Volume XX, Issue (1) (2024) 101-110

International May Conference on Strategic Management

DOI: *10.5937/IMCSM24010P* *Original research*

HARD HAT DETECTION FOR SAFETY PURPOSES BY USING YOLOV9

Natalija Perišić*, Radiša Jovanović

University of Belgrade, Faculty of Mechanical Engineering, Serbia

Abstract: Ensuring the safety of workers at workplaces is a crucial task for every company. The usage of personal protective equipment represents the basic form of protection. Hard hats are very useful in protecting head from injuries. However, workers often neglect the importance of wearing safety helmets and do not wear them. Systems for monitoring and detecting unsafe behaviors can be very helpful for maintaining security. For that purpose, this research examines the success of the application of the latest YOLO algorithm for detecting the presence of safety helmets on workers that can be applied in those systems. Two models with different numbers of parameters are trained for this purpose – YOLOv9c and YOLOv9e. The results showed that YOLOv9c model achieved mean average precision of 97.2%, 93%, and 92.9% in training, validation, and testing, respectively, while YOLOv9e reached slightly higher mean average precisions of 97.5% in training, 93.4% in validation and 93.4% in testing.

Keywords: Hard hat detection, deep learning, YOLOv9, safety helmet detection.

1. INTRODUCTION

The safety of workers at the workplace is important not only to workers but also to organizational leaders and policymakers. Worker injuries can be devastating to workers and their families and they cause significant financial costs for organizations (McGonagle & Kath, 2010). One way to minimize exposure to danger that threatens to cause workplace injuries is to use personal protective equipment (PPE). Hard hats are engineered to endure impact and prevent penetration from objects, along with shielding against electrical dangers. When worn correctly by workers, a reduction in fall-related fatalities can be expected, as well as a notable decrease in injuries caused by slips, trips, and being struck by falling objects (Shrestha et al., 2015). Shrestha et al. (2015) also stated in their work the result of the research which showed that 47.3% of fatally injured workers either had not used PPE or had not used it properly. It is clear that systems created to monitor workers and their usage of PPE can help to a large extent in maintaining workplace safety and preventing injuries. Those systems are often based on the usage of deep neural networks that can detect objects in real time with great accuracy. However,

* *Corresponding author: nperisic@mas.bg.ac.rs*

Natalija Perišić, ORCID: 0000-0002-8675-1934
Radiša Jovanović, ORCID: 0000-0002-8122-756X

researchers are always trying to improve those systems by using different deep learning algorithms, upgrading monitoring equipment, etc.

This paper deals with the detection of safety helmets based on images of construction workers during their work. For detection, YOLOv9 deep learning network is used for the first time for this purpose. Keeping the same training parameters, two different YOLOv9 models are trained and a comparison is made.

2. LITERATURE REVIEW

The system proposed in the paper by Shrestha et al. (2015) introduces a construction site monitoring system aimed at enhancing safety practices within organizations. This system integrates closed-circuit television cameras deployed across the construction site, interconnected with a central server, typically hosted on an office computer. The captured video data undergoes analysis, and upon detection of unsafe behavior, alerts are generated via various communication channels such as SMS messages, notifications on other computers, or through speakers. The detection process within the system employs two primary methods - face detection and hard hat detection. Initially, the system conducts face detection to identify individuals within the monitored area. Subsequently, it assesses whether the detected individuals are wearing protective hard hats. Notably, the hard hat detection process is initiated only upon successful face detection. While the proposed system offers a relatively straightforward implementation and is cost-effective, it acknowledges limitations in its detection methodology. Specifically, the reliance on face detection may result in instances where individuals are not properly identified due to occlusions or other environmental factors common in construction sites.

To overcome the mentioned limitation, researchers tried to use You Only Look Once (YOLO) algorithm for detection. It has the ability to process images very fast while containing a high accuracy rate. Wen et al. (2020) conducted a comparative study in which they evaluated an improved YOLOv3 network against YOLOv2 and the standard YOLOv3 for safety helmet detection. Their findings revealed that the improved YOLOv3 model attained a precision of 90.7%, surpassing precisions achieved by YOLOv2 and the conventional YOLOv3. Benyang et al. (2020) created an improved approach built upon the foundation of YOLOv4 for the identical task, achieving a mean Average Precision (mAP) value of 99.89%. This triumphs over the mAP values of both YOLOv3 and the conventional YOLOv4 model. Despite showing a lower mAP value when compared to Faster Region - Convolutional Neural Network, the proposed model demonstrated a substantially superior detection speed which is 75 times faster than the aforementioned alternative. Paper (Kisaezehra et al., 2023) offers a deep learning strategy for monitoring worker’s hard hats in real-time using five different YOLOv5 models – YOLOv5n, YOLOv5s, YOLOv5m, YOLOv5l and YOLOv5x. YOLOv5x network achieved the highest value of mAP, 95.8%, while YOLOv5n had the fastest detection speed of 70.4 frames per second. It can be said that these results were expected because YOLOv5x model is the largest, that is, it has the greatest number of parameters, while YOLOv5n is the smallest and, therefore, the fastest. YOLOv5 also overcame the performances of YOLOv6 with a mAP value of 75.7% in research conducted by Ludwika and Rifai (2024). The research focused on using YOLO models to detect seven different PPE objects and to ascertain whether each item was correctly utilized. An improved model based on YOLOv7 model is developed for detecting helmets in six different categories that include helmet, head with helmet, person with helmet, head, person no helmet, and face category. The improved model achieved slightly better results than all standard YOLOv7 models, achieving mAP value of 92.6% (Chen et al., 2023). Sometimes, mAP value is influenced not only by the model of neural network but also by the dataset that is used. An example of this statement can be found in (Chen & Xie, 2023). That research compared an improved YOLOv7 model with twelve other deep learning models applied to three different datasets for helmet detection. Lin (2024) developed improved YOLOv8 model, that has less parameters than YOLOv8n model, which is the smallest model of YOLOv8 networks. It achieved the highest mAP of 94.36% which is even higher than mAP value achieved by YOLOv8x, which is the second-largest YOLOv8 model.

3. DATASET

The dataset utilized for training and evaluating deep learning models was developed by Aydin and Bulut (2023) and it is available online. It comprises 23 637 distinct images depicting individuals wearing or not wearing safety helmets and vests. Additionally, there is a smaller subset of images exclusively featuring safety vests or hard hats. Each image within the dataset is resized to dimensions of 640 x 640 pixels and comes pre-labeled. Furthermore, the dataset is organized into separate subsets for training, testing, and validation. The training subset encompasses 78% of all images, while the test and validation subsets account for 11% of the total image count each. Random examples from the dataset are depicted in Figure 1.

Figure 1. Random samples of images from the dataset (Aydin & Bulut, 2023)

Even if there are three different classes – helmet, head, and vest, this research takes into consideration only two cases when individuals wear hard hats and when they do not, so vest class is neglected in further study. Figure 2 presents the number of instances for each class in every subset.

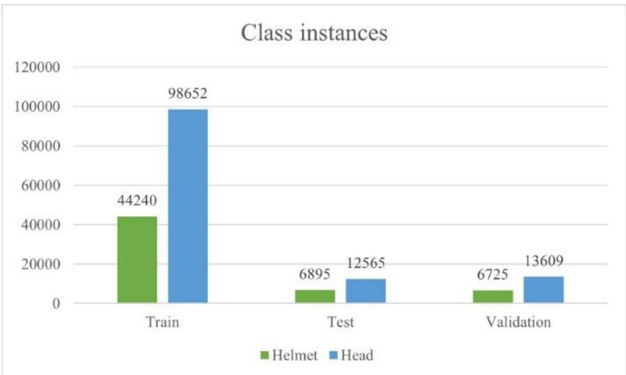

Figure 2. The number of class instances in all subsets

4. YOLOv9 DETECTION MODEL

YOLOv9 is a deep learning detection model developed to overcome the problem of information lost during the feedforward process. It combines the Programmable Gradient Information (PGI) concept and Generalized Efficient Layer Aggregation Network (GELAN), both created by Wang et al. (2024., preprint). The design of YOLOv9 model is based on the Information Bottleneck Principe and Reversible Functions.

Information Bottleneck Principe suggests that during the transformation of data through layers in neural networks information loss can be caused, as it is shown in Equation 1.

$$I(X,X) \geq I\big(X, f_\theta(X)\big) \geq I\left(X, g_\phi\big(f_{\theta(X)}\big)\right), \tag{1}$$

where I indicates mutual information, f, and g represent transformation functions and θ and ϕ represent parameters of transformation functions. In deep neural networks, functions $f_\theta(\cdot)$, and $g_\phi(\cdot)$ are the operations of two sequential layers. It is clear that increasing the number of layers in neural networks increases the probability that the original data will be lost. This means that using incomplete information in training of network can result in unreliable gradients and poor convergence. This disadvantage can be solved by using reversible functions. The reversible function, v, is an inverse transformation function of function r, as shown in Equation 2.

$$X = v_\zeta\left(r_\psi(X)\right), \tag{2}$$

where ζ and ψ are parameters of functions v and r, respectively. Equation 3 shows that if data X is converted by reversible function, then it will not come to the information lost.

$$I(X,X) = I\left(X, r_\psi(X)\right) = I\left(X, v_\zeta\left(r_\psi(X)\right)\right), \tag{3}$$

If network's transformation function includes reversible functions, then more dependable gradients can be obtained for updating the model. However, the repeated passing of original data through subsequent layers in deep learning architectures may lead to effective convergence but for difficult problems, it makes it harder to find simple mapping functions to map data to target. This especially affects the performance of lightweight models because they are under-parameterized when faced with a large amount of raw data.

PGI is a method that is developed to solve analyzed problems and is the first component of YOLOv9. It consists of three components – main branch, auxiliary reversible branch, and multi-level auxiliary information. PGI has only main branch for inference process which means that it does not require any additional computational costs. Next, auxiliary reversible branch serves to generate dependable gradients and update parameters of neural network. It can give reliable gradient information in learning process as main branch deep features might lose significant information due to information bottleneck. This gradient information guides parameter updates to help extract accurate and crucial information, thereby enabling the main branch to obtain features that are better suited for the target task. Nevertheless, PGI method does not compel the main branch to keep all original information but rather enhances it by generating valuable gradients through the auxiliary supervision mechanism. Also, the auxiliary reversible branch can be eliminated during the inference phase, so the original network's inference capabilities can remain intact. Finally, Multi-level auxiliary information has the task of combining gradient information by inserting integration network between feature pyramid layers of main and auxiliary supervision branch. The gathered information contains all target objects and it is forwarded to the main branch for updating parameters. The point is to mitigate the domination of specific object information to main branch's feature pyramid hierarchy. This technique is useful for making more accurate predictions for objects that have different sizes.

The second component of YOLOv9 model is GELAN. It is a neural network architecture created by combining two neural networks – Cross Stage Partial Network (CSPNet) and Efficient Layer Aggregation Network (ELAN). It considers lightweight design, accuracy, and inference speed. The main idea behind GELAN is to use ELAN architecture, which only uses a composition of convolutional layers, but to generalize it by using any computational blocks, which is characteristically for CSPNet. Figure 3 shows the PGI method and GELAN architecture, which represent the base of YOLOv9.

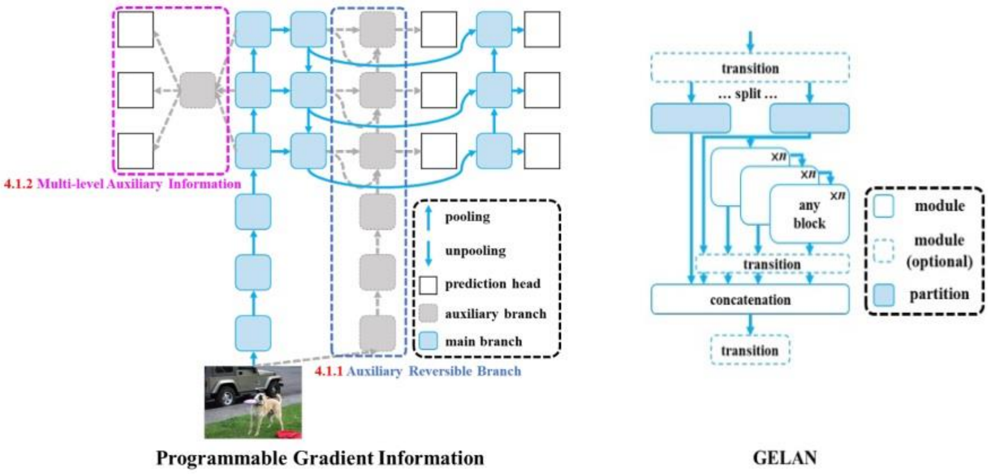

Figure 3. PGI and GELAN, (Wang et al., 2024, preprint)

5. TRAINING PARAMETERS

For the experiment, the image size is kept as the default size for YOLOv9 models. Two models are trained – YOLOv9c and YOLOv9e, with the same training parameters. As data augmentation methods translating, scaling, left-right flipping, mosaic, mixing up, and copy-pasting segments of image are used. Some of the parameters are given in Table 1.

Table 1. Training parameters

Parameter	Value
Optimizer	Stochastic Gradient Descent
Number of epochs	100
Batch size	8
Learning rate	0.01
Momentum	0.937
Weight decay	0.0005
Translate probability	0.1
Image scale	0.9
Flip left-right probability	0.5
Mosaic probability	1.0
Mix up probability	0.15
Segment copy-paste probability	0.3

6. RESULTS AND DISCUSSION

6.1. Evaluation criteria

One of four cases can be a result of each detection that describes if our model detected and classified an object correctly. Those cases are:

- True positive (TP), the case when the object is correctly detected
- False positive (FP), in the case when the detected bounding box is misplaced or in the case of incorrect object detection
- False negative (FN) when the object in the image is not detected
- True negative (when instances in pictures are not detected as objects, this case is not taken into consideration for object detection problems, because there is an infinite number of instances in the pictures that are not detected as objects)

Sums of all true positive, false positive, and false negative cases allow for calculating network's performance measures. In object detection, precision, recall, mAP, and F1 score are usually considered as performance measures.

Precision is the ability of a model to identify only objects that are relevant, in other words, the percentage of correct positive predictions (Padilla et al., 2020). It can be calculated as:

$$\text{Precision}_c = \frac{\text{TP}_c}{\text{TP}_c + \text{FP}_c}, \tag{4}$$

where c represents object's class. Recall represents the model's ability to find all relevant classes (Padilla et al., 2020). It can be represented via following equation:

$$\text{Recall}_c = \frac{\text{TP}_c}{\text{TP}_c + \text{FN}_c}, \tag{5}$$

F1 score is performance measure that takes into account precision and recall value, in the form as in the Equation 6.

$$F1_c = \frac{2 \cdot \text{Precision}_c \cdot Recall_c}{\text{Precision}_c + Recall_c}. \tag{6}$$

AP refers to the average value of all precisions obtained under all possible recall rates. The mAP is the average of the AP value in all categories (Jin et al., 2021), and it can be calculated as:

$$mAP \ = \frac{1}{N} \sum_{c=1}^{N} AP_c \ , \tag{7}$$

where *N* is the number of classes, c.

6.2. Obtained results

Results obtained after a hundred epochs on train, test, and valuation data for both models are shown in Table 2.

Table 2. Values of performance measures

	YOLOv9c			YOLOv9e		
	Train	Validation	Test	Train	Validation	Test
$\text{Precision}_{\text{helmet}}$	94.3%	93.2%	93.6%	94.6%	93.5%	93.6%
$\text{Precision}_{\text{head}}$	94.7%	91%	91.3%	95%	91.1%	91.2%
Overall Precision	94.5%	92.1%	92.5%	94.8%	92.3%	92.4%
$\text{Recall}_{\text{helmet}}$	94.8%	89.3%	88.1%	95.4%	89.3%	88.7%
$\text{Recall}_{\text{head}}$	93%	91%	91.9%	93.7%	91.8%	93.3%
Overall Recall	93.9%	90.1%	90%	94.6%	90.5%	91%
$\text{F1}_{\text{helmet}}$	94.5%	91.2%	90.8%	95%	91.4%	91.1%
F1_{head}	93.8%	91%	91.6%	94.3%	91.4%	92.2%
Overall F1	94.2%	91.1%	91.2%	94.7%	91.4%	91.7%
mAP	97.2%	93%	92.9%	97.5%	93.4%	93.4%

From Table 2 is clear that YOLOv9e, a larger network, that is, a network with more parameters achieved slightly better performance results. Even if YOLOv9c achieved the same or higher precision values in the test category, recall has smaller values. That means this model is capable of correct detection but there are more FN cases, cases when instances are presented in the image, but not recognized. However, as YOLOv9c has less number of parameters, learning is faster, as well as the detection process.

a b

Figure 4. Examples of detection on the test set of YOLOv9c – a) examples of correct detection b) examples of incorrect detection

Figure 4 represents YOLOv9c detection results on the test set, which has not taken part in the training process, meaning that this data is unknown for the trained model. Part a) shows some examples of correct detection, while part b) shows some mistakes made by using YOLOv9c model. It should be pointed out that, as the results show, there are a lot more cases of correct detection, but for the sake of comparison between proposed models, more examples of mistakes are given in this paper. As can be seen, the model can neglect the presence of heads and helmets, not recognizing them which can cause problems, especially in the case of not detecting heads. On the other hand, sometimes a mistake is caused by detecting the wrong object as the left top image of the b) part in Figure 4 shows, when glove is recognized as helmet. Figure 5 shows the detection results of YOLOv9e model, for the same image data.

a b

Figure 5. Examples of detection on the test set of YOLOv9e – a) examples of correct detection b) examples of incorrect detection

YOLOv9e model also correctly detected heads and helmets on the data showed in Figure 5, part a). However, when parts b) of Figures 4 and 5 are compared, it can be noted that YOLOv9e model did not make mistakes on the two left images, as YOLOv9c did. Of course, some mistakes are made, as shown in the right images of part b) in Figure 5, but they are the same that YOLOv9c made.

7. CONCLUSION

This study acknowledges the significance of employing deep learning models to enhance workplace safety for workers. Initially, it examines the deployment of earlier YOLO algorithms for identifying PSE, based on published researches. In continuation, it elucidates the advancements and innovations achieved by the YOLOv9 algorithm, representing the latest iteration of YOLO algorithm. Further, two large models, YOLOv9c and YOLOv9e, are trained for detecting the utilization of safety helmets by workers at workplaces, the first time for the mentioned purpose.

Results showed that both models achieve high value of mAP, but as it was expected, larger model, YOLOv9e, showed as a more precise model, achieving mAP of 97.5% during training, 93.4% during validation, and 93.4% during testing phases. It was observed that certain errors made by YOLOv9c could be mitigated by employing YOLOv9e. However, training of YOLOv9e is computationally demanding, so the training is more time-consuming and the detection process is slower.

In order to improve performances of models, it is recommended to use different hyperparameters, optimization, and regularization techniques. Further research may refer to training lighter models of YOLOv9, or implementing trained models on real time systems, as they are reliable and easy to use.

ACKNOWLEDGMENT

This paper is the result of research supported by the Ministry of Education, Technological Development and Innovations of the Republic of Serbia under contract 451-03-65/2024-03/ 200105 from 05.02.2024 and by the Science Fund of the Republic of Serbia, #GRANT No. 5151, Support Systems for Smart, Ergonomic and Sustainable Mining Machinery Workplaces – SmartMiner.

REFERENCES

Aydin, M. Z., & Bulut, B. (2023). HardHat-Vest Dataset. Retrieved from Kaggle: https://www.kaggle.com/datasets/muhammetzahitaydn/hardhat-vest-dataset-v3

Benyang, D., Xiaochun L., & Miao, Y. (2020). Safety Helmet Detection Method Based on YOLO v4. *2020 16th International Conference on Computational Intelligence and Security (CIS)* (pp. 155-158). Guangxi, China: IEEE. https://doi.org/10.1109/CIS52066.2020.00041

Chen, J., Zhu, J., Li, Z., & Yang, X. . (2023). YOLOv7-WFD: A Novel Convolutional Neural Network Model for Helmet Detection in High-Risk Workplaces. *IEEE Access,* 11, 113580-113592. https://doi.org/10.1109/ACCESS.2023.3323588

Chen, X., & Xie, Q. . (2023). Safety Helmet-Wearing Detection System for Manufacturing Workshop Based on Improved YOLOv7. *Journal of Sensors,* 2023. https://doi.org/10.1155/2023/7230463

Jin, Z., Qu, P., Sun, C., Luo, M., Gui, Y., Zhang, J., & Liu, H. (2021). DWCA-YOLOv5: An Improve Single Shot Detector for Safety Helmet Detection. *Journal of Sensors,* 2021, 1-12. https://doi.org/10.1155/2021/4746516

Kisaezehra, Farooq, M. U., Bhutto, M. A., & Kazi, A. K. . (2023). Real-Time Safety Helmet Detection Using Yolov5 at Construction Sites. *Intelligent Automation & Soft Computing,* 36(1), 911-927. https://doi.org/10.32604/iasc.2023.031359

Lin, B. (2024). Safety Helmet Detection Based on Improved YOLOv8. *IEEE Access,* 12, 28260-28272. https://doi.org/10.1109/ACCESS.2024.3368161

Ludwika, A.S., & Rifai, A., P. (2024). Deep Learning for Detection of Proper Utilization and Adequacy of Personal Protective Equipment in Manufacturing Teaching Laboratories. *Safety* , 10(1), 26. https://doi.org/10.3390/safety10010026

McGonagle, A. K., & Kath, L. M. (2010). Work-safety tension, perceived risk, and worker injuries: A meso-mediational model. *Journal od Safety Research,* 41(6), 475-479. https://doi.org/10.1016/j.jsr.2010.09.002

Padilla, R., Netto, S. L., & da Silva, E. A. B. (2020). A Survey on Performance Metrics for Object-Detection Algorithms. *2020 International Conference on Systems, Signals and Image Processing (IWSSIP).* Niteroi, Brasil: IEEE. https://doi.org/10.1109/IWSSIP48289.2020.9145130

Shrestha, K., Shrestha, P. P., Bajracharya, D., & Yfantis, E. A. (2015). Hard-Hat Detection for Construction Safety Visualization. *Journal of Construction Engineering,* 2015(1), 1-8. https://doi.org/10.1155/2015/721380

Wang, C.Y., Yeh, I.H., & Liao, H.Y. M. (2024). YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information. *arXiv preprint arXiv:2402.13616*. https://doi.org/10.48550/arXiv.2402.13616

Wen, P., Tong, M., Deng, Z., Qin, Q., & Lan, R. (2020). Improved Helmet Wearing Detection Method Based on YOLOv3. *Artificial Intelligence and Security. ICAIS 2020. Lecture Notes in Computer Science().* 12239, pp. 670-681. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-57884-8_59