Avrupa Bilim ve Teknoloji Dergisi
Özel Sayı 26, S. 343-351, Temmuz 2021

Araştırma Makalesi

www.ejosat.com ISSN:2148-2683

European Journal of Science and Technology
Special Issue 26, pp. 343-351, July 2021

Research Article

Object Detection for Safe Working Environments using YOLOv4 Deep Learning Model

Oğuzhan Önal[1], Emre Dandıl[2*]

[1] Bilecik Seyh Edebali University, Vocational School, Electronic and Automation, Bilecik, Turkey (ORCID: 0000-0002-4336-5064)
[2] Bilecik Seyh Edebali University, Faculty of Engineering, Department of Computer Engineering, Bilecik, Turkey (ORCID: 0000-0001-6559-1399)

(International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA) 2021 – 11-13 June 2021)

(**DOI:** 10.31590/ejosat.951733)

ATIF/REFERENCE: Önal, O. & Dandıl, E. (2021). Object Detection for Safe Working Environments using YOLOv4 Deep Learning Model. *European Journal of Science and Technology*, (26), 343-351.

Abstract

The health and safety of employees in workplaces maintains its importance since the concept of production emerged. Recent developments in computer vision and deep learning have made it widespread to be used in work environments as a secondary tool in ensuring occupational safety from surveillance videos. Thus, an important performance is achieved by minimizing human-induced errors in working environments. In this study, a method based on the YOLOv4 deep learning model is proposed to control the use of personal protective equipment from videos and to detect unsafe movements in the working environments of facilities operating in the field of industrial production. In the study, a dataset is created with videos collected from different working environments. In the study, later, on the prepared video dataset, the detection of personal protective equipment such as helmets, vests, masks, gloves, eyeglasses used by workers in factories operating in industrial areas and whether they use the appropriate equipment correctly is determined using the YOLOv4 framework. In the experimental studies conducted within the scope of the study, the mean average precision (mAP) value is achieved as 91.18% as a result of the training performed in the YOLOv4 network. In addition, results of 0.89, 0.91, 0.90, 70.35 and 1.1147 are obtained for other measurement metrics such as precision, recall, F1-score, intersection over union (IoU), and average loss, respectively. As a result, in the proposed study, instant inspection of the videos collected from the cameras installed in the factories, the meaning of the scene and the control of safe working environments are successfully achieved.

Keywords: Object detection, Safe working environment, Personal protective equipments, Deep learning, YOLOv4.

Güvenli İş Ortamı İçin YOLOv4 Derin Öğrenme Modeli Kullanarak Nesne Tanıma

Öz

İşyerlerinde çalışanların sağlığı ve güvenliği üretim kavramı ortaya çıktığından bu yana önemini korumaktadır. Bilgisayarlı görü ve derin öğrenme konusunda son yıllarda kaydedilen gelişmeler, çalışma ortamlarında gözetim videolarından iş güvenliğinin sağlanmasında ikincil bir araç olarak kullanılmaya başlamıştır. Böylece çalışma ortamlarında insandan kaynaklı hataların minimuma indirilerek önemli bir başarım elde edilmesi sağlanmaktadır. Bu çalışmada, endüstriyel üretim alanında faaliyet gösteren tesislerin çalışma ortamlarında, videolardan kişisel koruyucu donanımların kullanımın denetlenmesi ve güvensiz hareketlerin tespiti için YOLOv4 derin öğrenme modeli tabanlı bir yöntem önerilmektedir. Çalışmada, öncelikle farklı çalışma ortamlarından toplanan videolar ile bir veriseti oluşturulmuştur. Çalışmada daha sonra, hazırlanan video veriseti üzerinde, sanayi bölgelerinde faaliyet gösteren fabrikalarda işçilerin kullandığı baret, yelek, maske, eldiven, gözlük gibi kişisel koruyucu ekipmanların tanınması ve uygun

[*] Corresponding Author: Bilecik Seyh Edenali University, Faculty of Engineering, Department of Computer Engineering, Bilecik, Turkey (ORCID: 0000-0001-6559-1399), emre.dandil@bilecik.edu.tr

donanımları doğru kullanıp kullanmadıkları YOLOv4 altyapısı kullanılarak tespit edilmiştir. Çalışma kapsamında yürütülen deneysel çalışmalarda, YOLOv4 ağında yapılan eğitim sonucunda mean average precision (mAP) değeri %91.18 olarak başarılmıştır. Ayrıca, diğer ölçüm metrikleri kesinlik, duyarlılık, F1-skoru, kesiştirilmiş bölgeler (IoU) ve ortalama kayıp için sırasıyla 0.89, 0.91, 0.90, 70.35 ve 1.1147 sonuçları elde edilmiştir. Sonuç olarak, önerilen çalışmada, fabrikalarda tesis edilmiş kameralardan gelen videoların anlık olarak denetlenmesi ve sahnenin anlamlandırılması sağlanarak, güvenli çalışma ortamlarının kontrolü başarılı bir şekilde sağlanmıştır.

Anahtar Kelimeler: Nesne tanıma, Güvenli iş ortamı, Kişisel koruyucu ekipmanlar, Derin öğrenme, YOLOv4.

1. Introduction

Today, real-time video monitoring is performed from a large number of cameras in airports, hospitals, kindergartens, traffic, construction, large enterprises, factory environments, in short, in almost all critical areas open to the public. It is necessary to recognize, interpret and make sense of the complex events that occur by viewing the images in these video sequences. Due to the difficulty of observing these videos, systems that can recognize complex scenarios on the basis of semantic models representing monitored situations are needed. Recognition of complex events from surveillance videos obtained from work environments is very important in terms of creating safe working environments and tracking employees.

Ensuring the safety of people is a common and highly demanding task in workplaces due to the dynamic and complex working conditions that exist in working environments. Despite regulatory reforms, laws and efforts by industry associations, and extensive research to address this problem, accidents and fatalities in the workplace remain a worldwide problem (Ceylan & Ceylan, 2012). Occupational health and safety fulfills a fundamental function that contributes to the proper functioning of the productive structures of regions. In addition, occupational health and safety promotes the development of safer work environments and thus helps to stimulate safety policies, social welfare and regional economies (Ruser & Butler, 2010).

Thanks to the understanding and semantic segmentation of videos, important studies are carried out on occupational health and safety (Yu *et al.*, 2020). In a research, it was stated that approximately 90% of all accidents that occur in work environments are caused by unsafe behaviors (Heinrich & Granniss, 1959). Unsafe behavior can occur when an employee does not comply with safety rules, standards, procedures, instructions and specified project criteria. Such actions may adversely affect an employee's performance and / or endanger others in the workplace (Ding *et al.*, 2018). If unsafe behavior in workplaces can be reduced or prevented, safety performance will naturally increase. Traditional methods for determining behavior in work environments are predominantly based on observational methods. While such methods offer useful information, they are time consuming, labor intensive, and subjective in nature. Because of these limitations, computer vision technologies used for object recognition can be applied to identify unsafe actions of employees in the workplace.

In recent years, society has demanded improved mechanisms to prevent occupational hazards. In particular, the industrial sector is a field that needs to be carried out on specific studies. Workers in the industry have to obey certain working rules at the work sites. Compliance and implementation of these safety and health procedures are directly under the responsibility of the employee and the company. The use of protective elements such as helmets, gloves, boots, safety goggles and seat belts is one of the most important rules to be followed in working environments in a production enterprise. These elements are collectively known as personal protective equipment (PPE).

One of the most important elements of occupational health and safety is the use of PPE. After all the accident prevention measures in practice, PPEs are the only elements in ensuring the safety of the employees. For these reasons, the use of PPEs in workplaces is of great importance. The use of PPEs in industries varies according to the area of study. For example, the use of hard hats (helmet) is not mandatory in some manufacturing enterprises, while it is compulsory in some enterprises. In the some researches, it has been reported that thousands of casualties have occurred in employees without PPE as a result of being hit by falling objects (H. Wu & Zhao, 2018). In another study on the causes and prevention of occupational accidents experienced by technical personnel, it was stated that the accidents are mostly caused by unsafe behavior (Aybek *et al.*, 2003). Therefore, minimizing unsafe behaviors in industries requires uninterrupted and important measures.

In previous studies, it is seen that there are two main ways of using PPEs and detecting unsafe movements. The first and older of these ways is to track employees by placing RFID-like sensors on them and to detect unsafe movements (Kelm *et al.*, 2013; Lee *et al.*, 2012). Second one, because of its high performance in recent years and its ability to work in real time, it is ensured by computer vision that the PPE usage and unsafe movements of the employees can be controlled directly by using computers and cameras (B. H. Guo *et al.*, 2021; Nath *et al.*, 2020). In recent years, studies in the field of computer vision and accordingly developments in this field have increased very rapidly. These developments are reshaping many industrial areas. Computer vision develops widely on subjects such as recognizing and tracking objects, detecting anomalies, activity recognition and video understanding, and is mainly based on deep learning. Computer vision has two basic parameters, namely accuracy and speed (Barro-Torres *et al.*, 2012). The YOLO (You Only Look One) algorithm, which emerged as a deep learning network, has become very popular in real-time object detection because it performs the detection of objects in a single step, unlike previous studies (Redmon *et al.*, 2016).

Another factor that the studies proposed on the use of PPEs can be classified is the study areas where they are applied. Industries are classified according to their hazard rates depending on their working conditions and environments, and the usage patterns of PPEs vary accordingly. There are many studies on the construction industry in previous (Ding *et al.*, 2018; Kelm *et al.*, 2013; Lee *et al.*, 2012; Nath *et al.*, 2020; Nill, 2019). Since this industry is very common and many occupational accidents occur in this area, it is necessary to carry out studies based on new technologies. Another sector in which similar studies are carried out, though not as much as the construction industry, is the production sector. In the production sector, there is a need for work safety studies as much as the construction sector. Occupational accidents, injuries and even

deaths can occur, no matter how many precautions are taken with conventional methods. The reason we distinguish these two sectors from each other in terms of computer vision is the difference in working environments. Production environments are closed environments that contain more objects and require lighting. Therefore, it is difficult to work in these areas in terms of computer vision.

In previous studies, there are many studies in which different methods were used for the use of PPEs for the establishment of a safe work environment. In a 2012 study (Barro-Torres *et al.*, 2012), a system using Zigbee and RFID was proposed to monitor the use of PPEs in real time. However, the equipment for this system has to be specially produced and processed. In a study proposed in 2018 (Ding *et al.*, 2018), computer vision and pattern recognition approaches were applied to identify unsafe behaviors on construction sites. For this purpose, a hybrid model was developed using convolutional neural networks (CNN) and long-short term memory (LTSM). Another study (H. Wu & Zhao, 2018) conducted in 2018 focused on determining whether workers were wearing only helmets and the color of helmets. In the study, a hierarchical support vector machine was created for classification, and a certain accuracy performance was obtained in evaluation. Likewise, in a study conducted in 2019 (J. Wu *et al.*, 2019), it is based on the determination of whether the helmets used are suitable for employees with colors. The single shot multibox detector (SSD) algorithm was used to determine the final detection results, and a significant mAP value was achieved. Another study (Balakreshnan *et al.*, 2020) focused on the conformity determination of PPEs in factories. A combination of cloud-based and artificial intelligence was used to provide a real-time vision-based in-house security system that can detect and record potential security breaches, encourage compliance and ultimately prevent accidents before they happen. In the study, it was stated that the hybrid artificial intelligence architecture approach provides flexibility. In another study in 2021 (Chen & Demachi, 2021), a new solution was proposed to identify the inappropriate use of PPE with a combination of deep learning-based object detection and individual perception using geometric relations analysis by construction workers.

In this study, a YOLOv4 deep learning algorithm-based method is proposed to control the use of PPEs from videos and to detect unsafe movements in industrial production facilities. In the study, it is ensured that the videos coming from the cameras installed in the factories were instantly inspected by YOLOv4 deep learning methods and the scene is recognized meaningful. In the study, a dataset is created with videos collected from different working environments. Within the scope of the study, then, on the video dataset, it is determined whether the workers in the factories operating in the industrial areas are using the appropriate equipment correctly or not using the YOLOv4 deep learning algorithm. The rest of the work is organized as follows. Section 2 introduces the YOLOv4 architecture and processes of image/video dataset pre-processing, including image acquisition, image enhancement, and image dataset preparation. In Section 3, the results and discussion of the findings obtained in the experimental studies within the scope of the study are detailed. In the last Section, the inferences and expectations obtained from the results of the study are explained.

2. Material and Method

2.1. Dataset

The image and video data used in this study were obtained from two different factories named “Kafaoğlu Metal Plastik Makine San. ve Tic. A.Ş.” and “Tek Metal ve Plastik Endüstriyel Mamulleri San. Tic. Ltd. Şti.” operating in Eskisehir Organized Industrial Zone. In order to achieve higher accuracy performance in deep learning algorithms, it is necessary to train the network with many images / videos. For this reason, in order to expand the video / image dataset and accelerate the experimental studies, videos and images in different resolutions were collected from the working environments in Machine Workplace of Vocational School of Bilecik Şeyh Edebali University by using several different cameras at various times. Necessary permissions were obtained from the relevant institutions / organizations for all working environments where data were collected. Sample frames for some of the videos collected for the dataset prepared within the scope of the study are presented in Figure 1.

Figure 1. Sample images from the dataset prepared within the scope of the study

Especially, the biggest problem in terms of computer vision in production areas is lighting. Generally, these areas are constantly illuminated by artificial lighting sources. Therefore, the most accurate angles in terms of illumination were used while obtaining image / video data. A dataset was created for our study by obtaining a total of 2200 images of different sizes from the videos collected from different scenes. The images were selected in a way that they could determine the PPE usage of the

employees. Therefore, the images focused on helmets, protective glasses, gloves, work vests and masks while working.

2.2. Object Labelling in Working Environments

A total of 4000 images were obtained from the videos collected within the scope of the study, with 4 frames per second. 2200 of these images were labeled using the Labelimg program, in accordance with the format of the structure of the YOLOv4 algorithm, where the network was trained. The labeling process is done in the form of boxing the object in the image to determine its position. Accordingly, the x-y coordinates, height and width of the object are determined in the generated label file. The high number of objects in the images caused the labeling process to take a long time. After the labeling process, the classes used in object recognition and detection and the number of objects in the labeled classes are shown in Table 1. The inequality in object numbers here is due to the different classes of objects in the scenes. However, this has helped to determine what effect the differences in the number of objects have on the training of the network.

Table 1. Classes in object detection phase and the number of objects in labeled classes

Class number	Class name	The number of object in class
1	Helmet	1891
2	Glove	3408
3	Protective glass	2295
4	Mask	3214
5	No Helmet	3010
6	No Mask	370
7	Vest	2621

2.2. YOLOv4 Deep Learning Model

YOLOv4 algorithm is a deep learning algorithm published by (Bochkovskiy *et al.*, 2020) in April 2020 and used in object detection. In this study, YOLOv4 algorithm was used for object recognition in working environments. YOLO deep learning algorithm is one of the best proposed algorithms for real-time object detection by improving its speed and performance (Long *et al.*, 2020). Bochkovskiy *et al.* stated that by optimizing the existing YOLO structure, training of the network can be done easily with a single GPU graphics processing card and also high accuracy performance can be achieved with the Tesla V100 graphics card (Bochkovskiy *et al.*, 2020). The block diagram of the YOLOv4 algorithm is denoted in Figure 2. The block diagram structure in this architecture consists of three sub-layers. Feature extraction is performed in the backbone layer, which is the first layer. As with the YOLOv4 architecture, CSPDarknet53 is used as a backbone in this study. In the second layer, the neck layer, information is extracted from neighboring feature maps with bottom-top and top-down flows in order to achieve higher performance in predicting objects. Spatial pyramid pooling (SPP) and path aggregation network (PAN) are used in this layer (F. Guo *et al.*, 2021). In the last layer, the head layer, there are bounding boxes and the class of each box is estimated. In this layer, the estimation procedure of the YOLOv3 algorithm is used (D. Wu *et al.*, 2020).

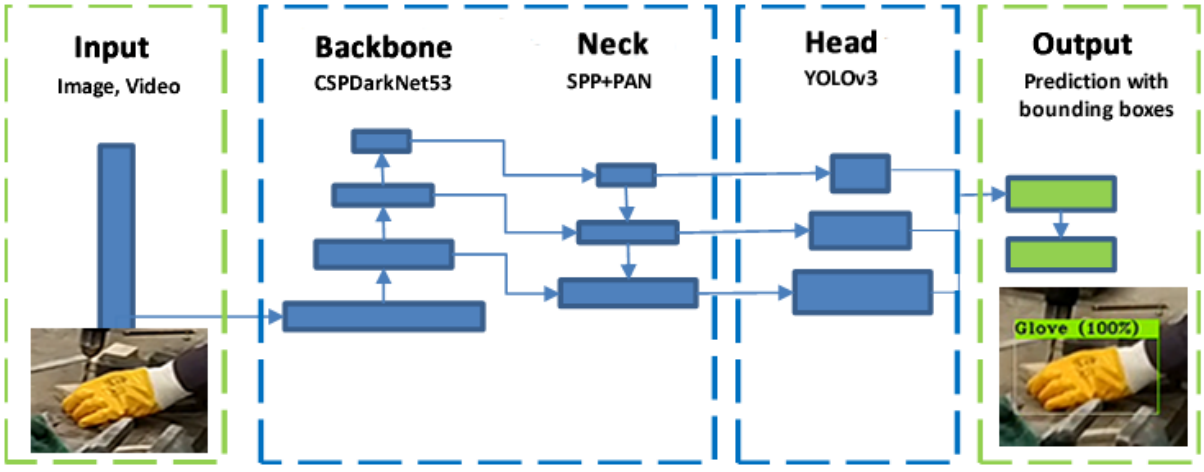

Figure 2.Block diagram of YOLOv4 model used for object detection in this study

3. Results and Discussion

In this study, a workstation with NVIDIA GeForce GTX 1080 double GPU was used for the experimental studies conducted to train the obtained data. In the study, firstly 1000 images were labelled and these images were edited and trained with the YOLOv4 algorithm. Afterwards, the number of the labelled images was increased two more times, and the network was trained a few more times and the required optimum weights were obtained. In this study, the training durations and other information obtained in different trainings performed according to various parameters of the YOLOv4 network, which is proposed to detect objects in working environments, are given in Table 2. There are many parameters that affect the training duration of the network. The most important parameter in terms of training time is GPU usage and number. The high number of layers the YOLO algorithm has increases the need for powerful hardware. Thus, training without the use of a GPU becomes very difficult. In fact, it is possible to shorten the training times thanks to the use of more than one GPU. In this study, 2 NVIDIA GeForce 1080 GPU were used for training the YOLOv4 network. As can be seen from Table 2, the training

times were compared by running the trainings with a single GPU and double GPU using the same parameters. According to the results, training time is shortened by approximately 30% in double GPU usage. In addition, the size of the image whose values can be adjusted in the pre-training configuration file shown in this table, Subdivision, Random values and the number of iterations significantly affect both the training time and the mAP value.

Table 2. The proposed YOLOv4 network training parameters and durations for object detection from working environments in this study

Training number	The number of training images	The number of validation images	The number of class	The number of iteration	The number of GPU	Image size (wxh)	Random	Batch size	Subdivisions	Training Duration (min)
1	2670	0	3	6000	1	416x416	0	64	32	423
2	2670	0	3	6000	2	416x416	0	64	32	246
3	888	112	7	15000	2	416x416	0	64	32	845
4	888	112	7	15000	1	416x416	0	64	32	1210
5	888	112	7	15000	2	416x416	0	64	16	605
6	888	112	7	15000	2	416x416	0	64	32	700
7	888	112	7	15000	2	416x416	1	64	32	640
8	1374	148	7	16000	2	608x608	0	64	32	1404
9	2006	203	7	18000	2	416x416	0	64	16	746
10	2000	178	7	14000	1	416x416	1	64	32	1560

For the experimental studies conducted within the scope of this study, Precision (P), Recall (R), F1-score (F1), Average Precision (AP), Mean Average Precision (mAP) and Intersection over Union (IoU) metrics were measured to evaluate the performance of the proposed YOLOv4 network model in object recognition in working environments. These metrics are denoted in Eq. (1), Eq. (2), Eq. (3), Eq. (4), Eq. (5) and Eq. (6), respectively. Precision refers to how many of the values predicted as positives are actually positives. Recall is a metric that shows how much of the data that have to be predicted as positive is predicted as positive. The F1-score criterion denotes the harmonic mean of the precision and recall scores. The mAP is calculated by mean of the average precision values of the classes. IoU is expressed as the area where two rectangles intersect divided by the area of the union of these two rectangles. In the IoU, B represents the predicted value and B_{gt} the reference value.

$$\text{Precision (P)} = \frac{\text{TP}}{\text{TP + FP}} \tag{1}$$

$$\text{Recall (R)} = \frac{\text{TP}}{\text{TP + FN}} \tag{2}$$

$$\text{F1 - score (F1)} = 2\,x\,\frac{P\,x\,R}{P + R} \tag{3}$$

$$\text{Average Precision (AP)} = \int_0^1 p(r)\ dr \tag{4}$$

$$\text{Mean Average Precision (mAP)} = \frac{1}{n}\sum_{i=1}^{n} AP_i \tag{5}$$

$$\text{Intersection over Union (IoU)} = \frac{|B \cap B_{gt}|}{|B \cup B_{gt}|} \tag{6}$$

In this study, the YOLOv4 algorithm used for object detection was trained with different sized datasets, and the performance of the network was measured with a test dataset. The results were compared with mAP, P, R, F1, TP, FP, FN, IoU and Average Loss parameters and the obtained results presented in Table 3. The results in this table were obtained by specifying the 0.5 threshold value for the IoU value in the trainings {3, 4, 5, 6, 7, 8, 9}. When Table 3 is examined in detail, it is seen that the highest mAP value was obtained in training number 5 with 91.18%. The variation of average loss and mAP values for the number 5 training of the proposed YOLOv4 network is shown in Figure 3. In training number 5 for the YOLOv4 network, the results of 0.89, 0.91, 0.90, 665, 79, 68, 70.35 and 1.1147 were achieved for the other measurement metrics P, R, F1, TP, FP, FN, IoU and Average Loss, respectively. In this training, YOLOv4 algorithm was trained with a dataset with 888 training and 112 verification images. When this graph is examined in detail, it is observed that the learning process took place quickly until the 4500th iteration, the average loss decreased dramatically, but there was no noticeable improvement after this point. In addition, it is seen that the mAP score became more stable after the 10000th iteration and this determination settled well in the last 2000 iterations.

Table 3.Comparison of performance metric results obtained with the YOLOv4 model for different trainings

Training number	mAP (@0.50)	P	R	F1	TP	FP	FN	IoU	Average Loss
3	85.29	0.76	0.84	0.80	619	196	114	57.15	1.5283
4	86.50	0.80	0.84	0.82	619	155	114	61.0	0.9543
5	91.18	0.89	0.91	0.90	665	79	68	70.35	1.1147
6	83.98	0.79	0.83	0.81	610	161	123	59.88	1.4929
7	90.63	0.89	0.90	0.89	661	84	72	70.54	1.5082
8	83.17	0.79	0.84	0.81	618	169	115	58.10	1.5425
9	76.54	0.68	0.79	0.73	580	269	153	50.36	1.8101

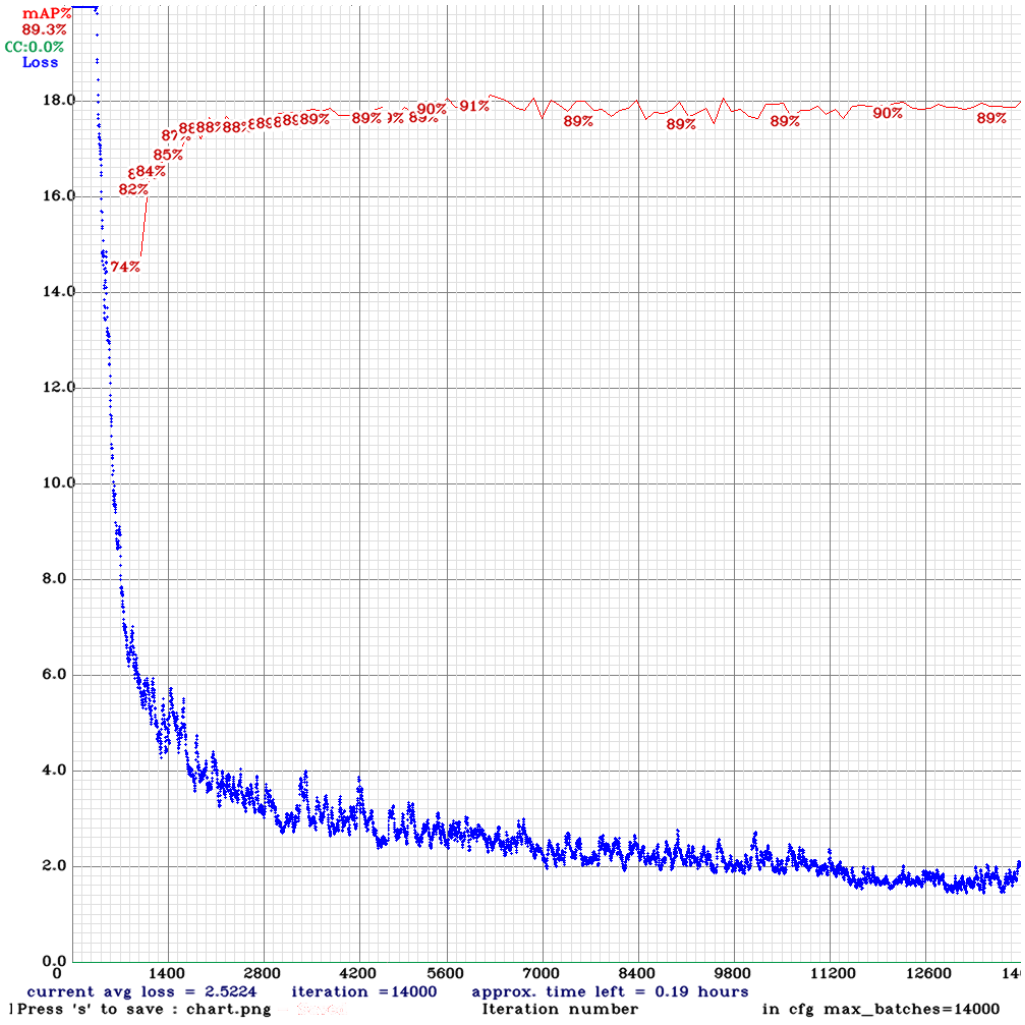

Figure 3.Variation of average loss and mAP scores for training number 5 of the proposed YOLOv4 network according to the number of iterations

Table 4 shows the AP, TP and FP scores obtained in the object recognition process using the YOLOv4 network after training number 5, where the highest mAP score was obtained at 0.50 threshold value for IoU. In addition, the scores achieved for each of the helmet, glove, protective glasses, mask, no-helmet, no-mask and vest classes can be seen from this table. As can be seen from this table, high performance was achieved in detecting the objects using the proposed YOLOv4 method.

Table 4. AP, TP and FP scores obtained for each class in training number 5 using the proposed YOLOv4 network model

Class name	AP (%)	TP	FP
Helmet	97.62	90	2
Glove	96.80	122	4
Eyeglasses	92.97	61	10
Mask	90.35	115	14
No Helmet	78.71	138	41
No Mask	82.69	18	5
Vest	99.09	121	3

In Figure 4, the visual results obtained in recognizing the objects in the working environments by using the proposed YOLOv4 algorithm in the frames, not used in the training phase, obtained from a video in the dataset prepared within the scope of

the study. From this figure, it can be seen that the objects in the working environment are detected successfully using the YOLOv4 algorithm and the accuracy rate is marked as the object label. The images used for the test in this video were not labelled and not included in the training set. These images were obtained using different angles at a different time from the other images.

Figure 4. Object detection in the working environment on video frames using the proposed YOLOV4 model

In Figure 5, visual results of object detection can be seen in a still image obtained from the work environment and not used in the training of the YOLOv4 network. Over 88% accuracy was achieved in the recognition of all objects except the glove object in work environments.

Figure 5.Object detection using the proposed YOLOv4 architecture in a still image obtained from the workplace environment

The results of the AP, TP and FP scores obtained in the recognition of the objects in the working environment using the YOLOv4 network in the experimental study performed by using the weights formed after the training numbered 5 with a threshold value of 0.75 for IoU are presented in Table 5. As can be seen from this table, the results of measurement metrics decreases due to the increase of the threshold value for IoU. In this performance measurement where the mAP value is obtained as 53.88%, while the TP values for the object classes decrease, the FP values increase. Here, P, R and F1 scores were measured as 0.62, 0.63 and 0.62, respectively .

Table 5. AP, TP and FP scores of each class for the IoU threshold value of 0.75 in training number 5 using YOLOv4 network

Class name	AP (%)	TP	FP
Helmet	81.39	78	14
Glove	55.77	85	41
Eyeglasses	46.59	37	34
Mask	47.57	70	59
No Helmet	31.75	66	113
No Mask	20.69	8	15
Vest	93.43	115	9

4. Conclusions

In this study, object detection was carried out in order to control the use of PPEs of employees in the manufacturing industry and to establish a safe working environment by using the YOLOv4 deep learning model. In this context, the detection and recognition of personal protective equipment such as helmets, vests, masks, gloves, protective glasses in working environments and whether the employees use the appropriate equipment correctly or not were determined using the YOLOv4 framework. Within the scope of the study, a dataset was created using images / videos from "Kafaoğlu Metal Plastik Makine San. ve Tic. A.Ş.", "Tek Metal ve Plastik Endüstriyel Mamulleri San. Tic. Ltd. Şti." and "Machine Workplace of Vocational School of Bilecik Şeyh Edebali University". In order to achieve higher accuracy performance in deep learning algorithms, it is necessary to train the architecture with a lot of images / videos. For this reason, videos and images in different resolutions were collected from the working environments using several different cameras at various times. In the experimental studies conducted within the scope of the study, the YOLOv4 network was trained using different network parameters and different numbers of GPUs, and the performance of the proposed method on video images obtained from similar environments was tested using the weights obtained as a result of the training. As a result of specifying the IoU threshold value as 0.5, the mean average precision (mAP) value reached the highest value as 91.18% as a result of the training performed on the proposed YOLOv4 network. Also, scores of 0.89, 0.91, 0.90, 70.35 and 1.1147 were achieved for P, R, F1, IoU and average loss, respectively. As a result, in the proposed study, the control of safe working environments was successfully achieved by object detection and recognition, by instantaneously inspecting the video streaming from the cameras and understanding of the scene. Within the scope of the study, experimental studies are able to be conducted on object detection and interpretation of video content from real-time videos from the working environments.

Acknowledgements

We would like to thank "Kafaoğlu Metal Plastik Makine San. ve Tic. A.Ş.", "Tek Metal ve Plastik Endüstriyel Mamulleri San. Tic. Ltd. Şti." and "Vocational School of Bilecik Şeyh Edebali University" for allowing us to use the image / video data used in this study. In addition, this study is financially supported by the Administration of Scientific Research Projects of Bilecik Şeyh Edebali University with the project number 2019-02.BŞEÜ.01-03.

References

Aybek, A., Güvercin, Ö., & Hurşitoğlu, Ç. (2003). Teknik personelin iş kazalarının nedenleri ve önlenmesine yönelik görüşlerinin belirlenmesi üzerine bir araştırma. *KSÜ Fen ve Mühendislik Dergisi, 6*(2), 91-100.

Balakreshnan, B., Richards, G., Nanda, G., Mao, H., Athinarayanan, R., & Zaccaria, J. (2020). PPE Compliance Detection using Artificial Intelligence in Learning Factories. *Procedia Manufacturing, 45*, 277-282.

Barro-Torres, S., Fernández-Caramés, T. M., Pérez-Iglesias, H. J., & Escudero, C. J. (2012). Real-time personal protective equipment monitoring system. *Computer Communications, 36*(1), 42-50.

Bochkovskiy, A., Wang, C.-Y., & Liao, H.-Y. M. (2020). Yolov4: Optimal speed and accuracy of object detection. *arXiv preprint arXiv:2004.10934*.

Ceylan, H., & Ceylan, H. (2012). Analysis of occupational accidents according to the sectors in Turkey. *Gazi University Journal of Science, 25*(4), 909-918.

Chen, S., & Demachi, K. (2021). Towards on-site hazards identification of improper use of personal protective equipment using deep learning-based geometric relationships and hierarchical scene graph. *Automation in construction, 125*, 103619.

Ding, L., Fang, W., Luo, H., Love, P. E., 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, 86*, 118-124.

Guo, B. H., Zou, Y., Fang, Y., Goh, Y. M., & Zou, P. X. (2021). Computer vision technologies for safety science and management in construction: A critical review and future research directions. *Safety science, 135*, 105130.

Guo, F., Qian, Y., & Shi, Y. (2021). Real-time railroad track components inspection based on the improved YOLOv4 framework. *Automation in construction, 125*, 103596.

Heinrich, H. W., & Granniss, E. (1959). *Industrial Accident Prevention*: McGraw-Hill Book Company.

Kelm, A., Laußat, L., Meins-Becker, A., Platz, D., Khazaee, M. J., Costin, A. M., Helmus, M., & Teizer, J. (2013). Mobile passive Radio Frequency Identification (RFID) portal for automated and rapid control of Personal Protective

Equipment (PPE) on construction sites. *Automation in construction, 36*, 38-52.

Lee, H.-S., Lee, K.-P., Park, M., Baek, Y., & Lee, S. (2012). RFID-based real-time locating system for construction safety management. *Journal of Computing in Civil Engineering, 26*(3), 366-377.

Long, X., Deng, K., Wang, G., Zhang, Y., Dang, Q., Gao, Y., Shen, H., Ren, J., Han, S., & Ding, E. (2020). PP-YOLO: An effective and efficient implementation of object detector. *arXiv preprint arXiv:2007.12099*.

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, 112*, 103085.

Nill, R. J. (2019). How to Select and Use Personal Protective Equipment. *Handbook of Occupational Safety and Health*, 469-494.

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). *You only look once: Unified, real-time object detection.* Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 779-788.

Ruser, J., & Butler, R. (2010). *The economics of occupational safety and health*: Now Publishers Inc.

Wu, D., Lv, S., Jiang, M., & Song, H. (2020). Using channel pruning-based YOLO v4 deep learning algorithm for the real-time and accurate detection of apple flowers in natural environments. *Computers and Electronics in Agriculture, 178*, 105742.

Wu, H., & Zhao, J. (2018). An intelligent vision-based approach for helmet identification for work safety. *Computers in Industry, 100*, 267-277.

Wu, J., Cai, N., Chen, W., Wang, H., & Wang, G. (2019). Automatic detection of hardhats worn by construction personnel: A deep learning approach and benchmark dataset. *Automation in construction, 106*, 102894.

Yu, W.-D., Liao, H.-C., Hsiao, W.-T., Chang, H.-K., Tsai, C.-K., & Lin, C.-C. (2020). *Automatic Safety Monitoring of Construction Hazard Working Zone: A Semantic Segmentation based Deep Learning Approach.* Paper presented at the Proceedings of the 2020 the 7th International Conference on Automation and Logistics (ICAL). pp. 54-59.