

Article

Not peer-reviewed version

Enhancing Construction Safety: A deep-learning based Personal Protective Equipment Detection System

Daniel Oluwatise Owolabi and [Desmond Moru](#) *

Posted Date: 12 December 2024

doi: 10.20944/preprints202412.1040.v1

Keywords: Deep learning; Personal Protective Equipment; Object Localization; Workplace safety



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Enhancing Construction Safety: A Deep-Learning Based Personal Protective Equipment Detection System

Daniel Owolabi [†], Desmond Moru

Department of Computer Science, Pan-Atlantic University, Ibeju-Lekki, Lagos; Pan-Atlantic University

* Correspondence: dmoru@pau.edu.ng; Tel.: +234-806-731-0924

† These authors contributed equally to this work.

Abstract: Worker safety is notably improved through the application of personal protective equipment (PPE), which effectively reduces the severity of injuries or fatal incidents in environments like construction sites, chemical facilities and hazardous areas. PPE is extensively mandated to ensure an acceptable level of safety, addressing not just accidents at the mentioned sites but also the risks posed by chemical hazards. Due to various factors or oversights, workers may intermittently fail to adhere to safety regulations regarding wearing protective equipment. Traditional manual monitoring is both labor-intensive and prone to errors. Thus, there is a pressing need for the advancement of intelligent monitoring systems capable of providing automated and precise detection of such safety equipment. As a solution, we present a deep learning approach for the real-time detection of PPE components, including helmets, safety boots, vests and gloves. The proposed deep learning model exhibited a remarkable mean average precision of 97.1%, indicating the model's proficiency in object localization and recognition. These results not only underline the effectiveness of the deep learning -based PPE detection system but also emphasize its practicality in diverse industrial and occupational settings. By surpassing established benchmarks attained in literature, this research contributes significantly to enhancing safety standards and reducing the risk of workplace accidents.

Keywords: deep learning; personal protective equipment; object localization; workplace safety

1. Introduction

The construction industry serves as a vital driver of economic growth and development, fostering infrastructural expansion and employment opportunities. However, amidst its exponential growth, the sector grapples with inherent challenges that impact the safety and well-being of its workforce. According to a comprehensive study conducted by the National Institute for Occupational Safety and Health (NIOSH) in collaboration with local authorities, this industry experiences a concerning rise in accidents, with an average of 6,000 reported incidents annually over the past five years [14]. These alarming statistics highlight the urgent need for innovative safety measures to safeguard the lives of construction workers and mitigate potential risks on construction sites. Safety in workplaces is of paramount importance, especially in industries that involve high-risk activities and potential hazards [21]. The construction industry is one such sector where ensuring the well-being of workers is crucial due to the inherent risks associated with the nature of work [24]. In many developing countries, construction sites have been witnessing a rising number of accidents and injuries, highlighting the need for robust safety measures [11]. Personal Protective Equipment (PPE) serves as a primary line of defense against workplace hazards. PPE includes equipment like hard hats, safety vests, goggles, and gloves, which protect workers from potential injuries and health risks [18]. However, the effectiveness of PPE relies on its proper usage and compliance with safety regulations. Traditional methods of inspecting and enforcing PPE compliance often suffer from subjectivity and human errors, leading to lapses in safety standards [15]. In recent years, advancements in deep learning and computer vision technologies have opened up new possibilities for automating safety inspections and ensuring

real-time compliance monitoring [19]. This study aims to leverage the power of deep learning based systems to develop an intelligent Personal Protective Equipment Detection System tailored for the construction industry. By addressing the challenges of manual inspection methods and improving PPE compliance, this system seeks to significantly enhance workplace safety, reduce accidents, and foster a culture of proactive safety practices in the construction sector. As a case study, Craneburg Construction, a leading construction company with diverse projects, will be utilized to implement and evaluate the efficacy of the developed PPE detection system. The outcomes of this research have the potential to revolutionize safety management in the construction industry, benefiting not only Craneburg Construction but also serving as a model for broader industry-wide safety improvement. This article's remaining sections are organized as follows: In Section 2, existing studies on Personal Protective Equipment detection are examined. Section 3 presents the study's materials and methods. Section 4 presents the results and findings. Conclusions are provided in Section 5.

2. Materials and Methods

The central objective of this research revolves around the identification of personal protective equipment (PPE) usage within a given environment. This was accomplished through a structured process involving dataset acquisition, model training, algorithm development for detection, and a comprehensive assessment that encompassed precision, recall, average precision, and mean average precision metrics shown in Figure 1. Additionally, an assessment of the system's real-time performance, measured in frames per second, was conducted.

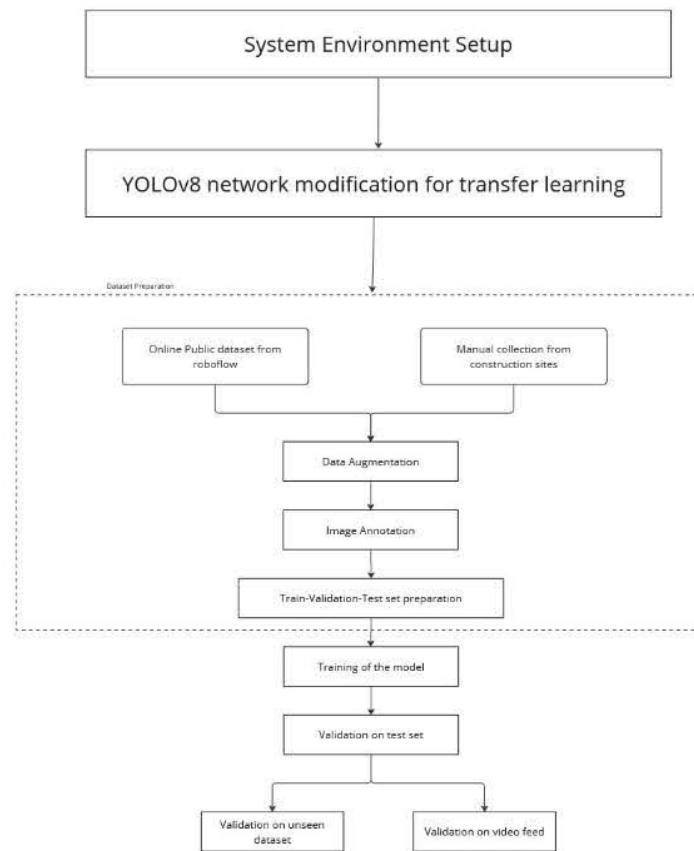


Figure 1. The study's research method

2.1. Environmental Setup

The hardware was a laptop running Windows 11 and powered by an Intel(R) Core(TM) i7-10750H processor clocked at 2.60 GHz together with 16GB of RAM and an NVIDIA GeForce GTX graphics card. These extra libraries CVZone, and OpenCV were combined with Python3. Due to its user friendly interface and expanded tools for programming in Python, PyCharm was used to train the dataset. The training process has been accelerated by using the system GPUs and the system's FPS can also be increased by a strong GPU. The model was trained using the Darknet framework (<https://github.com/AlexeyAB/darknet>). The decision was made based on yolo compatibility. Part of the dataset was gathered using a 16-megapixel smartphone.

2.2. Dataset Preparation

A model that has been trained using a dataset is used by a detection and recognition system to anticipate the input to the system. A data set of photos containing objects such as hard hats, safety vests, safety boots, and gloves was used to train the model. The objects have labels that indicate to which class they belong. A picture of a hardhat would be labeled "Hardhat" in this instance. Models can be trained to recognize specific types of things. A large data set can enhance the algorithm by increasing the amount of variation. More precise object detection has been made possible by photos taken at various angles and in variable lighting. Deep convolutional processing makes little things disappear, making it more challenging to detect objects, such as safety goggles. Hardhats come in different colors, but they all have around the same proportions, thus they probably do not need a big dataset. Roboflow is an online platform that offers a diverse range of machine learning datasets. In one specific dataset, which includes images featuring hardhats, gloves, safety boots and safety vests, there were 3,958 images (<https://universe.roboflow.com/faisal-lazuardy-fc9nq/deteksi-apd-ppfux/dataset/4>). Given the specific nature of these classes, it was challenging to locate pre-existing datasets that adequately covered them all. As a solution, 800 more images were collected manually using a mobile phone in various construction sites and were annotated using an open-source tool named LabelImg which can be found on GitHub (www.github.com/tzutalin/labelImg). Additionally, frames were extracted from Personal Protective Equipment (PPE) recordings to provide supplementary context regarding the equipment and the surrounding environment. These recordings and the captured images underwent augmentation, which involved adjusting parameters like saturation, brightness, and contrast. This process effectively resulted in the creation of additional images. Augmenting images offers the benefit of preserving the object coordinates from the original images. This enables a straightforward duplication of bounding box coordinates, resulting in a time-efficient method for expanding the dataset. Additionally, manual checks were performed on all images to ensure there were no inconsistencies. The distribution of the training dataset for the testing phase is detailed in the results section.

2.3. Model Training

In the process of object detection, a crucial step involves training a model because without training, the model would generate random predictions. To achieve this, the dataset is typically split into two parts: a training set (70%) and a testing set (30%). It's worth noting that an 80% training to 20% testing split is not ideal when dealing with a limited amount of data. For smaller datasets, a more suitable division typically falls within the range of 50% training to 50% testing and 70% training to 30% testing. This approach helps ensure a good balance between having a diverse set of data for training and a sufficient amount for testing, ultimately leading to better model performance. It's also important to emphasize that achieving highly accurate predictions hinges on having high-quality labeled data. The dataset was trained using Darknet, and to expedite the process, Google Colab was employed, making use of high-performance GPUs. When training a model, hyperparameters can be configured, and these values are determined through practical experimentation. Finding the perfect

settings for these hyperparameters beforehand is an exceedingly challenging task. The input image will be 640 pixels in both width and height. Increasing image resolution offers the benefit of better accuracy, but it comes at the cost of reduced frame rate and greater demands on hardware resources. The MaxBatches parameter, determining the maximum number of processed images before updating network weights, is calculated using equation 3.1. The learning rate is adjusted at specific intervals after processing several batches, with values set at 8,000 and 9,000, following the formula in equation 3.2. The filter value in convolutional layers ranges from 64 to 512, so the number of filters in each layer is variable. These calculations are based on a project titled "Automatic Road Traffic Signs Detection and Recognition using 'You Only Look Once' version 4 (YOLOv4)" (Srivastava et al.) in 2022, and these formula values are recommended.

$$\text{MaxBatches} = \text{number of classes} \times 2000 \quad (1)$$

$$\text{Steps} = (\text{MaxBatches} \times 0.8, \text{MaxBatches} \times 0.9) \quad (2)$$

$$\text{Filters} = (\text{number of classes} + 5) \times 3 \quad (3)$$

The number of classes used in the training process is represented by the variable called number of classes.

2.4. Conducting Tests and Experiments

The process involved two testing phases, which were conducted iteratively to improve the results, culminating in the final results. Within each testing phase, various conditions were assessed:

- a. Images taken in various settings
- b. Photographs taken at various distances
- c. Pictures featuring various individuals
- d. Pictures containing a variety of items
- e. Images in which the protective gear is present in the picture but not worn
- f. Images captured of individuals in various positions and perspectives

The list provided outlines the conditions assigned to each test phase, and each condition is clarified by the preceding list.

1. Phase 1 of testing covers items a, b, c, d, and f
2. Phase 2 of testing also encompasses items a, b, c, d, e, and f

The evaluations were conducted under various conditions, with each class being assessed for precision and recall in all tests. Nevertheless, the initial test only involved assessing precision and recall to gauge the model's performance. In contrast, the second test additionally measured the frames per second to enable a comparison with other models designed for PPE detection. The test phase employed models that underwent 50 training iterations, with each model taking approximately 2.9 hours for training. The duration of training varied according to factors such as the size of the dataset and the GPU resources accessible on Google Colab. Numerous examinations were conducted to assess the system's performance. Each test underwent analysis, and modifications were implemented to meet the necessary standards stated in Section 3.5. These tests encompassed diverse environments, varying distances, and different lighting conditions to enhance the system's robustness. In addition, a variety of individuals in various postures were examined to create a comprehensive solution. The evaluation of the detection model was designed to verify its ability to exclusively identify personal protective equipment worn by humans.

2.5. Requirement Standards

1. Attain a precision and recall rate of at least 95% for detecting hardhats and safety vests.
2. Achieve a minimum of 90% average precision for each individual object.
3. Attain a mean average precision (MAP) of at least 70% when considering all objects collectively.
4. Immediate identification with a requirement of at least 5 video frames per second.
5. Attain a minimum precision rate of 95% and a recall rate of 85% when it comes to identifying safety gloves and safety boots.

Performance testing of the system within an environment similar to the training data was anticipated to yield good results. However, factors such as lighting conditions and variable object distances were expected to have varying impacts on each object, with safety glasses being particularly challenging to detect. Detecting objects like hardhats and safety vests, characterized by distinct colors and shapes, was not expected to pose significant difficulties. However, the system was not initially designed to detect cases where personal protective equipment (PPE) was not worn. Refer to Figure 2 for a visualization of the system workflow.

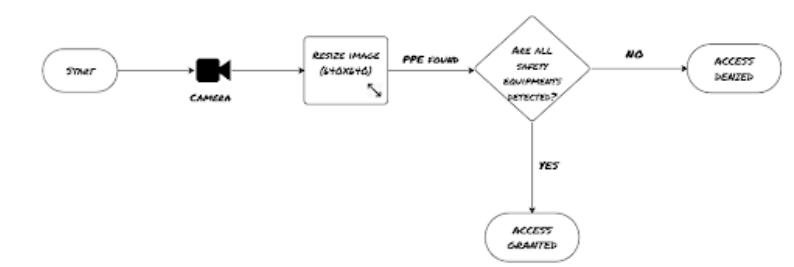


Figure 2. Schematic representation of the system pipeline

2.6. Model Pipeline

The system depicted in Figure 2 represents the outcome of multiple testing phases. The main objective was to establish a reliable mechanism for identifying personal protective equipment (PPE) and enhancing safety by minimizing accidents. The PPE detection model served as a prerequisite for gaining access to the construction site. It's important to note that this system evolved gradually during the course of the research and was not preconceived. Figure 3 illustrates the system's state after undergoing two testing phases. The system relies on a camera to capture an image of a person, presumably within a construction site or similar environment. This image serves as the input for the subsequent object detection process. The captured image is then resized to a fixed dimension of 640x640 pixels. This resizing is typically performed to ensure consistency and efficiency in the subsequent object detection process. Resizing the image to a predefined dimension can also be a requirement of the specific object detection model used. The resized image is fed into the object detection model. This model is responsible for identifying and locating personal protective equipment (PPE) within the image. The model has been trained to recognize specific PPE items, such as helmets, vests, goggles, or gloves. It analyzes the image and provides information on the presence and locations of these PPE items within the frame. Overall, the pipeline involves capturing an image, resizing it to a standardized dimension, and then using an object detection model to identify personal protective equipment within that image. This process is crucial for ensuring safety and compliance with PPE requirements in the construction site, as it allows for the automatic monitoring of whether workers are wearing the necessary safety gear.



Figure 3. Schematic representation of the system pipeline

3. Results

Several evaluations were carried out to assess the system's capability to identify objects. These assessments involved the use of a dataset comprising individuals wearing full personal protective equipment (PPE) to gauge the system's object detection accuracy across a range of factors like distance, angles, and lighting conditions. The performance of the system was analyzed, and a series of testing phases were conducted iteratively to enhance the results. The initial test phases employed precision and recall as metrics to evaluate the outcomes, with the second phase also considering the system's processing speed, measured in frames per second. A confusion matrix is a tool used to evaluate the precision and recall of a model. It involves four key variables: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). TP corresponds to objects correctly predicted by the model, TN represents accurately predicted negative objects, FP is when the model wrongly predicts a positive object, and FN signifies incorrect predictions of negative objects. These values are obtained from the results of the object detector. Precision gauges the proportion of positive predictions that were accurate, while Recall measures the proportion of actual positives that were correctly predicted. Precision and recall rates are computed using equations 1 and 2 as mentioned in reference.

$$P = \frac{TP}{TP + FP} \quad (4)$$

$$R = \frac{TP}{TP + FN} \quad (5)$$

The Average Precision (AP) is a metric that assesses how well a model performs in a dataset by considering both precision and recall rates. Precision evaluates the model's accuracy in making correct predictions, while average precision also considers the Intersection over Union. This means that average precision measures the model's capability to locate objects. AP calculates the average recall rates at different precision levels between 0 and 1, and its calculation is defined by equation 3.

$$AP = \sum_{k=1}^n P(k) \cdot \Delta R(k) \quad (6)$$

where:

1. $P(k)$ is the precision at the k -th threshold
2. $\Delta R(k)$ is the change in recall at the k -th threshold
3. n is the total number of thresholds

In the context of this calculation, 'N' represents the total number of images in the dataset. The variable 'P(k)' corresponds to the precision rate associated with the k image, and $\Delta r(k)$ stands for the

change in recall rate when moving from (K-1) to K image in the sequence. The average precision was computed separately for each class, and then these individual averages were combined to calculate the mean average precision (mAP) across all classes. This value represents the overall average precision for the entire set of classes.

$$mAP = \frac{1}{n} \sum_{i=1}^n AP_i \quad (7)$$

The variable 'n' refers to the total number of classes, and 'AP(k)' denotes the average precision for a specific class. The measurement of frames per second (FPS) involved recording the time difference during system operation within a specific interval. Equation 4.8 can be used to calculate and determine the average FPS.

$$FPS = \frac{n}{t_2 - t_1} \quad (8)$$

The variable 'n' represents the count of intervals, while the variables 't2 - t1' indicate the time gap between two measurement points.

3.1. Test Phases

The first and second testing phases have served as benchmarks to determine what is effective and what is not. Basic analytical calculations were employed during these phases. The test images featured various individuals wearing personal protective equipment (PPE) at various distances from the camera. Lighting conditions varied, resulting in a more comprehensive testing scenario. Anticipated limitations in the initial tests, specifically with regard to safety boots and gloves due to factors like size, shape, and transparency, resulted in expectedly subpar results. Subsequently, the training dataset was modified in response to the outcomes of the first and second tests. The testing dataset featured different individuals in a less distracting setting, with higher-resolution images captured from distances of 3 and 5 meters.

3.1.1. First Test Phase

The outcomes of the initial data set can be found in Table 1. This test involved individuals wearing complete personal protective equipment (PPE) and was conducted under varying conditions, including different camera angles, lighting settings, and distances. The table presents the frequency of each class within the training data set and provides precision, recall, and mean average precision metrics for the test set.

Table 1. Result of the first test phase. The training set instances represents the number of occurrences of each object in the dataset to train the model.

Class	Training set instance	Precision %
Gloves	387	92.4
Helmet	500	96.7
Non-Helmet	587	92.3
Person	1698	95
Safety boots	200	94.7
Safety vest	845	98.8
Class	Recall %	Average Precision %
Gloves	73.2	89.7
Helmet	97.2	98.7
Non-Helmet	93.5	93.7
Person	93.2	97.6
Safety boots	82.5	95.3
Safety vest	97.2	97.9

The findings indicate that the precision and recall metrics for safety gloves and boots fall short of the standard specified in Section 3.5. The training dataset exhibits skewness, indicating that the occurrences are unevenly distributed among the objects within the dataset. The dataset had a noticeable impact on the outcomes and necessitated modifications to enhance the performance of the detection model.

3.1.2. Second Test Phase

The dataset underwent modifications subsequent to the results presented in Table 1. Specifically, images predominantly featuring safety boots and gloves were subjected to augmentation techniques, including adjustments to sharpness, blurriness, and contrast, in order to expand the dataset and increase the diversity of the data. The specific augmentation settings applied to each object were as follows:

1. Safety boots (170 images): +40% brightness, +20% contrast, sharpened
2. Gloves (200 images): +35% brightness, -40% contrast.
3. Helmet (117 images): +10% brightness, +15% contrast, sharpened.

The test employed during the first testing phase was also utilized in the computation of the results presented in Table 2. Additional data was incorporated for all categories, resulting in slight enhancements in precision and recall metrics. However, despite these improvements, the precision and recall of safety boots and gloves failed to meet the necessary standards. The extreme variation in lighting made detection difficult. Variations in lighting conditions can create reflections on surfaces, leading to interference with object detection.

Table 2. Result of the second test phase.

Class	Training set instance	Precision %
Gloves	1112	92.4
Helmet	1202	97.6
Non-Helmet	1023	94.3
Person	1698	95.8
Safety boots	1458	95.2
Safety vest	1232	98.7
Class	Recall %	Average Precision %
Gloves	75.2	90.2
Helmet	98.2	95.4
Non-Helmet	97.5	93.7
Person	91.2	97.6
Safety boots	81.7	96.5
Safety vest	96.2	98.8

3.1.3. Final Test Phase

The lighting conditions were optimized to ensure the accurate detection of safety gloves and boots. Additionally, an environment was established to replicate a real-world scenario in the final test case. Different augmentation setting was applied to the same number of images for each object.

1. Safety boots (170 images): +40% brightness, +20% contrast, sharpened
2. Gloves (200 images): +20% brightness, gaussian blur ($\sigma=2$)
3. Helmet (117 images): +5% brightness, 10% contrast

The process for obtaining these results is depicted in Figure 3, encompassing both the image resizing and object detection phases. The training dataset employed in generating these results was obtained from the final test phase. Furthermore, additional augmented data was incorporated in accordance with the guidelines outlined in the final test phase. The outcomes were obtained

by assessing the precision, recall, and average precision of individual objects in an airlock room, with testing conducted under various conditions involving different angles and individuals, and observations were recorded. Every image included the PPE items mentioned such as gloves, helmet, safety vest and safety boots and the assessment process followed the sequence depicted in Figure 2. Precision, recall, and mean average precision were computed using the formulas in Equations (4), (5), and (7).

Each class had a precision score above the requirement standards in Section 3.7 While most classes had recall rates nearing 100%, there was a notable improvement in the recall rates for safety boots and gloves, primarily attributed to the implementation of data augmentation techniques. Table 3 displays results that surpass the current state-of-the-art, and it elaborates on how the system is employed. The papers referenced in Section 2 treat PPE detection as a form of surveillance. In surveillance systems, object detection can typically span a range of 10 to 100 meters. Detecting small items like safety glasses at these extended distances is a challenging task (even identifying a hardhat at 100 meters can be challenging). This research, however, conducted experiments at limited distances of 3 and 5 meters, which considerably facilitates the accurate detection of objects. The average precision is presented in Figure 4. It has been calculated according to Equation (4.7) and compares the average precision between various personal protective equipment.

Table 3. Result of the Final test phase.

Class	Training set instance	Precision %
Gloves	1112	98.3
Helmet	1202	97.7
Non-Helmet	1023	93.6
Person	1698	96
Safety boots	1458	96.5
Safety vest	1232	99.1
Class	Recall %	Average Precison %
Gloves	92.6	96.7
Helmet	98.5	99.2
Non-Helmet	94.7	98.2
Person	97.7	99.3
Safety boots	98.3	99.2
Safety vest	98.7	99.5

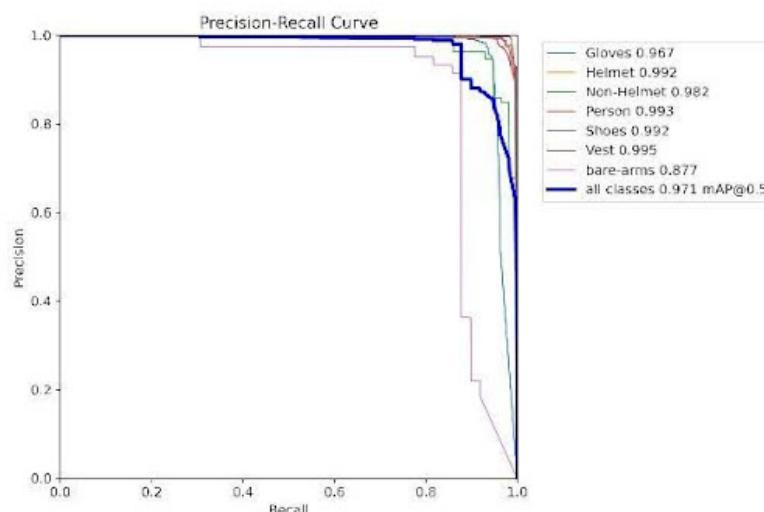


Figure 4. precision-recall curve comparing the average precision scores between various PPE components

In the previous test, the hardhat and safety vest consistently achieved good precision and recall rates at various distances. For safety gloves, the recall rates improved from 75.2% to 92.6%. In comparison, Zhafran et al achieved 100% accuracy in detecting hardhats, 73% accuracy in identifying safety vests, and 68% accuracy in recognizing gloves when observing them from a close distance. The training dataset for the final model, which generated the outcomes presented in Table 3, had an equivalent quantity of instances as seen in Table 2 for different personal protective gear. The sole contrast lies in the adjustments made to their augmentation settings. All objects met their specified precision and recall requirements, and the mean Average Precision (mAP) also met its specified requirements. (Delhi et al., 2020) created a system that bears resemblance to the one described in this thesis. Table 4 illustrates the interconnections among the results, enabling a comparison of different types of personal protective equipment.

Table 4. Result of the Final test phase.

Class	Precision %
Gloves	98.3
Gloves (Delhi et al., 2020)	68
Helmet	97.7
Helmet (Delhi et al., 2020)	100
Safety boots	96.5
Safety boots (Delhi et al., 2020)	Nil
Safety vest	99.1
Safety vest (Delhi et al., 2020)	73

4. Discussion and Conclusions

The ultimate outcome shows that the system performs effectively, achieving a mean average precision of 97.1% across different distances. It excels when used at close distances and in settings with a neutral background. However, its performance diminishes slightly when used at greater distances in more distracting environments due to the reduced size and increased difficulty in detecting objects. This conclusion is substantiated by the tests conducted in the first and second test phase. Items like hardhats and safety vests are consistently identifiable in various settings and at varying distances, whereas safety boots and safety gloves pose a detection challenge due to their smaller size and intricate contours. Distinguishing objects that share a similar size and appearance becomes more difficult when their midpoints coincide within the same grid. Ideally, improving detection results could be achieved by training the model using the user's specific PPE. The desired image processing speed was not attainable when using a CPU, but it can be reached if the task is performed on a GPU. To put it succinctly, this system functions effectively within its intended environment. The outcomes demonstrate seamless object detection within an enclosed room setting.

Author Contributions: All authors contributed equally to all aspects of the work, including conceptualization, methodology, software, validation, formal analysis, investigation, resources, data curation, writing the original draft, review and editing, visualization, supervision, project administration, and funding acquisition. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data used in this study come from the following sources: 1. Roboflow Available on the official website <https://roboflow.com> (accessed on 23 August 2023). 2. PPE 2023: The data are confidential and cannot be shared.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

PPE Personal Protective Equipment
NIOSH National Institute for Occupational Safety and Health

References

1. Roboflow. Available online: <https://roboflow.com> (accessed on 23 August 2023).
2. Balakrishnan, B.; Richards, G.; Nanda, G.; Mao, H.; Athinarayanan, R.; Zaccaria, J. PPE compliance detection using artificial intelligence in learning factories. *Procedia Manufacturing* **2020**, *45*, 277–282. DOI: 10.1016/j.promfg.2020.04.017.
3. Buckland, M. K.; Gey, F. C. The relationship between recall and precision. *Journal of the Association for Information Science and Technology* **1994**, *45*, 1, 12. DOI: 10.1002/(SICI)1097-4571(199401)45:1<12::AID-ASI2>3.0.CO;2-L.
4. Daghan, A.T.A.A.; Vineeta; Kesh, S.; Manek, A.S. A deep learning model for detecting PPE to minimize risk at construction sites. In Proceedings of the 2021 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT), Bangalore, India, 9–11 July 2021; pp. 1–6. Available online: <https://api.semanticscholar.org/CorpusID:244955755>.
5. 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. *Frontiers* **2020**.
6. Dünsch, I.; Gediga, G. Confusion matrices and rough set data analysis. *Journal of Physics: Conference Series* **2019**, *1229*, 012055. DOI: 10.1088/1742-6596/1229/1/012055.
7. Encord: Mean average precision for object detection. Available online: <https://encord.com/blog/mean-average-precision-object-detection/> (accessed on 28 November 2024).
8. Fang, Q.; Li, H.; Luo, X.; Ding, L.; Luo, H.; Rose, T. M.; An, W. Detecting non-hardhat-use by a deep learning method from far-field surveillance videos. *Automation in Construction* **2018**, *85*, 1–9. DOI: <https://doi.org/10.1016/j.autcon.2017.09.018>. URL: <https://www.sciencedirect.com/science/article/pii/S0926580517304429>.
9. Fernando, W.H.D.; Sotheeswaran, S. Automatic road traffic signs detection and recognition using ‘you only look once’ version 4 (yolov4). In: Proceedings of the 2021 International Research Conference on Smart Computing and Systems Engineering (SCSE), Location of Conference, Country, 2021; vol. 4, pp. 38–43. DOI: 10.1109/SCSE53661.2021.9568285.
10. Gong, C.; Wang, D.; Li, M.; Chandra, V.; Liu, Q. Keep-augment: A simple information-preserving data augmentation approach. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Online, 19–25 June 2021; pp. 1055–1064. DOI: 10.1109/CVPR46437.2021.00111.
11. Isah, H. Health and safety issues in the Nigerian construction industries: Prevention and mitigation. *EPMS* **2019**, *12*, 74. DOI: 10.15405
12. Joseph, V. R. Optimal ratio for data splitting. *Statistical Analysis and Data Mining: The ASA Data Science Journal* **15** (2022). DOI 10.1002/sam.11583.
13. Jung, M.; Lim, S.; Chi, S. Impact of work environment and occupational stress on safety behavior of individual construction workers. *Int. J. Environ. Res. Public Health* **2020**, *17*, 28304. DOI: 10.3390/ijerph17228304.
14. Kiersma, M. National institute for occupational safety and health. *Encyclopedia of Toxicology* **2014**, 10.1016/B978-0-12-386454-3.00340-7.
15. Kilgour, D. The use of costless inspection in enforcement. *Theory and Decision* **1994**, *36*, 207–232. DOI: 10.1007/BF01079928.
16. Liginlal, D.; Sim, I.; Khansa, L.; Fearn, P. Hipaa privacy rule compliance: An interpretive study using Norman’s action theory. *Comput. Secur.* **2012**, *31*, 206–220. DOI 10.1016/j.cose.2011.12.002.
17. Lo, J.H.; Lin, L.K.; Hung, C.C. Real-time personal protective equipment compliance detection based on deep learning algorithm. *Sustainability* **2020**, *12*, 3584.
18. Mckinnon, R. Personal protective equipment program. pp. 319–321 (2019), DOI: 10.1201/9780429280740-63.
19. Nain, M.; Sharma, S.; Chaurasia, S. Safety and compliance management system using computer vision and deep learning. *IOP Conference Series: Materials Science and Engineering* **2021**, *1099*, 012013. DOI 10.1088/1757-899X/1099/1/012013.

20. Nath, N.; Behzadan, A.; Paal, S. Deep learning for site safety: Real-time detection of personal protective equipment. *Automation in Construction* **2020**, *112*, 103085. DOI: 10.1016/j.autcon.2020.103085.
21. Olutuase, S. A study of safety management in the Nigerian construction industry. *IOSR Journal of Business and Management* **2014**, *16*, 01–10. DOI 10.9790/487X-16350110.
22. Padilla, R.; Netto, S.L.; da Silva, E.A.B. A survey on performance metrics for object-detection algorithms. *Proc. 2020 Int. Conf. on Systems, Signals and Image Processing (IWSSIP)* **2020**, pp. 237–242, DOI: 10.1109/IWSSIP48289.2020.9145130.
23. Scherzer, D.; Yang, L.; Mattausch, O.; Nehab, D.F.; Sander, P.; Wimmer, M.; Eisemann, E. Temporal coherence methods in real-time rendering. *Computer Graphics Forum* **2012**, *31*. DOI 10.1111/j.1467-8659.2012.03075.x
24. Suárez, F.; Carvajal, G.; Alís, J. Occupational safety and health in construction: A review of applications and trends. *Industrial Health* **2017**, *55*, 10.2486/indhealth.2016-0108.
25. Mean average precision. Available online: <https://www.v7labs.com/blog/mean-average-precision> (accessed on 23 May 2024).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.