Proceedings of the 3rd International Conference on Signal Processing and Machine Learning
DOI: 10.54254/2755-2721/6/20230838

Mask recognition in computer vision technology

Yaoyun Lu[1,3,†], **Yueyang Ma** [2,†]

[1] Suzhou Science and Technology Town Foreign Language High School, Jiangsu, 215000, China
[2] RKCS King's college Hangzhou, Zhejiang, 313200, China

[3]15010240226@xs.hnit.edu.cn
[†]These authors contributed equally.

Abstract. In order to lessen the strain on employees and the potential for carelessness, intelligent identification is required in China where there is a scarcity of staff to monitor the wearing of masks in public areas. In this work, we use a mask recognition technique to determine which members of the population weren't wearing masks by using the CNN and the VGG16 model. The ideas of data augmentation, dropout, non malicious, and transfer learning are used in the proposed study. This method may be used at hospitals, retail centers, transit hubs, dining establishments, and other community gatherings that require monitoring.

Keywords: CNN, VGG16, intelligent identification.

1. Introduction

Many nations have implemented total openness policies in response to the most recent outbreak of a novel coronavirus. But China continues to try to keep a tight rein. It is necessary for people to show their health and travel codes when they go to public places or other cities. Of course, wearing a mask is also the most basic rule. While most people follow these basic principles, there will always be a minority of people who go into public places without wearing a mask. When there are too many people and not enough managers, it's hard for managers to make sure they can check that everyone is wearing masks. So we decided to solve this problem by creating a system that could automatically identify if people were wearing masks.

In the week of august, the number of new weekly cases decreased by 9 percent as compared to the last week [1]. The accuracy has since increased. The first choice among these biometric modalities is face recognition, which has historically been seen as being more reliable. for example, these biometric modalities are fingerprint or iris recognition [2]. The most difference between face recognition and other biometric modalities are the distance. For example, fingerprint recognition needs people to put their finger into a sensor. But face recognition does not require people have physical contact with sensor, only stand in front of sensor. That is to say, during the pandemic, face recognition is very useful. Without any physical touch, it can recognize a person and determine whether or not they are wearing a mask. So, it can reduce the probability to get infected.

Face recognition technology scans and records the face characteristics of a portrait, then enters the data and processing system with these facial characteristics. To locate and compute the data of these face traits, the system will employ data analysis. The face characteristics of each individual have specific feature information, too.

DOI: 10.54254/2755-2721/6/20230838

The foundation of face recognition technology is feedback from the most fundamental face data. Having adequate information to identify a person is the first step. A computer's initial step in processing input data is to see if there are any faces. Succedent data analysis cannot be done to determine the features of the data until this premise is proven to be true. The background host then receives the feature data and uses big data to determine whether it agrees with the previously gathered data information. In order to avoid identity fraud and other issues, simultaneously check their database to see if their identity information matches. In the smart home market, for instance, the well-liked face-recognition intelligent door lock now incorporates facial recognition, fingerprint unlocks, mechanical unlock, and other capabilities. Although facial recognition technology may imitate fingerprints and replicate keys as part of its market appeal, your face contains your biological information.

2. Method

Our method process is shown in Figure 1. First we use dataset to train our AI model. After the training part finishes, we use the test data to test the model. Then we can get the final result to evaluate our model.

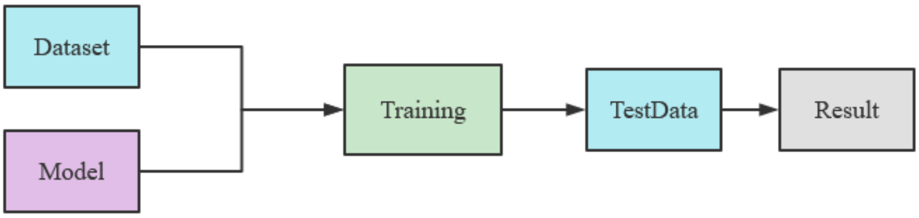

Figure 1. Research Method.

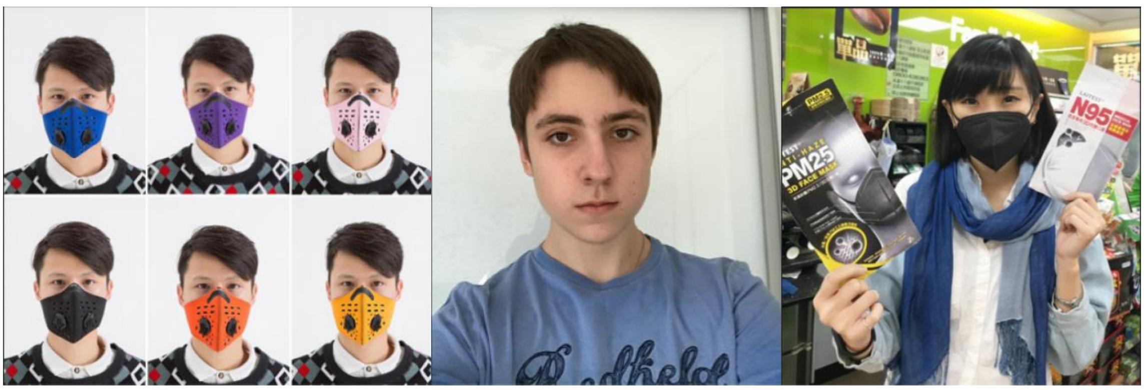

Figure 2. Those three are the face models.

Convolutional Neural Network (CNN) technology can increase face recognition's precision. It is a component of the forward-feedback neural network. Layer-by-layer convolution is used in the feature extraction procedure, followed by multi-layer nonlinearity to extract the features. The network employs mapping to autonomously learn from training data without any pre-processing in order to produce feature extractors and classifiers suited for the recognition job.

One of CNN's most significant innovations, VGG16, opened the door for much later study. It is a convolutional neural network (CNN) model put forth by Oxford University's Karen Simonyan and Andrew Zisserman. The idea for this model was initially proposed in 2013, but the final model wasn't unveiled until 2014 as a part of the ILSVRC ImageNet Challenge. The ImageNet Large-scale Visual Recognition Challenge, an annual competition, evaluates methods for large-scale picture categorization (including object detection). Despite performing well in the challenge, they lost. The model was compared to AlexNet-2012 and ZFNet-2013. First, the model suggests using a 3 3

DOI: 10.54254/2755-2721/6/20230838

convolution check to execute convolution throughout the entire network and setting the step size to 1. This is in contrast to the first convolutional layer's huge receptive field. In AlexNet, the 11x11 convolution kernel is used for step 4. 7 x 7 in ZFNet 2 steps. Vgg uses a 3 × 3 convolution kernel, which can extract features from images of larger size while making up for the deficiencies of 7x7 and 11x11. The combination of multiple 3x3 convolution kernels allows us to accept images of larger sizes. At the same time, he reduced the weight parameters of the model.

Since its creation, Vgg16 has been extensively applied in the field of image recognition, where it has had considerable success. Image identification, image detection, semantic segmentation, and other elements are the key application and enhancement aspects. In literature [3], the author combines the Inception and residual ideas are integrated, and four modules are designed, named Inception ResNet module 1, Inception ResNet module 2, ResNet module 3, And ResNet module 4. At the same time, the application of this network to commodity image classification has achieved great success. Experiments show that this model has significant improvement compared with the traditional vgg16. In the literature [4], the author proposed a fast identification and classification method for common radio and rational CHM based on improved VGG16. At the same time, the convolution module and attention module were added to Vgg16. After that, the average pooling of the whole play reduced the parameters of Vgg16. The experimental results show that it is higher than ResNet50, MobileNetV2, DenseNet121 and other networks. In literature [5], the author first applied Vgg16 to transfer learning, proving the reliability and availability of using Vgg16 in transfer learning. In reference [6], the author introduced a deep classifiable network to reduce convolution operations, and also introduced the SE module, which has proved to have a stronger feature extraction capability. At the same time, this work also reduced Vgg16 network parameters. In the literature [7], according to the characteristics of the tumor data set, Vgg16-SSD method is used to extract the tumor data. The experiment shows that the result is higher than that of traditional Vgg16. In the literature [8], multi-layer feature fusion technology is used to improve Vgg16 network, taking into account not only the semantic features of the picture, but also the texture features of the picture. This method can better extract the features. The literature [9] uses the improved Vgg16 to classify and predict the pictures, The original activation function is replaced by softmax, and the weighted activation is carried out in the softmax layer. Finally, good results are achieved. In literature [10], Vgg16 combined with DenseNet network was proposed to extract features. At the same time, attention mechanism was added to the model, and the model was improved. In the chest X-ray image classification task, 98% accuracy was achieved. In literature [11], the author still used Vgg16 for migration learning, and made a series of fine-tuning measures for downstream tasks. At the same time, he added a variety of data enhancement technologies, and finally achieved good results. The literature [12] has made a series of improvements to Vgg16 to solve the problems of the existing model that is difficult to extract features from the oven food dataset and low recognition accuracy. To boost the performance of the benchmark model, an asymmetric convolution block (ACB) is used in place of the original convolution kernel, and a batch normalization (BN) layer is added after the pooling layer to improve standardization. According to the experimental findings, the suggested model's recognition accuracy is 98.2% better than that of the conventional neural network. In view of the above work, we can see that the existing researchers' improvements to Vgg16 mainly focus on the following two aspects: 1. improving the internal model, adding activation layer and attention mechanism, 2. transferring learning to the model, adding fine-tuning measures in downstream tasks.

In order to implement CNN, different sizes of 3 x 3 filters—16, 16, 32, 64, 66, 96, 96, and 128—were used. To accomplish 10 convolution layers, Relu was used as the activation function, and lv (x) = Max (0, x) was computed. Five max-pooling layers, stride 2, and then flat layers make up the model. A fourth dense layer with two hidden nodes and the softmax activation function are used to produce the output. For the VGG16 implementation, fully connected layers from the original VGG16 design are taken out and additional dense layers are put on top of the model. Three fully connected layers are replaced in this article by two dense layers. The output of the second dense layer uses the softmax activation function. Overfitting is prevented by using pre-trained weights, transfer learning, dropout, augging, and normalization.

DOI: 10.54254/2755-2721/6/20230838

The system employs neural networks and SSD for face identification to identify faces and a detector to determine whether someone is wearing a mask as part of security measures. The door opens and closes if someone is wearing a mask, and if someone enters without one, a phone app sends out a warning. The previously mentioned 3D input matrix is used by the convolutional layer, which then puts it via a filter that is taken into account by the image. Up until the entire picture is scanned, this filter is built in a little window of pixels (3x3) at a time. The convolution operation computes the dot product generation of pixel values in the active filter window, taking into account the weights supplied in the filter. The final convolution picture is the result of this technique. Convolutional neural networks are the most often used neural networks for image processing (CNN). CNN receives the image in the form of a 3D matrix. The image's width and height are represented by the first two dimensions, while each pixel's RGB value is represented by the third dimension. The following successive modules make up a CNN: (each module can contain more than one layer). Convolution Function of activation. The pool, 4 completely joined layers 5. The last layer

3. Result

The table has been evaluated for several scenarios with various numbers of epochs (20) and batch sizes (20). In order to obtain smooth photos, we additionally employ average poles and maximum pools. In the first figure, the loss of training and validation accuracy is shown along with the loss of the diagram, while in the second figure, the training, validation, loss and hours are shown along with the accuracy of the diagram. As can be seen from the image, the loss gradually decreases as the accuracy increases. We utilized CNN to determine whether a person was hiding behind a mask. A "green rectangle" with the words "Thank you." Mask will be displayed if CNN points to "0," indicating that the individual is wearing a mask. The person's face is identified and a "red rectangle" with the message "no mask detected" is displayed if the CNN indicates that the individual is not wearing a mask, which is indicated by the number "1". Precision training: Training model with precision.

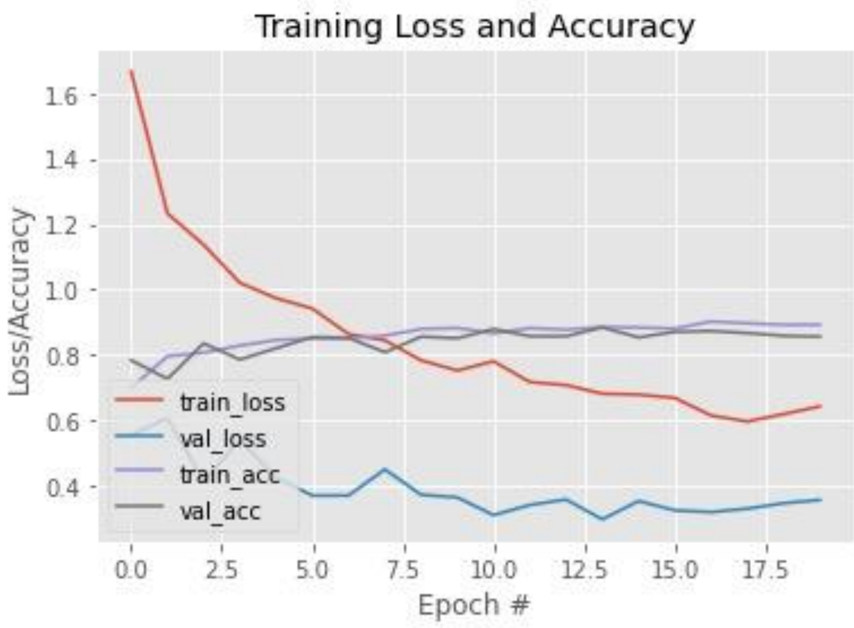

Figure 3. This graph represents the correct rate.

4. Limitations

Although the model in use has great accuracy in mask recognition, it has several drawbacks. Accurate findings cannot be produced without appropriate light and an appropriate picture. When a person's mask must be identified in low-quality images or in dim lighting, the model's accuracy declines and it is unable to produce reliable findings. In the framework of our next work, we will make an effort to

DOI: 10.54254/2755-2721/6/20230838

get around this restriction. We utilize the data to train a convolutional neural network that can distinguish between masked and unmasked pictures. Next, the classification of the established models is done using a convolutional neural network. According to experimental findings, this approach can successfully identify the face's mask in a picture. The findings of our model are highly positive since they showed that it was able to identify data more accurately than any other technique. Our model is more accurate than other models when compared to them. This model's accuracy score is 94.2.

5. Conclusion

This paper using CNN and VGG16 model for face recognition which provides a technique for identifying a person who wear mask or not in many different situations.

In the future, we can use some functions to improve the limitations. For example, using Gamma Intensity Correction, GIC).

References

[1] https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports
[2] K. Delac and M. Grgic, "A survey of biometric recognition methods," in 46th International Symposium Electronics in Marine, vol. 46, pp. 16– 18, 2004.
[3] Kim, M., Jain, A. K., & Liu, X. (2022). AdaFace: Quality Adaptive Margin for Face Recognition. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 18750–18759).
[4] An, X., Deng, J., Guo, J., Feng, Z., Zhu, X., Yang, J., & Liu, T. (2022). Killing Two Birds with One Stone: Efficient and Robust Training of Face Recognition CNNs by Partial FC. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 4042–4051).
[5] Wang, K., Wang, S., Zhang, P., Zhou, Z., Zhu, Z., Wang, X., … & You, Y. (2022). An efficient training approach for very large scale face recognition. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 4083–4092).
[6] Phan, H., & Nguyen, A. (2022). DeepFace-EMD: Re-Ranking Using Patch-Wise Earth Mover's Distance Improves Out-of-Distribution Face Identification. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 20259–20269).
[7] Wang, Z., Wang, Z., Yu, Z., Deng, W., Li, J., Gao, T., & Wang, Z. (2022). Domain Generalization via Shuffled Style Assembly for Face Anti-Spoofing. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 4123–4133).
[8] Liu, C., Yu, X., Tsai, Y. H., Faraki, M., Moslemi, R., Chandraker, M., & Fu, Y. (2022). Learning to Learn across Diverse Data Biases in Deep Face Recognition. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 4072–4082).
[9] Athar, S., Xu, Z., Sunkavalli, K., Shechtman, E., & Shu, Z. (2022). RigNeRF: Fully Controllable Neural 3D Portraits. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 20364–20373).
[10] Ukil A. Support Vector Machine[J]. Computer Science, 2002, 1(4):1-28.
[11] He K, Zhang X, Ren S, et al. Delving Deep into Rectifiers: Surpassing Human-Level Performance on Image Net Classification[J]. 2015:1026-1034.
[12] Hinton G E, Srivastava N, Krizhevsky A, et al. Improving neural networks by preventing co-adaptation of feature detectors[J]. Computer Science, 2012, 3(4):págs. 212-223.