electronics

MDPI

Review

Masked Face Recognition Using Deep Learning: A Review

Ahmad Alzu'bi [1,*], Firas Albalas [1], Tawfik AL-Hadhrami [2], Lojin Bani Younis [1] and Amjad Bashayreh [1]

[1] Department of Computer Science, Jordan University of Science and Technology, Irbid 22110, Jordan; faalbalas@just.edu.jo (F.A.); lhbaniyounis19@cit.just.edu.jo (L.B.Y.); amalbashayreh20@cit.just.edu.jo (A.B.)
[2] School of Science and Technology, Nottingham Trent University, Nottingham NG11 8NS, UK; tawfik.al-hadhrami@ntu.ac.uk
* Correspondence: agalzubi@just.edu.jo

Abstract: A large number of intelligent models for masked face recognition (MFR) has been recently presented and applied in various fields, such as masked face tracking for people safety or secure authentication. Exceptional hazards such as pandemics and frauds have noticeably accelerated the abundance of relevant algorithm creation and sharing, which has introduced new challenges. Therefore, recognizing and authenticating people wearing masks will be a long-established research area, and more efficient methods are needed for real-time MFR. Machine learning has made progress in MFR and has significantly facilitated the intelligent process of detecting and authenticating persons with occluded faces. This survey organizes and reviews the recent works developed for MFR based on deep learning techniques, providing insights and thorough discussion on the development pipeline of MFR systems. State-of-the-art techniques are introduced according to the characteristics of deep network architectures and deep feature extraction strategies. The common benchmarking datasets and evaluation metrics used in the field of MFR are also discussed. Many challenges and promising research directions are highlighted. This comprehensive study considers a wide variety of recent approaches and achievements, aiming to shape a global view of the field of MFR.

Keywords: masked face recognition; deep learning; neural networks; occluded face detection; secure authentication

Citation: Alzu'bi, A.; Albalas, F.; AL-Hadhrami, T.; Younis, L.B.; Bashayreh, A. Masked Face Recognition Using Deep Learning: A Review. *Electronics* **2021**, *10*, 2666. https://doi.org/10.3390/electronics10212666

Academic Editor: Martin Reisslein

Received: 7 October 2021
Accepted: 28 October 2021
Published: 31 October 2021

1. Introduction

Face recognition (FR) systems are conventionally presented with primary facial features such as eyes, nose, and mouth, i.e., non-occluded faces. However, a wide range of situations and circumstances impose that people wear masks in which faces are partially hidden or occluded. Such common situations include pandemics, laboratories, medical operations, or immoderate pollution. For instance, according to World Health Organization (WHO) [1] and Centers for Disease Control and Prevention (CDC) [2], the best way to protect people from the COVID-19 virus and avoid spreading or being infected with the disease is wearing face masks and practicing social distancing. Accordingly, all countries in the world require that people wear a protective face mask in public places, which has driven a need to investigate and understand how such face recognition systems perform with masked faces.

However, implementing such safety guidelines earnestly challenges the existing security and authentication systems that rely on FR already put in place. Most of the recent algorithms have been proposed towards determining whether a face is occluded or not, i.e., masked-face detection. Although saving people's lives is compelling, there is an urgent demand to authenticate persons wearing masks without the need to uncover them. For instance, premises access control and immigration points are among many locations where subjects make cooperative presentations to a camera, which raises a problem of face recognition because the occluded parts are necessary for face detection and recognition.

Moreover, many organizations have already developed and deployed the necessary datasets in-house for facial recognition as a means of person authentication or identification

use. Facial authentication, known as 1:1 matching, is an identity proofing procedure that verifies whether someone is who they declare to be. In the performance of a secure authentication, a personal facial image is taken, from which a biometric template is created and compared against an existing facial signature. In contrast, facial identification, known as 1:N matching, is a form of biometric recognition in which a person is identified by comparing and analyzing the individual pattern against a large database of known faces. Unfortunately, occluded faces complicate the subjects to be recognized accurately, thus threatening the validity of current datasets and making such in-house FR systems inoperable.

Recently, the National Institute for Standards and Technology (NIST) [3] presented the performance of a set of face recognition algorithms developed and tuned after the COVID-19 pandemic (post-COVID-19), which follows their first study on pre-COVID-19 algorithms [4]. They concluded that the majority of recognition algorithms evaluated after the pandemic still show a performance degradation when faces are masked. Additionally, the recognition performance deteriorates when both the enrolment and verification images are masked. This imposes the demand to tackle such authentication concerns using more robust and reliable facial recognition systems under different settings. For example, the concerted efforts to apply important facial technologies, e.g., people screening at immigration points, are undefended. Consequently, many leading vendors of such biometric technologies, including NEC [5] and Thales [6], have been forced to adapt their existing algorithms after the coronavirus pandemic in order to improve the accuracy of FR systems applied on persons wearing masks.

In recent years, deep learning technologies have made great breakthroughs in both theoretical progress and practical applications. The majority of FR systems have been shifted to apply deep learning models since the MFR has become a frontier research direction in the field of computer vision [7,8]. However, research efforts had been under way, even before the COVID-19 pandemic, on how deep learning could improve the performance of existing recognition systems with the existence of masks or occlusions. For instance, the task of occluded face recognition (OFR) has attracted extensive attention, and many deep learning methods have been proposed, including sparse representations [9,10], autoencoders [11], video-based object tracking [12], bidirectional deep networks [13], and dictionary learning [14].

Even though it is a crucial part of recognition systems, the problem of occluded face images, including masks, has not been completely addressed. Many challenges are still under thorough investigation and analysis, such as the large computation cost, robustness against image variations and occlusions, and learning discriminating representations of occluded faces. This made the effective utilization of deep learning architectures and algorithms one of the most decisive tasks toward the feasible face detection and recognition technologies. Therefore, facial recognition with occluded images will remain highly controversial for a prolonged period, and great research works will be increasingly presented for MFR and OFR. More implementations will be also continuously enhanced to track the movement of people wearing masks in real time [15–17].

Over the last few years, a rapid growth in the amount of research works has been witnessed in the field of MFR. The task of MFR or OFR has been employed in many applications such as secure authentication at checkpoints [18] and monitoring people with face masks [19]. However, the algorithms, architectures, models, datasets, and technologies proposed in the literature to deal with occluded or masked faces lack a common mainstream of development and evaluation. Additionally, the diversity of deep learning approaches devoted to detecting and recognizing people wearing masks is absolutely beneficial, but there is a demand to review and evaluate the impact of such technologies in this field.

Such important advances with various analogous challenges motivated us to conduct this review study with the aim of providing a comprehensive resource for those interested in the task of MFR. This study focuses on the most current progressing face recognition methods that are designed and developed on the basis of deep learning. The main contribution of this timely study is threefold:

1. To shape and present a generic development pipeline, which is broadly adopted by the majority of proposed MFR systems. A thorough discussion of the main phases of this framework is introduced, in which deep learning is the baseline.
2. To comprehensively review the recent state-of-the-art approaches in the domain of MFR or OFR. The major deep learning techniques utilized in the literature are presented. In addition, the benchmarking datasets and evaluation metrics that are commonly used to assess the performance of MFR systems are discussed.
3. To highlight many advances, challenges, and gaps in this emerging task of facial recognition, thereby providing important insights into how to utilize the current progressing technologies in different research directions. This review study is devoted to serving the community of FR and inspiring more research works.

The rest of this paper is organized as follows: Section 2 prefaces the study scope and some statistics on the existing works, Section 3 introduces the generic pipeline of MFR that is widely adopted in literature, Section 4 summarizes the benchmarking datasets used for MFR, Section 5 presents and discusses the recent state-of-the-art methods devoted for MFR, Section 6 presents the common metrics used in literature to evaluate the performance of MFR algorithms, Section 7 highlights the main challenges and directions in this field with some insights provided to inspire the future research works, and Section 8 concludes this comprehensive study. A taxonomy of the main issues covered in this study is provided in Appendix A and all the abbreviations are listed in Appendix B.

2. Related Studies

Face recognition is one of the most important tasks that has been extensively studied in the field of computer vision. Human faces provide largely better characteristics to recognize the person's identity compared to other common biometric-based approaches such as iris and fingerprints [20]. Therefore, many recognition systems have employed facial recognition features for forensics and security check purposes. However, the performance of FR algorithms is negatively affected by the presence of face disturbances such as occlusions and variation in illumination and facial expressions. For the task of MFR, the traditional methods of FR are confounded with complicated and occluded faces and therefore heighten the demand of adapting them to learn effective masked-face representations.

Since the COVID-19 pandemic, the research efforts in the domain of MFR have been dramatically increased, which have extended the existing FR or OFR methods and achieved promising accuracy results by a large margin. Most importantly, deep learning approaches have increasingly been developed to tackle the challenges of MFR. A search query is performed on the major digital libraries to track the growth of research interest in the tasks of MFR and OFR. A set of search strings is formulated on the leading repositories to find papers in which the use of deep learning techniques in a facial-based recognition context. The search results of MFR articles have been retrieved from IEEE Xplore, Scopus, ACM digital library, Web of Science, Wiley, Ei Compendex, and EBSCOhost. These repositories include popular symposiums, journals, workshops, and conference articles over the last five years.

However, the search queries were tuned on the basis of the goal of this study. The manuscripts and references retrieved from these repositories have been filtered further to generate a list of the most related articles to the task of MFR or OFR. Despite that, our main aim is to review and discuss the deep learning techniques used in the domain of MFR; the rapid evolution in the research works dealing with the task of OFR are highlighted, as demonstrated in Figure 1. It is important to note that the previous works dedicated for the OFR consider the general objects that hide the key facial features such as scarfs, hair styles, eyeglasses, as well as face masks. This work is focused on the face masks as a challenging factor for OFR in the wild using deep learning techniques.

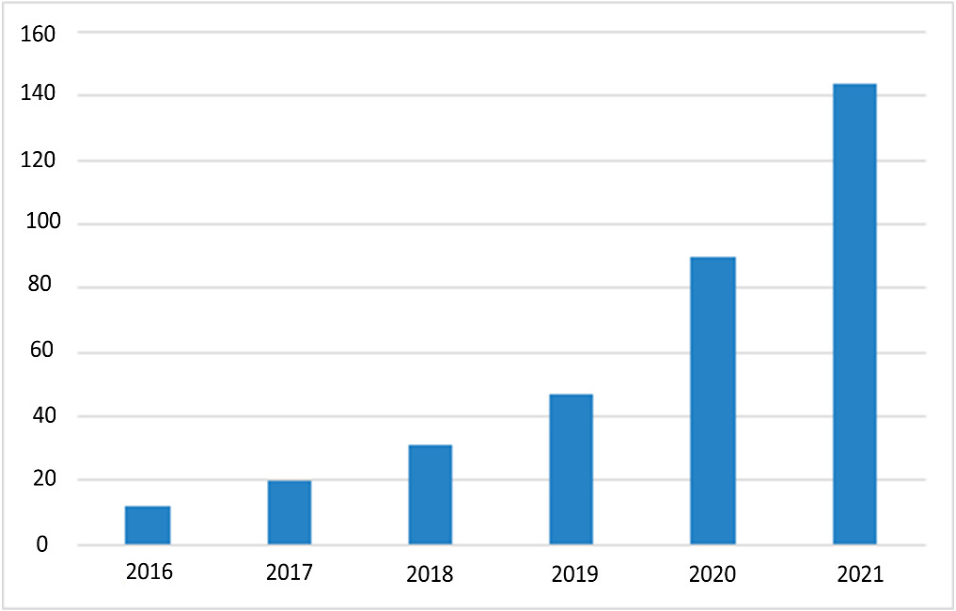


Figure 1. A demonstration of research efforts on MFR from 2016 to 2021.

Exhaustive surveys on FR [20–25] and OFR [3,26–28] have been published in recent years. These studies have set standard algorithmic pipelines and highlighted many important challenges and research directions. However, they focused on traditional methods and deep learning developed for recognizing the face with or without occlusions. To the best of our knowledge, there are no studies that have recently reviewed the domain of MFR with deep learning methods. Moreover, the surveys on OFR have focused on some issues, challenges, and technologies.

The NIST [3] recently reviewed the performance of FR algorithms before and after the COVID-19 pandemic. They evaluated the existing algorithms (pre-pandemic) after tuning them to deal with masked or concluded faces. They showed how these algorithms still tend to perform lower than the satisfactory level. However, this report is a quantitative study limited by reporting the accuracy of recognition algorithms on faces occluded by synthetic masks using two photography datasets collected in U.S. governmental applications, e.g., border crossing photographs. These algorithms were also submitted to NIST with no prior information on whether or not designed with the expectation of occluded faces.

Zeng et al. [26] reviewed the existing face recognition methods that only consider the occluded faces. They categorized the evaluated approaches into three main phases, which are the occlusion feature extraction, occlusion detection, and occlusion recovery and recognition. However, this study considered face masks as one of many occlusion objects and restricted the task of MFR by one dataset. It also generally assessed many algorithms and implementations, including deep learning approaches. Zhang et al. [27] presented a thorough review of facial expression analysis algorithms with partially occluded faces. Face masks were only used in this survey as examples of objects challenging the recognition system of facial expressions. Moreover, deep learning approaches were among six techniques evaluated in the presence of partial occlusion. Lahasan et al. [28] discussed three challenges that affect the FR systems, which are face occlusion, face expressions, and dataset variations. They classified the state-of-the-art approaches into holistic and part-based approaches. The role of several datasets and competitions in tackling these challenges has been also discussed.

Our study is distinguished from the existing surveys by providing a comprehensive review of recent advances and algorithms developed in the scope of MFR. It focuses on deep learning techniques, architectures, and models utilized in the pipeline of MFR, including face detection, unmasking, restoration, and matching. Additionally, the benchmarking datasets and evaluation metrics are presented. Many challenges and insightful research

directions are highlighted and discussed. This timely study would inspire more research works toward providing further improvements in the task of MFR.

3. The MFR Pipeline

This section presents how MFR systems are typically developed through a set of sophisticated phases, as depicted in Figure 2. The generic methodology is mainly based on deep learning models that are widely adopted to learn discriminating features of masked faces. As can be observed from this pipeline, several crucial steps are typically put in place toward developing the final recognition system, as discussed in the following subsections.

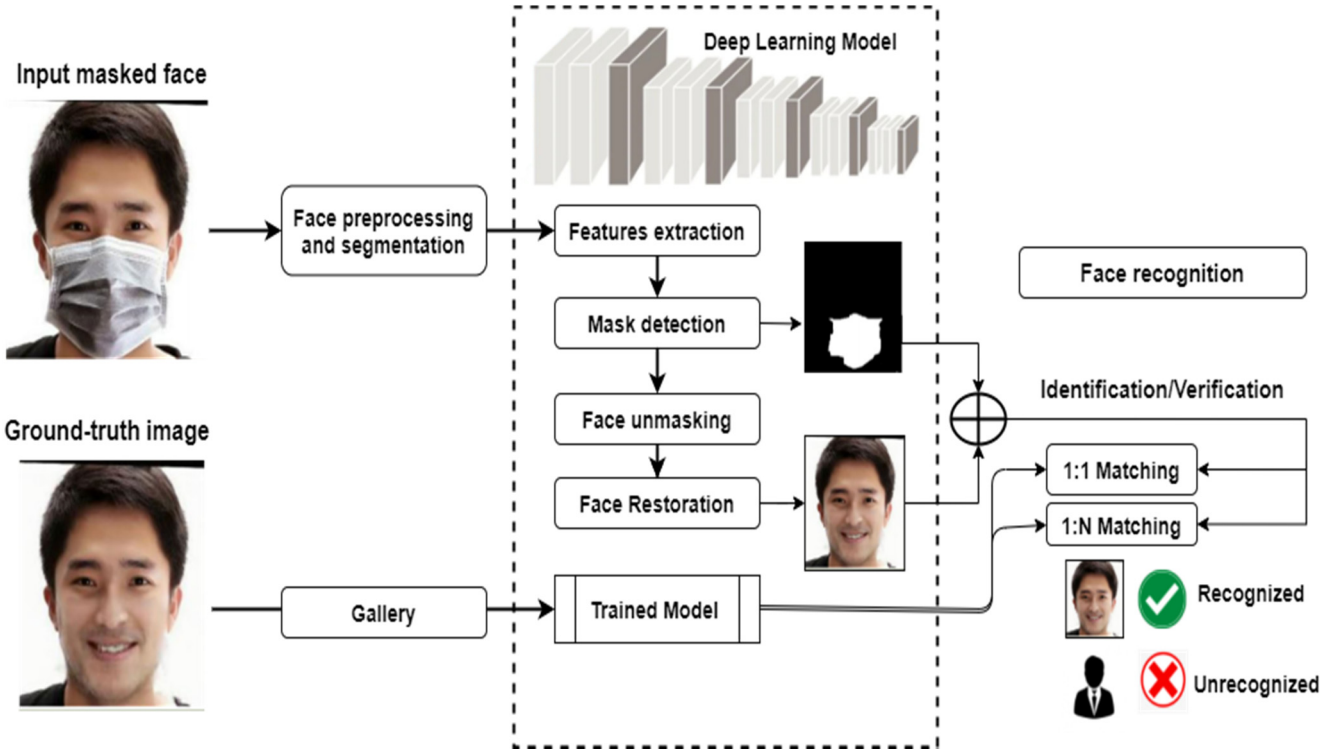


Figure 2. A pictorial representation of the masked face recognition framework.

Firstly, a collection of original masked images with corresponding ground-truth images are prepared. This usually includes splitting them into categorical directories for the purpose of model training, validation, and testing. This is followed by some preprocessing operations such as data augmentation and image segmentation. Then, a set of key facial features are extracted using one or more deep learning models usually pretrained on general-purpose images and fine-tuned on a new collection, i.e., masked faces. Such features should be discriminative enough to detect the face masks accurately. A procedure of face unmasking is then applied in order to restore the masked face and return an estimation of the original face. Finally, the predicted face is matched against the original ground-truth faces to decide whether or not a particular person is identified or verified.

3.1. Image Preprocessing

The performance of FR systems, with or without masks, is largely influenced by the nature of face images used in the training, validation, and testing stages. There are few publicly available datasets that include facial image pairs with and without mask objects to sufficiently train the MFR system in a progressing manner. Therefore, this strengthens the requirement of enriching the testbed by additional synthetic images with various types of face masks [29,30], as well as improving the generalization capability of deep learning models. Among the most popular methods used to synthesize the face masks are

MaskTheFace [31], MaskedFace-Net [32], deep convolutional neural network (DCNN) [33], CYCLE-GAN [34], Identity Aware Mask GAN (IAMGAN) [35], and starGAN [36].

Images have also been widely pre-processed using data augmentation, by which many operations could be applied to enrich the amount and variation of images such as image cropping, flipping, rotation, and alignment. Other augmentation processes are also applied to improve the quality of image representation such as image re-scaling, segmentation, noise removal, or smoothing. Moreover, image adjustment can be performed to improve its sharpness, and the variance of Laplacian [37] is one of the commonly adopted approaches.

To obtain better image representations, several methods segment the image into local parts instantly or semantically, then represent them by the ordered property of facial parts [38] or a set of discriminative components [39,40]. However, some techniques feed the image to an existing tool to detect the facial landmarks [41], while others represent the input still-image by a generic descriptor such as low-rank regularization [42] and sparse representation [43].

3.2. Deep Learning Models

Many well-known methods have been proposed and attempted to recognize human faces by hand-crafted local or global features, such as LBP [44], SIFT [45], and Gabor [46]. However, these holistic approaches suffer from the ability to maintain the uncontrolled facial changes that deviate from their initial assumptions [20]. Later, shallow image representations were introduced, e.g., learning-based dictionary descriptors [47], to improve the distinctiveness and compactness problems of previous methods. Although the accuracy improvements are achieved, these shallow representations still tend to show low robustness against real-world applications and instability against facial appearance variations.

After 2010, deep learning methods were rapidly developed and utilized in a form of multiple deep layers for feature extraction and image transformation. With time, they proved a superiority in learning multiple levels of facial representations that correspond to different levels of abstraction [48], showing solid invariance to the face changes, including lighting, expression, pose, or disguise. Deep learning models are able to combine low-level and high-level abstraction to represent and recognize stable facial identity with strong distinctiveness. In the remaining part of this section, common deep learning models used for masked face recognition are introduced.

3.2.1. Convolutional Neural Networks

Convolutional neural network (CNN) is one of the most effective neural networks that has shown its superiority in a wide range of applications, including image classification, recognition, retrieval, and object detection. CNNs typically consist of cascaded layers to control the degree of shift, scale, and distortion [49], which are input, convolutional, subsampling, fully connected, and output layers. They can efficiently learn various kinds of intra-class differences from training data, such as illumination, pose, facial expression, and age [50]. CNN-based models have been widely utilized and trained on numerous large-scale face datasets [48,51–55].

One of the most popular pretrained architectures that has been successfully used in FR tasks is AlexNet [56]. With the availability of integrated graphics processing units (GPUs), AlexNet decreased the training time and minimized the errors, even with large-scale datasets [57]. VGG16 and VGG19 [58] are also very common CNN-based architectures that have been utilized in various computer vision applications, including face recognition. The VGG-based models typically provide convolution-based features or representations. Despite the remarkable achieved accuracy, they suffer from the training time and complexity [59].

Over time, the task of image recognition became more complex and therefore it should be handled by deeper neural networks. However, if more layers are added to the networks, it becomes more complicated and difficult to train; hence, an accuracy decay is usually encountered. To overcome this challenge, residual network (ResNet) [60] was introduced, which stacks extra layers and accomplishes higher performance and accuracy. The added

layers can learn complex features; however, adding more layers must be empirically determined to control any degradation in the model performance. MobileNet [61] is one of the most important lightweight deep neural networks that mainly depends on a streamline architecture, and it is commonly used for FR tasks. Its architecture showed a high performance with hyperparameters, and the calculations of the model are faster [62].

Inception and its variations [63–65] are also popular CNN-based architecture; their novelty lies in using modules or blocks to build networks that contain convolutional layers instead of stacking them. Xception [66] is an extreme version of inception that replaces the modules of inception with depth-wise separable convolutions. Table 1 summarizes the main characteristics of the popular CNN-based models used in the domain of MFR.

Table 1. A summary of pre-trained CNN-based models.

Model	Variants	Trainable Parameters	Convolutional Layers	Other Layers	Total Layers
AlexNet	-	62 M	5	3	8
VGG	VGG16	138 M	13	3	16
	VGG19	143 M	16	3	19
ResNet	ResNet50	25 M	48	2	50
	ResNet101	44 M	99	2	101
MobileNet	MobileNet	13 M	28	2	30
	MobileNet-v2	3.5 M	-	-	53
Inception	GoogleNet	7 M	22	5	27
	InceptionV2	56 M	22	26	48
	InceptionV3	24 M	22	26	48
	InceptionV4	43 M	-	-	164
	Inception-ResNet-V2	56 M	-	-	164
Xception	-	23 M	36	35	71

3.2.2. Autoencoders

Autoencoder is a popular deep neural network that provides an unsupervised feature learning-based paradigm to efficiently encode and decode the data [67]. Due to its ability in learning robust features from a huge size of unlabeled data automatically [68], noticeable research efforts have been paid to encode the input data into low-dimensional space with significant and discriminative representations, which is accomplished by a decoder. Then, a decoder reverses the process to generate the key features from the encoded stage with backpropagation at the training time [69]. Autoencoders have been effectively utilized for the task of OFR, such as LSTM-autoencoders [70], double channel SSDA (DC-SSDA) [71], de-corrupt autoencoders [72], and 3D landmark-based variational autoencoder [73].

3.2.3. Generative Adversarial Networks

Generative adversarial networks (GANs) [74] are used to automatically explore and learn the regular patterns from the input data without extensively annotated training data. GAN consists of a pair of neural networks: generator and discriminator. The generator uses random values from a given distribution as noisy data and produces new features. The discriminator represents a binary classifier that classifies the generated features and decides whether they are fake or real. GANs are called adversarial due to their adversarial trained setting since the generator and the discriminator seek to optimize an opposing loss function in a minimax game (i.e., a zero-sum game). Another important fact that should be stressed is that the common problems of FR such as face synthesis [75], cross-age face recognition [76], pose invariant face recognition [77], and makeup-invariant face recognition [78] have been addressed using GANs.

3.2.4. Deep Belief Network

Deep belief network (DBN) is a set of multiple hidden units of different layers that are internally connected without connecting the units in the same layer. It typically includes a series of restricted Boltzmann machines (RBMs) or autoencoder where each hidden sub-layer acts as a visible layer for the next hidden sub-layer and the last layer is a softmax layer used in the classification process. DBNs have also utilized in the domain of FR [79] and OFR [80].

3.2.5. Deep Reinforcement Learning

Reinforcement learning learns from the nearby environment; therefore, it emulates the procedure of human decision making by authorizing the agent to choose the action from its experiences by trial and error [81]. An agent is an entity that can perceive its environment through sensors and act upon that environment through actuators. The union of deep learning and reinforcement learning is effectively applied in deep FR such as attention-aware [82] and margin-aware [15] methods.

3.2.6. Specific MFR Deep Networks

Many deep learning architectures have been specifically developed or tuned for the task of FR or OFR, and they noticeably contributed to the performance improvement. FaceNet [83] maps images to Euclidean space via deep neural networks, which builds face embeddings according to the triplet loss. When the images belong to the same person, the distance between them will be small in the Euclidean space while the distance will be large if those images belong to different people. This feature enables FaceNet to work on different tasks such as face detection, recognition, and clustering [84]. SphereFace [8] is another popular FR system that is rendering geometric interpretation and enabling CNNs to learn angularly discriminative features, which makes it efficient in face representation learning. ArcFace [7] is also an effective FR network based on similarity learning that replaces softmax loss with an angular margin loss. It calculates the distance between images using cosine similarity to find the smallest distance.

Deng et al. [85] have also proposed MFCosface as a MFR algorithm on the basis of the large margin cosine loss. It efficiently overcomes the problem of low recognition rates with mask occlusions by detecting the key facial features of masked faces. MFCosface also relies on the large margin cosine loss. It optimizes the representations of facial features by adding an attention mechanism to the model. VGGFace [48] is a face recognition system that includes a deep convolution neural network for recognition based on VGG-Very-Deep-16 CNN architecture. It also includes a face detector and localizer based on a cascade deformable parts model. DeepID [86] was introduced to learn discriminative deep face representation through classifying large-scale face images into a large number of identities, i.e., face identification. However, the learned face representations are challenged by the significant intrapersonal variations that have been reduced by many DeepID variants, such as joint face identification-verification presented in DeepID + 2 [87].

3.3. Feature Extraction

Feature extraction is a crucial step in the face recognition pipeline that aims at extracting a set of features discriminative enough to represent and learn the key facial attributes such as eyes, mouth, nose, and texture. With the existence of face occlusions and masks, this process becomes more complicated, and the existing face recognition systems need to be adapted to extract representative yet robust facial features. In the context of masked face recognition, the feature extraction approaches can be divided into shallow and deep representation methods.

Shallow feature extraction is a traditional method that explicitly formulates a set of handcrafted features with low learning or optimization mechanisms. Some methods use the handcrafted low-level features to find the occluded local parts and dismiss them from the recognition [88]. LBPs [44], SIFT [45], HOG [89], and codebooks [90] are among the popular

descriptors that represent holistic learning, local features, and shallow learning approaches. In the non-occluded face recognition tasks, they have achieved a noticeable accuracy and robustness against many face changes such as illumination, affine, rotation, scale, and translation. However, the performance of shallow features has shown a degradation while dealing with occluded faces, including face masks, which have been largely outperformed by the deep representations obtained by deep learning models.

Many methods were created and evaluated to extract features from faces using deep learning. Li et al. [91] assumed that the features of masked faces often include mask region-related information that should be modeled individually and learned two centers for each class instead of only one, i.e., one center for the full-face images and one for the masked face images. Song et al. [92] introduced a multi-stage mask learning strategy that is mainly based on CNN, by which they aimed at finding and dismissing the corrupted features from the recognition. Many other attention-aware and context-aware methods have extracted the image features using an additional subnet to acquire the important facial regions [93–95].

Graph image representations with deep graph convolutional networks (GCN) have also been utilized in the domain of masked face detection, reconstruction, and recognition [96–98]. GCNs have shown high capabilities in learning and handling face images using spatial or spectral filters that are built for a shared or fixed graph structure. However, learning the graph representations is commonly restricted with the number of GCN layers and the unfavorable computational complexity. The 3D space features have been also investigated for the task of occluded or masked 3D face recognition [34,99,100]. The 3D face recognition methods mimic the real vision and understanding of the human face features, and therefore they can help to improve the performance of the existing 2D recognition systems. The 3D facial features are robust against many face changes such as illumination variations, facial expressions, and face directions.

3.4. Mask Detection

Recently, face masks have become one of the common objects that occlude the facial parts, coming in different styles, sizes, textures, and colors. This strengthens the requirement of training the deep learning models to accurately detect the masks. Most of the existing detection methods, usually introduced for object detection, are tuned and investigated in the task of mask detection. Regions with CNN features (R-CNN) [101] has had a global adoption in the domain of object detection, in which a deep ConvNet is utilized to classify object proposals. In the context of occluded faces, R-CNN extracts thousands of facial regions by feeding them to a CNN network and applying a selective search algorithm, which generates a feature vector for each region. Subsequently, the presence of an object within that candidate facial region proposal from the extracted feature will be classified by support vector machine (SVM). Fast R-CNN [102] and Faster R-CNN [103] were also introduced to enhance the performance by transforming the R-CNN architecture. However, these methods have notable drawbacks such as the training process is a multi-stage pipeline and therefore expensive in terms of space and time. Moreover, the R-CNN slowly performs a ConvNet forward pass for each object proposal without sharing computation. Zhang et al. [104] proposed a context-attention R-CNN as a detection framework of wearing face masks. This framework is used to expand the intra-class distance and reduce the inter-class distance by extracting distinguishing features.

Consequently, more research efforts have been concentrated on using the segmentation-based deep networks for mask detection. Fully convolutional neural network (FCN) [105] is a semantic segmentation architecture that is mainly used with a CNN-based autoencoder that does not contain any dense layers. It is a developed version of a popular classification module by modifying the fully connected layers and replacing them with 1×1 convolution. U-Net [106] has also been firstly introduced for the biomedical image segmentation but widely applied in many computer vision applications, including face detection [107,108]. It includes an encoder that captures the image context using a series of convolutional and

max-pooling layers while a decoder up-samples the encoded information using transposed convolutions. Then, feature maps from the encoder are concatenated to the feature maps of the decoder. This helps in better learning of contextual (relationship between pixels of the image) information.

Other effective methods for MFR or OFR have also been proposed in the literature. Wang et al. [109] introduced a one-shot-based face detector called face attention network (FAN), which utilizes the feature pyramid network to address the occlusion and false positive issue for the faces with different scales. Ge et al. [110] proposed an LLE-CNN to detect masked faces through combining pre-trained CNNs to extract candidate facial regions and represent them with high dimensional descriptors. Then, a locally linear embedding module forms the facial descriptors into vectors of weights to recover any missing facial cues in the masked regions. Finally, the classification and regression tasks employ the weighted vectors as input to identify the real facial regions. Lin et al. [111] introduced the modified LeNet (MLeNet) by increasing the number of units in the output layer and feature maps with a smaller filter size, which in turn further reduces overfitting and increases the performance of masked face detection with a small amount of training images. Alguzo et al. [98] presented multi-graph GCN-based features to detect face masks using multiple filters. They used the embedded geometric information calculated on the basis of distance and correlation graphs to extract and learn the key facial features. Negi et al. [112] detected face masks on the Simulated Masked Face Dataset (SMFD) by proposing CNN- and VGG16-based deep learning models and combining AI-based precautionary measures.

Local features fusion-based deep networks have also been applied to a nonlinear space for masked face detection, as introduced by Peng et al. [113]. Many other detection-based works [19,114,115] have utilized the conventional local and global facial features based on the key face parts, e.g., nose and mouth.

The concept of face mask assistant (FMA) has recently been introduced by Chen et al. [116] as a face detection method based on a mobile microscope. They obtained micro-photos of the face mask, then the globally and locally consistent image completion (GLCM) is utilized to extract texture features and to choose contrast, correlation, energy, and homogeneity as facial features. Fan et al. [117] proposed a deep learning-based single-shot light-weight face mask detector to meet lower computational requirements for embedded systems. They introduced the single-shot light-weight face mask detector (SL-FMDet), which worked effectively due to its low hardware requirements. The lightweight backbone caused a lower feature extraction capability, which was a big obstacle. To solve this problem, the authors extracted rich context information and focused on the crucial face mask-related areas to learn more discriminating features for faces with and without masks. Ieamsaard et al. [118] studied and developed a deep learning model for face mask detection and trained it on YoloV5 at five different epochs. The YoloV5 was used with CNN to verify the existence of face mask and if the mask is placed correctly on the face.

3.5. Face Unmasking

There are various approaches adopted in the literature for object removal, which is the mask in this study context. Several common methods are presented here into learning-based object removal and non-learning-based object removal algorithms.

For learning-based approaches, Shetty et al. [119] proposed a GAN-based model that receives an input image, then removes the target object automatically. Li et al. [120] and Iizuka et al. [121] introduced two different models to learn a global coherency and complete the corrupted region by removing the target object and reconstructing the damaged part using a GAN setup. Khan et al. [122] used a coarse-to-fine GAN-based approach to remove the objects from facial images.

For mask removal, Boutros et al. [123] presented an embedding unmasking model (EUM) that takes a feature embedding extracted from the masked face as input. It generates a new feature embedding similar to an embedding of an unmasked face of the same identity

with unique properties. Din et al. [29,30] used a GAN setup with two discriminators to automatically remove the face mask.

For non-learning approaches, Criminisi et al. [124] introduced a model that removes the undesired part of an image and creates a new region that suits the missing region then matches what is left of the image synthetically. Wang [125] proposed a regularized factor that adjusts the curve of the patch priority function in order to compute the filling order. Park et al. [126] used principal component analysis (PCA) reconstruction and recursive error compensation to remove eyeglasses from facial images. Hays et al. [127] presented an image completion algorithm that depends on a large database of images to search for similar information and embed it into the corrupted pixel of input sample.

3.6. Face Restoration

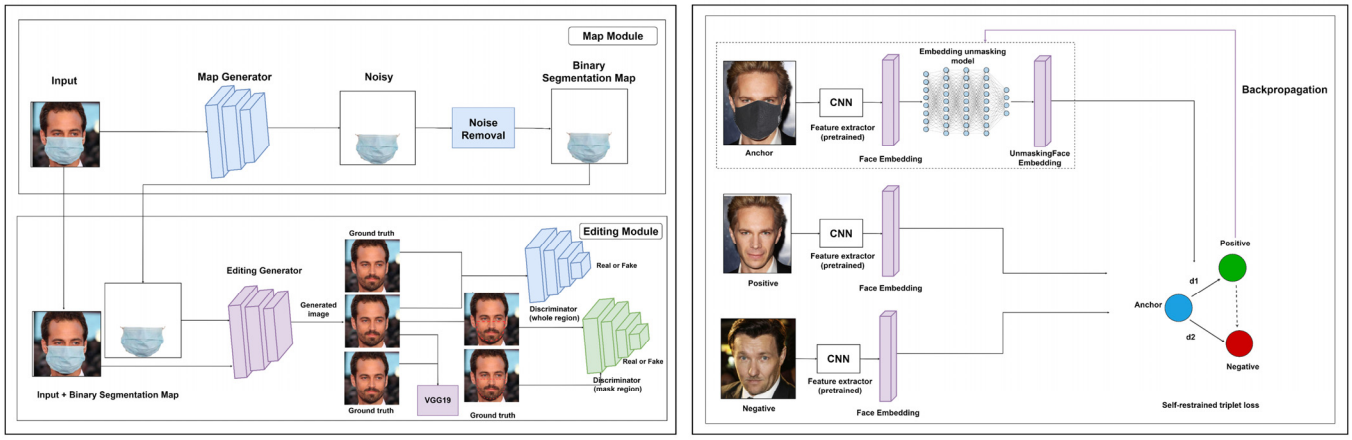
After unmasking the face, any missing parts should be estimated and restored in order to conduct the identity matching process to make the recognition decision, i.e., recognized or unrecognized identity.

One of the pioneering works in image reconstruction is sparse representation-based classification (SRC) [128] for robust OFR. Various variants of SRC were introduced for specific problems in FR, such as the extended SRC (ESRC) for the task of under-sampled FR [129] and the group sparse coding (GSC) [130] for increasing the discriminative ability of face reconstruction. Many other methods have been proposed to reconstruct the missing parts of occluded faces. Yuan et al. [131] used support vector discrimination dictionary and Gabor occlusion dictionary-based SRC (SVGSRC) for OFR. Sparse representation and particle filtering were combined and investigated by Li et al. [132]. Cen et al. [133] also presented a classification scheme based on a depth dictionary representation for robust OFR. A 2D image matrix-based error model named nuclear norm-based matrix regression (NMR) for OFR was also discussed in [134]. A sparse regularized NMR method by introducing L1-norm constraint instead of L2-norm on the representation of the NMR framework was introduced in [135]. However, the image reconstruction methods showed many well-known drawbacks such as the need for an overcomplete dictionary and a large increase in gallery images leading to a complexity problem, as well as their limitation in the generalization capability.

Deep learning methods have addressed such challenges in order to recover the missing part in the facial image. In the last few years, GAN-based methods [120,121,136] have been utilized with global and local discriminators to handle the task of face reconstruction. Yeh et al. [137] used the semantic image inpainting-based data to compute the missing pixels and regions. Yet, they cannot preserve facial identity. Consequently, Zhao et al. [138] introduced a model to retrieve the missing pixel parts under various head poses while trying to preserve the identity on the basis of an identity loss and a pose discriminator in network training. Duan et al. [139] proposed an end-to-end BoostGAN network that consists of three parts: multi-occlusion frontal-view generator, multi-input boosting network, and multi-input discriminator. This approach is equipped with a coarse-to-fine face de-occlusion and frontalization network ensemble. Yu et al. [140] proposed a coarse-to-fine GAN-based approach with a novel contextual attention module for image inpainting. Din et al. [29,30] used GAN-based image inpainting for image completion through an image-to-image translation approach. Duan et al. [141] used GANs to handle the face frontalization and face completion tasks simultaneously. They introduced a two-stage generative adversarial network (TSGAN) and proposed an attention model that is based on occluded masks. Moreover, Luo et al. [142] used GANs to introduce the EyesGAN framework, which is mainly used to construct the face based on the eyes.

Ma et al. [143] presented a face completion method, called learning and preserving face completion network (LP-FCN), to parse face images and extract the features of face identity-preserving (FIP) concurrently. This method is mainly based on CNN, which is trained to transform the FIP features. These features are fused to feed them into a decoder

that generates the complete image. Figure 3 shows two approaches that have been recently proposed to unmask faces and restore the missing facial parts.



(**a**) (**b**)

Figure 3. The general approach of (**a**) GAN-based network with two discriminators and (**b**) EUM model. (**a**) GAN-based face unmasking and restoration [29]. (**b**) EUM-based face unmasking [123].

3.7. Face Matching and Recognition

Face matching by deep features for FR and MFR can be considered as a problem of face verification or identification. In order for this task to be accomplished, a set of images of identified subjects is initially fed to the system during the training and validation phase. In the testing phase, a new unseen subject is presented to the system to make a recognition decision. For a set of deep features or descriptors to be effectively learnt, an adequate loss function should be implemented and applied. There are two common matching approaches adopted by the community of MFR: 1-to-1 and 1-to-N (1-to-many). In both approaches, common distance measures are usually used, such as Euclidean-based L2 and cosine. The procedure of 1-to-1 similarity matching is typically used in face verification, which is applied between the ground-truth image collection and the test image to determine whether the two images refer to the same person, whereas the procedure of 1-to-N similarity matching is employed in face identification that investigates the identity of a specific masked face.

Many methods have been introduced to enhance the discrimination level of deep features with the aim at making the process of face matching more accurate and effective, e.g., metric learning [144] and sparse representations [145]. Deep learning models for matching face identities have widely used the softmax loss-based and triplet loss-based models. Softmax-loss-based models rely on training a multi-class classifier regarding one class for each identity in the training dataset using a softmax function [92,93]. On the other hand, triplet loss-based models [83] are characterized in learning the embedding immediately by matching the results of various inputs to minimize the intra-class distance and therefore maximize the inter-class distance. However, the performance of softmax loss-based and triplet loss-based models suffer from the facemask occlusions [146,147].

Recently, numerous research works have also been presented in the literature to solve the MFR tasks. For instance, effective approaches have shown high FR performance either by GAN-based methods to unmask faces before feeding them to the face recognition model [29,91], by extracting features only from the upper part of the face [147], or by training the face recognition network with a combination of masked and unmasked faces [31,35]. Anwar et al. [31] combined the VGG2 dataset [55] with augmented masked faces and trained the model using the original pipeline defined in FaceNet [83], which in turn enabled the model to distinguish if a face is wearing a mask or not on the basis of the features of the upper half of the face. Montero et al. [148] introduced a full-training pipeline of ArcFace-based face recognition models for MFR. Geng et al. [35] were able to

identify two centers for each identity that match the full-face images and the masked face images sequentially using the domain constrained ranking (DCR).

4. Standard Datasets

This section introduces the common benchmarking datasets used in literature to evaluate the MFR methods. The Synthetic CelebFaces Attributes (Synthetic CelebA) [29] dataset consists of 10,000 synthetic images available publicly. CelebA [149] is a large-scale face attributes dataset with more than 200,000 celebrity images. It was built using 50 types of synthetic masks of various sizes, shapes, colors, and structures. In the building of the synthetic samples, the face was aligned using eye-coordinates for all images, and then the mask was put randomly on the face using Adobe Photoshop.

The Synthetic Face-Occluded Dataset [30] was created using the publicly available CelebA and CelebA-HQ [150] datasets. CelebA-HQ is a large-scale face attribute dataset with more than 30,000 celebrity images. Each face image is cropped and roughly aligned by eye position. The occlusions were synthesized by five popular non-face objects: hands, mask, sunglasses, eyeglasses, and microphone. More than 40 various kinds of each object were used with a variety in sizes, shapes, colors, and structures. Moreover, non-face objects were randomly put on faces.

The Masked Face Detection Dataset (MFDD), Real-World Masked Face Recognition Dataset (RMFRD), and Masked Face Recognition Dataset (SMFRD) were also introduced in [17]. MFDD includes 24,771 images of masked faces to enable the MFR model to detect the masked faces accurately. RMFRD includes 5000 images of 525 people with masks, and 90,000 images of the same people without masks. This dataset is the largest dataset available for MFR. To make the dataset more diverse, researchers introduced SMFRD, which consists of 500,000 images of synthetically masked faces of 10,000 people on the Internet. The RMFRD dataset was used in [151], as the unconscionable face images resulting from incorrect equivalence were manually eliminated. Furthermore, the right face regions were cropped with the help of semi-automatic annotation tools, such as LabelImg and LabelMe.

The Masked Face Segmentation and Recognition (MFSR) dataset [35] consists of two parts. The first part includes 9742 images of masked faces that were collected from the Internet with masked region segmentation annotation that is labeled manually. The second part includes 11,615 images of 1004 identities, where 704 of them are real-world collected and the rest of images were collected from the Internet, in which each identity has at least one image of both masked and unmasked faces. Celebrities in Frontal-Profile in the Wild (CFP) [152] includes faces from 500 celebrities in frontal and profile views. Two verification protocols with 7000 comparisons to each are presented: one compares only frontal faces (FF) and the other compares FF and profile faces (FP).

AgeDB dataset [153] is the first manually gathered dataset in the wild. It includes 16,488 images from 568 celebrities of various ages. It also contains four verification protocols where the compared faces have an age difference of 5, 10, 20, and 30 years. In [154], they created a new dataset by aligning their data with a 3D Morphable Model. It consists of 3D scans for 100 females and 100 males. In [155], they prepared 200 images and classified them then performed their model on two datasets of masked face recognition. They used 100 pictures for a masked face and 100 pictures for an unmasked face.

The MS1MV2 [7] dataset is a refined version of the MS-Celeb1M dataset [52]. MS1MV2 includes 58 million images of 85,000 various identities. Boutros et al. [123] produced a masked version of MS1MV2 noted as MS1MV2-Masked. The mask type and color were randomly chosen for each image to provide the mask color and cover more variations in the training dataset. A subset of 5000 images was randomly chosen from MS1MV2-Masked to verify the model during the training phase. For the evaluation phase, the authors used two real masked face datasets: Masked Faces in Real World for Face Recognition (MFR2) [31] and the Extended Masked Face Recognition (EMFR) datasets [146]. MFR2 includes 269 images of 53 identities taken from the internet. Hence, the images in the

MRF2 dataset can be considered to be captured under in-the-wild conditions. The database includes images of masked and unmasked faces with an average of five images per identity.

The EMFR is gathered from 48 participants using their webcams under three varied sessions: session 1 (reference), session 2, and session 3 (probes). The sessions were captured on three distinct days. The baseline reference (BLR) includes 480 images from the first video of the first session (day). The mask reference (MR) holds 960 images from the second and third videos of the first session. The baseline probe (BLP) includes 960 images from the first video of the second and third sessions and holds face images with no mask. The mask probe (MP) includes 1920 images from the second and third videos of the second and third sessions.

The Labeled Faces in the Wild (LFW) dataset [156] includes 50,000 images approximately. For training, Golwalkar et al. [157] used masked faces of 13 people and 204 images. For testing, they used the same face images but with 25 images of each person. Moreover, the LFW-SM variant dataset was introduced in [31], which extends the LFW dataset with simulated masks, and it contains 13,233 images of 5749 people. Many MFR methods also used the VGGFace2 [55] dataset for training, which consists of 3 million images of 9131 people with nearly 362 images per person. The Masked Faces in the Wild (MFW) mini dataset [37] was created by gathering 3000 images of 300 people from the Internet, containing five images of masked faces and five of unmasked faces for every person. The Masked Face Database (MFD) [158] includes 45 subjects with 990 images of females and males.

In [159], two datasets for MFR were introduced: Masked Face Verification (MFV) that consists of 400 pairs for 200 identities, and Masked Face Identification (MFI) that consists of 4916 images of 669 identities. The Oulu-CASIA NIR-VIS dataset [160] includes 80 identities with six expressions per identity and consists of 48 NIR and 48 VIS images per identity. CASIA NIR-VIS 2.0 [161] contains 17,580 face images with 725 identities, and the BUAA-VisNir dataset [162] consists of images of 150 identities, including nine NIR and nine VIS images for each identity.

VGG-Face2_m [85] is a new version of the VGG-Face dataset. It contains over 3.3 million images of 9131 identities. CASIA-FaceV5_m [85] is a refined version of CASIA-FaceV5, which contains 2500 images of 500 Asian people, with five images for each person.

The Webface dataset [51] is collected from the IMBb and consists of 500,000 images of 10,000 identities. The AR dataset [163] contains 4000 images of 126 identities. It is widely used in various OFR tasks. The Extend Yela B dataset [164] contains 16,128 images of 28 identities under nine poses and 64 illumination conditions. It is widely used in face recognition tasks. Table 2 shows the main characteristics of the datasets used in the masked face recognition task. Figure 4 also shows some sample images taken from common benchmarking MFR datasets.

Table 2. Summary of the common MFR benchmarking datasets.

Dataset	Size (Images)	Identities	Types of Masks
RMFRD	95,000	525	Real-world
SMFRD	500,000	10,000	Synthetic
MFSR	11,615	1004	Real-world/synthetic
LFW-SM	13,233	5749	Synthetic
MFR2	269	53	Synthetic
MFV	400	200	Synthetic
MFI	4916	669	Synthetic
MFD	990	45	Synthetic
MFW-mini	3000	300	Synthetic
CelebA	>200 K	10,177	Synthetic
CelebA-HQ	>30 K	307	Synthetic
MS1MV2-Masked	5.8 M	85,000	Synthetic
CFP-FP	7000	500	Synthetic
CFP-FF	7000	500	Synthetic
CASIA NIR-VIS 2.0	17,580	725	Synthetic

Table 2. *Cont.*

Dataset	Size (Images)	Identities	Types of Masks
Oulu-CASIA NIR-VIS	7680	80	Synthetic
BUAA-VisNir	2700	150	Synthetic
CASIA-FaceV5	2500	500	Synthetic
VGG-Face2_m	3.3 M	9131	Synthetic
Webface	500 K	10,000	Synthetic
AR	4000	126	Synthetic
Extend Yela B	16,128	28	Synthetic
AgeDB	16,488	568	Real-World

Figure 4. Sample images from publicly available benchmarking datasets.

5. State of the Art

This section firstly introduces the existing works proposed for FR with occluded parts, i.e., OFR with any objects including face masks. Then, the research contributions that are specifically presented in the task of MFR are discussed.

5.1. Occluded Face Recognition

Afzal et al. [154] proposed a computationally efficient method to apply feature extraction, depth calculation, and 3D image formulation. They used SIFT to represent the facial features densely. Then, image depth is computed using a multivariate Gaussian distribution. Finally, they determined the shape by applying the shading technique that runs on Lambertian reflectance law, thus recovering high details such as dimples and wrinkles. Din et al. [30] introduced a face de-occlusion technique for facial images in which the user should decide which object to remove. They produced well-incorporated and visual-artifact-free content by using a merged operation of vanilla and partial convolutions in a single network. Moreover, to solve the data insufficiency problem, they built a large synthetic face-occluded paired dataset using openly obtainable CelebA and CelebA-HQ datasets. They concluded that even with a model trained on a synthetic face-occluded

dataset, it efficiently removes non-face objects and provides structurally and perceptually plausible facial content in challenging real images.

Wan et al. [165] proposed a deep trainable module, called MaskNet, to learn formulating the image features with unusual accuracy and neglect those deformed by occlusions. It can be involved in several CNN architectures with limited personal identity labels and less computations. They used real-life and synthetic occluded face images to demonstrate the effectiveness of MaskNet. They trained this network on CASIA-Webface [51], fine-tuned it on AR dataset [163], and finally tested it on the LFW dataset [156]. Song et al. [92] introduced a pairwise differential Siamese network (PDSN) framework that is used to find the equivalence between occluded facial blocks and damaged feature elements for deep CNN models. The system performance was evaluated on face datasets with real-world and synthesized occlusions.

Qiu et al. [166] proposed a face recognition method with occlusions based on a single end-to-end deep neural network, called Face Recognition with Occlusion Masks (FROM). It is used to learn accurate feature masks, to discover the corrupted features using deep CNNs, and then to clean them with dynamically learned masks. Furthermore, the authors train FROM effectively by creating huge, occluded face images. They examined many datasets with occluded or masked faces such as LFW, Megaface challenge 1, RMF2, and AR.

Wang et al. [167] proposed pairwise self-contrastive attention-aware (PSCA) models to extract different local features. The proposed attention sparsity loss (ASL) increases sparse responses in attention maps, thereby decreasing the focus on distracted areas while promoting a focus on discriminative facial parts. They evaluated the recognition performance on several datasets, including LFW, VGGFace2, MS-Celeb-1M, and RMFRD.

Biswas et al. [168] presented a perceptual hashing method, called one-shot frequency dominant neighborhood structure (OSF-DNS). This method showed improvements on the tasks of occluded face verification and face classification. The ability to match occluded faces with their non-occluded versions is beneficial for occluded face verification. Moreover, receiving the identity of an occluded face using a classifier, trained with non-occluded faces and perceptual hash codes as feature vectors, is beneficial for face classification. They created an AERO attacked version of six state-of-the-art datasets: LFW, CUHK [169], MEDS-II [170], CFPW, VGGFace2, and NIMH-ChEFS [171]. Table 3 summarizes the main characteristics of the recent OFR approaches discussed in this subsection.

Table 3. A summary of OFR approaches.

Ref.	Model	Method	Requirements	Dataset
[30]	GANs	Object detection and image completion	Encoder–decoder	CelebA, CelebA-HQ
[154]	Basel face model	2D image detection, 3D face reconstruction	BFM	LFW
[165]	MaskNet	Fusion of MaskNet and CNN architectures	Maxout, ResNet	CASIA-Webface, AR, LFW
[92]	CNN-based PDSN	Calculate equivalence between occluded facial blocks and damaged feature elements	MTCNN, FCN, ResNet	AR, LFW, RMF2
[166]	CNN-based FROM	Learn and remove corrupted features using feature pyramid extractor	Feature pyramid extractor	LFW, Megaface challenge 1, RMF2, AR
[167]	HSNet-61	Extract local features guided by PSCA and ASL	HSNet-61	LFW, VGGFace2, MS-Celeb-1M, RMFRD
[168]	OSF-DNS	Occluded face verification and classification	MTCNN	LFW, CUHK, MEDS-II, CFPW, VGGFace2, NIMH-ChEFS

5.2. Masked Face Recognition

Din [29] proposed a method to remove mask objects from the face automatically and to synthesize the corrupted regions while preserving the initial face structure. They preserved structural and shape consistency in the retrieved face using two discriminators to learn the general face structure of the deep removed area. A synthetic paired dataset is used on the basis of CelebA dataset to solve the data insufficiency problem. Their combined feed-forward model produces structurally and perceptually plausible facial images to challenge real images. Chandra et al. [151] performed a comparable analysis on four state-of-the-art deep learning models, namely, VGGFace, FaceNet, OpenFace, and DeepFace. They concluded that these models show high accuracies in the task of face verification.

Montero et al. [148] presented a deep model based on ArcFace with changes made on the backbone and loss function. From the original face recognition dataset, they generated a masked version using data augmentation and examined ResNet-50 on MFR with low computational cost. ArcFace loss is then combined with the mask-usage classification loss into a new function called Multi-Task ArcFace (MTArcFace).

Hariri [172] proposed deep learning-based features to discard masked regions for MFR. They used pre-trained deep CNNs to select the best features from the captured regions, mostly eyes and forehead regions. Then, the bag-of-features paradigm was applied on the feature maps of the last convolutional layer to quantize the representation. They also used the RMFRD dataset in which three pre-trained deep CNNs—VGG-16, AlexNet, and ResNet-50—are used to select deep features from the captured regions. Maharani et al. [155] presented the MFR approach based on Haar-cascade and MobileNet to detect masks, and then used VGG16 and Triplet loss FaceNet with a multi-threading technique for face identification. Boutros et al. [123] introduced the EUM model that worked on the head of current face recognition models. They used self-restrained triplet (SRT) that allowed EUM to create embeddings related to the unmasked faces of the related characters.

Golwalkar et al. [157] employed the FaceMaskNet-21 network trained using quadruplets with deep metric learning to immediately identify masked faces. The 128-d encodings were generated for every face in the dataset and the input image or live video stream. They also used HOG features to reach more active recognition of faces occluded with a mask. Wang et al. [17] proposed three datasets for masked faces, Real-world Masked Face Recognition Dataset (RMFRD), Masked Face Detection Dataset (MFDD), and Masked Face Recognition Dataset (SMFRD), to handle the MFR tasks that require a sufficient amount of masked and unmasked images. They applied attention weights to the important features available in the uncovered facial parts, such as eye details, forehead, and face contour.

Anwar et al. [31] proposed the MaskTheFace model that synthetically generates a collection of masked faces. MaskTheFace detects face landmarks to identify the key features and tilt of the face. They also used MaskTheFace to recognize the masked faces using the FaceNet system, which adds embeddings to the faces. To train the FaceNet, they used the VGGFace2 dataset and randomly sampled 42 images per person to create a sub dataset, called VGGFace2-mini. From the new subset, they created another subset to include the same images but with masks, called VGGFace2-mini-SM.

Hong et al. [37] presented a 3D model-based approach, called WearMask3D, to augment masked face images of different poses. It fits a 3D morphable model (3DMM) on the image then generates a 3D mask surface to overlay it on the face model. It maps a mask texture to the model and renders the 3D surface to the 2D image on the basis of the image resolution and brightness. They also introduced the Masked Faces in the Wild (MFW) mini dataset and evaluated the model performance on MFW-mini and MFR2. Mandal et al. [173] presented a ResNet-50-based framework to recognize the masked faces. For training, they used the domain adaptation in which they considered the unmasked faces as source domain and the masked faces as target domain. The first scenario was to train the model only on the source domain and test it on the target domain. The second scenario was to train the model on the source domain and a portion of the target domain and test it on the remaining portion of the target domain.

Ejaz et al. [158] proposed the Multi-Task Cascaded Convolutional Neural Network (MTCNN) to detect the masked and unmasked face portions and convert them into high dimensional descriptors. After that, they resized and cropped images using the bounding box as a post-processing step then extracted the main features using FaceNet. The SVM classifier is used to recognize identities. They performed two scenarios, the first being conducted with unmasked faces as an input for training and masked faces as an input for testing, with the second scenario being conducted with both masked and unmasked faces as an input for training and masked faces for testing.

Geng et al. [35] presented a dataset called Masked Face Segmentation and Recognition (MFSR) enriched with more masked faces synthetically as training subjects using Identity Aware Mask GAN (IAMGAN). It is based on the MFSR dataset and consists of a cyclic generator that converts images of full faces into masked faces. However, this module is not effective due to the huge difference in domains and the lack of pairing between masked and unmasked images, leading to generating images without identity recognition. Therefore, this challenging part was addressed by the multi-level identity preserve module. It considers the intra-class variations between masked and unmasked faces by learning class centers using a domain constrained ranking loss (DCR), which assumes that masked faces' features contain information related to the mask region and should be modeled separately. This enabled the model to learn extracting the specific feature of identity and separate identities simultaneously.

Li et al. [91] presented a framework based on de-occlusion distillation to improve the accuracy of MFR. This framework includes two modules: de-occlusion module, which applies a face completion network based on GANs to remove the ambiguity of the appearance in masked faces, which shows the full face without a mask using an attention mechanism to focus on the informative areas of the face. The second module is the distillation that takes a pre-trained face recognition model and adapts its knowledge of faces by knowledge distillation based on VGGFace2. Moreover, they trained the model to classify masks into four classes: simple, complex, human body, and hybrid masks.

Ding et al. [159] introduced two datasets for MFR, Masked Face Verification (MFV) and Masked Face Identification (MFI), which are considered for testing and evaluation purposes. For training, data augmentation was used to generate synthetic masked faces from existing face recognition datasets by aligning faces and masks and detecting predefined facial landmarks. The Delaunay triangulation algorithm was applied to divide images into small triangles where each triangle of a face image has a corresponding mask triangle. For testing, MFV and MFI datasets were used with data augmentation applied to the LFW dataset, called synthesized masked LFW. They also proposed a latent part detection (LPD) model that is inspired by the fact that the human eye focuses on the visible parts, called latent parts, of the masked or occluded faces. However, the features of latent parts need to be discriminative to identities. The LPD model is restricted by the assumption that the masks are always in the lower part of the face.

MFR is also invited to detect and identify criminals who cover their faces. Hong et al. [174] introduced a pedestrian Re-IDentification (ReID) approach that attempts to address the problem of finding an association between masked and unmasked images of the same identity. It re-identifies masked pedestrian images using local and global image features, and then it measures the similarity between the masked and unmasked pedestrian images. FaceNet is used to recognize the person's identity.

Du et al. [175] discussed the near-infrared to visible (NIR-VIS) MFR challenge in terms of training method and data model. They proposed a heterogeneous semi-Siamese training (HSST) approach that aims at maximizing the joint information between the face representation using the semi-Siamese networks. They also presented a face reconstruction-based approach that synthesizes masks for face images from existing datasets.

Wu [176] introduced an attention-based MFR algorithm that separates the mask from the face using a local constrained dictionary learning method. It improves the resolution of images using the dilated convolution and reduces the loss of information using the

attention mechanism. They used ResNet to extract features, which were evaluated on RMFRD and SMFRD datasets. Deng et al. [85] proposed the MFCosface MFR algorithm, based on large margin cosine loss, to detect the key facial features optimized by adding an attention-aware mechanism in the model. Li et al. [177] also proposed an attention-based algorithm and a cropping-based algorithm for MFR. They used the convolutional block attention module (CBAM) [178] in the attention-based part to highlight the region around the eyes. Table 4 summarizes the main characteristics of the recent MFR approaches discussed in this subsection.

Table 4. A summary of MFR approaches.

Work Ref.	Model	Method	Requirements	Dataset
[29]	GANs	Map and editing modules	VGG-19	CelebA
[151]	Pre-trained CNNs	Comparative study on CNNs	VGGFace, Facenet, OpenFace, DeepFace	RMFRD
[148]	MTArcFace	Combination of ArcFace loss and mask-usage classification loss	ArcFace	LFW, CFP, Agedb
[172]	VGG-16, AlexNet, ResNet-50	Deep features of facial areas	VGG-16, AlexNet, ResNet-50	RMFRD, SMFRD
[155]	VGG-16 and FaceNet	Learning cosine distance	-	Collected dataset
[123]	ResNet-50, MobileFaceNet	Embedding unmasking model	FCNN	MS1MV2, MFR, MRF2
[157]	FaceMaskNet-21	Deep metric learning	FaceMaskNet	Collected dataset
[17]	Attention-based	Face-eye-based multi-granularity	-	MFDD, RMFRD
[31]	MaskTheFace	MaskTheFace with FaceNet	FaceNet	VGGFace2-mini-SM, LFW-SM
[37]	WearMask3D	Normalized softmax loss	ResNet-50	MFR2, MFW-mini
[173]	ResNet-50	Domain adaptation	-	RMFRD
[158]	MTCNN	Multi-task cascaded CNN	FaceNet	MFD
[35]	GANs	IAMGAN with DCR	-	MFSR, CASIA-WebFace, VGGFace2
[91]	GANs	De-occlusion distillation	-	Celeb-A, LFW, AR
[159]	Two-branch CNN	Latent part detection	ResNet-50	MFV, MFI, LFW
[174]	MTCNN	Pedestrian re-identification	FaceNet	Pedestrian images
[175]	Siamese networks	Heterogeneous semi-Siamese training	ResNet-50	Oulu-CASIA NIR-VIS, BUAA-VisNir.
[176]	ResNet	Attention-wise	-	RMFRD, SMFRD
[85]	MFCosface	Learning large margin cosine loss	FaceNet	VGGFace2_m, LFW_m, CASIAFaceV5_m, MFR2, RMFD
[177]	CBAM	Face cropping	CBAM	Webface, AR, Yela B, LFW

6. Standard Evaluation Metrics

1 Accuracy: One of the most widely used evaluation metrics for recognition and classification problems. It represents the ratio between the correct number of predictions and the total number of samples, which can be defined as follows:

$$\text{Accuracy} = (\text{TP} + \text{TN})/(\text{TP} + \text{TN} + \text{FP} + \text{FN}), \quad (1)$$

2 Ranked accuracy: Rank-1, Rank-5, and Rank-N are used to measure the performance of DNNs in computer vision. Rank-1 accuracy finds the percentage of correctly classified labels. Rank-5 accuracy is mostly used when there are more than two class labels, which aims to check when the top five most probable labels have the ground truth value. The Rank-N accuracy is similar to Rank-5, but usually used with larger datasets [179].

3 Precision: The ratio of correctly classified positive predictions, which can be defined as follows:

$$\text{Precision} = \text{TP}/(\text{TP} + \text{FP}), \tag{2}$$

4. Mean average precision (mAP): A popular performance measurement metric used in computer vision, specifically for object detection, classification, and localization. It can be generally calculated by taking the mean average precision over all classes and the overall intersection over union (IoU) thresholds [180].
5. Structural similarity index (SSIM): Used to measure the observed quality of digital images and videos. Moreover, it is applied for estimating the similarity between two images. The measurement or prediction of image quality is based on an initial uncompressed or distortion-free image as a reference, and therefore the SSIM index is considered as a full reference metric. It can be defined as follows:

$$\text{SSIM}(x, y) = \frac{\left(2\mu_x\mu_y + c_1\right)\left(2\sigma_{xy} + c_2\right)}{\left(\mu_x^2 + \mu_y^2 + c_1\right)\left(\sigma_x^2 + \sigma_y^2 + c_2\right)}, \tag{3}$$

where μ denotes the mean value of a given image and σ is the standard deviation of the image; x and y represent the two images being compared; c_1 and c_2 are constants to guarantee stability when the divisor becomes 0 [181].

6. Peak signal-to-noise ratio (PSNR): Represents the ratio between the maximum achievable power of a signal and the power of corrupting noise that influences the accuracy of its representation. PSNR is regularly shown as a logarithmic quantity using the decibel scale due to the availability of many signals that have a very wide dynamic range. Moreover, it is widely applied to quantify reconstruction quality of images and video subject to lossy compression. The dimensions of the original image matrix and the degraded image matrix must be the same [182]. It can be defined as follows:

$$\text{PSNR} = 20\log_{10}\left(\frac{\text{MAX}_f}{\sqrt{\text{MSE}}}\right), \tag{4}$$

where MAX_f is the maximum signal value that exists in the original image, and mean squared error (MSE) is calculated as follows:

$$\text{MSE} = \frac{1}{mn}\sum_{0}^{m-1}\sum_{0}^{n-1} \| f(i,j) - g(i,j) \|^2, \tag{5}$$

where f represents the matrix data of the original image, g represents the matrix data of the degraded image, m denotes the numbers of pixel rows of the image, i denotes the index of each row, n represents the number of pixel columns of the image, and j represents the index of each column.

7. Fréchet inception distance (FID): A metric applied to evaluate the quality of images produced by a generative model such as GANs. As opposed to the earlier inception score (IS), which works exclusively on estimating the distribution of generated images, the FID matches the distribution of generated images with the distribution of real images used to train the generator [181], where the lower the FID the higher the quality of the image. It can be defined as follows:

$$\text{FID} = \| \mu_r - \mu_g \|^2 + \text{T}_r(\textstyle\sum_r + \sum_g - 2\left(\sum_r \sum_g\right)^{1/2}), \tag{6}$$

where r and g are the real and fake embeddings, and μ_r and μ_g are the magnitudes of the vectors r and g. T_r is the trace of the matrix, and $\sum_r$ and $\sum_g$ represent the covariance matrix of vectors [183].

8. Error rate (ERR): ERR or misclassification rate is the complement of the accuracy metric. This metric describes the number of misclassified samples from both positive and negative classes. It is sensitive to imbalanced data, which is the same as the accuracy metric. It can be calculated as follows:

$$\text{ERR} = 1 - \text{Accuracy}, \tag{7}$$

9. Equal error rate (EER): A biometric security algorithm applied to determine the common value of its false acceptance rate (FAR) and its false rejection rate (FRR). If the rates are equal, the average value is pointed to as the equal error rate. EER value shows that the proportion of false acceptances is equal to the proportion of false rejections. The lower the EER value, the higher the accuracy of the biometric system [184]. False positive rate (FPR) is a measure of the accuracy used to define the ratio between the wrongly classified negative samples to the total number of negative samples. False negative rate (FNR) is a measure of the accuracy used to define the ratio of positive samples that were wrongly classified [185].

$$\text{FAR} = \text{FPR} = \text{FP}/(\text{FP} + \text{TN}), \tag{8}$$

$$\text{FRR} = \text{FNR} = \text{FN}/(\text{FN} + \text{TP}), \tag{9}$$

$$\text{ERR} = (\text{FAR} + \text{FRR})/2, \tag{10}$$

10. False discovery rate (FDR): The predicted ratio of the number of false-positive classifications (false discoveries) to the total number of positive classifications (rejections of the null). The total number of rejections of the null involves both the number of FP and TP [181]. FDR can be simply computed as follows:

$$\text{FDR} = \text{FP}/(\text{FP} + \text{TP}), \tag{11}$$

11. Geometric mean (G-Mean): Estimates the balance between classification performances on both the majority and minority classes. A low G-Mean is evidence of bad performance in the classification of the positive cases, even though the negative cases are perfectly classified. This measure is necessary for the evasion of overfitting the negative class and underfitting the positive class. Sensitivity is used to measure the accuracy of positive cases. On the other hand, the specificity is used to measure the accuracy of negative cases [186].

$$\text{Sensitivity} = \text{TP}/(\text{TP} + \text{FN}), \tag{12}$$

$$\text{Specificity} = \text{TN}/(\text{TN} + \text{FN}), \tag{13}$$

$$\text{G} - \text{Mean} = \sqrt{\text{Sensitivity} \times \text{Specificity}}, \tag{14}$$

12. True positive rate (TPR): True positive rate, or recall, indicates the ratio of the correctly classified positive samples to the total number of positive samples. It can be calculated as follows:

$$\text{TPR} = \text{TP}/(\text{FN} + \text{TP}), \tag{15}$$

13. False alarm rate (FAR): Also known as false positive rate (FPR), it calculates the ratio between the negative samples that are incorrectly classified to the total number of the negative samples. It is the complement of specificity measure. True negative rate (TNR) is the inverse recall used to measure the ratio of the rightly classified negative

samples to the total number of negative samples [187]. The FPR and TNR are referred to as the verification accuracy that can be defined as follows:

$$\mathrm{TNR} = \mathrm{TN}/(\mathrm{FP} + \mathrm{TN}), \tag{16}$$

$$\mathrm{FAR} = 1 - \mathrm{TNR} = \mathrm{FP}/(\mathrm{TN} + \mathrm{FP}), \tag{17}$$

Table 5 summarizes the performance of MFR methods in terms of accuracy, and Table 6 summarizes the performance of MFR methods in terms of ranked accuracy. Table 7 lists various types of performance metrics applied by the MFR methods.

Table 5. A comparison of accuracies achieved by MFR approaches.

Ref.	Model	Value (%)	Ref.	Model	Value (%)
[151]	CNNs	68.17	[174]	MTCNN + FaceNet	64.23
[148]	MTArcFace	99.78	[165]	MaskNet	93.80
[172]	CNNs	91.30	[167]	HSNet-61	91.20
[155]	VGG16 + FaceNet	100	[168]	OSF-DNS	99.46
[157]	FaceMaskNet-21	88.92	[128]	MFCosface	99.33
[176]	Attention-based	95.00	[177]	Cropping-based	92.61
[17]	Attention-based	95.00	[35]	GANs	86.50
[31]	FaceNet	97.25	[91]	GANs	95.44
[173]	ResNet-50	47.00	[159]	LPD	97.94
[158]	MTCNN	98.50			

Table 6. A comparison of ranked accuracies achieved by MFR approaches.

Work Ref.	Model	Rank-1 (%)	Rank-5 (%)	Rank-10 (%)
[35]	GANs	68.10	77.40	80.60
[159]	LPD	87.12	93.70	94.97
[174]	MTCNN + FaceNet	91.46	95.51	-
[175]	Siamese networks	98.60	-	-
[92]	CNN-based PDSN	100	-	-

Table 7. A comparison of other evaluation measures achieved by MFR approaches.

Work Ref.	Model	Metric	Value (%)
[151]	CNNs	Precision	60.17
[35]	GANs	mAP	42.70
[159]	LPD	mAP	75.92
[174]	MTCNN + FaceNet	mAP	85.62
[29]	GANs	SSIM	93.00
[30]	GANs	SSIM	91.00
[154]	Basel face model (BFM)	SSIM	0.986
[29]	GANs	PSNR	28.241 dB
[30]	GANs	PSNR	28.727 dB
[29]	GANs	FID	6.102
[151]	CNNs	ERR	31.83
[154]	Basel face model (BFM)	ERR	1.36
[123]	ResNet-50 + MobileFaceNet	EER	7.82
[123]	ResNet-50 + MobileFaceNet	FDR	55.96
[123]	ResNet-50 + MobileFaceNet	G-Mean	0.85
[31]	FaceNet	TPR	86.00
[37]	ResNet-50	Verification acc.	88.70
[175]	Siamese networks	Verification acc.	98.58
[92]	CNN-based PDSN	Verification acc.	99.20
[166]	CNN-based FROM	Verification acc.	99.38

7. Research Challenges and Directions

7.1. Categorical MFR

The MFR backbone models have been designed to work with non-masked faces but tuned to deal with masked faces. Therefore, face verification or face identification are typically handled in the literature as binary classification tasks, i.e., recognized on unrecognized identity. Softmax loss has been widely applied to detect face masks or recognize the face itself by training a multi-class classifier by training one class for each identity. Triplet loss is another successful approach used to learn the embedding by comparing different input identities, thus maximizing the inter-class distance. However, more categories can be considered in the detection or recognition phases, such as considering a classifier to estimate the head pose by dividing the facial image with mask into the front and side parts [188], multi-pose masked face detection [114], or human body-part learning [189]. Additionally, specific descriptions of masked faces can be learnt by deep learning models in order to extend the decision of face identification to handle in combination the mask type, face pose, face occlusions, etc.

7.2. Dataset Variations

The use of real-world faces with masks in the benchmarking datasets remains a vital challenge for the effectiveness of MFR systems. Despite the availability of data augmentation and face masking tools that generate synthetic face masks, there is a demand to evaluate the MFR algorithms under different types of real masks including textured masks. It will usually be instructive to specifically measure the performance of real-time MFR algorithms on real-world images collected with actual masks. In addition to the variations that exist in dataset images, there is also a need to develop MFR algorithms dealing with multiple faces or subjects appear in the same image or scene. MFR designs and implementations mainly address a single face with an algorithmic sensitivity to masks. Therefore, more publicly available datasets with sufficient variation in mask types and subjects are expected to be offered to provide a confident decision on the accuracy of MFR algorithms. Moreover, it would be beneficial to consider enriching the training and benchmarking datasets by images with various facial expressions, as presented in [154], to stress the MFR system with operational subjects.

7.3. Non-Cooperative MFR

It is important to mention that most MFR methods do consider people operating in a cooperative manner and are looking at the camera with unconstrained facial imagery. However, acting uncooperatively to cameras is also a popular practice in many sites, e.g., hospitals and public facilities, which challenges the applicability of secure authentication systems. Therefore, matching masked images to unmasked identities requires considering the properties of non-cooperative subjects in which more occluded facial parts are in place. Hong et al. [174] considered the non-cooperative faces but only with the pedestrian images. However, some identification techniques based on other biometrics could be invited to deal with non-cooperative persons wearing masks, such as deep multi-task attention networks for non-cooperative iris recognition [190], heterogeneous palmprint recognition [191,192], and pupil shapes with GAN-generated faces [193].

7.4. Learning Mask Removal and Face Restoration

One of the common gaps that should be highlighted in the domain of MFR is the algorithm's ability in learning mask removal and face restoration efficiently. It is important to note that the non-learning-based MFR algorithms are limited to small object removal from images, while the learning-based algorithms, e.g., GLCM [121], complete the random damaged region in facial images. However, learning this procedure is limited to relatively low image resolutions, producing artifacts for the damaged part located at the margins of image. Even with GFCM [120], face completion suffers when dealing with large removed parts. Some deep learning-based methods [90,121,194,195] only use vanilla convolution as

the backbone of their deep-learning networks. Such convolutional networks apply the same filter weights throughout the image, regardless of whether the region is valid or affected. This helps in achieving well-incorporated predictions but leads to severe visual artifacts, especially at the boundaries of the valid and affected facial regions. Domain-specific deep models such as VGGFace, FaceNet, OpenFace, and DeepFace [151] can also be incorporated with robust face completion algorithms to improve the learning capability of MFR systems.

7.5. 3D Face Reconstruction

More attention will be shifted to the use of 3D facial reconstruction instead of 2D in the MFR systems. The 2D face recognition is still constrained by its sensitivity to pose and illumination of face occlusions. Many existing algorithms have used 3D representations of masked faces, including FID of MaskTheFace [31] and WearMask3D [37] masking methods. Other effective techniques could be also reinvited and investigated for the MFR task, such as masked adaptive projection [196], multi-view recognition [197], and 3D morphable models [198].

7.6. Algorithm Complexity

As in FR systems, MFR systems have to deal with high intra-class variances. Deep learning-based techniques for most MFR scenarios encounter enormous algorithmic complexities during the training phase and therefore require computational power during testing and operation, which is unfavorable for compact devices and real-time systems [199,200]. Feasible solutions are needed to address this challenge in order to achieve higher speed and lower memory at the cost of minimal performance drop. Many effective solutions can be revisited or employed in order to cope with the computational cost of MFR systems. Ge et al. [201] proposed a face recognition deep model trained on limited computational resources. It approximates only the most representative facial cues through feature regression, and it recovers the missing facial cues via a low-resolution face classification. A high-precision and low-latency face alignment network, called MaskFAN [202], is also proposed as a lightweight backbone for masked face alignment with resource-limited devices. It involves a modified loss function and data augmentation module to improve the model performance based on depth-wise separable convolution and group operation.

A pose-specific classification system has been also presented in [203] to provide better classification with low computational cost. Kang et al. [204] applied real-time pupil localization and tracking of drivers wearing facial accessories including masks. It considers the key requirements of low complexity and algorithm performance by classifying images then assigning the appropriate eye tracker. They used a regression-based algorithm for non-occluded faces, while the eye position estimation was applied for occluded face area tracking.

7.7. MFR Competitions

Since there is a lack of publicly available large-scale real-world MFR benchmarking datasets, there are many contests, workshops, and challenge reports that have been proposed with a goal to accelerate the progress of practical MFR. WebFace260M MFR challenge [205] was organized to evaluate the participating MFR algorithms on new large datasets according to a predefined performance threshold. This challenge was also extended by another MFR challenge, called InsightFace [206], in which children test sets including 14,000 identities where a multi-racial test set containing 242,000 identities were provided. MFR competition [207] was also designed to motivate new solutions in enhancing the MFR accuracy, which considered the deployability of the MFR model in terms of compactness. The submitted algorithms were evaluated on a private dataset representing multisession and real masked capture scenarios. Another grand challenge of lightweight 106-point facial landmark localization was organized with the aim at improving the robustness of facial landmark localization of real-world masked faces. The submitted solutions

were evaluated on a new dataset, called JD-landmark-mask. Such public events will contribute to offering more robust MFR solutions along with new benchmarking test sets.

8. Conclusions

This paper has presented a comprehensive survey of the recent MFR works based on deep learning techniques. This study has discussed the generic MFR pipeline adopted over the recent years and has identified the most recent advances that contributed to improving the performance of MFR methods. Many important issues that directly affect MFR systems have been discussed, including image preprocessing, feature extraction, face detection, and localization; face unmasking and restoration; and identity matching and verification. Additionally, some recent interesting and promising techniques have been introduced that are expected to motivate more research efforts to cope with the existing MFR challenges. Most importantly, it can be concluded that the MFR task will be investigated for a prolonged time, and more research and operational works will be continuously proposed in the literature. The adaptation of existing FR methods to be utilized for MFR still tends to show a noticeable performance drop. Considering effective and advanced techniques to pay more attention to the learning ability of deep learning models would be beneficial. The nature of images and test set variations need to be carefully considered in order to improve the generalization capabilities of MFR systems. Moreover, a successful employment of hybrid deep neural networks to learn concurrent tasks, e.g., mask detection and face reconstruction, is important for the MFR accuracy. Metric learning will also positively affect the performance of identity verification or identification.

Author Contributions: Conceptualization, A.A. and F.A.; methodology, A.A., F.A. and T.A.-H.; validation, A.A., F.A. and T.A.-H.; formal analysis, A.A. and F.A.; investigation, A.A., F.A. and T.A.-H.; resources, A.A., F.A. and T.A.-H.; data curation, A.A., L.B.Y. and A.B.; writing—original draft preparation, A.A., L.B.Y. and A.B.; writing—review and editing, F.A. and T.A.-H.; visualization, A.A., F.A., T.A.-H., L.B.Y. and A.B.; supervision, A.A. and F.A.; project administration, A.A. and F.A.; funding acquisition, A.A. and F.A.; correspondence author, A.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Jordan University of Science and Technology, grant number 20210166.

Institutional Review Board Statement: Ethical review and approval were waived for this study, due to that this study reviews the literature of MFR methods and does not involve ethical implications. It includes publicly available datasets with no ethics implications.

Informed Consent Statement: Patient consent was waived due to that this study does not involve ethical implications and reviews public available datasets with no ethics implications.

Data Availability Statement: The photos in this study were taken from the publicly archived dataset, CelebA, https://mmlab.ie.cuhk.edu.hk/projects/CelebA.html (accessed on 28 September 2021).

Acknowledgments: The authors would like to thank Jordan University of Science and Technology for supporting this research work under grant No. 20210166.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. A taxonomy of the main issues addressed in this MFR review.

Issue/Challenge	Related Works
Facial image preprocessing	MaskTheFace [31] MaskedFace-Net [32], DCNN [33], CYCLE-GAN [34], IAMGAN [35], starGAN [36], segments [39–41], regularization [42], sparse rep. [43]
Domain-specific models	FaceNet [83], SphereFace [8], MFCosface [85], VGGFace [48], DeepID [86], LSTM-autoencoders [70], DC-SSDA [71], de-corrupt autoencoders [72], 3D autoencoder [73], pose invariant FR [77], makeup-invariant [78], DBNs [79,80], attention-aware [82], margin-aware [15]
Feature extraction	LBPs [44], SIFT [45], HOG [89], codebooks [90], multi-stage mask learning strategy [92], attention-aware and context-aware [93–95], GCN [96–98]
Mask detection	R-CNN [101], Fast R-CNN [102], Faster R-CNN [103], context-attention R-CNN [104], FCN [105], U-Net [106], FAN [109], LLE-CNNs [110], MLeNet [111], multi-graph GCN-based features [98], FMA [116], SL-FMDet [117]
Face unmasking	GAN-based model [119], coarse-to-fine GAN-based [122], EUM [123], GAN discriminators [29,30], regularized factor [125], PCA reconstruction [126]
Face restoration	SRC [128], extended SRC [129], GSC [130], SVGSRC [131], depth dictionary representation [133], NMR [134], sparse regularized NMR [135], GAN-based methods [120,121,136], semantic inpainting [137], BoostGAN [139], coarse-to-fine GANs [140], GAN-based inpainting [29,30], TSGAN [141], EyesGAN [142], LP-FCN [143], h GFCM [120], GLCM [121]
Identity matching	Metric learning [144], sparse-representations [145], softmax-loss-based [92,93], triplet-loss-based [83], FaceNet [83], ArcFace-based [148], DCR [35]
Non-mask occlusions	GANs [30], Basel face model [154], MaskNet [165], CNN-based PDSN [92], CNN-based FROM [166], HSNet-61 [167], OSF-DNS [168]
Mask de-occlusion	GANs [30,91], pre-trained CNNs [151,172,173,176], MTArcFace [148], EUM [123], FaceMaskNet-21 [157], attention-based [17], MaskTheFace [31], WearMask3D [37], MTCNN [158,174], IAMGAN/DCR [35], LPD [159], Siamese Net. [175], MFCosface [85], CBAM [178]
Dataset setups	Data augmentation and masking tools [31,37], dividing image into front and side parts [188], multi-pose masked face detection [114], deep multi-task attention [190], pupil shapes [193]
3D face reconstruction	MaskTheFace [31], WearMask3D [37], adaptive projection [196], 3D morphable models [198]
Computational cost	MaskFAN [202], pose specific classification [203], real-time pupil localization [204]
MFR community	WebFace260M [205], InsightFace [206], MFR competition [207]

Appendix B

Table A2. The list of all abbreviations used in this manuscript.

Abbr.	Definition	Abbr.	Definition
3DMM	3D Morphable Model	LP-FCN	Learning and preserving face completion network
AI	Artificial intelligence	LPD	Latent part detection
ASL	Attention sparsity loss	LSTM	Long short-term memory
BFM	Basel face model	MEDS-II	Multiple encounter dataset
BLP	the baseline probe	MFCosface	Masked-face recognition of large margin cosine loss
BLR	the baseline reference	MFD	Masked Face Database
CBAM	convolutional block attention module	MFDD	Masked Face Detection Dataset
CDC	centers for disease control and prevention	MFI	Masked Face Identification Dataset
CelebA	Celebfaces attributes	MFR	Masked face recognition
CelebA-HQ	Celebfaces attributes-high quality	LLE-CNN	Locally linear embedding-Cnn
CFP	Celebrities In Frontal-Profile	MFSR	Masked Face Segmentation and Recognition Dataset
CFPW	Celebrities In Frontal-Profile in The Wild	MFV	Is Masked Face Verification Dataset
CNN	Convolutional neural network	MFW	Masked Faces in The Wild Dataset
COVID-19	Coronavirus disease	MLeNet	Modified LeNet

Table A2. *Cont.*

Abbr.	Definition	Abbr.	Definition
CUHK	Chinese University of Hong Kong	MP	The mask probe
DBN	Deep belief network	MR	The mask reference
DCNN	Deep convolutional neural network	MS-Celeb-1M	Microsoft Celeb
DCR	Domain constrained ranking loss	MS1MV2	A Refined Version of the MS-Celeb1M Dataset
DNN	Deep neural network	MTArcFace	Multi-Task Arcface
EER	Equal error rate	MTCNN	Multi-Task Cascaded Convolutional Neural Network
ERR	Error rate	NEC	Nippon Electric Company
ESRC	Extended sparse representation-based classification	NIMH-ChEFS	Nimh Child Emotional Faces Picture Set
EUM	Embedding unmasking model	NIR-VIS	Near-infrared to visible
FAN	Face attention network	NIST	The National Institute for Standards and Technology
FAR	False acceptance rate	NMR	Nuclear norm-based matrix regression
FCN	Fully convolutional neural network	OFR	Occluded face recognition
FDR	False discovery rate	OSF-DNS	One-shot frequency dominant neighborhood structure
FF	Frontal faces	PCA	Principal component analysis
FID	Fréchet inception distance	PDSN	Pairwise differential Siamese network
FIP	Face identity-preserving	PSCA	Pairwise self-contrastive attention-aware
FMA	Face mask assistant	PSNR	Peak signal-to-noise ratio
FNR	False negative rate	RBM	Restricted Boltzmann machines
FP	Profile faces	RCAM	Residual context attention module
FPR	False positive rate	RMFD	Real-World Masked Face Dataset
FR	Face recognition	RMFRD	Real-World Masked Face Recognition Dataset
FROM	Face recognition with occlusion masks	SGHR	Synthesized Gaussian heatmap regression
FRR	False rejection rate	SIFT	Scale-invariant feature transform
GAN	Generative adversarial networks	SL-FMDet	Single-shot light-weight face mask detector
GCN	Graph Convolutional Networks	SMFD	Simulated masked face dataset
GFCM	Generative face completion	SMFRD	Masked Face Recognition Dataset
GLCM	Globally and locally consistent image completion	SRC	Sparse representation-based classification
GPU	Graphics processing units	SRT	Self-restrained triplet
GSC	Group sparse coding	SSIM	Structural similarity index
HOG	Histograms Of oriented gradients	SVGSRC	Support vector and Gabor SRC
HSNET-61	Harmonious multi-scale network	SVM	Support vector machine
HSST	Heterogeneous semi-Siamese training	TNR	True negative rate
IAMGAN	Identity aware mask GAN	TP	True positive
KNN	k-nearest neighbors	TPR	True positive rate
LBP	Local binary pattern	TSGAN	Two-stage generative adversarial network
LFW	Labeled faces in the wild	WHO	World Health Organization

References

1. When and How to Use Masks. Available online: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public/when-and-how-to-use-masks (accessed on 3 August 2021).
2. Coronavirus Disease (COVID-19). Available online: https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/index.html (accessed on 15 September 2021).
3. Ngan, M.L.; Grother, P.J.; Hanaoka, K.K. *Ongoing Face Recognition Vendor Test (FRVT) Part 6A: Face Recognition Accuracy with Masks Using Pre-COVID-19 Algorithms*; NIST Interagency/Internal Report (NISTIR); National Institute of Standards and Technology: Gaithersburg, MD, USA, 2020. [CrossRef]
4. Ngan, M.L.; Grother, P.J.; Hanaoka, K.K. *Ongoing Face Recognition Vendor Test (FRVT) Part 6B: Face Recognition Accuracy with Face Masks Using Post-Covid-19 Algorithms*; NIST Interagency/Internal Report (NISTIR); National Institute of Standards and Technology: Gaithersburg, MD, USA, 2020. [CrossRef]
5. Face Recognition: Biometric Authentication. Available online: https://www.nec.com/en/global/solutions/biometrics/face/ (accessed on 20 August 2021).

6. Biometric Technology to Control COVID-19. Available online: https://www.thalesgroup.com/en/spain/magazine/biometric-technology-control-covid-19 (accessed on 2 September 2021).
7. Deng, J.; Guo, J.; Xue, N.; Zafeiriou, S. Arcface: Additive angular margin loss for deep face recognition. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 4690–4699.
8. Liu, W.; Wen, Y.; Yu, Z.; Li, M.; Raj, B.; Song, L. Sphereface: Deep hypersphere embedding for face recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 212–220.
9. Ouyang, Y.; Sang, N.; Huang, R. Accurate and robust facial expressions recognition by fusing multiple sparse representationbased classifiers. *Neurocomputing* **2015**, *149*, 71–78. [CrossRef]
10. Zeng, S.; Gou, J.; Deng, L. An antinoise sparse representation method for robust face recognition via joint l1 and l2 regularization. *Expert Syst. Appl.* **2017**, *82*, 1–9. [CrossRef]
11. Bian, X.; Li, J. Conditional adversarial consistent identity autoencoder for cross-age face synthesis. *Multimed. Tools Appl.* **2021**, *80*, 14231–14253. [CrossRef]
12. Camuñas-Mesa, L.A.; Serrano-Gotarredona, T.; Ieng, S.-H.; Benosman, R.; Linares-Barranco, B. Event-driven stereo visual tracking algorithm to solve object occlusion. *IEEE Trans. Neural Netw. Learn. Syst.* **2017**, *29*, 4223–4237. [CrossRef] [PubMed]
13. Chen, Y.; Lin, G.; Li, S.; Bourahla, O.; Wu, Y.; Wang, F.; Feng, J.; Xu, M.; Li, X. BANet: Bidirectional aggregation network with occlusion handling for panoptic segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 3793–3802.
14. Ou, W.; You, X.; Tao, D.; Zhang, P.; Tang, Y.; Zhu, Z. Robust face recognition via occlusion dictionary learning. *Pattern Recognit.* **2014**, *47*, 1559–1572. [CrossRef]
15. Liu, B.; Deng, W.; Zhong, Y.; Wang, M.; Hu, J.; Tao, X.; Huang, Y. Fair loss: Margin-aware reinforcement learning for deep face recognition. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Korea, 27–28 October 2019; pp. 10052–10061.
16. Draughon, G.T.; Sun, P.; Lynch, J.P. Implementation of a computer vision frame-work for tracking and visualizing face mask usage in urban environments. In Proceedings of the 2020 IEEE International Smart Cities Conference (ISC2), Piscataway, NJ, USA, 28 September–1 October 2020; pp. 1–8.
17. Wang, Z.; Wang, G.; Huang, B.; Xiong, Z.; Hong, Q.; Wu, H.; Yi, P.; Jiang, K.; Wang, N.; Pei, Y.; et al. Masked face recognition dataset and application. *arXiv* **2020**, arXiv:2003.09093.
18. Militante, S.V.; Dionisio, N.V. Real-time facemask recognition with alarm system using deep learning. In Proceedings of the 2020 11th IEEE Control and System Graduate Research Colloquium (ICSGRC), Shah Alam, Malaysia, 8 August 2020; pp. 106–110.
19. Prasad, S.; Li, Y.; Lin, D.; Sheng, D. MaskedFaceNet: A Progressive Semi-Supervised Masked Face Detector. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, 3–8 January 2021; pp. 3389–3398.
20. Wang, M.; Deng, W. Deep face recognition: A survey. *arXiv* **2018**, arXiv:1804.06655. [CrossRef]
21. Li, S.; Deng, W. Deep facial expression recognition: A survey. *IEEE Trans. Affect. Comput.* **2020**. [CrossRef]
22. Abate, A.F.; Nappi, M.; Riccio, D.; Sabatino, G. 2D and 3D face recognition: A survey. *Pattern Recognit. Lett.* **2007**, *28*, 1885–1906. [CrossRef]
23. Guo, G.; Zhang, N. A survey on deep learning based face recognition. *Comput. Vis. Image Underst.* **2019**, *189*, 102805. [CrossRef]
24. Oloyede, M.O.; Hancke, G.P.; Myburgh, H.C. A review on face recognition systems: Recent approaches and challenges. *Multimed. Tools Appl.* **2020**, *79*, 27891–27922. [CrossRef]
25. Singh, S.; Prasad, S. Techniques and challenges of face recognition: A critical review. *Procedia Comput. Sci.* **2018**, *143*, 536–543. [CrossRef]
26. Zeng, D.; Veldhuis, R.; Spreeuwers, L. A survey of face recognition techniques under occlusion. *arXiv* **2020**, arXiv:2006.11366.
27. Zhang, L.; Verma, B.; Tjondronegoro, D.; Chandran, V. Facial expression analysis under partial occlusion: A survey. *ACM Comput. Surv. (CSUR)* **2018**, *51*, 1–49. [CrossRef]
28. 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.* **2019**, *52*, 949–979. [CrossRef]
29. Din, N.U.; Javed, K.; Bae, S.; Yi, J. A novel GAN-based network for unmasking of masked face. *IEEE Access* **2020**, *8*, 44276–44287. [CrossRef]
30. Din, N.U.; Javed, K.; Bae, S.; Yi, J. Effective Removal of User-Selected Foreground Object from Facial Images Using a Novel GAN-Based Network. *IEEE Access* **2020**, *8*, 109648–109661. [CrossRef]
31. Anwar, A.; Raychowdhury, A. Masked face recognition for secure authentication. *arXiv* **2020**, arXiv:2008.11104.
32. Cabani, A.; Hammoudi, K.; Benhabiles, H.; Melkemi, M. MaskedFace-Net—A dataset of correctly/incorrectly masked face images in the context of COVID-19. *Smart Health* **2021**, *19*, 100144. [CrossRef] [PubMed]
33. Hooge, K.D.O.; Baragchizadeh, A.; Karnowski, T.P.; Bolme, D.S.; Ferrell, R.; Jesu-dasen, P.R.; Castillo, C.D.; O'toole, A.J. Evaluating automated face identity-masking methods with human perception and a deep convolutional neural network. *ACM Trans. Appl. Percept. (TAP)* **2020**, *18*, 1–20.
34. Zhu, J.-Y.; Park, T.; Isola, P.; Efros, A.A. Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2223–2232.

35. Geng, M.; Peng, P.; Huang, Y.; Tian, Y. Masked face recognition with generative data augmentation and domain constrained ranking. In Proceedings of the 28th ACM International Conference on Multimedia, Seattle, WA, USA, 12–16 October 2020; pp. 2246–2254.
36. Choi, Y.; Choi, M.; Kim, M.; Ha, J.-W.; Kim, S.; Choo, J. Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 8789–8797.
37. Hong, J.H.; Kim, H.; Kim, M.; Nam, G.P.; Cho, J.; Ko, H.-S.; Kim, I.-J. A 3D model-based approach for fitting masks to faces in the wild. *arXiv* **2021**, arXiv:2103.00803.
38. Wen, Y.; Zhang, K.; Li, Z.; Qiao, Y. A discriminative feature learning approach for deep face recognition. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016; pp. 499–515.
39. Kortylewski, A.; Liu, Q.; Wang, H.; Zhang, Z.; Yuille, A. Combining compositional models and deep networks for robust object classification under occlusion. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Snowmass, CO, USA, 1–5 March 2020; pp. 1333–1341.
40. Li, H.; Suen, C.Y. Robust face recognition based on dynamic rank representation. *Pattern Recognit.* **2016**, *60*, 13–24. [CrossRef]
41. Zhang, K.; Zhang, Z.; Li, Z.; Qiao, Y. Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Process. Lett.* **2016**, *23*, 1499–1503. [CrossRef]
42. Qian, J.; Yang, J.; Zhang, F.; Lin, Z. Robust low-rank regularized regression for face recognition with occlusion. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Columbus, OH, USA, 23–28 June 2014; pp. 21–26.
43. Yang, M.; Zhang, L.; Yang, J.; Zhang, D. Robust sparse coding for face recognition. In Proceedings of the CVPR 2011, Colorado Springs, CO, USA, 20–25 June 2011; pp. 625–632.
44. Ahonen, T.; Hadid, A.; Pietikainen, M. Face description with local binary patterns: Application to face recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2006**, *28*, 2037–2041. [CrossRef]
45. Geng, C.; Jiang, X. Face recognition using sift features. In Proceedings of the 2009 16th IEEE International Conference on Image Processing (ICIP), Cairo, Egypt, 7–10 November 2009; pp. 3313–3316.
46. Liu, C.; Wechsler, H. Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition. *IEEE Trans. Image Process.* **2002**, *11*, 467–476. [PubMed]
47. Lei, Z.; Pietikäinen, M.; Li, S.Z. Learning discriminant face descriptor. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *36*, 289–302.
48. Parkhi, O.M.; Vedaldi, A.; Zisserman, A. Deep face recognition. In Proceedings of the British Machine Vision Conference (BMVC), Swansea, UK, 7–10 September 2015; pp. 41.1–41.12.
49. Setiowati, S.; Franita, E.L.; Ardiyanto, I. A review of optimization method in face recognition: Comparison deep learning and non-deep learning methods. In Proceedings of the 2017 9th International Conference on Information Technology and Electrical Engineering (ICITEE), Phuket, Thailand, 12–13 October 2017; pp. 1–6.
50. Putra, I. Klasifikasi Citra Menggunakan Convolutional Neural Network (CNN) Pada Caltech101. Thesis, Institut Teknologi Sepuluh Nopember. 2016. Available online: https://repository.its.ac.id/id/eprint/48842 (accessed on 25 September 2021).
51. Yi, D.; Lei, Z.; Liao, S.; Li, S.Z. Learning face representation from scratch. *arXiv* **2014**, arXiv:1411.7923.
52. Guo, Y.; Zhang, L.; Hu, Y.; He, X.; Gao, J. Ms-celeb-1m: A dataset and benchmark for large-scale face recognition. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016; pp. 87–102.
53. 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, Honolulu, HI, USA, 21–26 July 2017; pp. 7044–7053.
54. Bansal, A.; Nanduri, A.; Castillo, C.D.; Ranjan, R.; Chellappa, R. Umdfaces: An annotated face dataset for training deep networks. In Proceedings of the 2017 IEEE International Joint Conference on Biometrics (IJCB), Denver, CO, USA, 1–4 October 2017; pp. 464–473.
55. Cao, Q.; Shen, L.; Xie, W.; Parkhi, O.M.; Zisserman, A. Vggface2: A dataset for recognising faces across pose and age. In Proceedings of the 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition, Xi'an, China, 15–19 May 2018; pp. 67–74.
56. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **2012**, *25*, 1097–1105. [CrossRef]
57. Lu, Y. Image Classification Algorithm Based on Improved AlexNet in Cloud Computing Environment. In Proceedings of the 2020 IEEE International Conference on Industrial Application of Artificial Intelligence (IAAI), Harbin, China, 25–27 December 2020; pp. 250–253.
58. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
59. Gwyn, T.; Roy, K.; Atay, M. Face Recognition Using Popular Deep Net Architectures: A Brief Comparative Study. *Future Internet* **2021**, *13*, 164. [CrossRef]
60. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
61. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; An-dreetto, M.; Adam, H. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv* **2017**, arXiv:1704.04861.
62. Xu, X.; Du, M.; Guo, H.; Chang, J.; Zhao, X. Lightweight FaceNet Based on MobileNet. *Int. J. Intell. Sci.* **2020**, *11*, 1–16.

63. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Van-houcke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1–9.
64. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the inception architecture for computer vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2818–2826.
65. Szegedy, C.; Ioffe, S.; Vanhoucke, V.; Alemi, A.A. Inception-v4, inception-resnet and the impact of residual connections on learning. In Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, San Francisco, CA, USA, 4–9 February 2017.
66. Chollet, F. Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1251–1258.
67. Alom, M.Z.; Taha, T.M.; Yakopcic, C.; Westberg, S.; Sidike, P.; Nasrin, M.S.; Hasan, M.; Van Essen, B.C.; Awwal, A.A.; Asari, V.K. A state-of-the-art survey on deep learning theory and architectures. *Electronics* **2019**, *8*, 292. [CrossRef]
68. Yuan, F.-N.; Zhang, L.; Shi, J.; Xia, X.; Li, G. Theories and applications of auto-encoder neural networks: A literature survey. *Chin. J. Comput.* **2019**, *42*, 203–230.
69. Fuad, M.; Hasan, T.; Fime, A.A.; Sikder, D.; Iftee, M.; Raihan, A.; Rabbi, J.; Al-rakhami, M.S.; Gumae, A.; Sen, O.; et al. Recent Advances in Deep Learning Techniques for Face Recognition. *arXiv* **2021**, arXiv:2103.10492.
70. Zhao, F.; Feng, J.; Zhao, J.; Yang, W.; Yan, S. Robust LSTM-autoencoders for face de-occlusion in the wild. *IEEE Trans. Image Process.* **2017**, *27*, 778–790. [CrossRef]
71. Cheng, L.; Wang, J.; Gong, Y.; Hou, Q. Robust deep auto-encoder for occluded face recognition. In Proceedings of the 23rd ACM International Conference on Multimedia, Brisbane, Australia, 26–30 October 2015; pp. 1099–1102.
72. Zhang, J.; Kan, M.; Shan, S.; Chen, X. Occlusion-free face alignment: Deep regression networks coupled with de-corrupt autoencoders. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 3428–3437.
73. Sharma, S.; Kumar, V. 3D landmark-based face restoration for recognition using variational autoencoder and triplet loss. *IET Biom.* **2021**, *10*, 87–98. [CrossRef]
74. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. *Adv. Neural Inf. Process. Syst.* **2014**, *27*, 2672–2680.
75. Zhao, J.; Xiong, L.; Karlekar, J.; Li, J.; Zhao, F.; Wang, Z.; Pranata, S.; Shen, S.; Yan, S.; Feng, J. Dual-Agent GANs for Photorealistic and Identity Preserving ProfileFace Synthesis. *NIPS* **2017**, *2*, 3.
76. Yang, H.; Huang, D.; Wang, Y.; Jain, A.K. Learning face age progression: A pyramid architecture of gans. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 31–39.
77. Na, I.S.; Tran, C.; Nguyen, D.; Dinh, S. Facial UV map completion for pose-invariant face recognition: A novel adversarial approach based on coupled attention residual UNets. *Hum.-Cent. Comput. Inf. Sci.* **2020**, *10*, 45. [CrossRef]
78. Li, Y.; Song, L.; Wu, X.; He, R.; Tan, T. Learning a bi-level adversarial network with global and local perception for makeup-invariant face verification. *Pattern Recognit.* **2019**, *90*, 99–108. [CrossRef]
79. Li, C.; Wei, W.; Wang, J.; Tang, W.; Zhao, S. Face recognition based on deep belief network combined with center-symmetric local binary pattern. *Adv. Multimed. Ubiquitous Eng.* **2016**, *393*, 277–283. [CrossRef]
80. Chu, J.L.; Krzyźak, A. The recognition of partially occluded objects with support vector machines, convolutional neural networks and deep belief networks. *J. Artif. Intell. Soft Comput. Res.* **2014**, *4*, 5–19. [CrossRef]
81. Littman, M.L. Reinforcement learning improves behaviour from evaluative feedback. *Nature* **2015**, *521*, 445–451. [CrossRef]
82. Zhang, L.; Sun, L.; Yu, L.; Dong, X.; Chen, J.; Cai, W.; Wang, C.; Ning, X. ARFace: Attention-aware and regularization for Face Recognition with Reinforcement Learning. *IEEE Trans. Biom. Behav. Identity Sci.* **2021**. [CrossRef]
83. Schroff, F.; Kalenichenko, D.; Philbin, J. Facenet: A unified embedding for face recognition and clustering. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 815–823.
84. Huang, D.; Wang, C.-D.; Wu, J.-S.; Lai, J.-H.; Kwoh, C.-K. Ultra-scalable spectral clustering and ensemble clustering. *IEEE Trans. Knowl. Data Eng.* **2019**, *32*, 1212–1226. [CrossRef]
85. Deng, H.; Feng, Z.; Qian, G.; Lv, X.; Li, H.; Li, G. MFCosface: A masked-face recognition algorithm based on large margin cosine loss. *Appl. Sci.* **2021**, *11*, 7310. [CrossRef]
86. Sun, Y.; Wang, X.; Tang, X. Deep learning face representation from predicting 10,000 classes. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 1891–1898.
87. Sun, Y.; Wang, X.; Tang, X. Deeply learned face representations are sparse, selective, and robust. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 2892–2900.
88. Min, R.; Hadid, A.; Dugelay, J.-L. Improving the recognition of faces occluded by facial accessories. In Proceedings of the 2011 IEEE International Conference on Automatic Face & Gesture Recognition (FG), Santa Barbara, CA, USA, 21–25 March 2011; pp. 442–447.
89. Dalal, N.; Triggs, B. Histograms of oriented gradients for human detection. In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, 20–26 June 2005; pp. 886–893.
90. Yuan, X.; Park, I.K. Face de-occlusion using 3d morphable model and generative adversarial network. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Korea, 27 October–2 November 2019; pp. 10062–10071.

91. Li, C.; Ge, S.; Zhang, D.; Li, J. Look through masks: Towards masked face recognition with de-occlusion distillation. In Proceedings of the 28th ACM International Conference on Multimedia, Seattle, WA, USA, 12–16 October 2020; pp. 3016–3024.
92. Song, L.; Gong, D.; Li, Z.; Liu, C.; Liu, W. Occlusion robust face recognition based on mask learning with pairwise differential siamese network. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Korea, 27 October–2 November 2019; pp. 773–782.
93. Li, D.; Chen, X.; Zhang, Z.; Huang, K. Learning deep context-aware features over body and latent parts for person re-identification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 384–393.
94. Wang, C.; Zhang, Q.; Huang, C.; Liu, W.; Wang, X. Mancs: A multi-task attentional network with curriculum sampling for person re-identification. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 365–381.
95. Li, W.; Zhu, X.; Gong, S. Harmonious attention network for person re-identification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 2285–2294.
96. Ye, Q.; Li, R. Mask Recognition Method Based on Graph Convolutional Network. *J. Phys. Conf. Ser.* **2021**, *1920*, 012117. [CrossRef]
97. Lin, J.; Yuan, Y.; Shao, T.; Zhou, K. Towards high-fidelity 3D face reconstruction from in-the-wild images using graph convolutional networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 5891–5900.
98. Alguzo, A.; Alzu'bi, A.; Albalas, F. Masked Face Detection using Multi-Graph Convolutional Networks. In Proceedings of the 2021 12th International Conference on Information and Communication Systems (ICICS), Valencia, Spain, 24–26 June 2021; pp. 385–391.
99. Dagnes, N.; Vezzetti, E.; Marcolin, F.; Tornincasa, S. Occlusion detection and restoration techniques for 3D face recognition: A literature review. *Mach. Vis. Appl.* **2018**, *29*, 789–813. [CrossRef]
100. Tran, A.T.; Hassner, T.; Masi, I.; Paz, E.; Nirkin, Y.; Medioni, G.G. Extreme 3D Face Reconstruction: Seeing through Occlusions. In Proceedings of the CVPR, Salt Lake City, UT, USA, 18–23 June 2018; pp. 3935–3944.
101. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 580–587.
102. Girshick, R. Fast r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1440–1448.
103. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. *Adv. Neural Inf. Process. Syst.* **2015**, *28*, 91–99. [CrossRef]
104. Zhang, J.; Han, F.; Chun, Y.; Chen, W. A Novel Detection Framework About Conditions of Wearing Face Mask for Helping Control the Spread of COVID-19. *IEEE Access* **2021**, *9*, 42975–42984. [CrossRef]
105. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 3431–3440.
106. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015; pp. 234–241.
107. Meenpal, T.; Balakrishnan, A.; Verma, A. Facial mask detection using semantic seg-mentation. In Proceedings of the 2019 4th International Conference on Computing, Communications and Security (ICCCS), Rome, Italy, 10–12 October 2019; pp. 1–5.
108. Tian, W.; Wang, Z.; Shen, H.; Deng, W.; Meng, Y.; Chen, B.; Zhang, X.; Zhao, Y.; Huang, X. Learning better features for face detection with feature fusion and segmentation supervision. *arXiv* **2018**, arXiv:1811.08557.
109. Wang, J.; Yuan, Y.; Yu, G. Face attention network: An effective face detector for the occluded faces. *arXiv* **2017**, arXiv:1711.07246.
110. Ge, S.; Li, J.; Ye, Q.; Luo, Z. Detecting masked faces in the wild with lle-cnns. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2682–2690.
111. Lin, S.; Cai, L.; Lin, X.; Ji, R. Masked face detection via a modified LeNet. *Neurocomputing* **2016**, *218*, 197–202. [CrossRef]
112. Negi, A.; Kumar, K.; Chauhan, P.; Rajput, R. Deep Neural Architecture for Face mask Detection on Simulated Masked Face Dataset against Covid-19 Pandemic. In Proceedings of the 2021 International Conference on Computing, Communication, and Intelligent Systems (ICCIS), Greater Noida, India, 19–20 February 2021; pp. 595–600.
113. Peng, X.-Y.; Cao, J.; Zhang, F.-Y. Masked Face Detection Based on Locally Nonlinear Feature Fusion. In Proceedings of the 2020 9th International Conference on Software and Computer Applications, Langkawi, Malaysia, 18–21 February 2020; pp. 114–118.
114. Dewantara, B.S.B.; Rhamadhaningrum, D.T. Detecting multi-pose masked face using adaptive boosting and cascade classifier. In Proceedings of the 2020 International Electronics Symposium (IES), Surabaya, Indonesia, 29–30 September 2020; pp. 436–441.
115. Liu, S.; Agaian, S.S. COVID-19 face mask detection in a crowd using multi-model based on YOLOv3 and hand-crafted features. *Multimodal Image Exploit. Learn.* **2021**, *11734*, 117340M. [CrossRef]
116. Chen, Y.; Hu, M.; Hua, C.; Zhai, G.; Zhang, J.; Li, Q.; Yang, S.X. Face mask assistant: Detection of face mask service stage based on mobile phone. *IEEE Sens. J.* **2021**, *21*, 11084–11093. [CrossRef]
117. Fan, X.; Jiang, M.; Yan, H. A Deep Learning Based Light-Weight Face Mask Detector with Residual Context Attention and Gaussian Heatmap to Fight Against COVID-19. *IEEE Access* **2021**, *9*, 96964–96974. [CrossRef]

118. Ieamsaard, J.; Charoensook, S.N.; Yammen, S. Deep Learning-based Face Mask Detection Using YoloV5. In Proceedings of the 2021 9th International Electrical Engineering Congress (iEECON), Pattaya, Thailand, 10–12 March 2021; pp. 428–431.
119. Shetty, R.; Fritz, M.; Schiele, B. Adversarial scene editing: Automatic object removal from weak supervision. *arXiv* **2018**, arXiv:1806.01911.
120. Li, Y.; Liu, S.; Yang, J.; Yang, M.-H. Generative face completion. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 3911–3919.
121. Iizuka, S.; Simo-Serra, E.; Ishikawa, H. Globally and locally consistent image completion. *ACM Trans. Graph. (ToG)* **2017**, *36*, 107. [CrossRef]
122. Khan, M.K.J.; Ud Din, N.; Bae, S.; Yi, J. Interactive removal of microphone object in facial images. *Electronics* **2019**, *8*, 1115. [CrossRef]
123. Boutros, F.; Damer, N.; Kirchbuchner, F.; Kuijper, A. Unmasking Face Embeddings by Self-restrained Triplet Loss for Accurate Masked Face Recognition. *arXiv* **2021**, arXiv:2103.01716.
124. Criminisi, A.; P'erez, P.; Toyama, K. Region filling and object removal by exemplar-based image inpainting. *IEEE Trans. Image Process.* **2004**, *13*, 1200–1212. [CrossRef] [PubMed]
125. Wang, J.; Lu, K.; Pan, D.; He, N.; Bao, B.-k. Robust object removal with an exemplar-based image inpainting approach. *Neurocomputing* **2014**, *123*, 150–155. [CrossRef]
126. Park, J.-S.; Oh, Y.H.; Ahn, S.C.; Lee, S.-W. Glasses removal from facial image using recursive error compensation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2005**, *27*, 805–811. [CrossRef]
127. Hays, J.; Efros, A.A. Scene completion using millions of photographs. *Commun. ACM* **2008**, *51*, 87–94. [CrossRef]
128. Wright, J.; Yang, A.Y.; Ganesh, A.; Sastry, S.S.; Ma, Y. Robust face recognition via sparse representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2008**, *31*, 210–227. [CrossRef]
129. Deng, W.; Hu, J.; Guo, J. Extended SRC: Undersampled face recognition via intraclass variant dictionary. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *34*, 1864–1870. [CrossRef]
130. Huang, J.; Nie, F.; Huang, H.; Ding, C. Supervised and projected sparse coding for image classification. In Proceedings of the Twenty-Seventh AAAI Conference on Artificial Intelligence, Bellevue, WA, USA, 14–18 July 2013; pp. 438–444.
131. Yuan, L.; Li, F. Face recognition with occlusion via support vector discrimination dictionary and occlusion dictionary based sparse representation classification. In Proceedings of the 2016 31st Youth Academic Annual Conference of Chinese Association of Automation (YAC), Wuhan, China, 11–13 November 2016; pp. 110–115.
132. Li, G.; Liu, Z.-y.; Li, H.-B.; Ren, P. Target tracking based on biological-like vision identity via improved sparse representation and particle filtering. *Cogn. Comput.* **2016**, *8*, 910–923. [CrossRef]
133. Cen, F.; Wang, G. Dictionary representation of deep features for occlusion-robust face recognition. *IEEE Access* **2019**, *7*, 26595–26605. [CrossRef]
134. Yang, J.; Luo, L.; Qian, J.; Tai, Y.; Zhang, F.; Xu, Y. Nuclear norm based matrix regression with applications to face recognition with occlusion and illumination changes. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *39*, 156–171. [CrossRef] [PubMed]
135. Chen, Z.; Wu, X.-J.; Kittler, J. A sparse regularized nuclear norm based matrix regression for face recognition with contiguous occlusion. *Pattern Recognit. Lett.* **2019**, *125*, 494–499. [CrossRef]
136. Chen, Z.; Nie, S.; Wu, T.; Healey, C.G. High resolution face completion with multiple controllable attributes via fully end-to-end progressive generative adversarial networks. *arXiv* **2018**, arXiv:1801.07632.
137. Yeh, R.A.; Chen, C.; Yian Lim, T.; Schwing, A.G.; Hasegawa-Johnson, M.; Do, M.N. Semantic image inpainting with deep generative models. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 5485–5493.
138. Zhao, Y.; Chen, W.; Xing, J.; Li, X.; Bessinger, Z.; Liu, F.; Zuo, W.; Yang, R. Identity preserving face completion for large ocular region occlusion. *arXiv* **2018**, arXiv:1807.08772.
139. Duan, Q.; Zhang, L. Look more into occlusion: Realistic face frontalization and recognition with boostgan. *IEEE Trans. Neural Netw. Learn. Syst.* **2020**, *32*, 214–228. [CrossRef]
140. Yu, J.; Lin, Z.; Yang, J.; Shen, X.; Lu, X.; Huang, T.S. Generative image inpainting with contextual attention. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 5505–5514.
141. Duan, Q.; Zhang, L.; Gao, X. Simultaneous Face Completion and Frontalization via Mask Guided Two-Stage GAN. *IEEE Trans. Circuits Syst. Video Technol.* **2021**. [CrossRef]
142. Luo, X.; He, X.; Qing, L.; Chen, X.; Liu, L.; Xu, Y. EyesGAN: Synthesize human face from human eyes. *Neurocomputing* **2020**, *404*, 213–226. [CrossRef]
143. Ma, R.; Hu, H.; Wang, W.; Xu, J.; Li, Z. Photorealistic face completion with semantic parsing and face identity-preserving features. *ACM Trans. Multimed. Comput. Commun. Appl. (TOMM)* **2019**, *15*, 1–18. [CrossRef]
144. Yu, J.; Hu, C.-H.; Jing, X.-Y.; Feng, Y.-J. Deep metric learning with dynamic margin hard sampling loss for face verification. *Signal Image Video Process.* **2020**, *14*, 791–798. [CrossRef]
145. Yang, S.; Wen, Y.; He, L.; Zhou, M.C.; Abusorrah, A. Sparse Individual Low-rank Component Representation for Face Recognition in IoT-based System. *IEEE Internet Things J.* **2021**. [CrossRef]

146. Damer, N.; Grebe, J.H.; Chen, C.; Boutros, F.; Kirchbuchner, F.; Kuijper, A. The Effect of wearing a mask on face recognition performance: An exploratory study. In Proceedings of the 2020 International Conference of the Biometrics Special Interest Group (BIOSIG), Darmstadt, Germany, 15–17 September 2020; pp. 1–6.
147. Lane, L. NIST finds flaws in facial checks on people with Covid masks. *Biom. Technol. Today* **2020**, *8*, 2. [CrossRef]
148. Montero, D.; Nieto, M.; Leskovsky, P.; Aginako, N. Boosting Masked Face Recognition with Multi-Task ArcFace. *arXiv* **2021**, arXiv:2104.09874,.
149. Liu, Z.; Luo, P.; Wang, X.; Tang, X. Deep learning face attributes in the wild. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–15 December 2015; pp. 3730–3738.
150. Karras, T.; Aila, T.; Laine, S.; Lehtinen, J. Progressive growing of gans for improved quality, stability, and variation. *arXiv* **2017**, arXiv:1710.10196.
151. A Comparative Analysis of Face Recognition Models on Masked Faces. Available online: https://www.ijstr.org/paper-references.php?ref=IJSTR-1020-42646 (accessed on 8 September 2021).
152. Sengupta, S.; Chen, J.-C.; Castillo, C.; Patel, V.M.; Chellappa, R.; Jacobs, D.W. Frontal to profile face verification in the wild. In Proceedings of the 2016 IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Placid, NY, USA, 7–9 March 2016; pp. 1–9.
153. Moschoglou, S.; Papaioannou, A.; Sagonas, C.; Deng, J.; Kotsia, I.; Zafeiriou, S. Agedb: The first manually collected, in-the-wild age database. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Honolulu, HI, USA, 21–26 July 2017; pp. 51–59.
154. Afzal, H.R.; Luo, S.; Afzal, M.K.; Chaudhary, G.; Khari, M.; Kumar, S.A. 3D face reconstruction from single 2D image using distinctive features. *IEEE Access* **2020**, *8*, 180681–180689. [CrossRef]
155. Maharani, D.A.; Machbub, C.; Rusmin, P.H.; Yulianti, L. Improving the Capability of Real-Time Face Masked Recognition using Cosine Distance. In Proceedings of the 2020 6th International Conference on Interactive Digital Media (ICIDM), Bandung, Indonesia, 14–15 December 2020; pp. 1–6.
156. Huang, G.B.; Mattar, M.; Berg, T.; Learned-Miller, E. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. In Proceedings of the Workshop on Faces in Real-Life Images: Detection, Alignment, and Recognition, Marseille, France, 17 October 2008.
157. Golwalkar, R.; Mehendale, N. Masked Face Recognition Using Deep Metric Learning and FaceMaskNet-21. *Soc. Sci. Res. Netw. (SSRN)* **2020**. [CrossRef]
158. Ejaz, M.S.; Islam, M.R. Masked Face Recognition Using Convolutional Neural Network. In Proceedings of the 2019 International Conference on Sustainable Technologies for Industry 4.0 (STI), Dhaka, Bangladesh, 24–25 December 2019; pp. 1–6.
159. Ding, F.; Peng, P.; Huang, Y.; Geng, M.; Tian, Y. Masked face recognition with latent part detection. In Proceedings of the 28th ACM International Conference on Multimedia, Seattle, WA, USA, 12–16 October 2020; pp. 2281–2289.
160. Chen, J.; Yi, D.; Yang, J.; Zhao, G.; Li, S.Z.; Pietikainen, M. Learning mappings for face synthesis from near infrared to visual light images. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 156–163.
161. Li, S.; Yi, D.; Lei, Z.; Liao, S. The casia nir-vis 2.0 face database. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Portland, OR, USA, 23–28 June 2013; pp. 348–353.
162. Huang, D.; Sun, J.; Wang, Y. *The BUAA-VisNir Face Database Instructions*; Tech. Rep. IRIP-TR-12-FR-001; School of Computer Science and Engineering, Beihang University: Beijing, China, 2012; p. 3.
163. Martinez, A.; Benavente, R. The AR Face Database: CVC Technical Report, 24. 1998. Available online: https://www2.ece.ohio-state.edu/~{}aleix/ARdatabase.html (accessed on 5 October 2021).
164. Georghiades, A.S.; 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.* **2001**, *23*, 643–660. [CrossRef]
165. Wan, W.; Chen, J. Occlusion robust face recognition based on mask learning. In Proceedings of the 2017 IEEE international conference on image processing (ICIP), Beijing, China, 17–20 September 2017; pp. 3795–3799.
166. Qiu, H.; Gong, D.; Li, Z.; Liu, W.; Tao, D. End2End occluded face recognition by masking corrupted features. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**. [CrossRef]
167. Wang, Q.; Guo, G. DSA-Face: Diverse and Sparse Attentions for Face Recognition Robust to Pose Variation and Occlusion. *IEEE Trans. Inf. Forensics Secur.* **2021**, *16*, 4534–4543. [CrossRef]
168. Biswas, R.; González-Castro, V.; Fidalgo, E.; Alegre, E. A new perceptual hashing method for verification and identity classification of occluded faces. *Image Vis. Comput.* **2021**, *113*, 104245. [CrossRef]
169. CUHK Face Sketch Database (CUFS). Available online: http://mmlab.ie.cuhk.edu.hk/archive/facesketch.html (accessed on 21 September 2021).
170. Founds, A.P.; Orlans, N.; Genevieve, W.; Watson, C.I. *Nist Special Database 32-Multiple Encounter Dataset II (MEDS-II)*; National Institute of Standards and Technology: Gaithersburg, MD, USA, July 2011; p. 6.
171. Egger, H.L.; Pine, D.S.; Nelson, E.; Leibenluft, E.; Ernst, M.; Towbin, K.E.; An-gold, A. The NIMH Child Emotional Faces Picture Set (NIMH-ChEFS): A new set of children's facial emotion stimuli. *Int. J. Methods Psychiatr. Res.* **2011**, *20*, 145–156. [CrossRef] [PubMed]
172. Hariri, W. Efficient masked face recognition method during the covid-19 pandemic. *arXiv* **2021**, arXiv:2105.03026.

173. Mandal, B.; Okeukwu, A.; Theis, Y. Masked Face Recognition using ResNet-50. *arXiv* **2021**, arXiv:2104.08997.
174. Hong, Q.; Wang, Z.; He, Z.; Wang, N.; Tian, X.; Lu, T. Masked Face Recognition with Identification Association. In Proceedings of the 2020 IEEE 32nd International Conference on Tools with Artificial Intelligence (ICTAI), Baltimore, MD, USA, 9–11 November 2020; pp. 731–735.
175. Du, H.; Shi, H.; Liu, Y.; Zeng, D.; Mei, T. Towards NIR-VIS Masked Face Recognition. *IEEE Signal Process. Lett.* **2021**, *28*, 768–772. [CrossRef]
176. Wu, G. Masked Face Recognition Algorithm for a Contactless Distribution Cabinet. *Math. Probl. Eng.* **2021**, *2021*, 5591020. [CrossRef]
177. Li, Y.; Guo, K.; Lu, Y.; Liu, L. Cropping and attention based approach for masked face recognition. *Appl. Intell.* **2021**, *51*, 3012–3025. [CrossRef]
178. Woo, S.; Park, J.; Lee, J.-Y.; Kweon, I.S. Cbam: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19.
179. Ranked Accuracy. Available online: https://medium.com/analytics-vidhya/ranked-accuracy-11bdaef795e3 (accessed on 9 August 2021).
180. mAP (mean Average Precision) Might Confuse You! Available online: https://towardsdatascience.com/map-mean-average-precision-might-confuse-you-5956f1bfa9e2#:~{}:text=mAP%20stands%20for%20Mean%20Average (accessed on 9 August 2021).
181. All about Structural Similarity Index (SSIM): Theory + Code in PyTorch. Available online: https://medium.com/srm-mic/all-about-structural-similarity-index-ssim-theory-code-in-pytorch-6551b455541e (accessed on 9 August 2021).
182. Signal-to-Noise Ratio as an Image Quality Metric. Available online: https://www.ni.com/en-lb/innovations/white-papers/11/peak-signal-to-noise-ratio-as-an-image-quality-metric.html (accessed on 22 September 2021).
183. How to Evaluate GANs Using Frechet Inception Distance (FID). Available online: https://wandb.ai/ayush-thakur/gan-evaluation/reports/How-to-Evaluate-GANs-using-Frechet-Inception-Distance-FID---Vmlldzo0MTAxOTI (accessed on 22 September 2021).
184. Jahangir, R.; Teh, Y.W.; Nweke, H.F.; Mujtaba, G.; Al-Garadi, M.A.; Ali, I. Speaker Identification through artificial intelligence techniques: A comprehensive review and research challenges. *Expert Syst. Appl.* **2021**, *171*, 114591. [CrossRef]
185. Dalvi, S.; Gressel, G.; Achuthan, K. Tuning the false positive rate/false negative rate with phishing detection models. *Int. J. Eng. Adv. Technol.* **2019**, *9*, 7–13.
186. Akosa, J. Predictive accuracy: A misleading performance measure for highly imbalanced data. In Proceedings of the SAS Global Forum, Orlando, FL, USA, 2–5 April 2017; p. 12.
187. Tharwat, A. Classification assessment methods. *Appl. Comput. Inform.* **2021**, *17*, 168–192. [CrossRef]
188. Li, S.; Ning, X.; Yu, L.; Zhang, L.; Dong, X.; Shi, Y.; He, W. Multi-angle head pose classification when wearing the mask for face recognition under the COVID-19 coronavirus epidemic. In Proceedings of the 2020 International Conference on High Performance Big Data and Intelligent Systems (HPBD&IS), Shenzhen, China, 23–25 May 2020; pp. 1–5.
189. Rida, I. Towards human body-part learning for model-free gait recognition. *arXiv* **2019**, arXiv:1904.01620.
190. Wang, C.; Muhammad, J.; Wang, Y.; He, Z.; Sun, Z. Towards complete and accurate iris segmentation using deep multi-task attention network for non-cooperative iris recognition. *IEEE Trans. Inf. Forensics Secur.* **2020**, *15*, 2944–2959. [CrossRef]
191. Fei, L.; Zhang, B.; Xu, Y.; Tian, C.; Rida, I.; Zhang, D. Jointly Heterogeneous Palmprint Discriminant Feature Learning. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**. [CrossRef]
192. Rida, I.; Al Maadeed, N.; Al Maadeed, S. A novel efficient classwise sparse and collaborative representation for holistic palmprint recognition. In Proceedings of the 2018 NASA/ESA Conference on Adaptive Hardware and Systems (AHS), Edinburgh, UK, 6–9 August 2018; pp. 156–161.
193. Guo, H.; Hu, S.; Wang, X.; Chang, M.-C.; Lyu, S. Eyes Tell All: Irregular Pupil Shapes Reveal GAN-generated Faces. *arXiv* **2021**, arXiv:2109.00162.
194. Yang, C.; Lu, X.; Lin, Z.; Shechtman, E.; Wang, O.; Li, H. High-resolution image inpainting using multi-scale neural patch synthesis. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 6721–6729.
195. Pathak, D.; Krahenbuhl, P.; Donahue, J.; Darrell, T.; Efros, A.A. Context encoders: Feature learning by inpainting. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2536–2544.
196. Alyuz, N.; Gokberk, B.; Akarun, L. 3-D face recognition under occlusion using masked projection. *IEEE Trans. Inf. Forensics Secur.* **2013**, *8*, 789–802. [CrossRef]
197. Alzu'bi, A.; Abuarqoub, A.; Al-Hmouz, A. Aggregated Deep Convolutional Neural Networks for Multi-View 3D Object Retrieval. In Proceedings of the 2019 11th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT), Dublin, Ireland, 28–30 October 2019; pp. 1–5.
198. Egger, B.; Schönborn, S.; Schneider, A.; Kortylewski, A.; Morel-Forster, A.; Blumer, C.; Vetter, T. Occlusion-aware 3d morphable models and an illumination prior for face image analysis. *Int. J. Comput. Vis.* **2018**, *126*, 1269–1287. [CrossRef]
199. Alzu'bi, A.; Abuarqoub, A. Deep learning model with low-dimensional random projection for large-scale image search. *Eng. Sci. Technol. Int. J.* **2020**, *23*, 911–920. [CrossRef]
200. Alzu'bi, A.; Amira, A.; Ramzan, N.; Jaber, T. Improving content-based image retrieval with compact global and local multi-features. *Int. J. Multimed. Inf. Retr.* **2016**, *5*, 237–253. [CrossRef]

201. Ge, S.; Zhao, S.; Li, C.; Li, J. Low-resolution face recognition in the wild via selective knowledge distillation. *IEEE Trans. Image Process.* **2018**, *28*, 2051–2062. [CrossRef]
202. Sha, Y.; Zhang, J.; Liu, X.; Wu, Z.; Shan, S. Efficient Face Alignment Network for Masked Face. In Proceedings of the 2021 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), Shenzhen, China, 5–9 July 2021; pp. 1–6.
203. Asad, M.; Hussain, A.; Mir, U. Low complexity hybrid holistic–landmark based approach for face recognition. *Multimed. Tools Appl.* **2021**, *80*, 30199–30212. [CrossRef]
204. Kang, D.; Chang, H.S. Low-Complexity Pupil Tracking for Sunglasses-Wearing Faces for Glasses-Free 3D HUDs. *Appl. Sci.* **2021**, *11*, 4366. [CrossRef]
205. Zhu, Z.; Huang, G.; Deng, J.; Ye, Y.; Huang, J.; Chen, X.; Zhu, J.; Yang, T.; Guo, J.; Lu, J.; et al. Masked Face Recognition Challenge: The WebFace260M Track Report. *arXiv* **2021**, arXiv:2108.07189.
206. Deng, J.; Guo, J.; An, X.; Zhu, Z.; Zafeiriou, S. Masked Face Recognition Challenge: The InsightFace Track Report. *arXiv* **2021**, arXiv:2108.08191.
207. Han, D.; Aginako, N.; Sierra, B.; Nieto, M.; Erakin, M.; Demir, U.; Ekenel, H.; Kataoka, A.; Ichikawa, K.; Kubo, S.; et al. MFR 2021: Masked Face Recognition Competition. In Proceedings of the 2021 IEEE International Joint Conference on Biometrics (IJCB), Shenzhen, China, 4–7 August 2021; pp. 1–10.