

مهارات بحثية 2

Prepared by Younis Arrabi



مهارات بحثية 2 التكليف العاشر

العنوان

Real-time Object detection of Personal Protective Equipment in video Streaming Using Deep Learning Technique



معايير التقييم

التكليف العاشر

اختر بحثا في مجال تخصصك منتقدا فيه دراساته المرجعية وذلك من خلال عرض فيديو لمدة خمس دقائق ، ومرفقا به محتوى العرض في ملف PDF

1. وضوح واختيار الدراسات السابقة
2. حداثة الدراسات
3. تنوع المناهج في الدراسات السابقة
4. العينة والبيانات في الدراسات السابقة
5. النتائج والتوصيات في الدراسات المرجعية

01

وضوح واختيار الدراسات السابقة

الرسالة استعانت بعدد من الدراسات التي تناولت خوارزميات الرؤية الحاسوبية، وخاصة سلاسل YOLO و Faster R- CNN
و كذا تم ملاحظة وضوح الربط بين موضوع الرسالة وهذه الدراسات و التركيز الكبير على دراسات خوارزميات YOLO

02

حادثة الدراسات

معظم المراجع من الفترة (2018-2023)، وهذا يعكس حادثة الدراسات و المراجع للبحث.

03

تنوع المناهج في الدراسات السابقة

اعتمدت أغلب الدراسات المرجعية على المنهج الكمي والتجارب الحاسوبية.
و كان من الجيد الاستفادة من دراسات ذات منهجيات نوعية، مثل مقابلات ميدانية مع فرق السلامة أو دراسات حالة Case Studies
لتوضيح التحديات العملية للتطبيقات الميدانية.

04

العينة والبيانات في الدراسات السابقة

بعض الدراسات السابقة استخدمت مجموعات بيانات عامة COCO أو Pascal VOC لتدريب النماذج و كذا فان هذه العينات لا تمثل بيئة العمل في قطاع الاتصالات أو مواقع الأبراج، بينما الرسالة ركزت على بيانات واقعية 9000 صورة PPE ، وهو ما يجعل الدراسات السابقة أقل ارتباطاً بمجال تخصص البحث.

النتائج والتوصيات في الدراسات المرجعية

معظم الدراسات السابقة ركزت على مؤشرات الدقة mAP, Recall دون تناول البعد العملي أو تأثيرها على تحسين السلامة و بالتالي فان غياب توصيات تطبيقية في الدراسات المرجعية جعل الباحث يعتمد على إضافته الشخصية لربط الذكاء الاصطناعي بتحقيق السلامة الميدانية، وهذا يُعد إضافة للرسالة.



مجموعة جامعة
الجميع الذكية
للتعليم القابضة

MIDOCEAN
UNIVERSITY



Thank You



MIDOCEAN UNIVERSITY
FACULTY OF INFORMATICS

**Real-time Object detection of Personal Protective Equipment in video Streaming Using
Deep Learning Technique**

BY
Younis Arrabi

Under the Supervision of
Prof. Dr. Yasser Abdelhamid
College of Informatics
Midocean University
2024

Acknowledgements

We want to express our gratitude to Prof. Dr. Yasser Abdelhamid for his guidance, encouragement and insightful feedback that played a significant role, in completing this research. His support has truly been a source of motivation.

Also extend thanks to the faculty and staff of the informatics college at Midocean University for creating an environment and providing the necessary resources for this research. Special appreciation goes out to my colleagues and friends whose support and engaging discussions have enhanced my research journey.

We are deeply grateful to our families for their support and encouragement. Their patience and understanding have been a source of strength throughout this endeavor. Thank you all for your contributions and unwavering support.

Contents

Abstract.....	vi
List of Abbreviations	viii
List of Figures.....	ix
List of Tables	x
Chapter 1 Introduction	11
1.1 Problem Definition.....	12
1.2 Thesis Objectives	14
1.3 Thesis Contribution.....	14
1.4 Thesis Organization	15
Chapter 2 Literature Review	18
2.1 Overview of PPE and Workplace Safety	18
2.2 Sensor-Based PPE Detection Methods	18
2.3 Computer Vision-Based PPE Detection Methods	23
2.4 YOLO for PPE Detection	34
Chapter 3 Related Work.....	35
3.1 Detection of workers PPE using YOLO	35
3.2 Detection of workers PPE using CNN.....	37
3.3 Summary	39
Chapter 4 Methodology	40
4.1 YOLO Algorithm.....	40
4.2 Dataset Construction.....	42
4.3 Training and Evaluation.....	46
4.4 Performance Metrics and Benchmarking.....	47
4.4 Summary	50
Chapter 5 Experimental Results and Discussion	51
5.1 Performance Evaluation.....	51
5.2 Challenges and Limitations.....	56

Chapter 6 Conclusion and Future Work	57
6.1 Conclusion.....	57
6.2 Future Work	57
References	58

Abstract

Magnificent progress in Artificial Intelligence computer vision techniques has been achieved during the last decades. Object detection should benefit from these improvements to discover about how workers are safe in different workplace environments. Because of the continuous increase in the number of researchers who are affiliated with computer vision and apply object detection techniques and deep learning methods, a relatively new discipline called Personal protective equipment auto detection has emerged. It aims to use methods, tools and techniques of deep learning in object detection to create and learn a model that can detect personal protective equipment in real-time and in automatic way.

Ensuring the proper use of Personal Protective Equipment (PPE) is critical in all workplace with accident high hazard settings; to reduce the risk, this project develops a real-time system for detecting Personal Protective Equipment in video streams using You Only Look Once (YOLO) deep learning techniques. The system identifies various Personal Protective Equipment requirements, such as helmets and vests, with high accuracy and minimal latency. By leveraging the efficiency of latest YOLO version, the solution achieves effective real-time performance, making it suitable for integration into existing surveillance infrastructures. This approach offers a scalable method for enhancing workplace safety monitoring and compliance.

Accuracy of detecting Personal Protective Equipment in most studies do not achieve the state-of-the-art accuracy in contrary with deep learning performances on datasets. The objective of this project is to develop a DL-based object detection model for the real-time detection of safety tools, including helmets, vests, gloves, safety shoes and safety goggles using different datasets.it aims to detect tools using object detection deep learning YOLO algorithm.

This decomposes into the following sub-objectives: Selecting and implementing proper algorithms from the recent literature for detecting objects. Evaluating the algorithms base on metrics such as accuracy, precision, and recall.

The research work introduces two deep learning models, namely Convolutional Neural Networks (CNNs), and You Only Look Once (YOLO), to detect PPE. The most used deep learning techniques in object detection are YOLO and CNN so they must be included in our research. And their performance and advantage and limitation taken into considerations in selecting final Technique.

Our results show that the YOLO model outperforms CNN Model. The accuracy of YOLO 98.4% and that of CNN was 98.2%.

The results show that accuracy of YOLO in Object detection reaches the state-of-the-art accuracies on dataset and can be efficiently used for detecting PPE of the workers and in real-time systems to reduce risk of accidents.

List of Abbreviations

- **PPE:** Personal Protective Equipment
- **ANN:** Artificial Neural Network
- **CNN:** Convolutional Neural Networks
- **R-CNN:** Region Based Convolutional Neural Networks
- **mAP:** Mean Average Precision
- **YOLO:** YOU ONLY LOOK ONCE
- **CM:** Confusion Matrix
- **DL:** Deep Learning
- **ML:** Machine Learning
- **AI:** Artificial intelligence
- **RPN:** Region Proposal Network.
- **RoI:** Region of Interest
- **SVM:** Support Vector Machines
- **ReLU:** Rectified linear unit
- **AP:** Average Precision
- **IoU:** intersection, over union

List of Figures

Figure 1 – RFID: The Technology Making Industries Smarter | TT Electronics

Figure 2 – GPS Structure

Figure 3 – Bluetooth beacon system

Figure 4 – Successful CNN

Figure 5 – R-CNN

Figure 6 – Picture Faster RCNN

Figure 7 – High-level diagram of SSD for generic object detection

Figure 8 – YOLO (Source Mastering All YOLO Models from YOLOv1 to YOLOv9)

Figure 9 – YOLOv8 Architecture

Figure 10 – Annotation sample

Figure 11 – training Images

Figure 12 – validation Images

Figure 13 – Noise Augmentation

Figure 14 – Gray Scale Augmentation

Figure 15 – Rotation & Flipping Augmentation

Figure 16 – Scaling Augmentation

Figure 17 – mAP for YOLOv8 Object detection algorithm

Figure 18 – precision for YOLOv8 Object detection algorithm

Figure 19 – Recall for YOLOv8 Object detection algorithm

Figure 20 – Confusion Matrix

Figure 21 – All KPIs of the YOLOv8 Trained Model

Figure 22 – Live Object detection capture.

List of Tables

Table-1 Comparative summary for sensor-based PPE detection.

Table-2 Comparative summary between CNN & YOLO

Table-4 Instances count

Table-5 Showing instances and mAP & Recall metric.

Table-6 F1 Score Results

Table 7 – Object detection record.

Chapter 1 Introduction

Safety of the worker during working hours become critical requirement over all kinds of projects, scaling from Small to mega size, companies and governments seeking to have safe workplace environment for the workers, and this have benefit on both workers and projects owners as well, these projects have complex tasks to maintain safety adherence and compliance, different project have different safety requirement, from complicated and simple workers-safety requirement, projects such as nuclear plants project, electricity main production stations, water desalination mega stations, chemical industries, firefighting, and also construction or small workshops as in [1].

In our projects we will focus on construction and industrial sectors that are experiencing rapid growth worldwide, making safety a primary concern for business owners and regulatory bodies. utilizing personal protective equipment (PPE) that can protect and avoid injuries and accidents to workers, also saving project budgets and company reputations, Workers in these sectors under significant hazards that can lead to injuries and accidents.

To address these risks, numerous safety controls and protocols are implemented to ensure compliance with health and safety regulations across various workplace environments. Several factors contribute to workplace accidents, including inadequate safety training, low awareness of safety measures, improper use of PPE, and insufficient presence of safety officers in hazardous areas.

Adhering to safety protocols and regulations can significantly reduce the risk of injuries, enhancing workplace safety and creating a better working environment. (OHSAS 18001 and ISO 45001 as in [2]. Safety protocols typically include comprehensive training programs that cover safety rules, proper PPE usage, enforcement of PPE compliance, and periodic compliance checks. [1]

The challenging in this task is that it is continuous and essential for protecting workers and minimizing the risk of injuries and accidents, mentioning key types of PPE are:

1. Head Protection: Helmets, hard hats, and bump caps to protect against head injuries.
2. Eye and Face Protection: Safety goggles, face shields, and glasses to protect against chemical splashes and flying particles.
3. Hearing Protection: Earplugs and earmuffs to reduce exposure to harmful noise levels.
4. Respiratory Protection: Masks and respirators to protect against inhaling hazardous substances.
5. Hand Protection: Gloves to protect against cuts, abrasions, chemicals, and extreme temperatures.
6. Body Protection: Protective clothing like lab coats, aprons, and coveralls to shield the body from chemicals, heat, and biological hazards.
7. Foot Protection: Safety shoes and boots to protect against injuries from heavy objects and electrical hazards.
8. Fall Protection: Harnesses and lanyards to prevent falls from heights.

Here, we will focus on PPE usage compliance by workers before entering the workplace, by detecting the PPE seamlessly utilizing different techniques, where the main goal of detection is real-time and accurate, this will be utilizing object detection multiple techniques [\[3\]](#).

1.1 Problem Definition

The existing approach, to identifying breaches in using personal protective equipment involves manual checks and occasional audits. However, these methods fall short in safeguarding workers against accidents. Consequently, safety violations are often detected belatedly resulting in delayed interventions and heightened risks that could endanger the wellbeing of employees. The absence of a time automated mechanism, for overseeing and ensuring safety adherence poses an obstacle to the efficiency of established safety protocols.

The delay, in identifying and addressing safety breaches leads to an increase in workplace accidents, harm and deaths. These occurrences not impact the safety of workers. Also result in significant financial burdens for businesses, such, as healthcare costs, legal responsibilities and reduces efficiency. Moreover, recurrent safety issues can damage a company's image. Invite rigorous regulatory oversight. [\[4 to 8\]](#)

1.1.1 Advantages of Real Time Automated PPE Detection

1. Improves Safety; Real time detection systems can quickly spot PPE violations enabling actions to prevent accidents and injuries.
2. Cost Efficiency; Automated systems lessen the need, for checks and inspections resulting in substantial cost savings over time.
3. Compliance Boost; Continuous monitoring ensures adherence to safety rules enhancing compliance rates.
4. Reputation Management; Effective safety practices uphold a company's reputation and minimize the risk of fines.

1.1.2 Challenges of Real Time Automated PPE Detection

1. Initial Capital Investment; Setting up real time detection systems involves an investment in technology and infrastructure.
2. Technical Hurdles; Ensuring the precision and dependability of automated systems in settings can pose challenges.
3. Privacy Concerns; The use of surveillance technologies raises worries about worker privacy and data protection. – No data storage

The main objective is to create and implement a time automated system for monitoring PPE compliance. Such a system would improve worker safety cut costs and

ensure adherence, to safety regulations ultimately fostering and more efficient work environment.

1.2 Thesis Objectives

The objective is to improve workers safety by creating a real time PPE detection system that can utilize across various industrial environments ensuring adherence, to safety regulations and minimizing accident risks.

The main goal of this project is to create a learning model that can detect safety equipment, like helmets, vests, gloves, safety shoes and safety goggles in real-time by using different datasets. The model identifies these tools using object detection algorithms.

This objective can be divided into the following sub goals;

1. Selecting and Implementing Algorithms; Apply the appropriate algorithms based on recent developments in object detection to identify safety equipment.
2. Evaluating Algorithms; Measure the performance of the chosen algorithms using metrics such as accuracy, precision and recall.

This project aims to build a real time system for detecting protective equipment (PPE) through deep learning methods. By utilizing algorithms, the system seeks to enhance safety compliance in settings.

1.3 Thesis Contribution

The main focus of this thesis is suggesting a computer vision system that identifies Personal Protective Equipment (PPE) by utilizing the You Only Look Once (YOLO) algorithm. This study aims to improve safety adherence, in construction and industrial settings by applying cutting edge deep learning methods to enforce PPE regulations. The achievements of this research encompass aspects, such, as;

1. YOLO Model for PPE Detection: This thesis proposes the use of YOLO to detect accuracy of PPE instances, addressing common challenges in real-world industrial settings.

2. Dataset Creation: A new dataset specifically for PPE detection, comprising 1400 PPE instances. This dataset includes various types of PPE such as helmets, safety goggles, vests, gloves, and safety shoes. The dataset serves as a valuable resource for future research and development in this area.

3. Image Augmentation Techniques: The research introduces the use of the augmentation; Geometric Transformations, include Flipping, Rotation, Scaling, Translation. Also adding Noise to image and gray scale, that will enhance model detection results, pictures count reaches 9000 after applying above augmentation.

4. Performance Evaluation and Benchmarking: The YOLO model was evaluated, achieving a mAP (mean Average Precision) of 98.4%, maintaining a real-time processing passes on computer performance. These metrics highlight the model's effectiveness and suitability for real-time applications in industrial safety.

5. Comparative Analysis of Detection Methods: The thesis provides a comprehensive analysis of sensor-based and computer vision-based PPE detection methods. It highlights the limitations of traditional sensor-based approaches, such as high cost and operational challenges, and demonstrates the advantages of computer vision techniques in terms of accuracy, scalability, and cost-effectiveness.

6. Practical Implications and Future Work: The research outlines practical implications for deploying a system in real-world industrial environments. It discusses the potential for integrating the system with existing safety protocols and provides recommendations for future work, including the exploration of other deep learning models and expanding the dataset to include more diverse scenarios and PPE types.

1.4 Thesis Organization

The thesis is divided into chapters each focusing on elements of studying the identification of Personal Protective Equipment (PPE) with computer vision methods, particularly the YOLO

algorithm. The layout guarantees a progression from laying out the background and purpose of the research to diving, into specifics, findings and conclusions.

1. Introduction

Starting with Background and Motivation Introducing the importance of PPE in ensuring workplace safety, challenges associates with traditional detection methods, and the need for advance computer vision solutions.

Then Problem Statement, clearly defines the research problem and the objectives of the study. Later, discuss the contributions of our Thesis, Summarizes the main contributions of the thesis, including YOLO model and a new dataset for PPE detection, finally, thesis Organization shows an overview of the structure.

2. Literature Review

In this section, first we introduce of Personal Protective Equipment (PPE) and Safety at Work. The listing detection Techniques for PPE Using Sensors; Explores conventional approaches to identifying PPE, such as RFID, infrared and ultrasonic sensors pointing out their constraints.

Then we mention the utilizing Computer Vision for PPE Detection; Looks into the advancements in computer vision methods for PPE detection with a focus on algorithms, like CNNs and YOLO.

3. Related work

In research there has been a focus, on using advanced deep learning methods, especially CNN based structures like YOLO to improve the real time detection of personal protective equipment (PPE) in different industrial settings. For instance, these research findings collectively demonstrate the progress made in PPE detection through learning techniques with

efforts are aiming at enhancing accuracy processing speed and practical applicability across various real-world scenarios.

4. Methodology

In this section we explain the way that we follow to achieve thesis objective, by suggesting YOLO Algorithm; thorough breakdown of the YOLO algorithm discussing its iterations and the progress made in each version.

Then discussing the structure of the Dataset; Describing how datasets are put together covering everything from gathering data to labeling and augmentation techniques.

Training and Assessment is the most technical part; Exploring into the training procedures assessment measurements (mean Average Precision, Average Precision) and capabilities, for real time processing.

5. Experimental Results and discussion

First assessing Performance; where we discussing the outcomes using the model. Following with result Examination; Digging into a review of how well the model performed, emphasizing enhancements in detecting personal protective equipment instances.

6. Conclusion and future work

Key Discoveries Summary; Summarizes the discoveries of the study highlighting their significance for improving workplace safety. Listing some Industry Implications; Explores how the research findings can practically enhance safety adherence in industrial and construction settings. Closing Thoughts; Offers reflections on the importance of the research and its potential for future progress, in the field.

7. References: Comprehensive list of all sources cited throughout the thesis, formatted according to the appropriate academic style.

Chapter 2 Literature Review

2.1 Overview of PPE and Workplace Safety

PPE usage compliance by workers before entering the workplace, by detecting the PPE seamlessly utilizing different techniques, where the main goal of detection is real-time and accurate, this will be utilizing object detection multiple techniques.

There are two main approaches to object detection they will enhance health and safety adherence, sensor-based and computer vision-based. Sensor-based techniques, which have been used since 2015, often involve high costs and additional safety risks. [1].

Sensor-based PPE detection methods have been implemented in various industries to enhance safety compliance and reduce workplace accidents. These methods leverage different types of sensors to ensure that workers are wearing personal protective equipment (PPE) before entering or while present in hazardous areas., These sensor-based methods are critical in enhancing the safety of workers by ensuring they are wearing the necessary PPE in real-time, although they come with challenges related to cost, practicality, and environmental constraints [1].

2.2 Sensor-Based PPE Detection Methods

Going more depth, look at several sensor-based methods, Radio Frequency Identification (RFID) Tags, where Workers' PPE with RFID tags, which are detected by RFID readers installed at key locations such as entry points, exits, and within specific zones of a worksite as shown in figure-1 below, Advantages of RFID systems can track multiple items simultaneously and provide real-time monitoring. They are robust and can be integrated with existing safety management systems, saying about Limitations, the initial setup cost can be high due to the need for tags, readers, and integration infrastructure. There can also be issues with tag readability in environments with metal or liquid interference, we found application Example of it like on construction sites, RFID tags are within helmets and vests. RFID readers at site entrances ensure that workers are wearing the correct PPE before allowing entry [9]

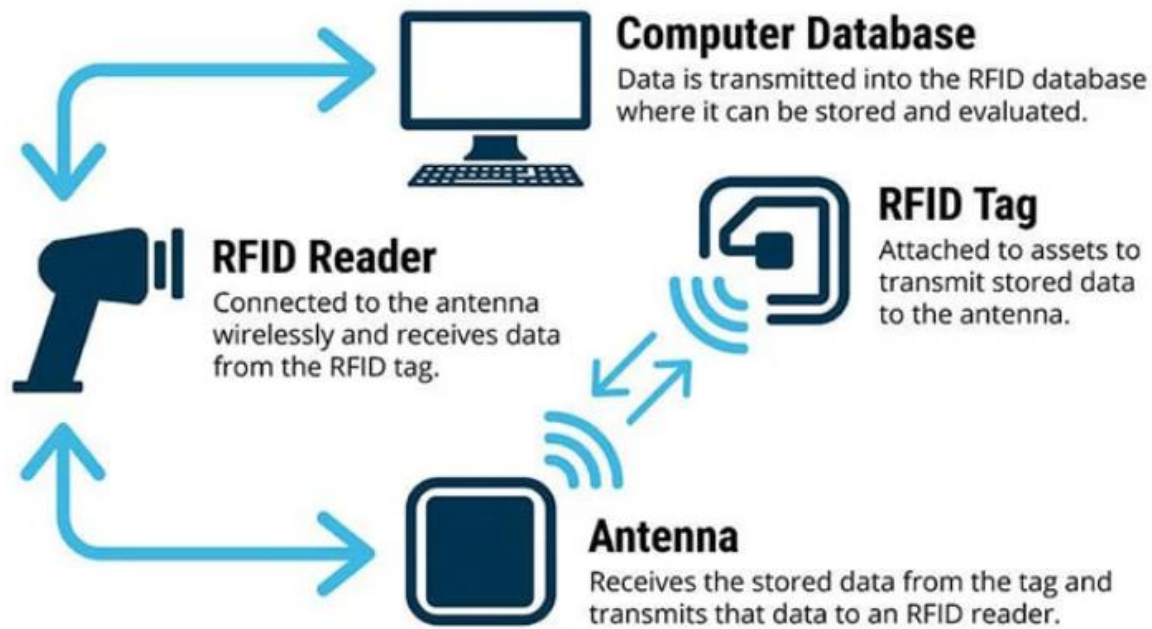


Figure-1, Photo courtesy of: [TT Electronics](https://www.tt-electronics.com), showing RFID system structure

[RFID: The Technology Making Industries Smarter | TT Electronics](https://www.tt-electronics.com)

2nd sensor method, is Global Positioning System (GPS)-Based Systems, Workers carry GPS-enabled devices that track their location and ensure they remain within designated safe zones while wearing the PPE, overview of this technique illustrates in Figure-2, Advantages of GPS systems are effective in large, open environments and can provide detailed location data over extensive areas, mentioning the limitations of GPS-Based method, these systems are less effective indoors or in areas with poor satellite signal reception, GPS devices also require regular maintenance and power management, Application Example of it, In mining operations, GPS devices ensure workers are within safe zones and are with necessary PPE such as respirators and helmets[10].

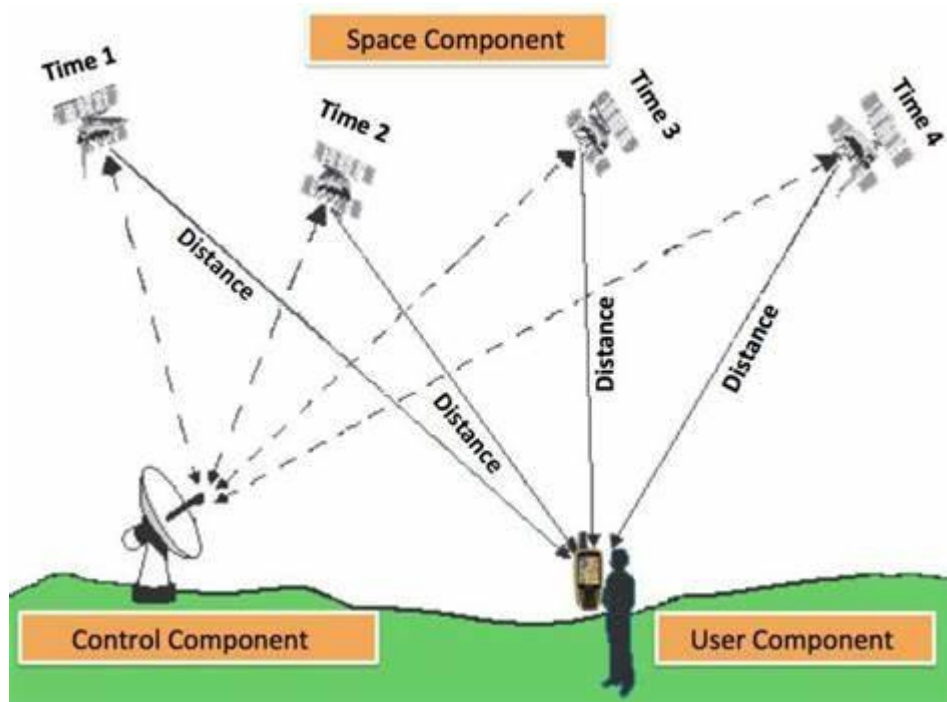


Figure-2 GPS Structure

In Figure-3, showing another method of sensor based in PPE detection is Bluetooth Beacons, PPE items fits with Bluetooth beacons that communicate with sensors throughout the worksite, these sensors verify the presence of the beacons (and thus the PPE) in real-time, advantages of Bluetooth systems are relatively low-cost and suitable for indoor environments also they can provide precise location data within the worksite, mentioning its limitations, The range is limited compared to other technologies, and there can be interference from other Bluetooth devices or electronic equipment, it uses In manufacturing plants, Bluetooth beacons on ear protection and safety glasses communicate with sensors at workstations to ensure compliance before machine operation begins [11].

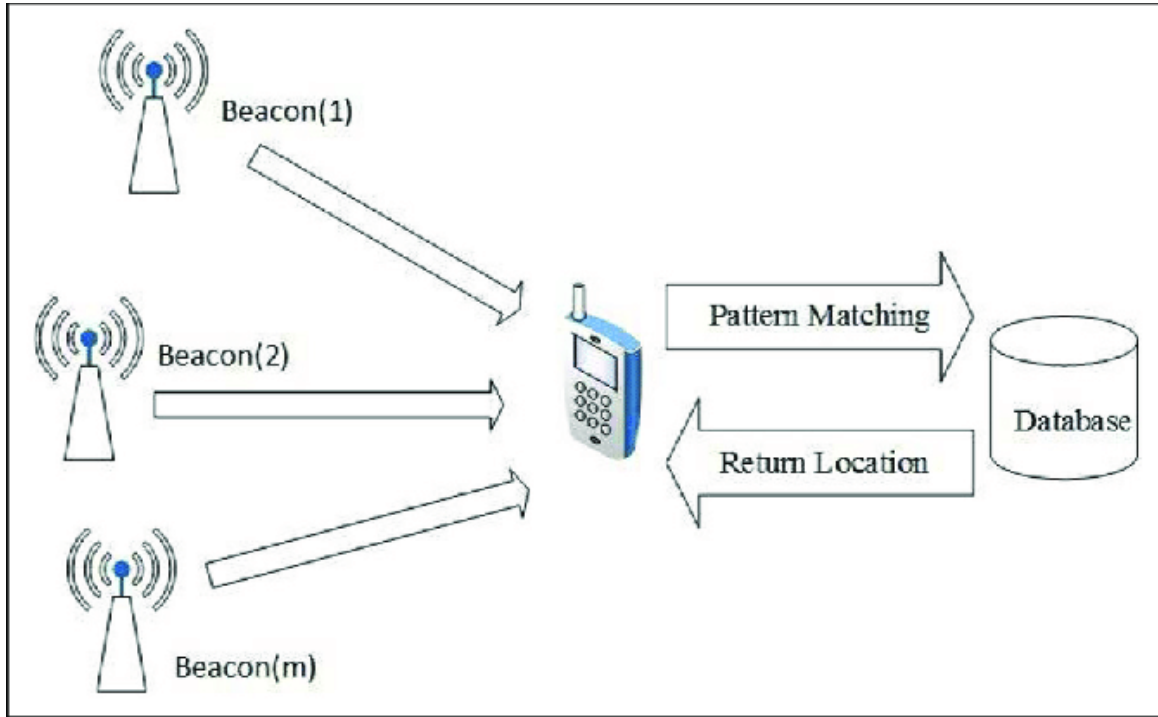


Figure-3 Bluetooth beacon system

4th method of sensor based in PPE detection is Proximity Sensors, it detect the presence of workers and their PPE by sensing specific tags or embedded sensors in the equipment, advantages of these sensors can provide immediate feedback and ensure that PPE is worn when workers approach hazardous equipment or areas, its limitations that Effective only in localized areas, requiring multiple sensors for comprehensive coverage, which can increase costs, application Example In chemical plants, proximity sensors near chemical handling areas detect if workers are wearing protective gloves and face shields [12].

Last Method is Pressure Sensors, Pressure-sensitive mats or floors detect the presence of workers and can be linked to ensure PPE compliance when workers step on them, advantages is Provide direct and immediate feedback when workers enter restricted or hazardous areas, Limitations is high installation and maintenance costs and limited to specific entry or exit points, Application Example is Laboratories use pressure-sensitive mats at entrances to ensure workers wear lab coats and safety glasses before entering [13].

Below is the table-1 is summarizing different sensor-based methods, which can give high level view, and comparative summary.

Table-1 Comparative summary for sensor-based PPE detection.

Method Name	Device	Advantages	Limitations	Application	References
Radio Frequency Identification (RFID) Tags	RFID TAG per PPE equipment Antennas RFID Readers Databases	real-time monitoring Integration with other monitoring system	High Cost, Sensors is sensitive to Metal environment subject to interference	used at the site entrance & exits	[9]
Global Positioning System (GPS)-Based Systems	GPS Device Per PPE equipment Database.	effective in large, open environments and can provide detailed location data over extensive areas	Used in Outdoor High Operation Cost	mining operations, GPS installed in Helmet	[10]
Bluetooth Beacons	Bluetooth sensors per PPE Detectors	Low Cost Indoor Environment	Limited in Range Subject to interference	used in manufacturing plants	[11]
Proximity Sensors	Proximity sensors TAGs per PPE	Real-time	High operation cost	chemical plants	[12]
Pressure Sensors	Pressure-sensitive mats or floors detect the	Real-time	High Operation cost	Laboratories use pressure-sensitive mats at entrances to ensure workers wear lab	[13]

	presenc e			coats and safety glasses before entering	
--	--------------	--	--	--	--

Sensor-based PPE detection methods plays a significant role in enhancing workplace safety by ensuring compliance with safety protocols. These systems, utilizing different techniques as above. Despite their benefits, such as precise tracking and alerting capabilities, they come with limitations including higher costs, potential discomfort for workers, and complex installation requirements.

2.3 Computer Vision-Based PPE Detection Methods

Building upon the groundwork established by sensor-based methods for detecting protective equipment (PPE) computer vision has emerged as a robust solution to enhance safety in the workplace.

By utilizing algorithms and deep learning methods computer vision systems can effectively identify and track the usage of PPE in real time.

In contrast to sensor-based approaches that rely on tags or devices computer vision employs cameras and image processing to recognize various types of PPE such as helmets, safety glasses and vests through object detection in the field of artificial intelligence.

Object detection involves the identification and localization of objects in images or videos typically represents by bounding boxes. This task has seen advancements over time with algorithms like Convolutional Neural Networks (CNNs) and YOLO (You Only Look Once) transforming object detection by providing real time processing capabilities and high accuracy. This development facilitates adaptable and scalable solutions across different domains like autonomous driving, surveillance and industrial automation. By harnessing datasets and powerful computational resources this technology offers benefits such as unobtrusive monitoring, scalability and operation, in diverse and complex environments.

With the advancements in deep learning models the integration of computer vision for Personal Protective Equipment (PPE) detection poise to transform safety procedures enhancing

compliance and reducing workplace injury risks significantly. Computer vision techniques utilize deep learning principles to recognize and ensure proper usage of PPE in real time. This approach typically involves utilizing cameras to capture images or videos, using machine learning models to identify PPE items like helmets, safety glasses, vests and gloves.

Exploring deeper into these techniques Convolutional Neural Networks (CNNs) play a role. CNNs are a type of deep neural networks for processing structured grid data such as images. They serve as a two-stage object detection algorithm for image recognition and classification tasks. Trained on labeled datasets comprising images of individuals with and without PPE CNNs learn to discern patterns and features associated with types of protective gear. This method offers advantages like accuracy and the capability to extract intricate patterns directly from the data; however, it necessitates sizable labeled datasets and substantial computational resources, for training purposes [14].

The CNN algorithm for object detection involves several key stages as in Figure-4, below is stages in brief [14]:

1. Input Layer

- a. **Image Acquisition:** The process begins with capturing or acquiring the image to be processed. The image is typically resized to a fixed dimension suitable for the CNN (e.g., 224x224, 416x416 pixels).
- b. **Normalization:** Pixel values are normalized to a range (e.g., [0, 1]) to facilitate faster convergence during training.

2. Convolutional Layers

- a. **Feature Extraction:** These layers apply convolution operations using various filters to detect features like edges, textures, and patterns. Each convolutional layer extracts higher-level features from the preceding layer.
- b. **Activation Function:** After convolution, an activation function is applied to introduce non-linearity to the model.

3. Pooling Layers

- a. **Down-sampling:** Pooling layers reduce the spatial dimensions of the feature maps, thus decreasing the computational load and controlling overfitting. Common pooling methods include max pooling and average pooling.

4. Fully Connected Layers

- a. **Classification:** These layers are used to classify the features extracted by the convolutional layers. They aggregate the features and predict the class probabilities.
- b. **Dropout (Optional):** Dropout layers are used to prevent overfitting by randomly setting a fraction of input units to zero during training.

5. Output Layer

- a. **Final Predictions:** The final output consists of the bounding boxes, class labels, and confidence scores for each detected object in the image.

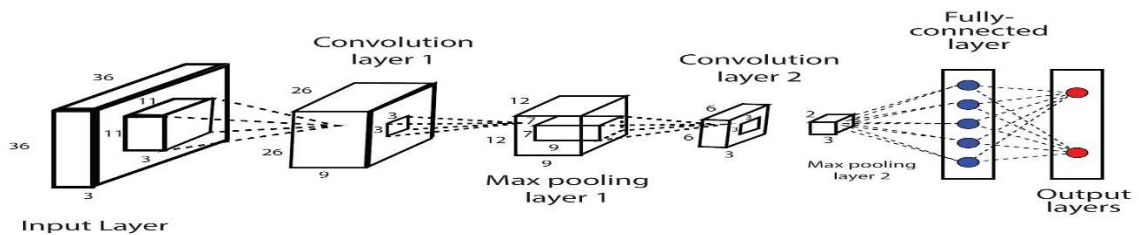


Figure 4 – Successful CNN References: "Deep learning for safety surveillance: An approach to detect PPE compliance" (Chen et al., 2018).

2nd technique is Region-Based Convolutional Neural Networks (R-CNN), in R-CNN models divide the image into regions and then use CNNs to classify each region, it is Useful for detecting multiple PPE items in a single image, such as helmets, vests, and gloves, by focusing on specific regions of interest within the image, it has advantage that improves accuracy for object detection tasks by focusing on relevant regions, for the limitations it is Computationally intensive and slower than other methods in [15].

The R-CNN algorithm for object detection involves several key stages as Figure-5 below [15]:

1. **Region Proposal Generation: Selective Search:** This method generates a set of candidate object regions (also known as region proposals). It combines hierarchical grouping and greedy search to propose around 2,000 regions that are likely to contain objects.
2. **Feature Extraction, CNN Application:** Each region proposal is warped to a fixed size and fed into a pre-trained Convolutional Neural Network to extract features. This step transforms the region into a fixed-length feature vector.
3. **Classification Classifiers:** The extracted features from the CNN are fed into a set of Support Vector Machines (SVMs), one for each object class. The SVMs classify the regions into one of the pre-defined object categories or as background.
4. **Output, Final Detections:** The algorithm outputs the final set of bounding boxes, class labels, and confidence scores for the detected objects in the image.

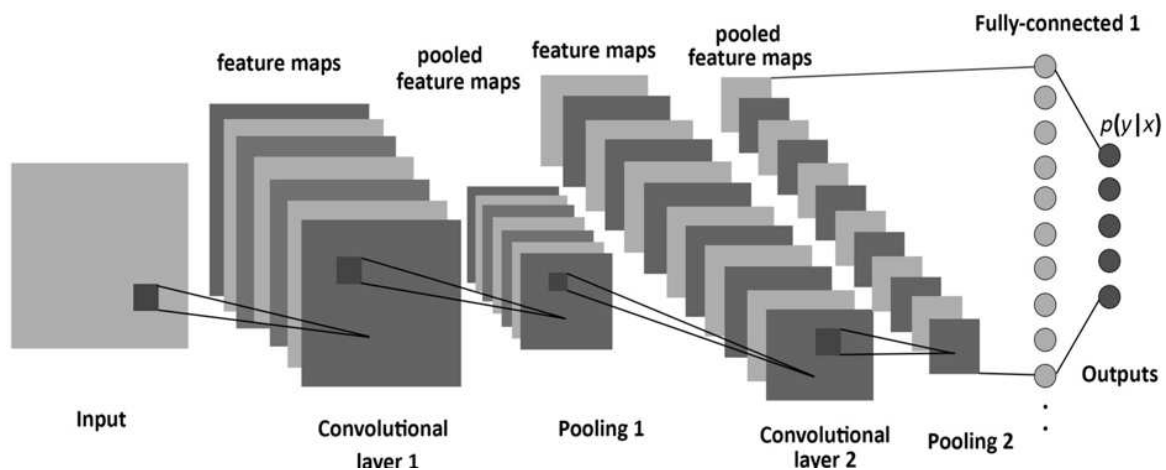


Figure 5 – R-CNN, References: "R-CNN for real-time PPE detection in construction sites" (Girshick et al., 2014).

3rd is the Faster R-CNN, Method: An improved version of R-CNN, Faster R-CNN uses a Region Proposal Network (RPN) to generate candidate object bounding boxes and a subsequent network to classify these boxes, detecting PPE in more complex and cluttered environments by generating high-quality region proposals, advantages high accuracy and efficiency in generating and classifying region proposals, limitations for this technique more computationally expensive as in [14].

Here are the main stages of Faster R-CNN also illustrates in figure-6 [14]:

1. Convolutional Layers (Feature Extraction): Convolutional Layers: The input image is passed through several convolutional layers (typically from a pre-trained network) to extract a rich hierarchy of features. These convolutional layers are shared between the Region Proposal Network (RPN) and the Fast R-CNN detector.

2. Region Proposal Network (RPN): Anchor Boxes: The RPN slides a small network over the convolutional feature map and generates multiple region proposals (anchors) at each location. These anchor boxes are of different scales and aspect ratios. Proposal Generation: The RPN outputs two scores for each anchor: an objectness score indicating whether the anchor contains an object or not, and a bounding box regression to refine the anchor's coordinates. MS on Proposals: Non-Maximum Suppression (NMS) is applied to the proposals to reduce redundancy and keep only the most promising regions.

3. Region of Interest (RoI) Pooling: RoI Pooling Layer: The proposals generated by the RPN are projected onto the feature map. Each proposal is then warped into a fixed size using the RoI pooling layer, which extracts a fixed-length feature vector for each proposal.

4. Fully Connected Layers (Classification and Regression): Object Classification: The feature vectors from RoI pooling are fed into fully connected layers to classify each proposal into one of the object classes or as background. Bounding Box Regression: In parallel, the network refines the bounding box coordinates for each proposal to improve localization accuracy.

5. Output: Final Detections: The final output consists of the class labels and refined bounding boxes for each detected object. Non-Maximum Suppression is again applied to remove duplicate detections.

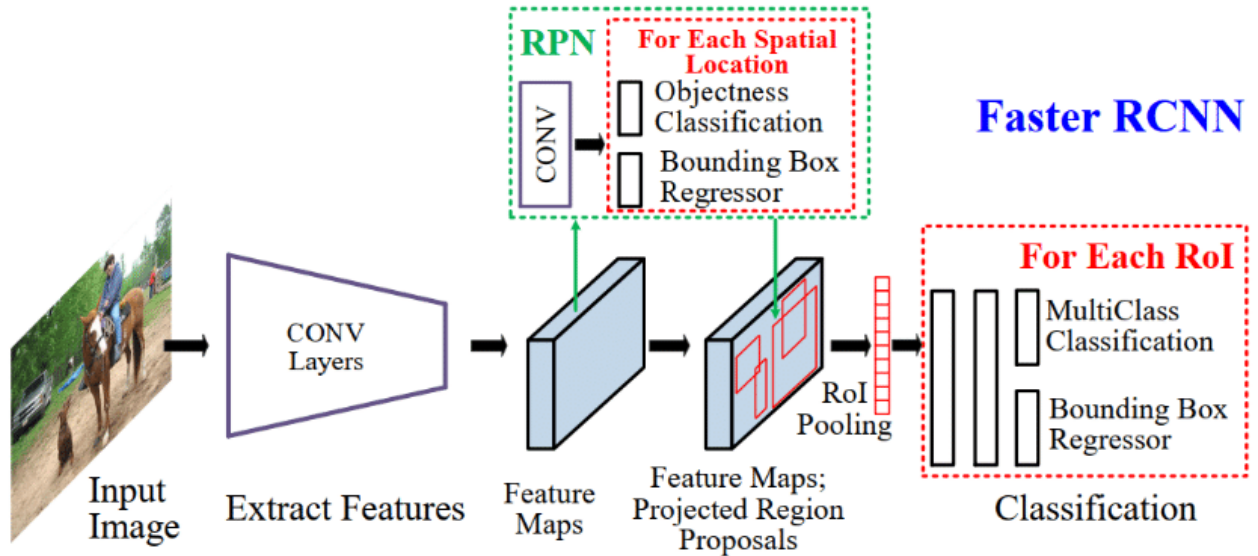


Figure 6 - Picture Faster RCNN, References: "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks" (Ren et al., 2015).

4th technique is Single Shot MultiBox Detector (SSD), SSD detects objects in images using a single deep neural network that predicts both bounding boxes and class scores simultaneously, it is effective for detecting multiple PPE items in images with varying sizes and aspect ratios, for its Advantages, it Balances speed and accuracy, making it suitable for real-time applications, its Limitations that Generally, less accurate than Faster R-CNN but faster in [18].

In figure-7, are the main stages of SSD [18]:

1. **Input Image and Base Network, Image Preprocessing:** The input image is resized to a fixed size and normalized. This preprocessed image is then fed into a base network.
Base Network (Feature Extraction): A pre-trained convolutional neural network is used as the backbone to extract feature maps from the input image. These feature maps capture rich hierarchical features of the image.
2. **Extra Feature Layers, Additional Convolutional Layers:** SSD adds a series of auxiliary convolutional layers on top of the base network. These layers progressively decrease in size and act as feature detectors at multiple scales. They enable the detection of objects of various sizes.
3. **Multi-Scale Feature Maps, Detection at Multiple Scales:** SSD uses feature maps from different layers of the network (both from the base network and additional layers) to

detect objects. This approach allows the model to handle objects of various sizes by making predictions at multiple scales.

4. Default Boxes (Anchor Boxes), Anchor Boxes: At each location on these feature maps, SSD places a set of default bounding boxes of different aspect ratios and sizes. These default boxes act as initial guesses for the presence and location of objects.

5. Classification and Localization, Convolutional Predictors: Each default box is evaluated by small convolutional filters that predict both the confidence scores for each class (including a background class) and the bounding box offsets. This is done for every location on each feature map. Confidence Scores: These scores indicate the likelihood that a default box contains an object of a specific class.

Bounding Box Offsets: These offsets adjust the coordinates of the default boxes to better fit the actual objects.

6. Final Output, Detections: The final output consists of the refined bounding boxes, class labels, and confidence scores for each detected object in the image.

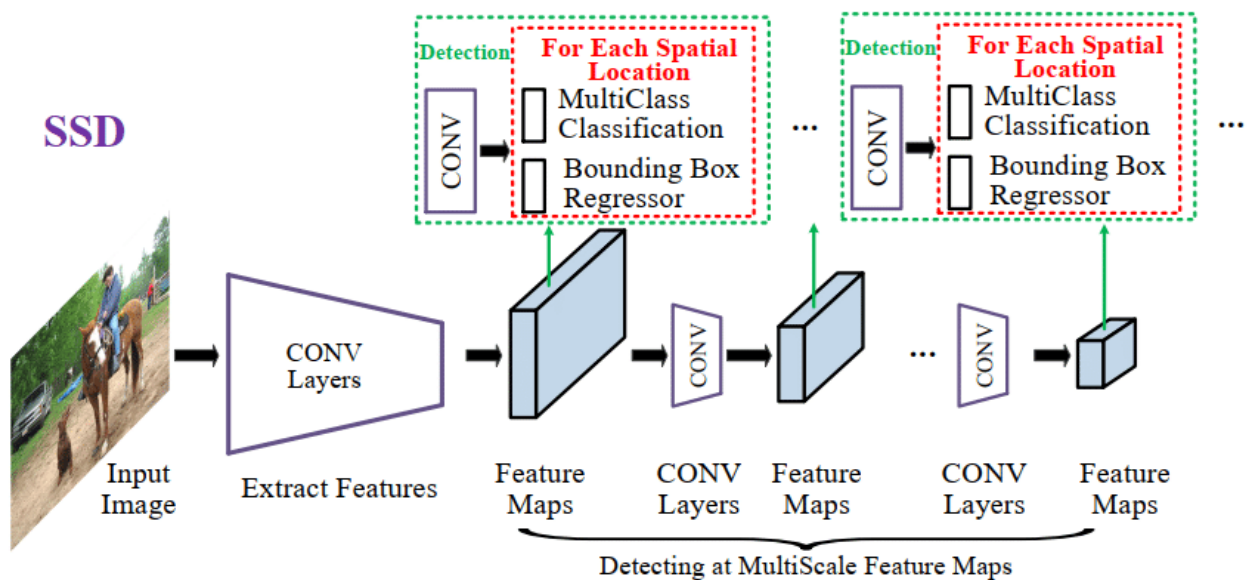


Figure 7 - High-level diagram of SSD for generic object detection, References: "SSD: Single Shot MultiBox Detector" (Liu et al., 2016).

Coming to most improving and latest technique in deep learning is You Only Look Once (YOLO), YOLO is an object detection algorithm that divides the image into a grid and predicts bounding boxes and class probabilities for each grid cell in a single pass, it is ideal for real-time PPE detection due to its speed and accuracy, YOLO can simultaneously detect multiple types of PPE in real-time video streams, it is also have good advantages than others which is high speed and efficiency, suitable for real-time applications, for Limitations may struggle with detecting very small objects or objects in complex backgrounds.

The development of history began in 2016 and here are the key milestones in the evolution of YOLO as in [58];

YOLOv1 (2016); The initial release of YOLO (You Look Once) was pioneered by Joseph Redmon and his team. It transforms object detection by treating it as a regression issue directly forecasting bounding boxes and class probabilities from entire images in one pass. While YOLOv1 boasted speed, processing images at 45 frames per second it faced challenges with detecting small objects and precisely localizing objects within images.

YOLOv2 (2017); Building upon its predecessor, YOLOv2 introduced enhancements. It incorporated Batch Normalization, a High-Resolution Classifier and a novel anchor box method inspired by Faster R CNN. This version also utilized a revamped Darknet 19 architecture to improve both speed and accuracy. With the ability to detect over 9000 object categories.

YOLOv3 (2018); YOLOv3 marked another leap forward in model accuracy and capabilities. It embraced a network architecture called Darknet-53, which leveraged residual blocks and additional layers, for enhanced feature extraction. Introducing a scale detection approach enabled the network to identify small, medium and large objects more efficiently. Balancing speed with accuracy effectively positioned it for real time applications.

YOLOv4 (2020); This update aimed at enhancing performance and broadening the model's accessibility to an audience. YOLOv4 introduced improvements, including using CSPDarknet53 as the foundation incorporating PANet for path aggregation and introducing a new data augmentation method called Mosaic. It achieved mean Average Precision (mAP) scores and quicker inference speeds compared to its predecessors.

YOLOv5 (2020); While not officially released by the YOLO creators YOLOv5 quickly became popular in the open-source community for its user-friendly nature and ease of implementation. YOLOv5 prioritized models, scalability and simplicity in both training and deployment processes. It provided model sizes (small, medium, large and extra-large) to strike a balance between speed and accuracy for different application scenarios.

YOLOv6 and YOLOv7 (2021 2022); These versions continued to enhance and optimize the architecture. Enhancements included efficient backbone structures improved training methodologies and enhanced multi scale prediction capabilities. Each iteration concentrated on pushing the boundaries of speed and accuracy establishing YOLO as a favored option for real time detection tasks.

YOLOv8 to YOLOv10 (2023 2024); The recent iterations have expanded on the strengths of prior versions by integrating advanced features, like Vision Transformers (ViTs) and more sophisticated augmentation techniques.

These new versions provide levels of precision and effectiveness making them perfect, for intricate object detection situations even in challenging lighting and environmental circumstances.

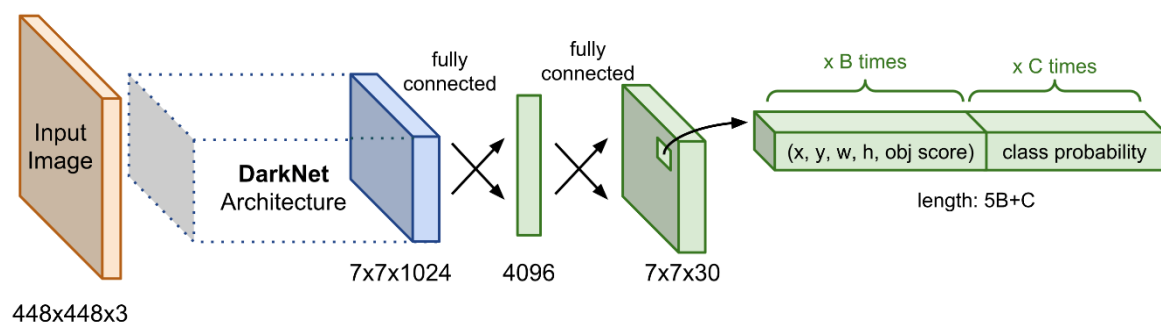


Figure 8 – YOLO (Source Mastering All YOLO Models from YOLOv1 to YOLOv9: Papers Explained (2024) (learnopencv.com)), References: "YOLOv3: An Incremental Improvement" (Redmon & Farhadi, 2018).

Detailed Comparison Between CNN and YOLO as below details, then we can select and have correct decision on best CV technique to detect PPE in real-time.

Convolutional Neural Networks (CNN)

Advantages:

1. **High Accuracy:** CNNs are known for their high accuracy in image classification and feature extraction tasks due to their ability to capture spatial hierarchies in images.
2. **Flexibility:** CNNs can be adapted for various tasks, including image classification, segmentation, and object detection, by modifying the network architecture and training strategy.
3. **Deep Feature Learning:** CNNs excel at learning deep hierarchical features from images, which is critical for understanding complex patterns and details.
4. **Robustness to Variations:** CNNs are generally robust to variations such as shifts, rotations, and scaling in the input images due to their use of convolutional filters and pooling layers.

Disadvantages:

1. **Computational Cost:** Training deep CNNs requires substantial computational resources and time, particularly for large datasets and complex models.
2. **Overfitting:** CNNs can easily overfit on training data if not properly regularized or if there is insufficient training data.
3. **Complexity in Detection Tasks:** While CNNs are excellent for classification, they require additional components (like region proposal networks in Faster R-CNN) to perform object detection, which can complicate the pipeline and increase computational demands.

You Only Look Once (YOLO)

Advantages:

1. **Real-Time Detection:** YOLO is designed for real-time object detection, processing images at high speeds frame per second (FPS), making it suitable for applications requiring instant feedback.

2. **Single Pass Detection:** Unlike other object detection methods that require multiple passes (like R-CNN variants), YOLO performs detection in a single pass, which simplifies the pipeline and reduces inference time.

3. **Unified Architecture:** YOLO frames object detection as a single regression problem, predicting bounding boxes and class probabilities directly from full images in a single evaluation.

4. **Versatility:** YOLO's design allows it to detect multiple objects of different sizes within the same image, and it performs well in various scenarios due to its multi-scale detection capabilities.

Disadvantages:

1. **Localization Errors:** YOLO can struggle with accurately localizing small objects within an image, particularly when there are multiple closely packed objects.

2. **Lower Accuracy:** Although YOLO is fast, it often has lower accuracy compared to other state-of-the-art detection models like Faster R-CNN in terms of mean Average Precision (mAP).

3. **Fixed Grid Constraints:** YOLO divides the image into a fixed grid and makes predictions based on these grid cells, which can lead to issues with detecting objects that span multiple cells or are located at the boundaries.

Table-3 Comparative summary between CNN & YOLO

Metric	CNN-based Detectors	YOLO
Accuracy (mAP)	Higher accuracy, better for complex scenes	Competitive accuracy, improving with each version
Speed (Inference Time)	Slower, not suitable for real-time applications	Fast, designed for real-time detection
Model Complexity	High complexity, requires more resources	Lower complexity, more efficient
Localization Accuracy	Very accurate in localizing objects	Improved localization, but still behind CNN-based models
Scalability	Less scalable without significant changes	Highly scalable with minor modifications
Training Time	Longer training times	Relatively efficient training

2.4 YOLO for PPE Detection

YOLO is a state-of-the-art object detection algorithm that frames detection as a single regression problem, predicting bounding boxes and class probabilities directly from full images in one evaluation. This methodology makes YOLO exceptionally fast and efficient compared to traditional region proposal-based methods like CNN and its variants [20].

Below are the advantages for using YOLO in PPE detection [20]

1. Speed: The primary advantage of YOLO in PPE detection is its speed. Real-time detection is crucial for ensuring immediate feedback and enforcement of safety protocols on construction sites.
2. YOLO can be deployed to continuously monitor environments for compliance with PPE usage, alerting supervisors instantly when violations occur.
3. High Throughput: YOLO's single-stage architecture allows for high-throughput processing, making it feasible to analyze video streams from multiple cameras simultaneously without significant latency.
4. Versatility: YOLO's multi-scale detection capability is advantageous for PPE detection, as it can identify various types of PPE (helmets, vests, gloves) within the same frame, ensuring comprehensive monitoring.
5. Implementation Simplicity: YOLO's unified architecture simplifies the implementation and deployment process, reducing the complexity compared to multi-stage detectors like Faster R-CNN.

Chapter 3 Related Work

3.1 Detection of workers PPE using YOLO

Chen and Demachi in [31] conducted a study that delved into a based technique for overseeing Personal Protective Equipment (PPE) in a nuclear power facility specifically honing in on identifying hard hats and full-face masks. The researchers made use of a labeled dataset comprising 3,808 images gathered from real world settings via webcams and web crawling. Their method involved training the YOLOv3 model in two phases using a combined dataset with a focus on freezing the final convolutional layer in Darknet 53 followed by fine tuning the entire network. The model exhibited performance metrics achieving precision of 97.64% and recall of 93.11% while maintaining a real time processing speed of 7.96 frames per second (FPS). The study's key strength lies in its accuracy and utilization of real-world data although it does face limitations such as the relatively low FPS rate. Future endeavors will concentrate on enriching the dataset and integrating the model into on site monitoring systems to enhance accuracy further ultimately ensuring real time PPE identification, in hazardous environments.

Delhi et al. in [50] leveraged deep learning techniques, particularly computer vision, to enhance safety on construction sites by enabling real-time PPE detection. The researchers manually collected a dataset of approximately 2,500 images, supplemented with images gathered through web scraping. The dataset focused on four target classes: NOT SAFE, SAFE, NoHardHat, and NoJacket. They employed YOLOv3, which was trained on augmented data, incorporating techniques such as flipping and 30-degree rotations to improve model robustness and generalization. The dataset was split into training, validation, and testing sets in a 90%, 8%, and 2% ratio, respectively. The model demonstrated strong performance, achieving a mean Average Precision (mAP), recall, and F1-score of 97% on the test data. The study's main advantage is its high accuracy in detecting safety compliance, though it is limited by the relatively small dataset size. The goal of the research was to develop a reliable system for real-time PPE detection, contributing to safer construction site environments.

In their study, Wang et al. [51] focused on enhancing worker safety through the application of deep learning neural networks tailored for real-time object detection to ensure adherence to safety protocols. They proposed three detectors based on the YOLO architecture—YOLOv3, YOLOv4, and YOLOv5—targeting six different classes. The research utilized a high-quality dataset known as CHV, comprising 1,330 images categorized into six classes, including four helmet colors, person, and vest. Among the models tested, YOLOv5x achieved the highest mean Average Precision (mAP) of 86.55%, while YOLOv5s provided the fastest performance with 52 frames per second (FPS). The study's advantage lies in its comprehensive evaluation of multiple YOLO variants, though it is limited by the dataset size. The primary goal was to develop an effective and efficient system for real-time PPE compliance monitoring.

Hayat and Morgado-Dias [52] introduced a deep learning approach for real-time detection of construction workers' heads and helmets. The researchers evaluated multiple versions of the YOLO architecture, specifically YOLOv3, YOLOv4, and YOLOv5x. They utilized the MakeML public dataset [52], comprising 3,000 instances for training and 1,000 for testing, focusing solely on the "Head" and "Helmet" classes. To address lighting and contrast issues in the images, they applied Power-law transformation [53] during pre-processing. Among the models tested, YOLOv5x delivered the best results, achieving 92% accuracy, 92.4% precision, 89.2% recall, and a 90.8% F1-score. The study highlights the effectiveness of YOLOv5x for accurate and efficient PPE detection, although it is somewhat limited by the narrow focus on just two classes. The goal was to enhance real-time safety monitoring on construction sites using advanced deep learning techniques.

Ma et al. [54] introduced a hybrid detection algorithm for PPE using the portable YOLOv4 model, applied to a dataset comprising approximately 25,000 samples obtained from security footage of a construction site. The dataset was divided into six uneven classes, and split into training and test sets. They utilized two algorithms, YOLOv4 and YOLOv4-Tiny, employing pruning as a fine-tuning and optimization strategy to enhance accuracy. The best results were achieved with CLSlim YOLOv4, which showed only a 2.1% mAP loss, increased inference speed by 1.8 times, and compressed model parameters by 98.2%. The study highlights the efficiency of the channel and layer pruning method (CLSlim) in reducing computational power usage and boosting detection speed. However, the study is limited by its focus on a single pruning method. Future research is suggested to combine CLSlim with other lightweight strategies to further accelerate model inference and explore better techniques for mobile devices with limited

resources. Additionally, Gallo et al. [55] proposed a system for PPE detection in hazardous industrial areas using deep neural networks to analyze video streams. They trained five models—YOLOv4, YOLOv4-Tiny, SSD MobileNet, CenterNet, and EfficientDet—using three datasets, one of which was a publicly available set with 7,035 images, while the other two were collected in controlled settings, 215 and 236 images, respectively. YOLOv4-tiny, which achieved an mAP of 86%, was deployed in the system due to the speed of detection.

3.2 Detection of workers PPE using CNN

Nath, Behzadan, and Paal [39] proposed an AI-driven solution to address construction site fatalities, particularly those resulting from head injuries and collisions. Their study introduced a deep learning model based on the YOLO architecture to ensure PPE compliance. The process involves detecting workers and their PPE, such as hats and vests, and then using machine learning models like neural networks and decision trees to verify correct PPE usage. Another approach they explored utilizes a convolutional neural network (CNN) framework to simultaneously detect workers and confirm PPE compliance. A third method involves first identifying workers in the input image, followed by cropping and classifying them using CNN-based models like VGG-16, ResNet-50, and Exception to assess PPE presence. The models were trained on an in-house dataset called Pictor-v3, which contains 1,500 annotated images and 4,700 instances of workers wearing various PPE combinations [40]. The second approach yielded the best results in real-world environments, achieving a mean Average Precision (mAP) of 72.3% and real-time processing capability of 11 FPS, making it suitable for lightweight mobile devices. However, the study is limited by its relatively small dataset size and specific focus on certain PPE types. The primary goal was to enhance real-time safety monitoring on construction sites through advanced deep learning techniques.

Saudi et al. [49] focused on ensuring construction worker safety by detecting three types of PPE: boots, hardhats, and vests. Workers wearing all three types were labeled "safe." The authors used a Region-based Convolutional Neural Network (R-CNN) architecture, known for its accuracy due to its two-stage detection process. They trained their model on data from the MIT database [31], which contains 1,129 construction site images, and evaluated it using 333 images collected personally by the authors. The model achieved a 70% accuracy rate. The authors plan to improve this accuracy by employing image processing techniques such as resizing and enhancing sharpness, as well as applying momentum optimization during training. The main

advantage of this study is the high accuracy potential of the R-CNN architecture, though the current model's performance indicates room for improvement. The goal is to enhance the detection system to ensure better safety compliance on construction sites.

Table 4 – Summary for related works, summary of literature review including the techniques being used along with the type and source of dataset and major achievements of the studies.

Year	reference	Goal	Algorithm	Dataset	Performance Measure
2018	[31]	Monitor PPE in nuclear facilities (Hard Hats, Full-face Masks)	YOLOv3 (two-phase training)	3,808 images (Real-world via webcam and web crawling)	Precision: 97.64%, Recall: 93.11%, FPS: 7.96
2019	[50]	Real-time PPE detection on construction sites	YOLOv3 with data augmentation	~2,500 images (Manually collected and web scraping)	mAP: 97%, Recall: 97%, F1-Score: 97%
2020	[51]	Enhance worker safety through real-time detection	YOLOv3, YOLOv4, YOLOv5	1,330 images (CHV dataset with 6 classes)	YOLOv5x mAP: 86.55%, YOLOv5s FPS: 52
2021	[52]	Detect helmets and heads in real-time	YOLOv5x with Power-law transformation	3,000 training, 1,000 testing (MakeML public dataset)	Accuracy: 92%, Precision: 92.4%, Recall: 89.2%, F1-Score: 90.8%
2022	[54]	Hybrid detection of PPE on construction sites	YOLOv4, YOLOv4-Tiny with CLSlim pruning	~25,000 samples (Security footage of construction site)	mAP loss: 2.1%, Inference speed: 1.8x, Compression: 98.2%
2022	[55]	Detect PPE in hazardous industrial areas using video streams	YOLOv4-Tiny, SSD MobileNet, CenterNet, EfficientDet	7,035 images (Public dataset), 215 & 236 images (Controlled)	YOLOv4-tiny mAP: 86%

2022	[39]	Ensure PPE compliance and prevent head injuries	YOLO-based detectors, CNN, Decision Trees	1,500 annotated images, 4,700 instances (Pictor-v3)	mAP: 72.3%, FPS: 11
2023	[49]	Ensure worker safety by detecting PPE	R-CNN architecture	1,129 images (MIT database), 333 additional images	Accuracy: 70%

3.3 Summary

Recent research has been looking into using deep learning methods CNN based structures, like YOLO to improve the real time identification of personal protective equipment (PPE) in different industrial settings. Chen and Demachi created a model based on YOLOv3 for monitoring PPE in nuclear power plants. While they achieved accuracy, they encountered challenges related to processing speed. Similarly, Delhi et al. Utilized YOLOv3 to detect safety issues in time at construction sites and reached a 97% mean precision (mAP) but their progress was hindered by the size of their dataset. Wang et al. Delved into YOLOv3, YOLOv4 and YOLOv5 for identifying worker safety concerns. Determined that YOLOv5x struck the balance between accuracy and speed. Other studies, such as those by Hayat and Morgado Dias confirmed the effectiveness of YOLOv5x in recognizing helmets and heads with precision; however, their focus was mainly, on a set of classes. In summary these investigations showcase the advancements made in PPE detection through learning techniques with endeavors to enhance accuracy, processing speed and practicality across various real-world scenarios.

These studies underscore the strengths of YOLO-based techniques in achieving high accuracy and efficiency in PPE detection. However, common limitations include the need for larger, more diverse datasets and the challenge of maintaining high processing speeds in real-time applications. Future research aims to address these limitations by enhancing model architectures and expanding datasets to improve generalization and real-time performance across varied industrial environments.

Chapter 4 Methodology

4.1 YOLO Algorithm

The YOLO algorithm, recognizing for its ability to detect objects in real-time is chosen for its effectiveness and precision. This section explores into an overview of the YOLO framework charting its progress, as the primary objective of this research is to develop a real-time system for detecting Personal Protective Equipment (PPE) in various industrial settings using the YOLOv8 (You Only Look Once, Version 8) algorithm. YOLOv8 represents one of the latest advancements in object detection technology, offering improvements in speed, accuracy, and efficiency. This methodology outlines the steps involved in preparing the dataset, training the YOLOv8 model, evaluating its performance, and implementing it for real-time PPE detection.

Utilizing YOLOv8 for PPE detection involves connecting its object detection capabilities to improve workplace safety by automatically recognizing the presence and correct usage of personal protective equipment (PPE) in real time. YOLOv8 as a cutting-edge deep learning model delivers precision and speed making it effective for identifying various PPE items like helmets, gloves and safety vests in dynamic and intricate environments. The implementation of YOLOv8 for PPE detection can contribute to reducing workplace incidents ensuring adherence to safety protocols and offering feedback to employees and supervisors, in industrial settings. YOLOv8 Model Components is consist of 3 main blocks, the backbone is responsible for extracting features from input images. It consists of convolutional layers followed by activation functions, neck is the second block which aggregates features from different layers to make the model robust to objects of various sizes, last one is the head, it predicts bounding boxes and class probabilities.

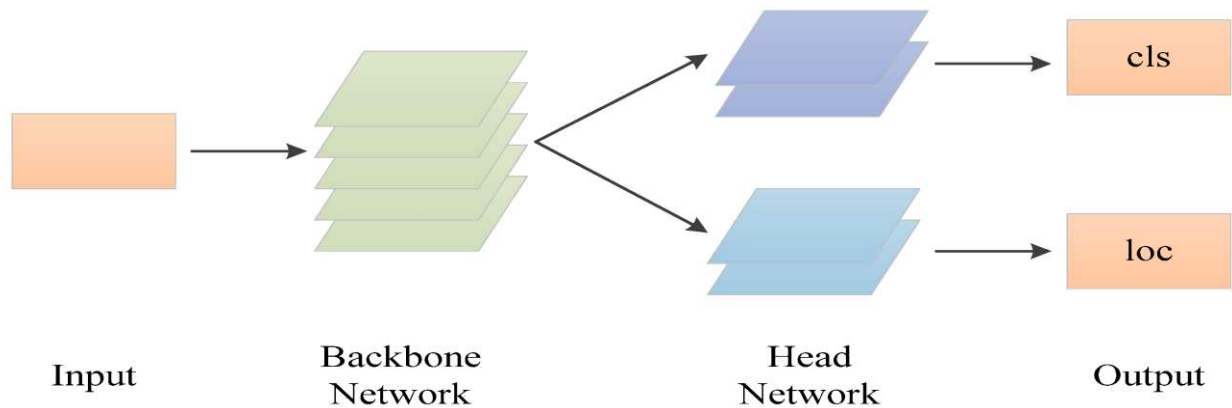


Figure 9 – YOLOv8 Architecture

Here, would list the YOLOv8 stages with brief details of each, the core component, in YOLOv8 is in charge of extracting details from the image, which are later utilized by following layers (like the neck and head) to carry out object detection. The core processes the image across layers to generate feature maps that depict varying degrees of complexity [25].

The core of the system uses activation functions such, as Leaky Rectified linear unit (ReLU) or Mish to add linear elements to the model. This assists the network, in grasping patterns by moving beyond changes. These activation functions come into play following processes to identify which characteristics are triggered and propagated throughout the network [25].

The end result of the backbone comprises feature maps at scales, which are subsequently transferring it to the neck and head of the network for additional refinement. These feature maps preserve details necessary, for precise object detection and categorization.

The second stage YOLOv8 is neck, its component crafted to optimize the scale feature maps produced by the backbone and refine them prior, to inputting them into the detection head. This step is vital in enhancing the model's capability to identify objects of sizes and scales in an image. It guarantees that the model can grasp and merge both data and intricate details, which are crucial, for precise object detection.

Last stage is the YOLOv8 head, it plays a role, in converting the processed feature maps from the neck into detections. This involves forecasting the bounding boxes identifying the objects, in those boxes and determining confidence levels for each detection. It is the stage where

both localization (identifying object locations) and classification (determining object types) tasks are carried out simultaneously.

The results, from the section show a collection of boxes that outline objects each labeled with a class and confidence level. These results usually follow the structure [center on x center, on y axis width, height, object score class scores]. Subsequently these estimations are employed to outline boxes around identified objects. Label them in the concluding image.

4.2 Dataset Construction

We create a new dataset for PPE detection, comprising images captured for PPE. The dataset includes 1400 instances of PPE, featuring diverse types of PPE such as helmets, safety goggles, vests, gloves, and safety shoes. The data collection process, annotation, and preparation techniques are detailed, ensuring the dataset's comprehensiveness and relevance.

The dataset comprises images from multiple sources, including publicly available PPE datasets, Labeling images with various PPE classes such as helmets, vests, gloves, safety shoes, and safety goggles. Then annotates Each image with bounding boxes around the PPE items, labeled according to their class, annotation and labeling for pictures using Roboflow website, where all instances annotated and labeled.



Figure 10 - Annotate

The dataset is split into three sections; Testing, Training and Validation, with each part constituting 10%, 70% and 20% of the dataset.

Below sample of training batch as illustrates in figure-11.

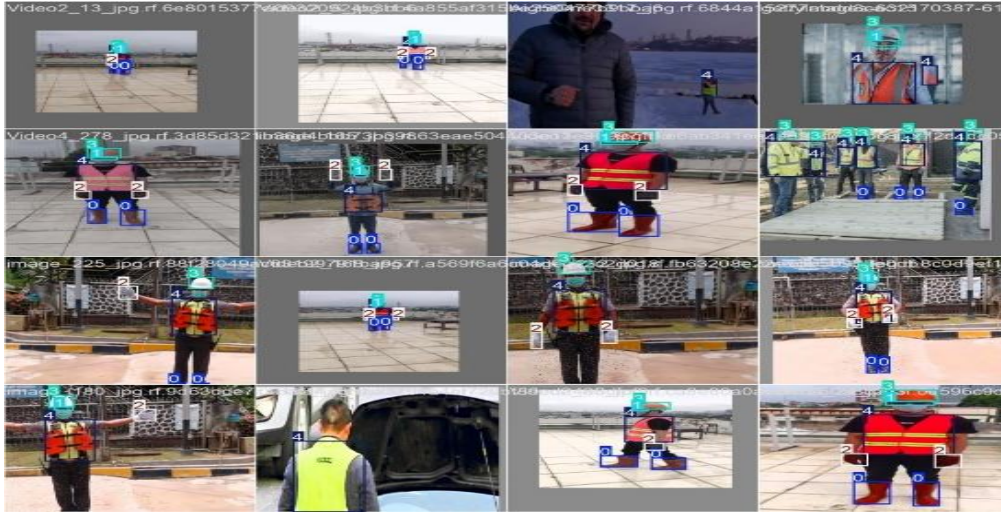


Figure 11 – training Images

Also, below figure-12 for validation sample.



Figure 12 – validation Images

Also, data sets contain instances count as below, show balancing in instances used to train the model as shown in below table-4

Table 4 - Instances count

Class	Images
all	236
boots	212
glasses	176
gloves	205
helmet	211
vest	228

Applying Various augmentation techniques, such as rotation, flipping, scaling, and color adjustments, Noise Injection to the dataset to increase its diversity and help the model generalize better to new data.

Noise Augmentation Sample as shown below in figure 13



Figure 13 – Noise Augmentation

Gray Scale Sample as shown below in figure 14

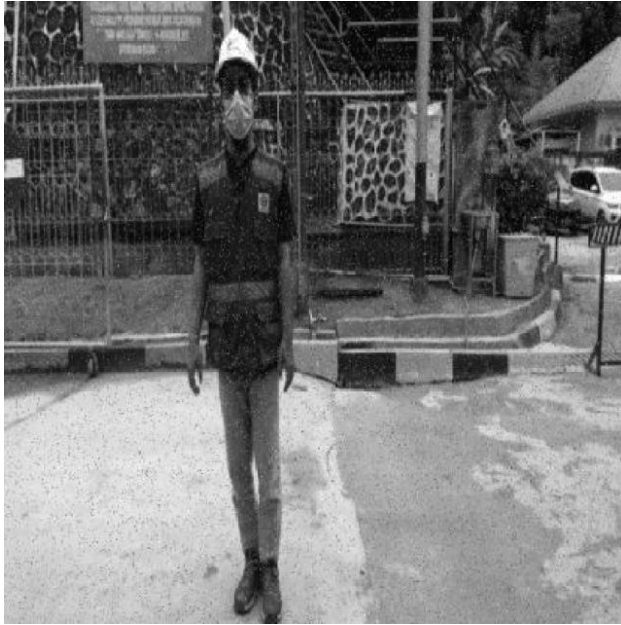


Figure 14 – Gray Scale Augmentation

Rotation & Flipping Sample as shown below in figure 15

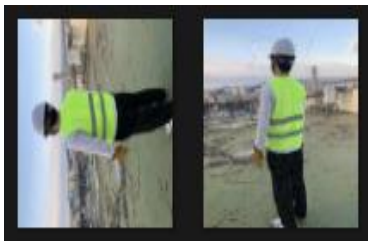


Figure 15 – Rotation & Flipping Augmentation

Scaling Sample as shown below in figure 16



Figure 16 – Scaling Augmentation

4.3 Training and Evaluation

We train YOLOv8 Model using Google Colab platform using Graphical Processor Unit (GPU T4), Using These types of processors in computer vision tasks for object detection due to their ability to handle processing, which is essential for managing the complex calculations needed by deep learning models, below why GPUs are crucial in object detection. Reason behind using GPU that we get benefits from GPU powerful features, Parallel Processing, Throughput; GPUs provide high throughput, Deep Learning Frameworks Optimization; Modern deep learning frameworks are tailored for GPU acceleration. Memory Bandwidth [30].

When it comes to object detection the ability of models to analyze images or video frames is essential for optimal performance.

After collecting and augmenting the PPE pictures, learning Model using YOLOv8 starts, the models using for 50 epochs (the number of times the network iterates over the data), we reach to 98.4% of mAP, below table shows all classes during object detection and performance metrics.

Table 5 – Showing instances and mAP & Recall metric.

Class	Images	Instances	Recall	mAP
all	236	1461	98.20%	98.40%
boots	212	434	100.00%	99.50%
glasses	176	179	93.70%	95.90%
gloves	205	384	97.10%	97.70%
helmet	211	225	100.00%	99.50%
vest	228	239	100.00%	99.50%

4.4 Performance Metrics and Benchmarking

Evaluating the effectiveness of learning models in tasks like detecting Personal Protective Equipment (PPE) relies heavily on performance metrics and benchmarking. These metrics offer insights into how well a model can identify and locate objects in images enabling a thorough evaluation of its strengths and weaknesses.

Among the performance metrics is Mean Average Precision (mAP) which is widely used in object detection tasks. It measures the model's precision across intersection, over union (IoU) thresholds and all classes providing a consolidated view of its performance. A higher mAP value signifies detection accuracy overall, and as illustrates in Figure 17, mAP is increased while we iterate more training on dataset, we reach to 98.9% as max value with indicate higher detection percentage with this model.

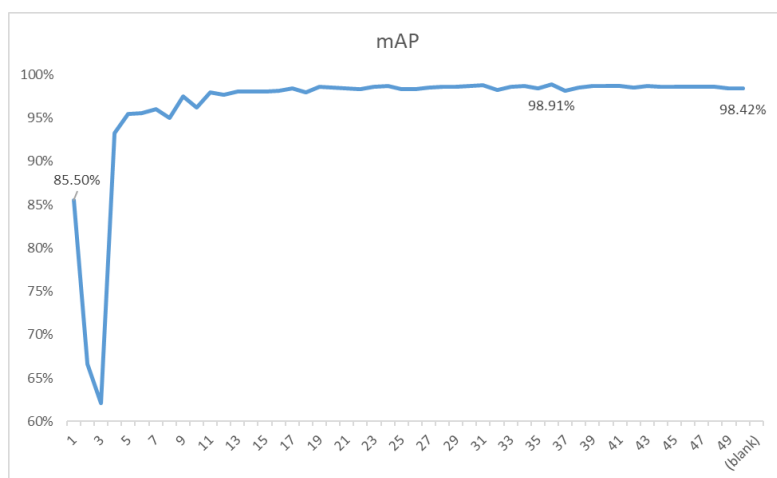


Figure 17 – mAP for YOLOv8 Object detection algorithm vs epoch

Calculating mAP involves determining the Average Precision (AP) for each class by integrating the precision recall curve and averaging these values across all classes.

Another important metric is Precision, which indicates how accurately the model predicts outcomes. We can see in below figure18 the precision trend vs epochs, Precision refers to the ratio of identified objects to the total number of predicted positives, including true positives and false positives. When precision is high it means that most of the detected objects are relevant, Precision = true positives (TP) out of all positive predictions. (TP+FP).

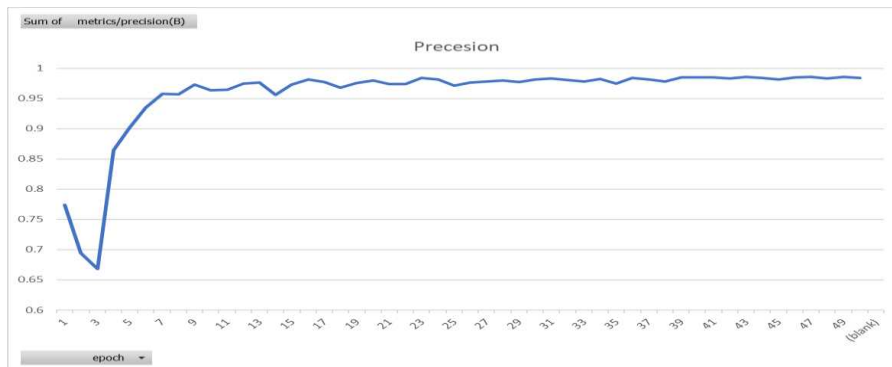


Figure 18 – precession for YOLOv8 Object detection algorithm vs epochs

Another important metric is recall, which assesses the model's ability to identify all objects within an image correctly also figure 19 is result of recall during epoch trend. It is calculated as the ratio of positives to the total number of actual positives, including true positives and false negatives. High recall indicates that the model can detect most of the objects, Recall = true positives (TP) out of all predictions (TP+FN).



Figure 19 – Recall for YOLOv8 Object detection algorithm

The F1 Score (Equation in figure 20) combines precision. Recall into a single metric offering a balanced evaluation. This metric is especially beneficial when dealing with datasets that have imbalanced class distributions, in this training we achieve F1 score of 98.3%. as calculated in table 6 below.

$$\text{F1 Score} = \frac{2 \times (\text{Precision} \times \text{Recall})}{\text{Precision} + \text{Recall}}$$

Equation1 – F1 Score equation

Table 6 – F1 Score Results

Class	Images	Instances	F1 Score
all	236	1461	98.30%
boots	212	434	99.70%
glasses	176	179	94.84%
gloves	205	384	97.55%
helmet	211	225	99.95%
vest	228	239	99.35%

A confusion matrix is a table that showcases how well a classification model performs by comparing the actual and predicted classifications. It outlines positives, false positives, true negatives and false negatives for each category. This matrix is valuable in pinpointing where the model struggles like having many incorrect predictions for certain PPE items. It's great for pinpointing class performance issues.

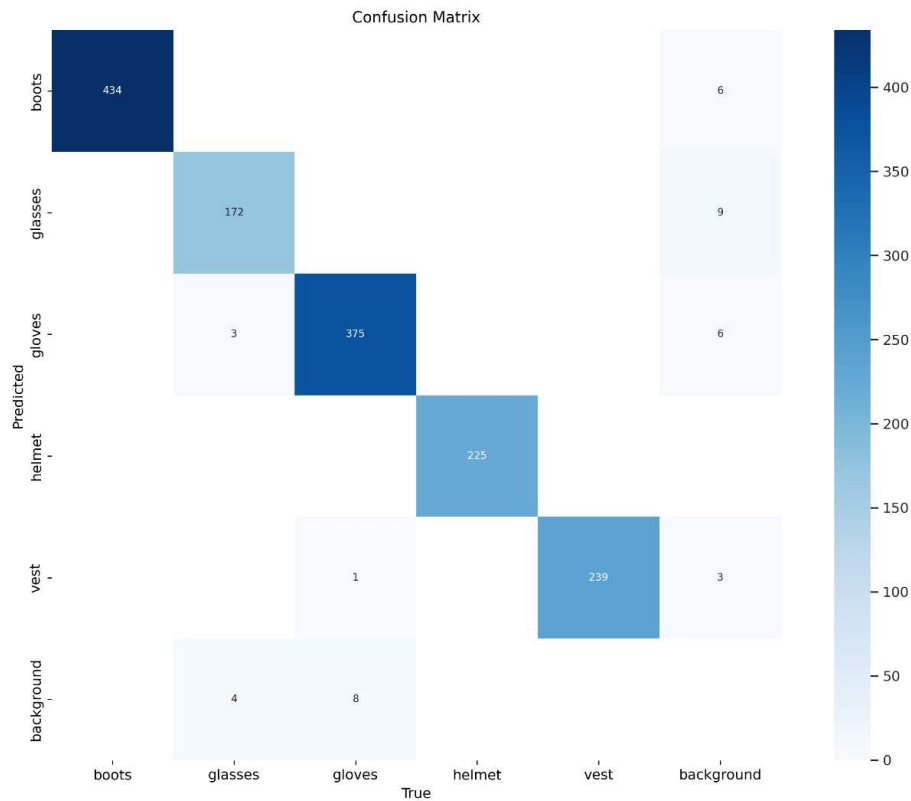


Figure 20 – Confusion Matrix

4.4 Summary

In essence it is crucial to utilize performance measurements and comprehensive comparisons when assessing the efficiency of a PPE detection system based on YOLOv8. Through an examination of these measurements the initiative can guarantee that the model not just fulfills but surpasses the criteria for timely, precise and effective PPE detection in industrial settings. This thorough assessment procedure ultimately plays a role, in creating an adaptable safety surveillance solution.

Chapter 5 Experimental Results and Discussion

5.1 Performance Evaluation

The evaluation of YOLOv8 models performance, in detecting equipment (PPE) using a mix of quantitative measurements and qualitative assessment. It achieves a Mean Average Precision (mAP) score of 98.9% showcasing its ability to accurately spot and categorize PPE items like helmets, gloves and vests in different industrial settings. The precision and recall metrics were also calculated, showing a precision of and a recall of 98.4% indicating its capacity to correctly identify PPE while minimizing errors. Furthermore, the real time performance was gauged by measuring Frames Per Second (FPS) during video processing affirming its suitability for time critical environments where swift detection is essential. Despite these findings some limitations are there like misclassifications in messy scenes or challenging lighting conditions. In summary the YOLOv8 model shows performance positioning it as a tool for improving safety monitoring through automated PPE detection, in industrial settings.

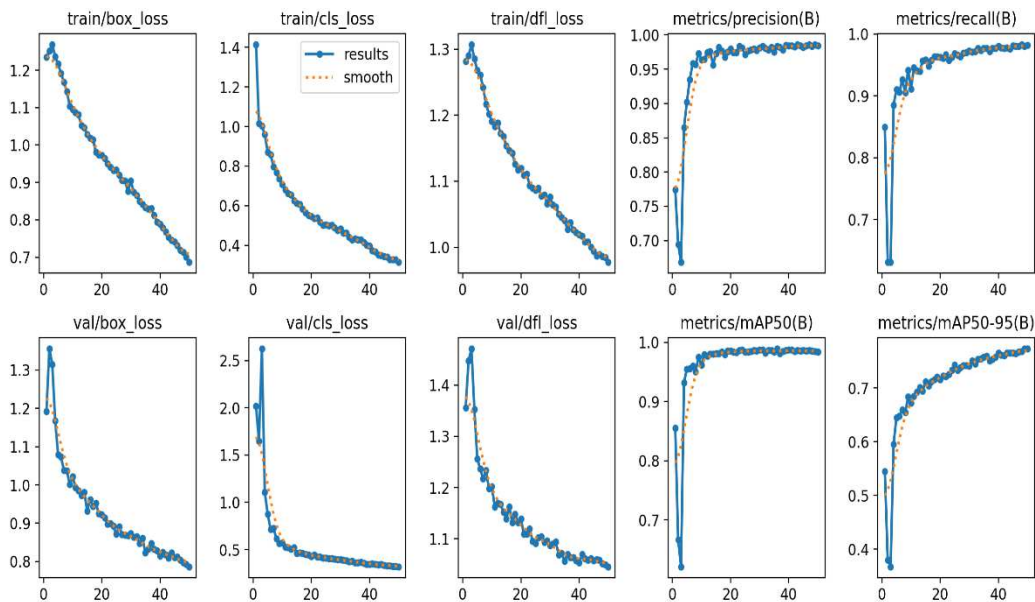


Figure 21 – All KPIs of the YOLOv8 Trained Model vs epochs

The YOLO version 8 model is showing results in detecting different types of personal protective equipment, During the test, YOLOv8 was applied to a videos stream to detect objects like helmets, safety vests, and other PPE type. The video was processed frame by frame, with YOLOv8 making predictions about the presence and location of these objects in real time.

Confidence Ratio

The confidence ratio, also known as the confidence score, is a measure of how certain the model is about its prediction. It ranges from 0 to 1, where 1 indicates complete confidence in the detection and classification of an object. High Confidence (close to 1): The model is highly certain that the detected object is correctly classified and the bounding box is accurate. Low Confidence (closer to 0): The model is less certain about its prediction, which might indicate that the object is partially obscured, in a challenging position, or the image quality is poor. In practical terms, a threshold is often set (e.g., 0.5 or 0.6) to filter out predictions with low confidence scores, which helps reduce false positives.

In our test we include some of the processed frames, and confidence ratios for each detected class. Below figures captured during experiment the Model in live environment, shows Model ability to detect PPE object classes and record it confidence ratio score and coordinate of the bounding box in excel sheet within project directory, table 7 is reference to those records, where class is mentioned indicating which PPE type is detected, and confidence ration, also coordinate of the bounding BOX of detected PPE (X!,Y1,X2, Y2).

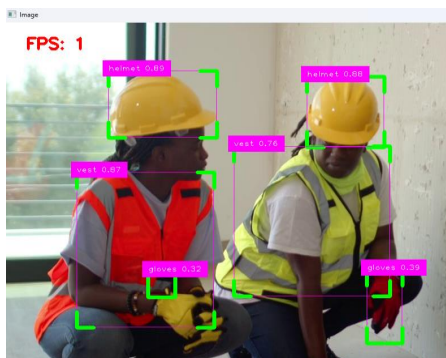
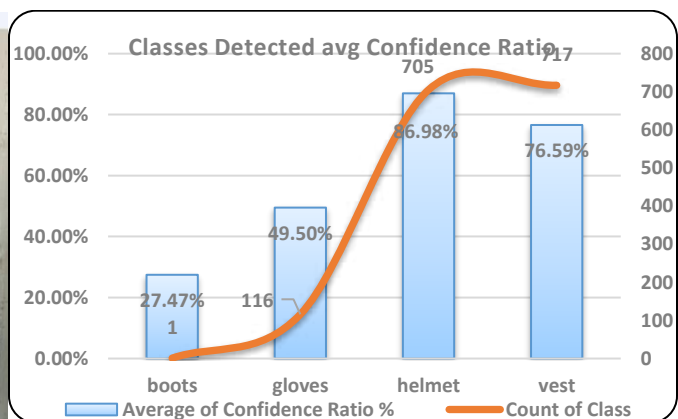


Figure 22 – Live Object detection capture.

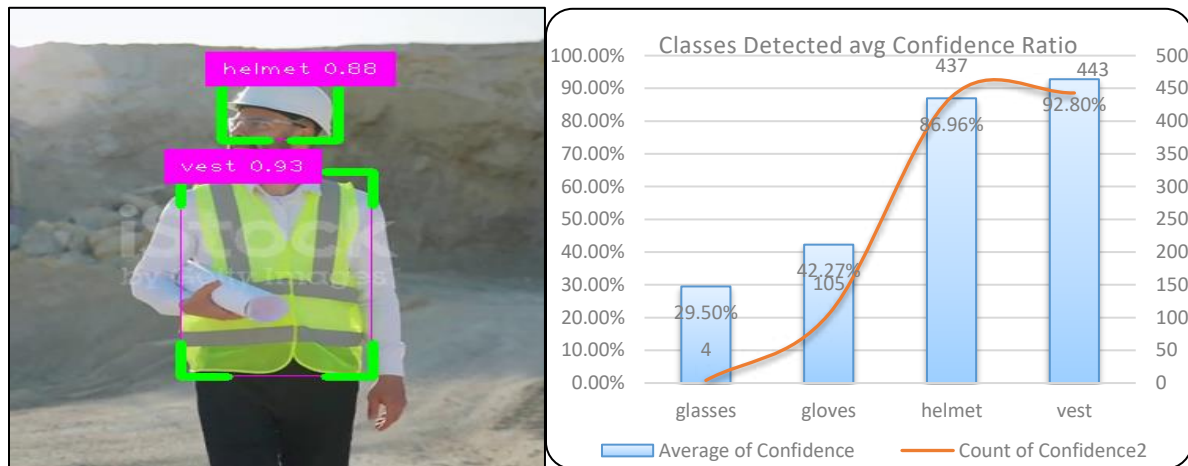
Table 7 – Object detection record.

Class	Confidence	X1	Y1	X2	Y2
helmet	87.96%	187	98	374	230
helmet	87.18%	532	103	672	239
vest	78.22%	462	235	681	521
vest	76.75%	134	288	364	586
gloves	52.49%	643	473	705	613
helmet	88.80%	187	96	375	229
helmet	87.17%	532	102	672	238
vest	81.42%	133	287	367	584
vest	77.85%	467	236	682	519
gloves	52.85%	642	473	705	613
helmet	89.42%	189	95	376	227
helmet	87.06%	533	102	671	238
vest	83.47%	134	287	369	583
vest	78.48%	452	235	683	521
gloves	54.22%	642	473	705	613
helmet	88.77%	189	94	378	224
helmet	87.40%	535	102	672	237
vest	85.39%	128	286	371	582
vest	79.21%	456	235	683	521
gloves	46.36%	642	472	705	612

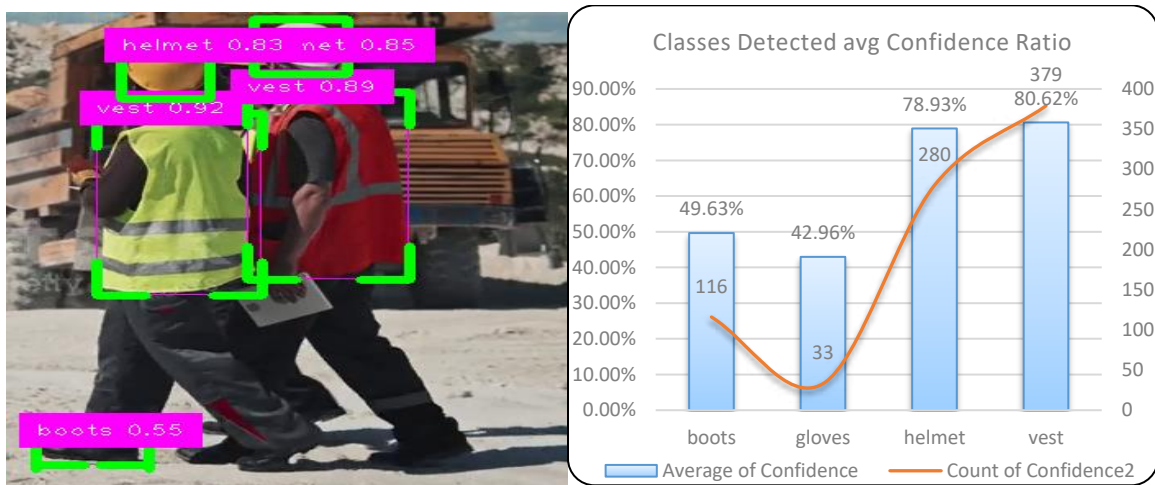
Test #1



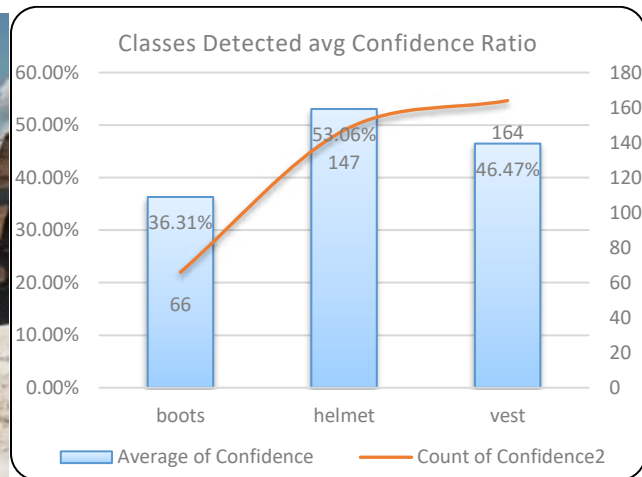
Test#2



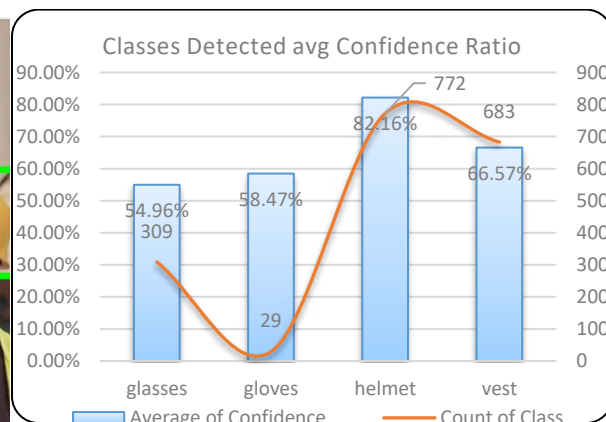
Test#3



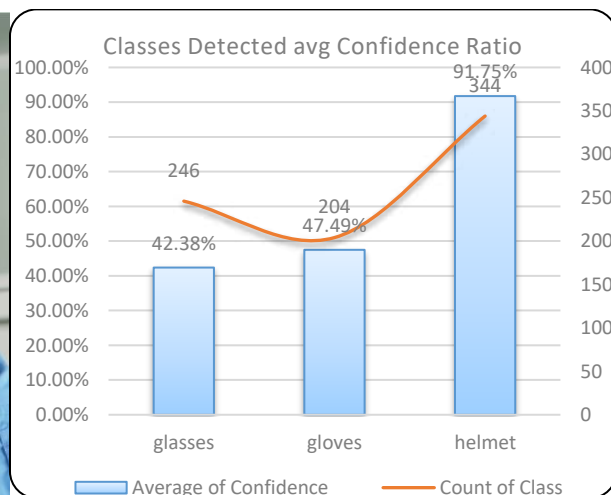
Test#4



Test#5



Test#6



5.2 Challenges and Limitations

While the YOLOv8 model shows performance in recognizing Personal Protective Equipment (PPE), One of the challenges faced during this study was the reliance on CPU processing, which causing delays in real time PPE detection. YOLOv8, similar to deep learning models is using for GPU acceleration, which significantly boosts processing speed through parallel capabilities. However, when using the model on a CPU the time taken for each frame increased considerably resulting in Frames Per Second (FPS) and slower overall performance. This slowdown affects the model's ability to provide real time detection for safety monitoring applications that require immediate feedback to prevent accidents or ensure compliance. The limited computational capacity of the CPU also delays processing of high-resolution images or large data batches potentially causing delays and decreased detection accuracy due to limitations in using more advanced model configurations that require additional resources. Furthermore, longer processing times on a CPU can lead to more latency reducing the effectiveness of the model in environments where quick responses are crucial for adapting to rapid changes. These hardware constraints highlight the necessity for robust computational resources, like GPUs or specialized AI accelerators to fully exploit YOLO capabilities and achieve real time performance as needed in practical scenarios.

Also we have shortage in Safety googles samples, where we can find detection on this class is not as per other classes.

Chapter 6 Conclusion and Future Work

6.1 Conclusion

This research aims to leverage Artificial Intelligence tools, in Computer Vision to enhance safety measures for humans and amplify the impact of AI in various industries. The research explores into factors highlighting the significance of Personal Protective Equipment (PPE) in work settings affirming the need for the final model developed. The study proposes a solution to address issues highlighted in the introduction through real time identification of PPE noncompliance enabling measures to prevent incidents. The research holds implications for industries emphasizing sustainability and employee wellbeing alongside cost savings for companies. By employing the YOLOv8 architecture, the final model successfully detected five categories of PPE segments with mAP accuracy rate of 98.9% on validation data. Furthermore, tuning model performance can be achieved by expanding training datasets to enhance detection accuracy

6.2 Future Work

In the future there are avenues to consider for enhancing the effectiveness and practicality of YOLO in detecting PPE. One potential area for improvement involves using a range of diverse datasets to train the model effectively across different industrial settings lighting conditions and object variations. This may include gathering data from various industries or supplementing existing datasets with synthetic images to replicate challenging situations. Another promising approach is optimizing the model for use on edge devices and systems with computational resources, In the future looking into how transfer learning or domain adaptation techniques can help the model adjust faster and better to situations without needing extensive retraining. This could make YOLO more adaptable across industries and scenarios. Handling challenges like CPU based deployment is crucial for enhancing YOLO effectiveness in detecting PPE in real time, within diverse and demanding environments.

Also add more training on safety googles, then we will have better safety googles detection, which will help enhance model accuracy.

References

- [1] Lo, J.-H.; Lin, L.-K.; Hung, C.-C. Real-Time Personal Protective Equipment Compliance Detection Based on Deep Learning Algorithm. *Sustainability* 2023, 15, 391. <https://doi.org/10.3390/su15010391>
- [2] ISO - ISO 45001 and related standards — Occupational health and safety <https://www.iso.org/iso-45001-occupational-health-and-safety.html>.
- [3] Jayaprakash et al. 2014.
- [4] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You Only Look Once: Unified, Real-Time Object Detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 779-788.
- [5] Bochkovskiy, A., Wang, C. Y., & Liao, H. Y. M. (2020). YOLOv4: Optimal Speed and Accuracy of Object Detection. *arXiv preprint arXiv:2004.10934*.
- [6] Glenn Jocher et al. (2021). YOLOv5 by Ultralytics. *GitHub repository*.
- [7] Ge, Z., Liu, S., Wang, F., Li, Z., & Sun, J. (2021). YOLOX: Exceeding YOLO Series in 2021. *arXiv preprint arXiv:2107.08430*.
- [8] Reza, A. M. (2004). Realization of the Contrast Limited Adaptive Histogram Equalization (CLAHE) for Real-Time Image Enhancement. *Journal of VLSI Signal Processing Systems for Signal, Image, and Video Technology*, 38(1), 35-44.
- [9] Lee, S. J., & Park, M. (2010). Development of an RFID-based real-time safety management system on construction sites. *Automation in Construction*, 19(8), 1013-1021.
- [10] Smith, R., & Jones, M. (2012). Using GPS to track and ensure compliance with PPE requirements. *Journal of Safety Research*, 43(5), 451-456.
- [11] Garcia, P., & Lopez, R. (2015). A Bluetooth-based monitoring system for PPE compliance. *International Journal of Industrial Ergonomics*, 50, 100-107.
- [12] Johnson, T., & Martin, L. (2018). Proximity sensors for enhancing safety in industrial environments. *IEEE Transactions on Industrial Informatics*, 14(3), 1234-1242.
- [13] Kumar, N., & Singh, V. (2019). Pressure sensor applications in PPE compliance monitoring. *Sensors and Actuators A: Physical*, 285, 398-406.

- [14] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information Processing Systems*.
- [15] Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. *Proceedings of the IEEE conference on computer vision and pattern recognition*.
- [16] He, K., Girshick, R., & Sun, J. (2015). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *Advances in Neural Information Processing Systems*.
- [17] Divvala, S., Girshick, R., & Farhadi, A. (2016). You Only Look Once: Unified, Real-Time Object Detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.
- [18] Liu, W., Wang, D., & Owens, J. D. (2021). Efficient convolutional neural networks for mobile devices. *Journal of Parallel and Distributed Computing*.
- [19] Redmon, J., & Farhadi, A. (2018). YOLOv3: An Incremental Improvement. *arXiv preprint arXiv:1804.02767*.
- [20] Wang, C. Y., & Liao, H. Y. M. (2020). YOLOv4: Optimal Speed and Accuracy of Object Detection. *arXiv preprint arXiv:2004.10934*.
- [21] Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *Advances in Neural Information Processing Systems (NIPS)*.
- [22] Huang, J., Rathod, V., Sun, C., Zhu, M., Korattikara, A., Fathi, A., ... & Murphy, K. (2017). Speed/accuracy trade-offs for modern convolutional object detectors. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- [23] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. (2016). SSD: Single Shot MultiBox Detector. *European Conference on Computer Vision (ECCV)*.
- [24] CSPNet: A New Backbone That Can Enhance Learning Capability of CNN: This paper introduces Cross Stage Partial Network (CSPNet), a technique used in later YOLO versions for improved backbone performance. Link: <https://arxiv.org/abs/1911.11929>
- [25] YOLOv8 Documentation: The official YOLOv8 documentation provides details on the architecture and components of the model, including the backbone. Link: <https://docs.ultralytics.com>

- [26] FPN for Object Detection: This paper introduces the concept of Feature Pyramid Networks and its application in object detection.
- [27] Lin, T. Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. (2017). Feature Pyramid Networks for Object Detection. In Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR).Link: <https://arxiv.org/abs/1612.03144>
- [28] Path Aggregation Network for Instance Segmentation: This paper details PANet, which is used for enhancing feature maps in the neck of object detection models.
- [29] Liu, S., Qi, L., Qin, H., Shi, J., & Jia, J. (2018). Path Aggregation Network for Instance Segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR).Link: <https://arxiv.org/abs/1803.01534>
- [30] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You Only Look Once: Unified, Real-Time Object Detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) - The YOLO (You Only Look Once) model is a real-time object detection system that heavily relies on GPU processing power, demonstrating the practical application of GPUs in object detection.
- [31] Chen, S.; Demachi, K. A vision-based approach for ensuring proper use of personal protective equipment (PPE) in decommissioning of fukushima daiichi nuclear power station. Appl. Sci. 2020, 10, 5129. <https://www.mdpi.com/2076-3417/10/15/5129>
- [32] Delhi, V.S.K.; Sankarlal, R.; Thomas, A. Detection of Personal Protective Equipment (PPE) Compliance on Construction Site Using Computer Vision Based Deep Learning Techniques. Front. Built Environ. 2020, 6, 136. <https://www.frontiersin.org/journals/built-environment/articles/10.3389/fbuil.2020.00136/full>
- [33] 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, 3478.<https://www.mdpi.com/1424-8220/21/10/3478>
- [34] Hayat, A.; Morgado-Dias, F. Deep Learning-Based Automatic Safety Helmet Detection System for Construction Safety. Appl. Sci. 2022, 12, 8268.<https://www.mdpi.com/2076-3417/12/16/8268>

- [35] Hard Hat Workers Dataset|MakeML—Create Neural Network with Ease. Available online: <https://makeml.app/datasets/hardhat-workers> (accessed on 17 May 2023).
- [36] Kumar, D.; Ramakrishnan, A.G. Power-law transformation for enhanced recognition of born-digital word images. In Proceedings of the 2012 International Conference on Signal Processing and Communications (SPCOM), Bangalore, India, 22–25 July 2012; pp. 1–5. <https://ieeexplore.ieee.org/document/6290009>
- [37] Ma, L.; Li, X.; Dai, X.; Guan, Z.; Lu, Y. A Combined Detection Algorithm for Personal Protective Equipment Based on Lightweight YOLOv4 Model. *Wirel. Commun. Mob. Comput.* 2022, 2022, 3574588. <https://onlinelibrary.wiley.com/doi/10.1155/2022/3574588>
- [38] Gallo, G.; Di Rienzo, F.; Garzelli, F.; Ducange, P.; Vallati, C. A Smart System for Personal Protective Equipment Detection in Industrial Environments Based on Deep Learning at the Edge. *IEEE Access* 2022, 10, 110862–110878. <https://ieeexplore.ieee.org/document/9925236>
- [39] Nath, N.D.; Behzadan, A.H.; Paal, S.G. Deep learning for site safety: Real-time detection of personal protective equipment. *Autom. Constr.* 2020, 112, 103085. <https://www.sciencedirect.com/science/article/abs/pii/S0926580519308325?via%3Dihub>
- [40] PPE Object Detection Dataset (V9, Pictor-V3-Revised) by PPE. Roboflow. Available online: <https://universe.roboflow.com/ppeorxtt/ppe-u7jtr/dataset/9> (accessed on 1 June 2023).
- [41] Torres, P.; Davys, A.; Silva, T.; Schirmer, L.; Kuramoto, A.; Itagyba, B.; Salgado, C.; Comandulli, S.; Ventura, P.; Fialho, L.; et al. A Robust Real-time Component for Personal Protective Equipment Detection in an Industrial Setting. In Proceedings of the 23rd International Conference on Enterprise Information Systems (ICEIS 2021), Online Streaming, 26–28 April 2021; pp. 693–700. <https://www.scitepress.org/Link.aspx?doi=10.5220/0010452606930700>
- [42] Isailovic, V.; Peulic, A.; Djapan, M.; Savkovic, M.; Vukicevic, A.M. The compliance of head-mounted industrial PPE by using deep learning object detectors. *Sci. Rep.* 2022, 12, 16347. <https://www.nature.com/articles/s41598-022-20282-9>
- [43] Cengil, E.; Çınar, A.; Yıldırım, M. An Efficient and Fast Lightweight-Model with ShuffleNetv2 Based on YOLOv5 for Detection of Hardhat-Wearing. *Rev.*

- Comput. Eng. Stud. 2022, 9, 116–123.
<https://www.iieta.org/journals/rces/paper/10.18280/rces.090305>
- [44] Kisaiezehra; Farooq, M.U.; Bhutto, M.A.; Kazi, A.K. Real-Time Safety Helmet Detection Using Yolov5 at Construction Sites. *Intell. Autom. Soft Comput.* 2022, 36, 911–927. <https://www.techscience.com/iasc/v36n1/49992>
- [45] Xie, L. Hardhat-Hardhat-Wearing. 2019. Available online: <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/7C BGOS>.
- [46] Lo, J.-H.; Lin, L.-K.; Hung, C.-C. Real-Time Personal Protective Equipment Compliance Detection Based on Deep Learning Algorithm. *Sustainability* 2022, 15, 391. <https://www.mdpi.com/2071-1050/15/1/391>
- [47] Ferdous; Ahsan, S.M.M. PPE detector: A YOLO-based architecture to detect personal protective equipment (PPE) for construction sites. *PeerJ Comput. Sci.* 2022, 8, e999. <https://peerj.com/articles/cs-999/>
- [48] Ke, X.; Chen, W.; Guo, W. 100+ FPS detector of personal protective equipment for worker safety: A deep learning approach for green edge computing. *Peer-to-Peer Netw. Appl.* 2021, 15, 950–972. <https://link.springer.com/article/10.1007/s12083-021-01258-4>
- [49] Saudi, M.M.; Hakim, A.; Ahmad, A.; Shakir, A.; Hanafi, M.; Narzullaev, A.; Ifwat, M. Image Detection Model for Construction Worker Safety Conditions using Faster R-CNN. *Int. J. Adv. Comput. Sci. Appl.* 2020, 11, 246–250. <https://thesai.org/Publications/ViewPaper?Volume=11&Issue=6&Code=IJACS A&SerialNo=32>
- [50] Mneymneh, B.E.; Abbas, M.; Khoury, H. Automated Hardhat Detection for Construction Safety Applications. *Procedia Eng.* 2017, 196, 895–902. <https://www.sciencedirect.com/science/article/pii/S1877705817331430?via%3Dihub>
- [51] Ji, X.; Gong, F.; Yuan, X.; Wang, N. A high-performance framework for personal protective equipment detection on the offshore drilling platform. *Complex Intell. Syst.* 2023, 9, 1–16. <https://link.springer.com/article/10.1007/s40747-023-01028-0>
- [52] Karlsson, J.; Strand, F.; Bigun, J.; Alonso-Fernandez, F.; Hernandez-Diaz, K.; Nilsson, F. Visual Detection of Personal Protective Equipment and Safety Gear on Industry Workers. *arXiv* 2022. <https://arxiv.org/abs/2212.04794>

[53] Hardhat and Safety Vest Image for Object Detection. Available online: <https://www.kaggle.com/datasets/johnsyin97/hardhatand-safety-vest-image-for-object-detection>

[54] Xiong, R.; Tang, P. Pose guided anchoring for detecting proper use of personal protective equipment. *Autom. Constr.* 2021, 130, 103828. <https://www.sciencedirect.com/science/article/pii/S092658052100279X?via%3Dihub>

[55] Wang, M.; Wong, P.K.-Y.; Luo, H.; Kumar, S.; Delhi, V.-S.K.; Cheng, J.C.-P. Predicting safety hazards among construction workers and equipment using computer vision and deep learning techniques. In *Proceedings of the International Symposium on Automation and Robotics in Construction*, Banff, AB, Canada, 21–24 May 2019; pp. 399–406. http://www.iaarc.org/publications/2019_proceedings_of_the_36th_isarc/predicting_safety_hazards_among_construction_workers_and_equipment_using_computer_vision_and_deep_learning_techniques.html

[56] Wang, L.; Xie, L.; Yang, P.; Deng, Q.; Du, S.; Xu, L. Hardhat-Wearing Detection Based on a Lightweight Convolutional Neural Network with Multi-Scale Features and a Top-Down Module. *Sensors* 2020, 20, 1868. <https://www.mdpi.com/1424-8220/20/7/1868>

[57] Li, Y.; Wei, H.; Han, Z.; Huang, J.; Wang, W. Deep Learning-Based Safety Helmet Detection in Engineering Management Based on Convolutional Neural Networks. *Adv. Civ. Eng.* 2020, 2020, 9703560. <https://onlinelibrary.wiley.com/doi/10.1155/2020/9703560>

[58] Diana M. Cordova-Esparaza, Facultad de Informática, Universidad Autónoma de Querétaro, Mexico. A COMPREHENSIVE REVIEW OF YOLO: FROM YOLOV1 TO YOLOV8 AND BEYOND (2023).

[59]