Signal, Image and Video Processing (2022) 16:605–612
https://doi.org/10.1007/s11760-021-02050-w

ORIGINAL ARTICLE

Efficient masked face recognition method during the COVID-19 pandemic

Walid Hariri[1]

Received: 15 July 2020 / Revised: 29 April 2021 / Accepted: 7 October 2021 / Published online: 15 November 2021

Abstract

The coronavirus disease (COVID-19) is an unparalleled crisis leading to a huge number of casualties and security problems. In order to reduce the spread of coronavirus, people often wear masks to protect themselves. This makes 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 occlusion removal and deep learning-based features in order to address the problem of the masked face recognition process. The first step is to remove the masked face region. Next, we apply three pre-trained deep Convolutional Neural Networks (CNN), namely VGG-16, AlexNet, and ResNet-50, and use them to extract deep features from the obtained regions (mostly eyes and forehead regions). The Bag-of-features paradigm is then applied to 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, Multilayer Perceptron (MLP) is applied for the classification process. Experimental results on Real-World-Masked-Face-Dataset show high recognition performance compared to other state-of-the-art methods.

Keywords Face recognition · COVID-19 · Masked face · Deep learning

1 Introduction

The COVID-19 can be spread through contact and contaminated surfaces; therefore, the classical biometric systems based on passwords or fingerprints are not anymore safe. Face recognition is safer without any need to touch any device. Recent studies on coronavirus have proven that wearing a face mask by a healthy and infected population reduces considerably the transmission of this virus. However, wearing the mask face causes the following problems: (1) fraudsters and thieves take advantage of the mask, stealing and committing crimes without being identified. (2) community access control and face authentication have become very difficult tasks when a grand part of the face is hidden by a mask. (3) existing face recognition methods are not efficient when wearing a mask which cannot provide the whole face image for description. (4) exposing the nose region is very important in the task of face recognition since it is used for face normalization [24], pose correction [18], and face matching [13]. 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 not. This can be applied in public places where the mask is compulsory. Masked face recognition, on the other hand, aims to recognize a face with a mask basing on the eyes and the forehead regions. In this paper, we handle the second task using a deep learning-based method. We use a 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: Sect. 2 presents the related works. In Sect. 3, we present the motivation and contribution of the paper. The proposed method is detailed in Sect. 4. Experimental results are presented in Sect. 5. The conclusion ends the paper.

✉ Walid Hariri
hariri@labged.net

[1] Labged Laboratory, Computer Science department, Badji Mokhtar Annaba University, Annaba, Algeria

2 Related works

Occlusion is a key limitation of real-world 2D face recognition methods. Generally, 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 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 occlusion removal 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 to each patch. Finally, a 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 and Aleix [20] 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. [30] proposed to recognize persons of interest from their partial faces. To accomplish this task, they firstly detected keypoints and extracted their textural and geometrical features. Next, point set matching is carried out to match the obtained features. Finally, the similarity of the two faces is obtained through the distance between these two aligned feature sets. Keypoint-based matching method is introduced in Duan et al. [7]. SIFT keypoint 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. [21] applied the 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. [3] 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) [31]. There are also several approaches that rely on the estimation of the occluded parts. Drira et al. [6] applied a statistical shape model to predict and restore the partial facial curves. Iterative closest point (ICP) algorithm has been used to remove occluded regions in [8]. The restoration is applied using a curve, which uses a 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.

Occlusion removal approach: In order to avoid a bad reconstruction 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 [26] 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, a mean-based weight matrix is used on the non-occluded regions for face recognition. Alyuz et al. [2] 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.

Since the publication of AlexNet architecture in 2012 by Krizhevsky et al. [14], deep CNN has become a common approach in face recognition. It has also been successfully used in face recognition under occlusion variation [1]. It is seen that the deep learning-based method is founded on the fact that the human visual system automatically ignores the occluded regions and only focuses on the non-occluded ones. For example, Song et al. [28] 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 an occlusion removal approach and deep CNN-based model to address the problem of masked face recognition during the COVID-19 pandemic. Motivations and more details about the proposed method are presented in the following sections.

3 Motivation and contribution of the paper

Motivated by the efficiency and the facility of the occlusion removal approaches, we apply this strategy to discard the masked regions. Experimental results are carried out on Real-world Masked Face Recognition Dataset (RMFRD) and Simulated Masked Face Recognition Dataset (SMFRD) presented in [29]. We start by localizing the mask region. To do so, we apply a cropping filter in order to obtain only the informative regions of the masked face (i.e., forehead and eyes). Next, we describe the selected regions using a pre-trained deep learning model as a feature extractor. This strategy is more suitable in real-world applications comparing to restoration approaches. Recently, some works have applied supervised learning on the missing region to restore

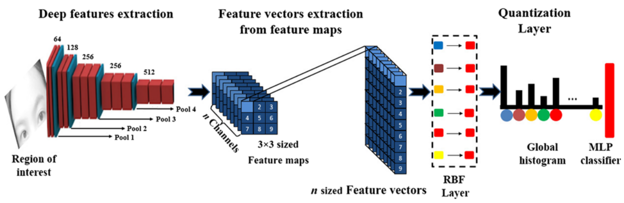

Fig. 1 Overview of the proposed method

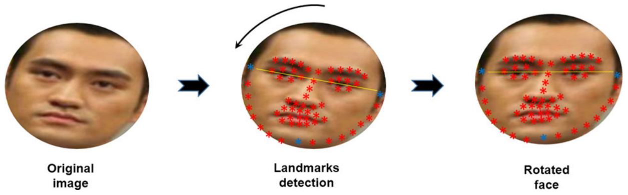

Fig. 2 2D Face rotation

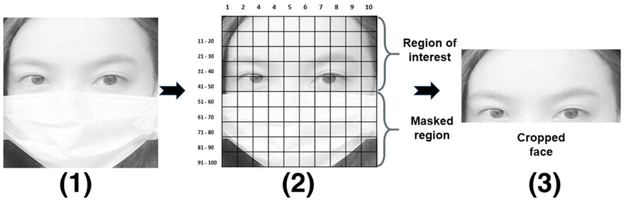

Fig. 3 (1): Masked face. (2): Sampling the masked face image into 100 regions of the same size. (3): Cropping filter

them such as in [4]. 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 connected 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 three pre-trained models. To do so, we only consider the feature maps at the last convolutional layers (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 features are then assigned to the nearest code in the codebook to be represented by a histogram. In the literature, the BoF paradigm has been largely used for handcrafted feature quantization [16] to accomplish image classification tasks. A comparative study between BoF and deep learning for image classification has been made in Loussaief and Abdelkrim [17]. To take full advantage of the two techniques, in this paper we can consider BoF as a pooling layer in our trainable convolutional layers. This aims to reduce the number of parameters and makes it possible to classify masked face images.

This deep quantization technique presents many advantages. It ensures a lightweight representation that makes the real-world masked face recognition process a feasible task. Moreover, the masked regions vary from one face to another, which leads to informative images of different sizes. The proposed deep quantization allows classifying images from different sizes in order to handle this issue. Besides, 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 [23]. It is worth stating that our proposed method doesn't need to be trained on the mission region after removing the mask. It instead improves the generalization of the face recognition process in the presence of the mask during the pandemic of coronavirus.

4 The proposed method

Fig. 1 presents an overview of the proposed method. It has four steps:

4.1 Preprocessing and cropping filter

The images of the used 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 [12]. According to the eye locations, we apply a 2D rotation to make them horizontal as presented in Fig. 2.

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 × 240 pixels. Next, we partition a face into blocks. The principle of this technique is to divide the image into 100 fixed-size square blocks (24 × 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 blocks as presented in Fig. 3.

4.2 Feature extraction layer

To extract deep features from the informative regions, we have employed three pre-trained models as feature extractors:

VGG-16: [27] is trained on the ImageNet dataset which has over 14 million images and 1000 classes. Its name VGG-16 comes from the fact that it has 16 layers. It contains different layers including convolutional layers, Max Pooling layers, Activation layers, and Fully Connected (fc) layers. There are 13 convolutional layers, 5 Max Pooling layers, and 3 Dense

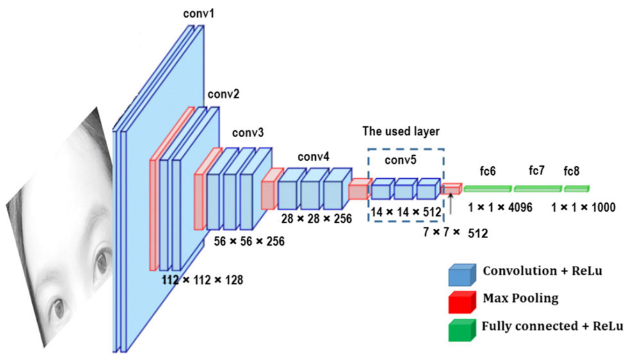

Fig. 4 VGG-16 network architecture introduced in [27]

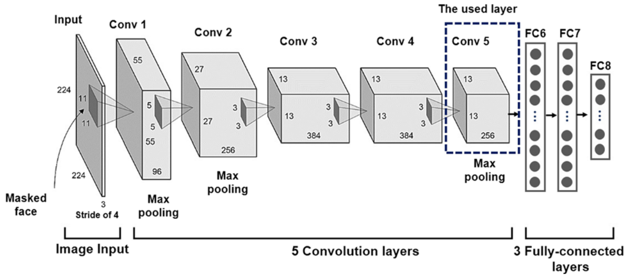

Fig. 5 AlexNet network architecture introduced in [15]

layers which sum up to 21 layers but only 16 weight layers. In this work, we choose the VGG-16 as the base network, and we only consider the feature maps (FMs) at the last convolutional layer, also called channels. This layer is employed as a feature extractor and will be used for the quantization in the following stage. Fig. 4 presents VGG-16 architecture.

AlexNet: has been successfully employed for image classification tasks [15]. This deep model is pre-trained on a few millions of images from the ImageNet database through eight learned layers, five convolutional layers and three fully connected layers. The last fully connected layer allows to classify one thousand classes. The fifth convolutional layer is used in this paper to extract deep features (See Fig. 5).

ResNet-50: [10] has been successfully used in various pattern recognition tasks, such as face and pedestrian detection [22]. It containing 50 layers trained on the ImageNet dataset. This network is a combination of Residual network integrations and Deep architecture parsing. Training with ResNet-50 is faster due to the bottleneck blocks. It is composed of five convolutional blocks with shortcuts added between layers. The last convolution layer is used to extract Deep Residual Features (DRF). Fig. 6 shows the architecture of the ResNet-50 model.

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 [23]. Thus, the first sub-layer will be composed of RBF neurons, each neuron is referred to a codeword.

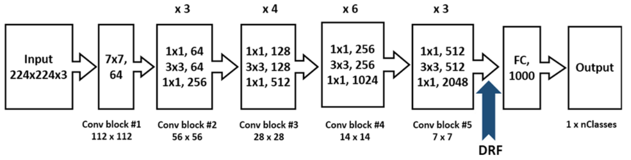

Fig. 6 ResNet-50 network architecture introduced in [10]. The extracted DRF are shown

As presented in Fig. 1, 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 $i^t h$ image. For example, if we have 10×10 feature maps from the last convolutional layer of the chosen pre-trained model, we will have 100 feature vectors to feed the quantization step using the 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: $F = \{V_{ij}, i = 1 \ldots V, j = 1 \ldots V_i\}$ 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 sub-layers:

(I) **RBF layer**: measures the similarity of the input features of the probe faces to the RBF centers. Formally: the j^{th} RBF neuron $\phi(X_j)$ is defined by Eq.(1):

$$\phi(X_j) = \exp(\|x - c_j\|_2/\sigma_j), \tag{1}$$

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 Eq.(2), where $\phi(V)$ is the output vector of the RBF layer over the c_k bins.

$$h_i = V_j \sum_{k}^{N^k} \phi(V_{jk}) \tag{2}$$

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 the 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 the 10-fold cross-validation strategy is applied in our experiments on the RMFRD dataset. We note $V = [v1, \ldots, v_k]$ 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 output (y). Note that the number of hidden layers used in our experiments is given by: $\frac{m+t}{2}$.

5 Experimental results

To evaluate the proposed method, we carried out experiments on very challenging masked face datasets. In the following, we present the datasets' content and variations, the experimental results using the quantization of deep features obtained from three pre-trained models, and a comparative study with other state-of-the-arts.

5.1 Dataset description

Real-World-Masked-Face-Dataset [29] is a masked face dataset devoted mainly to improve the recognition performance of the existing face recognition technology 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). In this paper, we focus on the last two datasets described in the following.

a) **RMFRD** is one of the richest real-world masked face datasets. It contains 5,000 images of 525 subjects with masks, and 90,000 images without masks which represent 525 subjects. A semi-automatic annotation strategy has been used to crop the informative face parts. Fig. 7 presents some pairs of face images from RMFRD dataset.

b) **SMFRD** contains 500,000 simulated masked faces of 10,000 subjects from two known datasets Labeled Faces in the Wild (LFW) [11] and Webface [32]. The simulation is carried out using Dlib library [5]. This dataset is balanced but more challenging since the simulated masks are not necessarily in the right position. Fig. 8 shows some examples of simulated masked faces from SMFRD dataset.

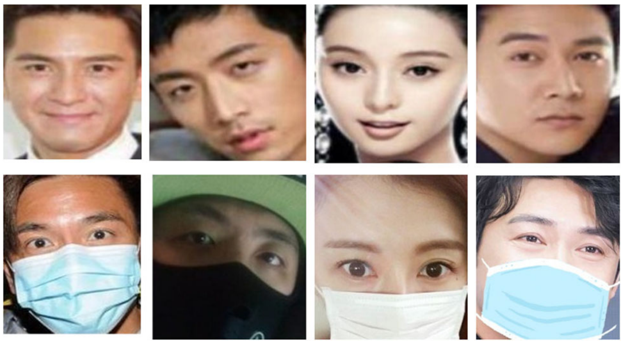

Fig. 7 Pairs of face images from RMFRD dataset: face images without a mask (up) and with a mask (down)

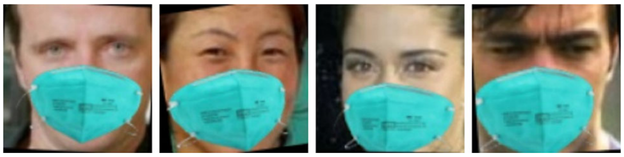

Fig. 8 Masked faces from SMFRD dataset

5.2 Method performance

The face images were firstly preprocessed as described in Sect. 4.1. In contrast to SMFRD dataset, RMFRD is imbalanced (5,000 masked faces vs 90,000 non-masked faces). Therefore, we have applied an over-sampling by cropping some non-masked faces to get an equivalent number of cropped and full faces. Next, using the normalized 2D faces, we employ the three pre-trained models (VGG-16, AlexNet and ResNet-50) separately to extract deep features from their last convolutional layers as presented in Sect. 4.2. The output features are (14 × 14 × 512, 13 × 13 × 256, 7 × 7 × 2048) dimensional, respectively.

The quantization is then applied to extract the histogram of a number of bins as presented in Sect. 4.3. Finally, MLP is applied to classify faces as presented in Sect. 4.4. In this experiment, the 10-fold cross-validation strategy is used to evaluate the recognition performance. The experiments are repeated ten times in RMFRD and SMFRD datasets separately, 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 the RMFRD dataset using four different sizes of the codebook (i.e., number of codewords in RBF layer) by (i.e., 50, 60, 70, 100 term vectors per image). We can see that the best recognition rate is obtained using the third FMs in the last convolutional layer

Table 1 Recognition performance on RMFRD dataset using four codebook sizes

Method term vectors	Size 1 50	Size 2 60	Size 3 70	Size 4 100
	VGG-16	Model		
Conv5 FM1 14×14×512	88.5%	89.2%	87.1%	87.5%
Conv5 FM2 14×14×512	90.8%	87.4%	87.2%	88.0%
Conv5 FM3 14×14×512	91.0%	**91.3%**	90.1%	89.8%
	AlexNet	Model		
Conv5 FM 13 × 13 × 256	84.3%	85.7%	85.9%	86.6%
	ResNet-50	Model		
Conv5 FM 7×7×2048	87.4%	87.9%	89.5%	89.3%

Table 2 Recognition performance on SMFRD dataset using four codebook sizes

Method term vectors	Size 1 50	Size 2 60	Size 3 70	Size 4 100
	VGG-16	Model		
Conv5 FM1 14×14×512	82.4%	83.7%	84.5%	84.7%
Conv5 FM2 14×14×512	83.1%	83.5%	85.0%	85.4%
Conv5 FM3 14×14×512	81.7%	81.3%	84.4%	85.6
	AlexNet	Model		
Conv5 FM 13 × 13 × 256	83.7%	83.9%	84.2%	86.0%
	ResNet-50	Model		
Conv5 FM 7×7×2048	83.5%	84.7%	**88.9%**	88.5%

from VGG-16 with 60 codewords by 91.3%. The second FMs achieved 90.8% with 50 codewords and outperformed the first FMs over the four codeword sizes. AlexNet, on the other hand, realized 86.6% with 100 codewords where the best recognition rate achieved by ResNet-50 was 89.5% with 70 codewords. In this experiment, it is clear that VGG-16 outperformed the AlexNet and ResNet-50 models.

Table 2 reports the classification rates on the SMFRD dataset. The highest recognition rate is achieved by the ResNet-50 through the quantization of DRF features by 88.9%. This performance is achieved using 70 codewords that feed an MLP classifier. AlexNet model realized good recognition rates comparing to the VGG-16 model (86.0% vs 85.6% as highest rates).

5.3 Performance comparison

To further evaluate the performance of our proposed method, we have compared the obtained experimental results with those of other face recognizers on the RMFRD and SMFRD datasets as follows:

Comparison with transfer learning-based technique: We have tested the face recognizer presented in [19] that achieved a good recognition accuracy on two subsets of the FERET database [25]. This technique is based on transfer learning (TL) which employs pre-trained models and fine-tuning them to recognize masked faces from RMFRD and SMFRD datasets. The reported results in Table 3 show that the proposed method outperformed the TL-based method on the RMFRD and SMFRD datasets.

Comparison with covariance-based technique: Covariance-based features have been applied in [9] and achieved high recognition performance on 3D datasets in the presence of occluded regions. We have employed this method using 2D-based features (texture, gray level, LBP) to extract covariance descriptors. The evaluation on the RMFRD and SMFRD datasets confirms the superiority of the proposed method as shown in Table 3.

Comparison with deep feature extractor: Another efficient face recognition method using the same pre-trained models (AlexNet and ResNet-50) is proposed in [1] and achieved a high recognition rate on various datasets. Nevertheless, the pre-trained models are employed in a different manner. It consists of applying a TL technique to fine-tune the pre-trained models to the problem of masked face recognition using an SVM classifier. We have tested this strategy on the masked faces. The results in Table 3 further demonstrate the efficiency of the BoF paradigm compared to the use of a machine learning-based classifier directly.

Table 3 Performance comparison with state-of-the-art methods

Method	Dataset	Technique	Masks	Accuracy
Luttrell et al. [19]	RMFRD	TL	yes	85.7%
Hariri et al. [9]	RMFRD	Covariance	yes	84.6%
Almabdy et al. [1]	RMFRD	CNN+SVM	yes	87.0%
Our	RMFRD	CNN+BoF	yes	***91.3%***
Luttrell et al. [19]	SMFRD	TL	yes	83.3%
Hariri et al. [9]	SMFRD	Covariance	yes	83.8%
Almabdy et al. [1]	SMFRD	CNN+SVM	yes	86.1%
Our	SMFRD	CNN+BoF	yes	***88.9%***

Table 4 Training and testing time on the RMFRD dataset in milliseconds

Method	AlexNet	VGG-16
Almabdy et al. [1]	train:550	train:930
	test:34	test:120
Our	train: 308	Train:605
	test:21	test:84

5.4 Computation and training times comparison

The comparison of the computation times between the proposed method and Almabdy et al.'s method [1] shows that the use of the BoF paradigm decreases the time required to extract deep features and to classify the masked faces (See Table 4). Note that this comparison is performed using the same pre-trained models (VGG-16 and AlexNet) on the RMFRD dataset. AlexNet is the lowest training and testing time compared to VGG-16 with less GPU memory usage.

5.5 Discussion

The obtained high accuracy compared to other face recognizers is achieved due to the best features extracted from the last convolutional layers of the pre-trained models, and the high efficiency of the proposed BoF paradigm that gives a lightweight and more discriminative power comparing to classical CNN with softmax function. Moreover, dealing with only the unmasked regions, 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 [4]. This strategy is a greedy task and not preferable for real-world applications.

The efficiency of each pre-trained model depends on its architecture and the abstraction level of the extracted features. When dealing with real masked faces, VGG-16 has achieved the best recognition rate, while ResNet-50 outperformed both VGG-16 and AlexNet on the simulated masked faces. This behavior can be explained by the fact that VGG-16 features fail to ensure a high discriminative power comparing to the DRF features that are still relatively steady compared to their results on the real masked faces. When dealing with other state-of-the-art recognizers, one of them applied the same pre-trained models with a different strategy. The proposed method outperformed TL-based method using the same pre-trained models. This performance is explained by the fact that the fc layers of the pre-trained models are more dataset-specific features (generally pre-trained on ImageNet dataset) which is a very different dataset, thus, this strategy is not always suitable for our task. Moreover, the proposed method outperformed previous methods in terms of training time. The achieved performance further confirms that the BoF paradigm is a slight representation that further reinforces the high discrimination power of the deep features to feed a machine learning-based classifier.

6 Conclusion

The proposed method improves the generalization of the 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. We employed three pre-trained models and used their last convolutional layers as deep features. The Bof paradigm is applied to quantize the obtained features and to feed an MLP classifier. The proposed method outperformed other state-of-the-art methods in terms of accuracy and time complexity. The proposed method can also be extended to richer applications, such as video retrieval and surveillance. In future work, we look at the application of deep ensemble models with additional pre-trained models to enhance the accuracy.

Acknowledgements The authors would like to thank the Agency for Research Results Valuation and Technological Development, Algeria (DGRSDT).

References

1. Almabdy, S., Elrefaei, L.: Deep convolutional neural network-based approaches for face recognition. Appl. Sci. **9**(20), 4397 (2019)
2. Alyuz, N., Gokberk, B., Akarun, L.: 3-d face recognition under occlusion using masked projection. IEEE Trans. Inf. Forensics Secur. **8**(5), 789–802 (2013)
3. Bagchi, P., Bhattacharjee, D., Nasipuri, M.: Robust 3d face recognition in presence of pose and partial occlusions or missing parts. *arXiv preprint* arXiv:1408.3709 (2014)
4. Din, N.U., Javed, K., Bae, S., Yi, J.: A novel gan-based network for unmasking of masked face. IEEE Access **8**, 44276–44287 (2020)
5. Dlib library: http://dlib.net/. [Accessed 02-January-2021]
6. Drira, H., Ben Amor, B., Srivastava, A., Daoudi, M., Slama, R.: 3d face recognition under expressions, occlusions, and pose variations. Pattern Anal. Mach. Intell. IEEE Trans. **35**(9), 2270–2283 (2013)
7. Duan, Y., Lu, J., Feng, J., Zhou, J.: Topology preserving structural matching for automatic partial face recognition. IEEE Trans. Inf. Forensics Secur. **13**(7), 1823–1837 (2018)
8. Gawali, A.S., Deshmukh, R.R.: 3d face recognition using geodesic facial curves to handle expression, occlusion and pose variations. Int. J. Computer Sci. Inf. Technol. **5**(3), 4284–4287 (2014)
9. Hariri, W., Tabia, H., Farah, N., Benouareth, A., Declercq, D.: 3d face recognition using covariance based descriptors. Pattern Recogn. Lett. **78**, 1–7 (2016)
10. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778 (2016)
11. Huang, G.B., Mattar, M., Berg, T., Learned-Miller, E.: Labeled faces in the wild: A database forstudying face recognition in unconstrained environments. (2008)
12. King, D.E.: Dlib-ml: A machine learning toolkit. J. Mach. Learn. Res. **10**, 1755–1758 (2009)
13. Koudelka, M.L., Koch, M.W., Russ, T.D.: 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*, pp. 168–168. IEEE (2005)
14. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: *Advances in neural information processing systems*, pp. 1097–1105 (2012)
15. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. Commun. ACM **60**(6), 84–90 (2017)
16. Lobel, H., Vidal, R., Mery, D., Soto, A.: Joint dictionary and classifier learning for categorization of images using a max-margin framework. In: *Pacific-Rim Symposium on Image and Video Technology*, pp. 87–98. Springer (2013)
17. Loussaief, S., Abdelkrim, A.: Deep learning vs. bag of features in machine learning for image classification. In: *2018 International Conference on Advanced Systems and Electric Technologies (IC_ASET)*, pp. 6–10. IEEE (2018)
18. Lu, X., Jain, A.K., Colbry, D.: Matching 2.5 d face scans to 3d models. IEEE Trans. Pattern Anal. Mach. Intell. **28**(1), 31–43 (2005)
19. Martínez, A.M.: Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class. IEEE Trans. Pattern Anal. Mach. Intell. **24**(6), 748–763 (2002)
20. Martínez, A.M.: Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class. IEEE Trans. Pattern Anal. Mach. Intell. **24**(6), 748–763 (2002)
21. McLaughlin, N., Ming, J., Crookes, D.: Largest matching areas for illumination and occlusion robust face recognition. IEEE Trans. Cybernet. **47**(3), 796–808 (2016)
22. Mliki, H., Dammak, S., Fendri, E.: An improved multi-scale face detection using convolutional neural network. *Signal Image and Video Processing* (2020)
23. Passalis, N., Tefas, A.: Learning bag-of-features pooling for deep convolutional neural networks. In: *Proceedings of the IEEE international conference on computer vision*, pp. 5755–5763 (2017)
24. Peng, X., Bennamoun, M., Mian, A.S.: A training-free nose tip detection method from face range images. Pattern Recogn. **44**(3), 544–558 (2011)
25. Phillips, P.J., Wechsler, H., Huang, J., Rauss, P.J.: The feret database and evaluation procedure for face-recognition algorithms. Image Vis. Comput. **16**(5), 295–306 (1998)
26. Priya, G.N., Banu, R.W.: Occlusion invariant face recognition using mean based weight matrix and support vector machine. Sadhana **39**(2), 303–315 (2014)
27. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. *arXiv preprint* arXiv:1409.1556 (2014)
28. Song, L., Gong, D., Li, Z., Liu, C., Liu, W.: Occlusion robust face recognition based on mask learning with pairwise differential siamese network. In: *Proceedings of the IEEE International Conference on Computer Vision*, pp. 773–782 (2019)
29. Wang, Z., Wang, G., Huang, B., Xiong, Z., Hong, Q., Wu, H., Yi, P., Jiang, K., Wang, N., Pei, Y., et al.: Masked face recognition dataset and application. *arXiv preprint* arXiv:2003.09093 (2020)
30. Weng, R., Lu, J., Tan, Y.-P.: Robust point set matching for partial face recognition. IEEE Trans. Image Process. **25**(3), 1163–1176 (2016)
31. Wold, S., Esbensen, K., Geladi, P.: Principal component analysis. Chemom. Intell. Lab. Syst. **2**(1–3), 37–52 (1987)
32. Yi, D., Lei, Z., Liao, S., Li, S.Z.: Learning face representation from scratch. *arXiv preprint* arXiv:1411.7923 (2014)