

# A NOVEL METHOD FOR PROTECTIVE FACE MASK DETECTION USING CONVOLUTIONAL NEURAL NETWORKS AND IMAGE HISTOGRAMS

E. Ryumina, D. Ryumin, D. Ivanko, A. Karpov

St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences, St. Petersburg  
Federal Research Center of the Russian Academy of Sciences (SPC RAS), St. Petersburg, Russian Federation –  
ryumina\_ev@mail.ru, dl\_03.03.1991@mail.ru, denis.ivanko11@gmail.com, karpov@iias.spb.su

Commission II, WG II/5

**KEY WORDS:** Computer Vision, Object Detection, Mask Detection, Face in a Medical Mask, CNN, Image Histograms.

## ABSTRACT:

This paper proposes a new hybrid method for automatic detection and recognition of the presence/absence of a protective mask on human's face. It combines visual features extracted using Convolutional Neural Network (CNN) with image histograms that convey information about pixel intensity. Several pre-trained models for building feature extraction systems using a CNN and several types of image histograms are considered in this paper. We test our approach on the Medical Mask Dataset and perform cross-corpus analysis on two other databases named Masked Faces (MAFA) and Real-World Masked Face Dataset (RMFD). We demonstrate that the proposed hybrid method increases the Unweighted Average Recalls (UARs) of recognition of the presence/absence of a protective mask on human's face in comparison with traditional CNNs on the MAFA and RMFD databases by 0.96% and 1.32%, respectively. The proposed method can be generalized and used for other tasks of biometry, computer vision, machine learning and automatic face recognition.

## 1. INTRODUCTION

The global spread of the COVID-19 coronavirus pandemic has led to numerous consequences in our everyday life, including the emergency of a social distance and a "mask" culture. Monitoring and evaluation of the security level of individuals and the entire society is one of the greatest challenges of the modern world. Initially, the mask culture has emerged and proved its usefulness in densely populated Asian countries (such as China, Japan and Korea), where people used to wear masks in public places to protect themselves from air pollution, allergens and respiratory diseases long before the current pandemic. To date, numerous scientific studies have been carried out on the topic of the usefulness of the "mask" culture (Cheng et al., 2020, Wang et al., 2020). Nowadays, all countries around the world use this positive experience of Asian countries to tackle the serious new challenge - protecting the health of entire human population from COVID-19. Thus, as of today, medical face masks are a part of every person's dress code.

In order to fight COVID-19 epidemic effectively we need new information technologies that are able to break the spread of infection by minimizing the threats of outbreaks. Such technologies include digital methods for automating preventive measures against COVID-19 by intelligently tracking the presence or absence of protective masks on human's faces. Currently, leading foreign scientific institutes and global industrial corporations have actively started research and development of intelligent technologies in the field of artificial intelligence (AI) Loey et al., 2021; Jiang and Fan, 2020). These technologies are based on AI techniques such as traditional and deep machine learning to create effective solutions to prevent the spread of the COVID-19 coronavirus infection.

This paper presents a method for automatic detection and recognition of the presence or absence of a protective mask on human's face, which combines visual features extracted using

Convolutional Neural Network (CNN) with image histograms that convey information about pixel intensity.

## 2. RELATED WORK

### 2.1 Face detectors

Recognition of the face of a masked person is impossible without detecting a region-of-interest (ROI), i.e. face region in our case. Face detection allows recognizing and detecting bounding boxes in images. For a long time, state-of-the-art face detection algorithms were considered the Viola-Jones cascade classifier (Viola et al., 2004) and Histogram of Oriented Gradients (HOG (Deniz et al., 2011)). However, the rapid development of neural networks has influenced the development of various face detectors based on CNN architectures. All the detectors based on neural networks can be one-staged or two-staged. The main idea of one-stage methods is that the search of ROI and classification of objects is carried out in one pass. One-stage methods include: Single Shot MultiBox Detector (SSD (Liu et al., 2016)), Max-Margin Object Detection (MOD (King, 2015)), as well as the recently proposed RetinaFace (Deng et al., 2020) and Multi-task Cascaded Convolutional Networks (MTCNN (Zhang et al., 2019)), and others. In two-stage methods, the purpose of the first stage is to form the assumptions about the possible location of objects, and the goal of the second stage (usually done by different neural network) is to make the final decision about object recognition and their final meta-location. Two-stage methods are R-CNN (Girshick et al., 2014), Faster-RCNN (Ren et al., 2016), and others. One-stage methods surpass two-stage counterparts both in speed and, in most cases, in accuracy. However, most detectors were trained on images of partially occluded faces, while a mask overlaps the face by more than 50%. Therefore, it is necessary to efficiently evaluate the performance of face detectors in severe occlusion conditions.

## 2.2 Masked face databases

At the moment there are few databases containing images of masked faces. Conventionally, such databases can be divided into those collected in uncontrolled or laboratory conditions and artificially generated (artificial imposition of masks on human's faces). Within the framework of this research, only databases collected in uncontrolled conditions are analyzed. The publicly available Medical Mask Dataset (MMD (Humans in the Loop, 2020)) contains 6,024 images depicting people of various nationalities, ages, and regions. Each human face was annotated for 20 various accessory classes (glasses, mask, hood, etc.), including faces with a protective facial mask, without a mask, and with an incorrectly worn protective mask. Masked Faces (MAFA (Ge et al., 2017)) contains 30,811 images, 35,806 detected faces belong to the "masked face" class. MAFA is the largest masks database at the moment. The images were collected on the Internet and were annotated by age, race, degree and type of occlusion, face position, and three classes of "masked face", "unmasked face" и "invalid face". MAFA was divided into training and testing sets. The Real-World Masked Face Dataset (RMFD (Wang et al., 2020)) contains 5,000 000 images of faces with a mask for 525 people and 90,000 images of the same people without a mask. However, only 2,203 images of faces with a mask and 90,468 images of faces without a mask are available for public access.

## 2.3 State-of-the-art Face Mask Recognition approach

The authors (Ejaz et al., 2019) propose a mask detection approach using a traditional machine learning method combined with principal component analysis. Experimental results show that this approach is not effective in recognizing the "masked face" class. The combination of traditional and transfer learning approach presents in the work (Loey et al., 2021a). This hybrid method is effective in recognizing both "masked face" and "unmasked face" classes. Then, in (Loey et al., 2021b) use a YOLOv2 detector with ResNet-50 to extract and detect features during training, validation, and testing phases. In (Chowdary et al. (2020)), the use of transfer learning by fine-tuning a pre-trained InceptionV3 model were proposed. Also, data augmentation (rotation, horizontal flip and so on) is used to increase the robustness of the system. When analyzing the state-of-the-art face mask recognition approaches, it can be seen that all the proposed approaches were tested only on a single corpus (when the same database is used for training, testing and validation, pre-divided into appropriate sets), in other words, there are no information about how the proposed approaches work on new data. In addition, it can be noted that CNNs are mainly used for feature extraction, while there are hand-crafted features that can improve the efficiency of systems, for example, images histograms, Gabor filters and other features that are actively used in other recognition tasks such as objects (Zheng et al., 2015), gestures (Ryumin et al., 2019, Ivanko et al., 2018), emotions detection (Ryumina and Karpov, 2020a), etc.

## 3. PROPOSED METHOD

In this paper, we propose a new method to automatically detect and recognize the presence/absence of protective mask on human's faces. The presented method is based on combining two feature sets: 1) features extracted using CNN; 2) characteristics of the pixel intensity distribution in the images (image histograms). Figure 1 shows a functional diagram of the proposed method.

### 3.1 CNN-based feature extraction

Due to the lack of a large training dataset, transfer learning is usually applied. Transfer learning is the transfer of knowledge from a model trained on one single task to a new one. Keras open source machine learning library (Keras, 2021) proposes several models trained on ImageNet (Deng et al., 2009) to solve the problem of recognition of 1000 various objects. Also, in the public domain (Malli, 2021) several models trained on VGGFace2 (Cao et al., 2018) for 8631 faces exists to tackle recognition problem. In the current work, several pre-trained models are considered both on ImageNet (Deng et al., 2009) and VGGFace2 (Cao et al., 2018). While training the models, data generation techniques such as random rotation, horizontal flip, scaling, and moving were used.

### 3.2 Features based on Image histograms

To increase the efficiency of the models, the features obtained from the last layer of the neural network were combined with normalized (in the range from 0 to 1) image histograms, i.e., the features of different origin are combined on the average level feature fusion. For more information on the construction of image histograms, see (Zheng et al., 2015).

### 3.3 Cross-corpus analysis

As mentioned earlier, to confirm the effectiveness of the masked face recognition system, it is necessary to provide results not only on a single corpus, but also to perform cross-corpus analysis. So the MMD database was selected for training and validating the systems, and the other two MAFA and RMFD databases were selected for testing.

## 4. DATA PRE-PROCESSING

### 4.1 Medical Mask Dataset - MMD

For the purposes of current research, images belonging to the "masked face" and "unmasked face" classes were selected from the database. When analyzing the inter-class distribution, an imbalance of classes was revealed, most of the images belong to the "masked face" class (6769 vs. 2086). This problem has a negative impact on machine learning algorithms (Ryumina and Karpov, 2020b). To solve it, we searched for face images "unmasked face" using the Google search engine and the Internet until the classes were well balanced. Since the MMD

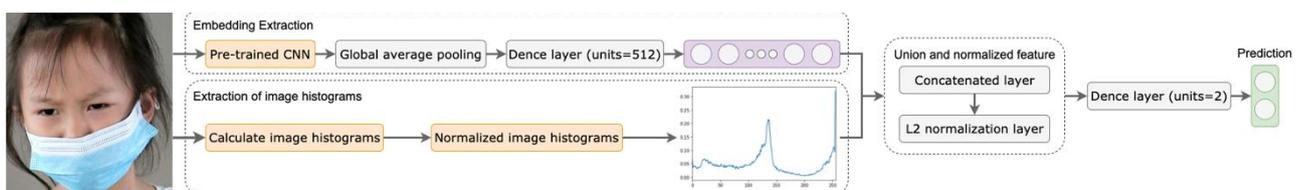


Figure 1. A diagram of the proposed method for facial mask detection.

already provides face area coordinates (bounding boxes), face detection for new samples was performed using the SSD. The SSD is implemented in the Deep Neural Networks (DNN) module of Open-Source Computer Vision Library (OpenCV, 2021, Pulli et al., 2012). Then the detected areas were manually checked on correctness of areas with faces. In this way, the Medical Mask Extended Dataset (MMED) was collected and annotated.

#### 4.2 Masked Faces - MAFA

Despite the fact that to date the MAFA database is the largest among counterparts, it has a serious drawback: presented labels are incorrect since any overlap of the face is considered as the "masked face" class, it is also worth noting that there are no clear labels, for example, "protective mask", "microphone" or "clown nose", which means that for more correct use of it, repeated manual annotation is necessary. The annotation of the test set was performed according to the following rules:

- "Unmasked face", i.e. mouth and nose are opened (0 class);
- "Masked face", i.e. mouth and nose are closed with a "protective mask" (1 class);
- "Incorrectly masked face", i.e. mouth or nose is closed with a "protective mask" (2 class);
- "Overlap with another object", i.e. it can be a phone, a fruit, a book, a hand, etc. (3 class).

The "protective mask" class includes various medical masks, respirators, colorful masks, scarves, etc. Thus, according to the original annotations, the MAFA test set had 4934 samples for the "masked face" class and 1 sample for the "unmasked face" class.

After re-annotating according to the rules set above, the following distribution of samples by class were obtained: 0 class has 447 samples, 1 class - 3707, 2 class - 128, 3 class - 653.

#### 4.3 General data pre-processing

When fed to the input of CNN, image pixels values must be normalized in accordance with normalization that was performed for a particular pre-trained neural network - this can be linear normalization, channel normalization, etc.

### 5. EXPERIMENTAL RESULTS

#### 5.1 CNN based visual features

Within the framework of this research, the following models of CNNs were considered: 1) ResNet-50 (He et al., 2016), SeNet-50 (Hu et al., 2017) and VGG16 (Simonyan and Zisserman, 2014) pre-trained on the VGGFace2 database (Cao et al., 2018); 2) MobileNetV2 (Sandler et al., 2018), ResNet101V2 (He et al., 2016), EfficientNetB0 (Tan and Le, 2019) pre-trained on ImageNet database (Deng et al., 2009). The input image size for neural networks was set to  $224 \times 224$  pixels. The classification layer in the pre-trained neural networks was frozen, then a global average pooling layer and two fully connected layers with 512 (+ 50% dropout) and 2 units were added. The training was performed for 20 epochs using the Adam optimizer. The initial learning rate was set to 0.00001 and was constantly decreasing according to the formula (1):

$$lr = lr_{start} \times \frac{1}{1 + \frac{lr_{start} \times epoch_{curr} \times N}{epoch \times bs}} \quad (1)$$

where  $lr_{start}$  = initial learning rate  
 $epoch$  = total number of training epochs  
 $epoch_{curr}$  = number of the current epoch at the time of the learning rate update  
 $N$  = total number of samples in the training set  
 $bs$  = batch size

For experimental evaluation, the MMED database was divided into training and validation set in a ratio of 80:20, observing a full balance of classes. Accuracy (A) and recall (R) metrics were used to score the efficiency. Table 1 presents efficiency scores of the models.

| Model          | Pre-training database | A, %  | R1, % | R2, % |
|----------------|-----------------------|-------|-------|-------|
| ResNet-50      | VGGFace2              | 98.74 | 98.59 | 98.89 |
| SeNet-50       | VGGFace2              | 98.11 | 98.48 | 97.74 |
| VGG16          | VGGFace2              | 98.00 | 97.37 | 98.63 |
| MobileNetV2    | ImageNet              | 98.08 | 98.11 | 98.04 |
| ResNet101V2    | ImageNet              | 98.15 | 98.19 | 98.11 |
| EfficientNetB0 | ImageNet              | 95.08 | 96.49 | 93.67 |

**Table 1.** Comparative table of model's efficiency: R1 - Recall for the "masked face", R2 - Recall for the "unmasked face".

As can be seen from Table 1, the best accuracy values were achieved using models pre-trained on the VGGFace2 database (Cao et al., 2018). In particular, the ResNet-50 model is the most efficient. The efficiency scores of this model will be used as a baseline for comparison with the efficiency scores obtained using hybrid methods.

#### 5.2 Image histograms

Since the presence of a mask on the face leads to a significant overlap, using image histograms allow adding additional features about the presence of a mask on the face. In this paper, we consider the effectiveness of applying image histograms of different types: (1) for each channel individually (the histogram channel Red/Grey/Blue (HCR/G/B), vectors with dimension 256 components); (2) averaged histograms over three channels (HA); (3) interval representation of  $8 \times 8 \times 8$  bins over three channels (H512, vectors with dimension 512 components). All the histograms were obtained using (OpenCV, 2021). Then, the histograms were fed to the input of a fully connected neural network (two fully connected layers with 256 and 2 units). Table 2 shows efficiency scores for the histograms under consideration.

| Histogram type | A, %         | R1, %        | R2, %        |
|----------------|--------------|--------------|--------------|
| HCR            | 67.52        | 62.35        | 72.71        |
| HCG            | 70.49        | 69.38        | 71.60        |
| HCB            | 75.11        | 74.78        | 75.44        |
| HA             | -            | -            | 64.94        |
| H512           | <b>86.24</b> | <b>87.94</b> | <b>87.09</b> |

**Table 2.** Comparative table of histogram's efficiency.

Thus, Table 2 shows that using this approach, there is a more accurate recognition of the "masked face" class than the "unmasked face" class, i.e. the pixel intensity distribution in the images contains additional information that the mask is present on the face. It can also be noted that the best accuracy values are achieved with the help of the HCB and H512, so only these types of the histograms will be considered later.

### 5.3 Combining CNN and image histograms based features

The combination of features was made according to Figure 1. Then the final recognition of the presence/absence of masks on human's faces was performed. Table 3 shows the final detection results on human's faces for the MMED database using hybrid methods.

| Features             | Model         | A, %         | R1, %        | R2, %        |
|----------------------|---------------|--------------|--------------|--------------|
| ResNet-50 (Baseline) |               | 98.74        | 98.59        | 98.89        |
| RI+HCB               | ResNet-50+FCN | 98.93        | 98.67        | <b>99.19</b> |
| Growth               |               | 0.19         | 0.08         | 0.30         |
| RI+H512              | ResNet-50+FCN | <b>99.15</b> | <b>99.19</b> | 99.11        |
| Growth               |               | 0.41         | 0.60         | 0.22         |

**Table 3.** Comparative table of hybrid method's efficiency: RI - raw image.

As you can see from Table 3, by adding vectors that contain image histograms to the last layer of CNN, it is possible to increase the efficiency scores for each class individually, and for the entire recognition process.

### 5.4 Cross-corpus analysis

Cross-corpus analysis allows concluding whether the recognition process is robust to new data and is suitable for use in real uncontrolled conditions. Table 4 shows the efficiency scores on the RMFD and MAFA databases for different research approaches. The Unweighted Average Recall (UAR) metric was used for the estimation, because the classes in the databases under research are not balanced.

| Features                       | Model         | UAR, %       | R1, %        | R2, %        |
|--------------------------------|---------------|--------------|--------------|--------------|
| RMFD                           |               |              |              |              |
| ResNet-50                      |               | 97.16        | 99.82        | 94.51        |
| RI+HCB                         | ResNet-50+FCN | <b>98.12</b> | <b>99.86</b> | <b>96.37</b> |
| Growth                         |               | 0.96         | 0.04         | 1.86         |
| RI+H512                        | ResNet-50+FCN | 97.41        | <b>99.86</b> | 94.96        |
| Growth                         |               | 0.25         | 0.04         | 0.45         |
| MAFA (without classes 2 and 3) |               |              |              |              |
| ResNet-50                      |               | 96.36        | 93.29        | 99.43        |
| RI+HCB                         | ResNet-50+FCN | 94.99        | 90.38        | <b>99.60</b> |
| Growth                         |               | -1.37        | -2.91        | 0.17         |
| RI+H512                        | ResNet-50+FCN | <b>97.68</b> | <b>95.97</b> | 99.38        |
| Growth                         |               | 1.32         | 2.68         | -0.05        |

**Table 4.** Efficiency scores for cross-corpus analysis.

Thus, Table 4 shows that the application of the hybrid method RI + HCB on the RMFD database allows achieving a significant increase in the R2 value compared to the baseline by 1.86%. The overall increase in UAR was 0.96%. When using RI + H512 hybrid method on the MAFA database, the R2 value decreases by 0.05%, but the overall UAR increase was 1.32%. The cross-corpus analysis shows that systems with hybrid methods are effective on new data.

As can be seen from Table 4, the classes 2 and 3 are excluded from the MAFA database for the efficiency scores, since these classes are not presented in the training database and, therefore, are considered separately. Thus, Figure 2 presents the predictions for classes 2 and 3 of the MAFA database obtained using the RI + H512 hybrid method and also provides information on how many samples from class 2 or 3 were classified as class 0 or 1.



**Figure 2.** Predictions for 2 and 3 classes.

Figure 2 shows that the process of face recognition in a mask occurs with errors in complex cases and with high confidence.

### 5.5 Face detection and Intersection over Union

The final step is to determine the optimal frames per second (FPS) and the Intersection over Union (IoU) of the masked face detector. The IoU calculation was performed considering the 50% overlap of the detected area to the annotated one according to the formula (2):

$$IoU = \frac{TP}{TP+FP+FN} \quad (2)$$

where  $TP$  = True Positive  
 $FP$  = False Positive  
 $FN$  = False Negative

Table 5 shows the efficiency scores of the face detectors. The experiments were carried out on an NVIDIA GeForce GTX 3080 GPU on a test sample of the MAFA database with reduced images, so that the maximum side of the images does not exceed 300 pixels.

| Face detector                  | FPS | IoU, %       |
|--------------------------------|-----|--------------|
| SSD (Liu et al., 2016)         | ≈81 | 65.44        |
| MTCNN (Zhang et al., 2019)     | ≈38 | 72.15        |
| RetinaFace (Deng et al., 2020) | ≈76 | <b>85.44</b> |

**Table 5.** Efficiency scores of face detectors.

The experimental results demonstrated that the RetinaFace face detector is significantly superior to other detectors in the terms of IoU metric, but this detector is slightly inferior to the SSD in metric of FPS.

## 6. CONCLUSION

In this paper we have presented a new method for automatic detection of protective masks on human's faces. The proposed method is based on combining visual features extracted using CNN with image histograms. We demonstrate that the proposed hybrid method allows increasing UARs of recognition of the presence/absence of a protective mask on human's face on the MMED, MAFA and RMFD databases by 0.41%, 0.96% and 1.32%, respectively.

In the future research, we plan to use the proposed method of parametric representation in automatic system for detecting the presence/absence of a mask on human's faces. This method is able to work in real time and suitable for many applications. Along with this, due to its versatility, the proposed method can be used for many different tasks of biometry, computer vision, machine learning and automatic face recognition.

## ACKNOWLEDGEMENTS

This research is financially supported by the RFBR, project No. 20-04-60529.

## REFERENCES

- Cao, Q., Shen, L., Xie, W., Parkhi, O.M., Zisserman, A., 2018: Vggface2: A dataset for recognising faces across pose and age. *IEEE International Conference on Automatic Face and Gesture Recognition, FG 2018*, 67-74. doi.org/10.1109/FG.2018.00020.
- Cheng, V. C. C., Wong, S. C., Chuang, V. W. M., So, S. Y. C., Chen, J. H. K., Sridhar, S., Yuen, K. Y., 2020. The role of community-wide wearing of face mask for control of coronavirus disease 2019 (COVID-19) epidemic due to SARS-CoV-2. *Journal of Infection*, 81(1), 107-114. doi.org/10.1016/j.jinf.2020.04.024.
- Chowdary, G. J., Punn, N. S., Sonbhadra, S. K., Agarwal, S., 2020: Face mask detection using transfer learning of inceptionv3. *International Conference on Big Data Analytics*, Springer, Cham, 12581, 81-90. doi.org/10.1007/978-3-030-66665-1\_6.
- D'eniz, O., Bueno, G., Salido, J., De la Torre, F., 2011. Face recognition using histograms of oriented gradients. *Pattern Recognition Letters*. 32 (12), 1598-1603. doi.org/10.1016/j.patrec.2011.01.004.
- Deng, J., Guo, J., Ververas, E., Kotsia, I., Zafeiriou, S., 2020: RetinaFace: Single-Shot Multi-Level Face Localisation in the Wild. *IEEE Conference on Computer Vision and Pattern Recognition, CVPR-2020*, Seattle, WA, 5203-5212. doi.org/10.1109/CVPR42600.2020.00525.
- Deng, J., Dong, W., Socher, R., Li, L.J., Li, K. Fei-Fei, L., 2009: ImageNet: A large-scale hierarchical image database. *IEEE Conference on Computer Vision and Pattern Recognition, CVPR-2009*, Miami, FL, 248-255. doi.org/10.1109/CVPR.2009.5206848.
- Ejaz, M. S., Islam, M. R., Sifatullah, M., Sarker, A., 2019: Implementation of principal component analysis on masked and non-masked face recognition. *1st International Conference on Advances in Science, Engineering and Robotics Technology, ICASERT 2019*, 1-5. doi.org/10.1109/ICASERT.2019.8934543.
- Ge, S., Li, J., Ye, Q., Luo, Z., 2017: Detecting masked faces in the wild with lle-cnns. *IEEE Conference on Computer Vision and Pattern Recognition, CVPR-2017*, Honolulu, HI, 2682-2690. doi.org/10.1109/CVPR.2017.53.
- Girshick, R., Donahue, J., Darrell, T. Malik, J., 2014: Rich feature hierarchies for accurate object detection and semantic segmentation. *IEEE Conference on Computer Vision and Pattern Recognition, CVPR-2014*, Columbus, OH, 580-587. doi.org/10.1109/CVPR.2014.81.
- He, K., Zhang, X., Ren, S., Sun, J., 2016: Deep residual learning for image recognition. *IEEE Conference on Computer Vision and Pattern Recognition, CVPR-2016*, Las Vegas, NV, USA, 770-778. doi.org/10.1109/CVPR.2016.90
- Hu, J., Shen, L., Sun, G., 2017. Squeeze-and-Excitation networks. arXiv preprint arXiv:1709.01507.
- Humans in the Loop, 2021. Medical mask dataset, <https://humansintheloop.org/medical-mask-dataset> (25 February 2021).
- Ivanko, D., Ryumin, D., Axyonov, A., Železný, M., 2018: Designing Advanced Geometric Features for Automatic Russian Visual Speech Recognition. *Speech and Computer, SPECOM 2018*, Springer, Cham, 11096, 245-254. doi.org/10.1007/978-3-319-99579-3\_26.
- Jiang, M., Fan, X., 2020. RetinaMask: a face mask detector. arXiv preprint arXiv:2005.03950.
- Keras, 2021. Keras Documentation, <https://keras.io> (25 February 2021).
- King, D.E., 2015. Max-margin object detection. arXiv preprint arXiv:1502.00046.
- Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C-Y., Berg A., 2016: SSD: Single Shot MultiBox Detector. *Computer Vision – ECCV 2016*, Springer, Cham, 9905, 21-37. doi.org/10.1007/978-3-319-46448-0\_2.
- Loey, M., Manogaran, G., Taha, M. H. N., Khalifa, N. E. M., 2021a. A hybrid deep transfer learning model with machine learning methods for face mask detection in the era of the COVID-19 pandemic. *Measurement*, 167. doi.org/10.1016/j.measurement.2020.108288.
- Loey, M., Manogaran, G., Taha, M. H. N., Khalifa, N. E. M., 2021b. Fighting against COVID-19: A novel deep learning model based on YOLO-v2 with ResNet-50 for medical face mask detection. *Sustainable Cities and Society*, 65. doi.org/10.1016/j.scs.2020.102600.
- Malli, 2021. Keras-VGGFace, <https://pypi.org/project/keras-vggface/> (25 February 2021).
- OpenCV, 2021. OpenCV library, <https://opencv.org> (25 February 2021).
- Pullil, K., Baksheev, A., Korniyakov, K., Eruhimov, V., 2012. Real-time computer vision with OpenCV. *Communications of the ACM*, 55 (6), 61-69.
- Ren, S., He, K., Girshick, R. and Sun, J., 2016: Faster r-cnn: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39 (6), 1137-1149. doi.org/10.1109/TPAMI.2016.2577031.
- Ryumin, D., Kagirow, I., Ivanko, D., Axyonov, A., Karpov, A.A., 2019: Automatic detection and recognition of 3D manual gestures for human-machine interaction. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, ISPRS Archives, XLII-2/W12*, 179-183. doi.org/10.5194/isprs-archives-XLII-2-W12-179-2019.
- Ryumina, E., Karpov, A., 2020a: Facial expression recognition using distance importance scores between facial landmarks. *Graphicon 2020, CEUR Workshop Proceedings*, 2744, 1-10. ceur-ws.org/Vol-2744/paper32.pdf
- Ryumina, E.V., Karpov, A.A., 2020b. Comparative analysis of methods for imbalance elimination of emotion classes in video data of facial expressions. *Scientific and Technical Journal of*

*Information Technologies, Mechanics and Optics*, 20 (5(129)), 683-691. doi.org/10.17586/2226-1494-2020-20-5-683-691.

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L. C., 2018: Mobilenetv2: Inverted residuals and linear bottlenecks. *IEEE Conference on Computer Vision and Pattern Recognition, CVPR-2018*, Salt Lake City, UT, USA, 4510-4520. doi.org/10.1109/CVPR.2018.00474

Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

Tan, M., Le, Q.V., 2019. Efficientnet: Rethinking model scaling for convolutional neural networks. arXiv preprint arXiv:1905.11946.

Viola, P., Jones, M., 2004. Robust real-time face detection. *International Journal of Computer Vision*, 57 (2), 137-154. doi.org/10.1023/B:VISI.0000013087.49260.fb.

Wang, J., Pan, L., Tang, S., Ji, J. S., Shi, X., 2020. Mask use during COVID-19: A risk adjusted strategy. *Environmental Pollution*, 266 (1). doi.org/10.1016/j.envpol.2020.115099.

Wang, Z., Wang, G., Huang, B., Xiong, Z., Hong, Q., Wu, H., Yi, P., Jiang, K., Wang, N., Pei, Y., Chen, H., 2020. Masked face recognition dataset and application. arXiv preprint arXiv:2003.09093.

Zhang, L., Gui, G., Khattak, A. M., Wang, M., Gao, W., Jia, J., 2019. Multi-task cascaded convolutional networks based intelligent fruit detection for designing automated robot. *IEEE Access*, 7, 56028-56038. doi.org/10.1109/ACCESS.2019.2899940

Zheng, Z., Li Z., Nagar, A., Kang, W., 2015: Compact deep convolutional neural networks for image classification. *ICMEW*, Turin, Italy, pp. 1-6. doi.org/10.1109/ICMEW.2015.7169768.