Facial mask impact on human age and gender classification

Krzysztof Małecki[1] [0000−0002−8687−1119], Adam Nowosielski[1] [0000−0001−7729−7867], and Mateusz Krzak[2]

[1] West Pomeranian University of Technology, Faculty of Computer Science and Information Technology, Żołnierska 52 Str., 71-210 Szczecin, Poland
{kmalecki,anowosielski}@wi.zut.edu.pl

[2] Netcompany Poland Sp.z o.o., Puławska 182 Str., 02-670 Warsaw, Poland
mateuszkrzak5@gmail.com

Abstract. The human face contains important information enabling the social identification of the owner about the age and gender. In technical systems, the face contains a number of important information that enables the identification of a person. The COVID-19 pandemic made it necessary to cover the face with a mask and thus hide a significant part of information content in the face, important for social or technical purposes. The paper analyses how covering the face with a mask makes it difficult to identify a person in terms of age and gender determination. Analyzes with the employment of state of the art models based on deep neural networks are performed. Their effectiveness is investigated in the context of the limited information available, as with the case of the face covered with a mask.

Keywords: Face recognition · Deep learning · COVID-19

1 Introduction

The human face contains a lot of essential information. The basic information that can be read from the face is age and gender. From the face, it is also possible to read emotions. In everyday life a person can be identified by others on the basis of the face. All the information contained in the face is also more and more often used in technical systems [1–4], especially those based on multispectral data [5]. Face ID allows to unlock the device or provide access to a security restricted area. Face can be utilized as the medium in touchless human-computer interfaces [6]. Novel HCI interfaces can read emotions of its user and adapt their behaviour and react appropriately to the emotional state of a human. Finally, from facial expressions the recognition of student emotions during a lecture is possible [7], providing instant feedback to the lecturer and contributing to the increase in the quality of education.

COVID-19 pandemic has changed how the image of the face functions in a society. Obligatory wearing of a mask to stop the spread of the virus made a significant influence on how people engage in social activities. Significant part

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08751-6_51

of the information contained in the face is hidden behind a mask. The areas around the mouth is highly informative providing clues for emotional state, and also premises for determining age and sex. By covering the face with a mask these hints become unavailable. Face-oriented algorithms also assume a particular level of visibility of facial features. Subjected to faces covered with masks might fail to operate.

The problem of facial mask impact on reading information contained in the human face has been already addressed in the literature. One of the first publications were dedicated to detecting a face covered with a mask, e.g. [8–10]. Wearing a mask incorrectly makes the prevention method pointless therefore algorithms for detection when facemasks are being worn incorrectly emerge [11]. The National Institute of Standards and Technology (NIST) started examining the performance of face recognition algorithms on faces occluded by masks [12], run under the ongoing Face Recognition Vendor Test (FRVT). In the augmented report [13] (after the COVID-19 pandemic was declared) NIST observed that developers have adapted their algorithms to support face recognition on subjects potentially wearing face masks [13]. Despite this, the performance of the evaluated recognition algorithms is, in general, deteriorated. A detailed overview of the progress in masked face recognition can be found in [14].

In the paper, we analyze the impact of the face mask on reading age and gender information from a face. The problem alone is challenging. The appearance of the human face, apart from the congenital determinants, is influenced by a lifestyle, environmental conditions, health, emotional state and others. A facial mask, by concealing a large part of the information, makes the task even more difficult. In our research, we focus on deep learning algorithms. With the help of Tensorflow and Keras libraries, different models were utilized and trained. Their effectiveness was evaluated on prepared database.

2 Idea

In the model training process photos of computer-generated faces based on the artificial intelligence algorithm StyleGAN2 [15] developed by NVIDIA were used. For the purposes of the article, a computer program has been prepared that automatically downloads photos from the website [16] with the age and gender information. Then, the Face-Recognition-Model-with-Gender-Age-and-Emotions-Estimations computer program [17] available on GitHub was modified and used, and the photos were divided into appropriate age and gender categories, saving this data to a CSV file. The computer-generated photos present the faces of people aged 9 to 74, both male and female.

The photo database consists of 8,000 photos divided into gender categories: female and male. Then, of these 8,000 images, 4 age groups (0-12, 13-20, 21-50, 51+) of 1500 images each were separated. Groups of equal size were chosen to guarantee the same training and testing conditions for each case. Furthermore, both sex categories have been split into 3000 training images, 500 validation, and

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08751-6_51

500 pictures for testing. Similar division was applied to each of the age groups with distribution values of: 1000, 250, and 250 appropriately.

Fig. 1. Two types of masks used: (a) a textile, (b) a surgical.

In the next step of preparing the benchmark database, the MaskTheFace program was used to apply face protective masks [18] available on the GitHub [19] platform, under the MIT license. In order to differentiate the research material, two types of masks were used: a textile mask and a surgical mask (Fig. 1). It must be noticed, that procedure of adding artificial masks to face images, instead of acquiring real data, is found in the literature (e.g. [12, 13]). The process of obtaining real photos is expensive and time-consuming, the face databases already exist, and the mask only hides the information already available. Some researchers use the Internet community to obtain data (e.g. [11] use citizens by asking to take different selfies through an app and placing the mask in different positions).

Fig. 2. Sample images from benchmark database containing faces without masks and with different categories of masks.

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08751-6_51

To illustrate the established and diversified benchmark database, the authors presented selected images of the faces without protective masks, with textile protective masks in a uniform black color, with different types of masks and different colors and shades, and faces of people with masks of different colors and textures (Fig. 2).

Four different training models have been utilized in this paper: a deep neural network (DNN) model and three convolutional neural network models - containing one (CNN1), three (CNN3) and four convolutional layers (CNN4). The DNN model consisted of seven layers, including four Dense layers (fully-connected layers): 2nd, 4th, 6th and 7th. The CNN1 model contained only one Conv2D layer (layer 1) and one MaxPooling2D layer (layer 2) and one flatten layer and three dense layers. The CNN3 model consisted of three alternating layers Conv2D (layers 1, 3, 5) and MaxPooling2D (layers 2, 4, 6) and four final layers: flatten (layer 7), dense (layers: 8, 10), dropout (last layer). Finally, the CNN4 model consisted of four alternating Conv2D layers (layers: 1, 3, 5, 7) and MaxPooling2D (layers: 2, 4, 6, 8), one flatten layer and four dense layers.

All models have been executed with the same parameters. The number of epochs was 150 and the number of samples taken from the generator before the end of the epoch was 75. Uploaded images were scaled so that the input object size was 64x64x3. The output layer of each network was activated with the softmax function. For classifiers containing convolutional layers, the Adam optimizer with a value of 0.0001 was selected. The deep neural network (DNN) model had the RMSprop optimizer applied.

3 Baseline validation

For comparison purposes, the previously mentioned Face-Recognition-Model-with-Gender-Age-and-Emotions-Estimations trained model [17] was executed on faces without applied masks.The original model obtained values above 90% for both gender and age groups. The detailed results of this experiment when no masks were worn are shown in Figure 3 for age groups and in Figure 4 for gender.

For age groups, all models performed well in correct assignment for two age groups: 0-12 and 51+. For the 0-12 age group, the percentage results of correct assignment are presented as DNN: 82%, CNN1: 91.2%, CNN3: 89.2%, CNN4: 90%, respectively. In the 51+ group, the models obtained similar values, i.e. DNN: 84.8%, CNN1: 93.2%, CNN3: 87.2%, CNN4: 90.8%. For groups 13-20 and 21-50, the classifiers performed significantly less well in correctly assigning categories. In the 13-20 group, of all the models, CNN4 performed the best with an efficiency of 46.8%. In the 21-50 group, CNN1 had the best efficiency with a value of 72.8%.

The correctness of gender assignment is high. This means that all classifiers performed well with the following percentages. The DNN model achieved an efficiency of 83.2%, CNN1: 81.4%, CNN3: 87.6%, CNN4: 90.2% in correctly assigning women. For men, the values were from DNN: 85.8%, CNN1: 88.6%, CNN3: 88.8%, CNN4: 89%.

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08751-6_51

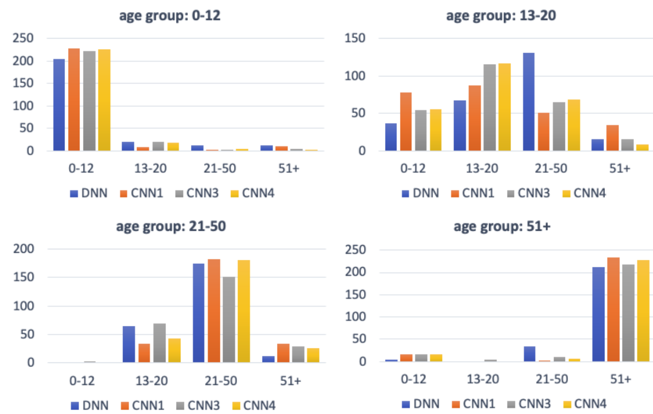

Fig. 3. Assigned categories for age groups with no masks applied

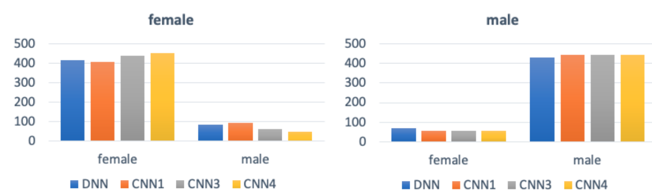

Fig. 4. Assigned categories for gender with no masks applied

The results obtained above will be considered as the baseline and the results of subsequent experiments will be compared to them. In the next step, the prepared models are evaluated for the correct assignment of people's faces with masks to the appropriate class categories.

4 The results of age groups and gender classification

This section presents the results of the study on the recognition performance of faces obscured by black cloth masks (sec. 4.1), cloth and surgical masks with different colors (sec. 4.2), and cloth and surgical masks with different colors and textures (sec. 4.3), by individual classifiers (DNN, CNN1, CNN3, CNN4).

4.1 Black fabric masks

In the task of classifying face images obscured by uniform black cloth masks (Fig. 5), models based on convolutional layers achieved the highest performance: 80.8% (CNN3) and 98.4% (CNN4). The DNN model was least effective in the task of classifying into the appropriate age group, generating the following results: 59.6% (for 0-12 group), 41.2% (for 13-20), 5.2% (for 21-50), 90.8% (for 51+).

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08751-6_51

In the gender classification task (Fig. 6), correct assignment for women equals: 96.4% (DNN), 98% (CNN1), 95.8% (CNN3) and 98% (CNN4). An interesting result is the score of 78% obtained by the DNN classifier in the task of classifying a face image into a group of men. The other classifiers obtained very close results as those of the female group.

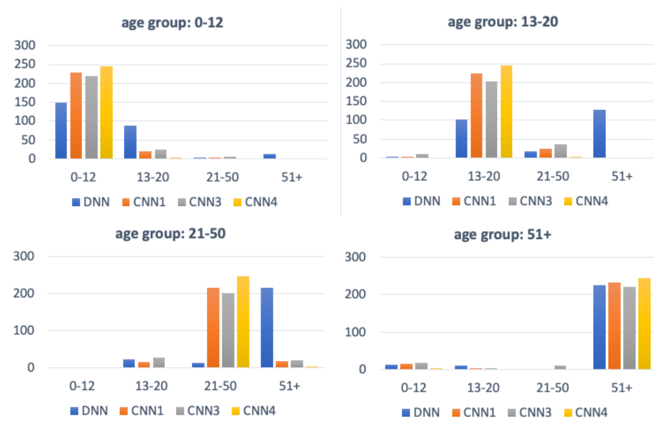

Fig. 5. Assigned categories for age groups with black masks applied.

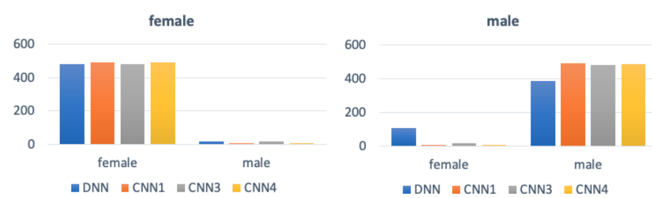

Fig. 6. Assigned categories for gender with black masks applied.

4.2 Masks in different colors

In the classification of face images covered by different types of masks of different colors, the best results were obtained for two age groups: 0-12 and 51+. These were respectively: 78% (DNN), 92% (CNN1), 89.6% (CNN3), 90.4% (CNN4) and 91.2% (DNN), 91.2% (CNN1), 90.4% (CNN3), 86.4% (CNN4). The effectiveness in the 21-50 age group for the DNN model was only 28.8%, while the other classifiers gave results ranging from 61.2% to 76%. The lowest recognition efficiency occurred in the group of 13-20. The DNN model obtained: 41.6%, CNN1: 28.4%, CNN3: 30.4%, CNN4: 44% (Fig. 7).

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08751-6_51

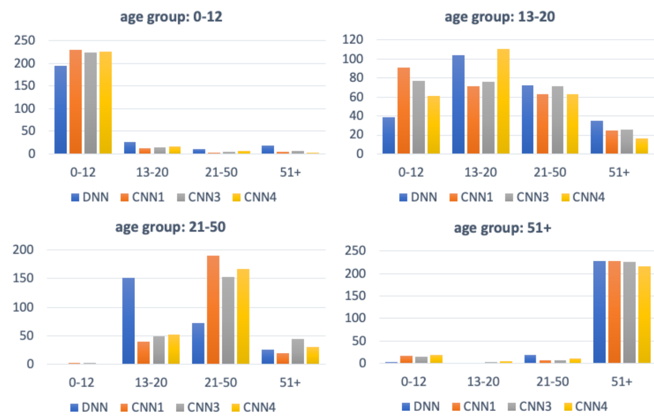

Fig. 7. Assigned categories for age groups with masks in different colors applied.

The effectiveness of correctly assigning men was high for all models (Fig. 8): 95.8% (DNN), 82.6% (CNN1), 92.2% (CNN3), 89.6% (CNN4). In the female group, the convolutional models performed equally well: 87.6% (CNN1), 83.6% (CNN3), 84.6% (CNN4). In contrast, the DNN model scored 48.4%. This classifier may not have learned to recognize certain features such as short hair, which can occur in both men and women.

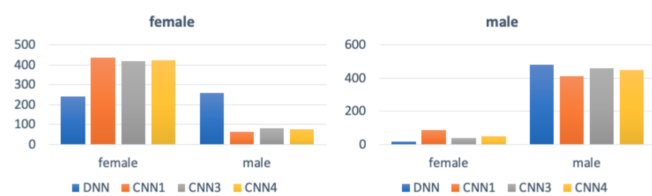

Fig. 8. Assigned categories for gender with masks in different colors applied.

4.3 Masks in different colors and textures

In the task of classifying images of faces covered by different types of masks of different colors and textures, the highest performance was observed for the age groups 0-12 and 51+ (Fig. 9). In detail, DNN: 97.6%, CNN1: 89.6%, CNN3: 91.6%, CNN4: 92.4%. The experiment highlighted the problem of image classification for the age group 13-20, for which the highest efficiency is 44.4% (CNN4) and for the age group 21-50, for which the effectiveness oscillates between 56.4% and 70%. It can be concluded that the patterns on the masks disrupted the correct learning process of the models.

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08751-6_51

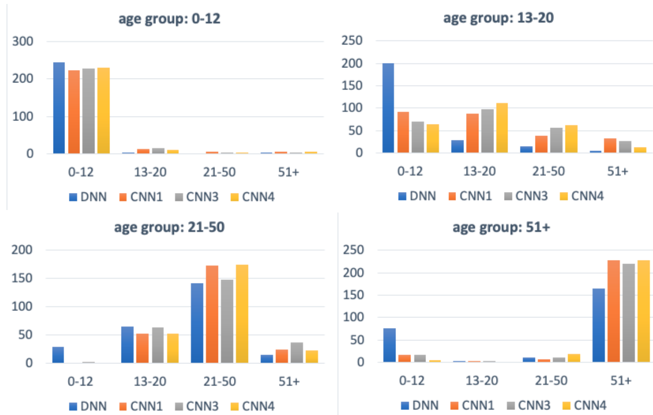

Fig. 9. Assigned categories for age groups with masks in different colors and textures applied.

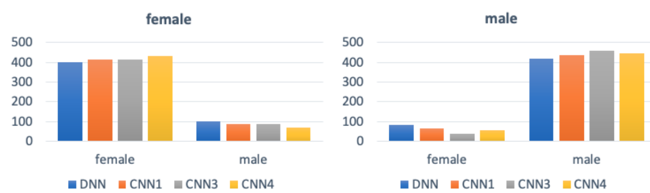

Fig. 10. Assigned categories for gender with masks in different colors and textures applied.

In the gender classification task (Fig. 10), for the women's group, the results were 80.2% (DNN), 83% (CNN1), 82.4% (CNN3), 86.4% (CNN4). The results for the men's group, on the other hand, are as follows: 83.6% (DNN), 87.4% (CNN1), 92% (CNN3), 89.2% (CNN4).

5 The performance results targeted at each model separately

In the following subsections we analyse the classification performance of different models separately.

5.1 Deep neural network model (DNN)

For the DNN classifier, the results for age recognition are shown in Fig. 11 and for gender recognition in Fig. 12.

In the 51+ age group, the classifier achieved face recognition performance: 84.8% (no masks), 90.8% (uniform black masks), 91.2% (color masks) and 65.6%

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08751-6_51

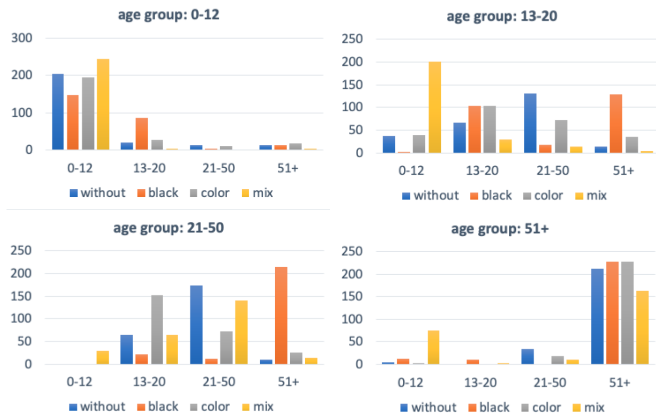

Fig. 11. The results of the DNN model for various age groups and masks.

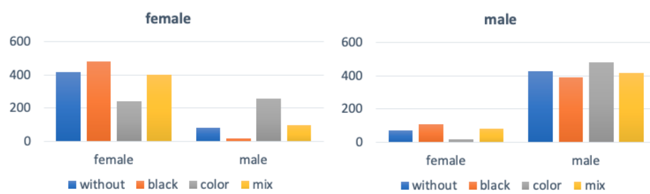

Fig. 12. The results of the DNN - gender classification task.

(color masks with different textures). For the 0-12 age group, the following results were achieved, respectively: 82%, 59.6%, 78% i 97.6%. In the case of the other age groups, the model achieved relatively low efficiency. In the male classification task, the DNN model achieved an efficiency of 85.8% (no masks), 78% (solid black masks), 95.8% (color masks), and 83.6% (color masks with different textures). In the assignment to the group of women, the results were as follows: for no masks - 83.2%, 96.4% for uniform black masks, 48.4% for color masks, 80.2% for color masks with different textures.

5.2 One convolutional layer model (CNN1)

For the 0-12 and 51+ groups, the type of mask applied, or lack thereof, had no significant effect on the correctness of imputation the assignment (Fig. 13). In the 0-12 age group the results are respectively: 91.2% (no masks), 91.6% (uniform black masks), 92% (color masks), 89.6% (color masks with different textures). Whereas for the group 51+ it will be respectively: 93.2%, 93.6%, 91.2%, 90.8%. Lower classification efficiency was observed for people in the age group of 21-50. Complication of the mask appearance by applying different colors, shades and textures causes a decrease in accuracy: 86.4% for black masks, 76% for color masks and finally 69.2% different texture masks. In the 13-20 age group, only

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08751-6_51

one case (face images with uniform black masks) had the highest classification of 90%. For the remaining cases, the results did not exceed 40%.

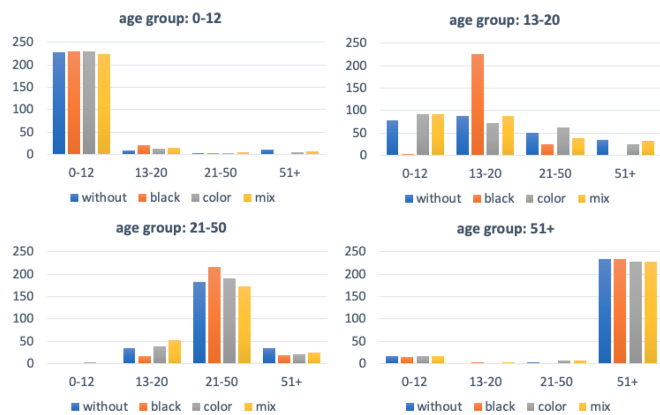

Fig. 13. The results of the CNN1 model for various age groups and masks.

In the case of gender classification (Fig. 14), in a group of women, varied appearance of the mask causes a slight decrease in accuracy: from 98% for black masks, through 87.6% for color masks, and ending on 83% for different textures masks. In the male classification, the matching to group is highest for face images masked with black masks and equal to 98.2%.

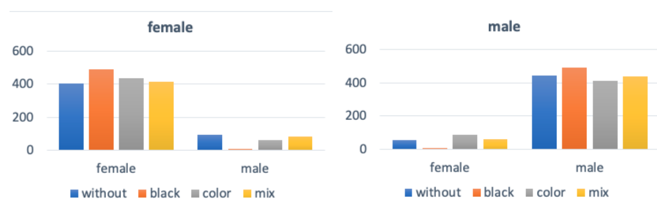

Fig. 14. The results of the CNN1 - gender classification task.

5.3 Three convolutional layers model (CNN3)

Face classification for all types of masks for groups: 0-12 and 51+ indicated that the type of mask applied had no significant effect on the CNN3 model. In the 0-12 group for no masks case, the model achieved an efficiency of 89.2%, for uniform black masks - 88%, color masks - 89.6%, and for mixed masks - 91.6%. For the 51+ age group, similar results were achieved. In the 13-20 group, the highest

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08751-6_51

efficacy was achieved for black masks (80.8%). Classification of the remaining cases ended with efficiencies ranging from 30.4% for color masks, through 39.2% for mixed masks, and obtaining 46.4% for the no masks case. In the 21-50 group, recognition of faces covered by uniform black masks ended with an efficiency of 80.8%. Classification efficiency for the remaining cases ranged from 59.2% to 61.2% (Fig. 15).

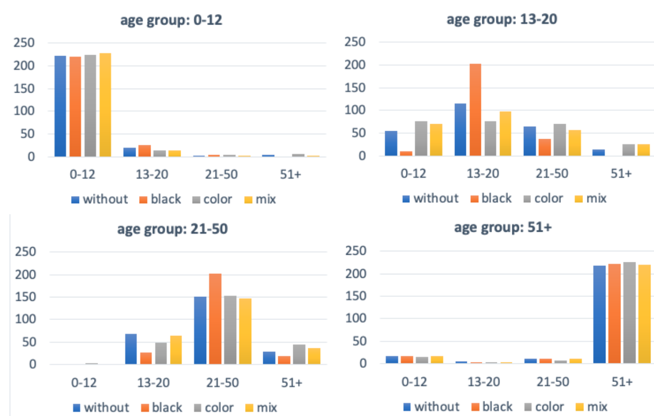

Fig. 15. The results of the CNN3 model for various age groups and masks.

Fig. 16 shows the results of correct and incorrect face assignment to the group of women and men. For the women's group the values are respectively: without mask applied - 87.6%, uniform black masks - 95.8%, color masks - 89.6% and mixed masks - 91.6%. For the group of men, the results are as follows: for photos of the face without the mask on - 88.8%, with black masks - 96.2%, for color masks - 92.2% and 92% for different textures masks.

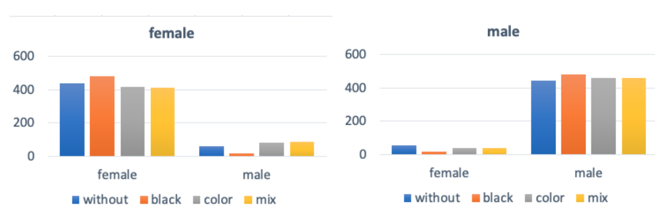

Fig. 16. The results of the CNN3 - gender classification task.

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08751-6_51

5.4 Four convolutional layers model (CNN4)

In the experiments for the classifier referred to as CNN4 results are presented in (Fig. 17) for age groups, and in in (Fig. 18) for gender.

For face classification to the 0-12 age group, all test cases showed high model performance: 90% (no masks), 98.4% (uniform black masks), 90.4% (color masks) and 92.4% (mixed masks). Almost identical results were obtained for the 51+ age group. For the other age groups, the CNN4 model achieved the best classification results for black masks. A large difference in classification effectiveness can be observed in the 13-20 age group: 98.4% (black masks), 46.8% (no masks), 44% (color masks), 44.4% (mixed masks). For the 21-50 group, and black masks the accuracy equals 98.4%, 72.4% for no masks case, 66.8% for color masks, and 70% for mixed masks case (Fig. 17).

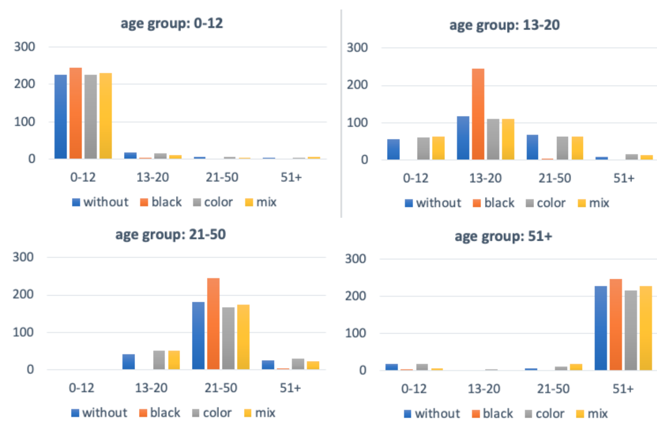

Fig. 17. The results of the CNN4 model for various age groups and masks.

Gender classification task (Fig. 18), for both groups ended with a high success rate. For a group of women: 90.2% (no masks), 98% (black masks), 84.6% (color masks), and 86.4% (mixed masks). For the male group, the classification results are similar and are as follows: for facial images without masks - 89%, black masks - 98%, color masks - 89.6%, and 89.2% for different textures masks.

6 Conclusion

The purpose of this study was to investigate the effect of facial masking on automatic estimation of a person's age and gender. The effectiveness of four deep learning models has been examined. From the study, some conclusions can be drawn regarding the human features contained in faces.

The main, rather expected, result is an overall deterioration in the age and gender classification of masked face subjects. Masks obscure facial hair, which

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08751-6_51

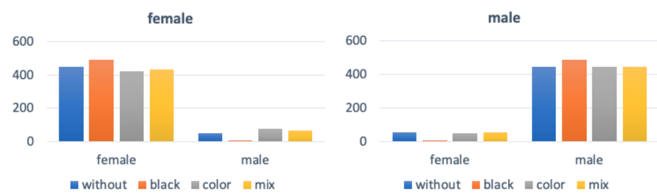

Fig. 18. The results of the CNN4 - gender classification task.

is a special feature for men, making it easy to quickly categorize them. Long hair, characteristic of women but found in men, influences misclassification into the wrong group. And in reverse, short hair in women provides a valid rationale for classification into the male group, especially when other facial features are obscured by a mask. The differences that occur between the different variants of the masks applied indicate the limited learning capabilities of the existing models. This is especially true for masks with miscellaneous textures. This is important from the point of view of the increasingly popular face monitoring and recognition systems. Nowadays, when it is common to mask parts of the face with a protective mask, it seems that detecting the mask itself and removing it programmatically, especially for different colored masks with complex textures, could improve the results of person identification or age and gender classification.

As for age, the situation is similar. For some age categories (13–20 and 21–50) all prepared models did not learn to recognize them, which could be due to the fact that the protective mask obscures large facial features such as lips, nose, and wrinkles making the amount of information the model processes limited as well. Interestingly, for convolutional models (CNNs) with different numbers of layers, in the vast majority of cases, the black masks performed better than the others.

It can be concluded from our study that DNN model was found to be more susceptible to face masks, especially those which additionally contained textures. Convolutional networks proved high performance in the task of classifying face images for the 0-12 and 51+ age groups, despite covering these faces with different masks.

References

1. Forczmański, P., Łabędź, P.: Recognition of occluded faces based on multi-subspace classification. In: Saeed, K., Chaki, R., Cortesi, A., Wierzchoń, S. (eds.) CISIM 2013. LNCS, vol. 8104, pp. 148—157. Springer, Heidelberg (2013).
2. Wojciechowska, A., Choraś, M., Kozik, R. The overview of trends and challenges in mobile biometrics. Journal of Applied Mathematics and Computational Mechanics **16**(2), 173–185 (2017).
3. Cyganek, B., Gruszczyński, S. Hybrid computer vision system for drivers' eye recognition and fatigue monitoring. Neurocomputing **126**, 78–94 (2014)

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08751-6_51

4. Małecki K., Forczmański P., Nowosielski A., Smoliński A., Ozga D. A New Benchmark Collection for Driver Fatigue Research Based on Thermal, Depth Map and Visible Light Imagery. In: Burduk R., Kurzynski M., Wozniak M. (eds) Progress in Computer Recognition Systems. CORES 2019. Advances in Intelligent Systems and Computing, vol 977, pp. 295–304. Springer, Cham (2020).
5. Małecki K., Nowosielski A., Forczmański P. Multispectral Data Acquisition in the Assessment of Driver's Fatigue. In: Mikulski J. (eds) Smart Solutions in Today's Transport. TST 2017. Communications in Computer and Information Science, vol 715. Springer, Cham (2017)
6. Nowosielski, A., Forczmański, P. Touchless typing with head movements captured in thermal spectrum. Pattern Analysis and Applications **22**(3), 841-855 (2019)
7. Tonguç, G., Ozkara, B. O. Automatic recognition of student emotions from facial expressions during a lecture. Computers Education **148**, 103797 (2020)
8. Sethi, S., Kathuria, M., Kaushik, T. Face mask detection using deep learning: An approach to reduce risk of Coronavirus spread. Journal of biomedical informatics, **120**, 103848 (2021)
9. Kansal I., Popli R., Singla C.: Comparative Analysis of various Machine and Deep Learning Models for Face Mask Detection using Digital Images. In: 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), pp. 1–5. (2021)
10. Joshi A. S. , Joshi S. S., Kanahasabai G., Kapil R., Gupta S.: Deep Learning Framework to Detect Face Masks from Video Footage. In: 12th International Conference on Computational Intelligence and Communication Networks (CICN), pp. 435–440. (2020)
11. Tomás, J., Rego, A., Viciano-Tudela, S., Lloret, J. Incorrect Facemask-Wearing Detection Using Convolutional Neural Networks with Transfer Learning. Healthcare **9**, 1050 (2021)
12. Ngan, M., Grother, P., Hanaoka, K. Ongoing Face Recognition Vendor Test (FRVT) Part 6A: Face recognition accuracy with masks using pre-COVID-19 algorithms, NIST Interagency/Internal Report (NISTIR), National Institute of Standards and Technology, Gaithersburg, MD (2020), [online], https://doi.org/10.6028/NIST.IR.8311 (Accessed January 10, 2022)
13. Ngan, M., Grother, P., Hanaoka, K. Ongoing Face Recognition Vendor Test (FRVT) Part 6B: Face recognition accuracy with face masks using post-COVID-19 algorithms, NIST Interagency/Internal Report (NISTIR), National Institute of Standards and Technology, Gaithersburg, MD (2020), [online], https://doi.org/10.6028/NIST.IR.8331 (Accessed January 11, 2022)
14. Alzu'bi, A., Albalas, F., AL-Hadhrami, T., Younis, L.B., Bashayreh, A. Masked Face Recognition Using Deep Learning: A Review. Electronics **10**, 2666 (2021)
15. Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., Aila, T.: Analyzing and Improving the Image Quality of StyleGAN. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 8110-8119 (2020)
16. Benchmark, https://www.thispersondoesnotexist.com [Access: 8.01.2022]
17. GitHub Face Recognition, https://www.github.com/rileykwok/Face-Recognition-Model-with-Gender-Age-and-Emotions-Estimations [Access: 8.01.2022]
18. Anwar, A., and Raychowdhury, A.: Masked Face Recognition for Secure Authentication. arXiv preprint arXiv:2008.11104 (2020)
19. MaskTheFace, https://www.github.com/aqeelanwar/MaskTheFace [Access: 8.01.2022]

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08751-6_51