Definition of Unique Objects by Convolutional Neural Networks using Transfer Learning

Rusakov K.D[1], Seliverstov D.E[2]
V.A. Trapeznikov Institute of Control Sciences of Russian Academy of Sciences, 65 Profsoyuznayastreet
Moscow 117997, Russia[1]
Plekhanov Russian University of Economics
36 Stremyannylane, Moscow, 117997, Russia[2]

Osipov V.V[3], Reshetnikov V.N[4]
Federal State Institution
"Scientific Research Institute for System Analysis of the Russian Academy of Sciences"
36 Nahimovskiipr-t, Moscow, 117218, Russia

***Abstract*—This article solves the problem of detecting medical masks on a person's face. Medical mask is one of the most effective measures to prevent infection with COVID-19, and its automatic detection is an actual task. The introduction of automatic recognition of medical masks in existing information security systems will allow quickly identify the violator of the mask regime, which in turn will increase security in a pandemic. The article provides a detailed analysis of existing solutions for face detection and automatic recognition of medical masks, method based on the use of convolutional neural networks was proposed. A distinctive feature of the new method is the use of two neural networks at once, using the RetinaFace neural network architecture at the face search stage and using the Resnet neural network architecture at the face mask recognition stage. It is shown that the use of transfer learning on scales, learned to work with faces, significantly accelerates learning and increases the accuracy of recognition. However, with this approach, there are some false positives, for example, when you try to cover your face with your hands, imitating a medical mask. Based on the study, we can conclude that the algorithm is applicable in the security system to determine the presence/absence of a medical mask on a person's face, as well as the need for additional research to solve the problems of false positives of the algorithm.**

***Keywords*—Recognition of medical masks; COVID-19; convolutional neural networks; retina face; Resnet**

I. Introduction

Nowadays the task of recognition the presence of a medical mask on a person's face has become very relevant in the condition of growing incidence of the new COVID-19 coronavirus infection. Medical mask is one of the effective measures for the prevention infection with COVID-19, its use minimizes the risk of spreading this disease.

Under the recognition task have a medical mask on the face of the man in this article means the following: on the input image to find metabologia all persons, for each person to determine the existence of face masks and give a certain confidence to this event.

The solution to the problem of recognizing a medical mask is also solved using convolutional neural networks. So, in [1], the authors solve the problem of face recognition in medical masks. In the real world, when a person tries to hide from systems such as video surveillance, having a face mask is one of the most common ways. If you have a medical mask on your face, the accuracy of facial recognition is reduced. The authors have conducted many studies on face recognition in various conditions, such as changes in posture or lighting, image degradation, and so on. The focus of this work is on medical masks, and especially improving the accuracy of facial recognition in medical masks. The authors solved the problem of detecting masked faces using a multitasking cascading convolutional neural network (MTCNN). Then, facial features are extracted using the Google FaceNet model. Finally, the classification task was performed by the authors using the Support Vector Machine (SVM). In [2], it is discussed that in recent years, face recognition has become a very difficult task due to various types of occlusion or masks, such as sunglasses, scarves, hats, and various types of makeup or disguise. All this affects the accuracy of facial recognition. Despite the fact that many algorithms for face recognition have been developed recently, which are widely used and provide better performance, little has been done in the field of face recognition in masks. Therefore, in this work, the authors chose a statistical procedure that is used in the recognition of unmasked faces, as well as used in the technique of face recognition in masks. RSA is a more efficient and successful statistical method that is widely used. For this reason, the RSA algorithm was chosen in this paper. Finally, there is also a comparative study for better understanding. In [3], the authors propose a new cascade structure based on convolutional neural networks, which consists of three carefully designed convolutional neural networks for detecting masked faces. The authors in their work talk about the applicability of the algorithm for tracking and identifying criminals or terrorists.

As you can see, the task confirms its relevance. Despite the relative simplicity of the task, recognizing medical masks on faces involves solving a number of non-trivial issues. Due to the widespread introduction and development of new information technologies, namely neural network approaches, in many areas of human life, there is a new task of automatic detection of a medical mask on a person's face, which will automatically monitor compliance with the mask mode.

II. Review of Existing Approaches to the Recognition of Medical Masks

From the analysis of various algorithms and methods for recognizing medical masks, it follows that in the structural

scheme of any image recognition method, as a rule, the following two typical types of mask recognition algorithms can be distinguished:

1) Search for a face in the image and then determine the mask on a person's face;

2) Search for a mask in an image without first identifying the face.

In addition, the construction of an algorithm for recognizing medical masks is based on a priori information about the subject area (in this case, on the characteristics of the person's face and the type of mask) and is corrected by empirical information that appears during the development of the algorithm.

The purpose of the face detection process (Fig. 1) is to localize all areas of the image that may contain a face, regardless of external lighting conditions, occlusion, etc. Despite the presence of distinctive features in the facial structure, this is quite a difficult task, since the features vary greatly depending on gender, skin color, and facial expression.

A significant number of factors can affect the detection of faces in a photo [4]:

1) Face position: the face in the image can be rotated at any angle of pitch, yaw, or roll.

2) Some parts of the face may be partially covered, which greatly complicates the ability to detect faces.

3) Different lighting conditions (the type the amount and direction of light sources, the color and brightness, the presence of shadows, color balance of camera, image distortion introduced by the optical system, etc.). For example, when lighting is used, the part of the face is very bright, while the other part is very dark, and it can influence the result.

4) The presence of normal or sunglasses, beards, moustaches, and various accessories make more errors in face detection.

5) The face size can change many times depending on the distance to the image.

6) Faces can be placed on different backgrounds: fixed, low contrast, noisy, etc., which can also make an error in the result of the face detection algorithm.

7) Different expressions and emotions: laughter, anger, surprise, etc., which can also affect the detection of faces.

Since face detection algorithms require a priori information about the face [5], the following categories of methods can be distinguished:

1) Feature-based methods (Fig. 2). For this category of methods, there are three areas, one of which they are based on: high-level feature analysis, low-level analysis, and methods based on form modeling. As a rule, the task of determining the face in an image is associated with the localization of this face on a complex background. Low-level analysis is based on feature segmentation based on image finite elements. Feature-based methods perform operations on image regions, taking into account the geometry and semantics of the face. The representativeness of these features is slightly higher in comparison with low-level analysis.

2) Image classification methods (Fig. 3). These methods are based on the construction of implicit facial patterns based on machine learning and modeling methods. This group of methods uses image recognition tools, presenting the face detection problem as a special case of the recognition problem.

After preliminary selection of the face on the image, it is binary classified for the presence or absence of a medical mask. In order to classify an image as «there is a mask/no mask», the face image is correlated with a feature vector calculated in some way. The most common way to get such a vector is to use the face image itself, with each pixel becoming a component of the vector. The main disadvantage of this method of representation is the extremely high dimension of the feature space, which increases the demand for computing power of the equipment. The advantage is a fairly high accuracy of classification. Thus, the task of automatically classifying images of people's faces by the values «there is a mask/no mask» is a typical task of learning from use cases [6].

The second method of determining the mask on the face is the direct determination of the medical mask on the image immediately, without first determining the faces.

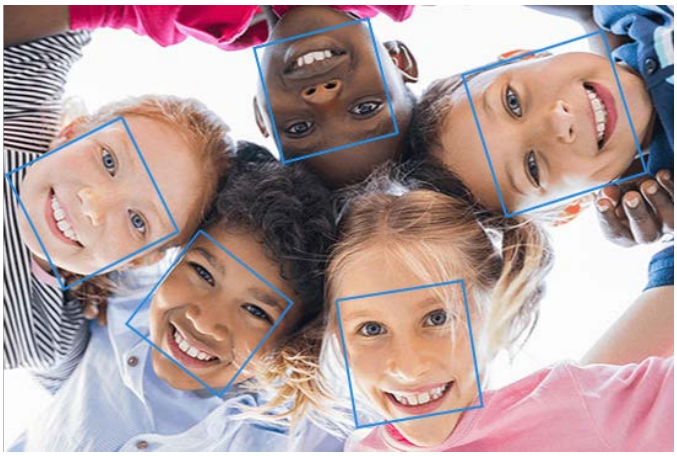

Fig. 1. Face Detection.

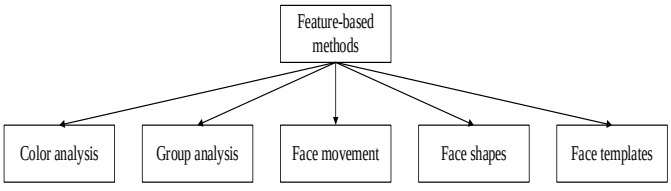

Fig. 2. Structure of Feature-based Methods.

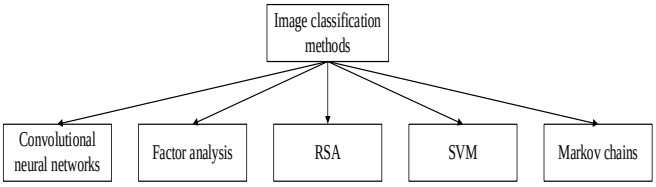

Fig. 3. Structure of Image Classification Methods.

From the analysis, it was revealed that to solve the problem of face detection, it is most preferable to choose the neural network approach as the main method. This choice was made based on the fact that this method provides:

1) The ability to detect a large number of faces in the image.

2) Ability to process images where the face angle is different from the front (with a yaw angle of up to 90°).

3) Ability to predict key points of the face to align it.

III. Architecture of the Medical Mask Recognition Solution

The proposed architecture consists of the following main elements: face detection by a fast one-step detector, face alignment, and face classification by a light convolutional neural network trained using transfer learning on scales for face recognition.

A. Detection and Face Alignment

At the first stage, the RetinaFace neural network architecture was chosen as the basis for the face search algorithm in the image. RetinaFace is a neural network architecture based on the detection and classification of objects at different levels of the "main highway" architecture. The Resnet50 network was used as the main neural network required for feature extraction. Distinctive feature: distinguishing faces in one run using the specified grid of Windows (default box) on the image pyramid (Fig. 4). In this way, both large and small faces are detected in a single network run.

In [7], it is shown that this solution solves the problem of finding and localizing a face quickly and accurately.

Since RetinaFace allows you to predict not only the localization of the face, but also the five key points of the face: eyes, nose and corners of the mouth, we can determine the angle of the face roll as follows:

$$\theta = \arctan(\frac{I}{M})$$

I – the distance between the eyes;

M – the distance between the corners of the mouth.

After getting information about the angle, we can align the image of the face so that the eye level is on a straight line parallel to the abscissa axis. In other words, as a result of this operation, all faces will be aligned.

B. Detection of Mask on Face

Finally, the process of recognizing the presence of a medical mask on the face s determined by learning the classifier in the form of a light convolutional neural network [8, 9, 10]. Convolutional layers transform the input image into a feature vector that is passed through a linear layer [11] of the following type (Fig. 5):

In the case of these classifications, the quality functional is obvious and, therefore, often used [12]:

$$Q(f_{\phi(x)}, I^{train}, I^{test})$$

which determines the percentage of correctly classified objects that are represented in the use cases of the I^{test} test sample constructed using the formalized φ(x) method of the I^{train} learning sample.

Algorithm 1 The Pseudo-code of Mask Detection on PyTorch

Input: image array (w,h,c)

```
dets=predict_face(image)
eye_angle=math.atan(dets[5][1]/dets[7][0])*180/math.pi
crop_image=ndimage.rotate(image,eye_angle)[dets[3]:dets[4],dets[1]:dets[2]]
out = predict_masks(img)
```

Output Class-wise affinity score

It is not common enough for training with random initialization of weights [13] to take place in practice, since a data set of sufficiently large and necessary size is rarely available. To speed up learning, the Transfer Learning technology [14, 15] is used: ready – made weights of the trained model are taken (usually ImageNet [16], containing 1.2 million images with 1000 categories), and then the weights of this trained model are used either as initialization or as a way to extract fixed features for the task of interest.

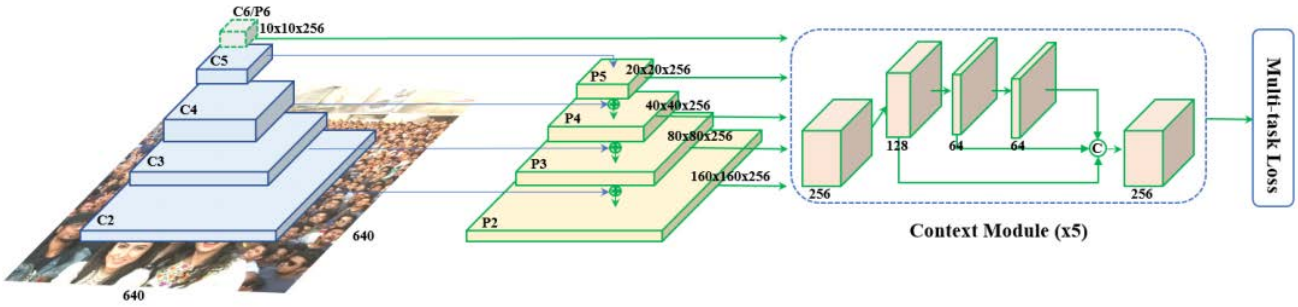

Fig. 4. Fully Connected Layer of the Medical Mask Recognition Neural Network Classifier.

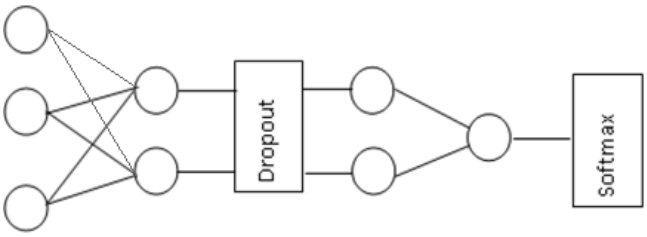

Fig. 5. Fully Connected Layer of the Medical Mask Recognition Neural Network Classifier.

IV. Result

The training was conducted through transfer learning. Two approaches were prepared: in the first approach, the weights of the ResNet neural network were initialized with weights trained on ImageNet, and the second approach: the weights of the ResNet neural network were initialized with weights trained to recognize faces on the ArcFace metric [17]. Comparative transfer learning graphs for different pre-trained weights are shown in Fig. 6.

The graphs show that the model initialized with ArcFace weights [8] converges faster than the model initialized with ImageNet weights.

The validation part of the data sets was used to measure the accuracy of facial recognition in medical masks. Table I shows the recognition accuracy. Accuracy and completeness parameters [18] are used for quality assessment at the selected threshold 0.5.

$$accurancy = \frac{TP}{TP + FP} \quad recall = \frac{TP}{TP + FN}$$

TP – true positive examples;

FP – false positive examples;

FN – false negative examples.

Fig. 7 shows examples of how medical mask recognition works.

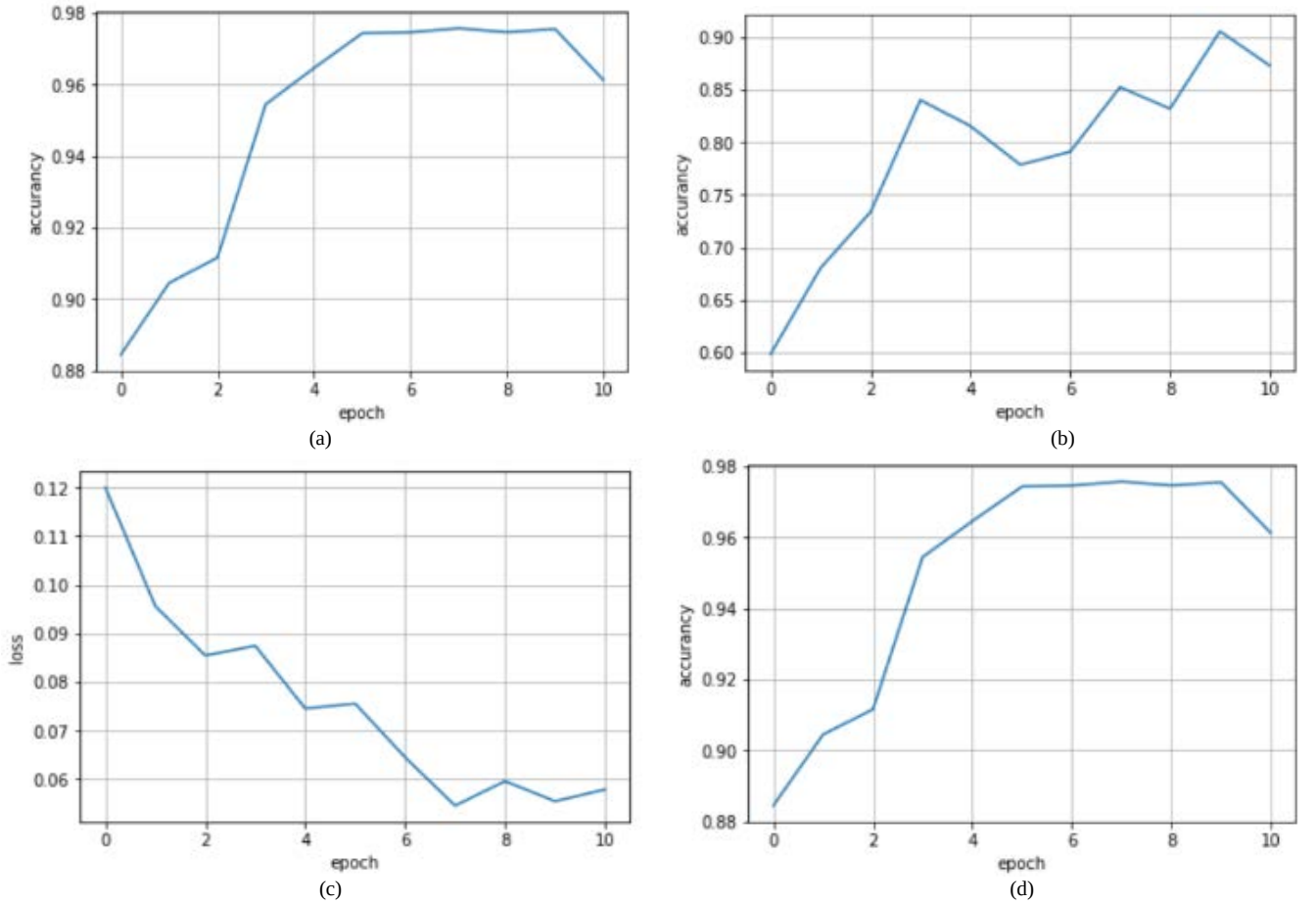

Fig. 6. Graphs of Changes during Validation: a) Errors when Initializing the Model with ImageNet Weights b) Accuracy when Initializing the Model with ImageNet Weights C) Errors when Initializing the Model with ArcFace Weights d) Accuracy when Initializing the Model with ArcFace Weights.

TABLE I. Quality Parameters of Recognition Medical Masks Model

	Accuracy of learning sample	Accuracy of test sample	Completeness of test sample
Recognition medical mask	99.8542	91.5621	92.4562

To demonstrate the performance and accuracy of the classification algorithm, there is an integral characteristic of the ROC curve [19] – another visualization method. The ROC-AUC curve shows what values accuracy and completeness take for different thresholds for making the «there is a mask/no mask» decision. The area under the curve [20] also characterizes the quality of the classification algorithm – the larger it is, the better. High accuracy determines a lower level of false positives.

Fig. 7 shows the ROC curve for the proposed algorithm. The area under the ROC curve is 0.98, which is a pretty good indicator.

Fig. 8 shows the examples of how medical mask recognition works. The left part of the drawing shows the original image with faces selected in the bounding box – the result of the RetinaFace neural network, the right part of the drawing shows the result of recognizing the presence/absence of a medical mask on the found face with some confidence.

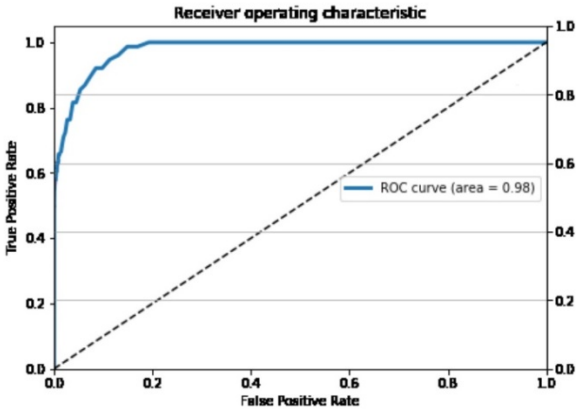

Fig. 7. ROC Curve of the Proposed Classifier.

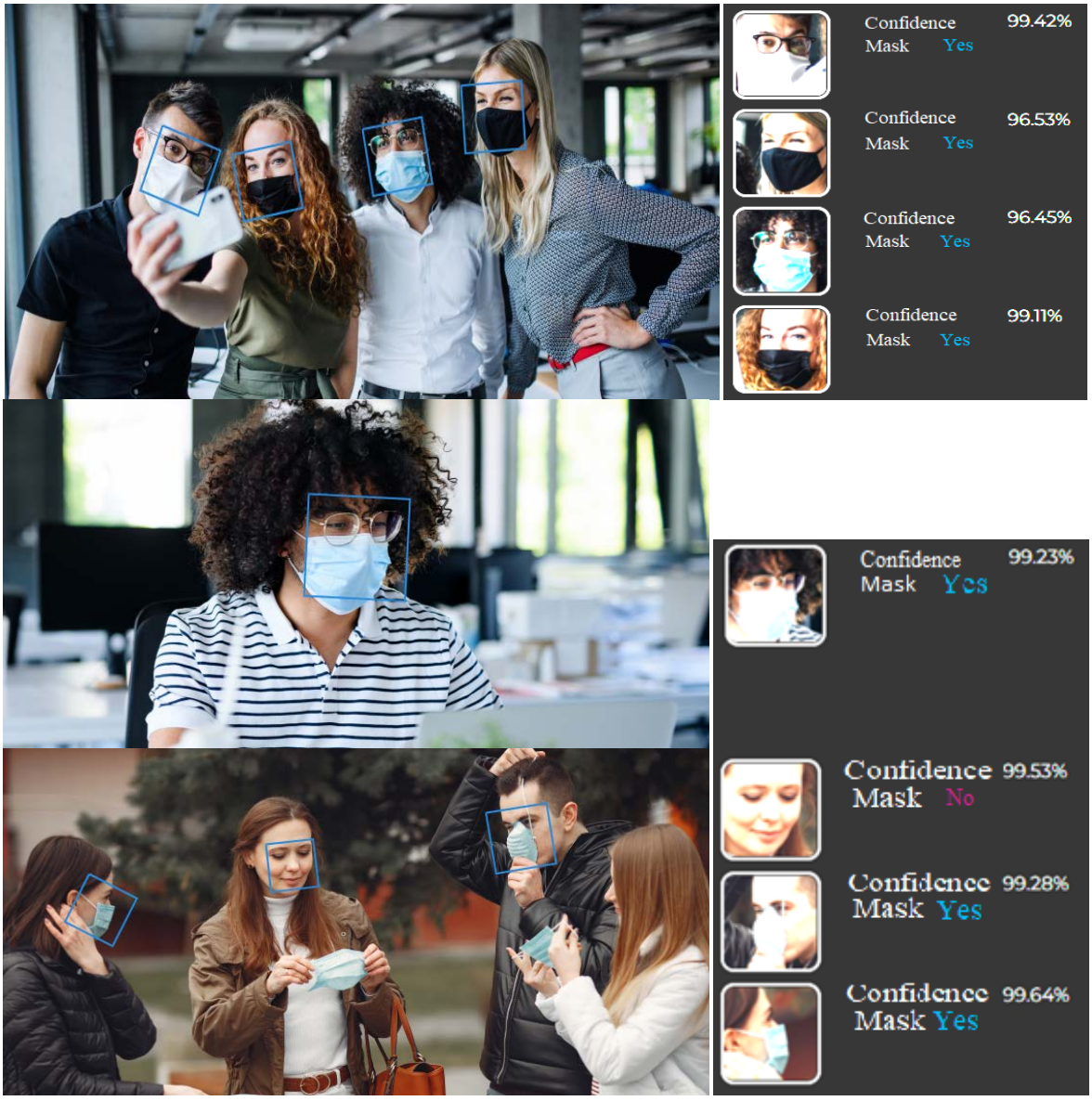

Fig. 8. Examples of Medical Masks Detector Functioning.

V. Comparison with other Methods

As can be seen from the analysis of existing developments in the field of medical mask recognition, most approaches are based on neural network methods [1, 2, 3]. The main difference between the proposed method and the existing ones is the use of pre-trained weights on faces to initialize the neural network before training, which significantly increases both the learning speed (up to 10 epochs) and the quality of the classifier. Table II shows the quality indicators of recognition of medical masks in the test sample.

For works [1,2], the accuracy is taken as the average for all test scenarios.

TABLE II. Verification results (%) of different Mask Detection Algorithms

	Accuracy of learning sample	Accuracy of test sample	Completeness of test sample
MTCNN + FaceNet + SVM	99.7862	82.4862	–
PCA	–	83.11	–
3 Cascade-CNN		86.6	87.8
RetinaFace + Resnet (TL)	99.8542	91.5621	92.4562

VI. Conclusion

This article presents a solution based on the use of two convolutional neural networks to predict the presence of a medical mask on a person's face. As can be seen from the results, the algorithm is able to accurately determine the presence of a medical mask, but there are some false positives, for example, when you try to cover your face with your hands, imitating a medical mask. The speed of the full processing cycle from obtaining a raw image to making a decision about the presence of a medical mask is less than 50 ms. The advantage of using this structure is its flexibility in terms of replacing classification algorithms with more efficient ones in the future. Also, in the course of the work, a unique database of faces in medical masks was collected, which is necessary for training a more effective classifier. The general disadvantage of the proposed algorithm is its computational complexity, and this factor cannot be ignored when building information systems that operate in real time. However, due to the rapid development of computer technology at the present time, these shortcomings can be overcome in the near future. In the future, it is planned to improve the accuracy of recognition of medical masks in more complex scenarios without loss in processing speed. Using graphics accelerators for neural networks is especially promising for solving such problems.

Acknowledgment

Publication is made as part of national assignment forSRISA RAS (fundamental scientific research 47 GP) on the topic No.0065-2019-0001 (AAAA-A19-119011790077-1).

References

[1] Ejaz, M. S., & Islam, M. R. (2019). Masked Face Recognition Using Convolutional Neural Network. 2019 International Conference on Sustainable Technologies for Industry 4.0 (STI). doi:10.1109/sti47673.2019.9068044.

[2] Ejaz, M. S., Islam, M. R., Sifatullah, M., &Sarker, A. (2019). Implementation of Principal Component Analysis on Masked and Non-masked Face Recognition. 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT).doi:10.1109/icasert.2019.8934543.

[3] Bu, W., Xiao, J., Zhou, C., Yang, M., & Peng, C. (2017). A cascade framework for masked face detection. 2017 IEEE International Conference on Cybernetics and Intelligent Systems (CIS) and IEEE Conference on Robotics, Automation and Mechatronics (RAM). doi:10.1109/iccis.2017.8274819.

[4] Chandrashekhar, P. Face Detection Techniques-A Review/ P. Chandrashekhar, N.D. Gopal // International Journal of Current Engineering and Technology, 2013. – P. 1809-1813.

[5] Erik, H. Face detection: A Survey/ H. Erik, K. Boon // Computer vision and image understanding, 2001. – Vol. 83, Issue 3. – P. 236-274.

[6] Cha, Z. A Survey of Recent Advances in Face Detection/ Z. Cha, Z. Zhengyou // Technical Report MSR-TR-2010-66, Microsoft research, Microsoft corporation, one Microsoft way Redmond, Multimedia, Interaction, and Communication (MIC) Group, 2010. – P. 1-17.

[7] Deng, Jiankang, J. Guo, Y. Zhou, Jinke Yu, I. Kotsia and S. Zafeiriou. "RetinaFace: Single-stage Dense Face Localisation in the Wild." *ArXiv* abs/1905.00641 (2019): n. pag.

[8] J. Deng, J. Guo, N. Xue and S. Zafeiriou, "ArcFace: Additive Angular Margin Loss for Deep Face Recognition," *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, CA, USA, 2019, pp. 4685-4694, doi: 10.1109/CVPR.2019.00482.

[9] K. D. Rusakov, "Automatic Modular License Plate Recognition System Using Fast Convolutional Neural Networks," 2020 13th International Conference "Management of large-scale system development" (MLSD), Moscow, Russia, 2020, pp. 1-4, doi: 10.1109/MLSD49919.2020. 9247817.

[10] Rusakov K.D., Genov A.A., Shil S.Sh. An anti-spoofing methodology for a limited number of photos. Software & Systems. 2020, vol. 33, no. 1, pp. 054–060 (in Russ.). doi: 10.15827/0236- 235X.129.054-060.

[11] S. Albawi, T. A. Mohammed and S. Al-Zawi, "Understanding of a convolutional neural network," 2017 International Conference on Engineering and Technology (ICET), Antalya, 2017, pp. 1-6, doi: 10.1109/ICEngTechnol.2017.8308186.

[12] S. Panigrahi, A. Nanda and T. Swarnkar, "Deep Learning Approach for Image Classification," 2018 2nd International Conference on Data Science and Business Analytics (ICDSBA), Changsha, 2018, pp. 511-516, doi: 10.1109/ICDSBA.2018.00101.

[13] L. M. Waghmare, N. N. Bidwai and P. P. Bhogle, "Neural Network Weight Initialization," 2007 International Conference on Mechatronics and Automation, Harbin, 2007, pp. 679-681, doi: 10.1109/ICMA.2007.4303625.

[14] P. Natrajan, S. Rajmohan, S. Sundaram, S. Natarajan and R. Hebbar, "A Transfer Learning based CNN approach for Classification of Horticulture plantations using Hyperspectral Images," 2018 IEEE 8th International Advance Computing Conference (IACC), Greater Noida, India, 2018, pp. 279-283, doi: 10.1109/IADCC.2018.8692142.

[15] M. Shaha and M. Pawar, "Transfer Learning for Image Classification," 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, 2018, pp. 656-660, doi: 10.1109/ICECA.2018.8474802.

[16] J. Deng, W. Dong, R. Socher, L. Li, Kai Li and Li Fei-Fei, "ImageNet: A large-scale hierarchical image database," 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, 2009, pp. 248-255, doi: 10.1109/CVPR.2009.5206848.

[17] J. Deng, J. Guo, N. Xue and S. Zafeiriou, "ArcFace: Additive Angular Margin Loss for Deep Face Recognition," *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, CA, USA, 2019, pp. 4685-4694, doi: 10.1109/CVPR.2019.00482.

[18] N. Seliya, T. M. Khoshgoftaar and J. Van Hulse, "Aggregating performance metrics for classifier evaluation," 2009 IEEE International Conference on Information Reuse & Integration, Las Vegas, NV, 2009, pp. 35-40, doi: 10.1109/IRI.2009.5211611.

[19] Jin Huang and C. X. Ling, "Using AUC and accuracy in evaluating learning algorithms," in IEEE Transactions on Knowledge and Data Engineering, vol. 17, no. 3, pp. 299-310, March 2005, doi: 10.1109/TKDE.2005.50.

[20] M. H. Ferris et al., "Using ROC curves and AUC to evaluate performance of no-reference image fusion metrics," 2015 National Aerospace and Electronics Conference (NAECON), Dayton, OH, 2015, pp. 27-34, doi: 10.1109/NAECON.2015.7443034.