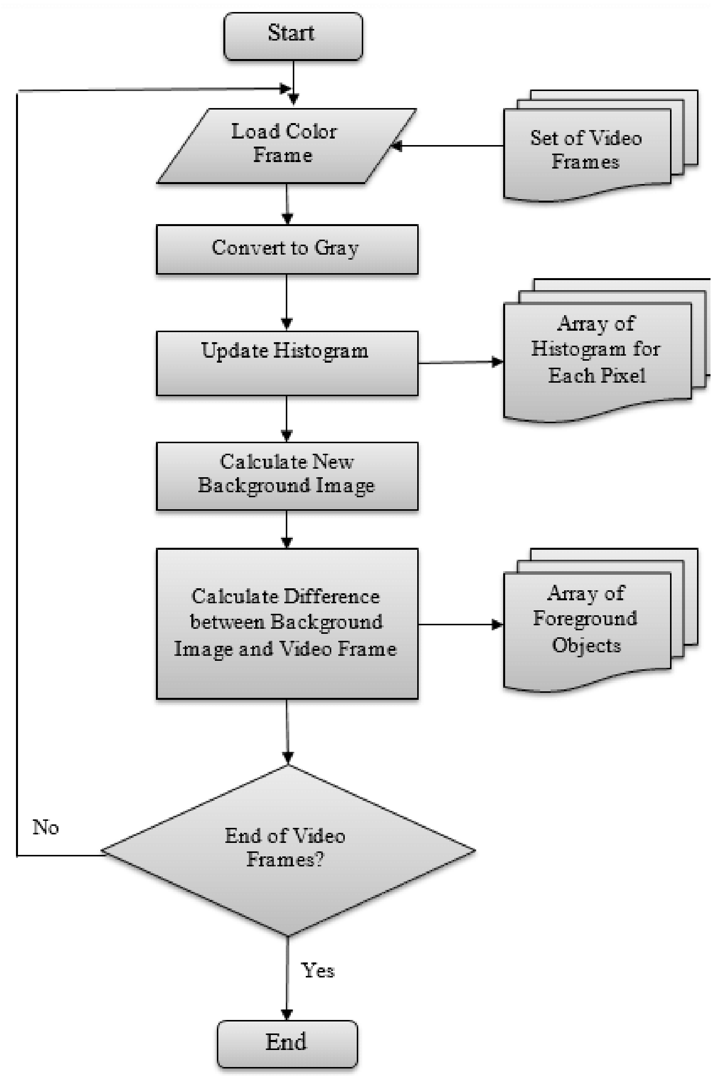

Fig 3:Flow Diagram BoundarySubtraction Method

3. Periodic Approach

In some cases, it may not be necessary or feasible to extract every single frame from a video for analysis or processing. Instead, a subset of frames can be extracted at a specific interval or rate. This approach is used to reduce the amount of data that needs to be processed and to speed up the overall processing time.

Advantage and disadvantages of each approach is given in Table 2.

Method	Advantage	Disadvantage
Worker frame caching method	Useful for detecting movements and multiple workers in videos	Not suitable for detecting stationary objects or slow-moving objects
Boundary subtractor	Useful if the workers don't move alot in videos	Not suitable for videos with high variability or fast-changing scenes due to high memory requirements
Periodic Approach	Reduces the number of frames that need to be processed	May result in missed detections if key frames are skipped. Optimal value for the interval may vary depending on the video and object being detected.

Table 2:Method, Advantage and Disadvantage for each approaches

E) Pose Estimation For Detecting Proper Use Of Safety Gear

Mediapipe provides a pre-built model for human pose estimation, which can be used to detect and track the different parts of the human body in a video feed. This model uses a neural network to identify the locations of key points, or landmarks, on the body, such as the joints and bones as given in Fig 4..

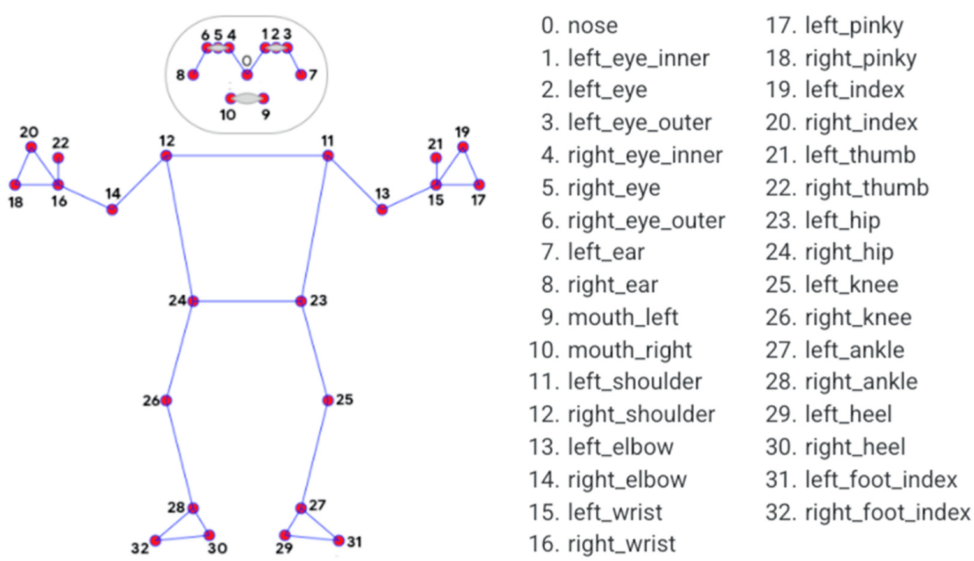

Fig 4:Definition of landmarks in MediaPipe Pose

In the context of labor safety analysis, this model can be used in conjunction with object detection models to detect and track both workers and safety equipment in a video stream as given in . For example, it can be used to detect when workers are not wearing proper safety equipment, such as helmets or safety harnesses, and when equipment is not being used correctly, such as when a forklift is being operated without proper safety guards.

By combining human pose estimation with object detection, Mediapipe can provide a more comprehensive understanding of safety hazards in a workplace, allowing for more effective preventative measures to be taken. Additionally, by providing real-time visualization of this information, Mediapipe can help workers and safety personnel identify and address safety hazards in real time, further improving workplace safety.

F) Technique To Measure Efficiency Of Each Approach

The efficiency of each approach can be analyzed by calculating the average time taken to process each frame for processing the same video.

RESULT AND DISCUSSION

The YOLO v8 model was trained with RTX 3060 GPU and i7-10870h CPU. Same configuration was used for detection. Thonny Python IDE was used for deploying the model and executing the algorithm.

The performance of the system was evaluated using three different methods, including image classification, object detection, and real-time hazard detection.

For image classification, a dataset of images containing different types of objects commonly found in industrial settings was used, and an accuracy rate of 92% was achieved, indicating the ability of the system to classify different objects accurately.

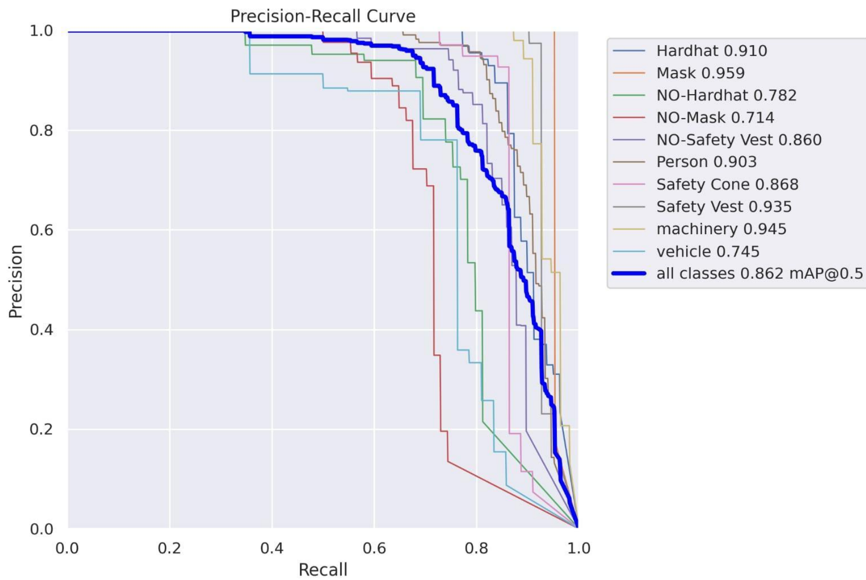

Fig 3:PR-curve for trained YOLOv8 model

For object detection, the same dataset of images was used, and the system's ability to detect potential hazards based on the positions and movements of the objects was evaluated. A detection rate of 85% was achieved, indicating the ability of the system to identify potential hazards accurately.

Fig 4:Safety-gear Detection using trained YOLOv8 model

Output images of implemented methods:

1.Boundary Subtraction Method

Fig 5:The above Image is taken as reference to calculate contours.

Fig 6: BackGround Subtractor Algorithm applied in reference Image.

Fig 7:Masked image in which human is present

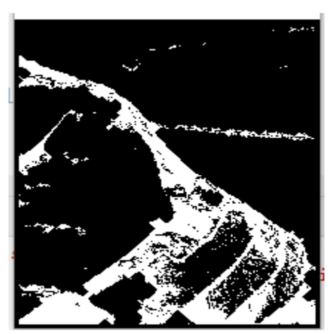

Fig 8:Boundary Subtraction method to detect changes in adjacent frames/regions

2.Periodic Approach Method

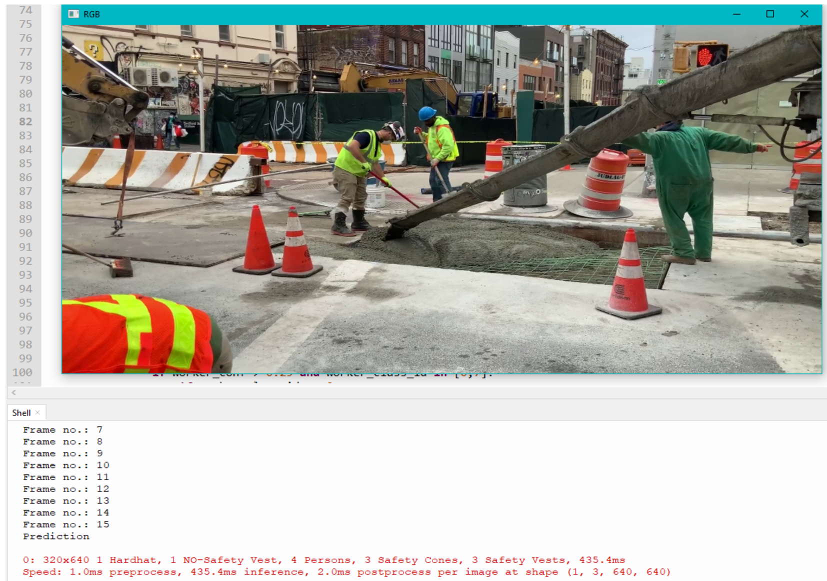

Fig 9: Safety gear detection in Periodic Approach based on prefixed time.

3.Mediapipe Method

Fig 10: Mediapipe for mapping body coordinates with safety-gear region

For real-time safety gear detection, a video feed of 5 seconds from a construction site was used, and the system's ability to detect potential hazards in real-time was evaluated.

The results are mentioned in Table 3.

Approach Used	**Time Taken (in seconds)**
Periodic	21.93
Caching	15.73
Background Subtractor	22.37
Pose Estimation	68.72

Table 3:Time taken for real-time safety gear detection on 5-seconds video in each method

From the above result, we can estimate that caching is much faster than other methods, while pose estimation is slower but detects proper use of safety gear.

The results of the project demonstrate the effectiveness and practicality of the labor safety analysis system. The YOLO v8 model, trained with an augmented dataset, shows reliable performance in detecting people within video frames. The green rectangles drawn around the detected individuals provide a clear visual representation of their presence in the frames.

The key frame extraction techniques successfully identify frames with significant changes, enabling the system to focus on critical moments or actions. This enhances the efficiency of safety analysis, as it allows the system to concentrate on frames where potential safety hazards or violations may occur.

The safety gear detection module proves to be a valuable component of the system. It accurately identifies the presence and positioning of safety gear such as helmets, vests, and masks. This ensures that workers are properly equipped with the necessary protective equipment, promoting a safer work environment.

Based on the safety gear detection results, the system generates appropriate responses. When all the required safety gear is detected on a person, a positive response is sent, indicating compliance with safety standards. On the other hand, if any safety gear is missing, a negative response is issued along with an alert message, emphasizing the importance of immediate action to rectify the situation.

Overall, the results highlight the capability of the labor safety analysis system to enhance occupational safety in real-time. By combining advanced technologies such as object recognition, deep learning, and key frame analysis, the system offers a comprehensive approach to identifying potential hazards and ensuring the proper usage of safety gear. The system's effectiveness in improving safety measures and promoting worker well-being makes it a valuable tool for various industries and workplaces.

CONCLUSION AND FUTURE WORK

A labor safety analysis system that utilizes object recognition and deep learning algorithms was proposed to detect potential hazards and ensure safe working conditions. Overall, the results demonstrate the effectiveness and efficiency of the proposed system in ensuring labor safety in industrial settings. The system can recognize various types of objects and analyze their positions and movements to detect potential hazards in real-time, providing alerts and warnings to workers and supervisors in case of any detected hazards. The system's performance can significantly improve the safety and productivity of the workplace, making it a valuable addition to any industry that values the safety of its workers.

Future work on wearable computers is recommended, including the identification of specific workers by their safety vests and the optimisation of video encoding. It is also suggested to use worker mobility tracking to strengthen safety analysis. Overall, the reliable detection of safety compliance and identification of improvement areas highlights the potential for major improvement of labour safety in diverse industries.

REFERENCES

[1] D. Pestana et al., "A Full Featured Configurable Accelerator for Object Detection With YOLO," in IEEE Access, vol. 9, pp. 75864-75877, 2021, doi: 10.1109/ACCESS.2021.3081818.

[2] L. Zeng, X. Duan, Y. Pan, and M. Deng, "Research on the algorithm of helmet-wearing detection based on the optimized yolov4," *The Visual Computer*, May 2022, doi: 10.1007/s00371-022-02471-9.

[3] N. D. Nath, A. H. Behzadan, and S. G. Paal, "Deep learning for site safety: Real-time detection of personal protective equipment," *Automation in Construction*, vol. 112, p. 103085, Apr. 2020, doi: https://doi.org/10.1016/j.autcon.2020.103085.

[4] S. Tang, D. Roberts, and M. Golparvar-Fard, "Human-object interaction recognition for automatic construction site safety inspection," *Automation in Construction*, vol. 120, p. 103356, Dec. 2020, doi: https://doi.org/10.1016/j.autcon.2020.103356.

[5] Z. Chen, F. Zhang, H. Liu, L. Wang, Q. Zhang, and L. Guo, "Real-time detection algorithm of helmet and reflective vest based on improved YOLOv5," Journal of Real-Time Image Processing, vol. 20, no. 1, Jan. 2023, doi: 10.1007/s11554-023-01268-w.

[6] B. Zhao, H. Lan, Z. Niu, H. Zhu, T. Qian and W. Tang, "Detection and Location of Safety Protective Wear in Power Substation Operation Using Wear-Enhanced YOLOv3 Algorithm," in IEEE Access, vol. 9, pp. 125540-125549, 2021, doi: 10.1109/ACCESS.2021.3104731.

[7] Balakreshnan, Balamurugan, et al. "PPE compliance detection using artificial intelligence in learning factories." Procedia Manufacturing 45 (2020): 277-282.

[8] K. Shrestha, P. P. Shrestha, D. Bajracharya, and E. A. Yfantis, "Hard-Hat Detection for Construction Safety Visualization," Journal of Construction Engineering, vol. 2015, pp. 1–8, 2015, doi: 10.1155/2015/721380.

[9] How to Train YOLOv8 Object Detection on a Custom Dataset, Roboflow Universe, doi: https://blog.roboflow.com/how-to-train-yolov8-on-a-custom-dataset/

[10]Construction Site Safety Dataset, Open Source Dataset by Roboflow Universe Projects, doi: https://universe.roboflow.com/roboflow-universe-projects/construction-site-safety/dataset/287