A Survey on Masked Facial Detection Methods and Datasets for Fighting Against COVID-19

Bingshu Wang, Jiangbin Zheng, and C. L. Philip Chen, *Fellow, IEEE*

***Abstract*—Coronavirus disease 2019 (COVID-19) continues to pose a great challenge to the world since its outbreak. To fight against the disease, a series of artificial intelligence (AI) techniques are developed and applied to real-world scenarios such as safety monitoring, disease diagnosis, infection risk assessment, and lesion segmentation of COVID-19 CT scans. The coronavirus epidemics have forced people wear masks to counteract the transmission of virus, which also brings difficulties to monitor large groups of people wearing masks. In this article, we primarily focus on the AI techniques of masked facial detection and related datasets. We survey the recent advances, beginning with the descriptions of masked facial detection datasets. A total of 13 available datasets are described and discussed in detail. Then, the methods are roughly categorized into two classes: conventional methods and neural network-based methods. The conventional methods are usually trained by boosting algorithms with hand-crafted features, which accounts for a small proportion. Neural network-based methods are further classified as three parts according to the number of processing stages. Representative algorithms are described in detail, coupled with some typical techniques that are described briefly. Finally, we summarize the recent benchmarking results, give the discussions on the limitations of datasets and methods, and expand future research directions. To our knowledge, this is the first survey about masked facial detection methods and datasets. Hopefully our survey could provide some help to fight against epidemics.**

***Impact Statement*—In the era of COVID-19, many AI techniques of masked facial detection have been proposed to determine whether one wears a mask, or provide masked face regions to help non-contact temperature measurement. However, it lacks of a review about these masked facial detection methods and datasets. In this survey paper, we review recent benchmarking efforts that primarily focus on the techniques of masked face detection to combat COVID-19. We have summarized thirteen open datasets and provided their available links that would help AI researchers and engineers use them quickly. We have presented several categories of representative techniques aimed for masked facial detection. Meanwhile, ten research directions have been identified to guide researchers for future research. It could offer a good reference for beginners, researchers and skilled AI engineers to develop more effective and efficient systems.**

***Index Terms*—Artificial intelligence, broad learning system, masked face datasets, masked facial detection, neural networks.**

I. Introduction

Since the first case was identified by COVID-19 in 2019, the coronavirus disease spread quickly and caused the outbreak all over the world in 2020 [1]–[3]. According to the data released by Ref. [4], by the end of December 8, 2021, more than 267.30 millions of humans have been identified by the COVID-19, with more being added every day. The coronavirus disease has caused more than 5.27 millions of deaths globally.

The COVID-19 epidemic has posed great challenge to the world. Artificial intelligence (AI) techniques are able to help people fight against the virus in many ways [5]–[10]. For example, detecting masked faces [11], [12], detecting COVID-19 patients [13], [14], assessing infection risks [15], building a disease monitoring and prognosis system [16], improving lesion segmentation of COVID-19 chest CT scans [17], etc. Among these techniques, this survey article primarily focuses on the techniques of masked facial detection.

Many doctors and epidemiologists have proved that wearing a mask is an effective means to counteract the spreading of coronavirus disease [18]–[20]. Detailed advice on the uses of masks was published by World Health Organization (WHO) [21]. As a consequence, people are suggested and even required by rules or laws to wear masks when entering public places. This brings demands to monitor large groups of people wearing masks. But it is not the goal of existing face detection methods that have been embedded in monitoring devices. To solve the problem, a series of masked facial detection methods and datasets have been proposed.

The objective of this article is to provide a detailed review of recent developments in the field of masked facial detection, in the hopes of providing reference or help for researchers and communities to develop more efficient and effective systems. Current methods employ hand-crafted features and neural networks to train detection models. In this survey article, we classify them according to the used feature and the number of processing stages. To our knowledge, this is the first survey about masked face detection methods.

Manuscript received May 11, 2021; revised September 10, 2021 and December 8, 2021; accepted December 25, 2021. Date of publication December 28, 2021; date of current version May 25, 2022. This work was supported in part by the National Natural Science Foundation of China, Youth Fund, under Grant 62102318, in part by the Fundamental Research Funds for the Central Universities under Grant G2020KY05113, in part by the National Key Research and Development Program of China under Grant 2019YFA0706200 and Grant 2019YFB1703600, in part by the National Natural Science Foundation of China under Grant 61702195, Grant 61751202, Grant U1813203, Grant U1801262, and Grant 61751205, and in part by the Science and Technology Major Project of Guangzhou under Grant 202007030006. This paper was recommended for publication by Associate Editor Pau-Choo Chung upon evaluation of the reviewers' comments. *(Corresponding author: C. L. Philip Chen.)*

Bingshu Wang and Jiangbin Zheng are with the School of Software, Taicang Campus, Northwestern Polytechnical University, Suzhou 215400, China (e-mail: wangbingshu@nwpu.edu.cn; zhengjb@nwpu.edu.cn).

C. L. Philip Chen is with the School of Computer Science and Engineering, South China University of Technology, Guangzhou 510641, China (e-mail: TCybernetics.EIC@gmail.com).

This article has supplementary downloadable material available at https://doi.org/10.1109/TAI.2021.3139058, provided by the authors.

Digital Object Identifier 10.1109/TAI.2021.3139058

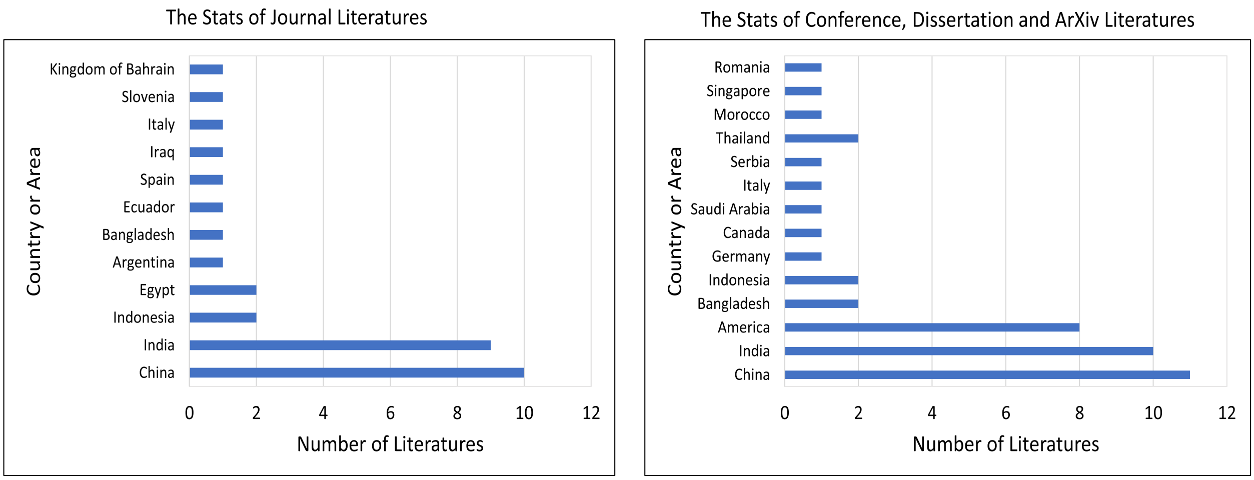

Fig. 1. Stats of state-of-the-art methods based on country or area of authors' affiliations. The literatures were surveyed by the end of September 1, 2021.

The aims of this review article are presented.

1) Describe the current open datasets of masked facial detection. Provide a detailed summary about the characteristics of datasets as well as the available links.
2) Present a division of masked facial detection methods. For each category of techniques, representative methods are outlined and commented.
3) Perform a comparison between different methods according to the results provided by the original literatures. Give discussions about the characteristics and limitations of methods and provide ten research directions in future.

The rest of this article is organized as follows. Section II presents the stats of related literatures in this article and how we surveyed literatures. Section III surveys the datasets of masked face detection. Details of 13 open datasets are outlined. Section IV gives the descriptions, main characteristics, and comparison analysis of masked facial detection methods. Limitations of datasets and methods and future research directions are discussed in Section V. Finally, Section VI concludes this article.

II. Stats and Analysis of Surveyed Literatures

Since the outbreak of COVID-19 epidemic, a series of works focus on how to use AI techniques to help fight against virus. The literatures of masked facial detection are springing up around the world. Many related international conferences were held with many solutions proposed for masked facial detection in recent two years. In this section, we shed light on the stats and analysis of state-of-the-art methods.

A. *Stats of Surveyed Literatures*

We surveyed literatures of masked facial detection by searching them in some large libraries or academic social websites such as Google Scholar, IEEEXplore, Elsevier, Springer, WebofScience, and ResearchGate. The searching key words are "masked face," "face mask," "masked facial" with the "document title" setting in the advanced search.

With hundreds of items obtained, all the searched journal papers [12], [22]–[51] are selected for review due to their detailed descriptions, experiments and discussions. Some conference papers are filtered out under the following conditions: 1) not written in English; 2) without experiments especially lacking of quantitative results; 3) unclear expressions or disordered organization; 4) without visual detection results shown; 5) number of images in dataset is too small, e.g., ≤ 500. Specially, a few literatures utilize very similar techniques and only test algorithms on different datasets. Only those with larger datasets and good performance are selected.

In total, more than 70 literatures are selected for this survey. They cover journal papers, conference papers, dissertations, and arXivs. In this article, we divide literatures into two classes for analysis: journal papers and conference papers. Particularly, dissertations and arXivs are assigned to conference class.

Stats is conducted based on two ways: Country or area of authors' affiliations; and published years. Fig. 1 outlines the number of papers for different countries or areas around the globe. For the stats of country or area of journal papers, it is clearly concluded that most of papers are proposed by Asia and Europe. China has published the largest number of journal papers with the ratio of 32.3%. The second largest is India with the ratio of 29.0%. These two Asia countries contribute to more than 60% journal papers. For the stats of country or area of conference papers, China and India are still the top two countries in accordance with the number of published literatures. American ranks third. More countries or areas bring out conference literatures than journal literatures.

For the stats of published years, Fig. 2 presents a direct representation. Before 2020, very few papers are published. In 2020, the number of literatures increases significantly, with 26% for journal class and 35% for conference class. Remarkably, in the first eight months of 2021, the ratio of journal literatures is much higher (71%) than that (26%) in 2020. Similar comparison is shown for conference literatures.

In summary, Asia countries take the lead in conducting the research and publish more papers than other countries or areas around the world. Since the large ratio of published literatures in

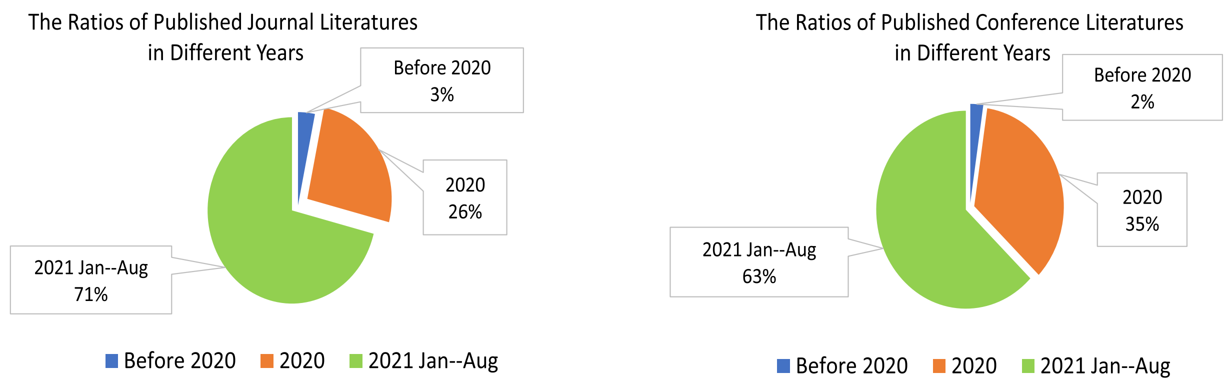

Fig. 2. Stats of state-of-the-art methods based on published date. Three parts are partitioned: before 2020, 2020, and 2021 Jan–Aug.

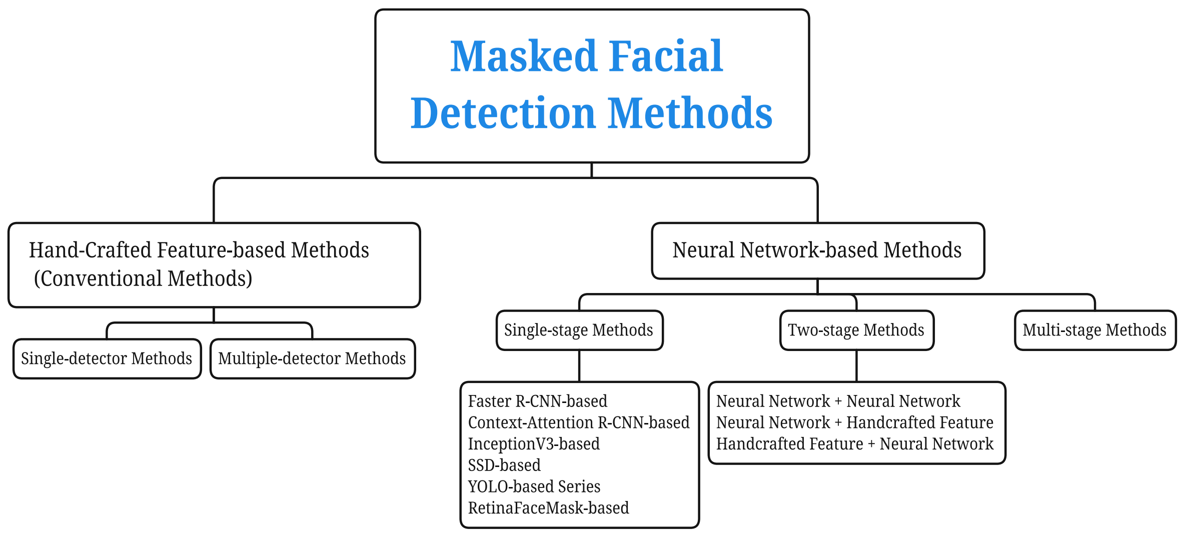

Fig. 3. Hierarchical representation of the state-of-the-art methods of masked facial detection.

2021 January–August, it is believed that more and more papers will come forth continuously.

B. Hierarchical Representation of Surveyed Literatures

To give a clear view of existing methods, a hierarchical representation is outlined in Fig. 3. According to the used features, all methods are divided into two classes: hand-crafted feature-based methods and neural network-based methods.

Hand-crafted feature-based methods are also usually regarded as conventional methods. They can be further classified as two categories in accordance with the number of detectors: single-detector methods and multiple-detector methods. Most detectors depend on AdaBoost algorithm. Different detectors, for example, face detector, facial mask detector, nose detector, mouth detector, nose and mouth detector, and eye detector, are selected or combined together. Details of hand-crafted feature-based methods are presented in Section IV.

Neural network-based methods attract many researchers' attentions. According to the number of stages, the methods can be classified into three categories: single-stage methods, two-stage methods, and multistage methods. For single-stage methods, they are mainly implemented by transfer learning of object detection algorithms. For example, YOLO series methods, YOLO, YOLOv2, YOLOv3, YOLOv4, YOLOv5, and corresponding tiny versions. For two-stage methods, they can be further divided into three kinds referring to the use of neural network: neural network + neural network, neural network + hand-crafted feature, and hand-crafted feature + neural network. Two-stage methods consist of two parts: face region predetection and face region classification. The former part is used to detect candidate facial regions, and the latter part is to classify the conditions of mask-wearing. For multistage methods, they include more and complex processing steps or make use of more than one models, which means more computation costs.

Notably, we also spend much time on the datasets of masked face detection, especially open-source datasets. Due to their accessibility, 13 datasets are reviewed in Section III.

III. Masked Facial Detection Datasets

To monitor the conditions of wearing masks, many datasets are proposed by researchers around the globe to train detection or classification models. These models will be deployed in monitoring systems or edge-nodes. In this section, detailed descriptions and discussions on these datasets are presented.

A. Description of Datasets

First, we present an earlier dataset about masked face detection. Ge *et al.* [52] proposed a large dataset called MAFA in 2017. It was claimed to be the largest wearing mask dataset before 2017. MAFA contains 30 811 images that are collected from the Internet, with 35 806 masked faces. The dataset is more likely to be an occluded face dataset because it covers many mask types; for example, man-made object with single color, hand, hair, neckerchief, medical mask, etc. The dataset is labeled with six attributes: location of face, location of eyes, location of mask, face orientation, occlusion degree, and mask type. It considers about 60 scenes of masked faces, and provides sufficient samples. However, many occlusions are noneffective to protect people from infection risks. This dataset is more suitable for occluded face detection. Preprocessing is required to reach the goal of wearing mask detection.

Wang *et al.* [53] created a Masked Face Detection Dataset (MFDD). The dataset only concentrates on single class: masked face. It has 4342 images with a total number of 24 771 masked faces. These images are captured from scenes of fighting coronavirus epidemics. They are divided into three sets in accordance with image size 256×256: equal to the size, smaller or larger than the size. The dataset can be used to train detection model to determine whether one wears a mask or not. However, it lacks of annotation information.

Cabani *et al.* [38] developed a MaskedFace-Net to generate simulated correct/incorrect masked faces called "MaskedFace-Net Image Dataset (MFNID)." The framework encompasses four steps: candidate face detection, facial landmarks detection, mask-to-face mapping, and manual image filtering. Original face images are derived from FFHQ dataset [54]. All the images have a fixed size of 1024×1024 and they are classified as two sets: Correct Masked Face Dataset (CMFD) and Incorrect Masked Face Dataset (IMFD). The authors presented a further division for IMFD: mask only covering nose and mouth (IMFD1), mask only covering mouth and chin (IMFD2), and mask only covering chin (IMFD3). The total number of this dataset is 137 016: 67 193 correctly masked (49%) and 69 823 incorrectly masked (51%) (IMFD1, IMFD2, IMFD3). This is a very large dataset in terms of image number. For each image, facial region accounts for a large ratio, making face detection easy. However, MFNID only contains one type of simulated mask and does not provide annotations.

Roy *et al.* [43] searched images from the Internet to build a dataset, namely Moxa3K. It consists of 3000 images. The dataset gives a careful consideration for boundary conditions; for example, if a face is covered by a handkerchief, it will be regarded as a "mask" class. Moxa3K includes a variety of samples such as blurred, rotated, crowded areas, and different illumination conditions. With 9161 faces and 2015 masked faces included, all the face regions are annotated by Pascal VOC format "LabelImg" [55] and YOLO format. Thus, it offers more choices for researchers to train their machine learning models. This setting is expected to improve the robustness of masked facial detectors.

Jiang *et al.* [50] proposed a properly wearing masked face detection (PWMFD) dataset. They collected 9205 images from several available datasets such as MAFA [52], MFDD [53], Wider Face [56], and the Internet. Although several datasets have their own annotations, PWMFD dataset provides uniform annotation manually for three classes "with_mask," "without_mask," and "Incorrect_mask". Specially, facial regions that are covered by other objects are labeled as "without_mask" so that trained models are not deceived. Face regions with nose uncovered are annotated as "Incorrect_mask" class. PWMFD dataset has 7695 "with_mask" faces, 10471 "without_mask" faces, and 366 "Incorrect_mask" faces.

Eyiokur *et al.* [57] proposed a Unconstrained Face Mask Dataset (UFMD) by collecting images from available datasets FFHQ [54], LFW [58], CelebA [59], Youtube videos, and the Internet. These publicly images allow UFMD be a complex dataset that covers ethnicity, age, gender, indoor, and outdoor scenarios. A large amount of head pose variations are also considered in UFMD, which helps improve robustness of masked face detectors. UFMD consists of 21 316 images with three classes: 10 618 images with masked faces, 10 698 images without masks, and 500 images with incorrect masks. The authors claimed that the website will be available soon.

Batagelj *et al.* [49] compiled a dataset called "Face-Mask-Label Dataset (FMLD)" by searching images from Wider Face [56] and MAFA [52] datasets. Real-world conditions are considered in FMLD: head pose, illumination, and image quality. Only when the faces are covered by nose, mouth, and chin, even the occlusions are something similar to a scarf or handkerchief, they are regarded as masked face class. Face samples are selected from Wider Face [56] to balance the classes, which requires a small size of 40 pixels for the height and width of each face, i.e., $min(width, height) > 40$. Thus, the face region size is not small. Incorrect masked faces are selected from those samples with nose uncovered in MAFA. Through inspecting samples carefully, a total number of 41 934 images (63 072 faces) are created in FMLD. It contains three classes of faces with labels: 32 012 faces without masks, 29 532 correct masked faces, and 1528 incorrect masked faces.

Dey *et al.* [60] created a dataset containing 4095 images that can be obtained from the available link in Table II. Most of the images have only one face. The images are selected from MFDD [53] and SMFD [61]. Dey's dataset consists of two classes: 1930 faces without masks and 2165 faces with masks. Head poses vary from frontal to profile. Most of the scenes are simple because face regions account for large ration in the whole image. However, annotations are not provided.

Singh *et al.* [48] generated a custom dataset manually, which includes 7500 images: 5191 training images, 1599 validation images, and 710 testing images. These images come from MAFA [52] and Wider Face [56]. Singh's dataset is labeled by two classes: "face" and "face_mask," which aims to train a model to determine whether one wears a mask or not. The detection results can be used to analyze the crowing extent. Bounding boxes are provided as annotations.

Wang *et al.* [44] proposed a Wearing Mask Detection (WMD) dataset with 7084 images. Most of the images are collected from the scenarios of combating COVID-19 in China, which allows the dataset be real-world scenarios. The dataset has a total number of 26 403 masked faces: 17 654 for train, 1936 for validation, and 6813 for test. It should be noted that for the test set is divided into three parts according to the difficulty of detection task and number of masked faces in one image: DS1, DS2, DS3. Every image in DS1 has only one masked face with a relative big size. Every image in DS2 has two to four masked faces. For DS3, over five masks are included in each image and the distance from face to camera is long (>2 m). Thus, the difficulty varies from easy to difficult for the three sets. In addition, the authors also present a self-built face detection dataset, which has 4054 images with 16 216 faces. Coupled with WMD, these datasets can be utilized together to train models of detecting the conditions of wearing masks.

Moreover, there are some datasets proposed with available links such as AIZOOTech [62], Kaggle [63], and SMFD [61]. The images of AIZOOTech [62] dataset are from MAFA [52] and Wider Face [56] datasets. The total number of images is 7959: 4034 masked faces and 12 620 faces. Notably, all selected images belong to scenes with medium-level difficulty.

Kaggle [63] dataset has three classes: faces without masks, correct wearing masks, and incorrect wearing masks. It consists of 853 images in total: 3232 faces with masks, 717 faces without masks, and 123 incorrect masked faces.

SMFD dataset was proposed by Prajnasb [61] and simulated totally by matching masks to faces. All the original images are captured from Web. It has two categories of faces with annotations: 690 with masks and 686 without masks. The head pose is from frontal to profile and the size of facial region is big. All these elements lead to a simple scene.

In summary, detailed information for the abovementioned datasets is illustrated in Table I. The corresponding available links are also provided in Table II. All the links had been verified to be effective before May 10, 2021.

B. *Discussions of Datasets*

In the previous section, we elaborated on datasets and their details of characteristics. Discussions about these datasets will be presented from four parts: image sources, reality of images, classes imbalance, and existing experimental results.

1) Image Sources: Almost all the datasets are created by collecting images from the Internet. A typical representative is MAFA [52], which is proposed as an earlier work. Most of the images in MFDD [53], WMD [44], Kaggle [63], and SMFD [61] are built through Internet search.

Some faces without masks are from some face datasets such as FFHQ [54] and Wider Face [56]. FFHQ is widely used in MFNID [38] and UFMD [57].

The masked face dataset MAFA [52] and face dataset Wider Face [56] are widely employed to create new masked face detection datasets such as PWMFD [50], FMLD [49], Singh's Dataset [48], and AIZOOTech [62]. This can give a good explanation about the high similarity between Singh's Dataset [48] and AIZOOTech [62].

Some datasets like PWMFD [50] and Dey's Dataset [60] are the combinations of several existing datasets. In realistic applications, combination of multiple datasets is an alternative way to build up a required dataset quickly. Thus, it is suggested for researchers to use this way to create their own datasets. Meanwhile, capturing a variety of images from the Web is beneficial to enrich the varieties of datasets.

2) Reality of Images: It is also notable from Table I that 9 of 13 datasets are constructed by real-world images. MFDD [53] and Dey's dataset [60] include both real and simulated images. For MFNID [38] and SMFD [61] datasets, the masked faces are created entirely by simulating images. Some samples are given in Fig. 4. Only one type of mask is used to synthesize masked faces in MFNID or SMFD.

Creating simulated samples requires a mask-to-face mapping technique. Large size of faces are always selected to synthesize masked faces because their landmarks can be located well, which helps generate proper samples. However, for small size of faces, it is hard to realize a good mapping due to the inaccurate landmarks and head pose variations. In addition, the number of mask types is inadequate. These factors allow masked face detection to be a simple problem. This has been verified by the method [44], which achieves an accuracy of 99.9% for incorrect masked faces on 4500 images randomly selected from MFNID [38].

In people's daily life, there are diverse mask types. It is not easy to collect enough images with a variety of masked faces. In this case, synthesizing samples can be regarded as a good choice to address this issue [64]. It illustrates that real mask looks more natural than the simulated masks. More details of synthesizing images are provided in supplementary materials. Another method of converting face dataset to masked dataset can be found in [65]. How to generate more natural masked faces is an interesting research in future.

3) Classes Imbalance: It is pretty clear that classes imbalance problem exists in the field of multiple categories of object detection. Table III sheds light on the problem for datasets PWMFD [50], UFMD [57], FMLD [49], and Kaggle [63]. High ratios of classes are denoted as "head classes," and low ratios of classes are denoted as "tail classes." Obviously, the ratios of incorrect face_mask in Table III are smaller than 3.1%. It implies that class distribution is extremely imbalanced. If a dataset with classes imbalance is used to train a model, it will easily lead to erroneous detections. The reason is that head classes can be learned well while tail classes are not learned well, as shown in Fig. 5.

How to solve the problem? Actually, it is not easy to obtain the incorrect or improperly masked faces. Two ways are suggested to solve the problem. One way is to collect images as many as possible from available datasets. The other way is to simulate images like MFNID [38].

4) Existing Experimental Results: Table IV shows original results of some methods on their own datasets. Herein, we first give some common evaluation metrics: $Recall$, $Precision$, $F1$,

TABLE I
DETAILED DESCRIPTIONS OF SOME OPEN DATASETS FOR MASKED FACIAL DETECTION

Dataset Name	Main Characteristics	Image Reality	Image Number	Category	Masks Number	Scale	Head Pose	Scene	Annotation	Open
MAFA [52]	All the images are from the Internet. Six attributes are manually annotated for each face region. More like occluded faces dataset.	Real	30811	Multiple mask types	35806 masked faces	Medium Large	Various	Complex	Yes	Yes
MFDD [53]	The images are from the Internet. Some images are collected from the scenarios of fighting against COVID-19.	Simulated Real	4342	One	24771 masked faces	Small Medium Large	Various	Complex	No	Yes
MFNID [38]	Face images are from FFHQ. All the masks are simulated by proposed MaskedFace-Net. It includes three classes of incorrect masked faces.	Simulated	137016	Two	67193 faces with correct masks; 69823 faces with incorrect masks	large	Frontal	Simple	No	Yes
Moxa3K [43]	The images are captured from Kaggle data set that are captured from Russia, Italy and China, India during the ongoing pandemic.	Real	3000	Two	9161 faces without masks; 3015 masked faces	Small Medium Large	Various	Complex	Yes	Yes
PWMFD [50]	Over half of the images are collected from WIDER Face, MAFA, RWMFD. “With mask” class requires faces with nose and mouth covered.	Real	9205	Three	10471 faces without masks; 7695 correct masked faces; 366 incorrect masked faces	Small Medium Large	Frontal to Profile	Medium	Yes	Yes
UFMD [57]	The images are captured from FFHQ, CelebA, LFW, YouTube videos, and the Internet. It covers ethnicity, age, gender, head pose variations.	Real	21316	Three	10698 faces without masks; 10618 correct masked faces; 500 incorrect masked faces	Large	Frontal to Profile	Medium	Yes	Soon Open
FMLD [49]	The images are from MAFA and Wider Face datasets. The annotations with a list of images publicly available are provided.	Real	41934	Three	32012 faces without masks; 29532 correct masked faces; 1528 incorrect masked faces	Medium Large	Various	Complex	Yes	Yes
Dey’s Dataset [60]	The images are real wearing masks and they come from Kaggle datasets, RMFD dataset and Bing Search.	Simulated Real	4095	Two	2165 images with masks; 1930 images without masks	Large	Frontal to Profile	Simple	No	Yes
Singh’s Dataset [48]	The dataset includes MAFA, WIDER FACE and captured images by surfing various sources.	Real	7500	Two	5191 training images; 1599 validation images; 710 testing images	Small Medium Large	Various	Complex	Yes	Yes
WMD [44]	Most of the images are collected from real scenarios of fighting against CoVID-19. It covers many long-distance scenes.	Real	7804	One	26403 masked faces	Small Medium Large	Various	Complex	Yes	Yes
AIZOO -Tech [62]	The dataset is created by modifying the wrong annotations from datasets of WIDER Face and MAFA.	Real	7959	Two	12620 faces without masks; 4034 masked faces	Small Medium Large	Various	Medium	Yes	Yes
Kaggle [63]	The images are all from the Internet for training two-class models.	Real	853	Three	717 faces without mask; 3232 correct masked faces; 123 incorrect masked face	Small Medium Large	Various	Complex	Yes	Yes
SMFD [61]	All the images are web-scrapped.	Simulated	1376	Two	686 faces without masks; 690 masked faces	Large	Frontal to Profile	Simple	Yes	Yes

Fig. 4. Some samples selected from the datasets in Table I. These samples are the representatives of different datasets.

TABLE II
AVAILABLE WEBSITES OF OPEN-SOURCE DATASETS

Dataset Name	Available Link	Access Date
MAFA [52]	https://drive.google.com/drive/folders/1nbtM1n0–iZ3VVbNGhocxbnBGhMau_OG	March 2, 2021
MFDD [53]	https://github.com/X-zhangyang/ Real-World-Masked-Face-Dataset	November 26, 2020
MFNID [38]	https://github.com/cabani/MaskedFace-Net	February 22, 2021
Moxa3K [43]	https://shitty-bots-inc.github.io/MOXA/index.html	April 22, 2021
PWMFD [50]	https://github.com/ethancvaa/Properly-Wearing-Masked-Detect-Dataset	April 22, 2021
UFMD [57]	https://github.com/iremeyiokur/COVID-19-Preventions-Control-System	August 30, 2021
FMLD [49]	https://github.com/borutb-fri/FMLD	April 23, 2021
Dey Dataset [60]	https://github.com/chandrikadeb7/Face-Mask-Detection	April 23, 2021
Singh Dataset [48]	https://drive.google.com/drive/folders/1pAxEBmfYLoVtZQlBT3doxmesAO7n3ES1?usp=sharing	April 24, 2021
WMD [44]	https://github.com/BingshuCV/WMD	April 29, 2021
AIZOO -Tech [62]	https://github.com/AIZOOTech/FaceMaskDetection	December 23, 2020
Kaggle [63]	https://www.kaggle.com/andrewmvd/face-mask-detection	April 22, 2021
SMFD [61]	https://github.com/prajnasb/observations	December 27, 2020

TABLE III
NUMBERS AND RATIOS OF DIFFERENT CLASSES FOR SEVERAL DATASETS

Dataset Name	Face	Correct Face_mask	Incorrect Face_mask
PWMFD [50]	10471 (56.50%)	7695 (41.52%)	366 (1.97%)
UFMD [57]	10698 (49.04%)	10618 (48.67%)	500 (2.29%)
FMLD [49]	32012 (50.75%)	29532 (46.82%)	1528 (2.42%)
Kaggle [63]	717 (17.61%)	3232 (79.37%)	123 (3.02%)

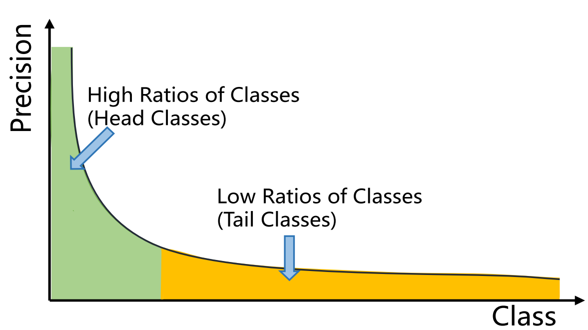

Fig. 5. Distribution example of detection precisions for head classes and tail classes.

$Accuracy$, AP, and mAP. They are defined as follows:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{1}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{2}$$

$$F1 = 2 \times \frac{\text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \tag{3}$$

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{4}$$

where TP represents the true positives, TN represents the true negatives, FP represents false positives, and FN represents false negatives. "Accuracy" represents the whole detection rate

$$\text{IoU} = \frac{|P \bigcap G|}{|P \bigcup G|} \tag{5}$$

where IoU means the overlap between predicted box P and ground truth box G. The term $\bigcap$ is defined as intersection, and $\bigcup$ is defined as union between two boxes.

The average precision (AP) is defined in (6) to evaluate the performance of object detection methods. It is calculated by finding the area under the Precision–Recall curve

$$\text{AP}_{\text{class}} = \int_0^1 P(r)dr \tag{6}$$

where "class" represents the object classes such as "face," "masked face," and "incorrect masked face," etc. mAP is the mean AP, as shown in the following:

$$\text{mAP} = \frac{1}{n} \sum_{k=1}^{k=n} \text{AP}_k. \tag{7}$$

It can be clearly concluded from Table IV that methods [57] achieve higher accuracy and mAP values on the given datasets. This also verifies the description in Table I that scenes of UFMD and Dey's datasets are simple. The method [49] largely gets benefit from two deep neural networks: RetinaFace and ResNet-152. Other datasets such as MAFA, Moxa3K, PWMFD, Singh's, and WMD are challenging in terms of quantitative results. As a consequence, these results can be treated as benchmarks for future comparison.

IV. MASKED FACIAL DETECTION METHODS

In this section, we primarily focus on masked facial detection methods. According to the used features, the methods can be divided into hand-crafted feature-based methods and neural network-based methods. Hand-crafted feature-based methods are regarded as conventional methods. Specially, neural network-based methods are sprouting up and they have achieved impressive and excellent results. Considering the high proportion of neural network-based methods, we classify them into three parts based on processing stages: single-stage methods,

TABLE IV
RESULTS OF ORIGINAL METHODS ON THEIR OWN AVAILABLE DATASETS

Literatures	Methods or Networks	Datasets	Results
Ge et al [52]	LLE-CNNs	MAFA	AP=76.4%
Roy et al [43]	SSD, Faster R-CNN, YOLOv3, YOLOv3Tiny	Moxa3K	SSD mAP=46.52%, Faster R-CNN mAP=60.5%, YOLOv3 mAP=63.99%, YOLOv3Tiny mAP=56.57%
Jiang et al [50]	Squeeze and Excitation-YOLOv3	PWMFD	Image size 608 × 608: AP=73.7%
Eyiokur et al [57]	InceptionV3, ResNet-50, MobileNetV2, EfficientNet-b3	UFMD	Three classes Accuracy: InceptionV3 98.28%, ResNet-50 95.44%,MobileNetV2 98.10%,EfficientNet-b3 98.00%
Batageli et al [49]	RetinaFace, ResNet152	FMLD	mAP=90.75 ± 0.99
Dey et al [60]	MobileNetV2	Dey's Dataset	IDS1 700 real images, Accuracy=93%, IDS2 276 simulated images, Accuracy=100%
Singh et al [48]	YOLOv3, Faster R-CNN	Singh's Dataset	YOLOv3 AP=55%, Faster R-CNN AP=62%
Wang et al [44]	Faster R-CNN, BLS	WMD	Recall=93.54%, Precision=94.84%, F1=94.19%

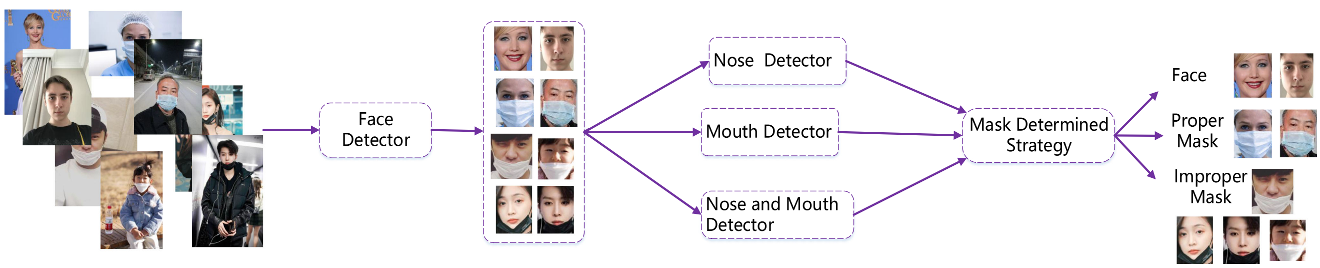

Fig. 6. Flowchart for some conventional methods aiming to identify wearing mask conditions. Several detectors are trained by self-built datasets or provided by Open-Source Computer Vision Library (OpenCV) with its link: "https://opencv.org/".

two-stage methods, and multistage methods. Detailed descriptions are given as follows.

A. Conventional Methods

Conventional face detection methods have been invested very well in past decades [66]. A face detector proposed by Viola and Jones [67] is trained by AdaBoost algorithm, which is the basis for face detection. Common hand-crafted features include Haar-like [67], local binary pattern (LBP) [70], histogram of orientation (HOG) [71], etc.

In this section, we mainly focus on masked face detection using conventional methods. Some published literatures recently are usually designed by hand-crafted features and boosting learning algorithms [30], [69], [72]–[76]. Most of the conventional methods for masked face detection are based on the observation that if one wears a mask well, the nose or mouth cannot be detected, and vice versa. One typical flowchart for conventional methods is shown in Fig. 6. One or several detectors are trained by self-built datasets or provided by OpenCV. Mask determined strategy is exploited to judge the mask-wearing conditions. According to the number of detectors, conventional methods can be divided into two parts: single-detector methods and multiple-detector methods.

Single-Detector Methods: Dewantara and Rhamadhaningrum [72] exploited to train a nose and mouth classifier to detect multipose masked faces. The authors create a dataset of nose and mouth. Haar-like, LBP, and HOG features are exploited for training models, respectively. If nose and mouth is not detected, the candidate facial region will be labeled "masked." Otherwise, it will be labeled "No mask." It is reported that the trained classifier of nose and mouth achieves an accuracy of 86.9% using Haar-like features, outperforming LBP and HOG. Obviously, there is further space to improve accuracy.

Multiple-Detector Methods: Petrovic and Kocic [73] developed an indoor safety IoT system which adopts multiple AdaBoost cadcade-classifiers. These classifiers are provided by OpenCV to detect frontal face, nose, and mouth, respectively. For a candidate face region, if no mouth and no nose is detected, it will be regarded as wearing a mask properly. If nose is detected, it will be labeled as "improper mask." If mouth is detected, it will be labeled as "no mask." This approach may work well in the access control system by OpenCV classifiers. However, it depends on OpenCV classifiers too much, and it does not provide details about accuracy.

Unlike methods in [73], Nieto-Rodriguez *et al.* [69] used two AdaBoost detectors to implement surgical mask detection. One detector is trained by LogitBoost for face detection, and the other is trained by GentleAdaBoost for mask detection. Then, two color filters in the HSV color space are employed to eliminate false positives. Considering the overlapping regions, cross-class removal strategy is designed to keep the region with higher confidence. The method is easy to implement and it achieves an accuracy of 95% on 496 faces and 181 masks. The process is illustrated in Fig. 7.

Fang *et al.* [75] developed a real-time system of masked facial detection that uses Haar-like features for face detection and mouth detection, respectively. Similar with [73], face region is

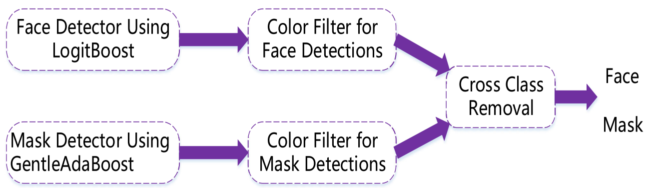

Fig. 7. Two AdaBoost classifiers used to detect faces and masks.

first located, and then mouth detection is used to determine the mask-wearing conditions. The designed algorithm is claimed to run on PYNQ-Z2 SoC platform with 0.13 s response of facial mask detection and 96.5% accuracy on given dataset.

In addition, He [76] employed skin color and eye detection to reach the goal of wearing mask detection. The first step is to locate face region using ellipse skin model and geometric relationship between eyes and other facial parts. Then, the coverage of skin color in the bottom half of facial region is calculated to judge mask-wearing conditions. However, this method can only be applied to specific scenes.

In summary, masked face detection methods based on AdaBoost algorithm and Haar-like features are typical conventional methods. They can work well for close-distance scenes that have evident features of face regions. However, due to limited learning ability, it is hard for these classifiers to adapt to complex scenes such as long distance and illumination changes. Neural network-based methods are data-driven framework that may provide feasible solutions.

B. Single-Stage (End-to-End) Methods

Single-stage methods based on deep learning techniques account for the largest proportion among the methods. They include Faster R-CNN [77], Context-Attention R-CNN [47], InceptionV3 [78], MobileNet [60], SSD [79], YOLO [80], YOLOv2 [42], YOLOv3 [50], YOLOv4 [51], YOLOv5 [85], and others [11], [41], [46], [52], [89]–[94]. It can be clearly concluded that YOLO and its variants are used widely. Representative methods are presented as follows.

Faster R-CNN-Based: Razavi *et al.* [77] employed Faster R-CNN structure to detect people who do not wear a mask or do not maintain a safety distance. It was applied to several road maintenance projects for monitoring workers, ensuring them wear masks and keep proper physical distance. However, the dataset is limited and it only focuses on construction scenes. Meivel *et al.* [23] used Faster R-CNN algorithm for mask detection and social distance measurement. This method achieves 93.4% accuracy for complex scenes such as facial poses, beard faces, multiple mask types, and scarf images. Notably, the effects need improvement when converting surveillance images into bird-view images.

Context-Attention R-CNN-Based: Zhang *et al.* [47] developed a new framework for masked facial detection called Contex-Attention R-CNN, which consists of multiple context feature extractor component, decoupling branches component, and attention component. It is able to enlarge intraclass difference and reduce interclass difference through extracting distinguishing features. They also created a dataset that includes 8635 faces with different conditions for experimental verification. The framework can achieve $mAP = 84.1\%$ on the given dataset, 6.8% higher than that of Faster R-CNN with ResNet-50. However, the dataset is classes imbalanced.

InceptionV3-Based: Chowdary *et al.* [78] exploited InceptionV3 pretrained model to classify one whether wears a mask or not. The last layer of InceptionV3 is replaced by five layers, which is regarded as a transfer learning model. It is reported to reach a 99.9% on a simulated dataset.

MobileNet-Based: Dey *et al.* [60] proposed a MobileNetMask to prevent the transmission of SARS-COV-2, which is a deep learning method of multiphase facial mask detection. The mask classifier depends on the ROI detection of SSD and ResNet-10. Due to the minimal processing capability and lightweight mobile-oriented model, MobileNet-V2 is a good selection for embedded systems. It is reported to achieve higher accuracy than other methods.

SSD-Based: Deng *et al.* [79] introduced attention mechanisms, inverse convolution, and feature fusion to SSD structure for the task of wearing mask detection. It achieves an mAP of 91.7%, outperforming SSD with 85.4% mAP.

YOLO-Based: Wang *et al.* [80] proposed a holistic edge-computing framework to detect masked faces. It is a serverless in-browser solution by integrating YOLO, CNN inference computing, and WebAssembly techniques. This design minimizes extra devices. It has easy deployment, low computation costs, fast detection speed, and achieves mAP=89%.

YOLOv2-Based: Loey *et al.* [42] developed a YOLOv2 with ResNet-50 detector for medical face mask detection. The method includes two parts. The first is designed by deep transfer learning for feature extraction. The second part is implemented by YOLOv2 for masked face detection. Specially, mean IoU is introduced to estimate the best number of anchor boxes and it can improve the accuracy. The method achieves AP=81% on a dataset with 1415 images.

YOLOv3-Based: Jiang *et al.* [50] designed Squeeze and Excitation (SE) YOLOv3 to balance the effectiveness and running speed for masked facial detection. It introduces SE into Darknet-53 as attention mechanism integration to extract essential feature, and adopts GIoUloss, focal loss to enhance stability and robustness. A new dataset called PWMFD Dataset is created for three categories of masked faces. It is reported that the method achieves mAP= 73.7% for 608×608 size of images. The method is expected to used in access control gate system and noncontact temperature measurement. However, the similarity between incorrect masks is high. It may bring confusions that masks only covering chin are regarded as without mask. Prusty *et al.* [26] proposed a data augmentation technique to expand dataset size. New dataset is used to train YOLOv3 model for masked facial detection. Average accuracy is more than 93% on given three datasets. However, only two kinds of data augmentation techniques (grayscale and Gaussian blur) are used. The number is very limited.

YOLOv4-Based: Kumar *et al.* [51] explored to test original and tiny variants of YOLO on a new face mask detection dataset, which encompasses 52 635 images. For the dataset, over 50 k labels are provided. Modified tiny YOLOv4 is recommended as an effective and efficient masked face detector because of its optimized feature extraction network. Yu and Zhang [31] improved YOLOv4 model by introducing a modified CSPDarkNet53 to reduce computation costs and enhance learning ability. An adaptive image scaling algorithm is designed to reduce redundancy and an improved PANet structure is used to learn more semantic information. It is reported to achieve 98.3% accuracy with 54.57 fps under the running environment of Windows 10, Inter(R)i7-9700 k and RTX 2070Super. One limitation is inconsideration of insufficient lighting samples.

YOLOv5-Based: Sharma [85] developed a model that uses YOLOv5 to detect whether one person is wearing a mask or not. However, if an individual does not face the camera, its performance will decrease. This is the method's limitation. Yang *et al.* [87] applied YOLOv5 in the supervision of wearing mask conditions. The authors design a man–machine interface for application and set the identifying time for 2 s with the consideration of complex scenes. A 97.9% recognition rate is achieved on the dataset [62]. It seems the response time is a bit longer. Ieamsaard *et al.* [88] tested the performance of YOLOv5-based model with 300 epochs, outperforming those models with less than 300 epochs.

RetinaFaceMask-Based: Jiang and Fan [11] proposed RetinaFaceMask for masked face detection, which is based on RetinaFace [95]. RetinaFaceMask is a single-stage detector. Its principle is to employ feature pyramid network to fuse high-level semantic information. A novel context attention module is presented to help RetinaFaceMask focus on the features of faces and masks. Moreover, a cross-class removal algorithm is proposed to remove those regions with low scores and high IoU values. Experiments demonstrate that RetinaFaceMask outperforms RetinaFace [95] in Recall and Precision.

Moreover, there are more experimental comparisons between methods. Singh *et al.* [48] utilized two object detection models, named Faster R-CNN and YOLOv3, for masked facial detection. They presented the comparison from visual and quantitative views, and gave detailed discussions about the application. Faster R-CNN outperforms YOLOv3 in the accuracy; however, for real-time application, it would be preferred to use YOLOv3, which runs faster than Faster R-CNN. The selection of model depends on the environment conditions. Similar conclusion is drawn in [96]. Roy *et al.* [43] used SSD, Faster R-CNN, YOLOv3, and YOLOv3Tiny to cope with the challenges of wearing medical mask detection. These methods are tested on Moxa3K dataset. Experimental results demonstrate that YOLOv3Tiny is the most suitable method for real-time inference among the methods.

In summary, object detectors such as Faster R-CNN and YOLO series attract more researchers' attentions, especially YOLOv3, YOLOv4, and YOLOv5. Tiny YOLO-based detectors with light-weighted models are expected to be deployed on real-time processing devices. Improved face detectors like RetinaFaceMask are also promising techniques. By transfer learning strategy, existing object detectors and face detectors can be applied for masked facial detection.

C. Two-Stage Methods

Two-stage methods mainly encompass two stages: face predetection and face class verification. The face predetection stage is usually implemented by many face detectors [66], [95], [97]–[102] or object detectors [103]–[109], etc. Notably, object detectors can also provide feature descriptors for candidate faces in the first stage. The second stage is designed by various classifiers or models [39], [110]–[114]. Its aim is to determine whether one wears a mask, correctly or incorrectly. The combination of object detector and classification model can realize masked face detection task.

According to the used features in literatures, two-stage methods can be divided into three groups: neural network + neural network [34], [36], [37], [44], [49], [115]–[117], neural network + hand-crafted feature [24], and hand-crafted feature + neural network [22], [28], [33], [45], [121]–[123].

Neural Network + Neural Network: One representative example refers to the method [44]. The first stage is designed by a deep learning transfer model: Faster R-CNN [103] and the second stage are designed by broad learning system (BLS) [110]. Input image is sent to the predetection stage. Then, many candidate regions are generated and they are further classified by trained BLS model which can remove false positives and keep masked faces. Finally, detected results are generated with labels. To train predetection model, annotated dataset is required, which is created using a tool called "LabelImg" [55]. The extracted faces and masks can be used to create classification datasets that are problem-dependent, for example, with/without mask, correct/incorrect mask.

The predetection in Faster R-CNN structure mainly includes four steps: extract feature maps, generate proposals by region proposal networks, obtain fixed dimension of feature map, and object classification and location regression. Faster R-CNN has advantages over SSD and YOLO in accuracy [96]. The verification stage employs BLS, which is a flat neural network structure with a very high training efficiency [110] and many variants have been proposed [125]. In practice, when a BLS model cannot learn a task well, one effective way is to add feature nodes that is called incremental learning. This ensures efficiency in training phase. It does not need to retrain from the scratch [44]. The combination of Faster R-CNN and BLS is verified to be effective on WMD dataset [44]. It achieves 97.32% accuracy for simple scene and 91.13% for complex scene. BLS can be a good selection for classification when training efficiency and small size of model are required in applications. Detailed descriptions can be found in supplementary materials.

Neural Network + Hand-Crafted Feature: Loy *et al.* [12] developed a hybrid method of deep learning and machine learning to detect facial mask. It includes two components (or stages): ResNet-50 is used as feature extractor, and SVM, decision tree, ensemble method are used as classification models. The authors claimed that SVM classifier achieves testing accuracy of 99.49%

in SMFD dataset [61], outperforming decision tree and ensemble method.

Similar with [12], the methods [118] also choose SVM as the classifier in the second stage. Buciu [118] took the ratio of color channels into account to discriminate mask and no-mask images. SSD is used to locate the positions of faces. Then, the lower part of face is considered to construct feature vector called color quotient feature, which will be classified by SVM model. A recognition rate of 97.25% is obtained. However, this method is sensitive to mask types, which is its potential weakness. Oumina *et al.* [119] presented several combinations of multiple CNNs and K-NN or SVM, and conducted experiments. It indicates that the combination of MobileNetV2 and SVM achieves the best performance among the combinations, 97.11% accuracy. More tests for the approach should be conducted on bigger datasets.

Zereen *et al.* [120] developed a two-stage approach to detect masked face and monitor the rule violations. It is based on the extraction of facial landmark. It first determines whether the target wears a multicolor mask or not by MTCNN, and second, it determines whether the target wears a skin-color mask or not. The method aims to detect five types of facial images, including no mask, beard and mustache, one-color-mask, multicolor mask, and skin-color mask. It achieves an accuracy of 97.13% and overcomes the problem of various-color mask detection, especially differentiates wearing skin-colored masks. However, the use of several techniques needs more computation costs, and the setting of empirical thresholds limits its adaptation ability.

Hand-Crafted Feature + Neural Network: Lin *et al.* [22] combined a sliding window algorithm with a modified LeNet (MLeNet) to locate masked faces. To improve performance with a small dataset, horizontal reflection is used to learn MLeNet via fine-tuning. MLeNet can be trained fast under CPU mode. It makes sense for real-world applications. However, sliding-window algorithm requires more computations for large size of images, which restricts its performance.

Rudraraju *et al.* [122] combined Haar-like cascade-classifiers and two MobileNet models for face mask detection. First, face regions are detected by Haar-like cascade-classifier. The first MobileNet model is used to classify masks and no masks. The second MobileNet model is used to distinguish correct or incorrect wearing masks. Experiments show that the system achieves around Accuracy $= 90$. It is expected to be deployed at fog gateway.

Tomás *et al.* [33] have also chosen Haar-like cascade classifier for rapid facial detection. CNN with transfer learning is used to determine whether one wears a mask or not. Multiple models are trained based on one dataset. VGG16 achieves the best performance with 0.834 accuracy, but its model size is also the largest. For deploying mobile device, MobileNetV2, with 0.812 accuracy, is selected as the classification model because it demands less computation costs and smaller storage. However, this method needs to be improved when detecting masked facials with alterations and sides.

In summary, most of two-stage methods are the combination of face detector and classification model. In many situations, predetection model and classification model are trained separately, which might require more time than those of single-stage methods. However, two-stage methods have advantages in coping with small object detection, multiclass classification, and cross classes removal. The combinations of "neural network + neural network" and "hand-crafted feature + neural network" are attached more importance, and they provide feasible solutions to solve real-world problems.

D. Multi-Stage Methods

Multistage methods always consist of multiple processing steps [32], [35], [40], [128]–[131]. For example, human detection or face region detection, ROI extraction or feature vector extraction, normalization, classification or prediction by sequences, and so on. Alternatively, multistage methods can be constructed by different combinations of those components.

The main idea of methods [128] is based on human posture estimation. First, a certain number of key points for one person are estimated. Then, some key points in face regions are analyzed to extract ROI from original image. After that, the ROI is normalized and sent to a trained classifier to predict class. In practice, some additional operations may be required to enhance performance.

Fig. 8 shows the process of the method [128]. It mainly includes five stages.

1) Human detection and location is implemented by YOLOv4 [108]. YOLOv4 is able to generate a series of candidates with a good tradeoff between speed and accuracy in the field of object detection.
2) Human pose is estimated by HRNet [132]. About 18 key points are generated for each individual. This is can be found in Fig. 8(a) and (b).
3) Face ROIs are determined by the points belonging to eyes and nose. Only those key points with higher confidence (>0.8) are selected to determine valid faces. Meanwhile, the size of valid faces is restricted by 20×20. Too small ROIs will be removed.
4) With valid faces obtained, they are classified by a transfer learning model $ResNet101 \times 1$ [133]. The model is trained on a data augmentation dataset.
5) For each person, it is assigned with an ID. DeepSort [134] is used to store some statistics. For each frame, the predicted label will be inserted into a buffer when the label's score is higher than 0.8. The final label is estimated from the buffer (size>3) by the most frequent label, i.e., majority voting.

YOLOv4 is trained on 1370 images containing face and masked face classes, and it achieves an mAP of 85.92% (IoU$=0.5$) on nearly 900 validated images. However, it does not perform well (mAP$=40.3\%$) on images with small size and low resolution. For classification, $ResNet101 \times 1$ reaches 99% for both classes: face and masked face.

The method proposed by Lin *et al.* [129] contains five stages: image data collection, human posture parsing, ROI selection, image normalization, and classification of masked face. Among these stages, human posture parsing is implemented by Openpose [135] that generates 25 key points for one individual. Five key points belonging to face region are used to extract ROI for

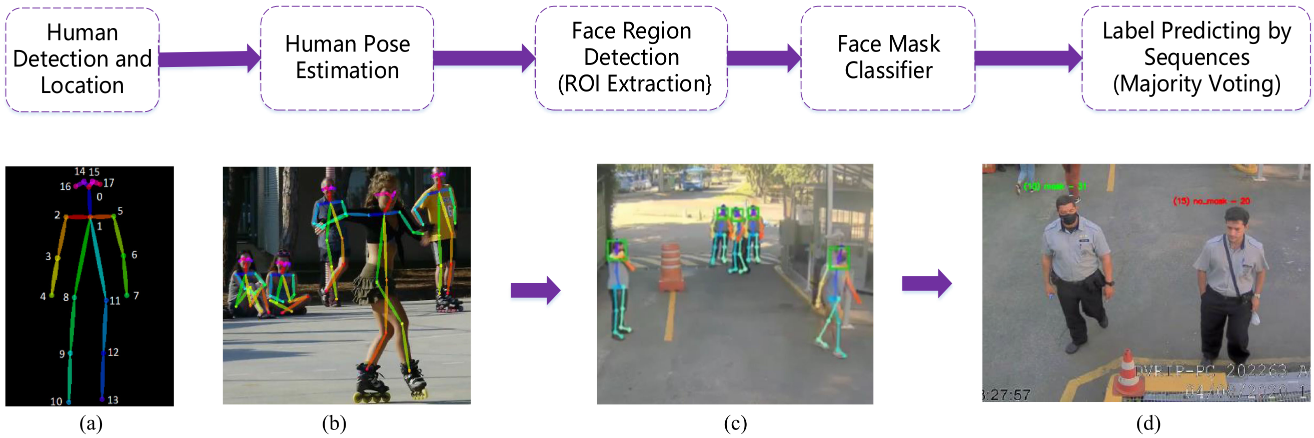

Fig. 8. Example of multistage method for masked facial detection [128]. (a) Keypoints joints set. (b) 2-D posture estimation. (3) Face region extraction. (d) Face/masked face classification and labeling.

image normalization. Then, the normalized image is classified by a face mask recognition network. It is reported that the method obtains 95.8% and 94.6% accuracy in daytime and nighttime, respectively.

Unlike [128], Qin *et al.* [40] proposed a multistage method including four steps: image preprocessing, face detection and cropping [98], image super-resolution, and wearing condition identification of face mask. The distinctiveness of this article is the introduction of super-resolution network (SRNet) in [40]. The goal of SRNet is to enhance face image. It helps improve the accuracy of subsequent classification network of mask-wearing condition. The method is claimed to achieve an accuracy of 98.7%. With the use of SRNet, it outperforms conventional deep learning method without SRNet by 1.5%. However, it needs many calculations when three networks are carried out. Meeting the requirement of real-time processing is still a challenging task.

Mohammed and Daood [35] developed a smart surveillance system to monitor one's mask-wearing condition and respecting social distancing. It includes three stages: the first stage is to detect humans using YOLOv3-tiny [136]; based on the regions of detected people, SSD with a ResNet is used to detect face regions; then, MobileNetV2 is used to determine one whether wears a mask or not; finally, the detected ones will be compared with identification database to finish the recognition process. Several networks are used in the process, which seems a complex framework. Although every network has good efficiency, it is inevitable to take more time to perform all the networks. How to reduce inference time and retain the performance is a worthy of study.

In summary, multistage methods mentioned above have at least two deep learning networks. The design of multiple stages is relative complex compared with one-stage and two-stage approaches. It primarily focuses on performance improvement of masked facial detection. Experimental results of original literatures also demonstrate this point. The drawback is also evident: Multiple networks require many computations and expensive processing devices such as GPU.

E. Discussions on the Results of Methods

Before the outbreak of COVID-19, very limited number of papers were proposed for masked facial detection [52]. One important reason is the lacking of masked face datasets. As one of occlusions, masks account for a low ratio in many face detection datasets. The COVID-19 epidemic accelerates the creations of masked facial detection datasets and lead to a rise in the research of masked facial detection methods.

This article presents a rough categorization for the masked face detection techniques according to the used features and the number of processing stages. An overview of some representative methods mentioned is listed in Table V. It can be concluded that most of these methods are tested on their own datasets. We try to analyze them from three parts.

1) *Detection classes:* One-class detection means that only masked face is the objective in image or video. Most of the masked face techniques are designed for two-class detection or three-class detection. Two-class detection methods determine whether one wears a mask or not. Three-class detection methods aim to detect face without mask, face with correct mask, face with incorrect mask. The four-class detection covers "mask area" class additionally. In real-world applications, two-stage methods or multistage methods always locate the face regions first, and then determine the mask-wearing conditions by further classification. In contrast, single-stage neural network methods are able to detect multiple classes through a forward pass process.
2) *Datasets and their sources:* For each family of approaches, thousands of images with annotations are used for training and testing except [69]. This is a common requirement for neural network-based methods with supervised learning. Images or datasets used in Ref. [78] are from some existing datasets. The rest of methods in Table V make use of their self-built datasets. In this article, we survey a series of available datasets in Table I. Different datasets can be combined for researchers to meet requirements and help solve the classes imbalanced problem. Additionally, simulating

TABLE V
BRIEF SUMMARY FOR THE REPRESENTATIVE METHODS

Category	Methods	Detection Classes	Datasets	Results	Experimental Environment and Runtime
Conventional	Dewantara et al [72]	2	1000 images, self-built	$Accuracy = 86.9\%$	Image size: 50×50 to 275×275, 25fps
	Nieto et al [69]	2	677 test cases, self-built	$Recall = 95\%$	VGA resolution 640×480, 10fps
	Petrovic et al [73]	3	Not provide the number	$Accuracy = 84\% - 91\%$	Intel i7 7700-HQ quad-core CPU 2.80 GHz with 16GB RAM, image size 320×240, 38.46fps
	Fang et al [75]	2	6024 images, self-built	$Precision = 96.5\%$	PYNQ-Z2 SoC platform, image size 1280×720, 45.79fps
Single-stage	Razavi et al [77]	3	1853 images, self-built	$Accuracy = 99.8\%$	Not provide runtime
	Zhang et al [47]	3	4672 images, self-built	$mAP = 84.1\%$	Geforce GTX TitanX with memory 12G, not provide runtime
	Chowdary et al [78]	2	1570 images, simulated from SMFD [61]	Train $Accuracy = 99.9\%$, Test $Accuracy = 100\%$	Google Colab, not provide runtime
	Dey et al [60]	2	3835 real images(IDS1), 1376 simulated images (IDS2)	IDS1 $Accuracy = 93\%$, IDS2 $Accuracy = 100\%$	Google Colab, not provide runtime
	Deng et al [79]	2	3656 images,self-built	$mAP = 91.7\%$	NVDIA GTX 1070Ti GPU, not provide runtime
	Wang et al [80]	2	9097 images,self-built	$mAP = 89\%$	Google Colab (Tesla V100-SXM2-16GB), not provide runtime
	Loey et al [42]	1	1415 images, Kaggle [63]	$AP = 81\%$	Not provide runtime
	Jiang et al [50]	3	9205 images, self-built	$mAP = 73.7\%$	RTX 2070 GPU with 8 GB memory, image size: 608×608, 64.0ms per image
	Kumar et al [51]	4	52635 images, self-built	$mAP = 71.69\%$	NVIDIA 1050i GPU with 8 GB memory, not provide runtime
	Yu et al [31]	3	10855 images created from RMFD [53] and MaskedFace-Net [38]	$mAP = 98.3\%$	Inter(R)i7-9700k and RTX 2070Super with 8G memory, image size 416×416, 54.57fps
	Sharma [85]	2	Not provide the number	$mAP \approx 60\%$	Not provide runtime
	Jiang et al [11]	2	7950 images, AIZOOTech [62]	Face $F1 = 93.73\%$, Masked Face $F1 = 93.95\%$	NVIDIA GeForce RTX 2080 Ti, not provide runtime
Two-stage	Wang et al [44]	1	7804 images, WMD [44]	$F1 = 94.19\%$	NVIDIA Geforce GTX 1660 super, 112.5ms per image
	Mercaldo et al [37]	2	4095 images from [53], [63]	$Accuracy = 98\%$	Intel Core i7 8th gen, equipped with 2 GPU and 16G RAM, 4.7s per image
	Loey et al [12]	2	DS1 [53], DS2 [61],LFW [58]	DS1 $Accuracy = 99.64\%$, DS2 $Accuracy = 99.49\%$	Intel Xeon processor 2 GHz, DS1 0.203s per image, DS2 0.031s per image
	Zereen et al [120]	2	5504 images, self-built	$Accuracy = 97.13\%$	Not provide runtime
	Rudraraju et al [122]	3	1270 images, self-built	$Accuracy = 90\%$	Not provide runtime
Multi-stage	Cota et al [128]	2	2270 images, self-built	$mAP = 85.92\%$	NVIDIA GeForce GTX 1650 Max-Q with 4G memory, image size 320×320, 15.7fps
	Lin et al [129]	2	992 images, self-built	Daytime $Accuracy = 95.8\%$, Nighttime $Accuracy = 94.6\%$	Daytime 1.826s per image, Nighttime 1.791s per image
	Qin et al [40]	3	3835 images, self-built	$Accuracy = 98.7\%$	A i7 CPU and P600 GPU with 4 GB memory, 0.03s per image
	Talahua et al [32]	2	13359 images, self-built	$Accuracy = 99.65\%$	Google Colab, image size 224×224 , 0.84s per image

In te third column, “1” means face mask; “2” means face with mask and face without mask; “3” means face without mask, face with correct mask, face with incorrect mask; “4” means face without mask, face with correct mask, face with incorrect mask, and “mask area.”

samples is an alternative way to enrich datasets. Diverse types of masks will make contribute to the performance improvement of models.

3) *Results:* It is hard to evaluate the best performance for all methods in Table V because they are tested on different datasets. It remains a work to compare these methods on a uniform dataset. Existing results with mAP and Accuracy can be regarded as a reference. On the other hand, various scenes can measure the adaptability of algorithms; for example, daytime and nighttime in [129]. In terms of current results, it is believed that the most promising detectors will be neural network-based techniques due to

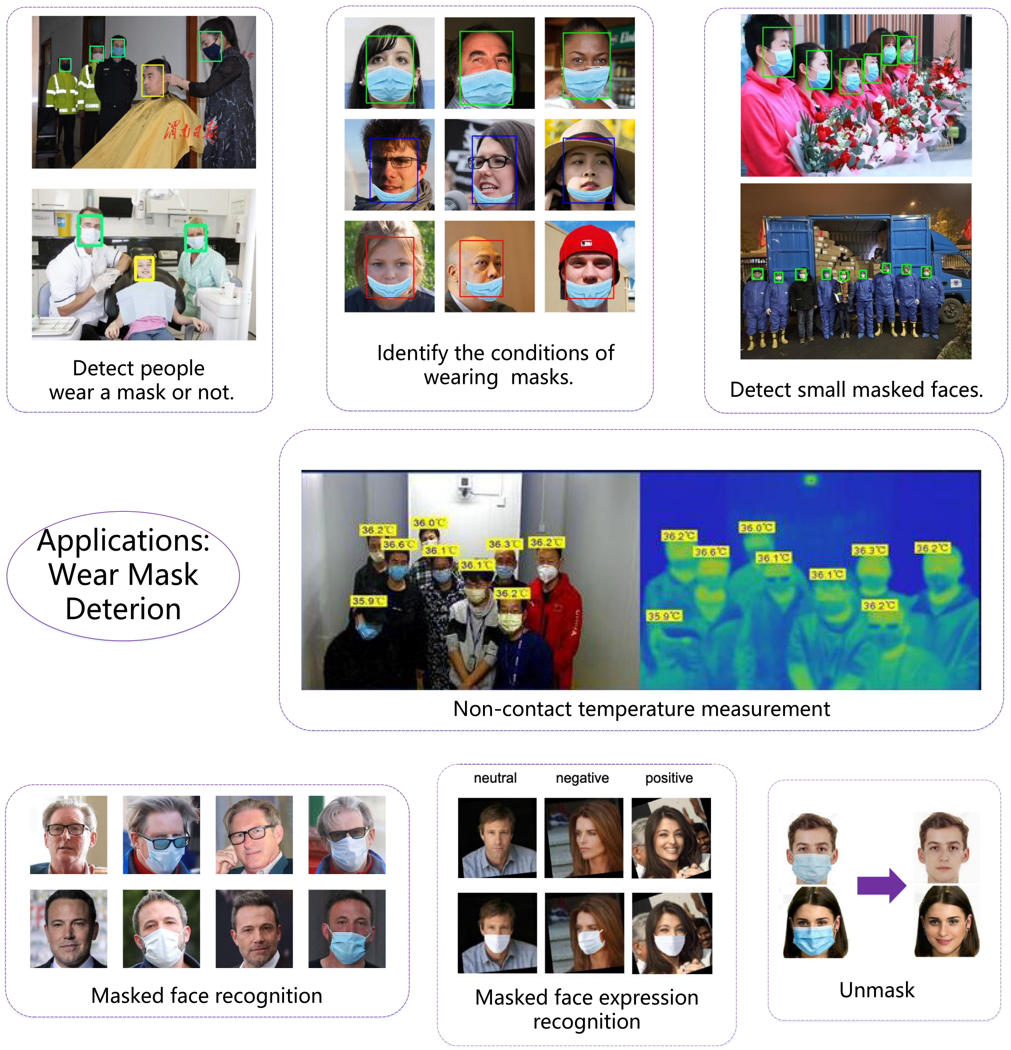

Fig. 9. Some applications of masked facial detection. The first row are generated by [44]. Noncontact temperature measurement is shown in the second row, which comes from "https://gongyi.gmw.cn/2020-03/23/content_33675137.htm." The images in the last row from left to right are collected from Ref.[65], respectively.

their strong learning ability and adaptability to significant variations in appearance of masks.

4) *Experimental environment and runtime:* Efficiency is an important metric to measure one approach. However, quite a number of methods do not provide detailed descriptions about efficiency in Table V. For example, no information about experimental environment and runtime is provided in methods [77]. The literatures [11], [47], [51], [60], [78]–[80] only give their environmental environments or test platforms, without runtime. One potential reason may be derived from that researchers attach more importance to the performance or accuracy. The rest of the methods in Table V shed better light on the runtime. Due to different running environments such as GPU types and various image sizes, it is inapplicable to give a fair comparison between methods. It is clearly shown in Table V that some conventional methods like [72] achieve the real-time processing effects without GPU. In contrast, some CNN-based methods [37] are time-consuming because of the operation of more than one networks. Neural network-based methods are expected to be optimized to reach real-time processing while maintaining high accuracy.

Moreover, there are many applications based on masked facial detection methods in the era of COVID-19. Some examples are shown in Fig. 9. Basic functions include the following: detect whether one wears a mask or not; identify the conditions of wearing a mask: correct or incorrect; and detect small masked faces from long-distance views [44]. These functions are helpful for access control system, crowd counting, social distance monitoring, etc. It should be mentioned that locating masked face regions can help infrared camera finish noncontact temperature measurement [139] and reduce the infection risks caused by close-contact. The results generated by masked facial detection methods can be sent to face recognition model to implement the identification verification [65]. Masked face expression recognition [137] is also an interesting application. Masked faces can be used for unmask or face restoration [138], which is promising in the field of safety protection.

V. Discussions and Future Research Directions

A. Discussions on the Limitations of Datasets

Generally, one benchmark dataset is required with large quantity, multiple classes of wearing mask conditions, versatile types of masked faces, proper ratio between realistic images and simulated images, and diverse scenes. However, there are some limitations for current datasets. Herein, we discuss five points about the limitations of datasets.

1) *Some datasets are (very) small in quantity:* For example, only several hundreds of images in the datasets are used for training deep learning models, which easily results in overfitting phenomenon. In most cases, the larger the dataset, the better the trained model.
2) *A fair proportion of datasets include only two classes:* mask and nonmask. These datasets are only designed for distinguishing masked face from nonmask face. Although some datasets include correct wearing mask and incorrect wearing mask, the number of incorrect wearing masks is very small.
3) *Some datasets are created by simulating masks:* Their quantities are always large. It makes for the training of masked face detection approaches. However, the mask type is always unitary when simulated. Versatile types of masked faces are required to enrich those datasets.
4) *Realistic and simulated images are both included in some datasets:* However, the ratios between realistic masks and simulated masks are imbalanced. The resolutions of images in some datasets may be in varied forms.
5) *Most of the images in some datasets are collected or captured from simple scenes:* They are easily biased toward to a special scene. Thus, trained model based on such datasets may be ineffective for a new scene. GAN-based techniques are expected to create various masked faces with different textures, colors and backgrounds.

B. Discussions on the Limitations of Methods

In the task of masked facial detection, there are some limitations for current methods.

1) *Masked facial wearing conditions:* Some methods only detect two classes: masked facial or no-mask facial, ignoring of mask wearing conditions. It is well known that incorrect wearing mask cannot counteract the spread of COVID-19. Only a few methods were proposed to detect the mask-wearing conditions. Thus, more algorithms should be verified on the detection of masked facial wearing conditions.
2) *Insufficiency of uniform evaluation for methods:* Although some literatures present an evaluation of several methods, there is still lack of uniform evaluation for so many masked facial detection methods. Different methods may be implemented on different platforms. The results provided by original literatures only give readers conceptual comparisons. It is not easy to give a fair judgment.
3) *Deficiency of computation cost:* Good performance is achieved by quite a number of methods. However, the cost-effectiveness and running environment are not detailed for some methods. In real applications, running time is an important measurement metric. Maintaining good performance with a light-weight equipment is a challenging task for existing techniques.
4) *Lacking of model size:* Many methods do not provide the size of trained models or the size of parameters. Actually, this is an important issue for real-time processing on edge devices with limited storage. Light-weight models are supposed to be highlighted because they are in the hopes of deploying in mobile devices or edge devices.
5) *Variation of image resolution:* Some deep neural networks need a fixed size of images as input. However, input images are always with various resolutions. To meet the requirement of fixed size, these images are resized to prepare them for subsequent steps. This may bring about low image quality and facial region distortion, decreasing detection performance.

C. Future Research Directions

In this section, we would like to highlight the future research directions. Even though it was demonstrated recently that neural network-based methods have achieved excellent results, there are still some issues should be invested further. We conclude ten directions as follows.

1) *Create more balanced datasets:* Classes imbalance problem exists as shown in Table III. Neural network-based methods are all appearance-based, which requires enough balanced data to train models. From the surveyed datasets, we find that the number of incorrect wearing mask is very limited. Thus, the category of images should be added significantly. Collecting sufficient samples is a time-consuming and expensive task. Two strategies can be taken into account. First, simulating techniques of matching an mask to face can be used to create samples [64], [140]. In this process, adding a variety of masked face types can enrich existing datasets. Second, GAN-based techniques can be used to produce a series of synthetic images that are very similar with real masks directly. Various environmental illuminations and head poses are expected to be generated. In addition, data augmentation techniques [26] can also be considered to add more masked facial face orientations. Hence, the mentioned techniques are all expected to make sense in the process of constructing more balanced datasets.
2) *Apply transfer learning techniques to masked facial detection:* In the past decades, various object detectors, like Faster R-CNN [103], SSD [104], YOLO [105], and MobileNet [106], were proposed, which achieved excellent results. These detectors are trained on multiclass detection datasets. They can also be used to detect masked faces by transfer learning techniques [9]. Masked face detection and segmentation based on Mask R-CNN [141] can also be considered as a way. It is expected to realize more multiclass detectors in future. Advanced works of object detection can also

be employed for the task of masked facial detection; for example, DEtection TRansformer (DETR) [142], anchor-free deep learning detectors CenterNet [143], and CornerNet [144]. In particular, how to implement knowledge transfer from current dataset to a special dataset is an interesting and promising direction.

3) *Combine predetector and verification model for masked facial detection:* Most of the two-stage methods are the combinations of face detector and classification model [44]. Predetection stage can be implemented by face detectors. Verification stage not only focuses on classification task but also solves some error detections like cross-classes problem. The combinations between two stages is feasible. Conventional models like AdaBoost cascade-classifiers can be combined with state-of-the-art CNN classification models for masked face detection. Multiple neural network models can be combined together to reach a high accuracy. It is an interesting research direction to make a proper selection with a good trade-off between accuracy and efficiency.
4) *Consider contextual information for masked facial detection:* Masked face is one part of body and it is linked with other body parts. Some literatures such as multistage methods [128] are designed to detect key points of body, e.g., 18 key points or 25 key points. Based on the points belonging to eyes and nose, face ROI can be estimated. Due to the occlusion of masks, the features are less in images that are captured from long distance [44]. Contextual information like key points can be utilized to improve the accuracy of small masked face detection. To our understanding, human pose estimation offers a powerful way and it is a very promising direction.
5) *Explore light-weight models and deploy them on mobile or edge devices:* A good light-weight model should be with fast inference and high accuracy. It is of importance to integrate real-world masked facial detection system with Internet of Things. Moreover, the proposed light-weight neural networks in the published literatures need to be conducted on the same dataset and platform. Uniform evaluation of these methods can make readers a good understanding of every method's performance, and guide users to select a proper algorithm to meet their requirements. This is a valuable research direction.
6) *Process various resolutions of images:* Some deep neural networks require a fixed size of images as input. In general, images with different resolutions need to be resized. Actually, resized images easily result in object distortion and information deficiency, which is a potential restriction. How to process various resolutions of images in a feasible manner is an important issue in future work.
7) *Masked face reconstruction:* This is also called "removing mask objects from facial images" [138]. It is a challenging task because more than half of face region is occluded by mask and it is nontransparent. To reach the goal of unmasking, two stages may be considered. First, mask regions need to be segmented very accurately. The second stage is to synthesize masked facial regions and it needs to keep whole coherency of face structure. GAN-based approaches are regarded to be effective because of its strong learning ability. Therefore, it is an interesting issue to explore image editing techniques or object removal techniques to attain global coherency and restore deep missing regions. This is of help for the tasks of masked face recognition [147] and masked facial expression recognition [137].
8) *Masked face recognition:* With the pandemic-driven continuous use of facial masks, it poses a huge challenge to conventional face recognition systems. This motivates researchers to develop a system that performs well with masked facials [147]. The requirement is more imperative than before. To solve the problem, two directions can be considered: The first is to recover masked regions for facial feature extraction; and the second is to generate occlusion-robust feature from masked faces. A competition of masked face recognition held at 2021 International Joint Conference on Biometrics (IJCB-MFR-2021) [151] attracted many participators around the world to submit their solutions [152]. It is reported to collect the largest masked face recognition dataset. In future, the deployability of innovative solutions proposed in IJCB-MFR-2021 will be considered to make sense in people's daily life. It is encouraged to propose excellent algorithms for masked face recognition further.
9) *Masked faces and other biometrics for multimodal identification:* In the era of COVID-19, people are required to wear a mask when entering public places. Single face recognition technique may fail when one wears a mask. Multimodal biometrics can help a lot. It is an interesting topic to combine masked facial with palm print, thumb, and finger vein to construct multimodal biometrics for object identification [155].
10) *Masked face alignment:* The goal of face alignment algorithms is to predict the positions of facial landmark or predefined key points on faces. When one wears a mask, much facial information is missing, which brings about huge challenge to existing face alignment algorithms. Although some researchers [156], [157] have proposed a few solutions based on neural networks to tackle the problem, there are still many worthwhile works to improve the accuracy and reduce inference time. It is believed to be a promising research direction.

VI. Conclusion

In this article, we survey recent advances in the field of masked facial detection. Masked facial datasets are first reviewed. Thirteen open datasets are concluded from various aspects and their valid links are provided. We analyze these datasets from image sources, reality of images, classes imbalance, and experimental results. They can be used to create new larger datasets. Simulating wearing masks is an alternative way to generate samples to enrich existing datasets and improve the robustness of deep learning models.

We review a series of masked face detection methods. They are classified into two categories: conventional methods and neural network-based methods. Five typical conventional algorithms are outlined briefly. For neural network-based methods, they account for the largest ratio and further classified into three classes according to the number of processing stages: single-stage methods, two-stage methods, and multistage methods. For each class, representative methods are described in detail and some typical techniques are introduced briefly. Moreover, we summarize the results of representative methods according to the original literatures. Limitations of datasets and methods are discussed. Neural network-based methods are the mainstream and promising techniques. Finally, we highlight 10 research directions about masked facial detection in the future. Our work is finished in the era of epidemics in the hopes of providing some help in the fighting against COVID-19.

Acknowledgment

We thank the authors of mentioned literatures for the sharings of their datasets.

References

[1] N. Zhu *et al.*, "A novel coronavirus from patients with pneumonia in China, 2019," *New England J. Med.*, vol. 382, no. 8, pp. 727–733, 2020.

[2] [Online]. Available: https://tinyurl.com/WHOPandemicAnnouncement

[3] M. Roser, H. Ritchie, E. Ortiz-Ospina, and J. Hasell, "Coronavirus pandemic (COVID-19)," Our World in Data, 2020.

[4] [Online]. Available: https://coronavirus.jhu.edu/map.html

[5] M. N. Islam, T. T. Inan, S. Rafi, S. S. Akter, I. H. Sarker, and A. N. Islam, "A systematic review on the use of AI and ML for fighting the COVID-19 pandemic," *IEEE Trans. Artif. Intell.*, vol. 1, no. 3, pp. 258–270, Dec. 2020.

[6] S. Latif *et al.*, "Leveraging data science to combat COVID-19: A comprehensive review," *IEEE Trans. Artif. Intell.*, vol. 1, no. 1, pp. 85–103, Aug. 2020.

[7] V. Ayumi, "Application of machine learning for SARS-CoV-2 outbreak," *Chaos, Solitons Fractals*, vol. 139, 2020, Art. no. 110059.

[8] A. Ulhaq, J. Born, A. Khan, D. P. S. Gomes, S. Chakraborty, and M. Paul, "Covid-19 control by computer vision approaches: A survey," *IEEE Access*, vol. 8, pp. 179 437–179 456, 2020.

[9] S. Niu, Y. Liu, J. Wang, and H. Song, "A decade survey of transfer learning (2010–2020)," *IEEE Trans. Artif. Intell.*, vol. 1, no. 2, pp. 151–166, Oct. 2020.

[10] F. Piccialli, V. S. di Cola, F. Giampaolo, and S. Cuomo, "The role of artificial intelligence in fighting the COVID-19 pandemic," *Inf. Syst. Frontiers*, vol. 23, pp. 1467–1497, 2021.

[11] M. Jiang and X. Fan, "Retinamask: A face mask detector," 2020, *arXiv:2005.03950*.

[12] M. Loey, G. Manogaran, M. H. N. Taha, and N. E. M. Khalifa, "A hybrid deep transfer learning model with machine learning methods for face mask detection in the era of the COVID-19 pandemic," *Measurement*, vol. 167, 2021, Art. no. 108288.

[13] W. M. Shaban, A. H. Rabie, A. I. Saleh, and M. Abo-Elsoud, "Detecting COVID-19 patients based on fuzzy inference engine and deep neural network," *Appl. Soft Comput.*, vol. 99, 2021, Art. no. 106906.

[14] C. Iwendi, K. Mahboob, Z. Khalid, A. R. Javed, M. Rizwan, and U. Ghosh, "Classification of COVID-19 individuals using adaptive neuro-fuzzy inference system," *Multimedia Syst.*, pp. 1–15, 2021.

[15] Y. Guo *et al.*, "Assessing and controlling infection risk with Wells–Riley model and spatial flow impact factor (SFIF)," *Sustain. Cities Soc.*, vol. 67, 2021, Art. no. 102719.

[16] A. Kallel, M. Rekik, and M. Khemakhem, "Hybrid-based framework for Covid-19 prediction via federated machine learning models," *J. Supercomput.*, pp. 1–28, 2021.

[17] T. Mahmud, M. A. Rahman, S. A. A. Fattah, and S.-Y. Kung, "CovSegNet: A multi encoder–decoder architecture for improved lesion segmentation of COVID-19 chest CT scans," *IEEE Trans. Artif. Intell.*, vol. 2, no. 3, pp. 283–297, Jun. 2021.

[18] E. Fischer, M. Fischer, D. Grass, I. Henrion, W. Warren, and E. Westman, "Low-cost measurement of face mask efficacy for filtering expelled droplets during speech," *Sci. Adv.*, vol. 6, no. 36, 2020, Art. no. eabd 3083.

[19] M. Klompas, C. A. Morris, J. Sinclair, M. Pearson, and E. S. Shenoy, "Universal masking in hospitals in the COVID-19 era," *New England J. Med.*, vol. 382, no. 21, p. e 63, 2020.

[20] S. Feng, C. Shen, N. Xia, W. Song, M. Fan, and B. J. Cowling, "Rational use of face masks in the COVID-19 pandemic," *Lancet Respir. Med.*, vol. 8, no. 5, pp. 434–436, 2020.

[21] W. H. Organization *et al.*, "Advice on the use of masks in the context of COVID-19: Interim guidance, 6 Apr. 2020," World Health Organization, Tech. Rep., 2020.

[22] S. Lin, L. Cai, X. Lin, and R. Ji, "Masked face detection via a modified LeNet," *Neurocomputing*, vol. 218, pp. 197–202, 2016.

[23] S. Meivel, K. I. Devi, S. U. Maheswari, and J. V. Menaka, "Real time data analysis of face mask detection and social distance measurement using MATLAB," *Mater. Today Proc.*, 2021.

[24] S. Meivel, K. I. Devi, T. M. Selvam, and S. U. Maheswari, "Real time analysis of unmask face detection in human skin using tensor flow package and IoT algorithm," *Mater. Today Proc.*, 2021.

[25] X. Fan, M. Jiang, and H. Yan, "A deep learning based light-weight face mask detector with residual context attention and Gaussian heatmap to fight against COVID-19," *IEEE Access*, vol. 9, pp. 96964–96974, 2021.

[26] M. R. Prusty, V. Tripathi, and A. Dubey, "A novel data augmentation approach for mask detection using deep transfer learning," *Intell.-Based Med.*, vol. 5, 2021, Art. no. 100037.

[27] S. Sethi, M. Kathuria, and T. Kaushik, "A real-time integrated face mask detector to curtail spread of coronavirus," *CMES Comput. Model. Eng. Sci.*, vol. 127, no. 2, pp. 389–409, 2021.

[28] F. D. Adhinata, D. P. Rakhmadani, M. Wibowo, and A. Jayadi, "A deep learning using DenseNet201 to detect masked or non-masked face," *JUITA: Jurnal Informatika*, vol. 9, no. 1, pp. 115–121, 2021.

[29] D. G. Dondo, J. A. Redolfi, R. G. Araguás, and D. Garcia, "Application of deep-learning methods to real time face mask detection," *IEEE Latin America Trans.*, vol. 19, no. 6, pp. 994–1001, Jun. 2021.

[30] A. Nowrin, S. Afroz, M. S. Rahman, I. Mahmud, and Y.-Z. Cho, "Comprehensive review on facemask detection techniques in the context of COVID-19," *IEEE Access*, vol. 9, pp. 106839–106864, 2021.

[31] J. Yu and W. Zhang, "Face mask wearing detection algorithm based on improved YOLO-v4," *Sensors*, vol. 21, no. 9, 2021, Art. no. 3263.

[32] J. S. Talahua, J. Buele, P. Calvopã, and J. Varela-Aldás, "Facial recognition system for people with and without face mask in times of the COVID-19 pandemic," *Sustainability*, vol. 13, no. 12, 2021, Art. no. 6900.

[33] J. Tomás, A. Rego, S. Viciano-Tudela, and J. Lloret, "Incorrect facemask-wearing detection using convolutional neural networks with transfer learning," in *Healthcare*. Multidisciplinary Digital Publishing Institute, vol. 9, no. 8, 2021.

[34] S. Hussain *et al.*, "IoT and deep learning based approach for rapid screening and face mask detection for infection spread control of Covid-19," *Appl. Sci.*, vol. 11, no. 8, 2021, Art. no. 3495.

[35] M. R. Mohammed and A. Daood, "Smart surveillance system to monitor the committed violations during the pandemic," *Int. J. Comput. Digit. Syst.*, 2021.

[36] P. Nagrath, R. Jain, A. Madan, R. Arora, P. Kataria, and J. Hemanth, "SSDMNV2: A real time DNN-based face mask detection system using single shot multibox detector and MobilenetV2," *Sustain. Cities Soc.*, vol. 66, 2021, Art. no. 102692.

[37] F. Mercaldo and A. Santone, "Transfer learning for mobile real-time face mask detection and localization," *J. Amer. Med. Informat. Assoc.*, vol. 28, no. 7, pp. 1548–1554, 2021.

[38] A. Cabani, K. Hammoudi, H. Benhabiles, and M. Melkemi, "Maskedface-net-A dataset of correctly/incorrectly masked face images in the context of COVID-19," *Smart Health*, vol. 19, 2020, Art. no. 100144.

[39] S. Chen, W. Liu, and G. Zhang, "Efficient transfer learning combined skip-connected structure for masked face poses classification," *IEEE Access*, vol. 8, pp. 209688–209698, 2020.

[40] B. Qin and D. Li, "Identifying facemask-wearing condition using image super-resolution with classification network to prevent COVID-19," *Sensors*, vol. 20, no. 18, 2020, Art. no. 5236.

[41] P. Mohan, A. J. Paul, and A. Chirania, "A tiny CNN architecture for medical face mask detection for resource-constrained endpoints," in *Innovations in Electrical and Electronic Engineering*. Berlin, Germany: Springer, 2021, pp. 657–670.
[42] M. Loey, G. Manogaran, M. H. N. Taha, and 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," *Sustain. Cities Soc.*, vol. 65, 2021, Art. no. 102600.
[43] B. Roy, S. Nandy, D. Ghosh, D. Dutta, P. Biswas, and T. Das, "MOXA: A deep learning based unmanned approach for real-time monitoring of people wearing medical masks," *Trans. Indian Nat. Acad. Eng.*, vol. 5, no. 3, pp. 509–518, 2020.
[44] B. Wang, Y. Zhao, and C. P. Chen, "Hybrid transfer learning and broad learning system for wearing mask detection in the COVID-19 era," *IEEE Trans. Instrum. Meas.*, vol. 70, 2021, Art. no. 5009612.
[45] A. Faizah, P. H. Saputro, A. J. Firdaus, and R. N. R. Dzakiyullah, "Implementation of the convolutional neural network method to detect the use of masks," *Int. J. Informat. Inf. Syst.*, vol. 4, no. 1, pp. 30–37, 2021.
[46] I. Omar and A. Rasha, "Automated realtime mask availability detection using neural network," *Int. J. Comput. Digit. Syst.*, vol. 10, pp. 1–6, 2021.
[47] J. Zhang, F. Han, Y. Chun, and W. Chen, "A novel detection framework about conditions of wearing face mask for helping control the spread of COVID-19," *IEEE Access*, vol. 9, pp. 42975–42984, 2021.
[48] S. Singh, U. Ahuja, M. Kumar, K. Kumar, and M. Sachdeva, "Face mask detection using YOLOv3 and faster R-CNN models: COVID-19 environment," *Multimedia Tools Appl.*, vol. 80, no.13, pp. 19753–19768, 2021.
[49] B. Batagelj, P. Peer, V. Štruc, and S. Dobrišek, "How to correctly detect face-masks for COVID-19 from visual information?," *Appl. Sci.*, vol. 11, no. 5, 2021, Art. no. 2070.
[50] X. Jiang, T. Gao, Z. Zhu, and Y. Zhao, "Real-time face mask detection method based on YOLOv3," *Electronics*, vol. 10, no. 7, p. 837, 2021.
[51] A. Kumar, A. Kalia, K. Verma, A. Sharma, and M. Kaushal, "Scaling up face masks detection with YOLO on a novel dataset," *Optik*, vol. 239, 2021, Art. no. 166744.
[52] S. Ge, J. Li, Q. Ye, and Z. Luo, "Detecting masked faces in the wild with LLE-CNNS," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 2682–2690.
[53] Z. Wang *et al.*, "Masked face recognition dataset and application," 2020, *arXiv:2003.09093* .
[54] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 4401–4410.
[55] [Online]. Available: https://github.com/tzutalin/labelImg/
[56] S. Yang, P. Luo, C.-C. Loy, and X. Tang, "Wider face: A face detection benchmark," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 5525–5533.
[57] F. I. Eyiokur, H. K. Ekenel, and A. Waibel, "A computer vision system to help prevent the transmission of COVID-19," 2021, *arXiv:2103.08773*.
[58] G. B. Huang and E. Learned-Miller, "Labeled faces in the wild: Updates and new reporting procedures," Dept. Comput. Sci., Univ. Massachusetts Amherst, Amherst, MA, USA, Tech. Rep., vol. 14, 2014.
[59] Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep learning face attributes in the wild," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 3730–3738.
[60] S. K. Dey, A. Howlader, and C. Deb, "Mobilenet mask: A multi-phase face mask detection model to prevent person-to-person transmission of SARs-COV-2," in *Proc. Int. Conf. Trends Comput. Cogn. Eng.*, Springer, 2021, pp. 603–613.
[61] [Online]. Available: https://github.com/prajnasb/observations
[62] [Online]. Available: https://github.com/AIZOOTech/FaceMaskDetection
[63] [Online]. Available: https://www.kaggle.com/andrewmvd/face-mask-detection
[64] [Online]. Available: https://github.com/zamhown/wear-a-mask
[65] A. Anwar and A. Raychowdhury, "Masked face recognition for secure authentication," 2020, *arXiv:2008.11104* .
[66] S. Zafeiriou, C. Zhang, and Z. Zhang, "A survey on face detection in the wild: Past, present and future," *Comput. Vis. Image Understanding*, vol. 138, pp. 1–24, 2015.
[67] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 1, 2001.
[68] P. Viola and M. J. Jones, "Robust real-time face detection," *Int. J. Comput. Vis.*, vol. 57, no. 2, pp. 137–154, 2004.
[69] A. Nieto-Rodriguez, M. Mucientes, and V. M. Brea, "System for medical mask detection in the operating room through facial attributes," in *Proc. Iberian Conf. Pattern Recognit. Image Anal.*, Springer, 2015, pp. 138–145.
[70] T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern Recognit.*, vol. 29, no. 1, pp. 51–59, 1996.
[71] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vision Pattern Recognit.*, 2005, pp. 886–893.
[72] B. S. B. Dewantara and D. T. Rhamadhaningrum, "Detecting multi-pose masked face using adaptive boosting and cascade classifier," in *Proc. Int. Electron. Symp.*, 2020, pp. 436–441.
[73] N. Petrovic and D. Kocic, "IoT-based system for COVID-19 indoor safety monitoring," *IcETRAN*, vol. 2020, pp. 1–6, 2020.
[74] Y. R. Arif, A. G. Putrada, and R. R. Pahlevi, "An evaluation of a modified Haar-like features based classifier method for face mask detection in the COVID-19 spread prevention," in *Proc. Int. Symp. Electron. Smart Devices*, 2021, pp. 1–5.
[75] T. Fang, X. Huang, and J. Saniie, "Design flow for real-time face mask detection using PYNQ system-on-chip platform," in *Proc. IEEE Int. Conf. Electro Inf. Technol.*, 2021, pp. 1–5.
[76] T. He, "Mask wearing detection method based on the skin color and eyes detection," in *Proc. J. Phys.: Conf. Ser.*, 2021, vol. 1906, no. 1, Art. no. 012012.
[77] M. Razavi, H. Alikhani, V. Janfaza, B. Sadeghi, and E. Alikhani, "An automatic system to monitor the physical distance and face mask wearing of construction workers in COVID-19 pandemic," 2021, *arXiv:2101.01373*.
[78] G. J. Chowdary, N. S. Punn, S. K. Sonbhadra, and S. Agarwal, "Face mask detection using transfer learning of InceptionV3," in *Proc. Int. Conf. Big Data Analytics*, Springer, 2020, pp. 81–90.
[79] H. Deng, J. Zhang, L. Chen, and M. Cai, "Improved mask wearing detection algorithm for SSD," in *Proc. J. Phys.: Conf. Ser.*, 2021, vol. 1757, no. 1, Art. no. 012140.
[80] Z. Wang, P. Wang, P. C. Louis, L. E. Wheless, and Y. Huo, "Wearmask: Fast in-browser face mask detection with serverless edge computing for COVID-19," 2021, *arXiv:2101.00784*.
[81] X. Ren and X. Liu, "Mask wearing detection based on YOLOv3," in *Proc. J. Phys.: Conf. Ser.*, 2020, vol. 1678, no. 1, Art. no. 012089.
[82] M. R. Bhuiyan, S. A. Khushbu, and M. S. Islam, "A deep learning based assistive system to classify COVID-19 face mask for human safety with YOLOv3," in *Proc. 11th Int. Conf. Comput., Commun. Netw. Technol.*, 2020, pp. 1–5.
[83] K. Bhambani, T. Jain, and K. A. Sultanpure, "Real-time face mask and social distancing violation detection system using YOLO," in *Proc. IEEE Bangalore Humanitarian Technol. Conf.*, 2020, pp. 1–6.
[84] S. Degadwala, D. Vyas, U. Chakraborty, A. R. Dider, and H. Biswas, "YOLO-v4 deep learning model for medical face mask detection," in *Proc. Int. Conf. Artif. Intell. Smart Syst.*, 2021, pp. 209–213.
[85] V. Sharma, "Face mask detection using yolov5 for COVID-19," master thesis, California State Univ., San Marcos, 2020.
[86] Y. Ding, Z. Li, and D. Yastremsky, "Real-time face mask detection in video data," 2021, *arXiv:2105.01816*.
[87] G. Yang *et al.*, "Face mask recognition system with YOLOv5 based on image recognition," in *Proc. IEEE 6th Int. Conf. Comput. Commun.*, 2020, pp. 1398–1404.
[88] J. Ieamsaard, S. N. Charoensook, and S. Yammen, "Deep learning-based face mask detection using YOLOv5," in *Proc. IEEE 9th Int. Elec. Eng. Congr.*, 2021, pp. 428–431.
[89] M. M. Rahman, M. M. H. Manik, M. M. Islam, S. Mahmud, and J.-H. Kim, "An automated system to limit COVID-19 using facial mask detection in smart city network," in *Proc. IEEE Int. IOT, Electron. Mechatronics Conf.*, 2020, pp. 1–5.
[90] S. Pooja and S. Preeti, "Face mask detection using AI," in *Predictive and Preventive Measures for COVID-19 Pandemic*. Berlin, Germany: Springer, 2021, pp. 293–305.
[91] J. Xiao, J. Wang, S. Cao, and B. Li, "Application of a novel and improved VGG-19 network in the detection of workers wearing masks," *J. Phys.: Conf. Ser.*, 2020, vol. 1518, no. 1, Art. no. 0 12041.

[92] S. Prasad, Y. Li, D. Lin, and D. Sheng, "Maskedfacenet: A progressive semi-supervised masked face detector," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis.*, 2021, pp. 3389–3398.

[93] W. Jian and L. Lang, "Face mask detection based on transfer learning and PP-YOLO," in *Proc. IEEE 2nd Int. Conf. Big Data, Artif. Intell. Internet Things Eng.*, 2021, pp. 106–109.

[94] M. M. Boulos, "Facial Recognition and Face Mask Detection Using Machine Learning Techniques," thesis, Dissertations and Culminating Projects, 2021.

[95] J. Deng, J. Guo, E. Ververas, I. Kotsia, and S. Zafeiriou, "RetinaFace: Single-shot multi-level face localisation in the wild," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 5203–5212.

[96] U. Alganci, M. Soydas, and E. Sertel, "Comparative research on deep learning approaches for airplane detection from very high-resolution satellite images," *Remote Sens.*, vol. 12, no. 3, p. 458, 2020.

[97] A. Kumar, A. Kaur, and M. Kumar, "Face detection techniques: A review," *Artif. Intell. Rev.*, vol. 52, no. 2, pp. 927–948, 2019.

[98] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, "Joint face detection and alignment using multitask cascaded convolutional networks," *IEEE Signal Process. Lett.*, vol. 23, no. 10, pp. 1499–1503, Oct. 2016.

[99] J. Wang, Y. Yuan, and G. Yu, "Face attention network: An effective face detector for the occluded faces," 2017, *arXiv:1711.07246* .

[100] [Online]. Available: https://github.com/ShiqiYu/libfacedetection

[101] Y. Chen, L. Song, Y. Hu, and R. He, "Adversarial occlusion-aware face detection," in *Proc. IEEE 9th Int. Conf. Biometrics Theory, Appl. Syst.*, 2018, pp. 1–9.

[102] X. Tang, D. K. Du, Z. He, and J. Liu, "Pyramidbox: A context-assisted single shot face detector," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 797–813.

[103] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2016.

[104] W. Liu *et al.*, "SSD: Single shot multibox detector," in *Proc. Eur. Conf. Comput. Vision*, 2016, pp. 21–37.

[105] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 779–788.

[106] A. G. Howard *et al.*, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," 2017, *arXiv:1704.04861* .

[107] Z.-Q. Zhao, P. Zheng, S.-T. Xu, and X. Wu, "Object detection with deep learning: A review," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 11, pp. 3212–3232, Nov. 2019.

[108] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal speed and accuracy of object detection," 2020, *arXiv:2004.10934*.

[109] J. Li *et al.*, "DSFD: Dual shot face detector," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 5060–5069.

[110] C. P. Chen and Z. Liu, "Broad learning system: An effective and efficient incremental learning system without the need for deep architecture," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 1, pp. 10–24, Jan. 2018.

[111] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 1–27, 2011.

[112] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.

[113] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Adv. Neural Inf. Process. Syst.*, vol. 25, pp. 1097–1105, 2012.

[114] B. Arslan, S. Memis, E. Battinisonmez, and O. Z. Batur, "Fine-grained food classification methods on the UEC food-100 database," *IEEE Trans. Artif. Intell.*, vol. 3, no. 2, pp. 238–243, 2021.

[115] A. Chavda, J. Dsouza, S. Badgujar, and A. Damani, "Multi-stage CNN architecture for face mask detection," in *Proc. 6th Int. Conf. Convergence Technol.*, 2021, pp. 1–8.

[116] A. S. Joshi, S. S. Joshi, G. Kanahasabai, R. Kapil, and S. Gupta, "Deep learning framework to detect face masks from video footage," in *Proc. 12th Int. Conf. Comput. Intell. Commun. Netw.*, 2020, pp. 435–440.

[117] S. E. Snyder and G. Husari, "Thor: A deep learning approach for face mask detection to prevent the COVID-19 pandemic," in *Proc. SoutheastCon*, 2021, pp. 1–8.

[118] I. Buciu, "Color quotient based mask detection," in *Proc. Int. Symp. Electron. Telecommun.*, 2020, pp. 1–4.

[119] A. Oumina, N. El Makhfi, and M. Hamdi, "Control the COVID-19 pandemic: Face mask detection using transfer learning," in *Proc. IEEE 2nd Int. Conf. Electron., Control, Optim. Comput. Sci.*, 2020, pp. 1–5.

[120] A. N. Zereen, S. Corraya, M. N. Dailey, and M. Ekpanyapong, "Two-stage facial mask detection model for indoor environments," in *Proc. Int. Conf. Trends Comput. Cogn. Eng.*, 2021, pp. 591–601.

[121] A. Das, M. W. Ansari, and R. Basak, "COVID-19 face mask detection using tensorflow, keras and openCV," in *Proc. IEEE 17th India Council Int. Conf.*, 2020, pp. 1–5.

[122] S. R. Rudraraju, N. K. Suryadevara, and A. Negi, "Face mask detection at the fog computing gateway," in *Proc. 15th Conf. Comput. Sci. Inf. Syst.*, 2020, pp. 521–524.

[123] A. Malakar, A. Kumar, and S. Majumdar, "Detection of face mask in real-time using convolutional neural networks and open-CV," in *Proc. 2nd Int. Conf. Emerg. Technol.*, 2021, pp. 1–5.

[124] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 2818–2826.

[125] C. P. Chen, Z. Liu, and S. Feng, "Universal approximation capability of broad learning system and its structural variations," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 4, pp. 1191–1204, Apr. 2019.

[126] Z. Liu, C. P. Chen, S. Feng, Q. Feng, and T. Zhang, "Stacked broad learning syst.: From incremental flatted structure to deep model," *IEEE Trans. Systems, Man, Cybern.: Syst.*, vol. 51, no. 1, pp. 209–222, Jan. 2021.

[127] C. P. Chen and B. Wang, "Random-positioned license plate recognition using hybrid broad learning system and convolutional networks," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, pp. 444–456, 2022.

[128] D. A. M. Cota, "Monitoring COVID-19 prevention measures on CCTV cameras using deep learning," Ph.D. dissertation, Politecnico di Torino, 2020.

[129] H. Lin, R. Tse, S.-K. Tang, Y. Chen, W. Ke, and G. Pau, "Near-realtime face mask wearing recognition based on deep learning," in *Proc. IEEE 18th Annu. Consumer Commun. Netw. Conf.*, 2021, pp. 1–7.

[130] W. Bu, J. Xiao, C. Zhou, M. Yang, and C. Peng, "A cascade framework for masked face detection," in *Proc. IEEE Int. Conf. Cybern. Intell. Syst. IEEE Conf. Robot., Autom. Mechatronics*, 2017, pp. 458–462.

[131] G. T. Draughon, P. Sun, and J. P. Lynch, "Implementation of a computer vision framework for tracking and visualizing face mask usage in urban environments," in *Proc. IEEE Int. Smart Cities Conf.*, 2020, pp. 1–8.

[132] K. Sun *et al.*, "High-resolution representations for labeling pixels and regions," 2019, *arXiv:1904.04514*.

[133] A. Kolesnikov *et al.*, "Big transfer (bit): General visual representation learning," in *Proc. Eur. Conf. Comput. Vis.,* 2020, pp. 491–507.

[134] N. Wojke, A. Bewley, and D. Paulus, "Simple online and realtime tracking with a deep association metric," in *Proc. IEEE Int. Conf. Image Process.*, 2017, pp. 3645–3649.

[135] Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei, and Y. Sheikh, "Openpose: Realtime multi-person 2D pose estimation using part affinity fields," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 1, pp. 172–186, Jan. 2021.

[136] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018, *arXiv:1804.02767*.

[137] B. Yang, J. Wu, and G. Hattori, "Facial expression recognition with the advent of face masks," in *Proc. 19th Int. Conf. Mobile Ubiquitous Multimedia*, 2020, pp. 335–337.

[138] N. U. Din, K. Javed, S. Bae, and J. Yi, "A novel GAN-based network for unmasking of masked face," *IEEE Access*, vol. 8, pp. 44276–44287, 2020.

[139] S. Khan *et al.*, "Comparative accuracy testing of non-contact infrared thermometers and temporal artery thermometers in an adult hospital setting," *Amer. J. Infection Control*, vol. 49, no. 5, pp. 597–602, 2020.

[140] Y. Hu and X. Li, "Covertheface: Face covering monitoring and demonstrating using deep learning and statistical shape analysis," 2021, *arXiv:2108.10430* .

[141] K. Lin *et al.*, "Face detection and segmentation based on improved mask R-CNN," *Discrete Dyn. Nature Soc.*, 2020.

[142] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, "End-to-end object detection with transformers," in *Proc. Eur. Conf. Comput. Vis.*, 2020, pp. 213–229.

[143] K. Duan, S. Bai, L. Xie, H. Qi, Q. Huang, and Q. Tian, "Centernet: Keypoint triplets for object detection," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 6569–6578.

[144] H. Law and J. Deng, "CornerNet: Detecting objects as paired keypoints," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 734–750.

[145] C. Li, S. Ge, D. Zhang, and J. Li, "Look through masks: Towards masked face recognition with de-occlusion distillation," in *Proc. 28th ACM Int. Conf. Multimedia*, 2020, pp. 3016–3024.

[146] F. Boutros, N. Damer, F. Kirchbuchner, and A. Kuijper, "Unmasking face embeddings by self-restrained triplet loss for accurate masked face recognition," 2021, *arXiv:2103.01716*.
[147] M. Geng, P. Peng, Y. Huang, and Y. Tian, "Masked face recognition with generative data augmentation and domain constrained ranking," in *Proc. 28th ACM Int. Conf. Multimedia*, 2020, pp. 2246–2254.
[148] H. Deng, Z. Feng, G. Qian, X. Lv, H. Li, and G. Li, "MFcosface: A masked-face recognition algorithm based on large margin cosine loss," *Appl. Sci.*, vol. 11, no. 16, p. 7310, 2021.
[149] F. Mo, Z. Zhang, T. Chen, K. Zhao, and X. Fu, "MFED: A database for masked facial expression," *IEEE Access*, vol. 9, pp. 96279–96287, 2021.
[150] H. Du, H. Shi, Y. Liu, D. Zeng, and T. Mei, "Towards NIR-VIS masked face recognition," *IEEE Signal Process. Lett.*, vol. 28, pp. 768–772, 2021.
[151] [Online]. Available: https://sites.google.com/view/ijcb-mfr-2021/home
[152] Z. Zhu *et al.*, "Masked face recognition challenge: The WebFace260 M track report," 2021, *arXiv:2108.07189*.
[153] F. Boutros *et al.*, "MFR 2021: Masked face recognition competition," in *Proc. IEEE Int. Joint Conf. Biometrics*, 2021, pp. 1–10.
[154] J. Deng, J. Guo, X. An, Z. Zhu, and S. Zafeiriou, "Masked face recognition challenge: The insightface track report," 2021, *arXiv:2108.08191* .
[155] M. I. Amin, M. A. Hafeez, R. Touseef, and Q. Awais, "Person identification with masked face and thumb images under pandemic of COVID-19," in *Proc. 7th Int. Conf. Control Instrum. Autom.*, 2021, pp. 1–4.
[156] Y. Sha, J. Zhang, X. Liu, Z. Wu, and S. Shan, "Efficient face alignment network for masked face," in *Proc. IEEE Int. Conf. Multimedia Expo. Workshops*, 2021, pp. 1–6.
[157] T. Wen *et al.*, "Towards efficient masked-face alignment via cascaded regression," in *Proc. IEEE Int. Conf. Multimedia Expo. Workshops*, 2021, pp. 1–5.

Bingshu Wang received the B.S. degree in computer science and technology from Guizhou University, Guiyang, China, in 2013, the M.S. degree in electronic science and technology (integrated circuit system) from Peking University, Beijing, China, in 2016, and the Ph.D. degree in computer science from the University of Macau, Macau, China, in 2020.

He is currently an Associate Professor with the School of Software, Northwestern Polytechnical University, Xi'an, China. His research interests include computer vision, intelligent video analysis, and machine learning.

Dr. Wang is a member of the Chinese Association of Automation (CAA).

Jiangbin Zheng received the Ph.D. degree in computer science and technology from Northwestern Polytechnical University, Xi'an, China, in 2002.

He is currently a Full Professor and Dean of the School of Software, Northwestern Polytechnical University. He has authored or co-authored more than 100 papers in the above related research area. His research interests include computer graphics, computer vision, and multimedia.

C. L. Philip Chen (Fellow, IEEE) received the M.S. degree in electrical engineering from the University of Michigan, Ann Arbor, MI, USA, in 1985, and the Ph.D. degree in electrical engineering from Purdue University, West Lafayette, IN, USA, in 1988.

He is currently the Chair Professor and Dean of the College of Computer Science and Engineering, South China University of Technology, Guangzhou, China. Being a Program Evaluator of the Accreditation Board of Engineering and Technology Education (ABET) in the US, for computer engineering, electrical engineering, and software engineering programs, he successfully architects the University of Macau's Engineering and Computer Science programs receiving accreditations from Washington/Seoul Accord through Hong Kong Institute of Engineers (HKIE), of which is considered as his utmost contribution in engineering/computer science education for Macau as the former Dean of the Faculty of Science and Technology. His research interests include cybernetics, systems, and computational intelligence.

Dr. Chen is a Fellow of AAAS, IAPR, CAA, and HKIE; a member of Academia Europaea (AE), European Academy of Sciences and Arts (EASA), and International Academy of Systems and Cybernetics Science (IASCYS). He was a recipient of the IEEE Norbert Wiener Award in 2018 for his contribution in systems and cybernetics, and machine learnings. Best Transactions Paper Award from IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS for his papers in 2014 and 2018. He is also a highly cited Researcher by Clarivate Analytics in 2018, 2019, and 2020. He is currently the Editor-in-Chief of the IEEE TRANSACTIONS ON CYBERNETICS, and an Associate Editor for the IEEE TRANSACTIONS ON AI and IEEE TRANSACTIONS ON FUZZY SYSTEMS. He was a recipient of the 2016 Outstanding Electrical and Computer Engineers Award from his alma mater, Purdue University in 1988. He was the President of IEEE SYSTEMS, MAN, AND CYBERNETICS SOCIETY from 2012 to 2013 and the Editor-in-Chief of the IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS: SYSTEMS from 2014 to 2019. He was the Chair of TC 9.1 Economic and Business Systems of International Federation of Automatic Control from 2015 to 2017.