EAI Endorsed Transactions
on Industrial Networks and Intelligent Systems
Research Article EAI.EU

Facial mask-wearing prediction and adaptive gender classification using convolutional neural networks

Mohamed Oulad-Kaddour[1,*], Hamid Haddadou[1], Daniel Palacios-Alonso[2], Cristina Conde[2], and Enrique Cabello[2]

[1]Ecole nationale Supérieure d'Informatique (ESI), Laboratoire de la Communication dans les Systèmes Informatiques (LCSI), BP, 68M Oued-Smar,16270 Alger, Algeria
[2]Escuela Tecnica Superior de Ingenier ıa Informatica, Rey Juan Carlos Univrsity, C/Tilupan, , s/n 28933, Mostoles, Madrid, Spain

Abstract

The world has lived an exceptional time period caused by the Coronavirus pandemic. To limit Covid-19 propagation, governments required people to wear a facial mask outside. In facial data analysis, mask-wearing on the human face creates predominant occlusion hiding the important oral region and causing more challenges for human face recognition and categorisation. The appropriation of existing solutions by taking into consideration the masked context is indispensable for researchers. In this paper, we propose an approach for mask-wearing prediction and adaptive facial human-gender classification. The proposed approach is based on convolutional neural networks (CNNs). Both mask-wearing and gender information are crucial for various possible applications. Experimentation shows that mask-wearing is very well detectable by using CNNs and justifies its use as a prepossessing step. It also shows that retraining with masked faces is indispensable to keep up gender classification performances. In addition, experimentation proclaims that in a controlled face-pose with acceptable image quality' context, the gender attribute remains well detectable. Finally, we show empirically that the adaptive proposed approach improves global performance for gender prediction in a mixed context.

Received on 06 November 2023; accepted on 09 March 2024; published on 13 March 2024

Keywords: Gender classification; face biometrics; facial occlusions; mask-wearing, convolutional neural networks; explainable artificial intelligence

doi:10.4108/eetinis.v11i2.4318

1. Introduction

The world has experienced a pandemic of coronavirus. Since the coronavirus initially appeared, the pandemic has had numerous veritable effects and changed our daily life. After some months of severe human movement restrictions, the government started to be adapted to live with the virus. This adaptation was done by giving a list of sanitary guidelines to be followed by people. One of the most common instructions is to wear a facial mask outside, especially in public places. In computer vision, for its sub-discipline of medical image analysis, researchers have been working on the deployment of automated Covid-19 detection solutions since the apparition of the pandemic [1]. Great results were achieved for Covid-19 detection from pneumatic image analysis [1] [2], in particular for cases presenting related symptoms.

Biometrics also is a computer vision and artificial intelligence sub-discipline significantly touched by the pandemic. For existing solutions designated to artificial intelligence tasks like auto-surveillance, recognition, and human-traces categorization, researchers are also working on their adaption and generalization to the Covid-19 context [40] [41] [12]. In this last context, the interesting and one of the most used modalities in the practice, facial data is highly concerned about the pandemic. Indeed, to limit the propagation and contaminations of the recent coronavirus, people are required to wear a mask outside. For existing biometric

*Corresponding author. Email: m_ouled_kaddour@esi.dz

systems, this fact can cause remarkable performance degradation [23]. Indeed, in facial data analysis, the mask-wearing sanitary instruction generates facial occlusion [7] [11]. Specially, lower face occlusion based on the oral region containing the mouth and nose. Accordingly, a considerable part containing important textures will be unavailable, such as lips, mustache, and chin. It makes human face detection, recognition, and categorization more challenging. Mainly for existing solutions exploiting lower-face components as part of their region of interest in local approaches.

Figure 1. Examples of person wearing real mask in uncontrolled context with various eventual multiple combined occlusions (Kaggle dataset). (a): masked faces with regular frontal pose, (b):masked faces with irregular poses, (c, d): Masked faces with combined occlusions.

For categorization tasks from facial data, by taking the case of the widely studied attribute of gender [21], several approaches were proposed in the literature. Traditional and recent artificial intelligence tools were exploited to predict human gender from facial data. At the same, several challenges were confronted and taken into consideration like as image quality, facial expressions, and partial occlusions [3]. As already mentioned, wearing a mask creates particular facial occlusion hiding lower-face. Moreover, unlike normal circumstances [7], in the Covid-19 context occlusion caused by mask-wearing will be more frequent and common. Indeed, a considerable part of the population wears a mask outside. As illustrated in Fig. 1, mask-wearing-caused occlusion can also be combined with other controlled or uncontrolled occlusions such as scarf, veil, glass, or hat-wearing. We suppose that facial gender classification could be more defying.

Knowing that almost existing facial gender classification solutions are designated for unmasked context (faces without masks) containing possible random partial occlusions with non-important incidence rates for each occlusion type. We aim in this work to investigate simultaneously both binary classification tasks of mask-wearing prediction and human-gender classification in masked and unmasked scenarios by taking into consideration of eventual occlusion-combination. We propose an approach based on deep learning. Automatic access control to deny the passage of person without a mask, video surveillance to monitor people's respect for mask-wearing in public places like airports and metro, gender-based demographics collections to fix the gender more respecting the mask-wearing sanitary instruction, robotic adaptation for Covid-19 context and face reconstruction aiming the retrieving of the whole face in occluded images are examples where both mask-wearing and human-gender information's can be exploited.

As main contributions in this paper, we cite:

- Proposition of an approach for facial mask-wearing prediction and gender classification using convolutional neural networks.
- Real-world experimental evaluation of facial mask-wearing prediction in the wild using convolutional neural networks.
- Experimental investigation and evaluation of facial gender classification in the wild under the mask-wearing occlusion.
- Convolutional neural networks based-classifiers' saliency visualization for facial mask-wearing prediction and gender classification in controlled and uncontrolled scenarios.

The rest of the present paper is organized as follows: In section 2 we give backgrounds for facial data, gender classification and masked face analysis, in section 3 we present the proposed approach for mask-wearing prediction and adaptive gender classification, used facial datasets are presented in section 4, the obtained results are discussed in section 5, and we terminate with a conclusion in section 6.

2. Backgrounds

Convolutional Neural Networks. Convolutional neural networks (CNNs) are deeper artificial neural networks. Over the last decade, CNNs emerged for pattern analysis, object detection, classification, and localization [28]. Based on convolutional operators, a CNN integrates a set of layers and blocs whose liaisons are performed with connectionism's philosophy. Convolutional, pooling, dropout, and fully connected are the most well-known types of CNN layers [28]. ImageNet classification with deep convolutional neural networks (AlexNet), inception Google net version-3 (Inpcetion-V3), very deep convolutional neural networks (VGG16), deep residual neural network (ResNet) [28], efficient convolutional neural networks for mobile vision applications (MobileNet) [32] and rethinking model scaling for convolutional neural networks (EfficientNet) [33] are popular CNN architectures which were largely exploited in pattern recognition, image classification and artificial vision generally.

In the case of facial data analysis, the state-of-the-are shows that CNNs are capable of learning implicit features hierarchically. Indeed, starting from a low pixel level, random facial segment feature maps are detected in the intermediary hidden layers from each interpretable facial region. Finally, whole facial mask' feature maps are perceivable at the end of trained CNNs. Face detection and localization [36] [37] [6], person identity recognition [8] [9], facial expression recognition [8], gender classification [14] [16] and age estimation [13] are examples of facial data analysis' biometric tasks where CNNs allow observable advancement in term robustness, performances and real-world applicability in comparison with priors existing methods.

Facial gender classification. Gender classification is a biometrics task that aims to predict human gender. This task was studied from various modalities in the practices. However, facial data is marked as the most studied and important. This is due to the fact that a lot of gender-related information is embedded in the human face. Getting a two-dimensional facial image as input, gender classification is generally realized in three steps. Namely, the preprocessing localizing the face in a bounding box, the feature extraction computing a discriminative vector, and binary classification for final decision-making.

Classical machine-learning techniques and image-processing operators were among the first methods to be used in automatic human gender prediction from faces. Various techniques were exploited for the key steps of feature extraction and classification. Artificial neural networks (ANN), support vector machine (SVM), Local binary pattern (LBP), boosting-adaptative algorithms (Adaboost), principal component analysis (PCA), scale-invariant feature transform (SIFT), histogram of oriented gradients (HOG), and local directional pattern (LDP) are examples of employed methods [3] [26]. The validation over smaller datasets and the targeting of controlled non-real-world scenarios are notable limitations for classical solutions. It restricts their real-life applicability and generalization ability in front of challenging data.

Recent approaches are almost based on deep learning methods. Indeed, emerged since 2015 for facial data analysis, convolutional neural network-based methods are becoming the stat-of-the-art for facial gender classification [16] [34] [21] [31] [30] [29] [25]. Convolutional neural networks can surpass-greatly the existing old techniques for gender classification [25]. They ensure best performances on larger test sets designated for real-world applications [31] [30] [29]. In addition, with fewer preprocessing operations and by taking a detected region of interest as input, convolutional neural networks can encompass both feature extraction and classification steps [28]. A well-designed and trained deep convolutional neural network can return promising results.

Facial data occlusion. Data occlusion is one of the most well-known and common challenges for pattern recognition and image processing [4] [35]. It is caused by the way when the region of interest (ROI) is superposed by another non-objective data. In facial data analysis, occlusion impact was studied for various tasks of localization, individual identification, and categorization [16] [22] [11] [13]. Depending on their period time of figuratively, facial occlusions can be qualified as permanent or temporary. At the same, depending on the capture conditions, it can be qualified as controlled or uncontrolled. Covering object or hand, mustache-presence, chin-presence, hairstyle, scarf-wearing, hat-wearing, and black-sunglass-wearing are examples of facial occlusion (Fig. 2) [7] [23]. FERET [6], FEI [27] and FRAV2D [42] are lab-collected datasets containing controlled occlusions and face poses with acceptable image quality. LFW [17] and Images of Groups [3] are challenging real-world datasets containing images with random occlusions.

Figure 2. Examples of random facial-occlusion in the wild (LFW dataset).

It was shown that data occlusion can decrease remarkably facial data analysis performances [8] [10]. Classical solutions remain more vulnerable in front of uncontrolled data occlusion. Recent machine learning-based solutions, especially deep learning techniques, are more solid in front of occlusions [11]. Geometric approaches dividing the input face into multiple sub-regions such as mouth, eyes, and nose were exploited to limit the facial variations [15] [16]. The occlusion problem can also be confronted by integrating facial data including random occlusion related to uncontrolled context [22]. Data augmentation of the training batch was also exploited to generate artificial facial occlusions like glass and hat-wearing [9]. Its goal is to force the trained networks to be more robust under real occlusion. It was empirically demonstrated that data augmentation improves significantly facial data analysis performances [9].

For the gender attribute, occlusion represents one of the most well-known obstacles that the impact still challenging in the literature [3] [16] [30]. Occlusion-robust system is proposed in [10]. The authors applied 2D-PCA on Gabor filters calculated for local facial blocs to extract features. SVM was used for the final classification. Natural and artificial occlusions generated by blacking local facial regions were exploited during the test. The best accuracy of 90.1% returned on occluded FERET faces. DeepGender, an approach for facial gender classification under occlusion and low-resolution, is proposed in [14]. In the goal to force CNN to be more watchful for the higher-profile region, progressive learning of an AlexNet-based classifier was realized. Three random artificial occlusions were applied, namely missing pixel, additive Gaussian noise, and contiguous occlusions. Performance degradations caused by occlusions can be remarkably observed in experimental results. Top accuracy of 93.12% and 97.95% were respectively obtained for occluded and less-challenging data. In [16], CNNs and Adaboost-classifiers were exploited to learn gender features for five sub-facial images. Competitive results were achieved by testing the approach on four challenging datasets rich in variations including occlusions. In the goal to confront the low resolution and partial occlusions challenges, authors analyzed in [30] the effect of using five separate facial components to train CNNs for gender classification. Promising results were achieved as they obtained an accuracy of 98.45 on the LFW real-world dataset. In [12] and [13], occlusion data augmentation's impact on facial gender classification was investigated. It was shown that exploiting data augmentation techniques can help to improve global performances. In [43], an approach using a convolutional neural network is proposed for gender classification from the oral region. Obtained results justify the importance of this facial region lost in the Covid-19. More recently, the ability of deepfake faces to be an alternative to real ones in biometrics tasks of categorization was experimentally investigated by studying the case of gender attribute [42]. It was observed that the training by exploiting the realistic deepfake faces allows getting promising results in various scenarios including random occlusion.

Masked face analysis. Since the Covid-19 pandemic appeared, researchers have been involved in the development of biometrics solutions for human identity analysis under facial mask occlusion. Indeed, diverse computer vision-based tasks are already taken into account in the state-of-the-art such as mask-usage prediction, facial mask localization, and masked face recognition. A simplified deep neural network is proposed in [5] to detect Covid-19 face mask. The network contains two convolutional layers followed by a flattened layer to collapse the spatial dimensions and a dense layer of 64 nodes. Respective mask-wearing detection rates of 95.77% and 94.58% were returned on two different datasets. In [40], authors exploit additive angular margin loss to boost masked face recognition. They modified the ResNet network to output the probability of mask usage. An average accuracy of 99.78% was achieved for mask-usage detection on a dataset containing artificially masked faces using the MaskTheFace tool. An architecture for real-time mask detection is adopted in [44] via the employment of a single shot multi-box detector (SSD). A MobileNet-based network was implemented as a mask detector by fine-tuning the SSD-MobileNetV2 network through the addition of costumed layers. The authors assert that well-performances were achieved after a large number of tests in different scenarios. An attention-enhanced face mask detection framework via an improved Yolo model is proposed in [45]. The framework is designated to detect the face mask in complicated real-world scenarios like small-size object detection and interference of similar occlusions. The model was evaluated using the mean average precision (mAP) metric and respective values of 94.1% and 94.1% were obtained on two different experimental datasets.

3. Methodology

The overview of our methodology is illustrated in the next figure. It consists of three main stages: face detection (a), mask-wearing prediction (b), and gender classification (c). After face detection, facial mask presence is predicted, then depending on the mask-wearing information an adaptive gender classification is performed.

Knowing that each mask-wearing prediction and gender classification generates two exclusive outputs, the exploited classifiers in the proposed approach also are binary. Four possible exclusive outputs are returned by the proposed approach: female not-wearing mask, female wearing mask, male not-wearing mask, and male wearing mask. Let a facial dataset containing a portion p of masked faces and let T_0, T_{11}, T_{12}, T_{21} and T_{22} respective accuracy for mask-wearing predictor, masked face gender classifier in masked context, masked face gender classifier in unmasked context, unmasked face gender classifier in masked context and unmasked face gender classifier in unmasked context. The theoretical accuracy T for the hybrid proposed gender classifier is:

$$T = pT_0T_{11} + p(1-T_0)T_{12} + (1-p)(1-T_0)T_{21} + (1-p)T_0T_{22} \quad (1)$$

As mush as the mask-wearing is well predicted, T_0 will be close to 1 and the difference 1-T_0 will be negligible approaching the 0. We get the following approximate theoretical accuracy:

$$T \approx pT_{11} + (1-p)T_{22} \quad (2)$$

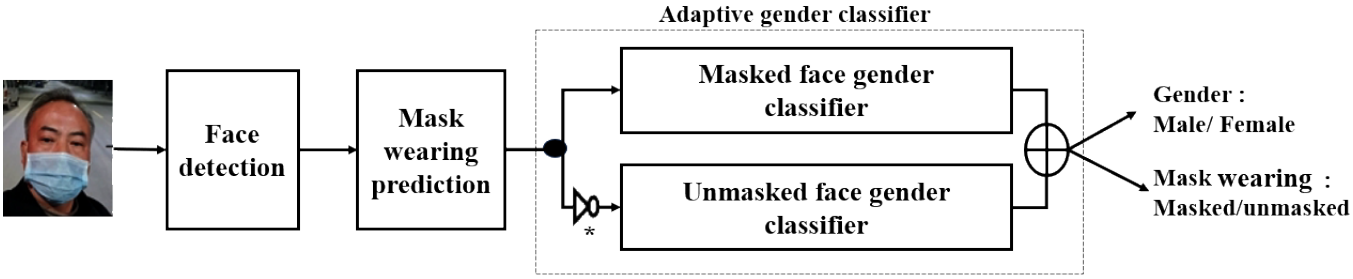

Figure 3. Overview of our proposed approach.
(*): We consider mask presence as positive class, the not gate is activated in the case where the input is predicted as unmasked.

We suppose that training the masked face gender classifier with masked faces allows keeping up the accuracy T_{11} value and kept up the global accuracy T as well.

The details of the proposed approach illustrating each classifier and adaptive one exploiting mask presence information are shown in Fig. 4. Both gender sub-classifiers and mask-wearing predictor were defined by fine-tuning and realizing domain adaptation of well-known convolutional neural networks. Two popular deep architectures were exploited in this work, MobileNet for the mask-wearing prediction and VGG16 for the gender classification task. We kept the convolutional section that we used as a features learner and extractor. In transfer learning, for mask-wearing and gender classifiers' better weights initialization with pre-trained CNNs, ImageNet benchmarked [28] models' hyper-parameters were weighted on the convolutional section layers. In a second classification section, we applied the max-pooling (Maxpool) operator on the last convolutional layer of the last pre-trained section. We employed a dense fully connected layer with 512 nodes. To limit overfitting, the dense layer is followed by a dropout operator randomly omitting 35% of neurons. We applied a dense binary softmax operator for domain adaptation and decision-making.

Face Detection. Object detection is a fundamental preprocessing step for an image processing system. Face detection or localization is one of the most explored and studied object detection tasks. In the proposed approach, face detection is the preprocessing step. In this stage, a facial bounding box containing the region of interest for mask-wearing prediction and gender classification tasks is computed. Almost of recent object detection in image processing and human-face localization especially' stat-of-the-art is founded on deep learning techniques. Indeed, deep convolutional neural networks achieve remarkable progress in real-world face detection in the wild. Existing solutions are principally designed to be used for unmasked contexts with eventual partially controlled and uncontrolled occlusions. R-CNN [36], YOLO [37], CornerNet [38] and RetinaFace [6] are reference existing face detection algorithms.

In this work, we used the RetinFace algorithm. Based on the famous ResNet networks, RetinaFace is a powerful algorithm for face detection. The advantage of this algorithm is its sufficient robustness to the occlusion caused by facial mask-wearing. Indeed, the model was trained by exploiting five facial annotations in the wild, namely: left and right eyes, nose, and left and right mouth commissures. The cited annotations are detectable even if the face is masked, with an irregular pose or very poor quality. Our preliminary experimentation shows that RetinaFace is more robust for masked face localization in comparison with other algorithms. After recuperating the five facial annotations returned by the RetinaFace, we defined a procedure for computing a bounding box including the objective face.

Mask-wearing prediction. The objective of this step is to detect the presence of a facial mask in the detected face from an input image. This is the first sub-decision in our proposed approach. It is also exploited for the next step of gender classification. We suppose that mask presence detection on the human face is not harder in comparison with other biometric tasks of categorization like facial expression recognition, age estimation, gender, and ethnic classification, etc. Indeed, the last tasks are performed on human faces sharing closer shapes and similar textures while predicting mask-wearing consists of implicitly separating the oral region containing mouth and nose from non-facial texture described by the mask image. In the proposed approach, we defined the mask-wearing classifier based on the MobileNet [32] CNN. It is a famous deep architecture that was proposed to be exploited for recognition and classification tasks destined for mobile vision applications. It is a lightweight CNN. To make it computationally intensive, authors introduce depth-wise separable convolutions (dw-Conv). Designed efficiently, the benchmarking of the MobileNet on the standard ImageNet dataset demonstrated its competitivity with other leaders CNNs like GoogleNet

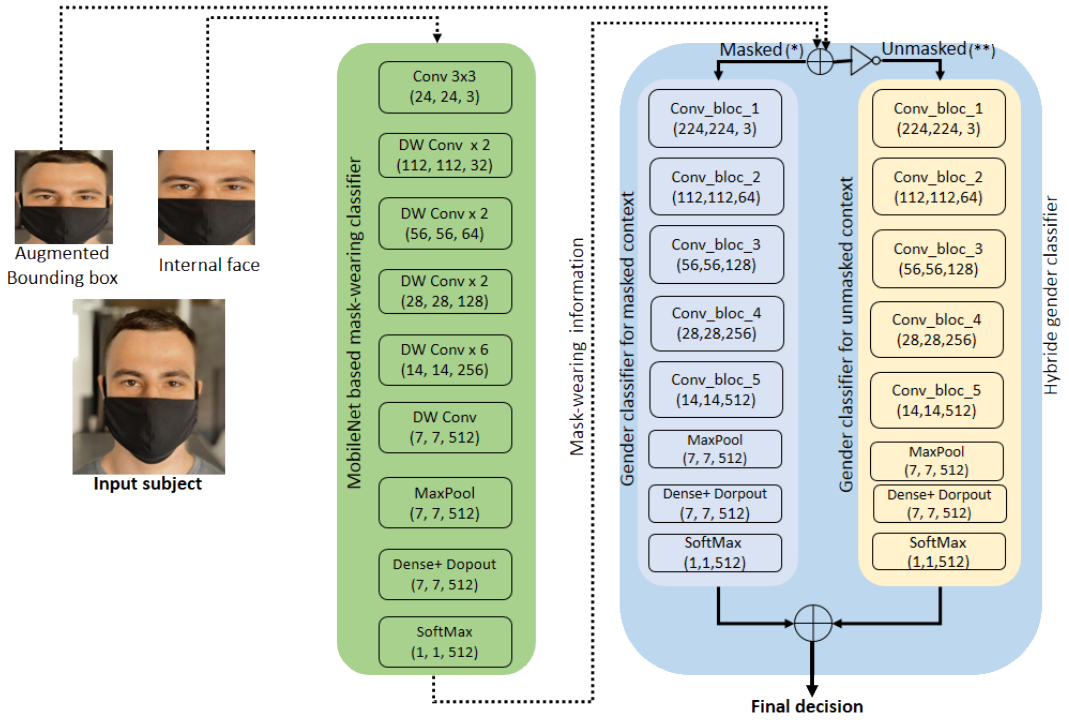

Figure 4. Details of the proposed approach and used classifiers.
(*): The masked faces classifier is used to predict gender only for subjects predicted as wearing mask (not gate disabled). (**): The unmasked faces classifier is used to predict gender only for subjects predicted as no wearing mask (not gate enabled).

and VGG16 for object detection and classification [32]. As input, we used the internal face. The details of the MobileNet-based classifier for mask-wearing prediction are illustrated in Fig. 4.

Gender classification. In this phase, adaptive gender classification is performed. The gender classification module is composed of two gender sub-classifiers. A first classifier is designated for masked faces and a second for unmasked faces. Depending on the mask-wearing information, an appropriate gender sub-classifier will be selected. Gender attribute prediction is the second sub-decision of our approach. Both gender sub-classifiers were trained by using related facial data of masked and unmasked scenarios. They were built inspired by the foremost VGG16 convolutional neural network architecture. Originally, the VGG16 network was made up of convolutional blocs composed principally of convolutional and max-pooling layers. Convolutional blocs are followed at the end of the networks by fully connected layers. Figure 4 illustrates the adopted gender classifiers architecture' details.

4. Databases

Talking the masked context, it should be noted that almost of available facial datasets are designated to the unmasked scenario. However, with the apparition of the Coronavirus pandemic, related datasets become available [19] [40] [39] for eventual use in masked face detection, recognition, and categorization. In this work, we exploited both masked and unmasked facial datasets for the implementation of our approach, namely: FERET, FRAV2D, LFW, Kaggle, and FFHQ. We short-describe each dataset below:

- **FERET dataset:** FERET (Facial Recognition Technology) [3] is an old and well-known dataset that was largely used in facial data analysis. It was created in 15 sessions during 3 years under controlled conditions. It contains 14126 facial images for 1199 subjects. FERET dataset images are of good quality and it is rich in variations like face pose, facial expression, illumination, age, and ethnicity. The dataset is gender-unbalanced and unlabeled. The dataset is delivered for free under engagement license .

- **FEI dataset**: Faculdade de Engenharia Industrial or FEI faces is a gender-balanced facial dataset. It was collected under controlled conditions like facial expression, face pose, image quality, and background. It is a multi-instance facial dataset where 14 acquisitions were established for each person from 200 volunteers of the white race with ages varying between 14 to 40 years. FEI is a publicly available dataset [27].

Table 1. Details of used dataset.

Dataset	Faces Number	Subjects number	Occlusion type	Gender-labeling	Facial mask occlusion
FERET	~14k	1199	Controlled	Labeled	Free
FRAV2D	3488	109	Controlled : semi- occlusion by covering hand	Unlabeled	Free
FEI	2800	200	Free	Unlabelad	Free
LFW	13233	5749	Random	Labeled	Free
FFHQ	~69k	~69k	Random	Labeled	Almost free
100k-generated	100K	-	Random	Unlabeled	Free
FFHQ-Masked	~69k	~69k	Artificial mask combined with random real occlusions	Labeled	Artificially masked
Kaggle	~5.23k	~5.23k	Real facial mask combined with random real-occlusions	Unlabeled	Mix

- **FRAV2D dataset:** The FRAV2D dataset is a facial dataset collected by experts under controlled conditions in the FRAV (Face recognition and Artificial Vision) laboratory. For 109 voluntaries (75 male and 34 female), 32 controlled facial images were acquired corresponding to multiples facial variations including facial expression, lighting conditions, background complexity, face pose, and covered hand occlusion. The FRAV2D is a free dataset delivered upon request for research purposes [42].

- **LFW dataset:** Labeled Face in the Wild is a real-world dataset. Used originally for face recognition and identification, the LFW dataset was formed by collecting faces with multi-instances, for more than 5470 persons, acquired in the wild. It was also exploited for categorization tasks, especially of the gender attribute which is marked as one of the widely used datasets. It contains around 13230 facial images with an uncontrolled pose, poor quality, random partial occlusions, and various ethnicities. Its gender labeling was recently proposed by Affifi [16].

- **FFHQ and 100k-generated datasets :** The FFHQ dataset is a recent facial database in comparison with other existing datasets. Collected by Karas et al. from the Filter website, this database was initially used for the training of the StyleGAN model [24]. One of the most well-known generative adversarial networks allowing the creation of extremely realistic artificial faces. It contains around 69K images. Originally, the FFHQ dataset was gender-unbalanced. The images are of high resolution. Identities are from various ethnicities, and ages, with random facial expression and occlusion, frontal and semi-profile face poses [24]. We also used the artificial 100k-generated dataset composed of StyleGAN-generated faces and an synthetic-mask' augmented version of the FFHQ dataset (FFHQ-Masked) [19].

- **Kaggle Dataset:** It is a publicly available dataset on the Kaggle framework that makes various challenging competitions for image classification. It is a real-world database composed of recent multi-faces images captured in the wild. Collected originally to be exploited for facial-mask detection, the first part of this database is formed with masked faces where people are wearing real masks. The second part is formed by facial images without a mask. The images of the Kaggle dataset are of random quality, face pose, various races, and age.

Table 01 resumes some details of used datasets by making, for each dataset, the size in number of faces and subjects, occlusion type, and eventual gender-labeling. We made an effort to realize gender labeling for unlabeled datasets, especially, Kaggle and 100k-generated ones. Fig. 5 shows samples of presented datasets.

5. Experiments, Results Discussion and Saliency Visualization

In this section, the obtained experimental results are presented and discussed for both mask-wearing prediction and gender classification tasks. The following scenarios are taken into consideration: unmasked context

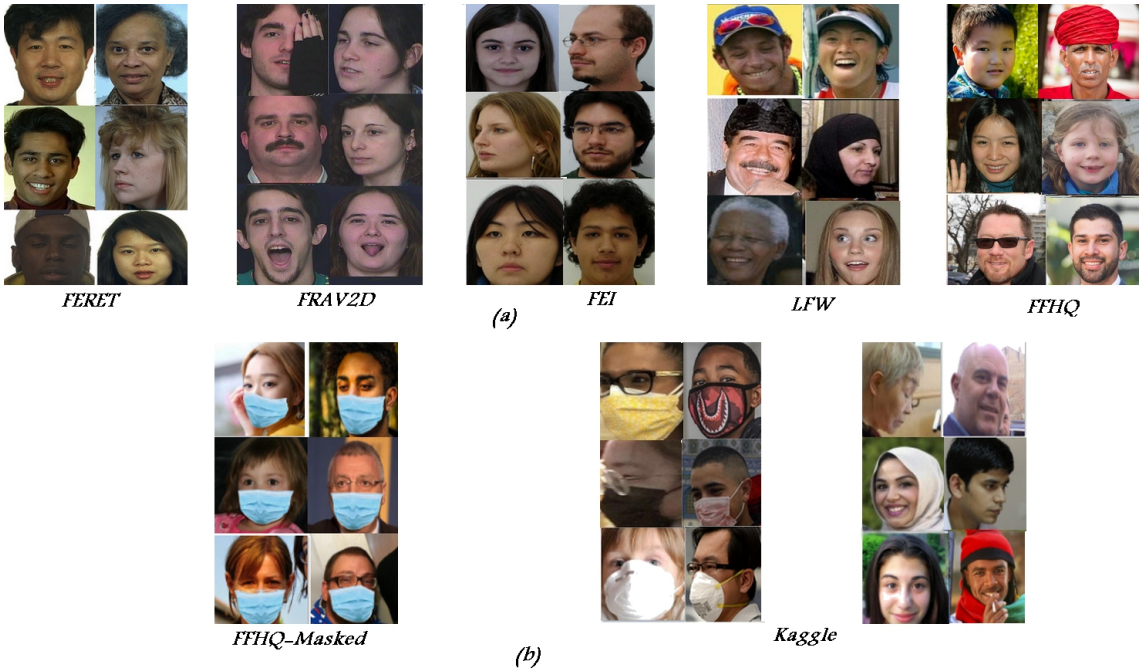

Figure 5. Samples of presented datasets.
(a): Unmaksed datasets. (b): Masked and mix datasets.

corresponding to the case where all subjects do not wear a facial mask; fully masked context corresponding to the case where all subjects wear a facial mask and mixed context combining both previous situations. Finally, we discuss saliency visualisation for both investigated tasks.

5.1. Mask-wearing prediction

A balanced set of 40k was used to realize the training of the mask-wearing classifier. Used faces are derived from the FFHQ and FFHQ-Masked dataset. The following table illustrates the prediction accuracy for mask-wearing on various facial datasets. The mask-wearing classifier was tested on the FRAV2D, LFW, FERET, FEI, and FFHQ datasets related to unmasked contexts. For a fully masked context, we used the FFHQ-Masked dataset composed of facial images with artificial masks. For mix context, we used the Kaggle dataset corresponding to mix context (facial images with and without mask).

As shown in Table 2, the mask presence can be predicted with very-well accuracy. Indeed, perfect accuracy was achieved for the FRAV2D, FEI, FERET, FFHQ, and FFHQ-Masked datasets containing images of acceptable quality and controlled face pose. For the real-world challenging datasets containing faces acquired in the wild, LFW, and Kaggle, respective accuracies of 99.65% and 99.5% were achieved. The obtained results validate the choice of using mask-wearing prediction as a pre-step preceding the adaptive gender classification of our approach. The appropriate gender sub-classifier can be selected quasi-perfectly.

Table 2. Obtained accuracy for mask-wearing prediction.

Dataset	Scenario	Accuracy
FRAV2D	Unmasked	100%
LFW	Unmasked	99.65%
FEI	Unmasked	100%
FERET	Unmasked	100%
FFHQ	Unmasked	100%
FFHQ-Masked	Artificial masks	100%
Kaggle	mix	99.5%

5.2. Gender classification

For the unmasked context, we used a balanced set composed of FFHQ and 100k-generated dataset faces to perform the training of the network. To evaluate the trained network, we exploited FEI, FERET, FRAV2D, FFHQ, LFW, and the Kaggle-Unmasked subset. We also used the Kaggle-Masked subset, collected by selecting faces with masks from the Kaggle database, to assess the performance of the masked data trained VGG16-based classifier in front of facial-mask occluded data. The details of used datasets for training steps in the unmasked scenario are resumed in the following Table 3.

Table 3. Training dataset for the unmasked scenario.

Training dataset	Female faces	Male faces
FFHQ	25k	25k
100k-generated	15k	15k

Obtained accuracies are recapitulated as shown in Table 4. Well results were acquired for facial gender classification in an unmasked context related to diverse facial variations. Indeed, state-of-the-art' accuracy of 99.04%, 99.02%, 98.32%, 97.79%, and 99.10% was achieved for the FEI, FERET, FRAV2D, LFW, and FFHQ datasets respectively. However, as can be observed in the last line of Table 5, the unmasked data-trained gender classifier's performance decreases remarkably in front of the masked Kaggle subset containing images with real masks of faces captured in the wild. Indeed, a lower accuracy of 91.20% was returned on the kaggle-masked subset. To affirm this remark, we trace in Fig. 6 ROC curves comparing the classifier's comportment in front of masked and unmasked datasets. We can clearly observe that the classifier converges better and covers a higher area in the case of testing on unmasked data. It also justifies the importance of the retraining of the network with masked faces.

Two datasets were exploited for gender classification performance evaluation in a fully masked context. The FFHQ-Masked, gender-balanced subset of 8k faces was used to assess results in controlled face pose, especially for frontal and semi-profile poses. The Kaggle-Masked dataset was used to evaluate gender classification results in the wild. We performed the network training by retraining the unmasked data pre-trained network with a balanced set of 50k images from the FFHQ-Masked dataset. The classifier also was tested on the Kaggle-Unmasked subset.

Table 4. Obtained accuracy for gender classification in unmasked context.

Dataset	Scenario	Test-set	Accuracy
FEI	Unmasked	Whole	99.04%
FERET	Unmasked	6233	99.02%
FRAV2D	Unmasked	Whole	98.32%
LFW	Unmasked	Whole	97.79%
FFHQ	Unmasked	8000	99.10%
Kaggle-Unmasked	Unmasked	Whole (1544)	97.30%
Kaggle-Masked	Masked	Whole (3690)	91.2%

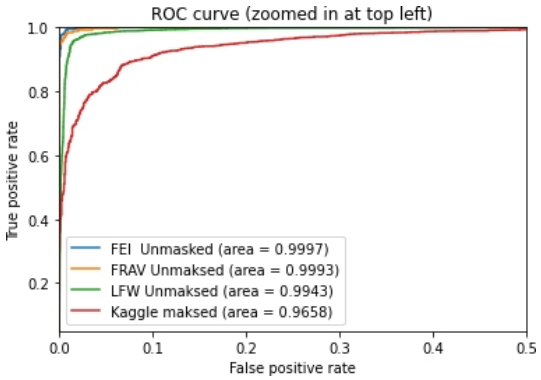

Figure 6. ROC curves for unmasked data trained network.

Table 5. Datasets used for gender classification evaluation in fully masked context.

Dataset	Male	Female	Total
FFHQ-Masked	4k	4k	8k
Kaggle-Masked	1555	2135	3690
Kaggle-Unmasked	830	714	1544

Table 6 recapitulates the obtained accuracy and EER error for presented datasets. Achieved accuracy shows that in controlled face pose and acceptable quality, facial mask wearing does not affect highly the gender classification performance as we obtain an accuracy of 97.20% on the gender-balanced FFHQ-Masked dataset composed of 8k images. At the same, gender classification remains well detectable in the wild as an accuracy of 95.90% was achieved by testing the classifier on the Kaggle-Masked set, composed of 3690 faces with real masks. Furthermore, low EERs (Equal Error Rates) were obtained. The best EER of 0.036 was returned width the FFHQ-Masked. The second EER of 0.053 with the Kaggle-Masked dataset. For Kaggle unmasked' faces, respective acceptable rates of 94.90%, 0.081, and 0.960 were obtained for the accuracy, EER, and AUC metrics. Discussed metrics allow asserting that the gender classification biometric task can be safely carried out from facial data in fully masked contexts.

Table 6. Dataset used for gender classification on fully masked context.

Dataset	Accuracy	EER	AUC
FFHQ-Masked	97.20%	0.036	0.994
Kaggle-Masked	95.90%	0.053	0.983
Kaggle-Unmasked	94.90%	0.081	0.960

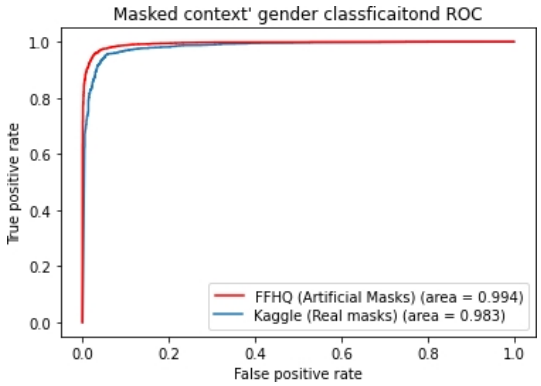

Figure 7. Fully-masked context gender classification ROCs curves.

In addition to the viewed metrics, we traced the ROC curves to interpret graphically the comportment of the gender classifier in a fully masked context. As illustrated in Fig. 7, we can observe that for both tested datasets the classifier converges normally like a good discriminator. However, it converges more rapidly in the less challenging context described by the FFHQ-Masked dataset. A larger area is covered (AUC= 0.994) in comparison with the challenging context described by the Kaggle-Masked dataset (AUC= 0.983).

After evaluating each gender sub-classifier in the targeted contexts, we finally performed the evaluation of the gender module in the mix and general case by exploiting mask-wearing information. We created an FFHQ-mix by combining the FFHQ and FFHQ-Masked subsets, used in the previously discussed scenarios of fully masked and unmasked context. We also used the whole Kaggle mix real-world dataset. As presented in the overview of the proposed approach, the mask-wearing information was predicted at first for the detected faces from the input image. The appropriate gender sub-classifier was selected based on the mask-wearing information. The details of the used mix test datasets are resumed in the next Table 7. The hybrid classifier also was evaluated on other datasets. Namely: FEI, FERET, FRAV2D, and LFW.

Table 7. Mix datasets used for the hybrid classifier' evaluation.

Dataset	Total subjects
FFHQ-Mix	16k
Kaggle	5232

Table 8 resumes the final gender attribute prediction results for the hybrid classifier by exploiting the mask-wearing information. For the FFHQ-mix gender-balanced dataset, an overall accuracy of 98.15% was obtained. For the challenging real-world Kaggle mix dataset composed of human faces captured in the wild

Table 8. Obtained accuracy for the hybrid gender classifier.

Dataset	Accuracy	Observation
FFHQ-Mix	98.15 %	mix
Kaggle	96.35%	mix, real-world
LFW	97.68%	Unmasked, real-world
FEI	99.04%	Unmasked
FERET	99.02%	Unmasked
FRAV2D	98.32%	Unmasked

where a portion of 0.71 wear facial masks, a global accuracy of 96.35% was achieved. An accuracy of 97.68% was returned by testing the hybrid classifier on the whole real-world LFW dataset containing images of poor quality and uncontrolled face pose for which the mask-wearing information was predicted with an order of 99.65%. For the rest unmasked less-challenging FEI, FERET, and FRAV2D datasets, knowing that we got a perfect accuracy for mask-wearing prediction, the unmasked context gender classification results were kept.

Table 9. Comparison with the SOTA approaches for gender classification in the wild.

Approach	Principle	Dataset	Result	Context
FairFace [45]	CNN	LFW (whole)	92.12%	Unmasked
FaceHop [18]	CNNs	LFW (2647)	94.63%	Unmasked
HyperFace [31]	CNN	LFW (whole)	94.00%	Unmasked
Afifi [16]	CNNs+ Adaboost	LFW (10283)	95.98%	Unmasked
VEGAC [34]	CNN	LFW (3739)	96.80%	Unmasked
Mohamed et al. [42]	CNN	LFW (13164)	96.97%	Unmasked
PANDA [29]	CNNs+ SVM	LFW (-)	99.00%	Unmasked
Proposed	**CNNs**	**LFW Kaggle** (whole)	97.68% 96.35%	**Unmasked Mix**

Finally, we performed a baseline comparison to highlight the contribution of our approach. Table 9 resumes a comparison with existing facial gender classification approaches designated for real-life application. The employed techniques, the used datasets, the best-attained accuracy on the largest test set, and the involved context aspects are underlined. By looking at the last table, we can acknowledge the competitiveness

of the obtained results with those achieved in the state-of-the-art (SOTA). In addition, unlike almost all existing works, the proposed approach presents the benefit of considering the masked context in the wild.

5.3. Saliency visualization

For a more interpretable study and to highlight the most important facial regions exploited by the MobileNet-based classifier for mask-wearing prediction making, we targeted the last convolutional layer of the network to compute the heatmaps for feature visualization using Grad-Cam technique [20]. We calculated a mask image resuming heatmaps for the facial image. The computed heatmaps highlight contributing features.

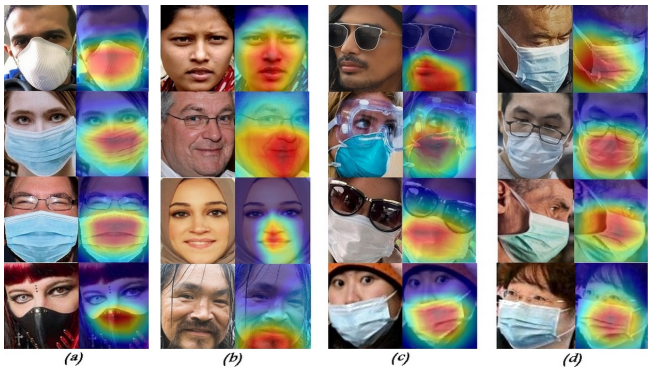

Figure 8. Mask-wearing classifier features visualization by computing the Grad-Cam heatmap.
(a): Masked scenarios (b): Unmasked scenarios (c): Combined occlusions (d): Irregular pose.

As shown in Fig. 8, by superposing the mask-wearing network' heatmaps mask on the input image we bring out the most pertinent and discriminative regions used by the CNN. The visualized features clearly illustrate the watchfulness of the trained CNN on the lower internal face containing the oral region for the mask-wearing prediction. The mask-wearing predictor focused on the lower face for both masked and unmasked cases even if the objective face contains additional occlusions or with an irregular pose.

At the same, to explore what the masked data trained gender sub-classifier is looking to make a decision for the gender attribute prediction for persons wearing a mask, we used the Grad-Cam technique on several subjects.

As illustrated in Fig. 9, we applied the Grad-Cam technique for both female and male gender class subjects. We observed that for the female gender and independently of the face pose, the classifier is more focused on the hair and forehead region, with a potential watch to other available facial parts. For the male gender, we noticed that the network watchfulness is mainly concentrated on the upper face including

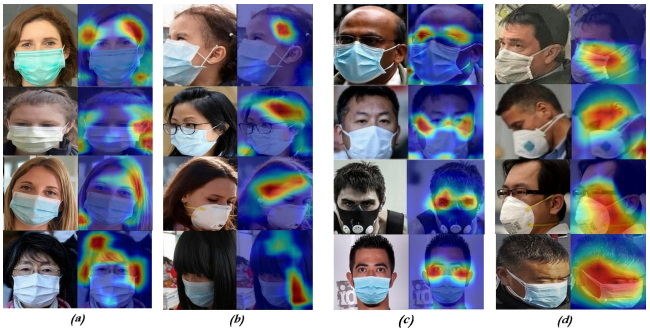

Figure 9. Gender classification under mask-wearing' feature visualization by computing the Grad-Cam' heatmap. (a): Female faces with regular pose (b): Female faces with irregular pose (c): Female faces with occluded hair (f): Male faces with frontal pose (d): Male faces with irregular pose.

primarily the periocular region when the objective face is acquired in a frontal regular pose. In the case when the face presents a profile or irregular pose, we observed that the network pays attention to the residual profile region containing the ear. Furthermore, unlike the mask-wearing classifier, we have noticed the forgetting of the masked lower face by the classier for both genders.

6. Conclusion

In this paper, we proposed an approach for facial mask-wearing prediction and hybrid human-gender classification using deep learning. For the mask-wearing prediction, we used a MobileNet-based classifier. Depending on the mask-usage information, we performed an adaptive gender classification by selecting an appropriate gender sub-classifier based on the VGG16 network. We trained each gender sub-classifier with data related to its targeted context. Experimentation shows that the mask-wearing information can be quasi-perfectly predicted in the wild by using CNNs. Using the Grad-Cam technique, we observed the watchfulness of the CNN to the oral region for mask-wearing prediction decision-making. For the gender attribute, we got a state-of-the-art on the unmasked datasets, and we empirically proved that the human gender remains well detectable from facial modality by training the network with masked faces, even if the person wears a facial mask. We also showed that the adaptive approach kept up the gender classification results. Computed Grad-Cam heatmaps for masked faces showed that the gender sub-classier paid attention to the residual facial parts like as periocular region, forehead, and hair.

References

[1] Bhattacharya, S., Maddikunta, P.K.R., Pham, Q., Gadekallu, T.R., Krishnan, S., Chiranji Lal Chowdhary, C.L., Alazab, M.and Piran, M. (2021) Deep learning and

medical image processing for coronavirus (COVID-19) pandemic: A survey. Sustainable Cities and Society 65, doi: 10.1016/j.scs.2020.102589.
[2] Wang, L., Zhong Qiu Lin. Z.Q. Wong, A. (2020) COVID-Net: a tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images. Sci Rep 10, doi: 10.1038/s41598-020-76550-z.
[3] Ng, CB., Tay, YH. and Goi, BM. (2015) A review of facial gender recognition. Pattern Anal Applic 18: 739–755, doi: 10.1007/s10044-015-0499-6.
[4] Benenson, R. (2014). Occlusion Detection. In: Ikeuchi, K. (eds) Computer Vision. Springer, doi: 10.1007/978-0-387-31439-6_135.
[5] Das, A., Ansari, W. and Basak, R. (2020) Covid-19 Face Mask Detection Using TensorFlow, Keras and OpenCV. IEEE 17th India Council International Conference (INDICON): 1-5, doi: 10.1109/INDICON49873.2020.9342585.
[6] Deng, J., Guo, J., Zhou, Y. Yu, J., Kotsia, I., and Zafeiriou, S. (2020) RetinaFace: Single-stage Dense Face Localisation in the Wild. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR): 5202-5211, doi: 10.1109/CVPR42600.2020.00525.
[7] Zhang, L., Verma, B., Tjondronegoro, D. and Chandran, V. (2018) Facial Expression Analysis under Partial Occlusion: A Survey. ACM Comput. Surv. 51 (2), doi: 10.1145/3158369.
[8] Ghazi, M. and Ekenel, K. (2016) A comprehensive analysis of deep learning based representation for face recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW): 102-109, doi: 10.1109/CVPRW.2016.20.
[9] Trigueros, D., Meng, L. and Hartnett, M. (2018) Enhancing convolutional neural networks for face recognition with occlusion maps and batch triplet loss. Image and Vision Computing 79: 99–108, doi: 10.1016/j.imavis.2018.09.011.
[10] Rai, P. and Khanna, P. (2014) A gender classification system robust to occlusion using Gabor features based (2D)2PCA, J. Vis. Commun. Image R. 25: 1118–1129, doi: 10.1016/j.jvcir.2014.03.009
[11] Wu, G., Tao, J., and Xu, X. (2019) Occluded Face Recognition Based on the Deep Learning. The 31th Chinese Control and Decision Conference, Nanchang, China: 793-797, doi: 10.1109/CCDC.2019.8832330.
[12] Lin, L.E. and Lin C.H. (2021) Data augmentation with occluded facial features for age and gender estimation. IET Biometrics, doi: 10.1049/bme2.12030
[13] Hsu, CY., Lin, LE. and Lin, C.H. (2021) Age and gender recognition with random occluded data augmentation on facial images. Multimed Tools Appl 80: 11631–11653, doi: 10.1007/s11042-020-10141-y.
[14] Juefei-Xu, F., Verma, E., Goel, P., Cherodian, A., and Savvides, M. (2016) DeepGender: Occlusion and Low Resolution Robust Facial Gender Classification via Progressively Trained Convolutional Neural Networks with Attention. IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW): 136-145, doi: 10.1109/CVPRW.2016.24.
[15] Li, Y., Zeng, J., Shan, S. and Chen, X. (2019) Occlusion Aware Facial Expression Recognition Using CNN With Attention Mechanism. IEEE Transactions on Image Processing 28: 2439-2450, doi: 10.1109/TIP.2018.2886767.
[16] Afifi, M. and Abdelhamed, A. (2019) AFIF4: Deep gender classification based on AdaBoost-based fusion of isolated facial features and foggy faces. J. Vis. Commun. Image R.62: 77-86, doi: 10.1016/j.jvcir.2019.05.001.
[17] Learned-Miller, E., Huang, G.B., RoyChowdhury, A., Li, H. and Hua, G. (2016) Labeled Faces in the Wild: A Survey. In Advances in Face Detection and Facial Image Analysis, Springer: 189-248, doi: 10.1007/978-3-319-25958-1_8.
[18] Rouhsedaghat, M., Wang, Y., Ge, X., Hu, Sh., You, S. and Kuo, C.J. (2021) Face-Hop: A light-weight low-resolution face gender classification method. In Proc. Int. Workshops Challenges, Springer: 169–183, doi: 10.1007/978-3-030-68793-9_12.
[19] Cabani, A., Hammoudi, K., Benhabiles, H. and Melkemi, M. (2021) MaskedFace-Net – A dataset of correctly/incorrectly masked face images in the context of COVID-19. Smart Health 19, 10.1016/j.smhl.2020.100144.
[20] Selvaraju, R R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D. (2017) Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization. IEEE International Conference on Computer Vision (ICCV): 618-626, doi: 10.1109/ICCV.2017.74.
[21] Jia, S., Lansdall-Welfare, T. and Cristianini, N. (2016) Gender Classification by Deep Learning on Millions of Weakly Labelled Images. IEEE 16th International Conference on Data Mining Workshops (ICDMW): 462-467, doi: 10.1109/ICDMW.2016.0072.
[22] Song, L., Gong, D., Li, Z., Liu, C. and Liu, W. (2019) Occlusion Robust Face Recognition Based on Mask Learning With Pairwise Differential Siamese Network. IEEE/CVF International Conference on Computer Vision (ICCV): 773-782, doi: 10.1109/ICCV.2019.00086.
[23] Zeng, D., Veldhuis, R., and Spreeuwers, L. (2021) A survey of face recognition techniques under occlusion. IET Biometrics, doi: 10.1049/bme2.12029.
[24] Karras, T., Laine, S. and Aila, T. (2019) A Style-Based Generator Architecture for Generative Adversarial Networks. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR): 4396-4405, doi: 10.1109/CVPR.2019.00453.
[25] Levi, G. and Hassncer T. (2015) Age and gender classification using convolutional neural networks. IEEE Conference on Computer Vision and Pattern Recognition Workshops, :34-42, doi: 10.1109/CVPRW.2015.7301352.
[26] Annalakshmi, M., Roomi, S.M.M. and Naveedh, A.S. (2019) A hybrid technique for gender classification with SLBP and HOG features. Cluster Comput 22 (Suppl 1): 11–20, doi: 10.1007/s10586-017-1585-x.
[27] [Online] FEI, Centro Universitario da FEI, FEI Face Database. Available online: fei.edu.br/ cet/facedatabase
[28] Alzubaidi, L., Zhang, J., Humaidi, A.J. et al. (2021) Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. J Big Data 8, doi: 10.1186/s40537-021-00444-8.
[29] Zhang, N., Paluri, M., Ranzato M.A. Darrell T., Bourdev L. (2014) Panda: Pose aligned networks for deep attribute modeling. EEE Conference on Computer Vision and Pattern Recognition: 1637-1644, doi:

10.1109/CVPR.2014.212.
[30] Lee, B., Gilani, S.Z., Hassan, G.M. and Mian, A. (2019) Facial Gender Classification — Analysis using Convolutional Neural Networks. Digital Image Computing: Techniques and Applications (DICTA): 1-8, doi: 10.1109/DICTA47822.2019.8946109.
[31] Rajeev R., Vishal M.P. and Rama C. (2019) HyperFace: A deep multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition. IEEE Trans. Pattern. Anal. Mach. Intell.41 (1): 121-135, doi: 10.1109/TPAMI.2017.2781233.
[32] Howard, A.G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., and Adam, H. (2017) MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. ArXiv preprint, arXiv:1704.04861.
[33] Tan, M and Le, Q V. (2019) EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. ArXiv preprint, arXiv:1905.11946.
[34] Gurnani, A., Gajjar, V., Mavani, V.and Khandhediya, Y. (2018) VEGAC: Visual Saliency-based Age, Gender, and Facial Expression Classification Using Convolutional Neural Networks. ArXiv preprint, arXiv:1803.05719.
[35] Dong, X., Shen, J., Yu,D., Wang, W., Liu, J. and Huang, H. (2017) Occlusion-Aware Real-Time Object Tracking. IEEE Transactions on Multimedia 19 (4): 763-771, doi: 10.1109/TMM.2016.2631884.
[36] Girshick, R. (2015) Fast R-CNN, Proceedings of the IEEE International Conference on Computer Vision (ICCV): 1440-1448, doi: 10.1109/ICCV.2015.169.
[37] Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016) You Only Look Once: Unified, Real-Time Object Detection. IEEE Conference on Computer Vision and Pattern Recognition (CVPR): 779-788, doi: 10.1109/CVPR.2016.91.
[38] Law, H. and Deng, J. (2020) CornerNet: Detecting Objects as Paired Keypoints. Int J Comput Vis 128: 642–656, doi: 10.1007/s11263-019-01204-1
[39] Wang, Z., Wang, G., Huang, B., Xiong, Z., Hong, Q., Wu, H., Yi, P., Jiang, K., Wang, N., Pei, Y., Chen, H., Miao, Y., Huang, Z., Liang, J. (2020) Masked face recognition dataset and application. ArXiv preprint, arXiv:2003.09093.
[40] Montero, D., Nieto, M., Leskovsky, P. and Aginako, N. (2021) Boosting Masked Face Recognition with Multi-Task ArcFace. International Conference on Signal-Image Technology & Internet-Based Systems (SITIS): 184-189, doi: 10.1109/SITIS57111.2022.00042.
[41] Vu, H.N., Nguyen, M.H. and Pham, C. (2022) Masked face recognition with convolutional neural networks and local binary patterns. Appl Intell 52: 5497–5512, doi: 10.1007/s10489-021-02728-1.
[42] Oulad-Kaddour, M., Haddadou, H., Conde, C., Palacios-Alonso, D., Benatchba, K. and Cabello, E. (2023) Deep Learning-Based Gender Classification by Training With Fake Data. IEEE Access 11: 120766-120779, doi: 10.1109/ACCESS.2023.3328210.
[43] Oulad-Kaddour, M., Haddadou, H., Conde, C., Palacios-Alonso, D., and Cabello, E. (2023) Real-world human gender classification from oral region using convolutional neural netwrok. ADCAIJ, 11(3): 249–261, doi:10.14201/adcaij.27797.
[44] Cheng, Ch. (2022) Real-Time Mask Detection Based on SSD-MobileNetV2. ArXiv preprint, arXiv:2208.13333.
[45] Zhang, H., Tang, J., Wu, P., Li, H., Zeng, N. (2023) A novel attention-based enhancement framework for face mask detection in complicated scenarios. Signal Processing: Image Communication 116, doi: 10.1016/j.image.2023.116985.
[46] Karkkainen, K. and Joo, J. (2021) FairFace: Face attribute dataset for balanced race, gender, and age for bias measurement and mitigation. In Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV): 1547–1557, doi:, doi: 10.1109/WACV48630.2021.00159.