Journal of Soft Computing and Computer Applications

Volume 1 | Issue 1 Article 1006

2024

Face Mask Detection Based on Deep Learning: A Review

Shahad Fadhil Abbas
University of Technology - Iraq, Baghdad, Iraq, cs.20.30@grad.uotechnology.edu.iq

Shaimaa Hameed Shaker
University of Technology - Iraq, Baghdad, Iraq, shaimaa.h.shaker@uotechnology.edu.iq

Firas. A. Abdullatif
Baghdad University, College of Education for Pure Science/Ibn-Al-Haithem, Baghdad, Iraq, firas.alobaedy@ihcoedu.uobaghdad.edu.iq

Recommended Citation
Abbas, Shahad Fadhil; Shaker, Shaimaa Hameed; and Abdullatif, Firas. A. (2024) "Face Mask Detection Based on Deep Learning: A Review," *Journal of Soft Computing and Computer Applications*: Vol. 1: Iss. 1, Article 1006.
DOI: https://doi.org/10.70403/3008-1084.1006

JOURNAL OF SOFT COMPUTING AND COMPUTER APPLICATIONS 2024;1:1006

REVIEW

Face Mask Detection Based on Deep Learning: A Review

Shahad Fadhil Abbas [a,*], **Shaimaa Hameed Shaker** [a], **Firas. A. Abdullatif** [b]

[a] University of Technology – Iraq, Baghdad, Iraq
[b] Baghdad University, College of Education for Pure Science/Ibn-Al-Haithem, Baghdad, Iraq

ABSTRACT

The coronavirus disease 2019 outbreak caused widespread disruption. The World Health Organization has recommended wearing face masks, along with other public health measures, such as social distancing, following medical guidelines, and thermal scanning, to reduce transmission, reduce the burden on healthcare systems, and protect population groups. However, wearing a mask, which acts as a barrier or shield to reduce transmission of infection from infected individuals, hides most facial features, such as the nose, mouth, and chin, on which face detection systems depend, which leads to the weakness of these systems. This paper aims to provide essential insights for researchers and practitioners interested in developing and implementing deep learning-based face mask detection systems. Although current deep learning models have made significant strides and tremendous advances in many applications, including security, access control, and identity verification, development efforts continue. This paper also discusses the importance of the datasets used to train and evaluate these models, emphasising the need for diverse data such as mask types, facial occlusions, lighting conditions, and high-quality data to enhance model performance. In addition to the challenges posed by real-world conditions that can affect detection accuracy, face detection and face mask recognition methods were compared to deep learning models, where the accuracy of the Multi-Task ArcFace model reached 99.78, which is the highest accuracy among other detection methods.

Keywords: Face mask detection, Face detection, Deep learning, Mask detection, Unmask detection

1. Introduction

Face mask possession varies with the extensive range of styles and sizes available globally, which makes it necessary for Deep Learning (DL) models to identify them worn on the face appropriately. Common detection algorithms often used for object detection are

Received 30 March 2024; accepted 12 May 2024.
Available online 27 June 2024

* Corresponding author.
E-mail addresses: cs.20.30@grad.uotechnology.edu.iq (S. Fadhil Abbas), shaimaa.h.shaker@uotechnology.edu.iq (S. Hameed Shaker), firas.alobaedy@ihcoedu.uobaghdad.edu.iq (F. A. Abdullatif).

https://doi.org/10.70403/3008-1084.1006

also being examined for discovering and recognizing faces with masks [1]. Face detection is particularly significant when used to determine whether a person has a mask on. Face detection in computer vision has been widely studied and has received much attention and research; therefore, better algorithms have been developed. The speed and accuracy with which face detection has been enhanced are advancements that people have seen over time. Previously, human face detection processes depended on manually extracted features using machine-polished features and conventional machine-learning algorithms [2]. Nevertheless, Convolutional Neural Networks (CNNs), as a deep learning branch, have revolutionized face detection. CNNs can extract hierarchical representations of facial features from raw pixel values in images using predefined kernels, thereby allowing them to better understand facial variations. Therefore, CNN-based approaches for face detection have been proven to be more effective than traditional methods because their performance is exceptionally high, with higher accuracy and robustness with respect to specific parameters such as lighting variations, pose, facial expression, and occlusion [3]. Face masks have proven effective in preventing the spread of the novel coronavirus. However, their improper disposal and prolonged use pose environmental and health risks. The mismanagement of face masks and other Personal Protective Equipment (PPE) is currently being studied owing to its environmental impact [4].

Facial recognition techniques first depend on detecting a face where all facial features are visible. However, wearing a mask confuses this technology because it covers approximately half of the face. The process of identifying masked and unmasked faces in an image involves two stages: face recognition and feature extraction. Object-detection methods can be used to detect multiple masked faces in an image. Subsequently, the features are extracted and matched with masked and unmasked images using a classifier or model [5].

Wearing a mask along with other precautions is crucial, even after receiving the COVID-19 vaccine, as per the World Health Organization (WHO). The use of masks helps prevent the spread of infectious viral particles through respiratory droplets [6]. Different types of masks, such as N95 respirators, KN95 masks, and surgical masks, provide various levels of protection. The WHO recommends wearing masks to reduce the risk of infection and other respiratory diseases, and provide protection from water splashes. Campaigns are being held to raise awareness about the importance of mask-wearing, and some public places have implemented “no mask, no entry” policies [7]. In large organizations and crowded areas, it is not feasible to manually monitor everyone to ensure they wear masks. The “Face Mask Detection System” can be an effective solution in such situations. Various studies are currently being conducted on the effects of wearing masks [8]. Artificial intelligence, particularly DL techniques, has made remarkable progress in detecting and recognizing faces, even when masks partially cover faces. Deep neural networks are among the most important technologies contributing to this progress, from CNN to You Only Look Once (YOLO) [9–11]. The main contributions of this paper are:

- Address the problems of detecting masked and unmasked faces, such as problems with different face sizes, the correct mask position, lighting, and angled faces.
- Reviews of recent works have been developed for face mask detection based on deep learning methods, providing details and discussions on the development of these systems.
- It provides a broad analysis of various deep-learning technologies and their performances.
- Identifies sources of multiple datasets and software needed for implementation.
- Different use cases of the system, challenges, and observations are also discussed, along with the conclusions and suggestions for future research.

This paper is organised as follows: Section 2 presents related work, and the techniques that have been used to improve detectors are presented in Section 3. Sections 4 and 5 present the standard datasets and evaluation matrices used in face-mask detection, respectively. Section 6 explains the challenges and future scope of face-mask detection techniques. Finally, conclusions and future work are presented in Section 7.

2. Related work

Computer Vision (CV) techniques are crucial for recognizing patterns and performing object detection, including image classification [12]. This process involves generating region proposals that are then classified into relevant categories [13]. Infrastructure transformations in detection approaches have resulted in newfangled one- and two-stage detectors. A single-stage detector predicts the probability and bounding boxes for each object candidate within a single pass through a network. Two-stage detectors aim to find regions of interest via classification and regression, which in turn, making them more precise and correct. Techniques to increase detection from the failure rate, throughput, or robustness via feature pyramid networks, anchor box optimization, or attention systems are used.

2.1. Single-stage detectors

The YOLO method has become popular because of its real-time predictions and fast detection speeds. Still, it has had difficulty accurately locating small objects compared to two-stage detectors. YOLO's ability of YOLO to predict bounding boxes and classes without requiring an additional bounding box proposal stage makes it suitable for this purpose [14]. To improve the YOLO network, researchers introduced YOLOv2 [15], which includes batch normalization, high-resolution classifiers, and fitting bins. Subsequently, YOLOv3 [16], YOLOv4, and YOLOv5 improved the previous model until they reached YOLOv8. The most important improvements are an increase in detection accuracy, greater efficiency in terms of memory use and speed, and an improvement in boundary boxes compared to previous versions. Other detectors exist, such as OverFeat [3] and DeepMultiBox [17].

YOLOv2 was developed using batch normalization, a high-resolution classifier, convolution with anchor boxes, dimension clusters for anchor boxes, and direct location prediction. YOLOv3 was created by building on YOLOv2's foundation, with Darknet-53 as the backbone, multiscale prediction, bounding box prediction, and improved training techniques (such as random scaling). However, YOLOv3 [16] does not outperform other methods, and it requires sufficient computer power for inference, which is sometimes unavailable on embedded or mobile devices. YOLOv3 is the fastest model, while other algorithms may put accuracy at the top with the admittance of speed, or vice versa. In contrast, Single-Shot Detectors (SSDs) [18] surpass YOLO networks because of their smaller convolutional filters, multiple feature maps, and multiple-scale predications. The primary distinction between these two architectures is that the SSD network utilizes convolutional layers of various sizes. Furthermore, RetinaNet [19] is a single-stage object detector that employs a feature image pyramid and focal loss, achieving remarkable accuracy and speed comparable to those of two-stage detectors. Fig. 1 shows the basic YOLO architecture, which then proceeds to the series of YOLOv4, YOLOv5, YOLOv6, YOLOv7, and YOLOv8.

2.2. Two-stage detectors

Two-stage detectors differ significantly from single-stage detectors in their methodology for predicting and classifying region proposals in CV. Two-stage detectors initially predict

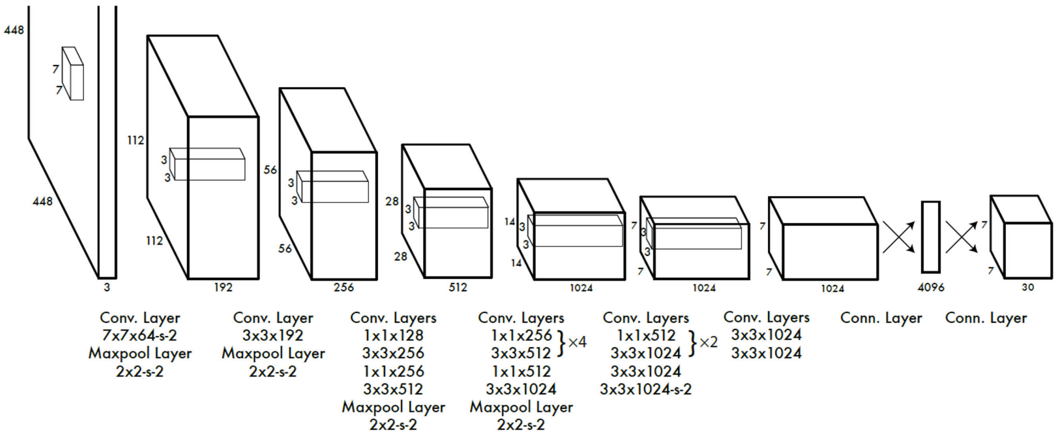

Fig. 1. YOLO architecture [14].

potential proposal regions and subsequently apply a classifier to determine the potential detections. Previous research has proposed several two-stage region-proposal models. One such model is Region-based CNN (R-CNN) [20]. The model achieved state-of-the-art results on benchmark datasets such as VOC-2012 [21].

The R-CNN model uses a selective search algorithm to generate object proposals that are classified using a Support Vector Machine (SVM) classifier at a later stage. The Spatial Pyramid Pooling Network (SPPNet) is where it can be seen how R-CNN has been upgraded for its poor computation performance. The SPPN's single-shot computing of feature maps significantly improved the object detection speed, making it nearly 20 times faster than R-CNN. Fast R-CNN [22] combines R-CNN and SPPNet, adding a Region Of Interest (ROI) pooling layer for improved detection. This allows for the simultaneous training of the detector and regressor without requiring any alterations to the network. However, it is slower than that of the single-stage detectors.

Furthermore, Faster R-CNN [23] is an improved version of Fast R-CNN that integrates the Region Proposal Network (RPN) into the object detection system, resulting in quicker detection, although at the cost of higher computational redundancy. In contrast, Region-based Fully Convolutional Networks (R-FCNs) allow complete backpropagation during training and inference, making it a popular choice among researchers. However, Feature Pyramid Networks (FPN) [24] are not widely used because of their high computational cost and memory usage despite their ability to detect non-uniform objects. Finally, Mask R-CNN enhances Faster R-CNN by predicting segmented masks for each ROI, making it a popular choice for object detection tasks. While two-stage models offer high accuracy, they are unsuitable for video surveillance's real-time inference speed. Fig. 2 shows the architecture of the CNN.

3. Improving detectors techniques

One proposal is to increase the performance of one- and two-phase detectors by upgrading the training data, employing hard negative sampling for high accuracy, and using context to improve detector speed or detection rate. Multi-scale CNN (MS-CNN) [26] and deconvolutional single-shot detector (DSSD) [27] models are superior in terms of feature representation and more efficient object detection, as they add an extra layer in a top-down manner, which enhances high-level features with detailed information from lower layers.

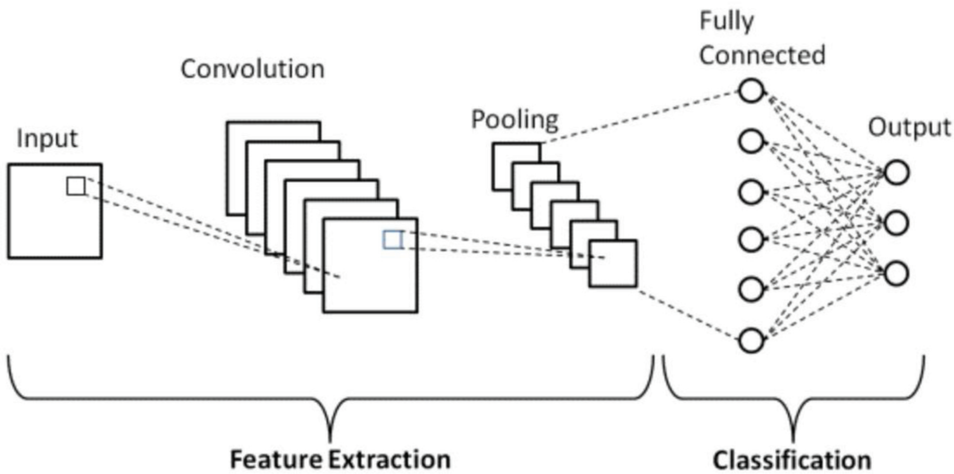

Fig. 2. CNN basic architecture [25].

The present BlitzNet [28] has a semantic segmentation method aims to improve SSD detection accuracy. The training of CNN based on open-source object detection models has significantly contributed to CV, with networks displaying high accuracy on ImageNet [29], Common Objects in COntext (COCO) [30], and ImageNet Large-Scale Visual Recognition Challenge 2013 (ILSVRC-2013), which have significantly advanced CV. These models can be adapted to address specific object recognition challenges without starting from scratch.

4. Standard dataset

Employing machine learning approaches with pre-trained models in public areas can enforce mask-wearing to reduce the spread of coronavirus in the community. These models should be fine-tuned using benchmark datasets. This section presents synthetic or real standard datasets commonly used in the literature because face mask detection methods require effort and accuracy. Table 1 shows the description of the datasets.

Generally, there are two types of face-related datasets: those focusing on masked faces and those focusing on face masks. Masked face datasets emphasize images with varying degrees of expression and landmarks, whereas face mask-focus datasets emphasize the use of masks.

The Real-World Masked Face Recognition Dataset (RMFRD) [35] is the largest available dataset for face-mask detection, which aims to improve the accuracy of face recognition models for masked faces by providing diverse images. Several operations were performed with RMFRD, which manually eliminated all facial images that resulted from incorrect parity and blurred images. Facial regions were isolated using semi-automated annotation methods. Each participant in the Masked Face Segmentation and Recognition (MFSR) dataset [37] had at least one photo of masked and unmasked faces. The AgeDB [38] dataset is a collection of handpicked images from the wild, encompassing a diverse range of 568 celebrities spanning various age groups. To increase versatility, it included four distinct verification protocols with age differences of 5, 10, 20, and 30 years. In a subsequent study, the researchers combined their data with a 3D Morphable Model to create a new dataset of 3D scans for 100 males and 100 females, offering a more comprehensive view of facial recognition. The Labelled Faces in the Wild (LFW) dataset [39] employed 13 individuals

Table 1. Description of list dataset between the total image, mask, unmask, class, and environment.

Dataset	Total of image	Mask	Unmask	Class	Environment
Synthetic CelebFaces [31]	10,000	–	10,000	1	Simulated
The Celebfaces Attributes (CelebA) [32, 33]	200,000	–	200,000	1	Simulated
Celebfaces Attributes-High Quality (CelebA-HQ) [34]	30,000	–	30,000	1	Simulated
RMFRD [35]	95,000	5000	90,000	2	Real
Masked Face Detection Dataset (MFDD) [35, 36]	24,771	24,771	0	1	Real
Masked Face Segmentation and Recognition (MFSR) [37]	21,357	9,742	11,615	2	Real/Simulated
Age DataBase (AgeDB) [38]	16,488	–	16,488	1	Real
Labeled Faces in the Wild (LFW) [39, 40]	50,000	–	50,000	1	Real
LFW-SM [41]	13,233	–	13,233	1	Simulated
Simulated Masked Face Dataset (SMFD) [41]	1570	785	785	2	Simulated
Moxa3K [42]	3000	–	–	2	Real
Face Mask Label Dataset (FMLD) [43]	41934	29532 mask/1528 incorrect masked faces	32012	3	Real
Medical Masks [44]	3835	3030/(134 incorrect masked faces)	671	3	Real
AIZOO-Tech [45]	7971	12,620	4034	2	Real

with masked faces. The Labeled Faces in The Wild Simulated Mask (LFW-SM) dataset was introduced in [41], extending the LFW dataset with simulated masks and containing 5,749 individuals. These are the most important datasets used in previous research on face mask detection. Fig. 3 shows images of the datasets.

5. Standard evaluation metrics

Evaluation metrics can differ based on the field or domain, but common metrics include the following:

1. Accuracy is a vital metric for assessing the performance of classification models. It measures the proportion of correctly classified instances in an entire dataset [46]. Mathematically, the accuracy is calculated as:

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}} \tag{1}$$

2. Precision is a measure of the performance used to assess binary classification models. It evaluates the percentage of accurate positive predictions that a model makes out of all its positive predictions [47]. Mathematically, the precision is calculated as:

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

where TP is True Positive and FP is False Positive.

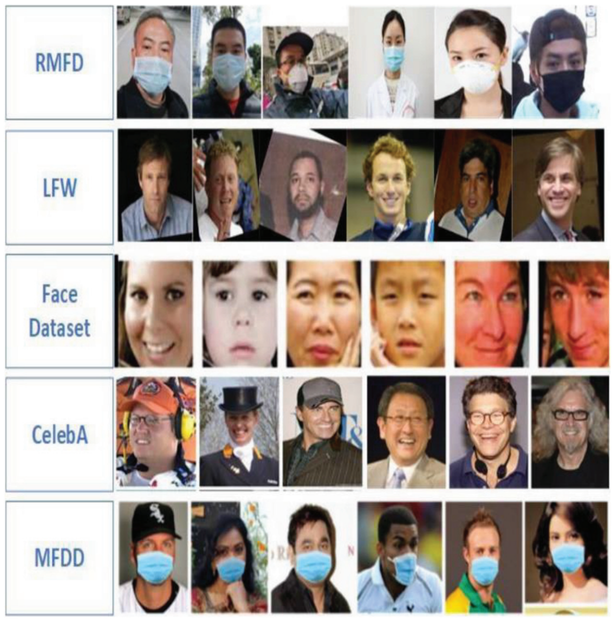

Fig. 3. Sample images from publicly available datasets [40].

3. Recall, also known as the sensitivity or True Positive Rate (TPR), is a key performance metric in binary classification tasks. It measures the proportion of correctly identified positive instances among all actual positive instances [48]. Mathematically, recall is calculated as:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{3}$$

where FN is False Negative.
4. The F1 score is a balanced performance measure that combines precision and recall [46].

$$\text{F1} = \frac{2 \times \text{Precision} \times \text{recall}}{\text{precision} + \text{recall}} \tag{4}$$

5. The error rate, often called ERR, is a performance indicator for classification models. This quantifies the proportion of misclassified instances within the entire dataset. Accuracy, however, evaluates the correct classification of instances. Therefore, the error rate is low when the accuracy is high, and vice versa. Mathematically, the error rate can be expressed as:

$$\text{ERR} = \frac{\text{Number of Missclassified Instances}}{\text{Total Number of Instances}} \tag{5}$$

6. Average Precision (AP) is a metric for ranking systems that calculates the average precision for each relevant document retrieved.
7. Mean Average Precision (mAP) is a metric that combines multiple Average Precision (AP) scores for various queries, providing a single value that represents the overall performance of the retrieval system. MAP is calculated by averaging the Average

Table 2. Performance of face-mask detection methods in terms of accuracy.

Ref.	Methods	Accuracy	Metric
[51]	Multi-Task Convolutional Neural Network (MTCNN)	98.50	mAP
[52]	FaceMaskNet-21	88.92	Precision
[37]	Generative Adversarial Network (GAN)	86.50	mAP
[53]	FaceNet	97.25	TPR
[54]	Multi-Task ArcFace (MTArcFace)	99.78	mAP
[55]	MFCosface	99.33	mAP
[56]	MaskNet	93.80	mAP
[57]	YOLOv3	78.6	mAP
[57]	Faster R-CNN	70.4	mAP
[58]	YOLO-v4	98.3	mAP

Precision scores across all queries [49]. The formula for MAP is:

$$\text{mAP} = \frac{\sum_{i=1}^{\text{N}} \text{AP}}{\text{N}} \tag{6}$$

where N is the total number of target-detection categories.

8. The Structural Similarity Index Measure (SSIM) is a measure of image similarity that considers both luminance and structural information, similar to how the human visual system perceives these aspects. It computes three comparison terms—luminance, contrast, and structure—and then calculates the SSIM index by multiplying these terms. SSIM is often used to assess the similarity between a reference image and distorted or compressed image [50]. Table 2 summarises the performance of the face mask detection methods in terms of accuracy.

6. Challenges and future scope

Challenges and potential areas for improvement in facial mask detection are currently being addressed. It is essential to consider these challenges and develop solutions to improve the efficiency of these systems.

1. Patterned masks are often stylish and feature details such as lips, nose, and chin, making it difficult for detection systems that were primarily designed to identify surgical masks. Consequently, recognizing personalized masks can be challenging, leading to increased false positives and decreased accuracy.
2. The challenge of face mask-based DL neural network methods is finding solutions to the issue of limited datasets. Still, current researchers use customized datasets to solve the problem of scarcity. The lack of sufficiently large datasets with images, which can hamper the performance of the detector, results in a few pictures. To enhance the performance of the life mask detector, future datasets should be consistent in having the same number of masked and unmasked images.
3. A significant process in model calibration is optimum parameter optimization, where the best values for model parameters, such as the number of epochs, learning rate, and batch size, and what leads to model performance have been found. Optimum learning algorithm performance is achieved by wisely selecting the best hyperparameters, which will ultimately determine the researcher's interest in what the algorithm can and cannot achieve. Specifying suitable hyperparameters is critical to ensure

successful training of the learning mechanism and to achieve the targeted performance metrics.

4- Ensuring the ethical implementation of face recognition technology in public surveillance while addressing privacy concerns and implementing face mask detection systems with adequate safeguards.

5. Researchers had previously emphasized two-class detection definitions: either “with_mask” or “without_mask” without_mask. Nevertheless, this process sidesteps the most crucial step in this move: wearing the mask correctly. Wearing a mask incorrectly is just as ineffective as not wearing a mask, which means that one needs to cover the nose and mouth when wearing the mask to prevent the emission of contagious diseases. Accordingly, in future research, a new category, “people_incorrect masks,” should be introduced, and a detector that can differentiate between correctly and improperly worn masks will be developed.

Thus, future research on face mask detection should target key areas to create more efficient and socially responsible technologies for improved public health and safety.

7. Conclusion

The COVID-19 pandemic has led to the widespread use of masks, posing challenges for face mask inspection. This paper examined face mask detection techniques and their datasets to address the issues in previous research. These techniques are commonly used in security and law enforcement. This review paper offers valuable insights to researchers by providing a comprehensive understanding of the current face mask detection algorithms. This allows for the enhancement or creation of innovative algorithms to construct a robust face-mask detection system. Despite notable advancements, there is still scope for further enhancements and advancements. In the future, various algorithms can be applied to widely recognized datasets, and their performance can be measured using a range of metrics.

References

1. R. Girshick, J. Donahue, T. Darrell, and J. Malik, “Rich feature hierarchies for accurate object detection and semantic segmentation,” in *Proc. 2014 IEEE Conf. on Computer Vision and Pattern Recognition*, Columbus, OH, USA, pp. 580–587, June 23–28, 2014, doi: 10.1109/CVPR.2014.81.
2. L. Nanni, S. Ghidoni, and S. Brahnam, “Handcrafted vs. non-handcrafted features for computer vision classification,” *Pattern Recognition*, vol. 71, pp. 158–172, Nov. 2017, doi: 10.1016/J.PATCOG.2017.05.025.
3. P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun, “OverFeat: integrated recognition, localization and detection using convolutional networks,” 2013, *arXiv: 0303149*.
4. S. Maocai *et al.*, “Neglected microplastics pollution in global COVID-19: disposable surgical masks,” *Science of The Total Environment*, vol. 790, Art. no. 148130, 2021, doi: 10.1016/J.SCITOTENV.2021.148130.
5. T. M. Saravanan, K. Karthiha, R. Kavinkumar, S. Gokul, and J. P. Mishra, “A novel machine learning scheme for face mask detection using pretrained convolutional neural network,” *Materialstoday: Proceedings*, vol. 58, part 1, pp. 150–156, 2022, doi: 10.1016/J.MATPR.2022.01.165.
6. World Health Organization, Mar. 2023, “Advice for the public: coronavirus disease (COVID-19),” [Online]: Available: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public.
7. World Health Organization, Dec. 2020 “Mask use in the context of COVID-19,” [Online]: Available: https://www.who.int/publications/i/item/advice-on-the-use-of-masks-in-the-community-during-home-care-and-in-healthcare-settings-in-the-context-of-the-novel-coronavirus-(2019-ncov)-outbreak.
8. M. Kumaraswamy, J. S. Shaji, V. S. Sowmya, and P. Paul, “General awareness regarding face masks to combat covid-19: a comprehensive review,” *Int. J. Pharm. Sci. Rev. Res.*, vol. 67, no. 1, pp. 24–29, 2021, doi: 10.47583/ijpsrr.2021.v67i01.004.

9. D. Dagar, Y. Dagar, and R. Sharma, "Face-mask recognition and detection using deep learning," in *Proc. 2022 Int. Conf. on Machine Learning, Big Data, Cloud and Parallel Computing (COM-IT-CON 2022)*, Faridabad, India, pp. 692–697, May 26–27, 2022, doi: 10.1109/COM-IT-CON54601.2022.9850742.
10. T. N. Pham, V. H. Nguyen, and J. H. Huh, "Integration of improved YOLOv5 for face mask detector and auto-labeling to generate dataset for fighting against COVID-19," *Journal of Supercomputing*, vol. 79, no. 8, pp. 8966–8992, 2023, doi: 10.1007/S11227-022-04979-2/TABLES/9.
11. A. Alzu'bi, F. Albalas, T. AL-Hadhrami, L. B. Younis, and A. Bashayreh, "Masked face recognition using deep learning: a review," *Electronics*, vol. 10, no. 21, Art. no. 2666, 2021, doi: 10.3390/electronics10212666.
12. M. Inamdar and N. Mehendale, "Real-time face mask identification using facemasknet deep learning network," *SSRN*, 2020, doi: 10.2139/SSRN.3663305.
13. S. Qiao, C. Liu, W. Shen, and A. Yuille, "Few-shot image recognition by predicting parameters from activations," in *Proc. 2018 IEEE/CVF Conf. on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, pp. 7229–7238, June 18–23, 2018, doi: 10.1109/CVPR.2018.00755.
14. J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: unified, real-time object detection," in *Proc. 2016 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, pp. 779–788, June 27–30, 2016, doi: 10.1109/CVPR.2016.91.
15. J. Redmon and A. Farhadi, "YOLO9000: better, faster, stronger," in *Proc. 2017 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, pp. 6517–6525, July 21–26, 2017, doi: 10.1109/CVPR.2017.690.
16. J. Redmon and A. Farhadi, "YOLOv3: an incremental improvement," 2018, *arXiv: 1804.02767*.
17. D. Erhan, C. Szegedy, A. Toshev, and D. Anguelov, "Scalable object detection using deep neural networks," in *Proc. 2014 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Columbus, OH, USA, pp. 2155–2162, June 23–28, 2014. doi: 10.1109/CVPR.2014.276.
18. W. Liu *et al.*, "SSD: single shot multibox detector," in *Proc. 14th European Conf. Computer Vision (ECCV 2016)*, Amsterdam, Netherlands, pp. 21–37, October 11–14, 2016, doi: 10.1007/978-3-319-46448-0_2.
19. A. Sharma, R. Gautam, and J. Singh, "Deep learning for face mask detection: a survey," *Multimedia Tools and Applications*, vol. 82, no. 22, pp. 34321–34361, 2023, doi: 10.1007/s11042-023-14686-6.
20. R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Region-based convolutional networks for accurate object detection and segmentation," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 1, pp. 142–158, 2016, doi: 10.1109/TPAMI.2015.2437384.
21. X. Zhan, H. Long, F. Gou, X. Duan, G. Kong, and J. Wu, "A convolutional neural network-based intelligent medical system with sensors for assistive diagnosis and decision-making in non-small cell lung cancer," *Sensors*, vol. 21, no. 23, Art. no. 7996, 2021, doi: 10.3390/S21237996.
22. R. Girshick, "Fast R-CNN," in *Proc. 2015 IEEE Int. Conf. on Computer Vision (ICCV)*, Santiago, Chile, pp. 1440–1448, December 7–13, 2015, doi: 10.1109/ICCV.2015.169.
23. S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analaysis & Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, 2017, doi: 10.1109/TPAMI.2016.2577031.
24. T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. 2017 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, pp. 936–944, July 21–26, 2017, doi: 10.1109/CVPR.2017.106.
25. K. O'Shea and R. Nash, "An introduction to convolutional neural networks," 2015, *arXiv: 1511.08458*.
26. Z. Cai, Q. Fan, R. S. Feris, and N. Vasconcelos, "A unified multi-scale deep convolutional neural network for fast object detection," in *Proc. 14th European Conf. on Computer Vision (ECCV 2016)*, Amsterdam, Netherlands, pp. 354–370, October 11–14, 2016, doi: 10.1007/978-3-319-46493-0_22.
27. R. Araki, T. Onishi, T. Hirakawa, T. Yamashita, and H. Fujiyoshi, "MT-DSSD: deconvolutional single shot detector using multi task learning for object detection, segmentation, and grasping detection," in *Proc. 2020 IEEE Int. Conf. on Robotics and Automation (ICRA)*, Paris, France, pp. 10487–10493, May 31–August 31, 2020, doi: 10.1109/ICRA40945.2020.9197251.
28. N. Dvornik, K. Shmelkov, J. Mairal, and C. Schmid, "BlitzNet: a real-time deep network for scene understanding," in *Proc. 2017 IEEE Int. Conf. on Computer Vision (ICCV)*, Venice, Italy, pp. 4174–4182, October 22–29, 2017, doi: 10.1109/ICCV.2017.447.
29. J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: a large-scale hierarchical image database," in *Proc. 2009 IEEE Conf. on Computer Vision and Pattern Recognition*, Miami, FL, USA, pp. 248–255, June 20–25, 2009, doi: 10.1109/CVPR.2009.5206848.
30. T. Y. Lin *et al.*, "Microsoft COCO: common objects in context," in *Proc. 13th European Conf. on Computer Vision (ECCV 2014)*, Zurich, Switzerland, pp. 740–755, September 6–12, 2014, doi: 10.1007/978-3-319-10602-1_48.
31. N. Ud 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, doi: 10.1109/ACCESS.2020.2977386.

32. Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep learning face attributes in the wild," in *Proc. 2015 IEEE Int. Conf. on Computer Vision (ICCV)*, Santiago, Chile, pp. 3730–3738, December 7–13, 2015, doi: 10.1109/ICCV.2015.425.
33. N. U. Din, K. Javed, S. Bae, and J. Yi, "Effective removal of user-selected foreground object from facial images using a novel GAN-based network," *IEEE Access*, vol. 8, pp. 109648–109661, 2020, doi: 10.1109/ACCESS.2020.3001649.
34. T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive growing of GANs for improved quality, stability, and variation," in *Proc. 6th Int. Conf. on Learning Representations (ICLR 2018)*, Vancouver, Canada, April 30–May 3, 2018.
35. Z. Wang, B. Huang, G. Wang, P. Yi, and K. Jiang, "Masked face recognition dataset and application," *IEEE Transactions on Biometrics, Behavior, and Identity Science*, vol. 5, no. 2, pp. 298–304, April, 2023, doi: 10.1109/TBIOM.2023.3242085.
36. Y. B. Chandra and G. K. Reddy, "A comparative analysis of face recognition models on masked faces," *International Journal of Scientific & Technology Research*, vol. 9, no. 10, pp. 175–178, 2020.
37. 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. on Multimedia*, New York, NY, USA, pp. 2246–2254, October, 2020, doi: 10.1145/3394171.3413723.
38. S. Moschoglou, A. Papaioannou, C. Sagonas, J. Deng, I. Kotsia, and S. Zafeiriou, "AgeDB: the first manually collected, in-the-wild age database," in *Proc. 2017 IEEE Conf. on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Honolulu, HI, USA, pp. 1997–2005, July 21–26, 2017, doi: 10.1109/CVPRW.2017.250.
39. G. Huang, M. Mattar, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: a database for studying face recognition in unconstrained environments," *Presented at Workshop on Faces in 'Real-Life' Images: Detection, Alignment, and Recognition*, Marseille, France, October, 2008.
40. S. F. Abbas, S. H. Shaker, and F. A. Abdullatif, "Masked face identification and tracking using deep learning: a review," *Journal of Computer Science*, vol. 19, no. 12, pp. 1423–1437, 2023, doi: 10.3844/jcssp.2023.1423.1437.
41. Z. Wang, B. Huang, G. Wang, P. Yi, and K. Jiang, "Masked face recognition dataset and application," *IEEE Transactions on Biometrics, Behavior, and Identity Science*, vol. 5, no. 2, pp. 298–304, 2023, doi: 10.1109/TBIOM.2023.3242085.
42. 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," *Transactions of the Indian National Academy of Engineering*, vol. 5, pp. 509–518, 2020, doi: 10.1007/S41403-020-00157-Z.
43. B. Batagelj, P. Peer, V. Štruc, and S. Dobrišek, "How to correctly detect face-masks for COVID-19 from visual information?," *Applied Sciences*, vol. 11, no. 5, Art. no. 2070, 2021, doi: 10.3390/APP11052070.
44. B. Qin and D. Li, "Identifying facemask-wearing condition using image super-resolution with classification network to prevent COVID-19," *Sensors* 2020, vol. 20, no. 18, Art. no. 5236, 2020, doi: 10.3390/S20185236.
45. X. Fan and M. Jiang, "RetinaFaceMask: a single stage face mask detector for assisting control of the COVID-19 pandemic," in *Proc. 2021 IEEE Int. Conf. on Systems, Man, and Cybernetics (SMC)*, Melbourne, Australia, pp. 832–837, October 17–20, 2021, doi: 10.1109/SMC52423.2021.9659271.
46. D. M. W. Powers, "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation," 2011, *arXiv: 2010.16061*.
47. L. Zhu, X. Geng, Z. Li, and C. Liu, "Improving YOLOv5 with attention mechanism for detecting boulders from planetary images," *Remote Sensing*, vol. 13, no. 18, Art. no. 3776, 2021, doi: 10.3390/RS13183776.
48. Z. Wang, W. Sun, Q. Zhu, and P. Shi, "Face mask-wearing detection model based on loss function and attention mechanism," *Comput. Intell. Neurosci.*, 2022, doi: 10.1155/2022/2452291.
49. A. Kumar, A. Kalia, and A. Kalia, "ETL-YOLO v4: a face mask detection algorithm in era of COVID-19 pandemic," *Optik.*, vol. 259, Art. no. 169051, 2022, doi: 10.1016/j.ijleo.2022.169051.
50. M. Ponomarenko, K. Egiazarian, V. Lukin, and V. Abramova, "Structural similarity index with predictability of image blocks," in *Proc. 2018 IEEE 17th Int. Conf. on Mathematical Methods in Electromagnetic Theory (MMET)*, Kyiv, Ukraine, pp. 115–118, July 2–5, 2018, doi: 10.1109/MMET.2018.8460285.
51. S. Ejaz and R. Islam, "Masked face recognition using convolutional neural network," in *Proc. 2019 Int. Conf. on Sustainable Technologies for Industry 4.0 (STI)*, Dhaka, Bangladesh, pp. 1–6, December 24–25, 2019, doi: 10.1109/STI47673.2019.9068044.
52. R. Golwalkar and N. Mehendale, "Masked-face recognition using deep metric learning and FaceMaskNet-21," *Applied Intelligence*, vol. 52, no. 11, pp. 13268–13279, 2022, doi: 10.1007/S10489-021-03150-3/TABLES/1.
53. A. Anwar and A. Raychowdhury, "Masked face recognition for secure authentication," 2020, *arXiv: 200811104*.

54. D. Montero, M. Nieto, P. Leskovsky, and N. Aginako, "Boosting masked face recognition with multi-task ArcFace," in *Proc. 2022 16th Int. Conf. on Signal-Image Technology & Internet-Based Systems (SITIS)*, Dijon, France, pp. 184–189, October 19–21, 2022, doi: 10.1109/SITIS57111.2022.00042.
55. J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210–227, 2009, doi: 10.1109/TPAMI.2008.79.
56. W. Wan and J. Chen, "Occlusion robust face recognition based on mask learning," in *Proc. 2017 IEEE Int. Conf. on Image Processing (ICIP)*, Beijing, China, pp. 3795–3799, September 17–20, 2017, doi: 10.1109/ICIP.2017.8296992.
57. L. Ramos and B. Morales, "SwiftFace: real-time face detection," in *Proc. XXI Int. Conf. on Human Computer Interaction*, Málaga, Spain, Art. no. 24, September 22–24, 2021, doi: 10.1145/3471391.3471418.
58. J. Yu and W. Zhang, "Face mask wearing detection algorithm based on improved YOLO-v4," *Sensors*, vol. 21, no.19, Art. no. 3263, 2021, doi: 10.3390/s21093263.