

Detection of Personal Protective Equipment (PPE) using an Anchor Free-Convolutional Neural Network

Honggang WANG¹

School of Electronic Information Engineering, Xi'an Siyuan University, Xi'an 710038, Shaanxi, China

Abstract—In industrial environments, the utilization of Personal Protective Equipment (PPE) is paramount for safeguarding workers from potential hazards. While various PPE detection methods have been explored in the literature, deep learning approaches have consistently demonstrated superior accuracy in comparison to other methodologies. However, addressing the pressing research challenge in deep learning-based PPE detection, which pertains to achieving high accuracy rates, non-destructive monitoring, and real-time capabilities, remains a critical need. To address this challenge, this study proposes a deep learning model based on the YOLOv8 architecture. This model is specifically designed to meet the rigorous demands of PPE detection, ensuring accurate results. The methodology involves the creation of a custom dataset and encompasses rigorous training, validation, and testing processes. Experimental results and performance evaluations validate the proposed method, illustrating its ability to achieve highly accurate results consistently. This research contributes to the field by offering an effective and robust solution for PPE detection in industrial environments, emphasizing the paramount importance of accuracy, non-destructiveness, and real-time capabilities in ensuring workplace safety.

Keywords—PPE detection; deep learning; YOLOv8; industrial environments; real-time detection

I. INTRODUCTION

Personal Protective Equipment (PPE) plays a pivotal role in ensuring the safety of workers in industrial environments [1], [2]. The first line of defense against potential hazards, PPE includes various gear such as aprons, gloves, helmets, masks, and shoes designed to protect workers from physical, chemical, and biological risks [3]. Proper detection and monitoring of PPE utilization in industrial settings are of paramount importance to guarantee workplace safety [4]. This paper delves into the realm of PPE detection, highlighting its significance and examining the latest advancements in technology for this purpose [5].

The accurate detection of PPE, including aprons, gloves, helmets, masks, and shoes, in industrial environments is crucial for safeguarding workers from potential hazards. PPE detection ensures that employees are equipped with the necessary gear, minimizing the risk of injuries and health issues. The ability to monitor PPE utilization is instrumental in maintaining a safe work environment, not only for individuals but also for the overall efficiency of industrial operations [6]. To address the need for effective PPE detection, a range of technologies has been developed over the years. In this paper, we review existing methods and explore the latest advances in PPE detection techniques. These advancements in technology are

essential for real-time monitoring of PPE usage and ensuring the utmost safety in industrial workplaces [8].

Deep learning-based approaches have gained significant attention in recent years for PPE detection [7]. These methods have become a focal point of research due to their exceptional ability to handle complex visual data and patterns. In comparison to traditional methods, deep learning offers superior accuracy and robustness in detecting PPE items. This paper provides insights into why deep learning-based approaches have garnered immense interest among researchers and have become a promising avenue for addressing PPE detection challenges [9].

Despite the potential of deep learning-based approaches, there exist certain limitations and research challenges. These include the demand for high accuracy and real-time requirements. In this paper, we emphasize the need for further exploration and research to overcome these challenges, underscoring the importance of striving for methods that can fulfill the rigorous demands of PPE detection in industrial settings.

In response to these challenges, this study proposes a deep learning method based on Convolutional Neural Networks (CNN) to address the complexities of PPE detection [10]–[12]. We provide a justification for adopting this deep learning approach and demonstrate how it can effectively resolve the research challenges and provide high-accuracy real-time PPE detection. The study involves the creation of a custom dataset and encompasses training, validation, and testing processes to ensure the robustness of the proposed method.

This research contributes to the field of PPE detection in several ways. First, it generates a custom dataset specifically designed for PPE detection challenges, enriching the available resources for future research in this domain. Second, it introduces an efficient deep-learning method tailored to the unique requirements of PPE detection. Lastly, extensive experiments and performance evaluations are conducted to validate the effectiveness of the proposed method, offering a comprehensive assessment of its capabilities and practical applications.

II. PREVIOUS STUDY

The realm of agriculture has seen remarkable progress, primarily propelled by the substantial contributions of machine learning and deep learning techniques. These state-of-the-art technologies have been instrumental in reshaping the prediction, classification, and identification of Personal Protective Equipment (PPE). Their integration offers a

multitude of benefits, including non-invasiveness, cost-effectiveness, speed, and reliable PPE detection. This transformative potential has sparked numerous research endeavors dedicated to enhancing PPE diagnosis and detection.

This paper in [13] presented a Convolutional Neural Network (CNN) method for identifying Personal Protective Equipment (PPE) usage compliance in manufacturing laboratory settings. The CNN model is trained on a custom dataset and demonstrates remarkable accuracy in recognizing various PPE items, including aprons, gloves, helmets, masks, and shoes. However, a limitation of the approach is its sensitivity to variations in lighting conditions, which can affect detection accuracy. Future work could involve the integration of advanced lighting normalization techniques to address this limitation and further enhance the model's robustness for real-time PPE compliance monitoring in dynamic industrial environments.

The authors in study [1] introduced a deep learning-based framework for monitoring the wearing of Personal Protective Equipment (PPE) on construction sites. The method employs convolutional neural networks (CNNs) to detect and identify PPE items, ensuring compliance. However, a limitation is the challenge of real-time monitoring due to computational demands, which may impact its practical applicability. Future research could focus on optimizing the model for real-time performance to enhance its effectiveness in the dynamic and time-sensitive construction site environment.

This paper in [2] presented a Substation Safety Awareness intelligence model employing a Graph Neural Network (GNN) approach for rapid detection of Personal Protective Equipment (PPE). The GNN model effectively identifies PPE items, ensuring safety in substation environments. However, one limitation is the need for high-quality data, which may not always be readily available for model training, potentially limiting its practical implementation. Future research should explore techniques for mitigating data scarcity and enhancing model generalization to diverse substation settings.

The authors in [14] introduced a video-based smart safety monitoring system to prevent industrial work accidents. The method utilizes computer vision and machine learning algorithms to analyze video footage, detecting potential safety hazards. A limitation is that it may require substantial computational resources for real-time monitoring across extensive industrial settings. Future work should focus on optimizing computational efficiency to enhance its practicality and scalability.

This paper in [15] presented an enhanced detection network model based on YOLOv5 for safety warnings in construction sites. The method employs YOLOv5, a state-of-the-art object detection architecture, to identify potential safety hazards. However, a limitation lies in the sensitivity of the model to variations in lighting and environmental conditions, which can affect detection accuracy. Future research could focus on refining the model's robustness to lighting and environmental changes, improving its performance as a safety warning tool in dynamic construction site environments.

The five papers discussed focus on improving safety by monitoring the compliance of the PPE in various industrial contexts. They primarily employ deep learning and computer vision technologies for this purpose. While they all share the goal of enhancing workplace safety, each paper addresses a different industrial setting and applies distinct methodologies. The [13] emphasizes PPE compliance in manufacturing labs, using CNNs for detection. Its limitation is sensitivity to lighting variations. The study in [1] targets construction sites but doesn't explicitly mention limitations, hinting at potential real-time computational challenges. The study in [2] uses GNNs for rapid PPE detection in substation environments, highlighting data availability as a limitation. The study in [14] focuses on video-based safety across various industrial settings, with computational resource requirements as the main limitation. The research in [15] refines a YOLOv5-based model for construction site safety, noting sensitivity to lighting and environmental conditions. In summary, these papers contribute to safety monitoring by adapting their approaches to different industrial contexts, each with its own set of challenges and limitations. Despite common concerns, such as sensitivity to lighting and real-time computational demands, each paper offers unique insights to enhance workplace safety.

III. METHODOLOGY

A. Data Collection

We constructed our dataset by gathering PPE images from both publicly available internet resources and the Roboflow platform, aiming to create a comprehensive and diverse collection of images for training and testing. The original dataset is a collection of 3897 images of workers wearing safety vests in various industrial settings¹. The images are annotated with bounding boxes that indicate the location and category of the safety vests. The dataset is intended for training computer vision models that can detect whether workers are wearing the PPE or not. The dataset is part of the Roboflow Universe Projects, which are open-source datasets for various computer vision tasks.

B. Data Augmentation

To ensure the dataset's richness and diversity, we employed data augmentation techniques. Data augmentation is a critical step in deep learning, especially in the context of the PPE detection, as it allows us to generate a more extensive and varied set of images, enhancing the model's ability to handle real-world scenarios. To augment our PPE dataset, we utilized common techniques such as rotation, translation, and scaling. These techniques simulate variations in object positions and orientations, which are essential for addressing real-world challenges like workers wearing PPE at different angles or in varying positions. We also applied techniques like horizontal and vertical flipping, which helps the model generalize better to PPE instances appearing both on the left and right sides of the frame.

Additionally, we used color adjustments, including brightness and contrast modifications, to account for varying

¹<https://universe.roboflow.com/roboflow-universe-projects/safety-vests>

lighting conditions in industrial settings. Finally, noise injection and blurring were employed to replicate scenarios where the PPE items might be partially obscured or subject to environmental interference. These data augmentation techniques collectively enhance the dataset's diversity, making it more robust and suitable for training models capable of handling a wide range of real-world PPE detection scenarios. After the data augmentation process the size of dataset is tripled and the new size of the dataset in this study is 11691.

C. PPE Model Object Detection

There are several compelling reasons to consider utilizing YOLOv8 in our computer vision project:

1) *Enhanced accuracy:* YOLOv8 boasts improved accuracy compared to its predecessors, making it an attractive choice for various computer vision tasks.

2) *Feature-rich implementation:* The latest YOLOv8 implementation introduces a wealth of new features, with a particularly user-friendly Command Line Interface (CLI) and a dedicated GitHub repository. These additions streamline development and project management.

3) *Versatility:* YOLOv8 supports multiple computer vision tasks, including object detection, instance segmentation, and image classification, providing a comprehensive solution for various applications.

4) *Efficient training:* YOLOv8 offers faster training times in comparison to some other two-stage object detection

models, making it a more time-efficient option for our computer vision projects.

The layout created by [16] offers an excellent visualization of the architecture, providing a clear and structured representation of the system. This visual aid can significantly enhance understanding and communication of complex concepts or technical architectures within the context of software development or any other field. In this study, a YOLOv8 based method is proposed for PPE object detection. Fig. 1 shows the proposed system architecture.

Anchor-free detection refers to a method where an object detection model directly predicts the center of an object, eliminating the need to calculate the offset from a predefined anchor box.

In contrast, anchor boxes are predetermined bounding boxes with specific height and width characteristics. These boxes are strategically chosen based on the size and aspect ratio of objects present in the training dataset. During the detection process, these anchor boxes are systematically arranged and tiled across the image.

The network's output includes probabilities and attributes for each of these tiled boxes, encompassing information like background, Intersection over Union (IoU), and offset values. These attributes are instrumental in adjusting the anchor boxes. Multiple anchor boxes can be created to accommodate objects of various sizes, functioning as fixed reference points for predicting bounding boxes. An illustration depicting bounding box predictions based on anchor boxes is depicted in Fig. 2.

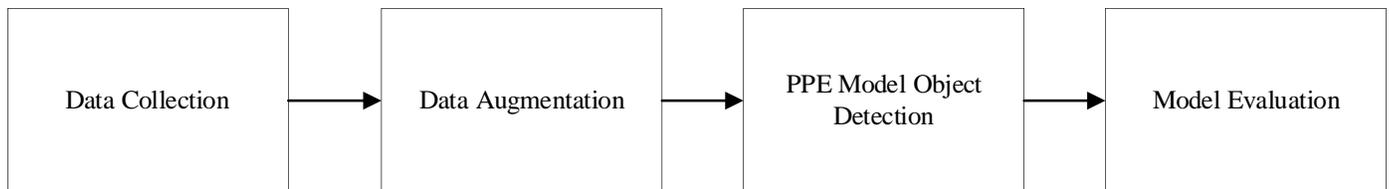


Fig. 1. The proposed system architecture.

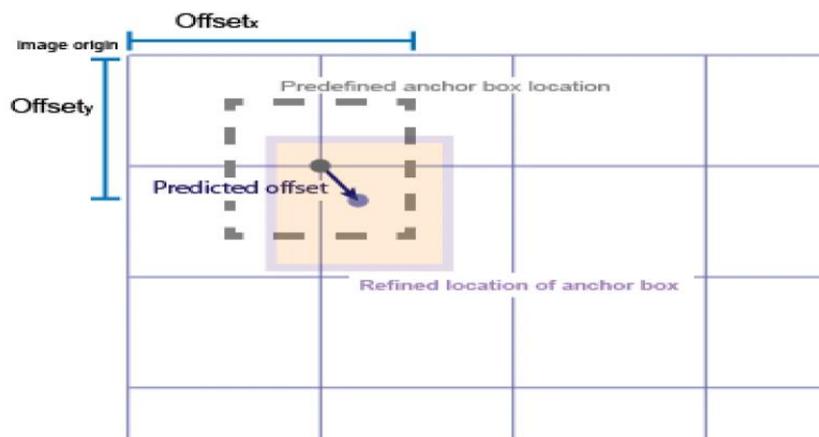


Fig. 2. An illustration depicting bounding box predictions based on anchor boxes.

Anchor-free detection offers notable advantages due to its flexibility and efficiency. Unlike methods that rely on predefined anchor boxes, anchor-free detection eliminates the need for manually specifying these anchors. In previous YOLO models like v1 and v2, selecting anchor boxes was a challenge and could result in suboptimal outcomes. Anchor-free detection simplifies this aspect, allowing for more adaptability in object detection tasks.

D. Model Evaluation

To evaluate the performance of a YOLOv8 model for PPE detection using precision, recall, and mean Average Precision (mAP) metrics, we'll assess the model's ability to detect and localize PPE items in images correctly. Here's how we calculate these metrics and their formulas:

Precision measures the accuracy of positive predictions made by the model.

$$\text{Precision} = \text{True Positives (TP)} / (\text{True Positives} + \text{False Positives})$$

True Positives (TP) are the number of correct PPE detections.

False Positives (FP) are the number of instances where the model incorrectly predicts PPE when there is none.

Recall, also known as sensitivity or true positive rate, measures the model's ability to identify all relevant PPE instances.

$$\text{Recall} = \text{True Positives (TP)} / (\text{True Positives} + \text{False Negatives})$$

True Negatives (TN) are the number of instances where the model correctly predicts the absence of PPE.

False Negatives (FN) are the number of actual PPE instances that the model misses.

(mAP) is a comprehensive metric used in object detection tasks to assess the model's performance. It calculates the average precision at various confidence thresholds and is often visualized as a precision-recall curve. The mAP formula involves the following steps:

- Calculate precision and recall for different confidence thresholds.
- Calculate the area under the precision-recall curve.
- Average the areas under the curve for different classes, resulting in the mAP.

mAP provides a holistic view of the model's performance across various levels of confidence in detecting PPE items. A higher mAP indicates a more reliable and accurate model.

These metrics collectively provides a quantitative assessment of the YOLOv8 model's ability to detect PPE items in terms of accuracy, completeness, and overall performance. Evaluating precision, recall, and mAP helps we understand the model's strengths and weaknesses, enabling us to fine-tune it for improved PPE detection accuracy.

In this study, a YOLOv8 model was generated for the purpose of PPE detection on a custom dataset. The dataset was initially divided into three sets: 70% for training, 20% for validation, and 10% for testing. This division ensures a robust evaluation of the model's performance while preventing overfitting and enabling efficient training. For training the YOLOv8 model, the dataset was utilized to teach the model to recognize PPE items. The training process involved passing the dataset through the model multiple times iteratively adjusting the model's parameters to improve accuracy. In the context of improving PPE detection accuracy, it is advisable to consider a few key details. Firstly, an optimal learning rate should be selected, typically through experimentation, to ensure the model converges effectively. It's recommended to start with a lower learning rate and gradually increase it as necessary. Batch size is another crucial factor; larger batches can accelerate training but may require more memory. Striking the right balance is essential. Augmentation techniques such as rotation, translation, and color adjustment can further improve accuracy by diversifying the training data.

Furthermore, the dataset should be balanced, meaning that it should contain an equal distribution of PPE and non-PPE examples. This avoids bias and helps the model achieve better accuracy in detection. The validation module is pivotal in monitoring the model's performance. It assesses the model's generalization capability and helps identify any potential overfitting. In the context of PPE detection accuracy improvement, the validation set should be representative of real-world scenarios. Hyperparameter tuning is typically performed here, experimenting with various learning rates, batch sizes, and data augmentation techniques. The aim is to find the configuration that optimizes PPE detection accuracy without compromising the model's ability to generalize to unseen data. The testing module evaluates the YOLOv8 model's performance on an independent dataset, ensuring that it can accurately detect PPE items in real-world situations. To enhance PPE detection accuracy, the testing set should be diverse and representative of the environments in which the model will be deployed. This module is instrumental in quantifying the model's accuracy and assessing its readiness for real-world applications. The model's performance can be measured using metrics like precision, recall, and F1-score, which should be optimized to achieve the desired PPE detection accuracy.

In conclusion, to generate a YOLOv8 model for PPE detection, it is crucial to carefully consider training details such as learning rate, batch size, and data augmentation techniques. Balancing the dataset and ensuring diversity in the validation and testing sets are essential for accurate model evaluation. These suggestions should be tailored to the specific requirements of improving the accuracy of PPE detection, ensuring the model performs effectively in real-world scenarios

The Precision-Confidence Curve is a critical evaluation tool for assessing the performance and efficiency of a PPE detection model, such as YOLOv8s. It provides valuable insights into how the model's confidence in its predictions relates to the precision it achieves for various object classes. This curve helps in understanding the model's effectiveness and its ability to make accurate predictions while maintaining

high precision. In this case, where we are detecting PPE items, including classes like aprons, gloves, helmets, masks, and shoes, achieving a high precision rate of approximately 0.95 for all classes is an exceptional accomplishment. It means that when the model makes a prediction for any of these classes, there is a very high likelihood that the prediction is correct, with very few false positives. This is particularly important in PPE detection, where ensuring the safety and compliance of individuals in industrial or medical settings is of utmost importance. The precision confidence curve of the model is depicted in Fig. 3.

The high precision values, nearly 0.95, across all classes in the Precision-Confidence Curve indicate the model's effectiveness in recognizing PPE items. It demonstrates the model's ability to confidently identify and localize these items, contributing significantly to safety and compliance. With such high precision, the model can be relied upon to provide accurate PPE detection, which is crucial for preventing accidents, maintaining safety standards, and ensuring that individuals are adequately protected. These values instill

confidence in the model's performance and highlight its efficiency in recognizing PPE classes, underlining its practical utility in real-world applications.

The Recall-Confidence Curve is a crucial tool for evaluating the efficiency of a YOLOv8s model in PPE detection. Fig. 4 illustrates how the model's confidence in its predictions correlates with the recall it achieves for various PPE classes. The model's exceptional performance is evident, with an approximate 0.84 recall rate at the maximum confidence level across all classes. This means that the model effectively captures the majority of actual PPE instances, reducing the risk of missing important items and enhancing safety compliance in industrial and medical settings. The YOLOv8s model, as demonstrated by the Recall-Confidence Curve, proves to be highly efficient in recognizing PPE classes. These high recall values emphasize its effectiveness in minimizing the risk of overlooking critical PPE items and contribute significantly to maintaining safety standards and preventing accidents in real-world applications.

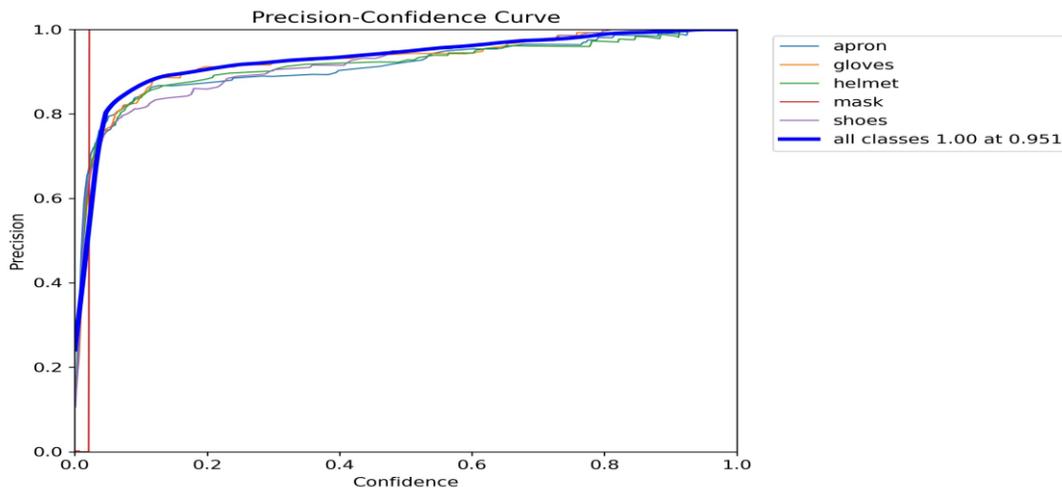


Fig. 3. Precision confidence curve of the model.

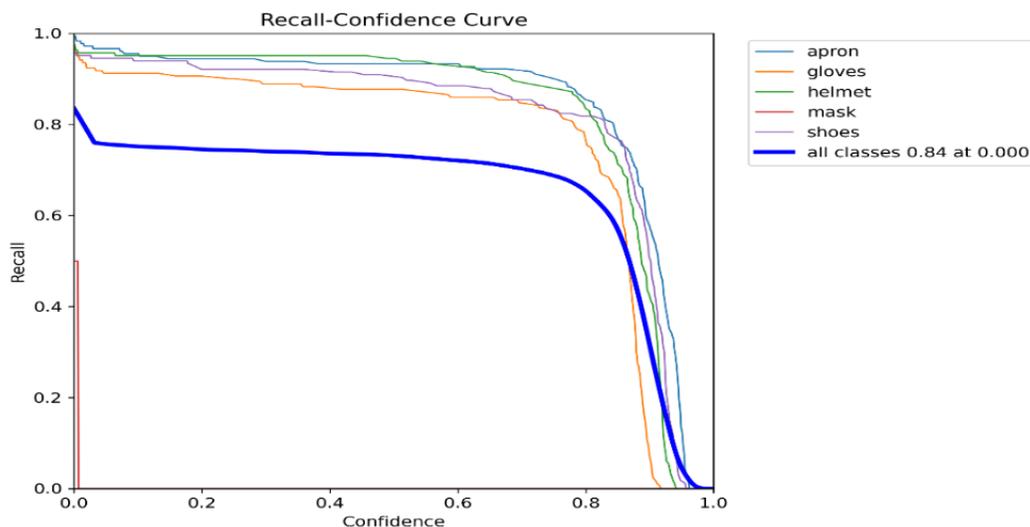


Fig. 4. Recall the confidence curve of the model.

The Precision-Recall curve is a crucial tool for evaluating the efficiency of a YOLOv8s model in PPE detection. Fig. 5 illustrates the trade-off between precision and recall, offering insights into the model's performance. The model's performance is notable, with an approximate 0.75 recall rate at the maximum confidence level across all PPE classes. This high recall rate indicates the model's effectiveness in recognizing these crucial PPE items, reducing the risk of overlooking important safety gear and enhancing safety compliance in industrial and medical settings. The YOLOv8s model, as shown by the Precision-Recall curve, is efficient and reliable in PPE detection, making it a valuable asset for maintaining safety standards and preventing accidents in real-world applications. These high recall values emphasize the model's effectiveness in minimizing the risk of missing important PPE items and contribute significantly to overall safety measures.

The F1 Confidence Curve is a vital tool for evaluating the efficiency of a YOLOv8s model in PPE detection. Fig. 6 provides insights into the model's confidence in its predictions and the resulting F1-score, which balances precision and recall. The model's performance is impressive, with an approximate 0.74 F1 score at the maximum confidence level across all PPE classes. This high F1-score indicates the model's effectiveness in accurately recognizing these essential PPE items, striking a balance between precision and recall. This balance is crucial for minimizing false alarms and missed detections, contributing significantly to safety compliance in various industrial and medical settings.

In conclusion, the YOLOv8s model, as demonstrated by the F1 Confidence Curve, is a robust and reliable solution for PPE detection. It excels in minimizing false alarms and ensuring that important PPE items are accurately identified, enhancing overall safety measures in real-world applications. The results of the model are depicted in Fig. 7.

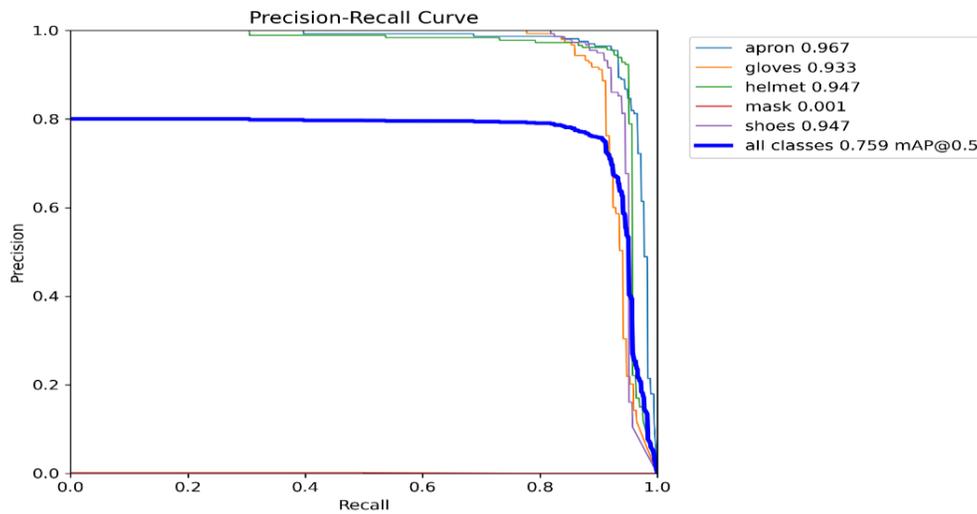


Fig. 5. Precision-Recall curve of the model.

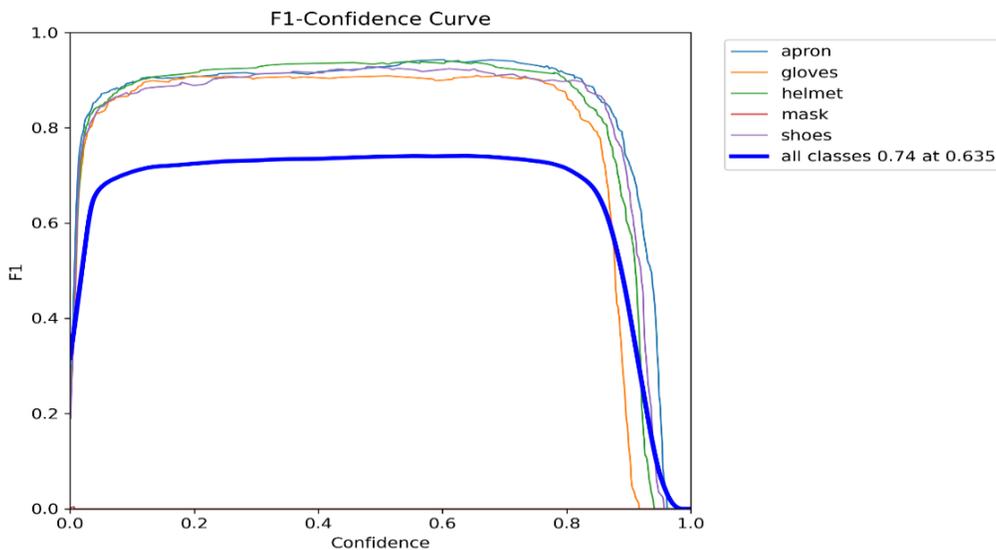


Fig. 6. F1-confidence curve of the model.



Fig. 7. The result of the model.

IV. RESULTS AND DISCUSSION

We have performance metrics for two object detection models, YOLOv5s and YOLOv8s, both applied to Personal Protective Equipment (PPE) detection. The metrics include Precision, Recall, mAP (mean Average Precision at a confidence threshold of 0.5), and F1-score. The table shows the comparison of the results of YOLOv8s and YOLOv5s. YOLOv8s has a higher precision (0.95) compared to YOLOv5s (0.88). This means YOLOv8s makes fewer false-positive predictions, resulting in more accurate detection of PPE items. High precision is essential to minimize false alarms, ensuring accurate PPE identification. YOLOv8s has a higher recall (0.84) compared to YOLOv5s (0.78). This suggests that YOLOv8s successfully identifies a higher proportion of actual PPE items within the dataset, indicating improved completeness in detection. A higher recall means fewer missed PPE items, enhancing safety. Both YOLOv8s and YOLOv5s have good mAP values at a confidence threshold of 0.5, with YOLOv5s slightly outperforming YOLOv8s (0.76 vs. 0.75). This indicates that both models perform well in terms of precision and recall, with YOLOv5s being slightly more consistent in precision and recall trade-offs. The F1-score for both models is the same (0.74). This score represents the balance between precision and recall. YOLOv5s achieves a better balance between these two metrics, whereas YOLOv8s has a higher precision but a slightly lower recall.

In summary, YOLOv8s exhibits higher precision and slightly lower recall compared to YOLOv5s. The choice between these models depends on specific requirements. YOLOv8s is more accurate in identifying PPE and reducing false positives, but it may miss some items. YOLOv5s provides a good balance between precision and recall, which might be preferable in scenarios where minimizing false alarms is crucial. Ultimately, the choice depends on the trade-off between precision and recall that aligns with the specific PPE detection objectives. A graph of the comparison result for the different algorithms is depicted in Table I and Fig. 8.

TABLE I. THE COMPARISON OF THE RESULTS FOR THE DIFFERENT ALGORITHMS

Model	Precision	Recall	mAP0.5
YOLOv5s	0.88	0.78	0.76
YOLOv8s	0.95	0.84	0.75
Faster R-CNN	0.92	0.82	0.78
SSD	0.89	0.80	0.73
RetinaNet	0.93	0.82	0.74

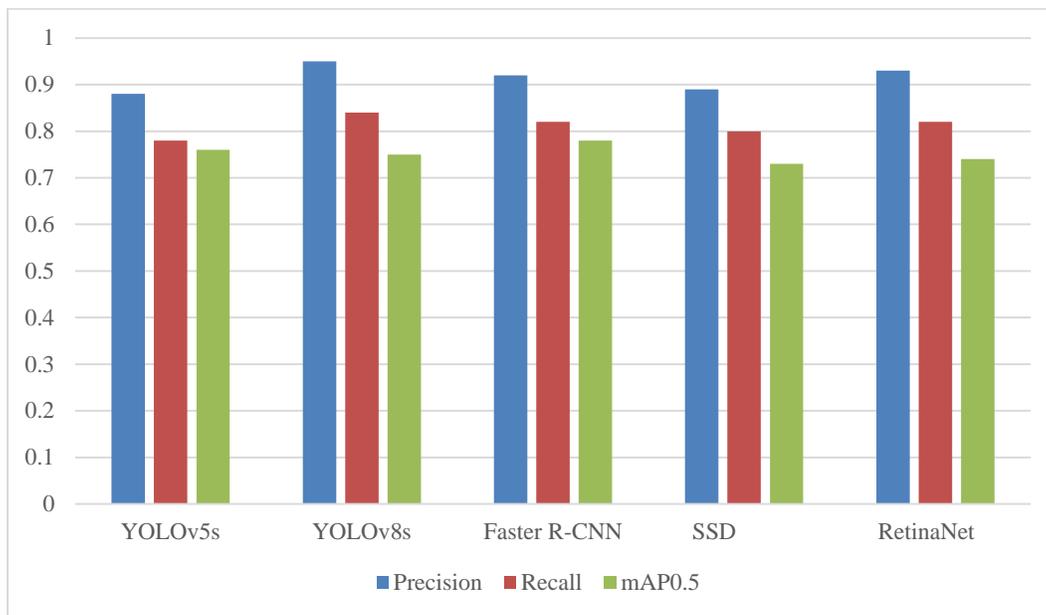


Fig. 8. Graph of the comparison result for the different algorithms.

The Table I and Fig. 9 present the performance metrics of various object detection algorithms, namely YOLOv5s, YOLOv8s, Faster R-CNN [17], RetinaNet [18], and SSD [19] in the context of PPE detection. YOLOv8s stands out with the highest precision of 0.95, indicating its ability to correctly identify PPE items with minimal false positives. This precision score suggests that YOLOv8s is adept at distinguishing between positive and negative predictions, crucial for applications where false alarms can have significant consequences. Additionally, YOLOv8s achieves a recall of 0.84, demonstrating its effectiveness in capturing a substantial portion of PPE instances present in the images. This balance between precision and recall is indicative of YOLOv8s' robust performance in detecting PPE across various scenarios.

Faster R-CNN also performs admirably with a precision of 0.92 and a recall of 0.82, closely trailing behind YOLOv8s. This indicates that Faster R-CNN is capable of achieving high accuracy in PPE detection, albeit with a slightly lower recall compared to YOLOv8s. Meanwhile, RetinaNet showcases competitive performance with a precision of 0.93 and a recall of 0.82, indicating its effectiveness in accurately identifying PPE items. However, SSD lags behind slightly with a precision of 0.89 and a recall of 0.80, suggesting that it may struggle with certain aspects of PPE detection compared to the other algorithms.

As result, YOLOv8s emerges as the algorithm with the overall best performance for PPE detection based on the provided metrics. Its combination of high precision and recall, along with a competitive mAP0.5 score, demonstrates its superiority over other models in accurately identifying PPE items. While Faster R-CNN and RetinaNet also exhibit strong performance, YOLOv8s' consistently high precision and recall make it the preferred choice for PPE detection tasks where both accuracy and reliability are paramount.

V. CONCLUSION AND FUTURE WORK

In industrial environments, the importance of Personal Protective Equipment (PPE) cannot be overstated, as it serves as a critical safeguard against potential hazards. While various methods for PPE detection have been explored in the literature, deep learning approaches have emerged as frontrunners, consistently delivering superior accuracy. However, a pressing research challenge persists in deep learning-based PPE detection, revolving around the need for even higher accuracy, non-destructiveness, and real-time capabilities, as evidenced by previous studies. In response to this challenge, this study proposes a deep learning model utilizing the YOLOv8 architecture, specifically designed to address the demanding accuracy, non-destructive, and real-time requirements. The model is trained on a custom dataset and validated to ensure robust performance, and extensive experiments and performance evaluations are conducted to demonstrate its effectiveness. In this study, a dataset is generated with extensive diversity of images, aiming to cover a wide array of scenarios relevant to PPE detection. From varying environments to different lighting conditions and types of the PPE), the dataset ensures robustness and adaptability in algorithm training. Its scalability enables extensive training and testing, enhancing the algorithm's accuracy and generalization capabilities. By including images from diverse settings and featuring different types of PPE, the dataset enriches the learning process, equipping algorithms to detect PPE across a range of real-world situations. Ultimately, this dataset serves as a valuable resource for developing accurate and reliable PPE detection algorithms, vital for workplace safety and compliance. The experimental results and performance evaluation underscore the proposed method's ability to achieve remarkable accuracy, making it a promising solution for PPE detection in industrial environments, satisfying the stringent demands of accuracy, non-destructiveness, and real-time functionality. Two limitations in PPE detection methods are, first, the potential difficulty in differentiating between visually

similar PPE items, such as gloves and protective sleeves, which could lead to misclassifications and compromise safety. Second, current PPE detection models may struggle in highly dynamic industrial environments, where the rapid movement of workers and changing scenes can hinder real-time tracking and detection accuracy. To address these limitations, two potential future research directions could be: Exploring advanced computer vision techniques, such as fine-grained object recognition and fusion with other sensor data, to improve the discrimination between visually similar PPE items, enhancing the model's precision. Investigating the integration of state-of-the-art real-time tracking algorithms and more sophisticated motion analysis to enhance PPE detection in dynamic industrial settings, ensuring accurate and reliable monitoring in high-speed, constantly changing work environments. Moreover, evaluation of different dataset is suggested to investigate the performance of the method on others scenarios.

REFERENCES

- [1] Y.-R. Lee, S.-H. Jung, K.-S. Kang, H.-C. Ryu, and H.-G. Ryu, "Deep learning-based framework for monitoring wearing personal protective equipment on construction sites," *J Comput Des Eng*, vol. 10, no. 2, pp. 905–917, 2023.
- [2] M. Zhao and M. Barati, "Substation Safety Awareness Intelligent Model: Fast Personal Protective Equipment Detection using GNN Approach," *IEEE Trans Ind Appl*, 2023.
- [3] X. Ji, F. Gong, X. Yuan, and N. Wang, "A high-performance framework for personal protective equipment detection on the offshore drilling platform," *Complex & Intelligent Systems*, pp. 1–16, 2023.
- [4] K. O. Monnikhof, P. Areerob, Z. Wu, T. Tanasitukul, and W. Kumwilaisak, "Novel Personal Protective Equipment Detection Technique with Attention-based YOLOv7 and Human Pose Estimation," *APSIPA Trans Signal Inf Process*, vol. 12, no. 1, 2023.
- [5] R. S. Aldossary, M. N. Almutairi, and S. Dursun, "Personal Protective Equipment Detection Using Computer Vision Techniques," in *SPE Gas & Oil Technology Showcase and Conference, SPE*, 2023, p. D021S031R001.
- [6] R. Wu and P. R. Selvaganapathy, "Porous biocompatible colorimetric nanofiber-based sensor for selective ammonia detection on personal wearable protective equipment," *Sens Actuators B Chem*, vol. 393, p. 134270, 2023.
- [7] A. M. Vukicevic, M. Djapan, V. Isailovic, D. Milasinovic, M. Savkovic, and P. Milosevic, "Generic compliance of industrial PPE by using deep learning techniques," *Saf Sci*, vol. 148, p. 105646, 2022.
- [8] K. Nisa, F. N. Fajri, and Z. Arifin, "Implementation of Personal Protective Equipment Detection Using Django and Yolo Web at Paiton Steam Power Plant (PLTU)," *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika (JITEKI)*, vol. 9, no. 2, pp. 333–347, 2023.
- [9] S. Bashir, R. Qureshi, A. Shah, X. Fan, and T. Alam, "YOLOv5-M: A Deep Neural Network for Medical Object Detection in Real-time," in *2023 IEEE Symposium on Industrial Electronics & Applications (ISIEA)*, IEEE, 2023, pp. 1–6.
- [10] L. Leite et al., "Prussian Blue Sensor for Bacteria Detection in Personal Protection Clothing," *Polymers (Basel)*, vol. 15, no. 4, p. 872, 2023.
- [11] M. C. Ang, E. Sundararajan, K. W. Ng, A. Aghamohammadi, and T. L. Lim, "Investigation of Threading Building Blocks Framework on Real Time Visual Object Tracking Algorithm," *Applied Mechanics and Materials*, vol. 666, pp. 240–244, 2014.
- [12] M. E. I. C. ANG, A. Aghamohammadi, K. O. K. W. NG, E. Sundararajan, M. Mogharrebi, and T. L. LIM, "Multi-Core Frameworks Investigation On A Real-Time Object Tracking Application.," *J Theor Appl Inf Technol*, vol. 70, no. 1, 2014.
- [13] K. O. P. P. Nugraha and A. P. Rifai, "Convolutional Neural Network for Identification of Personal Protective Equipment Usage Compliance in Manufacturing Laboratory," *Jurnal Ilmiah Teknik Industri*, vol. 22, no. 1, pp. 11–24, 2023.
- [14] J. Ahn, J. Park, S. S. Lee, K.-H. Lee, H. Do, and J. Ko, "SafeFac: Video-based smart safety monitoring for preventing industrial work accidents," *Expert Syst Appl*, vol. 215, p. 119397, 2023.
- [15] N. Ngoc-Thoan, D.-Q. T. Bui, C. N. N. Tran, and D.-H. Tran, "Improved detection network model based on YOLOv5 for warning safety in construction sites," *International Journal of Construction Management*, pp. 1–11, 2023.
- [16] M. Kang, C.-M. Ting, F. F. Ting, and R. C.-W. Phan, "BGF-YOLO: Enhanced YOLOv8 with Multiscale Attentional Feature Fusion for Brain Tumor Detection," *arXiv preprint arXiv:2309.12585*, 2023.
- [17] Saudi MM, Ma'arof AH, Ahmad A, Saudi AS, Ali MH, Narzullaev A, Ghazali MI. "Image detection model for construction worker safety conditions using faster R-CNN". *International Journal of Advanced Computer Science and Applications*. 2020.
- [18] Nath ND, Behzadan AH, Paal SG. "Deep learning for site safety: Real-time detection of personal protective equipment". *Automation in Construction*. 2020 Apr 1;112:103085.
- [19] Bhing NW, Sebastian P. "Personal Protective Equipment Detection with Live Camera". In *2021 IEEE International Conference on Signal and Image Processing Applications (ICSIPA) 2021 Sep 13* (pp. 221-226). IEEE.