

## Review

# A Comprehensive Survey of Masked Faces: Recognition, Detection, and Unmasking

Mohamed Mahmoud <sup>1,2</sup> , Mahmoud SalahEldin Kasem <sup>1,3</sup> and Hyun-Soo Kang <sup>1,\*</sup> <sup>1</sup> Department of Information and Communication Engineering, School of Electrical and Computer Engineering, Chungbuk National University, Cheongju-si 28644, Republic of Korea<sup>2</sup> Information Technology Department, Faculty of Computers and Information, Assiut University, Assiut 71526, Egypt<sup>3</sup> Multimedia Department, Faculty of Computers and Information, Assiut University, Assiut 71526, Egypt

\* Correspondence: hskang@cbnu.ac.kr

**Abstract:** Masked face recognition (MFR) has emerged as a critical domain in biometric identification, especially with the global COVID-19 pandemic, which introduced widespread face masks. This survey paper presents a comprehensive analysis of the challenges and advancements in recognizing and detecting individuals with masked faces, which has seen innovative shifts due to the necessity of adapting to new societal norms. Advanced through deep learning techniques, MFR, along with face mask recognition (FMR) and face unmasking (FU), represents significant areas of focus. These methods address unique challenges posed by obscured facial features, from fully to partially covered faces. Our comprehensive review explores the various deep learning-based methodologies developed for MFR, FMR, and FU, highlighting their distinctive challenges and the solutions proposed to overcome them. Additionally, we explore benchmark datasets and evaluation metrics specifically tailored for assessing performance in MFR research. The survey also discusses the substantial obstacles still facing researchers in this field and proposes future directions for the ongoing development of more robust and effective masked face recognition systems. This paper serves as an invaluable resource for researchers and practitioners, offering insights into the evolving landscape of face recognition technologies in the face of global health crises and beyond.



**Citation:** Mahmoud, M.; Kasem, M.S.; Kang, H.-S. A Comprehensive Survey of Masked Faces: Recognition, Detection, and Unmasking. *Appl. Sci.* **2024**, *14*, 8781. <https://doi.org/10.3390/app14198781>

Academic Editor: Douglas O'Shaughnessy

Received: 31 August 2024

Revised: 24 September 2024

Accepted: 26 September 2024

Published: 28 September 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

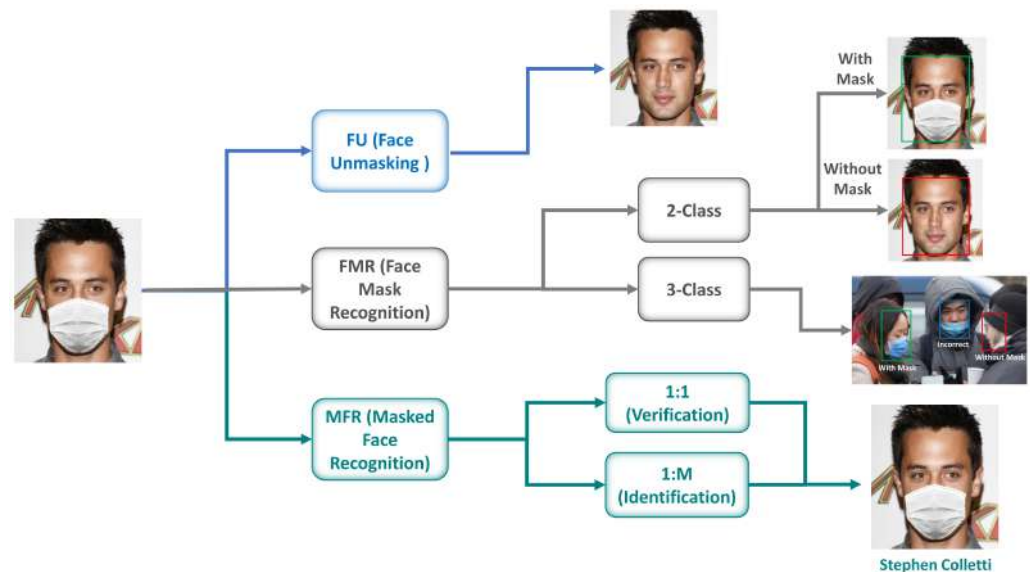
**Keywords:** masked face recognition; masked face identification; masked face verification; face mask removal; face unmasking; face mask recognition; face mask detection; deep learning

## 1. Introduction

In recent years, integrating facial recognition systems across diverse sectors, such as security, healthcare, and human–computer interaction, has revolutionized identity verification and access control. Nevertheless, the widespread adoption of face masks in response to the global COVID-19 pandemic has introduced unprecedented challenges to the remarkable performance of conventional facial recognition technologies. The masking of facial features has spurred research initiatives in masked face recognition, prompting the application of innovative deep learning techniques to address this novel challenge. This sets the stage for exploring advanced strategies in face recognition, particularly under challenging conditions involving small or partially obscured faces [1].

MFR poses a significant challenge in identifying and verifying individuals who wear face masks. This task is complex due to the partial occlusion and variations in appearance caused by facial coverings. Masks obscure critical facial features like the nose, mouth, and chin, and their diverse types, sizes, and colors add to the complexity. It is essential to distinguish between MFR and FMR. While FMR focuses on detecting mask presence, MFR aims to identify and verify individuals wearing masks. Additionally, FU endeavors to remove facial coverings and restore a clear facial representation. Figure 1 visually summarizes

our survey's essence, illustrating three distinct tasks—MFR, FMR, and FU—with different outputs for the same input, highlighting the task-specific outcomes.



**Figure 1.** Illustration showcasing the tasks of masked face recognition (MFR), face mask recognition (FMR), and face unmasking (FU) with varied outputs for the same input.

Deep learning has emerged as a promising avenue for addressing the challenges of MFR. Algorithms can be trained to discern facial features, even when partially obscured by masks. Proposed MFR methodologies based on deep learning encompass holistic approaches, mask exclude-based approaches, and mask removal-based approaches. Holistic approaches employ deep learning models to discern features of entire faces, leveraging attention modules. Mask exclude-based approaches train models to recognize features of the unmasked facial half, such as the eyes and head. Approaches based on mask removal leverage generative adversarial networks (GANs) to create lifelike facial images from their masked counterparts, as demonstrated in methods [2,3], facilitating subsequent recognition. The adaptability of deep learning underscores its crucial role as a transformative technology in the realm of MFR. Although deep learning-based MFR methods have demonstrated state-of-the-art performance on various public benchmark datasets, numerous challenges persist that require resolution before widespread deployment in real-world applications becomes feasible.

### 1.1. Challenges in MFR

The advent of face masks amid the COVID-19 pandemic has posed formidable challenges to facial recognition systems, leading to a significant decline in performance. The concealment of crucial facial features, including the nose, mouth, and chin, has triggered a cascade of obstacles, and addressing these challenges becomes imperative for advancing MFR technology.

- **Scarcity of Datasets:** The scarcity of datasets tailored for masked face recognition constitutes a pivotal challenge. Training any deep learning model requires a robust dataset, yet the shortage of publicly available datasets featuring masked faces complicates the development of effective MFR methods. Researchers tackling this challenge often resort to creating synthetic datasets by introducing masks to existing public face datasets like CASIA-WebFace [4], CelebA [5], and LFW [6]. To simulate masked–unmasked pairs, popular methods involve using deep learning-based tools such as MaskTheFace [7] or leveraging generative adversarial networks like CycleGAN [8]. Manual editing using image software, exemplified by the approach in [3], further supplements dataset generation efforts.

### Training on Synthetic Data:

While synthetic datasets offer a practical solution when real-world masked face datasets are scarce, relying solely on AI-generated data introduces certain challenges. Models trained exclusively on synthetic data may overfit to the specific features or artifacts inherent in the data generation process, rather than learning to recognize real-world occlusions. This can result in a diminished performance when these models are deployed in real-world scenarios involving masked faces. To mitigate this risk, it is essential to balance the use of synthetic datasets with real-world data or apply fine-tuning techniques on real-world samples, ensuring the model's generalizability. Additionally, models trained on AI-generated masks may develop an oversensitivity to synthetic artifacts, further necessitating the use of adversarial training approaches to enhance their robustness across a wide range of mask types, whether synthetic or real.

- **Dataset Bias:** In addition to the scarcity of publicly masked datasets, a prominent challenge lies in the bias inherent in existing benchmark datasets for MFR. Many widely used datasets exhibit a notable skew towards specific demographics, primarily favoring male and Caucasian or Asian individuals. This bias introduces a risk of developing MFR systems that may demonstrate reduced accuracy when applied to individuals from other demographic groups. To mitigate dataset bias in MFR, efforts should be directed towards creating more inclusive and representative benchmark datasets. This involves intentionally diversifying dataset populations to encompass a broader spectrum of demographics, including gender, ethnicity, and age.
- **Occlusion Complexity:** The complexity introduced by facial occlusion, particularly the masking of the mouth, poses a significant hurdle to existing face recognition methods. The diverse sizes, colors, and types of masks exacerbate the challenge, impacting the training of models for various masked face tasks, including recognition, detection, and unmasking. Strategies to address this complexity vary by task. Recognition methods may employ attention models [9,10] that focus on the upper half of the face or exclusively train on this region. Another approach involves using face mask removal methods as a pre-step before recognition. In unmasking tasks, researchers may introduce a pre-stage to detect the mask area, as demonstrated by generating a binary mask map in the first stage in [2]. Training datasets are further diversified by incorporating various mask types, colors, and sizes to enhance model robustness. These nuanced approaches aim to unravel the intricacies posed by occlusions, ensuring the adaptability of masked face recognition methodologies.
- **Real-Time Performance:** Integrating masked face recognition into real-world scenarios poses intricate challenges, given the variability in lighting conditions, diverse camera angles, and environmental factors. Maintaining a consistent performance amid these dynamic variables is a significant hurdle. Practical applicability across diverse settings necessitates real-time capabilities for MFR systems. However, the computational demands of deep learning-based MFR methods present a challenge, particularly when striving for real-time functionality on resource-constrained mobile devices. Addressing these real-time performance challenges involves a strategic optimization approach. Efforts focus on enhancing the efficiency of deep learning models without compromising accuracy.

### 1.2. Applications of MFR

MFR exhibits significant potential in numerous sectors, providing a secure and efficient means of identity verification in scenarios where individuals are wearing face masks. This potential translates into innovative solutions addressing contemporary challenges. This subsection explores the diverse applications where MFR can be leveraged, showcasing its adaptability and relevance.

- **Security and Access Control:** Strengthening security measures to achieve precise identification, especially in scenarios involving individuals wearing masks. Seamlessly

integrating with access control systems to guarantee secure entry across public and private spaces, including restricted areas like airports, government buildings, and data centers. Additionally, implementing facial recognition-based door locks for both residential and office settings, enhancing home and workplace security. Enabling employee authentication protocols for secure entry into workplaces.

- **Public Safety:** MFR plays a crucial role in safeguarding public safety in crowded spaces. Integrated seamlessly with surveillance systems, MFR empowers law enforcement with enhanced monitoring and rapid response capabilities. This technology aids in identifying suspects and missing persons involved in criminal investigations, proactively detects suspicious activity in public areas, swiftly pinpoints individuals involved in disturbances, and strengthens security measures at events and gatherings. MFR's potential to enhance public safety and create a secure environment is undeniable.
- **Healthcare:** Ensuring secure access to medical facilities and patient records, along with verifying the identity of both patients and healthcare workers. Implementing contactless patient tracking to elevate healthcare services while simultaneously fortifying security and privacy within healthcare settings.
- **Retail and Customer Service:** Delivering tailored and efficient customer service by recognizing individuals, even when their faces are partially obscured. Additionally, optimizing payment processes to elevate the overall shopping experience.
- **Human–Computer Interaction:** Facilitating secure and personalized interactions with user-authenticated devices while also improving the user experience across a spectrum of applications, including smartphones, computers, and smart home devices.
- **Workplace and Attendance Tracking:** Facilitating contactless attendance tracking for employees in workplace settings, thereby reinforcing security measures to grant access exclusively to authorized individuals in designated areas.
- **Education Institutions:** Overseeing and securing entry points in educational institutions to safeguard the well-being of students and staff. Streamlining attendance tracking in classrooms and campus facilities for enhanced efficiency.

By exploring these applications and more, it becomes evident that MFR has the potential to revolutionize diverse sectors, providing solutions that cater to the evolving needs of modern society.

Contributions of this survey encompass the following:

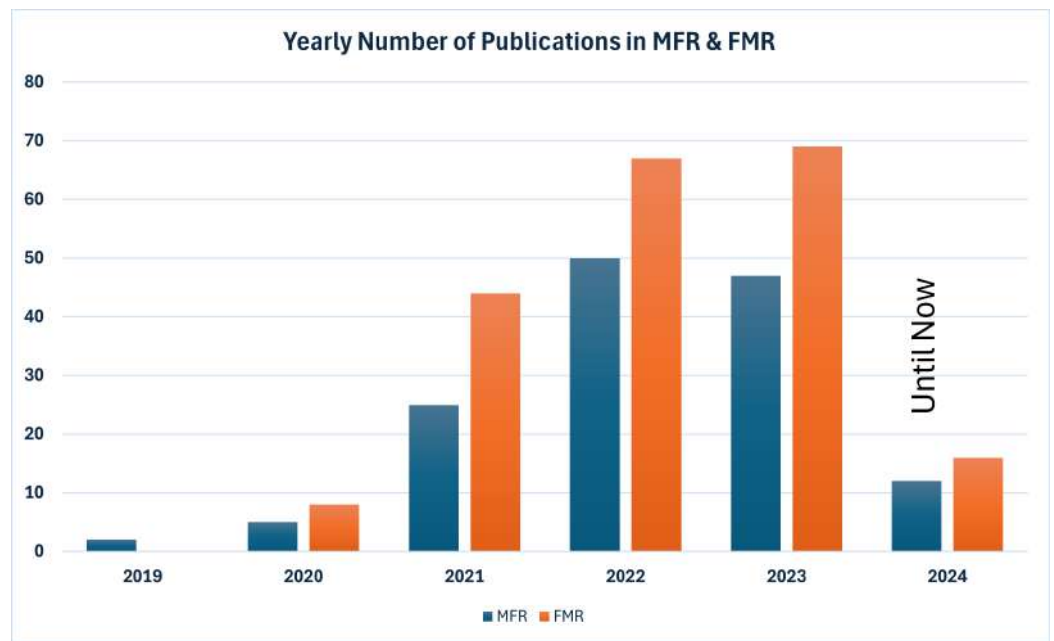
1. An in-depth exploration of MFR, FMR, and FU within the framework of deep learning methodologies, highlighting the challenges inherent in identifying individuals with partially obscured facial features.
2. A comprehensive exploration of evaluation metrics, benchmarking methods, and diverse applications of masked face recognition across security, healthcare, and human-computer interaction domains.
3. A detailed analysis of critical datasets and preprocessing methodologies essential for training robust masked face recognition models.
4. Tracing the evolutionary trajectory of face recognition within the deep learning paradigm, providing insights into the development of techniques tailored for identifying and verifying individuals under various degrees of facial occlusion.

This investigation into masked face recognition, face mask recognition, and face unmasking, grounded in the advancements of deep learning, aspires to furnish a foundational understanding for researchers. Serving as a roadmap, it delineates the current state-of-the-art methodologies and charts prospective avenues for continued research and development in this pivotal area.

## 2. Related Surveys

In this section, we look into previous surveys conducted in the fields of MFR and FMR, which serve as essential repositories of recent research and future directions. Despite the relatively short period following the onset of the COVID-19 pandemic, a considerable body of work and surveys has emerged. Figure 2 presents an analysis of related studies

per year for both MFR and FMR from 2019 to 2024, with data sourced from Scopus. For MFR, the search was conducted using keywords such as “Masked face recognition” and “Masked Faces”, while for FMR, keywords such as “Face mask detection”, “Face masks”, and “Mask detection” were utilized. These surveys offer invaluable insights into the evolution of methodologies, challenges encountered, and advancements made in tackling the intricacies associated with face masks. The objective of this subsection is to provide a succinct overview of select surveys in this domain, thereby situating the current study within the broader context of the existing literature.



**Figure 2.** Illustrates the evolving landscape of MFR and FMR studies from 2019 to 2024. The data were sourced from Scopus using keywords “Masked face recognition” for MFR and “Face mask detection”, “Face masks”, and “Mask detection” for FMR.

Face recognition under occlusion predates the COVID-19 era, indicating that masked face recognition is not a novel field but rather a specialized and intricate subset of occluded face recognition. The complexity of masks, ranging from size and color to shape, adds layers of intricacy to MFR. Several surveys have explored partial face recognition, such as the work by Lahasan Badr et al. [11], which investigates strategies addressing three core challenges in face recognition systems: facial occlusion, single sample per subject (SSPS), and nuances in facial expressions. While offering insights into recent strategies to overcome these hurdles, this survey lacks recent updates and focus on deep learning methods, limiting its applicability to MFR challenges. Zhang Zhifeng et al. [12] address real-world complexities like facial expression variability and lighting inconsistencies alongside occlusion challenges. Despite being conducted during the COVID-19 pandemic, this survey overlooks masked face recognition and lacks comprehensive coverage of existing methods and empirical results. Similarly, Zeng Dan et al. [13] tackle the persistent challenge of identifying faces obscured by various occlusions, including medical masks. While categorizing modern and conventional techniques for recognizing faces under occlusion, this survey lacks empirical results and comparative analyses of existing approaches addressing occlusion challenges.

Conversely, recent surveys have undertaken a comprehensive examination of MFR and FMR, with a focus on addressing the challenges encountered by face recognition and detection systems following the COVID-19 pandemic. Notably, Alzu’bi Ahmad et al. [14] conducted an exhaustive survey on masked face recognition research, which has experienced significant growth and innovation in recent years. The study systematically explores a wide range of methodologies, techniques, and advancements in MFR, with a specific



emphasis on deep learning approaches. Through a meticulous analysis of recent works, the survey aims to provide valuable insights into the progression of MFR systems. Furthermore, it discusses common benchmark datasets and evaluation metrics in MFR research, offering a robust framework for evaluating different approaches and highlighting challenges and promising research directions in the field. Moreover, Wang Bingshu et al. [15] address the pressing need for AI techniques to detect masked faces amidst the COVID-19 pandemic. Their comprehensive analysis includes an examination of existing datasets and categorization of detection methods into conventional and neural network-based approaches. By summarizing recent benchmarking results and outlining future research directions, the survey aims to advance understanding and development in masked facial detection. Similarly, Nowrin Afsana et al. [16] address the critical requirement for face mask detection algorithms in light of the global impact of the COVID-19 pandemic. Their study evaluates the performance of various object detection algorithms, particularly deep learning models, to provide insights into the effectiveness of face mask detection systems. Through a comprehensive analysis of datasets and performance comparisons among algorithms, the survey sheds light on current challenges and future research directions in this domain.

### 3. Masked Face Datasets

In the realm of MFR and FMR, the presence and quality of datasets play a pivotal role in shaping robust and accurate models. Datasets serve as the bedrock for unraveling the complexities associated with identifying individuals wearing face masks, making substantial strides in the progression of MFR methodologies. Furthermore, they establish the foundational framework for tasks such as face mask recognition by creating paired masked–unmasked face datasets—integral components for training algorithms in face mask removal. This section extensively explores widely adopted standard benchmark datasets across various masked face tasks, encompassing MFR, FMR, and FU. To ensure a comprehensive overview, the section is bifurcated into two subsections based on the type of mask, distinguishing between real and synthetic masks. While real-world datasets offer heightened realism, they may be noisy and lack control. Conversely, cleaner synthetic datasets may not entirely capture real-world scenarios' intricacies.

#### 3.1. Real Mask Datasets

In the domain of MFR and FMR, datasets that incorporate genuine face masks provide crucial insight into the complexities posed by real-world scenarios. These datasets meticulously capture the intricacies of diverse face masks worn by individuals across various settings, ranging from public spaces and workplaces to social gatherings. The authenticity embedded in these masks significantly enhances the realism of the training process, allowing models to effectively adapt to the challenges presented by authentic face coverings. Notably, addressing a challenge highlighted in the discussion of limitations—the scarcity of real masked face datasets, especially those utilized in face mask removal tasks where pairing masked and unmasked faces is essential—this subsection looks into prominent benchmark datasets featuring real face masks. Through an exploration of these datasets, we illuminate their characteristics, applications, and significance in propelling advancements within the field of MFR methodologies. Table 1 offers an overview of the primary benchmark datasets utilized in both MFR and FMR, elucidating their essential attributes and utility across various applications. Furthermore, Figure 3a showcases sample images from the RMFRD dataset, recognized as the most extensive dataset utilized for MFR. For comparison, Figure 3b displays samples from another significant dataset in the MFR domain, namely MFR2 [7]. Meanwhile, Figure 4 illustrates sample images from datasets employed in FMR, offering visual insights into the various real mask datasets discussed.



(a) RMFRD

(b) MFR2

(c) MDMFR

**Figure 3.** Samples of masked and unmasked faces from the real-mask masked face datasets used in masked face recognition.



**Figure 4.** Samples from real masked face datasets used in face mask recognition.

In the realm of real face mask datasets, the Real-World Masked Face Recognition Dataset (RMFRD) [17] stands out as one of the most extensive publicly available resources for MFR. RMFRD comprises 90,000 unmasked faces and 5000 masked faces for 525 individuals, serving as a valuable asset for training and evaluating MFR models. Figure 3a showcases samples from RMFRD, illustrating examples of both masked and unmasked faces for various individuals, thereby facilitating MFR applications and serving as ground truth data for training face mask removal models. Another notable dataset is the Face Mask Recognition Dataset [18], a large private repository comprising 300,988 images of 75,247 individuals. Each individual in this dataset is represented by four selfie images: one without a mask, one with a properly masked face, and two with incorrectly masked faces. These diverse scenarios offer rich training data for face mask recognition models. Additionally, MFR2 [7] is a popular dataset in the field, containing 269 images for 53 identities,

typically used to evaluate model performance trained on either real masked datasets or large synthetic datasets. Figure 3b showcases samples from MFR2.

**Table 1.** Summary of widely used benchmark datasets featuring real masks.

Dataset	Size	Identities	Access	Aim	Year
MAFA [19]	30,811	3	Public	FMR	2017
MD (Kaggle) [20]	853	3	Public	FMR	2019
RMFRD [17]	5000/90,000	525	Public	MFR	2020
MFR2 [7]	269	53	Public	MFR	2020
MASR-REC [21]	11,615	1004	Private	MFR	2020
MFI [22]	4916	669	Private	MFR	2020
MFV [22]	400	200	Private	MFR	2020
MFDD [23]	24,771	2	Private	FMR	2020
FMD (AIZOOTech) [24]	7971	2	Public	FMR	2020
Moxa3K [25]	3000	2	Public	FMR	2020
FMLD [26]	41,934	3	Public	FMR	2021
Sunil's custom dataset [27]	7500	2	Public	FMR	2021
Jun's practical dataset [28]	4672	3	Private	FMR	2021
ISL-UFMD [29]	21,316	3	Public	FMR	2021
PWMFD [30]	9205	3	Public	FMR	2021
WMD [31]	7804	1	Public	FMR	2021
WMC [31]	38,145	2	Public	FMR	2021
COMASK20 [32]	2754	300	Public	MFR	2022
MDMFR (FMD) [33]	2896	226	Public	MFR	2022
MDMFR (MFR) [33]	6006	2	Public	FMR	2022
TFM [34]	107,598	2	Private	FMR	2022
BAFMD [35]	6264	2	Public	FMR	2022
FMDD [18]	300,988	75,247	Private	MFR	-

The Masked Face Segmentation and Recognition (MFSR) dataset [21] comprises two components. The first part, MFSR-SEG, contains 9742 images of masked faces sourced from the Internet, each annotated with manual segmentation labels delineating the masked regions. These annotations are particularly useful in FU tasks, serving as an initial step, as seen in stage 1 of GANMasker [2]. The second part, MFSR-REC, encompasses 11,615 images representing 1004 identities, with 704 identities sourced from real-world collections and the remaining images gathered from the Internet. Each identity is represented by at least one image featuring both masked and unmasked faces. F. Ding, P. Peng, et al. [22] introduced two datasets to assess MFR models. The first dataset, known as Masked Face Verification (MFV), includes 400 pairs representing 200 distinct identities. The second dataset, Masked Face Identification (MFI), comprises 4916 images, each corresponding to a unique identity, totaling 669 identities. The COMASK20 dataset [32] employed a distinct approach. Video recordings of individuals in various settings and poses were captured and subsequently segmented into frames every 0.5 s, each stored in individual folders. Careful manual curation was conducted to remove any obscured images, ensuring data quality. Ultimately, a collection of 2754 images representing 300 individuals was assembled.

Figure 4 presents a selection of datasets specifically designed for Facial Mask Recognition (FMR), each offering samples showcasing correctly masked, incorrectly masked, and



unmasked faces. For instance, the Masked Face Detection Dataset (MFDD) [23] encompasses 24,771 images of masked faces. This dataset is compiled using two approaches: the first involves integrating data from the AI-Zoo dataset [24], which itself is a popular FMR dataset comprising 7971 images sourced from the WIDER Face [36] and MAFA [19] datasets. The second approach involves gathering images from the internet. MFDD serves the purpose of determining whether an individual is wearing a mask. The MAFA dataset [19] stands out as a prominent solution to the scarcity of large datasets featuring masked faces. Comprising 30,811 internet-sourced images, each possessing a minimum side length of 80 pixels, MAFA offers a substantial resource for research and development in MFR. The dataset contains a total of 35,806 masked faces, with the authors ensuring the removal of images featuring solely unobstructed faces to maintain the focus on occluded facial features.

N. Ullah, A. Javed, et al. [33] introduced a comprehensive dataset named the Mask Detection and Masked Facial Recognition (MDMFR) dataset, aimed at evaluating the performance of both face mask recognition and masked facial recognition methods. The dataset comprises two distinct parts, each serving a specific purpose. The first part is dedicated to face mask recognition and includes 6006 images, with 3174 images featuring individuals wearing masks and 2832 images featuring individuals without masks. Conversely, the second part focuses on masked facial recognition and contains a total of 2896 masked images representing 226 individuals. Figure 3c provides visual samples of both categories, with samples of masked and unmasked faces displayed on the left and masked face samples of three different individuals shown on the right.

Researchers continue to leverage existing large datasets like MAFA and Wider Face to develop new datasets tailored to specific research objectives. Batagelj, Peer, et al. [26] introduced the Face-Mask Label Dataset (FMLD), a challenging dataset comprising 41,934 images categorized into three classes: correct mask (29,532 images), incorrect mask (1528 images), and without mask (32,012 images). Similarly, Singh, S., Ahuja, U., et al. [27] amalgamated data from MAFA and Wider Face, and additional manually curated images sourced from various online platforms to create a dataset encompassing 7500 images, referred to here as Sunil's custom dataset. Additionally, J. Zhang, F. Han, et al. [28] developed a practical dataset known here as Jun's practical dataset, comprising a total of 4672 images. This dataset includes 4188 images sourced from the public MAFA dataset and 484 images sourced from the internet. The images in Jun's practical dataset are categorized into five types: clean face, hand-masked face, non-hand-masked face, masked incorrect face, and masked correct face, with the first three types grouped into the without-mask class and the remaining two classes designated as mask\_correct and mask\_incorrect. Likewise, the Interactive Systems Labs Unconstrained Face Mask Dataset (ISL-UFMD) [29] is compiled from a variety of sources, including publicly available face datasets like Wider Face [36], FFHQ [37], CelebA [5], and LFW [6], in addition to YouTube videos and online sources. This comprehensive dataset comprises 21,316 face images designed for face mask recognition scenarios, encompassing 10,618 images of individuals wearing masks and 10,698 images of individuals without masks.

Moreover, several other datasets cater to face mask recognition tasks. For instance, the Masked dataset (MD-Kaggle) [20] hosted on Kaggle comprises 853 images annotated into three categories. Similarly, the Properly Wearing Masked Face Detection Dataset (PWMFD) [30] contains 9205 images categorized into three groups. Additionally, datasets like Moxa3K [25] and TFM [34] focus on binary classifications, distinguishing between images with and without masks. Moxa3K features 3000 images, while TFM boasts a larger private collection of 107,598 images. Furthermore, the Wearing Mask Detection (WMD) [31] dataset provides 7804 images for training detection models. Meanwhile, the Wearing Mask Classification (WMC) [31] dataset consists of two classes, with 19,590 images containing masked faces and 18,555 background samples, resulting in a total of 38,145 images. The Bias-Aware Face Mask Detection (BAFMD) dataset [35] comprises 6264 images, featuring 13,492 faces with masks and 3118 faces without masks. Noteworthy is that each image contains multiple faces.

### 3.2. Synthetic Mask Datasets

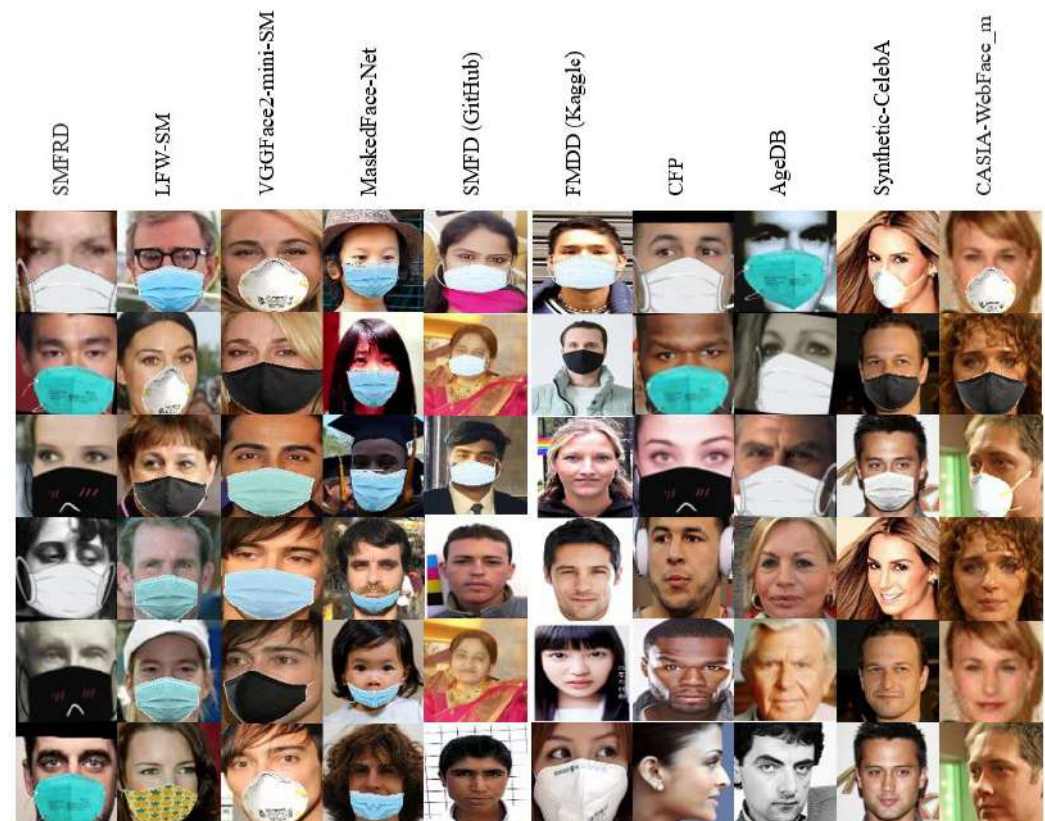
However, the accessibility of synthetic masked face datasets has seen a significant boost due to the abundance of public face datasets available for generation purposes. In contrast to datasets featuring authentic masks, synthetic mask datasets bring a unique perspective by incorporating artificially generated face coverings. These datasets offer a controlled environment, enabling researchers to explore various synthetic mask variations, including considerations such as style, color, and shape. The controlled nature of these datasets facilitates a systematic exploration of challenges in MFR, providing valuable insights into the response of models to different synthetic masks. Within this subsection, we look into benchmark datasets featuring synthetic masks, examining their creation processes, advantages, and potential applications in the field of masked face recognition and related tasks. Figure 5 showcases samples from some datasets with synthetic masks, while Table 2 provides a detailed summary of these synthetic datasets discussed in this subsection.

**Table 2.** Summary of popular synthetic mask benchmarking datasets.

Dataset	Size	Identities	Access	Aim	Year
LFW [6]	13,233	5749	Public	MFR	2008
CASIA-WebFace [4]	494,414	10,575	Public	MFR	2014
CelebA [5]	+200,000	10,000	Public	FMR and FU	2015
CFP [38]	7000	500	Public	MFR	2016
AgeDB [39]	16,488	568	Public	MFR	2017
IJB-C [40]	148,876	3531	Public	MFR	2018
CelebA-HQ [41]	30,000	-	Public	FMR and FU	2018
VGGFace2 [42]	3.31 M	9131	Public	MFR	2018
SMFRD [23]	536,721	16,817	Public	MFR	2020
LFW-SM [7]	64,973	5749	Public	MFR	2020
VGGFace2-mini-SM [7]	697,084	8631	Public	MFR	2020
SMFD(GitHub) [43]	1376	2	Public	FMR	2020
FMDD(Kaggle) [44]	7553	2	Public	FMR	2020
PS-CelebA [3]	10,000	-	Private	FU	2020
VGGFace2_m [45]	666,800	8335	Public	MFR	2021
LFW_m [45]	26,466	85,749	Public	MFR	2021
CF_m [45]	5000	500	Public	MFR	2021
MaskedFace-Net [46]	137,016	2	Public	FMR	2022
MS1MV2-Masked [47]	5.374 M	85,000	Public	MFR	2022
CASIA-WebFace_m [48]	789,296	10,575	Public	MFR	2022
Synthetic-CelebA [2]	30,000	-	Private	FU	2023

As highlighted in the preceding subsection, RMFRED stands out as one of the most expansive real-mask masked face datasets. Z. Wang, B. Huang, et al. [23] took a different approach by automatically applying masks to face images sourced from existing public datasets such as CASIA-WebFace [4], LFW [6], CFP-FP [38], and AgeDB-30 [39]. This effort resulted in creating a Simulated Mask Face Recognition Dataset (SMFRD), comprising 536,721 masked faces representing 16,817 unique identities. Aqeel Anwar and Arijit Raychowdhury [7] introduced MaskTheFace, an open-source tool designed to mask faces in public face datasets effectively. This tool was utilized to generate large masked face datasets like LFW-SM, which contains 64,973 images representing 5749 identities. It incorporates various types of masks, including surgical-green, surgical-blue, N95, and cloth,

derived from the original LFW dataset [6]. Utilizing the same tool, they developed another masked dataset called VGGFace2-mini-SM [7], extracted from the original VGGFace2-mini dataset, a subset of the VGGFace2 dataset [42], with 42 images randomly selected per identity. This augmentation expanded the total image count to 697,084, maintaining the same 8631 identities. Various mask types, akin to those employed in LFW-SM, were integrated into this dataset.



**Figure 5.** Samples of synthetic masked faces from benchmark datasets.

F. Boutros, N. Damer, et al. [47] developed a synthetic dataset named MS1MV2-Masked, derived from MS1MV2 [49], incorporating various mask shapes and colors. They utilized Dlib [50] for extracting facial landmark points. However, they noted that Dlib encountered difficulties in extracting landmarks from 426 images. Consequently, their synthetic dataset comprises approximately 5374 images representing 85,000 identities.

Building upon the FFHQ dataset [37], C. Adnane, H. Karim, et al. [46] introduced a comprehensive synthetic dataset called MaskedFace-Net. This dataset comprises two primary sub-datasets: the Correctly Masked Face Dataset (CMFD) and the Incorrectly Masked Face Dataset (IMFD). These datasets serve as the sole categories within the main dataset dedicated to face mask recognition, collectively containing a total of 137,016 images. Notably, the dataset comprises 49% correctly masked faces (67,193 images) and 51% incorrectly masked faces (69,823 images). The SMFD dataset [43] comprises two distinct classes: one with a mask, encompassing 690 images, and the other without a mask, comprising 686 images. Consequently, the dataset contains a total of 1376 images, which were utilized to train a classification CNN network tasked with distinguishing between faces with masks and those without. Moreover, the Face Mask Detection Dataset [44] comprises 7553 images divided into two categories: with masks and without masks. Specifically, there are 3725 images depicting faces with masks and 3828 images featuring faces without masks. These images were sourced from the internet and encompass the entirety of images found in the SMFD dataset [43]. Thus, the dataset represents a hybrid composition, incorporating both real masked faces and synthetic masked faces. They employed this dataset to train their

model and also utilized a similar approach to create additional synthetic datasets based on Labeled Faces in the Wild (LFW) [6] and IARPA Janus Benchmark -C (IJB-C) [40].

It is common practice among authors to create synthetic datasets either manually or automatically using deep learning tools. Therefore, various face datasets can be considered or referenced in synthetic masked face datasets. One such dataset is the Celebrities in Frontal-Profile (CFP) dataset [38], comprising 7000 images representing 500 identities. This dataset is divided into two sub-datasets based on the face angle: frontal faces and profile faces, both depicting the same 500 identities. The frontal faces subset consists of 10 images per identity, while each identity in the profile faces subset is represented by 4 images. The AgeDB [39] dataset comprises 16,488 images featuring 568 prominent individuals, including actors/actresses, politicians, scientists, writers, and more. Each image comes annotated with identity, age, and gender attributes. Moreover, CelebA [5] stands out as one of the largest face datasets, boasting over 200,000 images spanning 10,000 identities. Complementing this dataset is CelebA-HQ [41], a high-resolution variant derived from CelebA, featuring 30,000 meticulously crafted high-quality images. Furthermore, leveraging the CelebA dataset, M. Mohamed and K. Hyun-Soo [2] curated a Synthetic Masked Dataset comprising 30,000 images distributed across three subfolders: original unmasked faces, masked faces, and binary mask maps. Employing the MaskTheFace tool [7], they generated two masked datasets with varying sizes (256 and 512 pixels) tailored for face mask removal tasks. Additionally, N. Ud Din, K. Javed, et al. [3] engineered a synthetic dataset using Adobe Photoshop CC 2018, derived from CelebA, featuring 10,000 masked images alongside their original counterparts.

Deng H, Feng Z, et al. [45] introduced three composite datasets featuring mixed masked and unmasked faces. These datasets integrate original images with masked faces generated through their proprietary masked face image generation algorithm. The first dataset, VGGFace2\_m [45], originates from the VGGFace2 [42] face dataset, leveraging 8335 identities from the VGG-Face2 training set. From each identity, 40 pictures were randomly selected to construct VGGFace2\_mini, which was then combined with the generated masked faces to form VGGFace2\_m. Similarly, LFW\_m [45] was crafted using the LFW dataset [6], a widely used benchmark for facial recognition, comprising 13,233 face images and 5749 identities. The masked face images generated were merged with the original LFW dataset to produce LFW\_m. Lastly, CF\_m [45] was derived from the CASIA-FaceV5 dataset [51], which features images of 500 individuals, with five images per person totaling 2500 images. The original images from CASIA-FaceV5 were amalgamated with masked images generated by their algorithm to create CF\_m. Furthermore, Pann V and Lee HJ [48] devised CASIA-WebFace\_m as an extension of the CASIA-WebFace dataset [4], a comprehensive public face recognition dataset encompassing 494,414 images representing 10,575 distinct identities. However, due to limitations with the data augmentation tool, their generated masked faces amounted to 394,648, with 20% of face images remaining undetected. These generated masked face images were then integrated with the corresponding unmasked images from the original dataset, resulting in CASIA-WebFace\_m for model training purposes. Consequently, the combined dataset boasts a total of 789,296 training samples. Moreover, they produced modified versions of the original datasets LFW [6], AgeDB [39], and CFP [38], labeled as LFW\_m, AgeDB-30\_m, and CFP-FP\_m, respectively.

#### 4. Evaluation Metrics

This section will explain standard evaluation metrics, focusing specifically on those applied in MFR, FMR, and face unmasking (face mask removal). Evaluating model performance in these domains is essential for gauging their effectiveness in real-world applications. To this end, various evaluation metrics and benchmarking strategies are utilized to assess accuracy, robustness, and efficiency. In the following discussion, we explore the primary evaluation metrics and benchmarking approaches employed in these tasks.

1. **Accuracy** is a fundamental evaluation metric utilized across various domains, including facial recognition tasks. It represents the proportion of correct predictions



relative to the total number of samples and can be formally defined as illustrated in Equation (1).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

2. **ERR (Error Rate)** is a crucial metric utilized in diverse classification tasks, offering valuable insights into model accuracy by measuring misclassifications relative to dataset size. Unlike accuracy, ERR accounts for both false positives and false negatives, providing a comprehensive assessment of model performance. Its sensitivity to imbalanced data underscores its importance, making it an essential tool for evaluating classification accuracy. Mathematically, ERR is calculated by dividing the sum of false positive and false negative predictions by the total number of instances, as shown in Equation (2).

$$ERR = \frac{FP + FN}{TP + TN + FP + FN} = 1 - Accuracy \quad (2)$$

3. **Precision** quantifies the proportion of accurate positive identifications among all the positive matches detected, and it can be formally expressed as depicted in Equation (3).

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

4. **Recall** also known as **sensitivity** or **true positive rate** measures the proportion of true positive instances correctly identified by the system out of all actual positive instances. It is formally defined as shown in Equation (4).

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

5. **F1-Score** is a pivotal evaluation metric, that represents the harmonic mean of precision and recall. This metric offers a balanced measure of the facial recognition model's performance, accounting for both false positives and false negatives. Particularly valuable for imbalanced datasets, the F1-Score provides a comprehensive assessment of the model performance. Unlike accuracy, which may overlook certain types of errors, the F1-Score considers both false positives and false negatives, rendering it a more reliable indicator of a model's effectiveness. Its calculation is demonstrated in Equation (5).

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (5)$$

6. **ROC (Receiver Operating Characteristic) Curve** ROC curves [52] graphically represent the trade-off between sensitivity (true positive rate) and specificity (true negative rate) across various threshold values. This visualization aids in selecting an optimal threshold that strikes a balance between true positive and false positive recognition rates. By examining the ROC curve, decision makers can effectively assess the performance of a classification model and make informed decisions about threshold selection.
7. **AUC (Area Under the Curve)** is a pivotal evaluation metric in classification tasks, offering a comprehensive assessment of a model's performance. It quantifies the discriminative power of the model across varying threshold values, providing insights into its ability to correctly classify positive and negative instances. A higher AUC value signifies stronger discrimination, indicating a superior model performance. Conversely, an AUC value of 0.5 suggests that the model's predictive ability is no better than random chance. AUC is instrumental in gauging the effectiveness of classification models and is widely utilized in performance evaluation across diverse domains.
8. **Confusion Matrix** provides a detailed breakdown of the model's predictions, including true positives, true negatives, false positives, and false negatives. It serves as a basis for computing various evaluation metrics and identifying areas for improvement.

9. **FAR (False Acceptance Rate)** serves as a focused gauge of security vulnerabilities, offering precise insights into the system's efficacy in thwarting unauthorized access attempts. This pivotal metric plays a crucial role in evaluating the overall security effectiveness of biometric authentication systems, thereby guiding strategic endeavors aimed at bolstering the system reliability and mitigating security threats. Equation (6) delineates its formula, providing a quantifiable framework for assessing system performance.

$$FAR = \frac{FP}{FP + TN} \quad (6)$$

10. **FRR (False Rejection Rate)** is a crucial metric for evaluating system usability, representing the likelihood of the system inaccurately rejecting a legitimate identity match. Its assessment is integral to gauging the user-friendliness of the system, with a high FRR indicating diminished usability due to frequent denial of access to authorized individuals. Conversely, achieving a lower FRR is essential for improving user satisfaction and optimizing access procedures. The calculation of FRR is depicted in Equation (7).

$$FRR = \frac{FN}{FN + TP} \quad (7)$$

11. **EER (Equal Error Rate)** denotes the threshold on the ROC curve where the false acceptance rate (FAR) equals the false rejection rate (FRR), signifying the equilibrium point between false acceptance and false rejection rates. A lower EER signifies a superior performance in achieving a balance between these two error rates.
12. **Specificity** also known as the **true negative rate** gauges the system's proficiency in accurately recognizing negative instances. Specifically, it assesses the system's capability to correctly identify individuals who are not the intended subjects. Mathematically, specificity is calculated using Equation (8). This metric offers valuable insights into the system's performance in correctly classifying negatives, contributing to its overall effectiveness and reliability.

$$Specificity = \frac{TN}{TN + FP} = 1 - FAR \quad (8)$$

13. **Rank-N Accuracy** is a widely employed metric in facial recognition tasks, that assesses the system's capability to prioritize the correct match within the top N retrieved results. It quantifies the percentage of queries for which the correct match is positioned within the top-N-ranked candidates. In the Rank-N Identification Rate evaluation, the system's output is considered accurate if the true identity of the input is within the top N identities listed by the system. For instance, in a Rank-1 assessment, the system is deemed correct if the true identity occupies the top spot. Conversely, in a Rank-5 evaluation, the system is considered accurate if the true identity is among the top 5 matches. A higher Rank-N Accuracy signifies superior performance in identifying the correct match among the retrieved candidates, providing valuable insights into the system's efficacy in real-world scenarios. Mathematically, it is represented as depicted in Equation (9).

$$Rank - N Accuracy = \frac{\# \text{ correct matches within top } - N}{\text{Total number of queries}} \quad (9)$$

14. **Intersection over Union (IoU)** quantifies the extent of spatial overlap between the predicted bounding box (P) and the ground truth bounding box (G). Its mathematical representation is shown in Equation (10).

$$IoU = \frac{P \cap G}{P \cup G} \quad (10)$$

15. **AP (Average Precision)** serves as a crucial measure in assessing object detection systems. It provides insight into how effectively these systems perform across different confidence thresholds, by evaluating their precision–recall performance. AP computes the average precision across all recall values, indicating the model’s ability to accurately detect objects at varying confidence levels. This calculation involves integrating the precision–recall curve, as demonstrated in Equation (11). By considering the precision–recall trade-off comprehensively, AP offers a holistic evaluation of the detection method’s effectiveness.

$$AP = \int_0^1 P(r) dr \quad (11)$$

where  $P(r)$  represents the precision at a given recall threshold  $r$ , where  $r$  ranges from 0 to 1.

16. **mAP (Mean Average Precision)** enhances the notion of AP by aggregating the average precision values across multiple object classes. It offers a unified metric summarizing the overall performance of the object detection model across diverse object categories. Mathematically, mAP is calculated as the average of AP values for all classes, as illustrated in Equation (12).

$$mAP = \frac{1}{N} \sum_{i=1}^N AP_i \quad (12)$$

17. **Dice Score** also known as **Dice Coefficient** is a metric commonly used in image segmentation tasks to assess the similarity between two binary masks or segmentation maps. It quantifies the spatial overlap between the ground truth mask ( $A$ ) and the predicted mask ( $B$ ), providing a measure of segmentation accuracy. The Dice Score equation compares the intersection of  $A$  and  $B$  with their respective areas, as defined in Equation (13).

$$Dice\ Score = \frac{2 \times (|A \cap B|)}{|A| + |B|} \quad (13)$$

where  $|A \cap B|$  represents the number of overlapping pixels between the ground truth and predicted masks, while  $|A|$  and  $|B|$  denote the total number of pixels in each mask, respectively.

18. **PSNR (Peak Signal-to-Noise Ratio)** is widely employed in image inpainting to assess the quality of image generation or reconstruction. It quantifies the level of noise or distortion by comparing the maximum possible pixel value to the mean squared error (MSE) between the original and reconstructed images, as depicted in Equation (13).

$$PSNR = 10 \log_{10} \left( \frac{MAX^2}{MSE} \right) \quad (14)$$

19. **SSIM (Structural Similarity Index Measure)** The Structural Similarity Index Measure [53] evaluates the similarity between two images by considering their luminance, contrast, and structure. It provides a measure of perceptual similarity, accounting for both global and local image features. SSIM is calculated by comparing the luminance, contrast, and structure similarity indexes, as expressed in Equation (15).

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (15)$$

where  $x$  and  $y$  represent the two compared images,  $\mu_x$  and  $\mu_y$  denote the mean of  $x$  and  $y$ , respectively,  $\sigma_x^2$  and  $\sigma_y^2$  represent the variances of  $x$  and  $y$ , respectively,  $\sigma_{xy}$  is the covariance of  $x$  and  $y$ , and  $C_1$  and  $C_2$  are constants to stabilize the division, typically set to small positive values.

20. **FID (Fréchet Inception Distance)** [54] serves as a metric for assessing the likeness between two sets of images. It quantifies the disparity between feature representations of real and generated images within a learned high-dimensional space, typically modeled by a pre-trained neural network. A lower FID score denotes a higher degree of resemblance between the datasets. The calculation of FID involves the application of the Fréchet distance formula, as depicted in Equation (16).

$$FID = ||\mu_R - \mu_F||^2 + \text{Tr}(C_R + C_F - 2(C_R - C_F)^{\frac{1}{2}}) \quad (16)$$

where  $\mu_R$  and  $\mu_F$  are the mean feature vectors of the real and generated image sets,  $C_R$  and  $C_F$  are their covariance matrices, and  $\text{Tr}$  denotes the trace operator.

21. **NIQE (Naturalness Image Quality Evaluator)** [55] assesses the quality of an image based on natural scene statistics. It evaluates the level of distortions introduced during image acquisition or processing, providing a measure of image fidelity. NIQE computes the deviation of the image from the expected natural scene statistics, with higher scores indicating greater image distortion.
22. **BRISQUE (Blind/Referenceless Image Spatial Quality Evaluator)** [56] is a no-reference image quality assessment metric. It evaluates the perceived quality of an image by analyzing its spatial domain features, such as local sharpness and contrast. BRISQUE computes a quality score based on the statistical properties of these features, with lower scores indicating higher image quality.

## 5. Masked Face Methods

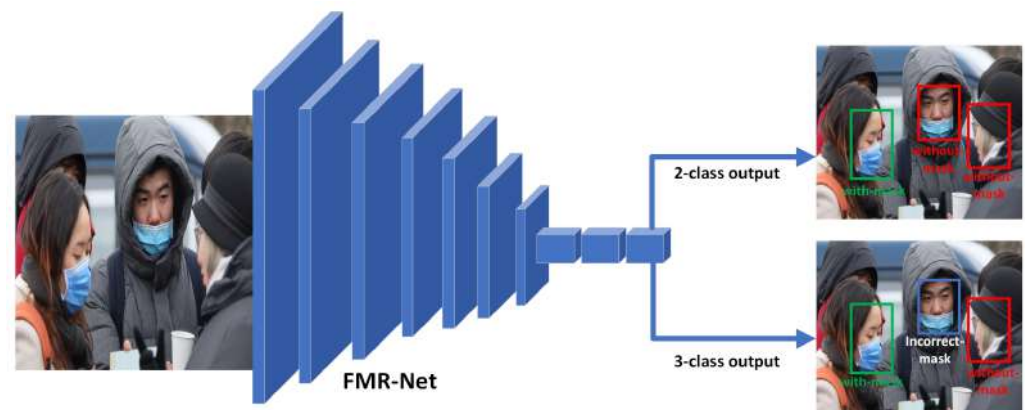
With the increased use of face masks in response to the COVID-19 outbreak, researchers have focused on the challenges given by masked faces, particularly in the domains of MFR, FMR, and FU. This section thoroughly analyzes the most recent breakthroughs in deep learning-based state-of-the-art (SOTA) approaches to overcoming these challenges. With three distinct subsections dedicated to each task, or preliminary steps toward them, ranging from face mask recognition and removal to masked face recognition approaches, researchers have introduced a diverse array of innovative solutions to improve the accuracy and reliability of masked face recognition systems. By thoroughly exploring these designs and approaches, this section aims to provide important insights into current advancements in deep learning-based approaches for masked face-related tasks and elucidate potential avenues for future research in these rapidly evolving fields.

### 5.1. Face Mask Recognition Approaches

In the realm of computer vision, face mask recognition has become crucial, especially during health crises like the COVID-19 pandemic. This technology relies on machine learning (ML) and deep learning (DL) techniques to automatically detect masks on human faces. DL, particularly convolutional neural networks (CNNs), excels by directly extracting features from raw input data, eliminating the need for manual feature engineering. Various backbone architectures, including multi-stage, YOLO-based, and transfer learning, hierarchically process data to distinguish masked from unmasked faces. The choice of backbone architecture significantly impacts the accuracy and computational efficiency [57,58].

Tables 3 and 4 summarize various face mask recognition models evaluated using different datasets and performance metrics. Furthermore, Figure 6 showcases the FMR-Net architecture, depicting examples of both the two-subtask two-class scenario, distinguishing between with-mask and without-mask, and the three-class scenario, discerning between with-mask, incorrect-mask, and without-mask.





**Figure 6.** Illustration of the FMR-Net architecture for face mask recognition, depicting two-subtask scenarios: 2-class (with and without mask) and 3-class (with, incorrect, and without mask).

#### 5.1.1. Convolutional Neural Network

Convolutional neural networks are crucial in computer vision due to their efficient pattern recognition and spatial feature extraction capabilities. By applying convolutional filters directly to input images, CNNs efficiently isolate high-level features, enhancing both the accuracy and computational speed for tasks like image classification and object detection. FMJM Shamrat et al. [59] exploring three deep learning techniques for face mask recognition: Max pooling, Average pooling, and MobileNetV2. MobileNetV2 achieved the highest accuracies—99.72% in training and 99.82% in validation—demonstrating a robust capability, while H Goyal et al. [60] developed an automated face mask recognition model to enforce mask wearing in public spaces. This model, capable of processing both static images and real-time video feeds, classifies subjects as “with mask” or “without mask.” Trained and evaluated using a dataset of approximately 4000 images from Kaggle, the model achieved an accuracy rate of 98%. It demonstrated computational efficiency and precision.

#### 5.1.2. Multi-Stage Detection

Multi-stage detection is a category of object detection algorithms where the detection process is divided into several sequential steps. In a typical multi-stage detector, such as RCNN, the first step involves identifying a set of potential regions of interest within an image, often through a technique like selective search. Subsequently, each region is individually processed to extract CNN feature vectors, which are then used to classify the presence and type of objects within those regions. This approach contrasts with single-stage detectors that perform detection in a single pass without a separate region proposal phase, trading some accuracy for increased processing speed.

S Sethi et al. [61] introduce a novel real-time face mask recognition technique. Combining one-stage and two-stage detectors, it accurately identifies individuals not wearing masks in public settings, supporting mask mandate enforcement. Utilizing ResNet50 as a baseline, transfer learning enhances feature integration across levels, and a new bounding box transformation improves localization accuracy. Experiments with ResNet50, AlexNet, and MobileNet optimize the model’s performance, achieving 98.2% accuracy.

Also, A Chavda et al. [62] develop a two-stage architecture combines RetinaFace face detection with a CNN classifier. Faces detected by RetinaFace are processed to determine mask presence. The classifier, trained using MobileNetV2, DenseNet121, and NASNet, ensures efficient real-time performance in CCTV systems. M Umer et al. [63] develop a new dataset called RILFD, consisting of real images annotated with labels indicating mask usage. Unlike simulated datasets, RILFD provides a more accurate representation for training face mask recognition systems. The researchers evaluate machine learning models, including YOLOv3 and Faster R-CNN, adapting them specifically for detecting mask-wearing individuals in surveillance footage. Enhancing these models with a custom

CNN and a four-step image processing technique, they achieved an impressive 97.5% accuracy on the RILFD dataset, as well as on two other publicly available datasets (MAFA and MOXA).

**Table 3.** Summary of different face mask recognition models.

Model	Year	Dataset	Accuracy	Precision	Recall	F1-Score
Fine-Tuning of InceptionV3 [64]	2020	SMFD	100%	100%	-	-
MobileNetV2 + SVM [65]	2020	Private Dataset	97.11%	95.08%	94.84%	-
Resnet50 + SVM + ensemble algorithm [66]	2021	RMFRD	99.64%	-	-	-
		SMFD	99.49%	-	-	-
		LFW	100%	-	-	-
Faster_RCNN + InceptionV2 + BLS [31]	2021	WMD Simple Scene	-	96.46%	98.20%	97.32%
		WMD Complex Scene	-	94.22%	88.24%	91.13%
Max pooling [59] Average pooling MobileNetV2	2021	RMFRD + SMFD + Own Dataset	98.67% 96.23% 99.82%	- - -	- - -	- - -
CNN [60]	2021	Kaggle Dataset	98%	98%	97%	98%
ResNet50 + bounding box transformation [61]	2021	MAFA dataset Face Detection	-	99.2%	99%	-
		MAFA dataset Mask Detection	-	98.92%	98.24%	-
RetinaFace + CNN(NASNetMobile)) [62]	2021	RMFRD + Larxel (Kaggle)	99.23%	98.28%	100%	99.13%
RetinaFace + CNN(Dense Net121))			99.49%	99.70%	99.12%	99.40%
SSDMNV2 [67]	2021	Self-made Dataset of Masked Faces	92.64%	-	-	93%
Fusion Transfer Learning [68]	2022	RMFRD and MAFA	97.84%	-	97.87%	98.13%
Customized CNN + Image Preprocessing Techniques [63]	2023	RILFD	97.25%	96.20%	97.34%	96.77%
		MAFA	95.74	-	94.29%	-
		MOXA	94.37%	-	95.28%	-
		RMFRD	99.63%	-	99.69%	-
SSD, ResNet-50, and DeepSiamese Neural Network [69]	2023	RMFRD + Larxel	98.24%	-	-	-
MobileNetV2 and Caffe-based SSD [70]	2023	Efficient Face Mask Dataset	97.81%	-	-	98%
CMNV2 [71]	2023	The Prajna Bhandary	99.64%	100%	99.28%	99.64%

**Table 4.** Summary of YOLO-based face mask recognition models.

Model	Year	Dataset	Accuracy	Precision	Recall	F1-Score	AP	mAP
YOLOv3 [27] Faster R-CNN	2021	MAFA and Wider Face	-	-	-	-	55% 62%	-
SE-YOLOv3 [30]	2021	PWMFD	-	-	-	-	73.7% AP <sub>50</sub> 99.5% AP <sub>75</sub> 88.7%	-
Improved YOLO-v4 (CSPDarkNet53) [72]	2021	RMFRD and Masked Face-Net	-	93.6%	97.9%	95.7%	84.7%	-
YOLOV5 [73]	2021	Kaggle and MakeM	96.5%	-	-	-	-	-
Efficient-YOLOv3 [68]	2022	Face Mask Dataset	-	-	-	-	98.18%	96.03%
FMDYolo [74]	2022	Kaggle (Face Mask Detection Dataset)	-	-	-	-	-	66.4%
		VOC Mask	-	-	-	-	-	57.5%
YOLOv5s-CA [75]	2023	Kaggle + Created Dataset from YouTube	-	95.9%	92.3%	94%	-	mAP@0.5 96.8%
AI-Yolo [76]	2023	Kaggle (WMD-1)	-	-	-	89.3%	-	94.1%
		Kaggle (WMD-2)	-	-	-	78.6%	-	90.7%
YOLOv8 [77]	2023	Face Mask Detection (FMD)	-	95%	95%	-	-	mAP@0.5 96%

### 5.1.3. Single Shot Detector

Single Shot Detector is an object detection technique that streamlines the process by using a single deep neural network. Unlike methods that rely on a separate region proposal network (which can slow down processing), SSD directly predicts object bounding boxes and class labels in a single pass. This efficiency allows SSD to process images in real

time with high speed and accuracy. S Vignesh Baalaji et al. [69] proposes an autonomous system for real-time face mask recognition during the COVID-19 pandemic. Leveraging a pre-trained ResNet-50 model, the system fine-tunes a new classification layer to distinguish masked from non-masked individuals. Using adaptive optimization techniques, data augmentation, and dropout regularization, the system achieves high accuracy. It employs a Caffe face detector based on SSD to identify face regions in video frames. Faces without masks undergo further analysis using a deep Siamese neural network (based on VGG-Face) for identity retrieval. The classifier and identity model achieve impressive accuracies of 99.74% and 98.24%, respectively.

B Sheikh et al. [70] presents the Rapid Real-Time Face Mask Detection System (RRFMDS) which is an automated method designed to monitor face mask compliance using video surveillance. It utilizes a Single-Shot Multi-Box Detector for face detection and a fine-tuned MobileNetV2 for classifying faces as masked, unmasked, or incorrectly masked. Seamlessly integrating with existing CCTV infrastructure, the RRFMDS is efficient and resource-light, ideal for real-time applications. Trained on a custom dataset of 14,535 images, it achieves high accuracy (99.15% on training and 97.81% on testing) while processing frames in just over 0.14 s. P Nagrath et al. [67] developed a resource-efficient face mask recognition model using a combination of deep learning technologies including TensorFlow, Keras, and OpenCV. Their model, SSDMNV2, employs a Single Shot Multibox Detector (SSD) with a ResNet-10 backbone for real-time face detection and uses the lightweight MobileNetV2 architecture for classifying whether individuals are wearing masks. They curated a balanced dataset from various sources, enhanced it through preprocessing and data augmentation techniques, and achieved high accuracy and F1-Scores.

#### 5.1.4. Transfer Learning

Transfer learning is a technique in deep learning where a model trained on one task is repurposed as the starting point for a model on a different but related task. This approach leverages pre-trained networks, such as InceptionV3, to improve learning efficiency and model performance, particularly when data is limited.

G Jignesh Chowdary et al. [64] propose an automated method for detecting individuals not wearing masks in public and crowded areas during the COVID-19 health crisis. They employ transfer learning with the pre-trained InceptionV3 model, fine-tuning it specifically for this task. Training is conducted on the Simulated Masked Face Dataset (SMFD), augmented with techniques like shearing, contrasting, flipping, and blurring, while A Oumina et al. [65] introduce a novel method for detecting whether individuals are wearing face masks using artificial intelligence technologies. They utilize deep convolutional neural networks (CNNs) to extract features from facial images, which are then classified using machine learning algorithms such as Support Vector Machine (SVM) and K-Nearest Neighbors (K-NN). Despite the limited dataset of 1376 images, the combination of SVM with the MobileNetV2 model achieved a high classification accuracy of 97.1%.

M Loey et al. [66] propose a hybrid model combining deep learning and classical machine learning techniques to detect face masks, a vital task during the COVID-19 pandemic. The model employs ResNet50 for extracting features from images in the first component, and uses decision trees, Support Vector Machine (SVM), and an ensemble algorithm for classification in the second component. The model was tested using three datasets: Real-World Masked Face Recognition Dataset (RMFRD), Simulated Masked Face Dataset (SMFD), and Labeled Faces in the Wild (LFW). It achieved high testing accuracies, notably 99.64% on RMFRD, 99.49% on SMFD, and 100% on LFW. Also, B Wang et al. [31] outline a two-stage hybrid machine learning approach for detecting mask wearing in public spaces to reduce the spread of COVID-19. The first stage uses a pre-trained Faster R-CNN model combined with an InceptionV2 architecture to identify potential mask-wearing regions. The second stage employs a Broad Learning System (BLS) to verify these detections by differentiating actual mask wearing from background elements. The method, tested on a new dataset comprising 7804 images and 26,403 mask instances, demonstrates high accuracy, achieving

97.32% in simple scenes and 91.13% in complex scenes. And X Su et al. [68] integrate transfer learning and deep learning techniques to enhance accuracy and performance. Firstly, the face mask detection component employs Efficient-YOLOv3 with EfficientNet as the backbone, using CIoU loss to improve the detection precision and reduce the computational load. Secondly, the classification component differentiates between “qualified” masks (e.g., N95, disposable medical) and “unqualified” masks (e.g., cotton, sponge masks) using MobileNet to overcome the challenges associated with small datasets and overfitting.

BA Kumar [71] develop a face detection system capable of accurately identifying individuals whether they are wearing masks or not. This enhancement addresses the increased use of face masks in public due to the COVID-19 pandemic. The system leverages a modified Caffe-MobileNetV2 (CMNV2) architecture, where additional layers are integrated for the better classification of masked and unmasked faces using fewer training parameters. The focus is on detecting facial features visible above the mask, such as the eyes, ears, nose, and forehead. The model demonstrated high accuracy, achieving 99.64% on static photo images and similarly robust performance on real-time video.

#### 5.1.5. YOLO (You Only Look Once)

YOLO is a real-time object detection system that recognizes objects with a single forward pass through the neural network. This one-stage detector efficiently combines the tasks of object localization and identification, making it ideal for applications requiring rapid and accurate object detection, such as face mask recognition. YOLO balances speed and precision, adapting to various scenarios where quick detection is crucial. S Singh et al. [27] focus on face mask recognition using two advanced deep learning models, YOLOv3 and Faster R-CNN, to monitor mask usage in public places during the COVID-19 pandemic. They develop a dataset of about 7500 images categorized into masked and unmasked faces, which they manually label and enhance with bounding box annotations. This dataset includes various sources and is accessible online. Both models are implemented using Keras on TensorFlow and trained with transfer learning. The models detect faces in each frame and classify them as masked or unmasked, drawing colored bounding boxes (red or green) around the faces accordingly. Also, X Jiang et al. [30] introduce SE-YOLOv3, an enhanced version of the YOLOv3 object detection algorithm, optimized for real-time mask detection by integrating Squeeze and Excitation (SE) blocks into its architecture. This modification helps focus the network on important features by recalibrating channel-wise feature responses, significantly improving detection accuracy. SE-YOLOv3 also employs GIoULoss for precise bounding box regression and Focal Loss to handle class imbalance effectively. Additionally, the model uses advanced data augmentation techniques, including mixup, to enhance its generalization capabilities.

J Yu and W Zhang [72] enhance the YOLO-v4 model for efficient and robust face mask recognition in complex environments, introducing an optimized CSPDarkNet53 backbone to minimize computational costs while enhancing model learning capabilities. Additionally, the adaptive image scaling and refined PANet structure augment semantic information processing. The proposed model is validated with a custom face mask dataset, achieving a mask recognition mAP of 98.3%. J Ieamsaard et al. [73] investigate an effective face mask recognition method using the YOLOv5 deep learning model during the COVID-19 pandemic. By leveraging a dataset of 853 images categorized into “With\_Mask”, “Without\_Mask”, and “Incorrect\_Mask”, the model is trained across different epochs (20, 50, 100, 300, and 500) to identify optimal performance. The results indicate that training the model for 300 epochs yields the highest accuracy at 96.5%. This approach utilizes YOLOv5’s capabilities for real-time processing. Also, TN Pham et al. [75] develop two versions: YOLOv5s-CA, with the CA module before the SPPF layer, and YOLOv5s-C3CA, where CA replaces the C3 layers. Tested on a new dataset created from YouTube videos, YOLOv5s-CA achieves a mAP@0.5 of 96.8%, outperforming baseline models and showing promising results for real-time applications in monitoring mask usage during the COVID-19 pandemic. The study also includes an auto-labeling system to streamline the creation of training datasets.

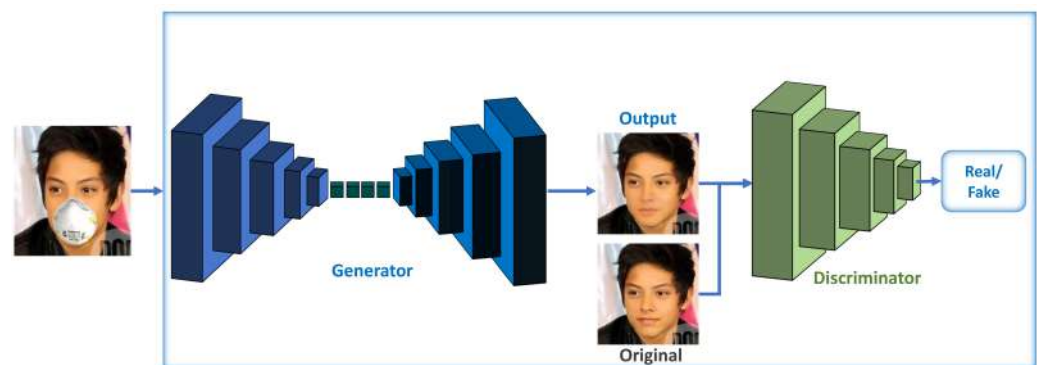


P Wu et al. [74] propose the FMDYolo framework which effectively detects whether individuals in public areas are wearing masks correctly, essential for preventing COVID-19 spread. It features the Im-Res2Net-101 as a backbone for deep feature extraction, combined with the En-PAN for robust feature fusion, improving the model generalization and accuracy. The localization loss and Matrix NMS in the training and inference stages enhance the detection efficiency. And H Zhang et al. [76] propose an enhanced object detection model named AI-Yolo, specifically designed for accurate face mask recognition in complex real-world scenarios. The model integrates a novel attention mechanism through Selective Kernel (SK) modules, enhances feature representation using Spatial Pyramid Pooling (SPP), and promotes effective feature fusion across multiple scales with a Feature Fusion (FF) module. Additionally, it employs the Complete Intersection over Union (CIoU) loss function for improved localization accuracy. Also, S Tamang et al. [77] evaluate the YOLOv8 deep learning model for detecting and classifying face mask-wearing conditions using the Face Mask Detector dataset. By employing transfer learning techniques, YOLOv8 demonstrates high accuracy in distinguishing between correctly worn mask, incorrectly worn mask, and no mask scenarios, outperforming the previous model, YOLOv5. The research highlights YOLOv8's enhancements in real-time object detection, making it suitable for applications requiring quick and reliable mask detection.

### 5.2. Face Unmasking Approaches

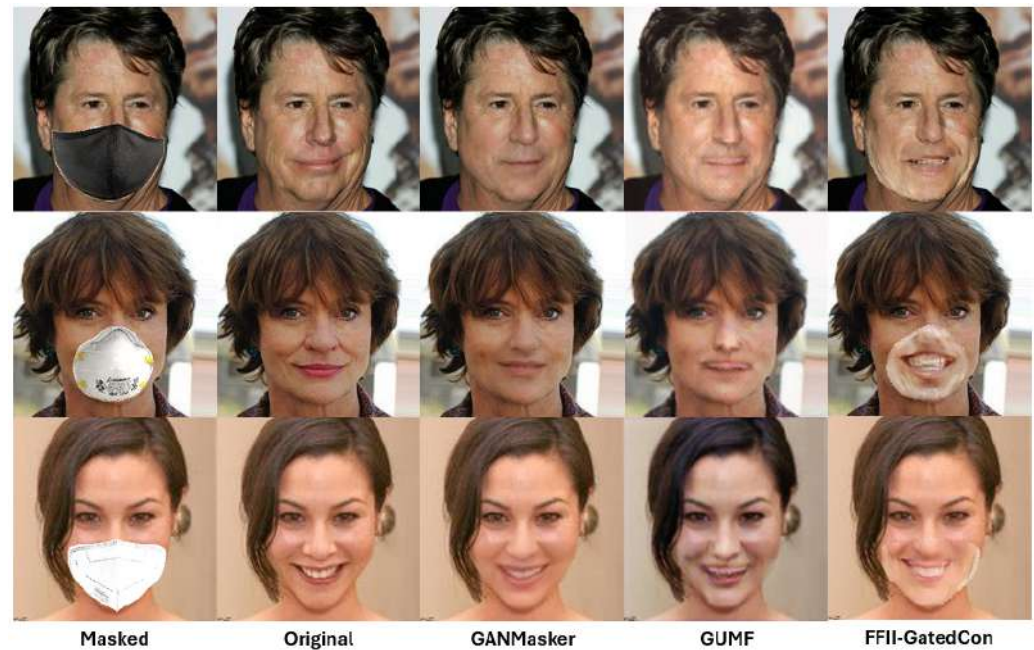
In this section, we explore recent progress in deep learning models designed for removing face masks, treating them as a specialized form of image inpainting, specifically focusing on “object removal.” This technique offers promising opportunities not only for mask removal and restoring unmasked faces but also for applications like verification and identification systems. Figure 7 provides an overview of the GAN network as a representative example of the FU-Net. Additionally, Tables 5 and 6 outline popular models utilizing the GAN and diffusion methods, respectively.

To further illustrate the capabilities of state-of-the-art face unmasking models, Figure 8 showcases a comparative example of three leading approaches: GANMasker, GUMF, and FFII-GatedCon. The first column presents the input masked face, while the second column displays the original unmasked face for reference. Each column demonstrates the output of a specific model, highlighting how these methods reconstruct essential facial features.



**Figure 7.** Overview of the GAN network as an example of FU-Net for face mask removal.

Most recent models in the field of object removal predominantly leverage GAN networks, a trend observed even before the emergence of COVID-19, as evidenced by works such as [78–86]. These methods are tailored for object removal tasks in general. However, there exists specific research dedicated to face mask removal, exemplified by works such as [2,3], along with diffusion-based models [87–90].



**Figure 8.** Face unmasking outputs from three state-of-the-art models: GANMasker, GUMF, and FFII-GatedCon. The first column shows the input masked face, while the second column displays the original unmasked face for reference.

**Table 5.** Summary of GAN-based face mask removal and image inpainting in general.

Model	Year	Dataset	PSNR	SSIM	FID	NIQ	BRISQUE	MAE	$L_1$	$L_2$
Context Encoders [78]	2016	Paris StreetView	18.58 dB	-	-	-	-	-	9.37%	1.96%
GFC (M5 and Q5) [84]	2017	CelebA	19.5	0.784	-	-	-	-	-	-
PConv (N/B) [81]	2018	Places2	18.21/19.04	0.468/0.484	-	-	-	-	6.45/5.72	-
FFII-GatedCon [80]	2019	Places2(rectangular mask)	-	-	-	-	-	-	8.6%	2.0%
		Places2(free-form mask)	-	-	-	-	-	-	9.1%	1.6%
EdgeConnect [83]	2019	Places2	21.75	0.823	8.16	-	-	-	3.86	-
MRGAN [85]	2019	Synthetic dataset	29.91 dB	0.937	-	3.548	29.97	-	-	-
ERFGOFI (Mask) [86]	2020	CelebA and CelebA-HQ	28.727	0.908	-	4.425	40.883	-	-	-
GUMF [3]	2020	CelebA	26.19 dB	0.864	3.548	5.42	37.85	-	-	-
		CelebA-HQ	40.40	0.94	3.09	-	-	31.91	-	-
R-MNet-0.4 [82]	2021	Paris Street View	39.55	0.91	17.64	-	-	33.81	-	-
		Places2	39.66	0.93	4.47	-	-	27.77	-	-
GANMasker [2]	2023	CelebA	30.96	0.95	16.34	4.46	19.27	-	-	-

Upon delving into the realm of image inpainting methods, one encounters a diverse landscape of approaches pioneered by various researchers. Among the earliest methodologies stands the work of P. Deepak, K. Philipp, et al. [78], who introduced a convolutional neural network (CNN)-based technique employing context encoders to predict missing pixels. Building upon this foundation, Iizuka et al. [79] proposed a generative adversarial network (GAN) framework equipped with two discriminators for comprehensive image completion. Similarly, Yu et al. [80] put forth a gated convolution-based GAN tailored specifically for free-mask image inpainting. Nazari et al. [83] devised a multi-stage approach involving an edge generator followed by image completion, facilitating precise inpainting. Additionally, Liu et al. [81] contributed to the field with their work on free-mask inpainting, leveraging partial convolutions to exclusively consider valid pixels and dynamically update masks during the forward pass.

Another subset of methods focuses on face completion or the removal of objects from facial images. Jam et al. [82] innovatively combined Wasserstein GAN with a Reverse Mask-

ing Network (R-MNet) for face inpainting and free-face mask completion. Similarly, Khan et al. [85] leveraged a GAN-based network to effectively remove microphones from facial images. Li et al. [84] devised a GAN architecture tailored specifically for generating missing facial components such as eyes, noses, and mouths. Further expanding the capabilities of inpainting, Ud Din et al. [86] introduced a two-stage GAN framework enabling users to selectively remove various objects from facial images based on their preferences, with the flexibility to remove multiple objects through iterative application.

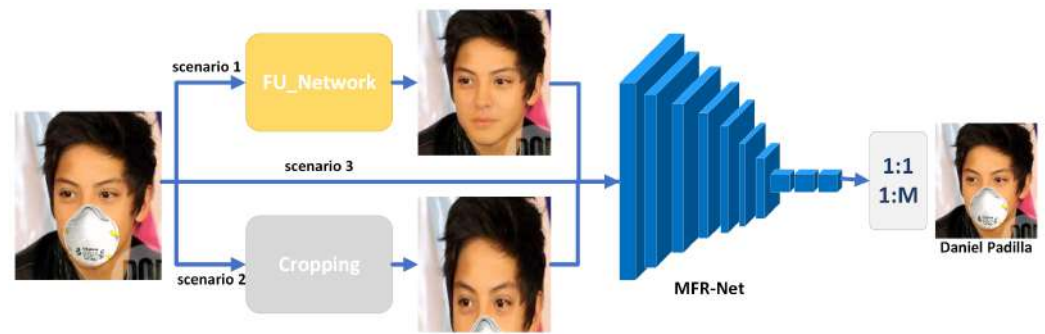
**Table 6.** Summary of diffusion-based face mask removal and image inpainting in general.

Model	Year	Dataset	Metric	Result
RePaint [87]	2022	CelebA-HQ	LPIPS (Half) LPIPS (Expand)	0.165 0.435
		ImageNet	LPIPS (Half) LPIPS (Expand)	0.304 0.629
DDRM-CC(SR) [89]	2022	ImageNet	PSNR SSIM KID LNFEs	26.55 0.74 6.56 20
DDNM [90]	2022	ImageNet	PSNR SSIM FID	32.06 0.968 3.89
		CelebA	PSNR SSIM FID	35.64 0.982 4.54
COPAINT-TT [88]	2023	CelebA-HQ	LPIPS (Half) LPIPS (Expand)	0.180 0.464
		ImageNet	LPIPS (Half) LPIPS (Expand)	0.294 0.636

After the COVID-19 pandemic, considerable attention has been directed towards the development of techniques for face mask removal, encompassing both GAN-based and diffusion-based approaches. Among the GAN-based methodologies, Mahmoud [2] introduced a two-stage network architecture, initially focusing on face mask region detection to guide the subsequent inpainting stage. Additionally, Mahmoud enhanced their results by integrating Masked–unmasked region Fusion (MURF) mechanisms. Furthermore, Din et al. [3] proposed a GAN-based network specifically designed for removing masks from facial images. Conversely, within the realm of diffusion models, Lugmayr et al. [87] introduced the RePaint method, which utilizes a DDPM [91] foundation for image inpainting tasks. Similarly, Zhang et al. [88] proposed the COPAINT method, enabling coherent inpainting of entire images without introducing mismatches. Broadly addressing image restoration, Kavar et al. [89] presented the Denoising Diffusion Restoration Model (DDRM), offering an efficient, unsupervised posterior sampling approach for various image restoration tasks. In a related context, Wang et al. [90] devised the Denoising Diffusion Null-Space Model (DDNM), a novel zero-shot framework applicable to diverse linear image restoration problems, including image super-resolution, inpainting, colorization, compressed sensing, and deblurring.

### 5.3. Masked Face Recognition Approaches

In this subsection, we delve into deep learning methodologies proposed to address the challenges faced by face recognition systems during the COVID-19 pandemic. The widespread use of masks has negatively affected the performance of traditional face recognition methods, encouraging authors to find novel approaches capable of effectively handling masked faces. Based on the important role of facial biometrics in various security systems and applications, it is important to address this issue by developing methods that perform robustly with both masked and unmasked faces. This subsection offers a comprehensive review of existing techniques for MFR, highlighting their diverse approaches and methodologies. The authors have pursued three distinct directions in masked face recognition, as illustrated in Figure 9, delineated as follows:



**Figure 9.** Three directions in masked face recognition (MFR): face restoration, masked region discarding, and deep learning-based approaches.

### 5.3.1. Face Restoration

Face restoration comprises two primary steps: initially, a model is employed for face unmasking to remove the mask and restore the hidden facial regions. Subsequently, another network is utilized to identify or verify the unmasked face. The primary objective within this category of the MFR task is to restore the face to its original, unmasked state. While the unmasking of faces and masked face recognition have traditionally been treated as distinct tasks, there are relatively few endeavors that amalgamate both within a single model. An example of such integration is evident in LTM [92], depicted in the final row of Table 7.

LTM [92] proposes an innovative approach to enhance masked face recognition through the utilization of amodal completion mechanisms within an end-to-end de-occlusion distillation framework. This framework comprises two integral modules: the de-occlusion module and the distillation module. The de-occlusion module leverages a generative adversarial network to execute face completion, effectively recovering obscured facial features and resolving appearance ambiguities caused by masks. Meanwhile, the distillation module employs a pre-trained general face recognition model as a teacher, transferring its knowledge to train a student model on completed faces generated from extensive online synthesized datasets. Notably, the teacher's knowledge is encoded with structural relations among instances in various orders, serving as a vital posterior regularization to facilitate effective adaptation. Through this comprehensive approach, the paper demonstrates the successful distillation and transfer of knowledge, enabling the robust identification of masked faces. Additionally, the framework's performance is evaluated across different occlusions, such as glasses and scarves. Notably, impressive accuracies of 98% and 94.1% are achieved, respectively, by employing GA [93] as an inpainting method.

**Table 7.** Summary of masked region discarding and face restoration-based methods for masked face recognition.

Model	Year	Dataset	Accuracy	mAP	Rank-1
IAMGAN-DCR (VGGFace2) [21]	2020	MFSR-REC	86.5%	42.7%	68.1%
IAMGAN-DCR (CAISA-Webface) [21]			82.3%	37.5%	67.4%
LPD [22]	2020	MFV	97.94%	-	-
		MFI	94.34%	49.08%	-
		Synthesized LFW	95.70%	75.92%	-
LTM [92]	2020	LFW	95.44%	-	-
		AR1	98.0%	-	-
		AR2	94.1%	-	-



Table 7. Cont.

Model	Year	Dataset	Accuracy	mAP	Rank-1
CA-MFR [94]	2021	Masked-Webface (Case 1)	91.525	-	-
		Masked-Webface (Case 2)	86.853%	-	-
		Masked-Webface (Case 3)	81.421%	-	-
		Masked-Webface (Case 4)	92.612%	-	-
Hariri [95]	2022	RMFRD	91.3%	-	-
		SMFRD	88.9%	-	-
UNMaskedArea-MFR (Cosine Similarity) [96]	2022	Custom (Indonesian people) Dataset	98.88%	-	-

### 5.3.2. Masked Region Discarding

Masked region discarding entails the removal of the masked area from the face image. This is achieved either by detecting the mask region and cropping out the masked portion or by using a predefined ratio to crop out a portion of the face that is typically unmasked. The remaining unmasked portion, usually containing facial features such as the eyes and forehead, is then utilized for training the recognition model. Table 7 provides a concise overview of the methodologies associated with this approach.

In one of these methods, Li et al. [94] explored the Occlusion Removal strategy by investigating various cropping ratios for the unmasked portion of the face. They incorporated an attention mechanism to determine the optimal cropping ratio, aiming to maximize accuracy. In their study, the optimal ratio was identified as  $0.9L$ , where  $L$  represents the Euclidean distance between the eye key points. The authors conducted experiments across four scenarios. In the first scenario, the model was trained using fully masked images and tested on masked images, achieving an accuracy of 91.529% with  $L = 0.9$ . The second and third scenarios involved training or testing with only one image masked, yielding accuracies of 86.853% and 82.533%, respectively, for  $L = 0.7$ . Finally, the fourth scenario adhered to the traditional face recognition approach, utilizing unmasked images for both training and testing purposes. Hariri [95] introduced a method that initially corrects the orientation of facial images and employs a cropping filter to isolate the unmasked areas. Feature extraction is conducted using pre-trained deep learning architectures including VGG-16, AlexNet, and ResNet-50. Leveraging pre-trained VGG-16 [97], AlexNet [98], and ResNet-50 [99] models for feature extraction in their masked face recognition approach, the effectiveness of these models in diverse image classification tasks has been well documented, showcasing their ability to attain high recognition accuracy. Feature maps are extracted from the final convolutional layer of these models, followed by the application of the bag-of-features (BoF) [100] methodology to quantize the feature vectors and create a condensed representation. The similarity between feature vectors and codewords is gauged using the Radial Basis Function (RBF) kernel. This approach demonstrates a superior recognition performance compared to other state-of-the-art methods, as evidenced by experimental evaluations on the RMFRD [17] and SMFRD [23] datasets, achieving accuracies of 91.3% and 88.9%, respectively.

Furthermore, G. Mengyue et al. [21] presented a comprehensive strategy aimed at overcoming the challenges associated with MFR through the introduction of innovative methodologies and datasets. Initially, the authors introduced the MFSR dataset, which includes masked face images annotated with segmentation and a diverse collection of full-face images captured under various conditions. To enrich the training dataset, they proposed the Identity Aware Mask GAN (IAMGAN), designed to synthesize masked face images from full-face counterparts, thereby enhancing the robustness of the dataset. Additionally, they introduce the Domain Constrained Ranking (DCR) loss to address intra-class variation, utilizing center-based cross-domain ranking to effectively align features between masked and full faces. Experimental findings on the MFSR dataset underscored the efficacy of the proposed approaches, underscoring their significance and contribution to the advancement of masked face recognition technologies. Fardause et al. [96] introduced

an innovative training methodology tailored for MFR, leveraging partial face data to achieve heightened accuracy. The authors curated their dataset, consisting of videos capturing faces across a range of devices and backgrounds, featuring 125 subjects. Drawing from established methodologies, such as employing YOLOv4 [101] for face detection, leveraging the pre-trained VGGFace model for feature extraction, and employing artificial neural networks for classification, the proposed system exhibits significant performance enhancements. While conventional training methods yielded a test accuracy of 79.58%, the adoption of the proposed approach resulted in a notable improvement, achieving an impressive test accuracy of 99.53%. This substantial performance boost underscores the effectiveness of employing a tailored training strategy for tasks related to masked face recognition.

Ding Feifei et al. [22] curated two datasets specifically tailored for MFR: MFV, containing 400 pairs of 200 identities for verification, and MFI, comprising 4916 images representing 669 identities for identification. These datasets were meticulously developed to address the scarcity of available data and serve as robust benchmarks for evaluating MFR algorithms. To augment the training data and overcome dataset limitations, they introduced a sophisticated data augmentation technique capable of automatically generating synthetic masked face images from existing facial datasets. Additionally, the authors proposed a pioneering approach featuring a two-branch CNN architecture. In this architecture, the global branch focused on discriminative global feature learning, while the partial branch was dedicated to latent part detection and discriminative partial feature learning. Leveraging the detected latent part, the model extracted discriminative features crucial for accurate recognition. Training the model involved utilizing both the original and synthetic training data, where images from both datasets were fed into the two-branch CNN network. Importantly, the parameters of the CNN in the two branches were shared, facilitating efficient feature learning and extraction.

### 5.3.3. Deep Learning-Based Masked Face Approaches

This subsection centers on leveraging deep learning techniques, often employing attention mechanisms to prioritize unmasked regions for feature extraction while attempting to mitigate the impact of the mask itself. Some authors opt to train models using a combined dataset of masked and unmasked faces, facilitating robustness to varying facial conditions. Unlike the previous methods, this approach does not require additional preprocessing steps such as face restoration or cropping of the upper face region. Table 8 presents a summary of the methodologies employed within this paradigm.

**Table 8.** Summary of deep learning-based approaches for masked face recognition.

Model	Year	Dataset	Accuracy	Precision	Recall	F1-Score	EER%
YOLOv3 [102]	2020	Custom dataset	93%	-	-	-	-
YOLO-face + VGGFace2			96.8%	-	-	-	-
RetinaFace + VGGFace2			94.5%	-	-	-	-
MFCosface [45]	2021	LFW_m	99.33%	-	-	-	-
		CF_m	97.03%	-	-	-	-
		MFR2	98.50%	-	-	-	-
		RMFRD	92.15%	-	-	-	-
MFR-CDC [103]	2021	SMFRD	95.31%	-	-	-	-
		RMFRD	95.22%	-	-	-	-
RggNet [104]	2021	Custom dataset	60.8%	77.7%	51.9%	-	-
MFCosface [45]	2021	LFW_m	99.33%	-	-	-	-
		CF_m	97.03%	-	-	-	-
		MFR2	98.50%	-	-	-	-
		RMFRD	92.15%	-	-	-	-

Table 8. Cont.

Model	Year	Dataset	Accuracy	Precision	Recall	F1-Score	EER%
MTArcFace [105]	2022	Masked-LFW	98.92%	-	-	-	-
		Masked-CFP_FF	98.33%	-	-	-	-
		Masked-CFP_FP	88.43%	-	-	-	-
		Masked-AGEDB_30	93.17%	-	-	-	-
		MFR2	99.41%	-	-	-	-
Deepmasknet [33]	2022	MDMFR	93.33%	93.00%	94.50%	93.74%	-
MFR-CNNandLBP [32]	2022	COMASK20	-	87%	87%	87%	-
		Essex dataset	-	99%	97%	98%	-
MFR-DML and FaceMaskNet-21 [106]	2022	User dataset	88.92%	-	-	-	-
		RMFRD	82.22%	-	-	-	-
		User dataset	88.186%	-	-	-	-
Att-Based-MFR (CASIA-Webface_m) [48]	2022	LFW_m	99.43%	99.30%	99.56%	99.43%	-
		AgeDB-30_m	95.86%	93.83%	97.82%	95.78%	-
		CFP-FP_m	97.74%	96.77%	98.69%	97.72%	-
		MFR2	96.75%	96.25%	97.22%	96.73%	-
Att-Based-MFR (VGG-Face2_m) [48]	2022	LFW_m	99.41%	99.26%	99.56%	99.40%	-
		AgeDB-30_m	95.38%	93.10%	98.11%	95.53%	-
		CFP-FP_m	96.98%	96.17%	98.40%	97.27%	-
		MFR2	99.00%	99.50%	98.45%	99.02%	-
Fine-Tuned MobileNet [107]	2022	MadFaRe (12 subjects)	78.41%	-	-	-	-
ResNet-100-MR-MP(SRT) [47]	2022	MFR	-	-	-	-	0.8270
		MRF2	-	-	-	-	3.4416
		LFW	-	-	-	-	0.9667
		IJB-C	-	-	-	-	2.9197
ResNet-50-MR-MP(SRT) [47]	2022	MFR	-	-	-	-	1.1207
		MRF2	-	-	-	-	6.2578
		LFW	-	-	-	-	1.2333
		IJB-C	-	-	-	-	3.0833
MobileFaceNet-MR-MP(SRT) [47]	2022	MFR	-	-	-	-	3.1866
		MRF2	-	-	-	-	7.8232
		LFW	-	-	-	-	2.2667
		IJB-C	-	-	-	-	4.6837
FaceNet + optimized SVM [108]	2023	CASIA + LWF + user dataset	99.912%	-	-	-	-
Lightweight CNN [109]	2024	HMFD (frontal image)	98.00%	98.00%	97.00%	98.00%	-
		HMFD (lateral image)	79.00%	83.00%	80.00%	79.00%	-

Building upon the ArcFace architecture, Montero David et al. [105] introduced an end-to-end approach for training face recognition models, incorporating modifications to the backbone and loss computation processes. Additionally, they implemented data augmentation techniques to generate masked versions of the original dataset and dynamically combine them during training. By integrating the face recognition loss with the mask-usage loss, they devised a novel function termed Multi-Task ArcFace (MTArcFace). Experimental results demonstrated that their model served as the baseline when utilizing masked faces, achieving a mean accuracy of 99.78% in mask-usage classification, while maintaining comparable performance metrics on the original dataset. On a parallel front, Deng Hongxia et al. [45] proposed their masked face recognition algorithm, leveraging large-margin cosine loss (MFCosface) to map feature samples in a space with a reduced intra-class distance and expanded inter-class distance. They further developed a masked face image generation algorithm based on the detection of key facial features, enabling the creation of corresponding masked face images. To enhance their model's performance and prioritize unmasked regions, they introduced an Att-inception module combining the Inception-Resnet module and the convolutional block attention module. This integration heightened the significance of unoccluded areas in the feature map, amplifying

their contribution to the identification process. Additionally, Wu GuiLing [103] proposed a masked face recognition algorithm based on an attention mechanism for contactless delivery cabinets amid the COVID-19 pandemic. By leveraging locally constrained dictionary learning, dilated convolution, and attention mechanism neural networks, the algorithm aimed to enhance the recognition rates of masked face images. The model, validated on the RMFRD and SMFRD databases, demonstrated a superior recognition performance. Furthermore, the algorithm addressed occlusion challenges by constructing subdictionaries for occlusion objects, effectively separating masks from faces. The network architecture incorporated dilated convolution for resolution enhancement and attention modules to guide model training and feature fusion. Overall, the proposed approach offers promising advancements in masked face recognition, crucial for ensuring the safety and efficiency of contactless delivery systems. Naeem Ullah et al. [33] introduced the DeepMasknet model, a novel construction designed for face mask recognition and masked facial recognition. Comprising 10 learned layers, the DeepMasknet model demonstrated effectiveness in both face mask recognition and masked facial recognition tasks. Furthermore, the authors curated a large and diverse unified dataset, termed the Mask Detection and Masked Facial Recognition (MDMFR) dataset, to evaluate the performance of these methods comprehensively. Experimental results conducted across multiple datasets, including the challenging cross-dataset setting, highlight the superior performance of the DeepMasknet framework compared to contemporary models.

Vu Hoai Nam et al. [32] proposed a methodology that leverages a fusion of deep learning techniques and Local Binary Pattern (LBP) features for recognizing masked faces. They employed RetinaFace, a face detector capable of handling faces of varying scales through joint extra-supervised and self-supervised multi-task learning, as an efficient encoder. Moreover, the authors extracted LBP features from specific regions of the masked face, including the eyes, forehead, and eyebrows, and integrated them with features learned from RetinaFace within a unified framework for masked face recognition. Additionally, they curated a dataset named COMASK20 comprising data from 300 subjects. Evaluation conducted on both the published Essex dataset and their self-collected COMASK20 dataset demonstrated notable improvements, with recognition results achieving an 87% F1-Score on COMASK20 and a 98% F1-Score on the Essex dataset. Golwalkar Rucha et al. [106] introduced a robust masked face recognition system, leveraging the FaceMaskNet-21 deep learning network and employing deep metric learning techniques. Through the generation of 128-dimensional encodings, the system achieves precise recognition from static images, live video feeds, and video recordings in real time. With testing accuracy reaching 88.92% and execution times under 10 ms, the system demonstrates high efficiency suitable for a variety of applications. Its effectiveness in real-world scenarios, such as CCTV surveillance in public areas and access control in secure environments, positions it as a valuable asset for bolstering security measures amid the widespread adoption of face masks during the COVID-19 pandemic. Kumar Manoj and Mann Rachit [104] explored the implications of face masks on the efficacy of face recognition methods, with a specific emphasis on face identification employing deep learning frameworks. Drawing from a tailored dataset derived from VGGFace2 and augmented with masks for 65 subjects, the research scrutinized the performance of prevalent pre-trained models like VGG16 and InceptionV3 after re-training on the masked dataset. Additionally, the study introduced a novel model termed RggNet, which capitalizes on a modified version of the ResNet architecture. This adaptation integrates supplementary layers within the shortcut paths of basic ResNet blocks, mirroring the structure of fundamental VGG blocks. This modification enables the model to effectively grasp an identity function, thereby fostering enhanced feature comprehension across layers. The proposed RggNet model architecture encompasses three sub-blocks organized akin to ResNet50v2, with customized identity blocks featuring convolution layers in lieu of direct shortcuts. Through meticulous experimental analysis, the study endeavored to offer valuable insights into bolstering masked face identification tasks amid the prevalent use of face masks in everyday contexts. Pann Vandet and Lee Hyo Jong [48] introduced an innovative



approach to MFR utilizing deep learning methodologies, notably the convolutional block attention module (CBAM) and angular margin ArcFace loss. By prioritizing the extraction of critical facial features, particularly around the eyes, essential for MFR tasks, their method effectively addressed the challenges posed by facial masks. To mitigate data scarcity, data augmentation techniques were employed to generate masked face images from traditional face recognition datasets. The refined ResNet-50 architecture acted as the backbone for feature extraction, augmented with CBAM to enhance efficiency in feature extraction. The resulting 512-dimensional face embeddings were optimized using the ArcFace loss function, leading to significant enhancements in MFR performance. Experimental findings corroborated the effectiveness of the proposed approach, underscoring its potential for practical applications within the realm of COVID-19 safety protocols. Kocacinar Busra et al. [107] presented a real-time masked detection service and mobile face recognition application aimed at identifying individuals who either do not wear masks or wear them incorrectly. Through the utilization of fine-tuned lightweight convolutional neural networks (CNNs), the system achieved a validation accuracy of 90.40% using face samples from 12 individuals. The proposed approach adopted a two-stage methodology: initially, a deep model discerns the mask status, categorizing individuals as masked, unmasked, or improperly masked. Subsequently, a face identification module employs traditional and eye-based recognition techniques to identify individuals. This system represents a significant advancement in masked face recognition, effectively addressing the challenges associated with masks in digital environments.

Deng Hongxia et al. [45] introduced MFCosface, an innovative algorithm tailored for masked face recognition amid the challenges posed by the COVID-19 pandemic. To mitigate the shortage of masked face data, the algorithm incorporated a novel masked face image generation method that utilizes key facial features for realistic image synthesis. Departing from conventional triplet loss approaches, MFCosface employed a large margin cosine loss function, optimizing feature mapping to bolster inter-class discrimination. Moreover, an Att-inception module was introduced to prioritize unoccluded facial regions, essential for precise recognition. Experimental findings across diverse datasets underscored the algorithm's notable enhancement in masked face recognition accuracy, presenting a promising solution for facial recognition in mask-wearing scenarios. Md Omar Faruque et al. [109] proposed a lightweight deep learning approach, leveraging the HSTU Masked Face Dataset (HMFD) and employing a customized CNN model to improve masked face identification. Integration of key techniques such as dropout, batch normalization, and depth-wise normalization optimized the model performance while minimizing complexity. In comparison to established deep learning models like VGG16 and MobileNet, the proposed model achieved a superior recognition accuracy of 97%. The methodology encompasses dataset preprocessing, model creation, training, testing, and evaluation, ensuring a robust performance in real-world scenarios. Transfer learning from pre-trained models such as VGG16 and VGG19, along with grid search for hyperparameter optimization, enhanced the model effectiveness. The architecture incorporated depthwise separable convolutions and carefully chosen layers to strike a balance between computational efficiency and accuracy, demonstrating exceptional performance even when facial features were partially obscured by masks. With an emphasis on simplicity and effectiveness, this lightweight CNN model offers a promising solution for recognizing masked faces, contributing to public health and safety efforts during the pandemic.

Putthiporn Thanathamathet et al. [109] conducted a study aimed at improving facial and masked facial recognition using deep learning and machine learning methods. Unlike previous research that often overlooked parameter optimization, this study employed a sophisticated approach. By integrating grid search, hyperparameter tuning, and nested cross-validation, significant progress was achieved. The SVM model, after hyperparameter tuning, achieved the highest accuracy of 99.912%. Real-world testing confirmed the efficacy of the approach in accurately identifying individuals wearing masks. Through enhancements in model performance, generalization, and robustness, along with improved data

utilization, this study offers promising prospects for strengthening security systems, especially in domains like public safety and healthcare. Vivek Aswal et al. [108] introduced two methodologies for detecting and identifying masked faces using a single-camera setup. The first method employed a single-step process utilizing a pre-trained YOLO-face/YOLOv3 model. Conversely, the second approach involved a two-step process integrating RetinaFace for face localization and VGGFace2 for verification. The results from experiments conducted on a real-world dataset exhibited robust performance, with RetinaFace and VGGFace2 achieving impressive metrics. Specifically, they attained an overall accuracy of 92.7%, a face detection accuracy of 98.1%, and a face verification accuracy of 94.5%. These methodologies incorporated advanced techniques such as anchor box selection, context attention modules, and transfer learning to enhance the efficiency and effectiveness of detecting masked faces and verifying identities. Fadi Boutros et al. [47] introduced an innovative method to improve masked face recognition performance by integrating the Embedding Unmasking Model (EUM) with established face recognition frameworks. Their approach incorporated the Self-Restrained Triplet (SRT) loss function, enabling the EUM to generate embeddings closely resembling those of unmasked faces belonging to the same individuals. The SRT loss effectively addressed intra-class variation while maximizing inter-class variation, dynamically adjusting its learning objectives to ensure a robust performance across various experimental scenarios. Leveraging fully connected neural networks (FCNN), the EUM architecture demonstrated adaptability to different input shapes, thereby enhancing its versatility. Rigorous evaluation of multiple face recognition models and datasets, including both real-world and synthetically generated masked face datasets, consistently revealed significant performance enhancements achieved by the proposed approach.

## 6. Limitations of Existing Works

However significant the progress in the domains of MFR, FMR, and FU, the current techniques still exhibit notable limitations. These limitations come from the challenges of data availability, model adaptability, and real-world performance, which hinder the practical application of these technologies. Below, we outline some key limitations:

- **Dependence on Synthetic Data:** Many state-of-the-art MFR and FU models are trained on synthetic datasets that simulate the presence of masks. While these datasets are essential due to the scarcity of real-world masked face data, they often fail to capture the complex variations in mask types, lighting conditions, facial structures, and occlusions seen in practice. As a result, models trained on synthetic data tend to perform well in controlled environments but show significant performance degradation when applied to real-world settings, where facial occlusions may vary unpredictably.
- **Lack of Robustness to Diverse Occlusions:** Current FU and MFR methods struggle to generalize across a wide range of occlusions. For instance, while a model may perform reasonably well when dealing with standard medical masks, it may falter when confronted with different types of face coverings (e.g., scarves, transparent masks, or masks with varying shapes and patterns). This limitation restricts the versatility and scalability of existing models, as their performance heavily depends on the types of occlusions present in the training data.
- **Weakness in Accurately Reconstructing Fine-Grained Features:** Face unmasking models typically focus on restoring critical facial features such as the mouth, nose, and chin. However, accurately reconstructing fine-grained details like lip color, facial hair, skin texture, and irregularities (e.g., scars or spots) remains a challenge. These subtleties are crucial for applications requiring high fidelity, such as identification and authentication systems.
- **Computational Complexity:** Advanced techniques such as GAN-based (generative adversarial networks) or diffusion-based models often require substantial computational power. While these models excel in generating visually plausible face reconstructions, their resource-intensive nature makes real-time processing difficult, especially on edge devices or mobile platforms. This limits the applicability of these models in

environments where computational resources are constrained, such as CCTV systems or on-device recognition systems.

- **Generalization to Unseen Data:** Transfer learning and domain adaptation techniques have been explored to enhance the generalization capabilities of MFR models. However, the effectiveness of these methods is limited when faced with real-world masked datasets that differ significantly from the training data. In many cases, the models are sensitive to variations in demographics, lighting, and pose, leading to a decreased performance when applied to diverse populations or environments outside of their training conditions.

## 7. Future Research Directions

Looking ahead, several avenues for future research hold promise for improving the current state of MFR, FMR, and FU technologies. Key areas of focus should include the following:

- **Real-World Dataset Availability:** A major challenge for both MFR and FU is the limited availability of real-world masked face data. Future research should focus on developing methods for collecting and organizing diverse, high-quality masked face datasets that represent a wide range of mask types, facial features, and environmental conditions. These efforts will be essential to enhance the generalization capabilities of models beyond controlled, synthetic datasets.
- **Synthetic Dataset Generation:** Addressing the challenge of dataset scarcity, generating high-quality synthetic masked face datasets can provide a significant solution. Advanced techniques like GAN-based augmentation, domain adaptation, and multi-task learning can be employed to produce more realistic masked face images, closely simulating real-world conditions, and occlusions.
- **Lightweight Models for Real-Time Applications:** To reconcile computational demands with real-time performance needs, future research should prioritize the development of efficient models for MFR and FU tasks. This involves exploring lightweight architectures and hardware-aware optimizations capable of running effectively on edge devices, while still achieving high accuracy.
- **Face Unmasking as a Preprocessing Step:** Combining face unmasking with masked face recognition could enhance the robustness of recognition systems. By treating FU as a preprocessing step, models can achieve better accuracy in scenarios with varying degrees of facial occlusion. This approach would also help mitigate the challenges posed by inconsistent facial coverings.
- **Cross-Disciplinary Integration:** Finally, the integration of MFR, FMR, and FU with other biometric systems (such as voice recognition or gait analysis) could offer multi-modal solutions that enhance the reliability and accuracy of identity verification systems, especially in security-critical applications.

By addressing these areas, future research can contribute significantly to advancing the state of MFR and FU systems, making them more adaptable, efficient, and reliable for real-world deployment.

## 8. Conclusions

This survey paper has conducted a thorough investigation into recent advancements and challenges within the realms of masked face recognition (MFR), face mask removal (FMR), and face unmasking (FU). By examining various methodologies, we have identified several significant strides made in these areas, particularly in addressing the occlusion challenges posed by masks and improving the generalization of models to different masked face conditions.

Throughout our exploration, we have highlighted the importance of refining synthetic data generation techniques, incorporating real-world masked datasets, and improving the performance of deep learning methods tailored to MFR and FU tasks. Researchers have also made notable advancements by developing innovative techniques for extracting and

reconstructing unmasked facial features from masked images, especially through deep learning models like GANs and diffusion-based architectures.

While substantial progress has been made, the limitations identified indicate that there are still considerable gaps that need to be addressed to ensure these technologies achieve reliable, real-time performance across diverse scenarios. Ethical and privacy concerns also remain a key aspect that must be considered as these technologies evolve and find broader applications.

**Author Contributions:** Conceptualization, M.M.; methodology, M.M.; software, M.M.; validation, M.M., M.S.K. and H.-S.K.; formal analysis, M.M., M.S.K. and H.-S.K.; investigation, M.M. and H.-S.K.; resources, M.M. and H.-S.K.; data curation, M.M. and M.S.K.; writing—original draft preparation, M.M.; writing—review and editing, M.M., M.S.K. and H.-S.K.; visualization, M.M. and H.-S.K.; supervision, M.M. and H.-S.K.; project administration, H.-S.K.; funding acquisition, H.-S.K. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (Ministry of Science and ICT) (No. 2023R1A2C1006944, 50%) and partly by Innovative Human Resource Development for Local Intellectualization program through the Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (IITP-2024-2020-0-01462, 50%).

**Data Availability Statement:** Not applicable.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

MFR	Masked face recognition
FMR	Face mask recognition
FU	Face unmasking
GANs	Generative adversarial networks
SSPS	Single sample per subject
RMFRD	Real-World Masked Face Recognition Dataset
MFSR	Masked Face Segmentation and Recognition Dataset
MFV	Masked Face Verification
MFI	Masked Face Identification
MFDD	Masked Face Detection Dataset
MDMFR	Mask Detection and Masked Facial Recognition Dataset
FMLD	Face-Mask Label Dataset
ISL-UFMD	Interactive Systems Labs Unconstrained Face Mask Dataset
MD-Kaggle	Masked Dataset
PWMFD	Properly Wearing Masked Face Detection Dataset
WMD	Wearing Mask Detection
WMC	Wearing Mask Classification
BAFMD	Bias-Aware Face Mask Detection
SMFRD	Simulated Mask Face Recognition Dataset
CMFD	Correctly Masked Face Dataset
IMFD	Incorrectly Masked Face Dataset
LFW	Labeled Faces in the Wild
IJB-C	IARPA Janus Benchmark-C
CFP	Celebrities in Frontal-Profile
ERR	Error rate
ROC	Receiver Operating Characteristic Curve
TP	True positive
TN	True negative
FP	False positive
FN	False negative
AUC	Area Under the Curve
FAR	False acceptance rate
FRR	False rejection rate



EER	Equal Error Rate
IoU	Intersection over Union
AP	Average precision
mAP	Mean Average Precision
PSNR	Peak Signal-to-Noise Ratio
MSE	Mean squared error
SSIM	Structural Similarity Index Measure
FID	Fréchet Inception Distance
NIQE	Naturalness Image Quality Evaluator
BRISQUE	Blind/Referenceless Image Spatial Quality Evaluator
SOTA	State-of-the-art
ML	Machine learning
DL	Deep learning
CNNs	Convolutional neural networks
RRFMDS	Rapid Real-Time Face Mask Detection System
SSD	Single Shot Multibox Detector
SMFD	Simulated Masked Face Dataset
SVM	Support Vector Machine
K-NN	K-Nearest Neighbors
BLS	Broad Learning System
CMNV2	Caffe-MobileNetV2
YOLO	You Only Look Once
SE	Squeeze and Excitation
SK	Selective Kernel
FF	Feature Fusion
CIoU	Complete Intersection over Union
R-MNet	Reverse Masking Network
MURF	Masked–unmasked region Fusion
DDRM	Denoising Diffusion Restoration Models
DDNM	Denoising Diffusion Null-Space Model
BoF	Bag-of-features
RBF	Radial Basis Function
IAMGAN	Identity Aware Mask GAN
DCR	Domain Constrained Ranking
MTArcFace	Multi-Task ArcFace
LBP	Local Binary Pattern
CBAM	Convolutional Block Attention Module
HMFD	HSTU Masked Face Dataset
EUM	Embedding Unmasking Model
SRT	Self-Restrained Triplet
FCNN	Fully Connected Neural Network

## References

1. Zhang, S.; Chi, C.; Lei, Z.; Li, S.Z. Refineface: Refinement neural network for high performance face detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *43*, 4008–4020. [[CrossRef](#)]
2. Mahmoud, M.; Kang, H.S. GANMasker: A Two-Stage Generative Adversarial Network for High-Quality Face Mask Removal. *Sensors* **2023**, *23*, 7094. [[CrossRef](#)]
3. 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](#)]
4. Yi, D.; Lei, Z.; Liao, S.; Li, S.Z. Learning face representation from scratch. *arXiv* **2014**, arXiv:1411.7923.
5. Liu, Z.; Luo, P.; Wang, X.; Tang, X. Deep Learning Face Attributes in the Wild. In Proceedings of the International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015.
6. 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, 7–20 October 2008.
7. Anwar, A.; Raychowdhury, A. Masked face recognition for secure authentication. *arXiv* **2020**, arXiv:2008.11104.
8. 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.

9. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 7132–7141.
10. 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.
11. 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. [\[CrossRef\]](#)
12. Zhang, Z.; Ji, X.; Cui, X.; Ma, J. A survey on occluded face recognition. In Proceedings of the 2020 9th International Conference on Networks, Communication and Computing, Tokyo, Japan, 18–20 December 2020; pp. 40–49.
13. Zeng, D.; Veldhuis, R.; Spreeuwiers, L. A survey of face recognition techniques under occlusion. *IET Biom.* **2021**, *10*, 581–606. [\[CrossRef\]](#)
14. 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. [\[CrossRef\]](#)
15. Wang, B.; Zheng, J.; Chen, C.P. A survey on masked facial detection methods and datasets for fighting against COVID-19. *IEEE Trans. Artif. Intell.* **2021**, *3*, 323–343. [\[CrossRef\]](#)
16. Nowrin, A.; Afroz, S.; Rahman, M.S.; Mahmud, I.; Cho, Y.Z. Comprehensive review on facemask detection techniques in the context of COVID-19. *IEEE Access* **2021**, *9*, 106839–106864. [\[CrossRef\]](#)
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. [\[CrossRef\]](#)
18. Pro, T.D. Face Masks Detection Dataset. Dataset. Available online: <https://trainingdata.pro/data-market/face-masks-detection/#header-form> (accessed on 25 January 2024).
19. 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.
20. MVD, A. Mask Dataset. Kaggle Dataset. 2020. Available online: <https://www.kaggle.com/datasets/andrewmvd/face-mask-detection/data> (accessed on 10 February 2024).
21. 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.
22. 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.
23. Wang, Z.; Huang, B.; Wang, G.; Yi, P.; Jiang, K. Masked Face Recognition Dataset and Application. *IEEE Trans. Biom. Behav. Identity Sci.* **2023**, *5*, 298–304. [\[CrossRef\]](#)
24. AIZOOTech. AIZOOTech-FaceMasksDetection. GitHub Repository. 2021. Available online: <https://github.com/AIZOOTech/FaceMaskDetection/tree/master> (accessed on 8 February 2024).
25. Roy, B.; Nandy, S.; Ghosh, D.; Dutta, D.; Biswas, P.; Das, T. MOXA: A deep learning based unmanned approach for real-time monitoring of people wearing medical masks. *Trans. Indian Natl. Acad. Eng.* **2020**, *5*, 509–518. [\[CrossRef\]](#)
26. Batagelj, B.; Peer, P.; Štruc, V.; Dobrišek, S. How to Correctly Detect Face-Masks for COVID-19 from Visual Information? *Appl. Sci.* **2021**, *11*, 2070. [\[CrossRef\]](#)
27. Singh, S.; Ahuja, U.; Kumar, M.; Kumar, K.; Sachdeva, M. Face mask detection using YOLOv3 and faster R-CNN models: COVID-19 environment. *Multimed. Tools Appl.* **2021**, *80*, 19753–19768. [\[CrossRef\]](#)
28. 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. [\[CrossRef\]](#)
29. Eyiokur, F.I.; Ekenel, H.K.; Waibel, A. Unconstrained face mask and face-hand interaction datasets: Building a computer vision system to help prevent the transmission of COVID-19. *Signal Image Video Process.* **2022**, *17*, 1027–1034. [\[CrossRef\]](#) [\[PubMed\]](#)
30. Jiang, X.; Gao, T.; Zhu, Z.; Zhao, Y. Real-time face mask detection method based on YOLOv3. *Electronics* **2021**, *10*, 837. [\[CrossRef\]](#)
31. Wang, B.; Zhao, Y.; Chen, C.P. Hybrid transfer learning and broad learning system for wearing mask detection in the COVID-19 era. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 1–12. [\[CrossRef\]](#)
32. Vu, H.N.; Nguyen, M.H.; Pham, C. Masked face recognition with convolutional neural networks and local binary patterns. *Appl. Intell.* **2022**, *52*, 5497–5512. [\[CrossRef\]](#)
33. Ullah, N.; Javed, A.; Ghazanfar, M.A.; Alsufyani, A.; Bourouis, S. A novel DeepMaskNet model for face mask detection and masked facial recognition. *J. King Saud-Univ.-Comput. Inf. Sci.* **2022**, *34*, 9905–9914. [\[CrossRef\]](#) [\[PubMed\]](#)
34. Benitez-Garcia, G.; Takahashi, H.; Jimenez-Martinez, M.; Olivares-Mercado, J. TFM a Dataset for Detection and Recognition of Masked Faces in the Wild. In Proceedings of the 4th ACM International Conference on Multimedia in Asia, Tokyo, Japan, 13–16 December 2022; pp. 1–7.
35. Kantarcı, A.; Ofli, F.; Imran, M.; Ekenel, H.K. Bias-Aware Face Mask Detection Dataset. *arXiv* **2022**, arXiv:2211.01207.
36. Yang, S.; Luo, P.; Loy, C.C.; Tang, X. Wider face: A face detection benchmark. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 5525–5533.
37. Karras, T.; Laine, S.; Aila, T. A style-based generator architecture for generative adversarial networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 4401–4410.

38. 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, 7–10 March 2016; pp. 1–9.
39. 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.
40. Maze, B.; Adams, J.; Duncan, J.A.; Kalka, N.; Miller, T.; Otto, C.; Jain, A.K.; Niggel, W.T.; Anderson, J.; Cheney, J.; et al. Iarpa janus benchmark-c: Face dataset and protocol. In Proceedings of the 2018 International Conference on Biometrics (ICB), Gold Coast, QLD, Australia, 20–23 February 2018; pp. 158–165.
41. Karras, T.; Aila, T.; Laine, S.; Lehtinen, J. Progressive growing of gans for improved quality, stability, and variation. *arXiv* **2017**, arXiv:1710.10196.
42. 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 (FG 2018), Xi'an, China, 15–19 May 2018; pp. 67–74.
43. SB, P. SMFD-GitHub. GitHub Repository. 2020. Available online: <https://github.com/prajnasb/observations/tree/master/experiments/data> (accessed on 8 February 2024).
44. Gurav, O. FMDD-kaggle. Kaggle Dataset. 2020. Available online: <https://www.kaggle.com/datasets/omkargurav/face-mask-dataset> (accessed on 8 February 2024).
45. 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. [\[CrossRef\]](#)
46. 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. [\[CrossRef\]](#) [\[PubMed\]](#)
47. Boutros, F.; Damer, N.; Kirchbuchner, F.; Kuijper, A. Self-restrained triplet loss for accurate masked face recognition. *Pattern Recognit.* **2022**, *124*, 108473. [\[CrossRef\]](#) [\[PubMed\]](#)
48. Pann, V.; Lee, H.J. Effective attention-based mechanism for masked face recognition. *Appl. Sci.* **2022**, *12*, 5590. [\[CrossRef\]](#)
49. 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.
50. King, D.E. Dlib-ml: A machine learning toolkit. *J. Mach. Learn. Res.* **2009**, *10*, 1755–1758.
51. Xiong, Z.; Wang, Z.; Du, C.; Zhu, R.; Xiao, J.; Lu, T. An asian face dataset and how race influences face recognition. In Proceedings of the Advances in Multimedia Information Processing—PCM 2018: 19th Pacific-Rim Conference on Multimedia, Hefei, China, 21–22 September 2018; pp. 372–383.
52. Zweig, M.H.; Campbell, G. Receiver-operating characteristic (ROC) plots: A fundamental evaluation tool in clinical medicine. *Clin. Chem.* **1993**, *39*, 561–577. [\[CrossRef\]](#)
53. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [\[CrossRef\]](#)
54. Heusel, M.; Ramsauer, H.; Unterthiner, T.; Nessler, B.; Hochreiter, S. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 6629–6640.
55. Mittal, A.; Soundararajan, R.; Bovik, A.C. Making a “completely blind” image quality analyzer. *IEEE Signal Process. Lett.* **2012**, *20*, 209–212. [\[CrossRef\]](#)
56. Mittal, A.; Moorthy, A.K.; Bovik, A.C. No-reference image quality assessment in the spatial domain. *IEEE Trans. Image Process.* **2012**, *21*, 4695–4708. [\[CrossRef\]](#) [\[PubMed\]](#)
57. Habib, S.; Alsanea, M.; Aloraini, M.; Al-Rawashdeh, H.S.; Islam, M.; Khan, S. An efficient and effective deep learning-based model for real-time face mask detection. *Sensors* **2022**, *22*, 2602. [\[CrossRef\]](#) [\[PubMed\]](#)
58. Boulila, W.; Alzahem, A.; Almoudi, A.; Afifi, M.; Alturki, I.; Driss, M. A deep learning-based approach for real-time facemask detection. In Proceedings of the 2021 20th IEEE International Conference on Machine Learning and Applications (ICMLA), Pasadena, CA, USA, 13–16 December 2021; pp. 1478–1481.
59. Shamrat, F.J.M.; Chakraborty, S.; Billah, M.M.; Al Jubair, M.; Islam, M.S.; Ranjan, R. Face Mask Detection using Convolutional Neural Network (CNN) to reduce the spread of COVID-19. In Proceedings of the 2021 5th International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, India, 3–5 June 2021; pp. 1231–1237.
60. Goyal, H.; Sidana, K.; Singh, C.; Jain, A.; Jindal, S. A real time face mask detection system using convolutional neural network. *Multimed. Tools Appl.* **2022**, *81*, 14999–15015. [\[CrossRef\]](#) [\[PubMed\]](#)
61. Sethi, S.; Kathuria, M.; Kaushik, T. Face mask detection using deep learning: An approach to reduce risk of Coronavirus spread. *J. Biomed. Inform.* **2021**, *120*, 103848. [\[CrossRef\]](#)
62. Chavda, A.; Dsouza, J.; Badgujar, S.; Damani, A. Multi-stage CNN architecture for face mask detection. In Proceedings of the 2021 6th International Conference for Convergence in Technology (i2ct), Maharashtra, India, 2–4 April 2021; pp. 1–8.
63. Umer, M.; Sadiq, S.; Alhebshi, R.M.; Alsubai, S.; Al Hejaili, A.; Nappi, M.; Ashraf, I. Face mask detection using deep convolutional neural network and multi-stage image processing. *Image Vis. Comput.* **2023**, *133*, 104657. [\[CrossRef\]](#)
64. Jignesh Chowdary, G.; Punna, N.S.; Sonbhadra, S.K.; Agarwal, S. Face mask detection using transfer learning of inceptionv3. In Proceedings of the Big Data Analytics: 8th International Conference, BDA 2020, Sonapat, India, 15–18 December 2020; pp. 81–90.

65. Oumina, A.; El Makhfi, N.; Hamdi, M. Control the COVID-19 pandemic: Face mask detection using transfer learning. In Proceedings of the 2020 IEEE 2nd International Conference on Electronics, Control, Optimization and Computer Science (ICECOCS), Kenitra, Morocco, 2–3 December 2020; pp. 1–5.
66. Loey, M.; Manogaran, G.; Taha, M.H.N.; Khalifa, N.E.M. A hybrid deep transfer learning model with machine learning methods for face mask detection in the era of the COVID-19 pandemic. *Measurement* **2021**, *167*, 108288. [\[CrossRef\]](#)
67. Nagrath, P.; Jain, R.; Madan, A.; Arora, R.; Kataria, P.; Hemanth, J. SSDMNV2: A real time DNN-based face mask detection system using single shot multibox detector and MobileNetV2. *Sustain. Cities Soc.* **2021**, *66*, 102692. [\[CrossRef\]](#)
68. Su, X.; Gao, M.; Ren, J.; Li, Y.; Dong, M.; Liu, X. Face mask detection and classification via deep transfer learning. *Multimed. Tools Appl.* **2022**, *81*, 4475–4494. [\[CrossRef\]](#)
69. Vignesh Baalaji, S.; Sandhya, S.; Sajidha, S.; Nisha, V.; Vimalapriya, M.; Tyagi, A.K. Autonomous face mask detection using single shot multibox detector, and ResNet-50 with identity retrieval through face matching using deep siamese neural network. *J. Ambient. Intell. Humaniz. Comput.* **2023**, *14*, 11195–11205. [\[CrossRef\]](#) [\[PubMed\]](#)
70. Sheikh, B.u.h.; Zafar, A. RRFMDs: Rapid real-time face mask detection system for effective COVID-19 monitoring. *SN Comput. Sci.* **2023**, *4*, 288. [\[CrossRef\]](#) [\[PubMed\]](#)
71. Kumar, B.A.; Bansal, M. Face mask detection on photo and real-time video images using Caffe-MobileNetV2 transfer learning. *Appl. Sci.* **2023**, *13*, 935. [\[CrossRef\]](#)
72. Yu, J.; Zhang, W. Face mask wearing detection algorithm based on improved YOLO-v4. *Sensors* **2021**, *21*, 3263. [\[CrossRef\]](#) [\[PubMed\]](#)
73. 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.
74. Wu, P.; Li, H.; Zeng, N.; Li, F. FMD-Yolo: An efficient face mask detection method for COVID-19 prevention and control in public. *Image Vis. Comput.* **2022**, *117*, 104341. [\[CrossRef\]](#)
75. Pham, T.N.; Nguyen, V.H.; Huh, J.H. Integration of improved YOLOv5 for face mask detector and auto-labeling to generate dataset for fighting against COVID-19. *J. Supercomput.* **2023**, *79*, 8966–8992. [\[CrossRef\]](#)
76. Zhang, H.; Tang, J.; Wu, P.; Li, H.; Zeng, N. A novel attention-based enhancement framework for face mask detection in complicated scenarios. *Signal Process. Image Commun.* **2023**, *116*, 116985. [\[CrossRef\]](#)
77. Tamang, S.; Sen, B.; Pradhan, A.; Sharma, K.; Singh, V.K. Enhancing COVID-19 safety: Exploring yolov8 object detection for accurate face mask classification. *Int. J. Intell. Syst. Appl. Eng.* **2023**, *11*, 892–897.
78. 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.
79. Iizuka, S.; Simo-Serra, E.; Ishikawa, H. Globally and locally consistent image completion. *Acm Trans. Graph. (Tog)* **2017**, *36*, 1–14. [\[CrossRef\]](#)
80. Yu, J.; Lin, Z.; Yang, J.; Shen, X.; Lu, X.; Huang, T.S. Free-form image inpainting with gated convolution. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 4471–4480.
81. Liu, G.; Reda, F.A.; Shih, K.J.; Wang, T.C.; Tao, A.; Catanzaro, B. Image inpainting for irregular holes using partial convolutions. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 85–100.
82. Jam, J.; Kendrick, C.; Drouard, V.; Walker, K.; Hsu, G.S.; Yap, M.H. R-mnet: A perceptual adversarial network for image inpainting. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, 3–8 January 2021; pp. 2714–2723.
83. Nazeri, K.; Ng, E.; Joseph, T.; Qureshi, F.Z.; Ebrahimi, M. Edgeconnect: Generative image inpainting with adversarial edge learning. *arXiv* **2019**, arXiv:1901.00212.
84. 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.
85. Khan, M.K.J.; Ud Din, N.; Bae, S.; Yi, J. Interactive removal of microphone object in facial images. *Electronics* **2019**, *8*, 1115. [\[CrossRef\]](#)
86. 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. [\[CrossRef\]](#)
87. Lugmayr, A.; Danelljan, M.; Romero, A.; Yu, F.; Timofte, R.; Van Gool, L. Repaint: Inpainting using denoising diffusion probabilistic models. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 11461–11471.
88. Zhang, G.; Ji, J.; Zhang, Y.; Yu, M.; Jaakkola, T.S.; Chang, S. Towards coherent image inpainting using denoising diffusion implicit models. In Proceedings of the 40th International Conference on Machine Learning, Honolulu, HI, USA, 23–29 July 2023.
89. Kawar, B.; Elad, M.; Ermon, S.; Song, J. Denoising diffusion restoration models. *Adv. Neural Inf. Process. Syst.* **2022**, *35*, 23593–23606.
90. Wang, Y.; Yu, J.; Zhang, J. Zero-shot image restoration using denoising diffusion null-space model. *arXiv* **2022**, arXiv:2212.00490.
91. Ho, J.; Jain, A.; Abbeel, P. Denoising diffusion probabilistic models. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 6840–6851.
92. 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. [\[CrossRef\]](#)



93. 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, 18–22 June 2018; pp. 5505–5514.
94. Li, Y.; Guo, K.; Lu, Y.; Liu, L. Cropping and attention based approach for masked face recognition. *Appl. Intell.* **2021**, *51*, 3012–3025. [[CrossRef](#)]
95. Hariri, W. Efficient masked face recognition method during the COVID-19 pandemic. *Signal Image Video Process.* **2022**, *16*, 605–612. [[CrossRef](#)] [[PubMed](#)]
96. Firdaus, F.; Munir, R. Masked face recognition using deep learning based on unmasked area. In Proceedings of the 2022 Second International Conference on Power, Control and Computing Technologies (ICPC2T), Raipur, India, 1–3 March 2022; pp. 1–6.
97. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
98. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. *Commun. ACM* **2017**, *60*, 84–90. [[CrossRef](#)]
99. 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.
100. Passalis, N.; Tefas, A. Learning bag-of-features pooling for deep convolutional neural networks. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 5755–5763.
101. Bochkovskiy, A.; Wang, C.Y.; Liao, H.Y.M. YOLOv4: Optimal speed and accuracy of object detection. *arXiv* **2020**, arXiv:2004.10934.
102. Aswal, V.; Tupe, O.; Shaikh, S.; Charniya, N.N. Single camera masked face identification. In Proceedings of the 2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA), Miami, FL, USA, 14–17 December 2020; pp. 57–60.
103. Wu, G. Masked face recognition algorithm for a contactless distribution cabinet. *Math. Probl. Eng.* **2021**, *2021*, 5591020. [[CrossRef](#)]
104. Kumar, M.; Mann, R. Masked face recognition using deep learning model. In Proceedings of the 2021 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N), Greater Noida, India, 17–18 December 2021; pp. 428–432.
105. Montero, D.; Nieto, M.; Leskovsky, P.; Aginako, N. Boosting masked face recognition with multi-task arcface. In Proceedings of the 2022 16th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS), Dijon, France, 19–21 October 2022; pp. 184–189.
106. Golwalkar, R.; Mehendale, N. Masked-face recognition using deep metric learning and FaceMaskNet-21. *Appl. Intell.* **2022**, *52*, 13268–13279. [[CrossRef](#)] [[PubMed](#)]
107. Kocacinar, B.; Tas, B.; Akbulut, F.P.; Catal, C.; Mishra, D. A real-time cnn-based lightweight mobile masked face recognition system. *IEEE Access* **2022**, *10*, 63496–63507. [[CrossRef](#)]
108. Thanathamath, P.; Sawangarereerak, S.; Kongkla, P.; Nizam, D.N.M. An Optimized Machine Learning and Deep Learning Framework for Facial and Masked Facial Recognition. *Emerg. Sci. J.* **2023**, *7*, 1173–1187. [[CrossRef](#)]
109. Faruque, M.O.; Islam, M.R.; Islam, M.T. Advanced Masked Face Recognition using Robust and Light Weight Deep Learning Model. *Int. J. Comput. Appl.* **2024**, *975*, 8887.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.