International Journal of Advanced Culture Technology Vol.10 No.1 302-309 (2022)
DOI https://doi.org/10.17703/IJACT.2022.10.1.302

IJACT 22-3-37

A study on Detecting the Safety helmet wearing using YOLOv5-S model and transfer learning

[1]NaeJoung Kwak, [2]DongJu Kim*

[1]Prof., Dept. of Cyber and Security, Baejae Univ., Korea
[2]Professor, Postech Institute of Artificial Intelligence, POSTECH
knj0125@pcu.ac.kr, kkb0320@postech.ac.kr

Abstract

Occupational safety accidents are caused by various factors, and it is difficult to predict when and why they occur, and it is directly related to the lives of workers, so the interest in safety accidents is increasing every year. Therefore, in order to reduce safety accidents at industrial fields, workers are required to wear personal protective equipment. In this paper, we proposes a method to automatically check whether workers are wearing safety helmets among the protective equipment in the industrial field. It detects whether or not the helmet is worn using YOLOv5, a computer vision-based deep learning object detection algorithm. We transfer learning the s model among Yolov5 models with different learning rates and epochs, evaluate the performance, and select the optimal model. The selected model showed a performance of 0.959 mAP.

Keywords: *Safety Helmet, Object Detection, Yolo, Safety Accidents, Personal Protective Equipment*

1. INTRODUCTION

The convergence of ICT technology and various industries enriches people's lives[1]. In addition, various disasters occur due to the urbanization of society due to the development of technology[2]. Since industrial sites such as construction sites, apartment facility management[3] are exposed to the outside environment, there are many risk factors than in other industries, so the accident rate is high. It is stipulated that personal protective equipment (PPE) that protects the body of workers from safety accidents must be worn[4].

According to the report on the occurrence of industrial accidents in Korea, the head accounts for 48% of the injuries among various body parts, showing the highest proportion. Therefore, if a safety helmet is worn when entering an industrial site, it is possible to reduce accidents by protecting the worker's head safely [5]. However, in the actual field, there are many cases where it is not worn or not properly worn for reasons such as stuffiness and annoyance. Also, in most construction sites, it is difficult for safety managers to continuously monitor whether construction workers are wearing helmet. In order to solve this problem, methods for automatically detecting the helmets of construction workers have been studied based on image data obtained from camera devices such as CCTV installed in the field.

The image-based helmet automatic detection technology is based on an object detection technique in the computer vision field. Methods for object detection have been studied based on traditional computer vision

Manuscript received: February 23, 2022 / revised: March 1, 2022 / accepted: March 8, 2022
Corresponding Author: kkb0320@postech.ac.kr
Tel: +*** - **** - ****
Professor, Postech Institute of Artificial Intelligence, POSTECH, Korea
Copyright©2022 by The International Promotion Agency of Culture Technology. This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0)

and machine learning techniques[6], but deep learning algorithms are currently being widely used for object detection[7]. The You Only Look Once (YOLO) algorithm is used in a representative way[8]. Compared to other deep learning-based object detection algorithms, YOLO shows fast processing speed and relatively high detection accuracy[9]. Among several versions of Yolo algorithm, YOLOv5[10] is a recently developed algorithm that can detect even small objects at high speed. And the algorithm is classified into four models: s, m, l, and x according to speed and performance.

In this study, we propose a method to automatically detect whether or not a helmet is worn using the s-model of YOLOv5 so that can find the target object by processing the image in real time. In addition, We transfer learning the s model among Yolov5 models with different learning rates and epochs, evaluate the performance, and select the optimal model.

2. RELATED WORKS

Detection of helmets by image recognition has already been continuously studied in the overseas computer vision field, and a study of applying machine learning based on geometrical features has been conducted. Recently, a method of classifying or detecting an object based on a neural network (NN) has emerged. As object detection methods, networks such as Faster R-CNN [11], R-FCN [12], YOLO [8], and SSD [13] achieve high accuracy. Various studies are being conducted to apply these algorithms to the detection of safety helmets.

Fang et al. [14] used Faster R-CNN [11] to detect people who are not wearing safety helmets. Wu et al. [15] proposed the use of SSD for the detection of wearing safety helmets. However, these algorithms have a problem that it takes a long time to learn, and real-time safety helmet detection is not easy. Due to this, the YOLO algorithm, which is easy to process in real time and to be lightweight, has begun to attract attention. Mistry et al. [16] implemented a system that detects whether a motorcycle rider is wearing a helmet using pretrained YOLOv2. [17] applied YOLOv4 to the detection of safety helmets. And the mAP was shown to be up to 81.81%.

3. METHODOLGY

This study uses to detect a safety helmet wearing by transfer learning and selects the model showing optimum performance by analyzing the performance according to the learning rate and epochs and a safety helmet wearing. The selected model was applied to images of various environments.

3.1 YOLOv5[10]

YOLOv5 is an extended and improved version of YOLOv4 and includes mosaic data augmentation and auto-learning bounding box anchors. It can predict objects in 0.007 seconds per image and process an average of 140 FPS. In addition, while YOLOv3 has a high FPS(Frame per second) and a relatively low mAP (mean average precision). YOLOv5 shows excellent performance in terms of FPS and mAP.

There are 4 types of backbone of YOLOv5, the smallest and lightest Yolov5s (small), Yolov5m (medium), YOlov5l (large), and Yolov5x (xlarge). All four models of YOLOv5 have the same backbone and head, but are divided according to the model depth multiple and the number of channels per layer.

Figure 1 shows the performance comparison result of s,m,l,x of YOLOv5 and EfficientDet[18], which has excellent performance among object detection algorithms. The four models of YOLOv5 have better performance than EfficientDet. In YOLOv5, the s model is the fastest, but the accuracy is lower, and the x model is the slowest, but the accuracy is improved.

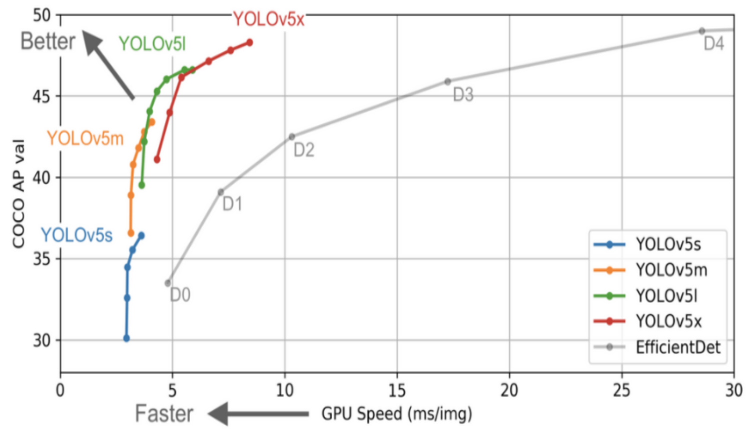


Figure 1. YOLOv5 performance[10]

In this paper, since real-time processing is targeted, the YOLOv5s model, which is the fastest model among the four models of YOLOv5, is used to determine whether or not a safety helmet is worn.

3.2 TRANSFER LEARNING

The structure of deep learning model is very complex. Therefore, as the number of training data is small, problems such as overfitting occur and its performance decreases. On the other hand, as the amount of data used for training of it increases, the performance of it is maximized[19]. Therefore, a transfer learning technique that trains data in a specific field with a pre-trained model in advance by abundant data in a similar field is widely used in various deep learning applications. A detailed description of transfer learning is introduced in [20]. Figure 2 is a basic conceptual diagram of the difference between the existing learning method and the transfer learning method.

In this study, the safety helmet is detected using transfer learning using the pre-trained Yolov5s model.

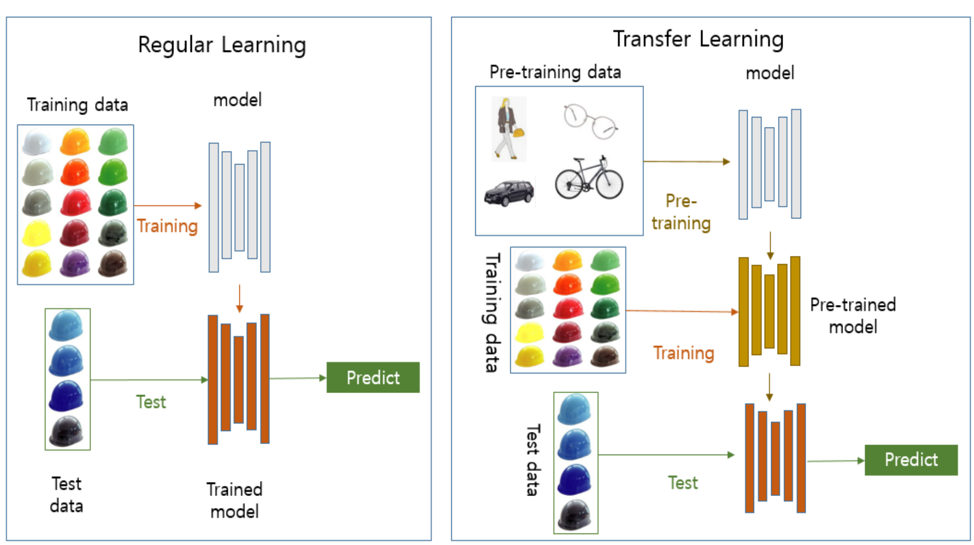


Figure 2. The difference of Regular learning and Transfer Learning

4. EXPERIMENTS AND ANALYSIS

YOLO divides the image into grids of the same size and judges the position and class of the object at once, so it is fast and easy to apply in real time. However, there was a disadvantage in that the accuracy was lowered and it was difficult to detect small objects. YOLOv5 is an improved technology of YOLO, which shows significantly improved performance with high speed and accuracy.

In this paper, we use the s-model of YOLOv5 to detect whether a helmet is worn.

4.1 Experimetal environments and datasets

Figure 3 shows the implementation process of safety helmet wearing detection in this study. The experimental environment was performed on Ubuntu 16.04.7 LTS and is shown in Table 1.

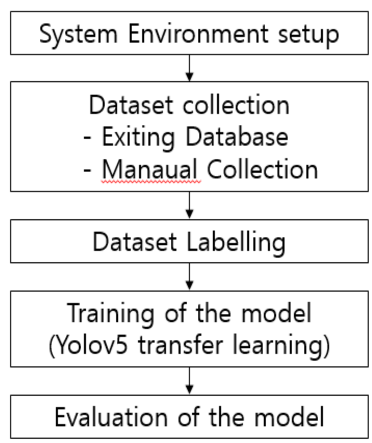


Figure 3. Implementation process of safety helmet wearing detection

Table 1. Experimental environment

OS	Ubuntu 18.04.5 LTS
GPU	Tesla V100-SXM2
CPU	16 core Intel(R) Xeon(R) Gold 5120 CPU @ 2.20GHz
RAM	177GB
CUDA	10.1
cuDNN	7.6.0
software	python/pytorch
Deep Learnig model	YOLOv5 S model

The dataset was used Kaggle's 'YOLO helemt/head' [21], It consists of 15887 training data, 4641 validation data, and 2261 test data. The class is labeled with a head and helmet, and the ground truth for each class is about 84,000 heads and 41000 helmets. In this paper, all heads not wearing safety helmet were labeled heads, and heads wearing safety helmet were labeled as heads. The ground truth area was set as a bounding box for the head area, including a human safety helmet, in order to avoid object detection when there is only a safety helmet.

4.2 Evaluation method

Mean Average Precision (mAP) was used as a criterion for evaluation of the experiment. mAP is an index used as an evaluation criterion in PASCAL VOC, and is the average of Average Precision (AP) for each classification class[22]. In general, the indicators used to evaluate the performance of an object detection model are precision and recall.

$$precision = \frac{True\ positive}{True\ positive + False\ positive} \tag{1}$$

$$Recall = \frac{True\ positive}{True\ positive + False\ negative} \tag{2}$$

The precision is in Eq. (1) and means the ratio of the case of detecting the true value to the total value of detecting the object. The recall is in Eq. (2) and refers to the ratio of cases in which the true value is detected to the value correctly detected. Since these two criteria generally have a negative correlation in each other, AP, which is defined as the area under the graph in the precision-detection rate graph, is used. The closer the AP is to 1, the higher the performance of the object detection algorithm.

4.3 Experimental result analysis

Experiments were compared by obtaining mAP according to batch-size and epoch at a learning rate of 0.01. Table 2 shows the experimental results of different epochs when the batch size is 64. The results in Table 2 show that the best performance is 0.9588 mAP at 100 epochs, and the mAP decreases as the epoch increases.

Table 2. mAPs by epochs for batch size 64

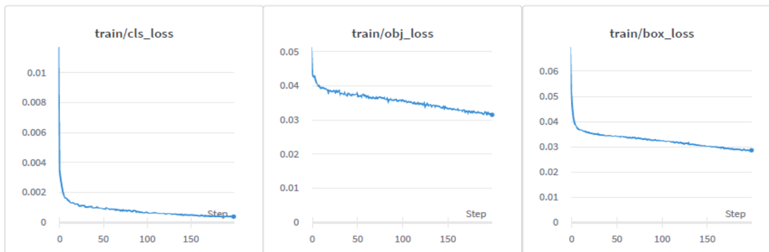
epochs	100	200	500	1000	2000
mAP	0.9588	0.9584	0.9563	0.9535	0.9526

Table 3 shows the results of the detection performance of wearing a safety helmet the batch sizes to 16, 32, 64, and 128 for the top three epochs in Table 2. 0.9588 mAP at batch size 64 for 100 epochs, 0.9593 mAP at batch size 32 for 200 epochs, and 0.9569 mAP at batch size 16 for 500 epochs, showed the best performance for each epoch. Among them, the case of batch at size 32 for 200 epoch showed the best result with 0.9593 mAP, and was selected as the optimal model of safety helmet wearing detectionl.

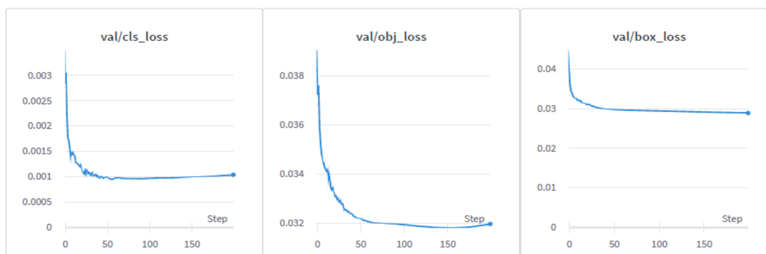
Table 3. mAPs by batch size in epochs 100/200/500

epochs	16	32	64	128
100	0.9584	0.9584	0.9588	0.9568
200	0.9591	0.9593	0.9584	0.9576
500	0.9569	0.9562	0.9563	0.9565

Figure 4 shows the class classification loss, object detection loss, and bounding box loss during training and validation. The class classification loss in training step shows a value lower than 0.001, and shows a value of 0.001 in validation step. This shows that it can be classified with little error. The object detection loss shows approximately 0.03 in training and validation step, and the bounding box loss also shows approximately 0.03 in training and validation step. This result shows that the object and its position can be detected almost accurately, although there is some error.



(a) train



(b) validation

Figure 4. class, object and box loss of train and validation

Figure 5 shows the results of detecting the head and helmet in various environments. In (a), it can be seen that the object is normally detected in the case of a person wearing a safety helmet, and that it is not detected in the case of a safety helmet that is not worn by a person. This is because, when setting the ground truth area, it does not specify the area for hard hats not worn by humans, so object detection is not performed. (b) is the result of detecting the head and helmet in a single object, and (c) and (d) show that multiple helmets and heads are detected well with relatively close objects. (e) and (f) are the detection results of a small helmet and head from a distance, showing that overlapping and distant objects are detected well.

(a) Detecting a helmet worn by a person (b) Helmet detection in a single object (c) Multiple helmet detection

(d) Multi-head detection (e) small helmet detection (f) small head detection

Figure 5. Head/helmet detection under various conditions

5. CONCLUSION

In this study, a model for determining whether or not wearing a safety helmet is selected and evaluated. First, transfer learning was performed by changing the batch size and learning rate of the YOLOv5s model. The performance of the trained model was evaluated by mAP, and it showed optimal performance when the batch size 32 and 200 epochs. Therefore, we select a model of 200 epochs and batch size 32 as a model for determining whether or not a safety helmet is worn. The selected model determined the case with and without a safety helmet well, and detected a small object wearing a safety helmet well. In addition, the case where a person wore a safety helmet and the case with only a safety helmet were classified well. In the future, we will conduct a study based on the processing speed and the size of the input image for a trade-off between speed and performance.

ACKNOWLEDGEMENT

This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government(MSIT) (No.2021-0-01972)

REFERENCES

[1] S. H. Moon, “A Study on ICT Conversion and Change of Industrial Society,” JCCT, Vol. 7, No. 4, pp.653-658, 2021.
[2] S. J. Moon, S. K. Cho, M. S. Jung, S. H. Park, “How to Respond to Complex Disasters on Future Megacities at the Government Level,” JCCT, Vol. 7, No.1, pp.211-215, 2021.
[3] M. H., Kim, H. S. Kong, “The Effects of Apartment Facility Maintenance on the Residential Satisfaction of Residents,” JCCT, Vol. 6, No. 3, pp. 175-183, 2020.
[4] Rules on OSH(Occupational Safety and Health) standards, Chapter 4.
[5] Y. W. Kim and K. S. Park, “A theoretical study on the shock-absorbing characteristics of safety helmet,” *Journal of the Ergonomics Society of Korea*, Vol. 9, No. 1, pp.29-33, 1990.
[6] M. W. Park, N. Elsafty and Z. Zhu, “Hardhat-wearing Detection for Enhancing On-site Safety of Construction Workers,” *J. Constr. Eng. Manag*. Vol. 141, No. 9, 2015.
[7] N.J. Kwak, D.J. Kim, “Object detection technology trend and development direction using deep learning,” IJACT, Vol.8, No. 4, pp.119-128, 2020.
[8] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only Look Once : Unified, Realtime Object Detection,” Proc. the IEEE Conf. on Comp. Vision and Pattern Recognition, pp.779-788, 2015.
[9] G. Liu, S. H. Lee, “Municipal waste classification system design based on Faster-RCNN and YoloV4 mixed model,” IJACT, Vol.9, No. 3, pp.305-314, 2021.
[10] https://github.com/ultralytics/yolov5
[11] S. Ren, K. He, R. Girshick, J. Sun, “Faster R-CNN: Towards Real-time Object Detection with Region Proposal Networks,” *IEEE Trans. Pattern. Anal. Mach. Intell*, Vol. 39, pp.1137–1149, 2017.
[12] J. Dai, Y. Li, K. He, J. Sun, “R-FCN: Object Detection via Region-based Fully Convolutional Networks,” Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems, Barcelona, Spain, pp. 379–387, 2016.
[13] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Y. Fu, A. C. Berg, “SSD: Single Shot MultiBox Detector,” In Computer Vision-ECCV 2016, Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2016.
[14] Q. Fang, H. Li, X. Luo, L. Ding, H. Luo,T. M. Rose, “An, W. Detecting non-hardhat-use by a deep

learning method from far-field surveillance videos," *Autom. Constr*. Vol. 85, pp. 1–9, 2018.

[15] J. Wu, N. Cai, W. Chen, H. Wang, G. Wang, "Automatic detection of hardhats worn by construction personnel: A deep learning approach and benchmark dataset," *Autom. Constr*., Vol. 106, 102894.

[16] J. Mistry, A. K. Misraa, M. Agarwal, A. Vyas, V. M. Chudasama, and K. P. Upla, "An automatic detection of helmeted and non- helmeted motorcyclist with license plate extraction using convolutional neural network." Proc. of 2017 Seventh International Conference on Image Processing Theory, Tools and Applications (IPTA), IEEE, pp. 1-6, 2017.

[17] J. Park, S. Jeon, J. Jeon, and J. Kim, "A study on deep learning based personal protective equipment detection," in Proc. 2020 The Korean Inst. of Broadcast and Media Eng. Summer Conf., pp. 650-651, online, Jul. 2020.

[18] M. Tan, R. Pang, Q. V. Le, "EfficientDet: Scalable and Efficient Object Detection," arXiv preprint arXiv:1911.09070v4, 2020.

[19] B. Liu, Y. Wei, Y. Zhang, and Q. Yang, "Deep neural networks for high dimension, low sample size data." Proc. of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17, Melbourne, pp. 2287-2293, 2017.

[20] L. Shao, F. Zhu, and X. Li, "Transfer learning for visual ategorization: A survey.*" IEEE Transactions on Neural Networks and Learning Systems*, Vol. 26, No. 5, pp. 1019-1034, 2015.

[21] Kaggle Helmet Dataset, https://www.kaggle.com/vodan37/yolo-helmethead

[22] M. Everingham, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes (voc) challenge." International Journal of Computer Vision, Vol. 88, No. 2, pp. 303-338, 2010.