Grenze International Journal of Engineering and Technology, June Issue

PPE Compliance Monitoring using Lightweight YOLO Object Detection Models

Srinithi S[1], Bandaru Pavana Sandhya[2] and J. Anitha[3]
[1-3]CSE, Karunya Institute of Technology and Sciences, Coimbatore, India
Email: srinithis22@karunya.edu.in, bandarupavana@karunya.edu.in, anitha_j@karunya.edu

***Abstract*— The industrial and construction sectors require Personal Protective Equipment (PPE) compliance to protect workers from accidents at work, but the existing manual monitoring systems face challenges because they struggle to detect violations accurately. You Only Look Once (YOLO) based object detection models are used to create an automated system for real-time safety monitoring. The models are trained and evaluated on a publicly available PPE dataset containing annotated instances of safety equipment. The evaluation process involves testing lightweight YOLO variants such as YOLOv8n, YOLOv11n, and YOLOv12n under the same training conditions. The models are assessed using precision, recall, mean Average Precision (mAP) at IoU threshold 0.5 (mAP@50), and mAP averaged across multiple IoU thresholds from 0.5 to 0.95 (mAP@50–95). The results show that YOLOv12n detects objects with maximum accuracy because it performs better in various testing conditions than the other tested variants. This implies that the adapted lightweight version of YOLO can effectively perform real-time PPE compliance detection and can hence be used efficiently to monitor safety in construction sectors.**

***Index Terms*— PPE Compliance Detection, YOLOv8n, YOLOv11n, YOLOv12n, Lightweight Object Detection, Real-Time Safety Monitoring, Construction Site Safety, Computer Vision, Deep Learning, mAP Evaluationwords.**

I. Introduction

Workplace safety remains a major concern within high-risk sectors, including construction, manufacturing, and mining, where employees are regularly subjected to perilous conditions. The Personal Protective Equipment (PPE) equipment, encompassing safety helmets, reflective vests, gloves, and masks, constitutes the initial safeguard against workplace injuries. The regulatory bodies have created strict safety requirements which enforce PPE usage, yet organizations face major difficulties with maintaining continuous compliance. In many real-world situations, workers either fail to use safety gear, or they use safety gear incorrectly which results in accidents and serious injuries that could have been avoided. Consequently, this highlights the requirement for intelligent monitoring systems capable of ensuring dependable and continuous compliance with safety standards.
Conventional PPE monitoring relies significantly on personal inspections of the safety personnel or on the retrospective analysis of surveillance footage. The methods are easy to understand but they cannot operate at higher levels of performance because their design restricts their operational capacity. The need to monitor individuals without stopping their work leads to a situation where workers become exhausted and their exhaustion creates a higher risk of detecting safety violations.

Grenze ID: 01.GIJET.12.2.529_26

The process of manual monitoring fails to deliver immediate notifications and alerts, which serve as essential components for maintaining safety in work environments that experience constant changes. The process of observing worker activities in industrial facilities especially those with many employees becomes increasingly complex as more time passes. The existing systems need development of automated technologies which can track safety violations through instant monitoring and quick violation detection.

Recent developments in computer vision and deep learning technology enable the creation of intelligent systems which automatically identify and classify objects through visual data examination. The YOLO model family has become a leading solution because it delivers fast real-time object detection which achieves high accuracy rates and operates differently from traditional multi-stage detection systems because it analyzes the complete image through one forward pass which decreases processing demands. And achieves high performance because it functions successfully in surveillance environments which require instant results. Modern YOLO versions now use architectural enhancements which boost their ability to extract features and detect objects in challenging situations that include hidden elements, different light conditions and multiple present objects.

A system based on YOLO object detection models is used to identify personal protective equipment compliance violations in real time, because of insufficient manual monitoring systems. It uses annotated images to detect safety equipment which includes helmets and safety vests. A comparative analysis is performed to evaluate modern detection systems under constant training settings. The system delivers continuous safety monitoring for industrial environments, which enables organizations to expand their operations while it instantly detects safety violations that occur in fast-paced industrial settings.

II. Literature Survey

Recent research demonstrates that YOLO-based object detection systems show successful results in different real-world scenarios. The research examines safety helmet detection through testing various YOLO versions which include YOLOv5 , YOLOv10 and their transfer learning capabilities to prove their performance in real safety assessment situations [1]. Experimental techniques are used to assess vehicle detection performance through testing various YOLO detection systems. The research demonstrates how detection methods produce different accuracy results because of the specific system design employed in each case [2]. The two-stage detectors which include Faster R-CNN reach similar or greater accuracy results, but YOLO-based systems deliver superior performance for real-time tasks [3].

Existing studies investigate how YOLO systems operate in both real-time situations and specific domain testing environments. The live-stream environment has been effectively used by YOLO-based methods to identify humans and estimate their body positions which shows their ability to process information with minimal time delay [4]. The CEM-YOLO system employs modified YOLO architectures to monitor infrastructure at large industrial facilities which improves its capability to detect defects and extract features from the monitored systems [5].

Multiple studies have tested different YOLO versions to assess their performance in various types of real-world applications. In agricultural applications, performance assessments based on palm oil ripeness detection and tree detection demonstrate how model variations affect their accuracy and computational efficiency [6], [7]. YOLO models demonstrate their capability to conduct real-time object detection for different difficulty levels through direct testing in autonomous driving evaluation [8].

The advancement of YOLO technology is further explained through various review and survey studies. The complete reviews of YOLO algorithm development show how architects improved system design through each new version release [9], [10]. The analysis compares different YOLO model variants to show how they differ in their ability to detect objects and their need for processing power and their real-world performance [11], [12]. The findings indicate that YOLO-based models work effectively across different fields because they can be adapted to different situations. The direct comparison between studies becomes difficult because previous studies assess applications and use different testing methods. Therefore, a unified testing method is required to compare different lightweight YOLO models for real-time safety monitoring applications.

III. Methodology

Lightweight YOLO-based object detection systems are evaluated in a controlled testing environment. All the models are tested with a single dataset after preprocessing and hyperparameter configuration. The methodology includes dataset preparation followed by standardized training of YOLOv8n, YOLOv11n and YOLOv12n models. The performance is evaluated using precision, recall, mAP@50 and mAP@50–95 metrics combining with GFLOPs and inference time to represent the computational performance. The visual inspection of detection outputs is used to learn about model behavior during real-world tests. The design enables analysis of how different architectural designs impact detection accuracy, processing power and real-time performance. Various research has explored PPE detection through YOLO-based models yet existing methods focus specifically on testing equipment which includes helmets and vests through multiple data sets and training setups. The existing models show this problem because their different testing methods make it impossible to compare them directly and their testing methods make it impossible to find the best model for active system use. The first method of detection performance improvement through architectural changes or extra processing methods increases the computational demands of some methods. The lightweight YOLO variations enable users to achieve accurate performance while maintaining operational efficiency through their design which makes them ideal for monitoring safety in real-time situations. The models are evaluated under the same experimental conditions to assess their performance in detecting PPE compliance.

A. Dataset Preparation and Configuration

The publicly available PPE dataset from Kaggle (Shlok Raval – PPE Dataset YOLOv8) was used for experiments which included 8,814 images and 22,077 annotated instances divided among 14 object classes. The annotations follow the YOLO format, which includes normalized bounding box coordinates together with their respective class labels. The models used the training validation and testing subsets which were created from the dataset because this method enabled researchers to compare their results in an unbiased manner across all experiments. The Ultralytics pipeline was used for the training process which required all images to be resized at 512 × 512 pixel dimensions. The model evaluation used data augmentation techniques that produced changes through horizontal flipping, scaling and through color space adjustments which included brightness, contrast, saturation changes and through mosaic augmentation and random translations for different object scales, positions and environmental conditions. A YAML file is used to specify dataset configuration. This information contains paths to datasets and class labels in them. The classes include safety equipment such as gloves, goggles, hard hats, masks, and safety vests; violation categories such as no-gloves, no-mask, and no-safety-vest; and additional objects including person, ladder, safety cone and fall-detected. Multiple classes testing was conducted for actual safety monitoring situations. Fig. 1. shows the representative samples from the PPE dataset depicting diverse real-world environments which include construction sites and indoor workplace environments. The images show two different situations which show how personal protective equipment was used correctly and how it was used incorrectly while the testing of different lighting conditions, background details and object sizes.

Fig. 1. Example images from PPE dataset

B. Model Architecture

The object detection task was performed using three variants of the YOLO architecture, namely YOLOv8n, YOLOv11n, and YOLOv12n. The models belong to the nano (n) category, making them suitable for lightweight applications that need real-time operation with less processing power. The YOLO architecture establishes a single-stage detection system as shown in Fig. 2 which enables simultaneous object location and identification through one integrated network. The model processes an input image through a backbone network for feature extraction followed by a neck module for feature aggregation and a detection head that predicts bounding boxes, class probabilities and confidence scores. YOLOv8n serves as the reference model for this proposed work. It uses an anchor-free detection head which enables output predictions without needing to use predetermined anchor box settings. The backbone uses the C2f module to achieve better gradient distribution and to allow multiple feature paths to be reused. The architectural improvements of YOLOv11n over YOLOv8n include C3k2 blocks and a PSA module which work together to enhance the model's ability to capture long-range feature dependencies while keeping the same parameter count. The latest version of YOLOv12n includes an attentionbased system design which uses R-ELAN blocks to create better feature collection methods that help detect objects with different sizes and shapes. The three different variants belong to the nano category because they have similar parameter counts and computation requirements, enabling a fair comparison of different architectural design approaches.

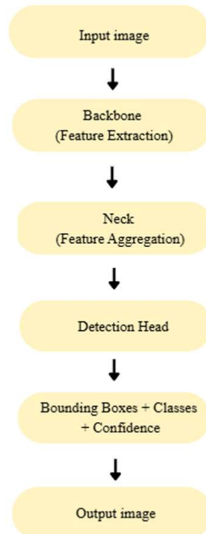

Fig. 2. YOLO object detection pipeline.

C. Training Configuration

All the models were experimented under same conditions to achieve fair evaluation. The training was conducted for 50 epochs, using a batch size of 16 on an NVIDIA Tesla T4 GPU, and was implemented with the Ultralytics YOLO framework.

The optimization process included three loss functions which were box loss, classification loss and distribution focal loss.

The Ultralytics framework automatically set the optimizer and learning rate. This meant the AdamW optimizer used model-specific learning rates. To keep things consistent, all models were trained using the same setup, including the same image size, batch size, and augmentation settings. Data augmentation followed the standard Ultralytics settings, which included horizontal flipping (0.5), mosaic augmentation, scaling transformations, HSV color changes, and random translations. The optimal model weights were determined by their validation performance.

D. Evaluation Metrics

Standard object detection metrics are used to assess the model performance which included precision, recall, mAP@50 and mAP@50–95. Precision measures the accuracy of positive predictions, while recall assesses how completely the model identifies relevant instances. Precision and recall are computed as shown in equation (1) and (2):

$$\text{Precision} = \frac{TP}{TP+F} \tag{1}$$

$$\text{Recall} = \frac{TP}{TP} \tag{2}$$

Equation (3) defines the Intersection over Union (IoU), which measures how much the predicted and ground truth bounding boxes overlap.

$$IoU = \frac{\text{Area of Overlap}}{\text{Area of Union}} \tag{3}$$

The mAP, which determines average detection efficiency across all categories, is computed according to equation (4). It defines true positives (TP), false positives (FP), false negatives (FN) while AP_irepresents the

(Average precision) of the i^{th}class. The mAP@50 evaluates performance at an IoU threshold 0.5 threshold while mAP@50–95 tests multiple IoU thresholds to determine performance, which enables the system to evaluate both localization and detection capabilities.

$$mAP = \frac{1}{N}\sum_{i=1}^{N} A\,P_i \quad (4)$$

IV. Results and Discussion

Model performance was evaluated through standard object detection metrics which included precision, recall, mAP@50 and mAP@50–95, along with computational efficiency measures such as GFLOPs and inference time. Table I presents the performance comparison results of YOLOv8n, YOLOv11n and YOLOv12n through their detection abilities and computational performance metrics. YOLOv12n achieves the highest detection performance with a mAP@50 of 0.720 and mAP@50–95 of 0.446 which enables it to surpass YOLOv8n by 3.7% in mAP@50 (0.720 vs. 0.694) and by 6.7% in mAP@50–95 (0.446 vs. 0.418). YOLOv12n achieves 0.747 recall performance which exceeds YOLOv8n 0.716 performance by 4.3%. YOLOv12n achieves its performance through 6.3 GFLOPs whereas YOLOv8n requires 8.1 GFLOPs which results in a 22% decrease of computational expenses while maintaining better performance efficiency. YOLOv11n delivers its fastest inference performance through 1.8 ms while matching YOLOv12n GFLOPs performance at 6.3 which makes it suitable for deployments with extreme resource limitations. Table II shows the epoch-wise development of YOLOv12n which demonstrates that mAP@50 improved from 0.424 at epoch 1 to 0.705 at epoch 30 showing an absolute increase of 28.1 percentage points. After epoch 30, it yielded only 1.4 percentage point gain (0.705 to 0.719 at epoch 50) which demonstrates that effective convergence happened between epoch 30 and 40.

TABLE I. YOLO MODELS PERFORMANCE COMPARISON

Model	mAP@50	mAP@50-95	Precision	Recall	GFLOPs	Inference (ms)
YOLOv12n	0.720	0.446	0.681	0.747	6.3	2.3
YOLOv11n	0.697	0.420	0.659	0.718	6.3	1.8
YOLOv8n	0.694	0.418	0.655	0.716	8.1	1.7

TABLE II. YOLOV12N PERFORMANCE ACROSS TRAINING EPOCHS

Epoch	Precision	Recall	mAP@50	mAP@50-95
1	0.493	0.503	0.424	0.214
10	0.561	0.647	0.599	0.325
20	0.644	0.692	0.672	0.394
30	0.656	0.728	0.705	0.425
40	0.681	0.739	0.717	0.439
50	0.677	0.747	0.719	0.445

The precision-recall relationship displays the results for all detected classes in Fig. 3. The classes gloves, goggles, and ladder show high detection performance because their visual features show complete uniqueness which matches the dataset representation. The two classes mask and no-safety-vest display lower precision results because their two categories show visual resemblance and partial occlusion of these items in real world actual scenes. The overall curve demonstrates an mAP@50 score of 0.720 because it shows balanced detection performance across all 14 detected classes. Fig. 4 displays the confusion matrix in class-wise prediction accuracy. The diagonal area shows correct classifications while the off-diagonal areas display misclassifications. The classes hardhat and safety cone show better performance because of their frequent occurrence together with their distinct visual characteristics. Confusion between visually similar class pairs occurs because the mask, no-mask, safety-vest and no-safety-vest pairs use subtle features that change according to different image resolutions and viewing angles. The training and validation losses show continuous decline in Fig. 5, while the precision and recall together with mAP metrics demonstrate steady improvement that reaches stability during the later training periods which shows successful model learning and convergence achievement. Fig. 6 demonstrates the model ability to identify various safety equipment items which include hard hats and safety vests through its accurate bounding box identification of individual workers in a controlled indoor space. The outdoor scenes in Fig. 7 show YOLOv12n detecting safety violations which include missing helmets and vests through its analysis of different background elements. The YOLOv12n-based approach successfully identifies most violations through its confidence scores, but it shows challenges for detecting violations because cluttered and occluded scenes create detection difficulties that result in reduced confidence levels.

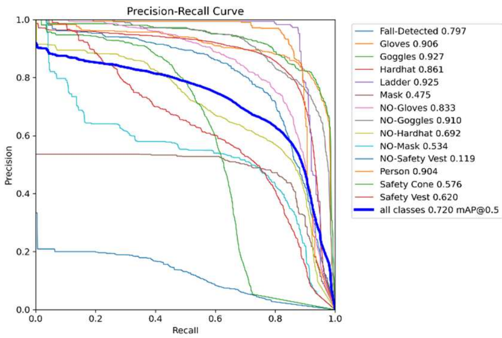

Fig. 3. Precision-recall curves for all classes with overall mAP performance

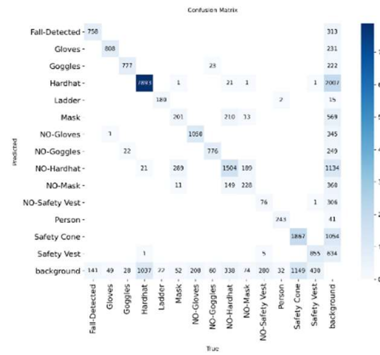

Fig. 4. Confusion matrix showing class-wise performance of YOLOv12n

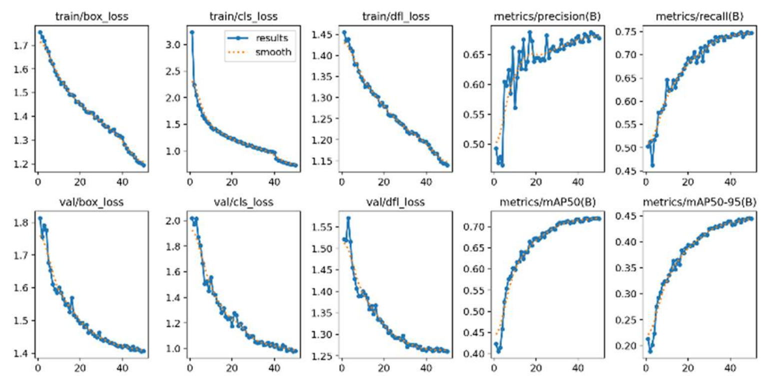

Fig. 5. Model performance trends across epochs

Fig. 6. Detection output of multiple safety equipment

Fig. 7. Sample detection output showing safety violation cases

V. CONCLUSION

The comparative study of YOLO based object detection models helps to understand their potential in PPE compliance monitoring with testing on a public dataset. YOLOv12n had the best detection capabilities as it achieves higher mAP@50 and mAP@50–95 which are better indicators of how good the model can localize objects accurately at different IoU thresholds. YOLOv11n offers a good compromise between precision and system resource usage while YOLOv8n is a stable but heavy resource consume benchmark system. The results illustrate that the architectural improvements incorporated in contemporary YOLO implementations yield increasingly stable and incremental performance gains, thus improving both the efficiency and reliability of systems operating under complex environmental conditions. The qualitative assessments also confirm the performance consistency towards various real-world factors namely varying illumination conditions, background challenge levels and scale of objects. For safety critical environments, lightweight YOLO models provide accurate detection results with less required computing power which ensures better monitoring of PPE compliance.

REFERENCES

[1] Chowdhury, Zayed Uddin, et al "Performance comparison of yolo models for safety helmet detection: Insights from yolov5 to yolov10 with transfer learning." In 2024 International Conference of Adisutjipto on Aerospace Electrical Engineering and Informatics (ICAAEEI), pp. 1-6. IEEE, 2024.

[2] Maity, Sourajit, Arpan Chakraborty, Pawan Kumar Singh, and Ram Sarkar. "Performance comparison of various yolo models for vehicle detection: An experimental study." In International conference on data analytics & management, pp. 677-684. Singapore: Springer Nature Singapore, 2023.

[3] Reswara, Evan, Sinung Suakanto, and Seno Adi Putra. "Comparison of Object Detection Algorithm using YOLO vs Faster R-CNN: A Systematic Literature Review." In Proceedings of the 2023 6th International Conference on Big Data Technologies, pp. 419-424. 2023.

[4] Sonawane, Achal, Varshini Dandam, Kiran Khamkar, Tanushri Wawge, and Priyanka More. "Leveraging YOLO for Real-Time Human Detection and Pose Estimation in Live Stream Environments." In 2025 International Conference on Computing and Communication Technologies (ICCCT), pp. 1-5. IEEE, 2025.

[5] Zhao, Xinyi, Fenghe Guo, Yongchun Wang, Chenyang Sun, Chen Gao, and Runjie Shen. "CEM-YOLO: Defect Detection in Large Hydropower Station Water Conveyance Pipelines Based on YOLOv7." In Journal of Physics: Conference Series, vol. 2988, no. 1, p. 012012. IOP Publishing, 2025.

[6] Chowdhury, Alvi Khan, et al. "Comparative Study of Different YOLO Variants for Ripeness Detection of Palm Oil Fresh Fruit Bunches." In 2025 IEEE 6th International Conference in Robotics and Manufacturing Automation (ROMA), pp. 18-23. IEEE, 2025.

[7] Shaikh, Istiyak M., Mohammad N. Akhtar, and Muhammad U. Gill. "Benchmarking YOLO-based models for palm oil tree detection: an in-depth analysis from YOLOv5X to YOLOv11X." In IET Conference Proceedings CP917, vol. 2025, no. 3, pp. 191-200. Stevenage, UK: The Institution of Engineering and Technology, 2025.

[8] Jin, Yichen. "Comparing YOLO Models for Self-Driving Car Object Detection." In 2025 IEEE 5th International Conference on Electronic Technology, Communication and Information (ICETCI), pp. 151-155. IEEE, 2025.

[9] Jiang, Peiyuan, Daji Ergu, Fangyao Liu, Ying Cai, and Bo Ma. "A Review of Yolo algorithm developments." Procedia computer science 199 (2022): 1066-1073.

[10] Pan, Yaohui, Gang Wang, and Jun Yu. "Overview of deep learning YOLO algorithm." In Fourth International Conference on Computer Vision, Application, and Algorithm (CVAA 2024), vol. 13486, pp. 622-630. SPIE, 2025.

[11] Chen, Aoxiang. "Comparative analysis of YOLO variants based on performance evaluation for object detection." In ITM Web of Conferences, vol. 70, p. 03008. EDP Sciences, 2025.

[12] Uzdiaev, Mikhail, Marina Astapova, and Elchin Khalilov. "A Comparative Study of YOLO." In Robotics in Agriculture: Proceedings of the Fifth International Conference on Agriculture Digitalization and Organic Production (ADOP 2025), Volume 1, vol. 1, p. 151. Springer Nature, 2026.