

A Survey on Computer Vision based Human Analysis in the COVID-19 Era

Fevziye Irem Eyiokur^{a,*}, Alperen Kantarcı^b, Mustafa Ekrem Erakin^b, Naser Damer^{c,d}, Ferda Offi^e,
Muhammad Imran^e, Janez Križaj^f, Albert Ali Salah^{g,h}, Alexander Waibel^{a,i}, Vitomir Štruc^f,
Hazim Kemal Ekenel^b

^a*Institute for Anthropomatics and Robotics, Karlsruhe Institute of Technology, Karlsruhe, Germany*

^b*Department of Computer Engineering, Istanbul Technical University, Istanbul, Turkey*

^c*Fraunhofer Institute for Computer Graphics Research IGD, Darmstadt, Germany*

^d*Department of Computer Science, TU Darmstadt, Darmstadt, Germany*

^e*Qatar Computing Research Institute, HBKU, Doha, Qatar*

^f*Faculty of Electrical Engineering, University of Ljubljana, Tržaška cesta 25, 1000 Ljubljana, Slovenia*

^g*Department of Information and Computing Sciences, Utrecht University, Utrecht, the Netherlands*

^h*Department of Computer Engineering, Boğaziçi University, Istanbul, Turkey*

ⁱ*Carnegie Mellon University, Pittsburgh, United States*

Abstract

The emergence of COVID-19 has had a global and profound impact, not only on society as a whole, but also on the lives of individuals. Various prevention measures were introduced around the world to limit the transmission of the disease, including face masks, mandates for social distancing and regular disinfection in public spaces, and the use of screening applications. These developments also triggered the need for novel and improved computer vision techniques capable of (i) providing support to the prevention measures through an automated analysis of visual data, on the one hand, and (ii) facilitating normal operation of existing vision-based services, such as biometric authentication schemes, on the other. Especially important here, are computer vision techniques that focus on the analysis of people and faces in visual data and have been affected the most by the partial occlusions introduced by the mandates for facial masks. Such computer vision based human analysis techniques include face and face-mask detection approaches, face recognition techniques, crowd counting solutions, age and expression estimation procedures, models for detecting face-hand interactions and many others, and have seen considerable attention over recent years. The goal of this survey is to provide an introduction to the problems induced by COVID-19 into such research and to present a comprehensive review of the work done in the computer vision based human analysis field. Particular attention is paid to the impact of facial masks on the performance of various methods and recent solutions to mitigate this problem. Additionally, a detailed review of existing datasets useful for the development and evaluation of methods for COVID-19 related applications is also provided. Finally, to help advance the field further, a discussion on the main open challenges and future research direction is given at the end of the survey. This work is intended to have a broad appeal and be useful not only for computer vision researchers but also the general public.¹

Keywords: Computer vision, COVID-19, human analysis, masked faces, survey

1. Introduction

The COVID-19 pandemic took the world by storm. Since the first large-scale outbreak in December 2019 in Wuhan, China, COVID-19, a highly infectious atypical (viral) pneumonia caused by the zoonotic

¹This is the uncorrected author proof, copyright with Elsevier. Please cite this paper as: Eyiokur, F.I., A. Kantarcı, M.E. Erakin, N. Damer, F. Offi, M. Imran, J. Križaj, A.A. Salah, A. Waibel, V. Štruc, and H.K. Ekenel, "A survey on computer vision based human analysis in the COVID-19 era," *Image and Vision Computing*, vol.130, p.104610, 2023.

*Corresponding author

Email address: fevziye.yaman@kit.edu (Fevziye Irem Eyiokur)

5 coronavirus SARS-CoV-2, spread throughout the globe and resulted in around 600 million recorded cases and over 6.5 million deaths by mid 2022 according to information from Worldometer² [1]. To contain the spread of the disease, minimize cases and limit the number of deaths, governments across the world started introducing prevention measures that had a profound impact on peoples' lives and changed their behavior and daily routines. Common prevention measures included (mandatory) face masks in public spaces, medical facilities and crowded areas, requests for social distancing, and restrictions on the allowed crowd size at
10 different events, among others [2, 3].

To help combat COVID-19, the computer vision community quickly took an active stance and initiated a wide range of research activities that resulted in novel techniques for COVID-19 detection and severity analysis from medical images [4, 5], monitoring solutions for assessing compliance with the given prevention measures [1, 6, 7], screening approaches for flagging potentially sick subjects [8, 9, 10], infection-risk
15 assessment methods [11], and efficient biometrics-based authentication schemes tailored towards the characteristics of the COVID-19 era [12, 13]. These solutions have been swiftly adopted in practice and were observed to have a critical role in the efforts towards containing the pandemic [14]. They allowed to automate many monitoring tasks, improved situation-awareness and facilitated large-scale screening efforts. Furthermore, themed workshops, such as the International Workshop on Face and Gesture Analysis for
20 COVID-19 (FG4COVID19)³, were organized in the scope of major computer vision conferences to provide a platform for discussion and presentation of the latest vision techniques related to COVID-19.

A key component of many of the solutions discussed above are, what we refer to in this survey as, *computer vision based human analysis* (CVHA) techniques that focus on the analysis of people and faces in visual data. While considerable progress has been made in the general area of vision based human analysis,
25 the COVID-19 pandemic introduced several new challenges that have been underexplored in the literature before, e.g.:

- **Mask-based occlusions:** One of the globally most prevalent prevention measures, introduced in response to COVID-19, are face masks. The presence of face masks represents a considerable obstacle with an adverse impact on the performance of many CVHA techniques, such as facial landmarking, face detection and recognition, but also related (auxiliary) tasks such as face image quality assessment (FIQA), presentation attack detection (PAD) and others. Dedicated mechanisms are, therefore, needed to handle this type of occlusion. It is important to note that partial occlusions of the facial area have been studied also in the pre-COVID-19 era [15, 16, 17]. However, most of the research from that period was not focused specifically on face masks and, as a result, techniques developed for more
30 general occlusions were observed to lead to suboptimal performance for many COVID-19 related human-centered vision tasks.
- **Relevant datasets:** The majority of modern vision techniques relies on machine learning and is, hence, trained using suitably annotated training data. Before the start of the pandemic, there was an obvious lack of datasets suitable for the development of CVHA techniques for combating COVID-19. Especially datasets with masked faces (and people) with high-quality annotations were not widely available. Several factors contributed to the lack of such datasets: (i) there was limited interest in vision problems involving masked-faces (and masked-people) only, (ii) the occlusions caused by the face masks made it difficult to generate annotations of reasonable quality (e.g., facial landmarks, accurate bounding boxes, segmentation/parsing maps, etc.), and (iii) a wide variety of facial masks with highly
35 diverse appearance was introduced during the pandemic, but was not available in the pre-COVID-19 era.
- **Ethical considerations and social impact:** While CVHA techniques for COVID-19 were developed with the goal of more efficiently combating the pandemic, the deployment of such techniques also raises ethical considerations and comes with a considerable societal impact. For example, installments of

²Accessible from: <https://www.worldometers.info/coronavirus/>

³URL: <https://fg4covid19.github.io/>

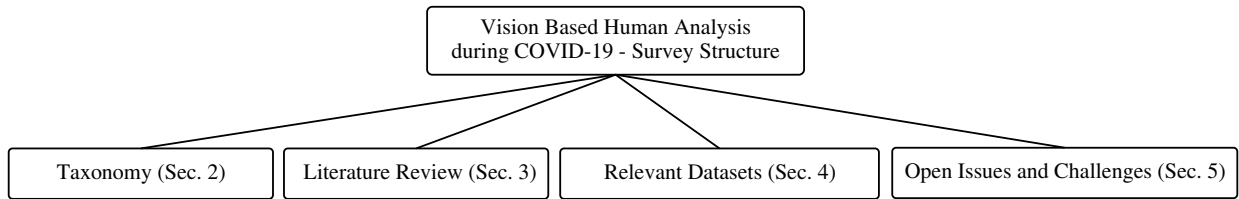


Figure 1: High-level structure of the survey. We provide a comprehensive review of recent human-centered computer vision techniques for combating COVID-19, discuss existing datasets and data-generation procedures, and present a list and discussion of the most important open issues and future research directions.

50 screening and monitoring applications can help to contain the spread of the coronavirus, but may also be extended into citizen surveillance and impact the privacy of individuals.

A considerable amount of work has been conducted over the course of the last three years to address the above challenges and has been covered partially in recent survey papers. Wang *et al.* [18], for example, reviewed techniques for masked facial detection and associated datasets. Alzubi *et al.* [19] as well as Utomo and Kusuma [20] discussed dedicated face recognition techniques for masked faces. Elbishlawi *et al.* [21] reviewed crowd-counting techniques and pointed to the importance of this technology for COVID-19. Related to these works is also the survey of Ulhaq *et al.* [4], which covers computer vision techniques applicable mostly to medical data, diagnostics and clinical management, and the deep-learning oriented review paper by Shorten *et al.* [22], where vision approaches, again mostly related to medical applications, are briefly discussed. Although the listed works provide some insight into CVHA research related to COVID-19, they focus on specific problems only, e.g., masked face detection or recognition, or provide a partial picture of the broader (and interconnected) research area. A well-structured and thorough survey on COVID-19 focused vision-based human analysis techniques, on the other hand, is, to the best of our knowledge, still missing from the literature.

65 In this work, we aim to address this gap and present a comprehensive overview of computer vision techniques that analyze visual data of people and faces with COVID-19 applications in mind. The goal of the survey is to: (i) provide a high-level taxonomy and background on vision techniques applied to human analysis relevant to COVID-19 (Sec. §2) (ii) present a consolidated summary of recent research activities in this area (Sec. §3), (iii) provide a review of dataset collection and generation efforts (Sec. §4), and (iv) elaborate on open problems and challenges with the goal of providing a basis for future research activities (Sec. §5). The overall structure of the survey is illustrated in Figure 1. The work is primarily intended for researchers looking for a broad overview of computer vision research for COVID-19, but also other stakeholders interested in this topic.

We make the following main contributions in this survey:

- 75 • We present a comprehensive review of computer vision techniques that analyze imagery of people and faces to support the COVID-19 containment efforts and discuss over 200 relevant references that cover diverse but relevant topics from this problem domain.
- We provide a taxonomy of existing solutions for the most relevant tasks studied as part of vision-based research for COVID-19, e.g., face mask detection, masked face detection, masked face recognition, crowd analysis, etc.
- 80 • We discuss issues beyond the technological solutions, such as ethics, social impact and elaborate on open problems and future research directions.

Since COVID-19 will not be the last pandemic the world faces, we believe this survey will help technological preparedness for similar situations, and ultimately improve the robustness and usability of relevant technologies.

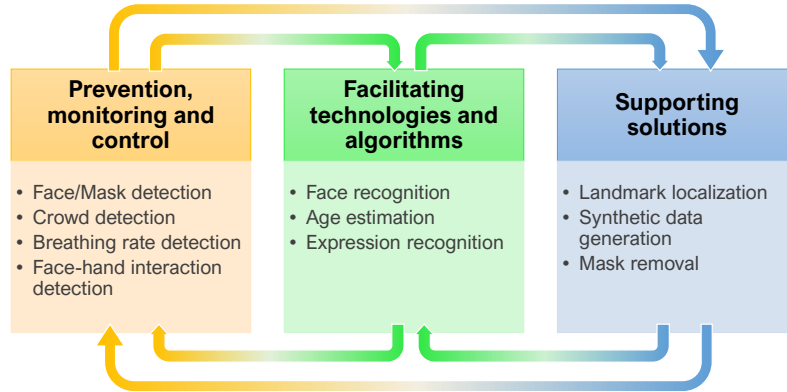


Figure 2: High-level taxonomy of Computer Vision based Human Analysis (CVHA) techniques surveyed in this paper with respect to the targeted goal.

2. Taxonomy on Computer Vision based Human Analysis (CVHA) for COVID-19

The COVID-19 pandemic triggered a need for efficient computer vision techniques related to different problems in visual human analysis that can broadly be categorized into three groups based on their overall goals, as also illustrated in Figure 2, i.e.:

- 90

Techniques for prevention, monitoring and control: The goal of the first group of CVHA techniques is to help prevent the spread of COVID-19 and monitor compliance with the given prevention measures and typically aim to detect/analyze some characteristics (e.g., the presence of masks, the crowd size, or physiological changes/abnormalities) of the subjects in the visual data. Techniques from this group are applicable for screening purposes and as a source of critical statistical data for governments, health organizations and regulatory bodies. CVHA solutions covered in this survey from this group include face/mask detection algorithms [23, 24, 25], crowd-counting solutions designed for COVID-19 characteristics [24, 26, 27, 28, 11], breathing rate detection techniques [29] and face-hand interaction detection approaches [30, 31].
- 95
- 100

Facilitating algorithms: The second group of techniques represents solutions that facilitate applications that are not immediately related to COVID-19 prevention, but whose performance is affected by the external circumstance caused by the pandemic, such as, the presence of face masks. A typical example of such an application is biometric identity inference from facial images, where face masks have been observed to have a considerable adverse effect on the overall recognition accuracy [32]. Many techniques and algorithms have, therefore, been proposed in the last few years to enable such critical applications also during COVID-19, but with minimal performance loss. CVHA techniques from the group of facilitating algorithms reviewed in this paper include face recognition solutions for masked faces [33, 32, 34, 35, 36, 37], as well age estimation [38, 39] and facial expression recognition approaches [40, 41] that were all extended recently with the goal of improving robustness with masked faces.
- 105
- 110

Supporting solutions: The last group of techniques in our taxonomy represents supporting solutions that do not address specific problems with real-world COVID-19-related applications, but are needed to enable techniques from the two groups above. The most important solutions from this group discussed in the survey are data generation techniques, capable of synthesizing artificial training data for the various computer vision models [33, 42, 43], mask removal techniques aiming to reconstruct the original (unoccluded) facial images [44, 45, 46] and landmark localization (or/and alignment) techniques [47, 48]
- 115

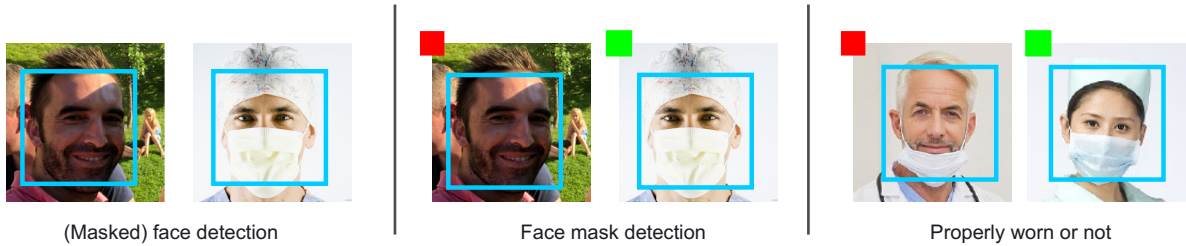


Figure 3: Masked face detection is a specific object detection problem where the objects to be detected (i.e., faces) can appear both, with and without masks (left). The different extensions that appeared during the COVID-19 pandemic also incorporate decisions on whether the masks are present in the images (middle) and whether the masks are worn correctly or not (right).

We note that there is no clear separation between these three groups and there are clear interdependencies that are present in the presented taxonomy, as also highlighted in the overview Fig. 2.

3. Survey of CVHA techniques in the COVID-19 era

In this section, we summarize research on the different CVHA techniques that emerged during the COVID-19 pandemic and are covered in this survey. Specifically, we discuss research efforts focused on (i) face/mask detection, (ii) face recognition and various auxiliary tasks needed for face recognition systems, such as presentation attack detection and face quality estimation, (iii) facial expression recognition, (iv) age classification, (v) landmark localization, (vi) crowd detection/counting, (vii) breathing rate detection, (viii) face-hand interaction detection, and (ix) synthetic data generation.

3.1. Face/Mask Detection

Among the various prevention measures introduced around the world, face masks were likely the most wide spread and, in fact, were mandatory in various countries [2, 3]. Facial masks were also supported by the World Health Organization (WHO), who published a detailed guide on this topic [49]. Computer vision based detection techniques for masked faces are typically needed to monitor whether people comply with the advice of health organizations and governments in public spaces and to facilitate situational awareness.

In general, masked face detection is a specialized object detection problem where, in addition to the standard detection of non-occluded faces, the goal is to also reliably identify the presence of faces with masks in an image (or video frame). This task typically includes a binary decision (face present/face absent) for a given sub-region of the input image, which also defines the approximate spatial location of the (masked) facial area. Extensions of this problem that emerged during the pandemic, in addition to the (masked) face detection task, also often detect the presence of masks in the image (mask present/mask absent) or/and determine whether the mask is placed/worn in accordance with regulations and general guidelines or not [50, 51, 52, 1, 30], as shown in Figure 3.

Before the pandemic, the masked face detection problem was mostly investigated as part of the more general detection tasks with partial occlusions, where the occlusions may have appeared due to the placement of the hands, the presence of sunglasses, gas masks, helmets, niqabs, and other objects that commonly cover some part of the face. In such settings, face detection methods are commonly observed to perform worse, with the performance degradations increasing as the occluded part of the face gets larger [53]. One of the earliest pre-COVID-19 works on detecting masked faces was presented by Nieto-Rodríguez et al. [54], where the authors introduced a system that checks whether medical staff wears mandatory medical masks in the operating room. They used two distinct detectors: one for the face and the other for the medical mask. They employed the Viola-Jones object detector [55] for both face and mask detection. They collected a dataset that contains faces with medical masks to train the detectors. Another work from this period [56] presented the first large-scale masked face detection dataset named MAFA and trained the locally linear embedding and convolutional neural networks (LLE-CNNs) for detecting faces with and without face masks. The proposed method first extracts (face) region proposals and describes them with a convolutional neural

network (CNN). After this, a k-nearest neighbor (KNN) module refines the descriptors for recovering missing facial cues of masked faces. As the last step, a unified CNN is used to perform classification and regression to identify candidate facial regions and their overall positions in images/frames.

155 In [57], the authors introduced a refined version of the MAFA [56] dataset, called MAFA-FMD, that included only images with medical face masks. Using this new dataset, the authors proposed a novel context attention module to extract highly descriptive contextual features, such as face mask wearing conditions, and showed that the proposed approach outperforms the benchmark RetinaFace [58] and YOLOv3 [59] face detectors. In [52], Nagrath et al. introduced the SSDMNV2 system, which combines a single-shot
160 multibox detector framework and a MobileNetV2 [60] based classifier for the detection of masked faces as well as face mask detection. This lightweight model is suitable for deploying on embedded devices and for real-time data processing. Another work that uses a single-stage face detector is [61]. Here, the authors used the YOLOv2 model [62] to detect masked faces and ResNet-50 [63] to detect face masks. To overcome the challenge of scarce labeled data, Cabani et al. [51] built a synthetic dataset of masked faces to train
165 robust face detection and face-mask detection models. The authors tried to imitate different mask-wearing conditions by using realistic image synthesis methods. However, they only used a single type of medical mask to simulate different wearing conditions, therefore, raising questions on the generalization capabilities of their models to real-world data, where facial masks may have different colors, shapes, and textures. Joshi et al. [50] proposed a framework to detect face masks from a video stream by using the MTCNN [64] face
170 detection model and classifying mask presence with MobileNetV2 [60]. They tested their framework on actual footage of public spaces, captured during the COVID-19 pandemic. The dataset contains multiple geographical locations and people from different ethnicities and the proposed method was demonstrated to outperform RetinaFaceMask [57] on the considered dataset. The authors of [65] proposed a two-stage Faster R-CNN [66] network with an InceptionV2 [67] model along with a novel wearing mask detection
175 (WMD) dataset to address the masked face detection task. Through comprehensive experiments, they show that the two-stage detector provides a good trade-off between accuracy and computational complexity. The work of Roy et al. [68] proposed using a YOLOv3 [59] model along with the Single Shot MultiBox Detector (SSD) [69] and Faster R-CNN [66] for masked face detection. The experimental results on the novel Moxa3K benchmark dataset [68] showed that YOLOv3 [59] achieves better performance than competing models while
180 having comparable runtime.

In [30], Eyiokur et al. studied an extended detection problem, where each face image was classified into one of three classes: no mask, mask, and incorrectly worn mask. The authors introduced a labeled large-scale face mask detection dataset and using the newly collected data trained a RetinaFace model [58] for face detection. Next, they employed several CNN models, namely, Inception-v3 [67], MobileNetV2 [60],
185 EfficientNet [70], etc., to classify the detected and cropped faces into the three above-mentioned classes. The authors also extensively tested their models, both on the proposed dataset as well as on other available datasets from the literature. Cross-dataset evaluations showed their dataset’s representation power, which is crucial for new face mask detection datasets. A similar problem was also studied in [71], where Jiang et al. proposed the Squeeze and Excitation (SE)- YOLOv3 [59] mask detector for the detection of properly worn
190 masks. The main idea behind the approach was to integrate the SE block with the YOLOv3 [59] model to teach the network to focus on the crucial features. The authors also utilized a focal loss to solve the extreme foreground-background class imbalance. Experimental results showed that the proposed network achieved better localization and detection performances than competing models on the considered dataset. Kantarci et al. [72] introduced a novel face mask detection dataset named Bias-Aware Face Mask Detection
195 (BAFMD) dataset. The dataset has been collected using Twitter images with a specific focus on mitigating dataset bias for ethnicity, age, and gender. In order to reduce such biases, their dataset contains real-world face mask images with a more balanced distribution across different demographics, e.g., gender, race, and age. They train a YOLOv5 [73] object detector, which shows superior performance over other detectors. In [1], Batagelj et al. compare different masked face detectors and correct face-mask placement classification
200 networks in detail using a dataset that they created using the MAFA [56] dataset. The reported experimental results provide insightful performance comparisons of various methods and show that the RetinaFace [58] model is the most stable masked face detection model among the considered techniques. Furthermore, the authors demonstrated that all face detection models’ performance deteriorate significantly, if face masks are

present in the image as compared to faces without masks. In [74], the authors proposed a face mask-wearing identification method by combining image super-resolution and classification networks (SRCNet). They used a standard face detector for detecting and cropping faces with and without masks. After the detection step, the authors evaluated the image size to choose the next step. If an image’s resolution was smaller than 150×150 pixels, i.e., the width or length was below 150 pixels, they applied the super-resolution model to enhance the high-frequency details of the image. If the image was already larger than 150×150 pixels, they skipped the super-resolution and subjected the image to a face mask-wearing classification network, which classify the mask-wearing conditions into one of three classes: no mask-wearing, incorrect mask-wearing, and correct mask-wearing. The reported experimental results show that applying super-resolution to low-resolution face crops boosts classification performance and that the presented model yielded competitive performance overall.

Most of the methods proposed for (masked) face detection and related problems, such as facial mask detection, build on advances made in the generic object detection problem domain. However, to adapt/extend the existing detectors to work reliably with partially occluded face data or, specifically, with masked faces, these solutions incorporate minor modifications to the overall detection pipelines and, more importantly, introduce new, large-scale datasets with masked faces that contain diverse data with rich appearance variations induced by facial masks for model training. Due to the importance of these datasets for the masked face detection problems, they are discussed separately in Section 4.

3.2. Face Recognition

Similarly to face detection, where the appearance and widespread usage of facial masks had an adverse impact on the performance of existing face detection models, face recognition is another area, where facial masks negatively impacted the applicability of face recognition technology. In this section, we therefore provide an in-depth discussion of the effect of face masks on different components of face recognition systems and then review the efforts made so far to mitigate such negative effects.

3.2.1. The effect of wearing masks on face recognition

Face recognition deployability is strongly affected by biometric sample capture and presentation, most prominently, face occlusions. Face recognition in the presence of occlusions has been studied widely within the computer vision community [17, 75, 76, 77, 78, 79, 80, 81]. However, most of the pre-COVID-19 work targeted general unstructured face occlusions. The effect of the specific occlusion induced by face masks gained attention at the start of the COVID-19 pandemic. An early work by Damer et al. [32] evaluated the verification performance drop of face recognition systems when verifying unmasked-to-masked faces, in comparison to verifying unmasked faces, all with real masks and in a collaborative environment. This was followed by an extended study [82] with a larger database and evaluation of both synthetic and real masks. As a part of the ongoing Face Recognition Vendor Test (FRVT), the National Institute of Standards and Technology (NIST) has released results (FRVT -Part 6A) on the effect of face masks on the performance of face recognition systems provided by vendors [83]. The results revealed that the verification accuracy with masked faces declined substantially. However, the study used simulated masked images under the assumption that their effect is representative of the effect of real face masks. Following NIST’s evaluation, the US Department of Homeland Security conducted a similar evaluation, however, on more realistic data [84]. They also identified a significant negative effect of facial masks on the accuracy of automatic face recognition solutions. A general conclusion by these studies was that the effect of masks was bigger on genuine pairs’ decisions, in comparison to imposter pairs’ decisions. A study comparing the effect of face masks on the human experts/verifiers in comparison to automatic face recognition models concluded with a set of comments on different aspects of the correlation between the verification performance of humans and machines [85]. The study showed a trend in the human experts’ verification performance drop similar to that of automatic face verification models [85]. In the next section, an overview of the solutions to mitigate this negative effect on face recognition performance is presented.

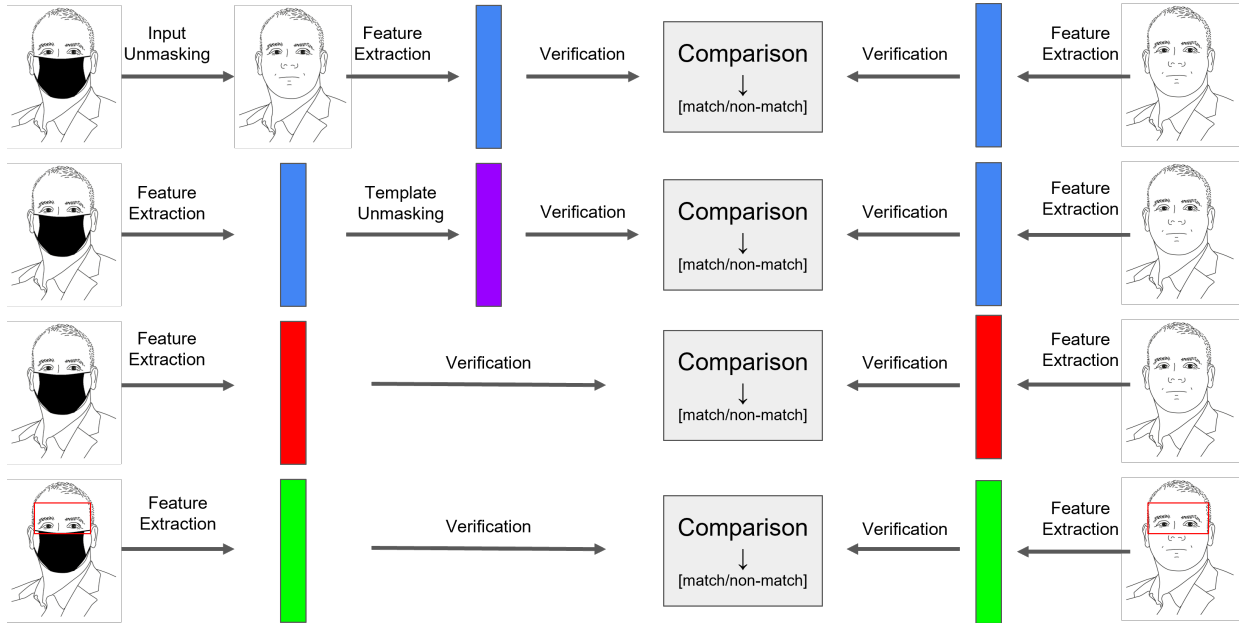


Figure 4: Masked face recognition solutions mainly fall under one of the four categories presented in this figure. From top to bottom, these solutions are face in-painting, template unmasking, model optimization, and periocular recognition. Blue rectangles are general-purpose face recognition models, purple rectangles are template unmasking models, red rectangles are face recognition models trained specifically to tolerate masked faces, and green rectangles are periocular recognition models.

3.2.2. Enhancing masked face recognition

As validated by the studies discussed above, wearing a face mask does significantly affect the performance of face recognition technology. This by itself is intuitive, as the mask covers part of the facial information that face recognition models can use to discriminate between individuals. However, the insights from the discussed studies also inspired many innovative solutions aiming at enhancing the performance of masked face recognition. In this work, we present an operational categorization of these solutions based on their conceptual modeling of the masked face recognition problem. These solutions can be categorized into four groups, (a) mask in-painting, (b) template unmasking, (c) model optimization, and (d) periocular recognition, and are presented in the following sections along with the main works that made significant contributions under each category. A graphical representation of these categories is also presented in Figure 4, where masked face probes are processed in four different processes to be compared to an unmasked face reference.

(A) *Mask in-painting.* Under this category, illustrated in the top of Figure 4, the main goal is to detect and in-paint the face area covered by the mask before processing the face with conventional face recognition models. Such a process will not necessarily add additional identity-specific information to the face, because such in-painting processes are trained to predict the occluded area details from the visible parts of the face, and thus they extract the initial identity information from the already visible parts of the face. However, such in-painting will transfer the image into a distribution (domain) that is more similar to what general-purpose face recognition models are trained for and bring it closer to the unmasked reference. Seen as a domain adaption process, this can have the potential in enhancing the performance of masked face recognition. The main advantage with in-painting based strategies is the possibility to maintain the use of well-performing general-purpose face recognition models. The main disadvantage is that the training of the generative in-painting process is commonly expensive in terms of the required training data and the training computational cost [86]. Such generative processes might also result in artifacts that are out of the normal face image distribution compared to the original masked faces themselves [87].

Although face in-painting is in general a well-studied field with the recent methods producing photo

realistic images [88, 89, 90], using this technique to enhance masked face recognition is still under-explored. Jiang et al. [91] recently addressed the specific issue of in-painting face mask areas without evaluating the effect on face recognition performance. Such aesthetic-driven face in-painting of the mask area, i.e. mask removal, is discussed more in details in Section 3.9. Similar in-painting approaches have been shown before to be beneficial, to some degree, in enhancing the recognition performance [92] of occluded faces.

(B) *Template unmasking*. Under this category, illustrated in the second row in Figure 4, the main goal is to transfer the extracted masked face template into a form where it behaves similarly to a template extracted from an unmasked face of the same identity. Here, both the masked probe and the unmasked reference are processed by a general-purpose face recognition model, however, the masked face template typically undergoes another processing step. Just like with in-painting, this process will not add identity information to the masked face template, but it rather will remove the template artifacts introduced by the mask information. The main advantage of such solutions is that it maintains the use of the well-performing general-purpose face recognition models and that the overhead computational cost of the template unmasking model is relatively negligible [93] when compared to the face recognition model itself or the generative in-painting model discussed under the first category.

Despite the clear operational benefits of this category of solutions, relatively few works targeted such a concept. The first to do so was Boutros et al. [93] that proposed to train a template unmasking model on top of any general-purpose face recognition model to transfer the masked face template to a form that behaves similarly to an unmasked face template of the same identity in the comparison operations. Based on the fact that the genuine comparisons are significantly more affected than the imposter ones when comparing masked to unmasked faces, the authors proposed the self-restrained triplet loss that assigns higher importance to positive pairs during training when the negatives pairs are deemed relatively distanced enough. Following a similar operational concept, a recent study also proposed to process the masked face template in a framework that utilizes contrastive learning [94].

(C) *Model optimization*. Under this category, illustrated in the third row in Figure 4, the main goal is to train a face recognition model that can produce comparable embeddings for both masked and unmasked faces. Understandably, training such a solution requires having masked and unmasked face samples in the training data. This also requires, in part, general-purpose face recognition training goals such as direct embedding learning [95] or embedding learning through classification [96, 97]. The main advantage of this category of solutions is that both the masked and unmasked faces are processed with the same model. However, such solutions induce the need for a tedious training process and considerable amounts of training data that also needs to include masked faces. Additionally, including masked faces in the training process might render the resulting model less accurate when comparing pairs of unmasked faces [98]. This shortfall was recently targeted in the literature with a high degree of success [99], where the authors forced the face recognition model to produce optimal templates for both, masked and unmasked faces by incorporating a template-level knowledge distillation loss between the trained network and a general-purpose face recognition network.

Most of the works addressing masked face recognition so far fall under this category. The authors in [12] combined the ArcFace loss [97] with a mask-usage classification loss and noted it as Multi-Task ArcFace [12]. Other work combined the traditional triplet loss and the mean squared error in an effort to improve the face recognition robustness to masks [13]. The authors in [98] theorized that the masked face recognition process requires a larger penalty margin when using the cosine loss. Others proposed improving the face template consistency using a pairwise loss [100]. Geng et al. [101] proposed to enhance masked face recognition performance through mask-like generative augmentation. Hsu et al. [102] experimented with different loss functions to determine their suitability for masked face recognition. A hybrid backbone of residual block and self-attention components was proposed by [103], an aspect that was also investigated in [104].

(D) *Periocular recognition*. Under this category, illustrated at the bottom of Figure 4, the main goal is to simply reduce the face recognition problem to be a partial face recognition problem. This assumes that the mask commonly covers the lower part of the face and maintains the visibility of the upper part of the face. This area that includes the eyes and the adjacent regions is commonly called the (peri)ocular region [105].

The biometric literature refers to the recognition of this area, when the iris is not exclusively targeted, as periocular recognition [106]. Periocular recognition can include the periocular region of one eye for some applications [107, 108]. However, in the masked face recognition scenario, both right and left periocular regions are typically considered.

330 A number of works proposed to crop the masked face and focus the recognition task on the periocular region when the mask is present [109, 110, 111]. The need to use the periocular region for recognition purposes when faces are masked was extensively studied in [112], including a detailed survey on periocular recognition technologies.

3.2.3. Masked face recognition competitions

335 Two major competitions were organized in an effort to attract novel solutions for masked face recognition. The first was the MFR2021 Masked Face Recognition Competition [113] organized as part of the International IEEE Joint Conference on Biometrics (IJCB) 2021 [114]. The competition examined the deployability of the solutions by considering the compactness of the face recognition models. A private dataset was used for evaluation. The dataset contained real masked faces and represented a collaborative capture scenario. 340 Out of 18 submitted solutions, 10 were able to outperform the widely used ResNet-100 baseline [63] trained using the ArcFace loss [97]. Most of the competition entries used synthetic or/and real masked face images in the training of their solutions.

The second competition was the Masked Face Recognition Challenge [34] organized within the Face Biometrics Under COVID? Masked Face Recognition (MFR) Workshop, one of the IEEE/CVF International Conference on Computer Vision Workshops [115]. The competition included three test sets and used an online model testing system and provided a detailed evaluation of the submitted face recognition models. 345 The results of the competition pointed to the effectiveness of augmentation strategies simulating facial masks when training recognition models for the targeted task of masked face recognition.

3.2.4. Masks and face recognition subsystems

350 *Presentation attack detection.* Presentation attacks on face recognition systems involve the presentation of an artifact or of human characteristics to a biometric capture subsystem in a fashion intended to interfere with system policy, as defined in ISO/IEC 30107-3 [116]. This can include attacks like face morphing [117, 118], makeup attacks [119], or even identification circumvention attacks [120]. However, given the attack scenarios, the attack that is most related to wearing masks is the spoofing attack, where an attacker presents an artifact 355 to a biometric capture subsystem with the aim of impersonating a different identity [116]. Presentation attack detection solutions (PAD) aim at differentiating between non-attack samples, i.e. bona fide, and spoofing presentation attacks [121]. Such solutions can be based on user challenge (user performing a specific task/move), on special sensor characteristics, e.g. light field camera or thermal sensor, or on software solutions [121]. The most widely spread software solutions depend on analyzing samples captured in the 360 visible domain given the high deployability of visible-spectrum cameras in personal devices. Such solutions can be texture-based [122], motion-based [123], frequency-based [124], or a combination of two or more of these technologies [125].

Wearing a face mask changes the nature of the sample processed by the face PAD algorithm. This was apparent in the wide-spread reporting of malfunctioning face logins into personal devices at the start of the COVID-19 pandemic, not only because of failing to match to the unmasked reference image, but also 365 because the masked face is seen as a spoofing attack by the PAD algorithm. This interesting fact was revealed by an extensive study presented by Fang et al. [126] where the authors collected a set of unmasked and masked bona fide and attack samples and tested both the vulnerability of face recognition to such attacks and the performance of established PAD algorithms when processing masked attacks. The study 370 also presented a novel kind of attack where the attacks, printed or shown on a screen, were covered with a real mask. The main findings of the study pointed out that PAD algorithms classify many of the masked bona fide samples as attacks. The study also found that face recognition algorithms are still vulnerable to masked face attacks, especially when a real face mask is placed on the attacks [126]. An effort to reduce this effect on PAD performance was successfully presented in [127] where the authors propose to train the 375 PAD using partial pixel-wise labels, where the real masks placed on the attacks are considered to be a bona

vide area in an attack sample. This was also supported by giving the non-covered parts of the face a higher influence in the PAD decision inference from the image, bringing the PAD behavior on masked faces closer to that of the unmasked faces [127]. Further efforts are required though to build publicly available masked face attack databases and mask-invariant masked face PAD algorithms.

380 *Quality assessment.* Face image quality (FIQ) measures the utility of an image to face recognition algo-
rithms [128, 129]. This utility is measured with an FIQ score as defined in ISO/IEC 2382-37 [130]. Various
methods have been developed for face image quality assessment (FIQA) whether by building quality pseudo
385 labels and learning to predict such labels [131, 132], by measuring different aspects of face recognition model
response to the investigated image [133, 134], or learning to predict the relative classifiability of a face by
385 predicting its class-relative placement in a face recognition training process [135]. As FIQA measures the
utility to face recognition algorithms, it does not necessarily reflect the perceived image quality (IQ) mea-
sured by conventional general image quality assessment (IQA) solutions [136, 137]. However, IQ measures
have been found to correlate to the face image utility, though to a much lower degree than FIQ [136]. As
390 mentioned earlier, wearing a mask does lower the accuracy of face recognition, and thus it is expected also to
be reflected in a lower FIQ. This issue was investigated by Fu et al. [138] where it was shown that even when
the perceptual quality and capture environment do not change, the FIQ drops substantially when a mask
is worn. This consistently correlates with the drop in face recognition performance, whether by machine or
human experts [138]. Additionally, the networks performing FIQA did shift their attention away from the
395 mask region and more towards the visible face region, more specifically the ocular region, as demonstrated
in [138].

3.3. Facial Expression Recognition

Another important application affected by the presence of face masks is Facial Expression Recogni-
tion (FER). FER is a longstanding computer vision problem, where the goal is to recognize specific facial
expression or emotional states based on changes in facial appearance. It is generally acknowledged that
400 different parts of the face are involved when expressing different expressions, as evidenced, for example, by
the facial action unit coding (FACS) system, one of the most widely conceptual frameworks to the FER
problem [139, 140, 141, 142]. Thus, occlusions of these areas, as caused, for example, by facial masks lead
to obvious performance degradations.

The problem of facial expression recognition under the presence of face masks was explored by Abate
405 et al. in [41]. Here, the authors studied class activation maps (CAM) for different expressions and found
that anger, happiness, sadness, and neutral expressions are most heavily represented around the nose and
mouth areas. As results of this observation the authors concluded that FER models struggle to extract
informative features from the face images when face masks are present. To address this problem two common
mitigation strategies were proposed in the literature, i.e.: (a) collecting/generating masked datasets with
410 facial expressions that can be used for fine-tuning of existing FER models, and (b) designing new models
capable of performing facial expression recognition despite the presence of masks.

(A) *Datasets and fine-tuning.* Collecting and labeling facial expressions is a difficult, time- and labor-
intensive task that might also be subjective. The difficulty of labeling facial expressions carries over to the
problem of Masked Facial Expression Recognition as well. Since there are no datasets publicly available for
415 this task, generally, simulated masks are utilized in the literature. Yang et al. [40], for example, developed a
mask simulation method that uses facial landmarks and their orientations to fit a mask. They also annotated
13000 images from the Labeled Faces in the Wild (LFW) dataset [143] for facial expression recognition and
compiled a new dataset, called LFW-FER. Finally, using the mask simulation methodology on the LFW-
FER dataset, they generated a synthetic dataset for FER containing simulated mask, called M-LFW-FER
420 that is publicly available for research purposes and can be used to fine tune FER models for expression
recognition under the presence of facial masks.

Similar ideas were also pursued by other works. Barros et al. [144], for example, first detected the
facial landmarks on images from the AffectNet dataset [145] and fit a mask to the faces covering all the
landmarks below the nose. Using the resulting MaskedAffectNet dataset, the authors then applied different

425 training strategies, e.g., transfer learning, to their FaceChannel model [146] to account for the presence of
face masks. When the authors trained the FER model from scratch using the MaskedAffectNet dataset,
the model performance drastically deteriorated for unmasked applications. However, when the model first
pretrained on a standard dataset and later fine-tuned, the FER performance was affected only so slightly,
making it useful for both masked and unmasked facial images.

430 (B) *Mask-agnostic FER.* Techniques from the second group aim to design models that are robust (agnostic)
with respect to the presence of facial masks and perform similarly for masked and unmasked faces. Yang et
al. [147] developed a new approach for masked face expression recognition along these lines. Their model
consists of two parts. The first part includes a classifier for masked and unmasked recognition that generates
435 a binary attention heatmap for the face masks. The second part of the model takes the binary attention
heatmaps and convolutional face features to classify the facial expression. The authors show that their
model outperforms other state-of-the-art occlusions robust facial expression recognition models, like region
attention network (RAN) [148] and CNN with attention mechanism (ACNN) [149].

3.4. Masked Face Age Classification

Similar to face and facial expression recognition systems, age estimation techniques also critically depend
440 on the visibility of the facial areas and struggle with performance when parts are occluded. As a result,
studies investigating age estimation with facial masks have also appeared during the COVID-19 pandemic.

Golwalker et al. [39] conjectured that using large prediction models in age estimation with occluded
faces makes it challenging due to the lack of suitable large-scale datasets. When wearing masks, the most
discriminative features for age estimation are largely hidden below the masks, like wrinkles on the cheeks
445 and mouth. Their approach to this problem was, therefore, using a shallow model that could be fine-tuned
easily based on a small set of images of people wearing masks. To this end, the authors used a simple
9-layer CNN architecture. For the age detection dataset, they collected faces wearing masks from various
age categories and augmented it with an auxiliary dataset of 4500 synthetic images of masked people using
a Generative Adversarial Network (GAN) [150]. Öztel et al. [38] developed a two-stage pipeline consisting of
450 a face mask detection and an age classification stage. With this approach, the authors first determine if the
person is wearing a face mask or not. Then, depending on the result, two separate age classification models
are utilized. If the person is not wearing a mask, a standard classification model in the form of a simple
CNN trained on UTKFace Large Scale Face Dataset [151] is used. If the person is wearing a face mask,
another simple CNN model is utilized, but this time trained with simulated face masks on the UTKFace
455 Large Scale Face Dataset [151]. The proposed pipeline included three different age classes, i.e., teenager
(12-20), middle-aged (21-64), and elderly (65+), and was shown to ensure competitive results.

3.5. Landmark Localization and Alignment

Facial landmark localization and alignment are essential components of various face-related applications,
such as face recognition, facial pose estimation, 3D face reconstruction, emotion recognition, face synthesis,
460 and face morphing, and falls into the category of supporting techniques given our taxonomy from Section 2.
The main goal of landmark localization is to locate key points of the given 2D face image, such as the nose
tip, eyebrow curve, mouth corners, eye centers, or eye corners among others. Until the pandemic, many
successful facial landmark localization approaches have been developed by using thousands of annotated
face images [152, 58, 153]. However, the widespread usage of face masks to prevent virus transmission
465 has brought new challenges to landmark localization and alignment similarly as to many other face-based
algorithms. As it was not possible to collect and label a new dataset for the task, most methods prefer to use
existing datasets, especially the JD-landmark dataset [154] and place virtual facial masks on the face images.
This is because labeling facial landmark points on images with facial masks is hard, especially for 68 or 106
points of landmarks, which is the most commonly used markup in the literature. The authors of [47] propose
470 MaskFan, which is a lightweight convolutional neural network that uses depthwise separable convolutions
and group operations. They also propose a novel loss function named Enhanced Wing loss, which gives
less importance to errors made near facial masks. Facial landmark localization methods generally adopt L1



Figure 5: Generic images of crowded street before and during the pandemic, respectively. (© JordyMeow and Benzoyl)

Table 1: Summary of the reviewed crowd counting methods

Method	Year	Handling of facial masks	Handling of social distancing	Mode of counting implementation
Al-Sa'd et al. [156]	2022	No	Yes	Detection based CNN
Valencia et al. [157]	2021	No	Yes	Detection based CNN
Somaldo et al. [158]	2020	No	Yes	Detection based CNN
Nguyen et al. [159]	2021	No	No	Regression based CNN
Almalki et al. [160]	2021	Yes	No	Detection based CNN
He et al. [161]	2022	No	No	Attention based CNN
Dosi et al. [162]	2021	No	No	Attention based CNN
Alvarez et al. [163]	2021	No	Yes	Detection based CNN
Jarraya et al. [164]	2021	No	No	Density based CNN
Amin et al. [165]	2021	Yes	Yes	Detection based CNN
Nguyen et al. [6]	2021	Yes	No	Detection based CNN

or L2 loss functions that focus on more considerable errors. Since predicting facial landmarks over facial masks is a hard task due to invisible parts of the face, applying L1 or L2 loss forces the model to pay more attention to large errors that most heavily impact performance. In [155], Wen et al. also propose a new architecture for the masked facial landmark localization problem. Their model consists of three different neural networks designed for: alignment, estimation, and refinement. They use downscaled face images in the alignment network and then align faces according to the reference pose. Then, the estimation network predicts 106 facial landmarks. Finally, their refinement model takes the non-masked region of the face, which is eyes and eyebrows and tries to generate more accurate predictions. Hu et al. [48] adopt multi-knowledge distillation and a pose-aware resampling strategy. They aim to increase data diversity by sampling images with different face poses.

All the presented works generate virtual masks and apply them on the JD-landmark dataset [154], while studies with real masks are still largely missing from the literature due to the obvious ground-truth issues associated with such work. Suggestions for evaluation strategies that only consider visible landmarks have also been made in prior publications [154].

3.6. Crowd detection and counting

One of the prevention measures to reduce the spread of the Coronavirus disease is physical/social distancing in public areas that can be monitored automatically using vision-based crowd counting techniques.

490 Such techniques are able to count or estimate the number of people in a given area from a single image or a
video acquired through surveillance cameras, CCTV or even drones. A plethora of research has been done
over the past years on the crowd counting problem [166, 167] dealing with challenges, such as mutual occlu-
sions, non-uniform people density, varying scale, perspective, illumination, weather conditions, crowd size,
and density, that can severely alter human appearance. In this section, we review crowd counting solutions
495 with the focus on approaches developed to assist policy measures against the COVID-19 pandemic.

Early crowd counting methods mainly rely on object detection with counting. Usually, these meth-
ods first extract image features, such as shapelets [168], Histograms of Oriented Gradients (HOGs) [169],
Haar wavelets [170] or other related descriptors from the image and then combine the computed represen-
tations with various classification methods, such as Support Vector Machines (SVMs) [169, 170], regression
500 forests [171] and alike, in order to detect people in images. These methods work well for detecting sparse
(masked) faces, but perform poorly on dense crowds where individual people are not clearly visible.

To alleviate the above-mentioned problems, some approaches rely on direct-count regression. Examples
of such methods in [172] use handcrafted features and some regression technique, such as linear regression, to
learn a mapping function between the features and the crowd count. These methods are able to accurately
505 estimate people counts even in the presence of occlusions and background clutter, but they ignore spatial
information. The solution to this problem is given by the density estimation based methods [173] that learn
a mapping between features in the local region and the corresponding object density maps, while integrating
over the densities to obtain crowd counts. The approaches of this type generally use dot-annotated images
for training that are transformed to density functions using kernel density estimation [174].

510 With the advent of deep learning, many crowd counting approaches have been proposed based on convo-
lutional neural networks [175]. Amin et al. [165], for example, proposed a solution that addresses both face
mask detection and crowd counting. With this approach, the YOLO-based algorithm [176] is used to detect
face masks, while the MobileNet single shot object detector [177] is used simultaneously for crowd counting.
Kammoun-Jarraya et al. [164] introduced a CNN based technique for crowd counting from a single image
515 for enforcing social distancing during the COVID-19 pandemic. The proposed model follows the structure
of VGG-19 [178] with small kernel sizes in the convolutional layers but without the fully connected layers.
Due to the fully convolutional structure, the model is able to process input images of arbitrary resolutions.
The reported results on a new large-scale crowd counting dataset from the Saudi public areas point to the
competitive performance compared to the state-of-art methods. Alvarez et al. [163] developed a software
520 to monitor the physical distance and the crowd density of a specified area or a region of interest. The
YOLOv3 model [59] was used in this work to detect humans in each video frame captured by a mobile
phone. Physical distancing is monitored through computing the interpersonal distance of a pair of centroids
of the detected bounding boxes, while the crowd density is computed by counting the number of people
present in the region of interest. The software is able to detect 83% of physical distancing and 84% of crowd
525 density violations. Dosi et al. [162] proposed a pipeline named Attentive EfficientNet (AECNet) for density
estimation in crowd counting that makes use of an encoder-decoder-based architecture. In the encoder block,
they use EfficientNet [70] and empirically show their superiority over other feature extraction architectures.

To mitigate the problem of huge scale variations, He et al. [161] proposed a novel approach for crowd
counting named Jointly Attention Network (JANet). They designed the Multi-order Scale Attention module
530 to extract meaningful high-order statistics with abundant scale details and also introduced the Multi-pooling
Relational Channel Attention module to investigate the global scope relations and structural semantics.
Various experiments illustrated the superiority of the JANet approach. Almalki et al. [160] introduced an
approach that detects, counts, and classifies the crowd’s masking condition and calculates spatiotemporal
safety index that can be used for assisting effective policy decisions and relief plans against COVID-19. The
535 approach uses YOLOv3 [59] to extract image features and a classification layer is added at the end of the
YOLOv3 extractor that classifies a detected face to either the mask or the no-mask group. A unified system
that allows the scale variation problem to be solved both directly and indirectly was described by Nguyen
et al. in [159]. Here, the dense scale information is learned directly through the main network, which is
designed with dense dilated convolution blocks and dense residual connections among the blocks. The scale
540 information is further incorporated into the features indirectly through learning depth information from an
auxiliary depth dataset.

Somaldo et al. [158] proposed a drone that has the ability of localization, navigation, people detection, crowd identifier, and social distancing warning. For this purpose they utilize YOLOv3 [59] to detect people and also define an adaptive social distancing detector. Valencia et al. [157] presented a desktop application that utilizes YOLOv4-tiny⁴ and the DeepSORT tracking algorithm [179] to monitor crowd counts and social distancing from a top-view camera perspective. A privacy-preserving adaptive social distance estimation and crowd monitoring solution for surveillance cameras was proposed by Al-Sa'd et al. [156]. The authors utilize OpenPose [180] to detect and localize people. Their approach is able to compute inter-personal distances in real-world coordinates, detect social distance infractions and identify overcrowded regions in a scene. The work presented in [6] investigated the effectiveness of different approaches to estimate the ratio of people wearing a mask within an observed crowd - a problem referred to by the authors as *mask-wearing ratio estimation*. Specifically, the authors compared detection-based and regression-based approaches to crowd counting, while also distinguishing between people with and without masks in the given crowd. Moreover, the authors improved the state-of-the-art face detector, RetinaFace [58], to be able to better estimate the mask-wearing ratio. A large-scale dataset with more than 580,000 face annotations was also introduced to facilitate the experiments.

3.7. Breathing rate measurements

Clinical studies on patients with COVID-19 disease showed that one of the most common symptoms are fever, respiratory and digestive symptoms [181]. In order to identify breathing abnormalities, which can be a symptom of COVID-19, multiple research works suggested measuring respiratory rate using wearable devices [182], non-contact radar signals [183] and thermal cameras [29].

Among these breathing rate measurement techniques, detecting breathing anomalies using thermal cameras represents a cheap and effective solution that can easily be implemented in practice, as many countries already deployed thermal cameras to detect people with high fever at airports and public buildings [184]. Following this line of research, Queiroz et al. [29] proposed to analyze the intensity of thermal images over time using deep learning techniques. The approach exploits the fact that the region covered by facial masks gets warmer when exhaling, which can be detected through the analysis of the pixel intensities in the thermal image. Similarly, when the pixel intensity within the mask region decreases, this indicates that the person is inhaling. To facilitate the research, the authors collected 33 videos of 11 subjects, with subjects breathing slowly, normal and fast. Their experimental results showed that breathing rate measurements can reach an accuracy of up to 91% on their dataset.

3.8. Face-hand interaction detection

To minimize the transmission of COVID-19, common advice issued by health organizations included limiting face-hand interaction. CVHA techniques were also developed to help monitor face-hand interaction in public spaces. A basic representation of the face-hand interaction detection task is presented in Figure 6.

One of the initial studies by Beyan et al. [31] investigated the face-hand touching behavior of people. The authors first manually annotated 64 video recordings, originally collected for the analysis of social interactions within a small group of people, for face-hand touching interaction. Next, they evaluated rule-based, hand-crafted feature-based, and learned CNN feature-based models for their performance in face-hand touching detection and found that the CNN model yielded the best overall results with an F1-score of 83.76%. In a more recent study, Eyiokur et al. [30] explored the applicability of several well-known CNN models, such as ResNet [63] and EfficientNet [70], for face-hand interaction detection. Here, the authors first introduced an unconstrained face-hand interaction dataset, named ISL-Unconstrained Face Hand Interaction Dataset (ISL-UFHD), to advance detection of face-hand interaction detection within a comprehensive prevention system for COVID-19 protection measurements, and then evaluated the considered classification models on the newly collected data. Experimental results showed that the highest classification accuracy of 93.35% was obtained with the EfficientNet-b2 model [70].

⁴<https://github.com/AlexeyAB/darknet>

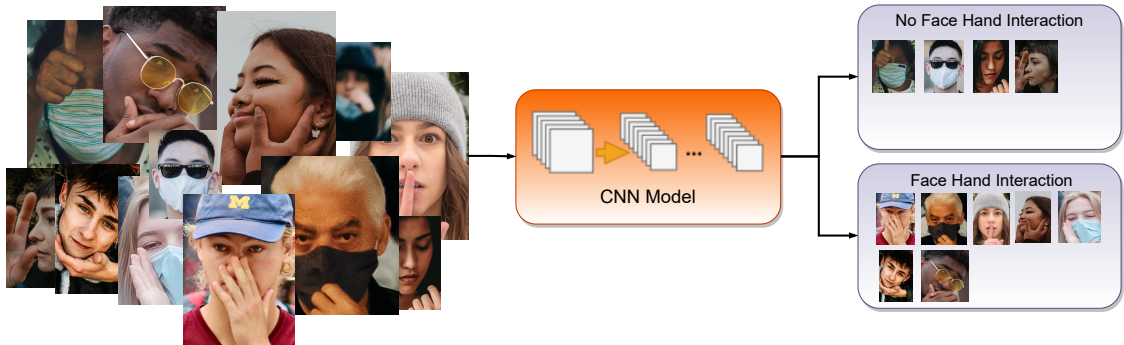


Figure 6: Illustration of the face-hand interaction detection task. One of the most common advice given to prevent virus transmission is not to touch the face with hands. During the COVID-19 pandemic, automatic detection of face-hand interaction has gained importance as a research topic in CVHA.

Both of the studies discussed above, proposed CVHA techniques that showed promise for the face-hand interaction problem. However, important challenges, such as the detection in extreme imaging conditions and under varying poses, or in the presence of ambiguity caused by the different depth levels of the face and hand, still persist. Face-hand interaction detection is, therefore, still considered an open research problem that requires further investigation.

3.9. Synthetic data generation and mask removal

One of the main challenges in CVHA at the beginning of the COVID-19 was the obvious lack of suitable datasets needed to train various CVHA techniques. In response to this challenge, generative approaches have been quickly adopted to build synthetic datasets to alleviate the need of collecting real-life masked face images as well as to develop methods based on data augmentation and generation for various tasks such as face recognition, identification, and landmark detection.

To artificially generate face images with masks, Anwar et al. [33] developed an open-source tool, MaskTheFace⁵, that can convert non-masked faces to masked faces effectively. The tool uses the Dlib-based face landmark detector [152] to identify the face tilt and six key features, i.e. landmarks, on the face to properly fit a face mask. Alternatively, Wang et al. [42] presented another open-source toolbox, FaceX-Zoo⁶, which implements a Facial Mask Adding (FMA-3D)⁷ method for adding a mask to a non-masked face image. Given a real masked face image I and a non-masked face image J , this method synthesizes a photo-realistic masked face image with the mask region coming from I and the facial area coming from J .

Encouraged by these initiatives, many studies have attempted to enrich existing datasets containing faces without masks, e.g. CelebA [185], CASIA-WebFace, LFW [143], CALFW [186], with synthetically generated masked face images to enable further research on masked face recognition [43, 187, 188, 154, 189, 190]. For instance, Wang et al. [43] and Karasugi et al. [187] generated synthetic face mask datasets using Dlib’s landmark detector [152] to properly align face mask templates on faces, whereas Mare et al. [188] relied on SparkAR Studio⁸, a developer program made by Facebook to create Instagram face filters, to create synthetic masks and overlay them on faces in the original images. On the other hand, Xiang et al. [154] targeted to improve the accuracy and robustness of facial landmark localization on masked faces by introducing a new dataset with generated masks that are largely varied in identity, head pose, facial expression, and occlusion based on the FMA-3D method in the FaceX-Zoo toolbox [42].

Some studies tackled the opposite problem and focused on removing the face masks from images [44, 45, 46, 25]. The idea with these studies is to bring the data closer to the real-world masked-free data, for

⁵<https://github.com/aeqelanwar/MaskTheFace>

⁶<https://github.com/JDAI-CV/FaceX-Zoo>

⁷https://github.com/JDAI-CV/FaceX-Zoo/tree/main/addition_module/face_mask_adding/FMA-3D

⁸<https://sparkar.facebook.com/ar-studio/>

which standard off-the shelf CVHA models for different tasks are readily available. For instance, Din et al. [44] investigated a two-stage method for unmasking of masked faces where the first stage detects and segments masks with a modified version of U-Net [191] and the second stage deploys a GAN-based network with global and local discriminators for mask-area inpainting. Similarly, Li et al. [45] proposed a method combining a GAN and a texture network to first inpaint the face after removing the mask and then to smooth out the texture to make the resulting face more realistic. Taking a step further, Coelho et al. [46] presented a generative approach for face mask removal using audio and appearance together. This approach estimated landmarks representing mouth structure from the audio, and feed these landmarks into a GAN to reconstruct the full face image with the mouth in a correct shape. Hu et al. [25] described a method to generate faces with properly worn masks, either by simply overlaying the mask if no mask is worn or by first removing and then overlaying the mask if the mask is incorrectly worn. The method employs Mean and Covariance Feature Matching GAN (MCGAN) [192] for the mask removal task and uses MaskTheFace [33].

4. Datasets

Table 2: Summary of COVID-19 related datasets reviewed in this paper. BB: Face bounding box, FM: Face Mask, FP: Frontal to Profile, MF: Masked face, FO:Face Occlusion

Dataset name	Dataset source & availability	Mask Types	Number of images	Head pose	Variation of subjects	Number of face mask classes	Purpose of data collection	Annotation type
MAFA [56]	Internet Publicly available	Real	30,811	Various	Medium Diversity Mostly Asian People	Multiple	MF detection	BB coordinates FM & occlusion classes
MAFA-FMD [193]	MAFA Not publicly available	Real	56,084	Various	Medium Diversity Mostly Asian People	3	FM detection	BB coordinates FM classes
FMLD [4]	MAFA,Wider Face Publicly available	Real	63,072	Various	High Diversity	3	MF detection	BB coordinates FM & occlusion classes
WearMask [194]	MAFA,Wider Face,Internet Publicly available	Real	9,097	Various	High Diversity	2	MF detection	BB coordinates FM classes
PWMFD [71]	MAFA,Wider Face,Internet Publicly available	Real	9,205	Various	High Diversity	3	MF detection	BB coordinates FM classes
ISL-UFMD [30]	Internet,CelebAq, FFHQ Publicly available	Real	20,891	Various	High Diversity	3	FM detection	FM classes
SMFRD [43]	Face recog. Datasets Publicly available	Artificial	500,000	Various	Various	2	MF recognition	FM classes Subject ids
MaskedFace-Net CMFD,IMFD [51]	FFHQ Publicly available	Artificial	133,783	Mostly Frontal	High Diversity	2	FM detection	FM classes Incorrect FM sub-classes
MS1MV2-Masked [93]	MS1MV2 Publicly available	Artificial	57.5m	Various	High Diversity	6	MF recognition	FM classes Subject ids
RMFRD [43]	Internet Publicly available	Real	92,671	FP	Low Diversity Asian People	2	MF recognition	FM classes Subject ids
Damer's MFRD [32]	Collected by researchers Not publicly available	Real	2,160	FP	Low Diversity	2	MF recognition	Subject ids
Damer's MFRD-Extended [82]	Collected by researchers Not publicly available	Real, Artificial	8,640	FP	Low Diversity	2	MF recognition	Subject ids
MFR2 [33]	Internet Publicly available	Real	269	FP	High Diversity	Multiple	MF recognition	FM classes Subject ids
FaceMask [195]	Internet Publicly available	Real	4,866	Various	Various	2	MF detection	BB coordinates FM classes
DS-IMF [196]	Collected by researchers Not publicly available	Real	3600	Frontal	Low Diversity Indian People	2	MF detection & recognition	BB coordinates Subject ids
Thermal-Mask [29]	SpeakingFaces Not publicly available	Artificial, Thermal	75,908	FP	Low Diversity	2	Thermal MF detection Breathing rate measurement	BB coordinates FM classes
Thermal mask dataset [197]	Collected by researchers Not publicly available	Real, Thermal	7,920	Various	Low Diversity	2	Thermal MF detection	BB coordinates FM classes
COVID-19 TFCD [198]	Collected by researchers Publicly available	Real, Thermal	261	FP	Low Diversity	2	Thermal MF detection	BB coordinates FM classes
Kaggle853 [199]	Internet Publicly available	Real	853	Various	Medium Diversity Mostly Asian	3	MF detection	BB coordinates FM classes
Kaggle12k [200]	Internet Publicly available	Real	12,000	FP	High Diversity	2	FM detection	FM classes
Kaggle FMLD [201]	Style Gan-2 Generated Publicly available	Artificial	20,000	Frontal	High Diversity	2	Not clarified	FM classes
WWMR-DB [202]	Collected by researchers Publicly available	Real	1222	FP	Low Diversity	Multiple	MF detection	BB coordinates FM classes
Medical Mask Dataset (MMD) [203]	Public domain Publicly available	Real	6,000	Various	High Diversity	3	MF detection	BB coordinates FM & occlusion classes
AIZOOTech face mask dataset [204]	MAFA,Wider Face Publicly available	Real	7,971	Various	High Diversity	2	MF detection	BB coordinates FM classes
BAFMD [72]	Twitter Publicly available	Real	6,264	Various	High Diversity	2	MF detection	FM classes FO classes
ROF [81]	Google Image Search Publicly available	Real	5,559	Various	High Diversity	3	MF recognition	Subject ids

An unprecedented number of datasets targeting various CVHA tasks related to COVID-19 have been introduced over the last few years. While masked face datasets were already collected back in 2017 [56], the need for suitable larger-scale collections of masked face images increased significantly during the COVID-19 pandemic. As a result, novel (real and simulated) datasets were introduced for masked face detection and



Figure 7: Illustrative example images from different datasets introduced for the development of CVHA techniques for the COVID-19 era.

635 recognition, face-hand interaction detection, (masked) crowd counting and as well as other related CVHA
problems. In Table 2, we summarize the main COVID-19 related datasets and compare their characteristics.

In [56], the first large-scale masked face dataset, named MAFA, was published. The dataset contains
30,811 images of multiple persons with various head poses, face occlusions, and ethnicity, collected from the
Internet. The MAFA dataset includes 35,806 masked face crops (there are multiple faces per image) with six
640 annotated attributes: face, eye, and mask coordinates, head pose, occlusion degree, and four different mask
types. The dataset is primarily intended for the development of face/mask detection models. However, it
needs to be noted that some of the masks present in the data are not worn correctly, e.g., they are not
covering the nose, so mask detection models developed on this dataset are generally considered less suitable
for monitoring applications aimed at preventing the spread of the COVID-19 disease. To address this issue,
645 some studies considered improper mask usage as an additional label for the facial images, i.e., next to the
mask and *no mask* labels. Such data labels helps conceive more appropriate systems with respect to health-
protective rules and usability in real-world conditions. In [1], Batagelj et al. pointed out that the actual
annotations of MAFA [56] are not suitable for training useful detectors to distinguish between correctly and
incorrectly worn masks. The authors, therefore, reannotated the MAFA images based on health-protective
650 rules. In addition to MAFA [56], they also annotated the Wider Face dataset and released the generated
annotations under the name Face Mask Label Dataset (FMLD). Thus, FMLD [1] includes images partitioned
into three groups: 29,532 images with correctly worn masks, 1,528 images with incorrectly worn masks, and
32,012 images with mask free faces. In addition to mask annotations, the FMLD also has bounding box
coordinates of faces, gender, ethnicity, and pose labels for each face. Wang et al. [194] utilized the same
655 benchmark datasets, Wider Face [53] & MAFA [56], to build a serverless edge face detection tool. Their
dataset included 4,065 images from MAFA, 3,894 images from Wider Face [53], and 1,138 additional images
from the Internet, for a total of 17,532 face crops with corresponding bounding boxes. In [71], a new dataset
called the Properly Wearing Masked Face Detection Dataset (PWFMD) is presented and consists of 3,615
newly collected images, 2,581 relabeled images from MAFA [56], 2,951 images from Wider Face [53], and
660 58 images from RMFRD [43]. Similar to the previous studies, Jiang et al. [71] considered three classes for
the labels of their dataset, i.e., correctly worn, incorrectly worn and mask-free. In total, there are 7,695
properly worn masked faces, 10,471 mask-free faces, and 366 incorrectly worn masked faces in PWFMD.
In [193], a face detector was first applied to the MAFA dataset [56], and the generated face crops were
then reannotated with respect to virus protection rules. This way, a new dataset, named MAFA-FMD,
665 was collected and includes 56,024 images belonging to correct, incorrect, and no mask-wearing classes.
Unfortunately, the MAFA-FMD dataset is not publicly available.

Eyiokur et al. [30] proposed an unconstrained masked face dataset ⁹, named ISL-Unconstrained Face
Mask Dataset (ISL-UFMD), to study CVHA techniques for COVID-19. ISL-UFMD [30] contains 11,075
mask-free, 9,300 proper, and 513 improper mask images, collected from the Internet, YouTube videos, and
670 well-known face datasets, such as CelebA-HQ [185] and LFW [143]. By relying on different resources during
data collection, the data in ISL-UFMD features highly diverse images captured in unconstrained conditions
with variability across ethnicity, age, gender, head pose, and environmental settings. Furthermore, Eyiokur
et al. [30] also presented the first unconstrained face-hand interaction dataset named ISL-Unconstrained Face
Hand Interaction Dataset (ISL-UFHD) to advance face-hand interaction detection with respect to COVID-
675 19 protection rules. In ISL-UFHD, there are 10,018 samples with face-hand interaction and 20,038 without.
Another related dataset, FaceMask, was described by Vrinks et al. in [195] and contains 4,866 images of
people with variations in gender and ethnicity, occlusions, and capture conditions, e.g., indoor/outdoor.
Some of the faces are blurred, have partial occlusions or are of low-resolution due to distance to the camera.
There are 15,419 and 12,262 face crops that belong to the *mask* and *no mask* classes, respectively. Moreover,
680 Kantarci et al. [72] proposed Bias Aware Face Mask Detection (BAFMD) dataset in order to create a
dataset that minimizes potential bias on ethnicity, age, and gender. Their dataset contains 6,264 images
from Twitter with more than 16,000 facial bounding boxes with and without facial masks.

The need for a large-scale masked face datasets motivated researchers to also generate simulated images
with artificial masks positioned on the face, as already discussed in Section 3.9. In [51], a large-scale

⁹<https://github.com/iremeiyokur/COVID-19-Preventions-Control-System>

685 simulated masked face dataset named MaskedFace-Net, which includes the Correctly worn Masked Face
Dataset (CMFD) and Incorrectly worn Masked Face Dataset (IMFD) subsets, was presented. MaskedFace-
Net was constructed from the Flickr-Faces-HQ3 (FFHQ) dataset [205] using a mask-to-face deformable
model and contains 137,016 images in total. In [99], the popular large-scale MS1MV2 dataset [206] with
5.8M images of 85k subjects was augmented with simulated face masks with a probability of 0.5. Similarly,
690 in [93], a face mask simulated version of the MS1MV2 dataset [206] was utilized for the training of the
presented masked face recognition system. To evaluate the proposed system, other well-known benchmarks
for face verification, namely IARPA Janus Benchmark-C (IJB-C) dataset [207] and LFW [143], were used
to generate face images with synthetic face masks. Moreover, in [43], three novel datasets, named Masked
Face Detection Dataset (MFDD), Real-world Masked Face Recognition Dataset (RMFRD), and Simulated
695 Masked Face Recognition Dataset (SMFRD) were published to investigate the face mask detection and face
recognition performance in the case of occlusion due to face masks. Wang et al. [43] proposed 500,000
simulated masked face images of 10,000 subjects constructed with an artificial mask generation tool.

In addition to the works that investigate prevention, monitoring and control CVHA techniques, e.g.,
for the detection and tracking of proper usage of face masks, some studies also examined the effect of
700 wearing face masks on the performance of face recognition systems. To facilitate this work, novel real-world
and simulated masked face recognition datasets were introduced. Wang et al. [43], for example, described
the Real-world Masked Face Recognition Dataset (RMFRD) which consists of 5,000 masked and 90,000
non-masked face images that belong to 525 celebrities. Although it is stated that the RMFRD dataset
contains 5,000 face images with masks, there are only 2,203 face images with masks in the publicly available
705 version. Damer et al. [32] presented a database that consists of 2,160 images of 24 participants from three
different sessions. For each session, three videos are collected from the participants; two of them containing
faces with and without a face mask in daylight and the third one containing faces with masks in different
lighting condition. Session one was considered as a reference, and sessions two and three were considered
as sources for the probe data. In their follow-up work [82], the same authors extended the initial dataset
710 with an additional 24 participants and a new type of face images with simulated masks. In [33], the authors
published a relatively small dataset, called Masked Faces in Real World for Face Recognition (MFR2), that
contains 53 identities with an average of five images, with or without face mask, per subject.

In [196], Mishra et al. focused on analyzing masked face detection, gender prediction, mask/no mask
classification, and masked face recognition on images that were acquired from Indian subjects. They in-
715 troduced the Dual Sensor Indian Masked Face Dataset (DS-IMF), which consists of 300 subjects with 300
mask-free and 1,500 mask images per class. The images were captured with a DSLR camera and a mobile
phone. Moreover, Fang et al. [126] presented the novel Collaborative Real Mask Attack Database (CRMA)
to investigate the effect of face masks on presentation attack detection. The CRMA dataset consists of
30% AM0 (unmasked face PA), 60% AM1 (masked face PA), and 10% AM2 (unmasked face PA with a
720 real masked placed on the PA) which are images of three different presentation attacks for analyzing both
print and replay attacks. In [81], a new real-world dataset named Real World Occluded Faces (ROF) with
3,195 neutral images, 1,686 sunglasses images (upper-face occluded) and 678 masked images (lower-face
occluded) is presented. In ROF dataset, collected images belong to 180 different identities and they are used
to explore effect of occlusions on face recognition performance.

725 Another notable group of works [29, 197, 198] focused on the collection of datasets for COVID-19 re-
lated applications using thermal imaging. Queiroz et al. [29], for example, utilized a large-scale multimodal
dataset known as SpeakingFaces [208] that consists of thermal images as well as visual and audio streams.
The original SpeakingFaces dataset [208] does not include faces with masks. Queiroz et al. [29], therefore,
generated thermal masked faces using artificial masks placed over the mouth and nose area. After prepro-
730 cessing, 42,460 thermal masked and 33,448 thermal mask-free faces were included in the final Thermal-Mask
Dataset (TMD) [29]. In [197], Glowacka et al. collected 7,920 thermal images with four different cameras
from various distances and subjects with and without facial masks. The captured dataset [197] in the final
form includes 10,555 faces, as some of the recorded images include multiple people. In [198], a small thermal
mask dataset (COVID-19 TFCD), with 250 images belonging to 20 participants, was collected.

735 There are many online dataset repositories such as Kaggle, IEEE DataPort, and Github that allow
researchers to publish their data collections. With growing interest in CVHA problems during the COVID-

19 times, several face masked datasets [199, 200, 201, 202, 203, 204] were published on these repositories. In [199], there are 853 images of 4,080 faces belonging to three face classes (present/not present/improperly worn). In [200], around 12k face images belonging to two main classes: mask, no mask, were published. 740 The dataset has variations in terms of resolution, mask type, diverse people. Another dataset named Face Mask Lite Dataset (Kaggle-FMLD) [201] on Kaggle contains 10,000 artificial face images generated using StyleGAN2 architecture. By adding artificial masks to the generated faces, the authors created a simulated dataset to address the masked face recognition problem. The Ways to Wear a Mask or a Respirator (WWMR-DB) dataset [202], published on the IEEE DataPort, consists of 1,222 images of a small number of people 745 with eight different mask usage annotations. The Medical Mask Dataset (MMD) [203] and AIZOOTech dataset [204] are publicly available masked face detection datasets and annotation efforts published by private companies.

5. Open Issues & Future Challenges

Significant progress has been made over the last couple of years to address the main challenges in CVHA techniques for the COVID-19 era, but several open issues still remain and need to be addressed in the future. 750 Below we provide a discussion of the most important topics in the opinion of the authors.

5.1. Self-adaptation of CVHA techniques

A desired ability of CVHA techniques is detecting new conditions and self-adaptation to these new conditions. For example, before the COVID-19 era it was not very common to wear masks. Therefore, most 755 of the pre-COVID-19 CVHA techniques were trained with datasets that do not at all contain or contain very few samples of people with masks. On account of this, the performance of many existing techniques deteriorated severely, once people started wearing face masks. As the changes in human appearance can occur over time due to several factors, e.g., due to fashion trends and health concerns, it is necessary and important to have CVHA techniques that adapt themselves to the current conditions. One way of performing 760 this is to benefit from online continual learning approaches [209, 210, 211, 212, 213, 214] that have been utilized in computer vision research. Detecting unseen cases is critical, as sometimes changes in appearance might not be as sudden. The adaption of models to new concepts and conditions by learning with few data [215, 216, 217, 218] and using self-supervised features [219] are other essential points that need to be considered in future works.

5.2. Generalization and robustness

The current generation of CVHA techniques is already exhibiting remarkable performance across diverse data characteristics. However, in unconstrained scenarios, large appearance variability may still adversely affect their performance [18, 19]. This, for example, includes low-resolution masked inputs for tasks such as face landmarking, face detection and face recognition. Significant pose variations and additional occlusions, 770 e.g., due to glasses, hats and scarfs, also still have an adverse effect on performance, especially with masked facial images. In CVHA solutions involving crowds, novel ideas and powerful techniques are needed that can differentiate between masked and non-masked people in various environments and across a range of viewing angles [6]. Thus, there is an imminent need to further improve the generalization capabilities and most of all robustness of CVHA techniques aimed at combating COVID-19.

5.3. Availability of large-scale benchmarks

As discussed in Section 4, a considerable amount of datasets, especially for the masked faces, appeared in response to the needs induced by the COVID-19 pandemic. However, many of these datasets are small, not well curated and come without a well-defined experimental protocol and/or performance indicators. In CVHA problems, such as facial landmarking and related tasks, for example, ground truth information is 780 usually also not readily available. As a result, research often combine datasets for their experiments and define in-house protocols for experimentation. It is, therefore, difficult to objectively evaluate progress and assess the merits and deficits of the CVHA techniques being proposed in the literature. Well-designed large-scale benchmarks with clear objectives and properly designed experimental protocols are critically needed to help advance the field further and provide a solid basis for research going forward.

785 5.4. Bias and fairness

Data-driven techniques that learn from labeled examples are today the most widely utilized solutions for various computer vision tasks. When such techniques are applied in automated decision-making systems that impact people’s lives, fairness and bias become critically important. As automated decisions need to be fair and equally accurate for all, regardless of race, gender, age and other demographic factors, it is paramount that CVHA techniques ensure unbiased performance for subjects with diverse demographic attributes [220]. The negative consequences of biased systems have, for example, made headlines for face recognition, prompting many of the largest software corporations, such as Microsoft, Amazon, and IBM, to reconsider their face recognition programs and policies [221, 222]. While several studies explored bias with standard CVHA techniques [223, 224, 225, 226, 227], this issue has seen far less attention with masked face images and the data characteristics induced by COVID-19 [228]. Therefore, studies are needed that help to better understand the behavior of CVHA techniques in terms of bias and fairness with masked face images, as well as targeted mitigation techniques that contribute towards fairer decisions for different CVHA tasks. Furthermore, as many of the existing datasets gathered for COVID-19 related CVHA techniques are not balanced across demographic groups, additional efforts are also required on the data collection and curation side to facilitate research into these topics.

5.5. Ethics and privacy

The ability of processing face images behind masks and identifying people raises certain questions about the surveillance capabilities enabled by such technologies. As in all ethics issues, a sound analysis should balance the potential benefits against the potential risks, and arrive at guidelines and recommendations that will mitigate the risks, while maximizing the benefits. The main risks of facial surveillance involve loss of privacy, especially in cases where privacy matters. Government surveillance, in particular during events that criticize the said government, is a major case in point, and there is widespread worry that the deployment of facial surveillance can jeopardize people’s rights of expression, can lead to prosecutions and harm. This is also linked to legitimate uses of facial surveillance and analysis technology, such as for public transport payments or health related screening purposes, which can then be extended into citizen surveillance, i.e. the “slippery slope” argument. Religious freedoms, freedom of opinion and expression, freedom of assembly and association are all fundamental human rights, and need to be protected. During the *Umbrella Movement* in Hong Kong, the protesters have used masks and other props to cover their faces to prevent the police from identifying them and singling out protesters for arrests¹⁰. While there may be legal barriers for governments to target protesters, they can be sidestepped quickly. A Chinese based company, Hanwang, announced in 2020 that its facial recognition software was identifying people with masks with 95% accuracy, as opposed to 99.5% for people without masks¹¹. When asked the possibility of this software being used to identify protesters in Hong Kong, the company spokesperson said that this use case is known, but the market is too small. The company reports having about 200 clients in Beijing using the technology, including the police.

The second potential risk for enhanced facial surveillance capabilities (even with masked faces) is associated with data use by private companies. Since identification and personalization content for potential customers is key for new marketing approaches, identity can be monetized easily. Privacy breaches in this sector can have drastic consequences¹². While personal data, including biometric data, such as facial imagery, are regulated in certain parts of the world, e.g., see GDPR in Europe¹³, the Japanese Act on the Protection of Personal Information in Japan [229], or the California Consumer Privacy Act (CCPA) [230] and the Biometric Information Privacy Act (BIPA) [231] in the US, technological safeguards are also critically needed to address ethics and privacy concerns. Along these lines, biometric privacy-enhancing technologies designed specifically for masked faces and capable of hiding part of the information contained in the data may become more important going forward [222].

¹⁰<https://www.nytimes.com/2019/07/26/technology/hong-kong-protests-facial-recognition-surveillance.html>

¹¹Martin Pollard, “Even mask-wearers can be ID’d, China facial recognition firm says,” Reuters, 9 March 2020, retrieved from <https://reut.rs/2TAwMux>

¹²“Before Clearview Became a Police Tool, It Was a Secret Plaything of the Rich,” The New York Times, March 2020, <https://www.nytimes.com/2020/03/05/technology/clearview-investors.html>

¹³Accessible from: <https://gdpr-info.eu/>

830 6. Conclusion

It is generally expected that COVID-19 will not simply disappear, and will remain an issue for years to come. Consequently, novel computer vision techniques adapted to societal developments and behavioral changes of people induced by prevention measures and health-related governmental policies will increasingly be needed. A significant amount of work has already been done to help prevent and control the spread of the disease and facilitate normal operation of identity management schemes and other relevant infrastructure using vision-based methods. As discussed in the survey paper, a large part of this work focused on computer vision techniques for human analysis (CVHA), which analyze visual data related to faces and people during the COVID-19 era, e.g., in the presence of occlusions with face masks.

In this survey paper, we presented a comprehensive review of existing CVHA solutions for the COVID-19 era. Specifically, we discussed the main challenges introduced to CVHA problems by the pandemic, presented a high-level taxonomy of existing methods, elaborated on relevant datasets and described, what we feel, are the most important open issues that need to be addressed in the future. The consolidated information presented in the survey is expected to help researchers working on similar problems to quickly get an overview of the work already done and the main challenges that require further research.

845 Acknowledgements

This research was supported in parts by the ARRS Research Programme P2-0250 (B) “Metrology and Biometric Systems” and the additional funding provided for COVID-19 related research as well as the bilateral ARRS-TUBITAK funded project: Low Resolution Face Recognition (FaceLQ), with TUBITAK project number 120N011. This research work has been also partially funded by the German Federal Ministry of Education and Research and the Hessen State Ministry for Higher Education, Research and the Arts within their joint support of the National Research Center for Applied Cybersecurity ATHENE, and EU COST project GoodBrother (19121). The project on which this report is based was also partially funded by the Federal Ministry of Education and Research (BMBF) of Germany under the number 01IS18040A.

References

- 855 [1] B. Batagelj, P. Peer, V. Štruc, S. Dobrišek, How to correctly detect face-masks for covid-19 from visual information?, *Applied Sciences* 11 (5) (2021) 2070. [2](#), [5](#), [6](#), [17](#), [19](#)
- [2] E. P. Fischer, M. C. Fischer, D. Grass, I. Henrion, W. S. Warren, E. Westman, Low-cost measurement of face mask efficacy for filtering expelled droplets during speech, *Science Advances* 6 (36) (2020) eabd3083. [2](#), [5](#)
- [3] S. Feng, C. Shen, N. Xia, W. Song, M. Fan, B. J. Cowling, Rational use of face masks in the covid-19 pandemic, *The Lancet Respiratory Medicine* 8 (5) (2020) 434–436. [2](#), [5](#)
- 860 [4] A. Ulhaq, J. Born, A. Khan, D. P. S. Gomes, S. Chakraborty, M. Paul, Covid-19 control by computer vision approaches: A survey, *Ieee Access* 8 (2020) 179437–179456. [2](#), [3](#)
- [5] A. Bhargava, A. Bansal, Novel coronavirus (covid-19) diagnosis using computer vision and artificial intelligence techniques: a review, *Multimedia tools and applications* 80 (13) (2021) 19931–19946. [2](#)
- 865 [6] K.-D. Nguyen, H. H. Nguyen, T.-N. Le, J. Yamagishi, I. Echizen, Effectiveness of detection-based and regression-based approaches for estimating mask-wearing ratio, in: 2021 16th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2021), IEEE, 2021, pp. 1–8. [2](#), [13](#), [15](#), [21](#)
- [7] N. Petrović, Đ. Kocić, Smart technologies for covid-19 indoor monitoring, in: *Viruses, Bacteria and Fungi in the Built Environment*, 2022, pp. 251–272. [2](#)
- 870 [8] S. Hussain, Y. Yu, M. Ayoub, A. Khan, R. Rehman, J. A. Wahid, W. Hou, Iot and deep learning based approach for rapid screening and face mask detection for infection spread control of covid-19, *Applied Sciences* 11 (8) (2021) 3495. [2](#)
- [9] W. Tan, J. Liu, Application of face recognition in tracing covid-19 fever patients and close contacts, in: 19th IEEE International Conference on Machine Learning and Applications (ICMLA), 2020, pp. 1112–1116. [2](#)
- [10] W. Tan, J. Liu, Y. Zhuo, Q. Yao, X. Chen, W. Wang, R. Liu, Y. Fu, Fighting covid-19 with fever screening, face recognition and tracing, *Journal of Physics: Conference Series* 1634 (1) (2020) 012085. [2](#)
- 875 [11] M. Rezaei, M. Azarmi, Deepsocial: Social distancing monitoring and infection risk assessment in covid-19 pandemic, *Applied Sciences* 10 (21) (2020) 7514. [2](#), [4](#)
- [12] D. Montero, M. Nieto, P. Leskovský, N. Aginako, Boosting masked face recognition with multi-task arcface, *CoRR abs/2104.09874*. [2](#), [9](#)
- 880 [13] P. C. Neto, F. Boutros, J. R. Pinto, M. Saffari, N. Damer, A. F. Sequeira, J. S. Cardoso, My eyes are up here: Promoting focus on uncovered regions in masked face recognition, in: *BIOSIG*, Vol. P-315 of LNI, Gesellschaft für Informatik e.V., 2021, pp. 21–30. [2](#), [9](#)

- [14] J. T. Widjaja, Developing Trustworthy Covid-19 Computer Vision Systems, <https://towardsdatascience.com/developing-trustworthy-covid-19-computer-vision-systems-c862767d0d50>, accessed: 2022-08-18 (2021). 2
- 885 [15] A. M. Martinez, Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24 (6) (2002) 748–763. 2
- [16] V. Štruc, S. Dobrišek, N. Pavešić, Confidence weighted subspace projection techniques for robust face recognition in the presence of partial occlusions, in: 20th International Conference on Pattern Recognition (ICPR), 2010, pp. 1334–1338. 2
- [17] H. K. Ekenel, R. Stiefelwagen, Why is facial occlusion a challenging problem?, in: *International Conference on Biometrics*, Springer, 2009, pp. 299–308. 2, 7
- 890 [18] B. Wang, J. Zheng, C. Chen, A Survey on Masked Facial Detection Methods and Datasets for Fighting Against COVID-19, *IEEE Transactions on Artificial Intelligence* 1 (01) (2021) 1–1. 3, 21
- [19] A. Alzu'bi, F. Albalas, T. Al-Hadhrami, L. B. Younis, A. Bashayreh, Masked face recognition using deep learning: A review, *Electronics* 10 (21) (2021) 2666. 3, 21
- 895 [20] Y. Utomo, G. P. Kusuma, Masked face recognition: Progress, dataset, and dataset generation, in: 3rd International Conference on Cybernetics and Intelligent System (ICORIS), 2021, pp. 1–4. 3
- [21] S. Elbishlawi, M. H. Abdelpakey, A. Eltantawy, M. S. Shehata, M. M. Mohamed, Deep learning-based crowd scene analysis survey, *Journal of Imaging* 6 (9) (2020) 95. 3
- [22] C. Shorten, T. M. Khoshgoftaar, B. Furht, Deep learning applications for covid-19, *Journal of big Data* 8 (1) (2021) 1–54. 3
- 900 [23] J. Tomás, A. Rego, S. Viciano-Tudela, J. Lloret, Incorrect facemask-wearing detection using convolutional neural networks with transfer learning, *Healthcare* 9 (8) (2021) 1050. 4
- [24] M. Razavi, H. Alikhani, V. Janfaza, B. Sadeghi, E. Alikhani, An automatic system to monitor the physical distance and face mask wearing of construction workers in covid-19 pandemic, *SN computer science* 3 (1) (2022) 1–8. 4
- 905 [25] Y. Hu, X. Li, Covertheface: face covering monitoring and demonstrating using deep learning and statistical shape analysis, arXiv preprint arXiv:2108.10430. 4, 16, 17
- [26] N. Petrović, Đ. Kocić, Iot-based system for covid-19 indoor safety monitoring, preprint), *IcETRAN*. 4
- [27] A. J. Sathyamoorthy, U. Patel, M. Paul, Y. Savle, D. Manocha, Covid surveillance robot: Monitoring social distancing constraints in indoor scenarios, *Plos one* 16 (12) (2021) e0259713. 4
- 910 [28] D. Yang, E. Yurtsever, V. Renganathan, K. A. Redmill, Ü. Özgüner, A vision-based social distancing and critical density detection system for covid-19, *Sensors* 21 (13) (2021) 4608. 4
- [29] L. Queiroz, H. Oliveira, S. Yanushkevich, Thermal-mask—a dataset for facial mask detection and breathing rate measurement, in: 2021 International Conference on Information and Digital Technologies (IDT), IEEE, 2021, pp. 142–151. 4, 15, 17, 20
- 915 [30] F. I. Eyiokur, H. K. Ekenel, A. Waibel, Unconstrained face mask and face-hand interaction datasets: building a computer vision system to help prevent the transmission of covid-19, *Signal, Image and Video Processing* (2022) 1–8. 4, 5, 6, 15, 17, 19
- [31] C. Beyan, M. Bustreo, M. Shahid, G. L. Bailo, N. Carissimi, A. Del Bue, Analysis of face-touching behavior in large scale social interaction dataset, in: *ICMI*, 2020, pp. 1–10. 4, 15
- 920 [32] N. Damer, J. H. Grebe, C. Chen, F. Boutros, F. Kirchbuchner, A. Kuijper, The effect of wearing a mask on face recognition performance: an exploratory study, in: 2020 International Conference of the Biometrics Special Interest Group (BIOSIG), IEEE, 2020, pp. 1–6. 4, 7, 17, 20
- [33] A. Anwar, A. Raychowdhury, Masked face recognition for secure authentication, arXiv preprint arXiv:2008.11104. 4, 16, 17, 20
- 925 [34] J. Deng, J. Guo, X. An, Z. Zhu, S. Zafeiriou, Masked face recognition challenge: The insightface track report, in: *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, 2021, pp. 1437–1444. 4, 10
- [35] K. Wang, S. Wang, J. Yang, X. Wang, B. Sun, H. Li, Y. You, Mask aware network for masked face recognition in the wild, in: *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, 2021, pp. 1456–1461. 4
- 930 [36] L. Zhang, L. Sun, L. Yu, X. Dong, J. Chen, W. Cai, C. Wang, X. Ning, Arface: attention-aware and regularization for face recognition with reinforcement learning, *IEEE Transactions on Biometrics, Behavior, and Identity Science* 4 (1) (2021) 30–42. 4
- [37] W. Wang, Z. Zhao, H. Zhang, Z. Wang, F. Su, Maskout: A data augmentation method for masked face recognition, in: *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, 2021, pp. 1450–1455. 4
- 935 [38] G. Yolcu, İ. Öztel, A multi-task deep learning system for face detection and age group classification for masked faces, *Sakarya University Journal of Science* 25 (6) (2021) 1394–1407. 4, 12
- [39] R. Golwalkar, N. Mehendale, Age detection with face mask using deep learning and facemasknet-9, *SSRN*. 4, 12
- 940 [40] B. Yang, J. Wu, G. Hattori, Facial expression recognition with the advent of face masks, in: 19th International Conference on Mobile and Ubiquitous Multimedia, 2020, pp. 335–337. 4, 11
- [41] A. F. Abate, L. Cimmino, B.-C. Mocuano, F. Narducci, F. Pop, The limitations for expression recognition in computer vision introduced by facial masks, *Multimedia Tools and Applications* (2022) 1–15. 4, 11
- [42] J. Wang, Y. Liu, Y. Hu, H. Shi, T. Mei, Facex-zoo: A pytorch toolbox for face recognition, 2021, pp. 1–8. 4, 16
- 945 [43] Z. Wang, et al., Masked face recognition dataset and application, arXiv preprint arXiv:2003.09093. 4, 16, 17, 19, 20
- [44] N. U. Din, K. Javed, S. Bae, J. Yi, A novel gan-based network for unmasking of masked face, *IEEE Access* 8 (2020) 44276–44287. 4, 16, 17

- [45] X. Li, C. Shao, Y. Zhou, L. Huang, Face mask removal based on generative adversarial network and texture network, in: 2021 4th International Conference on Robotics, Control and Automation Engineering (RCAE), IEEE, 2021, pp. 86–89. [4](#), [16](#), [17](#)
- [46] L. E. Coelho, R. Prates, W. R. Schwartz, A generative approach for face mask removal using audio and appearance, in: 2021 34th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI), IEEE, 2021, pp. 239–246. [4](#), [16](#), [17](#)
- [47] Y. Sha, J. Zhang, X. Liu, Z. Wu, S. Shan, Efficient face alignment network for masked face, in: 2021 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), IEEE, 2021, pp. 1–6. [4](#), [12](#)
- [48] H. Hu, C. Wang, T. Jiang, Z. Guo, Y. Han, Robust and efficient facial landmark localization, in: 2021 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), IEEE, 2021, pp. 1–7. [4](#), [13](#)
- [49] W. H. Organization, et al., Advice on the use of masks in the context of covid-19: interim guidance, 5 june 2020, Tech. rep., World Health Organization (2020). [5](#)
- [50] A. S. Joshi, S. S. Joshi, G. Kanahasabai, R. Kapil, S. Gupta, Deep learning framework to detect face masks from video footage, in: 2020 12th International Conference on Computational Intelligence and Communication Networks (CICN), IEEE, 2020, pp. 435–440. [5](#), [6](#)
- [51] A. Cabani, K. Hammoudi, H. Benhabiles, M. Melkemi, Maskedface-net—a dataset of correctly/incorrectly masked face images in the context of covid-19, *Smart Health 19* (2021) 100144. [5](#), [6](#), [17](#), [19](#)
- [52] P. Nagrath, R. Jain, A. Madan, R. Arora, P. Kataria, J. Hemanth, Ssdmnnv2: A real time dnn-based face mask detection system using single shot multibox detector and mobilenetv2, *Sustainable cities and society* 66 (2021) 102692. [5](#), [6](#)
- [53] S. Yang, P. Luo, C.-C. Loy, X. Tang, Wider face: A face detection benchmark, in: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR), 2016, pp. 5525–5533. [5](#), [19](#)
- [54] A. Nieto-Rodríguez, M. Mucientes, V. M. Brea, System for medical mask detection in the operating room through facial attributes, in: R. Paredes, J. S. Cardoso, X. M. Pardo (Eds.), *Pattern Recognition and Image Analysis*, Springer International Publishing, Cham, 2015, pp. 138–145. [5](#)
- [55] P. Viola, M. J. Jones, Robust real-time face detection, *International journal of computer vision* 57 (2) (2004) 137–154. [5](#)
- [56] S. Ge, J. Li, Q. Ye, Z. Luo, Detecting masked faces in the wild with lle-cnns, in: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR), 2017, pp. 2682–2690. [5](#), [6](#), [17](#), [19](#)
- [57] X. Fan, M. Jiang, Retinafacemask: A single stage face mask detector for assisting control of the covid-19 pandemic, in: 2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC), IEEE, 2021, pp. 832–837. [6](#)
- [58] J. Deng, J. Guo, E. Ververas, I. Kotsia, S. Zafeiriou, Retinaface: Single-shot multi-level face localisation in the wild, in: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR), 2020, pp. 5203–5212. [6](#), [12](#), [15](#)
- [59] J. Redmon, A. Farhadi, Yolov3: An incremental improvement, *arXiv preprint arXiv:1804.02767*. [6](#), [14](#), [15](#)
- [60] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, L.-C. Chen, Mobilenetv2: Inverted residuals and linear bottlenecks, in: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR), 2018, pp. 4510–4520. [6](#)
- [61] M. Loey, G. Manogaran, M. H. N. Taha, N. E. M. Khalifa, Fighting against covid-19: A novel deep learning model based on yolo-v2 with resnet-50 for medical face mask detection, *Sustainable cities and society* 65 (2021) 102600. [6](#)
- [62] J. Redmon, A. Farhadi, Yolo9000: better, faster, stronger, in: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR), 2017, pp. 7263–7271. [6](#)
- [63] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR), 2016, pp. 770–778. [6](#), [10](#), [15](#)
- [64] K. Zhang, Z. Zhang, Z. Li, Y. Qiao, Joint face detection and alignment using multitask cascaded convolutional networks, *IEEE Signal Proc. Letters* 23 (10). [6](#)
- [65] B. Wang, Y. Zhao, C. P. Chen, Hybrid transfer learning and broad learning system for wearing mask detection in the covid-19 era, *IEEE Transactions on Instrumentation and Measurement* 70 (2021) 1–12. [6](#)
- [66] S. Ren, K. He, R. Girshick, J. Sun, Faster r-cnn: Towards real-time object detection with region proposal networks, *Advances in neural information processing systems* 28 (2015) 1–9. [6](#)
- [67] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, Rethinking the inception architecture for computer vision, in: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR), 2016, pp. 2818–2826. [6](#)
- [68] B. Roy, S. Nandy, D. Ghosh, D. Dutta, P. Biswas, T. Das, Moxa: A deep learning based unmanned approach for real-time monitoring of people wearing medical masks, *Transactions of the Indian National Academy of Engineering* 5 (3) (2020) 509–518. [6](#)
- [69] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, A. C. Berg, Ssd: Single shot multibox detector, in: European conference on computer vision (ECCV), Springer, 2016, pp. 21–37. [6](#)
- [70] M. Tan, Q. Le, Efficientnet: Rethinking model scaling for convolutional neural networks, in: International conference on machine learning (ICML), PMLR, 2019, pp. 6105–6114. [6](#), [14](#), [15](#)
- [71] X. Jiang, T. Gao, Z. Zhu, Y. Zhao, Real-time face mask detection method based on yolov3, *Electronics* 10 (7) (2021) 837. [6](#), [17](#), [19](#)
- [72] A. Kantarcı, F. Ofli, M. Imran, H. K. Ekenel, Bias aware face mask detection dataset, *arXiv preprint arXiv:2211.01207*. [6](#), [17](#), [19](#)
- [73] G. Jocher, et. al., ultralytics/yolov5: v5.0 - YOLOv5-P6 1280 models, AWS, Supervise.ly and YouTube integrations (Apr. 2021). [doi:10.5281/zenodo.4679653](https://doi.org/10.5281/zenodo.4679653). [6](#)
- [74] B. Qin, D. Li, Identifying facemask-wearing condition using image super-resolution with classification network to prevent covid-19, *Sensors* 20 (18) (2020) 5236. [7](#)
- [75] J. Kim, J. Choi, J. Yi, M. A. Turk, Effective representation using ICA for face recognition robust to local distortion and partial occlusion, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27 (12) (2005) 1977–1981. [7](#)
- [76] W. Ou, X. You, D. Tao, P. Zhang, Y. Tang, Z. Zhu, Robust face recognition via occlusion dictionary learning, *Pattern*

- Recognition 47 (4) (2014) 1559–1572. 7
- [77] L. Song, D. Gong, Z. Li, C. Liu, W. Liu, Occlusion robust face recognition based on mask learning with pairwise differential siamese network, in: *IEEE/CVF International Conference on Computer Vision, ICCV, 2019*, pp. 773–782. 7
- [78] H. Qiu, D. Gong, Z. Li, W. Liu, D. Tao, End2end occluded face recognition by masking corrupted features, *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2021) 1–9. 7
- [79] P. C. Neto, J. R. Pinto, F. Boutros, N. Damer, A. F. Sequeira, J. S. Cardoso, Beyond masks: On the generalization of masked face recognition models to occluded face recognition, *IEEE Access* 10 (2022) 86222–86233. 7
- [80] P. C. Neto, F. Boutros, J. R. Pinto, N. Damer, A. F. Sequeira, J. S. Cardoso, M. Bengherabi, A. Bousnat, S. Boucheta, N. Hebbadj, M. E. Erakin, U. Demir, H. K. Ekenel, P. B. de Queiroz Vidal, D. Menotti, OCFR 2022: Competition on occluded face recognition from synthetically generated structure-aware occlusions, in: *IJCB, IEEE, 2022*. 7
- [81] M. E. Erakin, U. Demir, H. K. Ekenel, On recognizing occluded faces in the wild, in: *2021 International Conference of the Biometrics Special Interest Group (BIOSIG), IEEE, 2021*, pp. 1–5. 7, 17, 20
- [82] N. Damer, F. Boutros, M. Süßmilch, F. Kirchbuchner, A. Kuijper, Extended evaluation of the effect of real and simulated masks on face recognition performance, *Iet Biometrics* 10 (5) (2021) 548–561. 7, 17, 20
- [83] M. Ngan, P. Grother, K. Hanaoka, Ongoing face recognition vendor test (frvt) part 6b: Face recognition accuracy with face masks using post-covid-19 algorithms (2020). doi:<https://doi.org/10.6028/NIST.IR.8331>. 7
- [84] Arun Vemury and fake Hasselgren and John Howard and Yevgeniy Sirotnin, 2020 biometric rally results - face masks face recognition performance, <https://mdtf.org/Rally2020/Results2020>, last accessed: January 2, 2023(2020). 7
- [85] N. Damer, F. Boutros, M. Süßmilch, M. Fang, F. Kirchbuchner, A. Kuijper, Masked face recognition: Human versus machine, *IET Biometrics* 11 (5) (2022) 512–528. 7
- [86] Y. Li, S. Liu, J. Yang, M.-H. Yang, Generative face completion, in: *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR), 2017*, pp. 3911–3919. 8
- [87] L. Song, J. Cao, L. Song, Y. Hu, R. He, Geometry-aware face completion and editing, in: *AAAI, AAAI Press, 2019*, pp. 2506–2513. 8
- [88] C. Han, J. Wang, Face image inpainting with evolutionary generators, *IEEE Signal Processing Letters* 28 (2021) 190–193. 9
- [89] Z. Niu, H. Li, Y. Li, Y. Mei, J. Yang, An adaptive face image inpainting algorithm based on feature symmetry, *Symmetry* 12 (2) (2020) 190. 9
- [90] X. Zhang, X. Wang, C. Shi, Z. Yan, X. Li, B. Kong, S. Lyu, B. Zhu, J. Lv, Y. Yin, Q. Song, X. Wu, I. Mumtaz, De-gan: Domain embedded gan for high quality face image inpainting, *Pattern Recognition* 124 (2022) 108415. 9
- [91] Y. Jiang, F. Yang, Z. Bian, C. Lu, S. Xia, Mask removal: Face inpainting via attributes, *Multimedia Tools and Applications* (2022) 1–13. 9
- [92] C. Li, S. Ge, Y. Hua, H. Liu, X. Jin, Occluded face recognition by identity-preserving inpainting, in: *International Symposium on Artificial Intelligence and Robotics, Springer, 2018*, pp. 427–437. 9
- [93] F. Boutros, N. Damer, F. Kirchbuchner, A. Kuijper, Self-restrained triplet loss for accurate masked face recognition, *Pattern Recognition* 124 (2022) 108473. 9, 17, 20
- [94] P. C. Neto, F. Boutros, J. R. Pinto, N. Damer, A. F. Sequeira, J. S. Cardoso, Focusface: Multi-task contrastive learning for masked face recognition, in: *2021 16th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2021), IEEE, 2021*, pp. 01–08. 9
- [95] F. Schroff, D. Kalenichenko, J. Philbin, Facenet: A unified embedding for face recognition and clustering, in: *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR), 2015*, pp. 815–823. 9
- [96] F. Boutros, N. Damer, F. Kirchbuchner, A. Kuijper, Elasticface: Elastic margin loss for deep face recognition, in: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2022*, pp. 1578–1587. 9
- [97] J. Deng, J. Guo, N. Xue, S. Zafeiriou, Arcface: Additive angular margin loss for deep face recognition, in: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR), 2019*, pp. 4690–4699. 9, 10
- [98] H. Deng, Z. Feng, G. Qian, X. Lv, H. Li, G. Li, Mfcosface: a masked-face recognition algorithm based on large margin cosine loss, *Applied sciences* 11 (16) (2021) 7310. 9
- [99] M. Huber, F. Boutros, F. Kirchbuchner, N. Damer, Mask-invariant face recognition through template-level knowledge distillation, in: *2021 16th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2021), IEEE, 2021*, pp. 1–8. 9, 20
- [100] H. Qian, P. Zhang, S. Ji, S. Cao, Y. Xu, Improving representation consistency with pairwise loss for masked face recognition, in: *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), 2021*, pp. 1462–1467. 9
- [101] M. Geng, P. Peng, Y. Huang, Y. Tian, Masked face recognition with generative data augmentation and domain constrained ranking, in: *Proceedings of the 28th ACM International Conference on Multimedia, 2020*, pp. 2246–2254. 9
- [102] G.-S. J. Hsu, H.-Y. Wu, C.-H. Tsai, S. Yanushkevich, M. L. Gavrilova, Masked face recognition from synthesis to reality, *IEEE Access* 10 (2022) 37938–37952. 9
- [103] W. Chang, M. Tsai, S. Lo, Rissanet: A hybrid backbone of residual block and self-attention module for masked face recognition, in: *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), 2021*, pp. 1468–1476. 9
- [104] Y. Li, K. Guo, Y. Lu, L. Liu, Cropping and attention based approach for masked face recognition, *Appl. Intell.* 51 (5) (2021) 3012–3025. 9
- [105] H. M. Nguyen, N. Reddy, A. Rattani, R. Derakhshani, VISOB 2.0 - the second international competition on mobile ocular biometric recognition, in: *ICPR Workshops (8), Vol. 12668 of Lecture Notes in Computer Science, Springer, 2020*,

- pp. 200–208. [9](#)
- [106] F. Alonso-Fernandez, J. Bigün, A survey on periocular biometrics research, *Pattern Recognit. Lett.* 82 (2016) 92–105. [10](#)
- 1080 [107] F. Boutros, N. Damer, K. B. Raja, F. Kirchbuchner, A. Kuijper, Template-driven knowledge distillation for compact and accurate periocular biometrics deep-learning models, *Sensors* 22 (5) (2022) 1921. [10](#)
- [108] F. Boutros, N. Damer, K. B. Raja, R. Ramachandra, F. Kirchbuchner, A. Kuijper, Iris and periocular biometrics for head mounted displays: Segmentation, recognition, and synthetic data generation, *Image Vis. Comput.* 104 (2020) 104007. [10](#)
- 1085 [109] S. Dharanesh, A. Rattani, Post-covid-19 mask-aware face recognition system, in: 2021 IEEE International Symposium on Technologies for Homeland Security (HST), 2021, pp. 1–7. [10](#)
- [110] Ardiansyah, D. Y. Liliana, Facial biometric identification in the masked face, in: 2021 13th International Conference on Information Communication Technology and System (ICTS), 2021, pp. 129–133. [10](#)
- [111] M. Junayed, A. Sadeghzadeh, M. Islam, Deep covariance feature and cnn-based end-to-end masked face recognition, in: 2021 16th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2021), 2021, pp. 1–8. [10](#)
- 1090 [112] R. Sharma, A. Ross, Periocular biometrics and its relevance to partially masked faces: A survey, *CoRR* abs/2203.15203. [10](#)
- [113] F. Boutros, N. Damer, J. N. Kolf, K. B. Raja, F. Kirchbuchner, R. Ramachandra, A. Kuijper, P. Fang, C. Zhang, F. Wang, D. Montero, N. Aginako, B. Sierra, M. Nieto, M. E. Erakin, U. Demir, H. K. Ekenel, A. Kataoka, K. Ichikawa, S. Kubo, J. Zhang, M. He, D. Han, S. Shan, K. Grm, V. Struc, S. Seneviratne, N. Kasthuriarachchi, S. Rasnayaka, P. C. Neto, A. F. Sequeira, J. R. Pinto, M. Saffari, J. S. Cardoso, MFR 2021: Masked face recognition competition, in: International IEEE Joint Conference on Biometrics, IJCB, IEEE, 2021, pp. 1–10. [10](#)
- 1095 [114] International IEEE Joint Conference on Biometrics, IJCB 2021, Shenzhen, China, August 4-7, 2021, IEEE, 2021. doi: [10.1109/IJCB52358.2021](#). URL <https://doi.org/10.1109/IJCB52358.2021> [10](#)
- 1100 [115] IEEE/CVF International Conference on Computer Vision Workshops, ICCVW, IEEE, 2021. doi: [10.1109/ICCVW54120.2021](#). URL <https://doi.org/10.1109/ICCVW54120.2021> [10](#)
- [116] ISO/IEC JTC1 SC37 Biometrics, ISO/IEC 30107-3:2017 Information technology — Biometric presentation attack detection — Part 3: Testing and reporting, International Organization for Standardization, 2016. [10](#)
- 1105 [117] N. Damer, A. M. Saladie, S. Zienert, Y. Wainakh, P. Terhörst, F. Kirchbuchner, A. Kuijper, To detect or not to detect: The right faces to morph, in: 2019 International Conference on Biometrics (ICB), IEEE, 2019, pp. 1–8. [10](#)
- [118] N. Damer, C. A. F. López, M. Fang, N. Spiller, M. V. Pham, F. Boutros, Privacy-friendly synthetic data for the development of face morphing attack detectors, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2022. [10](#)
- 1110 [119] P. Drozdowski, S. Groberek, J. Schurse, C. Rathgeb, F. Stockhardt, C. Busch, Makeup presentation attack potential revisited: Skills pay the bills, in: 9th IEEE International Workshop on Biometrics and Forensics, IWBF, IEEE, 2021, pp. 1–6. [10](#)
- [120] N. Damer, Y. Wainakh, V. Boller, S. von den Berken, P. Terhörst, A. Braun, A. Kuijper, Crazyfaces: Unassisted circumvention of watchlist face identification, in: 9th IEEE International Conference on Biometrics Theory, Applications and Systems, BTAS, IEEE, 2018, pp. 1–9. [10](#)
- 1115 [121] R. Raghavendra, C. Busch, Presentation attack detection methods for face recognition systems: A comprehensive survey, *ACM Comput. Surv.* 50 (1) (2017) 8:1–8:37. [10](#)
- [122] F. Peng, L. Qin, M. Long, Face presentation attack detection using guided scale texture, *Multimedia Tools and Applications* 77 (7) (2018) 8883–8909. [10](#)
- 1120 [123] N. Damer, K. Dimitrov, Practical view on face presentation attack detection, in: R. C. Wilson, E. R. Hancock, W. A. P. Smith (Eds.), Proceedings of the British Machine Vision Conference 2016, BMVC, BMVA Press, 2016. [10](#)
- [124] R. Raghavendra, K. B. Raja, S. Marcel, C. Busch, Face presentation attack detection across spectrum using time-frequency descriptors of maximal response in laplacian scale-space, in: Sixth International Conference on Image Processing Theory, Tools and Applications, IPTA, IEEE, 2016, pp. 1–6. [10](#)
- 1125 [125] M. Fang, N. Damer, F. Kirchbuchner, A. Kuijper, Learnable multi-level frequency decomposition and hierarchical attention mechanism for generalized face presentation attack detection, in: IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), IEEE, 2022, pp. 1131–1140. [10](#)
- [126] M. Fang, N. Damer, F. Kirchbuchner, A. Kuijper, Real masks and spoof faces: On the masked face presentation attack detection, *Pattern Recognition* 123 (2022) 108398. [10](#), [20](#)
- 1130 [127] M. Fang, F. Boutros, A. Kuijper, N. Damer, Partial attack supervision and regional weighted inference for masked face presentation attack detection, in: 2021 16th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2021), IEEE, 2021, pp. 1–8. [10](#), [11](#)
- [128] ISO/IEC JTC1 SC37 Biometrics, ISO/IEC 29794-1:2016 Information technology - Biometric sample quality - Part 1: Framework, International Organization for Standardization, 2016. [11](#)
- 1135 [129] L. Best-Rowden, A. K. Jain, Learning face image quality from human assessments, *IEEE Transactions on Information Forensics and Security* 13 (12) (2018) 3064–3077. [11](#)
- [130] ISO/IEC JTC1 SC37 Biometrics, ISO/IEC 2382-37:2017 Information technology - Vocabulary - Part 37: Biometrics, International Organization for Standardization, 2017. [11](#)
- [131] J. Hernandez-Ortega, J. Galbally, J. Fierrez, R. Haraksim, L. Beslay, Faceqnet: Quality assessment for face recognition based on deep learning, in: 2019 International Conference on Biometrics (ICB), IEEE, 2019, pp. 1–8. [11](#)
- 1140 [132] F. Ou, X. Chen, R. Zhang, Y. Huang, S. Li, J. Li, Y. Li, L. Cao, Y. Wang, SDD-FIQA: unsupervised face image quality assessment with similarity distribution distance, in: IEEE/CVF Conference on Computer Vision and Pattern Recognition

- (CVPR), 2021, pp. 7670–7679. [11](#)
- 1145 [133] P. Terhörst, J. N. Kolf, N. Damer, F. Kirchbuchner, A. Kuijper, SER-FIQ: unsupervised estimation of face image quality based on stochastic embedding robustness, in: *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 5650–5659. [11](#)
- [134] Q. Meng, S. Zhao, Z. Huang, F. Zhou, Magface: A universal representation for face recognition and quality assessment, in: *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021, pp. 14225–14234. [11](#)
- 1150 [135] F. Boutros, M. Fang, M. Klemt, B. Fu, N. Damer, CR-FIQA: face image quality assessment by learning sample relative classifiability, *CoRR abs/2112.06592*. [11](#)
- [136] B. Fu, C. Chen, O. Henniger, N. Damer, A deep insight into measuring face image utility with general and face-specific image quality metrics, in: *IEEE/CVF Winter Conference on Applications of Computer Vision, WACV, IEEE*, 2022, pp. 1121–1130. [11](#)
- 1155 [137] P. Terhörst, J. N. Kolf, N. Damer, F. Kirchbuchner, A. Kuijper, Face quality estimation and its correlation to demographic and non-demographic bias in face recognition, in: *IEEE International Joint Conference on Biometrics, IJCB, IEEE*, 2020, pp. 1–11. [11](#)
- [138] B. Fu, F. Kirchbuchner, N. Damer, The effect of wearing a face mask on face image quality, in: *2021 16th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2021), IEEE*, 2021, pp. 1–8. [11](#)
- 1160 [139] J. F. Cohn, A. J. Zlochower, J. Lien, T. Kanade, Automated face analysis by feature point tracking has high concurrent validity with manual faces coding, *Psychophysiology* 36 (1) (1999) 35–43. [11](#)
- [140] M. Valstar, M. Pantic, Fully automatic facial action unit detection and temporal analysis, in: *2006 Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'06), IEEE*, 2006, pp. 149–149. [11](#)
- [141] R. Zhi, M. Liu, D. Zhang, A comprehensive survey on automatic facial action unit analysis, *The Visual Computer* 36 (5) (2020) 1067–1093. [11](#)
- 1165 [142] P. Ekman, W. V. Friesen, Facial action coding system, *Environmental Psychology & Nonverbal Behavior*. [11](#)
- [143] G. B. Huang, M. Mattar, T. Berg, E. Learned-Miller, Labeled faces in the wild: A database for studying face recognition in unconstrained environments, in: *Workshop on faces in 'Real-Life' Images: detection, alignment, and recognition*, 2008, pp. 1–11. [11](#), [16](#), [19](#), [20](#)
- 1170 [144] P. Barros, A. Sciutti, I only have eyes for you: The impact of masks on convolutional-based facial expression recognition, in: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021, pp. 1226–1231. [11](#)
- [145] A. Mollahosseini, B. Hasani, M. H. Mahoor, Affectnet: A database for facial expression, valence, and arousal computing in the wild, *IEEE Transactions on Affective Computing* 10 (1) (2017) 18–31. [11](#)
- 1175 [146] P. Barros, N. Churamani, A. Sciutti, The facechannel: a fast and furious deep neural network for facial expression recognition, *SN Computer Science* 1 (6) (2020) 1–10. [12](#)
- [147] B. Yang, W. Jianming, G. Hattori, Face mask aware robust facial expression recognition during the covid-19 pandemic, in: *2021 IEEE International Conference on Image Processing (ICIP), IEEE*, 2021, pp. 240–244. [12](#)
- [148] K. Wang, X. Peng, J. Yang, D. Meng, Y. Qiao, Region attention networks for pose and occlusion robust facial expression recognition, *IEEE Transactions on Image Processing* 29 (2020) 4057–4069. [12](#)
- 1180 [149] Y. Li, J. Zeng, S. Shan, X. Chen, Occlusion aware facial expression recognition using cnn with attention mechanism, *IEEE Transactions on Image Processing* 28 (5) (2018) 2439–2450. [12](#)
- [150] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. C. Courville, Y. Bengio, Generative adversarial nets, in: *NIPS*, 2014. [12](#)
- 1185 [151] Z. Zhang, Y. Song, H. Qi, Age progression/regression by conditional adversarial autoencoder, in: *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, 2017, pp. 5810–5818. [12](#)
- [152] D. E. King, Dlib-ml: A machine learning toolkit, *Journal of Machine Learning Research* 10 (2009) 1755–1758. [12](#), [16](#)
- [153] S. Honari, P. Molchanov, S. Tyree, P. Vincent, C. Pal, J. Kautz, Improving landmark localization with semi-supervised learning, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 1–9. [12](#)
- 1190 [154] M. Xiang, Y. Liu, T. Liao, X. Zhu, C. Yang, W. Liu, H. Shi, The 3rd grand challenge of lightweight 106-point facial landmark localization on masked faces, in: *2021 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), IEEE*, 2021, pp. 1–6. [12](#), [13](#), [16](#)
- [155] T. Wen, Z. Ding, Y. Yao, Y. Ge, X. Qian, et al., Towards efficient masked-face alignment via cascaded regression, in: *2021 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), IEEE*, 2021, pp. 1–5. [13](#)
- 1195 [156] M. Al-Sa'd, S. Kiranyaz, I. Ahmad, C. Sundell, M. Vakkuri, M. Gabbouj, A Social Distance Estimation and Crowd Monitoring System for Surveillance Cameras, *Sensors* 22 (2). [13](#), [15](#)
- [157] I. J. C. Valencia, E. P. Dadios, A. M. Fillone, J. C. V. Puno, R. G. Baldovino, R. K. C. Billones, Vision-based Crowd Counting and Social Distancing Monitoring using Tiny-YOLOv4 and DeepSORT, in: *2021 IEEE International Smart Cities Conference (ISC2)*, 2021, pp. 1–7. [13](#), [15](#)
- 1200 [158] P. Somaldo, F. A. Ferdiansyah, G. Jati, W. Jatmiko, Developing Smart COVID-19 Social Distancing Surveillance Drone using YOLO Implemented in Robot Operating System simulation environment, in: *2020 IEEE 8th R10 Humanitarian Technology Conference (R10-HTC)*, 2020, pp. 1–6. [13](#), [15](#)
- [159] M.-N. Nguyen, V.-H. Tran, T.-N. Huynh, Depth Embedded and Dense Dilated Convolutional Network for Crowd Density Estimation, in: *2021 International Conference on System Science and Engineering (ICSSE)*, 2021, pp. 221–225. [13](#), [14](#)
- 1205 [160] K. J. Almalki, M. Mohzary, B.-Y. Choi, S. Song, Y. Chen, Mosaic: Modeling Safety Index in Crowd by Detecting Face Masks against COVID-19 and Beyond, in: *2021 IEEE International Smart Cities Conference (ISC2)*, 2021, pp. 1–7. [13](#), [14](#)

- [161] Y. He, Y. Xia, Y. Wang, B. Yin, Jointly Attention Network for Crowd Counting, *Neurocomputing*. [13](#), [14](#)
- 1210 [162] M. Dosi, K. Thakral, S. Mittal, M. Vatsa, R. Singh, Aecnet: Attentive efficientnet for crowd counting, in: 2021 16th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2021), IEEE, 2021, pp. 1–8. [13](#), [14](#)
- [163] L. M. Alvarez, M. J. Llanos, J. R. Obrero, L. S. Felicuzo, Physical distance and crowd monitoring system using yolov3, in: 2021 1st International Conference in Information and Computing Research (iCORE), IEEE, 2021, pp. 139–144. [13](#), [14](#)
- 1215 [164] S. Kammoun Jarraya, M. Alotibi, M. Ali, A Deep-CNN Crowd Counting Model for Enforcing Social Distancing during COVID19 Pandemic: Application to Saudi Arabia’s Public Places, *Computers, Materials and Continua* 66 (2021) 1315–1328. [13](#), [14](#)
- [165] P. N. Amin, S. S. Moghe, S. N. Prabhakar, C. M. Nehete, Deep Learning Based Face Mask Detection and Crowd Counting, in: 2021 6th International Conference for Convergence in Technology (I2CT), 2021, pp. 1–5. [13](#), [14](#)
- 1220 [166] B. Li, H. Huang, Z. Ang, P. Liu, C. Liu, Approaches on crowd counting and density estimation: a review, *Pattern Analysis and Applications* 24. [14](#)
- [167] Q. Wang, J. Gao, W. Lin, X. Li, NWPU-Crowd: A Large-Scale Benchmark for Crowd Counting and Localization, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 43 (06) (2021) 2141–2149. [14](#)
- [168] P. Sabzmeydani, G. Mori, Detecting pedestrians by learning shapelet features, in: 2007 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2007, pp. 1–8. [14](#)
- 1225 [169] N. Dalal, B. Triggs, Histograms of oriented gradients for human detection, in: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), Vol. 1, 2005, pp. 886–893. [14](#)
- [170] S.-F. Lin, J.-Y. Chen, H.-X. Chao, Estimation of number of people in crowded scenes using perspective transformation, *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans* 31 (6) (2001) 645–654. [14](#)
- 1230 [171] L. Fiaschi, U. Koethe, R. Nair, F. A. Hamprecht, Learning to count with regression forest and structured labels, in: Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012), 2012, pp. 2685–2688. [14](#)
- [172] D. Ryan, S. Denman, C. Fookes, S. Sridharan, Crowd Counting Using Multiple Local Features, in: 2009 Digital Image Computing: Techniques and Applications, 2009, pp. 81–88. [14](#)
- [173] V. Pham, T. Kozakaya, O. Yamaguchi, R. Okada, COUNT Forest: CO-Voting Uncertain Number of Targets Using Random Forest for Crowd Density Estimation, in: 2015 IEEE International Conference on Computer Vision (ICCV), 2015, pp. 3253–3261. [14](#)
- 1235 [174] C. Arteta, V. S. Lempitsky, J. A. Noble, A. Zisserman, Interactive Object Counting, in: European Conference on Computer Vision (ECCV), 2014, pp. 504–518. [14](#)
- [175] Z. Fan, H. Zhang, Z. Zhang, G. Lu, Y. Zhang, Y. Wang, A survey of crowd counting and density estimation based on convolutional neural network, *Neurocomputing* 472 (2022) 224–251. [14](#)
- 1240 [176] C. Li, R. Wang, J. Li, L. Fei, Face Detection Based on YOLOv3, *Recent Trends in Intelligent Computing, Communication and Devices*. [14](#)
- [177] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, H. Adam, Mobilenets: Efficient convolutional neural networks for mobile vision applications, *arXiv preprint arXiv:1704.04861*. [14](#)
- [178] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, *arXiv preprint arXiv:1409.1556*. [14](#)
- 1245 [179] N. Wojke, A. Bewley, D. Paulus, Simple online and realtime tracking with a deep association metric, in: 2017 IEEE international conference on image processing (ICIP), IEEE, 2017, pp. 3645–3649. [15](#)
- [180] Z. Cao, G. Hidalgo Martinez, T. Simon, S. Wei, Y. A. Sheikh, OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields, *IEEE Transactions on Pattern Analysis and Machine Intelligence*. [15](#)
- 1250 [181] L. Pan, M. Mu, P. Yang, Y. Sun, R. Wang, J. Yan, P. Li, B. Hu, J. Wang, C. Hu, et al., Clinical characteristics of covid-19 patients with digestive symptoms in hubei, china: a descriptive, cross-sectional, multicenter study, *The American journal of gastroenterology* 115. [15](#)
- [182] A. Natarajan, H.-W. Su, C. Heneghan, L. Blunt, C. O’Connor, L. Niehaus, Measurement of respiratory rate using wearable devices and applications to covid-19 detection, *NPJ digital medicine* 4 (1) (2021) 1–10. [15](#)
- 1255 [183] A. T. Purnomo, D.-B. Lin, T. Adiprabowo, W. F. Hendria, Non-contact monitoring and classification of breathing pattern for the supervision of people infected by covid-19, *Sensors* 21 (9). [15](#)
- [184] D. Perpetuini, C. Filippini, D. Cardone, A. Merla, An overview of thermal infrared imaging-based screenings during pandemic emergencies, *International Journal of Environmental Research and Public Health* 18 (6) (2021) 3286. [15](#)
- 1260 [185] Z. Liu, P. Luo, X. Wang, X. Tang, Deep learning face attributes in the wild, in: IEEE International Conference on Computer Vision (ICCV), 2015, pp. 3730–3738. [16](#), [19](#)
- [186] T. Zheng, W. Deng, J. Hu, Cross-age lfw: A database for studying cross-age face recognition in unconstrained environments, *arXiv preprint arXiv:1708.08197*. [16](#)
- [187] I. P. A. Karasugi, Williem, Face mask invariant end-to-end face recognition, in: European Conference on Computer Vision Workshops (ECCVW), Springer, 2020, pp. 261–276. [16](#)
- 1265 [188] T. Mare, G. Duta, M.-I. Georgescu, A. Sandru, B. Alexe, M. Popescu, R. T. Ionescu, A realistic approach to generate masked faces applied on two novel masked face recognition data sets, in: Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks (NeurIPS Datasets and Benchmarks 2021), 2021, pp. 1–11. [16](#)
- [189] B. Huang, Z. Wang, G. Wang, K. Jiang, K. Zeng, Z. Han, X. Tian, Y. Yang, When face recognition meets occlusion: A new benchmark, in: ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2021, pp. 4240–4244. [16](#)
- 1270 [190] C. Wang, H. Fang, Y. Zhong, W. Deng, Mlfw: A database for face recognition on masked faces, *arXiv preprint arXiv:2109.05804*. [16](#)

- [191] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: International Conference on Medical image computing and computer-assisted intervention, Springer, 2015, pp. 234–241. 17
- 1275 [192] Y. Mroueh, T. Sercu, V. Goel, Mcgan: Mean and covariance feature matching gan, in: International conference on machine learning (ICML), PMLR, 2017, pp. 2527–2535. 17
- [193] M. Jiang, X. Fan, Retinamask: a face mask detector, arXiv preprint arXiv:2005.03950. 17, 19
- [194] Z. Wang, P. Wang, P. C. Louis, L. E. Wheless, Y. Huo, Wearmask: Fast in-browser face mask detection with serverless edge computing for covid-19, arXiv preprint arXiv:2101.00784. 17, 19
- 1280 [195] M. Vrigkas, E.-A. Kourfalidou, M. E. Plissiti, C. Nikou, Facemask: A new image dataset for the automated identification of people wearing masks in the wild, *Sensors* 22 (3) (2022) 896. 17, 19
- [196] S. Mishra, P. Majumdar, M. Dosi, M. Vatsa, R. Singh, Dual sensor indian masked face dataset, in: 2021 16th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2021), 2021, pp. 1–8. 17, 20
- [197] N. Głowacka, J. Rumiński, Face with mask detection in thermal images using deep neural networks, *Sensors* 21 (19) (2021) 6387. 17, 20
- 1285 [198] R. J. Ward, F. P. M. Jjunju, I. Kabenge, R. Wanyenze, E. J. Griffith, N. Banadda, S. Taylor, A. Marshall, Flunet: An ai-enabled influenza-like warning system, *IEEE sensors journal* 21 (21) (2021) 24740–24748. 17, 20
- [199] Face mask detection, <https://www.kaggle.com/andrewmvd/face-mask-detection>, accessed: August 8, 2022. 17, 21
- [200] Face mask detection 12k images dataset, <https://www.kaggle.com/ashishjangra27/face-mask-12k-images-dataset>, accessed: August 8, 2022. 17, 21
- 1290 [201] Face mask lite dataset, <https://www.kaggle.com/prasoonkottarathil/face-mask-lite-dataset>, accessed: August 8, 2022. 17, 21
- [202] A. Marceddu, R. Ferrero, B. Montrucchio, Ways to wear a mask or a respirator (wwmr-db), <https://dx.doi.org/10.21227/8atn-gn55> (2021). doi:10.21227/8atn-gn55. 17, 21
- 1295 [203] Face mask detection 12k images dataset, <https://humansintheloop.org/resources/datasets/medical-mask-dataset/>, accessed: August 8, 2022. 17, 21
- [204] Aizootech face mask detection dataset, <https://github.com/AIZ00Tech/FaceMaskDetection>, accessed: August 8, 2022. 17, 21
- [205] T. Karras, S. Laine, T. Aila, A style-based generator architecture for generative adversarial networks, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 4401–4410. 20
- 1300 [206] Y. Guo, L. Zhang, Y. Hu, X. He, J. Gao, Ms-celeb-1m: A dataset and benchmark for large-scale face recognition, in: European conference on computer vision (ECCV), Springer, 2016, pp. 87–102. 20
- [207] B. Maze, J. Adams, J. A. Duncan, N. Kalka, T. Miller, C. Otto, A. K. Jain, W. T. Niggel, J. Anderson, J. Cheney, et al., Iarpa janus benchmark-c: Face dataset and protocol, in: 2018 international conference on biometrics (ICB), IEEE, 2018, pp. 158–165. 20
- 1305 [208] M. Abdrakhmanova, A. Kuzdeuov, S. Jarju, Y. Khassanov, M. Lewis, H. A. Varol, Speakingfaces: A large-scale multi-modal dataset of voice commands with visual and thermal video streams, *Sensors* 21 (10) (2021) 3465. 20
- [209] Z. Li, D. Hoiem, Learning without forgetting, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 40 (12) (2017) 2935–2947. 21
- 1310 [210] D. Lopez-Paz, M. Ranzato, Gradient episodic memory for continual learning, *Advances in neural information processing systems* 30. 21
- [211] H. Shin, J. K. Lee, J. Kim, J. Kim, Continual learning with deep generative replay, *Advances in neural information processing systems* 30. 21
- [212] R. Aljundi, K. Kelchtermans, T. Tuytelaars, Task-free continual learning, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 11254–11263. 21
- 1315 [213] G. I. Parisi, R. Kemker, J. L. Part, C. Kanan, S. Wermter, Continual lifelong learning with neural networks: A review, *Neural Networks* 113 (2019) 54–71. 21
- [214] G. M. Van de Ven, A. S. Tolias, Three scenarios for continual learning, arXiv preprint arXiv:1904.07734. 21
- 1320 [215] O. Vinyals, C. Blundell, T. Lillicrap, D. Wierstra, et al., Matching networks for one shot learning, *Advances in neural information processing systems* 29. 21
- [216] S. Gidaris, N. Komodakis, Dynamic few-shot visual learning without forgetting, in: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR), 2018, pp. 4367–4375. 21
- [217] H. Yin, P. Molchanov, J. M. Alvarez, Z. Li, A. Mallya, D. Hoiem, N. K. Jha, J. Kautz, Dreaming to distill: Data-free knowledge transfer via deepinversion, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, pp. 8715–8724. 21
- 1325 [218] C. Zhang, N. Song, G. Lin, Y. Zheng, P. Pan, Y. Xu, Few-shot incremental learning with continually evolved classifiers, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021, pp. 12455–12464. 21
- [219] T. Ahmad, A. R. Dhamija, S. Cruz, R. Rabinowitz, C. Li, M. Jafarzadeh, T. E. Boult, Few-shot class incremental learning leveraging self-supervised features, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022, pp. 3900–3910. 21
- 1330 [220] P. Drodzowski, C. Rathgeb, A. Dantcheva, N. Damer, C. Busch, Demographic bias in biometrics: A survey on an emerging challenge, *IEEE Transactions on Technology and Society* 1 (2) (2020) 89–103. 22
- [221] R. Heilweil, Big tech companies back away from selling facial recognition to police. That’s progress., <https://www.vox.com/recode/2020/6/10/21287194/amazon-microsoft-ibm-facial-recognition-moratorium-police>, accessed: August 8, 2022 (2020). 22
- 1335 [222] B. Meden, P. Rot, P. Terhörst, N. Damer, A. Kuijper, W. J. Scheirer, A. Ross, P. Peer, V. Štruc, Privacy-enhancing face

- biometrics: A comprehensive survey, *IEEE Transactions on Information Forensics and Security*. [22](#)
- 1340 [223] A. Puc, V. Štruc, K. Grm, Analysis of race and gender bias in deep age estimation models, in: 2020 28th European Signal Processing Conference (EUSIPCO), 2021, pp. 830–834. [22](#)
- [224] J. P. Robinson, G. Livitz, Y. Henon, C. Qin, Y. Fu, S. Timoner, Face recognition: too bias, or not too bias?, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2020. [22](#)
- 1345 [225] J. G. Cavazos, P. J. Phillips, C. D. Castillo, A. J. O’Toole, Accuracy comparison across face recognition algorithms: Where are we on measuring race bias?, *IEEE transactions on biometrics, behavior, and identity science* 3 (1) (2020) 101–111. [22](#)
- [226] V. Albiero, K. Zhang, M. C. King, K. W. Bowyer, Gendered differences in face recognition accuracy explained by hairstyles, makeup, and facial morphology, *IEEE Transactions on Information Forensics and Security* 17 (2021) 127–137. [22](#)
- 1350 [227] Z. Babnik, V. Štruc, Assessing bias in face image quality assessment, in: 2022 30th European Signal Processing Conference (EUSIPCO), IEEE, 2022, pp. 1037–1041. [22](#)
- [228] J. Yu, X. Hao, Z. Cui, P. He, T. Liu, Boosting fairness for masked face recognition, in: Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops (CVPRW), 2021, pp. 1531–1540. [22](#)
- [229] Privacy measures of biometrics businesses, Technical report, NEC Technical Journal (2018). [22](#)
- 1355 [230] P. Bukaty, The California Consumer Privacy Act (CCPA): An Implementation Guide, IT Governance Ltd, 2019. [22](#)
- [231] 740 ILCS/14, Biometric Information Privacy Act (BIPA), Public act 095-994, Illinois General Assembly (2008). [22](#)