Check for updates

Received: 24 July 2020 | Revised: 7 October 2020 | Accepted: 10 November 2020

IET Biometrics

DOI: 10.1049/bme2.12029



REVIEW

# A survey of face recognition techniques under occlusion

Dan Zeng[1,2] | Raymond Veldhuis[1] | Luuk Spreeuwers[1]

[1]Faculty of EEMCS, University of Twente, Enschede, The Netherlands

[2]Southern University of Science and Technology, Shenzhen, China

**Correspondence**

Raymond Veldhuis, University of Twente, Faculty of Electrical Engineering, Mathematics and Computer Science, Zilverling 4067, P.O. Box 217, 7500 AE Enschede, The Netherlands.
Email: r.n.j.veldhuis@utwente.nl

**Funding information**

University of Twente spinoff company 20Face

**Abstract**

The limited capacity to recognise faces under occlusions is a long-standing problem that presents a unique challenge for face recognition systems and even 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 article, the scope to occluded face recognition is restricted and a systematic categorisation that new as well as classic methods fit into is presented. First, the authors explore the kind of the occlusion problem and the type of inherent difficulties that can arise. As a part of this review, face detection under occlusion, a preliminary step in face recognition. Second the authors analyse how the existing face recognition methods cope with the occlusion problem and classify them into three categories, which are given as: 1) occlusion robust feature extraction approaches, 2) occlusion aware face recognition approaches, and 3) occlusion recovery based face recognition approaches. Furthermore, the motivations, innovations, pros and cons, and the performance of representative approaches for comparison are analyzed. Finally, future challenges and method trends of occluded face recognition are thoroughly discussed.

## 1 | INTRODUCTION

Face recognition is a computer vision task that has been extensively studied for several decades [1]. Compared with other popular biometrics such as fingerprint, iris, palm, and vein, the face has a significantly better potential to recognise 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 [2–10] and the publicly available large-scale face datasets [11–16], face recognition performance has improved substantially [17, 18]. 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 induce a mismatch between the features available in probe faces and gallery faces.

Facial occlusion [19, 20] 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 Figure 1.

Table 1 presents a categorisation of occlusion challenges in different scenarios with their typical occlusion examples. Note that the self-occlusion problem due to pose variation in Table 1 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

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. *IET Biometrics* published by John Wiley & Sons Ltd on behalf of The Institution of Engineering and Technology.



**FIGURE 1** Examples of occluded face images from the MAFA dataset

**TABLE 1** A categorisation 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 |

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. Figure 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 obscured 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.

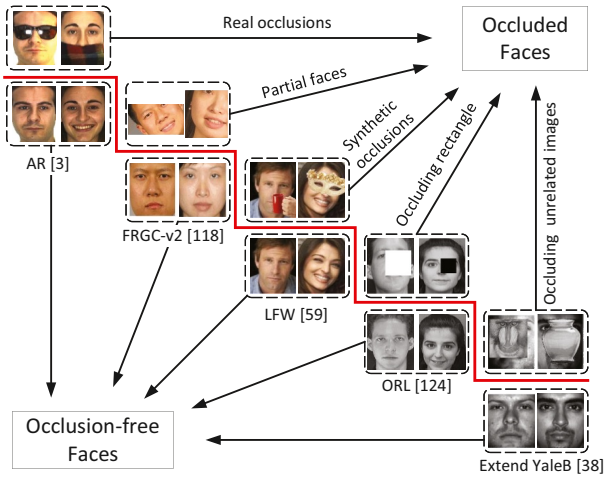


**FIGURE 2** Different occluded face recognition testing scenarios involved in occluded face recognition

Approaches for recognising faces under occlusions can be broadly classified into three categories (shown in Figure 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, discarding occluded parts 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

20474946, 2021, 6, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/bme2.12029 by Universitaet Duisburg-Essen, Wiley Online Library on [26/05/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

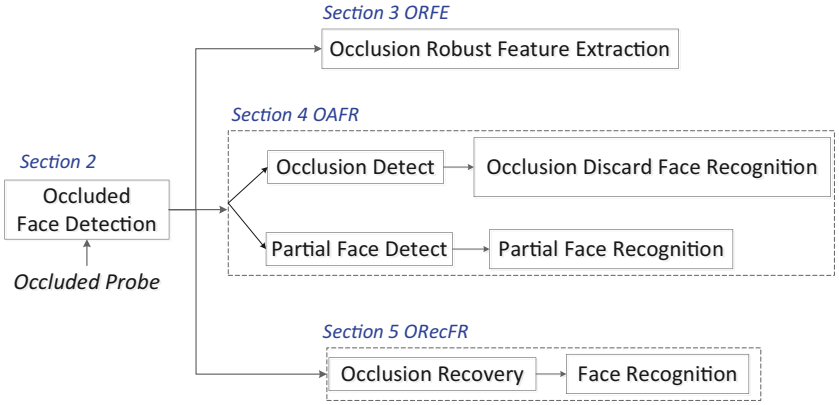


**FIGURE 3** The three categories of methods for face recognition under occlusion challenges

approaches as OAFR because they exclude occluded 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.

Numerous methods have been proposed to push the frontier of face recognition research. Several comprehensive surveys [21–23] have been published for face recognition under occlusion. Of the existing works, the survey by Lahasan et al. [23] on face recognition under occlusion challenges presents a thorough overview of the approaches before 2017, which is the most relevant to this article. 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 [24]. Recently, the MAFA dataset [25] is accessible for occluded face detection, and the IJB-C dataset [26] 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 and novel categorisation of methods for occluded 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. Despite that classic methods have been reviewed before, we deem it still valuable to relate classic methods and new methods in a new and high-level framework. Even in times of DNNs the older work is till valuable and illustrative for certain approaches.

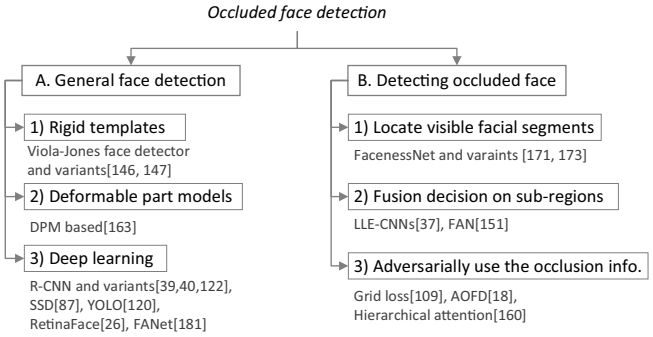


**FIGURE 4** Categorisation of methods used in occluded face detection

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 article is organised as follows: occluded face detection techniques are introduced in Section 2. Methods of occlusion robust feature extraction are described and analysed 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.

## 2 | 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 Figure 4.

20474946, 2021, 6, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/bme2.12029 by Universitaet Duisburg-Essen, Wiley Online Library on [26/05/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

## 2.1 | General face detection

Face detection generally intends to detect the face that is captured in an unconstrained environment. It is challenging due to the large-pose variation, varying illumination, low resolution, and occlusion etc. [27]. Approaches to general face detection can roughly be classified into three categories, which are given as: (1) Rigid templates-based face detection, (2) Deformable part models (DPM)-based face detection, and (3) Deep convolutional neural networks (DCNNs)-based face detection.

The Viola-Jones face detector and its variations [28, 29] are typical in the rigid templates-based category, which utilises 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 [30]. In contrast, DPM based face detection can achieve significantly better performance at the cost of a higher computational complexity [31] compared to Viola-Jones face detector [28].

A third, most promising category of research is DCNNs based face detection [32–42]. Some methods [33, 39, 40] join 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 [43], fast R-CNN [44], faster R-CNN [45]; (ii) region proposals free CNN (i.e.,one-stage detectors), such as the Single-Shot Multibox Detector (SSD) [46], YOLO [47]. 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 [34, 37, 41, 48] 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 a superior performance was achieved [32, 36, 42, 49].

Recent years have witnessed promising results of exploring DCNNs for face detection with the introduction of the Widerface benchmark [38], which offers a wide pose variation, significant scale difference (tiny face), expression variation, make-up, severe illumination, and occlusion. RetinaFace [50] is a single-stage pixel-wise face localization method, which employs extra-supervised and self-supervised multi-task learning and achieves 92.5% on the Widerface hard test set. Feature Agglomeration Networks (FANet) [48], a single-stage face detector, achieves the state-of-art performance on several face detection benchmarks include the FDDB [51], the PASCAL Face [52], and the Widerface benchmark [38]. 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 augmented 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.**

## 2.2 | 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 [53]. 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 [54–56] 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 [25] 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.

### 2.2.1 | 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 [57] 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 analysing 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 [51], PASCAL Face [52] and AFW [58] demonstrate that the proposed method is capable of yielding a good performance. In particular, they can achieve a high recall rate of 90.99% on FDDB. In short, Ref. [57] 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 [59] 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 [38], which is challenging in terms of severe occlusion and unconstrained pose variations. However, it requires the use of labelled attributes of facial data to train attribute-aware CNN, that impairs its practical use in some way.

20474946, 2021, 6, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/bme2.12029 by Universitaet Duisburg-Essen, Wiley Online Library on [26/05/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

### 2.2.2 | Fusing detection results obtained from face sub-regions

In paper [60], 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 14 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) [61, 62]. However, the use of simple architecture increases speed by compromising detection accuracy.

The introduction of MAFA [25] 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 [25] 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 [63] 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 [64]. 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 [38] with an 88.5% average precision on the hard subset as well as on an 88.3% average precision on MAFA [25] dataset.

### 2.2.3 | 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 minimise the adverse effects of face detection due to occlusions. One promising approach is to use a novel grid loss [65], 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 considers occluded face detection as a particular single-class object detection problem, inspired by other works on object detection [35, 66–68]. Furthermore, the proposed grid loss minimises the error rate on face sub-blocks independently rather than over the whole face to mitigate the adverse effects 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) [69] 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. [70] 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 [45].

## 3 | 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 Figure 5.

### 3.1 | 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

20474946, 2021, 6, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/bme2.12029 by Universitaet Duisburg-Essen, Wiley Online Library on [26/05/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

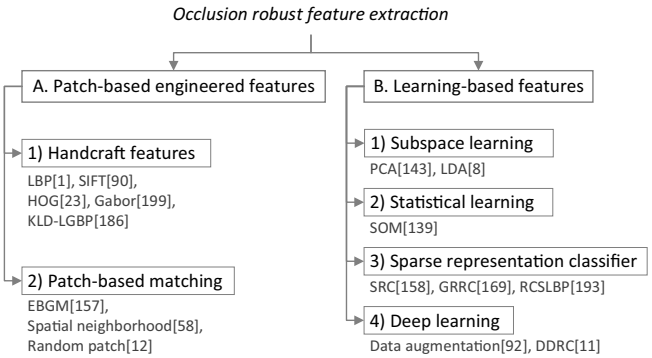


**FIGURE 5** Categorisation of methods used in occlusion robust feature extraction approaches

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.

### 3.1.1 | Handcraft features

Local Binary Patterns (LBP) [71, 72] is used to derive a novel and efficient facial image representation and has been widely used in various applications. LBP and variants [73, 74] 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 [75] is popular in object recognition and baseline matching and can also be applied to face recognition [76]. 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 [77] 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 characterise 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 [78]. Phase information instead of magnitude information from Gabor features contains discrimination and is thus widely used for recognition [79]. 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. [80] proposes KLD-LGBP, in which Kullback-Leibler Divergence (KLD) is introduced to measure the distance between the local Gabor binary patterns (LGBP) feature [81] 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.

### 3.1.2 | Patch-based matching

Besides 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) [82] uses a graph to represent a face, and each node of the graph corresponding to the Gabor jets extracted from facial landmarks. The matching method is used to calculate the distance between corresponding representations of two faces. To take an advantage of elastic and partial matching, Ref. [83] proposes a robust matching metric to match Difference of Gaussian (DoG) filter descriptor of a facial part against its spatial neighbourhood 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 neighbourhood descriptors and then the minimal distance is selected, which is effective in reducing the adverse effects of occlusion. A random sampling patch-based method [84] 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 [82],the dynamic graph representation [85] 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. The 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.

## 3.2 | Learning-based features

Compared with engineered features, learnt features are more flexible when various occlusion types at different locations are present. Features learnt from training data can be effective and have potentially high discriminative power for face recognition.

20474946, 2021, 6, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/bme2.12029 by Universitaet Duisburg-Essen, Wiley Online Library on [26/05/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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 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 as a weighted combination 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 [86], especially deep convolutional neural networks (DCNN) for uncontrolled face recognition applications.

### 3.2.1 | Subspace learning

Approaches such as the Principal component analysis (PCA) [87] and the Linear Discriminant Analysis (LDA) [88] are the two most representative methods in learning the subspace for feature extraction. Eigenface [87] uses PCA to learn the most representative subspace from training face images. Fisherface [88] 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 [89] 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. [90] uses an influence function technique to develop robust subspace learning for a variety of linear models within a continuous optimization framework. Ref. [91] uses the property of PCA to detect 'outliers"' and occlusions, and calculates the coefficients using the rest of the pixels. Inspired by robust PCA [91], a new approach combining discriminative power and reconstructive property has been proposed by Fidler et al. [92], 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 LDA coefficients in a robust manner. Besides, objects' orientation application, a regression task, is introduced to robustly estimate CCA coefficients. The authors of Ref. [93] propose 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. The authors of Ref. [94] propose kernel correlation feature analysis (KCFA) on facial regions, that is, eye-region, nose-region, and mouth-region to cope with partial face recognition. The proposed KCFA for feature extraction outperforms the linear approaches PCA [87], LDA [88] and variants of kernel methods [95, 96]. The authors of Ref. [97] propose the confidence weighted subspace method to weigh image pixels according to a measure, which quantifies the confidence for the validity of each image pixel. As a result, this method reduces the impact of occluded pixels on the subspace representation of the given face image.

### 3.2.2 | 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. [98], which takes a probabilistic approach, that is, a mixture of Gaussians to compensate for partially occluded faces. Specifically, they analyse 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. [99] extends the work by representing the face subspace with self-organising 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 neighbour 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. [100] 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. [101], 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 the article [102], 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 [103], which combines the learnt 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. [104] try to solve

20474946, 2021, 6, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/bme2.12029 by Universitaet Duisburg-Essen, Wiley Online Library on [26/05/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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 modelled 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.

### 3.2.3 | 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 [105], 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. Based on their former work [106], Yang et al. [107] 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 recognise 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 [108], 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) [109] 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 [110], 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 [111] 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. [112] proposes to learn a robust and discriminative low-rank representation (RDLRR) for optimal classification, which aiming to prepare the learnt 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 a better performance in case of illumination changes, real occlusion as well as block occlusion. Ref. [113] combines centre-symmetric local binary patterns (CSLBP) with enhanced centre-symmetric local binary patterns (ECSLBP) to build the SRC dictionary for occluded face recognition. In Ref. [114], 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 [115] 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) [116] 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, the authors of Ref. [117] propose 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 labelling 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 the few-shot occluded face recognition.

### 3.2.4 | 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 [86]. Face verification performance has been boosted due to advanced deep CNN architectures [2, 3, 6–8] and the development of loss functions [4, 5, 9, 10, 118–121]. From a practical point of view, we can obtain occlusion-robust face representation if a massive

20474946, 2021, 6, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/bme2.12029 by Universitaet Duisburg-Essen, Wiley Online Library on [26/05/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

training dataset is provided to contain sufficient occluded faces for a deep network to learn to handle occlusions [122]. 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 [123]. Lv et al. [124] synthesise 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 synthesise 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 Ref [125], 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, pre-defined occlusions may cause performance degradation when dealing with face images with an occlusion that is not of the same size. Cen et al. [126] 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.

# 4 | 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 categorisation of occlusion aware face recognition is shown in Figure 6.

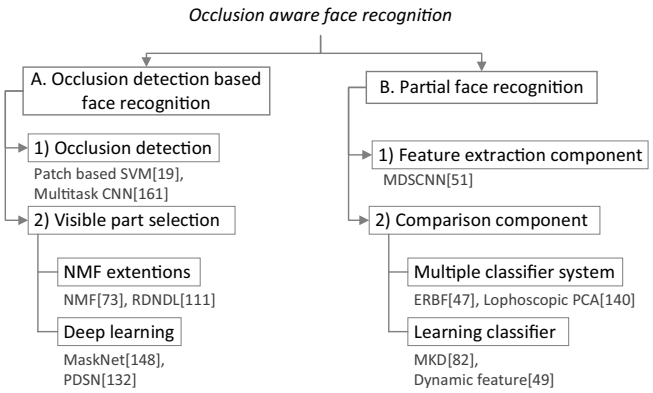


**FIGURE 6** Categorisation of methods classification used in occlusion aware face recognition approaches

## 4.1 | 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.

### 4.1.1 | Occlusion detection

An intuitive idea to deal with occlusions in face recognition is to detect the occlusions first and then recognise the face based on unoccluded facial parts. Methods use predefined occlusion types as a substitute for arbitrary occlusion in different locations to simplify the 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. [127] introduces a 1-NN (nearest neighbour) threshold classifier to detect the occluded face region in a PCA subspace learnt 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 [128, 129] 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. Min et al. [130] propose to first analyse the presence of potential occlusion on a face and then perform face recognition on the nonoccluded facial regions based on local Gabor binary pattern histogram sequences (LGBPHS) [81]. Specifically, an occlusion detection and segmentation module is proposed to identify potential

20474946, 2021, 6, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/bme2.12029 by Universitaet Duisburg-Essen, Wiley Online Library on [26/05/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

occluded facial components and generate an occlusion mask. During recognition, they propose a selective LGBPHS which selects features from the nonoccluded region only. Ref. [131] proposes to utilise 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. [132] 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.

#### 4.1.2 | Visible part selection

some works select visible facial parts for recognition and skip occlusion detection by assuming the prior knowledge of occlusion. Ref. [133] compares several subspace-based methods including PCA, non-negative matrix factorization (NMF) [134], local NMF [135] and spatially confined NMF [136] 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 reference [137] in this direction extends NMF to include adaptive occlusion estimation based on the reconstruction errors. Low-dimensional representations are learnt to ensure that features of the same class are close to that of the mean class centre. This method does not require prior knowledge of occlusions and can handle large continuous occlusions.

In reference [138], a proposed MaskNet is added to the middle layer of CNN models, aiming to learn image features with high fidelity and 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. [139] 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 synthesised and realistic occluded face datasets.

### 4.2 | 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 categorise the methods as feature-aware and comparison-aware methods.

#### 4.2.1 | Feature extraction component

As for feature-aware methods, Multiscale Double Supervision Convolutional Neural Network (MDSCNN) [140] 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 learnt 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.

#### 4.2.2 | 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. [141] 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 reference [142], they propose the Lophoscopic PCA method to recognise faces in the absence of a 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

20474946, 2021, 6, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/bme2.12029 by Universitaet Duisburg-Essen, Wiley Online Library on [26/05/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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. [143] conducts the first systematic study to recognise 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. [144] utilises 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 neighbour descriptors of all gallery images with the sparse constraint. As an improvement, Ref. [145, 146] 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) [146] 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.

Recently, there is a trend to combine the deep learning and SRC methods to tackle the partial face recognition [147, 148]. Dynamic feature learning [147] combines a fully convolutional network (FCN) with sparse representation classification (SRC) for partial face recognition. The sliding window matching proposed in paper [149] 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. [148] uses multi-scale representations in dynamic feature learning, which increase the ability to tolerate the misalignment between partial probe patches and the gallery face.

# 5 | OCCLUSION RECOVERY BASED FACE RECOGNITION

Apart from addressing the occlusion problem in feature space, one intuitive idea takes 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 Figure 7.

## 5.1 | 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.

### 5.1.1 | Linear reconstruction

As for reconstruction techniques, Ref. [150] utilises 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 [90–92] employing PCA to detect outliers or occlusion and then reconstruct occlusion-free face images. Ref. [151] 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.

### 5.1.2 | Sparse representation classifier

*The* sparse representation classifier [105] 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. [152]

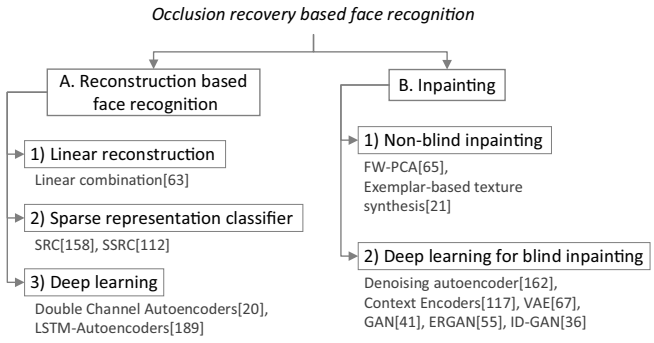


**FIGURE 7** Categorisation of methods used in occlusion recovery based face recognition approaches

20474946, 2021, 6, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/bme2.12029 by Universitaet Duisburg-Essen, Wiley Online Library on [26/05/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

improves sparse representation by including prior knowledge of pixel error distribution. In reference [153], 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. [154] proposes a reconstruction approach consisting of occlusion detection and face reconstruction. First, the downsampled SRC is used to locate all possible occlusions in the downsampled images, which has a low computing complexity compared to SRC based on original face images. Second, all discovered face pixels are imported into an overdetermined equation system to reconstruct an intact face. The authors in Ref. [155] propose scheme structured sparse error coding (SSEC) for occluded face recognition by combing structure information of error support and error distribution. Specifically, a morphological graph model is proposed to integrate the morphological information of occlusion into graph in order to explore the geometrical structural of the occlusion error. An exponential probabilistic model is further proposed to study the error distribution under different error metrics. An innovative solution for the occlusion challenge is presented by structured sparse representation based classification (SSRC) [156] to learn an occlusion dictionary. The regularisation 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 reference [157], 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. [158] 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. [159] 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.

#### 5.1.3 | Deep learning

A few works use deep learning techniques for occlusion reconstruction. One is work [160], that 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 learnt 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 [161] 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.

### 5.2 | 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.

#### 5.2.1 | 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 colour values from the source are widely used. In paper [162], 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. [163] 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 [164] has been proposed where a PDE method and modified exemplar inpainting is utilised to remark the occluded face region. However, the occlusion type of face images studied in this work is not representative of the real scenario.

#### 5.2.2 | Deep learning for blind inpainting

Generative models are known for the ability to synthesise or generate new samples from the same distribution of the

20474946, 2021, 6, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/bme2.12029 by Universitaet Duisburg-Essen, Wiley Online Library on [26/05/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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 [165, 166], or implicit density estimation [167]. The most prominent generative models are originated from the variational autoencoder (VAE) [165] and the generative adversarial network (GAN) [167, 168]. The traditional autoencoder is used to reconstruct data, and the VAE [165] 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 optimised. 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 [167] 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. [169] uses convolutional architectures in the GAN and can generate realistic samples. However, the GAN is notorious for unstable training characteristics. Makhzani et al. [170] 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 a 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 [171] that combines sparse coding [172] 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 regularised by a sparsity-inducing term to avoid over-fitting. 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 [173] 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 [173] are proposed; some extend it by defining global and local discriminators [174, 175] and some take the result of context encoders as the input and apply joint optimization of image content and texture constraints to avoid visible artefacts around the border of the hole [176]. Ref. [177] relies on the power of the generative network to complete the corrupted image without requiring an external training dataset. A partial convolution based network [178] 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 Maximising Generative Adversarial Networks (InfoGAN) [179] maximise 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 [180] is proposed to identify a corrupted image region with an associated corrupted region recovered using a GAN pretrained on occlusion-free faces. Ref. [181] employs GAN for eyes-to-face synthesis with only eyes visible. Very recently, AttGAN [182], 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) [183] 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 reference [184], ID-GAN (identity-diversity generative adversarial network) combines a CNN-based recogniser and GAN-based recognition to inpaint realism and identity-preserving faces. The recogniser is treated as the third player to compete with the generator.

## 6 | PERFORMANCE EVALUATION

In this section, we first evaluate the performance of occluded face detection on the MAFA dataset [25] 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 [185], CAS-PEAL [186], and Extended Yale B [187] datasets. Then we describe the representative algorithms based on the proposed categories. In addition, we also categorise them in the face recognition pipeline according to which component they work on to tackle the occlusion.

### 6.1 | Evaluation of occluded face detection

There are several datasets for face detection benchmarking, which are the FDDB [51], PASCAL Face [52], AFW [58], IJB-A [188], Widerface [38] and MAFA [25] 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 Figure 1. To the best of our knowledge, the MAFA dataset takes the occlusions as the main challenge in

20474946, 2021, 6, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/bme2.12029 by Universitaet Duisburg-Essen, Wiley Online Library on [26/05/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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 summarised in Table 2 (derived from paper [69]). The precision on the MAFA testing set with the iou threshold value 0.5 is shown. Only a few methods report the results below.

## 6.2 | 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 Figure 2, which are listed from most to least realistic, including real occlusions, partial faces, synthetic occlusions, occluding rectangles and occlusion by unrelated images. Since datasets including real occlusions are rare, synthetic occluded datasets are generated so that they can meet the requirements of OFR, where an occlusion-free gallery is queried using occluded faces. It is worth mentioning that the recently published DFW dataset (disguised faces in the wild) [189] also contains a few occluded faces. However, this dataset includes impersonator and genuine obfuscated face images for each subject in order to handle face authentication challenge that we are not addressing. Thus we leave it out of occluded face dataset.

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, synthesised occlusions are imposed to form 'D-occ' for establishment. Table 3 summarises each dataset by the number of subjects, the number of images, whether or not real occlusions are included, and if not, what kind of synthesised 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 Figure 2 first column). It consists of over 4000 faces of 126 individuals: 70 men and 56 women, taken in two sessions with a 2-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 person is captured under nine poses and 64 different illuminations without occlusion. It is widely used to evaluate the efficacy of the algorithms under synthesised occlusions. Occluding unrelated images are randomly superimposed on the occlusion-free faces (see Figure 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 Figure 2 fourth column).

## 6.3 | 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 categorise methods based on the face recognition pipeline and show evaluation results from this aspect.

**TABLE 2** Evaluation summary of different occluded face detection algorithms on MAFA

| Approach | Publication | Precision 'masked' | Precision 'w/o Ignored' (%) |
|---|---|---|---|
| Occlusion unaware | [40] | NA | 60.8 |
| Occlusion aware | [25] | NA | 76.4 |
| | [63] | 76.5% | 88.3 |
| Occlusion segmentation | [69] | 83.5% | 91.9 |

*Note*: 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.

**TABLE 3** Summary of existing benchmark datasets for occluded face recognition

| Dataset | # Sub | # Ims | Real Occ. | Syn. Occ. |
|---|---|---|---|---|
| AR [185] | 126 | 4, 000 | Yes | NA |
| Extend YaleB [187] | 38 | 161, 289 | No | Unrelated Ims |
| ORL [124] | 40 | 400 | No | Rectangle |
| FERET [113] | 1500 | 13, 000 | No | Rectangle |
| LFW [59] | 5749 | 13, 000 | No | Partial faces |
| CAS-PEAL [186] | 1040 | 9, 594 | Yes | NA |
| CMU-PIE [44] | 68 | 41, 368 | No | Rectangle |
| PubFig [70] | 200 | 58, 797 | No | Partial faces |
| NIR-distance [140] | 276 | 4300 | No | Partial faces |
| NIR-mobile [183] | 374 | 11, 286 | No | Partial faces |
| FRGC [105] | 276 | 943 | No | Partial faces |
| FRGC-v2 [118] | 466 | 4, 007 | No | Partial faces |
| IJB-C [26] | 3531 | 148, 800 | Yes | NA |

*Notes*: '# 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 synthesised occlusions imposed on the original dataset to obtain 'D-occ'.

20474946, 2021, 6, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/bme2.12029 by Universitaet Duisburg-Essen, Wiley Online Library on [26/05/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

### 6.3.1 | Evaluation based on the proposed categories

We summarise the identification performances of representative algorithms on the AR database in Table 4 based on the proposed categories: ORFE, OAFR, and ORecFR. 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-*X*ips' to abbreviate the situation when the training set contains *X* unoccluded images per person. For example, 'S-TR-8ips' indicates that for each person there are eight of unoccluded 13 images used for training. The rest of the face images occluded by sunglasses and the scarf are evaluated during testing. Specifically, 'S-TR-SSPP' single sample per person in the training set remains a challenging task. Because for each person there is only one unoccluded face image available for training. Usually, the gallery is equal to training set if not stated otherwise. We add asterisks to mark the gallery, only taking one neutral face for enrolment.
(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. For example, 'S-TR-Occ3' indicates that three occluded face images from one session are used for training, and the evaluation is done on the other session. It is plausible that 'S-TR-Occ*X*' setting is less difficult compared to 'S-TR-*X*ips', which has a higher accuracy. Typically, one neutral face image per individual is enroled during testing.
(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'.

As results on the AR dataset in Table 4 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.

The identification performance of representative algorithms on extended Yale B as well as other benchmarks is summarised in Table 5. The experimental settings include different occlusion types, such as an occluding rectangle and unrelated images, the occlusion ratio over a image (ranging from 0.3 to 0.9), the number of training/testing subjects are listed in detail. 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 (see Figure 2). 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. An evaluation summary of different categories of representative algorithms based on verification rates is shown in Table 6. Since partial faces are arbitrary patches, it is hard to be sure algorithms are on a level playing field. Therefore, these methods cannot be compared properly. For example, partial faces with eye regions are supposed to contain more discriminative features compared to those with no eyes involved.

### 6.3.2 | Evaluation based on the face recognition pipeline for categorisation

Generally, occlusion robust feature extraction (ORFE) methods deal with occlusion using feature extraction. However, they sometimes treat occluded faceRepresentative algorithms 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 categorise methods discussed so far based on a general face recognition pipeline to offer a fresh perspective. Representative algorithms regarding application-oriented purpose are listed in Table 8 and References classified based on their components are listed in Table 7. Each component is described below:

20474946, 2021, 6, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/bme2.12029 by Universitaet Duisburg-Essen, Wiley Online Library on [26/05/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

**TABLE 4** Evaluation Summary of Representative Algorithms on AR dataset regarding identification rates

| Category | Publication | Occlusions | Experiment settings | # of train subjects | # of test subjects | Identification rates (%) | |
|---|---|---|---|---|---|---|---|
| | | | | | | Session1 | Session2 |
| ORFE: patch-engineered features | [80] | Sunglasses | S-TR-8IPS | 50 | 50 | 84.0 | 80.0 |
| | | Scarf | | | | 100.0 | 96.0 |
| | [84] | Sunglasses | S-TR-SSPP | 100 | 100 | 89.0 | 98.0 |
| | | Scarf | | | | 73.0 | 92.0 |
| ORFE: learning-based features | [98] | Sunglasses | S-TR-SSPP | 50 | 50 | 52.0 | NA |
| | | Scarf | | | | 48.0 | NA |
| | [100] | Sunglasses | S-TR-8IPS | 40 | 40 | 80.0 | NA |
| | | Scarf | | | | 70.8 | NA |
| | [101] | Sunglasses | D-TR | 50 | 50 | 98.0 | NA |
| | | Scarf | | | | 90.0 | NA |
| | [105] | Sunglasses | S-TR-8IPS | 100 | 100 | 97.5 | NA |
| | | Scarf | | | | 93.5 | NA |
| | [102] | Sunglasses&Scarf | S-TR-3IPS | 100 | 100 | 89.0 | NA |
| | | Black block | | | | 94.0 | NA |
| | [107] | Sunglasses | S-TR-7IPS | 100 | 100 | 92.3 | 51.7 |
| | | Scarf | | | | 95.0 | 84.3 |
| | [156] | Sunglasses★ | S-TR-Occ1 | 100 | 100 | 93.0 | |
| | | Scarf★ | | | | 92.7 | |
| | | Sunglasses&Scarf★ | | | | 92.6 | |
| | [103] | Sunglasses★ | S-TR-3IPS | 100 | 100 | 95.7 | NA |
| | | Scarf★ | | | | 96.3 | NA |
| | [104] | Sunglasses★ | S-TR-Occ3 | 100 | 100 | TR | 98.0 |
| | | Scarf★ | | | | TR | 97.0 |
| | [112] | Sunglasses | S-TR-SSPP | 100 | 100 | TR | 92.0 |
| | | Scarf | | | | TR | 91.0 |
| | | Sunglasses & Scarf | | | | TR | 82.5 |
| | [113] | Sunglasses★ | S-TR-SSPP | 80 | 80 | 88.0 | 56.0 |
| | | Scarf★ | | | | 69.0 | 44.0 |
| | [114] | Sunglasses★ | EX-TR-SSPP | 80 | 80 | 95.8 | 78.3 |
| | | Scarf★ | | | | 90.0 | 77.9 |
| | [111] | Sunglasses & Scarf | D-TR-DL | Webface+20 | 80 | 100.0 | 96.3 |
| | [115] | Sunglasses & Scarf | S-TR-SSPP | 100 | 100 | 75.0 | |
| | [126] | Sunglasses | S-TR-7IPS | 50 | 50 | 94.7 | 85.3 |
| | | Scarf | | | | 99.3 | 98.7 |
| | [184] | Sunglasses | S-TR-3IPS | 80 | 80 | 99.0 | NA |
| | | Scarf | | | | 87.6 | NA |

20474946, 2021, 6, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/bme2.12029 by Universitaet Duisburg-Essen, Wiley Online Library on [26/05/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

**TABLE 4** (Continued)

| Category | Publication | Occlusions | Experiment settings | # of train subjects | # of test subjects | Identification rates (%) | |
|---|---|---|---|---|---|---|---|
| | | | | | | Session1 | Session2 |
| OAFR: Occlusion detection face recognition | [127] | Sunglasses | EX-TR-SSPP | 100 | 100 | NA | 49.0 |
| | | Scarf | | | | NA | 55.0 |
| | [129] | Sunglasses* | S-TR-3IPS* | 80 | 80 | TR | 54.2 |
| | | Scarf* | | | | TR | 81.3 |
| | [131] | Sunglasses* | S-TR-8IPS-Occ1 | 60 | 60 | 97.5 | NA |
| | | Scarf* | | | | 99.2 | NA |
| | [137] | Sunglasses | S-TR-RAND | 100 | 100 | 95.2 | |
| | | Scarf | | | | 94.2 | |
| | [139] | Sunglasses | D-TR-DL | Webface | 100 | 99.7 | NA |
| | | Scarf | | | | 100.0 | NA |
| OAFR: Partial face recognition | [144] | Sunglasses* | S-TR-7IPS | 100 | 100 | 94.3 | NA |
| | | Scarf* | | | | 98.0 | NA |
| | | Sunglasses&Scarf* | | | | 96.2 | NA |
| | [145] | Sunglasses* | S-TR-14IPS | 100 | 100 | 98.0 | NA |
| | | Scarf* | | | | 97.0 | NA |
| | | Sunglasses & Scarf* | | | | 97.5 | NA |
| | [146] | Sunglasses* | S-TR-7IPS | 100 | 100 | 100.0 | 92.0 |
| | | Scarf* | | | | 100.0 | 95.3 |
| ORecFR: Occlusion recovery face recognition | [151] | Sunglasses & Scarf* | S-TR-13IPS | 100 | 100 | TR | 90.6 |
| | [152] | Sunglasses | S-TR-8IPS | 100 | 100 | 100.0 | NA |
| | | Scarf | | | | 97.0 | NA |
| | [153] | Sunglasses* | S-TR-7IPS | 121 | 121 | 76.6 | NA |
| | | Scarf* | | | | 60.9 | NA |
| | [154] | Sunglasses* | S-TR-8IPS | 100 | 100 | 97.5 | NA |
| | [190] | Sunglasses | S-TR-8IPS | 100 | 100 | 94.5 | NA |
| | | Scarf | | | | 95.0 | NA |
| | [157] | Sunglasses | S-TR-3IPS-Occ3 | 120 | 120 | NA | 68.5 |
| | | Scarf | | | | NA | 70.7 |
| | [158] | Sunglasses | S-TR-2IPS | 100 | 100 | 89.8 | |
| | | Scarf | | | | 78.8 | |
| | [159] | Sunglasses | S-TR-4IPS | 120 | 120 | 99.2 | 99.7 |
| | | Scarf | | | | 87.5 | 83.6 |

*Notes:* Three categories are: (1) **ORFE**, (2) **OAFR**, and (3) **ORecFR**. 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.

Abbreviations: FR, face recognition; **OAFR**, Occlusion Aware Face Recognition; **ORFE**, Occlusion Robust Feature Extraction; **ORFE**, Occlusion Robust Feature Extraction.

- 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.

20474946, 2021, 6, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/bme2.12029 by Universitaet Duisburg-Essen, Wiley Online Library on [26/05/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

**TABLE 5** Evaluation Summary of Representative Algorithms on the other Benchmarks regarding identification rates

| Category | Publication | Occlusions | Occlusion arguments | Benchmark | # Of train subjects | # Of test subjects | Identification rates (%) |
|---|---|---|---|---|---|---|---|
| **ORFE** | [100] | Black or white rectangle | occlusion size:50 × 50 | ORL | 40 | 40 | 70.0 |
| | [101] | Black rectangle | occlusion size:50 × 50 | FERET | 240 | 960 | 78.5 |
| | [105] | Unrelated image | occlusion ratio:0.5 | Extended Yale B | 38 | 38 | 65.3 |
| | [107] | Unrelated image | occlusion ratio:0.5 | Extended Yale B | 38 | 38 | 87.4 |
| | [156] | Unrelated image | occlusion ratio:0.8 | Extended Yale B | 38 | 38 | 70.0 |
| | [104] | Unrelated image | occlusion ratio:0.8 | Extended Yale B | 38 | 38 | 53.0 |
| | [111] | Wild condition | NA | LFW | 108 | 50 | 86.0 |
| | | | NA | | 851 | 124 | 65.3 |
| | [114] | 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 |
| | [112] | Unrelated image | occlusion ratio:0.6 | Extended Yale B | 38 | 38 | 96.0 |
| | | Random pixels rectangle | occlusion ratio:0.6 | | | | 81.0 |
| | | Mixture noise | occlusion ratio:0.6 | | | | 25.0 |
| | [113] | Unrelated image | occlusion ratio:0.3 | Extended Yale B | 30 | 30 | 77.6 |
| | [137] | Pepper noise | portion:40% | Extended Yale B | 38 | 38 | 81.3 |
| | | White rectangle | NA | | | | 82.9 |
| | | Salt&Pepper noise | portion:40% | CMU-PIE | 68 | 68 | 98.5 |
| | | White rectangle | occlusion size:10 × 10 | | | | 98.8 |
| | [116] | Black rectangle | occlusion ratio:0.3 | Extended Yale B | 38 | 38 | 99.2 |
| | | Black rectangle | occlusion ratio:0.4 | | | | 97.6 |
| | | Black rectangle | occlusion ratio:0.5 | | | | 96.1 |
| | [110] | 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 |

20474946, 2021, 6, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/bme2.12029 by Universitaet Duisburg-Essen, Wiley Online Library on [26/05/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

**TABLE 5** (Continued)

| Category | Publication | Occlusions | Occlusion arguments | Benchmark | # Of train subjects | # Of test subjects | Identification rates (%) |
|---|---|---|---|---|---|---|---|
| **OAFR** | [144] | Arbitrary patches | NA | Partial-LFW | 158 | 158 | 34.8 |
| | [145] | Arbitrary patches | NA | Partial-LFW | 158 | 158 | 50.7 |
| | | Unrelated image | occlusion ratio:0.5 | Extended Yale B | 32 | 32 | 30.2 |
| | [146] | Unrelated image | occlusion ratio:0.5 | Extended Yale B | 32 | 32 | 56.7 |
| | | Arbitrary patches | NA | Partial-PubFig | 60 | 140 | 42.9 |
| | [147] | Arbitrary patches | NA | Partial-LFW | NA | 1000 | 27.3 |
| | | Arbitrary patches | NA | NIR-distance | NA | 276 | 94.9 |
| | | Arbitrary patches | NA | Partial-YTF | NA | 200 | 61.0 |
| | [148] | 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** | [190] | Unrelated image | occlusion ratio:0.7 | Extended Yale B | 38 | 38 | 62.3 |
| | [152] | Unrelated image | occlusion ratio:0.7 | Extended Yale B | 38 | 38 | 88.5 |
| | | Multiple patches | occlusion ratio:0.8 | | | | 96.0 |
| | [154] | Black or white rectangle | occlusion ratio:0.57 | Extended Yale B | 38 | 38 | 90.4 |
| | | Unrelated image | occlusion ratio:0.5 | | | | 87.7 |
| | [157] | Black rectangle on eyes | occlusion ratio:0.3 | Extended Yale B | 38 | 38 | 98.6 |
| | [158] | Unrelated image | occlusion ratio:0.6 | Extended Yale B | 38 | 38 | 95.8 |
| | | Unrelated image | occlusion ratio:0.9 | | | | 71.9 |

*Notes*: Three categories are given: (1) **ORFE**, (2) **OAFR**, and (3) **ORecFR**. In the 'Occlusion Arguments' column, details including the size of an 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 one 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.

Abbreviations: FR, face recognition; **OAFR**, Occlusion Aware Face Recognition; **ORFE**, Occlusion Robust Feature Extraction; **ORFE**, Occlusion Robust Feature Extraction.

**TABLE 6** Evaluation Summary of Representative Algorithms regarding verification rates

| Publication | Occlusions | Benchmark | # Of subjects in gallery | Verification rates |
|---|---|---|---|---|
| [94] | Eyeglasses | FRGC-v2 | 466 | 90%@FAR = 0.001 |
| [83] | Wild condition | LFW | 1680 | 60%@FAR = 0.1 |
| [104] | Wild condition | LFW | 1680 | 61%@FAR = 0.1 |
| [146] | Arbitrary patches | Partial-LFW | 1680 | 50%@FAR = 0.1 |
| | Arbitrary patches | Partial-PubFig | 140 | 63%@FAR = 0.1 |
| [147] | Arbitrary patches | Partial-LFW | 1000 | 29.8%@FAR = 0.001 |
| [148] | Arbitrary patches | Partial-LFW | 1000 | 37.9%@FAR = 0.001 |

*Notes*: 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.

20474946, 2021, 6, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/bme2.12029 by Universitaet Duisburg-Essen, Wiley Online Library on [26/05/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

**TABLE 7** Summary of Representative Algorithms based on components they work on during OFR

| Pipeline category | Publication |
|---|---|
| Data augmentation/Recovery | [90–92, 123–125, 150, 151, 153, 154, 159–164, 179, 182–184, 190] |
| Feature extraction | [71, 72, 75, 77, 78, 80, 93, 98–100, 102, 103, 105, 107, 109–114, 116, 117, 126, 137–139, 143, 147, 148, 151, 152, 156, 158] |
| Feature comparison | [82, 83, 101, 104, 128, 129, 131, 133, 144–146] |
| Fusion strategy | [84, 94, 114, 117, 127, 140–142, 157] |

*Notes*: In the 'data augmentation/recovery' category, data augmentation component means generating synthesised 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.

**TABLE 8** Summary of representative algorithms regarding application-oriented purpose from Figure 3

| Abbreviation | Application-oriented purpose | Publication |
|---|---|---|
| **OFD** | Occluded face detection | [25, 38, 51–53, 57–63, 65, 69, 70] |
| **ORFE** | Patch based engineered features | [71, 72, 75, 77, 78, 80, 81, 83, 84, 93, 94, 114] |
| | Learning-based features | [91–93, 98–105, 107, 109–114, 116, 117, 123–126, 156] |
| **OAFR** | Occlusion detection | [69, 132] |
| | Occlusion discard face recognition | [127–129, 131, 133, 137–139] |
| | Partial face detection | NA |
| | Partial face recognition | [140, 141, 143–148] |
| **ORFR** | Occlusion recovery | [160, 162–164, 171–183] |
| | Occlusion recovery face recognition | [24, 27, 32, 36, 61, 63, 74, 80, 91, 115, 152, 189] |

*Notes:* There is a category for occluded face detection with the abbreviation '**OFD**'. Three classes for occluded face recognition: (1) **ORFE**, (2) **OAFR**, and (3) **ORFR**. To our knowledge, there are no research works on detecting partial face; thus, we mark it as 'NA' (not available).

Abbreviations: FR, face recognition; **OAFR**, Occlusion Aware Face Recognition; **ORFE**, Occlusion Robust Feature Extraction; **ORFE**, Occlusion Robust Feature Extraction.

- 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 minimise the effect of features extracted from the occlusion.

# 7 | 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.

## 7.1 | 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 synthesised occlusions, occluding rectangles and unrelated images are commonly applied. In the literature, 'Baboon' is a typically used unrelated image to generate synthetic occluded faces [104, 158, 190]. However, these synthetically occluded faces like those using 'Baboon', or a black/white rectangle are not representative of real-life occluded faces. With regard to 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,

20474946, 2021, 6, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/bme2.12029 by Universitaet Duisburg-Essen, Wiley Online Library on [26/05/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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 [124]. 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., recognising 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 [26], 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.

## 7.2 | 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 Figure 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 [122–124]. Current solutions for data augmentation mainly rely on large-scale training data from the web, and use occlusion templates to generate synthesised 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 centres. 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 [184]. 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 [138, 139]. 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.

# 8 | CONCLUSION

In this article, we present a thorough survey of face recognition techniques under occlusion and systematically categorise 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.

## ACKNOWLEDGMENT

The authors would like to thank University of Twente spinoff company 20Face for supporting this research.

## ORCID

*Dan Zeng* https://orcid.org/0000-0002-9036-7791

20474946, 2021, 6, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/bme2.12029 by Universitaet Duisburg-Essen, Wiley Online Library on [26/05/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

## REFERENCES


1. Best-Rowden, L., Anil, K.J.: Longitudinal study of automatic face recognition. IEEE Trans. Pattern Anal. Mach. Intell. 40(1), 148–162 (2018)
2. He, K., et al.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 770–778 (2016)
3. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 1097–1105 (2012)
4. Liu, W., et al.: Sphereface: deep hypersphere embedding for face recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 212–220 (2017)
5. Liu, W., et al.: Large-margin softmax loss for convolutional neural networks. ICML. 2, 7 (2016)
6. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
7. Szegedy, C., et al.: Inception-v4, inception-resnet and the impact of residual connections on learning. In: Thirty-first AAAI Conference on Artificial Intelligence (2017)
8. Szegedy, C., et al.: Going deeper with convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1–9 (2015)
9. Wang, H., et al.: Cosface: large margin cosine loss for deep face recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 5265–5274 (2018)
10. Wen, Y., et al.: A discriminative feature learning approach for deep face recognition. In: European Conference on Computer Vision, 499–515. Springer (2016)
11. Bansal, A., et al.: Umdfaces: an annotated face dataset for training deep networks. In: 2017 IEEE International Joint Conference on Biometrics (IJCB), 464–473. IEEE (2017)
12. Cao, Q., et al.: Vggface2: a dataset for recognising faces across pose and age. In: 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018), 67–74. IEEE (2018)
13. Guo, Y., et al.: Ms-celeb-1m: a dataset and benchmark for large-scale face recognition. In: European Conference on Computer Vision, 87–102. Springer (2016)
14. Nech, A., Kemelmacher-Shlizerman, I.: Level playing field for million scale face recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 7044–7053 (2017)
15. Omkar, M., et al.: Deep face recognition. BMVC. 1, 6 (2015)
16. Dong Y., et al.: Learning face representation from scratch. arXiv preprint arXiv:1411.7923 (2014)
17. Learned-Miller, E., et al.: Labelled faces in the wild: a survey. Advances in face detection and facial image analysis, 189–248. Springer (2016)
18. Trigueros, D.S., Meng, L., Hartnett, M.: Face recognition: from traditional to deep learning methods. arXiv preprint arXiv:1811.00116, (2018)
19. Ekenel, H.K., Stiefelhagen, R.: Why is facial occlusion a challenging problem? In: International Conference on Biometrics, 299–308. Springer (2009)
20. Muhtahir, O.O., Hancke, G.P., Kapileswar, N.: Evaluating the effect of occlusion in face recognition systems. In: 2017 IEEE AFRICON, 1547–1551. IEEE (2017)
21. Azeem, A., et al.: A survey: face recognition techniques under partial occlusion. Int. Arab J. Inf. Technol. 11(1), 1–10 (2014)
22. Dagnes, N., et al.: Occlusion detection and restoration techniques for 3d face recognition: a literature review. Mach. Vis. Appl. 29, 789–813 (2018)
23. Lahasan, B., Lutfi, S.L., San-Segundo, R.: a survey on techniques to handle face recognition challenges: occlusion, single sample per subject and expression. Artif. Intell. Rev. 52, 949–979 (2019)
24. Ghazi, M.M., Ekenel, H.K.: A comprehensive analysis of deep learning based representation for face recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 34–41 (2016)
25. Ge, S., et al.: Detecting masked faces in the wild with lle-cnns. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2682–2690 (2017)
26. Maze, B., et al.: Iarpa janus benchmark-c: face dataset and protocol. In: International Conference on Biometrics (ICB), 158–165. IEEE (2018)
27. Zafeiriou, S., Zhang, C., Zhang, Z.: A survey on face detection in the wild: past, present and future. Comput Vis Image Understand. 138, 1–24 (2015)
28. Paul, V., Jones, M.: 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, 1, I–I. IEEE (2001)
29. Paul, V., Jones, M.J.: Robust real-time face detection. Int. J. Comput. Vis. 57(2), 137–154 (2004)
30. Yang, B., et al.: Aggregate channel features for multi-view face detection. In: IEEE International Joint Conference on Biometrics, 1–8. IEEE (2014)
31. Yan, J., et al.: The fastest deformable part model for object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2497–2504 (2014)
32. Bai, Y., et al.: Finding tiny faces in the wild with generative adversarial network. In: Proceedings of the IEEE Conference On Computer Vision and Pattern Recognition, 21–30 (2018)
33. Chen, D., et al.: Joint cascade face detection and alignment. In: European Conference on Computer Vision, 8694, 109–122. Springer (2014)
34. Jiang, H., Learned-Miller, E.: Face detection with the faster r-cnn. In: 12th IEEE 2017 International Conference on Automatic Face & Gesture Recognition (FG 2017), 650–657. IEEE (2017)
35. Li, H., et al.: A convolutional neural network cascade for face detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 5325–5334 (2015)
36. Li, Z., et al.: Pyramidbox++: high performance detector for finding tiny face (2019). arXiv preprint arXiv:1904.00386
37. Sun, X., Wu, P., Hoi, S.C.H.: Face detection using deep learning: an improved faster rcnn approach. Neurocomputing, 299, 42–50 (2018)
38. Yang, S., et al.: Wider face: a face detection benchmark. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 5525–5533 (2016)
39. Zhang, C., Zhang, Z.: Improving multiview face detection with multi-task deep convolutional neural networks. In: IEEE Winter Conference on Applications of Computer Vision, 1036–1041. IEEE (2014)
40. Zhang, K., et al.: Joint face detection and alignment using multitask cascaded convolutional networks. IEEE Signal Process. Lett. 23(10), 1499–1503 (2016)
41. Zhang, S., et al.: S3fd: single shot scale-invariant face detector. In: Proceedings of the IEEE International Conference on Computer Vision, 192–201 (2017)
42. Zhang, Z., et al.: Robust face detection via learning small faces on hard images. In: The IEEE Winter Conference on Applications of Computer Vision, 1361–1370 (2020)
43. Girshick, R., et al.: Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 580–587 (2014)
44. Girshick, R.: Fast r-cnn. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1440–1448 (2015)
45. Ren, S., et al.: Faster r-cnn: towards real-time object detection with region proposal networks. In: Advances in Neural Information Processing Systems, pp. 91–99 (2015)
46. Liu, W., et al.: Ssd: single shot multibox detector. In: European Conference on Computer Vision, 21–37. Springer (2016)
47. Joseph, R., et al.: You only look once: unified, real-time object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 779–788 (2016)
48. Zhang, J., et al.: Feature agglomeration networks for single stage face detection. Neurocomputing. 380, 180–189 (2020)
49. Hu, P., Ramanan, D.: Finding tiny faces. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 951–959 (2017)


20474946, 2021, 6, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/bme2.12029 by Universitaet Duisburg-Essen, Wiley Online Library on [26/05/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

50. Deng, J., et al.: Retinaface: single-stage dense face localisation in the wild. arXiv preprint arXiv:1905.00641 (2019)
51. Jain, V., Learned-Miller, E.: Fddb: a benchmark for face detection in unconstrained settings. Technical report. UMass Amherst Technical Report (2010)
52. Yan, J., et al.: Face detection by structural models. Image Vis. Comput. 32(10), 790–799 (2014)
53. Chen, J., et al.: Modification of the adaboost-based detector for partially occluded faces. In: 18th International Conference on Pattern Recognition (ICPR'06), 2, 516–519. IEEE (2006)
54. Guo, Z., et al.: Occlusion-robust detector trained with occluded pedestrians. SCITEPRESS-Science and Technology Publications (2018)
55. Zhang, S., et al.: Occlusion-aware r-cnn: detecting pedestrians in a crowd. In: Proceedings of the European Conference on Computer Vision, 637–653 (2018)
56. Zhou, C., Yuan, J.: Bi-box regression for pedestrian detection and occlusion estimation. In: Proceedings of the European Conference on Computer Vision, 135–151 (2018)
57. Yang, S., et al.: From facial parts responses to face detection: a deep learning approach. In: Proceedings of the IEEE International Conference on Computer Vision, 3676–3684 (2015)
58. Zhu, X., Ramanan, D.: Face detection, pose estimation, and landmark estimation in the wild. In: Conference on Computer Vision and Pattern Recognition (2012)
59. Yang, S., et al.: Faceness-net: face detection through deep facial part responses. IEEE Trans. Pattern Anal. Mach. Intell. 40(8), 1845–1859 (2018)
60. Mahbub, U., et al.: Partial face detection for continuous authentication. In: 2016 IEEE international conference on Image processing (ICIP), 2991–2995. IEEE (2016)
61. Samangouei, P., Patel, V.M., Chellappa, R.: Attribute-based continuous user authentication on mobile devices. In: IEEE 7th International Conference on Biometrics Theory, Applications and Systems (BTAS), 1–8. IEEE (2015)
62. Zhang, H., et al.: Touch gesture-based active user authentication using dictionaries. In: IEEE Winter conference on applications of computer vision, 207–214. IEEE (2015)
63. Wang, J., Yuan, Y., Yu, G.: Face attention network: an effective face detector for the occluded faces. arXiv preprint arXiv:1711.07246, (2017)
64. Lin, T.-Y., et al.: Feature pyramid networks for object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2117–2125 (2017)
65. Opitz, M., et al.: Grid loss: detecting occluded faces. In: European Conference on Computer Vision, 386–402. Springer (2016)
66. Farfade, S.S., Saberian, M.J., Li, L.-J.: Multi-view face detection using deep convolutional neural networks. In: Proceedings of the 5th ACM on International Conference on Multimedia Retrieval, 643–650. ACM (2015)
67. Jan, H., et al.: Taking a deeper look at pedestrians. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 4073–4082 (2015)
68. Sermanet, P., et al.: Pedestrian detection with unsupervised multi-stage feature learning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 3626–3633 (2013)
69. Chen, Y., Song, L., He, R.: Masquer hunter: adversarial occlusion-aware face detection. arXiv preprint arXiv:1709.05188 (2017)
70. Wu, S., et al.: Hierarchical attention for part-aware face detection. Int. J. Comput. Vis. 127, 560–578 (2019)
71. Ahonen, T., Hadid, A., Pietikäinen, M.: Face recognition with local binary patterns. In: European Conference on Computer Vision, 469–481. Springer (2004)
72. Ahonen, T., Hadid, A., Pietikainen, M.: Face description with local binary patterns: application to face recognition. IEEE Trans. Pattern Anal. Mach. Intell. 28(12), 2037–2041 (2006)
73. Liao, S., et al.: Learning multi-scale block local binary patterns for face recognition. In: International Conference on Biometrics, 828–837. Springer (2007)
74. Zhao, G., Pietikainen, M.: Dynamic texture recognition using local binary patterns with an application to facial expressions. IEEE Trans. Pattern Anal. Mach. Intell. 29(6), 915–928 (2007)
75. Lowe, D.G.: Object recognition from local scale-invariant features. In: Proceedings of the 7th IEEE International Conference on Computer Vision, 2, 1150–1157. IEEE (1999)
76. Zhang, G., Wang, Y.: Faceprint: fusion of local features for 3d face recognition. In: International Conference on Biometrics, 394–403. Springer (2009)
77. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: IEEE 2005 Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), 1, 886–893. IEEE (2005)
78. Zou, J., Ji, Q., Nagy, G.: A comparative study of local matching approach for face recognition. IEEE Trans. Image Process. 16(10), 2617–2628 (2007)
79. Zhang, B., et al.: Histogram of gabor phase patterns (hgpp): a novel object representation approach for face recognition. IEEE Trans. Image Process. 16(1), 57–68 (2006)
80. Zhang, W., et al.: Local gabor binary patterns based on kullback–leibler divergence for partially occluded face recognition. IEEE Signal Process. Lett. 14(11), 875–878 (2007)
81. Zhang, W., et al.: Local gabor binary pattern histogram sequence (lgbphs): a novel non-statistical model for face representation and recognition. In: 10th IEEE International Conference on Computer Vision (ICCV'05). 1, 786–791. IEEE (2005)
82. Wiskott, L., et al.: Face recognition by elastic bunch graph matching. IEEE Trans. Pattern Anal. Mach. Intell. 19(7), 775–779 (1997)
83. Hua, G., Akbarzadeh, A.: A robust elastic and partial matching metric for face recognition. In: IEEE 12th International Conference on Computer vision, 2082–2089. IEEE (2009)
84. Cheheb, I., et al.: Random sampling for patch-based face recognition. In: 2017 5th International Workshop on Biometrics and Forensics (IWBF), 1–5. IEEE (2017)
85. Ren, M., et al.: Dynamic graph representation for partially occluded biometrics. arXiv preprint arXiv:1912.00377 (2019)
86. Masi, I., et al.: Deep face recognition: a survey. In: 31st SIBGRAPI Conference On Graphics, Patterns and Images (SIBGRAPI), 471–478. IEEE (2018)
87. Turk, M., Pentland, A.: Eigenfaces for recognition. J. Cognit. Neurosci. 3(1), 71–86 (1991)
88. Belhumeur, P.N., Hespanha, J.P., Kriegman, D.J.: Eigenfaces vs. fisherfaces: recognition using class specific linear projection (Technical report). Yale University, New Haven (1997)
89. Gottumukkal, R., Vijayan, K.A.: An improved face recognition technique based on modular pca approach. Pattern Recogn. Lett. 25(4), 429–436 (2004)
90. Torre, F.D.L., Michael, J.B.: A framework for robust subspace learning. Int. J. Comput. Vis. 54(1–3), 117–142 (2003)
91. Leonardis, A., Bischof, H.: Robust recognition using eigenimages. Comput. Vis. Image Understand. 78(1), 99–118 (2000)
92. Fidler, S., Skocaj, D., Leonardis, A.: combining reconstructive and discriminative subspace methods for robust classification and regression by subsampling. IEEE Trans. Pattern Anal. Mach. Intell. 28(3), 337–350 (2006)
93. Kim, J., et al.: effective representation using ica for face recognition robust to local distortion and partial occlusion. IEEE Trans. Pattern Anal. Mach. Intell. 27(12), 1977–1981 (2005)
94. Savvides, M., et al.: Partial & holistic face recognition on frgc-ii data using support vector machine. In: CVPRW'06. Conference on Computer Vision and Pattern Recognition Workshop, 48–48. IEEE (2006)
95. Yang, M.-H.: Kernel eigenfaces vs. kernel fisherfaces: face recognition using kernel methods. In: fgr, 0215. IEEE (2002)
96. Zheng, W., Zou, C., Zhao, L.: Real-time face recognition using gram-schmidt orthogonalisation for lda. In: Proceedings of the 2004 17th International Conference on Pattern Recognition, 2, 403–406. IEEE (2004)


20474946, 2021, 6, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/bme2.12029 by Universitaet Duisburg-Essen, Wiley Online Library on [26/05/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

97. Štruc, V., Simon, D., Pavešić, N.: Confidence weighted subspace projection techniques for robust face recognition in the presence of partial occlusions. In: 20th International Conference on Pattern Recognition, 1334–1338. IEEE (2010)
98. Aleix, M.M.: Recognising imprecisely localised, partially occluded, and expression variant faces from a single sample per class. IEEE Trans. Pattern Anal. Mach. Intell. 24(6), 748–763 (2002)
99. Tan, X., et al.: Recognising partially occluded, expression variant faces from single training image per person with som and soft k-nn ensemble. IEEE Trans. Neural Network. 16(4), 875–886 (2005)
100. Hotta, K.: Robust face recognition under partial occlusion based on support vector machine with local Gaussian summation kernel. Image Vis. Comput. 26(11), 1490–1498 (2008)
101. Tan, X., et al.: Face recognition under occlusions and variant expressions with partial similarity. IEEE Trans. Inf. Forensics Secur. 4(2), 217–230 (2009)
102. Seo, J., Park, H.: A robust face recognition through statistical learning of local features. In: International Conference on Neural Information Processing, 335–341. Springer (2011)
103. Seo, J., Park, H.: Robust recognition of face with partial variations using local features and statistical learning. Neurocomputing. 129, 41–48 (2014)
104. McLaughlin, N., Ming, J., Crookes, D.: Largest matching areas for illumination and occlusion robust face recognition. IEEE Trans. Cyber. 47(3), 796–808 (2017)
105. Wright, J., et al.: Robust face recognition via sparse representation. IEEE Trans. Pattern Anal. Mach. Intell. 31(2), 210–227 (2009)
106. Yang, M., Zhang, L.: Gabor feature based sparse representation for face recognition with gabor occlusion dictionary. In: European Conference on Computer Vision, 448–461. Springer (2010)
107. Yang, M., et al.: Gabor feature based robust representation and classification for face recognition with gabor occlusion dictionary. Pattern Recogn. 46(7), 1865–1878 (2013)
108. Liu, H., et al.: Face recognition using training data with artificial occlusions. In: Visual Communications and Image Processing (VCIP), 1–4. IEEE (2016)
109. Wen, Y., et al.: Structured occlusion coding for robust face recognition. Neurocomputing, 178, 11–24 (2016)
110. Fu, Y., et al.: Efficient locality-constrained occlusion coding for face recognition. Neurocomputing. (260), 104–111 (2017)
111. Yang, M., et al.: Joint and collaborative representation with local adaptive convolution feature for face recognition with single sample per person. Pattern Recogn. 66, 117–128 (2017)
112. Gao, G., et al.: learning robust and discriminative low-rank representations for face recognition with occlusion. Pattern Recogn. 66, 129–143 (2017)
113. Chen, L., et al.: ubiquitous single-sample face recognition under occlusion based on sparse representation with dual features. J. Ambient Intell. Humaniz Comput. 1–11 (2017)
114. Yu, Y.-F. et al.: Discriminative multi-scale sparse coding for single-sample face recognition with occlusion. Pattern Recogn. 66, 302–312 (2017)
115. Cho, Y.W., Ding, J.J.: Occluded face recognition using low-rank regression with generalized gradient direction. Pattern Recogn. 80, 256–268 (2018)
116. Du, L., Hu, H.: Nuclear norm based adapted occlusion dictionary learning for face recognition with occlusion and illumination changes. Neurocomputing. 340, 133–144 (2019)
117. Zheng, W., Gou, C., Wang, F.-Y.: A novel approach inspired by optic nerve characteristics for few-shot occluded face recognition. Neurocomputing. 376, 25–41 (2020)
118. Deng, J., et al.: Arcface: additive angular margin loss for deep face recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 4690–4699 (2019)
119. Schroff, F., Kalenichenko, D., James, P.: Facenet: a unified embedding for face recognition and clustering. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 815–823 (2015)
120. Sun, Y., et al.: Deep learning face representation by joint identification-verification. In: Advances in Neural Information Processing Systems, 1988–1996 (2014)
121. Wang, F., et al.: Additive margin softmax for face verification. IEEE Signal Process. Lett. 25(7), 926–930 (2018)
122. Zhou, E., Cao, Z., Qi, Y.: Naive-deep face recognition: touching the limit of lfw benchmark or not? arXiv preprint arXiv:1501.04690 (2015)
123. Osherov, E., Lindenbaum, M.: Increasing cnn robustness to occlusions by reducing filter support. In: Proceedings of the IEEE International Conference on Computer Vision, 550–561 (2017)
124. Lv, J.-J., et al.: Data augmentation for face recognition. Neurocomputing. 230, 184–196 (2017)
125. Trigueros, D.S., Meng, L., Hartnett, M.: Enhancing convolutional neural networks for face recognition with occlusion maps and batch triplet loss. Image Vis. Comput. 79, 99–108 (2018)
126. Cen, F., Wang, G.: Dictionary representation of deep features for occlusion-robust face recognition. IEEE Access. 7, 26595–26605 (2019)
127. Oh, H.J., Lee, K.M., Lee, S.U.: Occlusion invariant face recognition using selective local non-negative matrix factorization basis images. Image Vis. Comput. 26(11), 1515–1523 (2008)
128. Chen, Z., Xu, T., Han, Z.: Occluded face recognition based on the improved svm and block weighted lbp. In: 2011 International Conference on Image Analysis And Signal Processing (IASP), 118–122. IEEE (2011)
129. Rui, M., Hadid, A., Dugelay, J.-L.: Improving the recognition of faces occluded by facial accessories. In: 2011 IEEE International Conference on Automatic face & Gesture Recognition and Workshops (FG 2011), 442–447. IEEE (2011)
130. Rui, M., Hadid, A., Dugelay, J.-L.: Efficient detection of occlusion prior to robust face recognition. Sci. World J. 2014(3), 519158 (2014)
131. Andrés, A.M., et al.: Face recognition on partially occluded images using compressed sensing. Pattern Recogn. Lett. 36, 235–242 (2014)
132. Xia, Y., Zhang, B., Coenen, F.: Face occlusion detection based on multi-task convolution neural network. In: 2015 12th International Conference on Fuzzy Systems And Knowledge Discovery (FSKD), 375–379. IEEE (2015)
133. Neo, H.F., Teo, C.C., Teoh, A.B.J.: Development of partial face recognition framework. InComputer graphics, imaging and Visualization (CGIV), 2010 Seventh international conference on, pp. 142–146. IEEE (2010)
134. Lee, D.D., Seung, H.S.: Learning the parts of objects by non-negative matrix factorization. Nature. 401(6755), 788 (1999)
135. Stan, Z., et al.: Learning spatially localised, parts-based representation. In: Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2001, I–I. IEEE (2001)
136. Neo, H.F., Teoh, A.B.J., Ngo, D.C.L.: A novel spatially confined non-negative matrix factorization for face recognition. In: MVA, 502–505. Citeseer (2005)
137. Ou, W., et al.: Robust discriminative nonnegative dictionary learning for occluded face recognition. Pattern Recogn. Lett. 107, 41–49 (2018)
138. Wan, W., Chen, J.: Occlusion robust face recognition based on mask learning. In: 2017 IEEE international conference on image processing (ICIP), 3795–3799. IEEE (2017)
139. Song, L., et al.: Occlusion robust face recognition based on mask learning with pairwise differential siamese network. arXiv preprint arXiv:1908.06290 (2019)
140. He, L., et al.: Multiscale representation for partial face recognition under near infrared illumination. In: 2016 IEEE 8th International Conference on Biometrics Theory, Applications and Systems (BTAS), 1–7. IEEE (2016)
141. Gutta, S., Philomin, V., Trajkovic, M.: An investigation into the use of partial-faces for face recognition. In: fgr. 33–38 (2002)
142. Tarrés, F., Rama, A., Torres, L: A novel method for face recognition under partial occlusion or facial expression variations. In: Proceedings of 47th International Symposium ELMAR, 163–166 (2005)
143. Liao, S., Jain, A.K., Stan, Z. L.: Partial face recognition: alignment-free approach. IEEE Trans. Pattern Anal. Mach. Intell. 35(5), 1193–1205 (2013)


20474946, 2021, 6, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/bme2.12029 by Universitaet Duisburg-Essen, Wiley Online Library on [26/05/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

144. Hu, J., Lu, J., Tan, Y.-P.: Robust partial face recognition using instance-to-class distance. In: Visual Communications and Image Processing, 1–6 (2013)
145. Weng, R., et al.: Robust feature set matching for partial face recognition. Proceedings of the IEEE International Conference on Computer Vision, 601–608 (2013)
146. Weng, R., Lu, J., Tan, Y.-P.: Robust point set matching for partial face recognition. IEEE Trans. Image Process. 25(3), 1163–1176 (2016)
147. He, L., et al.: Dynamic feature learning for partial face recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 7054–7063 (2018)
148. He, L., et al.: Dynamic feature matching for partial face recognition. IEEE Trans. Image Process. 28(2), 791–802 (2019)
149. Zheng, W.-S., et al.: Partial person re-identification. In: Proceedings of the IEEE International Conference on Computer Vision, 4678–4686 (2015)
150. Park, J.-S., et al.: Glasses removal from facial image using recursive error compensation. IEEE Trans. Pattern Anal. Mach. Intell. 27(5), 805–811 (2005)
151. Jia, H., Aleix, M.M.: Face recognition with occlusions in the training and testing sets. In: 8th IEEE 2008 International Conference on Automatic Face & Gesture Recognition, 1–6. IEEE (2008)
152. Zhou, Z., et al.: Face recognition with contiguous occlusion using markov random fields. In: IEEE 12th International Conference on Computer Vision, 1050–1057. IEEE (2009)
153. Deng, Y., Dai, Q., Zhang, Z.: Graph laplace for occluded face completion and recognition. IEEE Trans. Image Process. 20(8), 2329–2338 (2011)
154. Li, Y., Feng, J.: Reconstruction based face occlusion elimination for recognition. Neurocomputing. 101, 68–72 (2013)
155. Li, X.-X., et al.: Structured sparse error coding for face recognition with occlusion. IEEE Trans. Image Process. 22(5), 1889–1900 (2013)
156. Ou, W., et al.: Robust face recognition via occlusion dictionary learning. Pattern Recogn. 47(4), 1559–1572 (2014)
157. Zhao, S., Hu, Z.-P.: A modular weighted sparse representation based on Fisher discriminant and sparse residual for face recognition with occlusion. Inf. Process Lett. 115(9), 677–683 (2015)
158. Michael, I., et al.: Robust and low-rank representation for fast face identification with occlusions. IEEE Trans. Image Process. 26(5), 2203–2218 (2017)
159. Wang, M., et al.: Varying face occlusion detection and iterative recovery for face recognition. J. Electron. Imag. 26(3), 033009 (2017)
160. Cheng, L., et al.: Robust deep auto-encoder for occluded face recognition. In: Proceedings of the 23rd ACM International Conference on Multimedia, 1099–1102. ACM (2015)
161. Zhao, F., et al.: Robust lstm-autoencoders for face de-occlusion in the wild. IEEE Trans. Image Process. 27(2), 778–790 (2018)
162. Criminisi, A., Pérez, P., Toyama, K.: Region filling and object removal by exemplar-based image inpainting. IEEE Trans. Image Process. 13(9), 1200–1212 (2004)
163. Sandesh, V.K., Mundada, S.G.: an approach for restoring occluded images for face-recognition. Int. J. Adv. Res. Comput. Commun. Eng. 4(17), 571–575 (2015)
164. Vijayalakshmi, A.: Recognising faces with partial occlusion using inpainting. Int. J. Comput. Appl. 168(13), 20–24 (2017)
165. Kingma, D.P., Welling, M.: Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114 (2013)
166. van den Oord, V., Kalchbrenner, N., Kavukcuoglu, K.: Pixel recurrent neural networks. arXiv preprint arXiv:1601.06759 (2016)
167. Goodfellow, I., et al.: Generative adversarial nets. In: Advances in Neural Information Processing Systems, 2672–2680 (2014)
168. Goodfellow, I.: Nips 2016 tutorial: generative adversarial networks. arXiv preprint arXiv:1701.00160 (2016)
169. Radford, A., Metz, L., Chintala, S.: Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint arXiv:1511.06434 (2015)
170. Makhzani, A., et al.: Adversarial autoencoders. arXiv preprint arXiv:1511.05644 (2015)
171. Xie, J., Xu, L., Chen, E.: Image denoising and inpainting with deep neural networks. In: Advances in Neural Information Processing Systems, 341–349 (2012)
172. Elad, M., Aharon, M.: Image denoising via sparse and redundant representations over learnt dictionaries. IEEE Trans. Image Process. 15(12), 3736–3745 (2006)
173. Pathak, D., et al.: Context encoders: feature learning by inpainting. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2536–2544 (2016)
174. Iizuka, S., Simo-Serra, E., Ishikawa, H.: Globally and locally consistent image completion. ACM Trans. Graph. 36(4), 1–14 (2017)
175. Li, Y, et al.: Generative face completion. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 3911–3919 (2017)
176. Yang, C., et al.: High-resolution image inpainting using multi-scale neural patch synthesis. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 6721–6729 (2017)
177. Ulyanov, D., Vedaldi, A., Lempitsky, V.: Deep image prior. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 9446–9454 (2018)
178. Liu, G., et al.: Image inpainting for irregular holes using partial convolutions. In: Proceedings of the European Conference on Computer Vision, 85–100 (2018)
179. Chen, X., et al.: Infogan: interpretable representation learning by information maximising generative adversarial nets. In: Advances in neural information processing systems, 2172–2180 (2016)
180. Chen, Y.-A., et al.: Occlusion-aware face inpainting via generative adversarial networks. In: IEEE 2017 International Conference on Image Processing (ICIP), 1202–1206. IEEE (2017)
181. Chen, X., et al.: From eyes to face synthesis: a new approach for human-centred smart surveillance. IEEE access. 6, 14567–14575 (2018)
182. He, Z., et al.: Attgan: facial attribute editing by only changing what you want. IEEE Trans. Image Process. 28(11), 5464–5478 (2019)
183. Hu, B., Yang, W., Ren, M.: Unsupervised eyeglasses removal in the wild. arXiv preprint arXiv:1909.06989 (2019)
184. Ge, S., et al.: Occluded face recognition in the wild by identity-diversity inpainting. IEEE Trans. Circ. Syst. Video Technol. 30(10), 3387–3397 (2020)
185. Aleix, M., Robert, B.: The ar face database. CVC Techcal Report. 24 (1998)
186. Gao, W., et al.: the cas-peal large-scale Chinese face database and baseline evaluations. IEEE Trans. Syst. Man Cybern. Syst. Hum. 38(1), 149–161 (2008)
187. Athinodoros, S.G., Belhumeur, P.N., Kriegman, D.J.: From few to many: illumination cone models for face recognition under variable lighting and pose. IEEE Trans. Pattern Anal. Mach. Intell. 23(6), 643–660 (2001)
188. Klare, B.F., et al.: 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, 1931–1939 (2015)
189. Singh, M., et al.: Recognising disguised faces in the wild. IEEE Transactions on Biometrics, Behavior, and Identity Science (TBIOM). 1(2), 97–108 (2019)
190. Xiao, L., et al.: Extracting sparse error of robust pca for face recognition in the presence of varying illumination and occlusion. Pattern Recogn. 47(2), 495–508 (2014)

**How to cite this article:** Zeng, D., Veldhuis, R., Spreeuwers, L.: A survey of face recognition techniques under occlusion. IET Biom. 10(6), 581–606 (2021). https://doi.org/10.1049/bme2.12029

20474946, 2021, 6, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/bme2.12029 by Universitaet Duisburg-Essen, Wiley Online Library on [26/05/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

## APPENDIX

A glossary of abbreviations and expansions for terminologies is illustrated in Table 9.

**TABLE 9** A glossary of abbreviations and expansions for terminologies

| Abbrev | Expansion |
|---|---|
| AAE | Adversarial autoEncoder |
| AFC | Adaptive feature convolution |
| AOFD | Adversarial occlusion-aware face detection |
| CSLBP | Centre-symmetric local binary patterns |
| DA | Denosing autoencoder |
| DCNNs | Deep convolutional neural networks |
| DDRC | Deep dictionary representation based classification |
| DMSC | Discriminative multiscale sparse coding |
| DoG | Different of Gaussian filters |
| DPM | Deformable part models |
| DSCNNs | Double supervision convolutional neural network |
| EBGM | Elastic bunch graph matching |
| ECSLBP | Enhanced centre-symmetric local binary patterns |
| ELOC | Efficient locality-constrained occlusion coding |
| ERBF | Ensemble of radial basis function |
| ERGAN | Eyeglasses removal generative adversarial network |
| FAN | Face attention network |
| FANet | Feature agglomeration networks |
| FCN | Fully convolutional network |
| FW-PCA | Fast weighted-principal component analysis |
| GAN | Generative adversarial network |
| GRRC | Gabor based robust representation and classification |
| HOG | Histogram of oriented gradient |
| ICA | Independent component analysis |
| ID-GAN | Identity-diversity generative adversarial network |
| InfoGAN | Information maximising generative adversarial network |
| KCFA | Kernel correlation feature analysis |
| KLD-LGBP | Kullback-Leibler Divergence-local Gabor binary patterns |
| LBP | Local binary patterns |
| LDA | Linear discriminant analysis |
| LGBPHS | Local Gabor binary pattern histogram sequences |
| LLE | Locally linear embedding |
| LSTM | Long short term memory |
| MDSCNN | Multiscale double convolutional neural network |
| MKD | Multi-keypoint descriptors |
| NNAODL | Nuclear norm adapted occlusion dictionary learning |
| NMF | Non-negative matrix factorization |
| PCA | Principal component analysis |
| PDSN | Pairwise differential siamese network |
| R-CNN | Regions with convolutional neural networks |
| RCSLBP | Reinforced centrosymmetric local binary pattern |
| RDLRR | Robust discriminative low-rank representation |
| RLA | Robust LSTM-Autoencoders |
| RPSM | Robust point set matching |
| SIFT | Scale invariant feature transform |
| SOC | Structured occlusion coding |
| SOM | Self-organising maps |
| SRC | Sparse representation classifiers |
| SSD | Single-shot multibox detector |
| SSEC | Structured sparse error coding |
| SSRC | Structured sparse representation classification |
| SVM | Support vector machine |
| VAE | Variational AutoEncoder |
| YOLO | You only look once |

20474946, 2021, 6, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/bme2.12029 by Universitaet Duisburg-Essen, Wiley Online Library on [26/05/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License