ISSN 1816-6075 (Print), 1818-0523 (Online)
Journal of System and Management Sciences
Vol. 12 (2022) No. 2, pp. 367-377
DOI: 10.33168/JSMS.2022.0219

Implementation of Access Control System based on Face Prediction and Face Tracking

Bong-Hyun Kim

Professor, Dept. of Computer Engineering, Seowon University, Korea
bhkim@seowon.ac.kr

Abstract. Problems caused by COVID-19, which have been continuing for two years from 2020, are constantly being raised. In particular, time is wasted in measuring body temperature and verifying occupants every time they enter a building. In the existing access control method, manpower is placed at each entrance to check whether or not a mask is worn or not. Accordingly, there is a problem that there is a risk of wasting manpower and infection, and recognition takes a long time. Therefore, it is necessary to manage building access more effectively by providing a non-face-to-face environment suitable for the intact era. Therefore, in this paper, we designed and developed a system to control access by examining the condition of whether or not a mask is worn when entering a building. For this purpose, a face was detected using the Single Shot Multibox Detector (SSD) algorithm, which has a better recognition rate than the existing face detection algorithm. In addition, the detected face image was input to the deep learning model to examine whether it was worn. Finally, a system was implemented that allows the administrator to manage the building with one application by transmitting the result of whether or not the person is wearing a mask to the back-end server.

Keywords: Access control, face prediction, face tracking, CNN, SSD

1. Introduction

The emergence of artificial intelligence is being integrated in various fields of object detection, facial recognition, and image classification, taking over some of the human tasks. Artificial intelligence consists of a process of finding rules or features by collecting huge amounts of data and continuously training a model. It is the process of learning from samples by transforming input data into meaningful outputs with the generated model and comparing this output with known inputs.

In this way, face recognition deep learning models are created by finding and learning feature vectors from face images of many people. The face recognition model can identify the input image by comparing the similarity between the feature vectors. Among the numerous techniques of deep learning technology, research is conducted using a convolutional neural network (CNN) model (Nguyen, T. et al., 2020). The input data of an artificial neural network composed of only fully connected layers is limited to a one-dimensional form, and a single-color photograph is three-dimensional data. When it is necessary to train a neural network using image data, the 3D image data is flattened into one dimension. However, in the process of flattening the data, spatial information is inevitably lost. As a result, the artificial neural network is inefficient in extracting and learning features due to the lack of information due to the loss of image spatial information, and there is a limit to increasing the accuracy. Convolutional Neural Network (CNN) is a model that can learn while maintaining spatial information of images (Zhang, M. et al., 2019).

The outbreak of COVID-19 forced people to wear masks. Yet, there are still who are non-compliant to this health protocol. Currently, the recognition rate and speed to capture body temperature at entrance buildings are low (Lee H. J. et al., 2020). In addition, most of the manpower is placed at the entrance building to manually check whether a person is wearing a mask or not, to see their body temperature, and to control the entry. This leads to placing unnecessary manpower for each door. To solve these problems, it is necessary to implement a technology suitable for the untact era (Jun S. H. et al., 2020). Therefore, in this paper, we implemented a platform that controls access to the entire building by checking whether a mask is worn using deep learning technology. To this end, technology necessary for face detection and tracking was developed. The Single Shot Multibox Detector (SSD) algorithm and Convolutional Neural Network (CNN) model for face detection and tracking (Wang, Yuanyuan et al., 2018) will be applied in this study.

In addition, the entire development content was largely composed of front-end, signal transmission, back-end, and application. In the front-end stage, face tracking and mask wearing inspection technologies were developed and applied. The back-end stage developed a process of transmitting control signals to the server and controlling access from the server. Finally, in the application stage, building control was configured as a main function.

2. Technology Status

2.1. Object detection technology

To detect a face object, a standard template for the face is created and a search window is applied to the input image. It is a method to find a face region by comparing each search window image with a template. These methods can be broadly classified into two categories (Razali, M. N. et al., 2018). The first is to learn a face model and see how well a specific image matches the model and determine whether it is a face if it is above a certain level depending on the level of conformity. The second is to use a classification function that can distinguish faces and non-face images. In this case, the classification function is used by learning with a face image and a non-face image (Yu, Y. J. et al., 2020).

By detecting facial elements, positions of specific elements of the face, such as eyes, nose, and mouth, is extracted in advance. Then, a face part is detected by calculating a feature vector between these elements. However, this method requires high image quality and establishes rules for correlation between facial elements. However, it is difficult to develop an algorithm to apply it and the efficiency is lowered when the background is complicated. In this way, a method based on regional characteristics of a face has a complicated problem of multiple detection (Borji, A. 2015). Therefore, many facial region extraction techniques have been moving toward the former.

When face region detection is completed, facial components (eyes, eyebrows, nose, mouth, etc.) that can be used for recognition are found and the information is provided to the feature extraction step. Factors that make it difficult to extract facial features include problems in general object recognition such as lighting, size, and pose change (Zhang, L. et al., 2019). In addition, there are problems that only the face has, such as changes according to facial expression, glasses, hair shape, aging, and accessories. A high-performance face recognition system should be able to recognize the same person while accommodating these changes. However, the current level of technology is predominantly technology that performs recognition in a limited environment.

2.2. Face recognition technology

Face recognition technology is an applied technology that automatically identifies a person's face through a digital image. It refers to a technology that tracks a person's face in a photo or video and recognizes it through facial feature points (Wang, P. et al., 2021). Facial recognition technology consists of a face detection and face recognition functions. The main goal of face detection is to accurately find the location of a face in a photo. The key to the face recognition function is to identify whose face is the detected face (Hong, S. E. et al., 2020).

This is a feature that has already been commercialized or companies that develop mobile devices such as Apple and Samsung. Furthermore, it is a function that has

been implemented even for person identification (facial recognition) by recognizing a face. It is also possible for a user to determine where a face is located in an image or video and the specificity of that face (Malach, T. et al., 2020). For example, Amazon's Marineus Analytics uses facial recognition artificial intelligence to help agencies identify and rescue victims of human trafficking. Another example is the use of facial recognition technology in cameras. Smartphone cameras, DSLR cameras, mirrorless cameras, etc. have AF mode function. The AF mode is a function in which the camera automatically focuses on the subject. In AF mode, there is a function that tracks people's faces(Xu, S. et al., 2020). Even if you take a photo in an instant, the face recognition function can track and focus on a person's face.

3. Research Contents and Methods

The research content in this paper consists of a front-end part, a signal transmission part, a back-end part, and an application part. The front-end part checks through a CNN model whether a mask is worn. The signal transfer part transfers the control signals from the back-end to the server. The back-end part implements the access control process in the server. The application part implemented as a real-time management system for building access control. Figure 1 shows the overall system configuration.

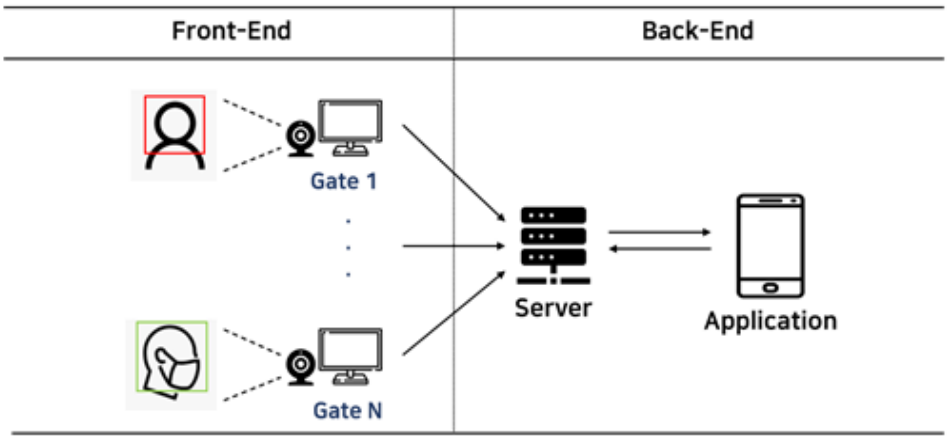

Fig. 1: System configuration

3.1. System design theory

There are various algorithms for face tracking algorithms including the Haar cascade classification algorithm and the Single Shot Multibox Detector (SSD) that detects faces with both eyes. The open-source face_recognition enables tracking of the face with a 99% probability without a worn mask. However, it was impossible to track if there are face obstructions including a mask. Thus, this study will use the SSD algorithm to enable face tracking on faces with masks. The SSD algorithm can extract

the location and size of a specific object class from an input image in real time.

Figure 2 shows the structure of the SSD algorithm. The SSD algorithm receives a 300x300 image and extracts features by passing it through to the Conv5_3 layer of VGG pretrained with an image set. Object detection is performed while passing the extracted feature map to the next layer through convolution. In the previous Fully Convolution Network, the problem of detailed information disappearing through convolution is solved by pulling the feature maps at the front. Based on this point, SSD is an algorithm that applies the method of performing object detection in each step-by-step feature map.

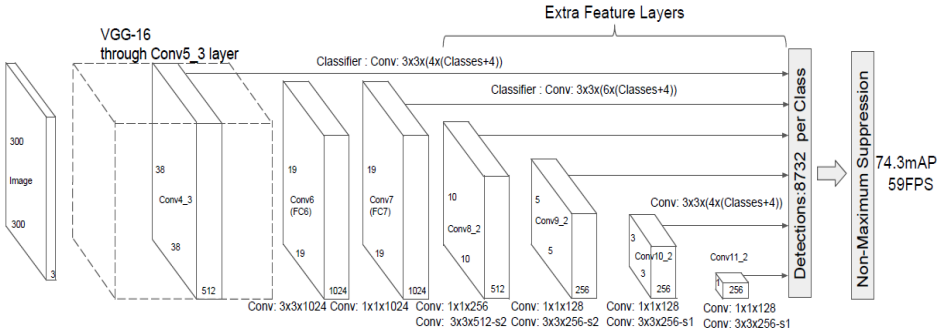

Fig. 2: SSD (Single Shot Multibox Detection) configuration.

In this paper, to implement the face detection function, the SSD face detection algorithm was mainly applied to development. Each frame image received from a camera connected to a computer was input into the deep learning module, and faces were detected for each frame. At the same time, it is implemented by drawing a rectangle at the corresponding position and outputting it as an image. The image for each frame will be blobbed, placed into the SSD model, and executed with forward propagation. It will get four coordinates around the detected face, draws a rectangle, and then produces an output.

Second, we design and implement a CNN model specialized for image processing. The CNN adds a new layer called the Convolutional Layer and the Pooling Layer before the Fully-Connected layer. Then, after applying the filtering technique to the original image, it is configured to perform a classification operation on the filtered image.

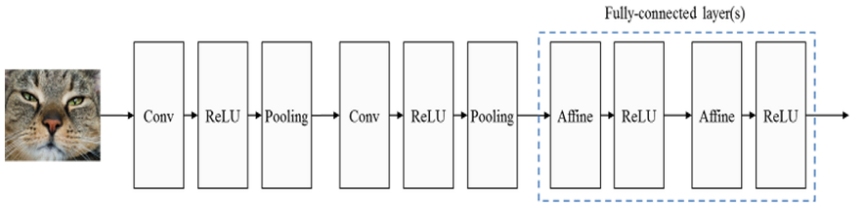

Fig. 3: CNN (Convolutional Neural Network) configuration

Figure 3 shows the structure of CNN. The CNN creates a new feature map that highlights the unique features of the image from the input image. The generated two-dimensional feature map is input to the artificial neural network, and the image is output by classifying which class label it belongs. Specifically, a convolutional neural network model is designed using an input image (a face wearing/not wearing a mask) as input data. Then, a feature map for wearing and not wearing a mask is created by itself. Then, with the generated features, the newly input face image will be classified if wearing a mask or not. Finally, class prediction values will be generated.

In this paper, the CNN model was designed by stacking a total of three layers. If the image features are derived through the convolutional layer, the pooling process should be performed. Unnecessary information is abstracted through the Max pooling process, which extracts the maximum value within a given area. Also, dropout technique is used to avoid overfitting. The DropOut technique prevents overfitting, which is learned by turning off the nodes at random, which is too biased to the training data. Table 1 shows the structure and parameters of the CNN model used in this paper.

Table 1. CNN model shape

Layer(Type)	Output Shape	Parameter
conv2d_1	(None, 126,126, 16)	448
MaxPooling2D	(None, 63, 63, 16)	0
Dropout	(None, 63, 63, 16)	0
conv2d_2	(None, 61, 61, 32)	4640
max_pooling2d_1	(None, 30, 30, 32)	0
Dropout	(None, 30, 30, 32)	0
conv2d_3	(None, 28, 28, 64)	18496
max_pooling2d_2	(None, 14, 14, 64)	0
flatten (Flatten)	(None, 12544)	0
Dense	(None, 512)	6423040
Dense2	(None, 2)	1026
Total	-	6,447,650

The activation functions of CNNs include a softmax regression function and a sigmoid regression function. Softmax regression function is used for multi-class classification, and sigmoid regression function is usually used for binary class classification. In the problem of classifying data into two groups, the most basic method is logistic regression analysis. The difference from regression analysis is that the range of the dependent variable is real because the desired value is the predicted value (real number). However, in logistic regression analysis, the dependent variable y has 0 or 1. Since this paper simply checks whether a mask is worn (O, X), a model

was designed using a sigmoid activation function that uses a binary classification technique.

3.2. Research methods

In this paper, a model is designed and executed to compose and classify 25,000 mask-wearing/non-masked human face images as data. The input data consisted of a total of 12,500 face photos with masks and 12,500 face photos without masks. Through Keras' ImageDataGenerator, the image divided into folders was pre-processed to 128X128 size and used as input data. Random transformation and normalization were performed on the image during data preprocessing using ImageDataGenerator. Also, it is possible to load the transformed image in batch units. In order to effectively train the CNN model, it was scaled to 1/255 and converted into a range of 0-1, and pre-processed as the training data of the model. After labeling the mask-wearing face photo and non-masking face photo data with (1, 0), it is used as input data. Figure 4 shows an example of an input data set.

Fig. 4: Example of face image with/without mask

Figure 5 shows the mask wearing inspection process.

Signals and images required for access control are delivered through a server/client program written in Python (Python3). When a person without a mask enters and exits, the deep learning model catches it. Then, a signal and a face image of a person not wearing a mask are detected and delivered to the server in real time. In addition, by blocking the entrance and generating a sound, the building occupants can check whether or not they are wearing a mask.

4. Results and Discussion

The research in this paper consists of 1) face recognition when wearing/not wearing a mask, and 2) checking the image classification result according to whether or not a mask is worn. Figure 6 shows the results of human face recognition when wearing and not wearing a mask.

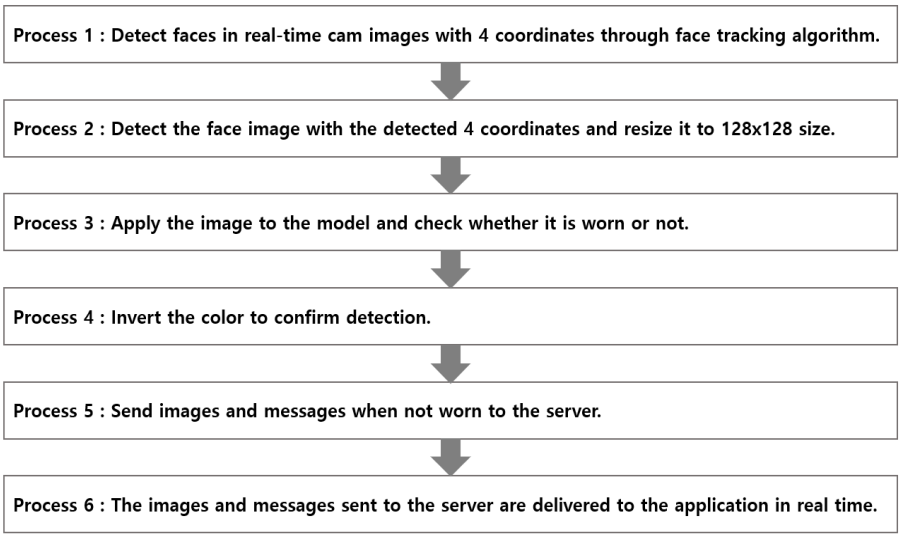

Fig. 5: Mask wearing inspection process

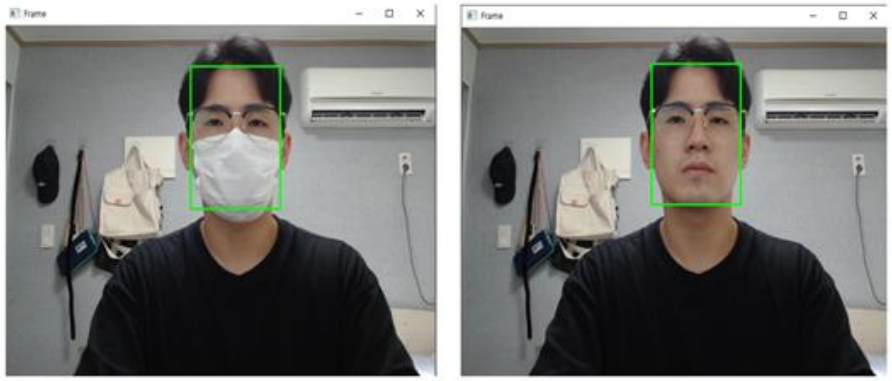

Fig. 6: Face recognition even when wearing a mask

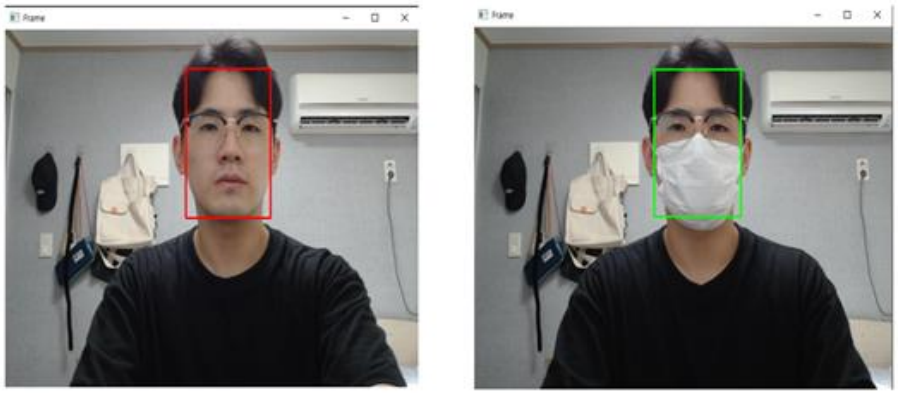

Fig. 7: Area indication according to wearing/not wearing a mask

The system detects the visitor's face through the web cam in real time, draws a rectangle around the visitor's face, and detects whether it is wearing a mask or not with deep learning model. In addition, a motion test will be conducted through varying distance to check the effectiveness in different environments. As a result, it was tested that the face recognition function and the mask wearing/non-wearing inspection function worked perfectly at which a face is identified regardless of the distance. Figure 7 shows the difference in area display according to wearing/not wearing a mask. That is, when the mask was worn, the RGB value was given as (0,255,0) and displayed in green, and when the mask was not worn, the RGB value was set as (255, 0, 0) and displayed in red.

In this paper, we developed and conducted a study on a mask-wearing CNN model, which was generated from the face image of a person wearing a mask and a face image of a person not wearing a mask. It showed 100% accuracy when tested with 20 random human face photos, and showed high accuracy of 90% when tested through a real-time web cam.

5. Conclusion

In recent years, rapidly developing artificial intelligence is effectively raising the function of face recognition and object detection. Wearing a mask has become a daily routine, but there are still people violating this health protocol upon building entry and in public transportation. Currently, the recognition rate and the speed of measuring body at each building entrance are very low. Unnecessary manpower is also placed at the entrance of each to control the entry of people.

Therefore, in this paper, we designed and developed a building access management system by examining whether a mask is worn or not in a real-time environment. This study used a Single Shot Multibox Detector (SSD) algorithm for better face recognition rate and face image detection using deep learning model to check whether a mask is worn or not. The results showed 100% accuracy when tested with 20 random human face photos and 90% high accuracy when tested through a real-time web cam.

Based on the research content of this paper, it is possible to easily inspect whether a mask is worn in a building in a CCTV environment. In addition, it can be extended to a system that easily detects when a person without a mask enters and exits a building.

References

Borji, A. (2015). What is a Salient Object? A Dataset and a Baseline Model for Salient Object Detection. *IEEE Transactions on Image Processing*, 24(2), 742-756. DOI: 10.1109/tip.2014.2383320

Hong, S. E. & Ryu, J. B. (2020). Unsupervised Face Domain Transfer for Low-Resolution Face Recognition. *IEEE Signal Processing Letters*, 27, 156-160. DOI: 10.1109/lsp.2019.2963001

Jun, S. H. & Kim, J. H. (2020). Theoretical background and Prospects for the Untact Industry. *Journal of New Industry and Business*, 38(1), 96-116. DOI: 10.30753/EMR.2020.38.1.005

Lee, H. J. & Lee, H. Y. (2020). COVID-19 Stress: Is the level of COVID-19 stress same for everybody? -Segmentation approach based on COVID-19 Stress level. *Korean Logistics Research Association*, 30(4), 75-87. DOI: 10.17825/klr.2020.30.4.75.

Malach, T. & Pomenkova, J. (2020). Optimal face templates: the next step in surveillance face recognition. *Pattern Analysis and Applications*, 23(2), 1021-1032. DOI: 10.1007/s10044-019-00842-y

Naimesh, P. J. & Jinan, F. (2021). Emotion Recognition using Facial Expression. *International Journal of IT-based Public Health Management*. 8(1), 9-18, doi:10.21742/IJIPHM.2021.8.1.02.

Nguyen, T., Nguyen, G. & Nguyen, B. M. (2020). EO-CNN: An Enhanced CNN Model Trained by Equilibrium Optimization for Traffic Transportation Prediction. *Procedia computer science*, 176, 800-809. DOI: 10.1016/j.procs.2020.09.075.

Razali, M. N. & Manshor, N. (2018). Object Detection Framework for Multiclass Food Object Localization and Classification. *Advanced Science Letters*, 24(2), 1357-1361. DOI: 10.1166/asl.2018.10749

Yu, Y. J., Moon, S. H., Sim, S. J. & Park, S. H. (2020). Recognition of License Plate Number for Web Camera Input using Deep Learning Technique. *Journal of*

Wang, Yuanyuan, Wang, Chao & Zhang, Hong. (2018). Combining a single shot multibox detector with transfer learning for ship detection using sentinel-1 SAR images. *Remote Sensing Letters*, 9(8), 780-788. DOI: 10.1080/2150704x.2018.1475770

Wang, P., Wang, P. & Fan, E. (2021). Violence detection and face recognition based on deep learning. *Pattern recognition letters*, 142, 20-24. DOI: 10.1016/j.patrec.2020.11.018

Xu, S., Song, X., Xia, L. & Xie, Z. (2020). Energy Efficiency Maximization for Energy Harvesting Bidirectional Cooperative Sensor Networks with AF Mode. *KSII Transactions on Internet and Information Systems*, 14(6), 2686-2708. DOI: 10.3837/tiis.2020.06.020.

Zhang, L., Yan, Y., Cheng, L. & Wang, H. (2019). Learning Object Scale With Click Supervision for Object Detection. *IEEE Signal Processing Letters*, 26(11), 1618-1622. DOI: 10.1109/lsp.2019.2937387

Zhang, M. & Geng, G. (2019). Adverse Drug Event Detection Using a Weakly Supervised Convolutional Neural Network and Recurrent Neural Network Model. *Information*, 10(9), 276. DOI: 10.3390/info10090276