JURNAL ILMIAH TEKNIK INDUSTRI
ISSN: 1412-6869 (Print), ISSN: 2460-4038 (Online)
Journal homepage: http://journals.ums.ac.id/index.php/jiti/index
doi: 10.23917/jiti.v22i1.21826

Convolutional Neural Network for Identification of Personal Protective Equipment Usage Compliance in Manufacturing Laboratory

Khania O.P.P. Nugraha[1a], Achmad Pratama Rifai[1b]◆

Abstract. *Data from the Badan Penyelenggara Jaminan Sosial (BPJS) Ketenagakerjaan Indonesia from 2019 to 2021 shows that the number of work accident victims who claimed Work Accident Insurance (Jaminan Kecelakaan Kerja / JKK) continues to increase. The high number of work accidents is mostly caused by unsafe behavior at work sites, one of which is in terms of compliance with the use of Personal Protective Equipment (PPE). One tool that is considered important as a step in reducing work accidents is an identification system for compliance of personal protective safety equipment use that can detect PPE used by visitors or workers. This study develops an automatic identification system that is built using the Convolutional Neural Network (CNN) to identify the use of PPE in the manufacturing technology laboratory. The CNN models used are the 4th and 5th versions of You Only Look Once (YOLO) which are then compared based on two methods: train from scratch and transfer learning. The dataset used for building the detection system has 11,579 images consisting of six classes of PPE objects. Overall performance of the proposed models shows very good results. Moreover, the comparison result among the three models shows that YOLOv5 transfer learning has the best performance with the best precision (94.2 %), recall (91.8 %), and mAP (88.6%).*

Keywords: *personal protective equipment; Convolutional Neural Network; deep learning; object detection; YOLO.*

I. INTRODUCTION

Every work done by humans has work accident potentials. In Indonesia, the number of work accidents increases every year, by 21,27% in 2020 (BPJS Ketenagakerjaan, 2021) and 5,69% in 2021 (BPJS Ketenagakerjaan, 2022). Even during pandemic, the numbers of work accident victims who claimed JKK keep increasing whereas there is reduction in workforce and work site capacity.

Work accidents are classified into 2 types based on the causal factors, which are work environment factors and human factor itself. Heinrich & Granniss (1980) concluded that 88% of all accidents that occurs in the work environment are caused by unsafe behavior, 10% are due to unsafe conditions, and 2% are the work accidents that cannot be avoided. If unsafe behavior in the work environment can be reduced or prevented, safety performance will naturally improve. Therefore, hazard control is an important factor to ensure workers' safety and occupational health, the use of personal protective equipment as one of the essential elements in this case. Several studies also reported that there have been thousands of fatalities in workers without using personal protective equipment (PPE) (Wu & Zhao, 2018). Based on the data from BPJS Ketenagakerjaan 2022, 65,89% of work accidents occurred in work site and 63% of them occurred, specifically in the manufacturing and construction industries.

This study observes a case in the manufacturing laboratory as one of manufacturing industries with low monitoring of PPE usage compliance. Identification models based on Convolutional Neural Network with YOLO algorithms are proposed to provide automatic monitoring of PPE usage.

Basically, the techniques used to monitor compliance with the use of PPE can be classified into 2 types, namely sensor-based and vision-based (Nath et al.; 2020). Sensor-based systems use sensors mounted on certain equipment (either by installation or by default) that transmit signals, for example by using Radio Frequency

[1] Department of Mechanical and Industrial Engineering, Universitas Gadjah Mada, Jalan Grafika No. 2, Yogyakarta, Indonesia 55281.

[a] email: khania.o.p@mail.ugm.ac.id
[b] email: achmad.p.rifai@ugm.ac.id
◆ corresponding author

Submited: 24-02-2023 Revised: 19-05-2023
Accepted: 08-06-2023

Identification (RFID) technology. Then, the signal obtained from the device is analyzed as an ingredient in making decisions regarding compliance and how to use PPE for workers. Thus, a significant investment is required for the installation and maintenance of appropriate devices.

In contrast, a vision-based system uses a camera to obtain input data in the form of images or videos taken directly at the work site to then be analyzed for compliance with the use of PPE in the environment. In addition to its very rapid development, vision-based monitoring techniques are considered easier to use and control. The data is then analyzed for the identification of PPE. The system also provides more information on image and video input data that can be used to understand complex job sites more comprehensively.

Most of the previous research has analyzed and reviewed several types of deep learning algorithms used in deep learning system development to identify and even verify how to use APD. The selection of the algorithm is based on the method or technique in the system built to identify the existence of objects.

Over the past few years, CNN has become the dominant algorithm in monitoring techniques using computer vision for pattern and image recognition needs. The development of CNN science and applications is still being researched in order to advance related science and knowledge. Two basic parameters used in CNN for this case are accuracy and detection speed (Önal & Dandıl, 2021). Some examples of applications in previous studies are Krizhevsky et al. (2017) who adopted Deep CNN to classify 1.2 million high-resolution images (ImageNet dataset) into 1,000 different classes of various objects, Ding et al. (2018) which combines the Long Short-Term Memory (LSTM) model with CNN to detect potentially unsafe worker behavior, and Lecun et al. (1998) who used CNN to recognize numbers and letters in handwriting.

YOLO is an object detection system based on CNN which has become very popular in direct object detection because it is carried out in only 1 stage of the process (Redmon et al.; 2016). A research by Wang et al. (2021) highlights the performance of the YOLO algorithm which is quite good in the direct object detection process. This research was conducted to build a CNN model by adopting the YOLOv3, YOLOv4, and YOLOv5 algorithms. Wang et al. (2021) has a wider research object where the designed model is tested for identification of 6 classes, namely hard hats with 4 colors, humans (construction workers), and vests.

Similarly, the results of research by Delhi et al. (2020) who reported that the YOLO algorithm, specifically the YOLOv3 version (the third version), has good performance in terms of precision and recall when detecting objects. This study implements a model using YOLOv3 with a hard library that will be applied to the identification of hard hats and jackets. The YOLOv3 algorithm was chosen because it has a faster predictive ability and has been shown to be effective in detecting many workers at construction sites, as stated by Luo et al. (2019) in his research (Delhi et al.; 2020). As a result, the developed CNN model performed well with an average precision level of 98%.

Nath et al. (2020) in his research shows similar support, which resulted in an increase in the performance of the system model that had been tested using 3 different approaches through the YOLOv3 algorithm. In the first approach, all classes, namely hard hats and vests, were identified one by one and then verified whether the use of PPE was appropriate. In the second approach, all classes were identified and verified simultaneously to then be directly categorized into 4 conditions: W (not wearing a hat or vest), WH (wearing a hat only), WV (wearing a vest only), or WHV (wearing a hat and vest). vest). In the third approach, the model first detects all workers in the input image and then classifies the classes using CNN-based classifiers. As a result, the second approach provides the most accurate PPE detection results (with a mAP value of 72.3%).

The study conducted by Wang et al. (2021), Delhi et al. (2020), and Nath et al. (2020) is relevant to the research question because it addresses the same context: identification of compliance with the use of PPE at construction sites. The fundamental difference from each

previous research can be seen from the algorithm used and the number of classes detected by the model built. Meanwhile, the CNN model developed by Li et al. (2020) in his paper only detects one class, namely hard hats, using the Single Shot Detection (SSD)-MobileNet algorithm.

As a result, the proposed model has a detection accuracy rate of 95% and a recall of 77%. However, there are flaws and errors in the model, where the hard hat with incomplete shape and small size in the image is difficult to identify which causes the SSD-MobileNet algorithm to be judged not good enough for identification of compliance with the use of PPE at construction sites.

The research of Wu & Zhao (2018) is related to the research of Li et al. (2020) where both focused on developing hard hat detection systems. Li et al. (2020) in their study proposed a hard hat detection system at a construction site by implementing SSD-300. Similarly, Wu & Zhao (2018) introduced a hard hat detection system based on the Deep CNN algorithm with the main focus on workers in the industrial sector, but model training was carried out on datasets taken in pedestrian zones to detect the presence or absence of humans in the image segment. In the study of Wu & Zhao (2018), it was found that the proposed deep CNN-based model showed outstanding performance in pedestrian detection and hard hat identification.

In contrast to various studies Wang et al. (2021), Delhi et al. (2020), Nath et al. (2020), Li et al. (2020), and Wu & Zhao (2018), research conducted by Fang et al. (2018) aims to detect non-hardhat use (NHU) non-compliance in remote surveillance videos at construction sites. This journal builds a model with Faster R-CNN algorithm which also has high precision and speed. The aim is to verify field conditions at construction sites to help improve supervision of construction workers by ensuring proper use of hard hats. The experimental results in this study show that Faster R-CNN is quite good for various backgrounds on images and changes in worker posture in NHU detection. Both precision and recall, the value is above 90% which is enough to improve construction site security supervision.

Similar to the research described above, the research conducted by Önal & Dandıl (2021) in the work environment in the form of industrial production facilities (factories) also focuses on developing PPE detection models, but with the YOLOv4 algorithm. The aim of the study was to build a model to control the use of PPE (hard hats, vests, masks, gloves, and protective eyewear) from video datasets and to detect unsafe movements in factories. Based on the research experiment, the model's mAP score became more stable after the 10,000th iteration. Then, more than 88% accuracy rate was achieved in the recognition of all objects except the glove object in the work environment.

Chen & Demachi (2020) with a combination of OpenPose and YOLOv3, introduced a model in the form of an automated system with a vision-based approach to detect compliance with the use of PPE in Nuclear Power Plants (NPP). This research is based on the activity of deactivating the Fukushima Daichii nuclear power plant in Japan which is quite challenging and requires a high level of compliance with the use of PPE because it has the potential to be exposed to various risks. The experimental results show that the approach developed by Chen & Demachi (2020) is able to identify workers who do not wear PPE properly with a high level of precision (97.64%) and a recall rate of 93.11%.

Previous studies were more limited to detecting dangerous signs related to safety equipment such as helmets, protective clothing, hard hats, and other PPE. Therefore, Hung & Su (2021) through their research propose a method to classify 3 categories of unsafe behavior with vision-based deep learning technology. The goal is to be able to recognize the hazardous behavior of workers and accelerate risk analysis and assessment with high accuracy. The dataset is divided into 3 dangerous behaviors: (a) humans reach for something, (b) human foot positions out, and humans climb with wrong movements (Hung & Su, 2021). The model introduced by Hung & Su (2021) was built with the CNN classifiers: VGG19, Inception_V3, and InceptionResnet_V2. The experimental results show that InceptionResnet_V2 has a better

performance than VGG19 and Inception_V3 in classifying workers' hazardous behavior. The accuracy rate of InceptionResnet_V2 reaches 92.44%, slightly higher than VGG19 with a value of 91.16%, while the accuracy rate of Inception_V3 is 47.06%.

Taken together, the above studies not only demonstrate innovative deep learning applications in the design of automated object detection systems, but also encourage an interdisciplinary approach so that the system can be applied more broadly in many other types of locations. However, from the various studies that have been carried out, there is no research that discusses the application of CNN-based deep learning to identify compliance with the use of PPE, specifically in manufacturing technology laboratories.

Viewed from the supervision system for the use of PPE in certain work locations, such as industrial plants, construction sites, and nuclear power plants, the supervision tends to be strict because they have clear SOP and a structured activity schedule. This is different from the low supervision of the PPE uses in the manufacturing laboratory because there is no written SOP regarding the use of PPE, especially with the different activity schedules of each visitor so that identification of compliance with the use of PPE is not carried out. The identification process can be done through a detection system model that uses deep learning technology with CNN.

Therefore, this paper aims to build a CNN model with the YOLOv4 and YOLOv5 algorithms and comparing the performance of those two in detecting PPE of the manufacturing laboratory visitors. The object of the PPE class studied were hard hats, safety coats, shoes, masks, protective glasses, and gloves.

II. Research Method

The main objective is to propose a fast and reliable automatic identification system using CNN for PPE detection. We built three different CNN models to be compared and obtain the best model. There are two different model development methods and algorithms, which are train from scratch and transfer learning for the methods with YOLOv4 and YOLOv5 for the algorithms.

The proposed methods to build the PPE identification model are classified into six stages: data collection, data pre-processing, dataset making, CNN implementation, experiment and result analysis, and evaluation.

The data used for the experiment done in this research are collected by the authors. Since there are not so many object detection applications of PPE in manufacturing laboratory using deep learning and there were no datasets of PPE in manufacturing laboratory available, thus we collected real images of manufacturing laboratory visitors and created a novel dataset.

The dataset contains 1,235 images of 15 people with varied gestures, angles (left, right, front), distance (far, close), light exposure, and background. The collected images are at 1980 x 3520 pixel. There are 6 classes according to the types of PPE which are the focus of this paper.

The data pre-processing stage consists of image selection, data annotation and data augmentation. Image selection aims to detect and remove duplicates images on the dataset. The images were then pre-processed in such features: flip, blur, brightness, and noise, to increase the PPE detection performance using CNN. As a result, the number of collected images increased to 11,206 images as a whole dataset.

The dataset making stage is the final step of dataset preparation. The whole dataset contains 11,579 images which are separated into three set: training set, validation set, and testing set. The ratio used in the separation process is 7 : 1 : 2 as there is still no rule set yet related the ratio.

The state-of-the-art architecture used in this paper is CNN-based YOLOv4 and YOLOv5. All versions of YOLO model consist of three main parts, as of backbone, neck, and head.

III. Result and Discussion

Experiments

The research was carried out at the Mechanical Technology Laboratory of DTMI UGM, Faculty of Engineering, Gadjah Mada University,

precisely in Selowo, Sinduadi, Mlati, Sleman, Yogyakarta Special Region 55284. The mechanical laboratory is separated into 3 different rooms with their respective functions. There are different items and different machines in each room that affects the type of background in the image data in the form of photos taken.

The research in this thesis was conducted on 6 classes, namely hard hats, safety coats, shoes, masks, protective glasses, and gloves. All of these classes are PPE available at the TM DTMI UGM Laboratory, except for masks and shoes. Specifically for protective eyewear and gloves, each of them has 2 different types that are the object of research. The type of data used in this study is primary data, namely data taken directly. The primary data source is photos of 15 people using a random combination of PPE. The datasets are divided into 3 types while building the CNN model, namely training datasets (70%), validation datasets (15%), and testing datasets (15%). Training dataset is data used for the learning process or training by the model. Validation dataset is data used to validate the model and prevent overfitting. Testing dataset is data used for model testing as a simulation of the use of the model in real life. Testing datasets must be different from training and validation.

The raw dataset consists of 1,235 images taken at various distances, viewing angles, lighting, and backgrounds. There are 6 classes according to the type of PPE that is the focus of this study. The 6 classes are helmets, safety coats, protective goggles, masks, gloves, and shoes.

The next stage is data annotation. Data annotation is the determination of the class for each image in the dataset by assigning a class label to each PPE object detected in the image. Data annotation was carried out using the online software "Roboflow". The label results from the annotated data are exported in .txt format with label descriptions: 0 for "Helm" (helmets), 1 for "Jas Pengaman" (safety coats), 2 for "Kacamata Pelindung" (protective glasses), 3 for "Masker" (masks), 4 for "Sarung Tangan" (gloves), and 5 for shoes. From the whole dataset, there are 5,461 annotated objects. Gloves are the class with the greatest number of labels, while shoes are the fewest. Figure 1 presents an example of data annotation in an image.

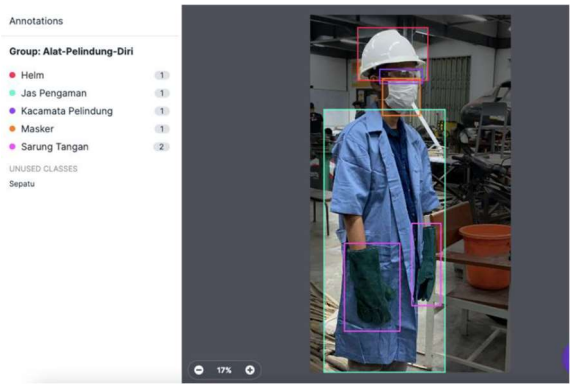

Figure 1. Data annotation

The next stage is data augmentation, which is a technique to increase the amount of data to increase image variations. The augmentation process uses 4 types of filters, such as flip (inverting the image vertically or horizontally), noise (adding grain to the image), blur (giving an unfocused effect on the image), and brightness (increasing or decreasing the brightness level of the image), as presented in Figure 2. As a result, the training dataset became 11,206 images and the overall dataset became 11,579 images.

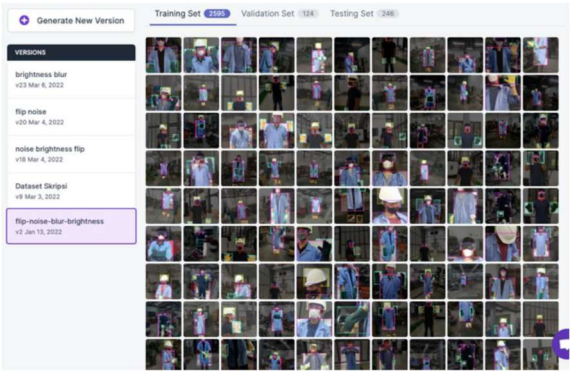

Figure 2. Data augmentation with flip-noise-blur-brightness

The object detection system testing in this study was conducted to find out that the developed system can detect objects with a good level of precision and recall (closer to a value of 1 or 100%). After successfully testing the model, an analysis of the model's performance is carried out based on the methods and algorithms used. The focus of the analysis of the testing results is to get a model with optimal algorithms and weights.

In training and testing the object detection system for the YOLOv4 algorithm, we use the transfer learning method with pretrained weights. In the YOLOv5 algorithm, two methods are used: transfer learning and train from scratch.

In the training process, the best weights are determined by comparing the performance of the training model in various iterations or epochs. The best iterations and epochs are obtained based on iterations or epochs that have superior levels of precision, recall, F1 score, and mAP compared to output metrics in other iterations or epochs. If the highest level of precision and recall is obtained by a model with different iterations or epochs (for example, epoch 10 has the highest precision but epoch 20 has the highest recall), then the best iteration or epoch is only determined based on the F1 score and mAP value. This is because the F1 score is a representation of the value of precision and recall because the value is obtained from a combination of the two metrics. Meanwhile, the mAP value represents the level of model accuracy because the value is obtained from the average AP results for each class.

YOLOv4 results

The training of YOLOv4 aims to obtain low precision, recall, mAP, and average loss values from each training weights result. With the YOLOv4 algorithm, the training process will save the set of weights automatically for every multiple of 1000 iterations and the best iteration. The YOLOv4 model is trained according to the max_batches value to determine the maximum number of iterations in training. The max_batches value is determined by calculating the number of classes multiplied by 2,000. In this study, the dataset used has 6 classes, so the max_batches value is 12,000. Therefore, the maximum number of iterations in this training model is 12000 iterations and resulting in 13 sets of weights from the training process.

The weights model of YOLOv4 with the highest level of precision is at iteration 8600, while the weights model with the highest recall rate is iteration 2000. According to Ghoneim (2019), the model with the highest f1 score can be the best choice because it represents 2 values, namely precision and recall. Figure 3 presents the training graph of YOLOv4. When viewed based on the highest f1 score, the best weights are obtained at iteration 8600. The difference in mAP values for iteration 2000 and iteration 8600 is also not very significant. So, the weights in the iteration 8600 will be used as weights to detect objects in the testing dataset.

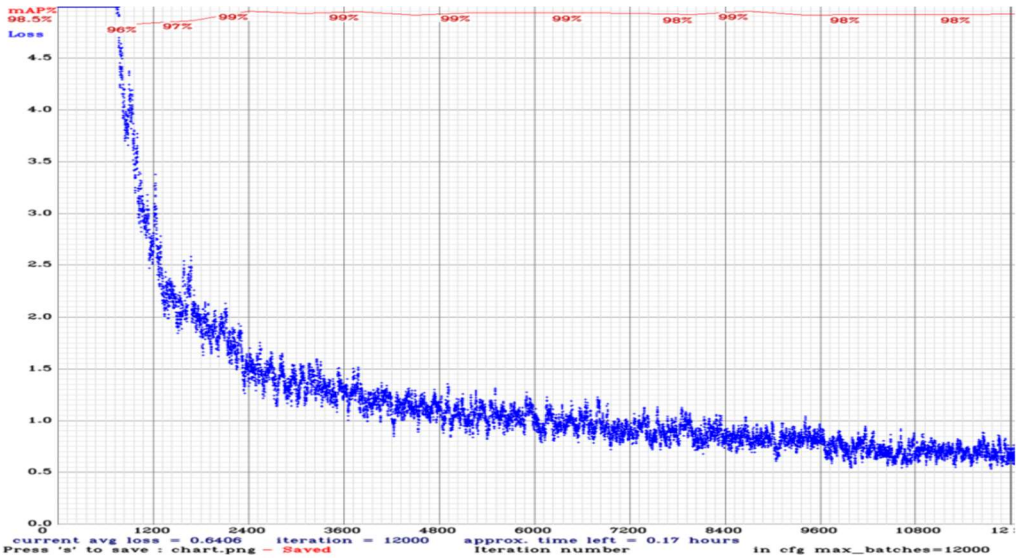

Figure 3. Training progress of YOLOv4

Based on the graph of the training results, the mAP value continues to increase until it reaches a value of 99% in the iteration 8600. After passing through the iteration 8600, there is no further increase in the mAP value so that the mAP value in that iteration is the maximum value during the training process. The mAP value of 99% can be said to be very good because the value is close to 100%. This shows that the level of accuracy of the model in detecting objects is quite high.

On the other hand, the average loss value in the graph of the training results decreases drastically until iterations of 5000. After passing through the 5000 iterations, the average loss value continues to decrease but not significantly. However, the decrease in the average loss value continues until the maximum iteration, which is 12000 iterations. The smaller average loss value is in line with the increasing number of training processes carried out. This shows that the longer the training causes the number of errors or errors to decrease so that the performance of the model continues to be better. The average loss value achieved during the maximum iteration is 0.640621.

Table 1. Testing result of YOLOv4

Detection Count	TP	FP	Average Precision (%)	mAP (%)
Helmet (0)	179	4	100	84.40
Safety suit (1)	194	1	100	
Safety Googles (2)	122	37	91.82	
Masker (3)	239	6	99.86	
Gloves (4)	343	7	95.77	
Shoes (5)	13	12	18.92	

TP	FP	FN	Precision (%)	Recall (%)	F1 Score (%)	mAP (%)
1090	67	100	94	91.60	92.88	84.4

The best weights obtained from the training process are used to detect objects in the testing dataset. In the detection process, a confidence threshold of 0.3 and an IoU threshold of 0.5 is applied. Table 1 shows the result of testing using trained YOLOv4 model. As a result, the AP value for each object class have AP scores above 90%, except Shoes class. The classes with the highest AP scores are safety suits and helmets, with 100% AP. Safety coats and helmets have the highest AP values because both objects have the same type and shape in all images in the dataset. In addition, safety suits and helmets are considered easier to identify when viewed from various sides. On the other hand, shoes have the lowest AP value because the types of shoes used by participants are quite varied with different shapes and colors, causing the training process to be less than optimal and the model to be less precise in detecting shoe objects. Most of the shoes used by the participants had dark colors so that they almost resembled the basic color or the floor which might cause the object of the shoes to be seen less clearly. In addition, the number of shoe pictures in the training dataset is the least compared to other classes so that the training process is not optimal. When collecting the dataset, a list of the combinations of PPE that must be used by the participants was not specified. The goal is to make the dataset more varied in terms of the number of objects of each type of PPE in the dataset. Thus, at the beginning of the study, it was not known the number of PPE objects in the dataset. Therefore, for further research it may be possible to use a dataset with a balanced number of PPE.

The overall performance of the PPE object detection process is acceptable with a fairly high level of precision and recall, namely 94% and 91.6%, respectively. Therefore, a high f1 score was obtained at 92.88%. The high level of precision means that this model can detect many objects exactly according to the ground truth bounding box, which is 94% objects (1090 TP objects) of all detected bounding boxes (1157 objects). In contrast, the model detects 6% of objects in the test dataset that don't actually exist (FP). The model has a high recall rate, meaning that this model can detect many objects correctly according to its ground truth bounding box, amounting to 91.6% (1090 TP objects) compared to all objects that have a ground truth bounding box (1190 objects). This means that the model failed to detect 8.4% of objects in the testing dataset. With this, the precision and recall values are considered good enough because they are close to 100% or 1.

In terms of detection speed, the YOLOv4 model successfully detects every 1 image in the testing dataset within 101.6 ms. The model successfully detects 1,631 objects before applying a confidence threshold of 0.3 and an IoU threshold of 0.5. The total number of objects that have been detected is 1,157 objects after the two thresholds are applied. Based on this, the detection speed of the YOLOv4 model is quite good because the detection speed is quite high.

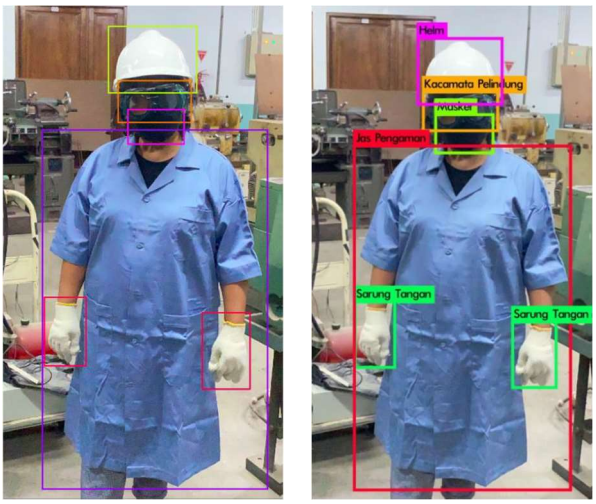

Figure 4. An example of actual (left) and predicted objects using YOLOv4 (right)

Based on the detection performance of the YOLOv4 model based on various metrics, the YOLOv4 algorithm can be said to be optimal in detecting testing datasets, meaning that this system is robust enough to detect various objects. Figure 4 illustrates an example of object detection using trained YOLOv4 model.

YOLOv5 results

The next experiment is done using the YOLOv5 model in two different methods, training from scratch and transfer learning. In the training process, both types of models are trained up to epoch 40. Figure 5 presents the training progress of YOLOv4 using training from scratch.

Based on the results of the training train from scratch, the best training results were obtained at epoch 25. Therefore, the weights in epoch 25 will be used as weights to detect objects in the testing dataset. In addition, it was found that the box loss and objectness loss values in the training and validation datasets decreased drastically until epoch 10. After passing through epoch 10, the loss value continued to decrease although not significantly. In contrast to the classification loss, the value decreased drastically only up to epoch 5 and continued to decline even though it was not significant. The loss value that continues to decrease indicates that the number of errors is decreasing so that the performance of the YOLOv5 train from scratch model is getting better.

The box loss, objectness loss, and classification loss values for the training dataset during epoch 25 are 0.0211, 0.017, and 0.0037,

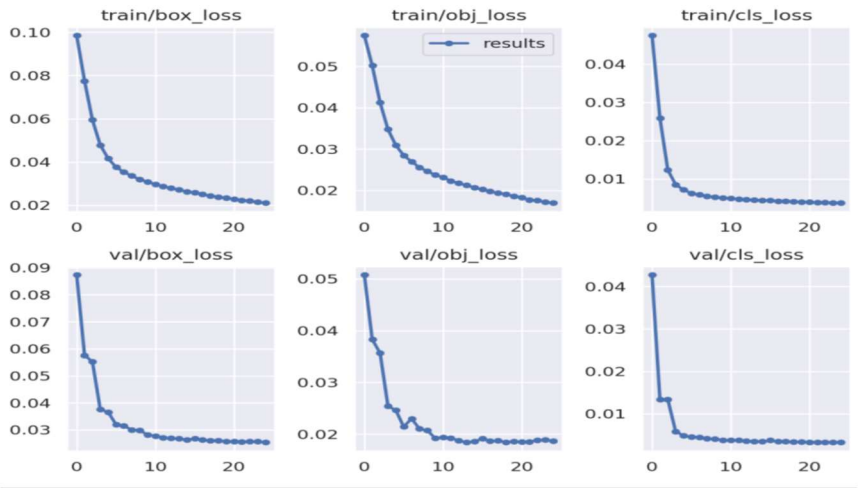

Figure 5. Training progress of YOLOv5 using from scratch approach

respectively, so the average train loss value is 0.0139. Then, the box loss, objectness loss, and classification loss values for the validation dataset are 0.0254, 0.0187, 0.00326, respectively, so the average validation loss is 0.0158. The following is a comparison of each loss in the training and validation datasets.

From the figure, it can be seen that the loss values for the training and validation datasets are not much different, such as the average loss values for each dataset which are also not much different. The comparison of loss values that are not much different indicates that the model fits the training and validation dataset, and the developed model is neither underfitting nor overfitting.

In this model, the mAP value continues to increase until it reaches a value of 95.9% in epoch 5. After passing through epoch 5, the value still continues to increase but not significantly. The mAP value obtained at epoch 25 is 97.3%. Furthermore, the precision and recall levels were 98.8% and 96.2%, respectively, indicating the robustness of the proposed model in detecting APD, supported by a high f1 score of 97.5%.

For the transfer learning model, the weights model with the highest precision was obtained at epoch 40, as shown in Figure 6. However, the weights model with the highest recall was at epoch 10. However, the model with epoch 20 had the highest f1 score and mAP values. Therefore, the transfer learning model with the best training results is epoch 20. Therefore, the weights at epoch 20 will be used as weights to detect

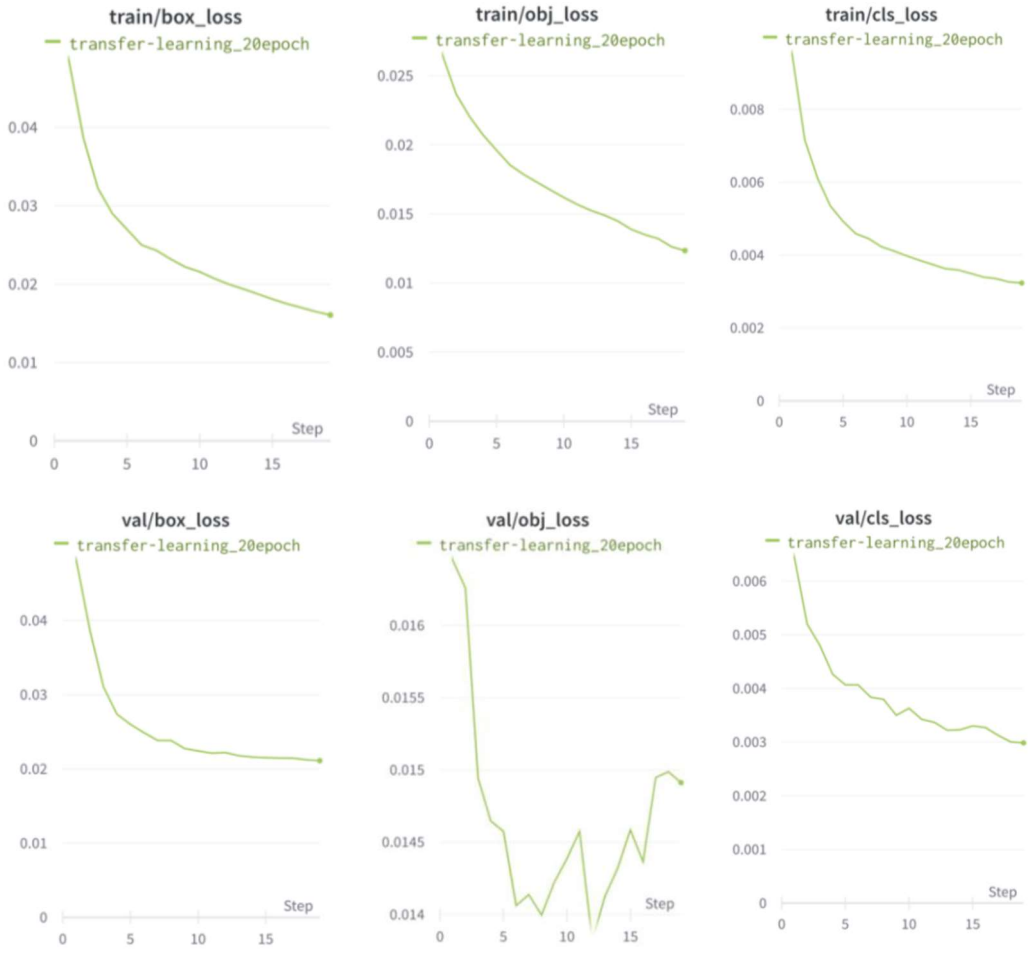

Figure 6. Training progress of YOLOv5 using transfer learning approach

objects in the testing dataset.

Based on the results of transfer learning training with 20 epochs, the value of box loss and classification loss in the training and validation datasets decreased drastically to epoch 10. After passing through epoch 10, the loss value continued to decrease although not significantly. Then, for classification loss, the value decreased drastically until epoch 5 and continued to decline even though it was not significant. This means that the number of errors continues to decrease during the transfer learning model training and the model performance continues to increase. The loss graph in this training model looks not very stable, but the training process is stopped or early stopping because the loss value obtained has exceeded the standard loss value limit and is considered very good because the loss value obtained is very small (close to 0). In addition, the performance of the model based on the other four metrics is considered very good and reliable to proceed to the testing process.

The box loss, objectness loss, and classification loss values for the training dataset during epoch 20 are 0.0161, 0.124, and 0.00324, respectively, so the average train loss value is 0.028. The box loss, objectness loss, and classification loss values for the validation dataset are 0.0211, 0.015, 0.00299, respectively, so that the average validation loss is 0.013. The following is a comparison chart for the loss.

The graph shows that the loss values for the training and validation datasets are not much different from the average loss values for each dataset. This result is the same as the comparison of loss values in the train from scratch model, which means that the transfer learning model is also suitable for application to the dataset and is neither underfitting nor overfitting.

The mAP model value increased to 97.6% in epoch 5. After passing through epoch 5, the value still continued to increase insignificantly. The mAP value obtained at epoch 20 is 98.2%. The overall model performance is acceptable with a very good mAP value and is supported by a fairly high level of precision and recall, which are 98.7% and 97.85%, respectively. This shows that the model is quite robust in detecting PPE objects.

Just like YOLOv4, the best weights of both methods (train from scratch and transfer learning) are used to detect objects. The detection process uses a confidence threshold with a value of 0.4 and an IoU threshold with a value of 0.3.

Table 2. Testing result of YOLOv5 using from scratch approach

Detection Count	TP	FP	Average Precision (%)	mAP (%)
Helmet (0)	174	3	99.1	79.0
Safety suit (1)	191	0	99.5	
Safety Googles (2)	106	44	78.9	
Masker (3)	220	8	94.6	
Gloves (4)	332	11	96.0	
Shoes (5)	3	23	5.80	

TP	FP	FN	Precision (%)	Recall (%)	F1 Score (%)	mAP (%)
1026	89	152	92.02	87.1	89.5	79.0

Table 3. Testing result of YOLOv4 using transfer learning approach

Detection Count	TP	FP	Average Precision (%)	mAP (%)
Helmet (0)	179	1	99.5	88.6
Safety suit (1)	192	2	99.5	
Safety Googles (2)	114	46	87.1	
Masker (3)	236	4	99.3	
Gloves (4)	343	8	97	
Shoes (5)	21	6	49	

TP	FP	FN	Precision (%)	Recall (%)	F1 Score (%)	mAP (%)
1090	67	97	94.2	91.80	93.0	88.6

From the results of testing with the train from scratch method, the AP values for each class were obtained, listed in Table 2 and 3. Similar to YOLOv4, the class with the highest AP value is a safety suit, while shoes are the class with the lowest AP value. In the transfer learning method, there are 2 classes that have the highest AP scores: helmets and safety suits, with both AP values of 99.5%. Meanwhile, shoes remained the class with the lowest AP score (49%).

The performance of the two YOLOv5 models is acceptable because the model is able to detect objects with a fairly high level of precision and recall. In YOLOv5 train from scratch, 92% and 87.1% of precision and recall were obtained, respectively. In terms of precision, this model can detect 92% of objects (1026 TP objects) objects in

the testing dataset correctly according to the ground truth bounding box and there are 8% objects (89 FP objects) in the testing dataset that do not actually exist. In terms of high recall, the model can detect 87.1% of objects (1026 TP objects) in the testing dataset correctly according to its ground truth bounding box compared to all objects that have a ground truth bounding box (1178 objects). This means that the model failed to detect 12.9% of objects.

In terms of detection speed, the YOLOv5 train from scratch model managed to detect every 1 image in the testing dataset within 24.7 ms with a total of 1,152 objects detected. This figure was obtained after the application of a confidence threshold of 0.4 and an IoU threshold of 0.3. Based on this, the detection speed of the YOLOv5 train from scratch model is good because the detection speed is high.

In the transfer learning method, the precision and recall levels are 94.2% and 91.8%, respectively. This means that the model can detect 94.2% of objects (1085 TP objects) in the testing dataset correctly according to the ground truth bounding box and there are 5.8% of objects (67 FP objects) in the testing dataset that do not actually exist. The precision level of 91.8% means that the model can detect 91.8% of objects (1085 TP objects) in the testing dataset correctly according to its ground truth bounding box compared to all objects that have a ground truth bounding box (1182 objects) and are unsuccessful to detect 8.2% of objects.

In general, the precision and recall values in both models are more than 85%, so they are considered good. Then, the f1 scores for train from scratch and transfer learning are 89.5% and 93%, respectively.

In terms of detection speed, the YOLOv5 transfer learning model succeeded in detecting every 1 image in the testing dataset within 23.8 ms with a total of 1,115 objects detected after applying a confidence threshold of 0.4 and an IoU threshold of 0.3. Based on this, the detection speed of the YOLOv5 transfer learning model is very good because the detection speed is very high.

Based on the performance of the two models, the performance of the YOLOv5 algorithm is quite optimal in detecting testing datasets and the system is quite robust when used to detect objects.

Comparisons

With the data obtained from the training performance parameters of the three YOLO models, a comparison can be made on the training results of each model. As a result, the F1 score and mAP of the YOLOv5 train from scratch model are the lowest values of the 3 existing models. However, in terms of precision, the YOLOv5 train from scratch model has the highest score compared to the other 2 models. In terms of mAP values, the YOLOv4 model has the highest value and is followed by the YOLOv5 transfer learning model. An illustration of the comparison of training results for all models can be seen in Figure 7.

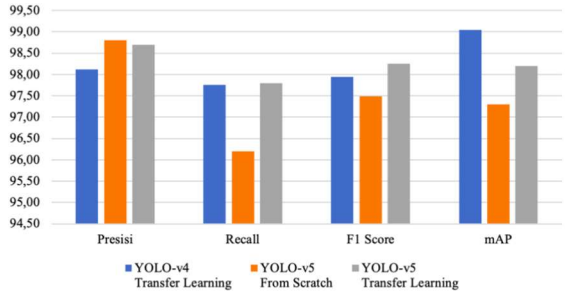

Figure 7. Comparison of the model performance during training

When viewed from the recall level, F1 score, and mAP value, the YOLOv5 train from scratch model has the lowest performance compared to all other models. The YOLOv5 transfer learning model tends to have the best performance, which is then followed by the YOLOv4 transfer learning model.

Overall, the training results in the transfer learning model tend to be better than the train from scratch model. This is because the transfer learning model has been previously trained using certain datasets that have very large or large data, both based on the number of classes and the number of objects in each class. Thus, the initial

network layer in the transfer learning model can extract basic features more quickly and precisely and the training process becomes better. On the other hand, the train from scratch method has no better performance than transfer learning because the training dataset is considered insufficient (too small or too few).

In the transfer learning model, the author uses the YOLOv4 and YOLOv5 algorithms. Based on the training results for the two models, the level of precision, recall, and F1 scores have almost the same value, but YOLOv5 has a better performance. The YOLOv4 model is superior only in terms of mAP values. But in general, YOLOv4 and YOLOv5 have very good performance and both models can be accepted and continued for testing.

In this study, all the proposed CNN models have data limitations, namely the PPE objects that can be detected by the model are the PPE objects studied during training or included in the training dataset and validation dataset. After further experiments were carried out, the result was that the model could not detect other types of PPE objects because the CNN model built in this study is a supervised learning where the model can detect objects based on what has been learned with training datasets and validation datasets. To increase the variation of detection ability in the CNN model that was built, it is possible to add variations of data or types of PPE to the training dataset and validation dataset to become the object of research.

Table 4. Comparison of the testing results

Testing			
Metric Evaluation	YOLOv4 Transfer Learning	YOLOv5 From Scratch	YOLOv5 Transfer Learning
Precision	94.2	92.0	94.2
Recall	91.6	82.1	91.8
F1 Score	92.9	89.5	93.0
mAP	84.4	79.0	88.6

In addition, a comparison is also made of the results of testing each model against the testing dataset, presented in Table 4. As a result, YOLOv5 transfer learning has the best performance in all metrics (precision, recall, F1 score, and mAP) compared to the other two models. In contrast, the YOLOv5 train from scratch model had the lowest performance across all metrics. This is in accordance with the comparison of the results of the training model which shows that the performance of the transfer learning model tends to be better than the train from scratch model.

To find out which algorithm has better detection results, then the performance of all transfer learning models for the two algorithms is compared. Previously, the training results showed that both algorithms had the same good performance. However, based on the testing results, YOLOv5 has a higher performance than YOLOv4. The difference in testing performance between the two algorithms is quite far, at least not like the training results where the two algorithms have comparable and equally good performance.

IV. CONCLUSION

This study presents a vision-based approach (image) to overcome difficulties in identifying compliance with the use of PPE in manufacturing technology laboratories. First, the authors created a data set using real-world images captured in the laboratory. Next, the author builds a CNN model for the needs of PPE identification. The CNN model was compared with different algorithms, namely YOLOv4 and YOLOv5. There are 3 model scenarios built in this research: YOLOv4 (transfer learning), YOLOv5 train from scratch, and YOLOv5 transfer learning. The performance of the proposed approach is evaluated based on the type of algorithm and method used to obtain the optimal CNN model with the best performance.

Based on experiments that uses a dataset of 11,579 images, the following conclusions can be drawn. First, the experimental results show that all scenarios of the proposed CNN model are able to detect various classes of PPE to identify compliance with the use of PPE in the manufacturing laboratory. Second, the results of training and testing have the same conclusion in which the CNN model development method with transfer learning has better performance than training from scratch. At last, the YOLOv5

algorithm has a better performance than YOLOv4 based on the comparison of the results of testing (detection) against the dataset. In terms of selecting the most optimal object detection system in this study, the CNN model with YOLOv5 transfer learning is the model chosen because it has the best performance compared to the other two models. Hence, this research has good prospects with various other applications, for example, compliance management of proper use of PPE in other laboratories, COVID-19 hospital facilities, production plants, and construction sites.

REFERENCES

Barro-Torres, S.; Fernández-Caramés, T.M.; Pérez-Iglesias, H.J.; Escudero, C.J. (2012). "Real-time Personal Protective Equipment Monitoring System", Computer Communications, Vol. 36(1), 42–50.

Bochkovskiy, A.; Wang, C.-Y.; Liao, H.-Y. M. (2020). "YOLOv4: Optimal Speed and Accuracy of Object Detection", arXiv preprint: 2004.10934

BPJS Ketenagakerjaan. (2021). Laporan Keuangan dan Pengelolaan Program BPJS Ketenagakerjaan 2020.

BPJS Ketenagakerjaan. (2022). Laporan Keuangan dazn Pengelolaan Program BPJS Ketenagakerjaan 2021.

Chen, S.; Demachi, K. (2020). "A Vision-based Approach for Ensuring Proper Use of Personal Protective Equipment (PPE) in Decommissioning of Fukushima Daiichi Nuclear Power Station", Applied Sciences, Vol. 10(15), 1–14.

Colares, R.A.L.; de Alencar, D.B.; Junior, J.D.A.B.; da Cruz, J.C.; Bezerra, C.M.V.O. (2019). "The Importance of PPE Use in Civil Construction: Case Study", Journal of Engineering and Technology for Industrial Applications, Vol. 5(20).

de Oliveira, C.S.; Sanin, C.; Szczerbicki, E. (2018). "Flexible Knowledge–Vision–Integration Platform for Personal Protective Equipment Detection and Classification Using Hierarchical Convolutional and Systems, Vol. 49(5–6), 355–367.

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, Vol. 6(September).

Ding, L.; Fang, W.; Luo, H.; Love, P.E.D.; Zhong, B.; Ouyang, X. (2018). "A Deep Hybrid Learning Model to Detect Unsafe Behavior: Integrating Convolution Neural Networks and Long Short-Term Memory", Automation in Construction, Vol. 86, 118–124.

Fang, Q.; Li, H.; Luo, X.; Ding, L.; Luo, H.; Rose, T.M.; dan An, W. (2018). "Detecting Non-hardhat-use by a Deep Learning Method from Far-field Surveillance Videos", Automation in Construction, Vol. 85, 1–9.

Ge, Z.; Liu, S.; Wang, F.; Li, Z.; Sun, J. (2021). "YOLOX: Exceeding YOLO Series in 2021", 1–7.

Hung, P.D.; Su, N.T. (2021). "Unsafe Construction Behavior Classification Using Deep Convolutional Neural Network", Pattern Recognition and Image Analysis, Vol. 31(2), 271–284.

Kasper-Eulaers, M.; Hahn, N.; Kummervold, P. E.; Berger, S.; Sebulonsen, T.; Myrland, Ø. (2021). "Short Communication: Detecting Heavy Goods Vehicles in Rest Areas in Winter Conditions Using YOLOv5", Algorithms, Vol. 14(4).

Krizhevsky, B. A.; Sutskever, I.; Hinton, G. E. (2017). "ImageNet Classification with Deep Convolutional Neural Networks", Communications of the ACM, Vol. 60(6), 84–90.

Lecun, Y.; Bottou, L.; Bengio, Y.; Ha, P. (1998). "Gradient-Based Learning Applied to Document Recognition", Proceedings of the IEEE, 1–46.

Li, Y.; Wei, H.; Han, Z.; Huang, J.; Wang, W. (2020). "Deep Learning-Based Safety Helmet Detection in Engineering Management Based on Convolutional Neural Networks", Advances in Civil Engineering, 2020.

Liu, H.; Fan, K.; Ouyang, Q.; Li, N. (2021). "Real-time Small Drones Detection Based on Pruned YOLOv4", Sensors, Vol. 21(10).

Luo, X.; Li, H.; Wang, H.; Wu, Z.; Dai, F.; Cao, D. (2019). "Vision-based Detection and Visualization of Dynamic Workspaces", Automation in Construction, Vol. 104, 1–13.

Nath, N.D.; Behzadan, A.H.; Paal, S.G. (2020). "Deep Learning for Site Safety: Real-time Detection of Personal Protective Equipment", Automation in Construction, Vol. 112, 103085.

Önal, O.; Dandıl, E. (2021). "Object Detection for Safe Working Environments using YOLOv4 Deep Learning Model", European Journal of Science and Technology, Vol. 26, 343–351.

Padilla, R.; Netto, S.L.; Da Silva, E.A.B. (2020). "A Survey on Performance Metrics for Object-Detection Algorithms", International Conference on Systems, Signals, and Image Processing, 237–242.

Padilla, R.; Passos, W.L.; Dias, T.L.B.; Netto, S.L.; Da Silva, E.A.B. (2021). "A Comparative Analysis of Object Detection Metrics with a Companion Open-source Toolkit", Electronics, Vol. 10(3), 1–28.

Patterson, J.; Gibson, A. (2017). Deep Learning: A Practitioner's Approach, In O'Reilly.

Rahman, E.U.; Zhang, Y.; Ahmad, S.; Ahmad, H.I.; Jobaer,

S. (2021). "Autonomous Vision-Based Primary Distribution Systems Porcelain Insulators Inspection Using UAVs", Sensors, Vol. 21(3), 1– 24.

Redmon, J.; Divvala, S.; Girshick, R.; dan Farhadi, A. (2016). "You Only Look Once: Unified, Real-Time Object Detection", Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 779– 788.

Wang, Z.; Wu, Y.; Yang, L.; Thirunavukarasu, A.; Evison, C.; Zhao, Y. (2021). "Fast personal protective equipment detection for real construction sites using deep learning approaches", Sensors, 21(10), 3478.

Wu, H. Zhao, J. (2018). "Automated Visual Helmet Identification Based on Deep Convolutional Neural Networks", Computer Aided Chemical Engineering, Vol. 44(2018), 2299–2304.

Zhafran, F.; Ningrum, E.S.; Tamara, M.N.; Kusumawati, E. (2019). "Computer Vision System Based for Personal Protective Equipment Detection, by Using Convolutional Neural Network", Proceedings of IES 2019 - International Electronics Symposium: The Role of Techno-Intelligence in Creating an Open Energy System Towards Energy Democracy, 516–521.