SAKARYA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ DERGİSİ

Sakarya University Journal of Science
SAUJS

e-ISSN 2147-835X Founded 1997 Period Bimonthly Publisher Sakarya University
http://www.saujs.sakarya.edu.tr/en/

Title: A Multi-task Deep Learning System for Face Detection and Age Group Classification for Masked Faces

Authors: Gozde YOLCU ÖZTEL, İsmail ÖZTEL

Recieved: 2021-08-12 00:00:00

Accepted: 2021-11-03 00:00:00

Article Type: Research Article

Volume: 25
Issue: 6
Month: December
Year: 2021
Pages: 1394-1407

How to cite
Gozde YOLCU ÖZTEL, İsmail ÖZTEL; (2021), A Multi-task Deep Learning System for Face Detection and Age Group Classification for Masked Faces. Sakarya University Journal of Science, 25(6), 1394-1407, DOI: https://doi.org/10.16984/saufenbilder.981927
Access link
http://www.saujs.sakarya.edu.tr/tr/pub/issue/66341/981927

Sakarya University Journal of Science 25(6), 1394-1407, 2021

A Multi-task Deep Learning System for Face Detection and Age Group Classification for Masked Faces

Gozde YOLCU ÖZTEL*[1], İsmail ÖZTEL[2]

Abstract

COVID-19 is an ongoing pandemic, and according to experts, using a face mask can reduce the spread of the disease. On the other hand, masks cause occlusion in faces and can create safety problems such as the recognition of the face and the estimation of its age. To prevent the spread of COVID-19, some countries have restrictions according to age groups. Also, in different countries, people in some age groups have safety restrictions such as driving and consuming alcohol, etc. But these rules are difficult to follow due to occlusion in faces. Automated systems can assist in monitoring these rules. In this study, a deep learning-based automated multi-task face detection and age group classification system is proposed for masked faces. The system first detects masked/no-masked-faces. Then, it classifies them according to age-groups. It works for multi-person regardless of indoor/outdoor environment. The system achieved 92.0% average precision score for masked face detection using YOLO with resnet50 network. Also, 83.87% accuracy for classifying age groups with masked faces and 84.48% accuracy for no-masked faces using densenet201 network have been observed. It produced better results compared to the literature. The results are significant because they show that a reliable age classification for masked faces is possible.

Keywords: age classification, computer vision, COVID-19, deep learning, pretrained networks

1. INTRODUCTION

The Novel Coronavirus Disease (COVID-19) is a worldwide emergency because of its rapid spread and the high death rate. Unfortunately, specific medications are not yet available. Despite some new vaccines, most scientists predict that COVID-19 vaccines will not be 100% effective like most other vaccines [1]. Many studies have been presented to assist the diagnosis of disease. In [2]–[5], the authors have studied on classifying of COVID-19 in chest X-ray images. Also, some studies [6], [7] worked on automatic COVID-19 lung infected region segmentation.

WHO reported that social distancing measures and wearing medical face mask help to slow the spread of disease. In many countries, wearing a medical face mask is obligatory. To prevent the spread of COVID-19, restrictions according to age groups are applied in some countries. For example, according to the restrictions applied in

* Corresponding author: gyolcu@sakarya.edu.tr
[1] Sakarya University, Faculty of Computer and Information Sciences, Software Engineering Department
ORCID: https://orcid.org/0000-0002-7841-2131
[2] Sakarya University, Faculty of Computer and Information Sciences, Computer Engineering Department
E-mail: ioztel@sakarya.edu.tr
ORCID: https://orcid.org/0000-0001-5157-7035

the past in Turkey, people over the age of 65 were only allowed to go out between 10:00 and 13:00. Similarly, people under the age of 20 were allowed to go out between 13:00 and 16:00. Also, public transportation was prohibited for these two age groups. Thus, to monitor whether compliance with these rules, there was a need to detect the age group of people wearing masks.

In addition, estimating the age groups of people wearing masks is also important for safety. For example, according to different countries, people in some age groups have restrictions such as driving, consuming alcohol, etc. After the pandemic, age has become hidden due to the mask, and it has become hard to check whether these rules are being followed.

It is quite difficult for the human observer to predict the age groups of masked people. An assistance system can make this observation easier. Thus, there is a need to develop non-invasive, quantitative, and objective automated systems for determining age groups.

This study proposes a deep learning-based multi-task system for face detection and age-group classification using masked face. It classifies human faces into 3 age groups which are teenager (12-20), middle (21-64), and elder (65+). The system can work for multi-person regardless of indoor/outdoor environment. The proposed system is useful for tracking restrictions according to age groups to prevent the spread of coronavirus disease. This tool also can be used for automated age determination of masked people for security purposes. Moreover, it can be an assistant system for suggesting videos/products according to the target age group. The main contributions of this study can be summarized as:

(1) Although the face masks cover the wrinkles in the mouth and cheeks, which play an important role in age determination, such a system that gives promising results has been developed.

(2) Reduced computational cost with fewer face attributes, making the approach advantageous for real-time or resource-limited systems.

The rest of the paper has been organized as follows: Section 2 has the information of related works for age classification. Section 3 includes preparation of the masked images dataset approach, detection and classification methods. Experimental results have been given in Section 4. Finally, the paper has been concluded in Section 5.

2. RELATED WORKS

Recently, many studies have been presented to help reduce the spread of COVID-19 or to assist the diagnosis of disease. In [8]–[10], tools were presented for automated mask detection on human faces. In [11]–[13], the researchers presented tools for automated social distance and mask detection. In [14], the authors presented a formally verified authentication protocol in secure framework for mobile healthcare during the COVID-19-like pandemic. In [15], the authors proposed a masked face recognition system.

This study proposes an assistance system for reducing the spread of COVID-19 using computer vision methods. Owing to computer vision technologies, dealing with many troublesome problems in daily life has become more effortless with automated systems [16]–[25]. The aging process on human facial appearances is affected by many factors such as health, weather, living environment, gender, race, etc. Thus, determining the age of a person may be difficult for an expert, and automated systems can be helpful for this task. Age group classification is one of the challenging tasks of computer vision due to changes in position and orientation, lighting conditions, image resolution, etc. [26].

Over the years, many age classification studies have been presented for different purposes. In the study of [27], the authors used age and gender classification for effective marketing analysis. Also, in [28], the authors used age classification for risk stratification in glioma patients. In [29], an age classification system was used for detecting illicit activity in suspicious sites.

Age group classification was first applied in [30]. In their study, authors classified images into three age groups which are babies, young adults, and senior adults. In the study, six ratios of distances between primary components (e.g., eyes, noses, mouth, etc.) and wrinkles on specific areas of a face were used.

In [31], Sobel edge operator and region labeling were used to locate the positions of eyes, noses, and mouths. Then, geometric and wrinkle features were extracted. Finally, the features were classified using two back-propagation neural networks. In [32], geometric features were extracted from faces and fused the results using five different classifiers. In [33], the authors used local binary patterns (LBP) and the k-nearest neighbor classifiers for the age group classification system. In [34], authors computed the geometric components of facial images; like wrinkle topography, face edge, left eye to right eye separation, eye to nose separation, eye to jaw separation, and eye to lip separation. Then K-means algorithm was used for the classification.

Different from the above studies, age group classification from masked faces due to COVID-19 is the motivation of this study. Because most of the important facial attributes, such as wrinkles on cheeks, nose, etc., are not visible due to the face mask; estimating the age group of people with facial masks becomes more difficult.

On the other hand, Haar Cascade can be another choice for mask detection. In the literature, some studies compare the Haar Cascade and deep learning networks. According to [35], CNN outperforms cascade classifiers in plenty of cases. Generally, Haar Cascade is fast because it uses fewer features. But detection rate is better in a deep learning model because the deep learning system uses many other features [35]. Usage of the Haar Cascade or deep learning model has advantages and disadvantages. A choice can be made based on the current requirements. In this study, deep learning models were preferred.

3. METHODOLOGY

3.1. Preprocessing

Unfortunately, there aren't any age datasets with masked faces. Thus, UTKFace Large Scale Face Dataset [36] images were preprocessed to be placed a medical face mask on each face image. Before placing the mask on the face, the width of the mask was resized to be the *x* distance between the two ears and the length of the mask to be the *y* distance between the nose and the chin. With this step, the system has been made flexible for each face size. Figure 1 shows these distances.

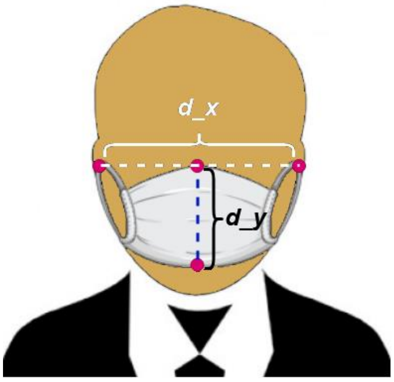


Figure 1 Distances used for mask placed on the face

In order to place the mask on the face with a transparent background, alpha-channel information of the mask image is used. Alpha channel has 0 values in regions associated with the transparent background.

To combine the transparent background mask image with a facial image, firstly, the transposition of the alpha channel (a in Figure 2) of the mask image is multiplied by the corresponding region on the face. The resulting image is then added to the multiplexing of the mask image (b) with its alpha channel (c). Thus, the composite image picks colors from the image when the alpha channel is 1 and picks colors from the facial region when the alpha channel is 0. Eq. 1 defines these processes. Also, these processes are illustrated in Figure 2.

$$C = alpha_m^T * fr + m * alpha_m \quad (1)$$

where C is composite image, $alpha_m$ is alpha channel of the mask image, fr is facial region which is adding the mask, m is the mask image and T represents the transpose process.

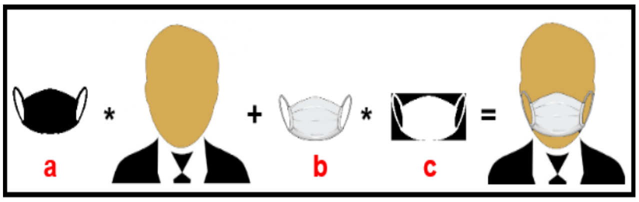


Figure 2 Combining the transparent background mask image with a facial image. a is the transposition of the alpha channel, b is the mask image, and c is the alpha channel of the mask image

3.2. Proposed System

The system pipeline of the proposed age group classification system for masked faces is illustrated in Figure 3. As can be seen in Figure 3, the system first detects faces. For this purpose, a Faster R-CNN and a YOLO detector have been tried, and the YOLO outperformed the Faster R-CNN. After the detection step in real-time, a classification network classifies the cropped faces as masked vs. no masked using the YOLO model.

3.3. Masked Face Detection using Faster R-CNN

The Faster R-CNN [37] architecture contains a pretrained network as a feature extractor and two subnets [37]. For pretrained network, Resnet-50 has been used. The task of the first subnet is to inform the second network of where the object will be searched. The first subnet is a fully convolutional network. It takes an image and outputs rectangular object proposals for that image. Every object proposals have objectness scores. The task of the second subnet is the prediction of the real class labels for object proposals.

For the masked face detection task, the maximum possible proposal number is determined as 3 empirically. That means 3 reference boxes, which are named anchor. Therefore, for all coordinates of 3 boxes, the classification layer outputs 6 score values for each proposal for the probability of object or not object. The loss function of the Faster R-CNN is defined in Eq. 2 [37].

$$L(\{p_i\},\{t_i\}) = \frac{1}{N_{cls}}\sum_i L_{cls}(p_i,p_i^*) + \lambda\frac{1}{N_{reg}}\sum_i p_i^* L_{reg}(t_i,t_i^*) \qquad (2)$$

where i is the number of anchor, p_i is anchor's predicted probability, p_i^* is 1 or 0 based on the situation if the anchor is positive or negative, respectively. t_i shows the coordinates of predicted bounding box. t_i^* shows the ground-truth box. L_{cls} represents the classification loss. $L_{reg}(t_i,t_i^*)$ represents the regression loss. The loss of regression can be active with just positive anchors, and this is performed using $p_i^* L_{reg}$ in the equation. N_{cls} and N_{reg} are performed to normalize the two term in the equation using a balancing parameter λ.

3.4. Masked Face Detection using YOLO

The You Only Look Once (YOLO) is an object detection algorithm. It uses a single stage CNN for detection. An object detector decodes the predictions and produces some bounding boxes [38], [39].

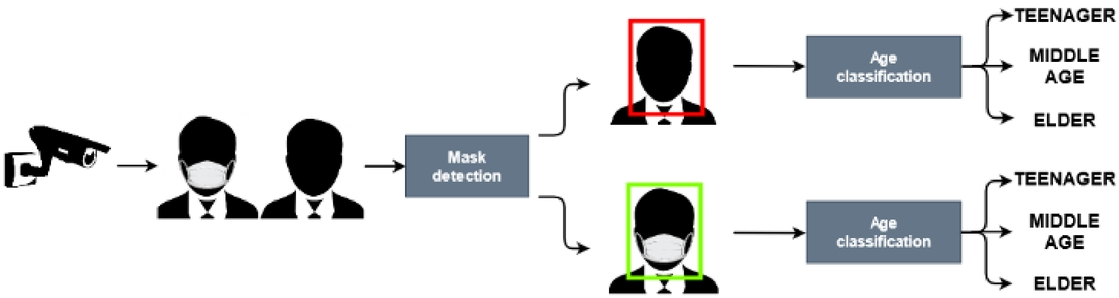


Figure 3 Proposed system pipeline

The anchor boxes are used to detect the object classes. For each anchor box; IoU, anchor box offsets, class probability attributes are predicted. IoU defines the objectness score of the anchor boxes. The anchor box offsets define anchor box position. Class probability defines the class label of each anchor box.

YOLO works on raw images. The input images are divided into grids. Each grid tries to find out if the object is in the field. So, a separate prediction vector is created for each grid. In this study, different feature extractor networks have been used for YOLO, and ResNet showed the best performance.

3.5. Age Group Classification using Deep Transfer Learning

After obtaining masked and no-masked faces from the first stage, the second task is classifying these faces according to age group. For this purpose, a Deep Convolutional Neural Network (DCNN) based approach is used. DCNN is one of the classes of deep neural network approaches used for computer vision problems [40]. It shows outstanding performance in different tasks [41]–[43].

In DCNN, information is transmitted through the layers, with the input of each layer being the output of the next layer. Learning is applied with the backpropagation [44] and stochastic gradient descent [40]. DCNN provides automated learning of discriminant features for a specific task. For the training process, data usually annotated by experts is used [45].

DCNNs are structured with different combinations of convolutions, pooling, dropout, and fully connected layers. In convolution layers, the input is convolved with the learned filters. The filters capture image features. Eq.3 shows the mathematical definition of the convolution process [40].

$$S(i,j) = (I * K)(i,j) = \sum_m \sum_n I(i+m, j+n) K(m,n) \quad (3)$$

where I is an input image, K is a filter, S is the output after the convolution process and i,j,m,n are the index values.

In pooling layers, inputs are subsampled for reducing the spatial size. Thus, the number of parameters is reduced, and overfitting is controlled. In fully connected layers, classification or regression are performed using the collected features from previous layers. Softmax layer determines which class it belongs to using the values obtained from the fully connected layer. The mathematical equation of the softmax function is given in Eq.4 [40].

$$softmax(x)_i = \frac{\exp(x_i)}{\sum_{j=1}^{n} \exp(x_j)} \quad (4)$$

where x is the output of the fully connected layer, i is the number of samples, and j is the number of the classes.

In dropout layers, input elements are randomly set to zero with a given probability. This is used for preventing overfitting.

Transfer learning allows the parameters of a network learned for a task to be reused as the starting point for another task [46]. Owing to the transfer learning, the performances of various machine learning tasks are improved [47]. In the proposed system, in order to take advantage of transfer learning, pre-trained networks developed for object detection task has been used. To apply these networks in the proposed study, just the fully connected and classification layers of the pretrained network have changed with new ones. The rest of the pretrained network has not been changed. To adapt the networks to the age group classification, the networks are retrained with facial age images. The properties of pretrained networks which are used in this study are given in Table 1.

Table 1 Properties of pretrained networks used for age group classification

Network	Depth	Parameters ($x10^6$)	Image input size	Size (MB)
alexnet [48]	8	61.0	227*227	227
vgg16 [49]	16	138	224*224	515
vgg19 [49]	19	144	224*224	535
squeezenet [50]	18	1.24	227*227	5.2
googlenet [51]	22	7	224*224	27
inceptionv3 [52]	48	23.9	299*299	89
densenet201 [53]	201	20	224*224	77
mobilenetv2 [54]	53	3.5	224*224	13
resnet-101 [55]	101	44.6	224*224	167
xception [56]	71	22.9	299*299	85
shufflenet [57]	50	1.4	224*224	5.4
darknet53 [38]	53	41.6	256*256	155

4. EXPERIMENTAL RESULTS

4.1. Datasets

In this study, two different databases were used for two different tasks. For the age group classification stage, UTKFace Large Scale Face Dataset was used. It has the age labels for no-masked human faces. In order to obtain training and test images for the masked age classification pipeline, firstly, masks were placed on the faces in this dataset as mentioned in the method section.

The database has 6631 face images. 1015 images contain faces from 12 to 20 years old, 4503 images from 21 to 64, and 1113 images from 65 to 110. The dataset was divided into 3 parts: training set (%75), validation set (%5), and test set (%20). For the masked faces detection step, the system was trained using Face Mask Dataset from Kaggle [58]. The dataset has 853 masked and no-masked face images. 767 images were used for the training, and 86 images were used for testing purposes. These images mostly include multiple human faces. Totally, the dataset has 3355 samples of faces with mask and 717 samples of faces without a mask. Also, a flipping process was applied to the images for data augmentation in order to increase the performance of the mask detection because of the small dataset size.

4.2. Evaluation of the System

Because classification results are usually evaluated using the Accuracy criteria index in the literature, this index was used for a reliable comparison in this study. The mathematical definition of Accuracy is given in Eq.5.

$$Accuracy = \frac{(TP+TN)}{(TP+TN+FN+FP)} \quad (5)$$

Also, in the literature, object detection performances are usually evaluated with Average Precision (AP), and mean Average Precision (mAP) metrics. Thus, this metric has been used for the mask detection task in this study. Average Precision (AP) is calculated using precision (Eq.6.) and Recall (Eq.7.) as can be seen in Eq 8. mAP is the mean of the APs for all classes.

$$Precision = \frac{TP}{(TP+FP)} \quad (6)$$

$$Recall = \frac{TP}{(TP+FN)} \quad (7)$$

$$Average\ Precision = \sum \frac{Precision}{Recall} \quad (8)$$

where TP is True-Positives, FP is False-Positives and FN is False-Positives.

In order to detect the faces with mask or no-mask, Faster R-CNN and YOLO with resnet50 feature extractor have been used in this study. The training parameters used by resnet50 are as follows: learning rate is 0.005, drop factor is 0.1, mini batch size is 8, momentum is 0.9, and the optimizer is sgdm. The masked face detection system showed 92% AP score for masked faces and 91.50% mAP score in Face Mask Detection Dataset.

For the age group classification task, some popular pretrained networks were retrained with masked and no-masked faces from the dataset. The parameters of the used CNN's are given as follows for 12 different pretrained networks. The optimizer is sgdm, the learning rate is 0.001, minibatch size is 32 and momentum is 0.9. Also, the training and validation data are shuffled once before training.

Classification results are shown in Table 2. As can be seen in this table, densenet201 showed the best performance when considering masked and unmasked face classification joint performance. Although a large part of the face has occlusion with the mask, age group classification in masked faces showed close results to unmasked faces (only slightly lower by 1%). As seen in the table, networks have generally produced similar results.

Table 2 Accuracy of pretrained networks used for age group classification

Pretrained Network	With mask (%)	Without mask (%)
alexnet	75.81	78.82
vgg16	78.37	84.48
vgg19	77.24	83.04
squeezenet	78.30	81.76
googlenet	81.24	81.54
inceptionv3	81.69	85.38
densenet201	83.87	84.48
mobilenetv2	80.78	81.54
resnet-101	83.35	83.57
xception	79.88	83.04
shufflenet	81.16	82.22
darknet53	81.69	83.95

In DCNNs, the first convolutional layers learn basic features. The same situation is observed on the first convolutional filters of densenet201, which has the best score for age group classification. The first convolutional filters of densenet201 are given in Figure 4. As can be seen in this figure, the filters generally learned edges and color information.

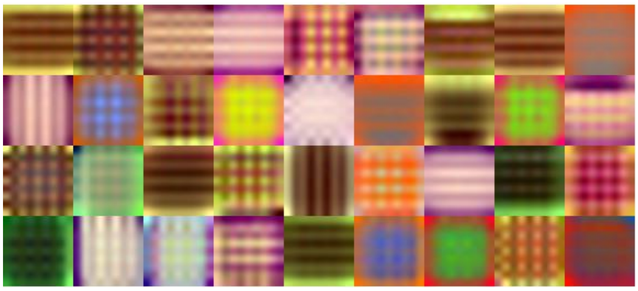
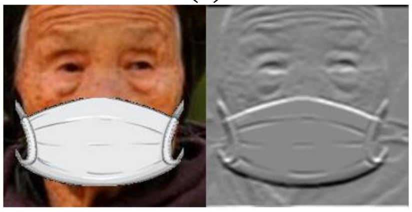


Figure 4 Samples of first convolutional layer filters from densenet201 network

In order to show that how to learn the networks, some output samples of the first convolutional layer of the densenet201 are given in Figure 5. Figure 5, can be interpreted that the filter ignores the mask region and highlight the remaining facial areas in (a). Situation (b) can be interpreted that the filter of the sample (b) becomes apparent in the lines of the figure and it can be related with wrinkles for the age classification.

(a)



(b)

Figure 5 Some samples of the outputs of the first convolutional layer from densenet201. (a) and (b) have an original image from UTKFace dataset with implemented face mask and corresponding output that is convolved a filter from the first convolutional layer of densenet201.

Figure 6 shows some sample frames from the system outputs. Figure 6 (a), (b), (c), (d) are from the Face Mask Dataset images and (e), (f) are from real time system. As can be seen in the subfigures, the system produces promising results. Also, Figure 6 (e) and (f), show that the system is not affected by poor lighting conditions and different background conditions in an environment.

4.3. Performance Comparison

The performance of the proposed system has been compared to masked face detection and age group classification systems in the literature, individually. Unfortunately, there is no unified benchmark test set for masked face detection and age group classification; different researchers use different test sets to report their results. Moreover, age classification studies in the literature work on the whole face without a mask. These factors make a fair comparison very hard.

The comparison of proposed masked face detection performance with the literature has been given in Table 3. According to the table, the proposed YOLO has the best performance in terms of AP. Also, in [59], the authors removed some images from the merged database, and the deleted images are not clear. Thus, a fair comparison can not be performed with this study.

Table 4 shows a comparison against age group classification systems using whole face information. In this study, the main motivation is detecting masked face groups in terms of assist controlling COVID-19 restrictions and other security restrictions. Thus, age groups between 12-20, 21-64, and 65+ are key points for this motivation. Unfortunately, there is not a benchmark dataset for this motivation. Thus, the comparison has been applied with-the studies that used the different number of age groups on unmasked faces. Although this is not a fair comparison, there are no other datasets to compare. It was especially shown in the table that the number of classes was different, and other studies worked on the unmasked face.

(a) (b) (c)

(d) (e) (f)

Figure 6 Some samples of the system output from Face Mask Dataset (a,b,c,d) and real time test environment (e,f)

Table 3 Comparison of masked face detection performance with the literature

Method	**Dataset**	**Average Precision**
LLE_CNNs [60]	MAFA	76.4%
YOLO with Resnet [59]	(MMD+FMD) *	81.0%
Proposed Faster R-CNN	FMD	79.0%
Proposed YOLO	FMD	92.0%

*: Note: The authors of Loey2020 removed some images in the merged database.

In this paper, two systems with the same processing steps, one with features from the whole face, one with features from the masked face, have been developed. The results in Table 4 demonstrate that successful age group classification is possible for masked faces. Using both whole face information and masked face information individually, the proposed system produced better results compared to previous studies. Also, using masked face information, the system produced comparable results against the system using whole face information (only slightly lower by ~1% compared to the whole face age group classification system while using only a small part of the face).

5. CONCLUSIONS

This study proposed a multi-task deep learning system for face detection and age group classification from masked faces. Recently, for fighting against COVID-19, people have used face masks. One of the disadvantages of face masks is causing occlusions in the face. Thus it produces difficulties for analyzing the face. But face analysis is essential for security. With this motivation, this study focuses on masked faces.

The proposed system first detects masked and no-masked faces in an environment. Then, it classifies these faces according to three age groups. It can classify age groups for people in front and side profiles, even in crowded areas. It can work regardless of the indoor/outdoor environment. It assists observers to more effortlessly and accurately detect the age group of the masked face. Also, it can quickly detect the people who do not obey the age-related rules in public areas such as markets, streets, shopping centers, airports, etc. and assist the authorities. Thus, more people can be provided to comply with the rules.

One of the study's limitations is that there is no database suitable for motivation, as explained in the experiment section. For this reason, the comparison could be made with studies of different classes of age groups. Also, errors are usually observed in the age borders separating the classes. For example, it is very difficult to distinguish two people's ages that are 20 and 21. But since these two are in separate classes, the correct classification of the system is expected. This is another limitation for the system

The proposed system achieved 92% AP score for masked face detection. Also, it produced 83.87% age group classification accuracy for masked faces. Comparison to the literature showed that the proposed system produces better results against the age group classification system with whole face analysis. These results are significant because they show that a reliable age classification for masked faces is possible. It may be one of the pioneering studies in terms of its motivation.

Acknowledgments

We would like to thank the reviewers and editors for their valuable time.

Funding

This study is supported by Sakarya University Scientific Research Projects Unit (Project Number: 2015-50-02-039).

The Declaration of Research and Publication Ethics

The authors of the paper declare that they comply with the scientific, ethical and quotation rules of SAUJS in all processes of the paper and that they do not make any falsification on the data collected. In addition, they declare that Sakarya University Journal of Science and its editorial board have no responsibility for any ethical violations that may be encountered, and that this study has not been evaluated in any academic publication environment other than Sakarya University Journal of Science.

Table 4 Age group classification accuracies for the proposed system and other studies

Method	Dataset	Age group number	Masked/no-masked	Acc(%)
MTCNN [61]	UTKFace	7	no-masked	70.1
FFNet [61]	UTKFace	7	no-masked	64.0
FaceNet [61]	UTKFace	7	no-masked	59.6
Local age group modelling [62]	Images of Groups	7	no-masked	59.8
Shape and TextureCharacteristic [63]	FG-NET	5	no-masked	79.2
Proposed	UTKFace	3	masked	83.87
Proposed	UTKFace	3	no-masked	84.48

REFERENCES

[1] “World Health Organization.” 2020https://www.who.int/.

[2] A. Abbas, M. M. Abdelsamea, and M. M. Gaber, "Classification of COVID-19 in chest X-ray images using DeTraC deep convolutional neural network," Appl. Intell., vol. 51, no. 2, pp. 854–864, Feb. 2021, doi:10.1007/s10489-020-01829-7.

[3] S. A. B. P and C. S. R. Annavarapu, "Deep learning-based improved snapshot ensemble technique for COVID-19 chest X-ray classification," Appl. Intell., Mar. 2021, doi:10.1007/s10489-021-02199-4.

[4] M. Turkoglu, "COVIDetectioNet: COVID-19 diagnosis system based on X-ray images using features selected from pre-learned deep features ensemble," Appl. Intell., vol. 51, no. 3, pp. 1213–1226, Mar. 2021, doi:10.1007/s10489-020-01888-w.

[5] H. Mittal, A. C. Pandey, R. Pal, and A. Tripathi, "A new clustering method for the diagnosis of CoVID19 using medical images," Appl. Intell., Jan. 2021, doi:10.1007/s10489-020-02122-3.

[6] A. Oulefki, S. Agaian, T. Trongtirakul, and A. Kassah Laouar, "Automatic COVID-19 lung infected region segmentation and measurement using CT-scans images," Pattern Recognit., vol. 114, p. 107747, 2021, doi:10.1016/j.patcog.2020.107747.

[7] D.-P. Fan et al., "Inf-Net: Automatic COVID-19 Lung Infection Segmentation From CT Images," IEEE Trans. Med. Imaging, vol. 39, no. 8, pp. 2626–2637, 2020, doi:10.1109/TMI.2020.2996645.

[8] M. Loey, G. Manogaran, M. H. N. Taha, and N. E. M. Khalifa, "A hybrid deep transfer learning model with machine learning methods for face mask detection in the era of the COVID-19 pandemic," Measurement, vol. 167, p. 108288, 2021, doi:10.1016/j.measurement.2020.108288.

[9] P. Mohan, A. J. Paul, and A. Chirania, "A Tiny CNN Architecture for Medical Face Mask Detection for Resource-Constrained Endpoints," ArVix, 2020http://arxiv.org/abs/2011.14858.

[10] B. Qin and D. Li, "Identifying facemask-wearing condition using image super-resolution with classification network to prevent COVID-19," Sensors (Switzerland), vol. 20, no. 18, pp. 1–23, 2020, doi:10.3390/s20185236.

[11] K. Bhambani, T. Jain, and K. A. Sultanpure, "Real-time Face Mask and Social Distancing Violation Detection System using YOLO," in Bangalore Humanitarian Technology Conference (B-HTC), 2020, pp. 1–6, doi:10.1109/B-HTC50970.2020.9297902.

[12] B. Sathyabama, A. Devpura, M. Maroti, and R. S. Rajput, "Monitoring Pandemic Precautionary Protocols using Real-time Surveillance and Artificial Intelligence," in 3rd International Conference on Intelligent Sustainable Systems (ICISS), 2020, pp. 1036–1041, doi:10.1109/ICISS49785.2020.9315934.

[13] S. Meivel, K. Indira Devi, S. Uma Maheswari, and J. Vijaya Menaka, "Real time data analysis of face mask detection and social distance measurement using Matlab," Mater. Today Proc., 2021, doi:10.1016/j.matpr.2020.12.1042.

[14] S. S. Ahamad and A.-S. Khan Pathan, "A formally verified authentication protocol in secure framework for mobile healthcare during COVID-19-like pandemic," Conn. Sci., pp. 1–23, 2020, doi:10.1080/09540091.2020.1854180.

[15] Y. Li, K. Guo, Y. Lu, and L. Liu, "Cropping and attention based approach for masked face recognition," Appl. Intell., Feb. 2021, doi:10.1007/s10489-020-02100-9.

[16] G. Yolcu, I. Oztel, S. Kazan, C. Oz, and F. Bunyak, "Deep learning-based face analysis

system for monitoring customer interest," J. Ambient Intell. Humaniz. Comput., vol. 11, no. 1, pp. 237–248, Jan. 2020, doi:10.1007/s12652-019-01310-5.

[17] S. Zhang, C. Chi, Z. Lei, and S. Z. Li, "RefineFace: Refinement Neural Network for High Performance Face Detection," IEEE Trans. Pattern Anal. Mach. Intell., p. 1, 2020, doi:10.1109/TPAMI.2020.2997456.

[18] M. Tao et al., "Smartphone-based detection of leaf color levels in rice plants," Comput. Electron. Agric., vol. 173, p. 105431, 2020, doi:10.1016/j.compag.2020.105431.

[19] P. Emami, P. M. Pardalos, L. Elefteriadou, and S. Ranka, "Machine Learning Methods for Data Association in Multi-Object Tracking," ACM Comput. Surv., vol. 53, no. 4, pp. 1–34, 2020, doi:10.1145/3394659.

[20] N. Al-Shakarji, F. Bunyak, H. Aliakbarpour, G. Seetharaman, and K. Palaniappan, "Multi-Cue Vehicle Detection for Semantic Video Compression in Georegistered Aerial Videos," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, 2019.

[21] G. V Santiago and A. J. Alvares, "Deployment framework for the Internet of water meters using computer vision on ARM platform," J. Ambient Intell. Smart Environ., vol. 12, no. 1, pp. 35–60, 2020, doi:10.3233/AIS-200544.

[22] K. Trejo, C. Angulo, S. Satoh, and M. Bono, "Towards robots reasoning about group behavior of museum visitors: Leader detection and group tracking," J. Ambient Intell. Smart Environ., vol. 10, no. 1, pp. 3–19, 2018, doi:10.3233/AIS-170467.

[23] H. Wu, L. Yang, S. Fu, and X. (Luke) Zhang, "Beyond remote control: Exploring natural gesture inputs for smart TV systems," J. Ambient Intell. Smart Environ., vol. 11, no. 4, pp. 335–354, 2019, doi:10.3233/AIS-190528.

[24] Y. Sun and Z. Yan, "Image target detection algorithm compression and pruning based on neural network," Comput. Sci. Inf. Syst., vol. 18, no. 2, pp. 499–516, 2021, doi:10.2298/csis200316007s.

[25] J. Ko and K. Cheoi, "A novel distant target region detection method using hybrid saliency-based attention model under complex textures," Comput. Sci. Inf. Syst., vol. 18, no. 2, pp. 379–399, 2021, doi:10.2298/csis200120001k.

[26] K. Ueki, T. Hayashida, and T. Kobayashi, "Subspace-based Age-group Classification Using Facial Images under Various Lighting Conditions," in 7th International Conference on Automatic Face and Gesture Recognition (FGR06), 2006, pp. 43–48, doi:10.1109/FGR.2006.102.

[27] K. S. Htet and M. M. Sein, "Market Intelligence Analysis on Age Estimation and Gender Classification on Events with deep learning hyperparameters optimization and SDN Controllers," in 2020 IEEE Conference on Computer Applications(ICCA), 2020, pp. 1–5, doi:10.1109/ICCA49400.2020.9022854.

[28] Z. Lin et al., "Establishment of age group classification for risk stratification in glioma patients," BMC Neurol., vol. 20, no. 1, p. 310, 2020, doi:10.1186/s12883-020-01888-w.

[29] E. Torres, S. L. Granizo, and M. Hernandez-Alvarez, "Gender and Age Classification Based on Human Features to Detect Illicit Activity in Suspicious Sites," in International Conference on Computational Science and Computational Intelligence (CSCI), 2019, pp. 416–419, doi:10.1109/CSCI49370.2019.00081.

[30] Y. H. Kwon and N. da V. Lobo, "Age Classification from Facial Images," Comput. Vis. Image Underst., vol. 74, no.

1, pp. 1–21, 1999, doi:10.1006/cviu.1997.0549.

[31] W. B. Horng, C. P. Lee, and C. W. Chen, "Classification of age groups based on facial features," Tamkang J. Sci. Eng., vol. 4, no. 3, pp. 183–192, 2001, doi:10.6180/jase.2001.4.3.05.

[32] P. Thukral, K. Mitra, and R. Chellappa, "A hierarchical approach for human age estimation," in IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2012, pp. 1529–1532, doi:10.1109/ICASSP.2012.6288182.

[33] A. Gunay and V. V Nabiyev, "Automatic age classification with LBP," in 23rd International Symposium on Computer and Information Sciences, 2008, pp. 1–4, doi:10.1109/ISCIS.2008.4717926.

[34] A. K. Soni, R. Kumar, and D. K. Kishore, "Estimation of age groups based on facial features," in 2015 International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT), 2015, pp. 681–687, doi:10.1109/ICATCCT.2015.7456970.

[35] D.-V. Bratu, S.-A. Moraru, and L. Georgeta Guseila, "A Performance Comparison between Deep Learning Network and Haar Cascade on an IoT Device," in 2019 International Conference on Sensing and Instrumentation in IoT Era (ISSI), 2019, pp. 1–6, doi:10.1109/ISSI47111.2019.9043714.

[36] Z. Zhang, Y. Song, and H. Qi, "Age Progression/Regression by Conditional Adversarial Autoencoder," in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 4352–4360, doi:10.1109/CVPR.2017.463.

[37] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 6, pp. 1137–1149, Jun. 2017, doi:10.1109/TPAMI.2016.2577031.

[38] J. Redmon, "Darknet: Open Source Neural Networks in C." 2016https://pjreddie.com/darknet/.

[39] J. Redmon and A. Farhadi, "YOLO9000: Better, Faster, Stronger," in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 6517–6525, doi:10.1109/CVPR.2017.690.

[40] I. Goodfellow, Y. Bengio, and A. Courville, Deep learning. MIT Press, 2016http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.672.7118%7B&%7Drep=rep1%7B&%7Dtype=pdf.

[41] R. Mostafiz, M. S. Uddin, N.-A.- Alam, M. Mahfuz Reza, and M. M. Rahman, "Covid-19 detection in chest X-ray through random forest classifier using a hybridization of deep CNN and DWT optimized features," J. King Saud Univ. - Comput. Inf. Sci., 2020, doi:https://doi.org/10.1016/j.jksuci.2020.12.010.

[42] R. Hafiz, M. R. Haque, A. Rakshit, and M. S. Uddin, "Image-based soft drink type classification and dietary assessment system using deep convolutional neural network with transfer learning," J. King Saud Univ. - Comput. Inf. Sci., 2020, doi:https://doi.org/10.1016/j.jksuci.2020.08.015.

[43] Y. M. Kassim et al., "Clustering-Based Dual Deep Learning Architecture for Detecting Red Blood Cells in Malaria Diagnostic Smears," IEEE J. Biomed. Heal. Informatics, p. 1, 2020, doi:10.1109/JBHI.2020.3034863.

[44] A. Vedaldi and K. Lenc, "MatConvNet," in Proceedings of the 23rd ACM international conference on Multimedia - MM '15, 2015, pp. 689–692, doi:10.1145/2733373.2807412.

[45] J. Roels, J. De Vylder, J. Aelterman, Y. Saeys, and W. Philips, "Convolutional neural network pruning to accelerate membrane segmentation in electron microscopy," in 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017), 2017, pp. 633–637, doi:10.1109/ISBI.2017.7950600.

[46] L. Torrey and J. Shavlik, "Transfer Learning," in Handbook of Research on Machine Learning Applications and Trends, IGI Global, pp. 242–264, doi:10.4018/978-1-60566-766-9.ch011.

[47] S. J. Pan and Q. Yang, "A Survey on Transfer Learning," IEEE Trans. Knowl. Data Eng., vol. 22, no. 10, pp. 1345–1359, 2010, doi:10.1109/TKDE.2009.191.

[48] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," Adv. Neural Inf. Process. Syst., pp. 1–9, 2012, doi:http://dx.doi.org/10.1016/j.protcy.2014.09.007.

[49] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," Sep. 2014http://arxiv.org/abs/1409.1556.

[50] F. N. Iandola, M. W. Moskewicz, K. Ashraf, S. Han, W. J. Dally, and K. Keutzer, "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <1MB model size," CoRR, vol. abs/1602.0, 2016.

[51] C. Szegedy et al., "Going deeper with convolutions," in 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 1–9, doi:10.1109/CVPR.2015.7298594.

[52] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 2818–2826, doi:10.1109/CVPR.2016.308.

[53] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks," in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 2261–2269, doi:10.1109/CVPR.2017.243.

[54] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," in 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 4510–4520, doi:10.1109/CVPR.2018.00474.

[55] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770–778, doi:10.1109/CVPR.2016.90.

[56] F. Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions," 2016http://arxiv.org/abs/1610.02357.

[57] X. Zhang, X. Zhou, M. Lin, and J. Sun, "ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices," 2017http://arxiv.org/abs/1707.01083.

[58] Kaggle, "Face Mask Detection," 2020. [Online]. Available: https://www.kaggle.com/andrewmvd/face-mask-detectionhttps://www.kaggle.com/andrewmvd/face-mask-detection.

[59] M. Loey, G. Manogaran, M. H. N. Taha, and N. E. M. Khalifa, "Fighting against COVID-19: A novel deep learning model based on YOLO-v2 with ResNet-50 for medical face mask detection," Sustain. Cities Soc., p. 102600, 2020, doi:10.1016/j.scs.2020.102600.

[60] S. Ge, J. Li, Q. Ye, and Z. Luo, "Detecting Masked Faces in the Wild with LLE-CNNs," in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 426–434, doi:10.1109/CVPR.2017.53.

[61] A. Das, A. Dantcheva, and F. Bremond, "Mitigating Bias in Gender, Age and Ethnicity Classification: a Multi-Task Convolution Neural Network Approach," in Proceedings of the European Conference on Computer Vision (ECCV) Workshops, 2018.

[62] S. H. Lee and Y. M. Ro, "Local age group modeling in unconstrained face images for facial age classification," in International Conference on Image Processing (ICIP), 2014, pp. 1395–1399, doi:10.1109/ICIP.2014.7025279.

[63] L. Liu, J. Liu, and J. Cheng, "Age-Group Classification of Facial Images," in 11th International Conference on Machine Learning and Applications, 2012, pp. 693–696, doi:10.1109/ICMLA.2012.129.