*Computers, Materials & Continua*
DOI: 10.32604/cmc.2023.035762
***Article***

Tech Science Press

# Detection of Worker's Safety Helmet and Mask and Identification of Worker Using Deeplearning

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

[1]Department of Information Security at Paichai University, Daejeon, 35345, Korea
[2]POSTECH Institute of Artificial Intelligence, Pohang, 24257, Korea
*Corresponding Author: DongJu Kim. Email: kkb0320@postech.ac.kr

Received: 02 September 2022; Accepted: 14 December 2022

**Abstract:** This paper proposes a method for detecting a helmet for the safety of workers from risk factors and a mask worn indoors and verifying a worker's identity while wearing a helmet and mask for security. The proposed method consists of a part for detecting the worker's helmet and mask and a part for verifying the worker's identity. An algorithm for helmet and mask detection is generated by transfer learning of Yolov5's s-model and m-model. Both models are trained by changing the learning rate, batch size, and epoch. The model with the best performance is selected as the model for detecting masks and helmets. At a learning rate of 0.001, a batch size of 32, and an epoch of 200, the s-model showed the best performance with a mAP of 0.954, and this was selected as an optimal model. The worker's identification algorithm consists of a facial feature extraction part and a classifier part for the worker's identification. The algorithm for facial feature extraction is generated by transfer learning of Facenet, and SVM is used as the classifier for identification. The proposed method makes trained models using two datasets, a masked face dataset with only a masked face, and a mixed face dataset with both a masked face and an unmasked face. And the model with the best performance among the trained models was selected as the optimal model for identification when using a mask. As a result of the experiment, the model by transfer learning of Facenet and SVM using a mixed face dataset showed the best performance. When the optimal model was tested with a mixed dataset, it showed an accuracy of 95.4%. Also, the proposed model was evaluated as data from 500 images of taking 10 people with a mobile phone. The results showed that the helmet and mask were detected well and identification was also good.

**Keywords:** Mask; PPE; safety helmet; Yolo; Facenet

## 1 Introduction

In today's large-scale construction/manufacturing and other complexes, diverse, and unsafe industrial sites, workers' activities are always exposed to many risks anytime, anywhere. Therefore, there are more risk factors than in other industries, so the frequency of accidents is high, and it is stipulated that personal protective equipment (PPE) that protects the body of workers from hazardous factors must be worn [1]. One of the most common safety accidents in industrial sites is accidents caused by not wearing personal protective equipment such as safety helmets, safety rings, and safety boots. According to an analysis of the types of accidents at industrial sites in Korea, 37.3% of the total deaths occurred without wearing a helmet [2]. 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 [3]. However, in the actual field, there are many cases where it is not worn or worn properly for reasons such as stuffiness and inconvenience. Currently, the system for accident prevention in industrial sites is being implemented as a video monitoring control system using CCTV to understand the work situation of workers in real-time. It is a method that is controlled by the human eye through the video information collected from CCTVs installed in major places or that responds by detecting an abnormality set in advance.

These CCTV monitoring systems and safety management systems have a problem in that they are not suitable for prevention due to the limitations of human control capabilities (decreased reliability due to detection errors). Therefore, smart safety management with cutting-edge technology that can immediately identify the situation of workers in the industrial site in real-time and manage them is required. 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 CCTVs in the workplace. Recently, with the development of deep learning technology, an automation method for preventing worker accidents by image recognition technology using deep learning is being actively studied. In a study regarding personal protective equipment detection, faster regions with convolutional neural networks features (faster R-CNN) [4] algorithm for the identification of wearing safety equipment in the construction safety field was applied to detect workers and equipment at the construction site to predict the possibility of collision [5].

A region-based object detection and classification algorithm based on a convolutional neural network (CNN) is a study to improve the safety of construction site workers by establishing an effective automatic safety helmet detection system using an object detection and classification algorithm based on construction site image data. Region-based fully convolutional networks (R-FCN) [6] were applied and were performed using the transfer learning technique [7]. Faster R-CNN and R-FCN are object recognition algorithms, which are two-step networks consisting of local proposal generation and proposal classification processes. The two-step object recognition algorithm has good object detection performance but is not suitable for real-time processing because of its slow speed. Among the object detection networks, the one-step network, which directly predicts the object bounding box and its class, has a lower detection rate than the two-step network but has a faster processing speed. As representative models of a one-step network include, we can take the examples as You Only Look Once (YOLO) [8] and Single Shot Multibox Detector (SSD) [9]. Because YOLO has a faster prediction speed as compared to other object detection algorithms and it is being used in various real-time object detection applications as well. However, there are some disadvantages which are, the prediction performance of objects being poor and small objects are not detected well. YOLOv5 [10] is a recently announced algorithm that can detect even small objects at high speed and is classified into four models: s, m, l, and x according to speed and performance. In this study, the s-model and m-model, which are fast among the four models, are transfer-learned for real-time processing. The two models change the learning rate, batch size, epoch, etc., and several models are created, and each model's performance is measured with mean average precision (mAP). The best model is selected by comparing the performance of each model. Currently, COVID-19 is prevalent around the world, and wearing a mask for prevention has become an important part of our lives. Wearing a mask indoors is an important factor to prevent from virus infection, so the mask on a face is detected when entering the workplace. The mask detection model is trained and evaluated at the same time as the helmet when generating the helmet detection model to select an optimal model.

Before the worker enters the workplace, the identification of the worker must be simultaneously confirmed for security, along with the confirmation of wearing personal protective equipment for personal safety. Current identity recognition systems are based on the face recognition method. Facial recognition systems based on deep learning have demonstrated excellent accuracy [11–13]. The accuracy of these systems depends on the characteristics of the available training images, and the existing systems learn important facial features such as eyes, nose, lips, face edges, etc. When learning facial features, face occlusion with glasses, masks, etc. becomes an obstacle to learning important facial features and causes serious performance degradation in the identification system. Therefore, if we are wearing a mask with the existing face recognition system, we have to take it off when verifying our identity. This not only makes the authentication process cumbersome for users but also increases the risk of virus transmission. Therefore, it is necessary to develop a system that allows identification even when wearing a mask. In this paper, we propose a system that can confirm identity while wearing a mask. The algorithm for extracting facial features while wearing a mask is generated by transfer learning of Facenet using a masked face dataset containing only masked faces and a mixed face dataset containing both masked and unmasked faces. The optimal model is selected by comparing the performance of the two models generated by the two datasets. For worker identification, the SVM classifier [14] is trained by applying the dataset from which the Facenet was trained and tested.

Therefore, in this paper, a system that can detect a helmet and mask and confirm workers' identities by face recognition before they enter the workplace is proposed. The proposed system combines the confirmation of compliance with the wearing of a safety helmet and mask and the confirmation of workers' identity in real-time. The paper is organized as follows: Section 2 describes the related work on which this study is based. Section 3 explains the proposed method and analyzes the experimental results. Section 4 summarizes the proposed method and performance.

## 2 Related Work

### *2.1 YOLOv5*

The YOLO algorithm [8] is one of the deep learning algorithms for object detection. As compared to other deep learning-based object detection algorithms, YOLO shows fast processing speed and YOLOv5 [10] is a recently announced algorithm, and the backbone, which plays a role in extracting important features from the input image, improves the learning ability by combining BottleNeck [15] and cross stage partial network (CSPNet) as well [16]. There are four types of backbones of YOLOv5 scuh as 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 (layer channel multiple).

The neck of the model that mixes the functions formed in the backbone uses the path aggregation network (PA-Net) [17] feature pyramid method. The model head that performs object detection follows the structure of the existing YOLOv4 [18] model. Fig. 1 shows the performance comparison results of s, m, l, x of YOLOv5 and EfficientDet [19], which has excellent performance among object detection algorithms, using the COCO dataset [20]. The four models of YOLOv5 have better performance than EfficientDet. As the characteristics of these models, the s-model is the fastest but has poor accuracy, and the x-model is the slowest but has improved accuracy.

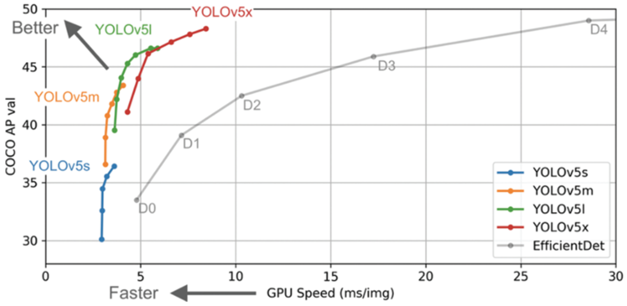


**Figure 1:** Performance of Yolov5

## 2.2 *Face Recognition*

Face recognition is a technology studied to identify people and is largely performed as face recognition or verification. Face verification is a 1:1 verification problem that determines whether two face images coming in as inputs are the same person. Face identification can be viewed as a 1:N problem in which one input face image corresponds to which of the N people registered in advance. Fig. 2 shows the flow of the face recognition system.

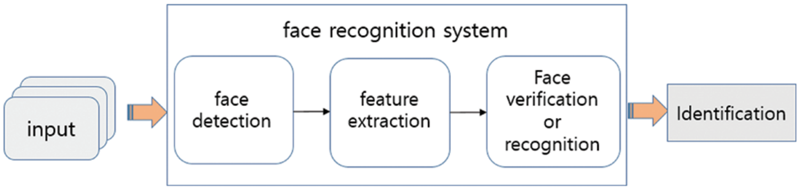


**Figure 2:** Flow of face recognition system

In the face recognition technology, features such as histogram of oriented gradient (HOG) [21], local binary pattern (LBP) [22], and Gabor [23], which were the main hand-crafted features in the face recognition field, have all been changed to deep learning-based features, and the face detection method has also been changed to a deep learning-based method, which improves the performance that was not possible. Face recognition can be divided into face region extraction, feature extraction from a face region, and recognition or verification through matching. The face recognition system receives multiple face images in advance to be trained and identifies the user by inputting images for user identification into the trained model. Face detection is to detect a face region from an input image, and the representative algorithm is multitask cascaded convolutional networks (MTCNN) [24], which passes through three CNNs, P-net, R-net, and O-net in turn. Facenet is a feature extractor that extracts facial features from detected face images as input. Facenet extracts feature from a face image and create a 128-element vector (face embedding vector) and verifies the identity using the distance between these embedding vectors or uses the embedded vector as an input to the classifier to verify the identity.

## 2.3 *SVM*

A support vector machine (SVM) is one of the fields of machine learning and is a supervised learning model for pattern recognition and data analysis, and is mainly used for classification and regression analysis. With a set of data belonging to either of the two categories, the SVM algorithm creates a non-stochastic binary linear classification model that determines which category the new data belongs to based on the given dataset. The classification model is expressed as a boundary in the space where the data is mapped, and the SVM algorithm is an algorithm that finds the boundary with the largest width. This separation boundary is also called a decision boundary which is a straight line in two dimensions. A decision boundary is a plane that cannot be visualized and has more than two dimensions which are called a hyperplane. The distance between the decision boundary and the support vector is called the margin, and the data that affect the decision of the margin are called support vectors. The optimal decision boundary of the support vector is the boundary that maximizes the margin. SVM shows good performance among classification algorithms as well. The margin data of SVM is shown in Fig. 3.

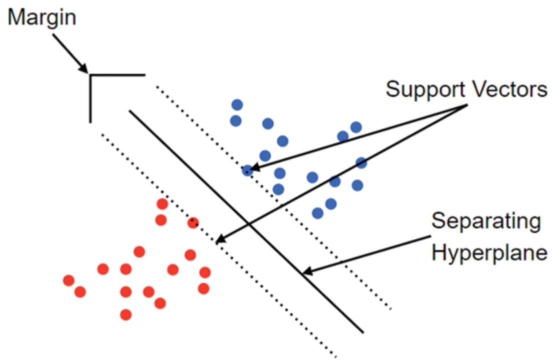


**Figure 3:** Support vector, margin, hyperplane of SVM

## 3 The Proposed Method

In this study, we implement a system that detects safety helmets and masks of the person and checks the his or her identity. For the detection of safety helmets and masks, Yolov5's s-model and m-model are transfer-learned, and their performance is evaluated to select the optimal model for object detection. The model of facial feature extraction for identification was derived by transfer learning of Facenet by inputting masked faces and mixed faces (masked faces + unmasked faces). The SVM classifier was trained using the embedded facial features for worker identification. Fig. 4 shows the system structure of the proposed method. The experiment was conducted in the environment shown in Table 1 on Ubuntu 16.04.7 LTS environment.

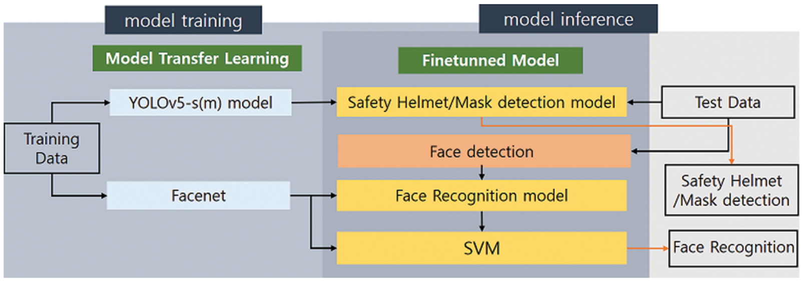


**Figure 4:** The system structure of the proposed method

**Table 1:** Test environments

| | |
|---|---|
| OS | Ubuntu 18.04.5 LTS |
| GPU | Tesla V100-SXM216 core Intel (R) |
| CPU | Xeon (R) Gold 5120 CPU @ 2.20 GHz |
| RAM | 177 GB |
| CUDA | 10.1 |
| cuDNN | 7.6.0 |
| Software | python/pytorch<br>YOLOv5/Facenet |

### *3.1 Safety Helmet/Mask Detection Experiment and Result Analysis*

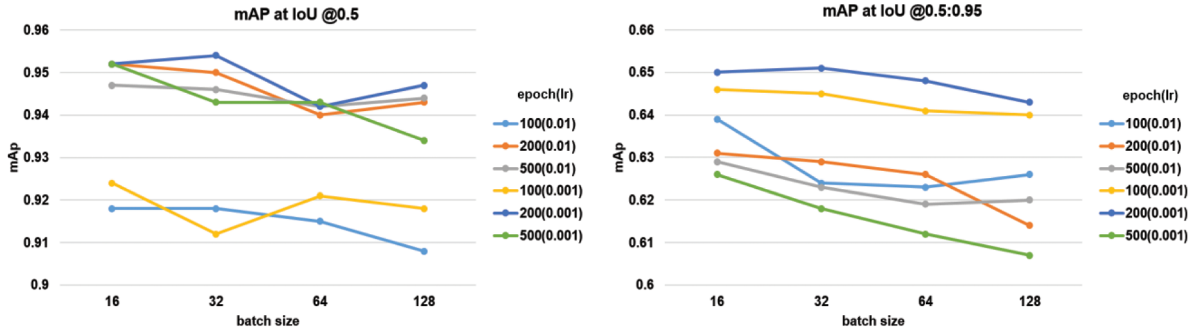
The safety helmet and mask detection models are trained by Kaggle's face mask dataset [25], Kaggle's PPE dataset [26], and data collected from the web. The data collected from the web were labeled as helmet/mask/head using a labeling tool [27]. The data used in this work are 5500 training data, 1500 validation data, and 1000 test data that gives total data of 8000.

In this study, in order to generate a model with optimal safety helmet/mask detection performance, mAP is obtained by changing the learning rate, batch size, and epoch, and the model with the best performance of the mAP is selected as the safety helmet/mask detection model. The learning rates were 0.01 and 0.001, the batch sizes were 16, 32, 64, and 128, and the epochs were 100, 200, and 500. Among the yolov5 models, the results were analyzed for the s-model and the m-model in consideration of the speed and performance for real-time processing.

Fig. 5 shows the experimental results of the Yolov5s model, and Fig. 6 shows the experimental results of the Yolov5m model. In Figs. 5 and 6, (a) is the case of IoU @0.5 and (b) is the case of IoU @0.5:0.95. In the s-model, at IOU @0.5, the results of the learning rate of 0.001 and batch sizes of 32 and 200 epochs were mAP of 0.954, which was the best, and in the m-model, the results of the learning rate of 0.001, 500 epochs, and the 64 batch size was mAP of 0.952, which was the best. In the case of IOU @0.5:0.95, the performance of the s-model was the best with a learning rate of 0.001 and a batch size of 32 and 200 epochs, mAP of 0.651. The performance of the m-model, at the learning rate of 0.001 and the batch size of 64 and 500 epochs, was the best with mAP of 0.651. In the results, there is no significant difference in the mAP results of the m model and the s model, but considering the speed, the s-model is selected as the optimal mode1. Also, among the results of @0.5 and @0.5:0.95, the optimal model is determined based on @0.5. Therefore, the s-model with a learning rate of 0.001, a batch size of 32, and an epoch of 200 based on @0.5 are selected as the optimal model. The determined model is shown in Fig. 7 which is the result of detecting the safety helmet/mask/head (head without both helmet and mask).

Fig. 7 shows that it detects well safety helmet/mask in a single image and multiple images. In addition, it detects safety helmets/masks well even in small-size images.

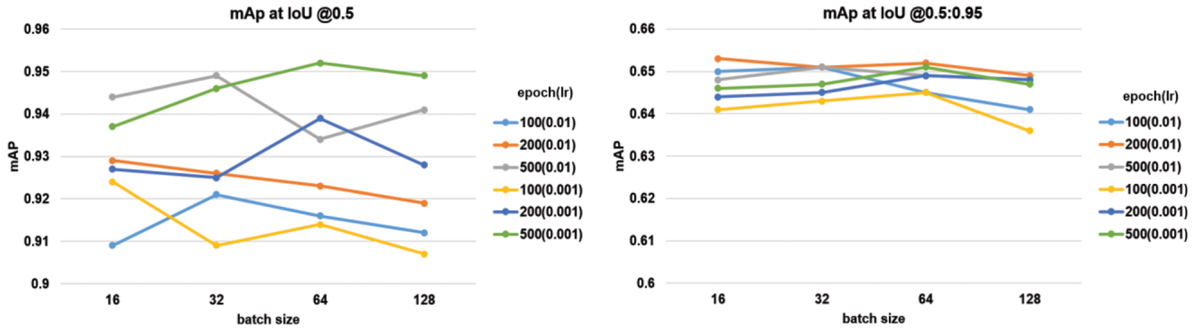
To verify the speed and the detection performance of the selected model, a mAP and an fps of YOLOv5s model and Yolov4 were measured at 200 epochs and @0.5. Table 2 shows that Yolov4 has 0.001 higher mAP than Yolov5s model, but there is a large difference in fps. Considering the speed and performance, the Yolov5s model is the best choice.



(a) mAP of IoU @0.5

(b) mAP of IoU@ 0.5:0.95.

**Figure 5:** mAP according to learning rate, batch size, and epoch (Yolov5s model)



(a) mAP of IoU @0.5

(b) mAP of IoU@ 0.5:0.95.

**Figure 6:** mAP according to learning rate, batch size, and epoch (Yolov6m model)

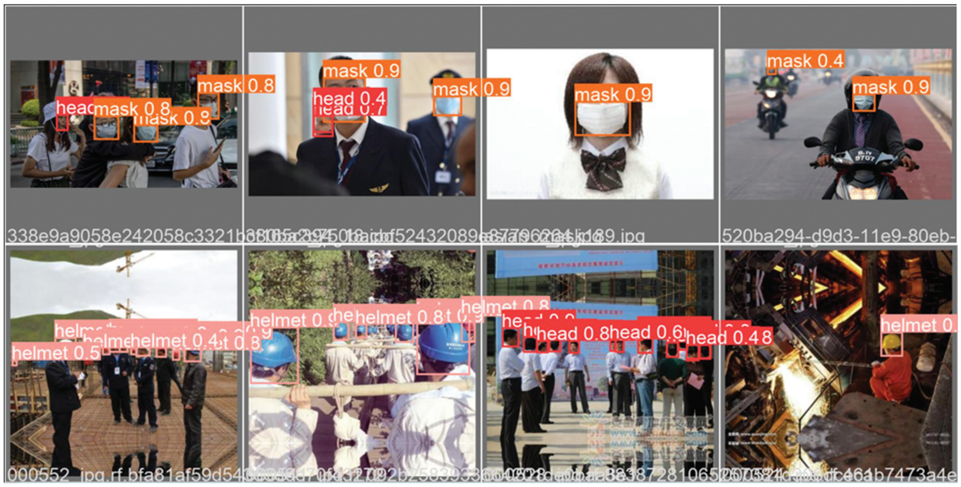


**Figure 7:** Safety helmet/mask/head detection result

**Table 2:** mAP and fps of Yolov4 and Yolov5-s model

| Model | Size | mAP | fps |
|---|---|---|---|
| Yolov4 | 512 | 0.944 | 95.7 |
| Yolov5s | 640 | 0.943 | 104.2 |

## *3.2 Identity Verification Experiment and Result Analysis*

### *3.2.1 Implementation of Dataset and Model*

In this paper, facial features are generated using Facenet among models for face recognition, and the identity of the worker is verified using the SVM classifier. The dataset is a masked face dataset [28] and a mixed dataset composed of a mixture of masked face [28] plus unmasked face [29], and the performance of models are compared after training two models using two datasets.

The masked face dataset in [28] was made by a tool call MaskTheFace [30] to simulate masked facial images based on some famous face image datasets. MaskTheFace is a computer vision-based tool to mask faces in images. It uses a dlib based face landmarks detector to identify the face tilt and six key features of the face necessary for applying a mask. Based on the face tilt, the corresponding mask template is selected from the library of the mask. The template mask is then transformed based on the six key features to fit perfectly on the face. In [28], several different masks are chosen such as green surgical mask (#44b4a8), blue surgical mask (#1ca0f4), white N95 mask (#FFFFFF), white KN-95 mask, and black cloth mask (#000000). The masked face creation procedure by the tool is shown in Fig. 8.

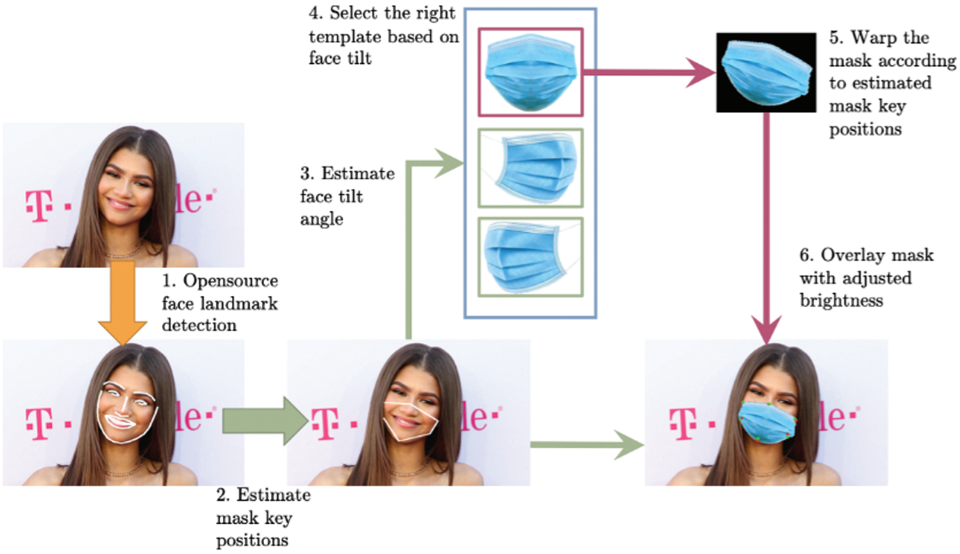

4. Select the right template based on face tilt
5. Warp the mask according to estimated mask key positions
3. Estimate face tilt angle
1. Opensource face landmark detection
6. Overlay mask with adjusted brightness
2. Estimate mask key positions

**Figure 8:** Masked face creation procedure by tool

In this paper, we used the LFW (eval plus test) dataset among the various datasets in [28] as a masked face dataset. Some of these data were chosen to compose the training data and to measure the performance of the trained model, data that did not overlap with the training data was selected to compose the test data. The masked face dataset consists of 7800 data and 446 classes. Fig. 9 is an example of a masked face dataset.

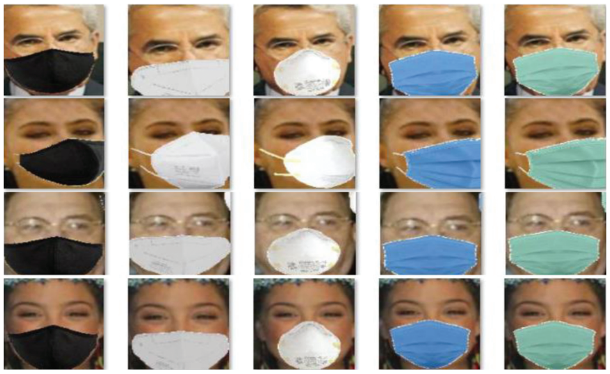


**Figure 9:** Examples of masked-face dataset

The mixed face dataset is made from the masked face dataset in [28] and the unmasked LFW face dataset in [29]. At this time, masked face data and unmasked face data are extracted at the same rate. The mixed face dataset has 446 classes and 9850 data. Fig. 10 is an example of a mixed face dataset.

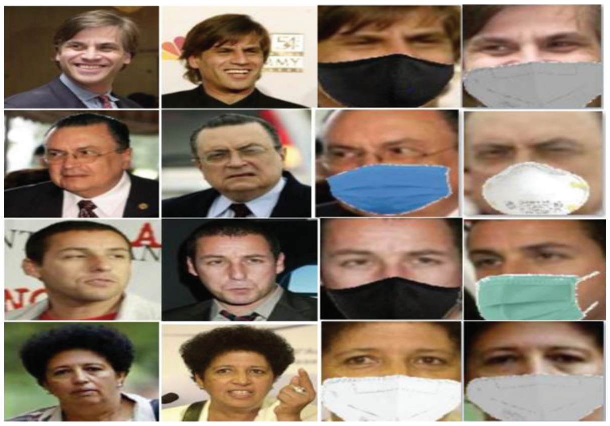


**Figure 10:** Examples of mixed face dataset

Facial feature extraction uses models trained by learning Facenet using a masked face dataset and a mixed face dataset. 20% of the entire training dataset is used as the validation dataset, and the facial feature extraction model is created by changing the learning rate, batch size, and epoch. Two models with good performance are selected from the model created with the masked face dataset and the model created with the mixed face dataset, respectively, and the performance of the models is evaluated with the test dataset. For each model, the test is performed as a masked face dataset and a mixed face dataset. Among the test results, we selected two models with good performance.

The worker's identity verification models also are created by learning the SVM classifier with the masked face dataset and the mixed face dataset. The SVM classifier is trained by changing the learning rate, batch size, and epoch using the same dataset as when learning the Facenet. The input of the classifier is the facial features generated by two selected models among the Facenet models. Among them, the model with the best performance is selected as the optimal model. Table 3 summarizes the model-selecting process using the model and dataset.

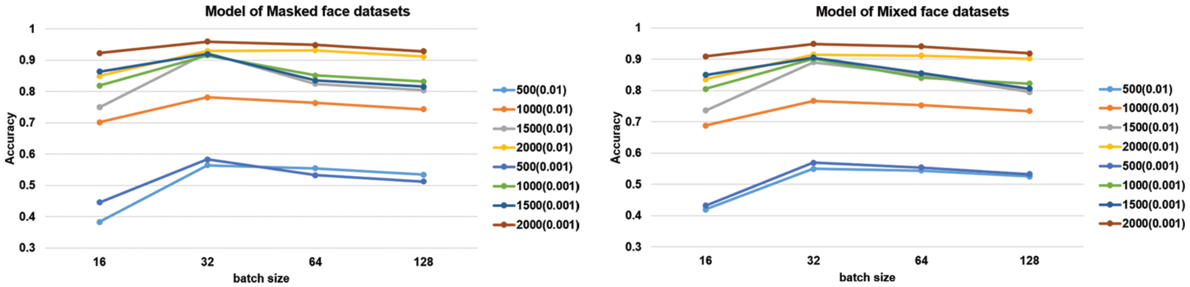
**Table 3:** Model summary

| Model | Training data | Test data |
|---|---|---|
| Facenet | Masked face dataset | Masked face dataset<br>Mixed face dataset |
| | Mixed face dataset | Masked face dataset<br>Mixed face dataset |
| Facenet+SVM | Masked face dataset | Masked face dataset<br>Mixed face dataset |
| | Mixed face dataset | Masked face dataset<br>Mixed face dataset |

### *3.2.2 Test Results of Feature Extraction Model*

The experiment of feature extraction model was performed at learning rates of 0.01 and 0.001, and batch sizes of 16, 32, 64, and 128, epochs of 500, 1000, 1500, and 2000 and 20% of the total training data was used as validation data.

Fig. 11 shows the accuracy of the validation data after learning the Facenet using the masked face dataset and the mixed face dataset. In this paper, two models of the masked face dataset and two models of the mixed face dataset were selected with good performance among the models. The four selected models were evaluated using the test dataset, and two models with good performance are selected among them and combined with the SVM model.



(a) Accuracy of model by masked face dataset.

(b) Accuracy of model by mixed face dataset.

**Figure 11:** Accuracy according to learning rate, batch size, and epoch

Both models (the masked face dataset and the mixed face dataset) achieved the best results at epochs of 2000, learning rate of 0.001, and batch sizes of 32 and 64 as shown in Fig. 11. In this paper, we give new names to the four models as model1~model4. Table 4 is a summary of the new names, learning conditions, and accuracy for the four models.

The performance of the selected four models was evaluated with the test dataset. The test dataset consists of a masked face dataset and a mixed face dataset, and the evaluation results are shown in Table 5.

**Table 4:** Summary of selected models from trained Facenet

| Model name | Training dataset | Learning rate | Epoch | Batch size | Accuracy |
|---|---|---|---|---|---|
| Model1 | Masked face dataset | 0.001 | 2000 | 32 | 0.959 |
| Model2 | Masked face dataset | 0.001 | 2000 | 64 | 0.948 |
| Model3 | Mixed face dataset | 0.001 | 2000 | 32 | 0.949 |
| Model4 | Mixed face dataset | 0.001 | 2000 | 64 | 0.941 |

**Table 5:** Evaluation results of the four models

| Model | Test dataset | |
|---|---|---|
| | Masked face dataset | Mixed face dataset |
| Model1 | 0.947 | 0.933 |
| Model2 | 0.941 | 0.924 |
| Model3 | 0.939 | 0.941 |
| Model4 | 0.928 | 0.933 |

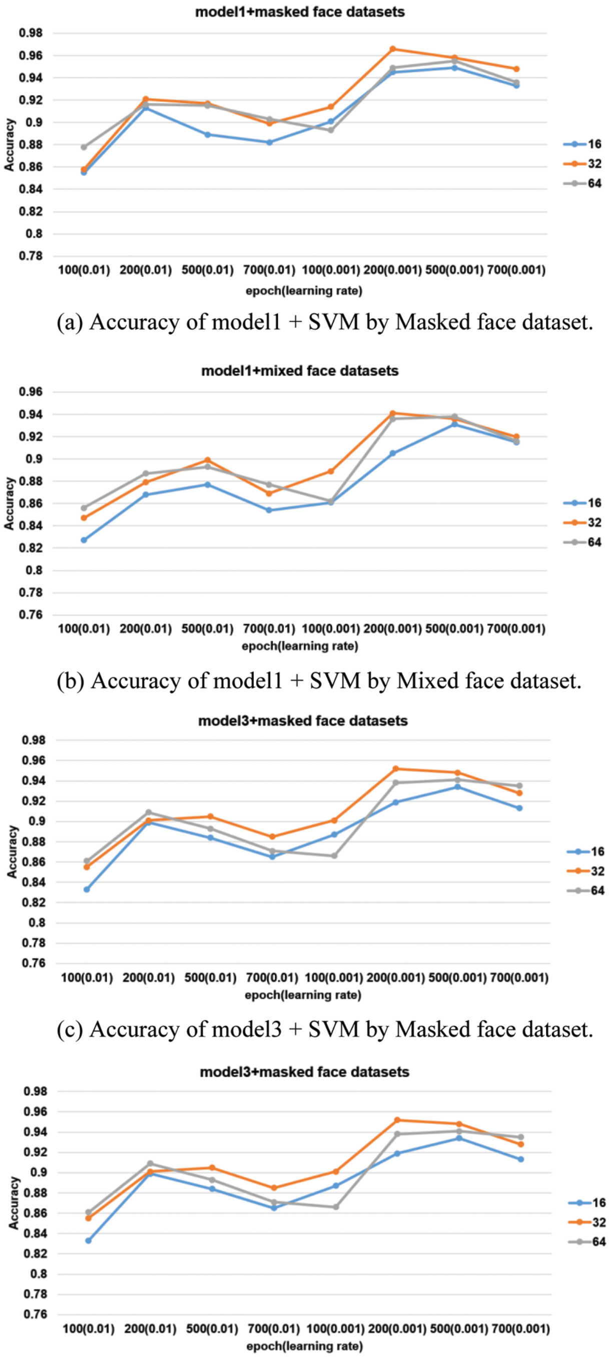
In Table 5, the results of model1 and model2 are the best for the masked face dataset, the results of the mixed face dataset are the best for model3, and the results of model1 and model4 are the same. We need a model that can identify both masked and unmasked faces, so we choose models with good performance on the mixed dataset. Therefore, we have selected a model1 and model3.

### *3.2.3 Test Results of SVM Model*

In this paper, we combine model1 and model3 selected in the previous section with the SVM classifier, train the combined models with the masked face dataset and the mixed face dataset, and then select the optimal model for identity recognition by evaluating the performance. During training, model1 and model3 were not trained, and only the SVM classifier was trained with learning rates of 0.01 and 0.001, and batch sizes of 16, 32, and 64, epochs of 100, 200, 500 and 700. Accuracy was used to evaluate the performance of each model.

Fig. 12 shows the performance of each model. In the Fig. 12, the four models show the best performance at a learning rate of 0.001 and batch sizes of 32 and 200 epochs. Also, the SVM classifier trained with the model3 and mixed face dataset showed the best performance with an accuracy of 0.975.

Among the four models, the model with the best performance was selected one by one. The performance of the selected four models was evaluated with the test dataset. Table 6 shows the evaluation results.



(a) Accuracy of model1 + SVM by Masked face dataset.

(b) Accuracy of model1 + SVM by Mixed face dataset.

(c) Accuracy of model3 + SVM by Masked face dataset.

(d) Accuracy of model3 + SVM by Mixed face dataset.

**Figure 12:** Accuracy according to learning rate, batch size, and epoch

**Table 6:** Evaluation results of the four models

| Model | Test dataset | |
|---|---|---|
| | Masked face dataset | Mixed face dataset |
| Model1+SVM by Masked face dataset | 0.948 | 0.912 |
| Model1+SVM by Mixed face dataset | 0.911 | 0.929 |
| Model3+SVM by Masked face dataset | 0.942 | 0.938 |
| Model3+SVM by Mixed face dataset | 0.941 | 0.954 |

In Table 6, the SVM classifier-trained mixed face dataset with model3 shows the best performance with an accuracy of 0.954 which is similar to the training results. Therefore, in this paper, the SVM classifier trained the mixed face dataset with model3 is determined as the optimal model.

Fig. 13 is the results of identification using the selected model. Both the images with the mask and the images without the mask were properly identified.

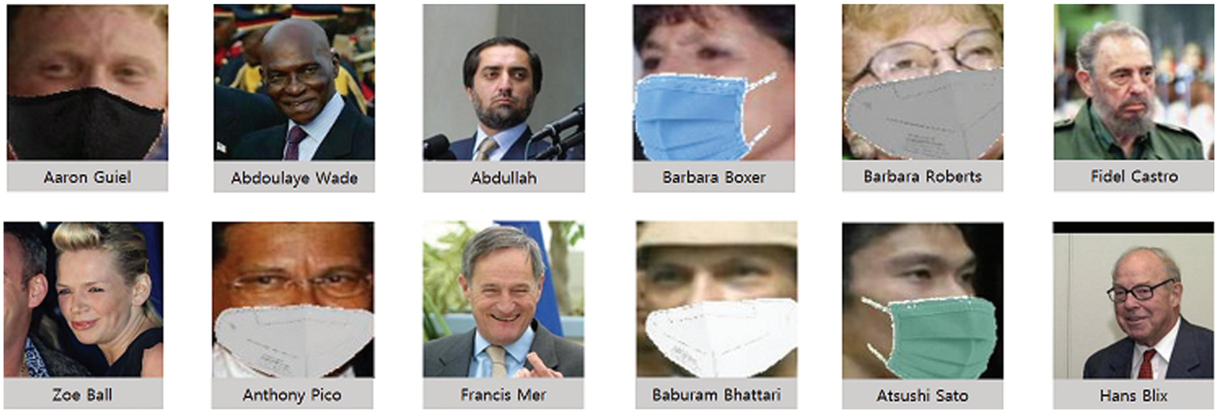


**Figure 13:** The results of identification using the selected model

### *3.3 Test of Overall System Combined with Safety Helmet/Mask Detection and Identification*

The proposed system was evaluated as data composed of 500 photos by taking 50 photos of 10 people each with a mobile phone. Since the number of classification classes is different, only the SVM classifier in the whole system is transfer-learned with the test dataset. Only 20% of the total data was used for testing. Table 7 shows the results of Accuracy, Precision, Recall, and F1 score of the SVM classifier of randomly extracted class from the test data.

The average values of accuracy, precision, and recall were found 0.98, 1.0, and 0.98 were found respectively out of 100 images. Fig. 14 shows the results of safety helmet/mask detection and identification using photos taken with a smartphone and the results shows that safety helmet/mask detection is also good and identity is accurately identified.

**Table 7:** Accuracy, Precision, Recall, and F1 score of the SVM classifier

| Class no | Accuracy | Precision | Recall | F1 score |
|---|---|---|---|---|
| 1 | 0.965 | 1.0 | 0.981 | 1.0 |
| 4 | 1.0 | 1.0 | 1.0 | 1.0 |
| 5 | 0.965 | 1.0 | 0.981 | 1.0 |
| 7 | 1.0 | 1.0 | 1.0 | 1.0 |
| 8 | 1.0 | 1.0 | 1.0 | 1.0 |

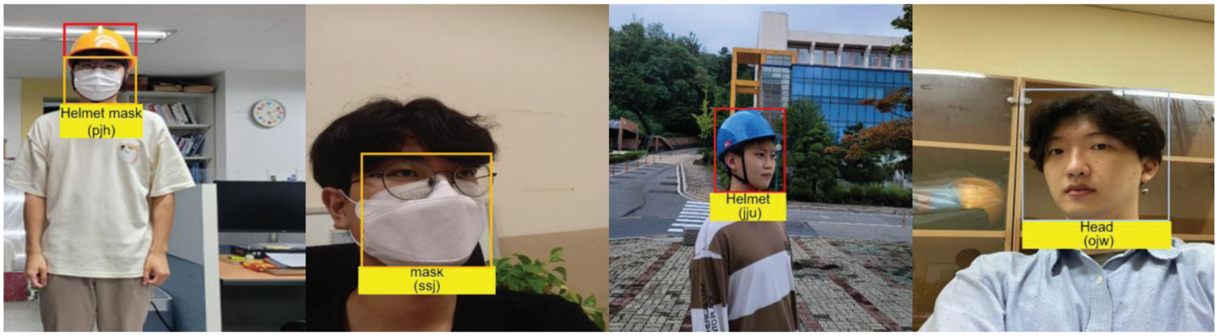


**Figure 14:** The results of safety helmet/mask detection and identification by images taken with a smartphone

## 4 Conclusion

In this paper, a model for safety helmet/mask detection and a model for worker identification is proposed and evaluated. The model for safety helmet and mask detection is generated by transfer learning of Yolov5's s-model and m-model. The two models were transfer-learned by changing the learning rate, batch size, and epoch, and their performance was measured with mAP, and the model with the best performance was selected as the optimal model. The learning rates were 0.01 and 0.001, the batch sizes were 16, 32, 64, 128, and the epochs were 100, 200, and 500. Among the yolov5 models, the results were analyzed for the s-model and the m-model in consideration of the speed and performance for real-time processing. At a learning rate of 0.001, a batch size of 32, and an epoch of 200, the s-model showed the best performance with a mAP of 0.95, and this was selected as the object detection model.

The identification part of the worker was composed of a facial feature extraction part and a classifier part for identification. Facial feature extraction is generated by transfer learning of Facenet, and the classifier for identification uses the SVM classifier. The dataset uses a masked face dataset containing only masked faces and then a mixed face dataset containing both masked and unmasked faces. The optimal model is selected by comparing the performance of the models generated by the two datasets. The combined model of Facenet and SVM classifier trained by the mixed face dataset showed the best performance. When this model was tested on the mixed face dataset, it showed an accuracy of 95.4%. The selected object detection model and identification model were tested with data from 10 people photographed with a smartphone. As a result, it was confirmed that safety helmets and masks were detected well with confirmation of identification.

This study shows good performance in detecting masks or helmets. However, protective equipment is more diverse than helmets. Therefore, research on algorithms for detecting various protective devices in the future can be conducted.

**Acknowledgement:** This research was supported by a grant (20015427) of Regional Customized Disaster-Safety R&D Program, funded by Ministry of Interior and Safety (MOIS, Korea), and was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (No. 2022R1A6A1A03052954).

**Funding Statement:** The authors received no specific funding for this study.

**Conflicts of Interest:** The authors declare that they have no conflicts of interest to report regarding the present study.

## References

[1] Ministry of Employment and Labor, "Industrial Safety Act, standards for wearing personal protective equipment, safety and health regulations Article 32", 2021.

[2] S. H. Jeung, Y. S. Lee and C. E. Kim, "Improved system for establishing a culture to wear personal protective Gear," *Journal of the Korea Institute of Construction Safety*, vol. 2, no. 1, pp. 16–20, 2019.

[3] 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.

[4] S. Ren, K. He, R. Girshick and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, 2017.

[5] D. S. Kim, J. S. Kong, J. H. Lim and B. C. Sho, "A study on data collection and object detection using faster R-CNN for application to construction site safety," *Journal of the Korean Society of Hazard Mitigation*, vol. 20, no. 1, pp. 119–126, 2020.

[6] J. Dai, Y. Li, K. He and J. Sun, "R-FCN: Object detection via region-based fully convolutional networks," in *Advances in Neural Information Processing Systems 29: in Proc. NIPS (Neural Information Processing Systems)*, Barcelona, Spain, pp. 379–387, 2016.

[7] S. Y. Park, S. H. Yoon and H. Joon, "Image-based automatic detection of construction helmets using R-FCN and transfer learning," *Journal of the Korean Society of Civil Engineers*, vol. 39, no. 3, pp. 399–407, 2019.

[8] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You only look once: Unified, real-time object detection," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, Nevada, USA, pp. 779–788, 2016.

[9] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed *et al.,* "SSD: Single shot multibox detector," in *European Conf. on Computer Vision (ECCV)*, Amsterdam, Netherlands, pp. 21–37, 2016.

[10] Ultralytics. YOLOv5. [Online]. Available: https://github.com/ultralytics/yolov5.

[11] F. Schroff, D. Kalenichenko and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA, pp. 815–823, 2015.

[12] V. Paul and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Kauai, Hi, USA, vol. 1, pp. 511–558, 2001.

[13] H. Li, Z. Lin, X. Shen, J. Brandt and G. Hua, "A convolutional neural network cascade for face detection," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA, pp. 5325–5334, 2015.

[14] Y. Li, J. Li and J. S. Pan, "Hyperspectral image recognition using SVM combined deep learning," *Journal of Internet Technology*, vol. 20, no. 3, pp. 851–859, 2019.

[15] J. Park, S. Woo, J. Y. Lee and I. S. Kweon, "Bam: Bottleneck attention module," arXiv preprint arXiv:1807.06514, 2018.

[16] C. Y. Wang, H. Y. M. Liao, Y. H. Wu, P. Y. Chen, J. W. Hsieh *et al.,* "CSPNet: A new backbone that can enhance learning capability of CNN," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) Workshops*, virtual site (https://www.kitware.com//demos/cvpr-2020-papers/workshops.html), pp. 390–391, 2020.

[17] S. Liu, L. Qi, H. Qin, J. Shi and J. Jia, "Path aggregation network for instance segmentation," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Salt Lake City, UT, USA, pp. 8759–8768, 2018.

[18] A. Bochkovskiy, C. Y. Wang and H. Y. M. Liao, "YOLOv4: Optimal speed and accuracy of object detection," arXiv:2004.10934v4, 2020.

[19] M. Tan, R. Pang and Q. V. Le, "EfficientDet: Scalable and efficient object detection," arXiv preprint arXiv:1911.09070v4, 2020.

[20] T. Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona *et al.,* "Microsoft coco: Common objects in context," in *European Conf. on Computer Vision (ECCV)*, Zurich, Switzerland, vol. 8693, pp. 740–755, 2014.

[21] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc CVPR*, San Diego, California, USA, pp. 886–893, 2005.

[22] T. Ojala, M. Pietikäinen and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971–987, 2002.

[23] J. G. Daugman, "Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters," *The Journal of the Optical Society of America A*, vol. 2, no. 7, pp. 1160–1169, 1985.

[24] K. Zhang, Z. Zhang, Z. Li and Y. Qiao, "Joint face detection and alignment using multitask cascaded convolutional networks," *IEEE Signal Processing Letters*, vol. 23, no. 10, pp. 1499–1503, 2016.

[25] Kaggle Mask Dataset. [Online]. Available: https://www.kaggle.com/aditya276/face-mask-dataset-yolo-format.

[26] Kaggle Helmet Dataset. [Online]. Available: https://www.kaggle.com/vodan37/yolo-helmethead.

[27] Labeling Tools. [Online]. Available: https://github.com/tzutalin/labelImg.

[28] Masked dataset. [Online]. Available: https://github.com/SamYuen101234/Masked_Face_Recognition.

[29] G. B. Huang, M. Ramesh, T. Berg and E. L. Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," Technical Report 07-49, University of Massachusetts, Amherst, 2007.

[30] A. Anwar and A. Raychowdhury, "Masked face recognition for secure authentication," arXiv:2008.11104, 2020.