A survey of face recognition techniques under occlusion

Dan Zeng, Raymond Veldhuis and Luuk Spreeuwers
University of Twente, The Netherlands
{d.zeng,r.n.j.veldhuis,l.j.spreeuwers}@utwente.nl

arXiv:2006.11366v1 [cs.CV] 19 Jun 2020

Abstract—The limited capacity to recognize faces under occlusions is a long-standing problem that presents a unique challenge for face recognition systems and even for humans. The problem regarding occlusion is less covered by research when compared to other challenges such as pose variation, different expressions, etc. Nevertheless, occluded face recognition is imperative to exploit the full potential of face recognition for real-world applications. In this paper, we restrict the scope to occluded face recognition. First, we explore what the occlusion problem is and what inherent difficulties can arise. As a part of this review, we introduce face detection under occlusion, a preliminary step in face recognition. Second, we present how existing face recognition methods cope with the occlusion problem and classify them into three categories, which are 1) occlusion robust feature extraction approaches, 2) occlusion aware face recognition approaches, and 3) occlusion recovery based face recognition approaches. Furthermore, we analyze the motivations, innovations, pros and cons, and the performance of representative approaches for comparison. Finally, future challenges and method trends of occluded face recognition are thoroughly discussed.

Index Terms—face recognition under occlusion, occluded face detection, occlusion detection, occlusion recovery, partial face recognition, survey

Fig. 1. Examples of occluded face images from the MAFA dataset.

TABLE I
A CATEGORIZATION OF OCCLUSION CHALLENGES

Occlusion Scenario	Examples
Facial accessories	eyeglasses, sunglasses, scarves, mask, hat, hair
External occlusions	occluded by hands and random objects
Limited field of view	partial faces
Self-occlusions	non-frontal pose
Extreme illumination	part of face highlighted
Artificial Occlusions	random black rectangles
	random white rectangles
	random salt & pepper noise

I. INTRODUCTION

FACE recognition is a computer vision task that has been extensively studied for several decades [9]. Compared with other popular biometrics such as fingerprint, iris, palm, and vein, the face has a significantly better potential to recognize the identity in a non-intrusive manner. Therefore, face recognition is widely used in many application domains such as surveillance, forensics, and border control. With the development of deep learning techniques [47], [68], [86], [87], [127], [131], [132], [145], [149] and the publicly available large-scale face datasets [7], [10], [44], [99], [113], [168], face recognition performance has improved substantially [71], [137]. However, face recognition systems still tend to perform far from satisfactory when encountering challenges such as large-pose variation, varying illumination, low resolution, different facial expressions, and occlusion. Generally, face images stored in a gallery are of high quality and free from the above degradations, while probe faces are suffering from what can be seen as *a missing data problem* due to these challenges. Consequently, fewer facial parts are available for recognition, which induces a mismatch between feature available in probe faces and gallery faces.

Facial occlusion [28], [104] is considered one of the most intractable problems because we do not have prior knowledge about the occluded part, which can be anywhere and of any size or shape in a face image. From a practical point of view, it is not feasible to collect a large-scale training dataset with all possible occlusions in a realistic scenario to use deep learning techniques. Therefore, the problem of face recognition under occlusions remains a challenge. Facial occlusion occurs when the subject wears accessories such as a scarf, a face mask, glasses, a hat, etc., or when random objects are present in front of the face. The recognition accuracy has been compromised in some way because of the higher inter-class similarity and the more considerable intra-class variations caused by occlusion. Facial appearance changes substantially due to occlusion, as illustrated in Fig. 1. We present a categorization of occlusion challenges in different scenarios with their typical occlusion examples (see Table I). Pose variation can partially be seen as a self-occlusion problem caused by a large rotation of the head. The self-occlusion problem due to pose variation is usually dealt with in pose correction and therefore not discussed here.

In most cases, occluded face recognition (OFR) involves querying a gallery consisting of occlusion-free faces using a probe image from an alternative test dataset of occluded faces. Occluded faces rely on either the collection of real occlusions

or synthetic occlusions. We first break down OFR research scenarios in the most obvious way by the pairs of images considered. Fig. 2 offers an illustration of the five categories regarding OFR testing scenarios. More specifically, five widely used testing scenarios for OFR, ranging from most real to least real, are:

- **Real occlusions**: gallery images are mugshots free from occlusion while probe images are faces occluded by realistic images such as sunglasses, or a scarf.
- **Partial faces**: gallery images are mugshots free from occlusion while test face images are partial faces; hence the name partial face recognition is given by researchers.
- **Synthetic occlusions**: gallery images are faces in the wild which are captured from uncontrolled scenarios while probe faces are blocked with synthetic occlusions to simulate real occlusions.
- **Occluding rectangle**: gallery images are occlusion-free mugshots while test face images are occluded with a rectangle such as white and black rectangles.
- **Occluding unrelated images**: gallery images are mugshots free from occlusion while test face images are occluded with unrelated images such as a baboon, or a non-square image.
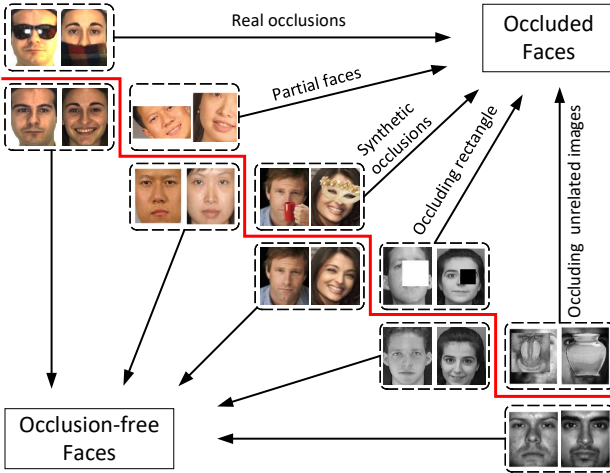

Fig. 2.   Different occluded face recognition testing scenarios involved in OFR.

Approaches to recognizing faces under occlusions can be broadly classified into three categories (shown in Fig. 3), which are 1) occlusion robust feature extraction (ORFE), 2) occlusion aware face recognition (OAFR) and 3) occlusion recovery based face recognition (ORecFR). An OFR system consists of three components, each corresponding to an important design decision: cross-occlusion strategy, feature extraction, and comparison strategy. Of these components, the second and third have analogues in general face recognition, while cross-occlusion strategy is unique to OFR.

- **ORFE category** searches for a feature space that is less affected by facial occlusions. Generally, patch-based engineered and learning-based features are used as the cross-occlusion strategy.
- **OAFR category** is explicitly aware where the occlusion

is. Generally, occlusion-discard is applied as the cross-occlusion strategy. As a result, only visible face parts qualify for face recognition (i.e., feature extraction, feature comparison). Furthermore, we classify partial face recognition approaches as OAFR because they exclude occlusion parts from face recognition, assuming that visible parts are ready in the beginning.
- **ORecFR category** intends to recover an occlusion-free face from the occluded face to meet the demands of conventional face recognition systems. In other words, it takes occlusion recovery as the cross-occlusion strategy.
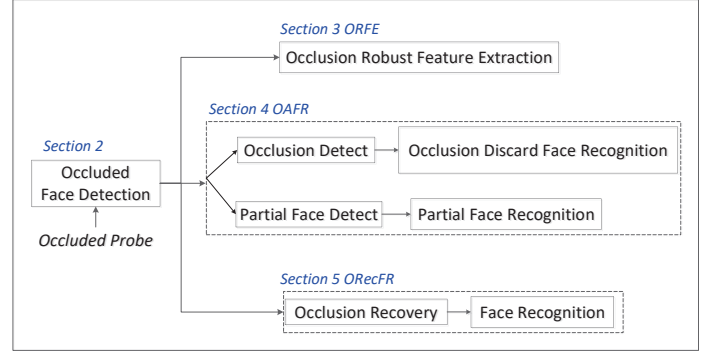

Fig. 3.   The three categories of methods for face recognition under occlusion challenges.

Numerous methods have been proposed to push the frontier of face recognition research. Several comprehensive surveys [5], [22], [70] have been published for face recognition under occlusion. Of the existing works, the survey by Lahasan et al. [70] on face recognition under occlusion challenges presents a thorough overview of the approaches before 2017, which is the most relevant to this paper. However, there are at least two reasons why a new survey on occluded face recognition is needed. *First, the explosive growth of face recognition techniques these years has stimulated many innovative contributions to handle occluded face recognition problems.* The increased number of publications over the last few years calls for a new survey for occluded face recognition, including up-to-date approaches, especially deep learning techniques. *Second, several large-scale occluded face datasets have become publicly available in recent years.* Without large-scale training data of occluded face images, deep learning models cannot function well [97]. Recently, the MAFA dataset [36] is accessible for occluded face detection, and the IJB-C dataset [95] is introduced as a general evaluation benchmark to include meta-information regarding the occlusion (i.e., occlusion location, occlusion degree). Predictably, these datasets would encourage occluded face recognition to develop faster. The proposed survey provides a systematic categorization of methods for face recognition. Specifically, occluded face detection techniques are briefly reviewed because an OFR system requires the application of occluded face detection as the first step. Moreover, newly published and innovative papers addressing occlusion problems are thoroughly reviewed. Finally, we represent comparative performance evaluations in terms of occluded face detection

and face recognition on widely used datasets as well as newly-developed large-scale datasets.

The remainder of the paper is organized as follows: occluded face detection techniques are introduced in Section 2. Methods of occlusion robust feature extraction are described and analyzed in Section 3. We review occlusion-aware face recognition approaches in Section 4. Then Section 5 briefly studies occlusion-recovery face recognition methods. A performance evaluation of the reviewed approaches is given in Section 6. In section 7, we discuss future challenges to datasets as well as to research. Finally, we draw an overall conclusion for occluded face recognition.

II. OCCLUDED FACE DETECTION

In this section, we break down occluded face detection into two parts containing *general face detection* methods which can be applied to detect occluded face, and *occluded face detection* methods which are designed specifically to tackle the occlusion issue in face detection. As for general face detection, we briefly study relevant methods for the simplicity. We then elaborate occluded face detection methods that are specifically designed to detect occluded faces. One way to classify the methods can be seen in Fig. 4.
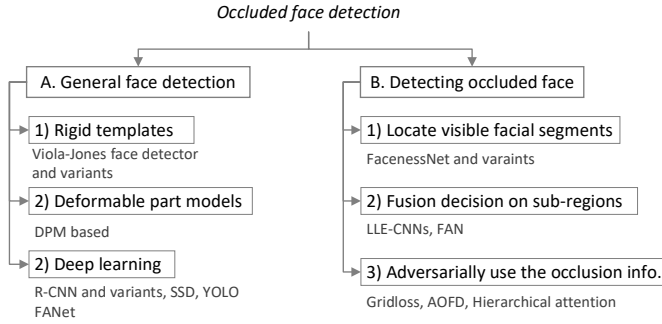

Fig. 4. Methods used in occluded face detection.

A. General face detection

Face detection generally intends to detect the face that is captured in an unconstrained environment. It is challenging due to large-pose variation, varying illumination, low resolution, and occlusion etc. [170]. Approaches to general face detection can roughly be classified into three categories, which are

- Rigid templates based face detection.
- Deformable part models (DPM) based face detection.
- Deep convolutional neural networks (DCNNs) based face detection.

The Viola-Jones face detector and its variations [141], [142] are typical in the rigid templates based category, which utilizes Haar-like features and AdaBoost to train cascaded classifiers and can achieve good performance with real-time efficiency. However, the performance of these methods can drop dramatically when handling real-world application [160]. In contrast, DPM based face detection can achieve significantly better performance based on the cost of high computational complexity [158]. A third, most promising category of research is DCNNs based face detection [6], [13], [63], [75], [79], [129], [166], [172], [176], [179], [182]. Some methods [13], [172], [176] joint face detection with face alignment to exploit their inherent correlation to boost the performance. There are two major branches of object detection framework: (i) region proposals based CNN (i.e.,two-stage detectors), such as R-CNN [39], fast R-CNN [38], faster R-CNN [119]; (ii) region proposals free CNN (i.e.,one-stage detectors), such as the Single-Shot Multibox Detector (SSD) [85], YOLO [117]. In short, two-stage detectors achieve higher performance but are time-consuming. One-stage detectors have significant computational advantages but compensate by less accurate detection results. Some methods [63], [129], [175], [179] treat a face as a natural object and adopt techniques from object detection in face detection. Most recently, finding tiny faces has become popular in face detection and superior performance has been achieved [6], [56], [79], [182].

Recent years have witnessed promising results of exploring DCNNs for face detection with the introduction of Widerface [166], which offers a wide pose variation, significant scale difference (tiny face), expression variation, make-up, severe illumination, and occlusion. Up to now, Feature Agglomeration Networks (FANet) [175], a single-stage face detector, achieves the state-of-art performance on several face detection benchmarks include the FDDB [61], the PASCAL Face [159], and the Widerface benchmark [166]. To exploit inherent multi-scale features of a single convolutional neural network, FANet introduced an Agglomeration Connection module to enhance the context-aware features and augment low-level feature maps with a hierarchical structure so that it can cope with scale variance in face detection effectively. Besides, Hierarchical Loss is proposed to train FANet to become stable and better in an end-to-end way. **In short, methods that achieve remarkable detection performance, for example on the Widerface dataset, also provide a solid solution for occluded face detection.**

B. Detecting occluded face

Detecting partially occluded faces aims to locate the face region in a given image where occlusion is present. Handling occlusion in face detection is challenging due to the unknown location and the type of occlusions [14]. Recently, occluded face detection is beginning to attract the attention of researchers; therefore a few publications are reviewed. At the same time, detecting occluded pedestrians is a long-standing research topic that has been intensively studied during the past few decades. Therefore, many researchers borrow techniques from pedestrian detection [45], [178], [189] to push the frontier of occluded face detection by treating occlusion as the dominating challenge during the detection. **Most occluded face detection methods report their performance on the MAFA dataset [36] while general face detection methods do not, which means it is not a level playing field for general face detection and occluded face detection.** Approaches to detect partially occluded faces are roughly clustered as 1) locating visible facial segments to estimate a full face;

2) fusing the detection results obtained from face sub-regions to mitigate the negative impact of occlusion; 3) using the occlusion information to help face detection in an adversarial way.

1) Locating visible facial segments to estimate face: If visible parts of a face are known, then difficulties in face detection due to occlusions are largely relieved. Observing that facial attributes are closely related to facial parts, the attribute-aware CNNs method [165] intends to exploit the inherent correlation between a facial attribute and visible facial parts. Specifically, it discovers facial part responses and scores these facial parts for face detection by the spatial structure and arrangement. A set of attribute-aware CNNs are trained with specific part-level facial attributes (e.g., mouth attributes such as big lips, open mouth, smiling, wearing lipstick) to generate facial response maps. Next, a scoring mechanism is proposed to compute the degree of face likeliness by analyzing their spatial arrangement. Finally, face classification and bounding box regression are jointly trained with the face proposals, resulting in precise face locations. The results on FDDB [61], PASCAL Face [159] and AFW [192] demonstrate that the proposed method is capable of yielding good performance. In particular, they can achieve a high recall rate of 90.99% on FDDB. In short, Ref. [165] is the first systematic study to attempt face detection with severe occlusion without using realistic occluded faces for training.

More recently, the extension faceness-net [167] improves the robustness of feature representations by involving a more effective design of CNN. As a result, it has achieved compelling results on the Widerface dataset [166], which is challenging in terms of severe occlusion and unconstrained pose variations. However, it requires the use of labeled attributes of facial data to train attribute-aware CNN, which impairs its practical use in some way.

2) Fusing detection results obtained from face sub-regions: In paper [91], a facial segment based face detection technique is proposed for mobile phone authentication with faces captured from the front-facing camera. The detectors are AdaBoost cascade classifiers trained with a local binary pattern (LBP) representation of face images. They train fourteen segments-based face detectors to help cluster segments in order to estimate a full face or partially visible face. As a result, this method could achieve excellent performance on the Active Authentication Dataset (AA-01) [120], [174]. However, the use of simple architecture increases speed by compromising detection accuracy.

The introduction of MAFA [36] offers plenty of faces wearing various masks, which contributes significantly to the occluded face detection, especially of masked faces. Based on this dataset, an LLE-CNNs [36] is proposed to benchmark the performance of masked face detection on the MAFA dataset. They extract candidate face regions with high-dimensional descriptors by pre-trained CNNs and employ locally linear embedding (LLE) to turn them into similarity-based descriptors. Finally, they jointly train the classification and regression tasks with CNNs to identify candidate facial regions and refine their position.

To avoid high false positives due to masks and sunglasses, a face attention network (FAN) detector [146] is proposed to highlight the features from the face region. More specifically, the FAN detector integrates an anchor-level attention mechanism into a single-stage object detector like Feature Pyramid Networks [82]. The attention supervision information is obtained by filling the ground-truth box and is associated with the ground-truth faces which match the anchors at the current layer. The attention maps are first fed into an exponential operation and then combined with feature maps. As a result, the method is capable of achieving impressive results on the Widerface [166] with an 88.5% average precision on the hard subset as well as on an 88.3% average precision on MAFA [36] dataset.

3) Occlusion information adversarially used for detection: Apart from selecting the visible facial parts and fusing results obtained from face sub-regions, it is a third way to minimize the adverse effects of face detection due to occlusions. One promising approach is to use a novel grid loss [106], which has been incorporated into the convolutional neural network to handle partial occlusion in face detection. It is based on the observation that partial occlusions would confuse a subset of detectors, whereas the remaining ones can still make correct predictions. To this end, this work regards occluded face detection as a particular single-class object detection problem, inspired by other works on object detection [30], [52], [75], [126]. Furthermore, the proposed grid loss minimizes the error rate on face sub-blocks independently rather than over the whole face to mitigate the adverse effect of partial occlusions and to observe improved face detection accuracy.

Using the occluded area as an auxiliary rather than a hindrance is a feasible solution to help face detection adversely. Adversarial occlusion-aware face detection (AOFD) [18] is proposed to detect occluded faces and segment the occlusion area simultaneously. They integrate a masking strategy into AOFD to mimic different occlusion situations. More specifically, a mask generator is designed to mask the distinctive part of a face in a training set, forcing the detector to learn what is possibly a face in an adversarial way. Besides, an occlusion segmentation branch is introduced to help detect incomplete faces. The proposed multitask training method showed superior performance on general as well as masked face detection benchmarks. To cope with different poses, scales, illumination, and occlusions, Wu et al. [155] introduce a hierarchical attention mechanism, applying long short-term memory (LSTM) to predict face-specific attention. In this way, it can further model relations between the local parts and adjust their contribution to face detection. The proposed method achieves promising performance compared with Faster R-CNN [119].

III. OCCLUSION ROBUST FEATURE EXTRACTION

If the extracted features are reasonably robust to the occlusion, then difficulties in face recognition due to occlusion are relieved. The aim is to extract features that are less affected by occlusions (outliers) while preserving the discriminative capability. We group the approaches into engineered features and learning-based features. The former generally extract

handcraft features from explicitly defined facial regions, which do not require optimization or a learning stage. The latter extract features by using learning-based methods such as linear subspace methods, sparse representation classification, or nonlinear deep learning techniques. One way to classify the methods can be seen in Fig. 5.
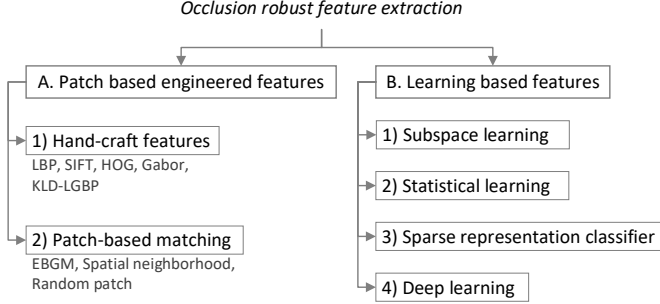
Occlusion robust feature extraction

A. Patch based engineered features
- 1) Hand-craft features
  LBP, SIFT, HOG, Gabor, KLD-LGBP
- 2) Patch-based matching
  EBGM, Spatial neighborhood, Random patch

B. Learning based features
- 1) Subspace learning
- 2) Statistical learning
- 3) Sparse representation classifier
- 4) Deep learning

Fig. 5.   Methods used in occlusion robust feature extraction approaches.

A. Patch-based engineered features

Facial descriptors obtained in an engineered way are rather efficient because they can: (i) be easily extracted from the raw face images; (ii) discriminate different individuals while tolerating large variability in facial appearances to some extent; (iii) lie in a low feature space so as to avoid a computationally expensive classifier. Generally, engineered features (i.e., handcraft features) are extracted from facial patches and then concatenated to represent a face. Therefore, a fusion strategy can be imported to reduce the adverse effects of occluded patches in some way. Alternatively, patch-based matching can be used for feature selection to preserve occlusion-free discriminative information.

These methods in general require precise registration such as alignment based on eye coordinates for frontal faces and integrate the decisions from local patches to obtain a final decision for face recognition. This is problematic because these methods rely on robust face alignment under occlusion, but the eyes are likely to be occluded. In short, these approaches are not realistic for application since most often the face images need to be aligned well to facilitate feature extraction of meaningful facial structure.

1) Handcraft features: Local Binary Patterns (LBP) [1], [2] is used to derive a novel and efficient facial image representation and has been widely used in various applications. LBP and variants [81], [184] retain popularity and succeed so far in producing good results in biometrics, especially face recognition. The main idea is to divide the face image into multiple regions, from which to extract LBP feature distributions independently. These descriptors are then concatenated to form an enhanced global descriptor of the face. For distance measurement between two faces, weighted Chi square distance is applied, accounting for some facial features being more important in human face recognition than others. The Scale Invariant Feature Transform (SIFT) descriptor [88] is popular in object recognition and baseline matching and

can also be applied to face recognition [173]. SIFT is largely invariant to changes in scale, translation, rotation, and is also less affected by illumination changes, affine or 3D projection. Similarly to SIFT, the Histograms of Oriented Gradient (HOG) descriptor [23] has been proposed to handle human detection and has been extended to cope with object detection as well as visual recognition. The main idea is to characterize local object appearance and shape with the distribution of local integrity gradients. After applying a dense (in fact, overlapping) grid of HOG descriptors to the detection window, the descriptors are combined to suit the further classifier. Contrary to the integrity oriented methods, Gabor filters and other frequency oriented approaches construct the face feature from filter responses. Generally, the filter responses computed for various frequencies and orientations from a single or multiple spatial locations are combined to form the Gabor feature [193]. Phase information instead of magnitude information from Gabor features contains discrimination and is thus widely used for recognition [171]. Features based on Gabor filters are versatile. By post-processing they can be converted, for example, to binary descriptors of texture similar to LBPs. Ref. [180] proposes KLD-LGBP, in which Kullback-Leibler Divergence (KLD) is introduced to measure the distance between the local Gabor binary patterns (LGBP) feature [181] of the local region of test images and that of the unoccluded local region of reference faces. They define the probability of occlusions of that area as the distance between two distributions of local regions and further use it as the weight of the local region for the final feature matching. The main drawback of this method is the high dimensionality of LGBP features, which are the combination of Gabor transform, LBP, and a local region histogram on local face regions.

2) Patch-based matching: Beside representation, the distance metric also plays an important role. Elastic and partial matching schemes bring in a lot of flexibility when handling challenges in face recognition. Elastic Bunch Graph Matching (EBGM) [152] uses a graph to represent a face, each node of the graph corresponding to Gabor jets extracted from facial landmarks. The matching method is used to calculate the distance between corresponding representations of two faces. To take advantage of elastic and partial matching, Ref. [57] proposes a robust matching metric to match Difference of Gaussian (DoG) filter descriptor of a facial part against its spatial neighborhood in the other faces and select the minimal distance for face recognition. Specifically, they extract N local descriptors from densely overlapped image patches. During the matching, each descriptor in one face is picked up to match its spatial neighborhood descriptors and then the minimal distance is selected, which is effective in reducing the adverse effects of occlusion. A random sampling patch-based method [12] has been proposed to use all face patches equally to reduce the effects of occlusion. It trains multiple support vector machine (SVM) classifiers with selected patches at random. Finally, the results from each classifier are combined to enhance the recognition accuracy. Similarly to elastic bunch graph matching [152], dynamic graph representation [118] is proposed to build feature graphs based on node representations. Each node corresponds to certain regions of the face

image and the edge between nodes represents the spatial relationship between two regions. Dynamic graph representation can enable elastic and partial matching, where each node in one face image is matched with adjacent nodes in the other face image for face recognition, which brings in robustness to occlusion especially when encountering partially occluded biometrics.

B. Learning-based features

Compared with engineered features, learned features are more flexible when various occlusion types at different locations are present. Features learned from training data can be effective and have potentially high discriminative power for face recognition. Unlike regular images, face images share common constraints, such as containing a smooth surface and regular texture. Face images are in fact confined to a face subspace. Therefore, *subspace learning* methods have been successfully applied to learn a subspace that can preserve variations of face manifolds necessary to discriminate among individuals. Taking the occlusion challenge as a major concern, it is natural to apply *statistical methods* on face patches allowing for the fact that not all types of occlusion have the same probability of occurring. *Sparse representation classifier* methods, which fully explore the discriminative power of sparse representation and represent a face with a combination coefficient of training samples, have been the mainstream approach to handle various challenges in face recognition for a long time. The last few years have witnessed a great success of *deep learning* techniques [94], especially deep convolutional neural networks (DCNN) for uncontrolled face recognition applications.

1) Subspace learning: Approaches such as the Principal component analysis (PCA) [138] and the Linear Discriminant Analysis (LDA) [8] are the two most representative methods in learning the subspace for feature extraction. Eigenface [138] uses PCA to learn the most representative subspace from training face images. Fisherface [8] uses LDA to explore a discriminant subspace that differentiates faces of different individuals by incorporating the class labels of faces images. One limitation for these appearance-based methods is that proper alignment specifically based on the eye location is required. Modular PCA [42] builds multiple eigen spaces (eigenfeatures) in the regions of facial components (e.g., eyes, nose, and mouth) to achieve variance tolerance, especially for face variances. Ref. [24] uses an influence function technique to develop robust subspace learning for a variety of linear models within a continuous optimization framework. Ref. [73] uses the property of PCA to detect "outliers" and occlusions, and calculates the coefficients using the rest of the pixels. Inspired by robust PCA [73], a new approach combining discriminative power and reconstructive property has been proposed by Fidler et al. [31], which can achieve good classification results and have the construction abilities to detect outliers (occlusions). Specifically, object recognition and face recognition applications are used to estimate the robustness of LDA coefficients. Besides, objects orientation application, a regression task, is introduced to estimate the

robustness of extended CCA coefficients. Ref. [65] proposes independent component analysis (ICA) to use locally salient information from important facial parts for face recognition. This part-based representation method could achieve better performance in case of partial occlusions and local distortions than PCA and LDA. Unlike linear subspace methods, nonlinear subspace methods use nonlinear transforms to convert a face image into a discriminative feature vector, which may attain highly accurate recognition in practical scenarios. Ref. [122] proposes kernel correlation feature analysis (KCFA) on facial regions, i.e., eye-region, nose-region, and mouth-region to cope with partial face recognition. The proposed KCFA for feature extraction outperforms the linear approaches PCA [138], LDA [8] and variants of kernel methods [164], [188].

2) Statistical learning: In real-world applications, not all types of occlusions have the same probability of occurring; for example, a scarf and sunglasses often have a higher probability of occurrence compared with others. Hence, it is natural to apply statistical learning on face patches to account for their occlusion possibility. One early work in this direction is Ref. [93], which takes a probabilistic approach, i.e., a mixture of Gaussians to compensate for partially occluded faces. Specifically, they analyze local regions divided from a face in isolation and apply the probabilistic approach to find the best match so that the recognition system is less sensitive to occlusion. Ref. [134] extends the work by representing the face subspace with self-organizing maps (SOM). It presents the similarity relationship of the subblocks in the input space in the SOM topological space. Furthermore, they use the soft k nearest neighbor ensemble classifier to efficiently capture the similarity relationship obtained by SOM projections, which in turn improves the whole system's robustness. However, this method assumes knowledge of the occluded parts in advance. Since partial occlusion affects only specific local features, Ref. [53] proposes a local Gaussian kernel for feature extraction and trains the SVM using the summation of local kernels as combine features. However, it is suboptimal to use the same size local kernel rather than select the appropriate size of the local kernel for local feature extraction. In Ref. [133], a non-metric partial similarity measure based on local features is introduced to capture the prominent partial similarity among face images while excluding the unreliable and unimportant features.

In paper [124], a face recognition method is proposed that takes partial occlusions into account by using statistical learning of local features. To this end, they estimated the probability density of the SIFT feature descriptors observed in training images based on a simple Gaussian model. In the classification stage, the estimated probability density is used to measure the importance of each local feature of test images by defining a weighted distance measure between two images. Based on this work, they extended the statistical learning based on local features to a general framework [125], which combines the learned weight of local features and feature-based similarity to define the distance measurement. However, feature extraction from the local region cannot code the spatial information of faces. Besides, the unreliable features from the

occluded area are also integrated into the final representation, which will reduce the performance. McLaughlin et al. [96] try to solve random partial occlusion in test images using the largest matching areas at each point on the face. They assume that the occluded test image region can be modeled by an unseen-data likelihood with a low posterior probability. More specifically, they de-emphasize the local facial area with low posterior probabilities from the overall score for each face and select only reliable areas for recognition, which results in improved robustness to partial occlusion.

3) Sparse representation classifier: Apart from these statistical learning methods, several algorithms use sparse representation classifiers (SRC) to tackle the occlusion in face recognition. Ever since its introduction [153], SRC attracts increasing attention from researchers. This method explores the discriminative power of sparse representation of a test face. It uses a linear combination of training samples plus sparse errors to account for occlusion or corruption as its representation. Yang et al. [163] propose a Gabor feature based Robust Representation and Classification (GRRC) scheme. They use Gabor features instead of pixel values to represent face images, which can increase the ability to discriminate identity. Moreover, the use of a compact Gabor occlusion dictionary requires less expensive computation to code the occlusion portions compared with that of the original SRC. To investigate the effectiveness of the proposed method, they conduct extensive experiments to recognize faces with block occlusions as well as real occlusions. A subset of the AR database was used in this experiment. It consists of 799 images (about eight samples per subject) of non-occluded frontal views with various facial expressions for training. The sunglasses test set contains 100 images with the subject wearing sunglasses (with a neural expression), and the scarves test set contains 100 images with the subject wearing a scarf (with a neural expression). The proposed GRRC achieves 93.0% recognition accuracy on sunglasses and 79% accuracy on scarves, which outperforms the original SRC by 5% and 20%, respectively. In paper [84], artificial occlusions are included to construct training data for training sparse and dense hybrid representation framework. The results show that artificially introduced occlusions are important to obtain discriminative features. Structured occlusion coding (SOC) [148] is proposed to employ an occlusion-appended dictionary to simultaneously separate the occlusion and classify the face. In this case, with the use of corresponding parts of the dictionary, face and occlusion can be represented respectively, making it possible to handle large occlusion, like a scarf. In paper [32], efficient locality-constrained occlusion coding (ELOC) is proposed to greatly reduce the running time without sacrificing too much accuracy, inspired by the observation that it is possible to estimate the occlusion using identity-unrelated samples.

Recently, another work [162] attempts face recognition with single sample per person and intends to achieve robustness and effectiveness for complex facial variations such as occlusions. It proposes a joint and collaborative representation with local adaptive convolution feature (ACF), containing local high-level features from local regular regions. The joint and collaborative representation framework requires ACFs extracted

from different local areas to have similar coefficients regarding a representation dictionary. Ref. [33] proposes to learn a robust and discriminative low-rank representation (RDLRR) for optimal classification, which aiming to prepare the learned representations optimally for classification as well as to capture the global structure of data and the holistic structure of each occlusion induced error image. The results demonstrate that the method can yield better performance in case of illumination changes, real occlusion as well as block occlusion. Ref. [74] combines center-symmetric local binary patterns (CSLBP) with enhanced center-symmetric local binary patterns (ECSLBP) to build the SRC dictionary for occluded face recognition. In Ref. [169], discriminative multi-scale sparse coding (DMSC) is proposed to address single-sample face recognition with different kinds of occlusion. More specifically, it learns the auxiliary dictionary to model the possible occlusion variations from external data based on PCA and proposes a multi-scale error measurement strategy to detect and disregard outlier pixels due to occlusion. Ref [154] proposes a hierarchical sparse and low-rank regression model and uses features based on image gradient direction, leading to a weak low-rankness optimization problem. The model is suited for occluded face recognition and yields better recognition accuracy. In another work, NNAODL (nuclear norm based adapted occlusion dictionary learning) [27] has been proposed to construct corrupted regions and non-corrupted regions for occluded face recognition. The same occlusion parts in training images are used to construct the occlusions while normal training images are used to reconstruct non-occlusion regions, leading to improved computing efficiency. To cope with occluded face recognition with limited training samples, Ref. [187] proposes a structural element feature extraction method to capture the local and contextual information inspired by the human optic nerve characteristics for face recognition. Besides, an adaptive fusion method is proposed to use multiple features consisting of a structural element feature, and a connected-granule labeling feature. It applies Reinforced Centrosymmetric Local Binary Pattern (RCSLBP) to better handle the robustness to the occlusion. Finally, few-shot sparse representation learning is applied for few-shot occluded face recognition.

4) Deep learning: Face representation obtained by DCNN is vastly superior to other learning-based methods in discriminative power, owing to the use of massive training sets [94]. Face verification performance has been boosted due to advanced deep CNN architectures [47], [68], [127], [131], [132] and the development of loss functions [25], [86], [87], [123], [130], [144], [145], [149]. From a practical point of view, we can obtain occlusion-robust face representation if a massive training dataset is provided to contain sufficient occluded faces for a deep network to learn to handle occlusions [190]. However, it is tough to collect such a dataset. Data augmentation seems to be a plausible solution to obtain a large number of occluded faces. Training with augmented occluded faces ensures the features are extracted more locally and equally [107]. Lv et al. [90] synthesize occluded faces with various hairstyles and glasses for data augmentation to enlarge the training dataset, which alleviates the impact of

occlusions. Specifically, 87 common hairstyle templates with various bangs and 100 glasses templates are collected to synthesize training face images with different hairstyles and face images with different glasses, to enable the CNN model is robust to various hairstyles and glasses. The method indeed relieves the data deficiency problem and results in improved performance. However, its use is limited to handling sunglasses and hair in recognition and does not solve the occlusion problem in general. In paper [136], instead of using synthetic occluded faces directly, they first identify the importance of face regions in an occlusion sensitivity experiment and then train a CNN with identified face regions covered to reduce the model's reliance on these regions. Specifically, they propose to augment the training set with face images with occlusions located in high-effect regions (the central part of the face) more frequently than in low effect regions (the outer parts of the face). This forces the model to learn more discriminative features from the outer parts of the face, resulting in less degradation when the central part is occluded. However, predefined occlusions may cause performance degradation when dealing with face images with an occlusion that is not of the same size. Cen et al. [11] propose a DDRC (deep dictionary representation based classification) scheme to alleviate the occlusion effect in face recognition, where the dictionary is used to linearly code the deep features that are extracted from a convolutional neural network. The dictionary is composed of deep features of the training samples and an auxiliary dictionary associated with the occlusion patterns of the testing face samples. However, the proposed DDRC assumes that the occlusion pattern of the test faces is included in the auxiliary dictionary, which limits its use.

IV. OCCLUSION AWARE FACE RECOGNITION

If only visible facial parts are used for recognition, then the occlusion problem is mitigated to some extent. These approaches that explicitly exclude the occlusion area are called occlusion aware face recognition (OAFR). There are two groups of methods that constitute OAFR. One is occlusion detection based face recognition, which detects the occlusions first and then obtains a representation from the non-occluded parts only. The other one is partial face recognition, which assumes that a partial face is available and aims to use a partial face for recognition. Occlusion detection is ignored during partial face recognition. A taxonomy of occlusion aware face recognition is shown in Fig. 6.

A. Occlusion detection based face recognition

To explicitly make use of facial parts for face recognition, some methods explicitly detect the occlusion and perform face recognition based on their results. The other techniques obtain visible facial parts for face recognition based on the prior knowledge of occlusion, which is called visible part selection.

1) Occlusion detection: An intuitive idea to deal with occlusions in face recognition is to detect the occlusions first and then recognize the face based on unoccluded facial parts. Methods use predefined occlusion types as a substitute for arbitrary occlusion in different locations to simplify the
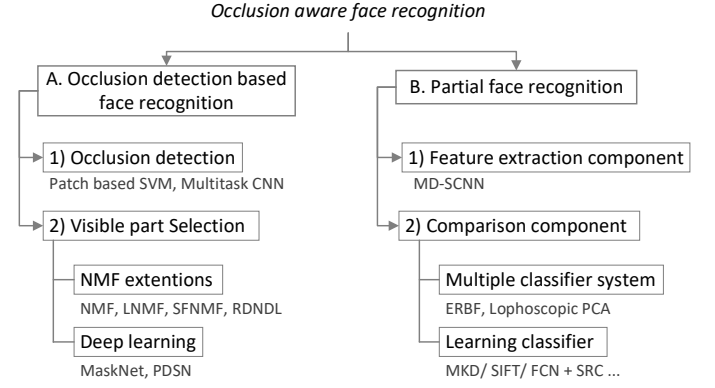

Fig. 6. Method classification used in occlusion aware face recognition approaches.

occlusion challenge. Usually, scarves and sunglasses are used as representative occlusions because of their high probability of appearance in the real world. Based on this idea, Ref. [103] introduces a 1-NN (nearest neighbor) threshold classifier to detect the occluded face region in a PCA subspace learned from training data of occlusion-free patches. Then they apply the selective local non-negative matrix factorization method to select features corresponding to occlusion-free regions for recognition. Some early works [19], [98] employ a binary classifier to search for the occluded area and incorporate only the unoccluded parts for comparison. Specifically, they first divide the face into multiple non-overlapping regions and then train an SVM classifier to identify whether the facial patch is occluded or not. By excluding occluded regions, improved overall recognition accuracy is observed. However, the performance is far from satisfactory and very sensitive to the training dataset. Ref. [4] proposes to utilize compressed sensing for occlusion detection by using occluded and occlusion-free faces of the same identity. For the recognition process, discriminative information is extracted by excluding the occluded areas detected.

Since occlusions can corrupt the features of an entire image in some way, deep learning techniques are developed to alleviate the problem by producing a better representation. Ref. [156] designs a convolutional neural network to detect the occlusion in a multitask setting. More specifically, there are four region-specific tasks for occlusion detection, and each aims to predict the occlusion probability of the specific component: left eye, right eye, nose, and mouth. However, predicting only predefined occlusions limits flexibility, and inaccuracy of occlusion detection can, in return, harm the recognition performance.

2) Visible part selection: Some works select visible facial parts for recognition and skip occlusion detection by assuming the prior knowledge of occlusion. Ref. [101] compares several subspace-based methods including PCA, Non-negative matrix factorization (NMF) [72], Local NMF [76] and Spatially Confined NMF [100] and uses the partial information available for face recognition. During face recognition, the eye region is selected when people are wearing masks or veils, and the bottom region is used when people are wearing glasses. This method has a deficiency in flexibility use because well-aligned

predefined subregions are hard to obtain in the real scenario. A paper [108] in this direction extends NMF to include adaptive occlusion estimation based on the reconstruction errors. Low-dimensional representations are learned to ensure that features of the same class are close to that of the mean class center. This method does not require prior knowledge of occlusions and can handle large continuous occlusions.

In paper [143], a proposed MaskNet is added to the middle layer of CNN models, aiming to learn image features with high fidelity and to ignore those distorted by occlusions. MaskNet is defined as a shallow convolutional network, which is expected to assign lower weights to hidden units activated by the occluded facial areas. Recently, Song et al. [128] propose a pairwise differential siamese network (PDSN) to build the correspondence between the occluded facial block and the corrupted feature elements with a mask learning strategy. The results show improved performance on synthesized and realistic occluded face datasets.

B. Partial face recognition

It is worth mentioning that we classify partial face recognition as occlusion aware methods because partial face recognition skips the occlusion detection phase and focuses on face recognition when arbitrary patches are available, which can be seen as implicit occlusion awareness. Partial faces frequently appear in unconstrained scenarios, with images captured by surveillance cameras or handheld devices (e.g., mobile phones) in particular. To the best of our knowledge, research for partial face detection has so far been ignored in literature reviews. It is essential to search for the semantic correspondence between the partial face (arbitrary patch) and the entire gallery face since it is meaningless to compare the features of different semantic facial parts. The semantic correspondence can be completed either in the feature extraction phase to extract invariant and discriminative face features, or in the comparison phase to construct a robust face classifier. Feature extraction and comparison methods can be developed to address the partial face recognition problem. Therefore, we categorize the methods as feature-aware and comparison-aware methods.

1) Feature extraction component: As for feature-aware methods, Multiscale Double Supervision Convolutional Neural Network (MDSCNN) [50] is proposed to have multiple networks trained with facial patches of different scales for feature extraction. Multiscale patches are cropped from the whole face and aligned based on their eye-corners. Each network is training with face patches of a scale. The weights of multiple networks are combined learned to generate final recognition accuracy. Even though this method can yield good feature representations, it is troublesome to train 55 different Double Supervision Convolutional Neural Networks (DSCNNs) according to different scaled patches. It is time-consuming in practice because window sliding is needed to generate multiscale patches for recognition.

2) Comparison component: Comparison-aware methods facilitate the semantic correspondence in the comparison phase of face recognition. Among the comparison based approaches, the multiple classifier systems use a voting mechanism to tolerate the misalignment problem to an extent. In this regard, Gutta et al. [46] present an Ensemble of Radial Basis Function (ERBF) Networks consisting of nine RBF components, and each of which is determined according to the number of clusters and overlap factors. A verification decision is based on the output generated by the RBFs. In paper [135], they propose the Lophoscopic PCA method to recognize faces in the absence of part of the relevant information. For this purpose, masks (black rectangles) that cover specific regions of the face, including left eye, right eye, both eyes, nose, mouth, and no mask are introduced to compute different subspaces. They learn the weights for each subspace during training. When classifying the subject, weights regarding each face subspace are considered and combined for recognition. These methods are not entirely satisfactory but need improvement, particularly in terms of their recognition rate.

Alternatively, a learning-based classifier compensates for the difficulty in the alignment of partial face recognition. Ref. [80] conducts the first systematic study to recognize an arbitrary patch of a face image without preliminary requirement on face alignment, which is regarded as a milestone in the field of partial face recognition. The proposed method employs the multi-keypoint descriptors (MKD) to represent a holistic or partial face with a variable length. Descriptors from a large gallery construct the dictionary, making it possible to sparsely represent the descriptors of the partial probe image and infer the identity of the probe accordingly. However, SRC requires a sufficient number of faces to cover all possible variations of a person, which hinders the realization of a practical application.

Even though most learning-based classifier papers are SRC based, there is a small group that develops similarity measures to address the partial face recognition problem. Ref. [55] utilizes the instance-to-class distance to address the partial face recognition with scale-invariant feature transform (SIFT) descriptors. The similarity between each probe patch and gallery image is obtained by comparing a set of local descriptors of one probe image to the nearest neighbor descriptors of all gallery images with the sparse constraint. As an improvement, Ref. [150] and [151] consider partial face recognition as a feature set matching problem, where they explicitly use geometric features and textural features for simultaneous matching. Robust point set matching (RPSM) [151] considers both geometric distribution consistency and textural similarity. Moreover, a constraint on the affine transformation is applied to prevent unrealistic face warping. However, these methods would fail to work if face keypoints are unavailable due to occlusions. Moreover, the computation complexity is high, which makes the recognition process slow.

Recently, there is a trend to combine the deep learning and SRC methods to tackle the partial face recognition [48], [49]. Dynamic feature learning [48] combines a fully convolutional network (FCN) with sparse representation classification (SRC) for partial face recognition. The sliding window matching proposed in paper [186] searches for the most similar gallery part by sliding the window of the same size as the partial probe. The combination of sliding window matching based on FCN and SRC brings a promising performance. As an improved version, Ref. [49] uses multi-scale representations in

dynamic feature learning, which increase the ability to tolerate the misalignment between partial probe patches and the gallery face.

V. OCCLUSION RECOVERY BASED FACE RECOGNITION

Apart from addressing the occlusion problem in feature space, one intuitive idea is to take occluded face recovery as a substitution to solve the occlusion in image space. Occlusion recovery methods recover a whole face from the occluded face, which allows the direct application of conventional face recognition algorithms. Existing occlusion recovery methods for face recognition use (i) reconstruction based techniques for face recognition, or (ii) inpainting techniques, which treat the occluded face as an image repairing problem. A possible way to classify the methods can be seen in Fig. 7.
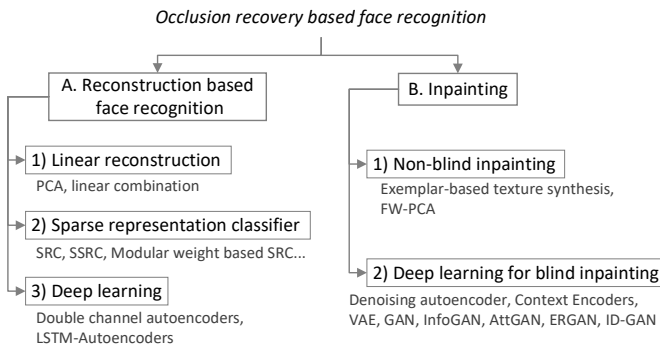

Fig. 7. Classification of methods used in occlusion recovery based face recognition approaches.

A. Reconstruction based face recognition

Image-based two-dimensional reconstructions carefully study the relationship between occluded faces and occlusion-free faces. The reconstruction techniques are classified as linear reconstruction, sparse representation classifier (dictionary learning), and deep learning techniques.

1) Linear reconstruction: As for reconstruction techniques, Ref. [112] utilizes PCA reconstruction and recursive error compensation to remove the eye occlusions caused by glasses. It combines a Markov Random Field model with a parse representation of occluded faces to improve the reconstruction of corrupted facial regions. There are many variants [24], [31], [73] employing PCA to detect outliers or occlusion and then reconstruct occlusion-free face images. Ref. [62] estimates an occluded test image as a linear combination of training samples of all classes, which allows having occlusions in training and testing sets. Distinct facial areas are weighted differently so that only non-occluded facial parts are used for reconstruction.

2) Sparse representation classifier: The sparse representation classifier [153] is considered to be the pioneering work on occlusion robust face recognition. This method explores the discriminative power of sparse representation of a test face. It uses a linear combination of training samples plus sparse errors accounting for occlusions or corruption as its

representation. To better tackle the occlusion, the SRC introduces an identity matrix as an occlusion dictionary on the assumption that the occlusion has a sparse representation in this dictionary. Ref. [191] improves sparse representation by including prior knowledge of pixel error distribution. In paper [26], they design a graph model-based face completion system for partially occluded face reparation. They leverage image-based data mining to find the best-matched patch to guide occluded face synthesis from the images, derived from the selection of sparse representation classification (SRC). The final face completion proceeds in the graph-based domain with the help of graph Laplace.

Similarly, Ref. [78] proposes a reconstruction approach consisting of occlusion detection and face reconstruction. First, the downsampled SRC is used to locate all possible occlusions at a low computing complexity. Second, all discovered face pixels are imported into an overdetermined equation system to reconstruct an intact face. An innovative solution for the occlusion challenge is presented by structured sparse representation based classification (SSRC) [109] to learn an occlusion dictionary. The regularization term of mutual incoherence forces the resulting occlusion dictionary to be independent of the training samples. This method effectively decomposes the occluded face image into a sparse linear combination of the training sample dictionary and the occlusion dictionary. The recognition can be executed on the recovered occlusion-free face images. Nevertheless, this method requires retraining of the model to handle the different occlusions.

In paper [185], a new criterion to compute modular weight-based SRC is proposed to address the problem of occluded face recognition. They partition a face into small modules and learn the weight function according to the Fisher rate. The modular weight is used to lessen the effect of modules with low discriminant and to detect the occlusion module. More recently, Ref. [60] proposes a robust representation to model contiguous block occlusions based on two characteristics. The first characteristic introduces a tailored potential loss function to fit the errors of distribution. Specifically, a Laplacian sparse error distribution or more general distributions based on M-Estimators. The second characteristic models the error image, which is the difference between the occluded test face and the unoccluded training face of the same identity, as low-rank structural. Wang et al. [147] proposed a method equipped with two stages: varying occlusion detection stage consisting of occlusion detection and elimination, and iterative recovery stage consisting of occluded parts being recovered and unoccluded parts being reserved. With the use of iteratively recovered strategy, this joint occlusion detecting and recovery method can produce good global features to benefit classification.

3) Deep learning: A few works use deep learning techniques for occlusion reconstruction. One is work [20], which extends a stacked sparse denoising autoencoder to a double channel for facial occlusion removal. It adopts the layerwise algorithm to learn a representation so that the learned encoding parameters of clean data can transfer to noisy data. As a result, the decoding parameters are refined to obtain a noise-free output. The other work [183] proposes to combine the LSTM and autoencoder architectures to address the face de-

occlusion problem. The proposed robust LSTM-Autoencoders (RLA) model consists of two LSTM components. One spatial LSTM network encodes face patches of different scales sequentially for robust occlusion encoding, and the other dual-channel LSTM network is used to decode the representation to reconstruct the face as well as to detect the occlusion. Additionally, adversarial CNNs are introduced to enhance the discriminative information in the recovered faces.

B. Inpainting

Image inpainting techniques are widely used to carefully obtain occlusion-free images and are not limited to face images. Inpainting techniques focus on repairing the occluded images and leave face recognition out of consideration. They can be divided into 1) non-blind inpainting and 2) blind inpainting categories, depending on whether the location information of corrupted pixels is provided or not. Deep learning is an effective approach to blind inpainting.

1) Non-blind inpainting: Those techniques fill in the occluded part of an image using the pixels around the missing region. Exemplar-based techniques that cheaply and effectively generate new texture by sampling and copying color values from the source are widely used. In paper [21], a non-blind inpainting method proposes a unified scheme to determine the fill order of the target region, using an exemplar-based texture synthesis technique. The confidence value of each pixel and image isophotes are combined to determine the priority of filling. Ref. [64] presents an image inpainting technique to remove occluded pixels when the occlusion is small. More specifically, it combines feature extraction and fast weighted-principal component analysis (FW-PCA) to restore the occluded images. More recently, a hybrid technique [140] has been proposed where a PDE method and modified exemplar inpainting is utilized to remark the occluded face region. However, the occlusion type of face images studied in this work is not representative of the real scenario.

2) Deep learning for blind inpainting: Generative models are known for the ability to synthesize or generate new samples from the same distribution of the training dataset. The core problem in generative models is to address density estimation by unsupervised learning, which can be carried out by explicit density estimation [66], [105], or implicit density estimation [41]. The most prominent generative models are originated from the variational autoencoder (VAE) [66] and the generative adversarial network (GAN) [40], [41]. The traditional autoencoder is used to reconstruct data, and the VAE [66] applies probabilistic spin to the traditional autoencoder to allow generating a new image. Having assumed that training data is generated from the underlying latent representation, the VAE derives a lower bound on the data likelihood that is tractable and can be optimized. This method is a principled approach to generative models, and useful features can be extracted by inference of $q(z|x)$. However, generated images are blurrier and of low quality. Instead, the GAN [41] learns to generate a new image from training distribution using a minimax game between the generator network and the discriminator network. The discriminator wants to distinguish between real and fake images, while the generator wants to fool the discriminator by generating real-looking images. Ref. [116] uses convolutional architectures in the GAN and can generate realistic samples. However, the GAN is notorious for unstable training characteristics. Makhzani et al. [92] propose an adversarial autoencoder (AAE) to use the GAN framework as a variational inference in a probabilistic autoencoder to ensure that generating from any part of prior space results in meaningful samples.

There are GAN variants for all kinds of applications. We focus on methods that are relevant to face image editing and image inpainting. One is a blind-inpainting work [157] that combines sparse coding [29] and deep neural networks to tackle image denoising and inpainting. In particular, a stacked sparse denoising autoencoder is trained to learn the mapping function from generated corrupted noisy overlapping image patches to the original noise-free ones. The network is regularized by a sparsity-inducing term to avoid overfitting. This method does not need prior information about the missing region and provides solutions to complex pattern removal like the superimposed text from an image. Context Encoders [114] combine the encoder-decoder architecture with context information of the missing part by regarding inpainting as a context-based pixel prediction problem. Specifically, the encoder architecture is trained on the input images with missing parts to obtain a latent representation while the decoder architecture learns to recover the lost information by using the latent representation. Pixel-wise reconstruction loss and adversarial loss are jointly used to supervise the context encoders to learn semantic inpainting results. Several variants of context encoders [114] are proposed: some extend it by defining global and local discriminators [59], [77] and some take the result of context encoders as the input and apply joint optimization of image content and texture constraints to avoid visible artifacts around the border of the hole [161]. Ref. [139] relies on the power of the generative network to complete the corrupted image without requiring an external training dataset. A partial convolution based network [83] is proposed to only consider valid pixels and apply a mechanism that can automatically generate an updated mask, resulting in robustness to image inpainting for irregular holes. Information Maximizing Generative Adversarial Networks (InfoGAN) [15] maximize the mutual information between latent variables and the observation in an unsupervised way. It decomposes the input noise vector into the source of incompressible noise z and the latent code c which will target the salient structured semantic features of data distribution. By manipulating latent codes, several visual concepts such as different hairstyles, presence or absence of eyeglasses are discovered. Occlusion-aware GAN [17] is proposed to identify a corrupted image region with an associated corrupted region recovered using a GAN pre-trained on occlusion-free faces. Ref. [16] employs GAN for eyes-to-face synthesis with only eyes visible. Very recently, AttGAN [51], a face image editing method, has imposed attribute classification constraints to the generated image so that the desired attributes are incorporated. Hairstyles and eyeglasses that may cause occlusion in a face image are treated as attributes which can be triggered to be present or

absent in the generated image. ERGAN (Eyeglasses removal generative adversarial network) [54] is proposed for eyeglasses removal in the wild in an unsupervised manner. It is capable of rendering a competitive removal quality in terms of realism and diversity. In paper [35], ID-GAN (identity-diversity generative adversarial network) combines a CNN-based recognizer and GAN-based recognition to inpaint realism and identity-preserving faces. The recognizer is treated as the third player to compete with the generator.

VI. Performance Evaluation

In this section, we first evaluate the performance of occluded face detection on the MAFA dataset [36] of partially occluded faces. Next, we present the performance of face recognition under occlusion in terms of the identification rate and the verification accuracy on multiple benchmarks such as the AR [3], CAS-PEAL [34], and Extended Yale B [37] datasets. Then we describe the representative algorithms based on the proposed categories. In addition, we also categorize them in the face recognition pipeline according to which component they work on to tackle the occlusion.

A. Evaluation of Occluded Face Detection

There are several datasets for face detection benchmarking, which are the FDDB [61], PASCAL Face [159], AFW [192], IJB-A [67], Widerface [166] and MAFA [36] datasets. MAFA is created for occluded face detection, involving 60 commonly used masks, such as simple masks, elaborate masks, and masks consisting of parts of the human body, which occur in daily life. Besides, it contains $35,806$ face annotations with a minimum size of 32×32 pixels. Some examples of occluded face images are shown in Fig. 1. To the best of our knowledge, the MAFA dataset takes the occlusions as the main challenge in face detection, so it is relevant to evaluate the capacity of occluded face detection methods. The performances of representative algorithms on the MAFA dataset are summarized in Table II (derived from paper [18]). The precision on the MAFA testing set with the IoU threshold value 0.5 is shown. Only a few methods report the results below.

TABLE II
EVALUATION SUMMARY OF DIFFERENT OCCLUDED FACE DETECTION ALGORITHMS ON MAFA. THE SETTING 'MASKED' ONLY COUNTS THE FACES ANNOTATED BY THE MAFA DATASET AND 'W/O IGNORED' COUNTS ALL DETECTED FACES, INCLUDING THE ONES THAT ARE MISSING ANNOTATION.

Approach	Publication	Precision 'masked'	Precision 'w/o Ignored'
Occlusion unaware	[176]	NA	60.8%
Occlusion aware	[36]	NA	76.4%
	[146]	76.5%	88.3%
Occlusion segmentation	[18]	83.5%	91.9%

B. Existing OFR benchmark datasets

There are numerous standard datasets for general face recognition, but they are not appropriate for OFR (occluded face recognition) because occluded faces are barely present in the datasets. Alternatively, researchers make the most of the general datasets to generate synthetic occluded datasets by incorporating synthetic occlusions, occluding rectangles, etc. Five categories regarding OFR testing scenarios are illustrated in Fig. 2. In this way, synthetic occluded datasets can meet the requirements of OFR, where an occlusion-free gallery is queried using occluded faces.

TABLE III
SUMMARY OF EXISTING BENCHMARK DATASETS FOR OCCLUDED FACE RECOGNITION. '# SUB' IS SHORT FOR THE NUMBER OF SUBJECTS. '# IMS' MEANS THE NUMBER OF IMAGES. 'REAL OCC.' SAYS WHETHER REAL OCCLUSIONS ARE INCLUDED IN THE ORIGINAL DATASET. 'SYN. OCC.' IS SHORT FOR SYNTHESIZED OCCLUSIONS IMPOSED ON THE ORIGINAL DATASET TO OBTAIN 'D-OCC'.

Dataset	# Sub	# Ims	Real Occ.	Syn. Occ.
AR [3]	126	$4,000$	Yes	NA
Extend YaleB [37]	38	$161,289$	No	Unrelated Ims
ORL [121]	40	400	No	Rectangle
FERET [110]	1500	$13,000$	No	Rectangle
LFW [58]	5749	$13,000$	No	Partial Faces
CAS-PEAL [34]	1040	$9,594$	Yes	NA
CMU-PIE [43]	68	$41,368$	No	Rectangle
PubFig [69]	200	$58,797$	No	Partial Faces
NIR-Distance [50]	276	4300	No	Partial Faces
NIR-Mobile [177]	374	$11,286$	No	Partial Faces
FRGC [102]	276	943	No	Partial Faces
FRGC-v2 [115]	466	$4,007$	No	Partial Faces

We make it a rule to state the 'D' as the original dataset, and 'D-occ' as the occluded dataset originated from 'D'. If there is real occlusion included in the 'D', then it is directly used for establishment. Otherwise, synthesized occlusions are imposed to form 'D-occ' for establishment. Table III summarizes each dataset by the number of subjects, the number of images, whether or not real occlusions are included, and if not, what kind of synthesized occlusions are used to obtain the 'D-occ' dataset. Here AR and Extended Yale B, the most widely used face datasets, are discussed.

AR face database is one of the very few datasets that contain real occlusions (see Fig. 2 first column). It consists of over 4000 faces of 126 individuals: 70 men and 56 women, taken in two sessions with a two week interval. There are 13 images per individual in every session, and these images differ in terms of facial expression, illumination, and partial occlusion, getting sunglasses and scarves involved. Index 8 and 11 of each session indicates the person wearing sunglasses or a scarf, respectively. Index $9-10$ and $11-12$ combine the sunglasses or the scarf with illuminations, respectively.

Extended Yale B face database contains $161,289$ face images from 38 persons, each captured under nine poses and 64 different illuminations without occlusion. It is widely used to evaluate the efficacy of the algorithms under synthesized

occlusions. Occluding unrelated images are randomly super-imposed on the occlusion-free faces (see Fig. 2 last column). Typically, gallery images are occlusion-free faces, and test images are randomly occluded with unrelated images such as a baboon, or a square image. Sometimes, occluded faces are occluded with, for example white or black rectangles to constitute a 'D-occ' dataset (see Fig. 2 fourth column).

C. Evaluation of Occluded Face Recognition

In this part, we demonstrate the results of the most representative methods based on the proposed categories. In addition, we also categorize methods based on the face recognition pipeline and show evaluation results from this aspect.

1) Evaluation based on the proposed categories: We summarize the identification performances of representative algorithms on the AR database in Table IV. To make sure these methods can be compared, we group the experimental settings and introduce the corresponding abbreviation for simplicity as follows:

(1) Face images for training and testing belong to the same individual and without overlapping. We use 'S-TR-Xips' to abbreviate the situation when the training set contains X unoccluded images per person. Specifically, 'S-TR-SSPP' single sample per person in the training set remains a challenging task. Usually, the gallery set keeps the same training set if there are no additional announcements. We add asterisks to mark the gallery, only taking one neutral face for enrollment.

(2) Face images for training and testing belong to the same individual, without overlapping. The training set consists of occluded faces for training, denoting as 'S-TR-Occ$X$', with X as the number of images. Typically, one neutral face image per individual is enrolled.

(3) Testing subjects are a subset of training subjects denoting as 'Ex-TR-SSPP'. Images of extra persons are included in the training set. Apart from that, the same subjects are used for training and testing. The training set also consists of a single sample per person.

(4) There is no overlapping between subjects for training and testing, which is represented by 'D-TR'. We use the setting 'D-TR-DL' to indicate if there are large-scale general data for training with deep learning techniques.

(5) Images of subjects are randomly selected for training, and the remaining ones are for testing, what we called 'S-TR-Rand'.

The identification performance of representative algorithms on extended Yale B as well as other benchmarks is summarized in Table V. An evaluation summary of different categories of representative algorithms as regards verification rates is exhibited in Table VI.

As results on the AR dataset in Table IV show, these experimental setups were slightly different from paper to paper, bringing a struggle to interpret these results at the first sight. This is because the AR dataset does not provide standard protocols for use, which are essential to compare methods in a fair ground. Despite the absence of the standard protocols, it is still possible to get some useful observations and findings.

First, most methods treat session1 of AR as the target and report the results on sunglasses and scarf individually. Second, there are some popular protocols gradually formed among these methods, which can be treated as a relatively fair ground for comparison. These are (i) S-TR-7IPS, (ii) S-TR-8IPS, and (iii) S-TR-SSPP. Specifically, S-TR-SSPP in particular has become a challenge and defines an unambiguous protocol to follow. Last but not least, thanks to the presence of deep learning methods, there is a trend to rely on general massive data for training rather than splitting the AR dataset to form the training set. There is still some room to improve the OFR performance, especially when we take the SSPP (single sample per person) protocol into consideration.

Results on Extended Yale B and other benchmark datasets are shown in Table V and Table VI. Even though the original occlusion-free datasets such as Extended Yale B and ORL are well addressed by current face recognition algorithms, the 'D-occ' (occluded version) originated from these datasets still remains a challenge. It is not surprising that algorithms suffer severe degradation when 90% of the face parts are occluded by a rectangle or unrelated images. Moreover, these testing scenarios are not similar to what we would find in a realistic scenario. Apart from an occluding rectangle and unrelated images, some methods intend to solve partial face issues by using arbitrary patches cropped from a face image. Since partial faces are arbitrary patches, it is hard to be sure algorithms are on a level playing field. For example, partial faces with eye regions are supposed to contain more discriminative features compared to those with no eyes involved.

2) Evaluation based on the face recognition pipeline for categorization: Generally, occlusion robust feature extraction (ORFE) methods deal with occlusion using feature extraction. However, they sometimes treat occluded face recognition as a semantic correspondence problem and address the occlusion challenge in the comparison component of the face recognition pipeline. Similarly, occlusion aware face recognition (OAFR) usually solves the occluded face recognition using comparison. Nevertheless, some occlusion-aware techniques use an occlusion-robust dictionary to represent an occluded face, which lies in the feature extraction component. Therefore, we categorize methods discussed so far based on a general face recognition pipeline to offer a fresh perspective (see Table VII). Each component is elaborated as follows:

- Data Augmentation or Data Recovery: augmentation based techniques cope with the occlusion challenge by augmenting training data to include possible real-life occlusion. With the use of deep learning techniques, occlusion-robust features can be extracted. Recovery based methods use local or global knowledge to explicitly fill the occlusion area or de-occlude the face in an implicit way, for example using an encoder-decoder structure.
- Feature extraction intends to make the most of the locality characteristic of occlusion using patch-based engineered feature extraction or applies machine learning or deep learning to obtain features. Specifically, statistical learning or sparse representation are two commonly used techniques in this approach.
- Feature comparison: features obtained for an entire face

are of fixed length, while features for occluded face images are determined by the occlusion area and are therefore of varied length. The comparison phase is used to find out the semantic correspondence between fixed-length and varied features because it is meaningless to match different semantic parts.

- Fusion strategy relies on the assumption that occlusions have less effect on identity recognition compared with other visible facial areas. Occlusion challenges can be met by fusing the decision of multiple classifiers or using the majority voting strategy to minimize the effect of features extracted from the occlusion.

VII. Discussion

In this section, future dataset challenges and research challenges are discussed. In many cases, a new research challenge means specific dataset requirements. On the other hand, datasets also reflect underlying challenges that need to be solved in real life.

A. Future Dataset Challenges

There are three major issues in the datasets referred to: **dataset scale**, **occlusion diversity**, and **standard protocol**. The datasets for occluded face recognition are of a small scale. AR is one of the very few that contain real occlusions and images of only 126 individuals are included. As for occlusion diversity, sunglasses and scarf occlusions are frequently considered. However, occlusions in real life are a lot more diverse than that. Regarding synthesized occlusions, occluding rectangles and unrelated images are commonly applied. In the literature, Baboon is a typically used unrelated image to generate synthetic occluded faces. However, these synthetically occluded faces like those using Baboon, or a black/white rectangle are not representative of real-life occluded faces. As regards a standard protocol, the results of representative approaches obtained from AR and other benchmark datasets cannot be compared objectively due to a wide diversity of occluded faces. Therefore, future research will have to overcome three weaknesses in the current datasets. Future benchmark datasets will contain more individuals, various and sufficient occlusions, and well-defined evaluation protocols.

Apart from specific requirements for benchmark datasets, collecting a massive occluded face dataset for training is also inevitable as we plunge into deep learning techniques. However, the main source of face images is usually the web, where labelled faces with occlusions are scarce. How to label occluded faces efficiently is an open issue that needs to be resolved because some severely occluded faces are difficult or impossible to be recognized by humans. From this aspect, occluded face recognition remains a challenge in the future. Given the combination of the pose variations, low resolution etc., recognizing unconstrained faces is far from being solved.

For the development of occluded face recognition techniques, new large-scale datasets are needed that respond to the occlusion challenges. Recently, IJB-C, a large-scale dataset, has been made publicly available. It not only contains many natural occlusions but also provides annotated occlusion locations. The public availability of IJB-C may usher in a new development of occluded face recognition. In the future, we expect to see the standard evaluation of fully efficient occluded face recognition algorithms on the AR dataset as well as on newly developed real-life and large-scale IJB-C face datasets.

B. Future Research Challenges

In the future, unconstrained occluded face recognition will become a challenging problem that needs to be addressed. Unlike the occluded faces we are handling, unconstrained occluded faces should conform to the future benchmark datasets, not limited by predefined occlusions. Currently used datasets such as AR and Extended Yale B are both problematic since they are aligned and captured in a constrained environment. Apart from unconstrained occluded face recognition, it remains an open issue to resolve face recognition burdened with unconstrained occlusions and other challenges such as large pose variations, age gap, low resolution, etc. It is worth mentioning that unconstrained face recognition usually does not take the occlusion challenge as the major challenge. Instead, unconstrained face recognition mainly takes pose and expression challenges into consideration. Face recognition will be free from constraints if real-life occlusion plays an equally important role among these challenges.

Since the existing occluded face recognition approaches cope with occlusion challenges from distinct perspectives, we expect the individual improvements in occlusion robust feature extraction, occlusion aware face recognition, and occlusion recovery based face recognition. The majority of methods in the literature so far follow the line of traditional machine learning techniques. Methods for occluded face recognition in the future will make use of deep learning techniques, especially CNN based deep networks. Based on the method categories in Fig. 3, we suggest a number of ways to develop a face recognition system that is better able to handle occlusions as follows.

Novel data augmentation techniques are needed that could help to learn more discriminative feature representations robust to occlusions if a CNN deep network is trained on the augmented dataset. Current solutions for data augmentation mainly rely on large-scale training data from the web, and use occlusion templates to generate synthesized occluded faces. However, manually designed occlusion templates heavily rely on accurate facial landmarking, and an acceptable occluded face can only be achieved if the occlusion template is properly aligned with the facial landmarks. For example, eyeglasses should be scaled and aligned to the eye centers. One potential solution is to take advantage of GANs to generate natural occluded faces by transferring the occlusion type from a real occluded face to an occlusion-free face.

More effective occlusion recovery models need to be devised to make the most of state-of-the-art unified face recognition systems. Current occlusion recovery solutions adopt an encoder-decoder architecture for recovery and are designed for visually pleasing results rather than accurate reconstruction. One potential way to resolve this issue is to combine the occlusion recovery task and identification or verification task in

a multi-task manner. The adversarial loss can be incorporated to ensure that the recovered face looks natural.

New occlusion detection techniques need to be designed to take advantage of massive public datasets. Better occlusion detection not only makes it possible to generate occlusion templates automatically but also provides an informative preprocessing for face recognition. Most methods reviewed address the occlusion detection problem by partitioning the face into patches using a binary classifier, resulting in a rough occlusion area. Combining occlusion detection and occluded face recognition in a unified framework seems a promising way, leading to an automatic face recognition system.

VIII. CONCLUSION

In this paper, we present a thorough survey of face recognition techniques under occlusion and systematically categorize methods up to now into 1) occlusion robust feature extraction, 2) occlusion aware face recognition, and 3) occlusion recovery based face recognition. Newly published and innovative papers, especially recent deep learning techniques for occluded face recognition, have been discussed. Furthermore, we report comparative performance evaluations in terms of occluded face detection and face recognition. In the end, we discuss future challenges in terms of dataset and research (including potential solutions) that move the field forward.

APPENDIX

A glossary of abbreviations and expansions for terminologies is illustrated in Table IX.

TABLE IV
EVALUATION SUMMARY OF REPRESENTATIVE ALGORITHMS ON AR DATASET REGARDING IDENTIFICATION RATES. THREE CATEGORIES ARE:
1) **ORFE**=OCCLUSION ROBUST FEATURE EXTRACTION, 2) **OAFR**=OCCLUSION AWARE FACE RECOGNITION, AND 3) **ORECFR**=OCCLUSION
RECOVERY FR. THE DETAIL OF 'EXPERIMENT SETTINGS' ABBREVIATIONS IS SHOWN. IN COLUMN 'OCCLUSIONS,' THE ASTERISK MEANS SPECIFIC
OCCLUSION UNDER ILLUMINATION. IN COLUMN 'IDENTIFICATION RATES,' TR MEANS THESE METHODS TRAIN ON SESSION ONE.

Category	Publication	Occlusions	Experiment Settings	# of Train Subjects	# of Test Subjects	Identification Rates (%) Session1	Session2
ORFE: Patch-engineered Features	[180]	Sunglasses / Scarf	S-TR-8IPS	50	50	84.0 / 100.0	80.0 / 96.0
	[12]	Sunglasses / Scarf	S-TR-SSPP	100	100	89.0 / 73.0	98.0 / 92.0
ORFE: Learning-based Features	[93]	Sunglasses / Scarf	S-TR-SSPP	50	50	52.0 / 48.0	NA / NA
	[53]	Sunglasses / Scarf	S-TR-8IPS	40	40	80.0 / 70.8	NA / NA
	[133]	Sunglasses / Scarf	D-TR	50	50	98.0 / 90.0	NA / NA
	[153]	Sunglasses / Scarf	S-TR-8IPS	100	100	97.5 / 93.5	NA / NA
	[124]	Sunglasses&Scarf / Black block	S-TR-3IPS	100	100	89.0 / 94.0	NA / NA
	[163]	Sunglasses / Scarf	S-TR-7IPS	100	100	92.3 / 95.0	51.7 / 84.3
	[109]	Sunglasses* / Scarf* / Sunglasses&Scarf*	S-TR-Occ1	100	100	93.0 / 92.7 / 92.6	
	[125]	Sunglasses* / Scarf*	S-TR-3IPS	100	100	95.7 / 96.3	NA / NA
	[96]	Sunglasses⋆ / Scarf*	S-TR-Occ3	100	100	TR / TR	98.0 / 97.0
	[33]	Sunglasses / Scarf / Sunglasses&Scarf	S-TR-SSPP	100	100	TR / TR / TR	92.0 / 91.0 / 82.5
	[74]	Sunglasses* / Scarf*	S-TR-SSPP	80	80	88.0 / 69.0	56.0 / 44.0
	[169]	Sunglasses* / Scarf*	EX-TR-SSPP	80	80	95.8 / 90.0	78.3 / 77.9
	[162]	Sunglasses&Scarf	D-TR-DL	Webface+20	80	100.0	96.3
	[154]	Sunglasses&Scarf	S-TR-SSPP	100	100	75.0	
	[11]	Sunglasses / Scarf	S-TR-7IPS	50	50	94.7 / 99.3	85.3 / 98.7
	[35]	Sunglasses / Scarf	S-TR-3IPS	80	80	99.0 / 87.6	NA / NA
OAFR: Occlusion Detection Face Recognition	[103]	Sunglasses / Scarf	EX-TR-SSPP	100	100	NA / NA	49.0 / 55.0
	[98]	Sunglasses* / Scarf*	S-TR-3IPS*	80	80	TR / TR	54.2 / 81.3
	[4]	Sunglasses* / Scarf*	S-TR-8IPS-Occ1	60	60	97.5 / 99.2	NA / NA
	[108]	Sunglasses / Scarf	S-TR-RAND	100	100	95.2 / 94.2	
	[128]	Sunglasses / Scarf	D-TR-DL	Webface	100	99.7 / 100.0	NA / NA
OAFR: Partial Face Recognition	[55]	Sunglasses* / Scarf* / Sunglasses&Scarf*	S-TR-7IPS	100	100	94.3 / 98.0 / 96.2	NA / NA / NA
	[150]	Sunglasses* / Scarf* / Sunglasses&Scarf*	S-TR-14IPS	100	100	98.0 / 97.0 / 97.5	NA / NA / NA
	[151]	Sunglasses* / Scarf*	S-TR-7IPS	100	100	100.0 / 100.0	92.0 / 95.3
ORecFR: Occlusion Recovery Face Recognition	[62]	Sunglasses&Scarf*	S-TR-13IPS	100	100	TR	90.6
	[191]	Sunglasses / Scarf	S-TR-8IPS	100	100	100.0 / 97.0	NA / NA
	[26]	Sunglasses* / Scarf*	S-TR-7IPS	121	121	76.6 / 60.9	NA / NA
	[78]	Sunglasses*	S-TR-8IPS	100	100	97.5	NA
	[89]	Sunglasses / Scarf	S-TR-8IPS	100	100	94.5 / 95.0	NA / NA
	[185]	Sunglasses / Scarf	S-TR-3IPS-Occ3	120	120	NA / NA	68.5 / 70.7
	[60]	Sunglasses / Scarf	S-TR-2IPS	100	100	89.8 / 78.8	
	[147]	Sunglasses / Scarf	S-TR-4IPS	120	120	99.2 / 87.5	99.7 / 83.6

TABLE V

Evaluation Summary of Representative Algorithms on the other Benchmarks regarding identification rates. Three categories: 1) **ORFE**=Occlusion Robust Feature Extraction, 2) **OAFR**=Occlusion Aware Face Recognition, and 3) **ORecFR**=Occlusion recovery FR. In the 'Occlusion Arguments' column, details including the size of occlusion, the occlusion ratio to the face image and portion of noise are listed. In '# of Train Subjects,' NA represents not available from the paper. The Rank 1 identification rates are shown. Some results are roughly estimated from the figure in the original papers. Since not all methods are of the same experimental settings, these methods cannot be compared properly.

Category	Publication	Occlusions	Occlusion Arguments	Benchmark	# of Train Subjects	# of Test Subjects	Identification Rates(%)
ORFE	[53]	Black or White Rectangle	occlusion size:50 × 50	ORL	40	40	70.0
	[133]	Black Rectangle	occlusion size:50 × 50	FERET	240	960	78.5
	[153]	Unrelated image	occlusion ratio:0.5	Extended Yale B	38	38	65.3
	[163]	Unrelated image	occlusion ratio:0.5	Extended Yale B	38	38	87.4
	[109]	Unrelated image	occlusion ratio:0.8	Extended Yale B	38	38	70.0
	[96]	Unrelated image	occlusion ratio:0.8	Extended Yale B	38	38	53.0
	[162]	Wild condition	NA	LFW	108	50	86.0
			NA		851	124	65.3
	[169]	Glasses&Sunglasses	NA	CAS-PEAL	350	350	96.6
		Hat	NA				90.3
		Unrelated image	occlusion ratio:0.5	Extended Yale B	30	30	78.4
		Unrelated image	occlusion ratio:0.6				96.0
	[33]	Random pixels Rectangle	occlusion ratio:0.6	Extended Yale B	38	38	81.0
		Mixture noise	occlusion ratio:0.6				25.0
	[74]	Unrelated image	occlusion ratio:0.3	Extended Yale B	30	30	77.6
		Pepper noise	portion:40%	Extended Yale B	38	38	81.3
	[108]	White Rectangle	NA				82.9
		Salt&Pepper noise	portion:40%	CMU-PIE	68	68	98.5
		White Rectangle	occlusion size:10 × 10				98.8
	[27]	Black Rectangle	occlusion ratio:0.3				99.2
		Black Rectangle	occlusion ratio:0.4	Extended Yale B	38	38	97.6
		Black Rectangle	occlusion ratio:0.5				96.1
	[32]	Unrelated image	occlusion ratio:0.5	Extended Yale B	38	38	96.9
		Random pixels Rectangle	portion:70%				98.9
		Black Rectangle	occlusion ratio:0.5	CMU-PIE	68	68	93.9
OAFR	[55]	Arbitrary Patches	NA	Partial-LFW	158	158	34.8
	[150]	Arbitrary Patches	NA	Partial-LFW	158	158	50.7
		Unrelated image	occlusion ratio:0.5	Extended Yale B	32	32	30.2
	[151]	Unrelated image	occlusion ratio:0.5	Extended Yale B	32	32	56.7
		Arbitrary Patches	NA	Partial-PubFig	60	140	42.9
		Arbitrary Patches	NA	Partial-LFW	NA	1000	27.3
	[48]	Arbitrary Patches	NA	NIR-Distance	NA	276	94.9
		Arbitrary Patches	NA	Partial-YTF	NA	200	61.0
	[49]	Arbitrary Patches	NA	Partial-LFW	NA	1000	32.4
		Arbitrary Patches	NA	NIR-Distance	NA	127	92.8
		Arbitrary Patches	NA	NIR-Mobile	NA	178	93.8
ORecFR	[89]	Unrelated image	occlusion ratio:0.7	Extended Yale B	38	38	62.3
	[191]	Unrelated image	occlusion ratio:0.7	Extended Yale B	38	38	88.5
		Multiple patches	occlusion ratio:0.8				96.0
	[78]	Black or White Rectangle	occlusion ratio:0.57	Extended Yale B	38	38	90.4
		Unrelated image	occlusion ratio:0.5				87.7
	[185]	Black rectangle on eyes	occlusion ratio:0.3	Extended Yale B	38	38	98.6
	[60]	Unrelated image	occlusion ratio:0.6	Extended Yale B	38	38	95.8
		Unrelated image	occlusion ratio:0.9				71.9

TABLE VI

Evaluation Summary of Representative Algorithms regarding verification rates. Occlusion Robust Feature Extraction and Occlusion Aware Face Recognition report the verification rates by cut-off rules. In the 'Benchmark' column, 'Partial-x' means the new partial faces originate from the database named 'x.' Some results are roughly estimated from the figure in the original papers. Since not all methods are of the same experimental settings, these methods cannot be compared properly.

Publication	Occlusions	Benchmark	# of Subjects in Gallery	Verification Rates
[122]	Eyeglasses	FRGC-v2	466	90%@FAR=0.1%
[57]	Wild condition	LFW	1680	60%@FAR=0.1
[96]	Wild condition	LFW	1680	61%@FAR=0.1
[151]	Arbitrary Patches	Partial-LFW	1680	50%@FAR=0.1
	Arbitrary Patches	Partial-PubFig	140	63%@FAR=0.1
[48]	Arbitrary Patches	Partial-LFW	1000	29.8%@FAR=0.1%
[49]	Arbitrary Patches	Partial-LFW	1000	37.9%@FAR=0.1%

TABLE VII
SUMMARY OF REPRESENTATIVE ALGORITHMS BASED ON COMPONENTS THEY WORK ON DURING OFR. IN THE 'DATA AUGMENTATION/RECOVERY'
CATEGORY, DATA AUGMENTATION COMPONENT MEANS GENERATING SYNTHESIZED OCCLUDED FACES WHILE THE DATA RECOVERY COMPONENT
INTENDS TO ELIMINATE THE OCCLUDED FACIAL PART. THE FUSION STRATEGY COMPONENT CONSISTS OF FEATURE-LEVEL FUSION AS WELL AS
DECISION-LEVEL FUSION.

Pipeline Category	Publication
Data Augmentation/ Recovery	[15], [20], [21], [24], [26], [31], [35], [51], [54], [62], [64], [73], [78], [89], [90], [107], [112], [136], [140], [147], [183]
Feature Extraction	[1], [2], [11], [23], [27], [32], [33], [48], [49], [53], [60], [62], [65], [74], [80], [88], [93], [108], [109], [124], [125], [128], [134], [143], [148], [153], [162], [163], [169], [180], [187], [191], [193]
Feature Comparison	[4], [19], [55], [57], [96], [98], [101], [133], [150]–[152]
Fusion Strategy	[12], [46], [50], [103], [111], [122], [135], [185], [187]

TABLE VIII
SUMMARY OF REPRESENTATIVE ALGORITHMS REGARDING APPLICATION-ORIENTED PURPOSE FROM FIG. 3. THERE IS A CATEGORY FOR OCCLUDED
FACE DETECTION WITH THE ABBREVIATION 'OFD'. THREE CLASSES FOR OCCLUDED FACE RECOGNITION: 1) ORFE=OCCLUSION ROBUST FEATURE
EXTRACTION, 2) OAFR=OCCLUSION AWARE FACE RECOGNITION, AND 3) ORFR=OCCLUSION RECOVERY FR. TO OUR KNOWLEDGE, THERE ARE NO
RESEARCH WORKS ON DETECTING PARTIAL FACE; THUS, WE MARK IT AS 'NA' (NOT AVAILABLE).

Abbreviation	Application-oriented Purpose	Publication
OFD	Occluded Face Detection	[14], [18], [36], [61], [61], [91], [106], [120], [146], [155], [159], [165], [165]–[167], [174], [192]
ORFE	Patch based engineered features	[1], [2], [12], [23], [57], [65], [88], [111], [122], [180], [181], [193]
	Learning based features	[11], [27], [31]–[33], [53], [65], [73], [74], [90], [93], [96], [107], [109], [124], [125], [133], [134], [136], [148], [153], [162], [163], [169], [187]
OAFR	Occlusion Detection	[18], [156]
	Occlusion Discard Face Recognition	[4], [19], [98], [101], [103], [108], [128], [143]
	Partial Face Detection	NA
	Partial Face Recognition	[46], [48]–[50], [55], [80], [150], [151]
ORFR	Occlusion Recovery	[15]–[17], [20], [21], [29], [51], [54], [59], [64], [77], [83], [114], [139], [140], [157], [161]
	Occlusion Recovery Face Recognition	[24], [26], [31], [35], [60], [62], [73], [78], [89], [112], [147], [183], [185], [191]

TABLE IX
A GLOSSARY OF ABBREVIATIONS AND EXPANSIONS FOR TERMINOLOGIES.

Abbreviation	Expansion	Abbreviation	Expansion
AAE	Adversarial AutoEncoder	LBP	Local Binary Patterns
AFC	Adaptive Feature Convolution	LDA	Linear Discriminant Analysis
AOFD	Adversarial Occlusion-aware Face Detection	LLE	Locally Linear Embedding
CSLBP	Center-Symmetric Local Binary Patterns	LSTM	Long Short Term Memory
DA	Denosing Autoencoder	MDSCNN	Multiscale Double Convolutional Neural Network
DCNNs	Deep Convolutional Neural Networks	MKD	Multi-Keypoint Descriptors
DDRC	Deep Dictionary Representation based Classification	NNAODL	Nuclear Norm based Adapted Occlusion Dictionary Learning
DMSC	Discriminative Multiscale Sparse Coding	NMF	Non-negative Matrix Factorization
DoG	Different of Gaussian filters	OAFR	Occlusion Aware Face Recognition
DPM	Deformable Part Models	OFR	Occluded Face Recognition
DSCNNs	Double Supervision Convolutional Neural Network	ORFE	Occlusion Robust Feature Extraction
EBGM	Elastic Bunch Graph Matching	ORecFR	Occlusion Recovery based Face Recognition
ECSLBP	Enhanced Center-Symmetric Local Binary Patterns	PCA	Principal Component Analysis
ELOC	Efficient Locality-constrained Occlusion Coding	PDSN	Pairwise Differential Siamese Network
ERBF	Ensemble of Radial Basis Function	R-CNN	Regions with Convolutional Neural Networks
ERGAN	Eyeglasses Removal Generative Adversarial Network	RCSLBP	Reinforced Centrosymmetric Local Binary Pattern
FAN	Face Attention Network	RDLRR	Robust Discriminative Low-Rank Representation
FANet	Feature Agglomeration Networks	RLA	Robust LSTM-Autoencoders
FCN	Fully Convolutional Network	RPSM	Robust Point Set Matching
FW-PCA	Fast Weighted-Principal Component Analysis	SIFT	Scale Invariant Feature Transform
GAN	Generative Adversarial Network	SOC	Structured Occlusion Coding
GRRC	Gabor based Robust Representation and Classification	SOM	Self-Organizing Maps
HOG	Histogram of Oriented Gradient	SRC	Sparse Representation Classifiers
ICA	Independent Component Analysis	SSD	Single-Shot multibox Detector
ID-GAN	Identity-Diversity Generative Adversarial Network	SSRC	Structured Sparse Representation Classification
InfoGAN	Information maximizing Generative Adversarial Network	SVM	Support Vector Machine
KCFA	Kernel Correlation Feature Analysis	VAE	Variational AutoEncoder
KLD-LGBP	Kullback-Leibler Divergence-Local Gabor Binary Patterns	YOLO	You Only Look Once

REFERENCES

[1] Timo Ahonen, Abdenour Hadid, and Matti Pietikäinen. Face recognition with local binary patterns. In *European conference on computer vision*, pages 469–481. Springer, 2004.

[2] Timo Ahonen, Abdenour Hadid, and Matti Pietikainen. Face description with local binary patterns: Application to face recognition. *IEEE transactions on pattern analysis and machine intelligence*, 28(12):2037–2041, 2006.

[3] Martínez Aleix and Benavente Robert. The AR face database. *CVC Tech. Rep*, 24, Jun 1998.

[4] A Morelli Andrés, Sebastian Padovani, Mariano Tepper, and Julio Jacobo-Berlles. Face recognition on partially occluded images using compressed sensing. *Pattern Recognition Letters*, 36:235–242, 2014.

[5] Aisha Azeem, Muhammad Sharif, Mudassar Raza, and Marryam Murtaza. A survey: Face recognition techniques under partial occlusion. *Int. Arab J. Inf. Technol.*, 11(1):1–10, 2014.

[6] Yancheng Bai, Yongqiang Zhang, Mingli Ding, and Bernard Ghanem. Finding tiny faces in the wild with generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 21–30, 2018.

[7] Ankan Bansal, Anirudh Nanduri, Carlos D Castillo, Rajeev Ranjan, and Rama Chellappa. Umdfaces: An annotated face dataset for training deep networks. In *2017 IEEE International Joint Conference on Biometrics (IJCB)*, pages 464–473. IEEE, 2017.

[8] Peter N Belhumeur, João P Hespanha, and David J Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. Technical report, Yale University New Haven United States, 1997.

[9] Lacey Best-Rowden and Anil K Jain. Longitudinal study of automatic face recognition. *IEEE transactions on pattern analysis and machine intelligence*, 40(1):148–162, 2018.

[10] Qiong Cao, Li Shen, Weidi Xie, Omkar M Parkhi, and Andrew Zisserman. Vggface2: A dataset for recognising faces across pose and age. In *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, pages 67–74. IEEE, 2018.

[11] Feng Cen and Guanghui Wang. Dictionary representation of deep features for occlusion-robust face recognition. *IEEE Access*, 7:26595–26605, 2019.

[12] Ismahane Cheheb, Noor Al-Maadeed, Somaya Al-Madeed, Ahmed Bouridane, and Richard Jiang. Random sampling for patch-based face recognition. In *Biometrics and Forensics (IWBF), 2017 5th International Workshop on*, pages 1–5. IEEE, 2017.

[13] Dong Chen, Shaoqing Ren, Yichen Wei, Xudong Cao, and Jian Sun. Joint cascade face detection and alignment. In *European conference on computer vision*, pages 109–122. Springer, 2014.

[14] Jie Chen, Shiguang Shan, Shengye Yang, Xilin Chen, and Wen Gao. Modification of the adaboost-based detector for partially occluded faces. In *18th International Conference on Pattern Recognition (ICPR'06)*, volume 2, pages 516–519. IEEE, 2006.

[15] Xi Chen, Yan Duan, Rein Houthooft, John Schulman, Ilya Sutskever, and Pieter Abbeel. Infogan: Interpretable representation learning by information maximizing generative adversarial nets. In *Advances in neural information processing systems*, pages 2172–2180, 2016.

[16] Xiang Chen, Linbo Qing, Xiaohai He, Jie Su, and Yonghong Peng. From eyes to face synthesis: a new approach for human-centered smart surveillance. *IEEE access*, 6:14567–14575, 2018.

[17] Yu-An Chen, Wei-Che Chen, Chia-Po Wei, and Yu-Chiang Frank Wang. Occlusion-aware face inpainting via generative adversarial networks. In *2017 IEEE International Conference on Image Processing (ICIP)*, pages 1202–1206. IEEE, 2017.

[18] Yujia Chen, Lingxiao Song, and Ran He. Masquer hunter: Adversarial occlusion-aware face detection. *arXiv preprint arXiv:1709.05188*, 2017.

[19] Zhaohua Chen, Tingrong Xu, and Zhiyuan Han. Occluded face recognition based on the improved svm and block weighted lbp. In *Image Analysis and Signal Processing (IASP), 2011 International Conference on*, pages 118–122. IEEE, 2011.

[20] Lele Cheng, Jinjun Wang, Yihong Gong, and Qiqi Hou. Robust deep auto-encoder for occluded face recognition. In *Proceedings of the 23rd ACM international conference on Multimedia*, pages 1099–1102. ACM, 2015.

[21] Antonio Criminisi, Patrick Pérez, and Kentaro Toyama. Region filling and object removal by exemplar-based image inpainting. *IEEE Transactions on image processing*, 13(9):1200–1212, 2004.

[22] Nicole Dagnes, Enrico Vezzetti, Federica Marcolin, and Stefano Tornincasa. Occlusion detection and restoration techniques for 3d face recognition: a literature review. *Machine Vision and Applications*, 29:789–813, 2018.

[23] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, volume 1, pages 886–893. IEEE, 2005.

[24] Fernando De La Torre and Michael J Black. A framework for robust subspace learning. *International Journal of Computer Vision*, 54(1-3):117–142, 2003.

[25] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4690–4699, 2019.

[26] Yue Deng, Qionghai Dai, and Zengke Zhang. Graph laplace for occluded face completion and recognition. *IEEE Transactions on Image Processing*, 20(8):2329–2338, 2011.

[27] Lingshuang Du and Haifeng Hu. Nuclear norm based adapted occlusion dictionary learning for face recognition with occlusion and illumination changes. *Neurocomputing*, 340:133–144, 2019.

[28] Hazım Kemal Ekenel and Rainer Stiefelhagen. Why is facial occlusion a challenging problem? In *International Conference on Biometrics*, pages 299–308. Springer, 2009.

[29] Michael Elad and Michal Aharon. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Transactions on Image processing*, 15(12):3736–3745, 2006.

[30] Sachin Sudhakar Farfade, Mohammad J Saberian, and Li-Jia Li. Multiview face detection using deep convolutional neural networks. In *Proceedings of the 5th ACM on International Conference on Multimedia Retrieval*, pages 643–650. ACM, 2015.

[31] Sanja Fidler, Danijel Skocaj, and Ales Leonardis. Combining reconstructive and discriminative subspace methods for robust classification and regression by subsampling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(3):337–350, 2006.

[32] Yuli Fu, Xiaosi Wu, Yandong Wen, and Youjun Xiang. Efficient locality-constrained occlusion coding for face recognition. *Neurocomputing*, 260:104–111, 2017.

[33] Guangwei Gao, Jian Yang, Xiao-Yuan Jing, Fumin Shen, Wankou Yang, and Dong Yue. Learning robust and discriminative low-rank representations for face recognition with occlusion. *Pattern Recognition*, 66:129–143, 2017.

[34] Wen Gao, Bo Cao, Shiguang Shan, Xilin Chen, Delong Zhou, Xiaohua Zhang, and Debin Zhao. The cas-peal large-scale chinese face database and baseline evaluations. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 38(1):149–161, 2008.

[35] Shiming Ge, Chenyu Li, Shengwei Zhao, and Dan Zeng. Occluded face recognition in the wild by identity-diversity inpainting. *IEEE Transactions on Circuits and Systems for Video Technology*, 2020.

[36] Shiming Ge, Jia Li, Qiting Ye, and Zhao Luo. Detecting masked faces in the wild with lle-cnns. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2682–2690, 2017.

[37] Athinodoros S Georghiades, Peter N Belhumeur, and David J Kriegman. From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (6):643–660, 2001.

[38] Ross Girshick. Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 1440–1448, 2015.

[39] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 580–587, 2014.

[40] Ian Goodfellow. Nips 2016 tutorial: Generative adversarial networks. *arXiv preprint arXiv:1701.00160*, 2016.

[41] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.

[42] Rajkiran Gottumukkal and Vijayan K Asari. An improved face recognition technique based on modular pca approach. *Pattern Recognition Letters*, 25(4):429–436, 2004.

[43] Ralph Gross, Iain Matthews, Jeffrey Cohn, Takeo Kanade, and Simon Baker. Multi-pie. *Image and Vision Computing*, 28(5):807–813, 2010.

[44] Yandong Guo, Lei Zhang, Yuxiao Hu, Xiaodong He, and Jianfeng Gao. Ms-celeb-1m: A dataset and benchmark for large-scale face recognition. In *European Conference on Computer Vision*, pages 87–102. Springer, 2016.

[45] Zhixin Guo, Wenzhi Liao, Peter Veelaert, and Wilfried Philips. Occlusion-robust detector trained with occluded pedestrians. SCITEPRESS-Science and Technology Publications, 2018.

[46] Srinivas Gutta, Vasanth Philomin, and Miroslav Trajkovic. An investigation into the use of partial-faces for face recognition. In *fgr*, pages 33–38, 2002.

[47] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[48] Lingxiao He, Haiqing Li, Qi Zhang, and Zhenan Sun. Dynamic feature learning for partial face recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7054–7063, 2018.

[49] Lingxiao He, Haiqing Li, Qi Zhang, and Zhenan Sun. Dynamic feature matching for partial face recognition. *IEEE Transactions on Image Processing*, 28(2):791–802, 2019.

[50] Lingxiao He, Haiqing Li, Qi Zhang, Zhenan Sun, and Zhaofeng He. Multiscale representation for partial face recognition under near infrared illumination. In *Biometrics Theory, Applications and Systems (BTAS), 2016 IEEE 8th International Conference on*, pages 1–7. IEEE, 2016.

[51] Zhenliang He, Wangmeng Zuo, Meina Kan, Shiguang Shan, and Xilin Chen. Attgan: Facial attribute editing by only changing what you want. *IEEE Transactions on Image Processing*, 28(11):5464–5478, 2019.

[52] Jan Hosang, Mohamed Omran, Rodrigo Benenson, and Bernt Schiele. Taking a deeper look at pedestrians. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4073–4082, 2015.

[53] Kazuhiro Hotta. Robust face recognition under partial occlusion based on support vector machine with local gaussian summation kernel. *Image and Vision Computing*, 26(11):1490–1498, 2008.

[54] Bingwen Hu, Wankou Yang, and Mingwu Ren. Unsupervised eyeglasses removal in the wild. *arXiv preprint arXiv:1909.06989*, 2019.

[55] Junlin Hu, Jiwen Lu, and Yap-Peng Tan. Robust partial face recognition using instance-to-class distance. In *Visual Communications and Image Processing (VCIP), 2013*, pages 1–6. IEEE, 2013.

[56] Peiyun Hu and Deva Ramanan. Finding tiny faces. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 951–959, 2017.

[57] Gang Hua and Amir Akbarzadeh. A robust elastic and partial matching metric for face recognition. In *Computer Vision, 2009 IEEE 12th International Conference on*, pages 2082–2089. IEEE, 2009.

[58] Gary B. Huang, Manu Ramesh, Tamara Berg, and Erik Learned-Miller. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical Report 07-49, University of Massachusetts, Amherst, October 2007.

[59] Satoshi Iizuka, Edgar Simo-Serra, and Hiroshi Ishikawa. Globally and locally consistent image completion. *ACM Transactions on Graphics (ToG)*, 36(4):1–14, 2017.

[60] Michael Iliadis, Haohong Wang, Rafael Molina, and Aggelos K Katsaggelos. Robust and low-rank representation for fast face identification with occlusions. *IEEE Transactions on Image Processing*, 26(5):2203–2218, 2017.

[61] Vidit Jain and Erik Learned-Miller. Fddb: A benchmark for face detection in unconstrained settings. Technical report, UMass Amherst Technical Report, 2010.

[62] Hongjun Jia and Aleix M Martinez. Face recognition with occlusions in the training and testing sets. In *2008 8th IEEE International Conference on Automatic Face & Gesture Recognition*, pages 1–6. IEEE, 2008.

[63] Huaizu Jiang and Erik Learned-Miller. Face detection with the faster r-cnn. In *2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)*, pages 650–657. IEEE, 2017.

[64] Sandesh V Khamele and Shyamal G Mundada. An approach for restoring occluded images for face-recognition. *Int J Adv Res Comput Commun Eng*, 4(17):571–575, 2015.

[65] Jongsun Kim, Jongmoo Choi, Juneho Yi, and Matthew Turk. Effective representation using ica for face recognition robust to local distortion and partial occlusion. *IEEE transactions on pattern analysis and machine intelligence*, 27(12):1977–1981, 2005.

[66] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

[67] Brendan F Klare, Ben Klein, Emma Taborsky, Austin Blanton, Jordan Cheney, Kristen Allen, Patrick Grother, Alan Mah, and Anil K Jain. Pushing the frontiers of unconstrained face detection and recognition: Iarpa janus benchmark a. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1931–1939, 2015.

[68] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[69] Neeraj Kumar, Alexander C Berg, Peter N Belhumeur, and Shree K Nayar. Attribute and simile classifiers for face verification. In *2009 IEEE 12th International Conference on Computer Vision*, pages 365–372. IEEE, 2009.

[70] Badr Lahasan, Syaheerah Lebai Lutfi, and Rubén San-Segundo. A survey on techniques to handle face recognition challenges: occlusion, single sample per subject and expression. *Artificial Intelligence Review*, pages 1–31, 2017.

[71] Erik Learned-Miller, Gary B Huang, Aruni RoyChowdhury, Haoxiang Li, and Gang Hua. Labeled faces in the wild: A survey. In *Advances in face detection and facial image analysis*, pages 189–248. Springer, 2016.

[72] Daniel D Lee and H Sebastian Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788, 1999.

[73] Aleš Leonardis and Horst Bischof. Robust recognition using eigenimages. *Computer Vision and Image Understanding*, 78(1):99–118, 2000.

[74] Chen Li, Shuai Zhao, Wei Song, Ke Xiao, and Yanjie Wang. Ubiquitous single-sample face recognition under occlusion based on sparse representation with dual features. *Journal of Ambient Intelligence and Humanized Computing*, pages 1–11, 2017.

[75] Haoxiang Li, Zhe Lin, Xiaohui Shen, Jonathan Brandt, and Gang Hua. A convolutional neural network cascade for face detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5325–5334, 2015.

[76] Stan Z Li, Xin Wen Hou, Hong Jiang Zhang, and Qian Sheng Cheng. Learning spatially localized, parts-based representation. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, volume 1, pages I–I. IEEE, 2001.

[77] Yijun Li, Sifei Liu, Jimei Yang, and Ming-Hsuan Yang. Generative face completion. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3911–3919, 2017.

[78] Yuelong Li and Jufu Feng. Reconstruction based face occlusion elimination for recognition. *Neurocomputing*, 101:68–72, 2013.

[79] Zhihang Li, Xu Tang, Junyu Han, Jingtuo Liu, and Ran He. Pyramidbox++: High performance detector for finding tiny face. *arXiv preprint arXiv:1904.00386*, 2019.

[80] Shengcai Liao, Anil K Jain, and Stan Z Li. Partial face recognition: Alignment-free approach. *IEEE Transactions on pattern analysis and machine intelligence*, 35(5):1193–1205, 2013.

[81] Shengcai Liao, Xiangxin Zhu, Zhen Lei, Lun Zhang, and Stan Z Li. Learning multi-scale block local binary patterns for face recognition. In *International Conference on Biometrics*, pages 828–837. Springer, 2007.

[82] Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. Feature pyramid networks for object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2117–2125, 2017.

[83] Guilin Liu, Fitsum A Reda, Kevin J Shih, Ting-Chun Wang, Andrew Tao, and Bryan Catanzaro. Image inpainting for irregular holes using partial convolutions. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 85–100, 2018.

[84] Hao Liu, Huiping Duan, Hongyu Cui, and Yunjie Yin. Face recognition using training data with artificial occlusions. In *2016 Visual Communications and Image Processing (VCIP)*, pages 1–4. IEEE, 2016.

[85] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C Berg. Ssd: Single shot multibox detector. In *European conference on computer vision*, pages 21–37. Springer, 2016.

[86] Weiyang Liu, Yandong Wen, Zhiding Yu, Ming Li, Bhiksha Raj, and Le Song. Sphereface: Deep hypersphere embedding for face recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 212–220, 2017.

[87] Weiyang Liu, Yandong Wen, Zhiding Yu, and Meng Yang. Largemargin softmax loss for convolutional neural networks. In *ICML*, volume 2, page 7, 2016.

[88] David G Lowe. Object recognition from local scale-invariant features. In *Proceedings of the seventh IEEE international conference on computer vision*, volume 2, pages 1150–1157. IEEE, 1999.

[89] Xiao Luan, Bin Fang, Linghui Liu, Weibin Yang, and Jiye Qian. Extracting sparse error of robust pca for face recognition in the presence of varying illumination and occlusion. *Pattern Recognition*, 47(2):495–508, 2014.

[90] Jiang-Jing Lv, Xiao-Hu Shao, Jia-Shui Huang, Xiang-Dong Zhou, and Xi Zhou. Data augmentation for face recognition. *Neurocomputing*, 230:184–196, 2017.

[91] Upal Mahbub, Vishal M Patel, Deepak Chandra, Brandon Barbello, and Rama Chellappa. Partial face detection for continuous authentication. In *Image Processing (ICIP), 2016 IEEE International Conference on*, pages 2991–2995. IEEE, 2016.

[92] Alireza Makhzani, Jonathon Shlens, Navdeep Jaitly, Ian Goodfellow, and Brendan Frey. Adversarial autoencoders. *arXiv preprint arXiv:1511.05644*, 2015.

[93] Aleix M Martínez. Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (6):748–763, 2002.

[94] Iacopo Masi, Yue Wu, Tal Hassner, and Prem Natarajan. Deep face recognition: A survey. In *2018 31st SIBGRAPI conference on graphics, patterns and images (SIBGRAPI)*, pages 471–478. IEEE, 2018.

[95] Brianna Maze, Jocelyn Adams, James A Duncan, Nathan Kalka, Tim Miller, Charles Otto, Anil K Jain, W Tyler Niggel, Janet Anderson, Jordan Cheney, et al. Iarpa janus benchmark-c: Face dataset and protocol. In *2018 International Conference on Biometrics (ICB)*, pages 158–165. IEEE, 2018.

[96] Niall McLaughlin, Ji Ming, and Danny Crookes. Largest matching areas for illumination and occlusion robust face recognition. *IEEE transactions on cybernetics*, 47(3):796–808, 2017.

[97] Mostafa Mehdipour Ghazi and Hazim Kemal Ekenel. A comprehensive analysis of deep learning based representation for face recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 34–41, 2016.

[98] Rui Min, Abdenour Hadid, and Jean-Luc Dugelay. Improving the recognition of faces occluded by facial accessories. In *Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on*, pages 442–447. IEEE, 2011.

[99] Aaron Nech and Ira Kemelmacher-Shlizerman. Level playing field for million scale face recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7044–7053, 2017.

[100] Han Foon Neo, Andrew Beng Jin Teoh, and David Chek Ling Ngo. A novel spatially confined non-negative matrix factorization for face recognition. In *MVA*, pages 502–505. Citeseer, 2005.

[101] HF Neo, CC Teo, and Andrew BJ Teoh. Development of partial face recognition framework. In *Computer Graphics, Imaging and Visualization (CGIV), 2010 Seventh International Conference on*, pages 142–146. IEEE, 2010.

[102] University of Notre Dame. University of notre dame biometrics database, 2008. http://www.nd.edu/~cvrl.

[103] Hyun Jun Oh, Kyoung Mu Lee, and Sang Uk Lee. Occlusion invariant face recognition using selective local non-negative matrix factorization basis images. *Image and Vision computing*, 26(11):1515–1523, 2008.

[104] Muhtahir O Oloyede, Gerhard P Hancke, and Nellore Kapileswar. Evaluating the effect of occlusion in face recognition systems. In *AFRICON, 2017 IEEE*, pages 1547–1551. IEEE, 2017.

[105] Aaron van den Oord, Nal Kalchbrenner, and Koray Kavukcuoglu. Pixel recurrent neural networks. *arXiv preprint arXiv:1601.06759*, 2016.

[106] Michael Opitz, Georg Waltner, Georg Poier, Horst Possegger, and Horst Bischof. Grid loss: Detecting occluded faces. In *European conference on computer vision*, pages 386–402. Springer, 2016.

[107] Elad Osherov and Michael Lindenbaum. Increasing cnn robustness to occlusions by reducing filter support. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 550–561, 2017.

[108] Weihua Ou, Xiao Luan, Jianping Gou, Quan Zhou, Wenjun Xiao, Xiangguang Xiong, and Wu Zeng. Robust discriminative nonnegative dictionary learning for occluded face recognition. *Pattern Recognition Letters*, 107:41–49, 2018.

[109] Weihua Ou, Xinge You, Dacheng Tao, Pengyue Zhang, Yuanyan Tang, and Ziqi Zhu. Robust face recognition via occlusion dictionary learning. *Pattern Recognition*, 47(4):1559–1572, 2014.

[110] S. A. Rizvi P. J. Phillips, H. Moon and P. J. RaussD. The feret evaluation methodology for face-recognition algorithms. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (10):1090–1104, 2000.

[111] Ke Pan, Shengcai Liao, Zhijian Zhang, Stan Z Li, and Peiren Zhang. Part-based face recognition using near infrared images. In *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–6. IEEE, 2007.

[112] Jeong-Seon Park, You Hwa Oh, Sang Chul Ahn, and Seong-Whan Lee. Glasses removal from facial image using recursive error compensation. *IEEE transactions on pattern analysis and machine intelligence*, 27(5):805–811, 2005.

[113] Omkar M Parkhi, Andrea Vedaldi, Andrew Zisserman, et al. Deep face recognition. In *bmvc*, volume 1, page 6, 2015.

[114] Deepak Pathak, Philipp Krahenbuhl, Jeff Donahue, Trevor Darrell, and Alexei A Efros. Context encoders: Feature learning by inpainting. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2536–2544, 2016.

[115] P Jonathon Phillips, Patrick J Flynn, Todd Scruggs, Kevin W Bowyer, Jin Chang, Kevin Hoffman, Joe Marques, Jaesik Min, and William Worek. Overview of the face recognition grand challenge. In *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, volume 1, pages 947–954. IEEE, 2005.

[116] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*, 2015.

[117] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 779–788, 2016.

[118] Min Ren, Yunlong Wang, Zhenan Sun, and Tieniu Tan. Dynamic graph representation for partially occluded biometrics. *arXiv preprint arXiv:1912.00377*, 2019.

[119] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems*, pages 91–99, 2015.

[120] Pouya Samangouei, Vishal M Patel, and Rama Chellappa. Attribute-based continuous user authentication on mobile devices. In *2015 IEEE 7th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, pages 1–8. IEEE, 2015.

[121] Ferdinando S Samaria and Andy C Harter. Parameterisation of a stochastic model for human face identification. In *Proceedings of 1994 IEEE Workshop on Applications of Computer Vision*, pages 138–142. IEEE, 1994.

[122] Marios Savvides, Ramzi Abiantun, Jingu Heo, S Park, Chunyan Xie, and BVK Vijayakumar. Partial & holistic face recognition on frgc-ii data using support vector machine. In *Computer Vision and Pattern Recognition Workshop, 2006. CVPRW'06. Conference on*, pages 48–48. IEEE, 2006.

[123] Florian Schroff, Dmitry Kalenichenko, and James Philbin. Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 815–823, 2015.

[124] Jeongin Seo and Hyeyoung Park. A robust face recognition through statistical learning of local features. In *International Conference on Neural Information Processing*, pages 335–341. Springer, 2011.

[125] Jeongin Seo and Hyeyoung Park. Robust recognition of face with partial variations using local features and statistical learning. *Neurocomputing*, 129:41–48, 2014.

[126] Pierre Sermanet, Koray Kavukcuoglu, Soumith Chintala, and Yann LeCun. Pedestrian detection with unsupervised multi-stage feature learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3626–3633, 2013.

[127] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[128] Lingxue Song, Dihong Gong, Zhifeng Li, Changsong Liu, and Wei Liu. Occlusion robust face recognition based on mask learning with pairwise differential siamese network. *arXiv preprint arXiv:1908.06290*, 2019.

[129] Xudong Sun, Pengcheng Wu, and Steven CH Hoi. Face detection using deep learning: An improved faster rcnn approach. *Neurocomputing*, 299:42–50, 2018.

[130] Yi Sun, Yuheng Chen, Xiaogang Wang, and Xiaoou Tang. Deep learning face representation by joint identification-verification. In *Advances in neural information processing systems*, pages 1988–1996, 2014.

[131] Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, and Alexander A Alemi. Inception-v4, inception-resnet and the impact of residual connections on learning. In *Thirty-first AAAI conference on artificial intelligence*, 2017.

[132] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.

[133] Xiaoyang Tan, Songcan Chen, Zhi-Hua Zhou, and Jun Liu. Face recognition under occlusions and variant expressions with partial similarity. *IEEE Transactions on information forensics and security*, 4(2):217–230, 2009.

[134] Xiaoyang Tan, Songcan Chen, Zhi-Hua Zhou, and Fuyan Zhang. Recognizing partially occluded, expression variant faces from single training image per person with som and soft k-nn ensemble. *IEEE Transactions on Neural Networks*, 16(4):875–886, 2005.

[135] Francesc Tarrés, Antonio Rama, and L Torres. A novel method for face recognition under partial occlusion or facial expression variations. In *Proc. 47th Intl Symp. ELMAR*, pages 163–166, 2005.

[136] Daniel Sáez Trigueros, Li Meng, and Margaret Hartnett. Enhancing convolutional neural networks for face recognition with occlusion maps and batch triplet loss. *Image and Vision Computing*, 79:99–108, 2018.

[137] Daniel Sáez Trigueros, Li Meng, and Margaret Hartnett. Face recognition: From traditional to deep learning methods. *arXiv preprint arXiv:1811.00116*, 2018.

[138] Matthew Turk and Alex Pentland. Eigenfaces for recognition. *Journal of cognitive neuroscience*, 3(1):71–86, 1991.

[139] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Deep image prior. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 9446–9454, 2018.

[140] A Vijayalakshmi. Recognizing faces with partial occlusion using inpainting. *International Journal of Computer Applications*, 168(13), 2017.

[141] Paul Viola and Michael Jones. Rapid object detection using a boosted cascade of simple features. In *Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001*, volume 1, pages I–I. IEEE, 2001.

[142] Paul Viola and Michael J Jones. Robust real-time face detection. *International journal of computer vision*, 57(2):137–154, 2004.

[143] Weitao Wan and Jiansheng Chen. Occlusion robust face recognition based on mask learning. In *2017 IEEE International Conference on Image Processing (ICIP)*, pages 3795–3799. IEEE, 2017.

[144] Feng Wang, Jian Cheng, Weiyang Liu, and Haijun Liu. Additive margin softmax for face verification. *IEEE Signal Processing Letters*, 25(7):926–930, 2018.

[145] Hao Wang, Yitong Wang, Zheng Zhou, Xing Ji, Dihong Gong, Jingchao Zhou, Zhifeng Li, and Wei Liu. Cosface: Large margin cosine loss for deep face recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5265–5274, 2018.

[146] Jianfeng Wang, Ye Yuan, and Gang Yu. Face attention network: an effective face detector for the occluded faces. *arXiv preprint arXiv:1711.07246*, 2017.

[147] Meng Wang, Zhengping Hu, Zhe Sun, Shuhuan Zhao, and Mei Sun. Varying face occlusion detection and iterative recovery for face recognition. *Journal of Electronic Imaging*, 26(3):033009, 2017.

[148] Yandong Wen, Weiyang Liu, Meng Yang, Yuli Fu, Youjun Xiang, and Rui Hu. Structured occlusion coding for robust face recognition. *Neurocomputing*, 178:11–24, 2016.

[149] Yandong Wen, Kaipeng Zhang, Zhifeng Li, and Yu Qiao. A discriminative feature learning approach for deep face recognition. In *European conference on computer vision*, pages 499–515. Springer, 2016.

[150] Renliang Weng, Jiwen Lu, Junlin Hu, Gao Yang, and Yap-Peng Tan. Robust feature set matching for partial face recognition. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 601–608, 2013.

[151] Renliang Weng, Jiwen Lu, and Yap-Peng Tan. Robust point set matching for partial face recognition. *IEEE transactions on image processing*, 25(3):1163–1176, 2016.

[152] Laurenz Wiskott, Norbert Krüger, N Kuiger, and Christoph Von Der Malsburg. Face recognition by elastic bunch graph matching. *IEEE Transactions on pattern analysis and machine intelligence*, 19(7):775–779, 1997.

[153] John Wright, Allen Y Yang, Arvind Ganesh, S Shankar Sastry, and Yi Ma. Robust face recognition via sparse representation. *IEEE transactions on pattern analysis and machine intelligence*, 31(2):210–227, 2009.

[154] Cho Ying Wu and Jian Jiun Ding. Occluded face recognition using low-rank regression with generalized gradient direction. *Pattern Recognition*, 80:256–268, 2018.

[155] Shuzhe Wu, Meina Kan, Shiguang Shan, and Xilin Chen. Hierarchical attention for part-aware face detection. *International Journal of Computer Vision*, pages 1–19, 2019.

[156] Yizhang Xia, Bailing Zhang, and Frans Coenen. Face occlusion detection based on multi-task convolution neural network. In *Fuzzy Systems and Knowledge Discovery (FSKD), 2015 12th International Conference on*, pages 375–379. IEEE, 2015.

[157] Junyuan Xie, Linli Xu, and Enhong Chen. Image denoising and inpainting with deep neural networks. In *Advances in neural information processing systems*, pages 341–349, 2012.

[158] Junjie Yan, Zhen Lei, Longyin Wen, and Stan Z Li. The fastest deformable part model for object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2497–2504, 2014.

[159] Junjie Yan, Xuzong Zhang, Zhen Lei, and Stan Z Li. Face detection by structural models. *Image and Vision Computing*, 32(10):790–799, 2014.

[160] Bin Yang, Junjie Yan, Zhen Lei, and Stan Z Li. Aggregate channel features for multi-view face detection. In *IEEE international joint conference on biometrics*, pages 1–8. IEEE, 2014.

[161] Chao Yang, Xin Lu, Zhe Lin, Eli Shechtman, Oliver Wang, and Hao Li. High-resolution image inpainting using multi-scale neural patch synthesis. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6721–6729, 2017.

[162] Meng Yang, Xing Wang, Guohang Zeng, and Linlin Shen. Joint and collaborative representation with local adaptive convolution feature for face recognition with single sample per person. *Pattern recognition*, 66:117–128, 2017.

[163] Meng Yang, Lei Zhang, Simon CK Shiu, and David Zhang. Gabor feature based robust representation and classification for face recognition with gabor occlusion dictionary. *Pattern Recognition*, 46(7):1865–1878, 2013.

[164] Ming-Hsuan Yang. Kernel eigenfaces vs. kernel fisherfaces: Face recognition using kernel methods. In *fgr*, page 0215. IEEE, 2002.

[165] Shuo Yang, Ping Luo, Chen-Change Loy, and Xiaoou Tang. From facial parts responses to face detection: A deep learning approach. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3676–3684, 2015.

[166] Shuo Yang, Ping Luo, Chen-Change Loy, and Xiaoou Tang. Wider face: A face detection benchmark. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5525–5533, 2016.

[167] Shuo Yang, Ping Luo, Chen Change Loy, and Xiaoou Tang. Facenessnet: Face detection through deep facial part responses. *IEEE transactions on pattern analysis and machine intelligence*, 40(8):1845–1859, 2018.

[168] Dong Yi, Zhen Lei, Shengcai Liao, and Stan Z Li. Learning face representation from scratch. *arXiv preprint arXiv:1411.7923*, 2014.

[169] Yu-Feng Yu, Dao-Qing Dai, Chuan-Xian Ren, and Ke-Kun Huang. Discriminative multi-scale sparse coding for single-sample face recognition with occlusion. *Pattern Recognition*, 66:302–312, 2017.

[170] Stefanos Zafeiriou, Cha Zhang, and Zhengyou Zhang. A survey on face detection in the wild: past, present and future. *Computer Vision and Image Understanding*, 138:1–24, 2015.

[171] Baochang Zhang, Shiguang Shan, Xilin Chen, and Wen Gao. Histogram of gabor phase patterns (hgpp): A novel object representation approach for face recognition. *IEEE Transactions on Image processing*, 16(1):57–68, 2006.

[172] Cha Zhang and Zhengyou Zhang. Improving multiview face detection with multi-task deep convolutional neural networks. In *IEEE Winter Conference on Applications of Computer Vision*, pages 1036–1041. IEEE, 2014.

[173] Guangpeng Zhang and Yunhong Wang. Faceprint: Fusion of local features for 3d face recognition. In *International Conference on Biometrics*, pages 394–403. Springer, 2009.

[174] Heng Zhang, Vishal M Patel, Mohammed Fathy, and Rama Chellappa. Touch gesture-based active user authentication using dictionaries. In *2015 IEEE Winter Conference on Applications of Computer Vision*, pages 207–214. IEEE, 2015.

[175] Jialiang Zhang, Xiongwei Wu, Steven CH Hoi, and Jianke Zhu. Feature agglomeration networks for single stage face detection. *Neurocomputing*, 380:180–189, 2020.

[176] Kaipeng Zhang, Zhanpeng Zhang, Zhifeng Li, and Yu Qiao. Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Processing Letters*, 23(10):1499–1503, 2016.

[177] Qi Zhang, Haiqing Li, Zhenan Sun, and Tieniu Tan. Deep feature fusion for iris and periocular biometrics on mobile devices. *IEEE Transactions on Information Forensics and Security*, 13(11):2897–2912, 2018.

[178] Shifeng Zhang, Longyin Wen, Xiao Bian, Zhen Lei, and Stan Z Li. Occlusion-aware r-cnn: Detecting pedestrians in a crowd. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 637–653, 2018.

[179] Shifeng Zhang, Xiangyu Zhu, Zhen Lei, Hailin Shi, Xiaobo Wang, and Stan Z Li. S3fd: Single shot scale-invariant face detector. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 192–201, 2017.

[180] Wenchao Zhang, Shiguang Shan, Xilin Chen, and Wen Gao. Local gabor binary patterns based on kullback–leibler divergence for partially

occluded face recognition. *IEEE signal processing letters*, 14(11):875–878, 2007.

[181] Wenchao Zhang, Shiguang Shan, Wen Gao, Xilin Chen, and Hongming Zhang. Local gabor binary pattern histogram sequence (lgbphs): a novel non-statistical model for face representation and recognition. In *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*, volume 1, pages 786–791. IEEE, 2005.

[182] Zhishuai Zhang, Wei Shen, Siyuan Qiao, Yan Wang, Bo Wang, and Alan Yuille. Robust face detection via learning small faces on hard images. In *The IEEE Winter Conference on Applications of Computer Vision*, pages 1361–1370, 2020.

[183] Fang Zhao, Jiashi Feng, Jian Zhao, Wenhan Yang, and Shuicheng Yan. Robust lstm-autoencoders for face de-occlusion in the wild. *IEEE Transactions on Image Processing*, 27(2):778–790, 2018.

[184] Guoying Zhao and Matti Pietikainen. Dynamic texture recognition using local binary patterns with an application to facial expressions. *IEEE transactions on pattern analysis and machine intelligence*, 29(6):915–928, 2007.

[185] Shuhuan Zhao and Zheng-ping Hu. A modular weighted sparse representation based on fisher discriminant and sparse residual for face recognition with occlusion. *Information Processing Letters*, 115(9):677–683, 2015.

[186] Wei-Shi Zheng, Xiang Li, Tao Xiang, Shengcai Liao, Jianhuang Lai, and Shaogang Gong. Partial person re-identification. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 4678–4686, 2015.

[187] Wenbo Zheng, Chao Gou, and Fei-Yue Wang. A novel approach inspired by optic nerve characteristics for few-shot occluded face recognition. *Neurocomputing*, 376:25–41, 2020.

[188] Wenming Zheng, Cairong Zou, and Li Zhao. Real-time face recognition using gram-schmidt orthogonalization for lda. In *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, volume 2, pages 403–406. IEEE, 2004.

[189] Chunluan Zhou and Junsong Yuan. Bi-box regression for pedestrian detection and occlusion estimation. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 135–151, 2018.

[190] Erjin Zhou, Zhimin Cao, and Qi Yin. Naive-deep face recognition: Touching the limit of lfw benchmark or not? *arXiv preprint arXiv:1501.04690*, 2015.

[191] Zihan Zhou, Andrew Wagner, Hossein Mobahi, John Wright, and Yi Ma. Face recognition with contiguous occlusion using markov random fields. In *2009 IEEE 12th international conference on computer vision*, pages 1050–1057. IEEE, 2009.

[192] Xiangxin Zhu and Deva Ramanan. Face detection, pose estimation, and landmark estimation in the wild. CVPR, 2012.

[193] Jie Zou, Qiang Ji, and George Nagy. A comparative study of local matching approach for face recognition. *IEEE Transactions on image processing*, 16(10):2617–2628, 2007.