Review

Masked Face Identification and Tracking Using Deep Learning: A Review

[1]Shahad Fadhil Abbas, [1]Shaimaa Hameed Shaker and [2]Firas A. Abdullatif

[1]*Department of Computer Science, University of Technology, Baghdad, Iraq*
[2]*Department of Computer Sciences, College of Education for Pure Science/Ibn-Al-Haithem, University of Baghdad, Baghdad, Iraq*

Article history
Received: 30-07-2023
Revised: 14-09-2023
Accepted: 19-09-2023

Corresponding Author:
Shahad Fadhil Abbas
Department of Computer Science,
University of Technology, Baghdad, Iraq
Email: shahad_fadhil94@yahoo.com

Abstract: Facial recognition systems are becoming more prevalent in our daily lives. Based on artificial intelligence, computers play a very important role in the issue of identifying and tracking. This technology is mostly used for security and law enforcement. In view of the COVID-19 pandemic, government directives have been issued to citizens to wear medical masks in crowded institutions and places, which has caused difficulties in identifying and tracking people who are wearing them. This study organizes and reviews work on facial identification and face tracking. Conventional facial recognition technology is unable to recognize people when they are wearing masks. This study proposes a Masked Face Identification and Tracking (MFIT) model using yolov5, attention mechanism, and FaceMaskNet-21 deep learning architectures. Standard datasets such as "CASIA-WEBFACE, Glint360K, and chokepoint, etc." are discussed and used to evaluate the criteria relevant to face mask detection and tracking. However, numerous difficulties such as "different size of facial when movement, identification with/without mask wear and Tracking in frames or cameras" have been encountered. Additionally, consideration of the system limits, observations, and several use cases are provided. This study aims to implement a facial recognition system capable of masked face identification and tracking using deep learning.

Keywords: Mask Face Identification, Deep Learning Tracking, Attention Mechanism, Identification with a Mask

Introduction

Facial recognition systems are becoming more prevalent in our daily lives. Many people are familiar with it by using mobile phones equipped with facial recognition technology even while wearing a mask to unlock the screen as well as to authenticate logins into certain applications. This technology is mostly used for security and law enforcement. The strategy of Facial Recognition (FR) systems usually focuses on the three essential facial features, namely the eyes, nose, and mouth. However, a variety of events and situations require that people wear masks that conceal or partially conceal their features. Such circumstances include epidemics, laboratory or medical procedures, or protection against excessive pollution (Alzu'bi *et al*., 2021). For example, according to the World Health Organization (WHO), wearing face masks and practicing social distancing is the most effective way to prevent the spread or infection of the COVID-19 virus (Ngan *et al*., 2020). Facemask usage is widespread worldwide owing to the COVID-19 pandemic. Facemasks aid in infection prevention, but there are worries about how they might affect social communication (Marini *et al*., 2021). The first stage of facial recognition is detecting the presence of a human face. Face detection differs from facial recognition in that the former involves only detecting the face within a digital image or video clip. An increasing number of research publications on the topic of face detection in people wearing masks have appeared, along with face masking research data sets (Ibrahim, 2022). Artificial intelligence algorithms are used to detect the presence of a human face. Since eyes are among the simplest features to recognize, they are usually the first thing a face detection system looks for. Then, it can look for the presence of the mouth, eyebrows, or nose (Razzaq *et al*., 2022). Face masks can obscure a person's features and make them difficult to identify (Marini *et al*., 2021).

When FR systems find a face in a photo or video, they utilize one of two ways to identify it, as follows:

- When identifying the face, an algorithm tries to guess its identity. It achieves this by checking whether the image matches any of the identifiers already saved in the system
- During face verification, the system tries to verify its identity, by determining whether the face in the image matches a particular face (identity) previously stored in the system

The primary difference between facial verification and facial identification is the type of matching: Facial identification uses one-to-many matching technologies, whereas, with face verification, the technology uses a one-to-one matching system, which is usually used for app login. We note that digital cameras are increasingly prevalent for use by government agencies, hospitals, schools, casinos, and other group events or centers that require monitoring for safety reasons. Face identification algorithms must be trained and tested on a huge set of images to work properly. In addition, these images must be taken from different angles and in different lighting situations. Many proposed FR algorithms, such as faster RCNN, CornerNe, YOLOv3, and YOLOv4, aim to detect and track masked faces. There is an urgent need to identify people wearing masks. Furthermore, many organizations have already created and implemented basic data sets within the company to use FR for person authentication or identification (Libby and Ehrenfeld, 2021). Unfortunately, faces covered by masks complicate knowing people's identities as they should, thus compromising the viability of modern datasets and rendering internal FR structures inoperable (Subramanian, 2011).

In videos, it is very important to strengthen facial recognition systems by tracking them in all frames. A critical issue in computer vision is face tracking, which seeks to predict where a face will appear in a video (Tathe *et al*., 2017).

Face tracking and face detection: While the two processes are similar, face tracking requires less processing power than a face detector per frame. Online tracking differs from offline tracking in that the former is more suitable for real-time applications because it only uses the previous and current frames of a single recorded video. The trace-by-detection architecture used in traditional methods seeks to customize detections across frame bases (Peng *et al*., 2020).

Recently, leading technologies powered by Deep Learning (DL) systems have greatly advanced their real-world applications (Ouyang *et al*., 2015). Deep learning-based techniques that have been introduced include Occluded Face Recognition (OFR), Masked Face Recognition (MFR), video-based object tracking, and Multiple Hypothesis Tracking (MHT). Despite these developments, the recognition of faces will remain very challenging for a long time for faces occluded in images. OFR has been used in many applications to, for example, track the movements of the mask wearer in real-time time (Prasad *et al*., 2021). However, regarding the development and evaluation of algorithms, techniques, datasets, and methodologies suggested in the literature for dealing with veiled or hooded faces, there is no widespread agreement. While there is no doubt that the set of DL algorithms dedicated to finding and identifying people wearing masks is valuable, it is also necessary to review and evaluate the consequences of these technologies in the field.

In this study we conduct this review study to create a complete resource for people interested in the problem of mask face recognition and tracking, considering these important developments and related challenges. The typical face-mask detection, face-identification, and face-tracking methods are discussed. This review study aims to provide ways of solving problems, such as illumination, pose, low resolution, identification, identification with a mask, and face tracking, which are reviewed by adding an attention mechanism in the YOLOv5 model to be higher accurate. Finally, we present the evaluation results for the video sequences along with some observations about each classifier's ability to discriminate.

This study is organized as follows: In section related work, the study begins by introducing the scope of the research and presenting some statistical information regarding prior works. In the section Standard Datasets, the benchmark datasets employed for Masked Face Recognition (MFR) are outlined, section Proposed MFIT presents the proposed Masked Face Identification and Tracking (MFIT) model, section Attentional Mechanism presents how the Attention Mechanism (ATM) works and how useful it is in the field of MFR, section discussion highlights the main discussion outlining the scenarios associated with the model and section conclusion concludes this comprehensive study.

Related Work

When it comes to computer vision, FR is a widely studied task that can take advantage of image categorization, object recognition, and object location, including social media. Spatial detection involves designing a bounding box that must be created around each object of interest in the image tag. On the other hand, image classification refers to identifying and labeling groups of pixels, or image vectors, that define the object class based on specific rules. Many face-tracking studies have suggested detection-tracking methods to take advantage of the high accuracy that contemporary face detectors can achieve. Manuscripts and references were retrieved from Scopus, IEEE Xplore, Wiley, the Web of Science, and Digital Library.

Object detection and face detection: The Convolutional Neural Network (CNN) (Kaur *et al*., 2022) achieves quite high performance in extracting features from images, after which those features are learned through multiple hidden layers. The technology recognizes the face in the photo or video and then determines whether it contains a mask or not. Regions-CNN (R-CNN) was able to attain fairly good performance, but it also had several flaws that were fixed in later iterations, such as:

(i) The training process requires multiple steps
(ii) Because the extracted features from the suggested areas are saved to a CD, this requires a considerable amount of space and a lot of time

As compared to earlier implementations, Fast R-CNN (Singh *et al*., 2021) was able to obtain relatively high performance in an object detection model in terms of speed, training times, and accuracy. Mask R-CNN (He *et al*., 2017) solved many of the problems by feeding the entire CNN image to the network, rather than just area suggestions, to get a single map of convolutional attributes. The specific architecture builds a frame to surround the detected objects while creating segmentation masks for them to reveal their exact features. This makes it more effective and has real-time detection capabilities (Safaa El-Din *et al*., 2020). The first and most crucial step in FR is face detection. Because human faces vary in terms of attitude, expression, position, orientation, skin color, the existence of glasses or facial hair, camera gain variations, lighting conditions, and image resolution, it is difficult to identify faces in an image. A lot of research has been done recently to improve the accuracy and complexity of FR and detection. The Viola-Jones real-time face detector, which can accurately recognize faces in real-time, has revolutionized the sector (Ephraim *et al*., 2009). Face detection is a computer vision problem that involves finding faces in a digital image, with the main challenges being occlusion, illumination, and a complicated background (Hasan *et al*., 2021). For face detection techniques, Yan, Kriegman and Ahuja (Yang *et al*., 2002) presented a categorization. The face detection algorithms can be categorized into two or more of the four categories that these methods fall under (Siddiqui *et al*., 2020). The following are the categories: Image-based approaches depend largely on image scanning, which is based on windows or sub-frames; feature-based approaches, which involve locating faces by extracting structural features of the face (image edges, corners); template-matching approaches, which use pre-described or parameterized face templates to find or locate the faces via the correlation among the templates and enter input images and face models, which can be constructed by way of edges just using the edge detection technique (Hasan *et al*., 2021).

Facial recognition and face identification: FR pertains to a streamlined image analysis and pattern recognition application that has attracted a lot of interest during the past three decades. The human face is not the best method when compared to other biometric features because it is often less accurate than fingerprint or iris methods and may be affected by cosmetics, disguise, and lighting. Adjabi *et al*. (2020), an analysis was presented that considered the most recent state-of-the-art methodology as well as the background of facial recognition technology and its potential future developments. According to the study, in the verification mode of the FR assessment protocols, the system verifies the identity of the person by matching it with a face stored in the database. To assess whether the demonstrated identity is genuine or untrue, the framework does a 1-1 analysis to verify it (Ahmed *et al*., 2019). Check meaning Work has been undertaken with an input camera that takes many pictures of the individual. Then, an application that generates many human templates uses the cascading classification algorithm, which finds facial features that are saved in a database defined by their identifier. Moreover, by comparing the models stored within the database, the verification process was started (Siddiqui *et al*., 2020). Facial features are among the most common physical features of a man or woman that reveal their gender identity and age (Bashbaghi *et al*., 2019). Two main methods can be utilized to verify people's identities using face images. The first is face comparison, often called “face mapping” which is a method of comparing and analyzing two or more images against a digital image of a person and requires little effort. The second method is a software application called Automated Facial Recognition (AFR), which compares a database of facial images with that of interest based on the individual’s biometric data using computer algorithms to select the most similar database images (Ning *et al*., 2022). The length, width (nose, chin, and mouth), eye spacing, and other crucial features are measured by facial verification systems. Several studies have already shown that approaches utilizing deep facial features perform better in facial verification than methods that use facial landmarks (Lee *et al*., 2019). The processes for automatically identifying and validating a person from either an image or a video are included in FR. However, there are still challenges to be addressed, including increasing the accuracy of facial recognition, different methodologies for misalignment, posture variability, illumination variance, and expression variability (Teoh *et al*., 2021).

Face-mask detection models: The author used ResNet50 as a starting point and pursued the idea of learning to integrate high-level semantic data into multiple feature maps. He also suggested modifying the bounding box to improve translation efficiency during mask recognition. Three well-known base models, ResNet50, AlexNet, and MobileNet, were used in the experiment (Sethi *et al*., 2021). He studied how to combine these models with the proposed one to obtain highly accurate results with shorter inference times. When used with ResNet50, the proposed technique was found to produce high accuracy (98.2%). In Chen and Sang (2018), a Gaussian Mixture Model (GMM) was used to build a human face model for face mask identification technology to prevent fraud. In the context of financial security precautions, the issue of distinguishing between a face and a mask was discussed. This information was utilized to determine how similar a sample of faces was to the model. Whether an image is a human face, or a mask was determined by dissecting and recognizing facial features. When utilized properly, a face mask detector may assist in ensuring everyone's safety (Dagar *et al*., 2022). A semi-Siamese network was established to define the issue of masked facial recognition in the visible-to-near-infrared range. The authors found that detecting a mysterious face in near-infrared probe images is a challenging task (Du *et al*., 2021). In Himeur *et al*. (2023), research focused on face mask detection to ensure safety in public areas of smart cities during social distancing and also to detect if masks are worn correctly. The focus was on deep learning techniques and modern developments for all stages of processing, in addition to confirming some existing problems, including the different degrees of obstruction (e.g., angle of view, illumination, and resolution), the lack of publicly available, annotated real-world datasets and the computational cost requirements for AI-based video analytics.

A deep-learning model was created to determine whether someone is wearing a mask. It was trained with the help of Open CV and Keras. MobilenetV2 image classification technology was used to reduce training time by using 2,000 images to train the model (Machiraju *et al*., 2021). The yolov5 model was used to identify people wearing masks in the streets and public places and a performance accuracy of 95% was achieved (Du *et al*., 2021). To detect face masks and person identities faster than other models, a new face mask detection and person recognition technology called Insight Face was applied, which was based on the SoftMax ArcFace Loss classification algorithm. Sheikh and Zafar (2023) proposed an RRFMDS system capable of processing 7 frames per second and processing 1 frame in 0.1320 sec of live video data. It uses a single-shot multi-box detection system to automatically detect the face mask in real time. This study (Sharma *et al*., 2023) recalls previous reviews about face mask detection and the data sets and techniques used where non-neural algorithms and neural algorithms are structured, giving a broader perspective to the researcher in face mask detection (Adhikari *et al*., 2016).

Face tracking: Hybrid method for face detection and tracking, where each frame starts with face detection, and then the tracker is applied to the detected faces in the next frames. The first enhancement is performed by using a manual thresholding approach to reduce type I false positive alarms. On the other hand, template matching based on face tracking is used to reduce the detection of errors (false negatives) for faces. The major goal is to enhance the video-based frontal face detector. Face tracking and face detection are two operations that are performed using different methods. The primary function of these actions is to identify and track the face when applying surveillance, even under difficult lighting conditions. This study discusses various face-tracking systems, their benefits, and their drawbacks, as well as provides a comparison of these technologies. In Tathe *et al*. (2017), a face-tracking system was presented that can recognize faces using Gabor feature extraction, match faces using a degree of correlation, and track faces using a Kalman filter. In Goyal *et al*. (2017), the aim was to illustrate ideal conditions for facial recognition. A tool for identifying and tracking faces in photos and videos that can be applied to many tasks was introduced. The goal of the study (Goyal *et al*., 2017) is to perform a comprehensive face detection examination using an open CV.

Our study stands out from other reviews because it provides a comprehensive analysis of recent advances and algorithms that have been made in the field of mask face recognition and tracking. It focuses on face identification, masking detection, matching, and tracking deep learning algorithms, with benchmarking data sets also being provided.

Standard Datasets

The key aspect of developing facial recognition systems is the training data used to learn face representations. Firstly, we present a dataset from previous work on masked face detection and tracking.

The dataset called CASIAWebFace (Yi *et al*., 2022) contains more than 500,000 photos of 10,000 people. The CASIA group (Zhang *et al*., 2012) collected them automatically and polished them manually. This group

DOI: 10.3844/jcssp.2023.1423.1437

displays a long distribution of images related to a topic, while others are well represented only by a small number of images. The Labeled Faces in the Wild (LFW) dataset comprises about 50,000 (Huang *et al*., 2008). The 3.31 million images that make up the VGGFace2 (Cao *et al*., 2018) dataset were used with 9,131 individuals, each representing a different personal identity, and less than 1% of the training set images had a resolution of lower than 32 pixels, with an average of 137×180 pixels (taking into account the shorter side).

The face data set (with/without mask data set) from (Face Mask Detection Dataset-Kaggle, 2022) is the data set used in this study (Kaur *et al*., 2022), A total of 3,832 images from the dataset were divided into two mask categories: Images taken in 1914 with a mask and images taken in 1918 without a mask. One of the most used datasets for masked people is the Realistic Masked Faces Dataset (RMFD) (Wang *et al*., 2023b; Das *et al*., 2019) which is available in the GitHub repository and is the largest masked face dataset in the world. Another dataset used is Celebrity Facial Attributes (CelebA) (Liu *et al*., 2015), which includes over 200,000 celebrity photos with 40 binary feature annotations. The Masked Face Simulation Dataset (SMFD), which includes 1,315 images (658 masked and 657 without), is used for training and validation, with a test set available in the public domain (Negi *et al*., 2021). An accurate masked face detection model can be trained using the Masked Face Detection Dataset (MFDD), which includes 24,771 disguised facial images and it can also be used to determine whether someone is wearing a mask. In Shahar and Mazalan (2021), two datasets were used, with the first consisting of 686 people without masks and 690 images of people wearing them. This dataset was used for data pre-processing and CNN training. It is publicly available and can be used for research. A self-generated dataset containing a set of 35 images was used as a second data set for the same identification technique. Two images, one with a mask and one without, were included for each person. The primary goal of these categories is to avoid problematic findings from the face mask detection and identification test. As for the tracking aspect, tracking faces in crowds is another area for which many datasets are available. There are two datasets that are widely used (MS1M) (Guo *et al*., 2016) and Glint360K (An *et al*., 2021); the MS1M dataset includes 5.1 million images from 93,000 IDs and the publicly available Glint360K suite includes 17 million images for 360,000 people (Deng *et al*., 2021a). In addition, there is a dataset from Chokepoint that presents 48 video clips, each of which presents a different topic that passes through one of two portals (Wong *et al*., 2011). Figure 1 presents some examples of datasets. (Table 1). Summary of some of the data sets in the studies mentioned above.

Proposed Masked Face Identification and Tracking (MFIT) Model

The proposed mask face detection and tracking model involves several complex steps, as shown in Fig. 2, which are described in this section. Deep learning models, which are frequently used to learn the distinctive characteristics of masked faces, are the basis of the general methodology. As described in the following subsections, many critical steps are put in place to establish the final identification and trace system.

Table 1: Summary of datasets in the studies

Dataset	Size	Method
Yi *et al*. (2022)	500,000	Learning large margin cosine loss
Wang *et al*. (2022)	95,000	Face cropping
LFW (Huang *et al*., 2008)	50,000	Combination of ArcFace loss and mask-usage classification loss
VGGFace2 (Cao *et al*., 2018)	3.31 million	IAMGAN with DCR
Face dataset (Face Mask Detection Dataset-Kaggle, 2022)	3,832	Mask the face with face net
Celebi (Liu *et al*., 2015)	Over 200,000	Object detection and image completion
MFDD (Wang *et al*., 2023a)	24,771	Deep metric learning
MS1M (Deng *et al*., 2021a)	5.1 million	Insight face track
Glint360K (Deng *et al*., 2021b)	images 17 million images	Insight face track
Chokepoint (Wong *et al*., 2011)	48 video clips	1) Score fusion method using pixel-based asymmetry analysis and two sharpness analyses 2) Asymmetry analysis with Gabor features 3) The classical Distance From the Face Space (DFFS) method

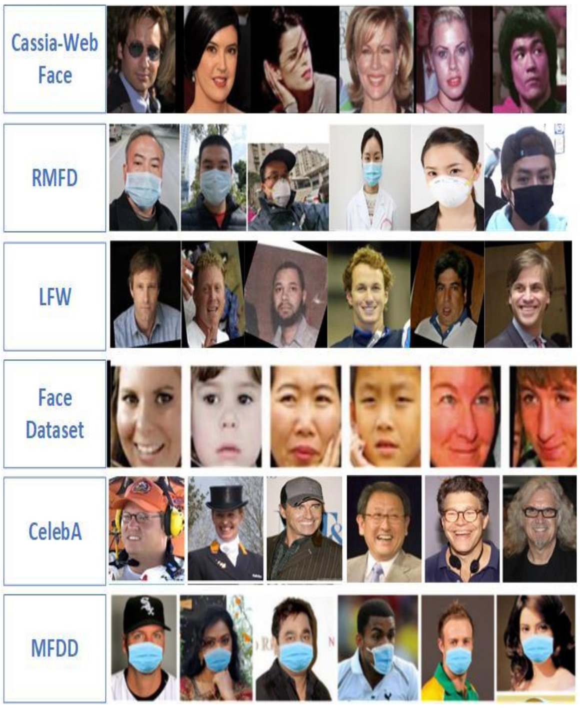

Fig. 1: Some typical images from common standard datasets

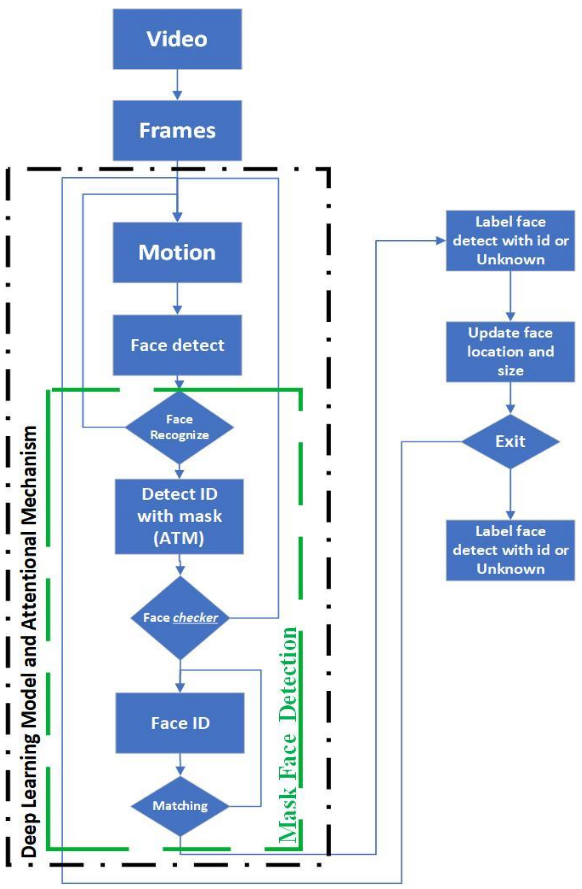

Fig. 2: Description of steps of mask face identification and tracking

Analysis of Face Video Image Databases

To develop effective methods for detecting faces and determine the characteristics, constraints, and factors that affect the recognition rate of existing databases. There is no publicly available facial video dataset with identified people wearing face masks. As a result, some simulation techniques are used to image face masks on popular face-tracking datasets, such as the Choke Point Video Database (Peng *et al*. 2020). Frames are collected in a set amount of time and face detection and tracking are used to find faces in the available video images. Faces are indexed after they are found in each frame. These indexed faces are then considered inputs to the later stages of the video-based facial recognition system. The next subsection provides reviews of current approaches at each of the abovementioned stages, as shown in the following Fig. 2, for a video-based face mask detection and tracking system.

Image Preprocessing (Frames)

The type of video-based face used in the training, validation, and testing phases has a huge impact on how face detection and tracking systems work and whether masks are used (Din *et al*., 2020). Data augmentation and data simulation have been widely used in image preprocessing, allowing for several operations to be performed to increase the quantity and variety of images, such as smart crop, crop by objects, flipping, skewing, rotating, and alignment. Other optimization techniques, such as segmentation, de-noising, or smoothing, are also used to enhance the accuracy of image representation. In addition, image correction can be performed to increase sharpness and one of the most widely used methods is Laplacian contrast (Hong *et al*., 2021). Several techniques divide an image into local parts known as segments to produce better representations of images. These local parts are, then, represented by the ordered characteristics of facial features or a set of discriminative components. Some methods, such as low-order organization (Qian *et al*., 2014) and sparse representation (Wright *et al*., 2008), describe the input of a still image with a generic descriptor, while others send the image to an existing function for facial feature recognition (Yang *et al*., 2011).

Motion

The main requirement for motion-based face detection is the presence of many continuous image frames or video sequences. Moving objects and targets provide useful information that can be used to find faces. Moving image contour features and frame contrast analysis are two of the most commonly used methods for detecting visual motion. Any type of background can be used in tire contrast analysis to determine the moving forepart. By defining the difference in the assembled frame, moving objects that include a face can be recognized. This method can be used to extract facial features in addition to the facial area (Thein and San, 2018; Crowley and Berard, 1997). A spatiotemporal Gaussian filter was used by Hasan *et al*. (2021). Visual flow analysis is an additional form of advanced motion analysis. We need to consider short-range and sensitive movements to detect faces. An accurate estimation of apparent brightness velocity is essential for optical flow analysis. The initial facial movement is determined and then the data is applied to identify the face, with a line clustering approach being proposed in which the picture velocity is thresholded to identify the moving parts of a face. Reliable and accurate tracking is provided by motion analysis analysis (Hasan *et al*., 2019). In addition, the analysis has a reduced search area because it primarily focuses on movement and excels in a real-time environment.

Face Detection

The first stage of the facial recognition system is face detection. This system of units uses image processing techniques to analyze a frame from a video sequence to identify the Candidate's facial areas (Siddiqui *et al*., 2020). The system can work on still images, in this case, the face localization procedure along with the face-tracking procedure is used while working with the videos. The goal of face localization and background extraction is to isolate the face area. Skin texture, movement, face/head shape, facial appearance, or a combination of these parameters can be used to perform face detection (Hasan *et al*., 2021). The sub-window scans an input image at all possible positions and scales. Classifying the pattern in the sub-window as either a face or a non-face is how face detection is suggested to work. Then, facial features tracking is initiated using face detection methods similar to those used in still images to determine the exact location of the faces in the current frame. Header tracking involves viewing the head as a solid object that can be translated and rotated (Razzaq *et al*., 2022). This is one of the face-tracking techniques. While using an estimate of the locations of a face or features in the previous frame, the tracker determines the exact position of the face features in the current frame(s). There are usually three main processes involved in detecting a face in videos. The first is frame-based detection, which uses many traditional methods for still images, including color-based face detection, statistical modeling, neural network-based methods, SVM-based methods, and others. The second method is combined detection and tracking, which involves detecting a face in the first frame and following it throughout the entire sequence. The third is called the temporal approach, which uses temporal relationships between frames to identify many human faces in a video sequence rather than detecting each frame.

A technique for identifying and following the characteristics of a face in a video sequence called edge pixel counting was proposed. The upright and frontal views of the images are entered in grayscale and the output is generated, with values ranging from 1 to -1; indicating the presence or absence of faces. The different lighting conditions and dynamic background were not included by the author. In the CMU database, this method works best with an acceptable discovery rate (Li *et al*., 2020; Najibi *et al*., 2017). Face detection has recently been improved in difficult conditions, such as partial or obscured faces, faces captured via depth sensors, and faces in violent settings (Yang *et al*., 2017).

Face Mask Detection

Face masks, which are available in a variety of designs, sizes, textures, and hues, have become one of the most popular items that cover facial features.

Facial Recognition

The performance of any FR algorithm is affected by the quality and variety of data used in the training and testing phases and in uncontrolled environments, as are the lighting conditions, camera views, posture, and face-to-face distance. In addition, wearing a face mask further reduces the effectiveness of the system. To automatically identify individuals and offer matching results from a given database without removing the face mask, CNN, AlexNet, and VGG16 were used for several mask detection tasks (Song *et al*., 2022). In terms of facial recognition challenges, a pipeline of disguised facial recognition using a modified SVM classifier and XGBoost, along with an updated FaceNet face detection device, was deployed. FR using Occlusion Masks (FROM) is a technique for recognizing faces with occlusions, as (Qiu *et al*., 2021) proposed, based on a single end-to-end deep neural network. Exact feature masks are recognized, deep CNNs are used to find damaged features and then the learned masks are applied dynamically to clean them up. In addition, the authors successfully trained by creating voluminous facial images. They looked at a variety of datasets that masked the face, including LFW, Mega face Challenge 1, RMF2, and AR. Leeway Hertz offers face mask detection (security checks) applications for surveillance camera systems. The program offers an interface for adding facial and identification details, such as phone numbers and ID numbers, as well as a control panel for real-time monitoring. A notification will be sent via the app when it detects someone not wearing a mask in real-time. Moreover, it can detect the absence of a mask and identify the person if they did not wear the mask; however, it cannot distinguish between the different types of masks or where they are properly in place.

Currently, it is possible to tell whether someone is wearing a mask with over 99% accuracy (Song *et al*., 2022).

Face ID

When the faces were first seen with the masks on, as opposed to when they were first seen without them, the accuracy of the re-identification (facial recognition techniques) task dropped dramatically, according to the data. As expected, people who appeared without masks before were easier to identify than those who previously wore masks. This is most likely explained by the fact that recovery and encryption were compatible under unmasked settings but not under disguised ones. When the mask faces are found in the search proposal, they are individually sent to a neural network for additional identity investigation. The step needs to enter a specific size, and for inputs of a fixed size, one option is to rearrange the face within the bounding box. The face might be staring in a different direction, which could be a problem with this solution. However, this problem is solved fairly easily using the affine transformation. The approach is similar to the deformed fragment models described in (Girshick *et al*., 2014). For identity recognition using Procrustes similarity analysis, only 31 landmark points at the nose, brows, and eyes were picked because they were unobstructed by the face mask (Shahar and Mazalan 2021; Wang *et al*., 2023a). Hybrid models are used in the two-step facial recognition method, according to Loey *et al*. (2021). ResNet 50 is used to extract features in the first stage, such as the Support Vector Machine (SVM), and decision trees are used for classification in the second stage. An overall performance of 99.49% was achieved for the classifier. Some studies have suggested alternative methods where incomplete facial characteristics are used for face detection. For example, one method is to simulate incomplete facial features using mask projection techniques (Alyuz *et al*., 2013).

Face Matching and Recognition

Face matching is a challenge for deep features FR and Masked Face Recognition (MFR) (Deng *et al*., 2021b). A set of images of many people must first be submitted to the program through the training and validation stages to complete this task. While the test is taking place, an image of a person known or unknown to the program is made available so that it can decide whether to recognize it or not. An appropriate loss function must be created and used for the successful learning of a set of deep features (Zhang *et al*., 2021). The MFR community often uses 1-1 and 1-many (N) matching techniques. For both strategies, widely used standard distance metrics, like cosine and L2, based on Euclidean geometry, are frequently employed. Several techniques have been utilized, such as metric learning (Yu *et al*., 2020) and sparse representations (Yang *et al*., 2021). These were developed to increase the amount of deep feature discrimination to increase the performance, speed, and efficiency of the face-matching process. Loss-based Triple and softmax models have been frequently used in deep technologies for matching and verifying facial identities. Loss-based Softmax models work by applying a softmax activation function to train a multi-class classification concerning one class for every identity in the training dataset (using a one-versus-all approach) (Song *et al*., 2019).

Conversely, triplet loss-based models (Schroff *et al*., 2015) have the advantage of instantly learning to embed by comparing the results of different inputs to reduce the intra-class distance and as a consequence, to increase the inter-class distance. However, face mask occlusion impairs the accuracy of softmax and triplet loss-based models Lane (2020). To address MFR tasks, a lot of research has recently been made available in the literature. Generative Adversarial Networks (GAN) based technologies have been utilized to detect faces before entering them with FR technology (Din *et al*., 2020a). The focus is on the upper part of the face because it is the only visible part for extracting the features (forehead and eyes) and a sticker is placed to cover the nose and mouth area, thus being neglected Lane (2020). Effective methods have shown high FR performance, for example (Du *et al*., 2021). Augmented mask faces were combined with the VGG2 dataset (Cao *et al*., 2018) and trained using the original pipeline defined in FaceNet (Schroff *et al*., 2015). It allowed the model to determine whether a person was wearing a mask or not without having to identify the person. A complete training pipeline for ArcFace-based facial recognition models for MFR was provided by Montero using Domain-Constrained Ranking (DCR). It is based on the MFSR dataset. They were able to identify two centers for each identity that corresponded to the images for the visible and masked faces, sequentially.

Attentional Mechanism (ATM)

In difficult scenes, people can easily identify or focus on important areas. Inspired by this observation, to simulate this feature of the human visual system, attention mechanisms have been added to computer vision. During feature extraction, the attention mechanism may force the system to focus more on Regions of Interest (ROIs) than on non-ROIs. Also, incorporating the attention mechanism into a deep learning network can greatly enhance feature extraction and simplify the model (Gan *et al*., 2022). The application of attention-based methods has gained significant attention in recent years for dealing with a variety of problems in vision tasks, including image classification (Woo *et al*., 2018), age-invariant face recognition (Yan *et al*., 2022; Li and Lee, 2022) and in particular, facial recognition using face masks. In addition, it should be highlighted that when compared with other methods, current attention-based MFR methods have shown great accuracy (Deng *et al*., 2021a). To verify people with masked faces using the ArcFace Attention and Loss of Angular Margin test, (Pann and Lee, 2022) propose a solution to the MFR problem. Through the proposed method, the model can obtain highly discriminatory features and enhance representations of facial features, effectively solving the problem of MFR recognition accuracy. To determine whether subjects are wearing masks, a high-performance portable mask detector based on YOLOv5 and attentional mechanisms is presented in this study (Xu *et al*., 2022). For a wider receptive field, ShuffleNetV2 can be used as a backbone to increase the size of the kernel in each deep convolution from three to five. In addition, the author incorporated coordination attention (Hou *et al*., 2021) into it to create the new ShuffleCANet backbone, which is still light but more efficient thanks to the attention mechanism.

Spatial Attention Module (SAM)

For people wearing masks, the problem of identification is related to certain parts because not all of the face is visible, so we need to focus on them. The features extracted from visible parts of the face inthese areas are very important (Pan *et al*., 2020). After the transfer of image characteristics in the feature extraction unit, the effect of the spatial attention unit is to give more weight to the parts that contain only useful information without giving any weight to the parts that do not contain useful information. The spatial attention module processes the image many times and the face will be broken up into several blocks. Above all, it will be possible to obtain many feature maps with important data using the initial compression and then, use these feature maps to determine the weights' significance for each spatial location. The results are quite different when SAM is combined with a feature extraction module (Chen *et al*., 2021). In Gan *et al*. (2022), Fig. (3) Shows the step-by-step feature processing and the short scheme of the spatial attention module. The spatial attention model receives the feature pattern as its first input. A 3×3 convolution layer 1×1 and average after features create the maximum F^{s}_{Max} average F^{s}_{Avg} pooling layers, followed by the action of a weighted convolution on the features generated by the sigmoid operation as a space attention map as a mask for attention action.

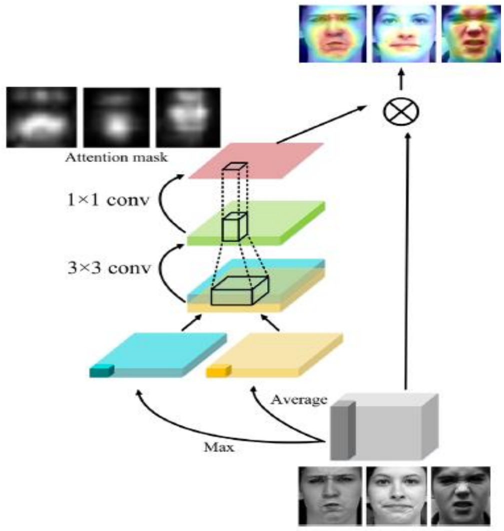

Fig. 3: Structure of the SAM (Gan *et al*., 2022)

Channel Attention Module

Each high-level feature channel map can be thought of as a class-specific answer and many semantic responses are related to each other (Gan *et al*., 2022). We can highlight correlated feature maps and enhance the representation of features for specific semantics by taking advantage of the interdependence between channel maps. Therefore, to formally represent the cross-correlations of channels, we construct the attention channel module (Li and Lee, 2022). In other words, the final feature representation of each pixel can be produced by adding the original features and the weighted total of all pixel properties, which is useful for simulating complex contextual connections across image regions. In addition, it can be used to improve feature distinguishability and strengthen category-dependent feature maps.

Discussion

Three scenarios were used in the tracking algorithm's experimental evaluation. The subjects approach the front camera in the first scenario. This is a mock-up of someone walking up to an immigration official or an ATM. The moment the face enters the camera's field of vision, it is recognized and then it constantly tracks it as it moves. The face is recognized by a tracking algorithm in the first step and then it is tracked by adjusting the angle and tilt of the camera so that the face is always in the center of the image. Also, it adjusts the camera's zoom by a factor of roughly 1.2 every time until the face's height roughly corresponds to the frame height. Keep in mind that the captured face's spatial resolution doubles after every two zoom iterations, increasing with each operation by the square of the increase in the face's height.

Table 2: A summary of FR approaches' accuracy

Dataset	Size	Model	Accuracy	Requirements
CASIA-WEBFACE Yi *et al*. (2022); (Huang *et al*., 2008)	500,000	MFCos face	99.33	FaceNet, CBAM, Maxout
LFW (Huang *et al*., 2008)	50,000	MTArc Face	99.78	ArcFace
Wang *et al*. (2022)	95,000	MF Cosface	98.14	ArcFace
VGGFace2 (Cao *et al*., 2018)	3.31 million	GANs	86.50	FaceNet
Face dataset (Face Mask Detection Dataset-Kaggle, 2022)	3,832	Mask TheFace	91.46	FaceNet
Celebi (Liu *et al*., 2015)	Over 200,000	GANs	95.44	Encoder-Decoder
MFDD (Wang *et al*., 2023a)	24,771	FaceMask Net-21	88.92	Face_MaskNet
SMFD (Negi *et al*., 2021)	1,315	ResNet-50	47.00	ResNet-50

Table 3: Performance of YOLO series methods with attention mechanisms for face recognition

Methods	Dataset	Metric	Accuracy
YOLOv5s (Al-Tamimi and Mohammed Ali, 2023)	1. AIZOO2. MoLaRGB_ CovSurv	mAP	With mask 0.953, without Mask 0.916
YOLOv5 (Xu *et al*., 2022)	AIZOO	Precision	95.2%
CBAM and the YOLO-v4 model are integrated (Zhao *et al*., 2023)	MAFA + WIDER	mAP	93.56%
YOLOv5 by adding coordinate attention (Pham *et al*., 2023)	Kaggle, YouTube	mAP	96.8%
YOLO-GBC (Wang *et al*., 2023a)	The mask detection	mAP	91.2%

In the second scenario, which is the same as the first, the subject's approach is in front of the camera with a spatial resolution of the captured face extraction feature. The camera detects the ID for the face as it appears in it. The identification algorithm detects the ID face in the first frame and then tracks it by changing the pan and tilt of the camera so that the face is always centered in the frame.

In the third scenario, the subjects move across a path. This is a reproduction of a person passing through an airport terminal or a mall's main entrance. In contrast to the first scenario, the camera's pan and tilt needs are lower in this situation because of the distance between it and the subject. Nevertheless, not only do additional zoom operations need to be performed, but the camera also needs to initially zoom in on the target before beginning to zoom out when the subject moves closer to the camera. For the time being, no standards have been established for the assessment of PTZ tracking algorithms. As a result, we use the ratio of the monitored face height (measured in pixels) to the frame height as our assessment criterion. Table 2 summarizes the MFR approaches' accuracy. Table 3. Shows the performance of yolo methods and attention mechanisms in MFR.

Conclusion

The emergence of the COVID-19 pandemic led to widespread mask-wearing, posing challenges in identifying and tracking individuals. this study provided a comprehensive examination of face recognition, masked face identification and tracking technologies, placing special emphasis on deep learning architectures and addressing certain challenges in identification and tracking. These technologies are predominantly employed in security and law enforcement contexts. Conventional facial recognition systems struggle to identify individuals when masks are worn. To address this issue, we proposed the Masked Face Identification and Tracking (MFIT) model, leveraging deep learning architectures including Yolov5, Attention Mechanism, and FaceMaskNet-21. We discussed the datasets on which previous research has relied and the problems they encountered, with a focus on standards related to face mask detection and tracking. However, we encountered various challenges, such as handling variations in facial size during movement, when the frame consisted of angled faces, and tracking individuals across frames or cameras. In addition, there are no video datasets that include the identity of individuals wearing masks to identify and track them. Hence, it is recommended to consider creating a dataset to meet this requirement. We also considered system limitations, observations, and multiple real-world use cases.

Author's Contributions

Shahad Fadhil Abbas: Participated in all experiments, coordinated the dada-analysis.

Shaimaa Hameed Shaker: Reviewed and audit identification.

Firas A. Abdullatif: Reviewed and audit tracking.

References

Adhikari, G., Sahani, S. K., Chauhan, M. S., & Das, B. K. (2016). Fast real time object tracking based on normalized cross correlation and importance of thresholding segmentation. In *2016 International Conference on Recent Trends in Information Technology (ICRTIT)* (pp. 1-5). IEEE. https://doi.org/10.1109/ICRTIT.2016.7569517

Adjabi, I., Ouahabi, A., Benzaoui, A. & Taleb-Ahmed, A. (2020). Past, Present, and Future of Face Recognition: A Review. *Electronics*, *9*(8). https://doi.org/10.3390/electronics9081188

Ahmed, A., Hasan, T., Abdullatif, F. A., Mustafa, S. T., & Rahim, M. S. M. (2019, October). A digital signature system based on real time face recognition. In *2019 IEEE 9th International Conference on System Engineering and Technology (ICSET)* (pp. 298-302). IEEE. https://doi.org/10.1109/ICSEngT.2019.8906410

Al-Tamimi, M. S., & Mohammed Ali, F. A. (2023). Face mask detection based on algorithm YOLOv5s. *International Journal of Nonlinear Analysis and Applications*, *14*(1), 679-697. https://doi.org/10.22075/ijnaa.2022.28178.3824

Alyuz, N., Gokberk, B., & Akarun, L. (2013). 3-D face recognition under occlusion using masked projection. *IEEE Transactions on Information Forensics and Security*, *8*(5), 789-802. https://doi.org/10.1109/TIFS.2013.2256130

Alzu'bi, A., Albalas, F., AL-Hadhrami, T., Younis, L.B., & Bashayreh, A. (2021). Masked Face Recognition Using Deep Learning: A Review. *Electronics (Switzerland)*. MDPI. https://doi.org/10.3390/electronics10212666

An, X., Zhu, X., Gao, Y., Xiao, Y., Zhao, Y., Feng, Z., ... & Fu, Y. (2021). Partial fc: Training 10 million identities on a single machine. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 1445-1449). https://doi.org/10.1109/ICCVW54120.2021.00166

Bashbaghi, S., Granger, E., Sabourin, R., & Parchami, M. (2019). Deep learning architectures for face recognition in video surveillance. *Deep Learning in Object Detection and Recognition*, 133-154. https://doi.org/10.1007/978-981-10-5152-4_6

Cao, Q., Shen, L., Xie, W., Parkhi, O. M., & Zisserman, A. (2018). Vggface2: A dataset for recognizing faces across pose and age. In *2018 13th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2018)* (pp. 67-74). IEEE. https://doi.org/10.1109/FG.2018.00020

Chen, L., Shi, W., & Deng, D. (2021). Improved YOLOv3 based on attention mechanism for fast and accurate ship detection in optical remote sensing images. *Remote Sensing*, *13*(4), 660. https://doi.org/10.3390/rs13040660

Chen, Q., & Sang, L. (2018). Face-mask recognition for fraud prevention using Gaussian mixture model. *Journal of Visual Communication and Image Representation*, *55*, 795-801. https://doi.org/10.1016/j.jvcir.2018.08.016

Crowley, J. L., & Berard, F. (1997). Multi-modal tracking of faces for video communications. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (pp. 640-645). IEEE. https://doi.org/10.1109/CVPR.1997.609393

Dagar, D., Dagar, Y., & Sharma, R., (2022). Face-Mask Recognition and Detection Using Deep Learning. In *2022 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COM-IT-CON)*, *1*, 692-97. https://doi.org/10.1109/COM-IT-CON54601.2022.9850742

Das, T. K., Banik, A., Chattopadhyay, S., & Das, A. (2019). FFT based classification of solar photo voltaic microgrid system. In *2019 2nd International Conference on Advanced Computational and Communication Paradigms (ICACCP)* (pp. 1-5). IEEE. https://doi.org/10.1109/ICACCP.2019.8882995

Deng, H., Feng, Z., Qian, G., Lv, X., Li, H., & Li, G. (2021a). MFCosface: A masked-face recognition algorithm based on large margin cosine loss. *Applied Sciences*, *11*(16), 7310. https://doi.org/10.3390/app11167310

Deng, J., Guo, J., An, X., Zhu, Z., & Zafeiriou, S. (2021b). Masked face recognition challenge: The insightface track report. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 1437-1444). http://arxiv.org/abs/2108.08191

Din, N. U., Javed, K., Bae, S., & Yi, J. (2020). A novel GAN-based network for unmasking of masked face. *IEEE Access*, *8*, 44276-44287. https://doi.org/10.1109/ACCESS.2020.2977386

Du, H., Shi, H., Liu, Y., Zeng, D., & Mei, T. (2021). Towards NIR-VIS masked face recognition. *IEEE Signal Processing Letters*, *28*, 768-772. https://doi.org/10.1109/LSP.2021.3071663

Ephraim, T., Himmelman, T., & Siddiqi, K. (2009). Real-time viola-jones face detection in a web browser. In *2009 Canadian Conference on Computer and Robot Vision* (pp. 321-328). IEEE. https://doi.org/10.1109/CRV.2009.48

Face Mask Detection Dataset-Kaggle. (2022). https://www.kaggle.com/datasets/andrewmvd/face-mask-detection

Gan, C., Xiao, J., Wang, Z., Zhang, Z., & Zhu, Q. (2022). Facial expression recognition using densely connected convolutional neural network and hierarchical spatial attention. *Image and Vision Computing*, *117*, 104342. https://doi.org/10.1016/j.imavis.2021.104342

Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 580-587). https://doi.org/10.1109/CVPR.2014.81

Goyal, K., Agarwal, K., & Kumar, R. (2017). Face detection and tracking: Using OpenCV. In *2017 International Conference of Electronics, Communication and Aerospace Technology (ICECA)* (Vol. 1, pp. 474-478). IEEE. https://doi.org/10.1109/ICECA.2017.8203730

Guo, Y., Zhang, L., Hu, Y., He, X., & Gao, J. (2016). Ms-celeb-1m: A dataset and benchmark for large-scale face recognition. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part III 14* (pp. 87-102). Springer International Publishing. https://doi.org/10.1007/978-3-319-46487-9_6/FIGURES/5

Hasan, M. K., Ahsan, M. S., Newaz, S. S., & Lee, G. M. (2021). Human face detection techniques: A comprehensive review and future research directions. *Electronics*, *10*(19), 2354. https://doi.org/10.3390/electronics10192354

Hasan, M. M., Yusuf, M. S. U., Rohan, T. I., & Roy, S. (2019). Efficient two stage approach to detect face liveness: Motion based and Deep learning based. In *2019 4th International Conference on Electrical Information and Communication Technology (EICT)* (pp. 1-6). IEEE. https://doi.org/10.1109/EICT48899.2019.9068813

He, Kaiming, Gkioxari, G., Ross Girshick, & P, D., (2017). Mask R-CNN. In *2017 IEEE International Conference on Computer Vision (ICCV)*, 2980-88. https://doi.org/10.1109/ICCV.2017.322

Himeur, Y., Al-Maadeed, S., Varlamis, I., Al-Maadeed, N., Abualsaud, K., & Mohamed, A. (2023). Face mask detection in smart cities using deep and transfer learning: lessons learned from the COVID-19 pandemic. *Systems*, *11*(2), 107. https://doi.org/10.3390/systems11020107

Hong, J. H., Kim, H., Kim, M., Nam, G. P., Cho, J., Ko, H. S., & Kim, I. J. (2021). A 3D model-based approach for fitting masks to faces in the wild. In *2021 IEEE International Conference on Image Processing (ICIP)* (pp. 235-239). IEEE. https://doi.org/10.1109/ICIP42928.2021.9506069

Hou, Q., Zhou, D., & Feng, J. (2021). Coordinate attention for efficient mobile network design. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 13713-13722). https://doi.org/10.1109/CVPR46437.2021.01350

Huang, G. B., Mattar, M., Berg, T., & Learned-Miller, E. (2008). Labeled faces in the wild: A database for studying face recognition in unconstrained environments. In *Workshop on Faces in 'Real-Life 'Images: Detection, Alignment, and Recognition*. http://vis-www.cs.umass.edu/lfw/

Ibrahim, L. M. (2022). Detecting A Medical Mask During The COVID-19 Pandemic Using Machine Learning: A Review Study. *Journal of Education and Science*, *31*(2), 55-68. https://doi.org/10.33899/edusj.2022.133181.1221

Kaur, G., Sinha, R., Tiwari, P. K., Yadav, S. K., Pandey, P., Raj, R., ... & Rakhra, M. (2022). Face mask recognition system using CNN model. *Neuroscience Informatics*, *2*(3), 100035. https://doi.org/10.1016/j.neuri.2021.100035

Lane, L. (2020). Nist finds flaws in facial checks on people with covid masks. *Biometric Technology Today*. https://doi.org/10.1016/S0969-4765(20)30101-6

Lee, W. J., Kim, D. M., Lee, U. Y., Cho, J. H., Kim, M. S., Hong, J. H., & Hwang, Y. I. (2019). A preliminary study of the reliability of anatomical facial landmarks used in facial comparison. *Journal of Forensic Sciences*, *64*(2), 519-527. https://doi.org/10.1111/1556-4029.13873

Li, L., Mu, X., Li, S., & Peng, H. (2020). A review of face recognition technology. *IEEE Access*, *8*, 139110-139120. https://doi.org/10.1109/ACCESS.2020.3011028

Li, S., & Lee, H. J. (2022). Effective Attention-Based Feature Decomposition for Cross-Age Face Recognition. *Applied Sciences*, *12*(10), 4816. https://doi.org/10.3390/app12104816

Libby, C., & Ehrenfeld, J. (2021). Facial recognition technology in 2021: Masks, bias, and the future of healthcare. *Journal of Medical Systems*, *45*(4), 39. https://doi.org/10.1007/S10916-021-01723-W

Liu, Z., Luo, P., Wang, X., & Tang, X. (2015). Deep learning face attributes in the wild. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 3730-3738). https://doi.org/10.1109/ICCV.2015.425

Loey, M., Manogaran, G., Taha, M. H. N., & Khalifa, N. E. M. (2021). A hybrid deep transfer learning model with machine learning methods for face mask detection in the era of the COVID-19 pandemic. *Measurement*, *167*, 108288. https://doi.org/10.1016/j.measurement.2020.108288

Machiraju, S., Urolagin, S., Mishra, R. K., & Sharma, V. (2021). Face mask detection using Keras, OpenCV and TensorFlow by implementing mobilenetv2. In *2021 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N)* (pp. 1485-1489). IEEE. https://doi.org/10.1109/ICAC3N53548.2021.9725546

Marini, M., Ansani, A., Paglieri, F., Caruana, F., & Viola, M. (2021). The impact of facemasks on emotion recognition, trust attribution and re-identification. *Scientific Reports*, *11*(1), 5577. https://doi.org/10.1038/s41598-021-84806-5

Najibi, M., Samangouei, P., Chellappa, R., & Davis, L. S. (2017). Ssh: Single stage headless face detector. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 4875-4884). https://doi.org/10.1109/ICCV.2017.522

Negi, A., Kumar, K., Chauhan, P., & Rajput, R. S. (2021). Deep neural architecture for face mask detection on simulated masked face dataset against covid-19 pandemic. In *2021 International Conference on Computing, Communication and Intelligent Systems (ICCCIS)* (pp. 595-600). IEEE. https://doi.org/10.1109/ICCCIS51004.2021.9397196

Ngan, M. L., Grother, P. J., & Hanaoka, K. K. (2020). Ongoing face recognition vendor test (frvt) part 6b: Face recognition accuracy with face masks using post-covid-19 algorithms. https://doi.org/10.6028/NIST.IR.8331

Ning, X., Xu, S., Nan, F., Zeng, Q., Wang, C., Cai, W., ... & Jiang, Y. (2022). Face editing based on facial recognition features. *IEEE Transactions on Cognitive and Developmental Systems*. https://doi.org/10.1109/TCDS.2022.3182650

Ouyang, Y., Sang, N., & Huang, R. (2015). Accurate and robust facial expressions recognition by fusing multiple sparse representation-based classifiers. *Neurocomputing*, *149*, 71-78. https://doi.org/10.1016/j.neucom.2014.03.073

Pan, Y., Zhang, G., & Zhang, L. (2020). A spatial-channel hierarchical deep learning network for pixel-level automated crack detection. *Automation in Construction*, *119*, 103357. https://doi.org/10.1016/j.autcon.2020.103357

Pann, V., & Lee, H. J. (2022). Effective attention-based mechanism for masked face recognition. *Applied Sciences*, *12*(11), 5590. https://doi.org/10.3390/app12115590

Peng, X., Zhuang, H., Huang, G. B., Li, H., & Lin, Z. (2020). Robust real-time face tracking for people wearing face masks. In *2020 16th International Conference on Control, Automation, Robotics and Vision (ICARCV)* (pp. 779-783). IEEE. https://doi.org/10.1109/ICARCV50220.2020.9305356

Pham, T. N., Nguyen, V. H., & Huh, J. H. (2023). Integration of improved YOLOv5 for face mask detector and auto-labeling to generate dataset for fighting against COVID-19. *The Journal of Supercomputing*, *79*(8), 8966-8992. https://doi.org/10.1007/s11227-022-04979-2

Prasad, S., Li, Y., Lin, D., & Sheng, D. (2021). maskedfacenet: A progressive semi-supervised masked face detector. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision* (pp. 3389-3398). https://doi.org/10.1109/WACV48630.2021.00343

Qian, J., Yang, J., Zhang, F., & Lin, Z. (2014). Robust low-rank regularized regression for face recognition with occlusion. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (pp. 21-26). https://doi.org/10.1109/CVPRW.2014.9

Qiu, H., Gong, D., Li, Z., Liu, W., & Tao, D. (2021). End2end occluded face recognition by masking corrupted features. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *44*(10), 6939-6952. https://doi.org/10.1109/TPAMI.2021.3098962

Razzaq, A. N., Ghazali, R., El Abbadi, N. K., & Dosh, M. (2022). A comprehensive survey on face detection techniques. *Webology*, *19*(1), 613-628. https://doi.org/10.14704/web/v19i1/web19044

Safaa El-Din, Y., Moustafa, M. N., & Mahdi, H. (2020). Deep convolutional neural networks for face and iris presentation attack detection: Survey and case study. *IET Biometrics*, *9*(5), 179-193. https://doi.org/10.1049/iet-bmt.2020.0004

Schroff, F., Kalenichenko, D., & Philbin, J. (2015). Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 815-823). https://doi.org/10.1109/CVPR.2015.7298682

Sethi, S., Kathuria, M., & Kaushik, T. (2021). Face mask detection using deep learning: An approach to reduce risk of Coronavirus spread. *Journal of Biomedical Informatics*, *120*, 103848. https://doi.org/10.1016/j.jbi.2021.103848

Shahar, M. S. M., & Mazalan, L. (2021, April). Face identity for face mask recognition system. In *2021 IEEE 11th IEEE Symposium on Computer Applications and Industrial Electronics (ISCAIE)* (pp. 42-47). IEEE. https://doi.org/10.1109/ISCAIE51753.2021.9431791

Sharma, A., Gautam, R., & Singh, J. (2023). Deep learning for face mask detection: A survey. *Multimedia Tools and Applications*, 1-41. https://doi.org/10.1007/s11042-023-14686-6

Sheikh, B. U. H., & Zafar, A. (2023). RRFMDS: Rapid real-time face mask detection system for effective COVID-19 monitoring. *SN Computer Science*, *4*(3), 288. https://doi.org/10.1007/s42979-023-01738-9

Siddiqui, M. F., Siddique, W. A., Ahmedh, M., & Jumani, A. K. (2020). Face detection and recognition system for enhancing security measures using artificial intelligence system. *Indian Journal of Science and Technology*, *13*(09), 1057-1064. https://doi.org/10.17485/ijst/2020/v13i09/149298

Singh, S., Ahuja, U., Kumar, M., Kumar, K., & Sachdeva, M. (2021). Face mask detection using YOLOv3 and faster R-CNN models: COVID-19 environment. *Multimedia Tools and Applications*, *80*, 19753-19768. https://doi.org/10.1007/s11042-021-10711-8

Song, L., Gong, D., Li, Z., Liu, C., & Liu, W. (2019). Occlusion robust face recognition based on mask learning with pairwise differential Siamese network. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 773-782). https://doi.org/10.1109/ICCV.2019.00086

Song, Z., Nguyen, K., Nguyen, T., Cho, C., & Gao, J. (2022). Spartan face mask detection and facial recognition system. In *Healthcare* (Vol. *10*, No. 1, p. 87). MDPI. https://doi.org/10.3390/healthcare10010087

Subramanian, N. (2011). Biometric Authentication. In: Van Tilborg, H.C.A., Jajodia, S. (eds) Encyclopedia of Cryptography and Security. Springer, Boston, MA. https://doi.org/10.1007/978-1-4419-5906-5_775

Tathe, S., Abhilasha,V., Narote, S., & Narote, S., P. (2017). Face Recognition and Tracking in Videos. *Advances in Science, Technology and Engineering Systems*, *2* (3), 1238-44. https://doi.org/10.25046/aj0203156

Teoh, K. H., Ismail, R. C., Naziri, S. Z. M., Hussin, R., Isa, M. N. M., & Basir, M. S. S. M. (2021). Face recognition and identification using deep learning approach. In *Journal of Physics: Conference Series* (Vol. 1755, No. 1, p. 012006). IOP Publishing. https://doi.org/10.1088/1742-6596/1755/1/012006

Thein, T., & San, K. M. (2018). Lip movements recognition towards an automatic lip-reading system for Myanmar consonants. In *2018 12th International Conference on Research Challenges in Information Science (RCIS)* (pp. 1-6). IEEE. https://doi.org/10.1109/RCIS.2018.8406660

Wang, C., Zhang, B., Cao, Y., Sun, M., He, K., Cao, Z., & Wang, M. (2023a). Mask Detection Method Based on YOLO-GBC Network. *Electronics*, *12*(2), 408. https://doi.org/10.3390/ELECTRONICS12020408

Wang, Z., Huang, B., Wang, G., Yi, P., & Jiang, K. (2023b). Masked face recognition dataset and application. *IEEE Transactions on Biometrics, Behavior and Identity Science*. https://doi.org/10.1109/TBIOM.2023.3242085

Wang, *et al*. (2022). RMFD (Real-World Masked Face Dataset) Papers with Code.

Wong, Y., Chen, S., Mau, S., Sanderson, C., & Lovell, B. C. (2011). Patch-based probabilistic image quality assessment for face selection and improved video-based face recognition. In *CVPR 2011 WORKSHOPS* (pp. 74-81). IEEE. https://doi.org/10.1109/CVPRW.2011.5981881

Woo, S., Park, J., Lee, J. Y., & Kweon, I. S. (2018). Cbam: Convolutional block attention module. In *Proceedings of the European Conference on Computer Vision (ECCV)* (pp. 3-19). https://doi.org/10.1007/978-3-030-01234-2_1

Wright, J., Yang, A. Y., Ganesh, A., Sastry, S. S., & Ma, Y. (2008). Robust face recognition via sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *31*(2), 210-227. https://doi.org/10.1109/TPAMI.2008.79

Xu, S., Guo, Z., Liu, Y., Fan, J., & Liu, X. (2022). An improved lightweight yolov5 model based on attention mechanism for face mask detection. In *International Conference on Artificial Neural Networks* (pp. 531-543). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-15934-3_44

Yan, C., Meng, L., Li, L., Zhang, J., Wang, Z., Yin, J., ... & Zheng, B. (2022). Age-invariant face recognition by multi-feature fusion and decomposition with self-attention. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, *18*(1s), 1-18. https://doi.org/10.1145/3472810

Yang, M. H., Kriegman, D. J., & Ahuja, N. (2002). Detecting faces in images: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *24*(1), 34-58. https://doi.org/10.1109/34.982883

Yang, M., Zhang, L., Yang, J., & Zhang, D. (2011). Robust sparse coding for face recognition. In *CVPR 2011* (pp. 625-632). IEEE. https://doi.org/10.1109/CVPR.2011.5995393

Yang, S., Luo, P., Loy, C. C., & Tang, X. (2017). Faceness-net: Face detection through deep facial part responses. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *40*(8), 1845-1859. https://doi.org/10.1109/TPAMI.2017.2738644

Yang, S., Wen, Y., He, L., Zhou, M., & Abusorrah, A. (2021). Sparse individual low-rank component representation for face recognition in the IoT-based system. *IEEE Internet of Things Journal*, *8*(24), 17320-17332. https://doi.org/10.1109/JIOT.2021.3080084

Yi, *et al*. (2022). CASIA-WebFace Dataset Papers with Code. https://paperswithcode.com/dataset/casia-webface

Yu, J., Hu, C. H., Jing, X. Y., & Feng, Y. J. (2020). Deep metric learning with dynamic margin hard sampling loss for face verification. *Signal, Image and Video Processing*, *14*, 791-798. https://doi.org/10.1007/s11760-019-01612-3

Zhang, Y., Yang, C., & Zhao, Q. (2021). Face mask recognition based on object detection. In *International Conference on Signal Image Processing and Communication (ICSIPC 2021)* (Vol. 11848, pp. 276-279). SPIE. https://doi.org/10.1117/12.2600460

Zhang, Z., Yan, J., Liu, S., Lei, Z., Yi, D., & Li, S. Z. (2012). A face antispoofing database with diverse attacks. In *2012 5th IAPR International Conference on Biometrics (ICB)* (pp. 26-31). IEEE. https://doi.org/10.1109/ICB.2012.6199754

Zhao, G., Zou, S., & Wu, H. (2023). Improved Algorithm for Face Mask Detection Based on YOLO-v4. *International Journal of Computational Intelligence Systems*, *16*(1), 104. https://doi.org/10.1007/s44196-023-00286-7