Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Efficient Masked Face Recognition Method during the COVID-19 Pandemic

Walid Hariri (hariri.walid@hotmail.com)
Badji Mokhtar Annaba University https://orcid.org/0000-0002-5909-5433

Research Article

Keywords: Face recognition, COVID-19, Masked face, Deep learning

Posted Date: July 7th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-39289/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Abstract

The COVID-19 is an unparalleled crisis leading to huge number of casualties and security problems. In order to reduce the spread of coronavirus, people often wear masks to protect themselves. This makes the face recognition a very difficult task since certain parts of the face are hidden. A primary focus of researchers during the ongoing coronavirus pandemic is to come up with suggestions to handle this problem through rapid and efficient solutions. In this paper, we propose a reliable method based on discard masked region and deep learning based features in order to address the problem of masked face recognition process. The first step is to discard the masked face region. Next, we apply a pre-trained deep Convolutional neural networks (CNN) to extract the best features from the obtained regions (mostly eyes and forehead regions). Finally, the Bag-of-features paradigm is applied on the feature maps of the last convolutional layer in order to quantize them and to get a slight representation comparing to the fully connected layer of classical CNN. Finally, MLP is applied for the classification process. Experimental results on Real-World-Masked-Face-Dataset show high recognition performance.

1 Introduction

The COVID–19 virus can be spread through contact and contaminated sur- faces, therefore, the classical biometric systems based on passwords or finger- prints are not anymore safe. Face recognition are more safe without any need to touch any device. Recent studies on coronavirus has proven that wearing a face mask by healthy and enfected population reduces considerably the trans- mission of this virus. However, wearing the mask face causes the following problems: i) fraudsters and thieves take advantage of the mask, stealing and committing crimes without being identified. ii) community access control and face authentication are become very difficult tasks when a grand part of the face is hidden by a mask. iii) existing face recognition methods are not ef- ficient when wearing a mask which cannot provide the whole face image for description. iv) exposing the nose region is very important in the task of face recognition since it is used for face normalization [18], pose correction [14], and face matching [9]. Due to these problems, face masks have significantly challenged existing face recognition methods.

To tackle these problems, we distinguish two different tasks namely: *face mask recognition* and *masked face recognition.* The first one checks whether the person is wearing a mask or no. This can be applied in public places where the mask is compulsory. Masked face recognition, in the other hand, aims to recognize a face with mask basing on the eyes and the forehead regions. In this paper we handle the second task using deep learning based method. We use pre-trained deep learning based model in order to extract features from the unmasked face regions (out of the mask region). It is worth stating that the occlusions in our case can occur in only one predictable facial region (nose and mouth regions), this can be a good guide to handle this problem efficiently.

The rest of this paper is organized as follows: Section 2 presents the re- lated works. In Section 3 we present the motivation and contribution of the paper. The proposed method is detailed in Section 4. Experimental results are presented in Section 5. Conclusion ends the paper.

2 Related Works

Occlusion is a key limitation of real world 2D face recognition methods. Gen- erally it comes out from wearing hats, eyeglasses, masks as well as any other objects that can occlude a part of the face while leaving others unaffected. Thus, wearing a mask is considered as the most difficult facial occlusion challenge since it occludes a grand part of the face including the nose. Many approaches have been proposed to handle this problem. We can classify them into three categories namely: local matching approach, restoration approach and discard occlusion based approach.

Matching approach: Aims to compare the similarity between images using a matching process. Generally, the face image is sampled into a number of patches of the same size. Feature extraction is then applied on each patch. Finally, matching process is applied between probe and gallery faces. The advantage of this approach is that the sampled patches are not overlapped, which avoids the influence of occluded regions on the other informative parts. For example, Martinez et aleix. [15] sampled the face region into a fixed number of local patches. matching is then applied for similarity measure.

Other methods detect the keypoints from the face image, instead of local patches. For instance, Weng et al. [23] proposed to recognize persons of interest from their partial faces. To accomplish this task, they firstly detected keypoints and extract their textural and geometrical features. Next, point set matching is carried out to match the obtained features. Finally, the similarity of two faces is obtained through the distance between these two aligned feature sets. Keypoint based matching method is introduced in Duan et al. [5]. SIFT key- point descriptor is applied to select the appropriate keypoints. Gabor ternary pattern and point set matching are then applied to match the local keypoints for partial face recognition. In contrast to the above mentioned methods based on fixed-size patches matching or keypoints detection, McLaughlin et al. [16] applied a largest matching area at each point of the face image without any sampling.

Restoration approach: Here, the occluded regions in the probe faces are restored according to the gallery ones. For instance, Bagchi et al. [2] proposed to restore facial occlusions. The detection of the occluded regions is carried out by thresholding the depth map values of the 3D image. Then the restoration is taken on by Principal Component Analysis (PCA). There are also several approaches that rely on the estimation of the occluded parts. Drira et al. [4] applied a statisticl shape model to predict and restore the partial facial curves. Iterative closest point (ICP) algorithm has been used to remove occluded re- gions in [6]. The restoration is applied using a curve, which uses statistical estimation of the curves to manage the occluded parts. Partially observed curves are completed by using the curves model produced through the PCA technique.

Discard occlusion based approach: In order to avoid a bad reconstruc- tion process, these approaches aim to detect regions found to be occluded in the face image, and discard them completely from the feature extraction and classification process. Segmentation based approach is one of the best methods that detect firstly the occluded region part, and using only the non-occluded part in the following steps. For instance, Priya and Banu [19] divided the face image into small local patches. Next, to discard the

occluded region, they applied the support vector machine classifier to detect them. Finally, Mean based weight matrix is used on the non-occluded regions for face recognition. Alyuz et al. [1] applied an occlusion removal and restoration. They used the global masked projection to remove the occluded regions. Next, the partial Gappy PCA is applied for the restoration using eigenvectors. Similarly, Yu et al. [24] carried out a partial matching mechanism to effectively eliminates the occluded regions and then using the non-occluded regions in the matching process.

Since the publication of AlexNet architecture in 2012 by krizhevsky et al. [10], deep CNN have become a common approach in face recognition. It has also been successfully used in face recognition under occlusion variation [7]. We find deep learning based method based on the fact that human visual system automatically ignores the occluded regions and only focuses on the non-occluded ones. For example, Song et al. [21] proposed a mask learning technique in order to discard the feature elements of the masked region for the recognition process.

Inspired by the high performance of CNN based methods that have strong robustness to illumination, facial expression and facial occlusion changes, we propose in this paper a discard occlusion based method and deep CNN based model to address the problem of masked face recognition during COVID–19 pandemic. Experimental results are carried out on Real-world Masked Face Recognition Dataset (RMFRD) presented in [22].

3 Motivation And Contribution Of The Paper

We start by localizing the mask region. To do so, we apply a cropping filter in order to obtain only he informative regions of the masked face (i.e. forehead and eyes ones). Next, we describe the selected regions using deep learning model. This strategy is more suitable in real-world applications comparing to restoration approaches. Recently, some works have applied a supervised learning on the missing region to restore them such as in [3]. This strategy, however, is a difficult and highly time-consuming process.

Despite the recent breakthroughs of deep learning architectures in pattern recognition tasks, they need to estimate millions of parameters in the fully con- nected layers that require powerful hardware with high processing capacity and memory. To address this problem, we present in this paper an efficient quantization based pooling method for face recognition using VGG–16 pre-trained model. To do so, we only consider the feature maps at the last convolutional layer (also called channels) using Bag-of-Features (BoF) paradigm.

The basic idea of the classical BoF paradigm is to represent images as orderless sets of local features. To get these sets, the first step is to extract local features from the training images, each feature represents a region from the image. Next, the whole features are quantized to compute a codebook. Test image's features are then assigned to the nearest code in the codebook to be represented by a histogram. In the literature, BoF paradigm has been largely used for handcrafted feature quantization [11, 12] to accomplish image classification tasks. A comparative study between BoF and deep learning for image classification has been made in Loussaief and Abdelkrim [13]. To take the advantages of the two

techniques, BoF is considered, in this paper, as a pooling layer in our trainable convolutional layers which aims to reduce the number of parameters and makes possible to classify masked face images.

This deep quantization technique present many advantages. It ensures a lightweight representation that makes real-world masked face recognition a feasible task. Moreover, the masked region vary from face to another, which leads to informative images from different sizes. The proposed deep quantiza- tion allows classifying images from different sizes in order to handle this issue. In addition, the Deep BoF approach uses a differentiable quantization scheme that enables simultaneous training of both the quantizer and the rest of the network, instead of using fixed quantization merely to minimize the model size[17]. It is worth stating that our proposed method doesn't need to learn on the mission region after removing the mask. It instead improves the generalization of face recognition process in the presence of the mask during the pandemic of coronavirus.

4 The Proposed Method

The Figure 3 presents an overview of the proposed method. It passes by four steps:

4.1 Preprocessing and cropping filter

The images of this dataset are already cropped around the face, so we don't need a face detection stage to localize the face from each image. However, we need to correct the rotation of the face so that we can remove the masked region efficiently. To do so, we detect 68 facial landmarks using Dlib-ml open source library introduced in [8]. According to the eyes location, we apply a 2D rotation to make them horizontal as presented in Figure 1.

The next step is to apply a cropping filter in order to extract only the non- masked region. To do so, we firstly normalize all face images into 240 x 240 pixels. Next, we use the partition into blocks. The principle of this technique is to divide the image into 100 fixed-size square blocks (24 x 24 pixels in our case). Then we extract only the blocks including the non-masked region (blocks from number 1 to 50). Finally, we eliminate the rest of the numbers of the blocks as presented in Figure 2.

4.2 Feature extraction layer

We extract deep features using VGG–16 face CNN descriptor [20] from the 2D images. It is trained on ImageNet dataset which has over 14 million images and 1000 classes. Its name VGG–16 comes from the fact that it has 16 layers. Its layers consists of convolutional layers, Max Pooling layers, Activation layers, Fully connected layers. There are 13 convolutional layers, 5 Max Pooling layers and 3 Dense layers which sums up to 21 layers but only 16 weight layers. Figure 4 presents VGG–16 architecture. In this work, we only consider the feature maps (FMs) at the last convolutional layer, also called channels. These features will be used in the following in the quantization stage.

4.3 Deep bag of features layer

From the i^{th} image, we extract feature maps using the feature extraction layer described above. In order to measure the similarity between the extracted feature vectors and the *codewords* also called *term vector,* we applied the RBF kernel as similarity metric as proposed in [17]. Thus, the first sublayer will be composed of RBF neurons, each neuron is referred to a codeword.

As presented in Figure 3, the size of the extracted feature map defines the number of the feature vectors that will be used in the BoF layer. Here we refer by V_i to the number of feature vectors extracted from the i^th image. For example, if we have 10 x 10 feature maps from the last convolutional layer of VGG–16 model, we will have 100 feature vectors to feed the quantization step using BoF paradigm. To build the *codebook,* the initialization of the RBF neurons can be carried out manually or automatically using all the extracted feature vectors overall the dataset. The most used automatic algorithm is of course k-means. Let F the set of all the feature vectors, defined by: $F =$

$\{V_{ij}, i = 1 \dots V, j = 1 \dots V_i\}$ and V_k is the number of the RBF neurons centers
referred by c_k. Note that these RBF centers are learned afterward to get the final codewords.

The quantization is then applied to extract the histogram with a predefined number of bins, each bin is referred to a *codeword.* RBF layer is then used as a similarity measure, it contains 2 sublayers:

(I) RBF layer: measures the similarity of the input features of the probe faces to the RBF centers.

Formally: the j^{th} RBF neuron $\varphi(X_j)$ is defined by:

$\varphi(X_j) = \exp(-\|x - c_j\|^2/\sigma_j^2)$, (1)

Where x is a feature vector and c_j is the center of the j^{th} RBF neuron.

(II) Quantization layer: the output of all the RBF neurons is collected in this layer that contains the histogram of the global quantized feature vector that will be used for the classification process. The final histogram is defined by:

$$h_i = \sum_{k}^{N^k} \varphi(V_{jk}) \qquad (2)$$

Where $\varphi(V)$ is the output vector of the RBF layer over the c_k bins.

4.4 Fully connected layer and classification

Once the global histogram is computed, we pass to the classification stage to assign each test image to its identity. To do so, we apply Multilayer perceptron classifier (MLP) where each face is represented by a term vector. Deep BoF network can be trained using back-propagation and gradient descent. Note that 10

cross validation strategy is applied in our experiments on RMFRD dataset. We note $V = [v1,..., v_k]$ the term vector of each face, where each v_i refers to the occurrence of the term i in the given face. t is the number of attributes, and m is the number of classes (face identities). Test faces are defined by their codeword V MLP uses a set of term occurrences as input values ($v_{i)}$ and associated weights ($w_{i)}$ and a sigmoid function (g) that sums the weights and maps the results to an output (y).. Note that the number of hidden layers used in our experiments is given by: $^{m+t}/_{2}$.

5 Experimental Results

We assess the efficiency of the proposed method on a publicly available masked face dataset. We present in the following the dataset description and the ob- tained results.

5.1 Dataset description

Real-World-Masked-Face-Dataset [22] is a masked face dataset devoted mainly to improve the recognition performance of the existing face recognition tech- nology on the masked faces during the COVID−19 pandemic. It contains three types of images namely : Masked Face Detection Dataset (MFDD), Real-world Masked Face Recognition Dataset (RMFRD) and Simulated Masked Face Recognition Dataset (SMFRD). Specifically, MFDD contains 24771 masked face images, it can be used to train an accurate masked face detection model, which serves for the subsequent masked face recognition task. RMFRD con- tains 5000 pictures of 525 people wearing masks, and 90000 images of the same 525 subjects without masks. SMFRD contains a simulated masked face dataset of 500000 face images of 10000 subjects. In this paper, we only focus on RMFRD dataset in order to address the problem of masked face recognition task during COVID−19 pandemic. Figure 5 presents some pairs of face images without mask (up) and their corresponding face images with mask (down).

Table 1: Recognition performance on RMFRD dataset using three codebook sizes.

Method	Size 1	Size 2	Size 3
term vectors	50	60	70
Conv5 FM1	88.5%	89.2%	87.1%
Conv5 FM2	90.8%	87.4%	87.2%
Conv5 FM3	91.0%	**91.3%**	90.1%

5.2 Method performance

Our dataset is imbalanced (5000 masked faces VS 90000 non-masked faces). Therefore, we need to apply the cropping filter on the masked faces before the feature extraction process.

RMFRD faces were firstly preprocessed as described in the Section 4.1. Using the normalized 2D faces of sizes 240 x 240 pixels, we apply VGG−16 pretrained model to extract the best features from the last convolutional layer as presented in Section 4.2. In this layer the feature maps are of size 14 x 14, with 512 channels. The quantization is then applied to extract the histogram of 70 bins as presented in Section 4.3. Finally, MLP is applied to classify faces as presented in Section 4.4. In this experiment, the 10 cross-

validation strategy is used to evaluate the recognition performance. The experiments are repeated ten times in the RMFRD dataset, where 9 samples are used as the training set and the remaining sample as the testing set, and the average results are calculated.

Table 1 reports the classification rates on RMFRD dataset using three different sizes of the codebook (i.e. number of codewords in RBF layer) by (i.e. 50, 60, 70 term vectors per image). We can see that the best recognition rate is obtained using the third FMs in the last convolutional layer with 60 codewords by 91.3%. Note that we don't compare the obtained results with state-of-the-art ones since there is no method has been assessed on the RMFRD dataset up to this point.

5.3 Discussion

This high accuracy achieved is due to the best features extracted from the last convolutional layer of VGG–16 model, and the high efficiency of the proposed BoF paradigm that gives a lightweight and more discriminative description comparing to classical CNN with softmax function. Moreover, dealing with only the informative regions (unmasked ones) and the high generalization of the proposed method makes it applicable in real-time applications. Other methods, however, aim to unmask the masked face using generative networks such as in [3]. This strategy is not preferable for real-world application since the image completion of the removed mask region is a greedy task, and often not efficient.

We can also notice that average size of codebook (RBF neurons) gives higher recognition rate comparing to 70 codebook size on RMFRD dataset. This behavior can be explained by the fact that the performance of the pro- posed deep BoF-based method depends on the number of the extracted deep features. Moreover, using different sizes of the pooling layer can increase the scale-invariance and bring more spatial information to the fully connected layer.

It is worth noting that other deep learning based methods apply a spatial region assignment in order to introduce more spatial information. Out method on the other hand, is full automatic since no spatial region assignment is applied before the feature vectors quantization.

6 Conclusion

In real-world scenarios (i.e. unconstrained environments), human faces might be occluded by other objects such as facial mask. This makes the face recog- nition process a very challenging task. Consequently, current face recognition methods will easily fail to make an efficient recognition. The proposed method improves the generalization of face recognition process in the presence of the mask. To accomplish this task, we proposed a deep learning based method and quantization based technique to deal with the recognition of the masked faces. The proposed method can also be extended to richer applications such as vio- lence video retrieval and video surveillance. The proposed method achieved a high recognition performance. For the best of our knowledge, this is the first work that addresses the problem of masked face recognition during COVID–19 pandemic. It is worth stating that this study is not limited to this pandemic period since a lot of people are self-aware constantly, they take care of their

health and wear masks to protect themselves against pollution and to reduce other pathogens transmission.

Declarations

The used faces belong to the Real-World-Masked-Face-Dataset. This dataset is freely available to industry and academia. It is available at this link: https://github.com/X-zhangyang/Real-World-Masked-Face-Dataset

Competing interests: The authors declare no competing interests.

References

1. Alyuz, B. Gokberk, and L. Akarun. 3-d face recognition under occlusion using masked projection. *IEEE Transactions on Information Forensics and Security*, 8(5):789–802, 2013.
2. Bagchi, D. Bhattacharjee, and M. Nasipuri. Robust 3d face recognition in presence of pose and partial occlusions or missing parts. *arXiv preprint arXiv:1408.3709*, 2014.
3. U. Din, K. Javed, S. Bae, and J. Yi. A novel gan-based network for unmasking of masked face. *IEEE Access*, 8:44276–44287, 2020.
4. Drira, B. Ben Amor, A. Srivastava, M. Daoudi, and R. Slama. 3d face recogni- tion under expressions, occlusions, and pose variations. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 35(9):2270–2283, 2013.
5. Duan, J. Lu, J. Feng, and J. Zhou. Topology preserving structural matching for automatic partial face recognition. *IEEE Transactions on Information Forensics and Security*, 13(7):1823–1837, 2018.
6. S. Gawali and R. R. Deshmukh. 3d face recognition using geodesic facial curves to handle expression, occlusion and pose variations. *International Journal of Computer Science and Information Technologies*, 5(3):4284–4287, 2014.
7. He, H. Li, Q. Zhang, and Z. Sun. Dynamic feature matching for partial face recog- nition. *IEEE Transactions on Image Processing*, 28(2):791–802, 2018.
8. E. King. Dlib-ml: A machine learning toolkit. *The Journal of Machine Learning Research*, 10:1755–1758, 2009.
9. L. Koudelka, M. W. Koch, and T. D. Russ. A prescreener for 3d face recognition using radial symmetry and the hausdorff fraction. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)-Workshops*, pages 168–168. IEEE, 2005.
10. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convo- lutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.
11. -C. Lian, Z. Li, B.-L. Lu, and L. Zhang. Max-margin dictionary learning for multiclass image categorization. In *European Conference on Computer Vision*, pages 157–170. Springer, 2010.

12. Lobel, R. Vidal, D. Mery, and A. Soto. Joint dictionary and classifier learning for categorization of images using a max-margin framework. In *Pacific-Rim Symposium on Image and Video Technology*, pages 87–98. Springer, 2013.
13. Loussaief and A. Abdelkrim. Deep learning vs. bag of features in machine learning for image classification. In *2018 International Conference on Advanced Systems and Electric Technologies (IC ASET)*, pages 6–10. IEEE, 2018.
14. Lu, A. K. Jain, and D. Colbry. Matching 2.5 d face scans to 3d models. *IEEE transactions on pattern analysis and machine intelligence*, 28(1):31–43, 2005.
15. M. Mart´ınez. Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class. *IEEE Transactions on Pattern analysis and machine intelligence*, 24(6):748–763, 2002.
16. McLaughlin, J. Ming, and D. Crookes. Largest matching areas for illumination and occlusion robust face recognition. *IEEE transactions on cybernetics*, 47(3):796–808, 2016.
17. Passalis and A. Tefas. Learning bag-of-features pooling for deep convolutional neural networks. In *Proceedings of the IEEE international conference on computer vision*, pages 5755–5763, 2017.
18. Peng, M. Bennamoun, and A. S. Mian. A training-free nose tip detection method from face range images. *Pattern Recognition*, 44(3):544–558, 2011.
19. N. Priya and R. W. Banu. Occlusion invariant face recognition using mean based weight matrix and support vector machine. *Sadhana*, 39(2):303–315, 2014.
20. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.
21. Song, D. Gong, Z. Li, C. Liu, and W. Liu. Occlusion robust face recognition based on mask learning with pairwise differential siamese network. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 773–782, 2019.
22. Wang, G. Wang, B. Huang, Z. Xiong, Q. Hong, H. Wu, P. Yi, K. Jiang, N. Wang,
23. Pei, et al. Masked face recognition dataset and application. *arXiv preprint arXiv:2003.09093*, 2020.
24. Weng, J. Lu, and Y.-P. Tan. Robust point set matching for partial face recognition. *IEEE transactions on image processing*, 25(3):1163–1176, 2016.
25. Yu, Y. Gao, and J. Zhou. 3d face recognition under partial occlusions using radial strings. In *2016 IEEE International Conference on Image Processing (ICIP)*, pages 3016–3020. IEEE, 2016.

Figures

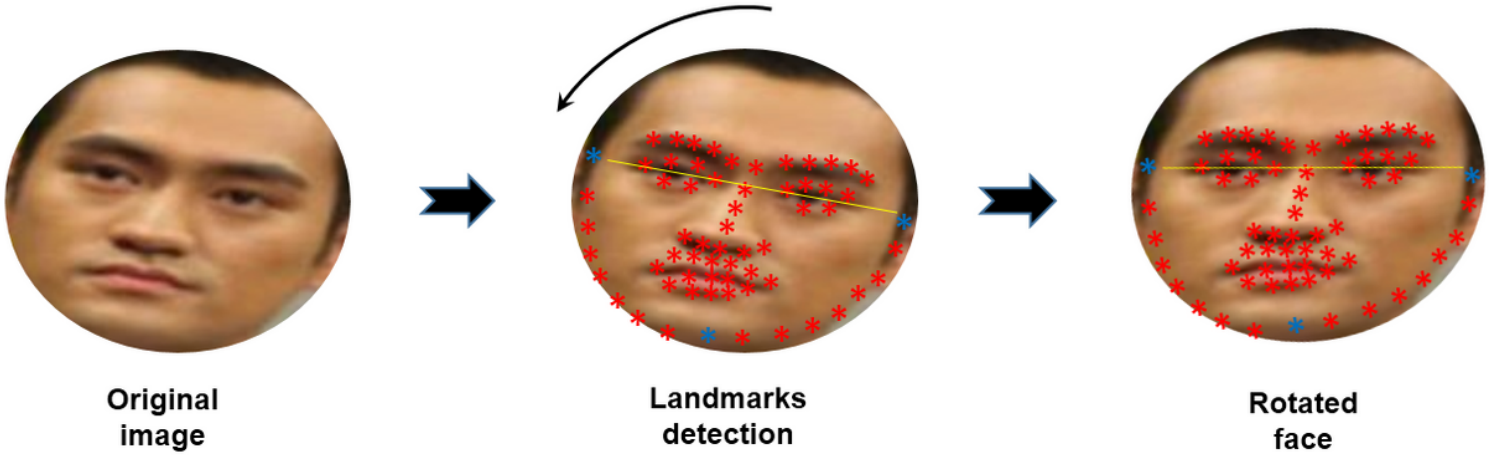

Figure 1

2D Face rotation.

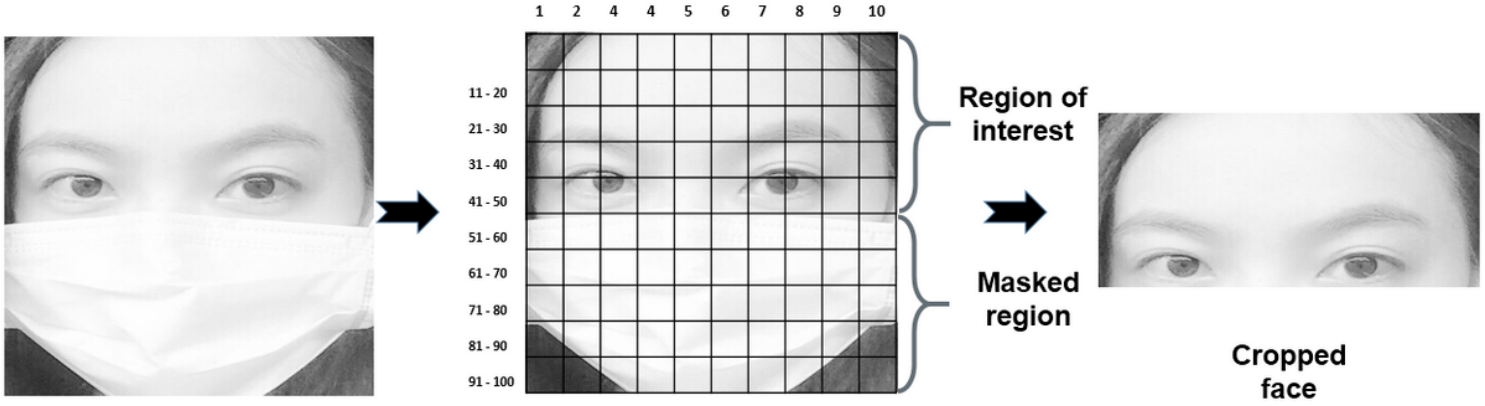

Figure 2

Sampling the masked face image into 100 regions of the same size and cropping filter.

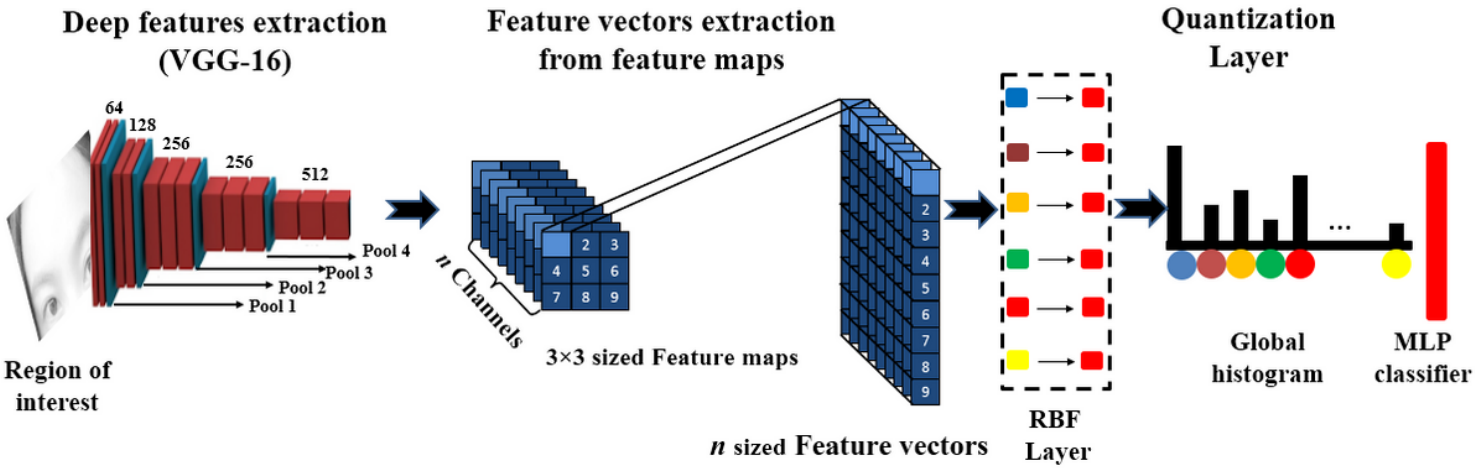

Figure 3

Overview of the proposed method.

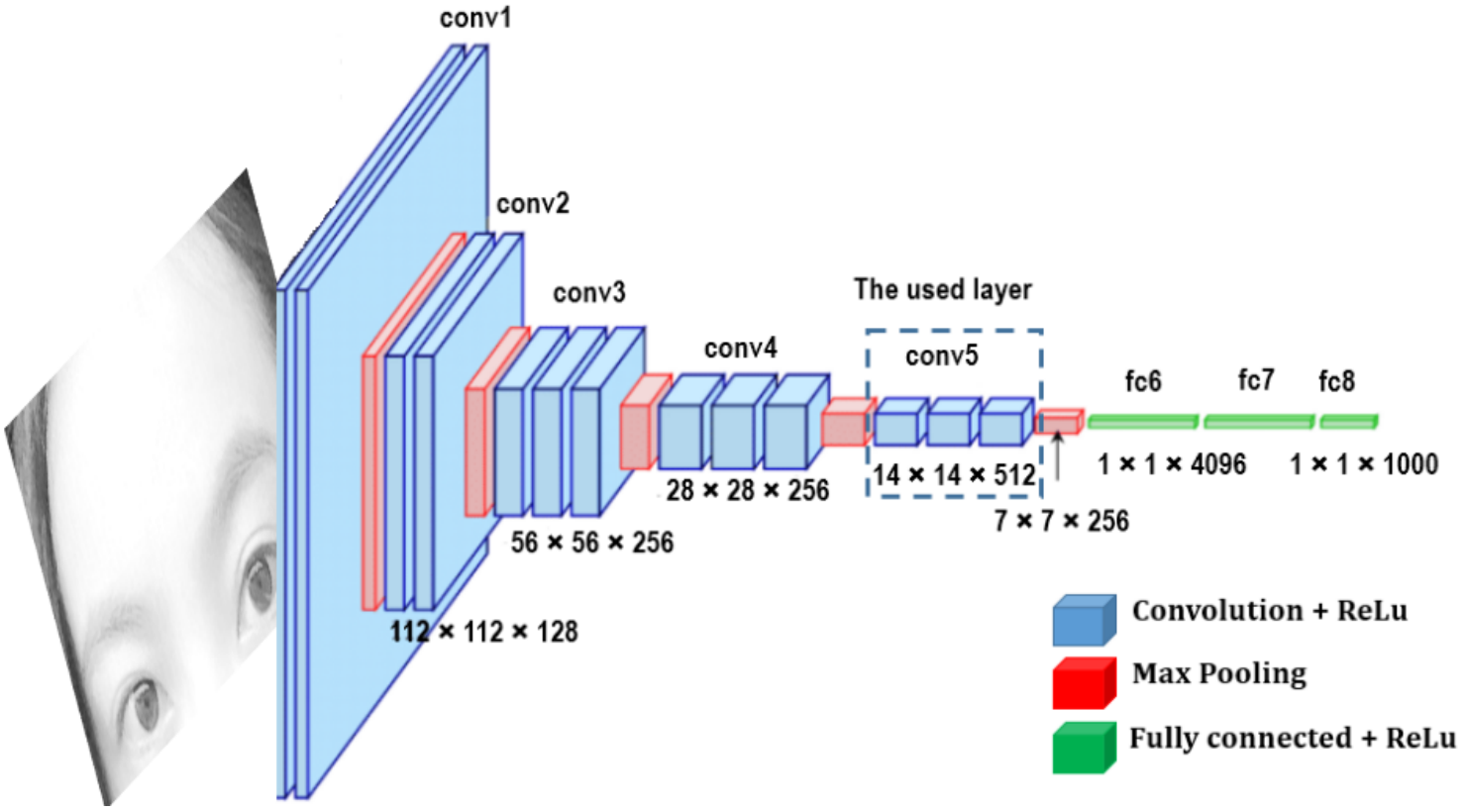

Figure 4

VGG-16 network architecture.

Figure 5

Pairs of face images from RMFRD dataset: face images without mask (up) and their corresponding face images with mask (down).