International Research Journal of Advanced Engineering and Science
ISSN (Online): 2455-9024

Implementation of Machine Learning for Personal Protective Equipment Detection Using Convolutional Neural Network

Mochammad Daffa Daniarfa Darmawan[1], Wahyu Kusuma Raharja[2]
[1]Department of Electrical Engineering, Gunadarma University, Depok, West Java, Indonesia-1624
[2]Department of Electrical Engineering, Gunadarma University, Depok, West Java, Indonesia-1624

Abstract— *Safety in construction is very important as the industry is prone to dangerous situations. Some of the factors that cause accidents on construction sites are personal protective equipment not provided to workers, workers not using the provided personal protective equipment and lack of adequate safety training for workers. Construction accidents sometimes occur because workers do not use the safety equipment provided. The frequency of accidents can be reduced if entities can observe workers continuously to immediately identify when safety equipment is not being used properly. Machine learning, part of the field of artificial intelligence (AI), is a technology that is widely researched and used for problem solving. The purpose of this research is to create a system that can detect whether workers use personal protective equipment or not, using machine learning and computer vision technology. This research applies an applied research approach, using the Convolutional Neural Networks (CNN) method which is commonly used in object detection using Machine Learning and Computer Vision. The results produced a precision value of 0.84 or 84%, a recall value of 0.81 or 81%, and an F1 score of 0.82 or 82%. Overall, the results show that the object detection model has a high level of accuracy in detecting personal protective equipment and to improve detection accuracy, diversity is needed in the dataset in terms of object type, background, pose, distance, orientation, and lighting conditions.*

Keywords— *Computer Vision, CNN, Construction Site, Machine Learning, Object Detection, Safety.*

I. Introduction

Construction site safety not only protects workers, but also keeps the public safe. Construction sites are usually located in busy areas where the general public passes by. Inadequate safety protocols can cause objects to fall and hit the general public passing by the construction site, putting the general public at high risk. Some of the factors that cause accidents on construction sites are personal protective equipment not provided to workers, workers not using the provided personal protective equipment and lack of adequate safety training for workers [1].

The effectiveness of the safety management system in the construction industry is to mitigate the hazards present in the workplace, reducing the risk of injury and property damage. To determine the effectiveness of the safety management system, a questionnaire survey was compiled containing supervisor attitudes, PPE use, construction complexity, heavy equipment use, etc. Statistical Product and Service Solutions (SPSS) software, a statistical software, was used to analyze the data. The results showed that coordination and control by subcontractors, control of subcontractors' safety behavior and provision of PPE by the company were the three main factors reflecting safety on construction sites [2].

Construction accidents sometimes occur because workers do not use the safety equipment provided. The frequency of accidents can be reduced if entities can observe workers continuously to immediately identify when safety equipment is not being used properly. Proper use of personal protective equipment is critical to the safety of construction workers and can be a critical factor between accidents and safety. 50.9% of workers who wore personal protective equipment (PPE) removed it while working. 75.4% of workers mentioned falling off their PPE while working and 56.1% of workers mentioned that PPE makes tasks more difficult to perform [3].

Machine Learning, which was reorganized as a separate field, began to develop in the 1990s. The area shifted its focus from developing artificial intelligence to solving practical problems. Machine Learning shifted focus away from symbolic approaches inherited from AI, and towards methods and models borrowed from statistics and probability theory. The field also benefited from the increasing availability of digital information, and the possibility to distribute it over the internet [4].

Machine learning, part of the field of artificial intelligence (AI), is a technology that is widely researched and used for problem solving. It involves presenting reviews from various fields in the form of algorithms for problem solving. Machine learning is categorized into three main groups: supervised learning, unsupervised learning, and reinforcement learning [5].

Computer vision can be utilized in a very wide variety of fields ranging from raw data recording to image pattern extraction and information interpretation. Most tasks in computer vision are related to the process of obtaining information about events or descriptions, from input scenes (digital images) and feature extraction. The nature of the data being processed and the application domain determine the approaches taken to solve computer vision challenges. Pattern recognition and image processing are combined to create computer vision. The development of computer vision depends on computer technology systems, whether it is about image quality enhancement, image recognition or object detection [6].

Mochammad Daffa Daniarfa Darmawan, Wahyu Kusuma Raharja “Implementation of Machine Learning for Personal Protective Equipment Detection Using Convolutional Neural Network,” *International Research Journal of Advanced Engineering and Science*, Volume 9, Issue 1, pp. 90-94, 2024.

Convolutional Neural Network (CNN), also called ConvNet, is a type of Artificial Neural Network (ANN), which has a deep feed-forward architecture and has outstanding generalization ability compared to other networks with fully connected layers, it can learn highly abstract features of objects, especially spatial data, and can identify them more accurately. A deep CNN model consists of a set of processing layers that can learn various features of input data (e.g. images) with different levels of abstraction. The initiation layer learns and extracts high-level features (with lower abstraction), and deeper layers learn and extract low-level features (with higher abstraction) [7].

Pattern recognition and image processing are combined to create computer vision. The output of the CV process is image understanding. By modifying the capacity of human eyesight to retrieve information, this field is developed. Unlike computer graphics, computer vision is the study of information extraction from images. The development of computer vision depends on computer technology systems, whether it is about image quality improvement or image recognition. There is an overlap with Image Processing in basic techniques, and some authors use the two terms interchangeably [8].

Image recognition is a branch of computer vision, which is a broad research area including techniques such as 3D modeling and object tracking. The history of computer vision dates back to the early 70s and unlike previous work in the field of image processing, the goal is to enable a thorough understanding of the scene by retrieving a 3D construction of the world from the image. Image recognition focuses on interpreting images and identifying different variables, such as locations, people and animals. For humans, the task of interpreting what humans see in an image is easy because humans know that objects can appear in different shapes, angles, and light. The human eye has no difficulty distinguishing between cats and dogs, and understanding that different breeds of dogs are still dogs [9].

In [10] designed an image detection model system regarding worker safety conditions based on compliance with the use of PPE using the Faster Region-based Convolutional Neural Networks (R-CNN) algorithm. This research was conducted using Tensorflow involving 1,129 images from the MIT Places Database as a training dataset, and 333 dataset images from actual construction sites were used for evaluation. The experimental results showed 276 images were detected as safe, and the average accuracy rate was 70%. This research is based on image detection of three PPE components, involving hardhat, safety vest, and safety boots.

In [11] developed a framework for real-time detection of construction workers' safety compliance related to PPE, which aims to be integrated into the safety workflow of an organization. This research uses a Convolutional Neural Networks model. By detecting whether workers are wearing hardhat and safety jacket, the model predicts into four categories such as NOT SAFE, SAFE, NoHardHat, and NoJacket. A dataset of 2,509 images was collected from video footage of several construction sites and this web-based collection was used to train the model.

These studies are very important as construction sites are hazardous environments where the risk of accidents is high. To minimize this risk, safety measures must be implemented and monitored regularly. This is where machine learning and computer vision can make a significant impact.

II. RESEARCH METHOD

The method used in this research is applied research method. Applied research is research that has practical reasons, the desire to know, aims to be able to do something much better, effectively and efficiently [12]. Applied research methods are used to solve problems related to worker safety at construction sites by designing a system that can detect whether workers use personal protective equipment or not. The framework of the research can be seen in Figure 1.

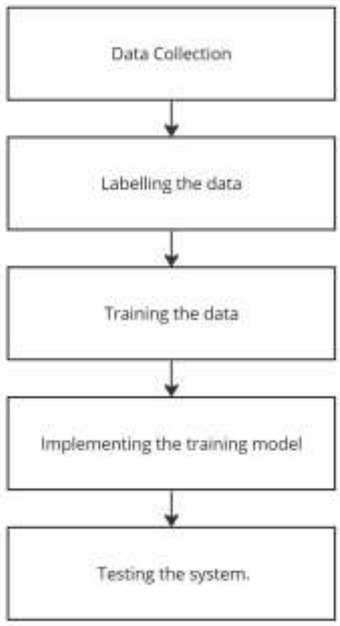

Fig. 1. Research Framework

Data collection is done by collecting images obtained from Google.com where each image will be labeled. Figure 2 is an example of some images that will be used as data for detection model training.

Fig. 2. Sample Image Used

Mochammad Daffa Daniarfa Darmawan, Wahyu Kusuma Raharja "Implementation of Machine Learning for Personal Protective Equipment Detection Using Convolutional Neural Network," *International Research Journal of Advanced Engineering and Science*, Volume 9, Issue 1, pp. 90-94, 2024.

International Research Journal of Advanced Engineering and Science
ISSN (Online): 2455-9024

The images obtained from Google.com are images related to personal protective equipment with the types of hardhat, safety-gloves, safety-goggles, safety-vest and mask. These types will be used as a reference in the personal protective equipment detection system. After collecting the images, the images will be labeled.

Creating meaningful labels is an important prerequisite for any machine learning-based computer vision application as it determines the quality of the model results [13]. In particular, prediction of multiple object classes in different environments requires thousands of high-quality labels for model training so that it has high accuracy [14]. The labeling process is needed in object detection so that the machine learning model can easily recognize objects, in this study labeling was carried out with 11 classes, namely hardhat, goggles, mask, safety vest, gloves, person, no-hardhat, no-goggles, no-mask, no-safety vest and no-gloves. The labeling process can be seen in Figure 3.

Fig. 3. Image Labeling with LabelImg.

After completing the labeling of each image, the next step is to divide the dataset into 3 folders, namely the test, train, and valid folders. The test set is used to measure the performance of the trained model. The division of folders into test set, training set, and validation set is an important strategy in machine learning to measure performance, perform model evaluation, and prevent overfitting. Overfitting is a fundamental problem in supervised machine learning that prevents perfectly generalizing the model to fit the observed data in the training data, as well as the unseen data in the test set. Due to the presence of noise, limited training set size, and classifier complexity, overfitting occurs [15]. The total dataset of the three folders is 1226 images and annotations of each image. Datasets that already have images and annotations will be trained using google colab. In training the detection model using 40 epochs with the YOLOv8 ultralytics library on google colab and using the T4 GPU as hardware that performs training for the detection model. The result of the training is a machine learning model that will be used in the system for personal protective equipment detection.

After the data training stage is complete, the model will be implemented on the system which aims to detect objects according to 11 predetermined classes, namely hardhat, goggles, mask, safety vest, gloves, person, no-hardhat, no-goggles, no-mask, no-safety vest and no-gloves. Then system testing is carried out to determine metrics such as training and validation loss, precision, F1 score, and recall. Overall metrics such as training and validation loss, precision, F1 score, and recall are important metrics in evaluating the performance of object detection models. These metrics help in understanding how well the model can detect and classify objects, as well as provide insight into how well the model avoids misclassification.

III. Results And Discussion

During the model training process using YOLOv8, the parameter used is the number of epochs. The model training process used 40 epochs and spent 17 minutes training. The benefit of epochs in training is to adjust or update the weights based on errors found during training and learn more complex patterns in the training data.

The number of epochs also affects the training loss and validation loss values of a model. The training loss and validation loss values provide important information about how learning performance changes with the number of epochs. The training loss and validation loss values can be seen in Table I.

TABLE I. Training and Validation Loss Values

Type	box_loss	cls_loss	dfl_loss
Training	0.692	0.707	1.091
Validation	1.177	0.810	1.401

The loss value of a model if the smaller the better, Figure 4 is a graph of the loss value of the model with 40 epochs.

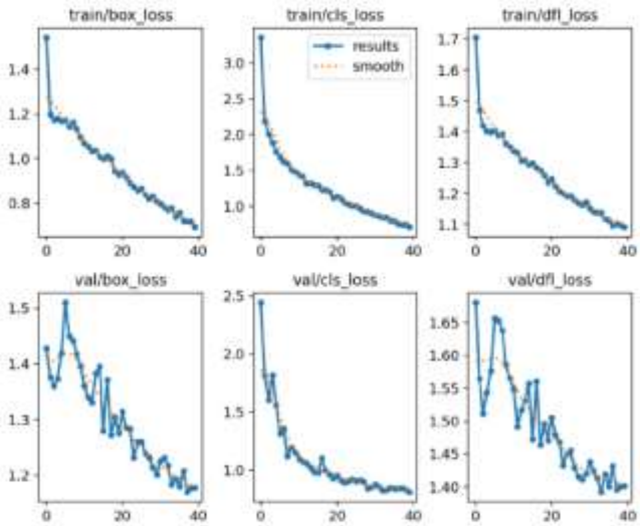

Fig. 4. Training and Validation Loss Graph of the Model.

In Figure 4 the x-axis is the number of epochs and the y-axis is the loss value. It can be seen based on Figure 4 that the increasing number of epochs, the smaller the loss value.

Model testing conducted on google colab provides the results of precision, recall and F1 score values. These metrics are important in evaluating the performance of object detection models and these metrics show how well a model performs in object detection. To get the value of these metrics, the true positive, false positive, and false negative values of a

International Research Journal of Advanced Engineering and Science
ISSN (Online): 2455-9024

model are needed. Table II shows the value of true positive, false positive, and false negative with 40 epochs.

TABLE III. True Positive, False Positive and False Negative Values of the Model

Type	Value
True Positive	256
False Positive	47
False Negative	59

Based on the true positive (TP), false positive (FP), and false negative (FN) values in Table II, the precision, recall, and F1 score values can be calculated.

Precision is the number of true positive predictions (correctly detected objects) of all positive predictions made by the model. The following is the formula for precision:

$$Precision = \frac{True\ Positive\ (TP)}{True\ Positive\ (TP) + False\ Positive\ (FP)}$$

Based on this formula, the value of precision is:

$$Precision = \frac{256}{256 + 47} = 0.84\ or\ 84\%$$

After calculating the precision value, the next step is to calculate the recall value. Recall is the proportion of true positive predictions from all true positive examples in a data set. Here is the formula for recall:

$$Recall = \frac{True\ Positive\ (TP)}{True\ Positive\ (TP) + False\ Negative\ (FN)}$$

Based on this formula, the value of recall is:

$$Recall = \frac{256}{256 + 59} = 0.81\ or\ 81\%$$

With the precision and recall values already known, we can calculate the next object detection model evaluation metric, namely F1 score.

F1 score is a commonly used evaluation metric that combines precision and recall into a single value. It provides a balanced assessment of the model's performance by considering the ability to correctly identify positive examples (precision) and the ability to find all positive examples (recall). F1 score ranges between 0 and 1, where higher values indicate better performance. The following is the formula for F1 score:

$$F1\ Score = 2\frac{Precision * Recall}{Precision + Recall}$$

Based on this formula, the value of the F1 score is:

$$F1\ Score = 2\frac{0.84 * 0.81}{0.84 + 0.81} = 0.82\ or\ 82\%$$

After evaluating the model by calculating metrics related to the performance in detecting objects. Then a test was conducted to detect an image. Figure 5 is the result of detection by the system.

Based on Figure 5, the system can detect images and classify objects accordingly and there is a bounding-box on each object where each bounding-box has the name of the object and the confidence value.

Fig. 5. System Detection Results

IV. Conclusion

The results of research on object detection modeling in personal protective equipment using YoloV8 ultralytics and 40 epochs as its parameter, provide a detection model with a precision level of 0.84 or 84%, recall of 0.81 or 81% and F1 score of 0.82 or 82%. This research was conducted using a 2.90 GHz AMD Ryzen 7 4800H processor laptop and 16 GB RAM with Windows 11 operating system.

It can be concluded from the results of the study that by training on 1226 images and annotations on google colab using YoloV8 ultralytics, it can produce an object detection model that has good performance in classifying and detecting objects and the model can be implemented in a system to detect whether someone is wearing personal protective equipment or not.

In the future, in order for an object detection model to have better performance in classifying and detecting objects, it is necessary to have diversity in the training dataset in terms of object type, background, pose, distance, orientation, and lighting conditions so that the system can be used and work well in various conditions.

References

[1] Toole, T. M. (2002). Construction Site Safety Roles. Journal of Construction Engineering And Management, 128, 203-210.
[2] Ajith, S., Sivapragasam, C., Arumugaprabu, V. (2020). SAFETY AND HAZARDS MANAGEMENT IN CONSTRUCTION SITES – A REVIEW. Journal of Technology, 35(03), 175-179.
[3] Sehsah, R., El-Gilany, A. H., Ibrahim, A. M. (2020). Personal protective equipment (PPE) use and its relation to accidents among construction workers. La Medicina del lavoro, 111(4), 285-295.
[4] Langley, P. (2011). The changing science of machine learning. Machine Learning, 82, 275-279.
[5] Roihan, A., Sunarya, P. A., Rafika, A. S. (2020). Pemanfaatan Machine Learning dalam Berbagai Bidang: Review paper. Indonesian Journal on Computer and Information Technology, 5(1), 75-82.
[6] Wiley, V., Lucas, T. (2018). Computer Vision and Image Processing: A Paper Review. International Journal Of Artificial Intelegence Research, 2(1), 28-36.

Mochammad Daffa Daniarfa Darmawan, Wahyu Kusuma Raharja “Implementation of Machine Learning for Personal Protective Equipment Detection Using Convolutional Neural Network,” *International Research Journal of Advanced Engineering and Science*, Volume 9, Issue 1, pp. 90-94, 2024.

[7] Ghosh, A., Sufian, A., Sultana, F., Chakrabarti, A. (2020). Fundamental Concepts of Convolutional Neural Network. Recent Trends and Advances in Artificial Intelligence and Internet of Things,172, 519-567.
[8] Babatunde, O. H., Armstrong, L. J., Leng, J., Diepeveen, D. (2015). A survey of computer-based vision systems for automatic identification of plant species. Journal of Agricultural Informatics. 6(1). 61-71.
[9] Cromsjö, M., Hallonqvist, L. (2021). Detection of safety equipment in the manufacturing industry using image recognition. Thesis. Linköping University.
[10] Saudi, M. M., Hakim, A., Ahmad, A., Saudi, A. S. M. (2020). Image Detection Model for Construction Worker Safety Conditions using Faster R-CNN. International Journal of Advanced Computer Science and Applications, 11(6).
[11] Delhi, V. S. K., Sankarlal, R., Thomas, A. (2020). Detection of Personal Protective Equipment (PPE) Compliance on Construction Site Using Computer Vision Based Deep Learning Techniques. Frontiers in Built Environment, 6.
[12] Irina, F. (2017). Metode penelitian terapan. Yogyakarta: Parama Ilmu.
[13] Sager, C., Janiesch, C., Zschech, P. (2021). A survey of image labelling for computer vision applications. Journal of Business Analytics, 4(2), 91-110.
[14] Szeliski, R. (2011). Computer vision: Algorithms and applications. Springer.
[15] Ying, X. (2019). An Overview of Overfitting and its Solutions. Journal of Physics Conference Series. 1168(2).